Multi-species space use by marine predators: defining biodiversity 'hotspots' in the marine realm

Rebecca Phillips

Submitted to Swansea University in fulfilment of the requirements for the Degree of Doctor of Philosophy (PhD)

Swansea University

2024

Abstract

Anthropogenic change of ecosystems has led to wide-scale changes in biodiversity globally, with declines across terrestrial, freshwater and marine realms. Global concern about the increasing anthropogenic impact on biodiversity has driven intense research into the drivers and consequences of change, alongside rapidly moving international policy and conservation development. Quantifying critical marine habitats is vital for protecting remaining biodiversity, and understanding areas of increased coexistence or biodiversity hotspots is particularly important if management and conservation methods continue to focus on spatial approaches (such as marine protected areas). A key challenge in biogeography is to understand and predict the potential impacts of climate change on the distribution of biodiversity, therefore identifying the environmental drivers that may impact richness may provide insight into future species richness patterns. Studies have highlighted the difficulty in mapping biodiversity at a large scale due to patchy data coverage, and this is further intensified using a particularly cryptic set of species that inhabit and spend much of their life beneath the surface. Top marine predators have essential ecological roles as ecosystem engineers and amplify trophic information across multiple spatiotemporal scales and have been identified as sentinel species which can exhibit clear responses to environmental variability and ecosystem health. However quantitative maps of marine predator coexistence are lacking, with the collection of new standardised data expensive, time-consuming and often focused on small-scale local studies.

This thesis utilises large available historical datasets to address this critical research gap in mapping the biodiversity of top marine predators across multiple taxa, around the UK. Chapter 2 summarises available data around the UK, which can be used to ascertain the status of the information available. This is useful for a wide range of stakeholders who are often under time pressure, under-resourced and trying to be proactive in a fast-paced legislative environment. It demonstrates available data exists to maximise temporal and spatial coverage of such a large-scale area, for large-scale research questions.

However, it is well-known that databases have inherent biases due to heterogeneous data sources and lack of standardisation. Chapter 3 adapts a well-recognised risk-

assessment matrix approach to quantify biases within four example datasets. The study highlighted the level of risk in using heterogeneous datasets is lower for assessing patterns of association in marine predators, rather than counts or abundances. Therefore, research questions should be carefully considered when utilising datasets not designed for a specific research hypothesis. The matrix method presented has two important roles to advance this field of research: the first is to aid people to analyse existing datasets and provide a standardised approach to enumerating bias quantitatively as opposed to just describing bias assumptions. The second is to guide research to design better surveys by understanding which risk factors are most influential on their study.

Critically, Chapter 4 provides maps of marine predator biodiversity hotspots around the UK, identifying areas of shared space use across taxa groups. It is the largest cross-taxa study of marine predator biodiversity around the UK to date. Sighting datasets were collated and species richness was determined across different spatial scales. Whilst species richness has been a keystone measure of biodiversity, it can be skewed by survey effort and therefore a new measure of 'species richness per unit effort' (SRPUE) was derived to identify relative areas of high and low richness. Patterns in species richness over time using seasonal-trend decomposition analysis revealed that the concept of carrying capacity becomes evident when survey effort is accounted for. This chapter demonstrates how the analysis of existing data can facilitate the mapping of the biodiversity of marine predators and allow areas of high shared space use to be prioritised in conservation and management.

Simply quantifying biodiversity hotspots is not enough, with the potential drivers of species richness patterns identified as important in the literature, particularly with anthropogenic climate change causing shifts in species coexistence. Chapter 5 uses three modelling approaches to look at associations of high species richness with a suite of environmental variables. Macro-ecological models (MEMs) using generalised additive models (GAMs) and stacked species distribution models (SSDMs) were utilised to give maps of species richness. SRPUE is a useful method to factor in effort in richness mapping, but user interpretation is not as intuitive as the well-known species richness integer scale, and therefore an alternative approach was derived using effort as a predictor. All variables were significant predictors for biodiversity. The GAM using raw species richness and SSDM model demonstrated sea surface

temperature has the highest relative contribution to richness hotspots and therefore is a concern in climate change with warming oceans.

Overall, this thesis demonstrates how analysis of existing and diverse data can be utilised cost-effectively to map biodiversity. While some caution is needed when using historical datasets, these distribution maps are the first available output at this spatial scale and taxonomic coverage and have widespread and immediate applications in identifying important areas of protection and providing a focus for marine management strategies.

Declarations and Statements

Declarations

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed......[redacted]..

Date.......18th February 2024...

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed......[redacted]..

Date.......18th February 2024...

I hereby give consent for my thesis, if accepted, to be available for electronic sharing.

Signed.......[redacted]..

Date.......18th February 2024...

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed..........[redacted]...

Date.......18th February 2024...

Collaborator Contributions

Chapter 1 Literature review and writing of manuscript by Rebecca Phillips (RP). Supervision and feedback on draft thesis were provided by Dr. Jim Bull (JB).

Chapter 2 Design, methodology conceived and written by RP. Supervision and feedback on the thesis were provided by JB. Datasets obtained from public data repositories and were freely available. The most recent European Seabird at Sea (ESAS) dataset kindly provided by Mark Lewis (ML) from JNCC on 22nd January 2019. Full ORCA dataset kindly provided by James Robbins (JR) on 29th August 2019. Feedback on draft chapter were provided by JB. Feedback on matrix method given by Chiara Bertelli (Scientific Officer Swansea University), ML (JNCC) and Rebecca Stone (Sustainability and Ecology Policy Officer at Carmarthen County Council) and presented in appendices.

Chapter 3 Conceived, designed, modelled and written by RP. Analysis by RP, with guidance from JB. Supervision and feedback on the thesis were provided by JB. Feedback on draft chapter were provided by JB.

Chapter 4 Conceived, designed, modelled and written by RP. Data analysis was conducted by RP with input and guidance from JB. Feedback on draft chapter provided by JB.

Chapter 5 Conceived, designed, modelled and written by RP. Data analysis was conducted by RP with input and guidance from JB. Feedback on draft chapter provided by JB.

Chapter 6 Written by RP. Feedback on draft chapter were provided by JB.

Contents page

Acknowledgements

People say the two most stressful things in life are buying a house and planning a wedding. I'm not sure they factored in doing a PhD during a global pandemic in that ranking, let alone doing all three, and I would not have survived without the support of some truly incredible people. There are so many people I would love to give a mention one by one, but that would be a thesis in itself – but know I am grateful to be surrounded by wonderful people who have kept me sane with their support, kindness and lately, lots of energy drinks.

Firstly, to Jim, thank you for giving me the freedom to explore ideas and run with them in many directions, for not putting pressure on me during some of the most difficult times, and for the constant support, encouragement and guidance. You have inspired me with your inquisitiveness for patterns and your statistics wizardry, and I feel very lucky to have had you as my supervisor.

To my parents, I feel so incredibly lucky to have travelled far and wide the way we did as a kid, and you are the reason I fell in love with the natural world and all its wacky creatures. Dad, you have always inspired me to follow in your footsteps, so thank you for being a constant fountain of knowledge, wisdom and calm. Mum: you have always been my top cheerleader (and provider of the best care packages), thank you for believing in me always, even when I have not. I hope I've made you both proud.

"Though she be but little she is fierce" My sister Ruth, who is the most inspiring person I know and constantly blows me away with her bravery and hard work. You have been a rock during these years and thank you for the many (many) pep talks. You astound me with being able to balance all you do, and still have time to listen to my ramblings about whales.

Pedro, I could not have had a better lab mate to go on this rather mad journey with. For providing humour amongst the stress, calm advice amongst the panic and pure camaraderie from start to finish, I can't thank you enough. As you always say, it's definitely now "time for a gin". To the other wonderful members of SPACEPOP, it was a pleasure to be part of such an enthusiastic talented bunch and what an achievement we kept it going throughout lockdown!

And finally, to Jake, you have been there through every high and low. When we first met 6 years ago, I was applying for this PhD. Six months later we were away on our first big trip, and I dialled in at 6am Pacific time for the interview, sat out on the balcony in Santa Barbara hoping the hotel WIFI would hold up. A good luck charm, you are here at the end. You have unequivocally relished my love for marine biology, even proposing on a boat in Tromsø after seeing orca. As Millie has always liked to say, “Becky sure loves the ocean”. I would not have survived the two years of Covid lockdown without you, or the kindness of your family taking me in when I was so far away from mine. I wish more than anything your dad was here to celebrate this achievement. Thank you for being there for me, even during the hardest year of your life, I will never take for granted that generosity even in grief. Thank you for your patience, always, but particularly the last few months during crunch time. A full-time job and a PhD write up, amongst all the other hard things we’ve faced, has been a test but I am grateful you have been there unwaveringly, always. Love you, husband.

Tables and Figures

Figures

Tables

List of key acronyms

Acronym	Description
AESs	Areas Of Ecological Significance
AOI	Area Of Interest
ATSR	Along Track Scanning Radiometer
AVHRR	Advanced-Very-High-Resolution-Radiometer
CBD	Convention on Biological Diversity
Chl-*a*	Chlorophyll-*a*
COFNOD	North Wales Environmental Information Service
COP	Conference Of the Parties of the UNFCCC
CPF	Central Place Forager
CRS	Coordinate Reference System
DEFRA	Department For Environment, Food & Rural Affairs
EBV	Essential Biodiversity Variable
ENSO	El Niño–Southern Oscillation
ESA	European Space Agency
ESAS	European Seabirds At Sea
ESDM	Ensemble Species Distribution Model
GAMs	Generalised Additive Models
GBIF	Global Biodiversity Information Facility
ICES	International Council for the Exploration of the Sea
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
JNCC	Joint Nature Conservation Committee
LPI	Living Planet Index
MEDIN	Marine Environmental Data and Information Network
MEMs	Macro-Ecological Models
MPA	Marine Protected Areas
NAO	North Atlantic Oscillation
NBN	National Biodiversity Network Atlas

NEODASS	NERC Earth Observation Data Acquisition and Analysis Service
NERC	Natural Environment Research Council
NOC	National Oceanography Centre
OBIS	Ocean Biogeographic Information System
OBIS-SEAMAP	Ocean Biodiversity Information System Spatial Ecological Analysis of Megavertebrate Populations
ORCA	Organisation Cetacea
PDO	Pacific Decadal Oscillation
SAC	Special Areas of Conservation
SDG	Sustainable Development Goals
SMRU	Sea Mammal Research Unit
SPA	Special Protection Areas
SRPUE	Species Richness Per Unit Effort
SSDMs	Stacked Species Distribution Models
SST	Sea Surface Temperature
STL	Seasonal-Decomposition of Time Series
UNFCCC	United Nations Framework Convention on Climate Change
WDC	Whale And Dolphin Conservation
WWF	Worldwide Fund for Nature

Chapter 1 General Introduction

1.1 The importance of biodiversity

Anthropogenic change of ecosystems has led to wide-scale changes in biodiversity globally, with declines across both terrestrial, freshwater and marine realms (Tittensor *et al.*, 2014, Hillebrand *et al.*, 2018, Butchart *et al.*, 2019, Diaz *et al.*, 2019) (Figure 1.1). According to the Convention on Biological Diversity (CBD), “biodiversity” or "Biological diversity" means the “variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems” (European Commission, 2016). Therefore, biodiversity is a multifaceted and complex concept.

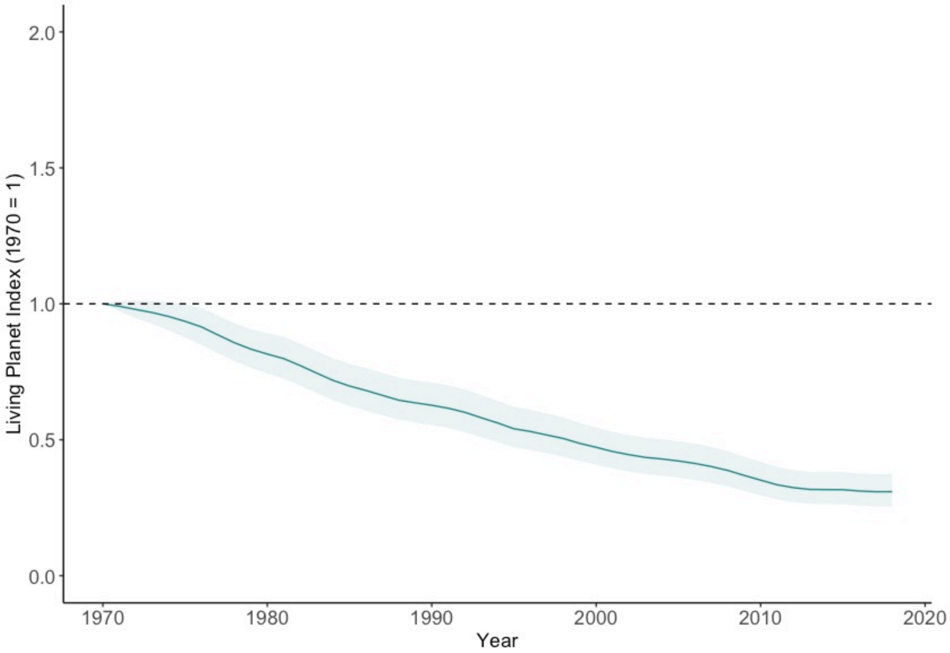


Figure 1.1 Global decline of biodiversity demonstrated using the Living Planet Index (LPI) from 1970. LPI uses wildlife population data to calculate average rates of change in 5,230 species, for terrestrial, freshwater and marine systems.

Further shifts in biodiversity are expected with such rapid environmental change (Sala *et al.*, 2000) driven by anthropogenic impacts such as land use-change, pollution, species over-exploitation (Sanderson *et al.*, 2002, Halpern *et al.*, 2008). The most recent Living Planet Report (2022) found that across Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) regions, habitat degradation and loss was the most commonly recorded threat to biodiversity, followed by species overexploitation (Figure 1.2) (threat types based upon classification IPBES

(Brondizio *et al.*, 2019), based upon (Salafsky *et al.*, 2008) also followed by IUCN Red list). These changes can lead to rising ocean temperatures (Hansen *et al.*, 2006), declining oxygen and ocean acidification (Breitburg *et al.*, 2018, Hill and Hoogenboom, 2022, Nagelkerken and Connell, 2022, Scherer *et al.*, 2022), thermal stratification (De Dominicis *et al.*, 2018), melting sheet ice (Stokes *et al.*, 2022), changes in salinity (Du *et al.*, 2019), and influences on global current systems such as El Niño–Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), and Pacific Decadal Oscillation (PDO) (Overland *et al.*, 2010). These changes have effects on ecosystem function and can trigger biological responses such as changes in phenology, productivity, population connectivity, local adaptation and speciation threatening biodiversity, as reviewed in Hoegh-Guldberg and Bruno (2010).

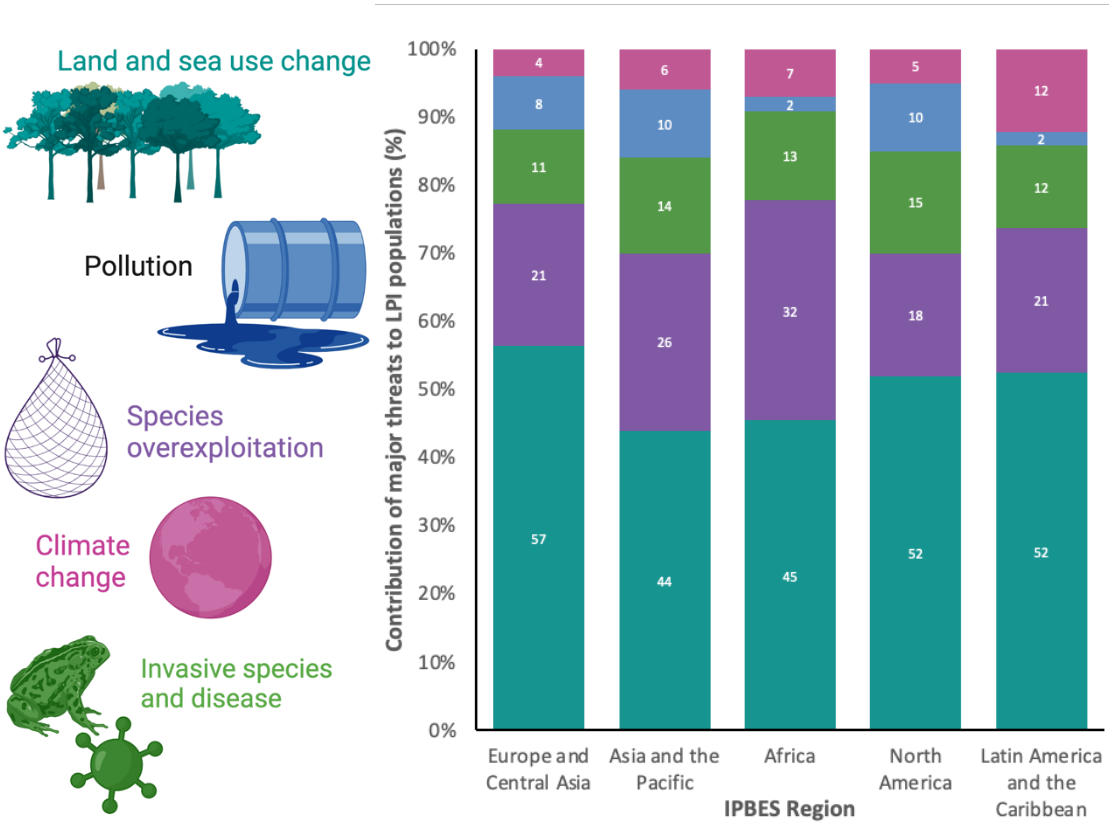


Figure 1.2 Contribution of the five main threats to biodiversity per IPBES regions reported in the Living Planet Report (2022). Data from WWF (2022).

Global concern about the increasing anthropogenic impact on biodiversity has driven both intense research into the drivers and consequences of change and movements in international policy and conservation (Hillebrand *et al.*, 2018). Documented changes in biodiversity, such as declining species abundance (Lotze *et al.*, 2006) or increased number of endangered species (Butchart *et al.*, 2010), have led to a "biodiversity crisis" with warnings that the rate of species extinction is exceptionally high (Mace and Baillie, 2007, Pimm *et al.*, 2014) and threatens a global mass extinction (Barnosky *et al.*, 2011, Barnosky *et al.*, 2012). A review of biodiversity indicators revealed most had negative trends (8 out of 10 indicators) with declining population trends in vertebrates, habitat specialist birds and shorebirds worldwide, alongside declining extent of forest, mangroves; seagrass beds, and the condition of coral reefs (Butchart *et al.*, 2010). In the latest Living Planet Index (LPI) (Westveer *et al.*, 2022) (which uses wildlife population data to calculate average rates of change in (at present) 5,230 species, for terrestrial, freshwater and marine systems) all five IPBES regions have shown biodiversity declines since 1970 (Figure 1.3). The Latin America and the Caribbean region show the most dramatic decline in biodiversity (94%). The importance of biodiversity for human well-being has been well documented (Geijzendorffer *et al.*, 2016, IPBES, 2019, Dasgupta, 2021) and is becoming much more apparent as the economic effects are revealed (Franceschi and Kahn, 2003, Gilbert, 2014) with the World Economic Forum rated biodiversity loss as one of the top five risks to the global economy (World Economic Forum, 2021).

It is important to understand the extent of the biodiversity crisis to be able to develop strategies to mitigate or slow the impact (Magurran, 2021). Both science and policy have responded to the need to measure and assess biodiversity change, with targets a key tool at both global and national level (Bane *et al.*, 2022). Within policy, global efforts have included the Aichi biodiversity targets (under the umbrella of the CBD) which aimed to halt further biodiversity decline by 2020 (Tittensor *et al.*, 2014). In an assessment of the Aichi Targets, Tittensor *et al.* (2014) suggested to be effective, actions towards the Aichi targets will have to be supported by updated information on regional and global patterns of biodiversity change, on drivers of biodiversity change, and on the effectiveness of conservation policies (Pereira and Cooper, 2006, Tittensor *et al.*, 2014).

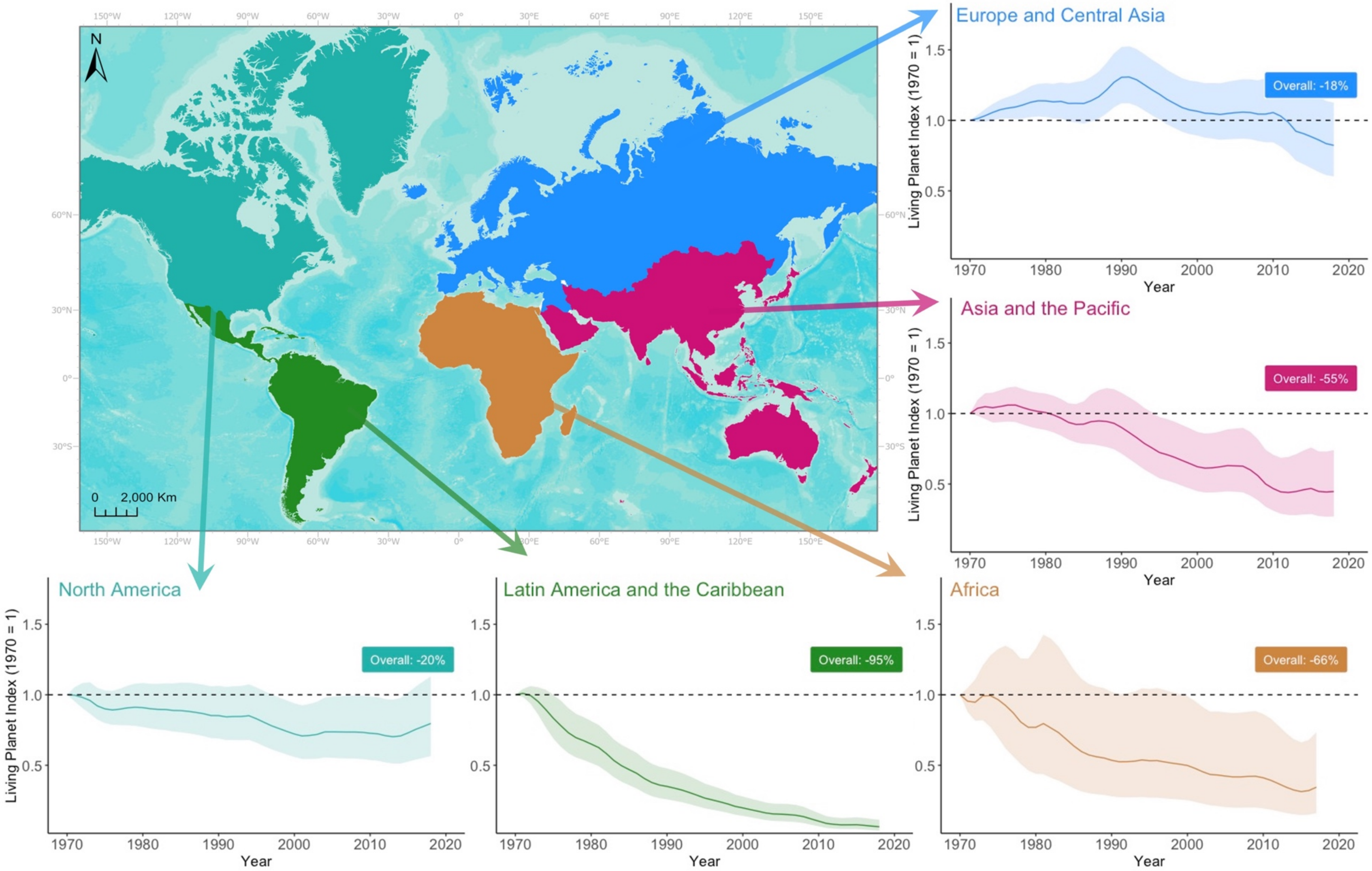


Figure 1.3 Trends in Living Planet Index (LPI) from 1970 to 2020 per Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) region. Data available from WWF and ZSL (2022).

The Aichi Targets are echoed in the United Nations' Sustainable Development Goals (SDGs), with goal 14 concerning the conservation of seas. The Post-2020 Global Biodiversity Framework (Convention on Biological Diversity, 2022) agreed new targets, intending to slow down the rate of biodiversity loss and include conserving 30% of land and sea areas by 2030. The most recent COP27 in 2022 (United Nations, 2022) highlighted the urgent need to address the global crisis of biodiversity loss with a dedicated Biodiversity Day at the conference. Legislative international agreements on monitoring and reporting on ecosystem biodiversity (Pereira and Cooper, 2006) have been instated. The most prominent examples in the EU are the Water Framework Directive and the Marine Strategy Framework Directive (Borja *et al.*, 2010, Hering, 2010).

Studies have highlighted that biodiversity includes aspects of identity, dominance and rarity and therefore biodiversity comprises more than a reduction in the number of species, it includes functional, phylogenetic and genetic diversity also (Pereira *et al.*, 2013). However, these aspects are rarely assessed despite increasing literature on their importance (Diaz *et al.*, 2013, Mace *et al.*, 2014, Steffen *et al.*, 2015) and indicators utilised depended on several other factors (such as example stakeholder interests, data availability and practical ad hoc solutions for immediate needs) rather than simply meeting scientific criteria (Noss, 1990, Feest, 2013). However, Pereira *et al.* (2013) developed a framework based on ecological principles for a set of essential biodiversity variables (EBVs), which has gained support particularly for facilitating collaboration. Species richness remains the most used dominant metric to detect biodiversity change (Hillebrand *et al.*, 2018). It is comparatively simple to observe and record, therefore affordable and applicable across the marine realm, and its ease of use means it remains an EBV. Pragmatic methods that rapidly capture biodiversity are important for spatial planning and management in this era of increased global policy effort. Species richness is defined as the number of species per unit area (Brown *et al.*, 2007). The marine realm is extremely diverse, with only 11% of species thought to be described (Jungblut *et al.*, 2018). However marine biodiversity is under serious threat, for example, marine fish abundance has declined by 38% compared to 1970 (Hutchings *et al.*, 2010) and areas of marine coastal habitats (e.g. seagrass beds and mangroves) have been depleted by over two-thirds (Lotze *et al.*, 2006). Loss of ocean biodiversity means loss of ecosystem services (the benefits humans derive from ecosystems (MEA, 2005, Cardinale *et al.*, 2012)) such as food resources, recreation and shoreline protection (MEA, 2005, Worm *et al.*, 2006, Palumbi *et al.*, 2009) as well as loss of ecosystem stability (the variability in ecosystem property through time (Schindler *et al.*, 2015)). Mechanisms to explain the link between biodiversity and ecosystem functioning and stability have included niche partitioning and "complementarity effects" where diverse community assemblages contain well-adapted species or species which disproportionately affect ecosystem function (Palumbi *et al.*, 2009).

Top marine predators include cetaceans, seabirds, sharks, pinnipeds and turtles (Hobday *et al.*, 2015), and have essential ecological roles as ecosystem engineers, with top-down regulation on trophic levels and food webs (Hunt and McKinnell, 2006, Estes *et al.*, 2016) and nutrient cycling (Doughty *et al.*, 2016). Top predators move across ocean basins and amplify trophic information across multiple spatiotemporal scales (Hazen *et al.*, 2019). They are however susceptible to challenges from environmental change (Hoegh-Guldberg and Bruno, 2010) anthropogenic activities (Halpern *et al.*, 2015) such as bycatch, habitat degradation and loss, noise disturbance, prey reductions, vessel traffic and collisions, pollution and energy extraction (Croxall *et al.*, 2012, Avila *et al.*, 2018), which is exacerbated by their life traits (e.g. smaller number of offspring, slow growth, and later maturity ages (McClellan *et al.*, 2014)). They are often identified as sentinel species which can exhibit clear responses to environmental variability (Sydeman *et al.*, 2015, Fleming *et al.*, 2016) and reflect ocean health (Evans *et al.*, 2005, Tittensor *et al.*, 2010, Sydeman *et al.*, 2012, Hazen *et al.*, 2019, Sydeman *et al.*, 2021). Hazen *et al.* (2019) defines 'sentinels' as a "species that responds to ecosystem variability and/or change in a timely, measurable, and interpretable way, and can indicate an otherwise unobserved change in ecosystem structure or function". Marine predators are often of high socio-economic and cultural importance given their charismatic status (Higham and Lück, 2007), and therefore a focus of public interest and long term-research programs (Chambers *et al.*, 2014).

Marine predators in particular also often fall under strict legislation in the UK and EU, with many species protected under Annex II of the Habitats Regulations (where core areas of their habitat are designated and included in the Natura 2000 framework), are European Protected Species under Annex IV of the European Commission habitats directive, listed under OSPAR Convention annex V and UK Biodiversity Action Plan

(BAP) priority species that continues to be regarded as a conservation priority in the subsequent UK Post-2010 Biodiversity Framework. Though there are thousands of species of conservation concern, marine management and conservation strategies often focus on particular species for conservation (Caro, 2010), and marine predators cover the terms ‘indicator’ species (those which respond quickly to changes in environment or biodiversity loss), ‘keystone’ species (important ecological roles in ecosystem structure and functioning), ‘umbrella’ species (protect biodiversity, due to wide home ranges) and ‘flagship species’ (species that people feel concerned about) (Albert *et al.*, 2018). Therefore, protecting marine predator biodiversity is beneficial for healthy ecosystems and ecosystems functioning overall.

Assessing changes in marine predator biodiversity requires mapping baseline data and modelling hotspots (Magurran *et al.*, 2010), and has become an increasingly growing field of study (Reese and Brodeur, 2006, Magurran *et al.*, 2010, Bouchet *et al.*, 2015). It is important to identify areas of high biodiversity where multiple species co-exist so that conservation and management, which are often limited in resources, can be focused appropriately (Evans and Hammond, 2004). Hays *et al.* (2016) also highlighted that the charismatic nature of marine predators and the dramatic voyages they undertake means studying the spatial ecology is a powerful way of gaining public support for conservation and winning decision-makers interest.

1.2 The “big data” push

There is a lack of quantitative data on biodiversity change, and this prevents effective monitoring of biodiversity (Bubb, 2013, Geijzendorffer *et al.*, 2016). Observations are inevitably biased; generally toward recent decades, large-bodied and charismatic species (Hudson *et al.*, 2014), often in terrestrial, temperate, economically developed, and easily-accessible environments (Boakes *et al.*, 2010, Pimm *et al.*, 2014, Meyer *et al.*, 2015, Meyer, 2015, Geijzendorffer *et al.*, 2016). Whilst the focus on charismatic species may be beneficial for marine predators, the challenge of monitoring such cryptic animals in the ocean environment means data collection is difficult and patchy and often spatially biased and many marine species and habitats are known to be poorly sampled (Costello *et al.*, 2010, Webb *et al.*, 2010, Mora *et al.*, 2011). With their wide-ranging nature, low detectability and relatively inaccessible habitat due to their marine lifestyle, it requires a large amount of offshore survey effort to have sufficient data to assess spatial usage (Kaschner *et al.*, 2012) and presents a considerable financial challenge for conservation (Betty *et al.*, 2019).

Further limited availability of robust data may be due to practical factors such as usage restrictions, limited accessibility, data confidentiality, data integration and quality issues (e.g. sampling bias, taxonomic inconsistencies, human error) (Henry *et al.*, 2008). These information gaps could be bridged through the mobilisation of existing historical data in combination with new data (Kot *et al.*, 2010). The need for data utilisation is internationally recognised, and increasing in focus and has led to several initiatives to increase sharing of biodiversity data such as the Global Biodiversity Information Facility (GBIF) (GBIF, 2019), TRY (Kattge *et al.*, 2011), GenBank (Benson *et al.*, 2005) and Biodiversity information system for Europe (BISE) (European Commission and European Environment Agency, 2023), facilitated by the advancement of information technology to handle, store and model big data (Saran *et al.*, 2022). In the last few decades, large volumes of data on the spatial ecology of marine megafauna have been collected (Grémillet *et al.*, 2022), allowing exploration and analysis for use in marine spatial planning, policy, research and legislation amongst others (Figure 1.4).

With the pressure to evaluate changes in biodiversity now, combining output available from ongoing monitoring initiatives is a priority and welcome option (Leonard *et al.*, 2006, Henry *et al.*, 2008), though careful consideration of the uncertainty of data collection and impacts of using heterogenous data sources (Boakes *et al.*, 2010). Biodiversity databases are patchy in coverage and inherently biased due to data collection limitation, lack of standardisation and heterogenous data quality (Jetz and Rahbek, 2001, Soberón *et al.*, 2007, Lobo, 2008, Boakes *et al.*, 2010, Rocchini *et al.*, 2011, Ladle and Hortal, 2013, Meyer *et al.*, 2015, Lobo *et al.*, 2018). Therefore, factoring in a measure of these biases is critical when interpreting patterns of biodiversity.

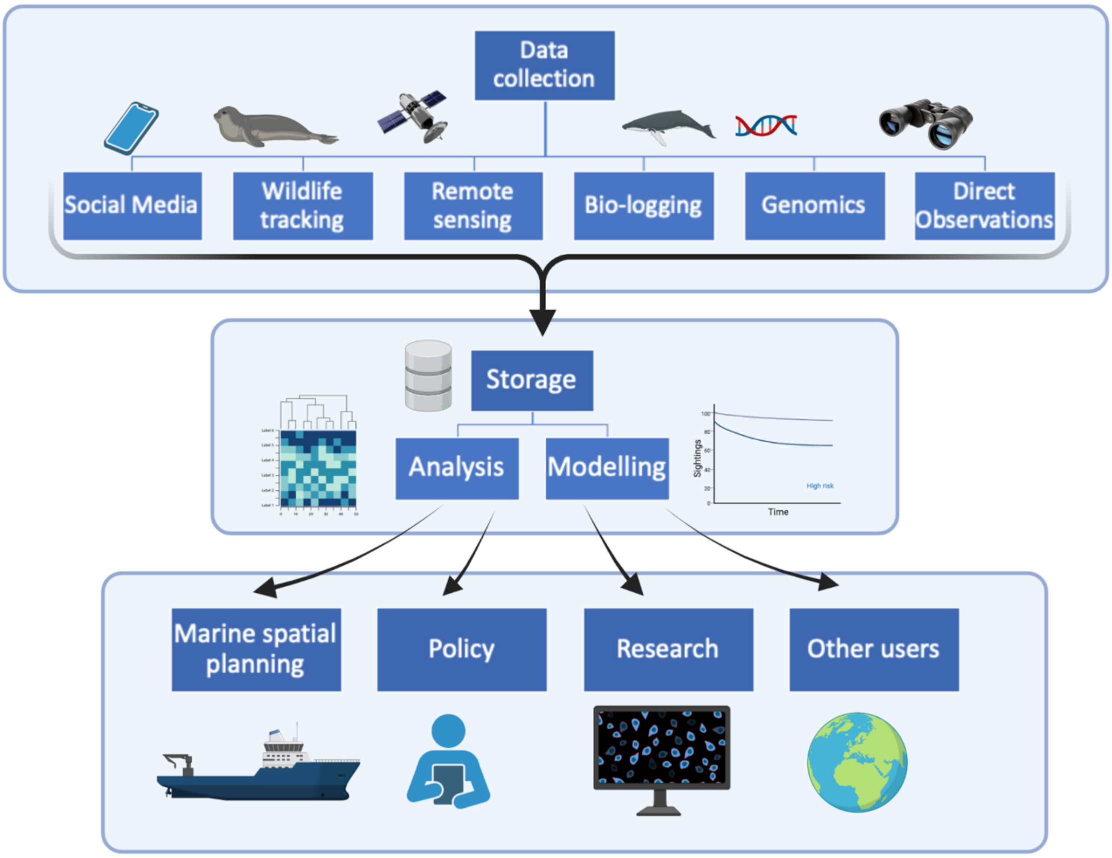


Figure 1.4 Diagram of how big data accumulated through a range of collection methods can then be stored, analysed and modelled to be used for a wide range of outputs and uses. Created using Biorender (2023).

Rapidly increasing data for marine predators means we are entering a realm of "big data science" (Grémillet *et al.*, 2022). Challenges of working with big datasets include having suitable data management and substantial-enough computing power for analyses and modelling with other sizeable information sources, such as considerable oceanographic data via satellite remote sensing (McMahon *et al.*, 2021). Marine predator data is often combined with abiotic or biotic environments to look for drivers of patterns (Grémillet *et al.*, 2022), such as biodiversity, and are often linked with remote-sensed variables such as sea surface temperature (SST) (Tittensor *et al.*, 2010, Kaschner *et al.*, 2011, Astudillo-Scalia *et al.*, 2020), chlorophyll *a* (Louzao *et al.*, 2006, O'Hara *et al.*, 2006, Grémillet *et al.*, 2008, Boertmann, 2011, Bennison and Jessopp, 2015, Grecian *et al.*, 2016), bathymetry (Hyrenbach *et al.*, 2007, Bailey and Thompson, 2009, Robinson *et al.*, 2009, Amelineau *et al.*, 2016, Retana and Lewis, 2017, Thorne *et al.*, 2017, Urmy and Warren, 2018, Astudillo-Scalia *et al.*, 2020, Evans *et al.*, 2021), slope (Cañadas *et al.*, 2002, Amorim *et al.*, 2009, Cafaro *et al.*, 2015, Waggitt *et al.*, 2019), fronts (Bost *et al.*, 2009, Nur *et al.*, 2011, Miller and Christodoulou, 2014, Scales *et al.*, 2014, Scales *et al.*, 2015, Cox *et al.*, 2018, Waggitt *et al.*, 2019), mixed layer thickness (Waggitt *et al.*, 2018, Evans *et al.*, 2021) and salinity (Bailey and Thompson, 2009, Astudillo-Scalia *et al.*, 2020).

Evaluating the associations between marine predator space use and environmental variables is important, as whilst these variables fluctuate naturally, they are also under the influence of anthropogenic impacts (most notably climate change) and big data approaches can facilitate studies into linkages between marine megafauna and their rapidly changing environment. With concerns over changes in marine biodiversity, there is considerable interest in modelling to generate spatial predictions (maps) from occurrence data (Guisan and Thuiller, 2005) and is increasingly important in marine spatial planning (Sequeira *et al.*, 2021).

1.3 Modelling approaches to assessing drivers of biodiversity

Approaches to assessing linkages between predators and their environment have commonly included modelling approaches such as macro-ecological models (MEMs) and stacked species distribution models (SSDMs), though approaches are reviewed in Thuiller *et al.* (2009) and Oppel *et al.* (2012). MEMs follow an “assemble first, predict later” approach, where species are combined into community-level measures such as species richness first before modelling prediction maps. They require only the measure values (e.g. species richness) therefore a computationally quick, but they lose species-specific information to interrogate. SSDMs instead utilise a “predict first, assemble later” approach, where species distributions are modelled individually first and then combined to a community level (Ferrier and Guisan, 2006, D'Amen *et al.*, 2015, Zhang, 2022). These are computationally more intensive but retain species-specific information such as individual distributions. The number and complexity of modelling techniques used to predict species distributions have amplified considerably in recent decades (Hegel *et al.*, 2010), with modellers making use of increasing volumes of presence-only data (Pearce and Boyce, 2006, Ready *et al.*, 2010) and comparisons of model performance have been carried out for terrestrial species (Segurado and Araujo, 2004, Elith *et al.*, 2006, Elith and Graham, 2009), but less so for the marine environment due to its complexity (Wakefield *et al.*, 2009, Ready *et al.*, 2010, Robinson *et al.*, 2011).

Modelling spatial patterns of biodiversity and their environmental drivers have historically concentrated on one or just a few taxa (Sutcliffe *et al.*, 2014), though studies on biodiversity patterns in marine species at a large cross-taxa scale do exist (e.g. Tittensor *et al.* (2010), McClellan *et al.* (2014), Sutcliffe *et al.* (2014) and Hamilton *et al.* (2022)). Cross-taxa studies on top marine predators are rarer, but studies have mapped the biodiversity of marine mammals and seabirds in the tropical Indian Ocean (Laran *et al.*, 2017), the Antarctic (Santora and Veit, 2013, Hindell *et al.*, 2020), the Arctic (Hamilton *et al.*, 2022), the North Pacific Ocean (Block *et al.*, 2011) and globally (Albouy *et al.*, 2017). Cross-taxa and cross-realm research can provide insights to advancing our understanding of processes or mechanisms of species survival (Pinsky *et al.*, 2022).

1.4 Thesis structure

Mapping biodiversity at a large scale is challenging due to patchy data coverage and a particularly cryptic set of species, however, there are methods available to utilise historical datasets to increase spatial and temporal coverage, whilst identifying and quantifying the inherent biases within them. Mapping and protecting biodiversity hotspots represent a cost-effective use of resources to protect planetary biota (Pimm and Raven, 2000, Roberts *et al.*, 2002, Mittermeier *et al.*, 2011, Marchese, 2015). My thesis aims to understand multi-species space use by marine predators from multiple taxa and define biodiversity ‘hotspots’ in the marine realm. I identified the lack of biodiversity mapping of these top predators around the UK and have provided a worked methodology to produce maps of hotspots and investigate drivers of those patterns.

The chapters in this thesis are written as individual chapters with stand-alone novel uses, however, they are linked together as one thesis that follows a logical novel methodology to define biodiversity hotspots of marine predators (Figure 1.5).

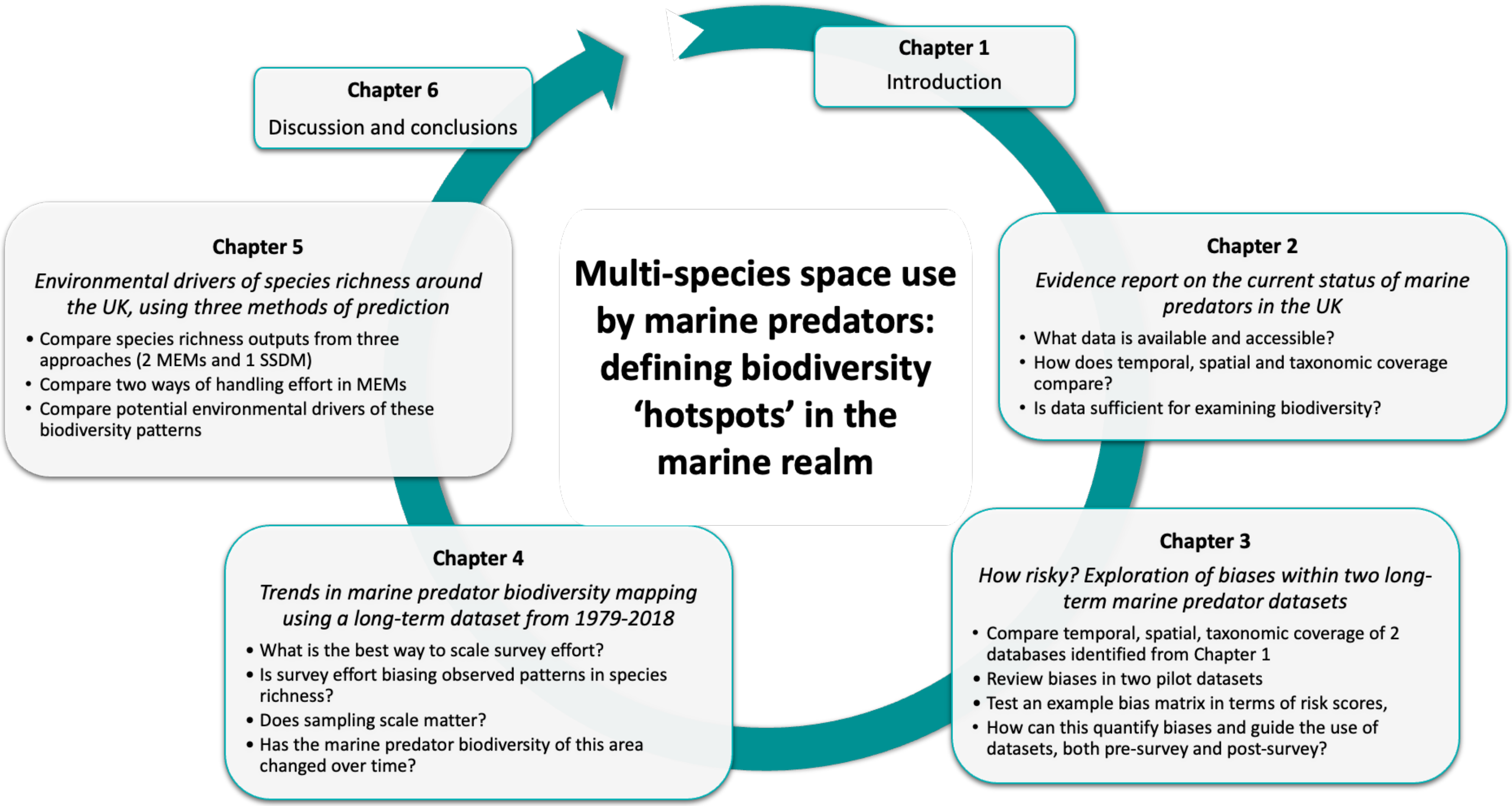
To summarise the data available and accessible for use in biodiversity mapping, Chapter 1 is an evidence report on the current status of marine predator datasets around the UK. It identifies the spatial, temporal and taxonomic coverage of data from portals and assesses if there is sufficient data available for large spatial biodiversity patterns.

Chapter 2 then further explores two large datasets identified in Chapter 1 to investigate and compare the level of bias within them. The well-known risk assessment matrix approach is adapted to quantify biases within datasets and guides the use of large-scale databases for further research, management and policy. This chapter identifies that the level of risk in using heterogeneous datasets is lower for assessing patterns of association in marine predators, rather than counts or abundances.

The two datasets are utilised in Chapter 3 produce maps of species richness around the UK. To address the uncertainty in factoring in effort post-survey when mapping biodiversity, a method is advocated to produce a new measure of species richness per unit effort (SRPUE). Patterns in raw species richness and SRPUE are then compared to assess the influence of survey effort standardisation on the conclusions we can draw from maps.

Understanding potential drivers of species richness patterns around the UK is important, particularly with anthropogenic climate change causing shifts in species coexistence. Therefore, to add to the research field surrounding potential drivers of top marine predator biodiversity hotspots, Chapter 4 uses three modelling approaches to look at associations of high species richness and a suite of environmental variables. MEMs in the form of Generalised Additive Models (GAMs) are used in two approaches: the first uses effort as a predictor variable in richness (therefore does not factor in survey effort in advance of modelling), whilst the second factors in effort in advance of modelling, utilising SRPUE richness maps from Chapter 3. The other modelling approach uses a SSDM to map species distributions in response to environmental variables and then combines them to model species richness and identify any common drivers of richness. It provides a comparison of modelling approaches with recommendations for specific use in future biodiversity modelling.

The general discussion (Chapter 5) synthesises the outcomes of this research and the implications for practical use in policy, management and conservation. I highlight the cost-effective approach to utilising big data to address large-scale biodiversity questions and provide recommendations on methodologies that may aid further research into biodiversity around the UK.



Multi-species space use by marine predators: defining biodiversity 'hotspots' in the marine realm

Chapter 1

Introduction

Chapter 2

Evidence report on the current status of marine predators in the UK

- What data is available and accessible?
- How does temporal, spatial and taxonomic coverage compare?
- Is data sufficient for examining biodiversity?

Chapter 3

How risky? Exploration of biases within two long-term marine predator datasets

- Compare temporal, spatial, taxonomic coverage of 2 databases identified from Chapter 1
- Review biases in two pilot datasets
- Test an example bias matrix in terms of risk scores,
- How can this quantify biases and guide the use of datasets, both pre-survey and post-survey?

Chapter 4

Trends in marine predator biodiversity mapping using a long-term dataset from 1979-2018

- What is the best way to scale survey effort?
- Is survey effort biasing observed patterns in species richness?
- Does sampling scale matter?
- Has the marine predator biodiversity of this area changed over time?

Chapter 5

Environmental drivers of species richness around the UK, using three methods of prediction

- Compare species richness outputs from three approaches (2 MEMs and 1 SSDM)
- Compare two ways of handling effort in MEMs
- Compare potential environmental drivers of these biodiversity patterns

Chapter 6

Discussion and conclusions

Figure 1.5 Thesis outline.

1.5 References

AGUILAR DE SOTO, N., ROGAN, E., Ó CADHLA, O., GORDON, J. C. D., MACKEY, M. & CONNOLLY, N. 2004. Volume III – Acoustic surveys for cetaceans. *Cetaceans and Seabirds of Ireland's Atlantic Margin. Report on research carried out under the Irish Infrastructure Programme (PIP): Rockall Studies Group (RSG) projects 98/6 and 00/13, Porcupine Studies Group project P00/15 and Offshore Support Group (OSG) project 99/38.*

AUBIN, I., CARDOU, F., BOISVERT-MARSH, L., GARNIER, E., STRUKELJ, M., MUNSON, A. D. & BELLO, F. 2020. Managing data locally to answer questions globally: The role of collaborative science in ecology. *Journal of Vegetation Science,* 31**,** 509-517.

BENSON, D. A., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & WHEELER, D. L. 2005. GenBank. *Nucleic Acids Res,* 33**,** D34-8.

BERROW, S. & O' CONNOR, I. 2013. Marine Mammals and Megafauna in Irish Waters - Behaviour, Distribution and Habitat Use. WP3 Biotelemetry of Marine Megafauna in Irish Waters (Project-based Award). *In:* GROUP, I. W. A. D. (ed.) *NDP Marine Research Sub-Programme 2007-2013.* Irish Whale and Dolphin Group.

BERROW, S. D., O'BRIEN, J., O'CONNOR, I. & MCGRATH, D. A. W., D. 2013. Marine Mammals and Megafauna in Irish Waters - Behaviour, Distribution and Habitat Use. Final Summary Report. *In:* MARINE INSTITUTE (ed.) *Marine Research Sub-Programme (NDP 2007-'13) Series.* Galway: Marine Institute,.

BETTY, E. L., BOLLARD, B., MURPHY, S., OGLE, M., HENDRIKS, H., ORAMS, M. B. & STOCKIN, K. A. 2019. Using emerging hot spot analysis of stranding records to inform conservation management of a data-poor cetacean species. *Biodiversity and Conservation,* 29**,** 643-665.

BLOCK, B. A., JONSEN, I. D., JORGENSEN, S. J., WINSHIP, A. J., SHAFFER, S. A., BOGRAD, S. J., HAZEN, E. L., FOLEY, D. G., BREED, G. A., HARRISON, A. L., GANONG, J. E., SWITHENBANK, A., CASTLETON, M., DEWAR, H., MATE, B. R., SHILLINGER, G. L., SCHAEFER, K. M., BENSON, S. R., WEISE, M. J., HENRY, R. W. & COSTA, D. P. 2011. Tracking apex marine predator movements in a dynamic ocean. *Nature,* 475**,** 86-90.

BUCKLAND, S., ANDERSON, D., BURNHAM, K., LAAKE, J., BORCHERS, D. & THOMAS, L. 2001. *Introduction to distance sampling: estimating abundance of biological populations.,* Oxford, Oxford University Press.

BULL, J. C., MASON, S., WOOD, C. & PRICE, A. R. G. 2013. Benthic marine biodiversity patterns across the United Kingdom and Ireland determined from recreational diver observations: A baseline for possible species range shifts induced by climate change. *Aquatic Ecosystem Health & Management,* 16**,** 20-30.

CARTER, M., BOEHME, L., CRONIN, M., DUCK, C., GRECIAN, W. J., HASTIE, G. D., M, J., MATTHIOPOULOS, J., MCCONNELL, B., MILLER, D. L., MORRIS, C. D., MOSS, S. E. W., THOMPSON, D., THOMPSON, P. & RUSSELL, D. J. F. 2022a. Supporting data for "Sympatric seals, satellite tracking and protected areas: habitat-based distribution estimates for conservation and management". *In:* ANDREWS, U. O. S. (ed.). Online: St Andrews.

CARTER, M. I., BOEHME, L., DUCK, C.D., GRECIAN, J., HASTIE, G.D., MCCONNELL, B.J., MILLER, D.L., MORRIS, C., MOSS, S., THOMPSON, D. AND THOMPSON, P. 2020. Habitat-based predictions of at-sea distribution for grey and harbour seals in the British Isles. *Report to BEIS.* University of St Andrews.

CARTER, M. I. D., BOEHME, L., CRONIN, M. A., DUCK, C. D., GRECIAN, W. J., HASTIE, G. D., JESSOPP, M., MATTHIOPOULOS, J., MCCONNELL, B. J., MILLER, D. L., MORRIS, C. D., MOSS, S. E. W., THOMPSON, D., THOMPSON, P. M. & RUSSELL, D. J. F. 2022b. Sympatric Seals, Satellite Tracking and Protected Areas: Habitat-Based Distribution Estimates for Conservation and Management. *Frontiers in Marine Science,* 9.

CHAMBERS, L. E., PATTERSON, T., HOBDAY, A. J., ARNOULD, J. P. Y., TUCK, G. N., WILCOX, C. & DANN, P. 2014. Determining trends and environmental drivers from long-term marine mammal and seabird data: examples from Southern Australia. *Regional Environmental Change,* 15**,** 197-209.

CONVENTION ON BIOLOGICAL DIVERSITY. 2022. *Preparations for the Post-2020 Biodiversity Framework* [Online]. UN Environment Programme. Available: https://www.cbd.int/conferences/post2020 [Accessed 16th February 2022].

CORNELL LAB OF ORNITHOLOGY. 2023. *eBird* [Online]. Ithaca, New York: Cornell Lab of Ornithology. Available: https://ebird.org/home [Accessed].

CRONIN, M. A., JESSOPP, M. & DEL VILLAR, D. 2011. Tracking grey seals on Irelands' continental shelf. Ireland: Coastal & Marine Research Centre, University College Cork.

CULINA, A., BAGLIONI, M., CROWTHER, T. W., VISSER, M. E., WOUTERSEN-WINDHOUWER, S. & MANGHI, P. 2018. Navigating the unfolding open data landscape in ecology and evolution. *Nat Ecol Evol,* 2, 420-426.

DOHERTY, P. D., BAXTER, J. M., GELL, F. R., GODLEY, B. J., GRAHAM, R. T., HALL, G., HALL, J., HAWKES, L. A., HENDERSON, S. M., JOHNSON, L., SPEEDIE, C. & WITT, M. J. 2017a. Long-term satellite tracking reveals variable seasonal migration strategies of basking sharks in the north-east Atlantic. *Sci Rep,* 7, 42837.

DOHERTY, P. D., BAXTER, J. M., GODLEY, B. J., GRAHAM, R. T., HALL, G., HALL, J., HAWKES, L. A., HENDERSON, S. M., JOHNSON, L., SPEEDIE, C. & WITT, M. J. 2017b. Testing the boundaries: Seasonal residency and inter-annual site fidelity of basking sharks in a proposed Marine Protected Area. *Biological Conservation,* 209, 68-75.

DOREMUS, G. 2016. *Observatoire Pelagis boat surveys 2003-2015* [Online]. Available: http://seamap.env.duke.edu/dataset/1403 [Accessed].

DOREMUS, G. 2018. *Observatoire Pelagis boat surveys 2003-2017* [Online]. Observatoire Pelagis UMS 3462, University La Rochelle - CNRS: University La Rochelle. Available: http://seamap.env.duke.edu/dataset/1403 [Accessed 10/1/19 2019].

DOUGHTY, C. E., ROMAN, J., FAURBY, S., WOLF, A., HAQUE, A., BAKKER, E. S., MALHI, Y., DUNNING, J. B., JR. & SVENNING, J. C. 2016. Global nutrient transport in a world of giants. *Proc Natl Acad Sci U S A,* 113, 868-73.

DOYLE, T. K., BENNISON, A., JESSOPP, M., HABERLIN, D. & HARMAN, L. A. 2015. A dawn peak in the occurrence of 'knifing behaviour' in blue sharks. *Animal Biotelemetry,* 3.

DRYAD. 2023. *Data Dryad* [Online]. Available: https://datadryad.org/stash [Accessed 6th January 2023].

EUROPEAN COMMISSION & EUROPEAN ENVIRONMENT AGENCY. 2023. *Biodiversity information system for Europe (BISE)* [Online]. Available: https://biodiversity.europa.eu/ [Accessed 21st January 2023].

EVANS, K., THRESHER, R., WARNEKE, R. M., BRADSHAW, C. J., POOK, M., THIELE, D. & HINDELL, M. A. 2005. Periodic variability in cetacean strandings: links to large-scale climate events. *Biol Lett,* 1, 147-50.

FAITH, D. P., COLLEN, B., ARIÑO, A. H., KOLEFF, P., GUINOTTE, J., KERR, J. T. & CHAVAN, V. 2013. Bridging biodiversity data gaps: recommendations to meet users' data needs. *Biodiversity Informatics,* 8, 41-58.

FROESE, R. & PAULY, D. 2023. FishBase. www.fishbase.org.

GBIF. 2019. *Global Biodiversity Information Facility* [Online]. GBIF. Available: https://www.gbif.org [Accessed 19th June 2019].

GROLEMUND, G. & WICKHAM, H. 2011. Dates and Times Made Easy with lubridate.

HALPERN, B. S., FRAZIER, M., POTAPENKO, J., CASEY, K. S., KOENIG, K., LONGO, C., LOWNDES, J. S., ROCKWOOD, R. C., SELIG, E. R., SELKOE, K. A. & WALBRIDGE, S. 2015. Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nat Commun,* 6, 7615.

HALPIN, P. N., READ, A. J., FUJIOKA, E., BEST, B. D., DONNELLY, B., HAZEN, L. J., KOT, C., URIAN, K., LABRECQUE, E., DIMATTEO, A., CLEARY, J., GOOD, C., CROWDER, L. B. & HYRENBACH, K. D. 2009. OBIS-SEAMAP: The world data center for marine mammal, sea bird, and sea turtle distributions. *Oceanography,* 22, 104-115.

HAMMOND, P. S., BERGGREN, P., BENKE, H., BORCHERS, D.L., COLLET, A., HEIDE-JØRGENSEN, M.P., HEIMLICH, S., HIBY, A.R., LEOPOLD, M.F. AND ØIEN, N 2002. Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology,* 39, 361-376.

HAMMOND, P. S., LACEY, C., GILLES, A., VIQUERAT, S., BÖRJESSON, P., HERR, H., MACLEOD, K., RIDOUX, V., SANTOS, M., SCHEIDAT, M. AND TEILMANN, J. 2021. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. *In:* SMRU (ed.). St Andrews, Scotland, UK: Sea Mammal Research Unit (SMRU).

HAMMOND, P. S., MACLEOD, K., BERGGREN, P., BORCHERS, D. L., BURT, L., CAÑADAS, A., DESPORTES, G., DONOVAN, G. P., GILLES, A., GILLESPIE, D., GORDON, J., HIBY, L., KUKLIK, I., LEAPER, R., LEHNERT, K., LEOPOLD, M., LOVELL, P., ØIEN, N., PAXTON, C. G. M., RIDOUX, V., ROGAN, E., SAMARRA, F., SCHEIDAT, M., SEQUEIRA, M., SIEBERT, U., SKOV, H., SWIFT, R., TASKER, M. L., TEILMANN, J.,

VAN CANNEYT, O. & VÁZQUEZ, J. A. 2013. Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation,* 164**,** 107-122.

HAYS, G. C., SCOTT, R. & HIGHAM, T. 2013. Global patterns for upper ceilings on migration distance in sea turtles and comparisons with fish, birds and mammals. *Functional Ecology,* 27**,** 748-756.

HEBERLING, J. M., MILLER, J. T., NOESGAARD, D., WEINGART, S. B. & SCHIGEL, D. 2021. Data integration enables global biodiversity synthesis. *Proc Natl Acad Sci U S A,* 118.

HIBY, L. & LOVELL, P. 1998. Using aircraft in tandem formation to estimate abundance of harbour porpoise. *Biometrics***,** 1280-1289.

HIGHAM, J. E. & LÜCK, M. 2007. *Marine wildlife and tourism management: Insights from the natural and social sciences,* Wallingford, UK, CABI Publishing.

HOBDAY, A. J., ARRIZABALAGA, H., EVANS, K., NICOL, S., YOUNG, J. W. & WENG, K. C. 2015. Impacts of climate change on marine top predators: advances and future challenges. Elsevier.

HOEGH-GULDBERG, O. & BRUNO, J. F. 2010. The impact of climate change on the world's marine ecosystems. *Science,* 328**,** 1523-8.

HUNT, G. L. & MCKINNELL, S. 2006. Interplay between top-down, bottom- up, and wasp-waist control in control in marine ecosystems. *Progress in Oceanography,* 68**,** 115–124.

JCDP. 2022. *Joint Cetacean Data Programme* [Online]. Copenhagen, Denmark. https://cetaceans.ices.dk. [Accessed 26th March 2023].

JONES, E., SMOUT, S., BLIGHT, C., SPARLING, C. & MCCONNELL, B. 2016. Fine-scale harbour seal at-sea usage mapping around Orkney and the North coast of Scotland. *Scottish Marine and Freshwater Science,* 7**,** 1-37.

JONES, J. P., COLLEN, B., ATKINSON, G., BAXTER, P. W., BUBB, P., ILLIAN, J. B., KATZNER, T. E., KEANE, A., LOH, J., MCDONALD-MADDEN, E., NICHOLSON, E., PEREIRA, H. M., POSSINGHAM, H. P., PULLIN, A. S., RODRIGUES, A. S., RUIZ-GUTIERREZ, V., SOMMERVILLE, M. & MILNER-GULLAND, E. J. 2011. The why, what, and how of global biodiversity indicators beyond the 2010 target. *Conserv Biol,* 25**,** 450-7.

JONES, P. G., ASNER, G. P., BUTCHART, S. H. M. & KARANTH, K. U. 2013. The 'why', 'what' and 'how' of monitoring for conservation. *Key Topics in Conservation Biology 2***,** 327-343.

KACPRZAK, E., KOESTEN, L., IBÁÑEZ, L.-D., BLOUNT, T., TENNISON, J. & SIMPERL, E. 2019. Characterising dataset search—An analysis of search logs and data requests. *Journal of Web Semantics,* 55**,** 37-55.

KASCHNER, K., QUICK, N. J., JEWELL, R., WILLIAMS, R. & HARRIS, C. M. 2012. Global coverage of cetacean line-transect surveys: status quo, data gaps and future challenges. *PLoS One,* 7**,** e44075.

KATTGE, J., DÍAZ, S., LAVOREL, S., PRENTICE, I. C., LEADLEY, P., BÖNISCH, G., GARNIER, E., WESTOBY, M., REICH, P. B., WRIGHT, I. J., CORNELISSEN, J. H. C., VIOLLE, C., HARRISON, S. P., VAN BODEGOM, P. M., REICHSTEIN, M., ENQUIST, B. J., SOUDZILOVSKAIA, N. A., ACKERLY, D. D., ANAND, M., ATKIN, O., BAHN, M., BAKER, T. R., BALDOCCHI, D., BEKKER, R., BLANCO, C. C., BLONDER, B., BOND, W. J., BRADSTOCK, R., BUNKER, D. E., CASANOVES, F., CAVENDER-BARES, J., CHAMBERS, J. Q., CHAPIN III, F. S., CHAVE, J., COOMES, D., CORNWELL, W. K., CRAINE, J. M., DOBRIN, B. H., DUARTE, L., DURKA, W., ELSER, J., ESSER, G., ESTIARTE, M., FAGAN, W. F., FANG, J., FERNÁNDEZ-MÉNDEZ, F., FIDELIS, A., FINEGAN, B., FLORES, O., FORD, H., FRANK, D., FRESCHET, G. T., FYLLAS, N. M., GALLAGHER, R. V., GREEN, W. A., GUTIERREZ, A. G., HICKLER, T., HIGGINS, S. I., HODGSON, J. G., JALILI, A., JANSEN, S., JOLY, C. A., KERKHOFF, A. J., KIRKUP, D., KITAJIMA, K., KLEYER, M., KLOTZ, S., KNOPS, J. M. H., KRAMER, K., KÜHN, I., KUROKAWA, H., LAUGHLIN, D., LEE, T. D., LEISHMAN, M., LENS, F., LENZ, T., LEWIS, S. L., LLOYD, J., LLUSIÀ, J., LOUAULT, F., MA, S., MAHECHA, M. D., MANNING, P., MASSAD, T., MEDLYN, B. E., MESSIER, J., MOLES, A. T., MÜLLER, S. C., NADROWSKI, K., NAEEM, S., NIINEMETS, Ü., NÖLLERT, S., NÜSKE, A., OGAYA, R., OLEKSYN, J., ONIPCHENKO, V. G., ONODA, Y., ORDOÑEZ, J., OVERBECK, G., OZINGA, W. A., et al. 2011. TRY - a global database of plant traits. *Global Change Biology,* 17**,** 2905-2935.

KONIG, C., WEIGELT, P., SCHRADER, J., TAYLOR, A., KATTGE, J. & KREFT, H. 2019. Biodiversity data integration-the significance of data resolution and domain. *PLoS Biol,* 17**,** e3000183.

LOFFLER, F., WESP, V., KONIG-RIES, B. & KLAN, F. 2021. Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? *PLoS One,* 16**,** e0246099.

MACKEY, M., Ó CADHLA, O., KELLY, T.C., AGUILAR DE SOTO, N. & CONNOLLY, N 2004. Cetaceans and Seabirds of Ireland's Atlantic Margin. Volume I – Seabird distribution, density & abundance. *Report on research carried out under the Irish Infrastructure Programme (PIP): Rockall Studies Group (RSG) projects 98/6 and 00/13, Porcupine Studies Group project P00/15 and Offshore Support Group (OSG) project 99/38.*

MANNOCCI, L., ROBERTS, J. J., HALPIN, P. N., AUTHIER, M., BOISSEAU, O., BRADAI, M. N., CANADAS, A., CHICOTE, C., DAVID, L., DI-MEGLIO, N., FORTUNA, C. M., FRANTZIS, A., GAZO, M., GENOV, T., HAMMOND, P. S., HOLCER, D., KASCHNER, K., KEREM, D., LAURIANO, G., LEWIS, T., NOTARBARTOLO DI SCIARA, G., PANIGADA, S., RAGA, J. A., SCHEININ, A., RIDOUX, V., VELLA, A. & VELLA, J. 2018. Assessing cetacean surveys throughout the Mediterranean Sea: a gap analysis in environmental space. *Sci Rep,* 8**,** 3126.

MCCONNELL, B., SMOUT, S. & WU, M. 2017. Modelling Harbour Seal Movements. *Scottish Marine and Freshwater Science,* 8**,** 1-33.

MICROSOFT CORPORATION 2018. Microsoft Excel. 16.70 ed.

MURPHY, S., EVANS, P. G. H., PINN, E. & PIERCE, G. J. 2019. Conservation management of common dolphins: Lessons learned from the North-East Atlantic. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 31**,** 137-166.

NATIONAL BIODIVERSITY NETWORK. 2022. *NBN Atlas* [Online]. Available: http://www.nbnatlas.org [Accessed 14th February 2022].

NATIONAL OCEANOGRAPHY CENTRE (NOC). 2022. *Marine Environmental Data and Information Network (MEDIN)* [Online]. Online. Available: https://medin.org.uk/ [Accessed 14th February 2022].

NATURE 2020. New biodiversity targets cannot afford to fail. *Nature,* 578**,** 338.

NUR, N., JAHNCKE, J., HERZOG, M. P., HOWAR, J., HYRENBACH, K. D., ZAMON, J. E., AINLEY, D. G., WIENS, J. A., MORGAN, K., BALLANCE, L. T. & STRALBERG, D. 2011. Where the wild things are: predicting hotspots of seabird aggregations in the California Current System. *Ecol Appl,* 21**,** 2241-57.

Ó CADHLA, O., MACKEY, M., AGUILAR DE SOTO, N., ROGAN, E. & CONNOLLY, N. 2004. Volume II – Cetacean distribution & abundance. *Cetaceans and Seabirds of Ireland's Atlantic Margin. Report on research carried out under the Irish Infrastructure Programme (PIP): Rockall Studies Group (RSG) projects 98/6 and 00/13, Porcupine Studies Group project P00/15 and Offshore Support Group (OSG) project 99/38.*

O'HARA, C. C., VILLASEÑOR-DERBEZ, J. C., RALPH, G. M. & HALPERN, B. S. 2019. Mapping status and conservation of global at-risk marine biodiversity. *Conservation Letters,* 12.

PEREIRA, H. M., LEADLEY, P. W., PROENÇA, V., ALKEMADE, R., SCHARLEMANN, J. P. W., FERNANDEZ-MANJARRÉS, J. F., ARAÚJO, M. B., BALVANERA, P., BIGGS, R., CHEUNG, W. W. L., CHINI, L., COOPER, H. D., GILMAN, E. L., GUÉNETTE, S., HURTT, G. C., HUNTINGTON, H. P., MACE, G. M., BERDORFF, T., REVENGA, C., RODRIGUES, P., SCHOLES, R. J., SUMAILA, U. R. & WALPOLE, M. 2010. Scenarios for Global Biodiversity in the 21st Century. *Science***,** 1496-1501.

R CORE TEAM 2022. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

ROBERTS, J. J., BEST, B. D., MANNOCCI, L., FUJIOKA, E., HALPIN, P. N., PALKA, D. L., GARRISON, L. P., MULLIN, K. D., COLE, T. V., KHAN, C. B., MCLELLAN, W. A., PABST, D. A. & LOCKHART, G. G. 2016. Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Sci Rep,* 6**,** 22615.

ROCHE, D. G., KRUUK, L. E., LANFEAR, R. & BINNING, S. A. 2015. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLoS Biol,* 13**,** e1002295.

SANTORA, J. A. & VEIT, R. R. 2013. Spatio-temporal persistence of top predator hotspots near the Antarctic Peninsula. *Marine Ecology Progress Series,* 487**,** 287-304.

SARAN, S., CHAUDHARY, S. K., SINGH, P., TIWARI, A. & KUMAR, V. 2022. A comprehensive review on biodiversity information portals. *Biodiversity and Conservation,* 31**,** 1445-1468.

SURYAN, R. M., SANTORA, J. A. & SYDEMAN, W. J. 2012. New approach for using remotely sensed chlorophyll a to identify seabird hotspots. *Marine Ecology Progress Series,* 451**,** 213-225.

SUŠKEVIČS, M., RAADOM, T., VANEM, B., KANA, S., ROASTO, R., RUNNEL, V. & KÜLVIK, M. 2021. Challenges and opportunities of engaging biodiversity-related citizen science data in environmental decision-making: Practitioners' perceptions and a database analysis from Estonia. *Journal for Nature Conservation,* 64.

TITTENSOR, D. P., MORA, C., JETZ, W., LOTZE, H. K., RICARD, D., BERGHE, E. V. & WORM, B. 2010. Global patterns and predictors of marine biodiversity across taxa. *Nature,* 466**,** 1098-101.

VAN CANNEYT, O. 2016. *Observatoire Pelagis aerial surveys 2002-2015.* [Online]. Available: http://seamap.env.duke.edu/dataset/1404 [Accessed].

WAGGITT, J. J., EVANS, P. G. H., ANDRADE, J., BANKS, A. N., BOISSEAU, O., BOLTON, M., BRADBURY, G., BRERETON, T., CAMPHUYSEN, C. J., DURINCK, J., FELCE, T., FIJN, R. C., GARCIA-BARON, I., GARTHE, S., GEELHOED, S. C. V., GILLES, A., GOODALL, M., HAELTERS, J., HAMILTON, S., HARTNY-MILLS, L., HODGINS, N., JAMES, K., JESSOPP, M., KAVANAGH, A. S., LEOPOLD, M., LOHRENGEL, K., LOUZAO, M., MARKONES, N., MARTÍNEZ-CEDEIRA, J., Ó CADHLA, O., PERRY, S. L., PIERCE, G. J., RIDOUX, V., ROBINSON, K. P., SANTOS, M. B., SAAVEDRA, C., SKOV, H., STIENEN, E. W. M., SVEEGAARD, S., THOMPSON, P., VANERMEN, N., WALL, D., WEBB, A., WILSON, J., WANLESS, S., HIDDINK, J. G. & PUNT, A. 2019. Distribution maps of cetacean and seabird populations in the North-East Atlantic. *Journal of Applied Ecology,* 57**,** 253-269.

WICKHAM, H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

WICKHAM, H. 2019. stringr: Simple, Consistent Wrappers for Common String Operations.

WICKHAM, H., FRANÇOIS, R., HENRY, L. & MÜLLER, K. 2022. dplyr: A Grammar of Data Manipulation.

Chapter 2 Evidence report on the current status of marine predator datasets in the UK

2.1 Abstract

Marine predators such as marine mammals, seabirds and sharks are keystone sentinel species with crucial ecological roles. Whilst they are of high socio-economic and cultural importance given their charismatic status, they require a large amount of offshore survey effort to have sufficient data for large-scale analysis of distribution, coexistence and biodiversity. Targets for biodiversity are of global importance but monitoring of highly mobile species in relatively inaccessible habitats presents a considerable and financial challenge for conservation, thus exploiting available historical surveys and databases may aid in research and conservation planning, with significant financial benefit if used appropriately. Obtaining baselines of biodiversity is crucial for future conservation and though knowledge of distributions of marine predators in the North Atlantic has improved in recent decades, multi-taxa studies are lacking. This evidence report summarises the status and knowledge of datasets of marine predators (seabirds, sharks, turtles, cetaceans, pinnipeds) around the UK to identify available suitable surveys, ascertaining ease of access to data and selecting surveys with potential for examining large spatial and temporal biodiversity patterns. The study found that whilst there are many historical datasets of marine predator records in data portals, access to them is often limited or requires direct contact with data providers. However, from the 246 datasets interrogated, 14 datasets were identified that would be suitable for multi-taxa study, and though the provision of this data varies in quality, temporal and spatial coverage and accuracy, historical datasets exist that can be a key resource for assessing biodiversity, as it allows cumulative data to increase temporal and spatial coverage of an area at a low financial cost.

2.2 Introduction

Top marine predators such as cetaceans, seabirds, sharks, pinnipeds and turtles (Hobday *et al.*, 2015) have crucial ecological roles as ecosystem engineers, with top-down regulation on trophic levels (Hunt and McKinnell, 2006) and nutrient cycling (Doughty *et al.*, 2016), but face challenges from environmental change (Hoegh-Guldberg and Bruno, 2010) and anthropogenic activities (Halpern *et al.*, 2015). Identified as sentinel species which can reflect ocean health (Evans *et al.*, 2005, Tittensor *et al.*, 2010), they have high socio-economic and cultural importance given their charismatic status (Higham and Lück, 2007), and so are often a focus of public interest and long-term research programs (Chambers *et al.*, 2014). However these species are often wide-ranging and have low detectability due to their marine lifestyle, and therefore require a large amount of offshore survey effort to have sufficient data (Kaschner *et al.*, 2012). Monitoring of highly mobile species in relatively inaccessible habitats presents a considerable financial challenge for conservation (Betty *et al.*, 2019). Therefore, exploiting available historical surveys and databases may aid in research and conservation planning, with significant financial benefit if used appropriately.

Historical datasets can be key for assessing biodiversity, as it allows cumulative data to increase the temporal and spatial coverage of an area. Targets for biodiversity are of global importance, with most of the 2010 Aichi Biodiversity Targets adopted by the United Nations Convention on Biodiversity (CBD) to counteract the decline of global biodiversity, not met (Nature, 2020). Effective protection of 10% of marine areas important to biodiversity and ecosystem services by 2020 was set by Aichi Target 11 (Pereira *et al.*, 2010). New targets are being agreed upon in the Post-2020 Global Biodiversity Framework (Convention on Biological Diversity, 2022), with the aim to slow down the rate of biodiversity loss and conserve 30% of land and sea areas by 2030. The advancement of information technology has allowed many open-source platforms to develop (Saran *et al.*, 2022), with initiatives such as the Global Biodiversity Information Facility (GBIF) (GBIF, 2019), TRY (Kattge *et al.*, 2011), GenBank (Benson *et al.*, 2005) and Biodiversity Information System for Europe (BISE) (European Commission and European Environment Agency, 2023) now providing access to large collections of biological data, facilitating comprehensive analyses across taxa that was not previously possible.

To conserve marine biodiversity, we must first understand the spatial distribution and status of marine species (O'Hara *et al.*, 2019). Monitoring is universally recognised as vital for biodiversity conservation (Jones *et al.*, 2013), and large-scale approaches are necessary, which may require integrating data (Jones *et al.*, 2011), as drivers of biodiversity loss operate at larger scales. Obtaining baselines of biodiversity is crucial for future conservation, and there is a need for quantitative designation of 'hotspots' - areas characterised by persistently elevated abundance or species richness (Nur *et al.*, 2011, Suryan *et al.*, 2012, Santora and Veit, 2013). Knowledge of distributions of marine predators in the North Atlantic has improved in recent decades, stimulated by both legislative requirements and public concern (Murphy *et al.*, 2019), but species are widely distributed and occupy numerous ecological niches with diverse demographics and population dynamics. This means that whilst species-specific surveys can give fine-scale use of an area, they may not be sufficient to capture the full biodiversity of marine predators around the United Kingdom.

For marine predators, many methods of data collection exist, including fine-scale tracking and tagging (Block *et al.*, 2011) or shipboard or aerial surveys which record abundance in an area (Suryan *et al.*, 2012), but obtaining multiple datasets that have the temporal and spatial coverage to reveal large scale patterns, highlight areas of protection or quantify overlap with anthropogenic activities demands an alternative approach. To maximise effort in time and space, survey data can be collated from as many different sources and suppliers as possible (Roberts *et al.*, 2016, Mannocci *et al.*, 2018), maximising coverage with heterogeneous data. Obtaining such data requires an understanding the routes to find data and examining accessibility and availability both in terms of species and survey type.

2.2.1 Aims

This evidence report aims to summarise the status and knowledge of datasets of marine predators around the UK to identify surveys, ascertain the ease of access and select datasets with the spatial and temporal potential for examining large scale biodiversity patterns.

Specific aims included:

- Quantify the number of available and accessible datasets containing marine predator taxa, recording parameters such as the number of observations, taxa, and survey methods.
- Compare the temporal coverage of datasets.
- Assess if there is sufficient data available for examining large spatial and temporal biodiversity patterns.

2.3 Methodology

2.3.1 Data portals

The search strategy began with identifying databases that collate species distribution or abundance data, primarily ‘Ocean Biodiversity Information System Spatial Ecological Analysis of Megavertebrate Populations’ (OBIS-SEAMAP) (Halpin *et al.*, 2009), National Biodiversity Network Atlas (NBN) (National Biodiversity Network, 2022) and Marine Environmental Data and Information Network (MEDIN) (National Oceanography Centre (NOC), 2022). These central data centres provide ease of access to numerous datasets at a high level. Where data did not occur in the waters surrounding the UK, they were excluded.

At the time of searching (19th June 2019), the MEDIN portal contains approximately 13,634 UK marine datasets from over 400 providers, through a network of Data Archive Centres, but is structured to consistent data standards with standardised field names and controlled vocabularies to allow for cross-comparison. The portal has a search feature with results downloadable as CSV files, with incorporated parameters such as survey type and survey length. It also allows datasets to be viewed as shapefile layers on a map to visualise location and coverage. The spatial focus of the MEDIN portal is UK territorial waters-based, rather than global coverage (unlike OBIS-SEAMAP and NBN ATLAS).

OBIS-SEAMAP is an open-access data platform, that focuses on marine mega-vertebrate populations. It is part of the Ocean Biodiversity Information System (OBIS) which integrates occurrence records and quality controls datasets, for over 135,000 marine species. OBIS-SEAMAP is a temporally and spatially interactive database for marine mammals, sea turtles, seabirds, rays, and sharks. Datasets were global, but searches were limited to the area of interest (AOI) around UK waters using the polygon tool (i.e. a rectangular search area with top-left corner = 72.8175 °N, 26.3348 °W to bottom-right corner 33.9372 °N, 18.4813 °E, given in Appendix A, Figure A.1) and only datasets that had coverage within that AOI were included.

The NBN Atlas aggregates UK-based biodiversity records from data partners and compiles them to make available online. In 2022 it has data on 46,856 species from 165 data partners with over 200 million species records. It is not limited to marine species, unlike MEDIN and OBIS-SEAMAP, and has many terrestrial species in the database, so records returned are not limited to searched taxa terms only, and surveys contain other non-marine predator species which would subsequently need filtering out manually.

Within these three databases, searches were carried out for the terms ‘cetaceans’, ‘seabirds’, ‘sharks’, ‘pinnipeds’ and ‘turtles’, and datasets returned were recorded. Where available the start and end date of the dataset was recorded, along with survey type category and accessibility online. If only the year was given, this was recorded, or if the exact start or end date was not available, the report publication date was used. The majority of land-based datasets with little sea coverage or primarily focused on terrestrial species were removed.

2.3.2 Non-portal

Other searches on web search engines were carried out to check for additional datasets that may not be held within the data portals. Search terms included combinations of: “marine predator”, ‘cetaceans’, ‘seabirds’, ‘sharks’, ‘pinnipeds’, ‘turtles’, “occurrence data”, “sighting data”, “biodiversity database”, “database” and datasets that had not been previously found in databases were recorded. Where multiple datasets were obtained from one data provider, the provider was further explored to see if additional datasets were available, but not held in the portal.

2.3.3 Analysis from database searches

Datasets from the database searches were compiled in a central database in Excel (version 16.7) (Microsoft Corporation, 2018), and information about each dataset stored within the fields was recorded and presented in Table 2.1, to allow consistency for comparison between datasets. Summary statistics and graphs for visualisation of the datasets were produced using R version 4.2.2 (R Core Team, 2022), using packages ggolot2, stringr, dplyr and lubridate (Grolemund and Wickham, 2011, Wickham, 2016, Wickham, 2019, Wickham *et al.*, 2022).

Table 2.1 Description of the data fields contained in the compiled central database.

Field	Description
Taxa contained	The taxa category contained within the dataset. Taxa included: Cetaceans (C), Pinnipeds (P), Seabirds (S), Sharks (Sh), Turtles (T).
Final name field	Condensed name of the dataset. Some datasets had very long descriptions, so this has been condensed in the database.
Freely available	Determined if the dataset was freely available to download, or whether it was held privately or behind paywalls.
Link	Website link to the dataset.
Date accessed	Date of last successful access to the dataset.
Survey type	A category for survey type was given for each dataset which best described the type of data e.g. Aerial, boat, report / review, tagging, net, shore (land-based sightings), stranding.
Start date	The start date of the survey, or where this cannot be given, the year the survey began.
End date	The end date of the survey, or where this cannot be given, the year of the last entry of the survey.
Online database	The database in which the dataset was found.
Custodian	The custodian of the data where applicable.
Originator	The author of the dataset.
Observations	Number of observations within the dataset.
Notes	Any other notes of use.

2.4 Results

2.4.1 Initial data portal search

Accessibility

Combined database searches resulted in 247 datasets that contained marine predator taxa observations, 122 of which were not freely accessible and largely from private consultancy surveys for energy developers (e.g. RPS Ltd, APEM, EON, Marine Atlas Consultants Ltd., SSE Renewables). Links to the data are provided in Appendix A (Table A-1) noting that these were the web addresses at the time of search (19th June 2019) and therefore may have altered since. Most available data were boat based, whilst fine-scale tracking and tagging data were noticeably less freely accessible (Figure 2.2).

Available online portals were a logical starting point for identifying datasets but were not without flaws. At the time of this study, MEDIN often cited data as 'accessible' but upon clicking the links lead to error pages or subsequent data holders which required further searching or specific data requests from suppliers (noting that the search was carried out 19th June 2019 and broken links may have since been fixed). Whilst it had metadata including parameters such as start dates, end dates and survey type, it did not provide number of observations or records in each dataset. Instead, data had to be examined individually to find number of records which proved time-consuming. When the aim is to identify datasets for good spatial and temporal coverage for biodiversity questions, number of observations would be a critical parameter for inclusion.

Temporal coverage

Datasets spanned from 1491 to 2022 but 93% of the datasets were post-1900 (see Appendix A, Figure A.6). There is a clear increase in the number of datasets post-1960 for all portals (Figure 2.4), particularly post-2000 for MEDIN, with a 15-fold increase from 1990 to 2020. MEDIN held 151 datasets, OBIS-SEAMAP held 50, and NBN held 45, however, there were datasets present in multiple data platforms that were allocated to MEDIN as that was the first database searched.

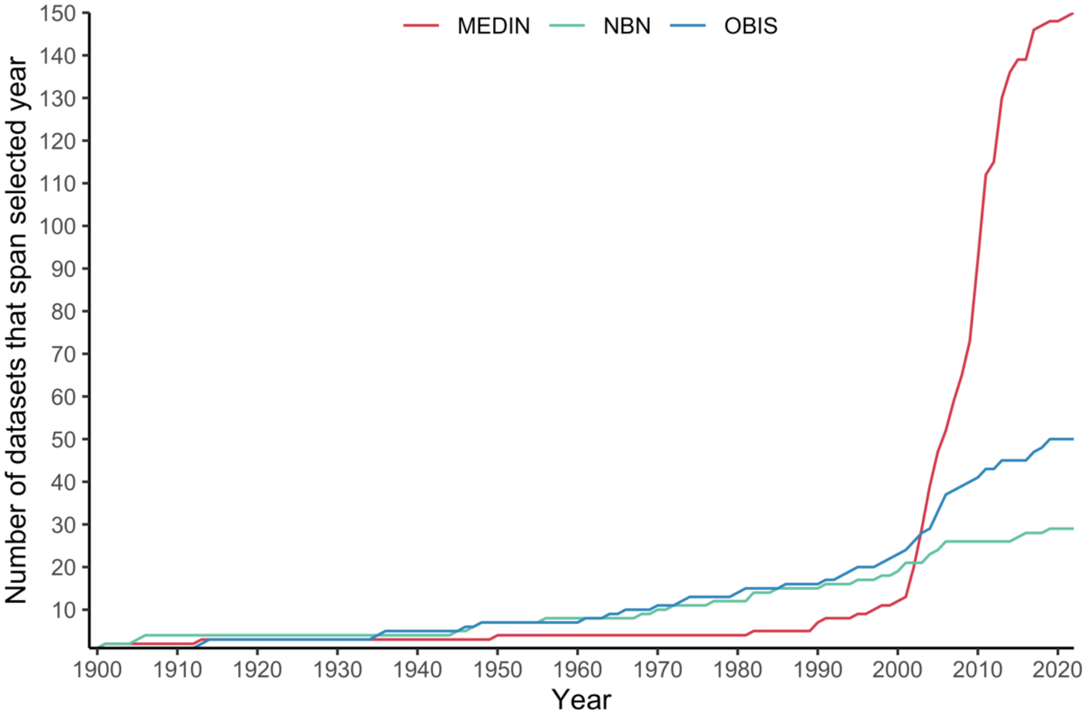


Figure 2.1 Number of datasets that cover each year, from 1900 to 2022.

Data types

A summary of data types and methodologies is presented in Table 2.2. Survey type was varied and included acoustic monitoring, aerial surveys, boat surveys, land-based surveys or shore watches, tagging, net surveys, and diving surveys (Figure 2.2).

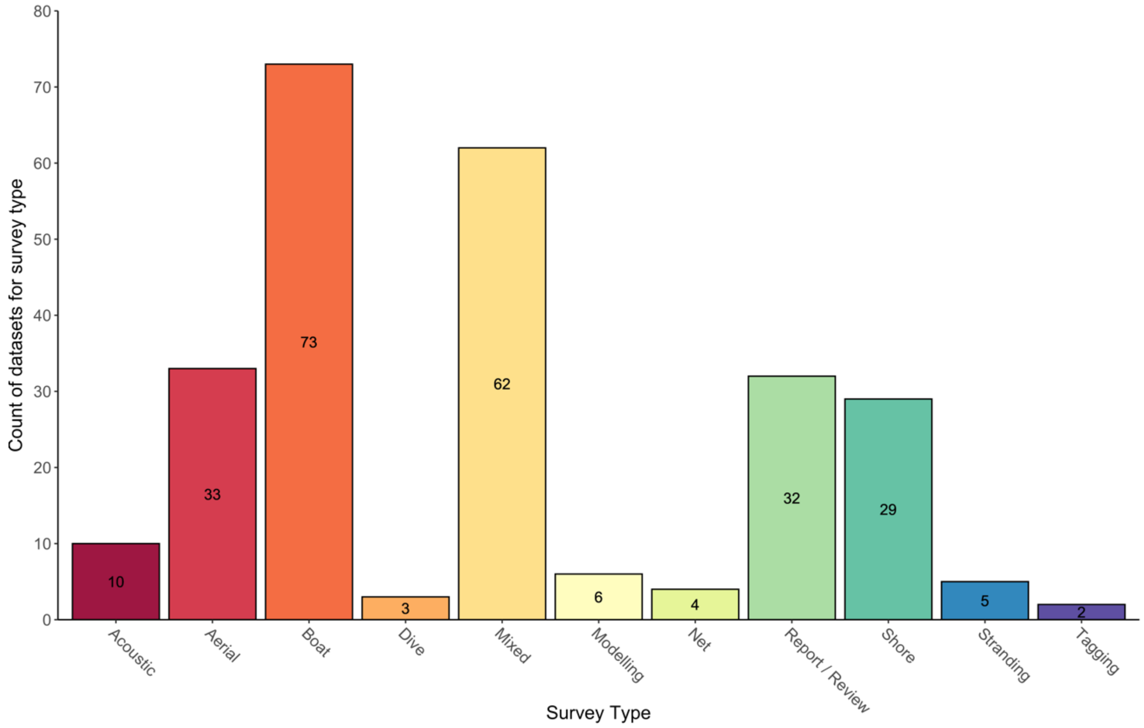


Figure 2.2 Number of datasets which utilised each survey type.

Acoustic monitoring of marine species involves the use of hydrophones or underwater microphones to detect and analyse sounds emitted by marine species, such as porpoise echolocation (Macaulay *et al.,* 2023). Hydrophones can be deployed in various ways, such as being towed behind a vessel or placed as stationary devices on the ocean floor (Sigourney et al, 2022) and collected acoustic data can be used to identify species-specific vocalisations, movement patterns, and estimate population parameters.

Aerial survey monitoring of marine mammals involves using aircraft such as airplanes or helicopters to systematically observe and record marine species over large areas of ocean. Boat surveys involve the use of vessels at a closer range, to provide detailed visual observations of marine mammals, often recording number of species, abundance, and behaviour. Some datasets were from fish net surveys, which comprised gillnets, ghost nets and trawling surveys which are commonly used to assess the abundance, species composition, and size distribution of fish in a particular area but may have bycatch of higher trophic level predators (Caretta *et al.*, 2019). One dataset from Manx Basking Shark Watch was a tagging dataset, whereby various types of tags, including satellite tags, acoustic tags, and archival tags, can be used to collect data on the predator's location, depth, temperature preferences, and migration patterns (Andrews *et al.*, 2019). Four datasets comprised close-up methods such as diving surveys, which provide fine scale identification of marine species. Fourteen surveys were land-based sighting datasets, which involve observing marine predators from fixed vantage points on land, often using binoculars or telescopes to record species behaviour, distribution, and abundance (Dolman *et al.,* 2014).

Effort or sighting based

Table 2.2 also presents whether surveys are effort-based or sighting based, with 36 of the available datasets being effort-based, and 77 sighting based. Effort-based surveys allocate effort consistently (with a predetermined amount of time or effort allocated to data collection), providing information about the likelihood of encountering marine mammals. Sighting-based surveys, record actual sightings focus on actual observations with emphasis on numbers and characteristics of individuals rather than recording periods of time with no observations.

Systematic versus opportunistic

Systematic surveys are characterised by a structured and predetermined methodology, such as predefined transects or grid cells. It can allow for comparisons over time or regions and provide a representative sample of an area, however, can be resource intensive to carry out dedicated surveys and limited in spatial coverage. Opportunistic surveys involve collecting data whenever opportunities arise such as during routine ship movements such as ferry crossings, or other activities that may not focused on dedicated data collection (such as wildlife sighting trips). These surveys can be cost-effective, as data collection is integrated into existing activities, and can cover a broad spatial region, but can lack standardisation and has a non-uniform effort.

Transect surveys

Eleven of these effort-based datasets were transect line surveys (ID 12, 13, 16, 28, 34, 38, 40, 184, 185, 186, 187). These involve a systematic approach using predetermined linear paths or grids known as transects, to cover specific study areas. Transects can be straight lines or follow geographic features, providing a standardised framework for data collection with the aim to have sufficient consistency in method and effort across all transects. Transects are often used for comparisons of abundance and density between study areas and study periods, with the measure of effort allowing density to be derived. A key example are the SCANS (Small Cetaceans in the European Atlantic and North Sea) surveys (ID 184 to 187) carried out periodically (SCANS I in 2002, SCANS II in 2005, SCANS III in 2016, and SCANS IV in 2022 (Gilles *et al, 2023,* Hammond *et al., 2002; 2007; 2016*) in summer months to allow comparison of cetacean's abundance, distribution and density in North-East Atlantic waters.

Spatial scale

Table 2.2 also presents the spatial scale of datasets in terms of the level of geographic coverage or extent. From accessible datasets, 14 were labelled global scale (covered multiple oceans and continents), 64 were regional (comprising an intermediate extent such as UK or country coverage) and 35 were local, which targeted very focused areas of coverage (e.g. ID 34 targeted a very specific area for West of Duddon Sands Offshore Wind Farm).

Table 2.2 Summary of survey type and methodologies for accessible datasets. ID column links to full dataset details in Appendix A.

ID	Final name field	Survey type	Location	Methodology	Systematic / Opportunistic	Spatial Scale
206	SMRU Ltd., SeaGen Strangford Lough, Monitoring of Harbour Porpoise in Strangford Lough with focus on the Narrows (2006-2010)	Acoustic	At-sea	Effort	Systematic	Local
208	SMRU Ltd., SeaGen Strangford Lough, Sonar Monitoring of Marine Mammals around SeaGen (2008-2009)					
12	Baltic Porpoise Acoustic Surveys 01-02					Regional
217	The East Coast Marine Mammal Acoustic Study (ECOMASS)					
222	United Kingdom Integrated Marine Observing Network (UK-IMON) initiative ocean Autumn 2013 glider deployments in the Celtic Sea.			Sighting	Opportunistic	Local
157	Observatoire Pelagis aerial surveys 2002-2021	Aerial	At-sea	Effort	Systematic	Local
185	SCANS II cetacean sightings from aerial surveys 2005					Regional
60	DHI, Offshore Renewables Joint Industry Programme (ORJIP), Bird Collision Avoidance Study (2014 - 2016)			Sighting	Systematic	Local
110	JNCC European Seabirds at Sea collected by INBO	Aerial / Boat	At-sea	Effort	Systematic	Regional
16	BLM CETAP OPP Sightings	Boat	At-sea	Effort	Opportunistic	Regional
226	Visual sightings from Song of the Whale 1993-2013				Systematic	Global
49	CRRU, Cetacean sighting in Scotland waters 1997-2015					Local
73	FishDAC - 2013 West Scotland Deepwater Trawl Survey (2013)					Local
89	Harbour porpoises, white-beaked dolphins and minke whales in North Sea - Boat surveys					Local
120	Marine Awareness North Wales, Wildlife Trust harbour porpoise baseline surveys on the north coast of Anglesey, Wales, UK					Local
158	Observatoire Pelagis boat surveys 2003-2021					Local

ID	Final name field	Survey type	Location	Methodology	Systematic / Opportunistic	Spatial Scale
189	SEA4 seabird and marine mammal observations onboard R.V. Kommandor Jack, during a multibeam survey of the seabed in the Faroe (2002)					Local
225	Visual sightings data set 2003-2020					Local
13	Baltic Porpoise Sightings 01-02					Regional
38	CODA cetacean sightings on primary platform of vessel surveys 2007					Regional
159	OCEAMM harbour porpoise sightings in the North Sea					Regional
163	PIROP Northwest Atlantic 1965-1992					Regional
184	SCANS I cetacean sightings 1994					Regional
186	SCANS II cetacean sightings on primary platform of vessel surveys 2005					Regional
187	SCANS II cetacean sightings on tracker platform of vessel surveys 2005					Regional
194	SEA7 seabird and marine mammal observations onboard the RSS Charles Darwin during British Geological Survey cruise CD174 in the Rockall Trough (2005)					Regional
34	CMACS West of Duddon Sands Offshore Wind Farm, Boat-based Ornithological Survey (2012)					Local
221	UK Royal Navy Marine Mammal Observations			Sighting	Opportunistic	Global
91	Hebridean Dolphin and Whale Trust killer whale sightings (1990-2006)					Regional
193	SEA7 Marine Mammal Observations Kommandor Jack, July August 2005 (NE Atlantic west of Scotland) (2005)					Regional
202	Seatrust Cetacean Records West Wales					Regional
2	Allied Humpback Whale Catalogue (1976 - 2003)					Regional
77	Happywhale - Atlantic White-sided Dolphin in North Atlantic Ocean	Boat and shore	At-sea / Land-based	Sighting	Opportunistic	Global
78	Happywhale - Common Bottlenose Dolphin in North Atlantic Ocean					Global
79	Happywhale - Cuvier's Beaked Whale in North Atlantic Ocean					Global
80	Happywhale - Fin Whale in North Atlantic Ocean					Global

ID	Final name field	Survey type	Location	Methodology	Systematic / Opportunistic	Spatial Scale
81	Happywhale - Harbour Porpoise in North Atlantic Ocean					Global
82	Happywhale - Humpback Whale in North Atlantic Ocean					Global
83	Happywhale - Killer Whale in North Atlantic Ocean					Global
84	Happywhale - Long-finned Pilot Whale in North Atlantic Ocean					Global
85	Happywhale - Minke Whale in North Atlantic Ocean					Global
86	Happywhale - Risso's Dolphin in North Atlantic Ocean					Global
87	Happywhale - Short-beaked Common Dolphin in North Atlantic Ocean					Global
88	Happywhale - Sperm Whale in North Atlantic Ocean					Global
199	Seasearch Marine Surveys in England	Dive	At-sea	Sighting	Opportunistic	Regional
200	Seasearch Marine Surveys in Ireland					Regional
201	Seasearch Marine Surveys in the Isle of Man					Regional
61	Diveboard - Scuba diving citizen science observations					Regional
219	The UK Archive for Marine Species and Habitats Data	Mixed	At-sea	Effort	Opportunistic	Regional
112	JNCC seabird distribution and abundance data (all trips) from ESAS database				Systematic	Regional
133	National Inventory of the Natural Heritage: Data from the air monitoring campaigns of marine megafauna (SAMM) in the French metropolitan area					Regional
15	Biodiversity of the North Sea - Sylt			Sighting	Opportunistic	Local
39	Commissioned surveys and staff surveys and reports for Scottish Wildlife Trust reserves					Local
52	DASSH Data Archive Centre volunteer survey data					Regional
58	DFO Maritimes Region Cetacean Sightings					Regional
71	Environment Agency Rare and Protected Species Records					Regional
72	ERIC NE Combined dataset to 2017					Regional
100	Incidental sightings of marine mammals					Regional

ID	Final name field	Survey type	Location	Methodology	Systematic / Opportunistic	Spatial Scale
216	The Belgian Marine Mammals Database (BMM)					Regional
131	Monitoring of the effects of Belgian wind mill parks on benthic macro-invertebrates and the fish fauna of soft substrates				Systematic	Local
152	Northern Seas Programme Data Set (2001-2007)					Local
196	Seabirds at Sea Evidence Base					Regional
50	DAERA Marine and Fisheries Division Marine Survey Data		At-sea / Land-based	Sighting	Opportunistic	Regional
51	DASSH Data Archive Centre volunteer sightings records					Regional
123	Marine List of Ireland		Land-based	Effort	Systematic	Local
116	Mammal records for Dumfries and Galloway			Sighting	Opportunistic	Local
125	Marine records from Pembrokeshire Marine Species Atlas					Local
127	Marine Species Records from Skomer Marine Conservation Zone (MCZ) Marine Monitoring Programme					Local
130	Miscellaneous records held on the Cofnod database					Local
211	Stackpole National Nature Reserve Species Inventory and Ad-hoc Sightings from Across Pembrokeshire					Local
213	Suffolk Biodiversity Information Service (SBIS) Dataset					Local
224	Vertebrates Outer Hebrides					Local
27	CEDaR Online Recording					Regional
105	Isle of Man historical wildlife records 1995 to 1999					Regional
106	Isle of Man wildlife records from 01/01/2000 to 13/02/2017					Regional
107	Isle of Man wildlife records from 01/01/2000 to 13/02/2017					Regional
108	Isle of Wight Notable Species					Regional
117	Mammal records from Britain from the Atlas of Mammals (1993), with some subsequent records					Regional

ID	Final name field	Survey type	Location	Methodology	Systematic / Opportunistic	Spatial Scale
135	National Trust for Scotland Species Records					Regional
143	NBIS Records to December 2016					Regional
144	NE Scotland marine mammal records 1800-2010					Regional
148	Norman and Florence Hammond records. Seawatch and coastal survey records					Regional
153	NRW Regional Data: all taxa (sensitive species only), West Wales					Regional
154	NRW Regional Data: North Wales					Regional
155	NRW Regional Data: South East Wales Non-sensitive species					Regional
170	RECORD Mammal Data					Regional
203	SEWBReC Mammals (South East Wales)					Regional
234	WTSWW Data: All Taxa (West Wales)					Regional
240	Yorkshire Naturalists Union Marine and Coastal Section Records					Regional
138	Natural England Marine Monitoring surveys				Systematic	Regional
28	CEFAS Gillnet retrieval survey 2006/07 - Fisheries Science Partnership	Net	At-sea	Effort	Systematic	Local
30	CEFAS Western Edge ghost nets 2005/06 - Fisheries Science Partnership					Local
53	Deepwater Elasmobranch Species Data From MSS Trawling Surveys (1996 - 2019)					Local
132	Monitoring the Consequences of the Northwestern North Sea Sandeel Fishery Closure	Shore	At-sea	Effort	Systematic	Local
229	WDC Lewis survey data					Local
204	Shetland Oil Spill Sensitivity Maps_Black Guillemot (2016)			Sighting	Opportunistic	Local
233	Wintering Birds Oct - March_oil spill response					Local
90	Harbour porpoises, white-beaked dolphins and minke whales in North Sea - Land surveys		Land-based	Effort	Systematic	Local
40	Copeland Island Bird Observatory SeaWatch Survey			Sighting	Opportunistic	Local

ID	Final name field	Survey type	Location	Methodology	Systematic / Opportunistic	Spatial Scale
124	Marine mammal monitoring from coastal sites in Cardigan Bay, UK, 2004-2009					Local
119	Manx Biological Recording Partnership VERIFIED Isle of Man records between 14/02/2017 and 05/09/2019					Regional
122	Marine Life Information Network (MarLIN) marine survey data (Professional)					Regional
134	National Mammal Atlas Project, online recording					Regional
145	Non-avian taxa (BTO+partners)					Regional
166	present Whale and Dolphin Conservation (WDC) Shorewatch Programme records (2005)					Regional
19	Breeding Coastal Seabirds for oil spill contingency plans					Regional
230	WDC Shorewatch Sightings					Regional
9	Argyll Biological Records Dataset	Sighting	At-sea	Sighting	Opportunistic	Regional
37	Coastwatch Marine Species Records from Northern Ireland		Land-based	Sighting	Opportunistic	Regional
156	Observatoire Pelagis - Reseau National Echouage (French stranding network) strandings 1934-2020	Stranding	Land-based	Sighting	Opportunistic	Regional
220	UK Cetacean Strandings Investigation Programme (CSIP)- strandings around the UK coast (1990-present)					Regional
223	United Kingdom National Whale Stranding Database 1913-2008					Regional
118	Manx Basking Shark Watch 2017-2018	Tagging	Tracking	Effort	Systematic	Regional

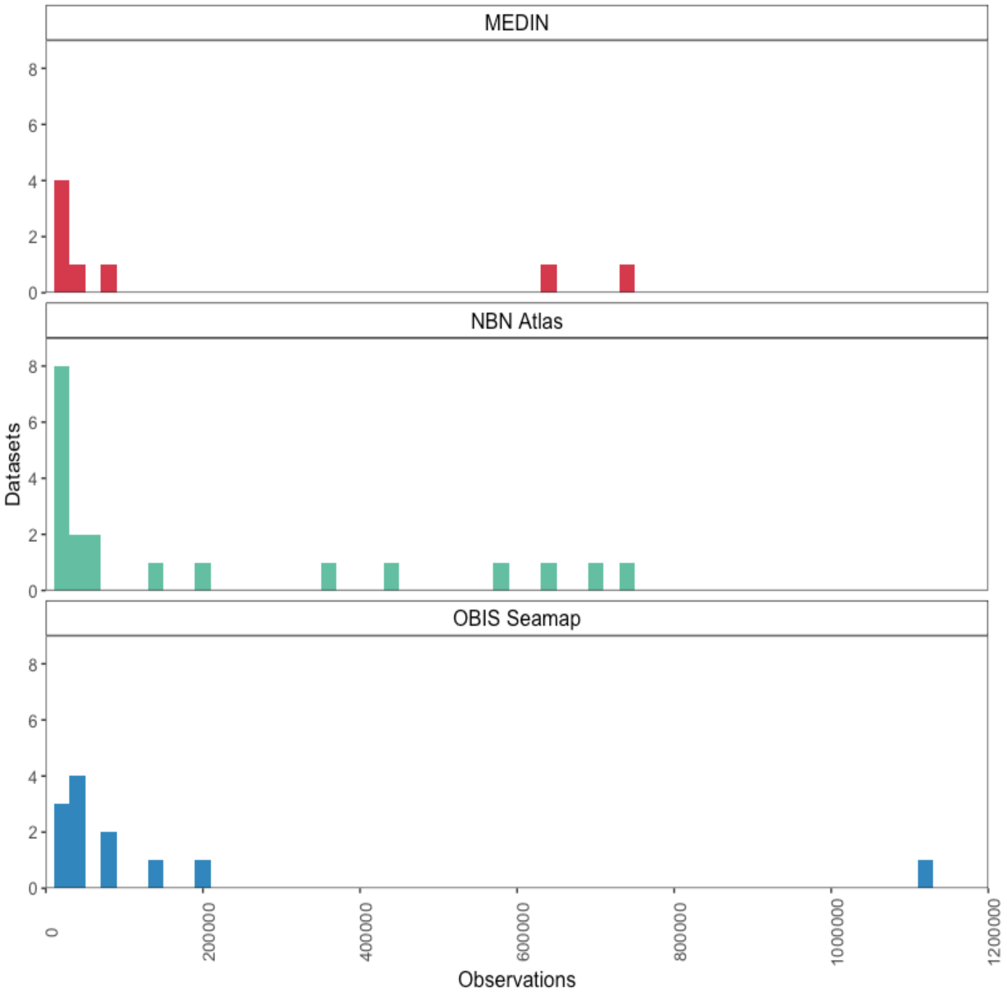


Figure 2.3 Histogram to visualise the distribution in number of observations per dataset.

Taxa coverage

Datasets that were freely available (i.e. those that can be directly downloaded) were then interrogated as to the taxa contained (cetaceans, pinnipeds, seabirds, sharks, turtles), and subsequently the temporal coverage of the datasets that contained multiple taxa. The number of datasets which contained cetaceans, pinnipeds, seabirds, sharks, and turtles is presented in Figure 2.1. Fourteen datasets contained all five taxa, but temporal coverage was varied (Figure 2.5). Some datasets had full temporal coverage from the 1900s to current day (e.g. Argyll Biological Records Dataset, Bristol

Regional Environmental Records Centre (BRERC), North Wales Environmental Information Service (COFNOD) database, Vertebrates of the Outer Hebrides).

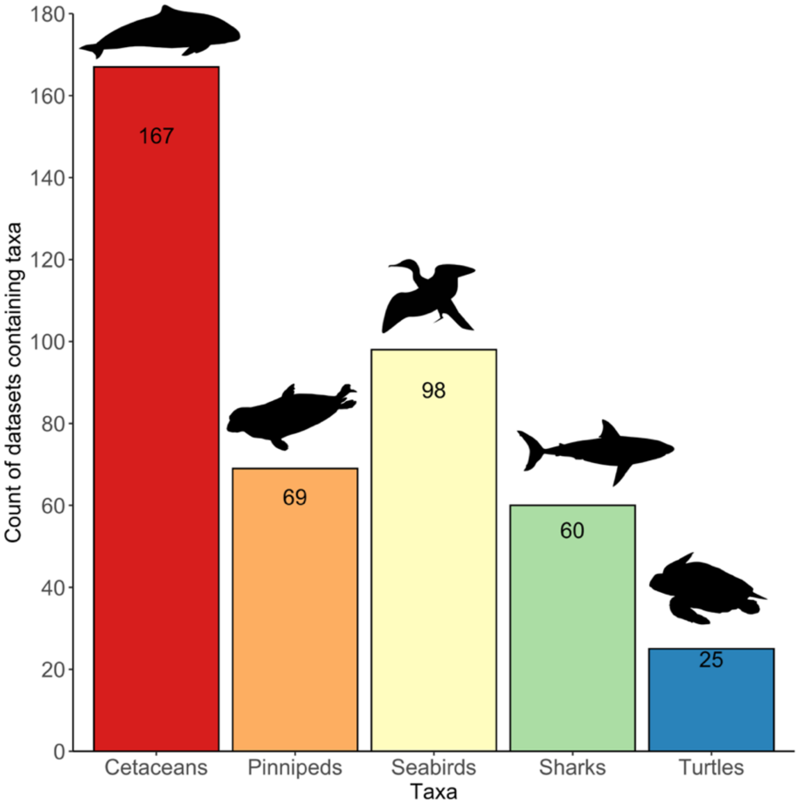


Figure 2.4 Number of datasets which contained records for each taxa category.

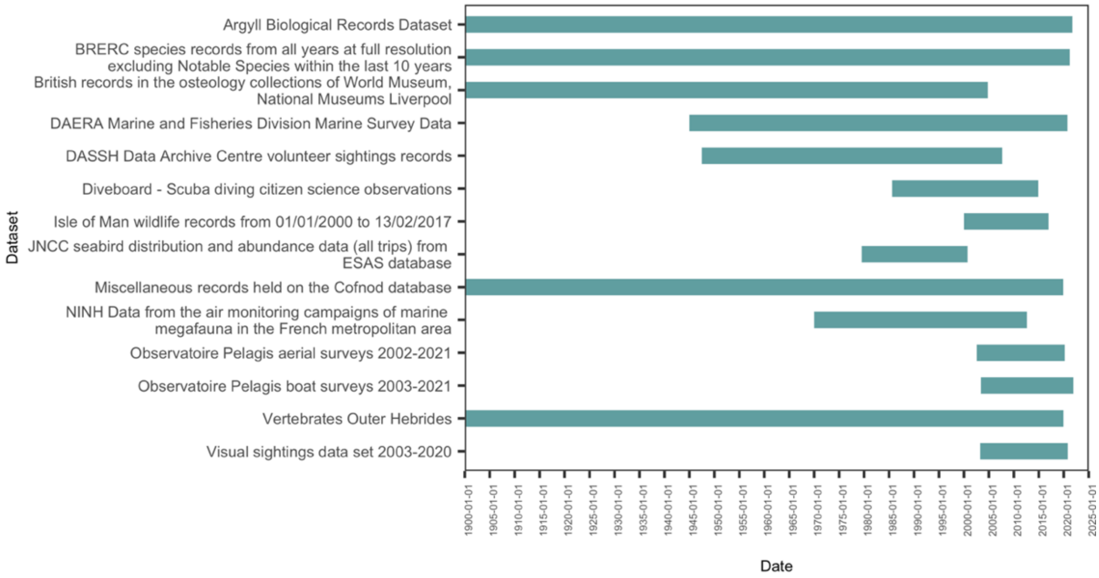


Figure 2.5 Gantt chart of available datasets which contain all five taxa (cetaceans, pinnipeds, seabirds, sharks, turtles) in the records.

Of datasets which had all five taxa, when comparing the number of records within each dataset (Figure 2.6), the JNCC European Seabirds At Sea (ESAS) database and BRERC species records had the maximum observations. However, the BRERC dataset was from NBN Atlas, and not limited to specific marine species, instead the 'count' is from classes Aves, Elasmobranchii, Mammalia, Reptilia and therefore is an overestimate of potential marine species. When results were limited by relevant families for marine predators (Alcidae, Balaenopteridae, Delphinidae, Hydrobatidae, Laridae, Molidae, Phalacrocoracidae, Phocidae, Procellariidae, Scolopacidae, Stercorariidae, Sulidae, Ziphiidae), observations were refined to 30,119, from 738,881 records, highlighting how the NBN Atlas may provide difficulty for the user for selecting specific taxa (e.g. if having to filter by each family).

The ESAS database provided an optimum combination of good temporal coverage, enough taxa and high record count. The OBIS-SEAMAP version of the ESAS dataset contains 1,123,100 records, with 181 species (153 bird species, 18 cetacean species, five pinniped species, four shark and one fish to species level (ocean sunfish *mola mola*)), alongside 17 generalised groups with less accurate non-species level identification (e.g. "Phalaropes, "Shearwaters", "Unidentified gull"). The ESAS dataset was heavily dominated by bird species sightings, likely due to the seabird focus of the dataset from its conception; however, it is still a valuable resource for other species given the lack of long-term data on marine predators. Though data from 1979 to 2000 for the ESAS database was available freely via OBIS-SEAMAP, JNCC manages the data on behalf of the ESAS Co-ordinating Group, and it requires direct contact with the JNCC to get the most up-to-date data available.

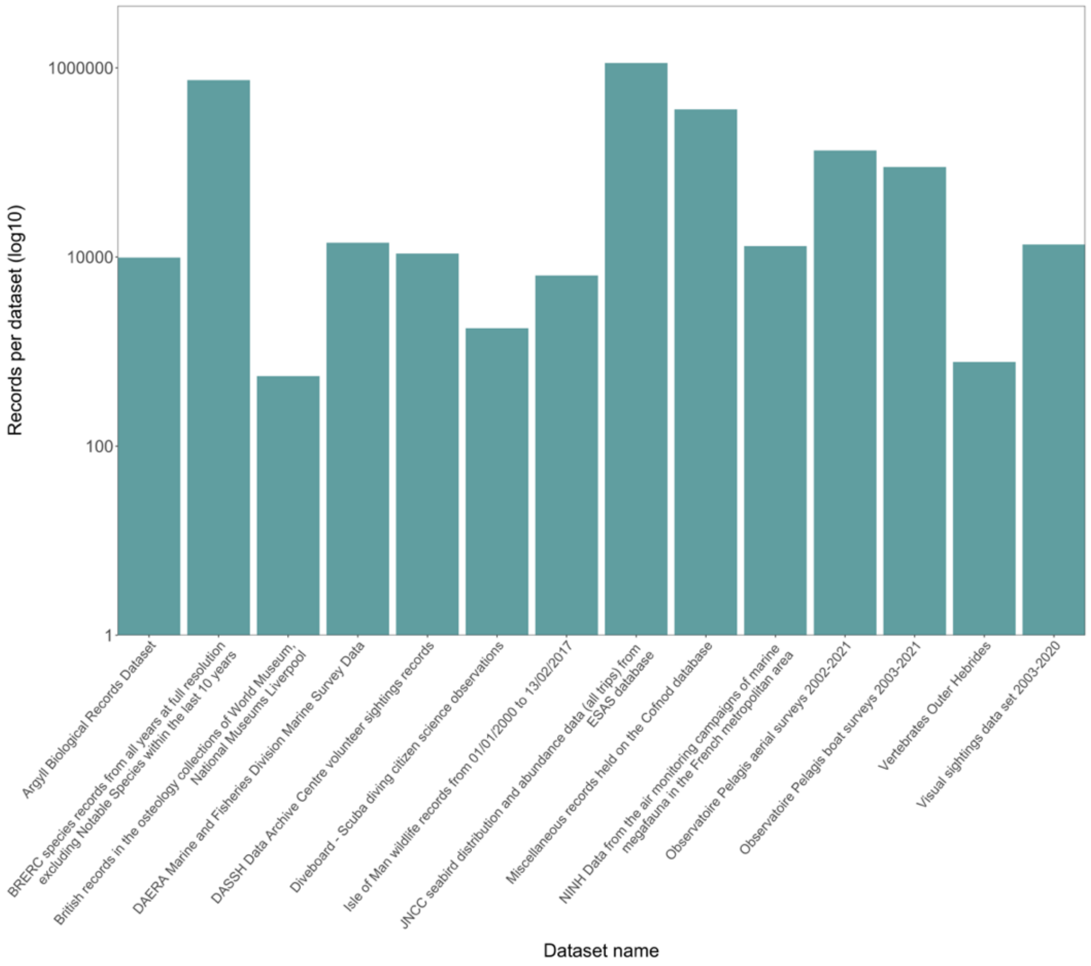


Figure 2.6 Bar chart of available datasets with all five taxa, with the number of records per dataset (on a log scale).

2.4.2 Non-portal searches

Some key organisations held multiple datasets and identifying these primary data holders allowed for further exploration of available data beyond those on the portals. Key data holders included the Sea Mammal Research Unit (SMRU) (responsible for the Small Cetaceans in European Atlantic Waters and the North Sea (SCANS) surveys (Hammond, 2002; Hammond *et al.*, 2013; Hammond, 2021; Gilles *et al.*, 2023), the Institute of Marine Research (IMR), the Marine Biological Association of the UK (MBA) and Observatoire Pelagis (the marine mammals and seabird observatory coordinator of the French stranding network). Datasets were identified outside of portals, via further looking into common data providers and research institutions (identified in Table 2.2). Research institutions held marine predator focused datasets with the majority of datasets were tracking studies that were not freely available (Table 2.3).

Trusts, charities, and other non-governmental organisations also provided data, such as the British Trust for Ornithology (BTO), Worldwide Fund for Nature (WWF), Hebridean Whale and Dolphin Trust, Seawatch Foundation, SeaTrust, The Mammal Society, Whale and Dolphin Conservation (WDC) and Wildlife Conservation Society. Government bodies also aid with data collection and provision with datasets from Marine Scotland, Natural England, Natural Resources Wales, Joint Nature Conservation Committee (JNCC), the Environment Agency and Department for Environment, Food & Rural Affairs (DEFRA), as well as local councils. Citizen science records were also frequent (Table 2.2), and often contained annual datasets e.g. Happywhale, Seasearch and Shorewatch. The search for non-portal data also identified the Organisation Cetacea (ORCA) charity, which has Marine Mammal Survey Teams on board ferries and cruise ships in European waters that conduct monthly scientific surveys of cetacean species. Data were not available to download, but could be requested, and there was broad spatial coverage with over 25 species of cetacean recorded and 18,591 sightings. Utilising the ORCA dataset provides a key valuable resource in increasing coverage of non-bird taxa.

Table 2.3: Datasets from key academic institutions focused on marine predators, with details on method and accessibility.

Institution	Group	Description	Method	Accessible
The Marine Centre (MaREI)	Marine Ecology Group (MEG)	Cetaceans and Seabirds at Sea' (CSS) Project - comprised 3 parallel studies which reported "Seabirds-at-sea" sighting surveys, cetacean sighting surveys; and acoustic surveys for cetaceans (Aguilar de Soto *et al*., 2004, Mackey, 2004, Ó Cadhla *et al*., 2004)	Boat / Acoustic	No
		BIM Seal - Seal depredation study utilising boat hauls at sea that ran in June and July 2012, observing the fish caught in Irish fisheries and any bycatch incidences (seals, cetaceans, elasmobranchs, birds).	Bycatch	No
		Biotelemetry of marine mammals in Irish waters' (BioToMM) project - telemetry technologies were used to study the distribution, movements and behaviour of grey seals. GPS Phase II tags were deployed on eight female seals for approximately 7-8 months (Cronin *et al*., 2011).	Tracking	No
		RAMSSI: Risk assessment for marine mammal and seabird populations in southwest Irish waters	Tracking	No
		GREY Seal telemetry - tracking grey seals off Ireland.	Tracking	No
		GREY Seal survey' - aerial survey techniques to census grey seals during breeding season around southwest and west Ireland.	Aerial	No
		GREY Seal census - grey seal breeding season surveys around southwest and west Ireland.	Land based	
		HARBOUR Seal Census: survey of the harbour seal using thermal imagery during the moult season	Aerial	

		Tracking Turtles: tracking turtles and their jellyfish prey	Tracking	No
Galway Mayo Institute of Technology	Marine Research Sub-Programme 2007-2013	Project exploring the behaviour, distribution, and habitat use of marine mammals and megafauna in Irish Waters, including two deployments of MK10 Archival tags on basking shark (Berrow and O' Connor, 2013).	Tracking	No
University of Exeter	Marine Predator Ecology and Conservation group	Project deploying 70 satellite tags on basking sharks over four years (2012-2015) (Doherty *et al*., 2017a, Doherty *et al*., 2017b).	Tracking	No

2.5 Discussion

For wide-ranging marine predators that often disperse or migrate across oceans (Hays *et al.*, 2013), maximising survey effort in space and time is critical (Waggitt *et al.*, 2019). Carrying out repeat surveys over the same areas with consistent methodologies are high in resource demand, cost and time (Bull *et al.*, 2013), and often limits research. Access to data can often be behind barriers or paywalls and accessibility is a major hurdle (Aubin *et al.*, 2020), even with increasing demand for open data (Culina *et al.*, 2018). Recent studies have shown that retrieving appropriate data for integrating heterogenous datasets is a time-consuming task in daily research practice (Loffler *et al.*, 2021), with datasets differing in format, size, and often scattered across numerous repositories. This study utilised three data portals, focused on specific search terms for specific taxa and returned 247 datasets alone. Ease of use of data portals could be improved, with Kacprzak *et al.* (2019) reporting that 40% of users searching within two open data portals could not find the data required, and ended up requesting the data directly from the repository manager. Data portals that utilised standardised common rules, such as MEDIN and OBIS-SEAMAP, enabled easier discovery and cross-comparison, and various top-down efforts and incentives have been implemented to encourage data sharing and integration (Aubin *et al.*, 2020). Recently within statutory bodies, it has been recognised that the process of accessing, collating and utilising survey data holds value in increasing our understanding of marine predators, for example, JNCC and International Council for the Exploration of the Sea (ICES) have released the JCDP Platform this year (JCDP, 2022). Though in its infancy and therefore the number of datasets available are still low, this JCDP platform aims to develop an international platform to host cetacean survey datasets from the North-East Atlantic. Critically, like OBIS-SEAMAP and MEDIN, the JCDP provides data standard to standardise data contributions. It must be noted however that increasingly these portals are hosting the same datasets (e.g. SCANS-I, SCANS-II, SCANS-III is found in all data portals in this study, and the JCDP Platform), and therefore a measure of unique datasets may be advantageous when comparing across platforms. However, this recent JDCP Platform highlights that this collaborative, standardised approach to storing data is a pressing issue across groups, in the scientific community, public bodies and private consultancy spaces.

Freely available high-quality data on species occurrences are vital in tracking changes in biodiversity (Faith *et al.*, 2013), and the accessibility of global biodiversity data has surged in the past two decades with a big data revolution (Konig *et al.*, 2019, Heberling *et al.*, 2021). Collaborative databases appear to be ever more in demand, for example, data available through GBIF increased 12-fold since 2007, with roughly two publications using GBIF-mediated data per day in 2019 (Heberling *et al.*, 2021). Public and private institutions are increasingly sharing global biological data through data portals, with the aim in aiding with cataloguing, analysis and publishing raw data on species, in an open and free manner. Furthermore, there have been widespread funding initiatives for digitising and logging museum specimens, and growing emergence of citizen-science initiatives, with the latter growing exponentially in recent years (Suškevičs *et al.*, 2021). Whilst the availability of ecological data is still proportionally low compared to other fields (Roche *et al.*, 2015, Aubin *et al.*, 2020), ecologists are increasingly asking large-scale questions that require vast datasets, and there is a consensus on the critical need for more open ecological data (Aubin *et al.*, 2020).

Provision of this data varies in quality, temporal and spatial coverage, and accuracy. Particularly with databases that are mixed methods, it is not clear at a high level the quality control processing that has been carried out for each dataset. Development of an agreed system is needed for effective integration (Saran *et al.*, 2022), and some portals have understood that maximising standardisation for cross-comparison is vital to make data useful and available for users. Many users for example now specify data must compliant with MEDIN Standards, which are also compliant with the UK Gemini and European INSPIRE standards. Actions have been recommended to improve observation databases, such as addressing data gaps and quality and aggregating new data types to support emerging applications (Faith *et al.*, 2013). It is also crucial to promote ease of use for wider applications, particularly to facilitate the needs of current and timely national and international biodiversity initiatives (e.g. Post-2020 Global Biodiversity Framework). In particular, the development of geospatial data portals which have powerful spatial analysis tools (often with R capability) and geocoded information (such as eBird Cornell Lab of Ornithology (2023), Atlas of Living Australia portal (Cornell Lab of Ornithology, 2023)) rather than purely taxonomic information (e.g. FishBase Froese and Pauly (2023)) or more basic occurrence data (e.g. GBIF, OBIS, iNaturalist) is critical for tackling broad-scale ecosystem level questions. However, despite its caveats, online available data represent an underused resource to study large spatial and temporal patterns of marine predators.

2.6 Conclusion

Whilst there are numerous historical datasets of marine predator records in data portals, access to them is often limited or requires direct contact with data providers. Resulting datasets vary in terms of spatial and temporal coverage, but datasets do exist that contain multiple taxa groups at a wide scale and thus could be crucial in biodiversity studies. Data portals could be improved (particularly by providing the number of observations per dataset as standard) but do provide a good entry point to investigating historical data available. Combining heterogeneous data sources can be a vital cost-effective method to aid filling in research gaps, where new studies are too expensive in less easily accessible environments such as the ocean, and this evidence report highlights that there is sufficient data available for examining large spatial and temporal biodiversity patterns.

2.7 References

AGUILAR DE SOTO, N., ROGAN, E., Ó CADHLA, O., GORDON, J. C. D., MACKEY, M. & CONNOLLY, N. 2004. Volume III – Acoustic surveys for cetaceans. *Cetaceans and Seabirds of Ireland's Atlantic Margin. Report on research carried out under the Irish Infrastructure Programme (PIP): Rockall Studies Group (RSG) projects 98/6 and 00/13, Porcupine Studies Group project P00/15 and Offshore Support Group (OSG) project 99/38.*

ANDREWS, R. D., BAIRD, R. W., CALAMBOKIDIS, J., GOERTZ, C. E., GULLAND, F. M., HEIDE-JORGENSEN, M. P., HOOKER, S. K., JOHNSON, M., MATE, B., MITANI, Y. & NOWACEK, D. P. 2019. Best practice guidelines for cetacean tagging. *Journal of Cetacean Research and Management,* 20**,** 27-66.

AUBIN, I., CARDOU, F., BOISVERT-MARSH, L., GARNIER, E., STRUKELJ, M., MUNSON, A. D. & BELLO, F. 2020. Managing data locally to answer questions globally: The role of collaborative science in ecology. *Journal of Vegetation Science,* 31**,** 509-517.

BENSON, D. A., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & WHEELER, D. L. 2005. GenBank. *Nucleic Acids Res,* 33**,** D34-8.

BERROW, S. & O' CONNOR, I. 2013. Marine Mammals and Megafauna in Irish Waters - Behaviour, Distribution and Habitat Use. WP3 Biotelemetry of Marine Megafauna in Irish Waters (Project-based Award). *In:* GROUP, I. W. A. D. (ed.) *NDP Marine Research Sub-Programme 2007-2013.* Irish Whale and Dolphin Group.

BERROW, S. D., O'BRIEN, J., O'CONNOR, I. & MCGRATH, D. A. W., D. 2013. Marine Mammals and Megafauna in Irish Waters - Behaviour, Distribution and Habitat Use. Final Summary Report. *In:* MARINE INSTITUTE (ed.) *Marine Research Sub-Programme (NDP 2007-'13) Series.* Galway: Marine Institute,.

BETTY, E. L., BOLLARD, B., MURPHY, S., OGLE, M., HENDRIKS, H., ORAMS, M. B. & STOCKIN, K. A. 2019. Using emerging hot spot analysis of stranding records to inform conservation management of a data-poor cetacean species. *Biodiversity and Conservation,* 29**,** 643-665.

BLOCK, B. A., JONSEN, I. D., JORGENSEN, S. J., WINSHIP, A. J., SHAFFER, S. A., BOGRAD, S. J., HAZEN, E. L., FOLEY, D. G., BREED, G. A., HARRISON, A. L., GANONG, J. E., SWITHENBANK, A., CASTLETON, M., DEWAR, H., MATE, B. R., SHILLINGER, G. L., SCHAEFER, K. M., BENSON, S. R., WEISE, M. J., HENRY, R. W. & COSTA, D. P. 2011. Tracking apex marine predator movements in a dynamic ocean. *Nature,* 475**,** 86-90.

BUCKLAND, S., ANDERSON, D., BURNHAM, K., LAAKE, J., BORCHERS, D. & THOMAS, L. 2001. *Introduction to distance sampling: estimating abundance of biological populations.,* Oxford, Oxford University Press.

BULL, J. C., MASON, S., WOOD, C. & PRICE, A. R. G. 2013. Benthic marine biodiversity patterns across the United Kingdom and Ireland determined from recreational diver observations: A baseline for possible species range shifts induced by climate change. *Aquatic Ecosystem Health & Management,* 16**,** 20-30.

CARRETTA, J. V., MOORE, J. E. & FORNEY, K. A. 2019. Estimates of marine mammal, sea turtle, and seabird bycatch from the california large-mesh drift gillnet fishery: 1990-2017. *NOAA Techcnial Memorandum NMFS.* National Marine Fisheries Service Southwest Fisheries Science Center: NOAA.

CARTER, M., BOEHME, L., CRONIN, M., DUCK, C., GRECIAN, W. J., HASTIE, G. D., M, J., MATTHIOPOULOS, J., MCCONNELL, B., MILLER, D. L., MORRIS, C. D., MOSS, S. E. W., THOMPSON, D., THOMPSON, P. & RUSSELL, D. J. F. 2022a. Supporting data for "Sympatric seals, satellite tracking and protected areas: habitat-based distribution estimates for conservation and management". *In:* ANDREWS, U. O. S. (ed.). Online: St Andrews.

CARTER, M. I., BOEHME, L., DUCK, C.D., GRECIAN, J., HASTIE, G.D., MCCONNELL, B.J., MILLER, D.L., MORRIS, C., MOSS, S., THOMPSON, D. AND THOMPSON, P. 2020. Habitat-based predictions of at-sea distribution for grey and harbour seals in the British Isles. *Report to BEIS.* University of St Andrews.

CARTER, M. I. D., BOEHME, L., CRONIN, M. A., DUCK, C. D., GRECIAN, W. J., HASTIE, G. D., JESSOPP, M., MATTHIOPOULOS, J., MCCONNELL, B. J., MILLER, D. L., MORRIS, C. D., MOSS, S. E. W., THOMPSON, D., THOMPSON, P. M. & RUSSELL, D. J. F. 2022b. Sympatric Seals, Satellite Tracking and Protected Areas: Habitat-Based Distribution Estimates for Conservation and Management. *Frontiers in Marine Science,* 9.

CHAMBERS, L. E., PATTERSON, T., HOBDAY, A. J., ARNOULD, J. P. Y., TUCK, G. N., WILCOX, C. & DANN, P. 2014. Determining trends and environmental drivers from long-term marine mammal and seabird data: examples from Southern Australia. *Regional Environmental Change,* 15**,** 197-209.

CONVENTION ON BIOLOGICAL DIVERSITY. 2022. *Preparations for the Post-2020 Biodiversity Framework* [Online]. UN Environment Programme. Available: https://www.cbd.int/conferences/post2020 [Accessed 16th February 2022].

CORNELL LAB OF ORNITHOLOGY. 2023. *eBird* [Online]. Ithaca, New York: Cornell Lab of Ornithology. Available: https://ebird.org/home [Accessed].

CRONIN, M. A., JESSOPP, M. & DEL VILLAR, D. 2011. Tracking grey seals on Irelands' continental shelf. Ireland: Coastal & Marine Research Centre, University College Cork.

CULINA, A., BAGLIONI, M., CROWTHER, T. W., VISSER, M. E., WOUTERSEN-WINDHOUWER, S. & MANGHI, P. 2018. Navigating the unfolding open data landscape in ecology and evolution. *Nat Ecol Evol,* 2**,** 420-426.

DOHERTY, P. D., BAXTER, J. M., GELL, F. R., GODLEY, B. J., GRAHAM, R. T., HALL, G., HALL, J., HAWKES, L. A., HENDERSON, S. M., JOHNSON, L., SPEEDIE, C. & WITT, M. J. 2017a. Long-term satellite tracking reveals variable seasonal migration strategies of basking sharks in the north-east Atlantic. *Sci Rep,* 7**,** 42837.

DOHERTY, P. D., BAXTER, J. M., GODLEY, B. J., GRAHAM, R. T., HALL, G., HALL, J., HAWKES, L. A., HENDERSON, S. M., JOHNSON, L., SPEEDIE, C. & WITT, M. J. 2017b. Testing the boundaries: Seasonal residency and inter-annual site fidelity of basking sharks in a proposed Marine Protected Area. *Biological Conservation,* 209**,** 68-75.

DOLMAN, S. J., HODGINS, N. K., MACLEOD, C. D., PIERCE, G. J. & WEIR, C. R. 2013. Harbour porpoises (*Phocoena phocoena*) and minke whales (*Balaenoptera acutorostrata*) observed during land-based surveys in The Minch, north-west Scotland. *Journal of the Marine Biological Association of the United Kingdom,* 94**,** 1185-1194.

DOREMUS, G. 2016. *Observatoire Pelagis boat surveys 2003-2015* [Online]. Available: http://seamap.env.duke.edu/dataset/1403 [Accessed].

DOREMUS, G. 2018. *Observatoire Pelagis boat surveys 2003-2017* [Online]. Observatoire Pelagis UMS 3462, University La Rochelle - CNRS: University La Rochelle. Available: http://seamap.env.duke.edu/dataset/1403 [Accessed 10/1/19 2019].

DOUGHTY, C. E., ROMAN, J., FAURBY, S., WOLF, A., HAQUE, A., BAKKER, E. S., MALHI, Y., DUNNING, J. B., JR. & SVENNING, J. C. 2016. Global nutrient transport in a world of giants. *Proc Natl Acad Sci U S A,* 113**,** 868-73.

DOYLE, T. K., BENNISON, A., JESSOPP, M., HABERLIN, D. & HARMAN, L. A. 2015. A dawn peak in the occurrence of 'knifing behaviour' in blue sharks. *Animal Biotelemetry,* 3.

DRYAD. 2023. *Data Dryad* [Online]. Available: https://datadryad.org/stash [Accessed 6th January 2023].

EUROPEAN COMMISSION & EUROPEAN ENVIRONMENT AGENCY. 2023. *Biodiversity information system for Europe (BISE)* [Online]. Available: https://biodiversity.europa.eu/ [Accessed 21st January 2023].

EVANS, K., THRESHER, R., WARNEKE, R. M., BRADSHAW, C. J., POOK, M., THIELE, D. & HINDELL, M. A. 2005. Periodic variability in cetacean strandings: links to large-scale climate events. *Biol Lett,* 1**,** 147-50.

FAITH, D. P., COLLEN, B., ARIÑO, A. H., KOLEFF, P., GUINOTTE, J., KERR, J. T. & CHAVAN, V. 2013. Bridging biodiversity data gaps: recommendations to meet users' data needs. *Biodiversity Informatics,* 8**,** 41-58.

FROESE, R. & PAULY, D. 2023. FishBase. www.fishbase.org.

GBIF. 2019. *Global Biodiversity Information Facility* [Online]. GBIF. Available: https://www.gbif.org [Accessed 19th June 2019].

GILLES, A., HAMMOND, P., TAYLOR, N., GEELHOED, S. C. V., SVEEGAARD, S., CARLSTRÖM, J., AUTHIER, M., SVEEDRA, C. & ARAÚJO, M. B. 2023. *SCANS-IV survey* [Online]. Available: https://www.tiho-hannover.de/en/clinics-institutes/institutes/institute-for-terrestrial-and-aquatic-wildlife-research-itaw/scans-iv-survey [Accessed 15th February 2023].

GROLEMUND, G. & WICKHAM, H. 2011. Dates and Times Made Easy with lubridate.

HALPERN, B. S., FRAZIER, M., POTAPENKO, J., CASEY, K. S., KOENIG, K., LONGO, C., LOWNDES, J. S., ROCKWOOD, R. C., SELIG, E. R., SELKOE, K. A. & WALBRIDGE, S. 2015. Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nat Commun,* 6**,** 7615.

HALPIN, P. N., READ, A. J., FUJIOKA, E., BEST, B. D., DONNELLY, B., HAZEN, L. J., KOT, C., URIAN, K., LABRECQUE, E., DIMATTEO, A., CLEARY, J., GOOD, C., CROWDER, L.

B. & HYRENBACH, K. D. 2009. OBIS-SEAMAP: The world data center for marine mammal, sea bird, and sea turtle distributions. *Oceanography,* 22**,** 104-115.

HAMMOND, P. S., BERGGREN, P., BENKE, H., BORCHERS, D.L., COLLET, A., HEIDE-JØRGENSEN, M.P., HEIMLICH, S., HIBY, A.R., LEOPOLD, M.F. AND ØIEN, N 2002. Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology,* 39**,** 361-376.

HAMMOND, P. S., LACEY, C., GILLES, A., VIQUERAT, S., BÖRJESSON, P., HERR, H., MACLEOD, K., RIDOUX, V., SANTOS, M., SCHEIDAT, M. AND TEILMANN, J. 2021. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. *In:* SMRU (ed.). St Andrews, Scotland, UK: Sea Mammal Research Unit (SMRU).

HAMMOND, P. S., MACLEOD, K., BERGGREN, P., BORCHERS, D. L., BURT, L., CAÑADAS, A., DESPORTES, G., DONOVAN, G. P., GILLES, A., GILLESPIE, D., GORDON, J., HIBY, L., KUKLIK, I., LEAPER, R., LEHNERT, K., LEOPOLD, M., LOVELL, P., ØIEN, N., PAXTON, C. G. M., RIDOUX, V., ROGAN, E., SAMARRA, F., SCHEIDAT, M., SEQUEIRA, M., SIEBERT, U., SKOV, H., SWIFT, R., TASKER, M. L., TEILMANN, J., VAN CANNEYT, O. & VÁZQUEZ, J. A. 2013. Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation,* 164**,** 107-122.

HAYS, G. C., SCOTT, R. & HIGHAM, T. 2013. Global patterns for upper ceilings on migration distance in sea turtles and comparisons with fish, birds and mammals. *Functional Ecology,* 27**,** 748-756.

HEBERLING, J. M., MILLER, J. T., NOESGAARD, D., WEINGART, S. B. & SCHIGEL, D. 2021. Data integration enables global biodiversity synthesis. *Proc Natl Acad Sci U S A,* 118.

HIBY, L. & LOVELL, P. 1998. Using aircraft in tandem formation to estimate abundance of harbour porpoise. *Biometrics***,** 1280-1289.

HIGHAM, J. E. & LÜCK, M. 2007. *Marine wildlife and tourism management: Insights from the natural and social sciences,* Wallingford, UK, CABI Publishing.

HOBDAY, A. J., ARRIZABALAGA, H., EVANS, K., NICOL, S., YOUNG, J. W. & WENG, K. C. 2015. Impacts of climate change on marine top predators: advances and future challenges. Elsevier.

HOEGH-GULDBERG, O. & BRUNO, J. F. 2010. The impact of climate change on the world's marine ecosystems. *Science,* 328**,** 1523-8.

HUNT, G. L. & MCKINNELL, S. 2006. Interplay between top-down, bottom- up, and wasp-waist control in control in marine ecosystems. *Progress in Oceanography,* 68**,** 115–124.

JCDP. 2022. *Joint Cetacean Data Programme* [Online]. Copenhagen, Denmark. https://cetaceans.ices.dk. [Accessed 26th March 2023].

JONES, E., SMOUT, S., BLIGHT, C., SPARLING, C. & MCCONNELL, B. 2016. Fine-scale harbour seal at-sea usage mapping around Orkney and the North coast of Scotland. *Scottish Marine and Freshwater Science,* 7**,** 1-37.

JONES, J. P., COLLEN, B., ATKINSON, G., BAXTER, P. W., BUBB, P., ILLIAN, J. B., KATZNER, T. E., KEANE, A., LOH, J., MCDONALD-MADDEN, E., NICHOLSON, E., PEREIRA, H. M., POSSINGHAM, H. P., PULLIN, A. S., RODRIGUES, A. S., RUIZ-GUTIERREZ, V., SOMMERVILLE, M. & MILNER-GULLAND, E. J. 2011. The why, what, and how of global biodiversity indicators beyond the 2010 target. *Conserv Biol,* 25**,** 450-7.

JONES, P. G., ASNER, G. P., BUTCHART, S. H. M. & KARANTH, K. U. 2013. The 'why', 'what' and 'how' of monitoring for conservation. *Key Topics in Conservation Biology 2***,** 327-343.

KACPRZAK, E., KOESTEN, L., IBÁÑEZ, L.-D., BLOUNT, T., TENNISON, J. & SIMPERL, E. 2019. Characterising dataset search—An analysis of search logs and data requests. *Journal of Web Semantics,* 55**,** 37-55.

KASCHNER, K., QUICK, N. J., JEWELL, R., WILLIAMS, R. & HARRIS, C. M. 2012. Global coverage of cetacean line-transect surveys: status quo, data gaps and future challenges. *PLoS One,* 7**,** e44075.

KATTGE, J., DÍAZ, S., LAVOREL, S., PRENTICE, I. C., LEADLEY, P., BÖNISCH, G., GARNIER, E., WESTOBY, M., REICH, P. B., WRIGHT, I. J., CORNELISSEN, J. H. C., VIOLLE, C., HARRISON, S. P., VAN BODEGOM, P. M., REICHSTEIN, M., ENQUIST, B. J., SOUDZILOVSKAIA, N. A., ACKERLY, D. D., ANAND, M., ATKIN, O., BAHN, M., BAKER, T. R., BALDOCCHI, D., BEKKER, R., BLANCO, C. C., BLONDER, B., BOND, W. J., BRADSTOCK, R., BUNKER, D. E., CASANOVES, F., CAVENDER-BARES, J., CHAMBERS, J. Q., CHAPIN III, F. S., CHAVE, J., COOMES, D., CORNWELL, W. K.,

CRAINE, J. M., DOBRIN, B. H., DUARTE, L., DURKA, W., ELSER, J., ESSER, G., ESTIARTE, M., FAGAN, W. F., FANG, J., FERNÁNDEZ-MÉNDEZ, F., FIDELIS, A., FINEGAN, B., FLORES, O., FORD, H., FRANK, D., FRESCHET, G. T., FYLLAS, N. M., GALLAGHER, R. V., GREEN, W. A., GUTIERREZ, A. G., HICKLER, T., HIGGINS, S. I., HODGSON, J. G., JALILI, A., JANSEN, S., JOLY, C. A., KERKHOFF, A. J., KIRKUP, D., KITAJIMA, K., KLEYER, M., KLOTZ, S., KNOPS, J. M. H., KRAMER, K., KÜHN, I., KUROKAWA, H., LAUGHLIN, D., LEE, T. D., LEISHMAN, M., LENS, F., LENZ, T., LEWIS, S. L., LLOYD, J., LLUSIÀ, J., LOUAULT, F., MA, S., MAHECHA, M. D., MANNING, P., MASSAD, T., MEDLYN, B. E., MESSIER, J., MOLES, A. T., MÜLLER, S. C., NADROWSKI, K., NAEEM, S., NIINEMETS, Ü., NÖLLERT, S., NÜSKE, A., OGAYA, R., OLEKSYN, J., ONIPCHENKO, V. G., ONODA, Y., ORDOÑEZ, J., OVERBECK, G., OZINGA, W. A., et al. 2011. TRY - a global database of plant traits. *Global Change Biology,* 17**,** 2905-2935.

KONIG, C., WEIGELT, P., SCHRADER, J., TAYLOR, A., KATTGE, J. & KREFT, H. 2019. Biodiversity data integration-the significance of data resolution and domain. *PLoS Biol,* 17**,** e3000183.

LOFFLER, F., WESP, V., KONIG-RIES, B. & KLAN, F. 2021. Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? *PLoS One,* 16**,** e0246099.

MACAULAY, J. D. J., ROJANO-DONATE, L., LADEGAARD, M., TOUGAARD, J., TEILMANN, J., MARQUES, T. A., SIEBERT, U. & MADSEN, P. T. 2023. Implications of porpoise echolocation and dive behaviour on passive acoustic monitoring. J Acoust Soc Am, 154, 1982-1995.

MACKEY, M., Ó CADHLA, O., KELLY, T.C., AGUILAR DE SOTO, N. & CONNOLLY, N 2004. Cetaceans and Seabirds of Ireland's Atlantic Margin.

Volume I – Seabird distribution, density & abundance. *Report on research carried out under the Irish Infrastructure Programme (PIP): Rockall Studies Group (RSG) projects 98/6 and 00/13, Porcupine Studies Group project P00/15 and Offshore Support Group (OSG) project 99/38.*

MANNOCCI, L., ROBERTS, J. J., HALPIN, P. N., AUTHIER, M., BOISSEAU, O., BRADAI, M. N., CANADAS, A., CHICOTE, C., DAVID, L., DI-MEGLIO, N., FORTUNA, C. M., FRANTZIS, A., GAZO, M., GENOV, T., HAMMOND, P. S., HOLCER, D., KASCHNER, K., KEREM, D., LAURIANO, G., LEWIS, T., NOTARBARTOLO DI SCIARA, G., PANIGADA, S., RAGA, J. A., SCHEININ, A., RIDOUX, V., VELLA, A. & VELLA, J. 2018. Assessing cetacean surveys throughout the Mediterranean Sea: a gap analysis in environmental space. *Sci Rep,* 8**,** 3126.

MCCONNELL, B., SMOUT, S. & WU, M. 2017. Modelling Harbour Seal Movements. *Scottish Marine and Freshwater Science,* 8**,** 1-33.

MICROSOFT CORPORATION 2018. Microsoft Excel. 16.70 ed.

MURPHY, S., EVANS, P. G. H., PINN, E. & PIERCE, G. J. 2019. Conservation management of common dolphins: Lessons learned from the North-East Atlantic. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 31**,** 137-166.

NATIONAL BIODIVERSITY NETWORK. 2022. *NBN Atlas* [Online]. Available: http://www.nbnatlas.org [Accessed 14th February 2022].

NATIONAL OCEANOGRAPHY CENTRE (NOC). 2022. *Marine Environmental Data and Information Network (MEDIN)* [Online]. Online. Available: https://medin.org.uk/ [Accessed 14th February 2022].

NATURE 2020. New biodiversity targets cannot afford to fail. *Nature,* 578**,** 338.

NUR, N., JAHNCKE, J., HERZOG, M. P., HOWAR, J., HYRENBACH, K. D., ZAMON, J. E., AINLEY, D. G., WIENS, J. A., MORGAN, K., BALLANCE, L. T. & STRALBERG, D. 2011. Where the wild things are: predicting hotspots of seabird aggregations in the California Current System. *Ecol Appl,* 21**,** 2241-57.

Ó CADHLA, O., MACKEY, M., AGUILAR DE SOTO, N., ROGAN, E. & CONNOLLY, N. 2004. Volume II – Cetacean distribution & abundance. *Cetaceans and Seabirds of Ireland's Atlantic Margin. Report on research carried out under the Irish Infrastructure Programme (PIP): Rockall Studies Group (RSG) projects 98/6 and 00/13, Porcupine Studies Group project P00/15 and Offshore Support Group (OSG) project 99/38.*

O'HARA, C. C., VILLASEÑOR-DERBEZ, J. C., RALPH, G. M. & HALPERN, B. S. 2019. Mapping status and conservation of global at-risk marine biodiversity. *Conservation Letters,* 12.

PEREIRA, H. M., LEADLEY, P. W., PROENÇA, V., ALKEMADE, R., SCHARLEMANN, J. P. W., FERNANDEZ-MANJARRÉS, J. F., ARAÚJO, M. B., BALVANERA, P., BIGGS, R., CHEUNG, W. W. L., CHINI, L., COOPER, H. D., GILMAN, E. L., GUÉNETTE, S., HURTT,

G. C., HUNTINGTON, H. P., MACE, G. M., BERDORFF, T., REVENGA, C., RODRIGUES, P., SCHOLES, R. J., SUMAILA, U. R. & WALPOLE, M. 2010. Scenarios for Global Biodiversity in the 21st Century. *Science*, 1496-1501.

R CORE TEAM 2022. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

ROBERTS, J. J., BEST, B. D., MANNOCCI, L., FUJIOKA, E., HALPIN, P. N., PALKA, D. L., GARRISON, L. P., MULLIN, K. D., COLE, T. V., KHAN, C. B., MCLELLAN, W. A., PABST, D. A. & LOCKHART, G. G. 2016. Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Sci Rep,* 6, 22615.

ROCHE, D. G., KRUUK, L. E., LANFEAR, R. & BINNING, S. A. 2015. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLoS Biol,* 13, e1002295.

SANTORA, J. A. & VEIT, R. R. 2013. Spatio-temporal persistence of top predator hotspots near the Antarctic Peninsula. *Marine Ecology Progress Series,* 487, 287-304.

SARAN, S., CHAUDHARY, S. K., SINGH, P., TIWARI, A. & KUMAR, V. 2022. A comprehensive review on biodiversity information portals. *Biodiversity and Conservation,* 31, 1445-1468.

SIGOURNEY, D. B., DEANGELIS, A., CHOLEWIAK, D. & PALKA, D. 2023. Combining passive acoustic data from a towed hydrophone array with visual line transect data to estimate abundance and availability bias of sperm whales (Physeter macrocephalus). *PeerJ,* 11, e15850.

SURYAN, R. M., SANTORA, J. A. & SYDEMAN, W. J. 2012. New approach for using remotely sensed chlorophyll a to identify seabird hotspots. *Marine Ecology Progress Series,* 451, 213-225.

SUŠKEVIČS, M., RAADOM, T., VANEM, B., KANA, S., ROASTO, R., RUNNEL, V. & KÜLVIK, M. 2021. Challenges and opportunities of engaging biodiversity-related citizen science data in environmental decision-making: Practitioners' perceptions and a database analysis from Estonia. *Journal for Nature Conservation,* 64.

TITTENSOR, D. P., MORA, C., JETZ, W., LOTZE, H. K., RICARD, D., BERGHE, E. V. & WORM, B. 2010. Global patterns and predictors of marine biodiversity across taxa. *Nature,* 466, 1098-101.

VAN CANNEYT, O. 2016. *Observatoire Pelagis aerial surveys 2002-2015.* [Online]. Available: http://seamap.env.duke.edu/dataset/1404 [Accessed].

WAGGITT, J. J., EVANS, P. G. H., ANDRADE, J., BANKS, A. N., BOISSEAU, O., BOLTON, M., BRADBURY, G., BRERETON, T., CAMPHUYSEN, C. J., DURINCK, J., FELCE, T., FIJN, R. C., GARCIA-BARON, I., GARTHE, S., GEELHOED, S. C. V., GILLES, A., GOODALL, M., HAELTERS, J., HAMILTON, S., HARTNY-MILLS, L., HODGINS, N., JAMES, K., JESSOPP, M., KAVANAGH, A. S., LEOPOLD, M., LOHRENGEL, K., LOUZAO, M., MARKONES, N., MARTÍNEZ-CEDEIRA, J., Ó CADHLA, O., PERRY, S. L., PIERCE, G. J., RIDOUX, V., ROBINSON, K. P., SANTOS, M. B., SAAVEDRA, C., SKOV, H., STIENEN, E. W. M., SVEEGAARD, S., THOMPSON, P., VANERMEN, N., WALL, D., WEBB, A., WILSON, J., WANLESS, S., HIDDINK, J. G. & PUNT, A. 2019. Distribution maps of cetacean and seabird populations in the North-East Atlantic. *Journal of Applied Ecology,* 57, 253-269.

WICKHAM, H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

WICKHAM, H. 2019. stringr: Simple, Consistent Wrappers for Common String Operations.

WICKHAM, H., FRANÇOIS, R., HENRY, L. & MÜLLER, K. 2022. dplyr: A Grammar of Data Manipulation.

Chapter 3 How risky? Exploration of biases within two long-term marine predator datasets

3.1 Abstract

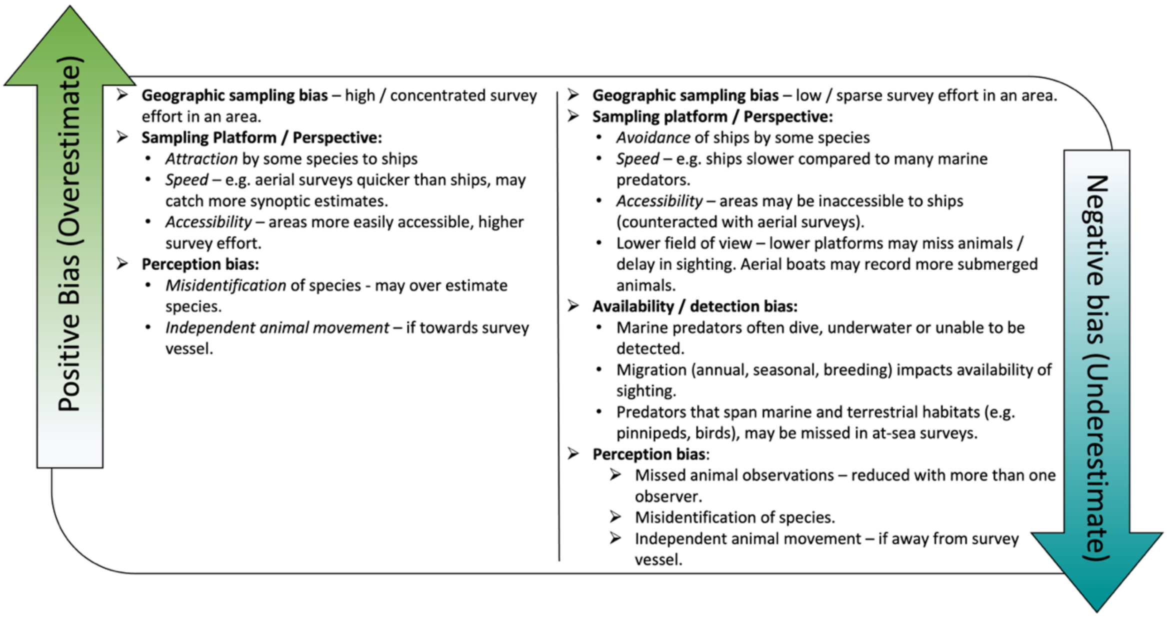
Marine predators such as marine mammals and seabirds are sentinels of marine ecosystems, and knowledge of their distribution and coexistence is critical for marine spatial management and conservation. It is critical to utilise historical datasets to obtain the required temporal and spatial coverage required for studying such wide-ranging mobile species, which individual surveys cannot. Availability of species distribution data is ever-increasing due to rising awareness and concern about the continuing loss of global biodiversity, but it is likely they each include different spatial and temporal biases, and currently a quantitative method of measuring the level of bias across datasets is absent. This study firstly compared the level of data collection effort in terms of temporal, spatial and taxonomic coverage of two major UK datasets (the European Seabirds at Sea (ESAS) and Organisation Cetacea (ORCA) databases) as case studies. Second, inherent biases in data collection were assessed and quantified for four example datasets, using a novel approach that adopts the well-used risk assessment matrix, combining likelihood and severity to give an overall risk score. Whilst scoring is subjective, it provides a valuable tool for helping end-users to assess and present levels of risk in a concise and visual way and cross-compare datasets. The study demonstrated that research question influences how "risky" each dataset is, with lower risk scores for all four datasets when used to answer relative questions about animal presence in an area, rather than using the datasets to look at absolute abundance in an area. The overall risk scores ranked from lowest to highest were SCANS-IV, digital aerial survey (DAS) from APEM, ORCA, and ESAS. The matrix approach has two important roles, the first is to aid end-users in analysing existing datasets, and standardising data submissions to encourage authors to highlight boundaries of their data. The other is to allow researchers to design better surveys by understanding which factors are most important to their study, and thus which biases are high risk to their research question. If the matrix method is used appropriately, it may allow the utilisation of large low-cost historical datasets in meeting ecological and conservation objectives that require wide temporal and spatial coverage, such as understanding space use in cryptic marine predators.

3.2 Introduction

Marine predators such as marine mammals and seabirds inhabit the largest ecosystem on earth and are sentinels of ocean ecosystem state (Tittensor *et al.*, 2010, Sydeman *et al.*, 2012), thus knowledge of their distribution and coexistence is vital for marine spatial management and conservation (Santora *et al.* 2012, Sigler *et al.* 2012). For these wide-ranging highly mobile species which often disperse or migrate across oceans (Hays *et al.*, 2013), maximising survey effort in space and time is critical (Waggitt *et al.*, 2019). Repeat surveys over the same areas with consistent methodologies are high in resource demand, cost and time (Bull *et al.*, 2013), so often limits research. Combining existing long-term datasets from multiple sources can give the wide temporal and spatial coverage required for marine predators, which individual fine-scale studies cannot, and overcome the technical challenges of marine species data collection (Pacifici *et al.*, 2016). With biological responses to climate change increasingly used for legislation, increased efficient and cost-effective use of existing historical datasets could be welcome (Leonard *et al.*, 2006). Particularly with the time delays in data processing and even more so in analysis, significant shifts in species space use resulting in changes in species richness may be, or already have been, missed. The availability of distribution data from online data portals and databases is ever-increasing due to increasing awareness and concern about the continuing loss of global biodiversity, but it is likely they each include different spatial and temporal biases (Boakes *et al.*, 2010), diverse data standards and varying levels of detail. Therefore, combining heterogeneous datasets must be done with consideration of data uncertainty.

3.2.1 Sources of bias

Bias is a form of uncertainty. Uncertainty can be non-directional, thus making researchers less confident in results, or it can be directional, known as biases. Bias is introduced when the animals sampled do not reflect the characteristics of the total population of interest (Stuber *et al.*, 2013), and be introduced into datasets in a variety of ways that may overestimate (positive bias) or underestimate parameters (negative bias) (summarised in Figure 3.1). Bias may not impede studies, but may alter what question can be asked, therefore quantifying the level of impact bias has on hypotheses is important. This may be an issue for conservation or management implications, for example if prioritisation of areas based upon measures such as abundance or species richness are misdirected. It is impossible to eliminate all forms of bias, but proper study design can help to mitigate or minimise bias (Sica, 2006).



- **Geographic sampling bias** – high / concentrated survey effort in an area.
- **Sampling Platform / Perspective:**
  - *Attraction* by some species to ships
  - *Speed* – e.g. aerial surveys quicker than ships, may catch more synoptic estimates.
  - *Accessibility* – areas more easily accessible, higher survey effort.
- **Perception bias:**
  - *Misidentification* of species - may over estimate species.
  - *Independent animal movement* – if towards survey vessel.

- **Geographic sampling bias** – low / sparse survey effort in an area.
- **Sampling platform / Perspective:**
  - *Avoidance* of ships by some species
  - *Speed* – e.g. ships slower compared to many marine predators.
  - *Accessibility* – areas may be inaccessible to ships (counteracted with aerial surveys).
  - Lower field of view – lower platforms may miss animals / delay in sighting. Aerial boats may record more submerged animals.
- **Availability / detection bias:**
  - Marine predators often dive, underwater or unable to be detected.
  - Migration (annual, seasonal, breeding) impacts availability of sighting.
  - Predators that span marine and terrestrial habitats (e.g. pinnipeds, birds), may be missed in at-sea surveys.
- **Perception bias**:
  - Missed animal observations – reduced with more than one observer.
  - Misidentification of species.
  - Independent animal movement – if away from survey vessel.

Figure 3.1 Directionality of biases within datasets.

Geographic / spatial sampling bias

Spatial or geographic sampling bias is widespread in occurrence datasets (Hijmans *et al.*, 2000, Reddy and Dávalos, 2003, Kadmon *et al.*, 2004, Aiello-Lammens *et al.*, 2015). The geographic location and intensity of collecting data can be influenced by the accessibility to the area (Reddy and Dávalos, 2003), particularly for boat based surveys, which affects animal distribution data (Davis *et al.*, 1990, Rondinini *et al.*, 2006, McBride-Kebert *et al.*, 2019). In marine studies higher effort may be skewed towards the coastline or in shipping routes or fishing areas which are repeatedly visited. In areas of high sampling intensity, it may lead to overestimates of measures such as species richness or abundance whilst low sampling intensity may lead to an underestimate. Thus if making comparisons between areas, the relationship between sample size and observed taxonomic measures must be assumed to hold for all areas (Williams *et al.*, 2002). If survey effort is recorded, then spatial bias may be more easily factored in. Furthermore, if sampling bias can be identified, it can identify target geographical areas for further data collection. Statistical tests to reflect true distribution rather than collection effort are needed (Reddy and Dávalos, 2003), with some existing such as rarefaction to homogenise sampling in original locality data, range modelling or extrapolation to poorly sampled areas (Colwell and Coddington, 1994, Duckworth, 1997, Williams *et al.*, 2002).

Sampling perspective / platform

The method of collection is important when integrating data from different sampling perspectives (Phillips *et al.*, 2019). Ships are slow in transit relative to many marine predators (Wahl and Heinemann, 1979, van Franeker, 1994), and animals may exhibit avoidance (e.g. porpoises avoiding marine traffic (Palka and Hammond, 2001, Díaz López and Methion, 2018)) or attraction behaviours (e.g. Hector's dolphins *Cephalorhynchus hectori* (Martinez *et al.*, 2010) or common dolphins *Delphinus delphis* (Cañadas *et al.*, 2009, Hammond *et al.*, 2013)). Aerial surveys move quickly compared to seabirds or marine predators so can provide more synoptic estimate of distribution or abundance (Briggs *et al.*, 1985a, Buckland *et al.*, 2001, Certain and Bretagnolle, 2008), but may lead to short-term behavioural disturbances (Hughes *et al.*, 2008). Different survey perspectives can indicate comparable high use areas, but fine-scale patterns depend on survey perspective and spatiotemporal extent (Phillips *et al.*, 2019). Sampling platform is interrelated with geographic sampling bias; planes or helicopters can reach survey areas that are inaccessible to ships, but then may not be able to reach areas far offshore (Henkel *et al.*, 2007, Hodgson *et al.*, 2016).

Lower survey platforms may have a lesser field of view for marine mammals, so have a delay in spotting animals until close or miss animals with avoidance behaviours, but higher platforms (such as ship bridges on ferries, or aerial surveys) may have an extended field of view and see marine mammals quicker or as they are swimming away (Williams *et al.*, 2017). This varies by species life history, for example high flying birds may be missed due to lower platform surveys focused on surface birds (Bemmelen *et al.*, 2011). Survey perspective and platform should be influenced by study objectives, accessibility of the area, sampling logistics, and available resources (Ainley *et al.*, 2012).

Availability bias

Availability bias (or detection probability) is particularly prevalent in marine vertebrates who are often underwater or diving so not visible to observers (Marsh and Sinclair, 1989). Availability for marine mammals is largely a function of the animal's depth-use patterns and water turbidity (Thomson *et al.*, 2013); an animal must be at a depth to be visible and detectable, it is the probability of being 'available'. Availability bias leads to an underestimate in abundance estimation (Williams *et al.*, 2017). Variations by species in terms of dive and surface times can also lead to heterogenous availability patterns so knowledge of dive-surfacing patterns is important for analyses and applications of sightings data (Thomson *et al.*, 2013).

Availability bias is complex, as this can be interlinked with survey perspective. Sampling marine predators from boat or aerial survey is already biased towards visible surface activities. Boat observers cannot see deep into the water thus only surfacing animals are recorded, but aerial surveys may record some more submerged animals. For example, sooty tern *Onychoprion fuscatus* and white tern *Gygis alba* were more prevalent in mounted binocular surveys rather than the strip transect method due to easier detection due to flight behaviour (high above the water unless actively feeding) and ship avoidance (Borberg *et al.*, 2005a), thus would be negatively biased if strip transects are used.

Temporary migration (annual, seasonal, breeding) of animals during different periods affects availability bias, as they may not be in the area surveyed temporarily (Valente *et al.*, 2017, Yamaura and Royle, 2017). Migrations may lead to a mismatch between high sampling survey seasons and lower winter months. There is often a clear increase in data collection in summer months, likely due to the better sea state for surveying or more successful sightings of animals in these conditions. Sightings have been higher in better sea state conditions (Beaufort 0 – 3), with Beaufort 4 or above reducing the ability to detect marine mammals (Jefferson and Lynn, 1994, Harwood *et al.*, 2004). Similarly, animals that span the boundary between marine and terrestrial habitats may influence detection bias. Pinnipeds regularly haul out at terrestrial or ice locations, and whilst this allows for population monitoring when they come ashore for resting, pupping, moulting, and to avoid predators (Boyd *et al.*, 2010), it gives little information on at-sea counts (Whitlock *et al.*, 2020). Conversely, at-sea counts may be underrepresenting hauling out species. It is rare that count data is analysed in combination with in-water sighting or remote tracking data (Herr *et al.*, 2009, Himes Boor and Small, 2012, Vincent *et al.*, 2017).

Furthermore, animals are freely moving with their own behaviours that vary both inter-specifically and intra-specifically. Responsive movement towards platforms may affect availability bias – via either attraction or displacement. For example with seabirds that follow fishing vessels (Briggs *et al.*, 1985b, Pollock *et al.*, 2000, Clarke *et al.*, 2003, Votier *et al.*, 2013), a density overestimation in counting birds may result (Hyrenbach, 2001, Spear *et al.*, 2004), but with species that avoid ships (Pollock *et al.*, 2000, Clarke *et al.*, 2003) an underestimation may be made. The level of displacement varies within taxa, species and even individuals.

Perception bias / misclassification bias

Perception bias is principally when observers did not notice animals present (Buckland *et al.*, 2001, Hobbs and Waite, 2010), and differs with factors such as weather conditions, sea state, observer experience (Thomson *et al.*, 2013). Observers may not detect animals even though they are present in the area, particularly elusive ones. Harbour porpoises *Phocoena phocoena* for example are difficult to detect in sea states greater than Beaufort Force 2, so relative abundances may be underestimated (Hammond *et al.*, 1995), and have been deemed hard to detect with detection curves, as they stay mostly underwater (Bemmelen *et al.*, 2011). For measures such as abundance, estimation depends on the estimation of detection probability under a particular sampling method (Pollock *et al.*, 2004, Southwell *et al.*, 2013).

Detection probability of marine predators can be affected by other factors such as size, colour, and behaviour of species (Tasker *et al.*, 1984, Van der Meer and Camphuysen, 1996). For example, smaller alcids such as puffins and auks have been reported harder to detect on the water as are often by themselves and are dark backed, so harder to see in poorer weather conditions and at greater distances. There is also likely misidentification of species, particularly in similar looking species – for example discrimination between common guillemot *Uria aalge* and razorbill *Alca torda* is often difficult from aerial surveys at considerable altitude (Buckland *et al.*, 2012), and species-level identification of phalaropes, storm petrels *Hydrobates pelagicus*, small jaegers, small alcids, and small terns is deemed tricky (Spear *et al.*, 2004). In these cases, species may be incorrectly recorded or not recorded to species level detail. Species detectability can be an issue, with smaller, rare or cryptic species less represented in sighting datasets (Barbraud and Thiebot, 2009, Monk, 2014).

Detection is rarely perfect, due to factors such as observer error (Nichols *et al.*, 2000), species rarity, survey method or variation with confounding variables (e.g. habitat, environmental conditions) (Gu and Swihart, 2004), and almost all species may be present in a study area, but simply not detected – known as imperfect detection (Monk, 2014). Imperfect detection is common when the population is small, individuals are hard to sample, or effort is limited.

Perception bias may be reduced by the presence of two or more observers at a time. Spear *et al.* (2004) determined during at-sea surveys a single observer missed approximately 26% of the birds present, whereas two or more observers on watch simultaneously are required to obtain a detection rate of 95%.

Independent animal movement

Independent movement of animals, not responsive movement or behaviour to observers, may generate biases in abundance estimates in particular (Glennie *et al.*, 2015, Lennert-Cody *et al.*, 2018).

In distance sampling methods, a major assumption is that all animals are detected at their initial location, however, these sampling methods are a ‘snapshot method’ (Buckland *et al.*, 2001) and in reality, animals are unlikely to be fixed in space and time during the survey. Animals may respond to vessels but also fundamentally move independently of the observer; there is a non-linear relationship between animal movement, detection, and bias. This bias varies depending on speed of the animal and the platform type. If animals travel faster than the observer, significant bias is caused (Buckland *et al.*, 2001). Animal speed is often slow relative to aircrafts (Marsden, 1991), thus minimal bias exists, however with ship surveys the animal may be travelling faster than the vessel. This varies by method; simulations using line transects on abundance estimates of whales showed that bias was reduced when animal speed was less than observer speed (negligible if mean animal speed is one quarter of that of the observer), and bias was smaller in line transect sampling over strip transect sampling (Glennie *et al.*, 2015). Bias was deemed more significant if moving at half of observer speed, and when mean animal speed exceeded observer speed, bias becomes higher and may exceed bias of strip transect sampling. As such, for fast-moving animals such as seabirds or small cetaceans, methods for correcting for movement may be required if estimating parameters such as abundances or densities.

Some suggestions include excluding animals entering or exiting strip transects or excluding animals that overtake observers (Granholm, 1983). However, post survey completion there is no clear analytical way currently to reduce bias in abundance estimates using the standard methods (Glennie *et al.*, 2015), so using historical datasets that have not factored in these movement corrections from the onset may be subject to bias. This independent animal movement exists irrespective, but highlights the importance of proper study design, quantification of bias within datasets and crucially when working with historical datasets; choosing the correct research questions to suit the data available.

3.2.2 Risk assessment matrix approach

Spatial patterns of marine predators are inextricably linked to collection methods, and no synthesis of bias patterns or their consequences exists currently (Hughes *et al.*, 2021). Whilst the various biases in datasets are discussed in section 3.2.1, a measurable method of quantifying the level of bias within a dataset is currently lacking and this study aims to address this gap.

A common tool used within industry and management (Duijm, 2015), a risk assessment matrix evaluates particular risks in terms of the likelihood of risk occurrence, and the severity of the risk (Cox, 2008, Smith *et al.*, 2009) and outputs a risk value (often likelihood score * severity score). Versions of risk matrices have been used in the biology field to evaluate for example the assessments of extinction risk (Wallace *et al.*, 2011) or climate future models (Clarke *et al.*, 2011), but at present there appears to be no studies using a matrix approach for assessing the level of biases in historical heterogenous datasets. Inherent biases in datasets will lead to consequences for use in further analysis but may vary depending on the desired use of the dataset.

3.2.3 Aims

The aims and hypotheses of this chapter are:

1. Compare quantitatively the temporal, spatial and taxonomic coverage of two example major UK databases of marine predator distribution, to understand collection effort.

2. Review the biases inherent in these two datasets (alongside two additional pilot datasets), with the hypotheses that:
    a. Organisation Cetacea (ORCA) datasets will have higher geographic sampling bias than the European Seabirds at Sea (ESAS) due to its utilisation of opportunistic ferry routes.
    b. Both datasets will be impacted equally by sampling perspective given their heterogenous data sources, availability bias given the focus on at-sea data.
    c. Perception bias and human error will be present equally in both datasets but minimised by training and quality control procedures.
    d. Independent animal movement exists as an inherent bias for both datasets.
3. Test an example bias matrix in terms of risk scores, using four pilot datasets for two research questions: the first using the data to obtain counts/abundances of species in an area versus the second using the data to identify species presence in an area. It is hypothesised that overall risk is lower (with fewer high risk scores) when using historical datasets for determining species presence/absence questions rather species abundances.
4. Discuss how the bias matrix approach can be used to quantify biases and guide the use of large-scale observation datasets, both pre-survey and post-survey.

3.3 Method

3.3.1 Pilot datasets used in this study

I prepared an evidence report (see Chapter 2) to summarise the status and knowledge of datasets of marine predators around the UK, reviewing data and identifying gaps or further data required. Predators are wide ranging and have low detectability due to their lifestyle therefore require a large amount of survey effort to have sufficient data (Kaschner *et al.*, 2012). Exploiting surveys and databases to date may aid in research and conservation planning and assist in planning future survey efforts to fill gaps in existing data. Given fine-scale tagging and tracking studies do not often collect information from more than one or two colonies or species at a time, large scale studies of multiple species may take a large amount of effort and resources (Block *et al.*, 2011, Grecian *et al.*, 2016), with significant financial burden if collecting new data. The evidence report summarised that suitable historic data were available via data portals, though varied extensively in spatial, temporal, and taxonomic coverage.

The European Seabirds at Sea (ESAS) and Organisation Cetacea (ORCA) datasets were identified in Chapter 2 as having suitable broad spatial, temporal and taxonomic coverage for use in further biodiversity studies, and therefore were carried forward as two of the pilot datasets (from Chapter 2) for this risk matrix approach. Detail on these two datasets is given in section 3.4.2 as they form the basis for subsequent chapters (Chapter 4).

However, given data access is not required for pilot matrix assessments, two additional datasets are included for comparison of the breath of the matrix approach in section 3.4.4. The two datasets are SCANS IV aerial surveys[1] (Gilles *et al.,* 2022) which focuses on cetaceans only, and Digital Aerial Surveys (DAS) from APEM Ltd (APEM Ltd, 2023) which focuses on marine megafauna and birds. ESAS and ORCA data

Occurrence records for species of marine predator were obtained using aerial and vessel survey data from ESAS database (Version 5) (JNCC, 2019) and cruise distance sampling and wildlife officer sightings datasets from ORCA (ORCA, 2019). The most up to date versions of both datasets were requested (in March 2020) from the Seabirds and Cetaceans Information Manager (Mark Lewis) at Joint Nature Conservation Committee (JNCC) and Science Officer (James Robbins) at ORCA. For research collaboration ORCA required a data request form process, and I requested any data for the area of interest (AOI) region for all species of cetacean available, including both distance sampling and survey effort datasets. Comparison of survey methodologies are summarised in section 3.4.2 (with additional detail in Appendix C).

ESAS data was extracted from the Microsoft Access database into Excel spreadsheets per species, with only species deemed to be marine predators selected (Appendix B). All ORCA data was in Excel spreadsheet format upon delivery. Datasets were then imported into R (R Core Team, 2022) and graphical analysis was carried out using

---

[1] Whilst SCANS IV surveys use aerial surveys for most blocks, two blocks are surveyed by boat. Only aerial surveys are considered in this chapter.

ggplot2 (Wickham, 2016). Maps were produced in ArcGIS Pro (v.3.0.3) (ESRI, 2022), with density carried out using the density geoprocessing tool.

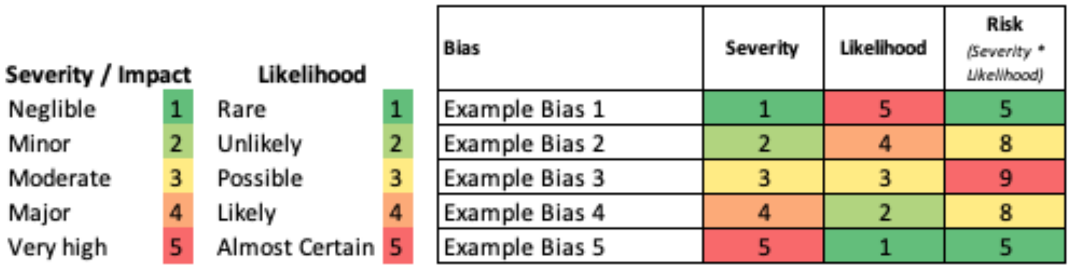
3.3.2 Biases within datasets

General survey approach with advantages and limitations and sources of potential bias were defined. Biases that may exist within the two key datasets (ESAS and ORCA) were identified in more detail (see section 3.2.1), and then a summary given as to the source of each bias and how it exists within the two datasets. I identified the biases that may exist within the datasets, and provided some explanation given as to the sources of bias and how it exists within the two datasets. Is it critical to choose the correct research question to suit available historical data? The biases identified will vary depending on hypotheses and questions asked of the data source.

3.3.3 Risk matrix

I developed a novel way of approaching the risks associated with uncertainty and bias in biological monitoring datasets, creating a unique risk assessment matrix approach.

I created a template risk assessment matrix for the use in quantifying biases (Figure 3.2). The '*severity* or *impact'* score is expressed as the severity each bias may have on the desired use of the dataset – for example, the severity of biases in datasets may be different depending on whether actual counts are required or just species presence. The '*likelihood'* score is the probability of the bias occurring, due to data collection methods and nature of the dataset. The risk score is the severity score multiplied by the likelihood score, to give a '*risk value'* that quantifies the bias for the specific dataset and research aim. These scores are subjective by the user, but allows comparison across datasets, methods, and research questions.



Severity / Impact		Likelihood	
Neglible	1	Rare	1
Minor	2	Unlikely	2
Moderate	3	Possible	3
Major	4	Likely	4
Very high	5	Almost Certain	5

Bias	Severity	Likelihood	Risk *(Severity * Likelihood)*
Example Bias 1	1	5	5
Example Bias 2	2	4	8
Example Bias 3	3	3	9
Example Bias 4	4	2	8
Example Bias 5	5	1	5

Figure 3.2 Template risk assessment matrix for assessing biases in datasets. The 'severity or impact' score is expressed as the severity each bias may have on the desired use of the dataset. The 'likelihood' score is the probability of the bias occurring, due to data collection methods and nature of the dataset.

The risk matrix was used to estimate the risk of using the two publicly available sightings databases (ESAS and ORCA) (identified in Chapter 2), alongside SCANS-IV (aerial surveys only, not vessel blocks) and a DAS survey (APEM Ltd, 2023) for two types of research question:

1. How risky is it using the dataset to obtain counts/abundances of species in an area?
2. How risky is it using the data to identify species presence in an area?

The first matrix was used to quantify the impact of biases on determining **abundances** of species, for the four datasets. So, in this instance, the '*severity* or *impact'* score is expressed as the severity each bias may have on the ability to get precise and accurate **counts** of species abundances. The '*likelihood'* score is the probability of the bias occurring due to data collection methods and nature of the dataset. The 'risk score' will be the risk of using the dataset for the aim of getting abundances or counts of species.

The second matrix (for the same four datasets) uses the matrix with the aim of answering **relative** questions, in this case **presence** of species in an area at any one time. As such, in this instance, the '*severity* or *impact'* score is expressed as the severity each bias may have on the ability to get precise and accurate **presence confirmation or accounts** of species sightings (not counts, or abundance measures) – i.e., how does the bias affect the severity of the sighting of a species being true. The '*likelihood'* score is the probability of the bias occurring due to data collection methods and nature of the dataset.

A pilot example of this risk assessment approach was sent out to various organisations that utilise historical datasets for feedback on approach, and whilst does not from part of this chapter, is presented in Appendix C.

3.4 Results

3.4.1 General approaches

In surveys of marine predator species, various methodologies are employed, each with unique advantages and limitations. Table 3.1 summarises general approaches to surveys (line-transect versus strip-transect surveys versus ESAS methodology, observational versus photographic/video approaches, vessel versus airplane platforms, systematic versus opportunistic strategies, and single-platform versus double-platform approaches) and provides advantages and limitations per method. The ESAS dataset follows the ESAS methodology, whilst ORCA, SCANS IV (aerial surveys) and the DAS APEM survey (APEM Ltd, 2023) use a line transect method.

Table 3.1 General survey approach with advantages and limitations, sources of potential bias.

Category	Subcategory	Description	Advantages	Limitations	Reference
Method	**Line transect**	Survey along predetermined paths (transects) and record sightings of animals at a known perpendicular distance from the transect line.	Standardised survey method, allows for comparison over time / between study areas *(Geographical/spatial bias)*	Assumes objects on track line are certain to be detected at original location (*Sampling perspective)* Resource intensive *(Availability bias)* Limited spatial coverage – fixed paths *(Geographical/spatial bias)* Resource intensive *(Geographical/spatial bias)* Dependent on environmental conditions such as visibility (*Perception bias)*	Buckland *et al.* (2001) Thomas *et al.* (2010)
	Strip transect	Survey along a survey line recording animals within parallel strips adjacent to the survey line, e.g. using defined widths on either side of the track (effective strip widths (ESWs).	Standardised survey method allows for comparison over time / between study areas. *(Geographical/spatial bias)* Wide coverage compared to line transect – covers a larger area. (*Sampling perspective)*	Reduced precision compared to line *(Geographical/spatial bias)* Dependent on environmental conditions such as visibility – detectability drops off rapidly with increasing distance from the line (*Perception bias)*	Glennie *et al.* (2015)

Category	Subcategory	Description	Advantages	Limitations	Reference
	ESAS method	Distance band method – 90-degree angle of view to be selected from bow to beam, then a 300-metre-wide strip transect of birds in the water is recorded. Birds in flight within the 90° area are recorded. Range/bearing recorded for non-bird marine species. A 180° or 90° forward scan operated at the same time is sometimes present.	Standardised survey method allows for comparison over time / between study areas. *(Geographical/spatial bias)* Wider field of view (*Sampling perspective)*	Dependent on environmental conditions such as visibility (*Perception bias)*	JNCC (2020)
Recording method	**Observational sightings**	Visually detecting and recording animals by human eye (with binoculars) without using recording equipment	Documentation of real-time data (*Availability bias)* Quick deployment (*Sampling perspective)* Low cost *(Geographical/spatial bias)*	Limited time for documentation (*Availability bias)* Subject to observer bias / subjective species identification (*Perception bias)* No permanent record for future reference (*Perception bias)*	Torres *et al* (2018) Bröker *et al.* (2019)
	Photographic / video recording	Capture footage of animals using camera or video equipment.	Detailed documentation for accurate ID. *(Perception bias)* Permanent record for future reference. *(Perception bias)* Reduced observer bias, can allow for second opinion *(Perception bias).*	High cost *(Geographical/spatial bias)* Limited field of view – equipment dependent, but lack of periphery or ability to move field of view. (*Sampling perspective)* Post-processing time required (*Perception bias)*	

Category	Subcategory	Description	Advantages	Limitations	Reference
Platform	**Vessel**	Using boats/ ships as platforms for observation.	Allows close proximity to ocean surface *(Survey perspective).* Stable platform *(Survey perspective).* Flexibility in navigating environments *(Geographical/spatial bias)*	Limited field of view (horizontal perspective) – distance / underwater challenging to detect *(Survey perspective)*. Potential for extended disturbance of marine species due to vessel noise – avoidance behaviour *(Availability Bias)*. Costs for maintenance, fuel, trained crew (*Geographical/spatial bias)* Potential for attraction behaviours – e.g. fishing vessels for seabirds *(Availability Bias)*.	Torres *et al (2018)* Fettermen *et al.* (2022)
	Aerial	Using aircraft (planes / helicopter) as platforms for observation.	Large coverage area *(Geographical/spatial bias)*. Can record submerged animals (more than vessels) *(Survey perspective).* Elevated position for visibility (vertical perspective) *(Survey perspective).* Can reach areas that may be inaccessible to ships *(Geographical/spatial bias)*. Less long-term disturbance for marine species than vessel noise *(Availability Bias)*	Limited detail – lack close range observations *(Survey perspective)*. Highly weather dependent *(Availability Bias)*. Costs for maintenance, fuel, trained crew (*Geographical/spatial bias)* May be limited in extent can travel offshore *(Geographical/spatial bias)*. Potential for short-term disturbance – avoidance, due to noise *(Availability Bias)*.	

Category	Subcategory	Description	Advantages	Limitations	Reference
			Move quickly compared to animals, synoptic estimate *(Survey perspective).* No attraction behaviours from animals (*Availability Bias)*.		
Approach	**Systematic**	Surveys following a structured and predetermined methodology – covering predefined transects or grid cells.	Standardised survey method, allows for comparison over time / between study areas, detection of trends *(Geographical/spatial bias)* Uniform sampling effort *(Geographical/spatial bias)*	Resource intensive to provide repeated coverage *(Geographical/spatial bias)* Limited spatial coverage – defined by survey design *(Geographical/spatial bias)*	McBride *et al.* (2019) Kiszka *et al*. (2007, Robbins *et al. (*2019) Ban *et al.* (2013) Harvey *et al.* (2018)
	Opportunistic	Survey whenever opportunities arise, such as during routine operations not specifically aimed at data collection of marine predators (e.g. ferry routes, whale watching trips).	Cost-effective, no additional costs for data collection *(Geographical/spatial bias)* Wide geographic range – not limited to predetermined path *(Geographical/spatial bias)*	Limited by other factors such as survey design – e.g. routes cannot be adapted according to surveyors needs *(Geographical/spatial bias)* Subject to spatial bias *(Geographical/spatial bias)* Non-uniform effort in sampling *(Geographical/spatial bias)*	
Number of platforms	**Single platform**	One observer positioned on the vessel/aircraft.	Simple logistically *(Geographical/spatial bias)* Lower cost *(Geographical/spatial bias)*	Higher chance of perception bias (*Perception bias)* Less coverage for detection and data acquisition. Limited field of view *(Availability bias)*	Lambert *et al.* (2019)

Category	Subcategory	Description	Advantages	Limitations	Reference
	Double platform	Two independent observers positioned at two different locations (often heights above sea level) on the same vessel/aircraft	More data acquisition per survey *(Survey perspective).* Higher chance of animal detection (*Perception bias)* Extended field of view *(Availability bias)*	Requires more observers - higher cost, potential for higher fuel demand *(Geographical/spatial bias)* Additional equipment and analysis time required *(Geographical/spatial bias)*	

3.4.2 ESAS and ORCA comparison

ESAS and ORCA were used as two different datasets, used as an example to how a dataset would be assessed and scored. Firstly, datasets are interrogated in terms of extent, taxonomic coverage, methodology, temporal and spatial coverage. Comparison of the two key datasets is summarised in Table 3.2, with detailed information on survey methodology given in Appendix C.

Two sets of data were provided by ORCA (see Appendix C for detailed methodology). The first is Distance Sampling data from the bridge of ferries during dedicated surveys and cruise ships, whilst the second is collected by Wildlife Officers on set ferry routes daily for up to six months of the year. The Wildlife Officers and Cruise teams operate with similar methods to Distance Sampling, but there may be some differences – such as the platform may not be forward facing, so they may be looking from the sides of the ship, and no bearing is taken of the sighting as such.

Table 3.2 Summary comparison of ESAS and ORCA datasets.

Parameter	ESAS	ORCA
Database manager	JNCC	ORCA
Spatial Extent (degrees)	Top left: 79.270°N, 55.522°W Bottom right: 12.800°N, 22.252°E	Top left: 72.649 °N, 26.065°W Bottom right: 33.896°N, 17.788°E
Total predator records	1,827,463	18,595
Birds	1,806,758	0
Cetaceans	16,924	18,382
Sharks	97	0
Pinnipeds	2,993	213
Turtles	344	0
Fish	181	0
Is there a standardised survey protocol?	Yes	Yes
Is a quality control procedure in place?	Yes	Yes
Temporal coverage of supplied dataset	1979 to 2011	2009 to 2018
Number of years of data in dataset	32	9

Parameter	ESAS	ORCA
Do surveyors undergo training in survey protocol, species ID?	Yes – only ESAS accredited surveyors used.	Yes – all surveyors undergo ORCA training course.
Sighting area (degrees)	90°	100°
Transect width	300 m	
Recording intervals	Intervals recorded – no sightings still recorded.	Only recorded when sighting.
Codes recorded	Time, species, age, plumage, number, distance band, direction, fly / sea, prey, Transect, Notes (e.g. behaviours, associations, floating matter)	Time, sighting number, watch number, GPS reading, reticle distance, angle, eye distance (m), species, school size, number of calves, cue, behaviour, animal heading, observer, comments (e.g. mixed group, behaviour)
Environmental data recorded	Wind force, sea state, swell height, visibility, cloud, rain, sun strength, sun direction	Sea state beaufort, swell, precipitation, glare, visibility (km), observer, comments. Taken every 30 minutes.
Navigation	GPS recorded on Navigation sheet.	Event time, effort, watch number, GPS reading, speed (knots), course of vessel.

Spatial coverage

Spatial coverage for the ESAS and ORCA datasets was similar (Figure 3.3) with density of data increasing around the UK. Shipping routes are more evident in the ORCA data, with clear paths of travel matching ferry routes (e.g. across the Bay of Biscay, around Portugal, returns to Iceland). Shipping routes are also visible in the ESAS dataset, along with specific transect surveys, but the longer temporal scale of the data means more coverage over less-well-surveyed areas.

Higher effort is skewed towards the coastline for the datasets, with 35.45% of the occurrence records being located within 12 nautical miles (inshore waters) from the coastline, thus coastal areas have high sampling intensity and repeated coverage, likely due to ease of access.

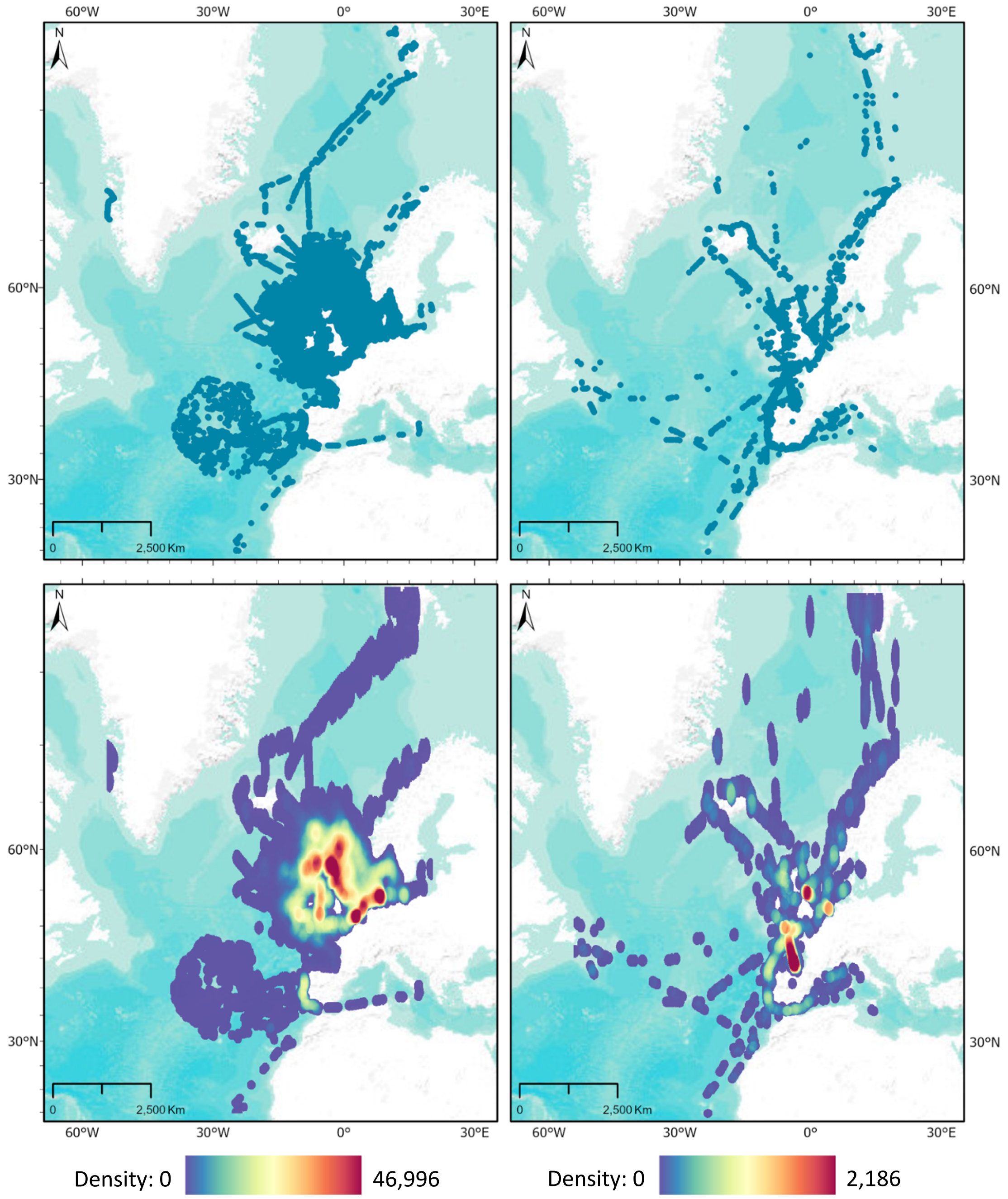


Figure 3.3 Sightings (top row) and heat map of data density (bottom row) for the ESAS and ORCA datasets.

Temporal coverage

The ESAS dataset has greater temporal coverage and more data entries than the ORCA dataset, and observations vary by year (Figure 3.4), with maximum data collection in 1987 (n = 116,388) and minimum in 2011 (n = 1,893). ORCA data shows a substantial increase in data collection from 2014 (mean count from 2006 to 2013 = 291.5, mean from 2014 to 2018 = 3249.8), with maximum data collection in 2017 (n = 3,679) and minimum in 2009 (n = 166). Patterns in the seasonal distribution of data collection are clear in both the ESAS and ORCA datasets (Figure 3.5), peaking in UK summer months (April to September) (maximum count for ORCA = 3,572 in June, ESAS = 311,909 in July), with less collected during winter months (October to March) (minimum count for ORCA = 2 in December, ESAS = 61,365 in December).

Survey method

Survey by boat is the most abundant data collection method for the ESAS dataset (Figure 3.6) and has the most oceanic and complete coverage (Figure 3.7). Plane and helicopter conversely tend to circumnavigate the coastline of the UK, with several aerial strip surveys North Sea around Denmark and in the Moray Firth. In ORCA data 'Distance sampling' has a much wider spatial footprint than 'WO sightings', due to distance sampling being carried out on ferries and cruise ships which are more oceanic, whereas WO is ferry only (Figure 3.8).

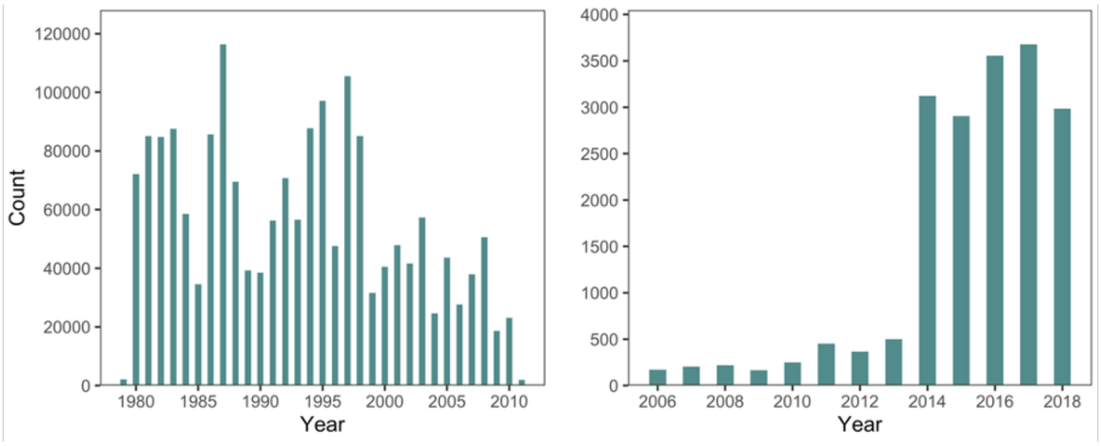


Figure 3.4 **Data counts by years for ESAS and ORCA.**

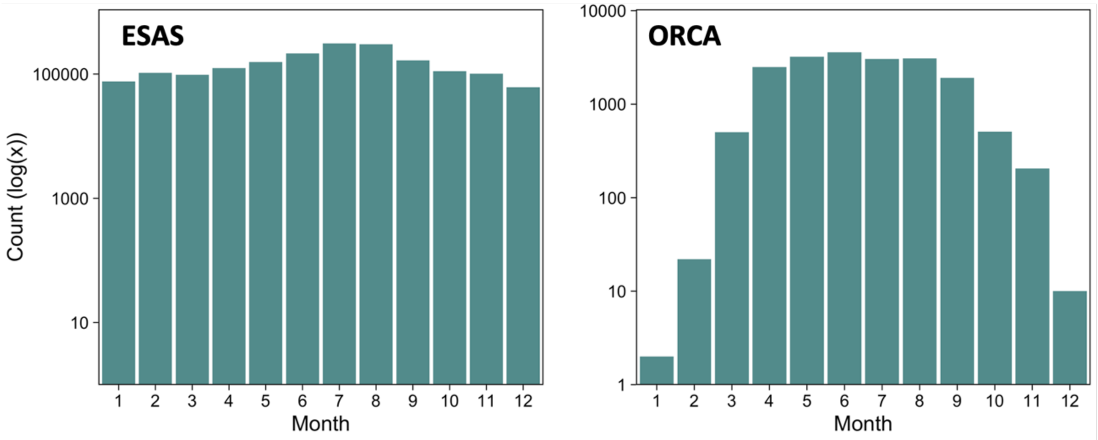


Figure 3.5 Data counts by months for ESAS and ORCA.

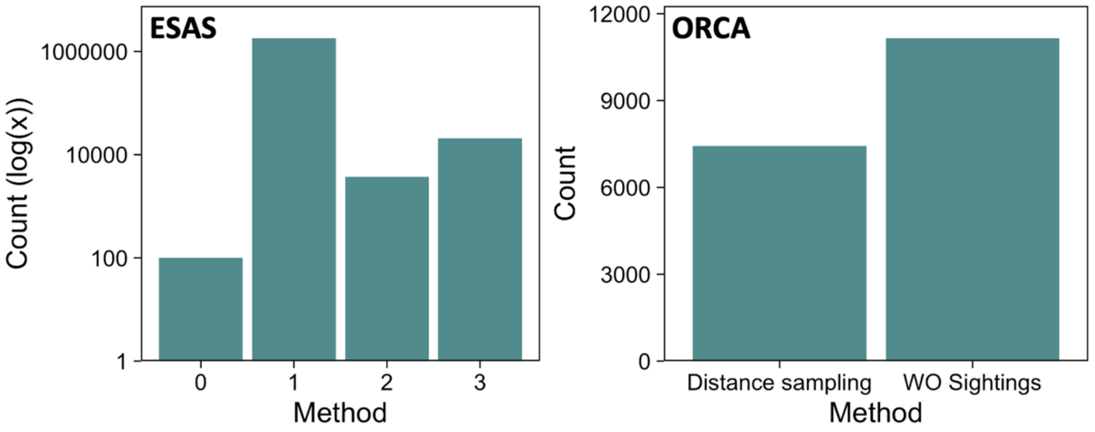


Figure 3.6 ESAS and ORCA data count by survey type. ESAS (on log scale) = 0) Null 1) Boat 2) Helicopter 3) Plane. ORCA data = Distance Sampling versus Wildlife Officer sightings.

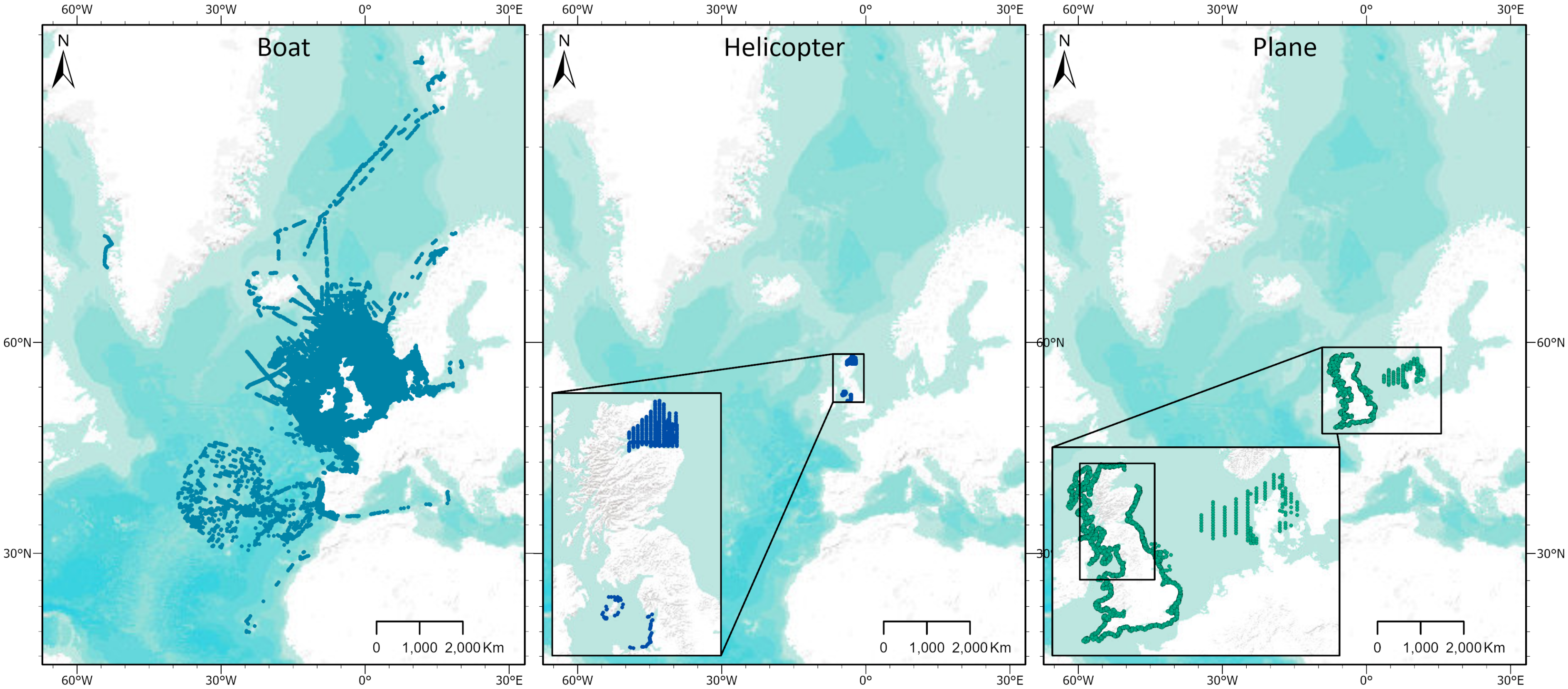


Figure 3.7 Spatial coverage of ESAS data count by survey type – 1) Boat 2) Helicopter 3) Plane.

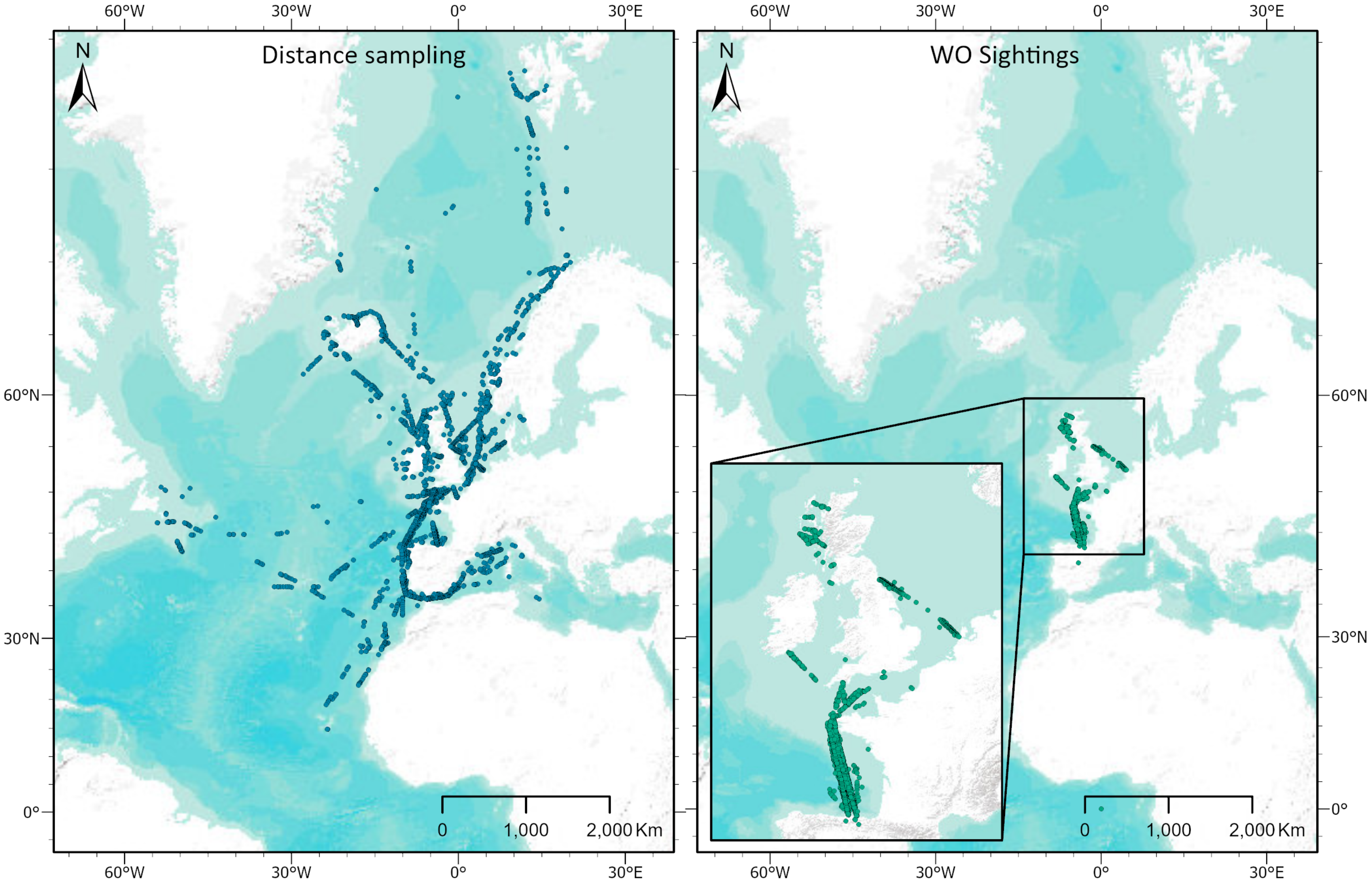


Figure 3.8 Spatial coverage of ORCA data count by survey type – 1) Distance Sampling 2) WO sightings.

Taxonomic coverage

Bird sightings dominate the ESAS dataset (Figure 3.9), comprising 99% of the data entries, which is anticipated due to the seabird focus of the dataset from its conception. It is however, still a valuable resource, with other taxa recorded and a lack of long-term data on marine predators. Sightings of Northern gannets *Morus bassanus* (n = 492,140) and fulmar *Fulmarus glacialis* (n = 486,981) were highly abundant, followed by gull species (e.g. lesser black-backed gull *Larus fuscus* = 124,620, great black-backed gull *Larus marinus* = 82,651, common gull *Larus canus* = 79,098). Ross's gull *Rhodostethia rosea* and yellow-legged gull *Larus michahellis* had just one sighting each in the dataset. For cetaceans, harbour porpoise was most abundant (n = 8160), with common dolphins (n = 2,259), white-beaked *Lagenorhynchus albirostris (*n = 2,160), and minke whale *Balaenoptera acutorostrata* (n = 1,364) also high in numbers. Right whale *Eubalaena glacialis* was lowest, with just one sighting. Seven species of pinnipeds were recorded, with harbour seal *Phoca vitulina* the most common pinniped in the ESAS database (n = 1,734), with walrus *Odobenus rosmarus* lowest with one sighting. For shark taxa, the most observations were basking shark *Cetorhinus maximus* (n = 63), and just one sighting per species for mako *Isurus oxyrinchus* and porbeagle *Lamna nasus*. There were 343 loggerhead turtle *Caretta caretta* sightings, but just one green turtle *Chelonia mydas*, and 181 sunfish *Mola mola* were sighted.

In the ORCA dataset, only cetaceans and marine mammals are present, with common dolphin highest in observations (n = 6,212) and rough-toothed dolphin *Steno bredanensis* lowest (n = 1). Some records were not detailed down to species level (n = 4,934) and were instead labelled more vaguely (e.g. 'big cetacean', 'medium cetacean', 'patterned dolphin). For grey seal *Halichoerus grypus*, the only pinniped in this dataset, there were 212 sightings.

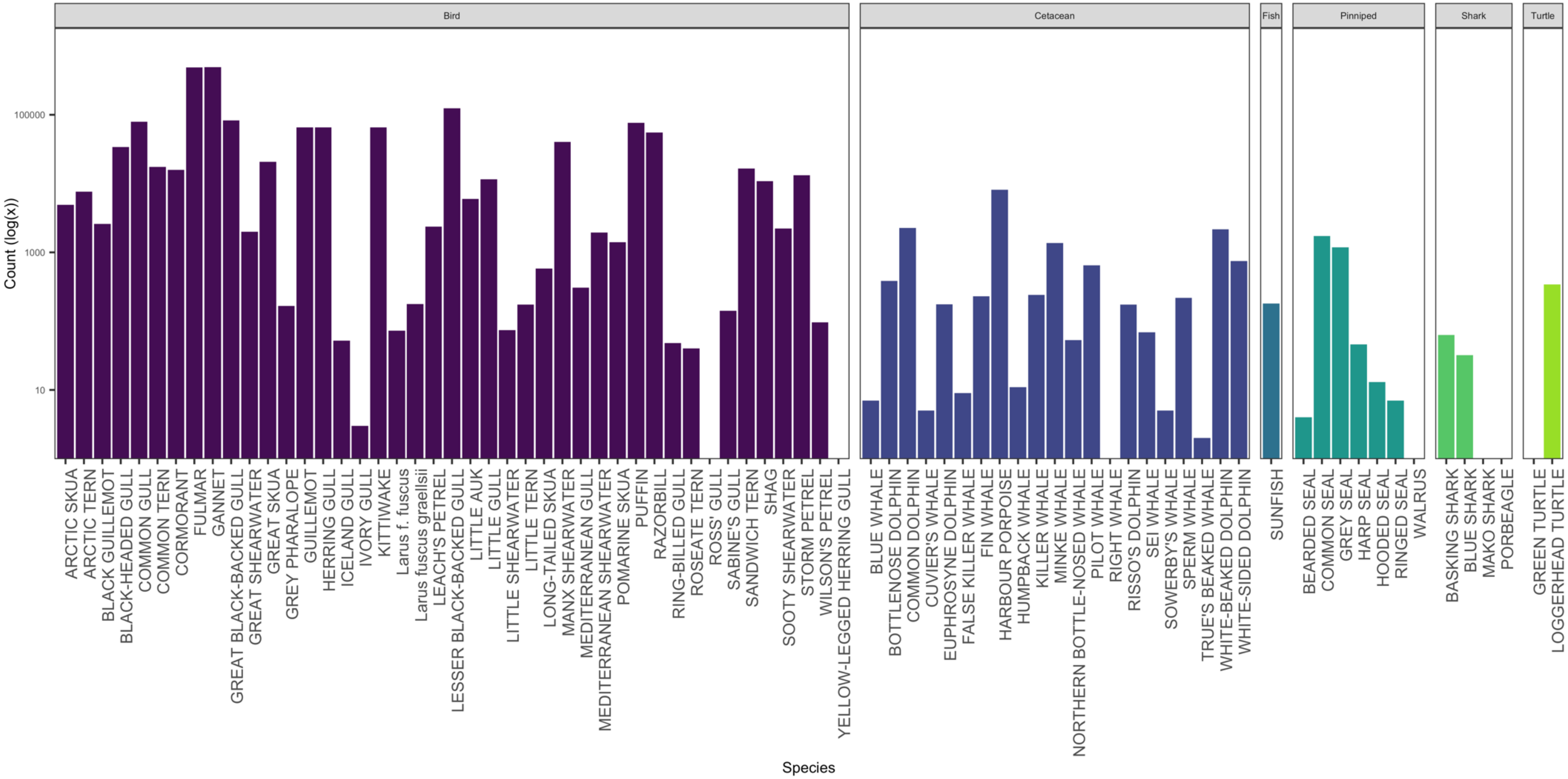


Figure 3.9 Species counts in ESAS database.

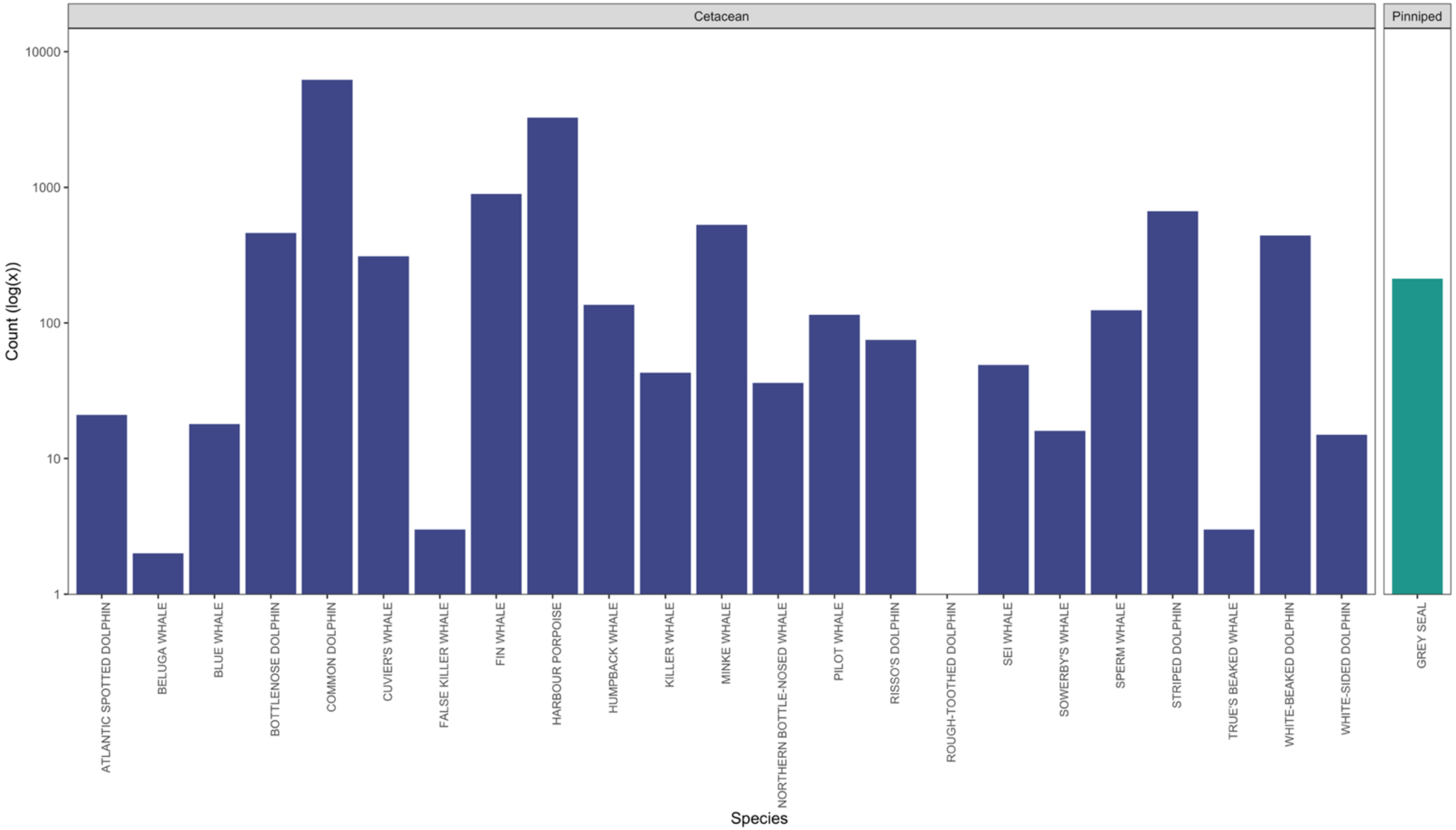


Figure 3.10 Species counts in ORCA database.

3.4.3 Biases within ESAS and ORCA datasets

Inherently, the two key datasets contain all the biases discussed in section 3.2.1 to some degree. Geographic or spatial sampling biases is clearly visible in the ESAS and ORCA datasets (Figure 3.3) with higher effort distributed along coastlines where areas are more easily accessible or more frequently traversed. Particularly for opportunistic surveys like the ORCA dataset, surveys are often concentrated over specific areas and at fixed season due to focus on ferry or cruise ship routes.

Sampling perspective and platform vary both between and within datasets. The ESAS database combines aerial and ship survey data, whilst the ORCA dataset is entirely ship based, but perspectives vary depending on ferry or cruise ship platforms. Both datasets have variations in platform within the datasets themselves that lead to different levels of bias: ESAS combines platforms of boats of many vessel types and aerial surveys from both plane and helicopter (Figure 3.6), and though ORCA uses boat only, which may reduce some differences in survey platforms, it uses both ferry and cruise ships for sightings which have different perspectives (as discussed in section 3.2.1).

In terms of availability bias, for both datasets, at-sea counts may be underrepresenting species that haul out, particularly pinnipeds. These species are often recorded by land, for example harbour seals are counted while they are on land during their August moult whilst grey seals pups are counted during the Autumn when females congregate on land to give birth (SCOS, 2021). Given marine predators have highly responsive behavioural adaptations to impacts such as underwater noise from vessels (and both datasets are boat-heavy in platform), the likelihood of availability bias is high. Both ESAS and ORCA databases include codes for behaviour, to allow for some estimation of displacement, either via avoidance or attraction behaviours. These behaviours are highly vessel dependent, and ESAS guidance considers data from around commercial fishing vessels not suitable for accurate estimates of abundance, given high discards that may attract higher amounts of seabirds and any associated animals attracted to increased activity.

There is clear evidence of perception bias given in the way species are recorded in the datasets. In the ESAS dataset for example, sightings of pilot whales are recorded simply as ‘Pilot whale’ and not as long-finned pilot whale *Globicephala melas* or the short-finned pilot whale *Globicephala macrorhynchus* as in the ORCA dataset, but then in the ORCA dataset there are many codes that are not detailed to species level (e.g. ‘patterned dolphin’, ‘small dolphin’). Whilst detection to species level may not be possible, it is still beneficial to record those sightings under such a high-level label to increase coverage. The temporal coverage of the ESAS and ORCA data, with lots of years of sampling over an area, may increase the chance of a species being sighted.

Human error is also possible when distance recording by observers, such as in the ESAS and ORCA datasets that use banding or reticule methods. With estimating densities, overestimating distances may negatively bias values but underestimating distances may positively bias estimates. As discussed in 3.2.1, the use of two or more observers can reduce perception bias, and both the ESAS and ORCA protocol consist of two observers at any one time. Both ESAS and ORCA surveyors undergo training and follow a protocol to minimise misidentification or missing animals, with a quality control process afterwards to minimise recording errors.

Independent animal movement exists irrespective of the two datasets, but highlights the importance of proper study design, quantification of bias within datasets and crucially when working with historical datasets. Post survey completion there is no clear analytical way currently to reduce bias in abundance estimates using the standard methods (Glennie *et al.*, 2015), so using historical datasets such as ESAS or ORCA that have not factored in these movement corrections from the onset may be subject to bias.

3.4.4 Matrix approach

Choosing the correct research question to suit the data source of choice is critically important, as the level of risk changes depending on the intended use of the data.

Four examples were assessed using the matrix approach: ESAS, ORCA, SCANS-IV aerial data (Gilles *et al.,* 2023) and APEM DAS data (APEM Ltd, 2023). Table 3.3 and Table 3.4 demonstrates using the risk assessment matrix to quantify biases that exist in four datasets, when two different measures are wanted. Table 3.3 quantifies the level of bias that exists when using both datasets to get counts or abundances of species in an area. Table 3.4 quantifies the level of bias that exists when using both datasets to identify species presence in an area.

When comparing both count (Table 3.3) and presence (Table 3.4) risk assessment matrices, the overall risk scores are lower for biases in the dataset when looking at using them for relative questions such as whether a species is present in an area or not recorded (overall risk score = 71 ESAS, 70 ORCA, 48 SCANS-IV, 53 DAS APEM) than for count/abundance questions (overall risk score = 107 ESAS, 99 ORCA, 63 SCANS-IV, 71 DAS APEM).

Risk scores for each bias for the two research question types, for all four datasets, are given in Figure 3.11 – with each bias labelled A to M in the table beneath. Figure 3.11 aids quick visualisation of key differences between the two matrices.

For independent animal movement to observer (Bias A), level of training of observers (Bias E), probability detection (Bias H) and sampling perspective (Bias L) there was no difference between the research questions for the four datasets (i.e. they were the same risk score for both count and presence). For platform height (Bias B) the ESAS and ORCA dataset had a higher risk score than SCANS-IV and DAS APEM, for both the 'count/absence and 'presence' questions. For observers missing animals (Bias C) ESAS had the highest risk score, followed by DAS APEM, for both count and presence. For misidentification of species (Bias D), ESAS had the highest risk, with others equal for count, but risk was reduced for presence for SCANS-IV and DAS APEM datasets. For behaviour towards platforms (positive) (Bias F) SCANS-IV and DAS APEM was lowest, followed by ORCA and then ESAS, with risk scores for all four datasets less for presence questions. For seasonal patterns of usage (Bias I) ESAS and SCANS were lower than SCANS-IV and APEM DAS (which was highest) for both count and presence, though risk scores reduced for presence. The ORCA dataset gave higher risk scores for behaviour away from platforms (Bias G) and geographic sampling bias for both high use area (Bias J) and low use area (Bias K) than the others, for count. For presence, risk reduced for all datasets. For misidentification of species (Bias D) and failure to record null findings (Bias M) using the datasets for obtaining 'counts' resulted in higher risk scores than for 'presence'.

The number of biases scored per risk score (which ranges from 1 to 20) is presented in Figure 3.12. When scoring biases for "presence purpose" the number of biases for very low risk scores (1 to 4) was generally greater than when scoring biases for 'count purposes'. The number of biases at higher risk score was greater for 'count purposes', suggesting that there is a greater risk of using the four datasets for identifying counts or abundance of species in an area than identifying areas of high species presence or absence. This is highlighted in the trend lines (calculated using best-fit trendline in Excel version 16.70 (Microsoft Corporation, 2018)), demonstrating the number of biases is more evenly spread across risk scores for ‘count purposes’ but there are more low risk biases and less high-risk scores when using the datasets for ‘presence’ research questions.

Table 3.3 Bias risk assessment matrix for using datasets for count or abundance purposes.

Risks			ESAS				ORCA				SCANS IV (aerial surveys)				DAS APEM			
Bias	**Description**	**Directionality**	**Severity**	**Likelihood**	**Risk**	**Justification**	**Severity**	**Likelihood**	**Risk**	**Justification**	**Severity**	**Likelihood**	**Risk**	**Justification**	**Severity**	**Likelihood**	**Risk**	**Justification**
Perception bias - observers do not notice animals present.	Independent animal movement to observer - can affect abundance estimates.	Underestimate	2	5	10	Cannot be controlled, independent to any steps that can be taken by surveyors so whilst severity of bias is low as any dataset is susceptible to this, likelihood is almost certain for any survey,	2	5	10	Cannot be controlled, independent to any steps that can be taken by surveyors so whilst severity of bias is low as any dataset is susceptible to this, likelihood is almost certain for any survey,	2	4	8	Cannot be controlled, independent to any steps that can be taken by surveyors so whilst severity of bias is low as any dataset is susceptible to this, likelihood is almost certain for any survey. Speed relative to species means less chance of missing species.	2	4	8	Cannot be controlled, independent to any steps that can be taken by surveyors so whilst severity of bias is low as any dataset is susceptible to this, likelihood is almost certain for any survey. Speed relative to species means less chance of missing species.
	Platform height - low survey platforms such as boats rather than aeroplanes or helicopters, more likely to miss marine mammals or higher flying birds further from field of view.	Underestimate	1	2	2	Low platform is better than no platform, so severity of bias is reduced. Majority of data is vessel based, but some aerial surveys included. Likelihood reduced due use of aerial surveys as well as vessels.	1	2	2	Low platform is better than no platform, and it is the most available method of survey, so severity of bias is reduced. Likelihood higher due to majority of data being vessel based, but from ferries / cruise ships means platform height and field of view slightly increased.	1	1	1	High platform (aircraft) likely to observe more animals present and can record those submerged.	1	1	1	High platform (aircraft) likely to observe more animals present and can record those submerged, however image collection is 2 km apart between image points (only 3% coverage of area flown).

	Observers missing the animals when they are present due to factors such as human error, number of observers, fatigue.	Underestimate	5	3	15	Severity is high as missing crucial sightings for abundances. Likelihood higher if only 2 observers used as opposed to 4 (in ORCA), and if multiple species observed at once (as is likely in feeding assemblages around ships).	5	1	5	Severity is high as missing crucial sighting for abundances. likelihood reduced due to ORCA using 4 person observer teams in rotation.	5	1	5	Severity is high as missing crucial sighting for abundances. Likelihood reduced due to 2 observers from high vantage point in addition to one member recording data.	5	2	10	Severity is high as missing crucial sighting for abundances. Likelihood lower as data is captured via camera system, not manually recorded via observers however highly dependent on node aligning with animals, could have many blank images.
	Misidentification of species - similar family but incorrect species, such as between phalaropes, storm petrels, small jaegers, alcids and terns.	N/A	3	3	9	Severity reasonable for counts as trying to quantify abundance so need numbers, and likelihood is higher due to difficulty in detecting species levels for some birds.	3	1	3	Severity reasonable for counts as trying to quantify abundance so need numbers, but likelihood of misidentifying cetaceans is less than misidentifying similar bird species.	3	1	3	Severity reasonable for counts as trying to quantify abundance so need numbers, but whilst likelihood of misidentifying species higher from aerial surveys, digital pictures are analysed by a human operator to validate species identification and group size of visual sightings, especially in the case of unidentified common/striped dolphin sightings.	3	1	3	Severity reasonable for counts as trying to quantify abundance so need numbers, but digital pictures are analysed by a human operator to validate species identification and group size of visual sightings.
	Level of training of observers affecting identification- citizen science versus trained systemic surveys.	N/A	2	2	4	Hard to assume professional / experience surveyors are doing any better job than citizen scientists / less experienced people, so severity and likelihood is low.	2	2	4	Hard to assume professional / experience surveyors are doing any better job than citizen scientists / less experienced people, so severity and likelihood is low.	2	1	2	Hard to assume professional / experience surveyors are doing any better job than citizen scientists / less experienced people, so severity and likelihood is low. However, digital pictures are analysed by a	2	1	2	Hard to assume professional / experience surveyors are doing any better job than citizen scientists / less experienced people, so severity and likelihood is low. However, digital

														human operator to validate species identification and therefore less likelihood.				pictures are analysed and QA'd by a human operator to validate species identification and therefore less likelihood.
Availability bias - probability of animal being available to detect.	Behaviour towards platforms - if animals are **attracted** to boats (e.g. then density estimates may be positively biased.	Overestimate	4	3	12	Severity less than if displaying avoidance behaviours, but will still skew counts / abundances. likelihood higher for some taxa such as birds who follow ships.	4	2	8	Severity less than if displaying avoidance behaviours, but will still skew counts / abundances. likelihood less than ESAS due to cetacean based, less known attraction behaviour.	4	1	4	Severity less than if displaying avoidance behaviours, but will still skew counts / abundances. Less attraction behaviour likely for aircrafts than vessels.	4	1	4	Severity less than if displaying avoidance behaviours, but will still skew counts / abundances. Less attraction behaviour likely for aircrafts than vessels.
	Behaviour away from platforms - if animals display **avoidance** to platform (e.g. porpoises) then density estimates may be negatively biased.	Underestimate	5	3	15	Severity higher than if displaying attraction behaviours, as will skew counts / abundances negatively. Likelihood higher for some taxa such as porpoises.	5	4	20	Severity higher than if displaying attraction behaviours, as will skew counts / abundances negatively. likelihood higher for ORCA due to use of large ferries and cruise ships.	5	1	5	Severity higher than if displaying attraction behaviours, as will skew counts / abundances negatively. Likelihood higher for birds, but less for aircraft likely than vessel noise.	5	1	5	Severity higher than if displaying attraction behaviours, as will skew counts / abundances negatively. Likelihood higher for birds, but less for aircraft likely than vessel noise.
	The probability of the animal being available to detect - often a function of depth-use patterns and turbidity.	Underestimate	1	5	5	Severity is low as cannot avoid this bias, it is due to animal life history, however likelihood is maximum as it is inevitable.	1	5	5	Severity is low as cannot avoid this bias, it is due to animal life history, however likelihood is maximum as it is inevitable.	1	4	4	Severity is low as cannot avoid this bias, it is due to animal life history, however likelihood is high as inevitable, however high vantage point will increase chances of detection.	1	4	4	Severity is low as cannot avoid this bias, it is due to animal life history, however likelihood is high as inevitable, however high vantage point will increase chances of detection.

	Seasonal patterns of usage - movement at different times of the year (migrations, breeding seasons) may impact availability at different times of year and for different species.	Both	3	3	9	With survey effort increasing during summer months, may affect count estimations - over estimate of abundances compared to winter where survey effort is lower. Particularly with bird species who migrate for the winter.	3	3	9	With survey effort increasing during summer months, may affect count estimations - over estimate of abundances compared to winter where survey effort is lower. Particularly with any marine mammal species that have long migrations such as humpback whales.	3	4	12	Summer surveys only, therefore skewed toward summer species (e.g. minke whale move into Irish Sea for summer months).	3	5	15	Limited temporal coverage e.g. October to December 2022, winter focused.
Geographic sampling bias	Influenced by the accessibility to the area - e.g. proximity to the coastline, ports, shipping routes, areas of fishing / tourism activity - **High use area, areas frequently surveyed.**	Overestimate	4	3	12	Severity is high for counts and abundances can skew counts to overestimate in areas of high use. Likelihood medium due to following well used routes or along coastlines.	4	5	20	Severity is high for counts and abundances can skew counts to overestimate in areas of high use. Likelihood higher due to ORCA being opportunistic and using ferries / cruise ships with set routes.	4	3	12	Repeated survey areas and transects for each SCANS survey, therefore highly biased towards sampling blocks however good even coverage due to line survey approach - designed to provide equal coverage probability (using the equal spaced zig-zag option, and in coastal blocks the parallel option).	4	3	12	Only select areas surveyed with limited coverage aimed at set survey area, and image collection is 2 km apart between image points so allows even coverage, reducing sampling bias.
	Influenced by the accessibility to the area - e.g. proximity to the coastline, ports, shipping routes, areas of fishing / tourism activity - **Low use area, areas less visited.**	Underestimate	3	2	6	Severity is high for counts and abundances, can skew counts to appear low in areas that are comparatively under sampled. Likelihood medium due to surveys usually following well used routes or along	3	2	6	Severity is high for counts and abundances, can skew counts to appear low in areas that are comparatively under sampled. Likelihood medium due to surveys usually following very well	3	1	3	Severity is high for counts and abundances, can skew counts to appear low in areas that are comparatively under sampled. Likelihood low due to line transect approach designed to provide equal coverage probability (using the equal	3	1	3	Severity is high for counts and abundances, can skew counts to appear low in areas that are comparatively under sampled. Likelihood lower due to line transect approach designed to provide

						coastlines, rather than travelling into low use areas.				used routes or along coastlines and not heading into low use areas (particularly for cruise ships and ferries, large vessels).				spaced zig-zag option, and in coastal blocks the parallel option).				equal coverage with images captured whether animals sighted or not.
Sampling **perspective** / platform	Can affect extent, duration, frequency and sampling resolution of surveys. E.g., aerial methods reach more inaccessible areas than ships, but may not be able to travel as offshore.	Both	1	3	3	Severity is low as all methods of collection will still cover relatively similar scope. Inclusion of aerial surveys for ESAS means likelihood reduced, but as all survey platforms present there may be biases within the data.	1	2	2	Severity is low as ORCA just uses vessels, thus likelihood is also low.	1	1	1	Severity is low as only aerial surveys used in set survey methodology, rather than combined methods. High vantage point.	1	1	1	Severity is low as only aerial surveys used in set survey methodology, rather than combined methods. High vantage point.
Failure to record null findings	Trips where animals are not found are often not recorded, leading to presence only dataset.	N/A	1	5	5	Severity is low as counts are based upon sightings, however likelihood high as null findings are not recorded	1	5	5	Severity is low as counts are based upon sightings, however likelihood high as null findings are not recorded	1	3	3	Severity is low as counts are based upon sightings, lower severity as null findings are recorded.	1	3	3	Severity is low as counts are based upon sightings, lower severity as null findings are recorded.
			Total Risk Score		**107**		**Total Risk Score**		**99**		**Total Risk Score**		**63**		**Total Risk Score**		**71**	

Table 3.4 Bias risk assessment matrix for using datasets for relative measures (presence) purposes.

Risks			ESAS				ORCA				SCANS IV (aerial surveys)				DAS APEM			
Bias	Description	Directionality	Severity	Likelihood	Risk	Justification	Severity	Likelihood	Risk	Justification	Severity	Likelihood	Risk	Justification	Severity	Likelihood	Risk	Justification
Perception bias - observers do not notice animals present.	Independent animal movement to observer - can affect whether sighting is made or not.	Underestimate	2	5	10	Cannot be controlled, independent to any steps that can be taken by surveyors so whilst severity of bias is low as any dataset is susceptible to this, likelihood is almost certain for any survey,	2	5	10	Cannot be controlled, independent to any steps that can be taken by surveyors so whilst severity of bias is low as any dataset is susceptible to this, likelihood is almost certain for any survey,	2	4	8	Cannot be controlled, independent to any steps that can be taken by surveyors so whilst severity of bias is low as any dataset is susceptible to this, likelihood is almost certain for any survey. Speed relative to species means less chance of missing species.	2	4	8	Cannot be controlled, independent to any steps that can be taken by surveyors so whilst severity of bias is low as any dataset is susceptible to this, likelihood is almost certain for any survey. Speed relative to species means less chance of missing species.
	Platform height - low survey platforms such as boats rather than aeroplanes or helicopters, more likely to miss marine mammals or higher flying birds further from field of view.	Underestimate	1	2	2	Low platform is better than no platform, so severity of bias is reduced. Majority of data is vessel based, but some aerial surveys included. Likelihood reduced due use of aerial	1	2	2	Low platform is better than no platform, and it is the most available method of survey, so severity of bias is reduced. Likelihood higher due to majority of data being vessel based, but from ferries / cruise ships means platform height and	1	1	1	High platform (aircraft) likely to observe more animals present and can record those submerged.	1	1	1	High platform (aircraft) likely to observe more animals present and can record those submerged, however image collection is 2 km apart between image points (only 3% coverage of area flown).

						surveys as well as vessels.				field of view slightly increased.								
	Observers missing the animals when they are present due to factors such as human error, number of observers, fatigue.	Underestimate	3	3	9	Severity is medium as missing crucial sightings, will not be as critical as for numbers as it is more likely that at least one from all species present is recorded thus "present in area", rather than numbers. likelihood higher if only 2 observers used as opposed to 4 (in ORCA), and if multiple species observed at once (as is likely in feeding assemblages around ships).	3	1	3	Severity is medium as missing crucial sightings, will not be as critical as for numbers as it is more likely that at least one from all species present is recorded, rather than numbers. Likelihood reduced due to ORCA using 4 person observer teams in rotation.	3	1	3	Severity is medium as missing crucial sightings, will not be as critical as for numbers as it is more likely that at least one from all species present is recorded, rather than numbers. Likelihood reduced due to 2 observers from high vantage point in addition to one member recording data.	3	2	6	Severity is medium as missing crucial sightings, will not be as critical as for numbers as it is more likely that at least one from all species present is recorded, rather than numbers. Likelihood lower as data is captured via camera system, not manually recorded via observers however highly dependent on node aligning with animals, could have many blank images.

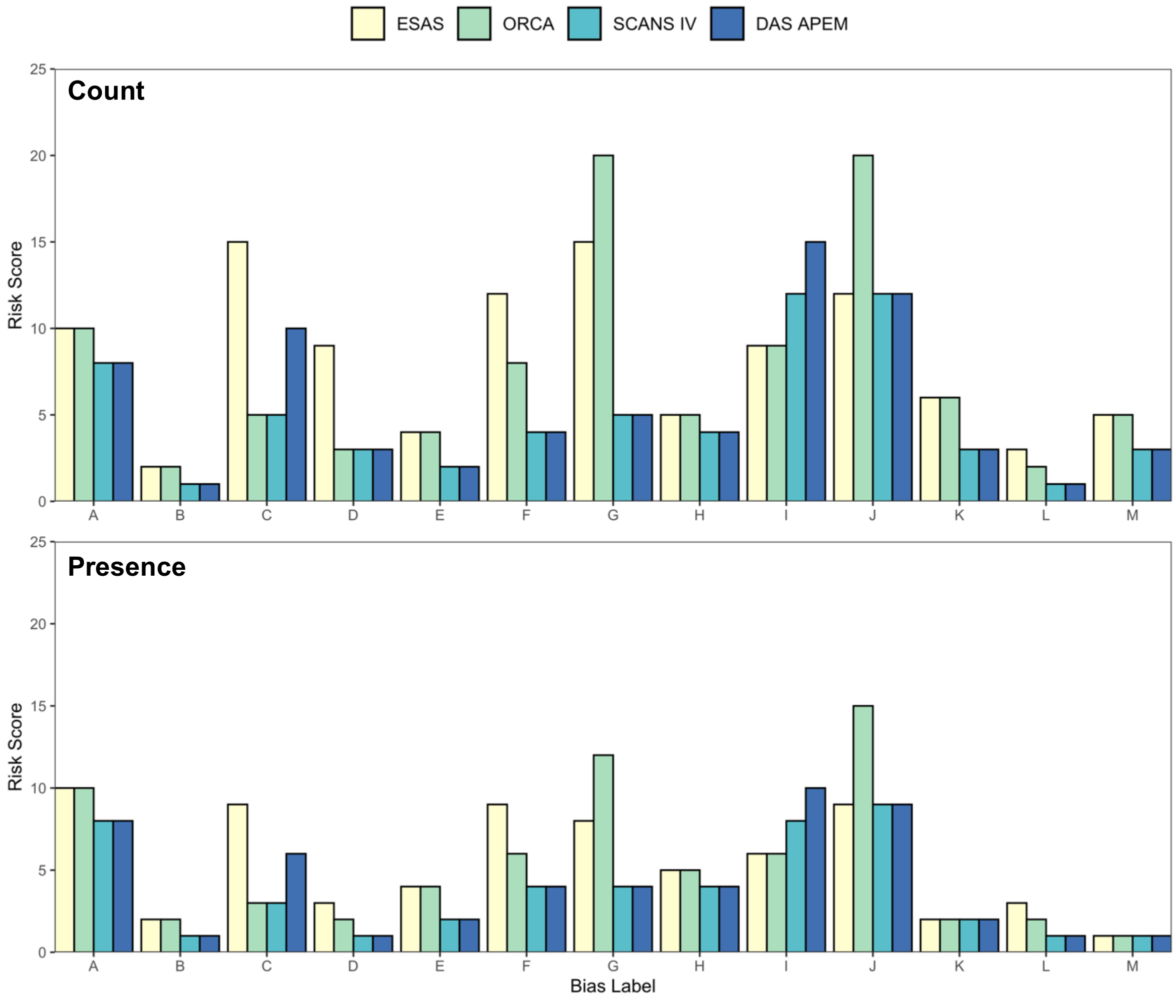
	Misidentification of species - similar family but incorrect species, such as between phalaropes, storm petrels, small jaegars, alcids and terns.	N/A	1	3	3	Severity lower for presence sightings, and likelihood is medium due to difficulty in detecting species levels for some birds.	1	2	2	Severity lower for presence than counts, but likelihood of misidentifying cetaceans is less than misidentifying similar bird species.	1	1	1	Severity reasonable for counts as trying to quantify abundance so need numbers, but whilst likelihood of misidentifying species higher from aerial surveys, digital pictures are analysed by a human operator to validate species identification and group size of visual sightings, especially in the case of unidentified common/striped dolphin sightings.	1	1	1	Severity reasonable for counts as trying to quantify abundance so need numbers, but digital pictures are analysed by a human operator to validate species identification and group size of visual sightings.
	Level of training of observers affecting identification- citizen science versus trained systemic surveys.	N/A	2	2	4	Hard to assume professional / experience surveyors are doing any better job than citizen scientists / less experienced people, so severity and likelihood is low. Birds and cetaceans to ID.	2	2	4	Hard to assume professional / experience surveyors are doing any better job than citizen scientists / less experienced people, so severity and likelihood is low. Cetaceans easier to detect and only 1 from each species requiring high level ID.	2	1	2	Hard to assume professional / experience surveyors are doing any better job than citizen scientists / less experienced people, so severity and likelihood is low. However, digital pictures are analysed by a human operator to validate species identification and therefore less likelihood.	2	1	2	Hard to assume professional / experience surveyors are doing any better job than citizen scientists / less experienced people, so severity and likelihood is low. However, digital pictures are analysed and QAd by a human operator post survey to validate species identification and therefore less likelihood.

Availability bias - probability of animal being available to detect.	Behaviour towards platforms - if animals are **attracted** to boats . then density estimates may be positively biased.	Overestimate	3	3	9	Severity medium as will affect whether sightings are made, but less severe than if wanting counts / abundances. likelihood higher for some taxa such as birds who follow ships.	3	2	6	Severity medium as will affect whether sightings are made, but less severe than if wanting counts / abundances. likelihood higher for some taxa who follow ships, but likely less for cetaceans than bird species.	4	1	4	Severity less than if displaying avoidance behaviours, but will still skew counts / abundances. Less attraction behaviour likely for aircrafts than vessels.	4	1	4	Severity less than if displaying avoidance behaviours, but will still skew counts / abundances. Less attraction behaviour likely for aircrafts than vessels.
	Behaviour away from platforms - if animals display **avoidance** to platform (e.g. porpoises) then density estimates may be negatively biased.	Underestimate	4	2	8	Severity higher than if displaying attraction behaviours, but less severe than for counts. Likelihood higher for some taxa such as porpoises.	4	3	12	Severity higher than if displaying attraction behaviours, but less severe than counts. likelihood higher for some taxa such as porpoises. Likelihood higher for ORCA due to use of large ferries and cruise ships.	4	1	4	Severity higher than if displaying attraction behaviours, as will skew counts / abundances negatively. Likelihood lower as cetacean only focus and therefore less likelihood for aircraft likely than vessel noise.	4	1	4	Severity higher than if displaying attraction behaviours, as will skew counts / abundances negatively. Likelihood higher for birds, but less for aircraft likely than vessel noise.
	The probability of the animal being available to detect - often a function of depth-use patterns and turbidity.	Underestimate	1	5	5	Severity is low as cannot avoid this bias, it is due to animal life history, however likelihood is maximum as it is inevitable.	1	5	5	Severity is low as cannot avoid this bias, it is due to animal life history, however likelihood is maximum as it is inevitable.	1	4	4	Severity is low as cannot avoid this bias, it is due to animal life history, however likelihood is high as inevitable, however high vantage point will increase chances of detection of subsurface animals	1	4	4	Severity is low as cannot avoid this bias, it is due to animal life history, however likelihood is high as inevitable, however high vantage point will increase chances of detection of subsurface animals

	Seasonal patterns of usage - movement at different times of the year (migrations, breeding seasons) may impact availability at different times of year and for different species.	Both	2	3	6	With survey effort increasing during summer months, may affect sightings but as not measuring abundances or counts, less severe. Particularly with bird species who migrate for the winter.	2	3	6	With survey effort increasing during summer months, may affect sightings but as not measuring abundances or counts, less severe. Particularly with any marine mammal species that have long migrations such as humpback whales.	2	4	8	Summer surveys only, therefore skewed toward summer species (e.g. minke whale move into Irish Sea for summer months).	2	5	10	Limited temporal coverage e.g. October to December 2022.
Geographic sampling bias	Influenced by the accessibility to the area - e.g. proximity to the coastline, ports, shipping routes, areas of fishing / tourism activity - **High use area, areas frequently surveyed.**	Overestimate	3	3	9	Severity is high still for presence sightings, but will not be as severe as if wanting counts. Likelihood medium due to following well used routes or along coastlines.	3	5	15	Severity is high still for presence sightings, but will not be as severe as if wanting counts. likelihood higher due to ORCA being opportunistic and using ferries / cruise ships with set routes.	3	3	9	Severity is high still for presence sightings, but will not be as severe as if wanting counts. Likelihood reduced due to survey design giving even coverage due to line survey approach - designed to provide equal coverage probability (using the equal spaced zig-zag option and in coastal blocks the parallel option).	3	3	9	Only select areas surveyed with limited coverage aimed at set survey area, and image collection is 2 km apart between image points so allows even coverage, reducing sampling bias.
	Influenced by the accessibility to the area - e.g. proximity to the coastline, ports, shipping routes, areas of fishing / tourism activity -	Underestimate	2	1	2	Severity is lower as sighting presence in low use areas still means species are present, rather than	2	1	2	Severity is lower as sighting presence in low use areas still means species are present, rather than needing counts. Likelihood medium due to surveys	2	1	2	Severity is lower as sighting presence in low use areas still means species are present, rather than needing counts. Likelihood low due to line transect approach designed to provide equal coverage	2	1	2	Severity is lower as sighting presence in low use areas still means species are present, rather than needing counts. Likelihood lower due to line transect

	Low use area, areas less visited.					needing counts. Likelihood medium due to surveys usually following well used routes or along coastlines, rather than travelling into low use areas, however a record of one species in any low use area is less of a risk than aiming to derive counts.				usually following very well used routes or along coastlines and not heading into low use areas (particularly for cruise ships and ferries, large vessels), however record of one species in any low use area less risk than aiming to derive counts.				probability (using the equal spaced zig-zag option, and in coastal blocks the parallel option). Records of absence (aka transects with no sightings) also but not as crucial for presence. May cover low use areas with aerial survey approach.				approach designed to provide equal coverage with images captured whether animals sighted or not. Records of absence (aka images with no sightings) also but not as crucial for presence. May cover low use areas with aerial survey approach.
Sampling **perspective** / platform	Can affect extent, duration, frequency and sampling resolution of surveys. E.g., aerial methods reach more inaccessible areas than ships, but may not be able to travel as offshore.	Both	1	3	3	Severity is low as all methods of collection will still cover relatively similar scope. Inclusion of aerial surveys for ESAS means likelihood reduced, but as all survey platforms present there may be biases within the data.	1	2	2	Severity is low as ORCA just uses vessels, thus likelihood is also low.	1	1	1	Severity is low as only aerial surveys used in set survey methodology, rather than combined methods and has a high vantage point.	1	1	1	Severity is low as only aerial surveys used in set survey methodology, rather than combined methods and has a high vantage point.

Failure to record null findings	Trips where animals are not found are often not recorded, leading to presence only dataset.	N/A	1	1	1	Severity is low as just wanting sightings, however likelihood also low as just want to know if species present in area or not (i.e.. records of null sightings is not required).	1	1	1	Severity is low as just wanting sightings, however likelihood also low as just want to know if species present in area or not (i.e.. records of null sightings not required).	1	1	1	Severity is low as just wanting sightings, however likelihood also low as just want to know if species present in area or not (i.e.. records of null sightings not required). Records of absence (aka transects with no sightings) also but not as crucial for presence.	1	1	1	Severity is low as just wanting sightings, however likelihood also low as just want to know if species present in area or not (i.e.. records of null sightings not required). Records of absence (aka transects with no sightings) also but not as crucial for presence.
			Total Risk Score		**71**		**Total Risk Score**		**70**		**Total Risk Score**		**48**		**Total Risk Score**		**53**	



A	**(Perception Bias) Independent animal movement to observer – can affect abundance estimates.**
B	(Perception Bias) Platform height – low survey platforms such as boats rather than aeroplanes or helicopters, more likely to miss marine mammals or higher-flying birds further from field of view.
C	(Perception Bias) Observers missing the animals when they are present due to factors such as human error, number of observers, fatigue.
D	(Perception Bias) Misidentification of species – similar family but incorrect species, such as between phalaropes, storm petrels, small jaegers, alcids and terns.
E	(Perception Bias) Level of training of observers affecting identification- citizen science versus trained systemic surveys.
F	(Availability bias) Behaviour towards platforms (attraction) – if animals are attracted to boats (e.g. then density estimates may be positively biased.
G	(Availability bias) Behaviour away from platforms – if animals display avoidance to platform (e.g. porpoises) then density estimates may be negatively biased.
H	(Availability bias) The probability of the animal being available to detect (avoidance) – often a function of depth-use patterns and turbidity.
I	(Availability bias) Seasonal patterns of usage – movement at different times of the year (migrations, breeding seasons) may impact availability at different times of year and for different species.
J	(Geographic sampling bias) Influenced by the accessibility to the area – e.g. proximity to the coastline, ports, shipping routes, areas of fishing / tourism activity – High use area, areas frequently surveyed.

K	(Geographic sampling bias) Influenced by the accessibility to the area – e.g. proximity to the coastline, ports, shipping routes, areas of fishing / tourism activity – Low use area, areas less visited.
L	(Sampling perspective) Can affect extent, duration, frequency, and sampling resolution of surveys e.g. aerial methods reach more inaccessible areas than ships but may not be able to travel as far offshore.
M	(Failure to record null findings) Trips where animals are not found are often not recorded, leading to presence only dataset.

Figure 3.11 Risk scores (Severity * Likelihood) for each bias (A – M), for two research questions – obtaining species counts for ESAS data and ORCA data, and for obtaining species presence with ESAS data and ORCA data.

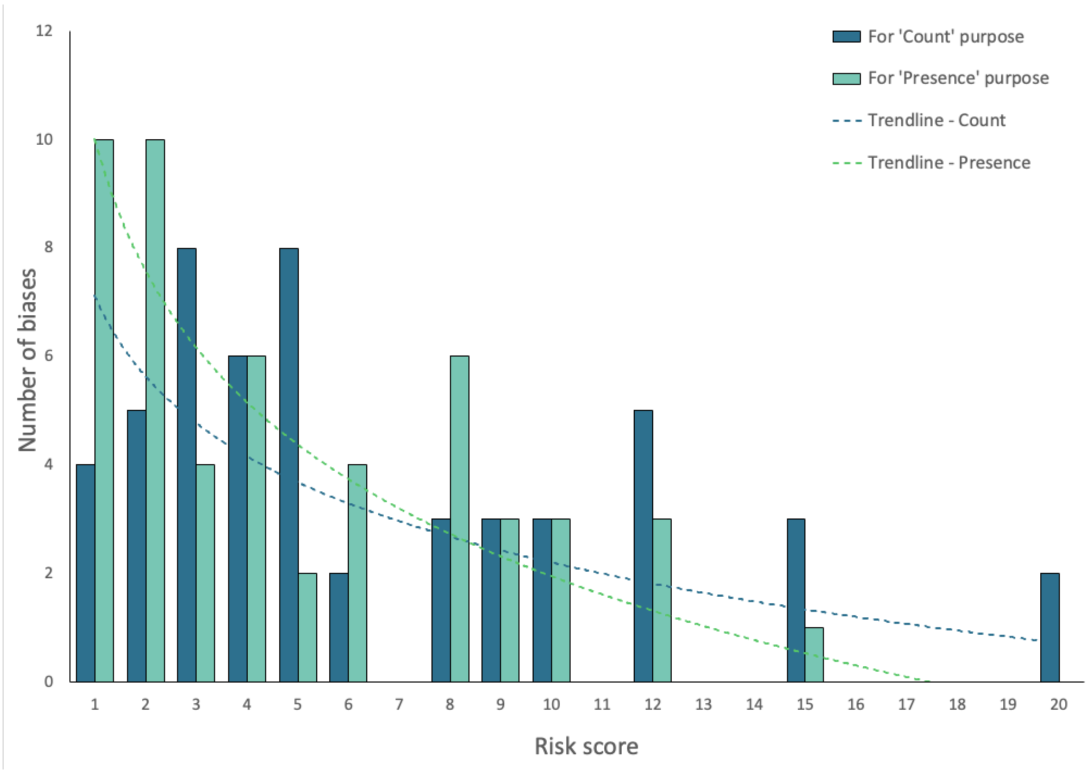


Figure 3.12 Number of biases scored per risk score 1-20 for ‘Count’ versus ‘Presence’ purposes.

3.5 Discussion

Robust sighting records, such as the ESAS and ORCA datasets, can be invaluable tools for revealing spatial and temporal patterns in marine predators which routinely migrate or disperse hundreds or thousands of kilometres (Hays *et al.*, 2013). The datasets were initially chosen due to their wide spatial, temporal and taxonomic coverage (identified in Chapter 2) but act as suitable case studies of much larger number of databases, with interesting differences between the four datasets. Whilst it is important to recognise inherent bias associated with opportunistic datasets, these datasets are important for research questions that require wide geographic scope or revealing large scale patterns (Dickinson *et al.*, 2010) and may be the best data source available. They can be used to create reliable estimates of parameters such as abundance, population trends, or taxonomic biodiversity measures, if the potential for error and bias is considered in relation to the hypothesis being tested (van Strien *et al.*, 2013). This study provides a quantitative method to quantify the level of bias in datasets, which is critical when using historical datasets that collate data from an assortment of different survey designs which may not have targeted the desired research hypothesis.

Using existing diverse data can meet conservation and management needs, where cost and time of new individual surveys may be expensive, particularly when they are standardised like ESAS, ORCA, ORCA and DAS surveys. Spatial or geographic sampling bias is widespread in occurrence datasets (Hijmans *et al.*, 2000, Reddy and Dávalos, 2003, Kadmon *et al.*, 2004, Aiello-Lammens *et al.*, 2015), so patterns of increased density around the UK coast visible in the ESAS and ORCA datasets are not uncommon. The geographic location and intensity of collecting data can be influenced by the accessibility to the area (Reddy and Dávalos, 2003), which affects animal distribution data (Davis *et al.*, 1990, Rondinini *et al.*, 2006, McBride-Kebert *et al.*, 2019).

Spatially biased survey effort is inherent to opportunistic surveys, compared to standardised surveys (such as SCANS-IV or the APEM DAS). In particular for ORCA surveys which utilise ferry or cruise routes, there is little deviation that would expand coverage. Surveys are often concentrated over specific areas and at fixed seasons, and whilst data can be provided at little expense with less time and cost, such effects of spatial bias on opportunistic data is still poorly understood (McBride-Kebert *et al.*, 2019). Furthermore, given the ESAS and ORCA datasets are presence-only datasets, where observations of no animals seen are not included, it cannot be determined if there is a true congregation of predators around the coast of the UK, or whether this is an artefact of survey effort. Systematic surveys such as the SCANS-IV or DAS survey from APEM that record null sightings are less susceptible to this bias.

The multi-year coverage in the ESAS and ORCA datasets is valuable for revealing temporal trends, which targeted focused research surveys (such as SCANS-IV which is a summer survey, or DAS APEM which is a one winter month of survey) may not be able to detect. They may prove to be critical for exposing deviation from baseline trends (Olson *et al.*, 2018), particularly with recent climate and oceanographic changes. Patterns of distributions at monthly or yearly scales are needed in marine ecosystems, particularly for top predators that have often diverse life histories and breeding seasons, so analysing long term datasets that span decades such as ESAS and ORCA could be vital in facilitating this.

The major bias toward summer data collection in three of the datasets was expected given the difficulty of sampling in marine environments during more harsh winter weather conditions. In some cases, such as when birds are absent from breeding colonies due to migrations, these temporal gaps may reflect true biological patterns but generally the lack of winter data will restrict some usage of the datasets to understand seasonal patterns, unless more data can be integrated to fill in the paucity of data collected during winter months. It must be reiterated however, that a low abundance of data during winter months is better than no data at all, and the ESAS and ORCA datasets had some coverage during all months of the year. SCANS-IV was carried out in June to August 2022, whilst APEM DAS was carried out in January 2022.

Birds dominated the ESAS database but it contained all five taxa (birds, cetaceans, pinnipeds, sharks, and turtles), the ORCA dataset contained cetaceans and pinnipeds only, SCANS-IV targeted cetaceans only whilst DAS APEM recorded marine megafauna and birds. Understanding the different contributions of species per datasets is vital for further analysis and understanding potential biases which may drive these contributions. Fulmar and gannets are larger seabird species and are also more aerial in behaviour so may be more easily spotted during surveys. Larger birds that flew highest have been shown to be detected most consistently during at-sea surveys (Spear *et al.*, 2004). Fulmar have been shown to be the most frequently recorded species in other at-sea studies, alongside common guillemots, kittiwakes, puffins and gannets (Pollock *et al.*, 2000). Species that are easier to identify are often more accurately recorded to species level, but smaller species such as storm-petrels, grebes, phalaropes, diving petrels and small alcids may be more inconspicuous and harder to detect or identify at sea (Briggs *et al.*, 1985b, Spear *et al.*, 2004) and are often recorded by taxon.

Some bird species are known ship followers, which may change the availability bias during sighting surveys, positively influencing estimates. Northern gannets are known to associate with fishing vessels (Votier *et al.*, 2013), with behaviour of birds shown to be influenced by boats up to 11 km away (Bodey *et al.*, 2014). Birds have even been shown to differentiate between vessel activity and adjust behaviour accordingly, with gannets switching to foraging behaviour when trawling vessels were fishing and commuting behaviour when vessels were steaming or drifting (Bodey *et al.*, 2014). Similarly, gulls have been shown to be attracted to ships (Camphuysen *et al.*, 1995, Garthe *et al.*, 1999, Pollock *et al.*, 2000, Clarke *et al.*, 2003, Kubetzki and Garthe, 2003, Schwemmer and Garthe, 2005), and gulls and terns have shown a low disturbance vulnerability index (Fliessbach *et al.*, 2019). As such vessels at sea attract many seabirds, even if they are not fish trawlers or providing food directly (Camphuysen *et al.*, 2004), so whilst vessels used in ESAS and ORCA do not include those that catch or discard fish (Pollock *et al.*, 2021), they may indirectly attract ship following species.

Conversely, species such as black guillemots, sooty terns, white terns and albatrosses may show avoidance behaviour responses to vessels (Clarke *et al.*, 2003, Borberg *et al.*, 2005b, Fliessbach *et al.*, 2019), and thus abundance counts become negatively biased. Approaching vessels may present a threatening stimulus to marine birds, with subsequent risk avoidance behaviour reducing the time available for other activities such as feeding, resting, or mating (Gill *et al.*, 1996, Frid and Dill, 2002, Beale and Monaghan, 2004). Observable responses by seabirds include flying off, escape diving, and increased alertness, which can result in loss of energy and opportunities, displacement, and net habitat loss (Béchet *et al.*, 2004, Fliessbach *et al.*, 2019). Aerial surveys (such as SCANS-IV and DAS) are largely immune to attraction or avoidance bias, avoiding the need for adjustment, for both cetaceans (Evans and Hammond, 2004) and birds (Henkel *et al.*, 2007).

Fulmar, gannets, and gulls are also widespread and abundant in the study area, which may account for high sighting rate. Lesser-black-backed gulls (*Larus fuscus*) for example are characteristically offshore in distribution (Schwemmer and Garthe, 2005) so more suited to at-sea sightings as in the four datasets. Black-headed gulls (*Chroicocephalus ridibundus*) are common residents in Britain (Flegg, 1986) whilst Northern fulmars (*Fulmarus glacialis)* and Northern gannets have a wide geographic range, having expanding their range in the North Atlantic and an increasing population trend overall (BirdLife International, 2018a, BirdLife International, 2018b). Similarly, puffins, storm petrels and fulmars are often more abundant in oceanic waters showing a preference for deeper waters, compared to common guillemots, Manx shearwaters *Puffinus puffinus* or razorbills which are found more inshore (Stone *et al.*, 2009).

Porpoises are the most encountered cetacean in the ESAS database, like many other sighting studies in the North Sea (Hammond *et al.*, 1995, Hammond *et al.*, 2002, Ó Cadhla *et al.* 2004*,* Robinson *et al.*, 2007). This may be due to their wide distribution over most of the deep North Atlantic (Teilmann and Sveegaard, 2019) and being the most abundant cetacean in UK waters (Evans and Prior, 2012), with estimates of 345,501 harbour porpoise in the North Sea Management Unit (MU) alone (with 28,936 in the West Scotland MU and 62,517 in the Celtic and Irish Sea MU), as defined by the Inter-Agency Marine Mammal Working Group (IAMMWG, 2021). In SCANS-IV surveys which were carried out to provide robust large-scale estimates of cetacean abundance, harbour porpoise was the most detected species in both aerial and ship surveys (Gilles *et al.*, 2023). The IAMMWG determined the next most common species around the UK to be bottlenose dolphin *Tursiops truncatus*, short-beaked common dolphin, white-beaked dolphin, Atlantic white-sided dolphin *Lagenorhynchus acutus*, Risso's dolphin *Grampus griseus* and minke whale (IAMMWG, 2021). Short-beaked common dolphin were highest in the ORCA surveys and second highest in sightings in ESAS data, matching the SCANS-IV report which demonstrated high numbers off the coast of Portugal, Spain and France (Gilles *et al.*, 2023), and this is where ORCA surveys often traversed. Colourisation of the common dolphin also makes it easy to identify both above and beneath the water, likely reducing perception bias. Sightings of fin whale *Balaenoptera physalus* were also abundant in both datasets, but likely as fin whales have multiple North Atlantic feeding grounds, including west Iceland and northwest Spain which both datasets transverse (Borrell *et al.*, 2018). Species which were vagrant around the UK (North Atlantic right whale, false killer whale *Pseudorca crassidens*) (Evans and Hammond, 2004) were expectedly low in sighting numbers.

Minke whale was commonly detected in ORCA, ESAS and the SCANS-IV survey and have been observed all year round in UK waters and the North Sea (Northridge *et al.*, 1995, Weir and Stockin, 2001, Dolman *et al.*, 2013), with 20,118 estimated in the Celtic and Greater North Seas (CGNS) MU. Minke whale are observed around the UK mainly from April to October (Reid *et al.*, 2003, Macleod *et al.*, 2004), but lack of winter survey effort means knowledge of winter distribution, feeding grounds and occurrence is limited. Only one right whale was observed in the ESAS dataset, with none in ORCA, and given their endangered status and global population of 200-250 mature individuals (Cooke, 2020), it is anticipated sightings are low. Other species such as orca, rough-toothed dolphin and beluga were low in sightings within the databases, but these species have been recorded rarely in UK waters (Reid *et al.*, 2003).

Lower pinniped species sightings compared to seabirds may be due to the primarily at-sea survey location of datasets , which may not lend itself to species that haul out or central place foragers, leading to geographic sampling bias. Grey seals were still abundant in the ESAS and ORCA datasets, but harbour seal was highest in the ESAS dataset. The high number of grey seal sightings in both databases is expected, as the larger of the two UK seal species for ease of identification and wide distribution. Recent population models provide UK estimates of 157,300 grey seals (approximate 95% CI 44,600-169,400) in 2019 (SCOS, 2021). Though grey seals tend to travel further from haul-out sites than harbour seals (Thompson *et al.*, 1996, McConnell *et al.*, 2017), recent telemetry studies have suggested grey seals may forage up to 448 km and harbour seals up to 273 km from haul-out sites (Carter *et al.*, 2022), so both are likely to be observed further offshore where boats are surveying. Harbour seals were the most sighted pinniped in the ESAS database, and the UK harbour seal population estimates are 43,750 approximately (95% CI: 35,800-58,300) (SCOS, 2021) thus both species are likely to be encountered during surveys of waters around the UK.

Basking shark was frequent in ESAS sightings and are found in all UK coastal waters during summer months, with regional hotspots in western Scotland, Isle of Man and southwest England have been identified (Southall *et al.*, 2005, Witt *et al.*, 2012) and survey effort was high in these areas for ESAS, ORCA and SCANS-IV surveys. There are also high numbers of blue shark *Prionace glauca*, and though having a global distribution (Nakano and Seki, 2003), there is evidence of a single stock of blue sharks in the North Atlantic Ocean, and the oceanic predator often undergoes long-distance trans-Atlantic migrations which are frequently observed in the tagging programs, thus potentially enhancing chances of sightings in offshore waters of the North Atlantic (Kohler *et al.*, 1998).

Whilst six species of marine turtle have been recorded in UK waters (Witt *et al.*, 2007b, Doyle *et al.*, 2008), just two were recorded in the ESAS database and are likely stray juveniles carried by currents from warmer seas (Witt *et al.*, 2007b). Past analysis of sighting and stranding data occurred in the western aspects of the UK and records decreased with increasing latitude, suggesting juvenile turtles arrive in these waters via the North Atlantic current systems (Botterell *et al.*, 2020). Records of marine turtles have significantly increased over time according to an analysis of the spatial and temporal occurrence of marine turtle sightings and strandings in the UK and Ireland between 1910 and 2018 (Botterell *et al.*, 2020) but records for some species, such as green turtle, remain very sparse, aligning with the low numbers in the ESAS database. The migratory behaviour, shorter juvenile oceanic phase and small size of juvenile marine turtles may make monitoring them in the open ocean difficult (Mansfield *et al.*, 2014), leading to negative availbility bias. Leatherback sea turtles *Dermochelys coriacea* were not recorded in any of the datasets, even though they migrate seasonally to UK waters to feed on jellyfish (Witt *et al.*, 2007a). Studies suggest sightings, strandings and capture of leatherbacks around the UK and Ireland have declined since the 1990s (Botterell *et al.*, 2020), with reduced reporting effort, and most leatherback sightings and captures were reported by fishers, which are not included in the ESAS database.

This study highlighted that the use of a novel matrix approach to quantify the level of bias in historical data can aid in analysing existing surveys and guiding research questions and future survey design. Whilst surveys should specify whether they are estimating absolute abundance or relative abundance indices for detecting trends

(Dawson *et al.*, 2008) ideally before data collection, this is not always possible using a heterogeneous historical dataset such as ESAS or ORCA post-survey (unlike systematic surveys designed for a specific research question such as SCANS-IV or DAS APEM). This study highlights that important research questions can still be answered using large-scale historical datasets if the biases are recognised and accounted for. Despite inherent biases highlighted in occurrence datasets, these temporally and spatially large datasets can provide valuable information on marine predators for questions or hypotheses that are relative in nature - such as associations, trends, residency patterns, and behaviours (McBride-Kebert *et al.*, 2019). This study showed that using the two example datasets for identifying the 'presence' of a species in an area gave a lower risk total risk score overall, and a lower number of 'high-risk scores' per bias, compared to using the datasets for identifying exact abundance (count) of animals in an area. Focusing on overlaps, similarities, and agreements between species' space use, rather than differences or gaps in distributions is practical. If data are limited temporally due to survey effort, or spatially due to accessibility, then truthful patterns may be hidden or influenced so it is important to highlight limitations from the onset. The absence of evidence is not evidence of absence, particularly with marine species with their highly cryptic lifestyles, and therefore the focus could be on species observed and areas of shared space use, rather than numbers or counts or predicting distributions.

The matrix methodology in this study has two important roles. The first is to aid people in analysing existing available datasets and standardising data submissions to encourage authors to highlight the boundaries of their data. Deciding scores on severity and likelihood are subjective, but still provides a tool for helping users to assess and present levels of risk concisely and visually. Data providers could quantify the level of biases in datasets submitted to public portals or organisations, and thus end users have information on the biases within the dataset. At present, portal data often has 'readme' or metadata documents with caveats in acknowledging that there is bias in the data but does not explicitly quantify the level. The matrix approach could be used to flag up issues with data, such as ageing data, the level of standardisation in data collection (if at all), the level of expertise etc. Some marine species datasets have confidence scales within them, allowing users to filter datasets on that bias or give quality control to surveys. For example, Sea Watch Foundation (SWF) asks volunteers to categorise sightings in terms of confidence of their sightings as “Definite, Probably, Possible” (Sea Watch Foundation, 2019). This is due to a high heterogeneity of observers, with most systematic observations at SWF being from public observers with varying experience submitting casual records, and only a small number of sightings from more experienced observers (Evans and Hammond, 2004).

Use of the matrix approach will maximise data-use, allowing end-users to confidently utilise widely available historical datasets that quantify biases within them. It raises awareness for end users about the potential risks of using the dataset, aiding in quick, transparent decision making when evaluating data sources. Virtually any industry or profession which relies on having robust datasets can benefit from a bias matrix approach.

End-users such as government organisations and Statutory Nature Conservation Bodies (SNCBs) are often limited in budget and time, so this matrix approach could aid in speeding up the time spend searching for suitable data sources. Accessing bias matrices can enable users to spend less time searching themselves and encourage more informed narrowing of data searches. It can enable such organisations to allocate resources effectively and prioritise use of datasets with lower risk scores. Similarly, for other NGOs, such as those involved in offshore development, such as renewable energy projects or oil exploration who produce detailed environmental impact assessments, often require robust scientific justification to present to stakeholders in their decisions in applications for consent, this could aid in providing justification for their chosen data sources and subsequent assessment.

The risk matrix provides a structured approach to identifying, evaluating, and if required mitigating potential risks (such as adjusting for detection probability as in SCANS-IV analyses) that can be easily repeated and by any user, ultimately contributing to more reliable, secure, and compliant data analyses. A standardised method of assessing the level of bias inherent in datasets facilitates comparability across sources, between organisations and across time periods. This therefore fosters collaboration by providing a common structure, that researchers and professionals from many different organisations can access quickly, interpret easily, and can be repeated over and over.

The other important role of this matrix method is to allow researchers to design better surveys by understanding which factors are most important to their study, and thus, which biases are of high risk to their research question. Even heavily biased datasets might be used to answer relative questions, where absolute questions cannot be answered, and still provide valuable contribution to the field of research. Bull *et al.* (2013) for example highlighted that volunteer data of benthic species records can still provide a valuable contribution to the understanding of biodiversity, climate change and aquatic ecosystem health, particularly when baseline data is lacking. For marine predators, obtaining baseline data is difficult but often needed in marine management and conservation, particularly with rapidly increasing offshore renewable energy developments around the UK. The bias matrix review suggests that answering relative questions may still be better than no insight at all, as long as data is inspected and evaluated in terms of bias alongside, and inferences are not made without considering how surveys have impacted data collection. Critically, this study has demonstrated how to design better surveys to minimise bias and given users a methodology which establishes which research questions existing datasets are most suitable for.

3.6 Conclusion

This chapter quantitatively compared publicly available datasets in terms of survey method, spatial, temporal, and taxonomic coverage, and reviewed the biases present in such occurrence datasets. This study presents a novel method to quantify biases within four pilot datasets using a risk assessment approach, and tests how the risk assessment matrix approach can guide the use of large-scale observation databases both pre-survey (by designing better surveys to reduce bias) and post-survey (by establishing which research questions are best suited to the data). Overall risk of using historical datasets (where the true level of bias is difficult to quantify post-survey) is reduced if answering relative research questions, rather than trying to estimate, for example, absolute abundance or counts. If the matrix method is used appropriately, it may allow the utilisation of large low-cost historical datasets by a wide variety of end users in meeting ecological and conservation objectives that require wide temporal and spatial coverage, such as understanding space use in cryptic marine predators.

3.7 References

AIELLO-LAMMENS, M. E., BORIA, R. A., RADOSAVLJEVIC, A., VILELA, B. & ANDERSON, R. P. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography,* 38**,** 541-545.

AINLEY, D. G., DUGGER, K. D., FORD, R. G., PIERCE, S. D., REESE, D. C., BRODEUR, R. D., TYNAN, C. T. & BARTH, J. A. 2009. Association of predators and prey at frontal features in the California Current: Competition, facilitation, and co-occurrence. *Marine Ecology Progress Series,* 389**,** 271–294.

AINLEY, D. G., RIBIC, C. A. & WOEHLER, E. J. 2012. Adding the ocean to the study of seabirds: a brief history of at-sea seabird research. *Marine Ecology Progress Series,* 451**,** 231-243.

APEM LTD. 2023. *POSEIDON, Celtic Sea seabird and marine mammal surveys* [Online]. https://www.marinedataexchange.co.uk/details/TCE-3789/2023-apem-poseidon-celtic-sea-seabird-and-marine-mammal-surveys/packages/11234?directory=%2F. [Accessed 3rd December 2023].

BAN, N. C., BODTKER, K. M., NICOLSON, D., ROBB, C. K., ROYLE, K. & SHORT, C. 2013. Setting the stage for marine spatial planning: Ecological and social data collation and analyses in Canada's Pacific waters. *Marine Policy,* 39**,** 11-20.

BARBRAUD, C. & THIEBOT, J.-B. 2009. On the importance of estimating detection probabilities from at-sea surveys of flying seabirds. *Journal of Avian Biology,* 40**,** 584-590.

BEALE, C. M. & MONAGHAN, P. 2004. Human disturbance: people as predation-free predators? *Journal of Applied Ecology,* 41**,** 335-343.

BÉCHET, A., GIROUX, J. F. & GAUTHIER, G. 2004. The effects of disturbance on behaviour, habitat use and energy of spring staging snow geese. *Journal of Applied Ecology,* 41**,** 689-700.

BEMMELEN, R. V., GEELHOED, S. C. V. & LEOPOLD, M. 2011. Ship-based monitoring of seabirds and cetaceans. Netherlands: IMARES - Institute for Marine Resources & Ecosystem Studies.

BERROW, S. 2001. Biological diversity of cetaceans (whales, dolphins and porpoises) in Irish waters. *Biodiversity in Ireland and adjacent waters***,** 115-120.

BIRDLIFE INTERNATIONAL. 2018a. *Fulmarus glacialis* [Online]. IUCN. Available: https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22697866A132609419.en [Accessed 20th February 2020].

BIRDLIFE INTERNATIONAL. 2018b. *Morus bassanus* [Online]. IUCN. Available: https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22696657A132587285.en. [Accessed 20th February 2020].

BLOCK, B. A., JONSEN, I. D., JORGENSEN, S. J., WINSHIP, A. J., SHAFFER, S. A., BOGRAD, S. J., HAZEN, E. L., FOLEY, D. G., BREED, G. A., HARRISON, A. L., GANONG, J. E., SWITHENBANK, A., CASTLETON, M., DEWAR, H., MATE, B. R., SHILLINGER, G. L., SCHAEFER, K. M., BENSON, S. R., WEISE, M. J., HENRY, R. W. & COSTA, D. P. 2011. Tracking apex marine predator movements in a dynamic ocean. *Nature,* 475**,** 86-90.

BOAKES, E. H., MCGOWAN, P. J., FULLER, R. A., CHANG-QING, D., CLARK, N. E., O'CONNOR, K. & MACE, G. M. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol,* 8**,** e1000385.

BODEY, T. W., JESSOPP, M. J., VOTIER, S. C., GERRITSEN, H. D., CLEASBY, I. R., HAMER, K. C., PATRICK, S. C., WAKEFIELD, E. D. & BEARHOP, S. 2014. Seabird movement reveals the ecological footprint of fishing vessels. *Curr Biol,* 24**,** R514-5.

BORBERG, J. M., BALLANCE, L. T., PITCHER, R. L. & AINLEY, D. G. 2005a. A test for bias attributable to seabird avoidance of ships during surveys conducted in the tropical pacific. *Marine Ornithology,* 33**,** 173-179.

BORBERG, J. M., BALLANCE, L. T., PITMAN, R. L. & AINLEY, D. G. 2005b. A test for bias due to seabird avoidance of ships when conducting surveys in the tropical Pacific. *Marine Ornithology,* 33**,** 173-179.

BORRELL, A., SAIZ, L., VIKINGSSON, G. A., GAUFIER, P., LOPEZ FERNANDEZ, A. & AGUILAR, A. 2018. Fin whales as bioindicators of multi-decadal change in carbon and oxygen stable isotope shifts in the North Atlantic. *Mar Environ Res,* 138**,** 129-134.

BOTTERELL, Z. L. R., PENROSE, R., WITT, M. J. & GODLEY, B. J. 2020. Long-term insights into marine turtle sightings, strandings and captures around the UK and Ireland (1910–2018). *Journal of the Marine Biological Association of the United Kingdom,* 100**,** 869-877.

BOYD, I. L., BOWEN, W. D. & IVERSON, S. J. 2010. *Marine Mammal Ecology and Conservation: A Handbook of Techniques,* Oxford, Oxford University Press.

BRIGGS, K. T., TYLER, W. B. & LEWIS, D. B. 1985a. Aerial surveys for seabirds: Methodological experiments. *The Journal of Wildlife Management,* 49**,** 412–417.

BRIGGS, K. T., TYLER, W. B. & LEWIS, D. B. 1985b. Comparison of ship and aerial surveys of birds at sea. *Journal of wildlife management***,** 405-411.

BUCKLAND, S., ANDERSON, D., BURNHAM, K., LAAKE, J., BORCHERS, D. & THOMAS, L. 2001. *Introduction to distance sampling: estimating abundance of biological populations.,* Oxford, Oxford University Press.

BUCKLAND, S., REXSTED, E., MARQUES, T. & OEDEKOVEN, C. 2015. *Distance sampling methods and applications.,* New York, NY, Springer.

BUCKLAND, S. T., BURT, M. L., REXSTAD, E. A., MELLOR, M., WILLIAMS, A. E. & WOODWARD, R. 2012. Aerial surveys of seabirds: The advent of digital methods. *Journal of Applied Ecology,* 49**,** 960–967.

BULL, J. C., MASON, S., WOOD, C. & PRICE, A. R. G. 2013. Benthic marine biodiversity patterns across the United Kingdom and Ireland determined from recreational diver observations: A baseline for possible species range shifts induced by climate change. *Aquatic Ecosystem Health & Management,* 16**,** 20-30.

BUSCH, M., KANNEN, A., GARTHE, S. & JESSOPP, M. 2013. Consequences of a cumulative perspective on marine environmental impacts: Offshore wind farming and seabirds at North Sea scale in context of the EU Marine Strategy Framework Directive. *Ocean & Coastal Management,* 71**,** 213-224.

CAMPHUYSEN, C. J., CALVO, B., DURINCK, J., ENSOR, K., FOLLESTAD, A., FURNESS, R. W., S., G., LEAPER, G., SKOV, F., TASKER, M. L. & WINTER, C. J. N. 1995. Consumption of discards by seabirds in the North Sea. . *Final report to the European Comm., study contr. BIOECO/93/lO. NIOZ-Report 1995-5.* Texel: Netherlands Institute for Sea Research.

CAMPHUYSEN, K., FOX, A., LEOPOLD, M. & PETERSEN, I. 2004. Towards standardised seabirds at sea census techniques in connection with environmental impact assessments for offshore wind farms in the UK. *Report to COWRIE (BAM-02–2002).* Texel: NIOZ.

CAMPHUYSEN, K. C. & GARTHE, S. 2004. Recording foraging seabirds at sea standardised recording and coding of foraging behaviour and multi-species foraging associations. *Atlantic Seabirds,,* 6**,** 1-32.

CAÑADAS, A., DONOVAN, G. P., DESPORTES, G. & BORCHERS, D. L. 2009. A short review on the distribution of short-beaked common dolphins (*Delphinus delphis*) in the central and eastern North Atlantic with an abundance estimate for part of this area. . *North Atlantic Marine Mammal Commission (NAMMCO) Scientific Publications* 7**,** 201-220.

CARTER, M. I. D., BOEHME, L., CRONIN, M. A., DUCK, C. D., GRECIAN, W. J., HASTIE, G. D., JESSOPP, M., MATTHIOPOULOS, J., MCCONNELL, B. J., MILLER, D. L., MORRIS, C. D., MOSS, S. E. W., THOMPSON, D., THOMPSON, P. M. & RUSSELL, D. J. F. 2022. Sympatric Seals, Satellite Tracking and Protected Areas: Habitat-Based Distribution Estimates for Conservation and Management. *Frontiers in Marine Science,* 9.

CERTAIN, G. & BRETAGNOLLE, V. 2008. Monitoring seabirds population in marine ecosystem: The use of strip-transect aerial surveys. *Remote Sensing of Environment,* 112**,** 3314-3322.

CLARKE, E., SPEAR, L., MCCRACKEN, M., MARQUES, F., BORCHERS, D., BUCKLAND, S. & AINLE, Y. D. 2003. Validating the use of generalized additive models and at-sea surveys to estimate size and temporal trends of seabird populations. *Journal of Applied Ecology,* 40**,** 278-292.

CLARKE, J. M., WHETTON, P. H. & HENNESY, K. J. Providing application-specific climate projections datasets: CSIRO's Climate Futures Framework. 19th International Congress on Modelling and Simulation, 2011 Perth, Australia. CSIRO.

COLWELL, R. K. & CODDINGTON, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society A,* B345**,** 101–118.

COOKE, J. G. 2020. *Eubalaena glacialis (errata version published in 2020)* [Online]. Available: https://www.iucnredlist.org/species/41712/178589687 [Accessed 05 April 2022].

COX, L. A., JR. 2008. What's wrong with risk matrices? *Risk Analysis,* 28**,** 497-512.

DAVIS, F. W., STOMS, D. M., ESTES, J. E., SCEPAN, J. & SCOTT, J. M. 1990. An information systems approach to the preservation of biological diversity. *International Journal of Geographical Information System,* 4**,** 55-78.

DAWSON, S., WADE, P., SLOOTEN, E. & BARLOW, J. 2008. Design and field methods for sighting surveys of cetaceans in coastal and riverine habitats. *Mammal Review,* 38**,** 19-49.

DÍAZ LÓPEZ, B. & METHION, S. 2018. Does interspecific competition drive patterns of habitat use and relative density in harbour porpoises? *Marine Biology,* 165.

DICKINSON, J. L., ZUCKERBERG, B. & BONTER, D. N. 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics,* 41**,** 149-172.

DOLMAN, S. J., HODGINS, N. K., MACLEOD, C. D., PIERCE, G. J. & WEIR, C. R. 2013. Harbour porpoises (*Phocoena phocoena*) and minke whales (*Balaenoptera acutorostrata*) observed during land-based surveys in The Minch, north-west Scotland. *Journal of the Marine Biological Association of the United Kingdom,* 94**,** 1185-1194.

DOYLE, T. K., HOUGHTON, J. D., O'SÚILLEABHÁIN, P. F., HOBSON, V. J., MARNELL, F., DAVENPORT, J. & HAYS, G. C. 2008. Leatherback turtles satellite-tagged in European waters. *Endangered Species Research,* 4**,** 23-31.

DUCKWORTH, J. W. 1997. Correcting avian richness estimates for unequal sample effort in atlas studies. *Ibis,* 139**,** 189– 192.

DUIJM, N. J. 2015. Recommendations on the use and design of risk matrices. *Safety Science,* 76**,** 21-31.

ESRI 2022. ArcGIS Pro. Redlands, CA: Environmental Systems Research Institute: ESRI.

EVANS, P. G. H. & HAMMOND, P. S. 2004. Monitoring cetaceans in European waters. *Mammal Review,* 34**,** 131-156.

EVANS, P. G. H. & PRIOR, J. S. 2012. Protecting the harbour porpoise in UK seas - Identifying a network of draft SACs for the harbour porpoise in the UK *A report to WWF UK.*

FLEGG, J. J. M. 1986. Black-headed gull. *In:* LACK, P. (ed.) *The Atlas of wintering birds in Britain and Ireland. .* Calton: T. & AD Poyser.

FLIESSBACH, K. L., BORKENHAGEN, K., GUSE, N., MARKONES, N., SCHWEMMER, P. & GARTHE, S. 2019. A ship traffic disturbance vulnerability index for Northwest European seabirds as a tool for marine spatial planning. *Frontiers in Marine Science,* 6.

FRID, A. & DILL, L. 2002. Human-caused disturbance stimuli as a form of predation risk. *Conservation Ecology,* 6**,** 11.

GARTHE, S., FREYER, T., HÜPPOP, O. & WÖLKE, D. 1999. Breeding lesser black-backed gulls *Larus graellsii* and herring gulls *Larus argentatus:* coexistence or competition? *Ardea,* 87**,** 227-236.

GILL, J. A., SUTHERLAND, W. J. & WATKINSON, A. R. 1996. Approaches to measuring the effects of human disturbance on birds: measuring the effects of human disturbance on birds. *Ibis,* 148**,** 9-14.

GILLES, A., AUTHIER, M., RAMIREZ-MARTINEZ, N.C., ARAUJO, H., BLANCHARD, A., CARLSTROM, J., EIRA, C., DORÉMUS, G., FERNANDEZ-MALDONAD, C., GEELHOED, S.C.V. AND KYHN, L. 2023. Estimates of cetacean abundance in European Atlantic waters in summer 2022 from the SCANS-IV aerial and shipboard surveys. Hannover: University of Veterinary Medicine.

GLENNIE, R., BUCKLAND, S. T. & THOMAS, L. 2015. The effect of animal movement on line transect estimates of abundance. *PLoS One,* 10**,** e0121333.

GRANHOLM, S. 1983. Bias in density estimations due to movement of birds. *Condor,* 85**,** 243–248.

GRECIAN, W. J., WITT, M. J., ATTRILL, M. J., BEARHOP, S., BECKER, P. H., EGEVANG, C., FURNESS, R. W., GODLEY, B. J., GONZALEZ-SOLIS, J., GREMILLET, D., KOPP, M., LESCROEL, A., MATTHIOPOULOS, J., PATRICK, S. C., PETER, H. U., PHILLIPS, R. A., STENHOUSE, I. J. & VOTIER, S. C. 2016. Seabird diversity hotspot linked to ocean productivity in the Canary Current Large Marine Ecosystem. *Biol Lett,* 12.

GU, W. & SWIHART, R. K. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation,* 116**,** 195-203.

HAMMOND, P. S., BENKE, H., BREGGREN, P., COLLET, A., HEIDE-JØRGENSEN, M. P., HEIMLICH-BORAN, S., LEOPOLD, M. & ØIEN, N. 1995. The distribution and abundance of harbour porpoises and other small cetaceans in the North Sea and adjacent waters. . *ICES.*

HAMMOND, P. S., BERGGREN, P., BENKE, H., BORCHERS, D. L., COLLET, A., HEIDE-JØRGENSEN, M. P., HEIMLICH, S., HIBY, A. R., LEOPOLD, M. F. & ØIEN, N. 2002. Abundance of harbour porpoises and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology,* 39**,** 361-376.

HAMMOND, P. S., LACEY, C., GILLES, A., VIQUERAT, S., BÖRJESSON, P., HERR, H., MACLEOD, K., RIDOUX, V., SANTOS, M. B., SCHEIDAT, M., TEILMANN, J., VINGADA, J. V. & ØIEN, N. 2021. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. Wageningen Marine Research.

HAMMOND, P. S., MACLEOD, K., BERGGREN, P., BORCHERS, D. L., BURT, L., CAÑADAS, A., DESPORTES, G., DONOVAN, G. P., GILLES, A., GILLESPIE, D., GORDON, J., HIBY, L., KUKLIK, I., LEAPER, R., LEHNERT, K., LEOPOLD, M., LOVELL, P., ØIEN, N., PAXTON, C. G. M., RIDOUX, V., ROGAN, E., SAMARRA, F., SCHEIDAT, M., SEQUEIRA, M., SIEBERT, U., SKOV, H., SWIFT, R., TASKER, M. L., TEILMANN, J., VAN CANNEYT, O. & VÁZQUEZ, J. A. 2013. Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation,* 164**,** 107-122.

HARVEY, G. K. A., NELSON, T. A., PAQUET, P. C., FERSTER, C. J. & FOX, C. H. 2018. Comparing citizen science reports and systematic surveys of marine mammal distributions and densities. *Biological Conservation,* 226**,** 92-100.

HARWOOD, L. A., MCLAUGHLIN, F., ALLEN, R. M., ILLASIAK, J. & ALIKAMIK, J. 2004. First-ever marine mammal and bird observations in the deep Canada Basin and Beaufort/Chukchi seas: expeditions during 2002. *Polar Biology,* 28**,** 250-253.

HAYS, G. C., SCOTT, R. & HIGHAM, T. 2013. Global patterns for upper ceilings on migration distance in sea turtles and comparisons with fish, birds and mammals. *Functional Ecology,* 27**,** 748-756.

HENKEL, L. A., FORD, R. G., TYLER, W. B. & DAVIS, J. N. 2007. Comparison of aerial and boat-based survey methods for marbled murrelets *Brachyramphus marmoratus* and other marine birds. *Marine Ornithology,* 35**,** 145–151.

HERR, H., SCHEIDAT, M., LEHNERT, K. & SIEBERT, U. 2009. Seals at sea: modelling seal distribution in the German bight based on aerial survey data. *Marine Biology,* 156**,** 811.

HIJMANS, R. J., GARRETT, K. A., HUAMAN, Z., ZHANG, D. P., SCHREUDER, M. & BONIERBALE, M. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology,* 14**,** 1755–1765.

HIMES BOOR, G. K. & SMALL, R. J. 2012. Steller sea lion spatial-use patterns derived from a Bayesian model of opportunistic observations. *Marine Mammal Science,* 28**,** E375–E403.

HOBBS, R. C. & WAITE, J. M. 2010. Abundance of harbor porpoise (*Phocoena phocoena*) in three Alaskan regions, corrected for observer errors due to perception bias and species misidentification, and corrected for animals submerged from view. *Fishery Bulletin,* 108**,** 251-267.

HODGSON, J. C., BAYLIS, S. M., MOTT, R., HERROD, A. & CLARKE, R. H. 2016. Precision wildlife monitoring using unmanned aerial vehicles. *Scientific Reports,* 6**,** 22574.

HUGHES, A. C., ORR, M. C., MA, K., COSTELLO, M. J., WALLER, J., PROVOOST, P., YANG, Q., ZHU, C. & QIAO, H. 2021. Sampling biases shape our view of the natural world. *Ecography,* 44**,** 1259-1269.

HUGHES, K. A., WALUDA, C. M., STONE, R. E., RIDOUT, M. S. & SHEARS, J. R. 2008. Short-term responses of king penguins *Aptenodytes patagonicus* to helicopter disturbance at South Georgia. *Polar Biology,* 31**,** 1521-1530.

HUNT, J., G.L. 1990. Occurrence of polar seabirds at sea in relation to prey concentrations and oceanographic factors. *Polar Research,* 10**,** 553-559.

HUNT JR, G. L., RENNER, M., KULETZ, K. J., SALO, S., EISNER, L., RESSLER, P. H., LADD, C. & SANTORA, J. A. 2018. Timing of sea-ice retreat affects the distribution of seabirds and their prey in the southeastern Bering Sea. *Marine Ecology Progress Series,* 593**,** 209-230.

HYRENBACH, K. D. 2001. Albatross response to survey vessels: implications for studies of the distribution, abundance, and prey consumption of seabird populations. *Marine Ecology Progress Series,* 212**,** 283–295.

IAMMWG 2021. Updated abundance estimates for cetacean Management Units in UK waters. *JNCC Report 680.* Peterborough: JNCC.

JEFFERSON, T. A. & LYNN, S. K. 1994. Marine mammal sightings in the Caribbean Sea and Gulf of Mexico, Summer 1991. *Caribbean Journal of Science,* 30**,** 83-99.

JNCC 2019. European Seabirds at Sea database (ESAS). *In:* JOINT NATURE CONSERVATION COMMITTEE (JNCC) (ed.). Peterborough, UK.

JNCC 2020. Volunteer Seabirds At Sea Survey Methods. *In:* JNCC (ed.). Peterborough: JNCC,.

KADMON, R., FARBER, O. & DANIN, A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications,* 14**,** 401-413.

KASCHNER, K., QUICK, N. J., JEWELL, R., WILLIAMS, R. & HARRIS, C. M. 2012. Global coverage of cetacean line-transect surveys: status quo, data gaps and future challenges. *PLoS One,* 7**,** e44075.

KISZKA, J., MACLEOD, K., VAN CANNEYT, O., WALKER, D. & RIDOUX, V. 2007. Distribution, encounter rates, and habitat characteristics of toothed cetaceans in the Bay of Biscay and adjacent waters from platform-of-opportunity data. *ICES Journal of Marine Science,* 64**,** 1033–1043.

KOHLER, N. E., CASEY, J. G. & TURNER, P. A. 1998. NMFS Cooperative Shark Tagging Program, 1962–93: an atlas of shark tag and recapture data. *Marine Fisheries Review,* 60**,** 1-87.

KUBETZKI, U. & GARTHE, S. 2003. Distribution, diet and habitat selection by four sympatrically breeding gull species in the south-eastern North Sea. *Marine Biology,* 143**,** 199-207.

LENNERT-CODY, C. E., BUCKLAND, S. T., GERRODETTE, T., WEBB, A., BARLOW, J., FRETWELL, P. T., MAUNDER, M. N., KITAKADO, T., MOORE, J. E., SCOTT, M. D. & SKAUG, H. J. 2018. Review of potential line-transect methodologies for estimating abundance of dolphin stocks in the eastern tropical Pacific. *Journal of Cetacean Resource Management,* 19**,** 9-21.

LEONARD, D. R., ROBERT CLARKE, K., SOMERFIELD, P. J. & WARWICK, R. M. 2006. The application of an indicator based on taxonomic distinctness for UK marine biodiversity assessments. *J Environ Manage,* 78**,** 52-62.

MACLEOD, K., FAIRBAIRNS, R., GILL, A., FAIRBAIRNS, B., GORDON, J., BLAIR-MYERS, C. & PARSONS, E. 2004. Seasonal distribution of minke whales *Balaenoptera acutorostrat*a in relation to physiography and prey off the Isle of Mull, Scotland. *Marine Ecology Progress Series,* 277**,** 263–274.

MANSFIELD, K. L., WYNEKEN, J., PORTER, W. P. & LUO, J. 2014. First satellite tracks of neonate sea turtles redefine the 'lost years' oceanic niche. *Proc Biol Sci,* 281**,** 20133039.

MARSDEN, S. J. 1991. Estimation of parrot and hornbill densities using a point count distance sampling method. *Ibis,* 141**,** 327-390.

MARSH, H. & SINCLAIR, D. F. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *Journal of Wildlife Management* 53**,** 1017−1024.

MARTINEZ, E., ORAMS, M. B. & STOCKIN, K. A. 2010. Swimming with an Endemic and Endangered Species: Effects of Tourism on Hector's Dolphins In Akaroa Harbour, New Zealand. *Tourism Review International,* 14**,** 99-115.

MATEAR, L., ROBBINS, J. R., HALE, M. & POTTS, J. 2019. Cetacean biodiversity in the Bay of Biscay: Suggestions for environmental protection derived from citizen science data. *Marine Policy,* 109**,** 103672.

MCBRIDE-KEBERT, S., TAYLOR, J. S., LYN, H., MOORE, F. R., SACCO, D. F., KAR, B. & KUCZAJ II, S. A. 2019. Controlling for survey effort is worth the effort: comparing Bottlenose dolphin (*Tursiops truncatus*) habitat use between standardized and opportunistic photographic-identification surveys. *Aquatic Mammals,* 45**,** 21-29.

MCCONNELL, B., SMOUT, S. & WU, M. 2017. Modelling Harbour Seal Movements. *Scottish Marine and Freshwater Science,* 8**,** 1-33.

MICROSOFT CORPORATION 2018. Microsoft Excel. 16.70 ed.

MONK, J. 2014. How long should we ignore imperfect detection of species in the marine environment when modelling their distribution? *Fish and Fisheries,* 15**,** 352-358.

NAKANO, H. & SEKI, M. P. 2003. Synopsis of biological data on the blue shark, *Prionace glauca* Linnaeus. *Bulletin of Fisheries Research Agency,* 6**,** 8–55.

NICHOLS, J. D., HINES, J. E., SAUER, J. R., FALLON, W. F., FALLON, J. E. & HEGLUND, P. J. 2000. A double-observer approach for estimating detection probability and abundance from point counts. *The Auk,* 117**,** 393-408.

NORTHRIDGE, S. P., TASKER, M. L., WEBB, A. & WILLIAMS, J. M. 1995. Distribution and relative abundance of harbour porpoises (*Phocoena phocoena* L.), white-beaked dolphins (*Lagenorhynchus albirostris* Gray), and minke whales (*Balaenoptera acutorostrata* Lacepède) around the British Isles. *ICES Journal of Marine Science,* 52**,** 55-66.

Ó CADHLA, O., MACKEY, M., AGUILAR DE SOTO, N., ROGAN, E. & CONNOLLY, N. 2004. Volume II – Cetacean distribution & abundance. *Cetaceans and Seabirds of Ireland's Atlantic Margin. Report on research carried out under the Irish Infrastructure Programme (PIP): Rockall Studies Group (RSG) projects 98/6 and 00/13, Porcupine Studies Group project P00/15 and Offshore Support Group (OSG) project 99/38.*

OLSON, J. K., WOOD, J., OSBORNE, R. W., BARRETT-LENNARD, L. & LARSON, S. 2018. Sightings of southern resident killer whales in the Salish Sea 1976-2014: the importance of a long-term opportunistic dataset. *Endangered Species Research,* 37**,** 105-118.

ORCA 2017. Protocol and Survey Guide. ORCA.

ORCA 2019. ORCA Full Dataset. Portsmouth, UK.

ORCA. 2020. *ORCA Survey Network Routes* [Online]. Available: https://www.orcaweb.org.uk/surveyor-network [Accessed 31st January 2020].

PACIFICI, K., REICH, B. J., DORAZIO, R. M. & CONROY, M. J. 2016. Occupancy estimation for rare species using a spatially-adaptive sampling design. *Methods in Ecology and Evolution,* 7**,** 285-293.

PALKA, D. L. & HAMMOND, P. S. 2001. Accounting for responsive movement in line transect estimates of abundance. *Canadian Journal of Fisheries and Aquatic Sciences,* 58**,** 777-787.

PHILLIPS, E. M., HORNE, J. K., ZAMON, J. E., FELIS, J. J. & ADAMS, J. 2019. Does perspective matter? A case study comparing Eulerian and Lagrangian estimates of common murre (Uria aalge) distributions. *Ecol Evol,* 9**,** 4805-4819.

POLLOCK, C. J., LANE, J. V., BUCKINGHAM, L., GARTHE, S., JEAVONS, R., FURNESS, R. W. & HAMER, K. C. 2021. Risks to different populations and age classes of gannets from impacts of offshore wind farms in the southern North Sea. *Mar Environ Res,* 171**,** 105457.

POLLOCK, C. M., MAVOR, R., WEIR, C. R., REID, A., WHITE, R. W., TASKER, M. L., WEBB, A. & REID, J. B. 2000. The distribution of seabirds and marine mammals in the Atlantic Frontier, north and west of Scotland. Aberdeen UK: Joint Nature Conservation Committee.

POLLOCK, K. H., MARSH, H., BAILEY, L. L., FARNSWORTH, G. L., SIMONS, T. R. & ALLDREDGE, M. W. 2004. Separating components of detection probability in abundance estimation: an overview with diverse examples. . *In:* THOMPSON, W. L. (ed.) *Sampling Rare or Elusive Species: concepts, designs, and techniques for estimating population parameters.* Washington DC, USA: Island Press.

R CORE TEAM 2022. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

REDDY, S. & DÁVALOS, L. M. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography,* 30**,** 1719-1727.

REID, J. B., EVANS, P. G. H. & NORTHRIDGE, S. P. 2003. Atlas of Cetacean distribution in north-west European waters. *JNCC.* Peterborough, UK: JNCC.

ROBBINS, J. R., BABEY, L. & EMBLING, C. B. 2019. Citizen science in the marine environment: a case-study estimating common dolphin densities in the north-east Atlantic. *PeerJ Preprints***,** e27569v1.

ROBINSON, K. P., BAUMGARTNER, N., EISFELD, S. M., CLARK, N. M., CULLOCH, R. M., HASKINS, G. N., ZAPPONI, L., WHALEY, A. R., WEARE, J. S. & TETLEY, M. J. 2007. The summer distribution and occurrence of cetaceans in the coastal waters of the outer southern Moray Firth in northeast Scotland (UK). *Lutra,* 50**,** 19.

RONDININI, C., WILSON, K. A., BOITANI, L., GRANTHAM, H. & POSSINGHAM, H. P. 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters,* 9**,** 1136-1145.

SANTORA, J. A., FIELD, J. C., SCHROEDER, I. D., SAKUMA, K. M., WELLS, B. K. & SYDEMAN, W. J. 2012. Spatial ecology of krill, micronekton and top predators in the central California Current: Implications for defining ecologically important areas. *Progress in Oceanography,* 106**,** 154-174.

SCHWEMMER, P. & GARTHE, S. 2005. At-sea distribution and behaviour of a surface-feeding seabird, the lesser black-backed gull *Larus fuscus*, and its association with different prey. *Marine Ecology Progress Series,* 285**,** 245-258.

SCOS 2021. Scientific Advice on Matters Related to the Management of Seal Populations: 2021. *In:* UNIT, S. M. R. (ed.) *Special Committee on Seals.* Natural Environment Research Council (NERC).

SEA WATCH FOUNDATION 2019. Sea Watch Foundation Recording Instructions. New Quay, UK: Sea Watch Foundation.

SICA, G. T. 2006. Bias in research studies. *Radiology,* 238.

SIGLER, M. F., KULETZ, K. J., RESSLER, P. H., FRIDAY, N. A., WILSON, C. D. & ZERBINI, A. N. 2012. Marine predators and persistent prey in the southeast Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography,* 65-70**,** 292-303.

SMITH, E. D., SIEFERT, W. T. & DRAIN, D. 2009. Risk matrix input data biases. *Systems Engineering,* 12**,** 344-360.

SOUTHALL, E. J., SIMS, D. W., METCALFE, J. D., DOYLE, J. I., FANSHAWE, S., LACEY, C., SHRIMPTON, J., SOLANDT, J.-L. & SPEEDIE, C. D. 2005. Spatial distribution patterns of basking sharks on the European shelf: preliminary comparison of satellite-tag geolocation, survey and public sightings data. *Journal of the Marine Biological Association of the United Kingdom,* 85**,** 1083-1088.

SOUTHWELL, C., MCKINLAY, J., LOW, M., WILSON, D., NEWBERY, K., LIESER, J. L. & EMMERSON, L. 2013. New methods and technologies for regional-scale abundance estimation of land-breeding marine animals: application to Adélie penguin populations in East Antarctica. *Polar Biology,* 36**,** 843-856.

SOUTHWELL, C., PAXTON, C. G. M., BORCHERS, D., BOVENG, P., ROGERS, T. & DE LA MARE, W. K. 2008. Uncommon or cryptic? Challenges in estimating leopard seal abundance by conventional but state-of-the-art methods. *Deep Sea Research Part I: Oceanographic Research Papers,* 55**,** 519-531.

SPEAR, L. B., AINLEY, D. G., HARDESTY, B. D., HOWELL, S. N. & WEBB, S. W. 2004. Reducing biases affecting at-sea surveys of seabirds: use of multiple observer teams. *Marine Ornithology,* 32**,** 147-157.

STONE, C. J., WEBB, A. & TASKER, M. L. 2009. The distribution of auks and Procellariiformes in north-west European waters in relation to depth of sea. *Bird Study,* 42**,** 50-56.

STUBER, E. F., ARAYA-AJOY, Y. G., MATHOT, K. J., MUTZEL, A., NICOLAUS, M., WIJMENGA, J. J., MUELLER, J. C. & DINGEMANSE, N. J. 2013. Slow explorers take less risk: a problem of sampling bias in ecological studies. *Behavioral Ecology,* 24**,** 1092-1098.

SYDEMAN, W. J., THOMPSON, S. A. & KITAYSKY, A. 2012. Seabirds and climate change: roadmap for the future. *Marine Ecology Progress Series,* 454**,** 107-117.

TASKER, M. L., JONES, P. H., DIXON, T. I. & BLAKE, B. F. 1984. Counting seabirds at sea from ships: a review of methods employed and a suggestion for a standardized approach. *The Auk,* 101**,** 567-577.

TEILMANN, J. & SVEEGAARD, S. 2019. Porpoises the World Over: Diversity in Behavior and Ecology. *In:* WÜRSIG, B. (ed.) *Ethology and Behavioral Ecology of Odontocetes. Ethology and Behavioral Ecology of Marine Mammals.* . Cham: Springer.

THOMPSON, P. M., MCCONNELL, B. J., TOLLIT, D. J., MACKAY, A., HUNTER, C. & RACEY, P. A. 1996. Comparative distribution, movements and diet of harbour and grey seals from Moray Firth, NE Scotland. *Journal of Applied Ecology,* 33**,** 1572-1584.

THOMSON, J. A., COOPER, A. B., BURKHOLDER, D. A., HEITHAUS, M. R. & DILL, L. M. 2013. Correcting for heterogeneous availability bias in surveys of long-diving marine turtles. *Biological Conservation,* 165**,** 154-161.

TITTENSOR, D. P., MORA, C., JETZ, W., LOTZE, H. K., RICARD, D., BERGHE, E. V. & WORM, B. 2010. Global patterns and predictors of marine biodiversity across taxa. *Nature,* 466**,** 1098-101.

VALENTE, J. J., HUTCHINSON, R. A. & BETTS, M. G. 2017. Distinguishing distribution dynamics from temporary emigration using dynamic occupancy models. *Methods in Ecology and Evolution,* 8**,** 1707-1716.

VAN DER MEER, J. & CAMPHUYSEN, K. J. 1996. Effect of observer differences on abundance estimates of seabirds from ship-board strip transect surveys. *Ibis* 138**,** 433–437.

VAN FRANEKER, J. A. 1994. A comparison of methods for counting seabirds at sea in the Southern Ocean. *Journal of Field Ornithology,* 65**,** 96–108.

VAN STRIEN, A. J., VAN SWAAY, C. A. M., TERMAAT, T. & DEVICTOR, V. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology,* 50**,** 1450-1458.

VINCENT, C., HUON, M., CAURANT, F., DABIN, W., DENIAU, A., DIXNEUF, S., DUPUIS, L., ELDER, J.-F., FREMAU, M.-H., HASSANI, S., HEMON, A., KARPOUZOPOULOS, J., LEFEUVRE, C., MCCONNELL, B. J., MOSS, S. E. W., PROVOST, P., SPITZ, J., TURPIN, Y. & RIDOUX, V. 2017. Grey and harbour seals in France: distribution at sea, connectivity and trends in abundance at haulout sites. . *Deep Sea Res. Part II Top. Stud. Oceanogr.,* 141**,** 294–305.

VOTIER, S. C., BICKNELL, A., COX, S. L., SCALES, K. L. & PATRICK, S. C. 2013. A bird's eye view of discard reforms: bird-borne cameras reveal seabird/fishery interactions. *PLoS One,* 8**,** e57376.

WAGGITT, J. J., EVANS, P. G. H., ANDRADE, J., BANKS, A. N., BOISSEAU, O., BOLTON, M., BRADBURY, G., BRERETON, T., CAMPHUYSEN, C. J., DURINCK, J., FELCE, T., FIJN, R. C., GARCIA-BARON, I., GARTHE, S., GEELHOED, S. C. V., GILLES, A., GOODALL, M., HAELTERS, J., HAMILTON, S., HARTNY-MILLS, L., HODGINS, N., JAMES, K., JESSOPP, M., KAVANAGH, A. S., LEOPOLD, M., LOHRENGEL, K., LOUZAO, M., MARKONES, N., MARTÍNEZ-CEDEIRA, J., Ó CADHLA, O., PERRY, S. L., PIERCE, G. J., RIDOUX, V., ROBINSON, K. P., SANTOS, M. B., SAAVEDRA, C., SKOV, H., STIENEN, E. W. M., SVEEGAARD, S., THOMPSON, P., VANERMEN, N., WALL, D.,

WEBB, A., WILSON, J., WANLESS, S., HIDDINK, J. G. & PUNT, A. 2019. Distribution maps of cetacean and seabird populations in the North-East Atlantic. *Journal of Applied Ecology,* 57**,** 253-269.

WAHL, T. R. & HEINEMANN, D. 1979. Seabirds and fishing vessels co-occurrence and attraction. *Condor,* 81**,** 390–396.

WALLACE, B. P., DIMATTEO, A. D., BOLTEN, A. B., CHALOUPKA, M. Y., HUTCHINSON, B. J., ABREU-GROBOIS, F. A., MORTIMER, J. A., SEMINOFF, J. A., AMOROCHO, D., BJORNDAL, K. A., BOURJEA, J., BOWEN, B. W., BRISENO DUENAS, R., CASALE, P., CHOUDHURY, B. C., COSTA, A., DUTTON, P. H., FALLABRINO, A., FINKBEINER, E. M., GIRARD, A., GIRONDOT, M., HAMANN, M., HURLEY, B. J., LOPEZ-MENDILAHARSU, M., MARCOVALDI, M. A., MUSICK, J. A., NEL, R., PILCHER, N. J., TROENG, S., WITHERINGTON, B. & MAST, R. B. 2011. Global conservation priorities for marine turtles. *PLoS One,* 6**,** e24510.

WEIR, C. R. & STOCKIN, K. A. 2001. The occurrence and distribution of bottlenose dolphins (Tursiops truncatus) and other cetacean species in the coastal waters of Aberdeenshire, Scotland. *SeaWatch Foundation.*

WHITLOCK, S. L., WOMBLE, J. N. & PETERSON, J. T. 2020. Modelling pinniped abundance and distribution by combining counts at terrestrial sites and in-water sightings. *Ecological Modelling,* 420**,** 108965.

WICKHAM, H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

WILLIAMS, P. H., MARGULES, C. R. & HILBERT, D. W. 2002. Data requirements and data sources for biodiversity priority area selection. *Journal of Bioscience,* 27**,** 327-338.

WILLIAMS, R., ASHE, E., GAUT, K., GRYBA, R., MOORE, J. E., REXSTAD, E., SANDILANDS, D., STEVENTON, J. & REEVES, R. R. 2017. Animal Counting Toolkit: a practical guide to small-boat surveys for estimating abundance of coastal marine mammals. *Endangered Species Research,* 34**,** 149-165.

WITT, M., BRODERICK, A., JOHNS, D., MARTIN, C., PENROSE, R., HOOGMOED, M. & GODLEY, B. 2007a. Prey landscapes help identify potential foraging habitats for leatherback turtles in the NE Atlantic. *Marine Ecology Progress Series,* 337**,** 231–243.

WITT, M. J., HARDY, T., JOHNSON, L., MCCLELLAN, C. M., PIKESLEY, S. K., RANGER, S., RICHARDSON, P. B., SOLANDT, J. L., SPEEDIE, C., WILLIAMS, R. & GODLEY, B. J. 2012. Basking sharks in the northeast Atlantic: spatio-temporal trends from sightings in UK waters. *Marine Ecology Progress Series,* 459**,** 121-134.

WITT, M. J., PENROSE, R. & GODLEY, B. J. 2007b. Spatio-temporal patterns of juvenile marine turtle occurrence in waters of the European continental shelf. *Marine Biology,* 151**,** 873–885.

YAMAURA, Y. & ROYLE, J. A. 2017. Community distance sampling models allowing for im- perfect detection and temporary emigration. . *Ecosphere,* 8.

Chapter 4 Trends in marine predator biodiversity mapping using a long-term dataset from 1979-2018

4.1 Abstract

Mapping hotspots of marine predator biodiversity is needed for conservation and marine management. A rapid global drive for biodiversity recovery requires accurate biodiversity baselines against which to measure change but inter-taxa studies of species biodiversity are lacking, particularly for wide-ranging cryptic marine species. This study collated diverse survey data for 74 marine predator species and provides novel comprehensive maps of biodiversity around the UK and Europe. The European Seabird at Sea (ESAS) database and sightings datasets from ORCA were used to generate maps of species richness at three cell sizes (100 km^2, 10 km^2 and 1 km^2). Whilst species richness has been a keystone measure of biodiversity, it can be skewed by survey effort and therefore a new measure of 'species richness per unit effort' (SRPUE) was derived to identify relative areas of high and low richness. SRPUE was calculated using a semi-log relationship between richness and survey effort to give a measure of species richness that accounts for survey effort. Patterns in species richness over time using seasonal-trend decomposition analysis revealed that the concept of carrying capacity becomes evident when survey effort is accounted for. Species richness appears to have fluctuated over four decades studied whilst SRPUE remained stable, suggesting artefacts of sampling effort can skew conclusions of temporal changes in richness. Failure to factor in survey effort when determining species richness hotspots can be detrimental to marine planning. This study demonstrates how the analysis of existing long-term diverse survey data can facilitate the mapping of cross-taxa biodiversity of marine predators and allow areas of high shared space use to be prioritised in an era where proactively protecting biodiversity is at the forefront of global policy.

4.2 Introduction

Biodiversity is a key topic globally, with both scientific research (Hillebrand *et al.*, 2018, Herberling *et al.*, 2019) and international law (CMS, 1979, ASCOBANS, 1994, European Commission, 2008, Convention on Biological Diversity, 2020, European Commission, 2020, Convention on Biological Diversity, 2021, OSPAR commission, 2021) alike focusing on the worsening "biodiversity crisis" and the narrow window for reversing the decline in biodiversity. The most recent COP27 in 2022 (United Nations, 2022) highlighted the urgent need to address the global crisis of biodiversity loss, with a themed Biodiversity Day, and new targets are being agreed upon in the Post-2020 Global Biodiversity Framework (Convention on Biological Diversity, 2022), aiming to slow down the rate of biodiversity loss and halt further decline (Tittensor *et al.*, 2014). Biodiversity has also been suggested to provide valuable ecosystem services to both the economy and society (IPBES, 2019, Dasgupta, 2021), with the World Economic Forum rating biodiversity loss as one of the top five risks to the global economy (World Economic Forum, 2021). Therefore, this universal drive for biodiversity recovery requires accurate baselines against which to measure change.

Within the marine realm, apex predators such as cetaceans, seabirds and pinnipeds are considered 'sentinels' of marine ecosystems (Moore, 2008, Reisinger *et al.*, 2018), with significant top-down effects on ecosystems (Hunt and McKinnell, 2006, Baum and Worm, 2009), as their distribution and life histories are influenced by those of their prey, and both are affected by physical and biological factors of the multidimensional environment they inhabit. These species, which include charismatic animals with high ecotourism and cultural importance (Higham and Lück, 2007) often face numerous anthropogenic threats such as bycatch, habitat degradation and loss, noise disturbance, prey reductions, vessel traffic and collisions, pollution and energy extraction (Croxall *et al.*, 2012, Avila *et al.*, 2018). Whilst many thousands of species are of conservation concern, it is common practice to focus on particular species for marine management and conservation strategies (Caro, 2010), with marine predators encompassing the terms 'indicator' species (those which respond quickly to changes in environment or biodiversity loss), 'keystone' species (important ecological roles in ecosystem structure and functioning), 'umbrella' species (protect biodiversity, due to wide home ranges) and 'flagship species' (species that people feel concerned about)

(Albert *et al.*, 2018). Therefore, protecting marine predator biodiversity is critical for healthy ecosystems and ecosystem functioning.

To assess changes in marine predator biodiversity, mapping baseline data and modelling hotspots is critical (Magurran *et al.*, 2010) and has become an increasingly growing field of study (Reese and Brodeur, 2006, Bouchet *et al.*, 2015). Whilst the distributions of different predator species in isolation are becoming better known, there is a critical need to consider species assemblages to identify important ecological areas or common drivers of space use amongst predators (Block *et al.*, 2011, Thiers *et al.*, 2016, Reisinger *et al.*, 2018), and inter-taxa studies are lacking (Cox *et al.*, 2018). Studies to date have largely been carried out on a species-by-species basis and there are limited studies that have considered multi-species inter-taxa datasets in the marine realm (Block *et al.*, 2011, Santora and Veit, 2013, Tittensor *et al.*, 2014, Waggitt *et al.*, 2019). Cross-taxa and cross-realm studies can provide unique insights and opportunities to advance our understanding of biodiversity decline. For wide-ranging marine predators such as cetaceans and seabirds that often disperse or migrate across oceans (Hays *et al.*, 2013) maximising survey effort in space and time is critical (Waggitt *et al.*, 2019). Repeat surveys over the same areas, with consistent methodologies, are high in resource demand, cost and time (Bull *et al.*, 2013), and can often limit research. Utilising available historical databases can provide the broad spatial and temporal coverage required to understand marine predator hotspots. Combining existing long-term datasets from multiple bodies may give the coverage necessary for studying marine predators, which individual fine-scale studies cannot, as long as biases with the heterogeneous datasets are accounted for (as discussed in Chapter 3). Additionally, subsequent time delays in data processing, cleaning, and even more so in analysis, hinder data progress, and significant shifts in species space use (such as richness) may have already occurred. With biological responses to climate change increasingly at the forefront of research and legislation, increased efficient and cost-effective use of existing historical datasets would be welcome to meet conservation and management needs (Leonard *et al.*, 2006). The Aichi Target 19 specifically encouraged the development of a shared advanced biodiversity knowledge base (Meyer *et al.*, 2015).

Whilst biodiversity is a fundamentally multidimensional concept which cannot be summarised easily by a single measure (Magurran, 1988, Purvis and Hector, 2000), considerable importance has been put on species richness (Colwell and Coddington, 1994, Rosenzweig, 1995, Gaston, 1996, Jetz and Rahbek, 2001, Rahbek and Graves, 2001) which continues to be the most frequent metric of biodiversity (Hillebrand *et al.*, 2018). Species richness is deemed a good indicator of community structure and ecosystem resilience (Magurran, 1988, Brewer and Williamson, 1994) and is easily observed and recorded, making it an affordable way to monitor ecosystem change (Hillebrand *et al.*, 2018). Areas of high concentrations of species richness (termed "biodiversity hotspots") have shaped conservation priorities in terrestrial habitats (Kremen *et al.*, 2008) but remain lacking or mismatched in marine habitats (Worm *et al.*, 2003, Bouchet *et al.*, 2015, Devillers *et al.*, 2015). With environmental changes and anthropogenic activity impacting ecosystems, understanding the macroecological patterns of species distributions and coexistence is important for assessing change, conservation, mitigation and management (Jewell *et al.*, 2012, Scales, 2016).

Species richness mapping over different temporal scales and at different resolutions is lacking for top marine predators, and utilising long-term datasets which have the data quantity and coverage could facilitate a critical understanding of how these patterns have changed over the past few decades and give insight into the future. Characteristically, these biodiversity databases are patchy in coverage and biased due to limitations with data collection, lack of standardisation and heterogenous data quality (Jetz and Rahbek, 2001, Soberón *et al.*, 2007, Lobo, 2008, Boakes *et al.*, 2010, Rocchini *et al.*, 2011, Ladle and Hortal, 2013, Meyer *et al.*, 2015, Lobo *et al.*, 2018) (discussed in detail in Chapter 3). When databases are combined from heterogeneous sources of data, with many different collecting methodologies within them, they often lack a standardised quantification or measure of survey effort (Lobo *et al.*, 2018). Unless survey effort is considered, areas of seemingly high biodiversity may be simply reflective of higher sampling coverage.

Furthermore, sighting databases are often presence-only data (crucially not presence-absence as commonly termed (Gotelli, 2011)), where species are only recorded when sighted, rather than recording when they are not. This makes it difficult to distinguish between actual true absences of species or a lack of surveying at the correct time and place to observe them, so knowledge of species distributions are typically incomplete (known as the Wallacean shortfall (Lomolino, 2004, Whittaker *et al.*, 2005)). The absence of evidence is not evidence of absence. However, unlike some other species diversity indices or distribution estimates, species richness is not confounded by patchy abundance distributions (Chazdon *et al.*, 1998, Brose *et al.*, 2003) so may be best suited to use for heterogenous datasets compiled from different providers to answer relative questions - such as associations, trends, residency patterns, behaviours (Phillips, 2009, McBride-Kebert *et al.*, 2019). It must be considered that species richness is always a downward-biased estimator (Gotelli, 2011), as the survey method and the cryptic lifestyle of marine predators means sighting data is always an underestimate of true richness.

However, it is important to factor in how survey effort can influence mapped richness hotspots to determine true areas of high species richness rather than as an artefact of survey effort. It is well accepted in ecological study that the greater the intensity of sampling, the more species will be recorded (Ugland *et al.*, 2003) (i.e. the more you search the more you will find). Maps of observed species richness often mimic maps of the number of records per unit (Hortal *et al.*, 2007b). Highly surveyed areas may insinuate high species richness but that may be entirely down to more survey effort meaning the chances of sighting animals, and new species are increased. As such, a measure of associated sampling effort should be factored in for future analysis of richness (Gaston, 1996).

Many studies have demonstrated the use of semi-log or log-log models to quantify species-area relationships or species accumulation curves (Tjørve, 2003, Ugland *et al.*, 2003), and whilst the discussion and debate between the two models are ongoing (Palmer, 1990, Colwell and Coddington, 1994, Rosenzweig, 1995, He and Legendre, 1996, Crawley and Harral, 2001, Gotelli and Colwell, 2001, Hubbell, 2001, Krebs, 2001, Gotelli and Colwell, 2011), it has been argued that the choice should be based upon the goodness-of-fit of the model to the data (He and Legendre, 1996). Several studies suggest the choice may be related to sampling design (Scheiner, 2003, Gray, 2004) or the spatial scale of the study (He and Legendre, 1996, Crawley and Harral, 2001, Lomolino, 2004). Species-area predicts a positive linear relationship between species richness and area when area is log-transformed (a semi-log scale) (Elliott *et al.*, 2020), with more species typically present in a larger area. Species-accumulation curves measure the rate of accumulation of different species as the area sampled is increased and has been shown to be semi-logarithmic (Gray, 2004). Recently, in their study of plant species richness in temperature semi-natural grassland, White *et al.*,

(2018) found that the semi-logarithmic function was the best fit for all diversity-area, diversity-time-area relationships and equally best for diversity-time relationships (with power function).

In this chapter, the relationship explored is between species richness and survey effort (number of occurrences in that area), rather than the more well-known species richness area or species accumulation curves. Whilst literature on this relationship is more limited, Ugland *et al.* (2003) used Norwegian continental shelf data and showed that species richness plotted against the log of the corresponding number of samples explained 99% of the variability. Gotelli and Colwell (2001) stated that for patchy distributions it is preferable to use sample-based species–accumulation curves that take account of sample heterogeneity, rather than an individual-based approach, and therefore sample-based curves are to be utilised in such heterogenous datasets as ESAS and ORCA. Unlike in controlled experiments or set habitat plots, it can be impossible to count all the species in the area, particularly in dynamic marine environments such as coral reefs or marine sediments or for highly mobile species such as marine predators. In these habitats, the approach is to estimate total species richness and factor in sampling effort to obtain reliable estimates. The more individuals are sampled, the more species will be recorded (Bunge and Fitzpatrick, 1993). The sampling curve rises rapidly asymptotically and then theoretically reaches a plateau as more rare species are found – however recording all marine species is realistically impossible in the field, and as such an asymptote is never reached (Pielou, 1966, Chao *et al.*, 2009).

4.2.1 Aims

This study aims to provide maps of species richness hotspots for top marine predators at different spatial resolutions (100 km^2, 10 km^2 and 1 km^2) around the UK at different temporal scales. Effort in time and space is maximised by using survey data from a database containing many different suppliers and data sources.

Key questions and hypotheses are:

1) What is the best way to scale survey effort when calculating species richness using heterogeneous databases?
2) Is survey effort biasing observed patterns in species richness? It is hypothesised that survey effort is concealing true species richness patterns, with areas of high

survey effort driving areas of high richness and areas of low survey coverage behind low richness values.

3) Does sampling scale matter? It is hypothesised that grid resolution impacts the maximum species richness recorded, with larger grids having higher mean and maximum species richness values due to the species-area relationship.
4) Has the marine predator biodiversity of this area changed over time? It is predicted that species richness has declined over the 39-year period.

4.3 Materials and Methods

4.3.1 Data sourcing, cleaning and preparation

Occurrence records for species of marine predator were obtained using aerial and vessel survey data from the ESAS database (Version 5) (JNCC, 2019) and cruise distance sampling and wildlife officer sightings datasets from ORCA (ORCA, 2019). ESAS assembles offshore monitoring data on seabirds and marine mammals from various partners during aerial or ship surveys (JNCC, 2022), whilst ORCA collects sightings of marine mammals using trained volunteers on ferries and cruise ships (ORCA, 2019). The spatial, temporal, taxonomical coverage, survey methods and biases in these datasets are discussed in detail in Chapter 3.

ESAS data was extracted from the Microsoft Access database into Excel spreadsheets per species (Microsoft Corporation, 2018), whilst ORCA data was provided in Excel spreadsheet format. Species names were made equivalent for both datasets, and only species considered to be marine predators selected (see Appendix B). A marine predator is defined broadly as species that actively prey on other individuals and play unique or irreplaceable functional roles (e.g. controlling trophic cascades, removing weak or diseased individuals and translocating nutrients between habitats) (Letessier *et al.*, 2019). Literature on each species' prey preference and primary feeding strategy was reviewed for determining qualification as a true marine predator for inclusion in the study.

The datasets were cleaned, with records that were not specific to species level or ambiguous (for example entries such as “small cetacean” or “unidentified dolphin”), or not geo-referenced removed, alongside sightings recorded as being on land (i.e. incorrect coordinates) removed. Discrimination between some species, such as common guillemot *Uria aalge* and razorbill *Alca torda* (Buckland *et al.*, 2012), is difficult in the field, but where species names were given, sightings were maintained whilst those recorded at higher taxonomic naming, for example as Alcidae, with no specific species name were removed. The original ESAS dataset supplied had 153 different "species labels", which included non-species-specific names such as: 'auk spp', 'aytha spp', 'diver spp', 'diving duck spp', 'grebe spp', 'large gull spp' etc and totalled 2,758,378 sightings. Of these, 103 (67%) species-specific labels were retained. The original ORCA dataset had 37 species labels including "big cetacean", "patterned dolphin", and "unidentified beaked whale", totalling 18,591 sightings, and of these 26 species names were retained (70%). Infraspecific taxa were merged to species level and scientific names were standardised according to the IUCN database (IUCN, 2023), within and between datasets. Finally, species were filtered down to predator species only, as defined by Letessier *et al*. (2019), using literature on prey and migration ranges. This led to 78 species retained in the ESAS dataset (50.9% of the original) and 26 species in the ORCA dataset (70% of the original).

This standardised, geo-referenced species data were then imported into ArcGIS Pro v2.7.0 (ESRI Inc., 2020), converted into shape files and merged into one feature class, leading to one final combined dataset from both the ESAS and ORCA data. The area of interest was narrowed down to an extent covering approximately 12,164,988 square kilometres (72.817°N, 26.334°W to 33.937°N, 18.481°E) selected using kernel density analysis in ArcGIS Pro, with the area that encompassed two standard deviations selected as the limiting area. This resulted in a dataset of 1,834,366 observations (Figure 1.3), comprising 74 species of marine predator, which included 40 bird species, 23 species of cetacean, four species of pinnipeds, four species of shark, two species of turtle and one species of fish. The final combined dataset spanned from 1979 to 2018, giving a long-term dataset of 39 years with 13,340 sightings from ORCA (71.8% of the original ORCA dataset provided), and 1,821,026 from ESAS (66% of the original ESAS dataset provided). The shapefile contained information in the attribute table on data source (ESAS or ORCA), species name, taxa group (Birds, Cetaceans, Fish, Pinnipeds, Sharks, Turtles), latitude and longitude, date (Day, Month, Year) and the methodology of collection (Ship, Plane, Helicopter for ESAS data and WO Sightings or Distance sampling for ORCA).

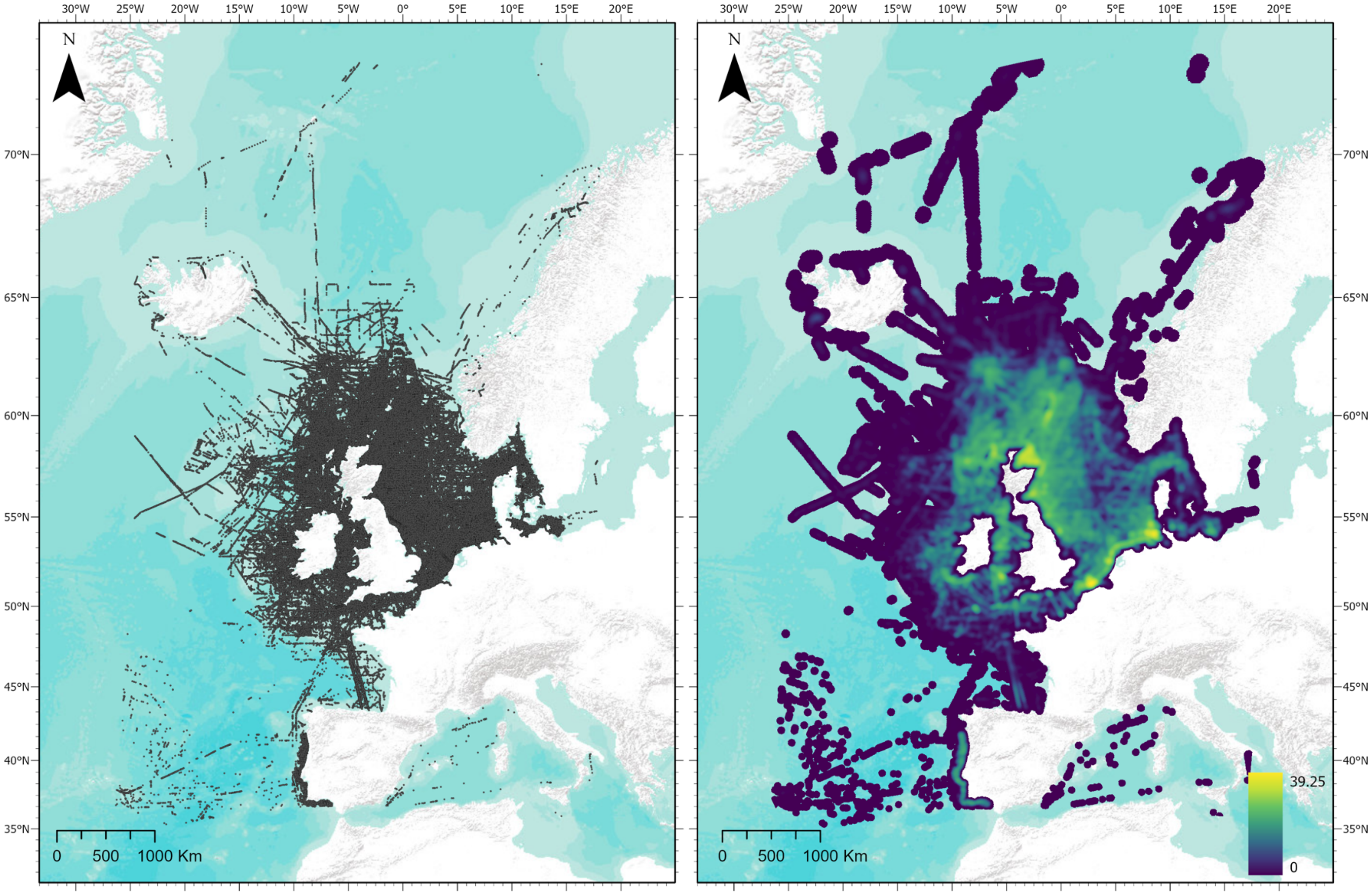


Figure 4.1 Raw sighting occurrences and density map from the combined ESAS and ORCA datasets.

4.3.2 Species richness using grids

To look at species richness at different spatial scales, as well as temporally, the sampling tool ‘fishnet’ was used in ArcGIS Pro, to create a grid of rectangular polygons based upon user-defined parameters for cell size and width. The use of square kilometres was chosen over degrees latitude or longitude for calculating richness, to allow more precise and small-scale changes in grid coarseness. Three grid sizes were chosen for calculating species richness: 100 km^2, 10 km^2 and 1 km^2 cells. The data were converted from WGS 1984 projection into an equal-area projection, to preserve area when mapping. ETRS 1989 LAEA – Lambert Azimuthal Equal Area was chosen and is the most recommended for statistics and mapping display in the EU (Annoni *et al.*, 2003), and suitable for mapping large areas. The subsequent X and Y LAEA coordinates for the sighting data points were added to the attribute table.

Once in the correct projection, shapefiles were exported for further analysis of changes over space and time in R (R Core Team, 2020) using packages ‘Raster’ (Hijmans, 2020), ‘Sp’ (Pebesma and Bivand, 2005, Bivand *et al.*, 2013) and ‘RGDAL’ (Bivand *et al.*, 2019). Raster layers of the extent of the dataset were generated with the cell resolution of each grid cell set to the desired resolution (100 km^2, 10 km^2, 1 km^2). Two measures were calculated using the ‘rasterize’ function: raw species richness and the number of observations per cell. To explore temporal differences in species richness, species richness and observations per grid cell were also calculated per year (from 1979 to 2018), and for each month across all years combined (January to December), and then finally split out into each month, per year – giving rasters of richness and observations for 468 contiguous months.

4.3.3 Using observations as a surrogate of survey effort

The samples were treated as randomised, rather than adding each data point chronologically, and plotted to ascertain the relationship between richness and observations. I used the number of observations within each corresponding grid cell as a proxy for sampling effort. I compared the log-log versus semi-log approximations (Appendix E) to find the most suitable way to scale survey effort for calculating species richness (discussed further in section 4.4.1).

Subsequently, for each cell, at each grid size resolution, a new measure was calculated:

Species richness / Log(Observations)

This new measure, hereafter termed "**Species Richness per Unit Effort**" (**SRPUE**) is a standardised measure of species richness controlled for by survey effort. This was written as a custom function in R, where rasters of species richness were divided by log rasters of observations: *effortfactor <- function(r1, r2){fun1 <- r1 / (log(r2))return(fun1)}*.

Rasters of SRPUE were created for the three grid sizes per year (1979 to 2018); per month (January to December all years combined) and again split out into each month, per year to give 468 contiguous months of SRPUE rasters.

4.3.4 Statistical analysis

Comparison of measures

Metrics of geometric mean, minimum, maximum, standard deviation and the coefficient of variation (CV) were calculated from the raw species richness rasters and SRPUE rasters, at the three different resolutions. CV is the ratio of the standard deviation to the mean and was chosen because it is independent of the unit in which the measurement was taken, and therefore can be used to compare data sets (such as species richness and SRPUE) with different units or widely different means.

To illustrate the magnitude and extent of differences between raw species richness maps and SRPUE maps, both rasters were standardised onto a standard deviation scale for comparison. Firstly data was centred, so for all values of species richness / SRPUE, the average was subtracted (so any pixel lower than average would be negative, and any pixel above average would be positive). The standard deviation (SD) was then obtained, and all data was divided by the SD to convert the data range scale to units of SD to allow comparison across datasets (i.e. both are on a SD unit scale). Differences between the two datasets was then calculated.

Time series decomposition

Metrics (described in section 4.3.4) were also calculated per monthly raster for use in the time series decomposition analysis. Arithmetic mean can be more easily distorted if the sample of observations contains outliers and the geometric mean is more suitable to the log-normal distribution (Limpert *et al.*, 2001), therefore both were calculated. Time series decomposition was carried out using the STLPlus package (Hafen, 2016) on these measures to identify any seasonal variation in the data and identify overall trends over the 39 years. In the SRPUE data, infinity values arose when dividing species richness value of one by log (effort one), so those cells where observations = 1 were removed for further analysis. For the 100 km^2 dataset 87% of the original data points were retained (14,016 data points removed), for 10 km^2 80% of the original data points were retained (48,380 data points removed) and for 1 km^2 74% of the original data points were retained (88,226 data points removed).

To test if the decline in ESAS data post 2011 (when the ESAS dataset ends) effects the diversity measures, a Shapiro-Wilk's test for normality was carried out on the raw species richness dataset and the SRPUE dataset. The raw species richness dataset did not meet Shapiro-Wilk's test for normality, and therefore the non-parametric Mann-Whitney U test (Wilcoxon rank sum test) was used to test for a significant difference between the full datasets from 1979 and 2018, and datasets that excluded data post 2011.

4.4 Results

4.4.1 Scaling survey effort for calculating species richness

Outputs from raw species richness, log and semi-log plots found the semi-log was most linear relationship (Figure **4.2**) and most suitable through inspection of the linear model residuals. The linear relationship did not vary with the size of the area sampled (grid cell grain). As such, this semi-log relationship was taken forward for further analysis and demonstrates the best way to scale survey effort across multiple grid size resolutions.

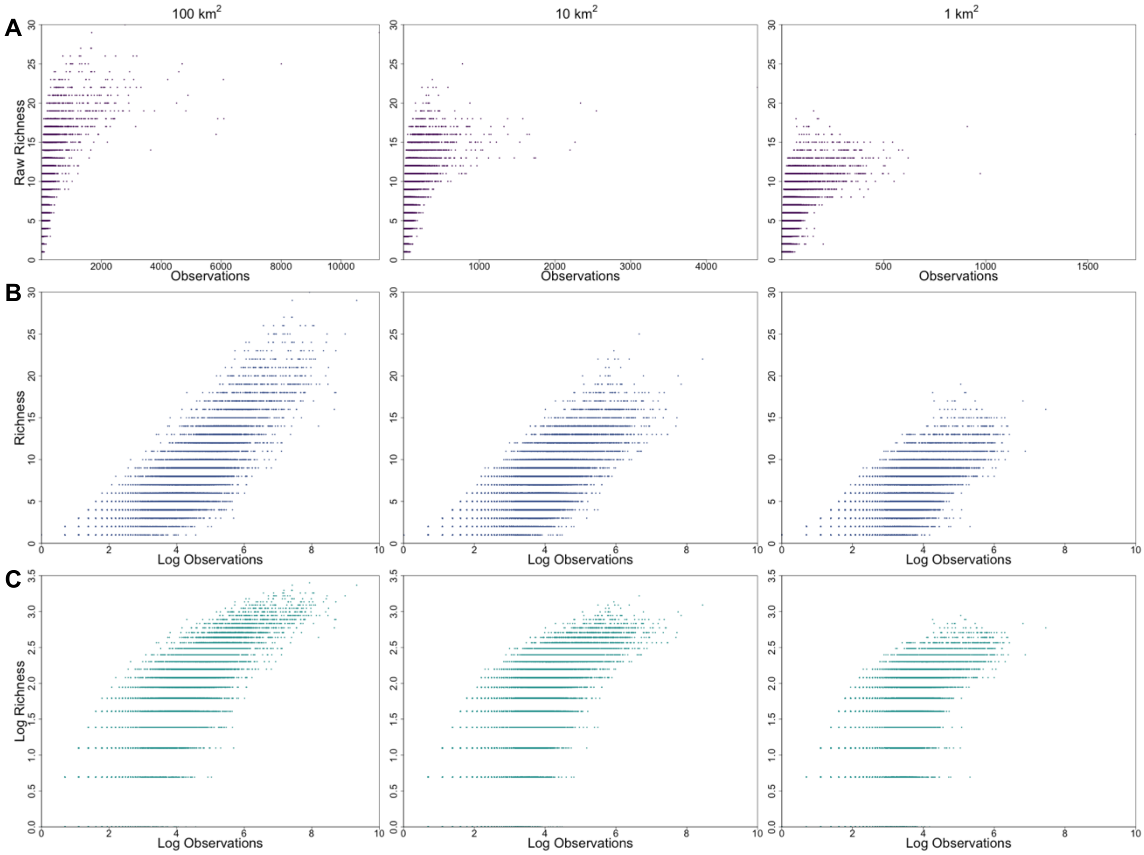


Figure 4.2 **Comparison of raw data (A), semi-log relationship (B) and log-log relationship (C) per grid cell at different grid size resolutions (100 km^2, 10 km^2, 1 km^2)**

4.4.2 Survey effort hiding patterns of species richness: comparison of species richness versus SRPUE maps

Metrics summarising species richness versus SRPUE are presented in Table 4.1, presenting the minimum, mean, maximum, standard deviation (SD) and coefficient of variation (CV) for each dataset (raw species richness and SRPUE). Mean species richness was highest for the 100 km^2 grid cell for both the species richness (geometric mean = 4.30) and SRPUE raster (geometric mean = 1.57), and lowest for 1 km^2. Raw species richness had more variation within the dataset (e.g. CV = 0.75 for 100 km^2) compared to SRPUE (e.g. CV = 0.38 for 100 km^2).

Table 4.1 Metrics summarising species richness versus SRPUE approach: minimum, mean (arithmetic and geometric), maximum, standard deviation (SD) and coefficient of variation (CV).

Measure	Scale (km^2)	Min	Max	Arithmetic Mean	Geometric mean	SD	CV
Species Richness	100	1	30	5.87	4.30	4.42	0.75
	10	1	25	3.31	2.57	2.56	0.77
	1	1	19	2.33	1.90	1.70	0.73
SRPUE	100	0.22	4.18	1.70	1.57	0.64	0.38
	10	0.22	4.01	1.53	1.40	0.62	0.40
	1	0.24	4.22	1.52	1.39	0.62	0.41

Factoring in effort alters the areas of high species richness observed in the raw richness. Figure 4.3 shows three rows of maps, the first illustrates raw species richness at different grid resolutions, the second row illustrates the density of observations, and the third shows species richness per unit effort (SRPUE). A clear reflection of species observations is seen in the maps of species richness, and well-sampled cells tended to be geographically aggregated, with areas of high effort and high species richness observed in Moray Firth, Shetland Islands, and the coast of Belgium, Netherlands and Denmark. With SRPUE, the areas of high richness have changed – for example, less of a hotspot in Moray Firth but increased relative richness along the east coast of the UK. It is important to highlight that this SRPUE is now a measure of “relative richness” over the area rather than actual species richness integer values.

Figure 4.4 illustrates that some hotspots that were seen in the raw richness maps do persist in the SRPUE values. Around the coast of the Irish Sea areas of high species richness for example that clearly reflected shipping routes from Dublin to Holyhead and Rosslare to Fishguard have dissipated, but some areas of high species richness remain suggesting true hotspots in the area. Around the coastline of France, Belgium and the Netherlands there is a clear density of observations due to high shipping traffic and repeated survey areas, but in the SRPUE inset, whilst there are relative hotspots still found along the coast, the relative strength and size of the hotspot has reduced. In the Moray Firth, this area was hugely well-surveyed in comparison to other areas and lead to high species richness counts, however in the SRPUE inset hotspots are much more localised to off the coast of Peterhead, towards the Orkney Islands and inshore towards Inverness.

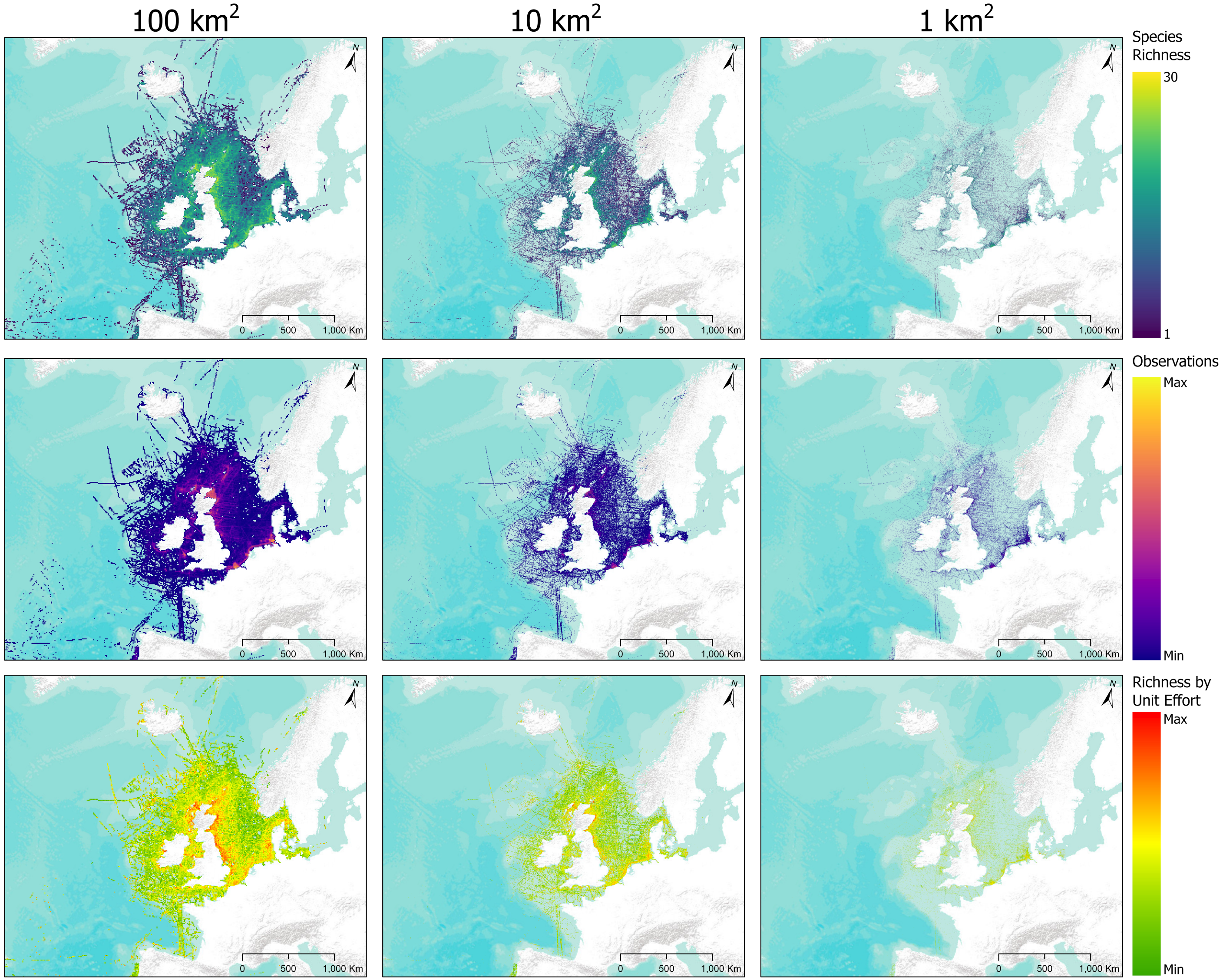


Figure 4.3 Panel plot of species richness (top row), observations (middle row) and species richness standardised by effort (bottom row) at 3 resolutions (100 km^2, 10 km^2, 1 km^2).

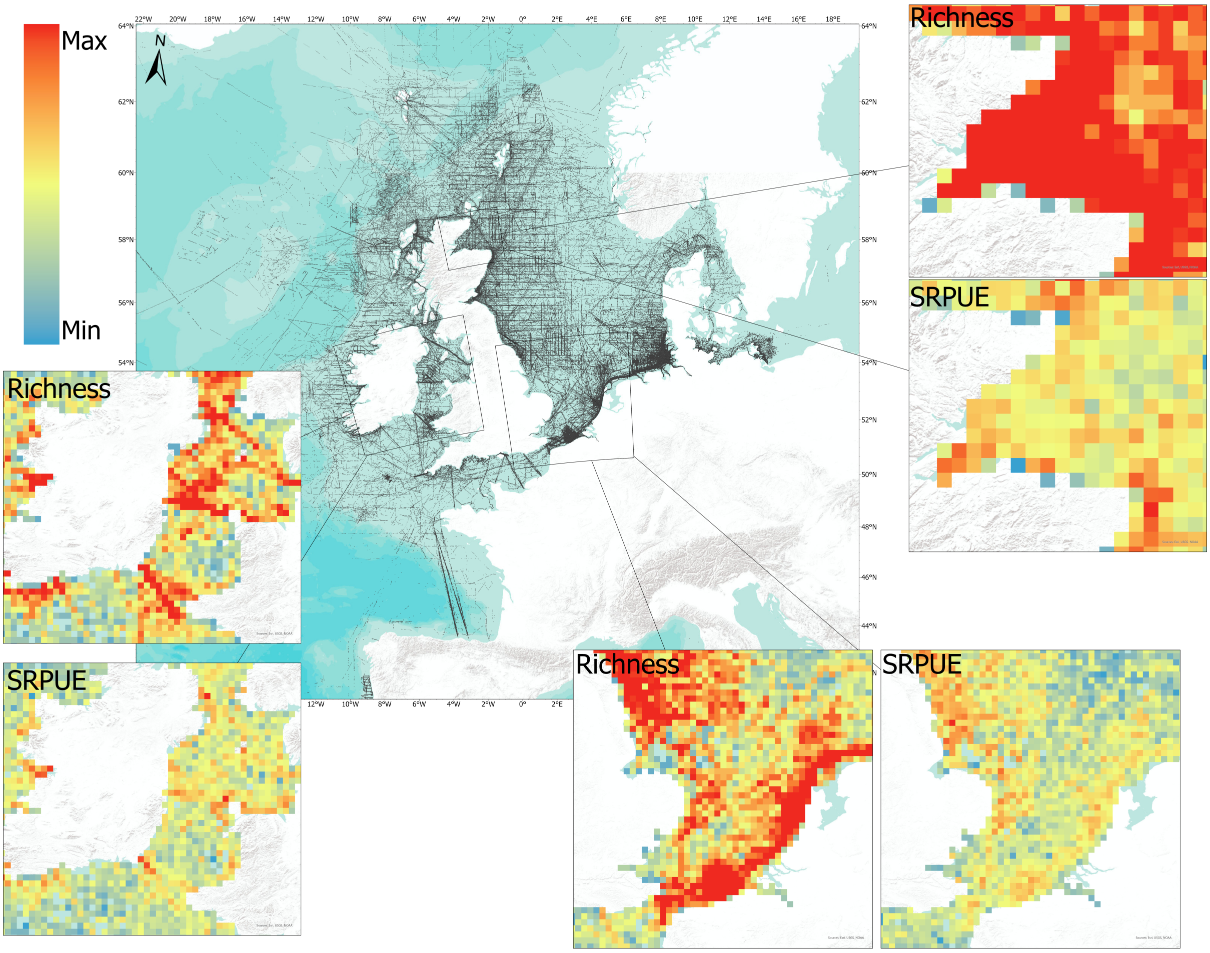


Figure 4.4 Selected insets of raw species richness against species richness by unit effort for areas around the study area, at 100 km^2 resolution.

4.4.3 Extent of differences

Rasters of species richness and SRPUE were centred and then standardised onto a SD scale to allow comparison of uncorrected and corrected diversity measures in terms of the magnitude and extent of differences. Figure 4.5 presents maps of raw species richness (Row A) and SRPUE (Row B) and the difference between measures (Row C) (all expressed on a SD unit scale) at three resolutions (100 km^2, 10 km^2, 1 km^2). Warmer colour cells represent areas of higher species richness (expressed on a scale of number of SD units), whilst cooler colour cells represent areas of lower species richness. Areas of higher richness in the raw species richness plots align with areas of higher observations identified in Figure 4.3 (e.g., around the Northeast coast of Scotland and along the coast of Belgium and the Netherlands), but these areas do not show as much variability in the SRPUE plots. Again, the areas of the highest difference between measures are those with high survey effort (the Moray Firth, along the coast of Belgium and the Netherlands, off the coast of west Wales).

The Pearson correlation coefficient was also calculated to assess the correlation between raw richness (in SD) and SRPUE (in SD). At 100 km^2, the correlation was found to be $r = 0.712$, indicating a moderately strong positive relationship between SD of raw richness and SD of SRPUE ($p < 0.001$). At 10 km^2, the correlation was found to be $r = 0.565$, indicating a moderately strong positive relationship between SD of raw richness and SD of SRPUE ($p < 0.001$). At 1 km^2, the correlation was found to be $r = 0.461$, indicating a moderately strong positive relationship between SD of raw richness and SD of SRPUE ($p < 0.001$).

Both mean and maximum species richness values (in both raw species richness and the relative measure SRPUE) become smaller as grid resolution becomes finer (Table 4.1), as expected by the species-area relationship (the bigger area to include, the more species to find) (Figure 4.3, Figure 4.6). However, across grid sizes, persistent patterns of hotspots remain in the same locations (such as along the east coast of Scotland and the coastline of Europe).

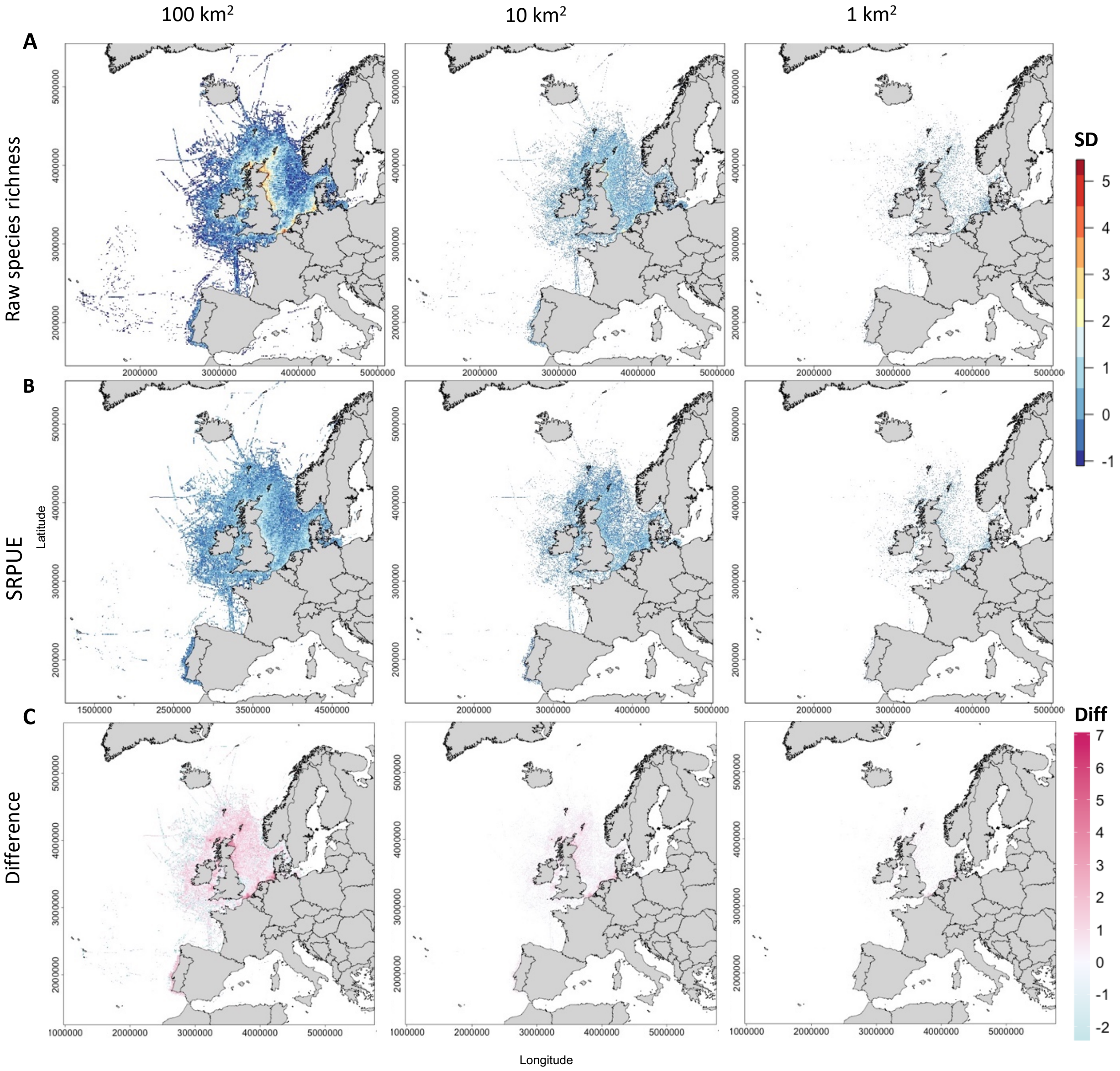


Figure 4.5 Maps of raw species richness (row A) and SRPUE (row B) expressed on the scale of number of SD units from the mean, at three resolutions (100 km^2, 10 km^2, 1 km^2). Row C demonstrates the difference between the two measures.

4.4.4 Time series decomposition of species richness

The time series decomposition over the monthly time series from 1979 to 2018 exhibited the same overall “trend” for mean, geometric mean and median, at all grid size resolutions, with maximum richness values smaller as resolution becomes finer, as expected by the species-area relationship (the bigger area to include, the more species to find) (Figure 4.6). The seasonal-trend decomposition (STL) plot of the geometric mean of raw species richness demonstrated a big peak from 1980 to around 1986, then a more stable period until 2012 and this is reflected in the 100 km^2, 10 km^2, 1 km^2 outputs. The values on the y-axis are taken as relative measures, looking at the trend in richness over time, rather than absolute integers of species richness. When effort is factored in (SRPUE), this trend is broadly removed. The ESAS dataset ends in 2011 (see Table 3.2), and this is reflected in Figure 4.6 in both the raw species richness (visible with lower species richness until 2018) and the corrected SPRUE data (more variability post 2011). The raw species richness dataset did not meet Shapiro-Wilk’s test for normality, and therefore the non-parametric Mann-Whitney U test (Wilcoxon rank sum test) is presented in Table 4.2. Removal of data post 2011 was significant for raw species richness (e.g. for 100 km^2 grid size, $U = 71916$, $p = 0.0001$) which does not account for survey effort, but not significant for SRPUE (e.g. for 100 km^2 grid size, $U = 85423.5$, $p = 0.362$) which accounts for survey effort.

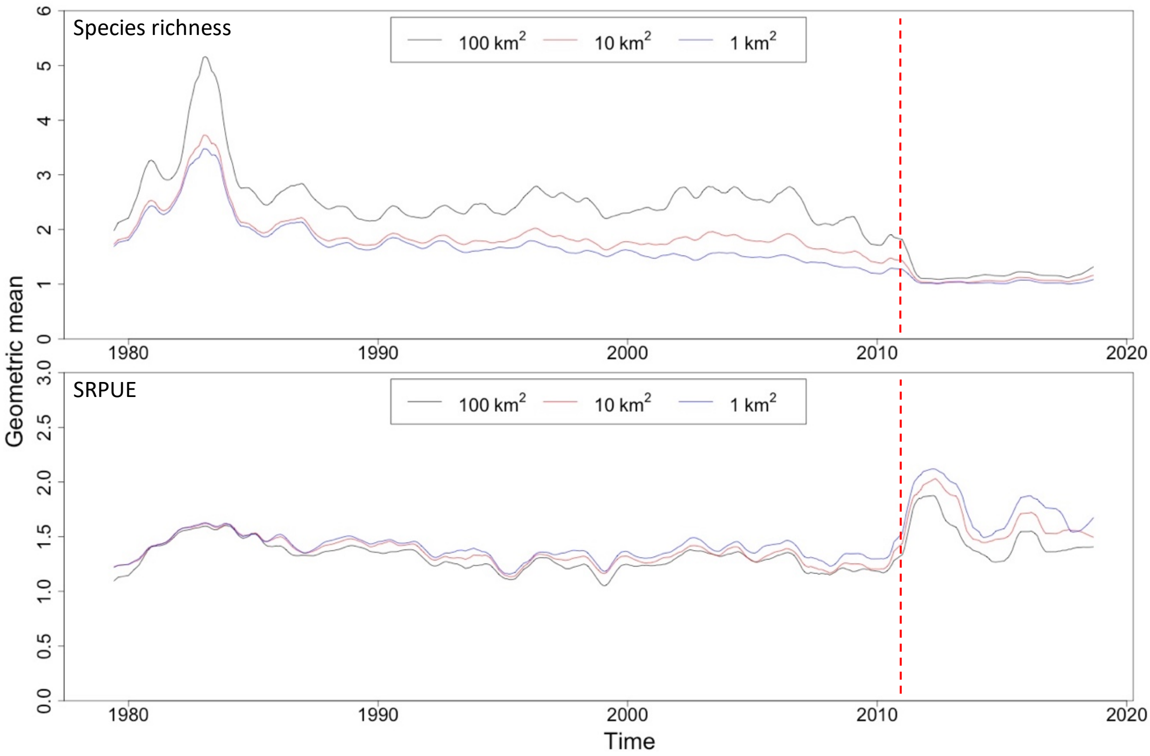


Figure 4.6 Comparison of seasonal-trend decomposition (STL) for trend for geometric mean raw richness versus mean SRPUE – for all resolutions. Red line at 2011 indicates end of ESAS dataset.

Table 4.2 Outputs of Mann-Whitney U test for a significant difference between full dataset (1979 to 2018) and pre-2012 dataset (1979 to 2011).

Measure	Scale ($km^{2)}$)	statistic	P value	Significance
Species Richness	100	71916.5	0.001	Significant
	10	71896.5	0.001	Significant
	1	71883	0.001	Significant
SRPUE	100	85423.5	0.362	Not significant
	10	86912.5	0.084	Not significant
	1	87102.5	0.056	Not significant

Boxplots of the geometric mean of species richness at the different scales (for both the full dataset and the dataset with post 2011 removed) demonstrate that mean species richness and variation in richness values decline as grid size is reduced (Figure 4.7), but SRPUE geometric mean values and variation remain very consistent over the scales.

Boxplots of the seasonal and trend components of the time-series decomposition for all five measures (Arithmetic mean, median and geometric mean of species richness and arithmetic and geometric mean of SRPUE) (for both the full dataset and the dataset with post 2011 removed) (Figure 4.8) demonstrated that whilst seasonal variation is still evident in SRPUE it is much smaller in variation, with less of an obvious impact on the overall trend visible. This is expected due to less survey effort in winter months, and poorer visibility to record sightings. The trend component is reduced in the SRPUE boxplots, confirming the impact survey effort has on the trends seen in Figure 4.6. Well-defined seasonal cycles are seen in the time-series decompositions for each measure at all scales for species richness but are virtually removed when using SRPUE (Appendix F). Comparison of the full dataset versus the dataset with post 2011 removed (Figure 4.8) showed very little difference between the SRPUE boxplots but some difference could be observed between the raw richness boxplots.

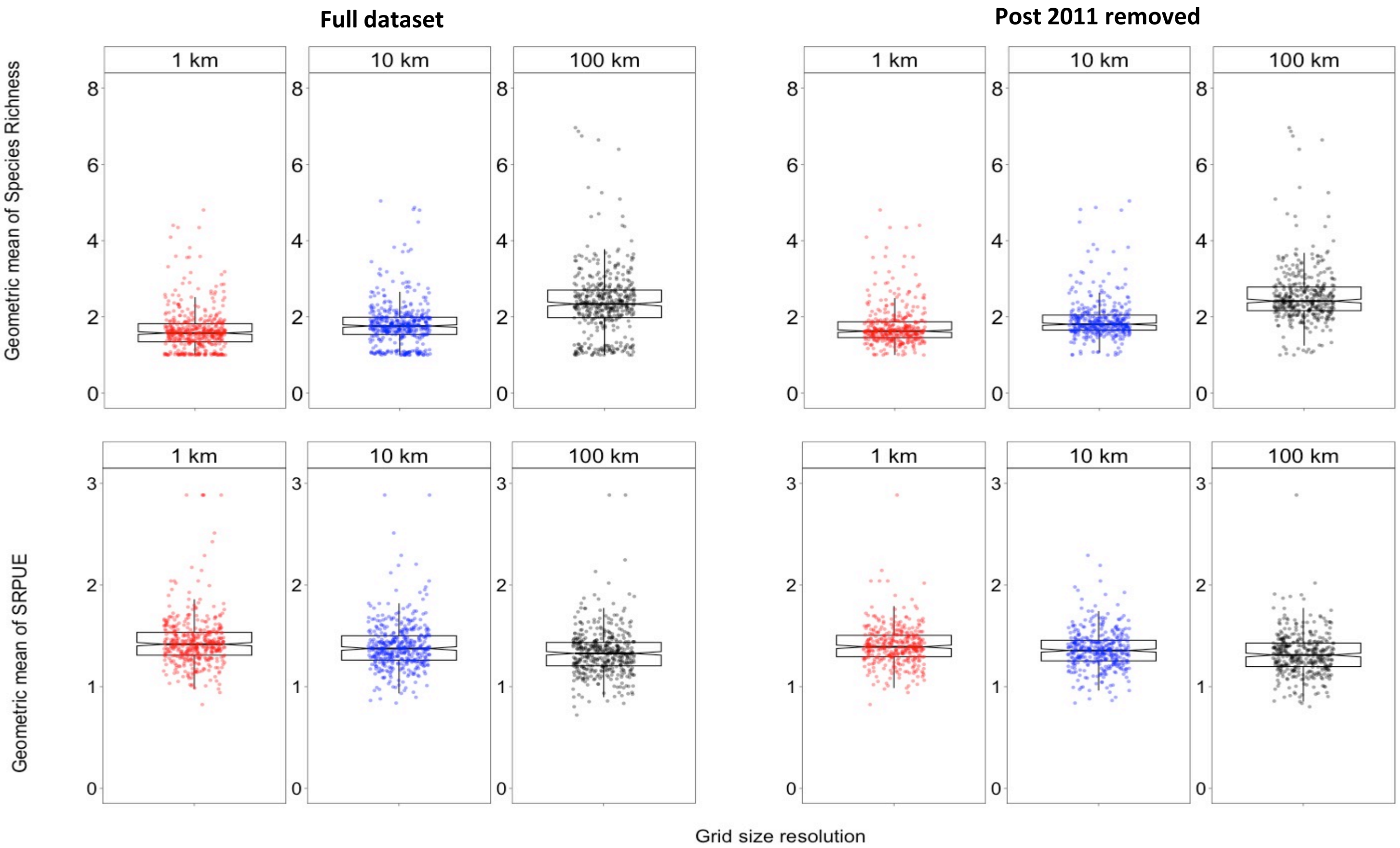
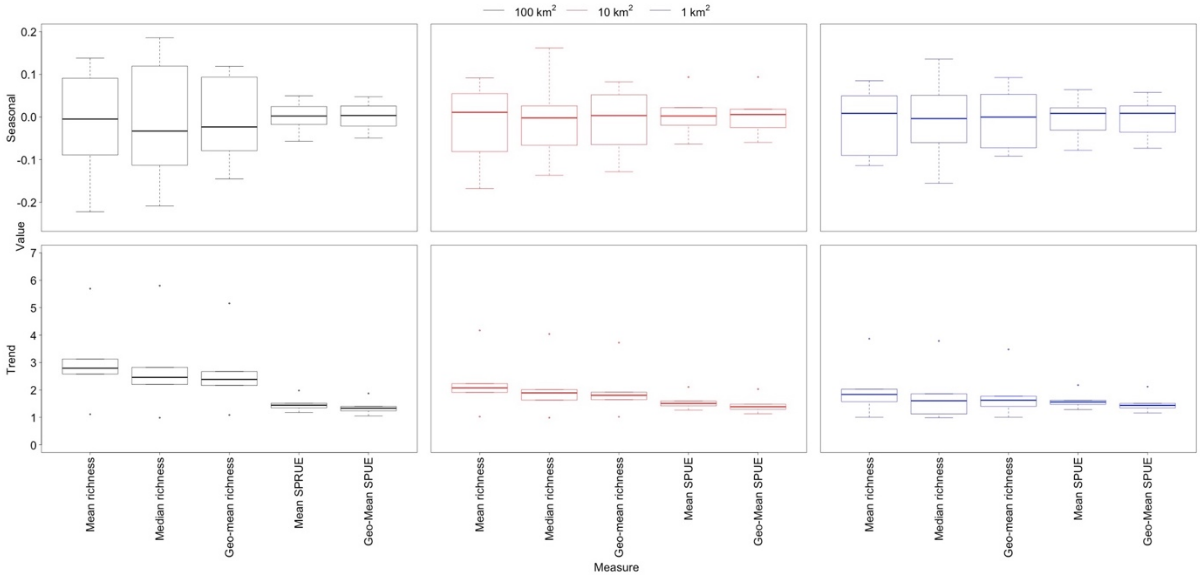


Figure 4.7: Top panel: boxplots of the geometric mean of raw species richness at three resolution scales (100 km^2, 10 km^2, 1 km^2) for the full dataset and data post 2011 removed. Bottom panel: boxplots of geometric mean values of species richness per unit effort (SRPUE) richness at 100 km^2, 10 km^2, 1 km^2 for the full dataset and data post 2011 removed.

Full dataset (1979 to 2018)



Post 2011 removed

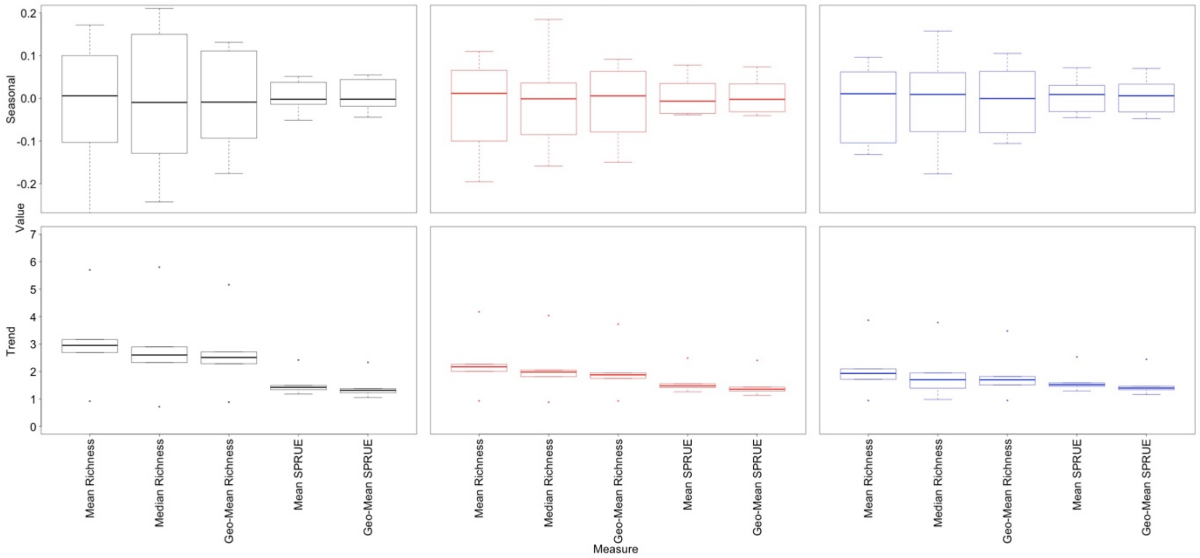


Figure 4.8 Boxplot showing the variation of data for the seasonal component of the STL for each parameter measured (mean species richness, median richness, geometric mean, mean SRPUE, geomean SRPUE) at 100 km^2, 10 km^2, 1 km^2 resolutions, against the actual trend values for the full dataset (1979 to 2018) and dataset with post 2012 removed (1979 to 2011).

4.5 Discussion

This study provides novel species richness maps of marine predators in the North-East Atlantic at different spatial scales, using a well-used approach to calculate species richness by counting the number of distinct species (derived from species occurrence data) for cells on a grid (Hijmans and Spooner, 2001, Graham and Hijmans, 2006). To my knowledge, studies on species richness that cross multiple marine taxa are limited (Tittensor *et al.*, 2010, Gladstone *et al.*, 2020), particularly for marine predators. This exploration of 74 species across six taxa provides the first broad indicative coverage of marine predator species, but it is still only a very small proportion of all named marine species (Costello *et al.*, 2013). A large proportion of the ESAS dataset is bird taxa (~98%) but incorporating the ORCA dataset into one heterogeneous dataset increases taxonomic coverage, allowing more sightings of marine mammals to be included in the mapping of hotspots. The study highlights the practical use of existing databases with wide spatial and temporal coverage to answer large-scale important ecological questions, such as locations of species richness hotspots and cold spots. The quantity and wide spatial extent of data from the combined databases provides a good representation of richness patterns in the study area.

This is particularly relevant given the focus on biodiversity-driven international targets and legal frameworks (Convention on Biological Diversity, 2020) and the Post-2020 Biodiversity Framework adopted in December 2022 (Convention on Biological Diversity, 2022). Species richness remains the most common measure of biodiversity (Brose *et al.*, 2003) and biodiversity plays a vital role in legislation and conservation designation to date (CMS, 1979, ASCOBANS, 1994, European Commission, 2008, Convention on Biological Diversity, 2020, European Commission, 2020, Convention on Biological Diversity, 2021, OSPAR commission, 2021). With a lack of inter-taxa studies of marine predator coexistence (Cox *et al.*, 2018), this study contributes to filling this knowledge gap and identifying a method of detecting areas of high biodiversity whilst factoring in survey effort bias.

4.5.1 Mapping biodiversity under the influence of survey effort and grid size

This study clearly shows that whilst spatial, temporal and taxonomic coverage is increased when combining multiple datasets and this allows for cross-taxa analysis, biases in sampling effort can impact the patterns of taxonomic species richness observed from heterogeneous datasets. This must be integrated into mapping so that future conservation and legislative actions protect the right areas of high biodiversity. There are clear areas of high species richness directly correlated with high survey effort in this study as seen in other papers (Gaston, 1996). Studies have shown that heterogeneous databases from combined data sources are prone to wide disparities in sampling effort that are not uniform in space (Meier and Dikow, 2004, Soberon and Peterson, 2004, Hortal *et al.*, 2007a) and subsequently can lead to distorted patterns of biodiversity or artefactual results (Hortal *et al.*, 2007a, Boakes *et al.*, 2010, Gotelli, 2011, Ballesteros-Mejia *et al.*, 2013, Yang *et al.*, 2013). Rahbek (1995) noted over a decade ago that many studies have had methodological problems because they did not account for the effect of sampling effort or area on patterns of species richness.

Detectability (particularly for cryptic and wide-ranging marine predators), poorly surveyed areas due to limited funding or accessibility and seasonality are all significant challenges in sighting datasets. Particularly for sighting data, the need for favourable meteorological and oceanographic conditions and wide coverage combined with the short duration of visual observation opportunity (e.g. surfacing cetaceans, diving seabirds) (Barreto *et al.*, 2021) can mean coverage is patchy. Identified areas of higher species richness and, even simpler, maps of observations, can also be used to direct future data efforts and focus survey directions to target gaps in coverage. Hotspots can also focus conservation efforts, such as ensuring protected areas include those identified areas of high biodiversity. Whilst using presence-only occurrence data will almost always underestimate the true area of occupancy for a species because of sampling bias and stochasticity of occupancy of locations over time (Graham and Hijmans, 2006), using richness maps that account for sampling effort (SRPUE maps) still provides a useful relative indication of areas of high and low richness around the UK. Whilst there is much discussion on the best-fit relationship between species richness and measures such as sampling effort or area, this study concluded that the log-linear was the best fit for the relationship between species richness and effort at all spatial scales, as in other studies (Palmer and Dixon, 1990, Ugland *et al.*, 2003). Nicolas *et al.* (2010) similarly standardised fish species richness by dividing richness by the log-transformed total sampled surface, making the relationship between species richness and surface sampled linear. Log-log has been found to overestimate true species richness (Palmer and Dixon, 1990), which could lead to incorrect designation of protective or management measures. In governance, decision makers often use the precautionary principle when there is uncertainty (European Parliament, 2015), and therefore using log-linear avoids this potential overestimation.

As grid cell size reduced, both the maximum and mean species richness reduced. This is expected with more species present in a larger area size, and consistent with the species-area curve, which demonstrates a positive linear relationship between species richness and log(area) (Elliott *et al.*, 2020). Graham and Hijmans (2006) suggests that estimated species richness accuracy increases as the resolution of the grid cell decreases (larger cells), due to the high influence of sampling effort in estimating richness. Whilst using a coarse resolution of grid cell (or grain) sizes can lead to higher richness due to broader coverage of potential assemblages (Soberón *et al.*, 2007) and therefore be useful to pinpointing areas of high biodiversity for protection, it can reduce opportunities for small-scale deduction of richness patterns (Meyer, 2016) (such as isolating small scale drivers of richness like depth or oxygen levels (Gagne *et al.*, 2020)) that may be identified in the maps using smaller grain size. Smaller spatial units (e.g. 1 km^2) may comprise fewer habitat types and thus capture species with closer related traits or life histories, whilst large-scale units may capture more habitat types and thus comprise a wider range of species or taxa (Sessa *et al.*, 2018), and therefore future study into the diversity of functional traits captured within these hotspots would be beneficial. However, this study demonstrates it is possible to use different cell sizes and factor in survey effort (using the SRPUE measure) to output maps at different scales that can reliably be used in subsequent analysis.

Some areas of high SRPUE match patterns observed in a recent distribution study using a similar ESAS-derived dataset by (Waggitt *et al.*, 2019). Waggitt *et al.* used a different approach to derive distribution maps (rather than richness) using detection functions to standardise effort and then used species distribution models (SDM) to overcome issues with heterogeneity and uneven coverage. Harbour porpoise *Phocoena phocoena* for example had high distribution year-round in the North Sea, particularly near the coast of Denmark (Waggitt *et al.*, 2019), whilst white-sided dolphin *Lagenorhynchus acutus* and several seabirds (great skua *Stercorarius skua*, common guillemot, black-legged kittiwake *Rissa tridactyla*) showed very high summer densities around the Orkney Islands, and similar hotspots persisted in SRPUE maps in this study. Therefore, whilst this study aimed to identify hotspots, there can be increased confidence in the approach when areas of high/low densities in Waggitt *et al.* (2020) align with areas of high/low species richness in this study. A clear coastal association is observed in many seabird species in Waggitt *et al.* (2019) particularly during summer months likely due to nesting and breeding sites, and this may drive high coastal hotspots observed in the SRPUE maps in this study. Understanding potential shared drivers of high-density hotspots (e.g. proximity to land, bathymetry, chlorophyll and sea surface temperature) is the logical next step with this analysis (Davidson *et al.*, 2011, Kaschner *et al.*, 2011, Harvey *et al.*, 2017) and is explored in Chapter 5.

4.5.2 Changes in biodiversity patterns over time

Whilst raw species richness appears to have fluctuated over the 39-year period, species richness standardised by unit effort has not, and thus artefacts of sampling effort can skew conclusions of temporal changes in richness. Peaks in data collation around 1980 were seen clearly in the raw richness time-series plots (likely due to an initial drive in data collection when ESAS was established), with collection effort diminishing in 2011 (with removal of data post 2011 significant in the species richness dataset, but not in SRPUE which considers effort in the adjusted measure). This decline could be perhaps due to a shift in funding allocations or a shift in research focus, with less funding given to traditional sighting data surveys and thus less new data being contributed, over improved telemetry or digital survey technology. For example, Laidre *et al.* (2022) describes the advancement in satellite telemetry technologies in 2011 which allowed release devices to be triggered by the user and the addition of salt-water sensors, thermometers, accelerometers, and video cameras: thus focus may have been directed towards advancing these technologies rather than wide-scale boat or aerial surveys. Compensating for these fluctuations in data collection is important when drawing further conclusions from trends in biodiversity.

SRPUE appears to have remained relatively stable over time, at all spatial scales. Recent studies have also found stable patterns in local-scale biodiversity, for mean, net changes and multi-year trends (Vellend *et al.*, 2013, Dornelas *et al.*, 2014), but fewer studies have enumerated changes at regional scales and have mostly focused on islands (Sax and Gaines, 2003, Sax and Gaines, 2008, Vellend *et al.*, 2017). There is less understanding within marine systems with studies demonstrating an increase in diversity in some areas and decline in others (Hiddink and Ter Hofstede, 2008, ter Hofstede *et al.*, 2010, Hiddink and Coleby, 2012). A key challenge with SRPUE is its lack of natural interpretability (and this is discussed further in Chapter 5), as species richness is well recognised on an absolute integer scale, but as a relative measure SRPUE is still useful to demonstrate areas of high and low richness from multi-taxa assemblages, which is important for management, legislation and conservation. Measures of determining relative rather than absolute patterns may be superior, particularly when sampling methods and magnitude have varied over time (Telfer *et al.*, 2002, Boakes *et al.*, 2010).

The stable species richness observed is consistent with several theories. The first is that the 39-year time span is not long enough to see real changes in species richness in such a large area of ocean. Paleontological studies have shown that relatively constant numbers of species for long periods of time are common (Ricklefs, 2010), with the number of mammal species for example varying very little over the last 60 million years (Alroy, 2000), so the patterns of richness over a few decades may be even further concealed. This is intrinsically linked to the second theory: the overall stable diversity reflects the concepts of carrying capacity and niche space. The idea of a carrying capacity (or diversity equilibria) in an ecosystem is well-known and well-deliberated (Storch *et al.*, 2019). Darwin considered carrying capacity 164 years ago (Darwin, 1859), reflecting on what limits the number of species, and theories have circulated since, including the idea of niche space limiting richness (Gavrilets and Vose, 2005). Species fill any vacant niche or “gaps” that appear in the environment depending on abiotic and biotic factors that drive their life history thus species richness levels may remain the same in an area, but the species assemblages vary within it. Partitioning is deemed a stipulation for species coexistence, to reduce interspecific competition (Chesson, 2000). Therefore, the stable patterns observed may echo the carrying capacity of the ecosystem. It has been recently suggested that more region-specific diversity limits (such as energy availability, environmental stability, temperature and topography) determine species richness and macroecological biodiversity patterns, independent of historical lineage diversification, and this has gathered increasing empirical support (Storch *et al.*, 2019). There are short-term examples of marine mammals tracing anthropogenic structures for new foraging areas (Russell *et al.*, 2014) so it is conceivable that if these mammals move out of an existing area, the niche may be filled by other predators that target the same resources. This could be further explored by looking at the functional richness of species assemblages, to assess if, whilst species numbers remain stable, the groupings within them have changed over time. The time-series outputs of this study provide scope for many future ideas: such as what infrastructure (such as offshore wind developments) or other changes in the marine seascape (e.g. marine conservation zones, protected areas) have appeared (or gone) within the four decades and how this could be correlated with changes in species richness or SRPUE.

High seasonal patterns in survey effort were expected and were reflected in the species richness values but were also still evident in SRPUE time series plots, albeit with much less variability. This could suggest there remains a seasonal turnover of species richness assemblages, with seasonal patterns of usage visible and some species migrating into the area for summer months and away for winter months (e.g. Procellariiformes, Mysticetes and transient odontocetes (Reid *et al.*, 2003, Snow and Perrins, 2004, Evans, 2009)). Alternatively, it could be an artefact, even after semi-log transformation, of lower winter sampling, or changes in seasonal sampling in general, due to accessibility, perception or species misidentification bias due to weather and visibility (Reddy and Dávalos, 2003, Shillinger *et al.*, 2012, Thomson *et al.*, 2013, Bennison and Jessopp, 2015).

Scaling unit (100 km^2, 10 km^2, 1 km^2) did not lead to big changes in patterns over time, all showing the same trends just with reduced maximum and mean richness values, anticipated by the species-area curve relationship (Ugland *et al.*, 2003). For practical use, this suggests that selecting the right location for legislative protection (e.g. marine protected areas or marine nature reserves) is critical, given relative hotspots of richness can be identified across grid sizes, and therefore making sure protection covers the centre of those hotspots, may be more important than the popular prevalent "total area protected" metric used to gauge conservation targets (Edgar *et al.*, 2014, Wilhelm *et al.*, 2014, Devillers *et al.*, 2015, Gill *et al.*, 2017, Roberts *et al.*, 2018).

4.5.3 Limitations

Considering both the mean species richness and dispersion of richness patterns can provide insight into species richness across space (Lynch *et al.*, 2014), and therefore assessing any aggregations or spread in the data would be a logical next step. However, grid-based diversity measures are almost always spatially autocorrelated so further study on dispersion using this dataset is limited (Ruggiero, 1999, Rahbek and Graves, 2001, Diniz-filho *et al.*, 2002, Van Rensburg *et al.*, 2002, Diniz-filho *et al.*, 2003), and delimiting the causes of spatial autocorrelation based on the analysis of species richness alone would be very difficult (Diniz-filho *et al.*, 2003). Grid cells are given spatial coordinates based on their centroid, so the cells will have uniform centroids in both the X and Y direction, giving strict regularity to coordinates rather than if points were taken with variable coordinates along a continuous scale. It has been shown that richness values in quadrats or grid systems coupled at a set distance often show autocorrelation, positive or negative (Diniz-filho *et al.*, 2003). For species richness measured in a grid system, spatial autocorrelation can also occur because of low species turnover. Species ranges may be beyond the cell limits (likely for wide-ranging marine predators) so nearby cells may have similar species richness because they share many of the same species as they are not constrained to one cell in their range (Diniz-filho *et al.*, 2003). This large geographical range size relative to chosen cell size means adjacent cells could be pseudo-replicated in space. However, it may also occur due to adjacent cells having similar environmental factors that display spatial autocorrelation, so the drivers of species diversity are in themselves autocorrelated. Additionally, it would be informative to model statistically any residual spatial correlation to explore potential animal movement in terms of aggregation or avoidance – and this provides a future avenue of study. If similar species richness values are caused by different groups of species in relation to chosen environmental factors, then there may be indication of an ecological process of interest. Thus, as aforementioned, understanding the drivers of species richness are crucial when such patterns are observed and is further explored in Chapter 5.

4.6 Conclusion

This study demonstrates the potential of using long-term datasets to answer macroecological scale questions such as revealing hotspots of biodiversity in large marine environments. With the high cost of marine surveys, the importance of utilising existing databases with broad spatial and temporal coverage is significant and can be used successfully to expose hotspots of biodiversity in large areas. Collating and combining survey data is essential for marine predator biodiversity research in the future, as surveys in isolation do not give the broad-scale spatial coverage necessary for such wide-ranging, migrating, marine predators or the temporal coverage to reveal long-term trends in biodiversity.

This study highlights the importance of evaluating and standardising presence-only data before using it in further analysis and provides an approach to factoring in a measure post-survey of effort – which is particularly important in heterogenous databases with patchy coverage where surveys were not designed or carried out to specifically derive richness measures. A new measure, SRPUE, was utilised whereby species richness is adjusted by effort and can be used to identify relative areas of high and low species richness. This has benefits for science and our ability to identify meaningful species richness areas, but also for conservation and management. Failure to factor in survey effort when determining species richness hotspots can be detrimental to marine planning, we need to make sure we are crucially targeting protection measures in the right areas, rather those skewed by increased survey effort. International biodiversity targets require information gaps to be identified and prioritised (Meyer *et al.*, 2015), and thus whilst maps of survey effort can be useful to highlight areas that need increased survey coverage, using the relative measure of SRPUE can allow areas of high biodiversity to be prioritised in an era where quickly and proactively protecting biodiversity is at the forefront of global policy (United Nations, 2022).

The study also revealed the difference between temporal patterns of species richness and SRPUE when derived from the same dataset. It demonstrated that when survey effort is factored in, the concept of carrying capacity is revealed. Species richness appears to have fluctuated over the four decades studied whilst SRPUE has remained stable, suggesting artefacts of sampling effort can skew conclusions of temporal changes in richness. Species may adapt to fill ecological niches that appear in the environment depending on abiotic and biotic factors that drive their life history; and therefore, as this study suggests, richness levels may remain the same in an area, but the species assemblages vary within it.

4.7 References

ALBERT, C., LUQUE, G. M. & COURCHAMP, F. 2018. The twenty most charismatic species. *PLoS One,* 13**,** e0199149.

ALROY, J. 2000. Successive approximations of diversity curves: Ten more years in the library. *Geology,* 28**,** 1023–1026.

ANNONI, A., BERNARD, L., LILLETHUN, A., IHDE, J., GALLEGO, J., RIVES, M., SOMMER, E., POELMAN, H., CONDÉ, S., GREAVES, M., UOTILA, P., PINTO, J. T., TORRES, J. O. A., LAMPINEN, R., KAYADJANIAN, M., SCHMIDT-SEIWERT, V., MEINER, A., LUZET, C. & WIRTHMANN, A. Short Proceedings of the 1st European Workshop on Reference Grids, Ispra, 27-29 October 2003. *In:* EUROPEAN ENVIRONMENT AGENCY, ed. 1st Workshop on European Reference Grids, 2004-03-03 2003 Ispra. JRC-Institute for Environment and Sustainability.

ASCOBANS 1994. Agreement on the Conservation of Small Cetaceans of the Baltic, North East Atlantic, Irish and North Seas. *In:* UNEP (ed.). New York, US: UNEP.

AVILA, I. C., KASCHNER, K. & DORMANN, C. F. 2018. Current global risks to marine mammals: Taking stock of the threats. *Biological Conservation,* 221**,** 44-58.

BALLESTEROS-MEJIA, L., KITCHING, I. J., JETZ, W., NAGEL, P. & BECK, J. 2013. Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. *Global Ecology and Biogeography,* 22**,** 586-595.

BARRETO, J., CAJAÍBA, L., TEIXEIRA, J. B., NASCIMENTO, L., GIACOMO, A., BARCELOS, N., FETTERMANN, T. & MARTINS, A. 2021. Drone-Monitoring: Improving the Detectability of Threatened Marine Megafauna. *Drones,* 5.

BAUM, J. K. & WORM, B. 2009. Cascading top-down effects of changing oceanic predator abundances. *J Anim Ecol,* 78**,** 699-714.

BENNISON, A. & JESSOPP, M. 2015. At-sea surveys confirm a North Atlantic biodiversity hotspot. *Bird Study,* 62**,** 262-266.

BIVAND, R., KEITT, T. & ROWLINGSON, B. 2019. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. *R package version 1.4-8.* .

BIVAND, R. S., PEBESMA, E. & GOMEZ-RUBIO, V. 2013. *Applied spatial data analysis with R* NY, Springer.

BLOCK, B. A., JONSEN, I. D., JORGENSEN, S. J., WINSHIP, A. J., SHAFFER, S. A., BOGRAD, S. J., HAZEN, E. L., FOLEY, D. G., BREED, G. A., HARRISON, A. L., GANONG, J. E., SWITHENBANK, A., CASTLETON, M., DEWAR, H., MATE, B. R., SHILLINGER, G. L., SCHAEFER, K. M., BENSON, S. R., WEISE, M. J., HENRY, R. W. & COSTA, D. P. 2011. Tracking apex marine predator movements in a dynamic ocean. *Nature,* 475**,** 86-90.

BOAKES, E. H., MCGOWAN, P. J., FULLER, R. A., CHANG-QING, D., CLARK, N. E., O'CONNOR, K. & MACE, G. M. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol,* 8**,** e1000385.

BOUCHET, P. J., MEEUWIG, J. J., SALGADO KENT, C. P., LETESSIER, T. B. & JENNER, C. K. 2015. Topographic determinants of mobile vertebrate predator hotspots: current knowledge and future directions. *Biol Rev Camb Philos Soc,* 90**,** 699-728.

BREWER, A. & WILLIAMSON, M. 1994. A new relationship for rarefaction. *Biodiversity and Conservation,* 3**,** 373–379.

BROSE, U., MARTINEZ, N. D. & WILLIAMS, R. J. 2003. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology,* 84**,** 2364-2377.

BUCKLAND, S. T., BURT, M. L., REXSTAD, E. A., MELLOR, M., WILLIAMS, A. E. & WOODWARD, R. 2012. Aerial surveys of seabirds: the advent of digital methods. *Journal of Applied Ecology,* 49**,** 960-967.

BULL, J. C., MASON, S., WOOD, C. & PRICE, A. R. G. 2013. Benthic marine biodiversity patterns across the United Kingdom and Ireland determined from recreational diver observations: A baseline for possible species range shifts induced by climate change. *Aquatic Ecosystem Health & Management,* 16**,** 20-30.

BUNGE, J. & FITZPATRICK, M. 1993. Estimating the number of species; a review. *Journal of the American Statistical Association,* 88**,** 364-373.

CARO, T. 2010. *Conservation by proxy,* Washington, Covelo, London, Island Press.

CHAO, A., COLWELL, R. K., LIN, C.-W. & GOTELLI, N. J. 2009. Sufficient sampling for asymptotic minimum species richness estimators. *Ecology,* 90**,** 1125-1133.

CHAZDON, R. L., COLWELL, R. K., DENSLOW, J. S. & GUARIGUATA, M. R. G. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of northeast Costa Rica. *In:* DALLMEIER, F. & COMISKEY, J. A. (eds.) *Forest biodiversity research, monitoring and modeling.* London, UK: UNESCO and The Parthenon Publishing Group.

CHESSON, P. 2000. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics,* 31:**,** 343–366.

CMS 1979. The Bonn Convention on the Conservation of Migratory Species of Wild Animals. Bonn, Germany: CMS.

COLWELL, R. K. & CODDINGTON, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society A,* B345**,** 101–118.

CONVENTION ON BIOLOGICAL DIVERSITY. 2020. *Aichi Biodiversity Targets* [Online]. Aichi Biodiversity Target,. Available: https://www.cbd.int/sp/targets/ [Accessed 11th March 2021].

CONVENTION ON BIOLOGICAL DIVERSITY. 2021. *Preparations for the Post-2020 Biodiversity Framework* [Online]. Online: Convention on Biological Diversity,. Available: https://www.cbd.int/conferences/post2020 [Accessed 12th March 2021].

CONVENTION ON BIOLOGICAL DIVERSITY. 2022. *Preparations for the Post-2020 Biodiversity Framework* [Online]. UN Environment Programme. Available: https://www.cbd.int/conferences/post2020 [Accessed 16th February 2022].

COSTELLO, M. J., BOUCHET, P., BOXSHALL, G., FAUCHALD, K., GORDON, D., HOEKSEMA, B. W., POORE, G. C., VAN SOEST, R. W., STOHR, S., WALTER, T. C., VANHOORNE, B., DECOCK, W. & APPELTANS, W. 2013. Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. *PLoS One,* 8**,** e51629.

COX, S. L., EMBLING, C. B., HOSEGOOD, P. J., VOTIER, S. C. & INGRAM, S. N. 2018. Oceanographic drivers of marine mammal and seabird habitat-use across shelf-seas: A guide to key features and recommendations for future research and conservation management. *Estuarine, Coastal and Shelf Science,* 212**,** 294-310.

CRAWLEY, M. J. & HARRAL, J. E. 2001. Scale dependence in plant biodiversity. *Science,* 291**,** 864–868.

CROXALL, J. P., BUTCHART, S. H. M., LASCELLES, B., STATTERSFIELD, A. J., SULLIVAN, B., SYMES, A. & TAYLOR, P. 2012. Seabird conservation status, threats and priority actions: a global assessment. *Bird Conservation International,* 22**,** 1–34.

DARWIN, C. 1859. *On the origin of species; Facsimile of the first edition,* Cambridge, MA, Harvard University Press.

DASGUPTA, P. 2021. The Economics of Biodiversity: The Dasgupta Review. London: HM Treasury.

DAVIDSON, R. J., BOYER, A. G., KIM, H., POMPA-MANSILLA, S., HAMILTON, M. J., COSTA, D. P., CEBALLOS, G. & BROWN, J. H. 2011. Drivers and hotspots of extinction risk in marine mammals. *PNAS,* 109**,** 3395-3400.

DEVILLERS, R., PRESSEY, R. L., GRECH, A., KITTINGER, J. N., EDGAR, G. J., WARD, T. & WATSON, R. 2015. Reinventing residual reserves in the sea: are we favouring ease of establishment over need for protection? *Aquatic Conservation: Marine and Freshwater Ecosystems,* 25**,** 480-504.

DINIZ-FILHO, J. A. F., BINI, L. M. & HAWKINS, B. A. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology & Biogeography,* 12**,** 53-64.

DINIZ-FILHO, J. A. F., DE SANT'ANA, C. E. R., DE SOUZA, M. C. & RANGEL, T. F. L. V. B. 2002. Null models and spatial patterns of species richness in South American birds of prey. *Ecology Letters,* 5**,** 47–55.

DORNELAS, M., GOTELLI, N. J., MCGILL, B., SHIMADZU, H., MOYES, F., SIEVERS, C. & MAGURRAN, A. E. 2014. Assemblage time series reveal biodiversity change but not systematic loss. . *Science,* 344**,** 296–299.

EDGAR, G. J., STUART-SMITH, R. D., WILLIS, T. J., KININMONTH, S., BAKER, S. C., BANKS, S., BARRETT, N. S., BECERRO, M. A., BERNARD, A. T., BERKHOUT, J., BUXTON, C. D., CAMPBELL, S. J., COOPER, A. T., DAVEY, M., EDGAR, S. C., FORSTERRA, G., GALVAN, D. E., IRIGOYEN, A. J., KUSHNER, D. J., MOURA, R., PARNELL, P. E., SHEARS, N. T., SOLER, G., STRAIN, E. M. & THOMSON, R. J. 2014. Global conservation outcomes depend on marine protected areas with five key features. *Nature,* 506**,** 216-20.

ELLIOTT, L. H., IGL, L. D. & JOHNSON, D. H. 2020. The relative importance of wetland area versus habitat heterogeneity for promoting species richness and abundance of wetland birds in the Prairie Pothole Region, USA. *The Condor,* 122.

ESRI INC. 2020. ArcGIS Pro (Version 2.7.0). https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview: ESRI Inc.

EUROPEAN COMMISSION 2008. Marine Strategy Framework Directive (2008/56/EC). *DIRECTIVE 2008/56/EC* European Commission, .

EUROPEAN COMMISSION 2020. EU Biodiversity Strategy for 2030. *In:* EUROPEAN COMISSION (ed.). Brussels.

EUROPEAN PARLIAMENT. 2015. *The precautionary principle: Definitions, applications and governance* [Online]. Online: European Parliament,. Available: https://www.europarl.europa.eu/thinktank/en/document/EPRS_IDA(2015)573876 [Accessed 24th March 2023].

EVANS, P. G. H. 2009. Whales, porpoises and dolphins. Order Cetacea. *In:* IS. HARRIS, D. W. Y. (ed.) *Mammals of the British Isles* 4ed. Southampton: The Mammal Society.

GAGNE, T. O., REYGONDEAU, G., JENKINS, C. N., SEXTON, J. O., BOGRAD, S. J., HAZEN, E. L. & VAN HOUTAN, K. S. 2020. Towards a global understanding of the drivers of marine and terrestrial biodiversity. *PLoS One,* 15**,** e0228065.

GASTON, K. J. 1996. Species richness: measure and measurement. *In:* GASTON, K. J. (ed.) *Biodiversity. A Biology of Numbers and Difference.* Oxford, UK: Blackwell Science.

GAVRILETS, S. & VOSE, A. 2005. Dynamic patterns of adaptive radiation. *Proc Natl Acad Sci U S A,* 102**,** 18040-5.

GILL, D. A., MASCIA, M. B., AHMADIA, G. N., GLEW, L., LESTER, S. E., BARNES, M., CRAIGIE, I., DARLING, E. S., FREE, C. M., GELDMANN, J., HOLST, S., JENSEN, O. P., WHITE, A. T., BASURTO, X., COAD, L., GATES, R. D., GUANNEL, G., MUMBY, P. J., THOMAS, H., WHITMEE, S., WOODLEY, S. & FOX, H. E. 2017. Capacity shortfalls hinder the performance of marine protected areas globally. *Nature,* 543**,** 665-669.

GLADSTONE, W., MURRAY, B. R. & HUTCHINGS, P. 2020. Promising yet variable performance of cross-taxon biodiversity surrogates: a test in two marine habitats at multiple times. *Biodiversity and Conservation,* 29**,** 3067-3089.

GOTELLI, N. J. & COLWELL, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters,* 4**,** 379-391.

GOTELLI, N. J. & COLWELL, R. K. 2011. Estimating species richness. *In:* MAGURRAN, A. E. & MCGILL, B. J. (eds.) *Frontiers in measuring biodiversity.* New York, USA: Oxford University Press.

GOTELLI, N. J. A. C., R.K. 2011. Estimating species richness. *In:* MAGURRAN, A. E. & MCGILL, B. J. (eds.) *Biological diversity: frontiers in measurement and assessment.* Oxford, UK: Oxford University Press.

GRAHAM, C. H. & HIJMANS, R. J. 2006. A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography,* 0**,** 060831043455001.

GRAY, J. S., UGLAND, K. I., & LAMBSHEAD, J. 2004. Species accumulation and species area curves: a comment on Scheiner (2003). *Global Ecology and Biogeography,* 13**,** 473-476.

HAFEN, R. 2016. stlplus: Enhanced Seasonal Decomposition of Time Series by Loess. *R package version 0.5.1.*

HARVEY, G. K. A., NELSON, T. A., FOX, C. H. & PAQUET, P. C. 2017. Quantifying marine mammal hotspots in British Columbia, Canada. *Ecosphere,* 8**,** e01884.

HAYS, G. C., SCOTT, R. & HIGHAM, T. 2013. Global patterns for upper ceilings on migration distance in sea turtles and comparisons with fish, birds and mammals. *Functional Ecology,* 27**,** 748-756.

HE, F. & LEGENDRE, P. 1996. On species–area relations. *The American Naturalist,* 148**,** 719–737.

HERBERLING, J. M., MILLER, J. T., NOESGAARD, D., WEINGART, S. B. & SCHIGEL, D. 2019. Data integration enables global biodiversity synthesis. *PNAS,* 118**,** 1-7.

HIDDINK, J. G. & COLEBY, C. 2012. What is the effect of climate change on marine fish biodiversity in an area of low connectivity, the Baltic Sea? *Global Ecology and Biogeography,* 21**,** 637-646.

HIDDINK, J. G. & TER HOFSTEDE, R. 2008. Climate induced increases in species richness of marine fishes. *Global Change Biology,* 14**,** 453-460.

HIGHAM, J. E. & LÜCK, M. 2007. *Marine wildlife and tourism management: Insights from the natural and social sciences,* Wallingford, UK, CABI Publishing.

HIJMANS, R. J. 2020. raster: Geographic Data Analysis and Modeling. *R package version 3.4-5.*

HIJMANS, R. J. & SPOONER, D. M. 2001. Geographic distribution of wild potato species. *American Journal of Botany,* 88**,** 2101-2112.

HILLEBRAND, H., BLASIUS, B., BORER, E. T., CHASE, J. M., DOWNING, J. A., ERIKSSON, B. K., FILSTRUP, C. T., HARPOLE, W. S., HODAPP, D., LARSEN, S., LEWANDOWSKA, A. M., SEABLOOM, E. W., VAN DE WAAL, D. B., RYABOV, A. B. & CADOTTE, M.

2018. Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring. *Journal of Applied Ecology,* 55**,** 169-184.

HORTAL, J., LOBO, J. M. & JIMENEZ-VALVERDE, A. 2007a. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conserv Biol,* 21**,** 853-63.

HORTAL, J., LOBO, J. M. & JIMÉNEZ-VALVERDE, A. 2007b. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). *Conservation Biology,* 21**,** 853–863.

HUBBELL, S. P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography,* Princeton, NJ, Princeton University Press.

HUNT, G. L. & MCKINNELL, S. 2006. Interplay between top-down, bottom- up, and wasp-waist control in control in marine ecosystems. *Progress in Oceanography,* 68**,** 115–124.

IPBES 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. *In:* S. DÍAZ, J. S., E. S. BRONDÍZIO E.S., H. T. NGO, M. GUÈZE, J., AGARD, A. A., P. BALVANERA, K. A. BRAUMAN, S. H. M. BUTCHART, K. M. A. CHAN, L. A. GARIBALDI, K. ICHII, J. LIU, S. M. SUBRAMANIAN,, G. F. MIDGLEY, P. M., Z. MOLNÁR, D. OBURA, A. PFAFF, S. POLASKY, A. PURVIS, J. RAZZAQUE, B. REYERS, R. ROY CHOWDHURY, Y. J. SHIN, & I. J. VISSEREN-HAMAKERS, K. J. W., AND C. N. ZAYAS (eds.). IPBES Secretariat.

IUCN. 2023. *The IUCN Red List of Threatened Species* [Online]. Available: https://www.iucnredlist.org [Accessed 1st February 2023].

JETZ, W. & RAHBEK, C. 2001. Geometric constraints explain much of the species richness pattern in African birds. *Proc Natl Acad Sci U S A,* 98**,** 5661-6.

JEWELL, R., THOMAS, L., HARRIS, C. M., KASCHNER, K., WIFF, R., HAMMOND, P. S. & QUICK, N. J. 2012. Global analysis of cetacean line-transect surveys: detecting trends in cetacean density. *Marine Ecology Progress Series,* 453**,** 227-240.

JNCC 2019. European Seabirds at Sea database (ESAS). *In:* JOINT NATURE CONSERVATION COMMITTEE (JNCC) (ed.). Peterborough, UK.

JNCC 2022. European Seabirds At Sea (ESAS). ICES, Copenhagen, Denmark.

KASCHNER, K., TITTENSOR, D. P., READY, J., GERRODETTE, T. & WORM, B. 2011. Current and future patterns of global marine mammal biodiversity. *PLoS One,* 6**,** e19653.

KREBS, C. J. 2001. *Ecology: The experimental analysis of distribution and abundance,* San Francisco, Pearson Education.

KREMEN, C., CAMERON, A., MOILANEN, A., PHILLIPS, S. J., THOMAS, C. D., BEENTJE, H., DRANSFIELD, J., FISHER, B. L., GLAW, F., GOOD, T. C., HARPER, G. J., HIJMANS, R. J., LEES, D. C., LOUIS JR, E., NUSSBAUM, R. A., RAXWORTHY, C. J., RAZAFIMPAHANANA, A., SCHATZ, G. E., VENCES, M., VIETES, D. R., WRIGHT, P. C. & ZJHRA, M. L. 2008. Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science,* 32**,** 222–226.

LADLE, R. & HORTAL, J. 2013. Mapping species distributions: living with uncertainty. *Frontiers of Biogeography,* 5.

LAIDRE, K. L., DURNER, G. M., LUNN, N. J., REGEHR, E. V., ATWOOD, T. C., RODE, K. D., AARS, J., ROUTTI, H., WIIG, Ø., DYCK, M., RICHARDSON, E. S., ATKINSON, S., BELIKOV, S. & STIRLING, I. 2022. The Role of Satellite Telemetry Data in 21st Century Conservation of Polar Bears (Ursus maritimus). *Frontiers in Marine Science,* 9.

LEONARD, D. R., ROBERT CLARKE, K., SOMERFIELD, P. J. & WARWICK, R. M. 2006. The application of an indicator based on taxonomic distinctness for UK marine biodiversity assessments. *J Environ Manage,* 78**,** 52-62.

LETESSIER, T. B., MOUILLOT, D., BOUCHET, P. J., VIGLIOLA, L., FERNANDES, M. C., THOMPSON, C., BOUSSARIE, G., TURNER, J., JUHEL, J. B., MAIRE, E., CALEY, M. J., KOLDEWEY, H. J., FRIEDLANDER, A., SALA, E. & MEEUWIG, J. J. 2019. Remote reefs and seamounts are the last refuges for marine predators across the Indo-Pacific. *PLoS Biol,* 17**,** e3000366.

LIMPERT, E., WERNER, A., STAHEL, W. A. & ABBT, M. 2001. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience,* 51**,** 341-352.

LOBO, J. M. 2008. Database records as a surrogate for sampling effort provide higher species richness estimations. *Biodiversity and Conservation,* 17**,** 873-881.

LOBO, J. M., HORTAL, J., YELA, J. L., MILLÁN, A., SÁNCHEZ-FERNÁNDEZ, D., GARCÍA-ROSELLÓ, E., GONZÁLEZ-DACOSTA, J., HEINE, J., GONZÁLEZ-VILAS, L. &

GUISANDE, C. 2018. KnowBR: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. *Ecological Indicators,* 91**,** 241-248.

LOMOLINO, M. V. 2004. Conservation biogeography. *In:* IN: LOMOLINO, M. V., HEANEY, L.R (ed.) *Frontiers of Biogeography: New Directions in the Geography of Nature.* Sinauer Associates.

LYNCH, H. J., THORSON, J. T. & SHELTON, A. O. 2014. Dealing with under- and over-dispersed count data in life history, spatial, and community ecology. *Ecology,* 95**,** 3173-3180.

MAGURRAN, A. E. 1988. *Ecological diversity and its measurement*, Princeton university press.

MAGURRAN, A. E., BAILLIE, S. R., BUCKLAND, S. T., DICK, J. M., ELSTON, D. A., SCOTT, E. M., SMITH, R. I., SOMERFIELD, P. J. & WATT, A. D. 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends Ecol Evol,* 25**,** 574-82.

MCBRIDE-KEBERT, S., TAYLOR, J. S., LYN, H., MOORE, F. R., SACCO, D. F., KAR, B. & KUCZAJ II, S. A. 2019. Controlling for Survey Effort Is Worth the Effort: Comparing Bottlenose Dolphin (Tursiops truncatus) Habitat Use Between Standardized and Opportunistic Photographic-Identification Surveys. *Aquatic Mammals,* 45**,** 21-29.

MEIER, R. & DIKOW, T. 2004. Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. *Conservation Biology,* 18**,** 478-488.

MEYER, C. 2016. Limitations in global information on species occurrences. *Frontiers in biogeography,* 8**,** e28195.

MEYER, C., KREFT, H., GURALNICK, R. & JETZ, W. 2015. Global priorities for an effective information basis of biodiversity distributions. *Nat Commun,* 6**,** 8221.

MICROSOFT CORPORATION 2018. Microsoft Excel. 16.70 ed.

MOORE, S. E. 2008. Marine mammals as ecosystem sentinels. *Journal of Mammalogy,* 89**,** 534-540.

NICOLAS, D., LOBRY, J., LEPAGE, M., SAUTOUR, B., LE PAPE, O., CABRAL, H., URIARTE, A. & BOËT, P. 2010. Fish under influence: A macroecological analysis of relations between fish species richness and environmental gradients among European tidal estuaries. *Estuarine, Coastal and Shelf Science,* 86**,** 137-147.

ORCA 2019. ORCA Full Dataset. Portsmouth, UK.

OSPAR COMMISSION 2021. Convention for the Protection of the Marine Environment of the North-East Atlantic - OSPAR Convention. *In:* COMMISSION, O. (ed.). London: OSPAR Commission.

PALMER, M. W. 1990. The estimation of species richness by extrapolation. *Ecology,* 71**,** 1195–1198.

PALMER, M. W. & DIXON, P. M. 1990. Small-scale environmental heterogeneity and the analysis of species distributions along gradients. *Journal of Vegetation Science,* 1**,** 57-65.

PEBESMA, E. J. & BIVAND, R. S. 2005. Classes and methods for spatial data in R. *R News,* 5.

PHILLIPS, S. J., DUDÍK, M., ELITH, J., GRAHAM, C. H., LEHMANN, A., LEATHWICK, J., & FERRIER, S. (2009). SAMPLE SELECTION BIAS AND PRESENCE-ONLY DISTRIBUTION MODELS: IMPLICATIONS FOR BACKGROUND AND PSEUDO-ABSENCE DATA. ECOLOGICAL APPLICATIONS, 19(1), 181-197. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. . *Ecological applications,* 19**,** 181-197.

PIELOU, E. C. 1966. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology,* 13**,** 131-144.

PURVIS, A. & HECTOR, A. 2000. Getting the measure of biodiversity. *Nature,* 405**,** 212.

R CORE TEAM 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

R CORE TEAM 2022. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

RAHBEK, C. 1995. The elevational gradient of species richness: a uniform pattern? *Ecography,* 18**,** 200-205.

RAHBEK, C. & GRAVES, G. R. 2001. Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences,* 98**,** 4534-4539.

REDDY, S. & DÁVALOS, L. M. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography,* 30**,** 1719-1727.

REESE, D. C. & BRODEUR, R. D. 2006. Identifying and characterizing biological hotspots in the northern California Current. *Deep Sea Research Part II: Topical Studies in Oceanography,* 53**,** 291-314.

REID, J. B., EVANS, P. G. H. & NORTHRIDGE, S. P. 2003. Atlas of Cetacean distribution in north-west European waters. *JNCC.* Peterborough, UK: JNCC.

REISINGER, R. R., RAYMOND, B., HINDELL, M. A., BESTER, M. N., CRAWFORD, R. J. M., DAVIES, D., DE BRUYN, P. J. N., DILLEY, B. J., KIRKMAN, S. P., MAKHADO, A. B., RYAN, P. G., SCHOOMBIE, S., STEVENS, K., SUMNER, M. D., TOSH, C. A., WEGE, M., WHITEHEAD, T. O., WOTHERSPOON, S., PISTORIUS, P. A. & SCHOEMAN, D. 2018. Habitat modelling of tracking data from multiple marine predators identifies important areas in the Southern Indian Ocean. *Diversity and Distributions,* 24**,** 535-550.

RICKLEFS, R. E. 2010. Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *Proc Natl Acad Sci U S A,* 107**,** 1265-72.

ROBERTS, K. E., VALKAN, R. S. & COOK, C. N. 2018. Measuring progress in marine protection: A new set of metrics to evaluate the strength of marine protected area networks. *Biological Conservation,* 219**,** 20-27.

ROCCHINI, D., HORTAL, J., LENGYEL, S., LOBO, J. M., JIMÉNEZ-VALVERDE, A., RICOTTA, C., BACARO, G. & CHIARUCCI, A. 2011. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography: Earth and Environment,* 35**,** 211-226.

ROSENZWEIG, M. L. 1995. *Species Diversity in Space and Time,* Cambridge, Cambridge University Press.

RUGGIERO, A. 1999. Spatial patterns in the diversity of mammal species: a test of the geographic area hypothesis in South America. *Ecoscience,* 6**,** 338–354.

RUSSELL, D. J. F., BRASSEUR, S., THOMPSON, D., HASTIE, G. D., JANIK, V. M., AARTS, G., MCCLINTOCK, B. T., MATTHIOPOULOS, J., MOSS, S. E. W. & MCCONNELL, B. 2014. Marine mammals trace anthropogenic structures at sea. *Curr Biol,* 24**,** R638-R639.

SANTORA, J. A. & VEIT, R. R. 2013. Spatio-temporal persistence of top predator hotspots near the Antarctic Peninsula. *Marine Ecology Progress Series,* 487**,** 287-304.

SAX, D. F. & GAINES, S. D. 2003. Species diversity: from global decreases to local increases. *Trends in Ecology & Evolution,* 18**,** 561-566.

SAX, D. F. & GAINES, S. D. 2008. Colloquium paper: species invasions and extinction: the future of native biodiversity on islands. *Proc Natl Acad Sci U S A,* 105 Suppl 1**,** 11490-7.

SCALES, K. L. 2016. *The application of ocean front metrics for understanding habitat selection by marine predators.* PhD, Plymouth University.

SCHEINER, S. M. 2003. Six types of species-area curves. *Global Ecology & Biogeography,* 12**,** 441–447.

SESSA, E. B., CHAMBERS, S. M., LI, D., TROTTA, L., ENDARA, L., BURLEIGH, J. G. & BAISER, B. 2018. Community assembly of the ferns of Florida. *Am J Bot,* 105**,** 549-564.

SHILLINGER, G. L., BAILEY, H., BOGRAD, S. J., HAZEN, E. L., HAMANN, M., GASPAR, P., GODLEY, B. J., WILSON, R. P. & SPOTILA, J. R. 2012. Tagging through the stages: technical and ecological challenges in observing life histories through biologging. *Marine Ecology Progress Series,* 457**,** 165-170.

SNOW, D. & PERRINS, C. M. 2004. *The Birds of the Western Palaearctic (Concise Edition). ,* Oxford, Oxford University Press.

SOBERÓN, J., JIMÉNEZ, R., GOLUBOV, J. & KOLEFF, P. 2007. Assessing completeness of biodiversity databases at different spatial scales. *Ecography,* 30**,** 152-160.

SOBERON, J. & PETERSON, A. T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philos Trans R Soc Lond B Biol Sci,* 359**,** 689-98.

STORCH, D., OKIE, J. G. & FIELD, R. 2019. The carrying capacity for species richness. *Global Ecology and Biogeography,* 28**,** 1519-1532.

TELFER, M. G., PRESTON, C. D. & ROTHERY, P. 2002. A general method for measuring relative change in range size from biological atlas data. *Biological Conservation,* 107**,** 99-109.

TER HOFSTEDE, R., HIDDINK, J. G. & RIJNSDORP, A. D. 2010. Regional warming changes fish species richness in the eastern North Atlantic Ocean. *Marine Ecology Progress Series,* 414**,** 1-9.

THIERS, L., DELORD, K., BOST, C.-A., GUINET, C. & WEIMERSKIRCH, H. 2016. Important marine sectors for the top predator community around Kerguelen Archipelago. *Polar Biology,* 40**,** 365-378.

THOMSON, J. A., COOPER, A. B., BURKHOLDER, D. A., HEITHAUS, M. R. & DILL, L. M. 2013. Correcting for heterogeneous availability bias in surveys of long-diving marine turtles. *Biological Conservation,* 165**,** 154-161.

TITTENSOR, D. P., MORA, C., JETZ, W., LOTZE, H. K., RICARD, D., BERGHE, E. V. & WORM, B. 2010. Global patterns and predictors of marine biodiversity across taxa. *Nature,* 466**,** 1098-101.

TITTENSOR, D. P., WALPOLE, M., HILL, S. L., BOYCE, D. G., BRITTEN, G. L., BURGESS, N. D., BUTCHART, S. H., LEADLEY, P. W., REGAN, E. C., ALKEMADE, R., BAUMUNG, R., BELLARD, C., BOUWMAN, L., BOWLES-NEWARK, N. J., CHENERY, A. M., CHEUNG, W. W., CHRISTENSEN, V., COOPER, H. D., CROWTHER, A. R., DIXON, M. J., GALLI, A., GAVEAU, V., GREGORY, R. D., GUTIERREZ, N. L., HIRSCH, T. L., HOFT, R., JANUCHOWSKI-HARTLEY, S. R., KARMANN, M., KRUG, C. B., LEVERINGTON, F. J., LOH, J., LOJENGA, R. K., MALSCH, K., MARQUES, A., MORGAN, D. H., MUMBY, P. J., NEWBOLD, T., NOONAN-MOONEY, K., PAGAD, S. N., PARKS, B. C., PEREIRA, H. M., ROBERTSON, T., RONDININI, C., SANTINI, L., SCHARLEMANN, J. P., SCHINDLER, S., SUMAILA, U. R., TEH, L. S., VAN KOLCK, J., VISCONTI, P. & YE, Y. 2014. A mid-term analysis of progress toward international biodiversity targets. *Science,* 346**,** 241-4.

TJØRVE, E. 2003. Shapes and functions of species-area curves: a review of possible models. *Journal of Biogeography,* 30**,** 827-835.

UGLAND, K. I., GRAY, J. S. & ELLINGSEN, K. E. 2003. The species–accumulation curve and estimation of species richness. *Journal of Animal Ecology,* 72**,** 888-897.

UNITED NATIONS. 2022. *COP27: Protecting biodiversity is protecting the Paris Agreement* [Online]. Online: United Nations,. Available: https://news.un.org/en/story/2022/11/1130677 [Accessed 6th February 2023].

VAN RENSBURG, B. J., CHOWN, S. L. & GASTON, K. J. 2002. Species richness, environmental correlates, and spatial scale: a test using South African birds. *American Naturalist,* 159**,** 566–577.

VELLEND, M., BAETEN, L., BECKER-SCARPITTA, A., BOUCHER-LAONDE, V., MCCUNE, J. L., MESSIER, J., MYERS-SMITH, I. H. & SAX, D. F. 2017. Plant biodiversity change across scales during the Athropocene. *Annual Review of Plant Biology,* 68**,** 563-586.

VELLEND, M., BAETEN, L., MYERS-SMITH, I. H., ELMENDORF, S. C., BEAUSEJOUR, R., BROWN, C. D., DE FRENNE, P., VERHEYEN, K. & WIPF, S. 2013. Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proc Natl Acad Sci U S A,* 110**,** 19456-9.

WAGGITT, J. J., EVANS, P. G. H., ANDRADE, J., BANKS, A. N., BOISSEAU, O., BOLTON, M., BRADBURY, G., BRERETON, T., CAMPHUYSEN, C. J., DURINCK, J., FELCE, T., FIJN, R. C., GARCIA-BARON, I., GARTHE, S., GEELHOED, S. C. V., GILLES, A., GOODALL, M., HAELTERS, J., HAMILTON, S., HARTNY-MILLS, L., HODGINS, N., JAMES, K., JESSOPP, M., KAVANAGH, A. S., LEOPOLD, M., LOHRENGEL, K., LOUZAO, M., MARKONES, N., MARTÍNEZ-CEDEIRA, J., Ó CADHLA, O., PERRY, S. L., PIERCE, G. J., RIDOUX, V., ROBINSON, K. P., SANTOS, M. B., SAAVEDRA, C., SKOV, H., STIENEN, E. W. M., SVEEGAARD, S., THOMPSON, P., VANERMEN, N., WALL, D., WEBB, A., WILSON, J., WANLESS, S., HIDDINK, J. G. & PUNT, A. 2019. Distribution maps of cetacean and seabird populations in the North-East Atlantic. *Journal of Applied Ecology,* 57**,** 253-269.

WHITE, H. J., MONTGOMERY, W. I., PAKEMAN, R. J. & LENNON, J. J. 2018. Spatiotemporal scaling of plant species richness and functional diversity in a temperate semi-natural grassland. *Ecography,* 41**,** 845-856.

WHITTAKER, R. J., ARAÚJO, M. B., JEPSON, P., LADLE, R. J., WATSON, J. E. M. & WILLIS, K. J. 2005. Conservation Biogeography: assessment
and prospect. *Diversity and Distributions,* 11**,** 3-23.

WILHELM, T. A., SHEPPARD, C. R. C., SHEPPARD, A. L. S., GAYMER, C. F., PARKS, J., WAGNER, D. & LEWIS, N. A. 2014. Large marine protected areas - advantages and challenges of going big. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 24**,** 24-30.

WOOD, S. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B),* 73**,** 3-36.

WORLD ECONOMIC FORUM 2021. The Global Risks Report 2021. *In:* WORLD ECONOMIC FORUM (ed.) *The Global Risks Report.* 16 ed.

WORM, B., LOTZE, H. K. & MYERS, R. A. 2003. Predator diversity hotspots in the blue ocean. *Proc Natl Acad Sci U S A,* 100**,** 9884-8.

YANG, W., MA, K., KREFT, H. & DANIEL KISSLING, W. 2013. Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *Journal of Biogeography,* 40**,** 1415-1426.

Chapter 5 Environmental drivers of species richness around the UK, using three methods of prediction

5.1 Abstract

Marine ecosystems have been rapidly impacted by accelerated climate change due to anthropogenic cumulative stressors, with effects on ecosystem function and marine biodiversity. Top marine predators serve as sentinels of ecosystem change, and spatial aggregations of marine predators identify not only areas that are important to individual predator species themselves, but also areas of broader ecosystem importance. It is important to identify areas of high biodiversity where multiple species co-exist so that conservation and management, which are often limited in resources, can be focused appropriately on the right areas. Additionally, understanding the potential environmental drivers of species richness hotspots can provide invaluable information for understanding both current and future patterns in species richness. This study uses two modelling approaches to identify patterns of species richness: macro-ecological models (MEMs) and stacked species distribution models (SSDMs). MEMs are useful tools for modelling overall relationships between biodiversity and environment at a community level, but loses any individual species responses, whilst SSDMs are computationally demanding but allows exploration of individual patterns on a species level. Two MEMs are utilised: a generalised additive model (GAM) utilising species richness per unit effort (SRPUE) maps from Chapter 4 (which corrects for effort in advance of modelling, but the output is not on the well-recognised interpretable integer scale) and a GAM which includes effort as a predictor and allows outputs on the integer scale (though a trade-off with amount of data retained is required). The study also identifies the potential environmental drivers (bathymetry, chlorophyll *a*, distance to land, fronts, mixed layer thickness, salinity, sea surface temperature and slope) of species richness from all methods. When quantifying cross-taxa species richness, the MEMs and SSDMs show similar spatial richness patterns when mapped comparatively. Investigating environmental drivers of species richness demonstrated all variables are significant predictors for biodiversity, and the SSDM model demonstrated sea surface temperature has the highest relative contribution to richness hotspots. This study demonstrates how analysis of existing and diverse data can be utilised cost-effectively to map biodiversity. While some caution is needed, these novel biodiversity maps have widespread and immediate applications in identifying important areas of protection and providing focus for marine management strategies.

5.2 Introduction

Marine ecosystems have been rapidly impacted by accelerated climate change due to anthropogenic cumulative stressors such as overfishing, land-based pollution, chemical pollution, direct disturbance and shipping (Croxall *et al.*, 2012, Halpern *et al.*, 2015, Avila *et al.*, 2018, Halpern *et al.*, 2019, O'Hara *et al.*, 2021) causing rising ocean temperatures (Hansen *et al.*, 2006), declining oxygen and ocean acidification (Breitburg *et al.*, 2018, Hill and Hoogenboom, 2022, Nagelkerken and Connell, 2022, Scherer *et al.*, 2022). These stressors drive other changes including increasing thermal stratification (De Dominicis *et al.*, 2018), melting sheet ice (Stokes *et al.*, 2022), changes in the salinity of surface waters (Du *et al.*, 2019) and influences on global current systems such as El Niño–Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), and Pacific Decadal Oscillation (PDO) (Overland *et al.*, 2010). These changes have effects on ecosystem function and can trigger biological responses such as changes in phenology, productivity, population connectivity, local adaptation and speciation (reviewed in Hoegh-Guldberg and Bruno (2010)) which threaten marine biodiversity.

Top marine predators such as seabirds, marine mammals and sharks can serve as sentinels of ecosystem change, with their long lifespans, wide-ranging mobility and ability to integrate impacts from the bottom to the top of the food web (Hazen *et al.*, 2019, Sydeman *et al.*, 2021). Thus, spatial aggregations of marine predators identify not only areas that are important to individual predator species themselves (who depend on prey at lower trophic levels (Cury *et al.*, 2011)) but also areas of broader ecosystem importance such as areas of high productivity or biomass (Nicol *et al.*, 2000, Hindell *et al.*, 2020). They have crucial ecological roles including top-down regulation of lower trophic levels (Hunt and McKinnell, 2006) and nutrient transport (Doughty *et al.*, 2016) and are often charismatic species with high socio-economic and cultural significance (Higham and Lück, 2007), therefore are of high conservation importance. However, apex predators are particularly vulnerable to human-related threats due to life history traits such as few offspring, slow growth, and late age to maturity (McClellan *et al.*, 2014). In addition, their consistent migration pathways can mean that adapting to changes in bottom-up effects caused by changes in the distribution of their prey (Luczak *et al.*, 2011, Evans and Bjørge, 2013) and/or to shifts in environmental conditions (MacLeod, 2009, Soldatini *et al.*, 2016) may be more challenging.

Several studies have looked at the impact of climate change on multiple taxa of marine predators. Poloczanska *et al.* (2013, 2016) carried out global metanalyses of marine climate impacts on biodiversity on seabirds and marine mammals, but highlighted observational data for marine mammals was sparse at the time of review, whilst Orgeret *et al.* (2022) conducted a systematic review using 484 studies to assess the effects of climate change on seabirds and marine mammals, finding that species with narrow thermal ranges and relatively long generation times were more often reported to be affected by climate change.

However, a rapidly changing marine environment may change the ecological niches species occupy. In response species can adjust to a changing climate by moving to new areas within their thermal niche (range shifts), acclimate or adapt in place their physiology or behaviour, or become extinct (Donelson *et al.*, 2019, Pinsky *et al.*, 2022). Therefore, it is vital to map current species biodiversity and identify areas of critical marine habitat utilised by many species. An understanding of species co-existence or shared space used is required if marine spatial management and conservation is to be effective (Gissi *et al.*, 2019, Sadykova *et al.*, 2020), particularly with the focus within legislative approaches to conserving biodiversity being based upon area protection (e.g. marine protected areas (MPAs), Special Areas of Conservation (SACs) and Special Protection Areas (SPAs)). It is important to identify areas of high biodiversity where multiple species co-exist so that conservation and management, which is often limited in resources, can be focused appropriately (Evans and Hammond, 2004).

Additionally, understanding potential environmental drivers of species richness hotspots can provide invaluable information for understanding both current and future patterns in species richness (Gagne *et al.*, 2020, Windsor *et al.*, 2022). A key challenge in biogeography is to understand and predict the potential impacts of global change on the distribution of biological diversity (Biber *et al.*, 2019) and identifying the environmental drivers that may impact richness may provide insight into future species richness patterns (Gagne *et al.*, 2020). García-Barón *et al.*, (2020) for example identified high-value biodiversity areas for seabird and cetaceans in Spanish waters and the Bay of Biscay and found the main drivers of species abundance were sea surface temperature (SST), distance to shelf-break and chlorophyll *a* concentration (Chl *a*). In a study aiming to identify important pelagic areas in the eastern Chukchi and western Beaufort seas (Kuletz *et al.*, 2015), locations of hotspots for seabirds and marine mammals are located over underwater canyons or continental shelf features or slopes, likely due to the upwelling and currents with areas of high euphausiid density in summer and autumn months.

Spatial patterns of biodiversity and their environmental drivers have historically concentrated on one or just a few taxa (Sutcliffe *et al.*, 2014) with analysis at species level using a wide range of methods (Hamilton *et al.*, 2022), though studies on biodiversity patterns in marine species at a large cross-taxa scale do exist (e.g. Tittensor *et al.* (2010), Sutcliffe *et al.* (2014) and McClellan *et al.* (2014)). Cross-taxa studies on top marine predators are rarer but Laran *et al.* (2017) mapped the taxonomic richness of marine mammals and seabirds (alongside distribution and density) using multi-taxa aerial surveys to reveal different "hotspots" in the tropical Southwest Indian Ocean. Santora and Veit (2013) used shipboard surveys off the Antarctic Peninsula to quantify species richness and abundance of seabirds and marine mammals to identify marine areas that are persistently attractive to top predators. Albouy *et al.* (2017) compiled geographic range maps from the IUCN database for 128 marine mammal species to identify areas of high species richness and phylogenetic and functional diversity. Focusing on marine predators in the North Pacific Ocean, Block *et al.* (2011) linked ocean features to multispecies hotspots in the California Current large marine ecosystem and the North Pacific transition zone, using tracking data for 23 species of fish, pinnipeds, sharks, seabirds, turtles and cetaceans. Integration of tracking data from 17 bird and mammal species has revealed areas of ecological significance (areas of high biodiversity and prey biomass) in sub-Antarctic islands in the Atlantic and Indian Oceans and over the Antarctic continental shelf (Hindell *et al.*, 2020). Recently Hamilton *et al.* (2022) calculated hotspots and areas of high species richness for Arctic marine mammals using available biotelemetry data, to serve as a benchmark for future shifts. There are even fewer studies that have focused on wide-scale marine predator space use around the UK, with Waggitt *et al*. (2019) stating their recent study provides the most comprehensive distribution maps of cetacean and seabird in the North-East Atlantic. The study collated diverse survey data to identify distributions of 12 cetacean and 12 seabird species but does not identify areas of shared-space use or biodiversity. Matear *et al.* (2019) used opportunistic data from vessels to investigate spatial patterns of cetacean biodiversity and identify important marine mammal areas in the Bay of Biscay, but the study was limited to one taxon. Pinsky *et al.* (2022) highlights that cross-taxa and cross-realm research can provide insights to advancing our understanding of processes or mechanisms of species survival, and large overall changes affecting multiple species may go undetected or undervalued (Hamilton *et al.*, 2022).

However, as discussed in detail in Chapter 2, the cost of data collection, the wide coverage required, and the complexity of monitoring large marine predators in the ocean environment can prevent straightforward mapping of biodiversity, with limited data at the appropriate spatial and temporal resolution required (Matear *et al.*, 2019). Large-scale data sources such as Global Biodiversity Information Facility (GBIF) or Ocean Biogeographic Information System (OBIS) are inherently biased (Meyer *et al.*, 2015) but more quality-controlled datasets such as the European Seabird at Sea (ESAS) and Organisation Cetacea (ORCA) data may facilitate the adequate required coverage for exploring biodiversity patterns in wide-ranging marine predators. Collecting data for marine megafauna species requires wide-scale coverage, but some large-scale repeated surveys (such as SCANS surveys (Hammond, 2021, Gilles *et al.*, 2023)) have allowed high-value biodiversity areas to be identified in consistent areas (McClellan *et al.*, 2014, Kuletz *et al.*, 2015, García-Barón *et al.*, 2020). McClellan *et al.* (2014) used stranding data gave an index of marine megafauna biodiversity in the Channel region, using umbrella groups of large pelagic fish, cetaceans, marine turtles, seabirds and pinnipeds. It must be noted that the study urges caution when interpreting the maps, with data being necessarily coarse in resolution due to patchy data and the stranding data reflecting an increase in observer effort.

Though biodiversity is multifaceted and comprises genetic, taxonomic, phylogenetic and ecological components, species richness remains a commonly used metric to detect biodiversity trends. For example, the importance of richness as a measure of biodiversity in monitoring and trend analyses was assessed through a systematic literature analysis (Hillebrand *et al.*, 2018), and based on a population of 14,720 articles with ‘biodiversity’ or ‘biological diversity’ in the title, species richness (or species number) dominated measures (22.4%). When further subdivided to include mention of ‘monitoring’ or ‘assessing’ or ‘trend’, it was the most frequently used measure of biodiversity (32.8%), over Simpson or Shannon Indices, species composition or species turnover. Species richness is easily observed and recorded, making it a relatively affordable way to monitor change and species richness remains a simple but essential measure to quantifying biodiversity around the UK. Species richness has been identified as one of the Essential Biodiversity Variables (EBVs) (Pereira *et al.*, 2013). EBVs aim to link raw data with indicators and provide robust coordinated data about global biodiversity change, to aid policymakers (Geijzendorffer *et al.*, 2016, Brummitt *et al.*, 2017).

Biodiversity spatial models tend to follow three strategies: “assemble first, predict later” where species are combined into community-level measures such as species richness first (e.g. macro-ecological models (MEMs)), “predict first, assemble later” where tends to model species individually first and then combine to a community level (e.g. Stacked Species Distribution Models (SSDMs)), and “assemble and predict together” where all species are modelled simultaneously in an integrated model and biodiversity is derived from species-level predictions (Ferrier and Guisan, 2006, D'Amen *et al.*, 2015, Zhang, 2022). A combination of MEMs (that used both uncorrected species richness and the corrected SRPUE measure) and SSDM approaches are used to map richness and identify potential environmental drivers of patterns in this study.

5.2.1 Aims of the study

In this study, I utilise species richness grids from Chapter 4 (quantifying cross-taxa species richness) and test hypotheses on environmental drivers of these patterns using two approaches: MEMs and SSDMs. The measure SRPUE discussed in Chapter 4 factors in survey effort bias, but outputs are more difficult to interpret in a natural way given they are not on an integer scale. Therefore, this chapter explores an alternative approach using a GAM with survey effort as a predictor to compare if outputs are comparable with SRPUE.

Further to this, I aim to examine the mechanisms underpinning species richness patterns, by looking at potential environmental drivers of biodiversity and how this varies by species assemblages.

Therefore, the specific hypotheses of this study are:

1. That SSDMs are computationally hard but retain important biological information and allows individualistic species responses to be interrogated, whilst MEMs allow for more rapid analysis at a community level but lose species-specific responses and detailed information, but crucially the two approaches provide equivalent maps of species richness and therefore are equally suitable for identifying hotspots. I will compare the outputs of relative species richness from the three approaches.

2. Output maps from MEMs that include survey effort as a predictor within modelling are equivalent to factoring in survey effort in advance of modelling (using SRPUE as response variable). To look at this I compare the two approaches for handling sampling effort using MEMs.
3. SSDM and MEMs will both identify significant environmental drivers but will differ with environmental variable importance due to method. Identifying drivers from MEMs involves identifying potential significant relationships between areas of high species richness and environmental variables. From SSDMs, this involves assessing variable importance both for overall species richness patterns, and when grouped by life history traits such as taxa, trophic level, central place forager, foraging depth, foraging strategy, diet and IUCN status.

   I hypothesise, from variables identified in the literature, that:

   - In terms of taxa: bathymetry and SST will be strongly associated with marine mammal and seabird richness. For seabirds, Chl *a*, mixed-layer thickness and fronts will be of high importance. For pinnipeds, salinity will be of high importance. Slope will be important to diving cetaceans.
   - No difference between trophic levels, as each occupies a different niche that is influenced by the environmental variables.
   - Distance to land is expected to be most important to species that haul out or are CPFs.

- Regarding foraging depth: SST, mixed-layer thickness, chlorophyll *a* is expected to be most important for surface foraging species, whilst bathymetry/slope is more important for subsurface foraging species.
- Bathymetry and slope are expected to be important for cetaceans that lunge or bubble feed, whilst chl *a* is expected to be most important filter-feeding species.
- For diet, it is expected that chl *a* and SST are most crucial for planktivores, but environment variables are otherwise equal between carnivores and omnivores.
- It is expected that the most endangered species will have high SST and salinity importance, which are under threat from climate change. Bathymetry, distance to land and slope are static variables and therefore are less tied to IUCN status.

5.3 Method

5.3.1 Predictive environmental variables

Both the MEM and SSDM methods require predictive environmental variables for exploring drivers of species richness.

Potential predictive oceanic environmental variables were identified from the literature (Table 5.1), and only plausible relationships showing recognised associations between animals and environmental conditions were chosen. Multidimensional raster layers for each environmental variable (Table 5.2) spanning the 39 years of species observation data (1979 to 2018) (detailed in Chapter 3 and Chapter 4) were imported into ArcGIS Pro (ESRI, 2022) from the various data providers detailed below in Table 5.2, and then aggregated to give a raster of mean values. Data was chosen to span as much of the period as possible, but satellite imagery was not available for some of the early years and so datasets were chosen based upon the maximum temporal coverage possible.

Bathymetry values (seabed depth in metres) were sourced from GEBCO and provided at approximately 15 arc-second grid resolution (GEBCO Compilation Group, 2021). Two further environmental conditions were derived from this product: distance to land and slope (as a proxy for seabed roughness). Distance to land was calculated using the ‘proximity analysis’ tool whilst slope was calculated using the ‘Slope tool’ in Spatial

Analyst in ArcGIS Pro, with gradients calculated from GEBCO bathymetry using the degrees between the focal cell and its neighbouring cells. A higher slope degree angle indicated a stronger gradient. From the Copernicus Marine Service database (Copernicus Marine Service, 2023), Chl *a* values were obtained from the ‘Global Ocean Physics Reanalysis’ Copernicus dataset (Lea, 2015) and salinity and mixed layer thickness values from ‘Global Ocean Chlorophyll, PP and PFT’ dataset (CMEMS, 2021). Monthly composite thermal front maps and SST data were obtained via formal data request from NERC Earth Observation Data Acquisition and Analysis Service (NEODAAS, 2022) at 1.1 km resolution. The composite front maps are processed from European Space Agency (ESA) SST-CCI L4 data and combines the location, gradient, persistence and proximity of all fronts observed over a period into a single map (Miller, 2009). Monthly composites of SST data use combined data from the Advanced-Very-High-Resolution-Radiometer (AVHRR) and Along Track Scanning Radiometer (ATSR) SST-CCI Climate Data Records, using a data assimilation method to provide SSTs where there were no measurements (Merchant *et al.*, 2019).

Rasters of mean values for the 39-year time period were created for each environmental variable, cropped to the area of interest (area containing species observations, see Appendix A (AOI)) and exported as GeoTiffs for statistical modelling (the SSDM package requires a rgdal compatible format (Bivand *et al.*, 2021)).

Table 5.1 Environmental variables to be used as explanatory variables identified from the literature for importance to marine predator species.

Driver	Description	Taxonomic association	Reference
Bathymetry	Depth of the ocean floor to the surface.	Has been strongly associated with patterns of marine mammal species richness, particularly species such as semi-aquatic pinnipeds or cetaceans.	Bailey and Thompson (2009) Robinson *et al*. (2009) Retana and Lewis (2017) Thorne *et al*. (2017) Astudillo-Scalia *et al*. (2020)
		Deemed important for seabird density hotspot – especially for diving birds where shallow bathymetry may push prey into near-surface foraging range.	Hyrenbach *et al*. (2007) Amelineau *et al*. (2016) Urmy and Warren (2018) Evans *et al*. (2021)
Chl *a*	Used as an indicator / proxy for the concentration of phytoplankton present in the ocean, an indicator of productivity.	Seabird diversity and abundance have been shown to correlate with chl *a* concentration with transitions to foraging behaviour over areas of high Chl *a*.	Louzao *et al*. (2006) O'Hara *et al*. (2006) Grémillet *et al*. (2008) Boertmann (2011) Bennison and Jessopp (2015) Grecian *et al*. (2016) Embling *et al*. (2012) Kane *et al*. (2020). Suryan *et al*. (2012) Sabarros *et al*. (2014) Scales *et al*. (2014)
		Evidence of chlorophyll driving patterns and behaviours in cetaceans (e.g. humpback whales, fin whales and orca).	Cox (2015) Trudelle *et al*. (2016) Zerbini *et al*. (2016) Chen *et al*. (2020) Salgado Kent *et al*. (2021)
		Evidence of driving patterns in basking shark.	Miller *et al*. (2015)
		Cross-taxa studies of top predators.	Scott *et al. (*2010)

Driver	Description	Taxonomic association	Reference
Ocean fronts	Physical interfaces between contrasting water masses and create hydrodynamic processes that increase the availability of pelagic prey.	Association with major frontal areas in top predators.	Bost *et al*. (2009) Miller and Christodoulou (2014) Scales *et al*. (2015) Cox *et al*. (2018) Waggitt *et al*. (2019)
		Seabirds exhibit scale-dependent associations with eddies and fronts.	Wakefield *et al*. (2021)
		High associations of fronts for cetaceans (particularly deep-diving species).	Virgili *et al*. (2019) Bestley *et al*. (2019),
		Intense frontal activity shown to attract turtles.	Scales *et al*. (2015)
		Strong predictor of basking shark presence – aggregate in tidal fronts.	Miller *et al*. (2015)
SST	Temperature of the ocean's surface measured at a certain depth (typically at ~1m) below the water's surface.	SST shown to be a significant predictor of marine mammal diversity.	Tittensor *et al*. (2010) Kaschner *et al*. (2011) Astudillo-Scalia *et al*. 2020)
		Important in determining seabird distribution (due to prey location).	Hyrenbach *et al*. (2007) Evans *et al*. (2021) Krüger *et al*. (2018)
Salinity	Concentration of dissolved salts in seawater.	Important variable for cetaceans and sirenians.	Astudillo-Scalia *et al.* (2020)
		Significant importance in seal sightings.	Bailey and Thompson (2009)
MLT	The depth that separates the homogenized mixed water above from the denser stratified water below.	Correlated with seabird distribution.	Evans *et al*. (2021)
		Encounters between prey and seabirds was highest in stratified water.	Waggitt *et al*. (2018b)
Slope	The steepness or incline of the underwater terrain.	Cetaceans targeting specific gradients – e.g. pilot whales targeting steep bathymetric gradients, Risso's dolphins targeting intermediate slopes, and bottlenose dolphin targeting gentler slopes.	Thorne *et al. (*2017) Cañadas *et al*. (2002) Cafaro *et al*. (2015)
		Seabird abundance correlated with slope.	Amorim *et al. (*2009) Wakefield *et al*. (2017)

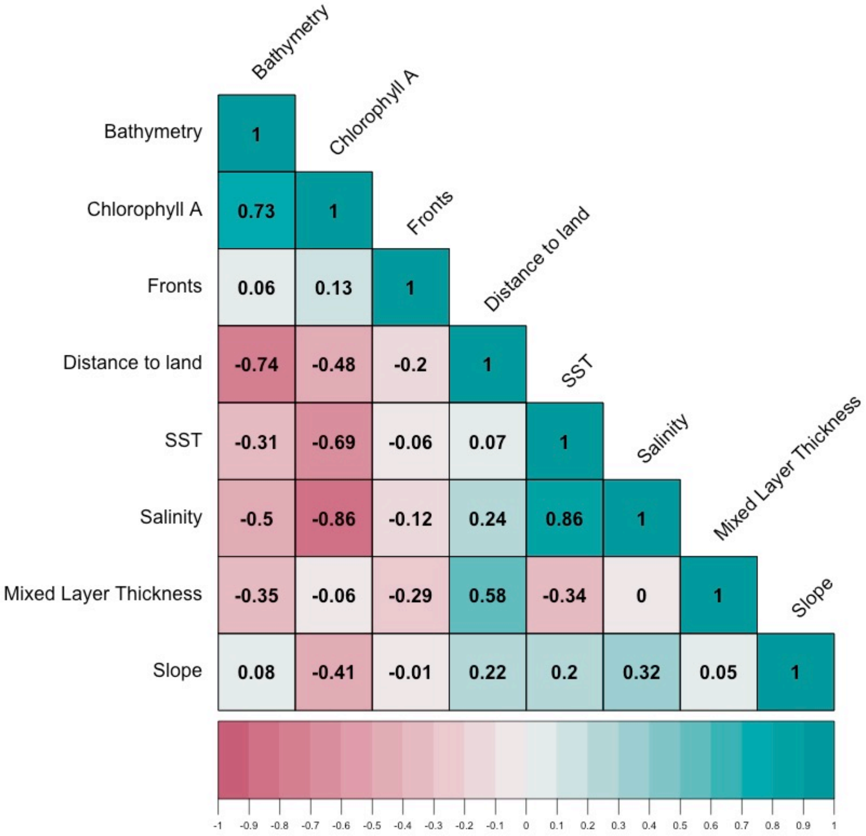
Driver	Description	Taxonomic association	Reference
Distance to land	Distance from a location to the nearest land mass or coastline.	Important for cetaceans and pinnipeds.	Bailey and Thompson (2009) Retana and Lewis (2017) Cronin *et al*. (2014)
		Foraging hotspots for seabirds have also been shown to be related to distance to land.	Nur *et al*. (2011)

Table 5.2 Explanatory environmental variables used in SSDM and GAMS, with source information.

Variable	Type	Unit	Temporal coverage	Spatial Resolution	Source
Bathymetry (Depth)	Static	m	Static	15 arc-second grid	GEBCO (GEBCO Compilation Group, 2021)
Chlorophyll	Dynamic	Mg m^{-3}	01/09/1997 to 01/12/2018, monthly means	4 km	Copernicus (CMEMS, 2021)
Fronts	Dynamic	Fcomp	01/09/1981 to 31/01/2016, monthly means	1.1 km	NEODAAS (NEODAAS, 2022)
Distance to land	Static	m	Static	m	Distance to Land ARCGIS Tool
Sea surface temperature	Dynamic	K	01/09/1981 31/01/2016, monthly means	1.1 km	(NEODAAS, 2022)
Salinity	Dynamic	Psu	16/01/1993 to 16/12/2018, monthly means	0.083° × 0.083°	Copernicus (Lea, 2015)
Density ocean mixed layer thickness	Dynamic	m	16/01/1993 to 16/12/2018, monthly means	0.083° × 0.083°	Copernicus (Lea, 2015)
Slope	Static	°	Static	15 arc-second grid	Calculated from GEBCO using Slope tool

5.3.2 Co-linearity between layers

Co-linearity between variables was calculated using the variance inflation factor (VIF) using the UDSM package (Naimi *et al.*, 2014) in R (version 4.2.2) (R Core Team, 2022) and correlation was assessed using the non-parametric Spearman's Rank Correlation Coefficient for complete pairwise observations (Figure 5.1). Where VIF was greater than 10, variables were removed, as this signals a collinearity problem. Primary productivity was removed due to its high collinearity with Chlorophyll *a* (VIF = 12.66, correlation coefficient = 0.9), as has been observed in other literature (Zhang *et al.*, 2004). When removed, the remaining variables had VIF values between 1.06 and 2.42, and all remaining variables were retained. Intrinsic collinearity often exists between predictor variables in reality, particularly in the environment where oceanographic processes are often closely linked.



	Bathymetry	Chlorophyll A	Fronts	Distance to land	SST	Salinity	Mixed Layer Thickness	Slope
Bathymetry	1							
Chlorophyll A	0.73	1						
Fronts	0.06	0.13	1					
Distance to land	-0.74	-0.48	-0.2	1				
SST	-0.31	-0.69	-0.06	0.07	1			
Salinity	-0.5	-0.86	-0.12	0.24	0.86	1		
Mixed Layer Thickness	-0.35	-0.06	-0.29	0.58	-0.34	0	1	
Slope	0.08	-0.41	-0.01	0.22	0.2	0.32	0.05	1

Figure 5.1 Spearman rank correlation of eight environmental variables – bathymetry, chlorophyll *a*, fronts, distance to land, sea surface temperature, salinity, mixed layer thickness, and slope.

5.3.3 Model selection: MEMs and SSDM

For MEMs, two approaches to modelling potential environmental drivers of species richness were carried out: the first modelling effort as a predictor within a GAM model, whilst the second GAM uses the SRPUE measure which has corrected for effort before modelling (see Chapter 4). This is to compare the value of factoring in effort prior to modelling in understanding associations with environmental variables. Furthermore, SRPUE factors in bias in advance of modelling but is more difficult to interpret given output values are on a relative scale of richness rather than integers of the number of species. This difficulty in natural interpretation led to this study exploring a second approach: incorporating survey effort in an alternative GAM where survey effort is factored into the model as a predictor.

SSDMs are common in ecology studies (Lee-Yaw *et al.*, 2021) and use species occurrences and environmental predictor variables to infer the environmental niche of species and identify drivers of richness (Guisan and Zimmermann, 2000, Guisan and Thuiller, 2005, Elith and Leathwick, 2009), by aggregating individual Species Distribution Models (SDMs) or Ensemble SDMs (ESDMs) into a SSDM which models species assemblages and species diversity and richness, outputting a single species richness layer (Schmitt *et al.*, 2017).

A summary of the MEM modelling approach, requirements, and the variables (response and predictors) for each model is presented in Table 5.3, with further specifics on the methodology for each model set up detailed below.

Table 5.3 MEM and SSDM modelling approach, requirements, and model variables (response and predictors) used in this study.

Model Approach	Strategy	Requirements	Advantages/Disadvantages	Model	Response variable	Predictor variables
MEM	“assemble first, predict later”	Total species richness values - no species-specific information required.	Less computationally intensive but no species level interpretation retained.	**GAM of species richness:** survey effort as a predictor variable	Species richness	Survey Effort (Number of observations) Smoothed predictors: Bathymetry, Chlorophyll, Fronts, Distance to land, SST, Salinity, mixed layer thickness, slope
				GAM of Species Richness per Unit Effort (SRPUE): Correcting for effort in advance	SRPUE (species richness per unit effort)	Smoothed predictors: Bathymetry, Chlorophyll, Fronts, Distance to land, SST, Salinity, mixed layer thickness, slope
SSDM	“predict first, assemble later”	Environmental variables to predict species distributions.	Computationally intensive but allows for more user interpretation at a species level.	SSDM	Occurrences	Bathymetry, Chlorophyll, Fronts, Distance to land, SST, Salinity, mixed layer thickness, slope

MEM: GAM of species richness: survey effort as a predictor variable

Species richness gridded rasters from Chapter 4 were utilised for GAMs of species richness. Rasters were imported and reprojected into WGS 84 (latitude and longitude) to match the coordinate reference system (CRS) of the environmental variable rasters, and to align with the SSDM method (which can only use latitude and longitude coordinate systems). When changing the CRS, as this involves resampling the grids, the method was set to 'ngb' or nearest neighbour, to keep the species richness as integers, rather than bilinear interpolation which treats them as continuous variables. A point shapefile with a species richness value for the centre coordinates of each gridded cell was created.

Environmental variables were resampled to match the smallest resolution using the bilinear method and then aligned to the same extent using extend and crop functions in the Raster package (Hijmans, 2020). Raster values were extracted by points, giving environmental values at each central coordinate point from the species richness grids, and attached to make a final data frame with each cell's central coordinates, species richness values, number of observations, bathymetry, Chl *a,* fronts, distance to land, SST, salinity, mixed layer thickness and slope values.

GAMs were then run to look at associations between environmental variables and species richness patterns. Model selection was performed (using the select = TRUE function), which allows selection of which explanatory values should be retained without requiring a stepwise selection procedure (stepwise model selection is inappropriate, as models are not nested owing to differences in the degree of nonlinearity (Isojunno *et al.* 2012), with shrinkage smoothers recommended instead Wood (2006)). All variables were retained as a result of model selection. A negative binomial family with a log link function was used, as species richness are integers with a lower boundary of zero. Species richness (with an offset) was the response variable, with survey effort (log observations) included as a predictor to account for variation in effort, and then the eight environmental variables as smooth predictor terms (the linear predictor depends on the smooth functions of predictors). The number of knots in GAMs smoothers was limited to k = 4, to provide ecologically interpretable relationships and avoiding the likelihood of overfitting. Given the GAM used log(observations), cells with observations of 2 or less were removed. Smooths were checked using gam.check function to check smooths did not have significant patterns in their residuals and have enough basis functions for the model.

A predicted species richness map was then generated using predict() function. Output richness values were also normalised onto a 0 to 1 scale, to cross-compare against outputs from SSDM, the GAM of species richness and the GAM of SRPUE.

MEM: GAM of Species Richness per Unit Effort (SRPUE): Correcting for effort in advance

For comparison when correcting for effort in advance, Species Richness per Unit Effort (SRPUE) gridded rasters from Chapter 4 were used for GAMs, following the same methodology as for raw species richness. However, SRPUE values are continuous and non-negative and are not integers like with raw species richness. Therefore, the Gamma distribution was chosen, with a log link function, with SRPUE as the response variable, and the eight environmental variables, as smooth predictor terms (the linear predictor depend s on the smooth functions of predictors). Model selection was performed (using the select = TRUE function), and all variables were retained. The number of knots in GAMs smoothers was limited to k = 4, to provide ecologically interpretable relationships and avoiding the likelihood of overfitting. Smooths were checked using gam.check function for robust models.

A predicted species richness map was then generated using predict() function. Output SRPUE values were also normalised onto a 0 to 1 scale, to cross-compare against outputs from SSDM, the GAM of species richness and the GAM of SRPUE.

SSDM model

For the SSDM (Schmitt *et al.*, 2017), occurrence data and environmental layers were imported into R and each of the eight environmental variables was tested individually that it loaded correctly, and then combined into a raster stack. Rasters of environmental data had to have the same coordinate reference system (WGS 84), but the spatial extent and resolution of the environmental layers can differ in the SSDM package. Variables were aggregated to the coarsest resolution, by clipping to the smallest common extent and then upscaling to the coarsest resolution. Occurrences were loaded without performing any optional geographical thinning (thinning removes unnecessary records to reduce the effect of sampling bias whilst retaining the greatest amount of information), as the size and scale of the dataset proved too computationally intensive to perform thinning. Ten species were removed during modelling due to having three or fewer occurrences (blue whale *Balaenoptera musculus*, false killer whale *Pseudorca crassidens*, green turtle *Chelonia mydas*, hooded seal *Cystophora cristata*, little shearwater *Puffinus assimilis*, mako shark *Isurus oxyrinchus*, Ross's gull *Rhodostethia rosea*, Sowerby's whale *Mesoplodon bidens*, spotted dolphin *Stenella frontalis* and yellow-legged herring gull *Larus michahellis*), which left 59 species in the model.

Several methods of stacking are available (pSSDM, Bernoulli, bSSDM, MaximumLikelihood, PRR.MEM, PRR.pSSDM, see Calabrese *et al.* (2014)) but the pSSDM method was used, which sums the continuous habitat suitability maps to give maps of local species richness and composition. Stack modelling of all species using GAM modelling algorithms was carried out, with one iteration set due to limits in computing power for carrying out SSDMs for 59 species over such a large AOI. Outputs include an overall raster file of diversity (species richness), as well as rasters of endemism (species being unique to a location) and uncertainty, and CSV files that contain a range of model evaluation metrics and overall variable importance (which summarises the relative contribution of environmental variables to determine species distribution). The SSDM package provides measures of the relative contribution of environmental variables at both an overall SSDM level and on a species-by-species basis. It averages the outputs from two approaches; a jacknife approach (Phillips *et al.*, 2006) and Pearson's correlation coefficient (Thuiller *et al.*, 2009) between a full model and a model with each environmental variable omitted in turn.

Therefore, rasters of binary presence/absence, probability of occurrence and uncertainty, model evaluation metrics and variable importance are also produced for each of the 59 species, thus allowing the importance of each environmental variable for a specific species to be examined. Maps of species richness were plotted in R, and species richness values were also normalised to give a scale of 0 to 1 to compare equivalently against maps using GAMs of species richness and GAMs of SRPUE (section 5.3.3).

The species-specific outputs of environmental variable importance were grouped by different factors presented in Table 5.4, and stacked bar charts for each category were produced to see if there are differences or similarities in variable importance within groupings. Trophic level position was obtained from available literature (birds (Hobson, 1993, Karpouzi, 2005), cetaceans and pinnipeds (Trites, 2019), turtles (Lemons *et al.*, 2020), shark and fish trophic level obtained from Fishbase (Froese and Pauly, 2000).

Table 5.4 Groupings for analysis of relative variable importance and the categories used per grouping.

Grouping	**Categories**
Taxa	Birds, cetaceans, fish, pinnipeds, sharks, turtles
Trophic level position	3.0-3.4, 3.41-3.8, 3.81-4.2, 4.21-4.6
Foraging depth	Subsurface, surface, surface and land, surface and subsurface
Foraging behaviour	Diving, filter, lunge, bubble, mixed, plunge, pursuit, shallow, kleptoparasitic, carrion, suction feeding, dabbler.
Feeding location	Inshore, land, offshore
Diet	Carnivore, Omnivore, Planktivore
Diet - prey species	Cephalopods, crustaceans, fish, molluscs, vegetation, marine mammals, seabirds, offal/rubbish, plankton, insects.
Nesting/Haul out location	Burrow, Cliff, Crevice, Ground
Central Place Forager	Yes / No
IUCN status (IUCN, 2023)	Critically endangered, endangered, near threatened, least concern, data deficient.

5.4 Results

5.4.1 Maps of predicted species richness from SSDM and GAMs

Species richness maps and normalised richness maps from the three methods (GAM of raw species richness with effort as a predictor, GAM of SRPUE and SSDM) are given in Figure 5.2 and histograms of normalised rasters given in Figure 5.3**.** Summary statistics are given in Table 5.5.

In the non-normalised maps (Figure 5.2, Panels A to C) predicted species richness has lower maximum species richness values from the GAM with effort as predictor, than the SSDM (Table 5.5). Areas of high species richness are fewer and are more localised (Figure 5.2, panel A). Mean species richness was 6.2 for the species richness GAM, 1.7 for SRPUE GAM (though noting this is not a comparable measure) and 13.8 for the SSDM model, demonstrating the SSDM predicts higher species richness values than the other two methods.

Table 5.5 Minimum, mean, median, and maximum of GAMs (species richness and SRPUE) and SSDM maps.

Measure	Species richness (SR)	SRPUE	SSDM	SR Normalised	SRPUE Normalised	SDM normalised
Minimum	0.6	1.0	0.2	0.000	0.000	0.000
Mean	6.2	1.7	13.8	0.254	0.296	0.439
Median	5.7	1.6	14.4	0.232	0.270	0.458
Maximum	22.6	3.3	31.1	1.000	1.000	1.000

When the three methods are normalised, the median richness value was similar between species richness (0.232) and SRPUE (0.270) (Figure 5.3) though the spread of data was wider for the GAM with effort as a predictor (mean = 0.254) versus the GAM using SRPUE (mean = 0.296). The median and mean value for the SSDM was 0.458 and 0.439 respectively, therefore much higher than GAMs, with a much more even spread of data up to the maximum value (1). The methodology for the GAM with effort as a predictor required removing values with 2 or less occurrences, and therefore there are less values overall, evidenced in the histogram comparison with lower peak and wider tail (Figure 5.3). The spread of species richness from the GAM model (where effort is a predictor) is skewed towards lower species richness values, SRPUE has a higher peak, whilst the richness values from SSDM are more evenly spread (Figure 5.3).

Figure 5.3 demonstrates that for predicted species richness from GAMs where effort it factored in, richness shows a non-normal distribution, skewed towards lower values, with a higher frequency of smaller species richness values. For SRPUE, a more symmetric normal distribution is observed, with a peak in species richness closer to the median and counts less frequent towards the lower and higher species richness values (Figure 5.3).

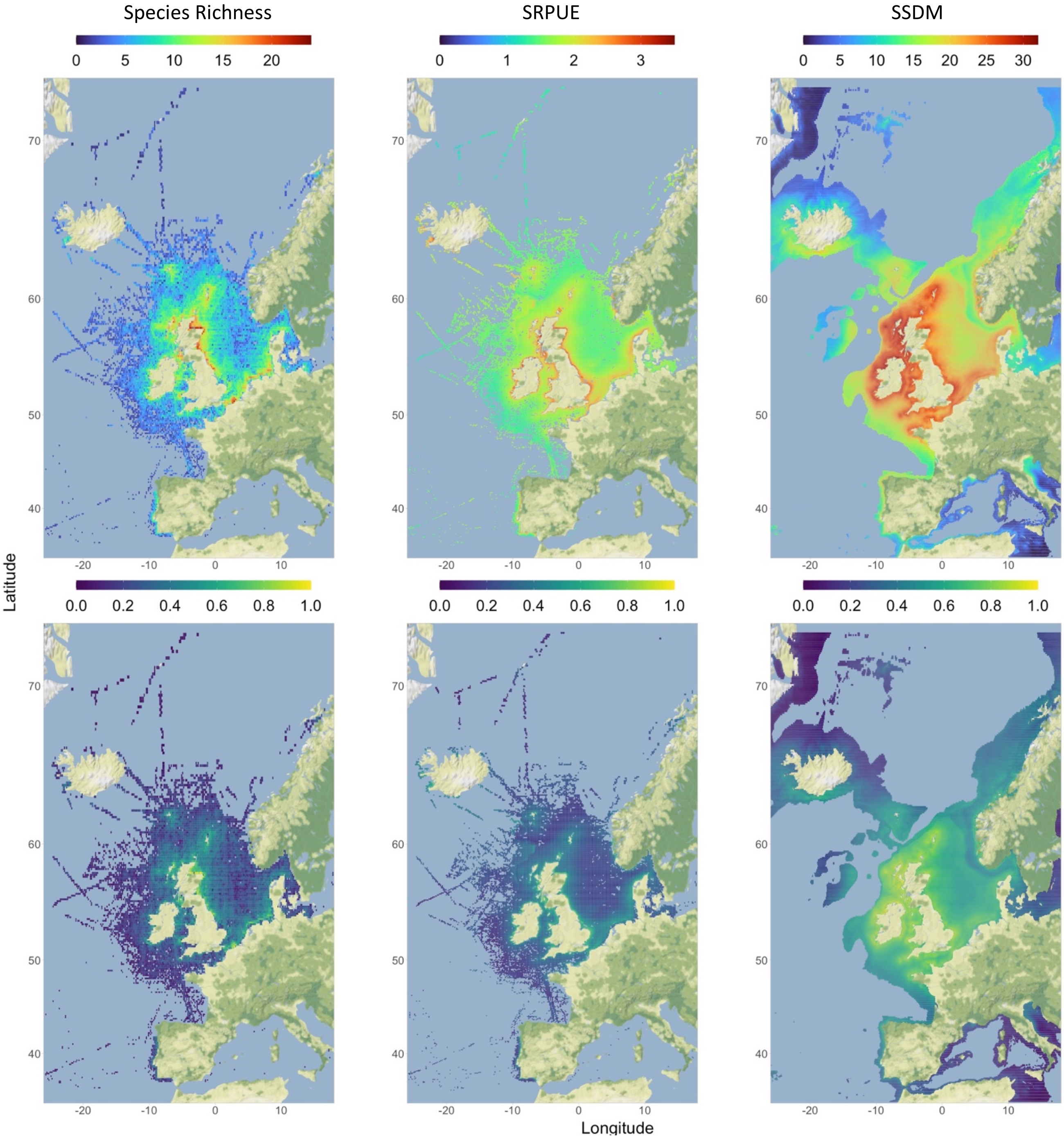


Figure 5.2 Top row: Predicted density surfaces from model outputs for richness. A) Species richness B) Species Richness Per Unit Effort (SRPUE) and C) SSDM. Bottom row: Predicted density surfaced, normalised for comparison between areas. D) Species richness normalised E) SRPUE normalised and F) SSDM normalised.

All three methods in the original scaled maps and when normalised (Figure 5.2, panels D-F) give patterns of higher species richness around the coastline, particularly the East Coast of the UK, Scotland and along the coastline of mainland Europe (Belgium, the Netherlands, Germany, and Denmark). However, the SSDM map also gives high areas of richness around the entire Irish coast, North and West Scotland and along the coast of Norway.

When maps are normalised (Figure 5.2, panels D-F), all three methods show comparable patterns of species richness, albeit the normalised SSDM map gives much higher relative values of species richness overall than in species richness and SRPUE maps (Figure 5.3).

Species richness remained high (compared to raw maps) across the North Sea in the SSDM maps (Figure 5.2, panel F), whereas in the GAM predict surfaces for species richness (Figure 5.2, Panel D) and SRPUE (Figure 5.2, Panel E) there is a noticeable drop in species richness further offshore into the North Sea and shows areas of low species richness in deeper waters (offshore from Norway). The two GAMS show highly similar patterns of species richness (Figure 5.2, panels D and E), demonstrating accounting for survey effort within models gives similar mapping outputs to using the SRPUE measure. However, the normalised GAM with effort as a predictor still shows some remnants of survey effort, with shipping routes evident in Panel D (Figure 5.2) particularly visible between Dublin and Liverpool, Wexford to Fishguard across St George’s Channel and routes out from Aberdeen towards Lerwick. Survey effort is a shown to be a highly significant predictor in the GAM model (see section 5.4.2), but simple maps of richness (Panel A, D) are not adjusted for this.

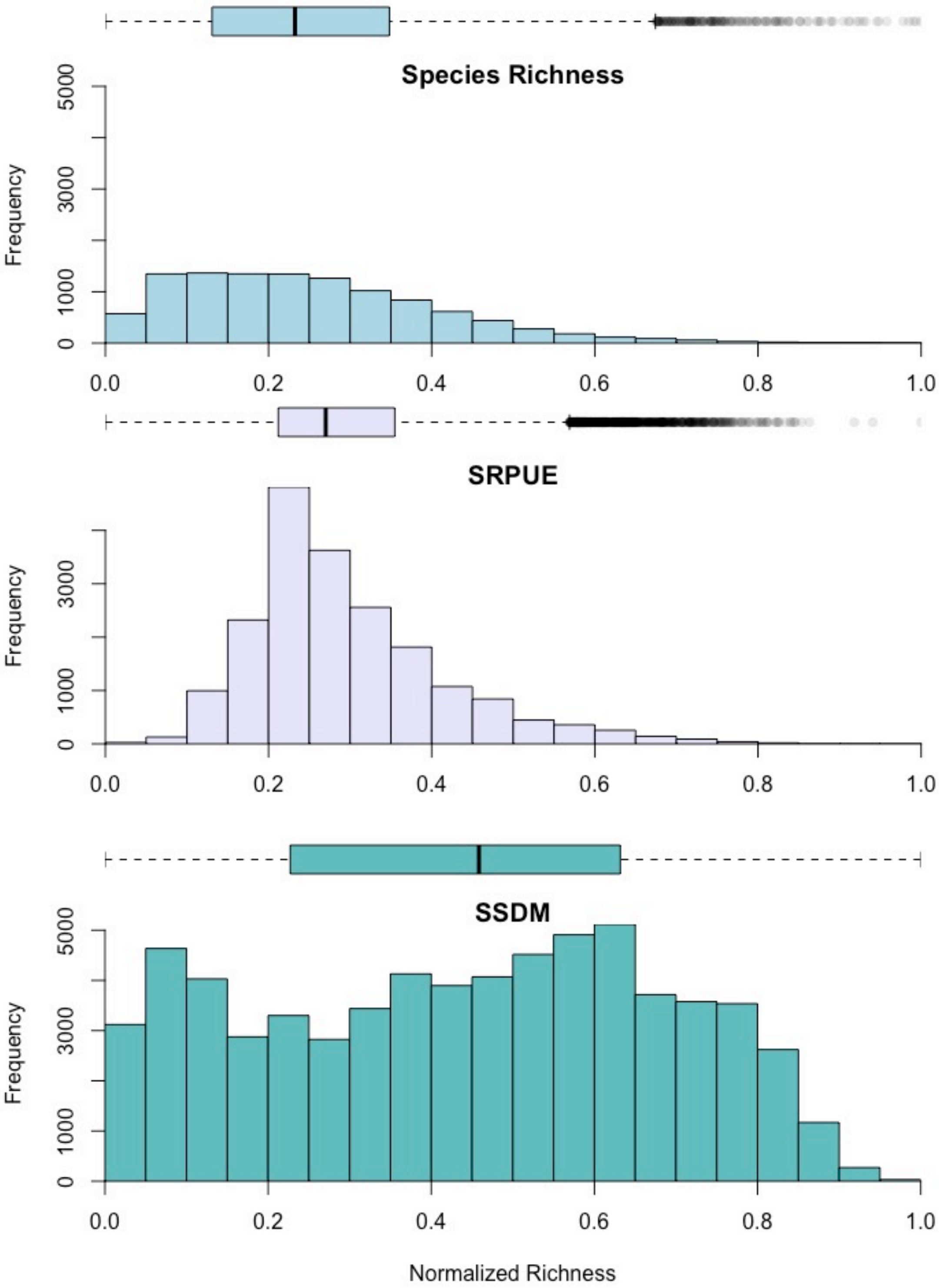


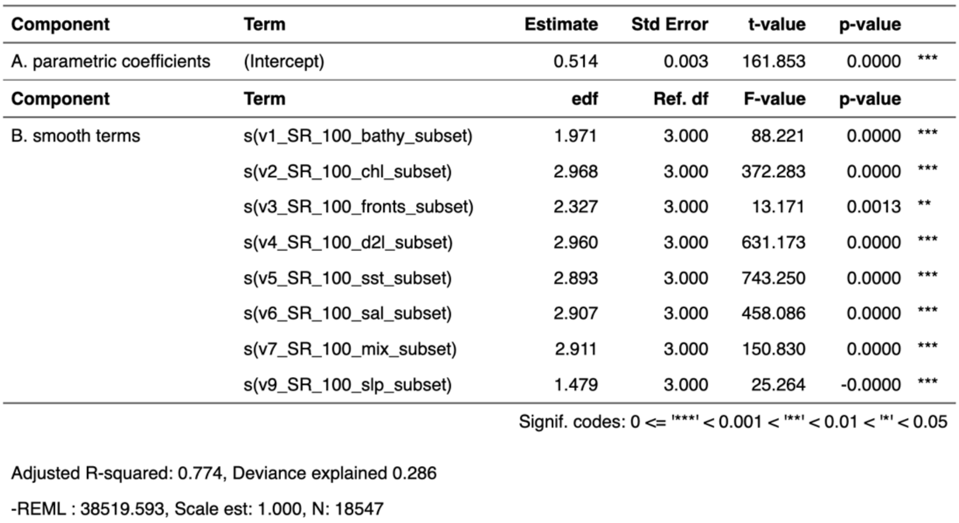
Figure 5.3 Histogram and boxplots of species richness values from rasters of predicted species richness, normalised for comparative scale from predict function of GAM of species richness (with effort as a predictor variable), GAM of SRPUE and SSDM models.

5.4.2 Comparison of predicted relationships with environmental variables in two MEMs

Both the GAM which include effort (as log(observations)) as a predictor (Table 5.6) and the GAM using SRPUE (pre-corrected for effort) (Table 5.7) showed all variables have a significant relationship with species richness.

For the GAM using survey effort as a predictor, all environmental variables except fronts showed highly significant relationships with species richness ($p < 0.0001$) (Table 5.6), with fronts still significant at $p = 0.01$. Log observations is also a highly significant driver of species richness, supporting conclusions in Chapter 4. The GAM using SRPUE showed all variables are highly significant ($p < 0.0001$) (Table 5.7).

Table 5.6 Statistical output from GAM of raw species richness.



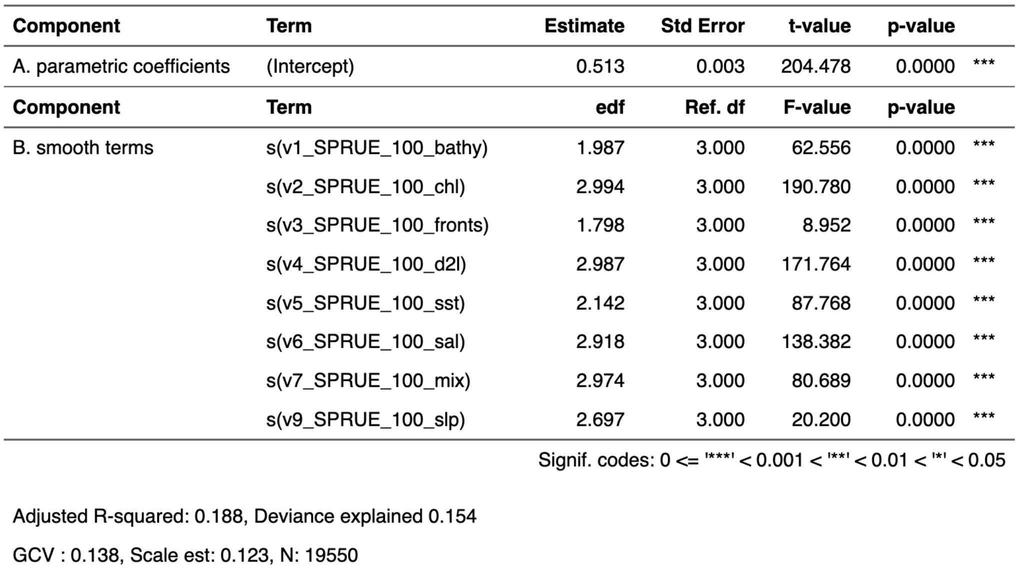
Component	Term	Estimate	Std Error	t-value	p-value	
A. parametric coefficients	(Intercept)	0.514	0.003	161.853	0.0000	***
Component	**Term**	**edf**	**Ref. df**	**F-value**	**p-value**	
B. smooth terms	s(v1_SR_100_bathy_subset)	1.971	3.000	88.221	0.0000	***
	s(v2_SR_100_chl_subset)	2.968	3.000	372.283	0.0000	***
	s(v3_SR_100_fronts_subset)	2.327	3.000	13.171	0.0013	**
	s(v4_SR_100_d2l_subset)	2.960	3.000	631.173	0.0000	***
	s(v5_SR_100_sst_subset)	2.893	3.000	743.250	0.0000	***
	s(v6_SR_100_sal_subset)	2.907	3.000	458.086	0.0000	***
	s(v7_SR_100_mix_subset)	2.911	3.000	150.830	0.0000	***
	s(v9_SR_100_slp_subset)	1.479	3.000	25.264	-0.0000	***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.774, Deviance explained 0.286

-REML : 38519.593, Scale est: 1.000, N: 18547

Table 5.7 Statistical output from GAM using SRPUE.



Component	Term	Estimate	Std Error	t-value	p-value	
A. parametric coefficients	(Intercept)	0.513	0.003	204.478	0.0000	***
Component	**Term**	**edf**	**Ref. df**	**F-value**	**p-value**	
B. smooth terms	s(v1_SPRUE_100_bathy)	1.987	3.000	62.556	0.0000	***
	s(v2_SPRUE_100_chl)	2.994	3.000	190.780	0.0000	***
	s(v3_SPRUE_100_fronts)	1.798	3.000	8.952	0.0000	***
	s(v4_SPRUE_100_d2l)	2.987	3.000	171.764	0.0000	***
	s(v5_SPRUE_100_sst)	2.142	3.000	87.768	0.0000	***
	s(v6_SPRUE_100_sal)	2.918	3.000	138.382	0.0000	***
	s(v7_SPRUE_100_mix)	2.974	3.000	80.689	0.0000	***
	s(v9_SPRUE_100_slp)	2.697	3.000	20.200	0.0000	***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Adjusted R-squared: 0.188, Deviance explained 0.154

GCV : 0.138, Scale est: 0.123, N: 19550

Overall, the two GAMs show similar outputs (likely due to limiting of knots) (Figure 5.4 and Figure 5.5), with the GAM with effort as a predictor in panels A and GAM for SRPUE in panel B). In both GAMs, there is a clear decline in richness gradually until about -1000 m below sea level (Figure 5.4, panel A1 and B1), with the decline more steep in the GAM of SRPUE. In both models, Chlorophyll *a* shows a peak in diversity between 2 and 4 mg m^{-3}, then a dip in diversity, and then a second peak between 10 and 12 mg m^{-3} (but confidence intervals are wider for this second peak). There is no strong direction relationship with fronts in the species richness GAM (Figure 5.4, panel A3) but a slight positive relationship in SRPUE (Figure 5.4, panel B3). Species richness appears to decline the further from land in both models, although there is a slight peak at around 70,000 metres in both (Figure 5.4 and Figure 5.5, Panel A4 and B4).

With SST (Figure 5.5, panels A5 and B5), the relationship between SST and richness appears to be relatively increase until ~283 kelvin (approximately 9.85 °C), and then steeply decrease at temperatures over ~285 kelvin (approximately 11.85 °C), suggesting an optimal temperature for marine predator species.

Species richness shows an initial stable relationship with salinity until 25 psu, and then a rapidly increasing relationship with salinity post 25 psu (Figure 5.5, panels A6 and B6), with the more saline the water, the more species observed. Mixed layer thickness shows a similar trend(Figure 5.5, Panel A7 and B7), with initial stable relationship until a clear increase in richness at 150 m, which is more evident in the GAM of SRPUE compared to the GAM of species richness (Figure 5.5, Panel B7). For slope, there is a very slight increase in species richness as slope increases (Figure 5.5, Panel A8 and B8) but the confidence intervals are much wider as slope increases so a distinct relationship cannot be identified.

None of the response curves change species richness by more than one or two species, suggesting that these drivers may not have large influence on species richness hotspots compared to other environmental or biological factors.

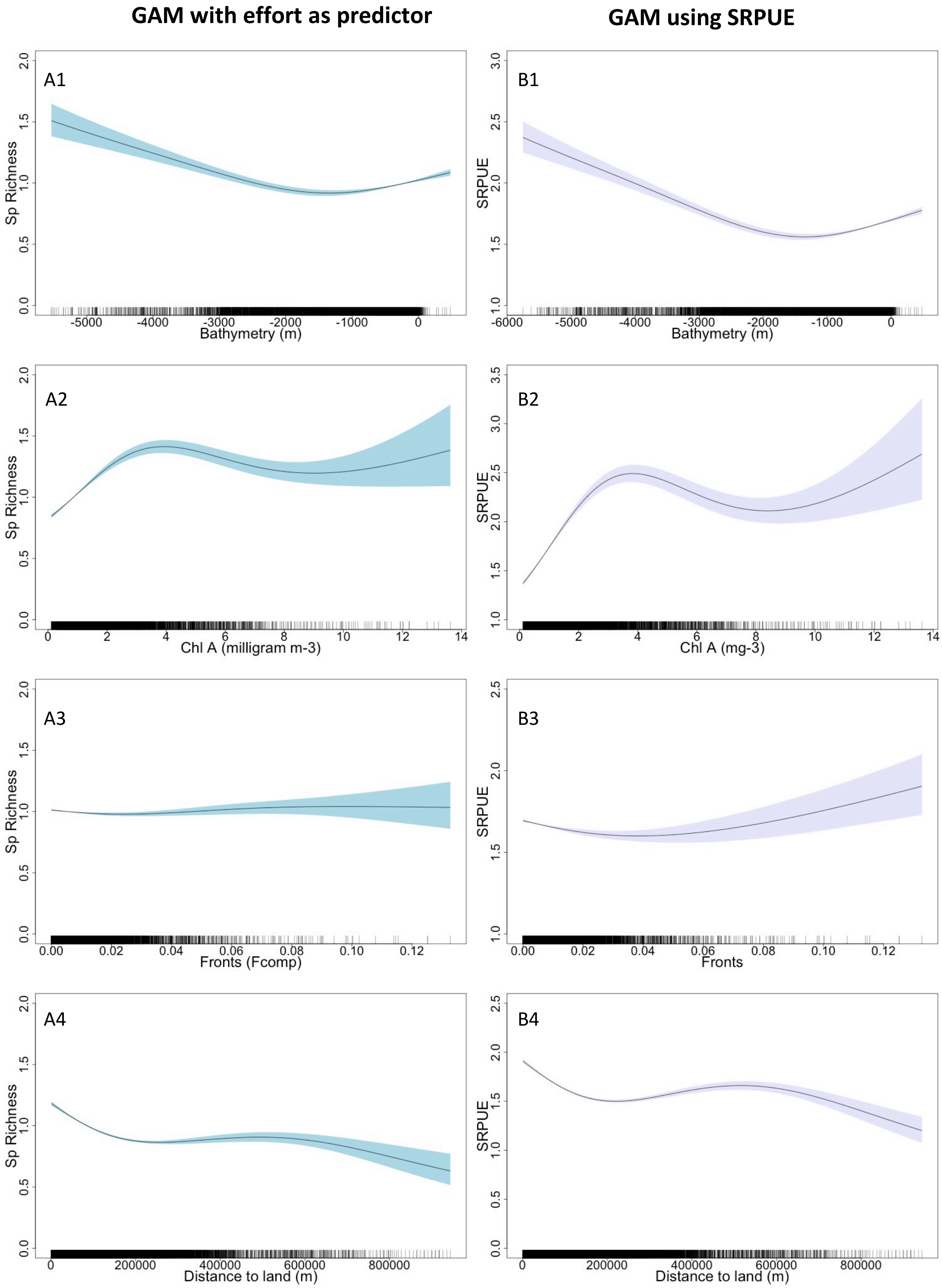


Figure 5.4 Output from MEMs for bathymetry (m), Chl *a* (mg^{-3}), fronts (F_{comp}) and distance to land (m) for both GAM using species richness with effort as a predictor (panels A1-A4, blue shading) and GAM using SRPUE (panels B1-B4, purple shading).

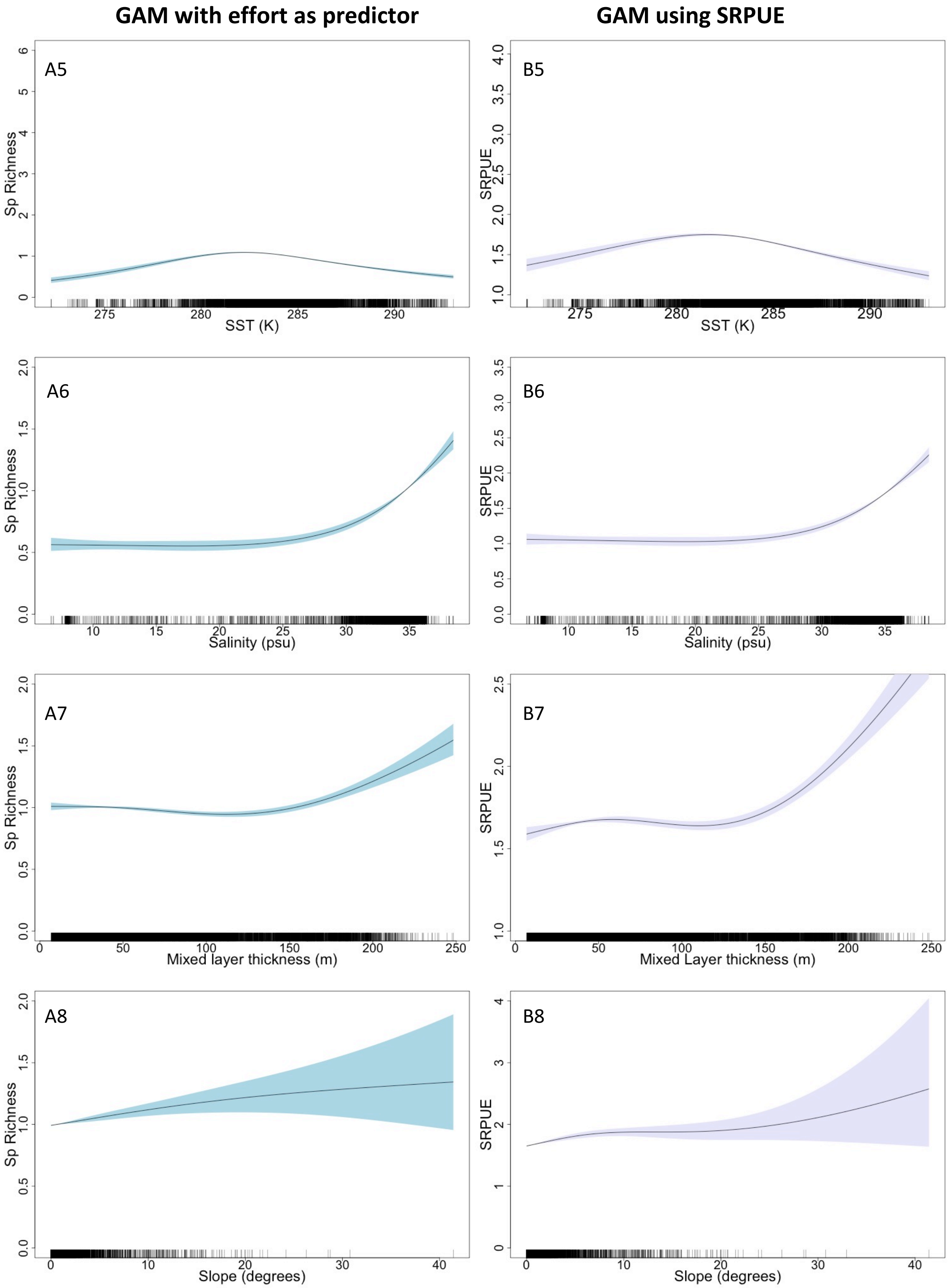


Figure 5.5 Output from MEMs for SST (K), salinity (psu), mixed layer thickness (m) and slope (degrees) for both GAM using species richness with effort as a predictor (panels A5-A8, blue shading) and GAM using SRPUE (panels B5-B8, purple shading).

5.4.3 Environmental drivers of species richness hotspots

A comparison of variable importance is presented in Table 5.8 and in Figure **5.6**. Importance was derived using chi-squared for GAMs and for the SSDM the model outputs overall variable importance (which is derived from Pearson correlation coefficient with normalisation to give a percentage).

SST was most important in the GAM of species richness (29.94%) and SSDM (46.36%), whilst Chl-*a* was most important in the GAM of SRPUE (25.07%). Fronts had least importance in all three models (0.53% in GAM species richness, 1.18% in GAM SRPUE and 2.5% in SSDM), slope (1.02% in GAM species richness, 2.65% in GAM SRPUE and 3.92% in SSDM). Distance to land was important in the GAMs (29.94% in the GAM species richness, 22.57% in GAM SRPUE) but far less important in the SSDM (12.87%). Salinity was of high importance in the SSDM (14.85%) and GAM of species richness (18.45%) and GAM of SPRUE (18.18%).

Table 5.8 Environmental variable importance (%) per model – GAM of species richness where effort included as predictor, GAM using SRPUE measure and SSDM.

	Variable importance per model (%)		
Variable	**GAM Species Richness**	**GAM SRPUE**	**SSDM**
Bathymetry	3.55	8.22	6.68
Chlorophyll-*a*	15.00	25.07	5.66
Fronts	0.53	1.18	2.50
Distance to land	25.43	22.57	12.87
Sea surface temperature (SST)	29.94	11.53	46.36
Salinity	18.45	18.18	14.85
Mixed layer thickness (MLT)	6.08	10.60	7.16
Slope	1.02	2.65	3.92

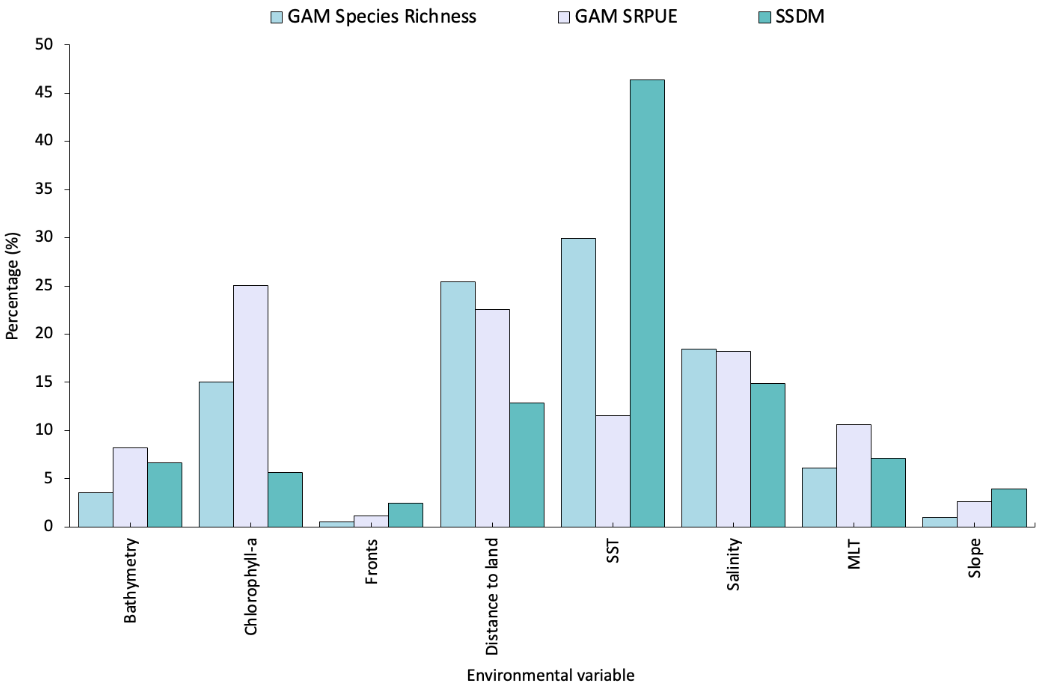


Figure 5.6 **Overall relative variable importance (%) from three models: GAM with species richness and effort as predictor, GAM using SRPUE measure and SSDM model.**

Taxa

The SSDM model allows for further breakdown of variable importance by groupings such as taxa, trophic level, foraging strategy etc, which the GAMs do not. Stacked bar charts in Figure 5.7 and Table 5.9 demonstrate that when variable importance contribution was split out by taxa from the SSDM, SST was the most important variable for birds (49.6%), cetaceans (41.83%), fish (54.16%), pinnipeds (39.80%) and sharks (35.12%), but salinity was most important for turtles (50.15%) (Figure 5.7). However, the number of species for turtles was one with 525 observation points so may be simply correlation.

Bathymetry had a moderate relative contribution for fish (13.89%), cetaceans (9.5%) and pinnipeds (8.47%). Chlorophyll *a* had a low contribution for most taxa, but contribution for birds was 11.19% and pinnipeds was 8.84%. Distance to land had a high relative proportion in sharks (34.51%), but was also important for birds (11.19%), cetaceans (14.95%), and pinnipeds (16.08%). Fronts had little contribution to all taxa (ranging from 0.3% in sharks to 3.8% for cetaceans). Mixed-layer thickness was important for sharks (26.4%). Salinity was very important for turtles (50.2%), and birds (17.6%), fish (15.2%), and pinnipeds (19.3) but was the proportion was small for sharks (1.2%) and cetaceans (7.3%). Slope had low relative contribution across all taxa (ranging 0.8% for sharks to 8.3% for cetaceans).

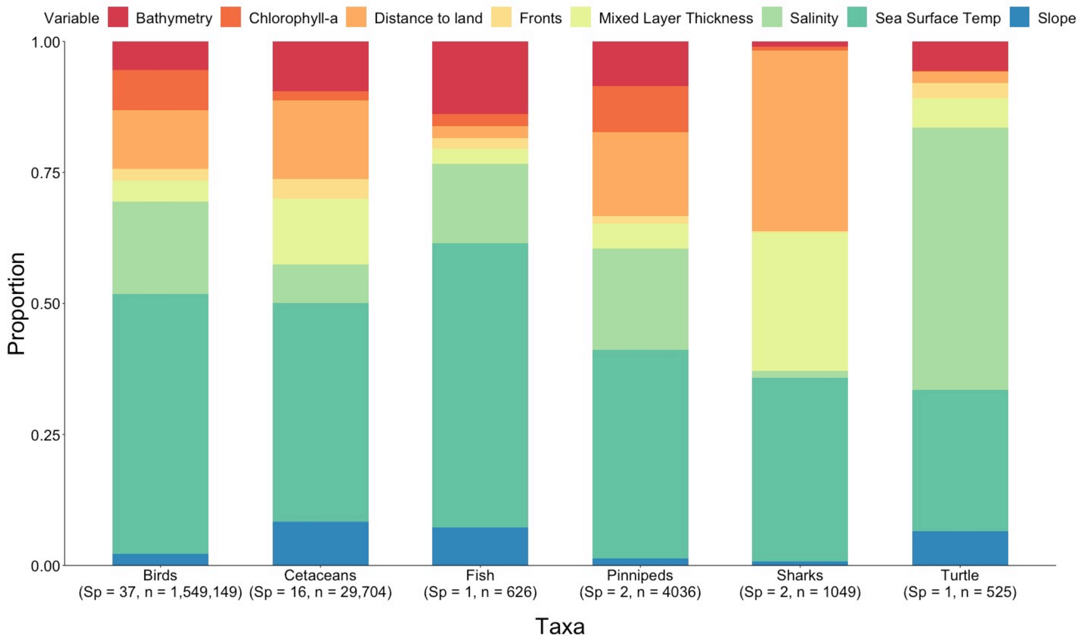


Figure 5.7 Proportions of variable importance, per taxa group.

Table 5.9 Relative contribution of environmental variables per taxa group. MLT = mixed layer thickness, SST = sea surface temperature.

Taxa	**Bathymetry**	**Chl *a***	**Distance to land**	**Fronts**	**MLT**	**Salinity**	**SST**	**Slope**
Bird	5.5%	7.7%	11.2%	2.1%	4.1%	17.6%	49.6%	2.2%
Cetacean	9.5%	1.8%	15.0%	3.8%	12.5%	7.3%	41.8%	8.3%
Fish	13.9%	2.3%	2.3%	1.9%	3.0%	15.2%	54.2%	7.3%
Pinniped	8.5%	8.8%	16.1%	1.4%	4.7%	19.3%	39.8%	1.4%
Shark	1.0%	0.7%	34.5%	0.3%	26.4%	1.2%	35.1%	0.8%
Turtle	5.6%	0.1%	2.3%	2.8%	5.6%	50.2%	27.0%	6.5%

Trophic level

When split by trophic level (Figure 5.8, Table 5.10), SST was the prevailing environmental driver for all trophic levels (ranging from 41.0% to 53.6%). There were no clear patterns in variable contribution by trophic level for other variables.

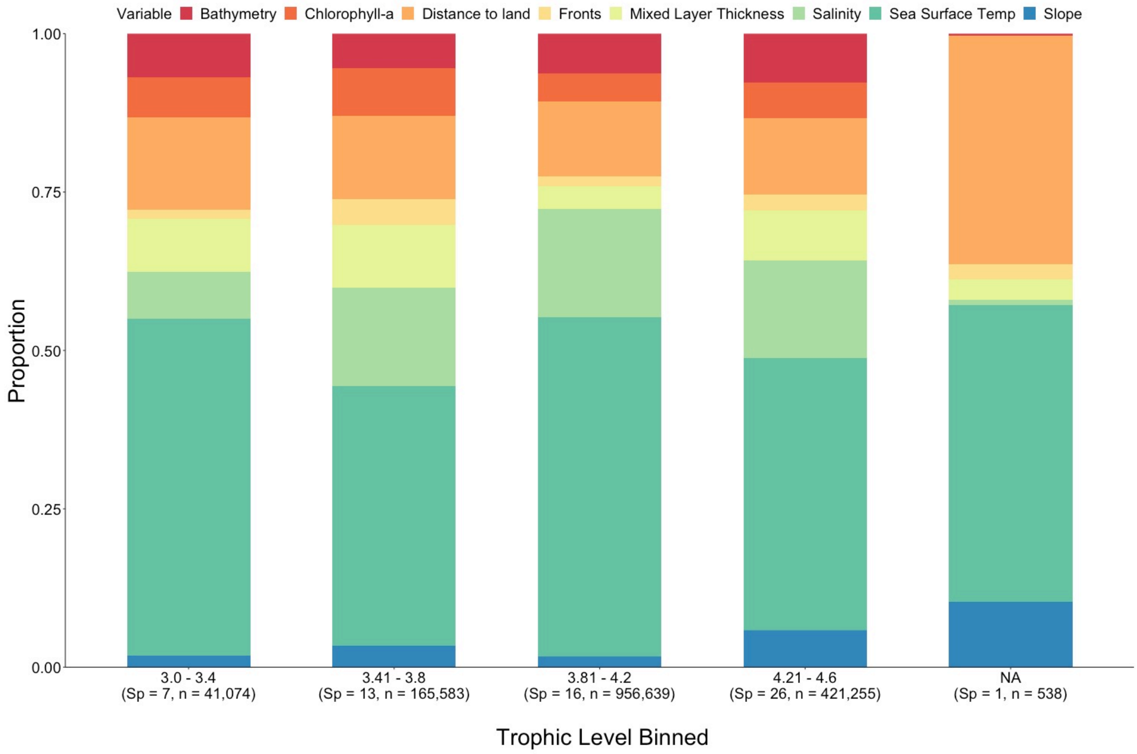


Figure 5.8 Proportions of relative variable importance by trophic level (binned).

Table 5.10 Relative contribution of environmental variables per trophic level (species specific trophic levels binned into categories). MLT = mixed layer thickness, SST = sea surface temperature.

Trophic Level (binned)	Bathymetry	Chl *a*	Distance to land	Fronts	MLT	Salinity	SST	Slope
3.0 - 3.4	6.8%	6.4%	14.5%	1.5%	8.3%	7.4%	53.1%	1.9%
3.41 - 3.8	5.5%	7.5%	13.1%	4.0%	10.0%	15.6%	41.0%	3.4%
3.81 - 4.2	6.3%	4.4%	11.8%	1.6%	3.6%	17.1%	53.6%	1.6%
4.21 - 4.6	7.7%	5.6%	12.0%	2.5%	7.9%	15.5%	43.0%	5.8%
N/A	0.3%	0.0%	36.0%	2.5%	3.2%	0.7%	46.9%	10.3%

Central place forager (CPF)

There were no clear differences between patterns of relative contribution of variables for those species that are CPFs and those that are not (Figure 5.9, Table 5.11). Salinity and SST had a higher relative contribution in CPFs (17.72% and 49.8% compared to 9.24% and 41%). Distance to land was expected to be higher in CPFs, however, was lower (15.64% in non-CPFs compared to 11.44% in CPFs).

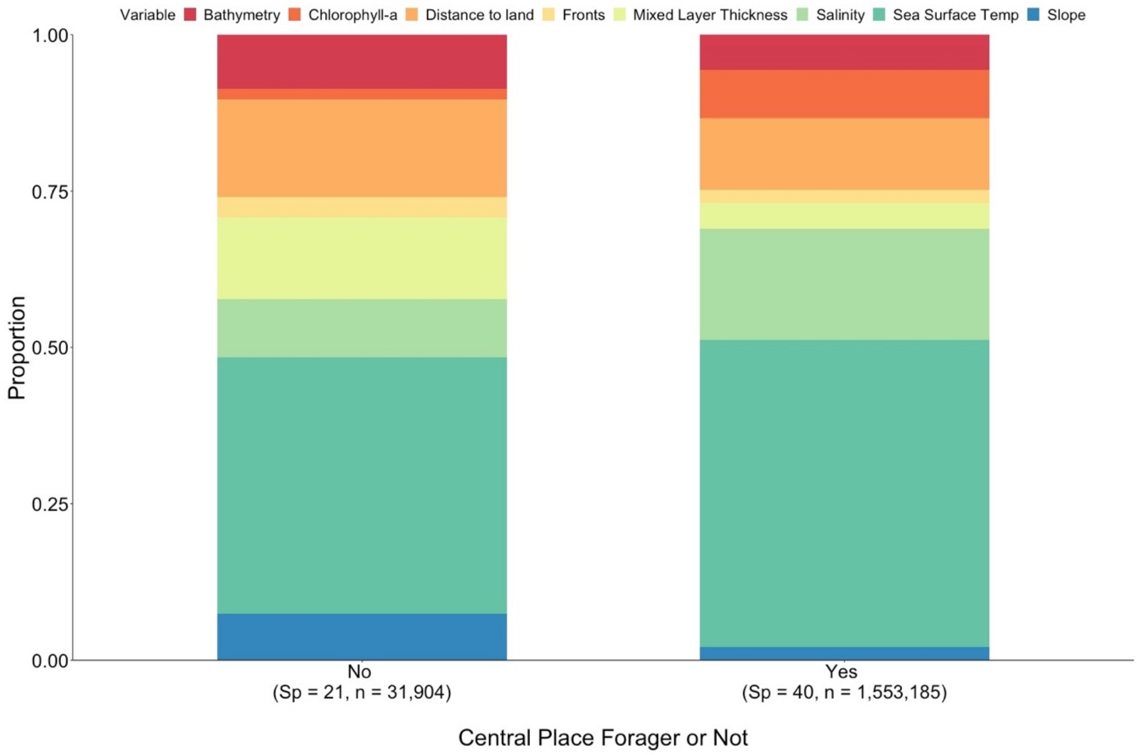


Figure 5.9 Proportions of relative variable importance split by whether central place forager or not.

Table 5.11 Relative contribution of environmental variables for species that are central place foragers (CPF). MLT = mixed layer thickness, SST = sea surface temperature.

CPF	Bathymetry	Chl *a*	Distance to land	Fronts	MLT	Salinity	SST	Slope
No	9%	2%	16%	3%	13%	9%	41%	7%
Yes	6%	8%	11%	2%	4%	18%	49%	2%

Foraging depth and location

When split by foraging depth (Figure 5.10, Table 5.12), SST was highest proportion for all species, and highest for surface/land foraging species. Bathymetry was of lower proportion in those species that forage at the surface only (5.29%) than those that forage subsurface only (8.30%) (Table 5.12). Chl *a* was also of lower proportion in subsurface species (3.21% subsurface only, 3.94% surface/subsurface) than surface and land-based species (7.86% for surface only, 8.36% for surface/land). Fronts were low importance for all foraging depths, particularly for surface/land-based species (0.11%). Mixed layer thickness was lowest for surface/land foraging species (3.29%) but similar across other foraging depths. Slope was of low importance for those surface foraging species (e.g. 2.54% surface only) compared to subsurface foragers (6.29% subsurface only).

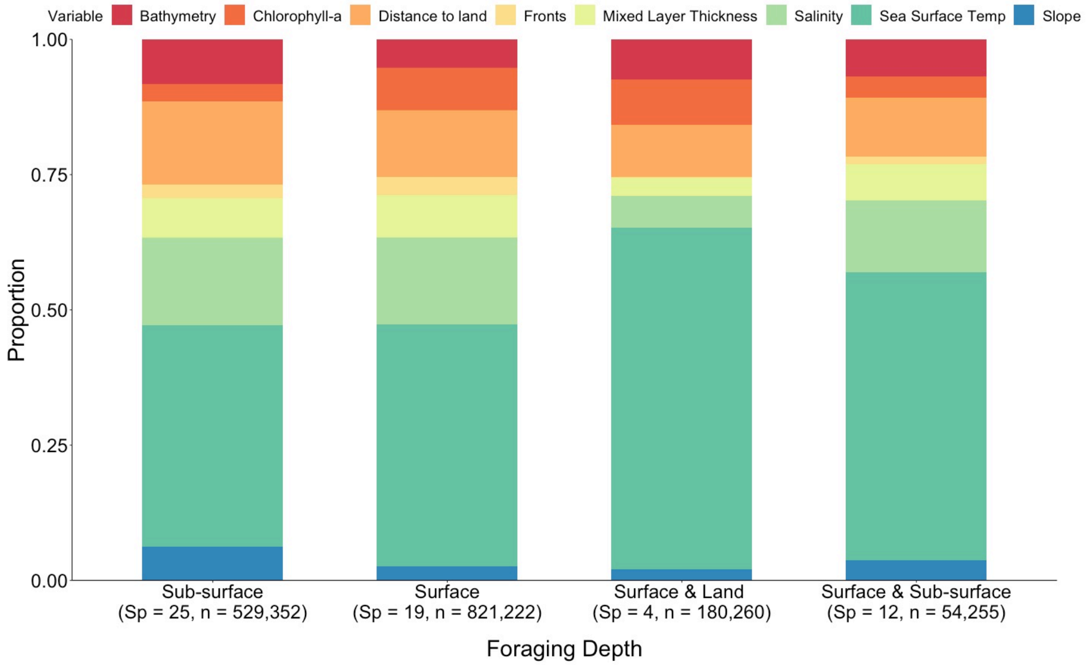


Figure 5.10 Proportions of relative variable importance by foraging depth.

Table 5.12 Relative contribution of environmental variables when species are split by their foraging depth. MLT = mixed layer thickness, SST = sea surface temperature.

Foraging depth	Bathymetry	Chl *a*	Distance to land	Fronts	MLT	Salinity	SST	Slope
Subsurface	8.30%	3.21%	15.35%	2.54%	7.30%	16.19%	40.82%	6.29%
Surface	5.29%	7.86%	12.32%	3.34%	7.85%	15.97%	44.84%	2.54%
Surface / Land	7.45%	8.36%	9.70%	0.11%	3.29%	5.94%	63.11%	2.04%
Surface / Subsurface	6.79%	3.94%	10.96%	1.39%	6.75%	13.22%	53.31%	3.65%

There were no distinctive differences in proportions of environmental variables when split by feeding location (inshore, offshore, inshore / land) (Figure 5.11). Bathymetry and chlorophyll *a* were more important in inshore/land feeding species (13.31% and 15.41%). Distance to land appeared to be lower in importance in inshore/land feeding species (6.35%) compared to inshore, inshore/offshore and offshore species. Offshore species had the highest proportions for fronts (3.08%), mixed layer thickness (8.86%), and slope (5.47%).

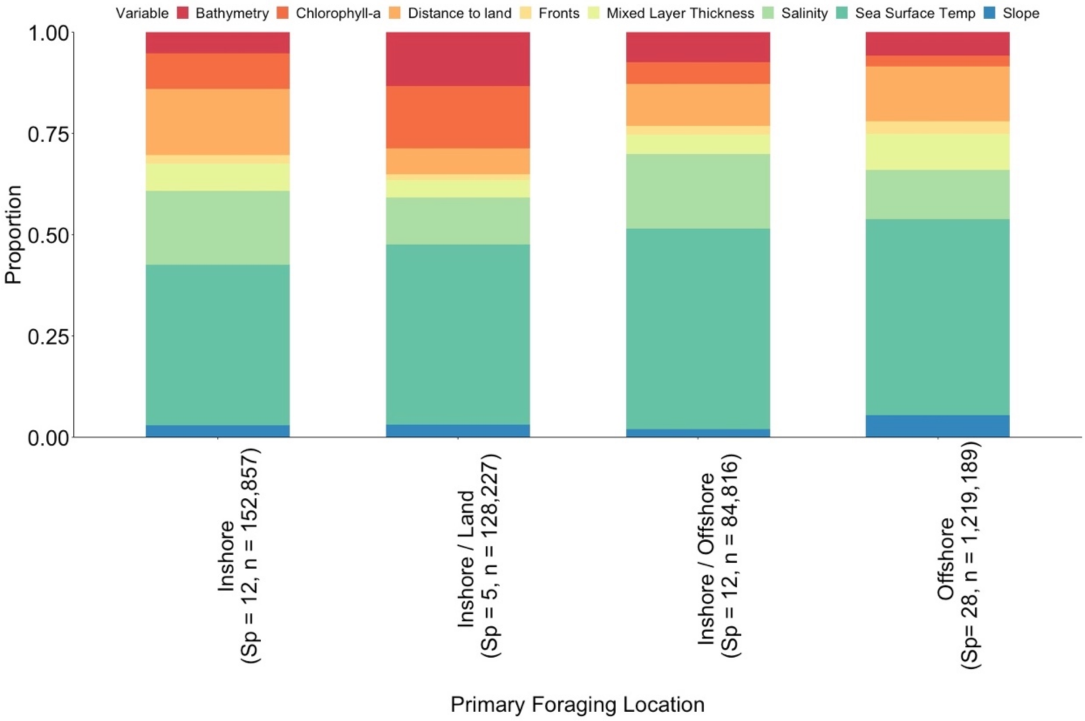


Figure 5.11 Proportions of relative variable importance by feeding location.

Table 5.13 Relative contribution of environmental variables when species are split by their foraging depth. MLT = mixed layer thickness, SST = sea surface temperature.

Foraging location	Bathymetry	Chl *a*	Distance to land	Fronts	MLT	Salinity	SST	Slope
Inshore	5.21%	8.80%	16.41%	1.94%	6.90%	18.20%	39.60%	2.93%
Inshore / Land	13.31%	15.41%	6.35%	1.46%	4.33%	11.52%	44.58%	3.05%
Inshore / Offshore	7.37%	5.33%	10.36%	2.22%	4.82%	18.36%	49.62%	1.91%
Offshore	5.87%	2.62%	13.54%	3.08%	8.86%	12.26%	48.31%	5.47%

Foraging strategy

SST had the highest contribution to all foraging strategies (ranging from 26.9% to 75.6%) except for lunge/filter feeding, which mixed layer thickness was most important (48.8%). Bathymetry was most important for diving (12.0%) and lunge feeding species (13.3%). Chlorophyll *a* had high important in surface plunge diving species (17.3%). Distance to land had high proportions in filter feeding species (37%) and lunge / filter feeding species (40.2%), and surface dabbler species. Fronts had minimal proportions across all strategies except for lunge feeding (12.9%) and mixed layer thickness was very important for lunge/filter feeding species (48.8%), though sample size is small for this category. Salinity had high proportion for plunge (32.8%) and shallow / kleptoparasite / carrion species (71.7%). Slope was important for suction feeding species (Cuvier's beaked whale *Ziphius cavirostris,* pilot whale, sunfish *Mola mola*) (24.4%), but had low percentage contribution for other foraging strategies.

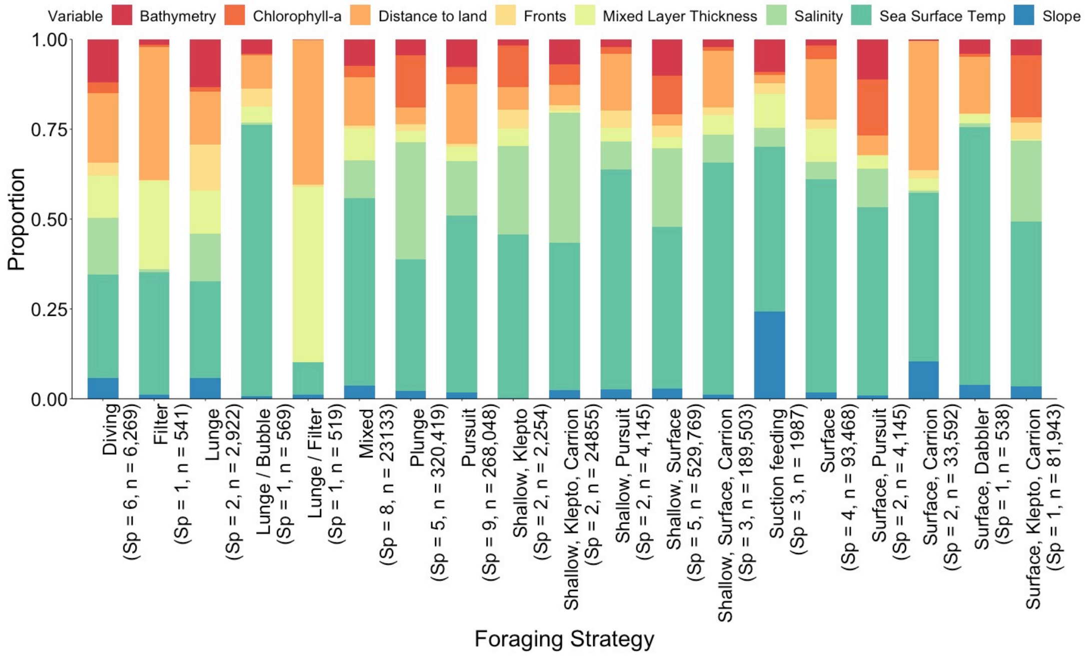


Figure 5.12 Proportions of relative variable importance by foraging behaviour.

Table 5.14 Relative contribution of environmental variables when species are split by foraging strategy. MLT = mixed layer thickness, SST = sea surface temperature.

Foraging strategy	Bathymetry	Chl *a*	Distance to land	Fronts	MLT	Salinity	SST	Slope
Diving	12.0%	3.0%	19.3%	3.6%	11.7%	15.8%	28.8%	5.8%
Filter	1.5%	0.6%	37.0%	0.3%	24.5%	0.9%	34.1%	1.1%
Lunge	13.3%	1.2%	14.7%	12.9%	11.9%	13.4%	26.9%	5.7%
Lunge / Bubble	4.1%	0.3%	9.3%	5.0%	4.5%	0.6%	75.6%	0.6%
Lunge / Filter	0.1%	0.0%	40.2%	0.7%	48.8%	0.1%	9.1%	1.1%
Mixed	7.3%	3.2%	13.3%	0.9%	8.8%	10.6%	52.2%	3.6%
Plunge	4.3%	14.6%	4.8%	1.8%	3.1%	32.8%	36.4%	2.3%
Pursuit	7.7%	4.7%	16.7%	0.6%	4.0%	15.1%	49.4%	1.7%
Shallow, Kleptoparasite	1.8%	11.5%	6.3%	5.2%	4.9%	24.6%	45.6%	0.1%
Shallow, Kleptoparasite, Carrion	6.9%	5.8%	5.7%	1.4%	0.7%	36.1%	41.1%	2.3%
Shallow, Pursuit	2.2%	1.8%	15.9%	4.8%	3.7%	7.8%	61.2%	2.6%

Shallow, Surface	10.1%	10.7%	3.3%	3.0%	3.3%	21.7%	45.1%	2.8%
Shallow, Surface, Carrion	2.1%	1.0%	15.9%	2.1%	5.3%	8.0%	64.5%	1.1%
Suction feeding	9.1%	0.9%	2.2%	3.0%	9.5%	5.2%	45.8%	24.4%
Surface	1.8%	3.7%	16.8%	2.7%	9.2%	4.9%	59.3%	1.7%
Surface, Carrion	11.1%	15.6%	5.5%	0.1%	3.6%	10.7%	52.3%	1.0%
Surface, Dabbler	0.3%	0.0%	36.0%	2.5%	3.2%	0.7%	46.9%	10.3%
Surface, Kleptoparasite, Carrion	3.9%	0.9%	15.9%	0.1%	2.5%	1.1%	71.7%	3.8%
Surface, Plunge	4.4%	17.3%	1.6%	4.5%	0.5%	22.5%	45.9%	3.3%

Diet

When split into diet category (carnivore, omnivore, planktivore) (Figure 5.13,Table 5.15), SST was the most important variable for all species diet types (Table 5.15), but distance to land and mixed layer thickness was more important in planktivores (27.34% and 25.56% respectively), but had low salinity (2.52%) whilst for carnivores and omnivores salinity had high contribution (15.90% and 13.46%).

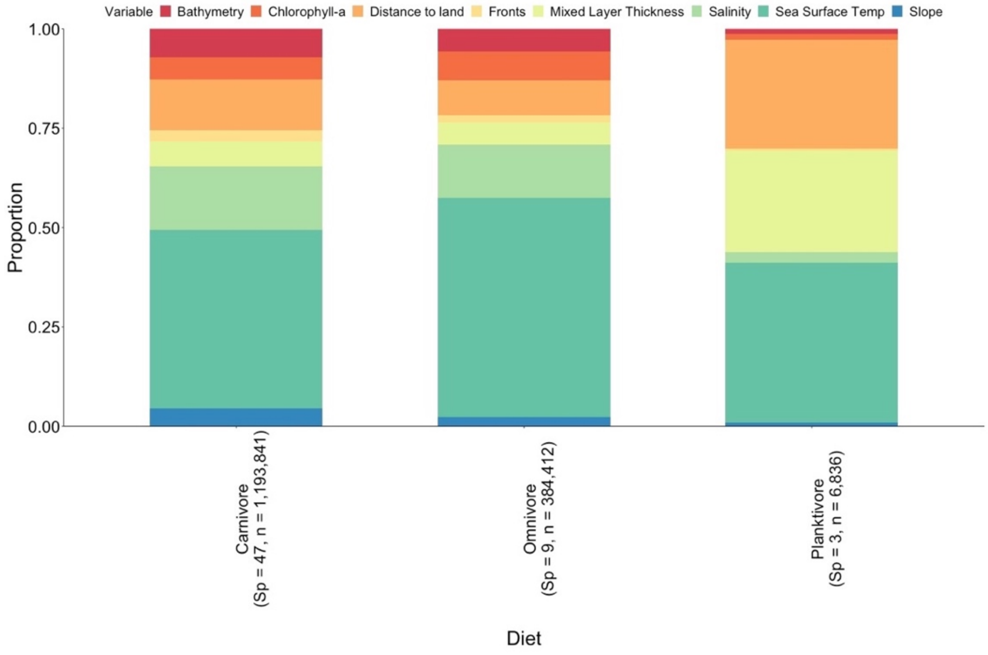


Figure 5.13 Proportions of relative variable importance by diet.

Table 5.15 Relative contribution of environmental variables when species are split by diet (carnivore, omnivore, planktivore). MLT = mixed layer thickness, SST = sea surface temperature.

Diet category	Bathymetry	Chl *a*	Distance to land	Fronts	MLT	Salinity	SST	Slope
Carnivore	7.20%	5.65%	12.72%	2.75%	6.28%	15.90%	45.06%	4.43%
Omnivore	5.79%	7.10%	8.78%	1.84%	5.59%	13.46%	55.22%	2.23%
Planktivore	1.24%	1.58%	27.34%	0.54%	25.56%	2.52%	40.27%	0.96%

When focusing on the prey type (Figure 5.14, Table 5.16), SST was most important for all prey types except those that fed on 'Cephalopods, Crustaceans, Fish, Molluscs, Vegetation' for whom salinity was most important (50.15%) although this was based upon one species of turtle. Slope was important for species that fed on cephalopods (22.57%), whilst distance to land was important for species (orca) that fed off a range of prey sources (Cephalopods, Fish, Mammals, Seabirds, Sea turtles). Mixed layer thickness had high importance for species that fed off plankton (25.56%).

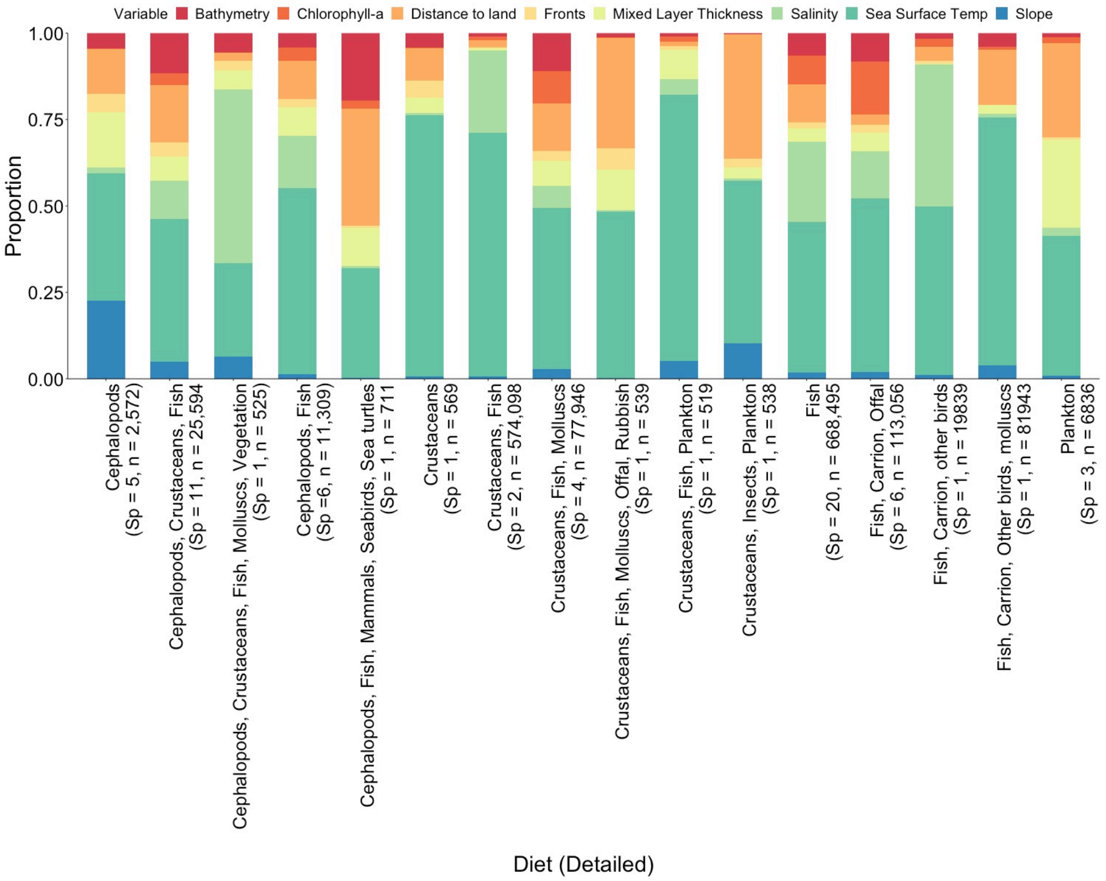


Figure 5.14 Proportions of relative variable importance by prey type.

Table 5.16 Relative contribution of environmental variables when species are split by primary prey choice. MLT = mixed layer thickness, SST = sea surface temperature.

Prey type	**Bathymetry**	**Chl *a***	**Distance to land**	**Fronts**	**MLT**	**Salinity**	**SST**	**Slope**
Cephalopods	4.51%	0.18%	12.80%	5.49%	15.90%	1.59%	36.87%	22.67%
Cephalopods, Crustaceans, Fish	11.56%	3.46%	16.64%	4.04%	7.01%	11.06%	41.35%	4.87%
Cephalopods, Crustaceans, Fish, Molluscs, Vegetation	5.64%	0.06%	2.28%	2.80%	5.61%	50.15%	26.98%	6.46%
Cephalopods, Fish	4.25%	3.83%	11.07%	2.33%	8.32%	15.02%	53.87%	1.31%
Cephalopods, Fish, Mammals, Seabirds, Sea turtles	19.40%	2.34%	33.83%	0.78%	10.93%	0.74%	31.70%	0.27%
Crustaceans	4.07%	0.29%	9.30%	5.04%	4.47%	0.59%	75.65%	0.60%
Crustaceans, Fish	1.01%	1.08%	2.00%	0.22%	0.70%	23.90%	70.47%	0.61%
Crustaceans, Fish, Molluscs	10.97%	9.47%	13.74%	2.88%	7.05%	6.52%	46.51%	2.86%
Crustaceans, Fish, Molluscs, Offal, Rubbish	1.28%	0.03%	32.12%	6.12%	11.57%	0.41%	48.34%	0.13%
Crustaceans, Fish, Plankton	0.93%	1.56%	1.21%	1.21%	8.49%	4.38%	76.98%	5.24%
Crustaceans, Insects, Plankton	0.34%	0.01%	36.00%	2.47%	3.24%	0.74%	46.90%	10.30%
Fish	6.48%	8.39%	11.04%	1.67%	3.89%	23.23%	43.48%	1.83%
Fish, carrion, offal	8.12%	15.44%	2.95%	2.34%	5.36%	13.68%	50.12%	1.97%
Fish, carrion, other birds	1.61%	2.44%	3.87%	0.86%	0.30%	41.06%	48.73%	1.15%
Fish, carrion, other birds, molluscs	3.94%	0.92%	15.87%	0.05%	2.53%	1.13%	71.75%	3.81%
Plankton	1.24%	1.58%	27.34%	0.54%	25.56%	2.52%	40.27%	0.96%

IUCN status

SST remained the highest variable importance across IUCN status group (21.58% to 50.16%) however for critically endangered species salinity was nearly equal in proportion (32.59%) suggesting SST and salinity are of high importance for those species. For endangered and near threatened species, distance to land was of high proportion (38.61% and 31.12% respectively).

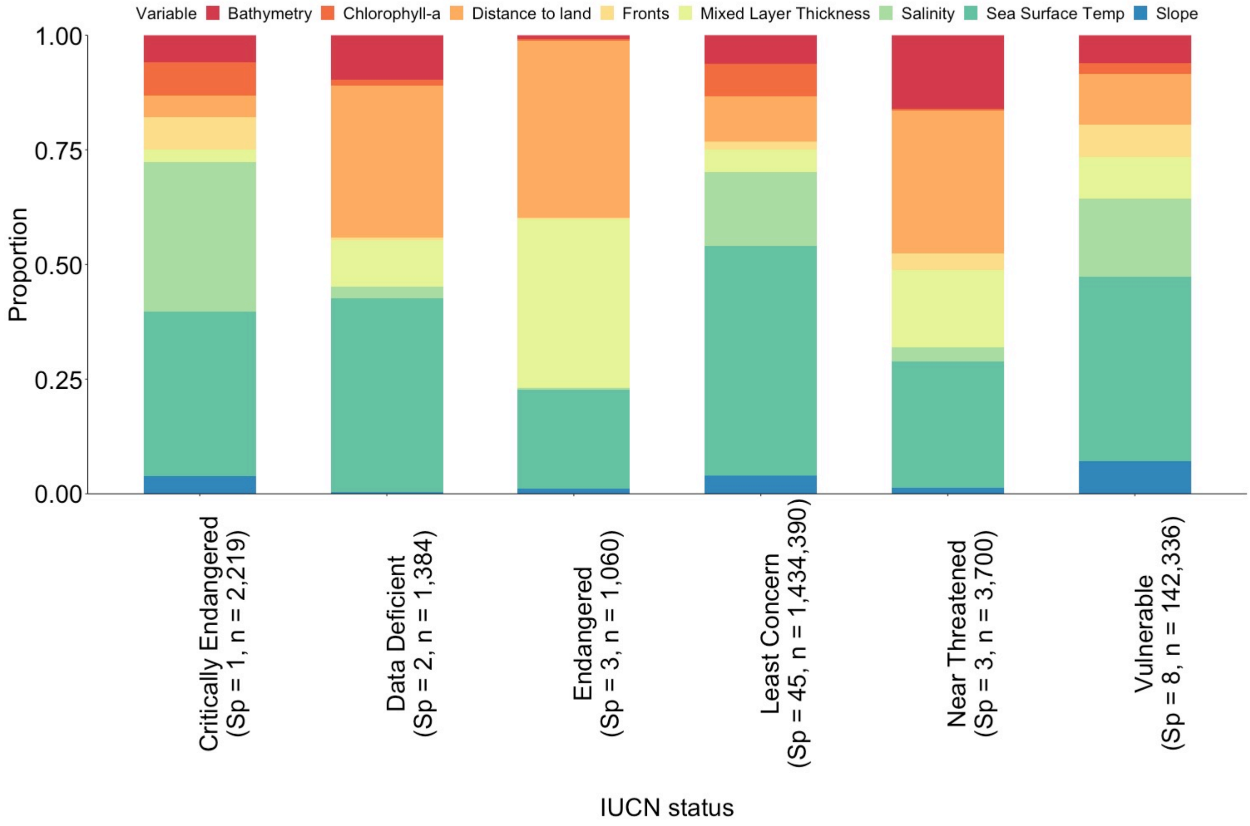


Figure 5.15 Proportions of relative variable importance by IUCN status.

Table 5.17 Relative contribution of environmental variables when species are split by IUCN status. MLT = mixed layer thickness, SST = sea surface temperature.

IUCN status	Bathymetry	Chl *a*	Distance to land	Fronts	MLT	Salinity	SST	Slope
Critically endangered	5.87%	7.24%	4.70%	7.04%	2.78%	32.59%	35.93%	3.83%
Data deficient	9.74%	1.32%	33.12%	0.42%	10.20%	2.50%	42.44%	0.26%
Endangered	0.78%	0.34%	38.61%	0.49%	36.66%	0.48%	21.58%	1.06%
Least Concern	6.28%	6.97%	9.99%	1.77%	4.83%	16.11%	50.16%	3.89%
Near Threatened	16.14%	0.32%	31.12%	3.69%	16.82%	3.12%	27.56%	1.23%
Vulnerable	6.08%	2.30%	11.13%	7.09%	8.99%	17.03%	40.23%	7.14%

5.5 Discussion

This study is the first to map species richness of marine predators from six different taxa at such a large spatial scale around the UK. It is the largest collation of its kind (69 species) using diverse survey data for marine mammals, seabirds, sharks, fish, and turtles, comparable in size only to recent studies on distribution for twelve species of cetacean and twelve of seabirds (Waggitt *et al.*, 2019). Waggitt *et al.* (2019) utilise a similar dataset which collates aerial and vessel survey data from the North-East Atlantic between 1980 and 2018 and a similar methodology (using SDMs) however the study focused on modelling monthly predicted distribution maps of cetacean and seabird populations at 10 km resolution, rather than measures of biodiversity such as species richness.

Marine predators face a high level of threat due to climate change and these results can help stakeholders and citizens to advocate for actions in priority areas by identifying areas of high and low species richness, which is a fundamental persistent measure of biodiversity. Predicting accurate species richness is a significant challenge and it is important to have a robust baseline of mapped biodiversity, particularly when wanting to model future changes in response to climate change. Amongst a lack of consensus globally of what to monitor to assess biodiversity change, species richness remains an Essential Biodiversity Variable (EBV) that can be used across taxa and for studies applicable in terrestrial, freshwater and marine realms (Pereira *et al.*, 2013). Having a measure which is easily applicable is crucial for marine species, where available baseline data is substantially lacking or of low resolution compared to terrestrial environments (UNEP, 2015, Shumway *et al.*, 2018) and may limit the accuracy of biodiversity mapping (Martin *et al.*, 2015). Oceans are dynamic and multi-dimensional meaning the boundaries of ecosystems move in space and time (Shumway *et al.*, 2018), which can create significant challenges for mapping and monitoring, over more fixed terrestrial systems. Additionally, compared to terrestrial systems, species inhabiting marine realms often have large-scale dispersal (at various life stages) and migrations. This is particularly true for marine predators which have long-distance movements that can span between countries and continents. Comprehensive monitoring is challenging across methods; with traditional ship-surveys being expensive, autonomous underwater vehicles and floats sparsely distributed, remote sensing lacking the ability to capture in 3-D and ecological monitoring often comprising extraction-based approaches thus being lethal (Hazen *et al.*, 2019). Therefore, whilst mapping species diversity in these habitats is complex, utilising available datasets in innovative and cost-effective ways that can rapidly assess biodiversity change are vital. This study compared different approaches in modelling species richness using historical heterogenous survey data, and demonstrated all are successful in mapping marine predator biodiversity around the UK, but models must be chosen based upon research question. MEMs are successful quick tools for modelling overall relationships between biodiversity and environment but cannot allow interrogation of species level patterns. SSDMs in comparison are computationally more intensive and less wieldy for understanding responses at a community level (e.g. they provide an overall importance value but no plots of relationships) but allow for exploration of individual species responses.

5.5.1 Maps of marine predator biodiversity

A key finding of this study is that quantifying cross-taxa species richness using MEMs and SSDMs show very similar spatial richness patterns when mapped comparatively. SSDM models give higher mean and maximum richness values than the comparative GAM, and distribution of richness is more evenly spread, with a wider uniform histogram. Modelling using SRPUE gave lower richness but with a unimodal distribution. Studies have asserted SSDMs may over-predict species richness (Guisan and Rahbek, 2011), potentially due to thresholding of individual occurrence probabilities (Calabrese *et al.*, 2014) whilst MEMs may better predict areas of richness (Harris *et al.*, 2018). However, Biber *et al.* (2019) compared SSDM and MEM model types for 14,000 species across multiple taxa globally, and found outputs from both model types over-predicted areas with low richness and underpredicted areas with high richness. In this chapter, histograms of frequency of species richness may agree with Biber’s theory, with histograms of raw species richness showing a left-skewed distribution with closer spread and SSDMs showing more even, wide spread of values. SRPUE therefore may be most appropriate to mitigate over-predicted low richness areas and underpredicted high richness areas, demonstrating a normal distribution compared to the two other models. Very similar shaped distribution curves have been observed in a wide range of studies on species richness including aquatic biodiversity of fish, mussel, and crayfish species in the US (Elkins *et al.*, 2019), global microlichens (Mendonça *et al.*, 2020), trees in jarrah forests (Standish *et al.*, 2015) and South American anurans (Villalobos *et al.*, 2013).

When comparing raw species richness (not SRPUE as this utilises a different non-integer scale), MEMs predicted lower maximum richness values than SSDMs, but hotspots appear are more localised. This may be due to patchiness of coverage (exacerbated by the methodology requirement of removing cells with observations of 2 or less for GAMS of raw species richness) with areas of very high survey effort over certain areas increasing the chances of detecting more new species, known as species-accumulation or collector's curve (Roswell *et al.*, 2021), but means lower-surveyed areas are likely underestimated. Maps using raw richness still appear to show remnants of survey effort, with high richness around the coastal areas and some shipping patterns still evident. Using effort as a predictor variable in the GAM highlighted that survey effort has a highly significant relationship with species richness (as discussed in detail in Chapter 4), but simply mapping outputs does not factor this in and therefore survey effort can remain an artefact in maps of richness. Whilst in theory, the two GAMS both use log(observations) to factor in effort, there are some important distinctions. GAMs of SRPUE adjusted the source of bias before modelling and did not require removing instances of two observations therefore allows inclusion of more data. Several studies have factored effort in advance of MEMs: Nicolas *et al.* (2010) standardised fish species richness by dividing by the log-transformed total sampled surface, whilst Gusmão *et al.* (2020) carried out individual-based rarefaction to control the effect of sampling effort (in this case, the number of collected birds) on mite richness. These studies highlight that species richness is a metric that is sensitive to differences in sampling effort (Gotelli and Colwell, 2001) and therefore comparing intensely with weakly sampled areas (such as using raw species richness GAMs) might bias richness comparisons. However, two advantages of the GAM of species richness are that predictions of richness can be produced on an interpretable integer scale which is well recognised and more practical for wider audience, and alternative distributions can be easily tested if required (e.g. poisson versus negative binomial). This study highlights model predictions should be carefully inspected and interrogated and used to infer the most plausible biological picture of species diversity depending on the research question.

However, all three methods still gave high species richness areas around the coastline of the UK, and particularly the East Coast of the UK, Scotland and along the coastline of mainland Europe (Belgium, the Netherlands, Germany, and Denmark), so as such these areas would be of good focus for conservation efforts if preserving areas of high species richness is a priority. These areas have been highlighted as high-density areas in distribution maps across the same spatial area (Waggitt *et al.*, 2019), with areas North of Scotland important for several cetacean and seabird species (Atlantic white-sided dolphin *Lagenorhynchus acutus*, Risso's dolphin, sperm whale *Physeter macrocephalus,* white-beaked dolphin *Lagenorhynchus albirostris*, Atlantic puffin *Fratercula arctica*) and high coastal coverage particularly along the north-east coast of the UK for seabird species (black-legged kittiwake *Rissa tridactyla*, common guillemot *Uria aalge*, and razorbill *Alca torda*).

5.5.2 Environmental drivers of species richness

Investigating environmental drivers of species richness demonstrated variables are significant predictors for marine predator biodiversity. MEMs showed all variables were significant drivers of species richness, with the GAM of species richness and SSDM methodology confirming the particular importance of SST, salinity and distance to land, whilst Chl-*a* was important for GAM of SRPUE. The SSDM in this study may still be susceptible to survey effort bias as, whilst the package always selects pseudo-absences for presence-only data, spatial thinning (i.e. removing the fewest possible records needed to reduce the effect of sampling bias, while retaining the greatest possible amount of information) was not utilised due to computational power and therefore may retain some level of sampling bias. Spatial sampling bias can lead to environmental bias in models because of the over-representation of certain environmental variables of the more accessible and well surveyed areas leading to overfitting (Kramer-Schadt *et al.*, 2013, Brown *et al.*, 2017). Where spatial filtering cannot be carried out, manipulation of background environmental datasets has been suggested. In a recent study on identifying marine invasion hotspots using SSDMs, Lyons *et al*. (2020) used an occurrence dataset that comprised records almost entirely from coastal areas (most < 30 m), most likely because of increased sampling near shore, and clipped environmental data layers to areas shallower than 100 m deep to reduce potential spatial bias in their models. Detailed environmental data manipulation unlikely to be possible for such a wide spatial study as this one, and therefore it is important to highlight the susceptibly of the model to survey bias in this instance and interpret results from SSDMs that do not include spatial thinning with caution.

Comparing the environmental parameters using all three methods allows more accurate delineation of drivers of species richness. SST appears to be of high importance in determining species richness hotspots in all models, with an optimum temperature for species richness (between 280 and 285 kelvin). SST has emerged a predictor of large scale patterns of biodiversity across taxa (Whitehead *et al.*, 2008). Evans *et al.* (2021) demonstrated an overall negative relationship of seabirds with SST, but relationship was non-linear and there were larger densities of birds at both cooler and warmer temperatures on the scale. Chaudhary *et al.* (2021) demonstrated richness is sensitive to temperature, with richness generally increasing as SST increases (similar to this study) until at plateau at ~20 °C for most taxa (noting this study was based at the equator, and therefore the peak for marine species around the UK is lower). Previously SST has been strongly correlated with marine mammal diversity (Astudillo-Scalia *et al.*, 2020) and seabird distribution (Krüger *et al.*, 2018, Serratosa *et al.*, 2020, Evans *et al.*, 2021), likely due to the fact they are considered indirect proxies for direct factors such as prey availability (Ainley *et al.*, 2009).

It must be noted that by combining all taxa into one GAM model, individual species patterns and nuances are lost, where this is retained in SSDMs. For example, short-tailed shearwaters *Puffinus tenuirostris* have demonstrated positive relationship with temperature, whilst high densities of albatross species have occurred in the coldest SST (Evans *et al.*, 2021) and so whilst SST appears to be critical in determining areas of high conservation value via species richness, it cannot be a blanket relationship for all species. Fullard *et al.* (2000) determined that even within a species (long-finned pilot whale *Globicephala melas*) population isolation occurs between areas of the ocean which differ in sea surface temperature. SST was important for those species that foraged at the surface, which is likely linked to targeting specific prey species e.g. gannets which are likely targeting epipelagic species of prey such as Atlantic mackerel (*Scomber scombrus*) or Atlantic herring (*Clupea harengus*) (Montevecchi *et al.*, 2009), or humpback whales targeting euphausiids which require sunlight in the upper layers (Trudelle *et al.*, 2016).

Given the influence of temperature on metabolic rates and physiological processes for all species (Helaouet *et al.*, 2013), it is unsurprising SST is a significant driver of richness. Therefore, ocean warming is a concern for species or biodiversity hotspots that are tied to SST, with marine species shifting their distributions in response (Burrows *et al.*, 2011, Poloczanska *et al.*, 2013, Poloczanska *et al.*, 2016). It may lead to species being unable to survive if they cannot move to new areas within their thermal niche or adapt their physiology or behaviour (Donelson *et al.*, 2019, Pinsky *et al.*, 2022). Whitehead *et al.* (2008) predicted a decline in cetacean diversity in warmer tropic waters and increases at higher latitudes, and this reduction at mid-latitudes and shift towards the poles has been observed across taxa (Chaudhary *et al.*, 2021).

Species richness shows an important and increasing relationship with salinity (the more saline the water, the more species observed) in both MEMs and confirmed by overall variable importance in the SSDM. Salinity has been identified as a critical driver of high conservation priority for marine mammals particularly cetaceans (Astudillo-Scalia *et al.*, 2020), and seabird and seal distribution (Serratosa *et al.*, 2020). Sea surface salinity, for example, has been deemed an important predictor in bottlenose dolphin (with the majority of dolphins preferring areas of high salinity) (White *et al.*, 2018), humpback whales and harbour porpoise *Phocoena phocoena* (Tynan *et al.*, 2005) habitat use, though it does appear to vary seasonally. Areas of higher salinity (maps in Appendix H) are in more enclosed areas (known as ‘concentration basins’) such as the Mediterranean Sea and particularly in the Adriatic Sea (Notarbartolo di Sciara, 2002), due to slow exchange of water between the Mediterranean, Atlantic Ocean and Black Sea. The Mediterranean Sea has been deemed one of the worlds biodiversity hotspots (Lejeusne *et al.*, 2010, Pace *et al.*, 2013) with increased endemism and high species richness (Pace *et al.*, 2013), with 21 species of cetacean have been recorded in the Mediterranean Sea for example (Reeves and Notarbartolo di Sciara, 2006, Notarbartolo di Sciara and Birkun, 2010). Therefore, whilst salinity has been highlighted in the literature as critical driver, the SSDM methodology describes associations and therefore may associate high areas of salinity with known areas of high biodiversity. Individual relationships may be hidden in large scale cross-taxa study, for example seals showed a bimodal relationship with salinity (Bailey and Thompson, 2009), but this is not visible at a multiple species scale.

Bathymetry was a significant predictor in MEMs, though lower in variable importance than other drivers in MEMs and SSDMs. Bathymetry has been deemed an important predictive variable for seabirds (Hyrenbach *et al.*, 2007, Nur *et al.*, 2011, Amelineau *et al.*, 2016) and marine mammals (Astudillo-Scalia *et al.*, 2020, Awbery *et al.*, 2022). Sea-floor depth has been shown to influence prey availability and foraging efficiency in marine predators (Guinet *et al.*, 2001, Yen *et al.*, 2004). When explored in further detail, bathymetry was important for diving and lunge feeding species. Colonies with shallower waters may be more suitable to diving seabirds where prey is easier to catch, compared to deeper waters (Chiaradia *et al.*, 2007, Meyer *et al.*, 2017, Sutton *et al.*, 2020). Slope held less variable importance in MEM and SSDM models, though significant in GAMs and showed richness increased with slope. Slope appeared important for suction feeding species that feed on cephalopods (Cuvier's beaked whale, pilot whale, sunfish), and this could result from prey distribution over slopes, rather than an association of top predators directly with slope. Squids that constitute the main prey in Ziphiidae for example, prefer areas of steep slope such as submarine canyons (Riedl, 1983, Guerra, 1992, Cañadas *et al.*, 2002). Slope was most important in offshore species when split by foraging location, which again directs towards the prey source rather than a direct association. Deep diving species such as sperm whales and beaked whales have shown highest relative densities along continental slopes (Virgili *et al.*, 2019). Behavioural flexibility such as adapting specific foraging techniques allows animals to adapt to prevailing environmental conditions throughout their foraging range, for example short-finned pilot whales have potentially developed specialized foraging behaviours for targeting steep bathymetric gradients where they use the barriers to herd prey (Thorne *et al.*, 2017), much like other cetaceans which steer prey against the surface (Vaughn *et al.*, 2008). Striped dolphins, long-finned pilot whales, sperm whales, Risso's dolphins and beaked whales have shown a preference for waters deeper than 600 m and areas with intermediate or high slopes (Cañadas *et al.*, 2002). There is evidence of seabird association with slope (Amorim *et al.*, 2009, Wakefield *et al.*, 2017) but varies by species, and slope was of low proportion in seabirds when split by taxa in this study.

Salinity, temperature, and bathymetry have important roles in determining water density (Astudillo-Scalia *et al.*, 2020), highlighting variables may combine to determine species distributions rather than single factors alone. The models used in this study do not look at interactions between environmental variables (only longitude and latitude) and indicates an important next step for this research study. Mauritzen *et al.* (2012) found that density mediates heat uptake of deeper oceans and have important temperature regulation properties in the deep North Atlantic Ocean. The prevalence of SST (most important in GAM of species richness and SSDM) and secondly salinity as important drivers of species richness in this study highlights a critical concern for the future. Global anthropogenic warming leading to rising ocean temperatures (Hansen *et al.*, 2006) and changes in salinity of surface waters (Du *et al.*, 2019) which increases near-surface stratification and amplifies surface salinity patterns (Zika *et al.*, 2018) may therefore change species distributions or reduce of suitable habitat.

Species richness declined with increasing distance to land in plots from the MEMs, suggesting coastal waters are important for multiple species. Both GAMs and the SSDM showed distance to land was a very important overall driver. For GAMs, distance to land was the second most important variable, reiterating there may be an element of survey effort that remains in this method (increased survey effort in accessible coastal areas and therefore increased species accumulation curve). However, whilst studies comparing inshore and offshore biodiversity gradients are rare and conflicting paradigms exist (Gray, 2001), richness has been demonstrated to be highest in coastal waters compared to offshore waters in demersal fish species (van Lleeuwen *et al.*, 1994) and deep sea benthos (Gray, 2001) and therefore likely to impact biomass up the food chain, with distance to shore shown to be a significant predictor in other studies on marine mammals and seabirds (Bailey and Thompson, 2009, Nur *et al.*, 2011, Retana and Lewis, 2017). The distance a species travels is also likely limited by travel costs (time and energy) associated with distance from colonies or haul out (Patenaude-Monette *et al.*, 2014) and will be highly species specific. For example grey seals *Halichoerus grypus* travel further out to forage than harbour seals *Phoca vitulina* (Thompson *et al.*, 1996). Nord and Forslund (2015) for example found distance to open sea had opposing effects on species groups, with the number of specialist species decreasing with increasing distance to open sea whilst the numbers of generalists increased. SSDMs have the benefit of being able to give the importance of species-specific drivers, by giving importance per species, whilst GAMs do not allow that capability.

It was expected that distance to land would be particularly important to those species that are CPFs such as pinnipeds or seabirds. Access to land is important for breeding and moulting of pinnipeds (Cronin *et al.*, 2014) whilst seabirds have colonies for nesting (Ballance, 2008). However, SSDMs showed that distance to land had a higher proportion of importance to non-CPFs over CPFs, and this aligns with studies who found richness were not significantly associated with distance to land, attributing persistence to other unique biological conditions and complex habitat heterogeneity (Santora and Veit, 2013).

GAMs predicted richness to increase with chl *a,* with peaks up to 4 mg^{-3}, and was most important in GAM of SRPUE, but was of low importance in SSDMs, though was more important for birds and pinnipeds than other taxa. There is an array of literature supporting association between chl *a* and high marine predator densities or diversity, across taxa e.g. cetaceans (Scott *et al.*, 2010, Trudelle *et al.*, 2016, Zerbini *et al.*, 2016, Chen *et al.*, 2020, Salgado Kent *et al.*, 2021), seabirds (Louzao *et al.*, 2006, Embling *et al.*, 2012, Sabarros *et al.*, 2014, Bennison and Jessopp, 2015, Grecian *et al.*, 2016, Kane *et al.*, 2020) and sharks (Miller *et al.*, 2015). However, the relationship between predators and chlorophyll levels has been demonstrated to vary considerably by species and by location, and therefore a large scale, cross-taxa study will not show such species-specific associations with chl *a.* For example even within the Sulidae family, Cape gannets *Morus capensis* foraged in response to chl *a* concentration (Sabarros *et al.*, 2014) but northern gannet *Morus bassanus* foraging locations were not related (Suryan *et al.*, 2012). Some studies have questioned the use of surface chl *a* as a driver of marine biodiversity (Grémillet *et al.*, 2008, Suryan *et al.*, 2012), as marine predators do not feed directly on phytoplankton but several trophic levels higher and therefore may explain why chl *a* may have not been a clear driver in species richness in this study. Studies exist which have shown fish species richness has been negatively associated with higher concentrations of chl *a* (Lazzari *et al.*, 2020).

Mixed layer thickness had less importance overall, but was important for shark taxa, but only two species were included in the SSDM: basking shark and blue shark *Prionace glauca*. These are both pelagic species, and fine-scale vertical movements of other epipelagic species of shark (oceanic white-tip *Carcharhinus longimanus*) have showed most of the time was spent in the mixed layer (Tolotti *et al.*, 2017), albeit with complex movement patterns. Basking sharks have been shown to switch behaviours to a tidal rhythm when encountering boundaries between thermally stratified and mixed waters (Shepard, 2006). Mixed layer thickness was very important for lunge and filter feeding species (though sample size was small and includes basking shark) and was also important in planktivores. Shallow mixed layer depth may act as a barrier to trap fish to within near-surface diving range of birds such as albatross and terns (Urmy and Warren, 2018, Waggitt *et al.*, 2018a) or plankton in the surface layer of the water column making them more available for surface feeders, such as Balaenopteridae species. Whales average dive depth has been shown to be strongly correlated with the average depth of the bottom mixed layer and influences the spatial distribution of North Atlantic right whales *Eubalaena glacialis* (Baumgartner *et al.*, 2003, Baumgartner and Mate, 2003). Prey availability at particular depths is likely to influence predator foraging and distribution, and for marine mammals foraging behaviour and physiology will dictate what aspects of the water column are important to each species e.g. mesopelagic prey for deep divers, epipelagic prey for visual hunters, benthic prey for bottom feeders, thin prey layers for ram filter feeders, dense prey schools for lunge feeders (Palacios *et al.*, 2013).

Unexpectedly, whilst still a significant predictor, fronts showed the lowest importance in GAMs and SSDM overall. This may have been due to low importance for all species overall or it may mask the possibility that fronts may be of very high importance for just a small number of species – for example fronts were more important for lunge feeding species and those that feed on cephalopods. Fronts have been demonstrated as a predictor of marine predator presence for sharks (Miller *et al.*, 2015), turtles (Scales *et al.*, 2015), seabirds (Bost *et al.*, 2009, Cox *et al.*, 2018) and cetaceans (Cox *et al.*, 2018, Bestley *et al.*, 2019, Virgili *et al.*, 2019). However, given fronts are highly dynamic and active, the method used in this study of averaging the rasters over the 39-year period may lose the required temporal detail needed to reveal associations between species, hence low importance overall which contrasts recent literature. Fronts can vary seasonally and interannually with changes in climate and wind (Bograd, 2017) and therefore have consequences on predators. For example, during years where upwelling intensity decreased, the availability of high-quality foraging habitats around frontal zones reduced and breeding seabirds showed signs of reduced body condition (Croll, 1990) or extensive foraging trips which may result in less reproductive success (Becker and Beissinger, 2003, Bograd, 2017). More fine scale temporal patterns may be required to determine whether there are high biodiversity associations with fronts.

An interesting finding from this study is that when variable importance was considered by IUCN status category, critically endangered species showed high proportions of SST and salinity drivers, further indicating increasing global warming and climate change is a specific threat to species already at extremely high risk of extinction. This pattern has been observed across marine predator taxa, with models identifying SST and salinity (alongside depth and chlorophyll *a* concentration) to be of importance in the critically endangered species: angel shark (*Squatina squatina*) in coastal waters of Wales and the central Irish Sea (Barker *et al.*, 2022) and in the Canary Islands (Meyers *et al.*, 2017), scalloped hammerhead (*Sphyrna lewini*) (Zhang, 2022), Balearic shearwaters *Puffinus mauretanicus* (Zhang, 2022).

5.5.3 Niche partitioning

The differences observed between the environmental drivers when split by groupings such as taxa, foraging strategy, diet etc. may allude to niche partitioning. Niche partitioning allows sympatric predators to coexist, via resource partitioning (species specialise in different prey items), spatial partitioning (forage or hunt in different areas) and temporal partitioning (sympatric species alternate their times of peak foraging on a diel or seasonal scale) (Lear *et al.*, 2021). These regimes can also shift based on environmental variation or changes in species assemblage (Kronfeld-Schor and Dayan, 2003, McCauley *et al.*, 2012, Monterroso *et al.*, 2014, Frey *et al.*, 2017). Particularly for marine predators, who have top-down control on ecosystems (Estes *et al.*, 2011), understanding drivers of coexistence may aid in understanding partitioning mechanisms. Intraguild interactions between top marine predators that co-occur in the same area with overlapping environments can lead to complex interactions and affect the distribution and abundance of the predator populations (Ritchie and Johnson, 2009). Furthermore , given the rate of anthropogenic environmental change causing shifts in weather (Harley, 2011, Hazen *et al.*, 2012, Sepulveda *et al.*, 2020), habitat destruction (Airoldi *et al.*, 2008, Jungblut *et al.*, 2018), invasive species (Essl *et al.*, 2020), prey overexploitation (Bearzi *et al.*, 2006) and predator mortality (Ripple *et al.*, 2013), understanding the current patterns and drivers of partitioning in predator guilds; will be beneficial for conservation. Several studies have examined the various mechanisms of marine predator guilds (Lear *et al.*, 2021), particularly in elasmobranchs (Papastamatiou *et al.*, 2006, Humphries *et al.*, 2016, Navarro *et al.*, 2016, Shipley *et al.*, 2019, Dwyer *et al.*, 2020).

The differences in the proportions of environmental drivers between taxa and foraging strategies may suggest that marine predators target different environments and prey species. Foraging range is limited by oceanographic conditions and movement capabilities of species (Scales *et al.*, 2016). For example, bathymetry and chlorophyll *a* had very little importance in sharks but is higher in birds and pinnipeds. Studies have suggested intraspecific differences in foraging behaviour target habitat differences on a local scale, and bathymetry depth of foraging (alongside water layer temperature, depth and distance to coast) is important in diving seabirds (Masello *et al.*, 2010), petrel taxa, and sympatric alcids (Gulka *et al.*, 2019).

Slope had little importance for pinnipeds, birds and sharks, but much higher in cetaceans. It is known that some deep-diving cetaceans target deep steep slopes (Cañadas *et al.*, 2002, Thorne *et al.*, 2017, Virgili *et al.*, 2019), and have adapted foraging behaviours such as herding prey against the bathymetric gradients. Equally, slope had little importance for bird taxa, which (in general) cannot forage as deep as marine mammals that have evolved refined aquatic adaptions to diving (Huelsmann *et al.*, 2019, Williams and Ponganis, 2021). For birds, sea surface temperature and salinity were instead very important, and is confirmed in the literature; Costa *et al.* (2020) have shown the most important factor explaining the density of flying seabirds was the difference between surface and bottom salinity.

However, the fundamental complexity of niche partitioning makes understanding direct associations difficult and the large temporal and spatial scale used in this study cannot unpick the detailed ecological interactions between predators. It has been shown for example that orca redistribute white shark *Carcharodon carcharias* foraging pressure on seals (Jorgensen *et al.*, 2019) and this level of interaction cannot be revealed in this study. Insight from this study into the various mechanisms of niche partitioning can vary. Resource partitioning can be partly explored by looking at the importance of environmental drivers by diet – for example mixed layer thickness was important for planktivores. Patterns in spatial partitioning may be evident when split by taxa, with different habitats being targeted. Influences of temporal partitioning are not reflected with this methodology, as it does not delimit patterns down to annual or monthly scales and compromises temporal resolution for spatial coverage (maximum occurrence data).

However, migration and niche partitioning have been shown to increase species richness (and species rarity) (Ai *et al.*, 2013), therefore the presence of hotspots of high species richness in itself suggests evidence of niche partitioning – where species go exist, they will reduce competition though niche partitioning or spatial segregation (Jessopp *et al.*, 2020). Though sighting data were aggregated over four decades, the sightings of those species still occurred within a relatively small geologic timescale and therefore likely to coexist spatially at some point. Whilst theories of environmental drivers being important different species can be elucidated to some extent, more detailed study is required to untangle any temporal partitioning.

Nonetheless, this study is the first to combine such large coverage and provide indication of areas of high species richness around the UK. It is important to highlight that my study is one of association between environmental drivers, and not causation. Limitations due to computational power can limit the conclusions drawn with the models including many occurrence data points over a vast spatial scale, and detailed layers of environmental parameters. Whilst MEMs can speed up modelling, they lose the ability to explore patterns on a species-level. SSDMs are conversely computationally difficult but retain more detail biological information. The 100 km^2 grid size was chosen due to computational limits of the SSDM modelling, however grid size will influence the associations and richness captured (especially in patchy presence-only data) (Volis, 2022). In regions where biological data are sparse, such as the marine environment, agreeing on a cell size that represents truthful patterns but keeps false-absence rates low is a fine balance. Future analysis may reveal different associations depending on grid size. For example, in Bailey and Thompson (2009) grid sizes affected whether variables were significant or not, with distance to shore being significant for porpoises in a 2 km^2 grid but not in a 1 km^2 grid, depth was instead. Further study looking at the impact of grid sizes on environmental drivers would be beneficial if suitable scale and resolution environmental data can be obtained. Chapter 4 used three grid sizes to explore species richness patterns, and therefore this approach could be applied to smaller grid sizes (10 km^2 and 1 km^2) if sufficient computational power was available.

As mentioned previously, further breakdown into a monthly temporal scale would be highly valuable for future study, as seasonality is a key factor in these marine predator species. Movement patterns such as migration for reproduction, foraging migrations for food resources, or commuting between home ranges and foraging grounds, means predator distributions are highly variable in both space and time. For example seabird assemblages have shown strong seasonality (Commins *et al.*, 2013), and whilst are predictable between years, they vary within and therefore an average over the study time period will not account for within-year variability. This chapter combined occurrences from a dataset spanning 39 years, for using in richness calculations, and used averages of environmental variables, which therefore removes any temporal variability. It would be possible to repeat this methodology for each year within the dataset, or even month (as discussed in Chapter 4) but with a trade-off to data coverage. This study maximised spatial coverage and survey effort to identify biodiversity hotspots but loses potential fine-scale temporal patterns.

This study highlights how using different approaches (MEMs and SSDM) to predict species richness can produce similar mapping outputs but has different advantages in terms of computational speed versus the ability to dig into species level detail, and this must be carefully considered regarding the aim of the study. The SSDM was computationally demanding for the spatial scale and number of species in this study but retains important biological information and allows individualistic species responses to be interrogated, whilst MEMs allow for more rapidly analysis at a community level but loses species specific responses. However, if the end goal is to producing maps of species richness, the maps were comparable and when effort was factored in, equally suitable for identifying relative hotspots. When looking at environmental drivers of richness, GAMs which factored in effort in advance (SRPUE) showed similar relationships with environmental variables to GAMs of species richness, but removed a major source of bias before modelling. There is a danger that using the wrong method may lead to under prediction or over prediction of richness hotspots, and this may have implications for conservation or designations (such as MPAs). With emphasis on global biodiversity challenges (e.g. COP15 targets (Convention on Biological Diversity, 2021) and the Biodiversity Strategy for 2030 (European Commission, 2020)) it is important to evaluate methods when planning future management strategies, particular those that focus on richness as a key measure of biodiversity. Mitigation currently focuses on the identification and protection of Areas of Ecological Significance (AESs), which have so far not been determined at the ocean-basin scale (Hindell *et al.*, 2020).

5.6 Conclusion

In conclusion, this study can provide the first insight into areas of high and low species richness around the UK, a key EBV and critical for meeting global marine management and conservation objectives. This study demonstrates how analysis of existing and diverse data can be utilised in a cost-effective manner, as such vast coverage is hard to obtain given data is less easily gathered in the marine realm, and use of existing historical observation datasets can be beneficial for mapping biodiversity, particularly those of wide-ranging marine predators. While some caution is needed, due to temporal averaging and reliance on historical data, these distribution maps have widespread and immediate applications in identifying important areas of protection and providing focus for marine management strategies and is the first available output at this scale and coverage. In addition, the study identifies potential drivers of patterns of biodiversity, which provide invaluable information for understanding both current and future patterns in species richness in an era of rapidly declining biodiversity.

5.7 References

AI, D., CHU, C., ELLWOOD, M. D. F., HOU, R. & WANG, G. 2013. Migration and niche partitioning simultaneously increase species richness and rarity. *Ecological Modelling,* 258**,** 33-39.

AINLEY, D. G., DUGGER, K. D., FORD, R. G., PIERCE, S. D., REESE, D. C., BRODEUR, R. D., TYNAN, C. T. & BARTH, J. A. 2009. Association of predators and prey at frontal features in the California Current: Competition, facilitation, and co-occurrence. *Marine Ecology Progress Series,* 389**,** 271–294.

AIROLDI, L., BALATA, D. & BECK, M. W. 2008. The gray zone: relationships between habitat loss and marine diversity and their applications in conservation. *Journal of experimental marine biology and ecology,* 366**,** 8-15.

ALBOUY, C., DELATTRE, V. L., MÉRIGOT, B., MEYNARD, C. N., LEPRIEUR, F. & HAWKES, L. 2017. Multifaceted biodiversity hotspots of marine mammals for conservation priorities. *Diversity and Distributions,* 23**,** 615-626.

ALGAR, A. C., KHAROUBA, H. M., YOUNG, E. R. & KERR, J. T. 2009. Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography,* 32**,** 22-33.

AMELINEAU, F., GREMILLET, D., BONNET, D., LE BOT, T. & FORT, J. 2016. Where to Forage in the Absence of Sea Ice? Bathymetry As a Key Factor for an Arctic Seabird. *PLoS One,* 11**,** e0157764.

AMORIM, P., FIGUEIREDO, M., MACHETE, M., MORATO, T., MARTINS, A. & SERRAO SANTOS, R. 2009. Spatial variability of seabird distribution associated with environmental factors: a case study of marine Important Bird Areas in the Azores. *ICES Journal of Marine Science,* 66**,** 29–40.

ASTUDILLO-SCALIA, Y., ALBUQUERQUE, F. S. & PERES-NETO, P. 2020. The geography of high-priority conservation areas for marine mammals. *Global Ecology and Biogeography,* 29**,** 2097-2106.

AVILA, I. C., KASCHNER, K. & DORMANN, C. F. 2018. Current global risks to marine mammals: Taking stock of the threats. *Biological Conservation,* 221**,** 44-58.

AWBERY, T., AKKAYA, A., LYNE, P., RUDD, L., HOOGENSTRIJD, G., NEDELCU, M., KNIHA, D., ERDOĞAN, M. A., PERSAD, C., AMAHA ÖZTÜRK, A. & ÖZTÜRK, B. 2022. Spatial distribution and encounter rates of delphinids and deep diving cetaceans in the eastern Mediterranean sea of Turkey and the extent of overlap with areas of dense marine traffic. *Frontiers in Marine Science,* 9.

BAILEY, H. & THOMPSON, P. M. 2009. Using marine mammal habitat modelling to identify priority conservation zones within a marine protected area. *Marine Ecology Progress Series,* 378**,** 279-287.

BALLANCE, L. T. 2008. Understanding Seabirds At Sea: Why And How? *Marine Ornithology,* 35**,** 127-135.

BARKER, J., DAVIES, J., GORALCZYK, M., PATEL, S., O'CONNOR, J., EVANS, J., SHARP, R., GOLLOCK, M., WOOD, F. R., ROSINDELL, J., BARTLETT, C., GARNER, B. J., JONES, D., QUIGLEY, D. & WRAY, B. 2022. The distribution, ecology and predicted habitat use of the Critically Endangered angelshark (*Squatina squatina*) in coastal waters of Wales and the central Irish Sea. *J Fish Biol,* 101**,** 640-658.

BAUMGARTNER, M. F., COLE, T. V., CAMPBELL, L. A., TEEGARDEN, G. J. & DURBIN, E. G. 2003. Associations between North Atlantic right whales and their prey, *Calanus finmarchicus*, over diel and tidal time scales. *Marine Ecology Progress Series,* 264**,** 155-166.

BAUMGARTNER, N. & MATE, B. R. 2003. Summertime foraging ecology of North Atlantic right whales. *Marine Ecology Progress Series,* 364**,** 123-135.

BEARZI, G., POLITI, E., AGAZZI, S. & AZZELLINO, A. 2006. Prey depletion caused by overfishing and the decline of marine megafauna in eastern Ionian Sea coastal waters (central Mediterranean). *Biological Conservation,* 127**,** 373-382.

BECKER, B. H. & BEISSINGER, S. R. 2003. Scale-dependent habitat selection by a nearshore seabird, the marbled murrelet, in a highly dynamic upwelling system. *Marine Ecology Progress Series,* 256**,** 243–255.

BELL, W. J. & BELL, W. J. 1990. Central place foraging. *Searching behaviour: The behavioural ecology of finding resources***,** 171-187.

BENNISON, A. & JESSOPP, M. 2015. At-sea surveys confirm a North Atlantic biodiversity hotspot. *Bird Study,* 62**,** 262-266.

BESTLEY, S., ANDREWS-GOFF, V., VAN WIJK, E., RINTOUL, S. R., DOUBLE, M. C. & HOW, J. 2019. New insights into prime Southern Ocean forage grounds for thriving Western Australian humpback whales. *Sci Rep,* 9**,** 13988.

BIBER, M. F., VOSKAMP, A., NIAMIR, A., HICKLER, T. & HOF, C. 2019. A comparison of macroecological and stacked species distribution models to predict future global terrestrial vertebrate richness. *Journal of Biogeography,* 47**,** 114-129.

BIVAND, R., KEITT, T. & ROWLINGSON, B. 2021. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. 1.5-27 ed.

BLOCK, B. A., JONSEN, I. D., JORGENSEN, S. J., WINSHIP, A. J., SHAFFER, S. A., BOGRAD, S. J., HAZEN, E. L., FOLEY, D. G., BREED, G. A., HARRISON, A. L., GANONG, J. E., SWITHENBANK, A., CASTLETON, M., DEWAR, H., MATE, B. R., SHILLINGER, G. L., SCHAEFER, K. M., BENSON, S. R., WEISE, M. J., HENRY, R. W. & COSTA, D. P. 2011. Tracking apex marine predator movements in a dynamic ocean. *Nature,* 475**,** 86-90.

BOERTMANN, D. 2011. Seabirds in the central North Atlantic, September 2006: further evidence for an oceanic seabird aggregation area. *Marine Ornithology,* 39**,** 183–188.

BOGRAD, S. J. 2017. Scale of inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data. *Ecography,* 49**,** 210–220.

BOST, C. A., COTTÉ, C., BAILLEUL, F., CHEREL, Y., CHARRASSIN, J. B., GUINET, C., AINLEY, D. G. & WEIMERSKIRCH, H. 2009. The importance of oceanographic fronts to marine birds and mammals of the southern oceans. *Journal of Marine Systems,* 78**,** 363-376.

BREITBURG, D., LEVIN, L. A., OSCHLIES, A., GREGOIRE, M., CHAVEZ, F. P., CONLEY, D. J., GARCON, V., GILBERT, D., GUTIERREZ, D., ISENSEE, K., JACINTO, G. S., LIMBURG, K. E., MONTES, I., NAQVI, S. W. A., PITCHER, G. C., RABALAIS, N. N., ROMAN, M. R., ROSE, K. A., SEIBEL, B. A., TELSZEWSKI, M., YASUHARA, M. & ZHANG, J. 2018. Declining oxygen in the global ocean and coastal waters. *Science,* 359**,** 1-11.

BROWN, J. L., BENNETT, J. R. & FRENCH, C. M. 2017. SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ,* 5**,** e4095.

BRUMMITT, N., REGAN, E. C., WEATHERDON, L. V., MARTIN, C. S., GEIJZENDORFFER, I. R., ROCCHINI, D., GAVISH, Y., HAASE, P., MARSH, C. J. & SCHMELLER, D. S. 2017. Taking stock of nature: Essential biodiversity variables explained. *Biological Conservation,* 213**,** 252-255.

BURGER, A. E. 2003. Effects of the Juan de Fuca Eddy and upwelling on densities and distributions of seabirds off southwest Vancouver Island, British Columbia. *Marine Ornithology,* 31**,** 113-122.

BURROWS, M. T., SCHOEMAN, D. S., BUCKLEY, L. B., MOORE, P., POLOCZANSKA, E. S., BRANDER, K. M., BROWN, C., BRUNO, J. F., DUARTE, C. M., HALPERN, B. S., HOLDING, J., KAPPEL, C. V., KIESSLING, W., O'CONNOR, M. I., PANDOLFI, J. M., PARMESAN, C., SCHWING, F. B., SYDEMAN, W. J. & RICHARDSON, A. J. 2011. The pace of shifting climate in marine and terrestrial ecosystems. *Science,* 334**,** 652-5.

CAFARO, V., ANGELETTI, D., BELLISARIO, B., MACALI, A., CARERE, C. & ALESSI, J. 2015. Habitat overlap between bottlenose dolphins and seabirds: a pilot study to identify high-presence coastal areas in the Tyrrhenian Sea. *Journal of the Marine Biological Association of the United Kingdom,* 96**,** 891-901.

CALABRESE, J. M., CERTAIN, G., KRAAN, C. & DORMANN, C. F. 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography,* 23**,** 99-112.

CAÑADAS, A., SAGARMINAGA, R. & GARCIA-TISCAR, S. 2002. Cetacean distribution related with depth and slope in the Mediterranean waters off southern Spain. *Deep Sea Research Part I: Oceanographic Research Papers,* 49**,** 2053-73.

CHAUDHARY, C., RICHARDSON, A. J., SCHOEMAN, D. S. & COSTELLO, M. J. 2021. Global warming is causing a more pronounced dip in marine species richness around the equator. *Proc Natl Acad Sci U S A,* 118.

CHEN, B., HONG, Z., HAO, X. & GAO, H. 2020. Environmental models for predicting habitat of the Indo-Pacific humpback dolphins in Fujian, China. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 30**,** 787-793.

CHIARADIA, A., ROPERT-COUDERT, Y., KATO, A., MATTERN, T. & YORKE, J. 2007. Diving behaviour of Little Penguins from four colonies across their whole distribution range: bathymetry affecting diving effort and fledging success. *Marine Biology,* 151**,** 1535-1542.

CMEMS 2021. Global Ocean Chlorophyll, PP and PFT (Copernicus-GlobColour) from Satellite Observations: Monthly and Daily Interpolated (Reprocessed from 1997). *In:* CMEMS (ed.). Online: Copernicus Marine Service.

COMMINS, M. L., ANSORGE, I. & RYAN, P. G. 2013. Multi-scale factors influencing seabird assemblages in the African sector of the Southern Ocean. *Antarctic Science,* 26**,** 38-48.

CONVENTION ON BIOLOGICAL DIVERSITY. 2021. *COP15: Fifteenth meeting of the Conference of the Parties to the Convention on Biological Diversity* [Online]. Online. Available: https://www.cbd.int/cop/#:~:text=The%20Fifteenth%20meeting%20of%20the,April%20to%208%20May%202022. [Accessed 9th February 2022].

COPERNICUS MARINE SERVICE. 2023. *Copernicus Marine Service: Ocean Products* [Online]. Available: https://marine.copernicus.eu/ [Accessed 15th February 2023].

COSTA, P. L., BUGONI, L., KINAS, P. G. & MADUREIRA, L. 2020. Seabirds, environmental features and the Argentine anchovy Engraulis anchoita in the southwestern Atlantic Ocean. *Marine Ecology Progress Series,* 651**,** 199-213.

COX, S. L. 2015. *The Role of Physical Oceanography on the Distributions and Foraging Behaviours of Marine Mammals and Seabirds in Shelf-Seas.* Doctor of Philosophy PhD, Plymouth.

COX, S. L., EMBLING, C. B., HOSEGOOD, P. J., VOTIER, S. C. & INGRAM, S. N. 2018. Oceanographic drivers of marine mammal and seabird habitat-use across shelf-seas: A guide to key features and recommendations for future research and conservation management. *Estuarine, Coastal and Shelf Science,* 212**,** 294-310.

CROLL, D. A. 1990. Physical and biological determinants of the abundance, distribution and diet of the common murre in Monterey Bay, California. *Studies in Avian Biology,* 14**,** 139–148.

CRONIN, M., GREGORY, S. & ROGAN, E. 2014. Moulting phenology of the harbour seal in south-west Ireland. *Journal of the Marine Biological Association of the United Kingdom,* 94**,** 1079-1086.

CROXALL, J. P., BUTCHART, S. H. M., LASCELLES, B., STATTERSFIELD, A. J., SULLIVAN, B., SYMES, A. & TAYLOR, P. 2012. Seabird conservation status, threats and priority actions: a global assessment. *Bird Conservation International,* 22**,** 1–34.

CURY, P. M., BOYD, I. L., BONHOMMEAU, S., ANKER-NILSSEN, T., CRAWFORD, R. J., FURNESS, R. W., MILLS, J. A., MURPHY, E. J., OSTERBLOM, H., PALECZNY, M., PIATT, J. F., ROUX, J. P., SHANNON, L. & SYDEMAN, W. J. 2011. Global seabird response to forage fish depletion--one-third for the birds. *Science,* 334**,** 1703-6.

D'AMEN, M., DUBUIS, A., FERNANDES, R. F., POTTIER, J., PELLISSIER, L. & GUISAN, A. 2015. Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *Journal of Biogeography,* 42**,** 1255-1266.

DE DOMINICIS, M., WOLF, J. & O'HARA MURRAY, R. 2018. Comparative Effects of Climate Change and Tidal Stream Energy Extraction in a Shelf Sea. *Journal of Geophysical Research: Oceans,* 123**,** 5041-5067.

DISTLER, T., SCHUETZ, J. G., VELÁSQUEZ-TIBATÁ, J., LANGHAM, G. M. & LADLE, R. 2015. Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *Journal of Biogeography,* 42**,** 976-988.

DONELSON, J. M., SUNDAY, J. M., FIGUEIRA, W. F., GAITAN-ESPITIA, J. D., HOBDAY, A. J., JOHNSON, C. R., LEIS, J. M., LING, S. D., MARSHALL, D., PANDOLFI, J. M., PECL, G., RODGERS, G. G., BOOTH, D. J. & MUNDAY, P. L. 2019. Understanding interactions between plasticity, adaptation and range shifts in response to marine environmental change. *Philos Trans R Soc Lond B Biol Sci,* 374**,** 20180186.

DOUGHTY, C. E., ROMAN, J., FAURBY, S., WOLF, A., HAQUE, A., BAKKER, E. S., MALHI, Y., DUNNING, J. B., JR. & SVENNING, J. C. 2016. Global nutrient transport in a world of giants. *Proc Natl Acad Sci U S A,* 113**,** 868-73.

DU, Y., ZHANG, Y. & SHI, J. 2019. Relationship between sea surface salinity and ocean circulation and climate change. *Science China Earth Sciences,* 62**,** 771-782.

DWYER, R. G., CAMPBELL, H. A., CRAMP, R. L., BURKE, C. L., MICHELI-CAMPBELL, M. A., PILLANS, R. D., LYON, B. J., FRANKLIN, C. E. & KILLEN, S. 2020. Niche partitioning between river shark species is driven by seasonal fluctuations in environmental salinity. *Functional Ecology,* 34**,** 2170-2185.

ELITH, J. & LEATHWICK, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics,* 40**,** 677-697.

ELKINS, D., SWEAT, S. C., KUHAJDA, B. R., GEORGE, A. L., HILL, K. S. & WENGER, S. J. 2019. Illuminating hotspots of imperiled aquatic biodiversity in the southeastern US. *Global Ecology and Conservation,* 19.

EMBLING, C. B., ILLIAN, J., ARMSTRONG, E., VAN DER KOOIJ, J., SHARPLES, J., CAMPHUYSEN, K. C. J. & SCOTT, B. E. 2012. Investigating fine-scale spatio-temporal predator-prey patterns in dynamic marine ecosystems: a functional data analysis approach. *Journal of Applied Ecology,* 49**,** 481-492.

ESRI 2022. ArcGIS Pro. Redlands, CA: Environmental Systems Research Institute: ESRI.

ESSL, F., LENZNER, B., BACHER, S., BAILEY, S., CAPINHA, C., DAEHLER, C., DULLINGER, S., GENOVESI, P., HUI, C., HULME, P. E., JESCHKE, J. M., KATSANEVAKIS, S., KUHN, I., LEUNG, B., LIEBHOLD, A., LIU, C., MACISAAC, H. J., MEYERSON, L. A., NUNEZ, M. A., PAUCHARD, A., PYSEK, P., RABITSCH, W., RICHARDSON, D. M., ROY, H. E., RUIZ, G. M., RUSSELL, J. C., SANDERS, N. J., SAX, D. F., SCALERA, R., SEEBENS, H., SPRINGBORN, M., TURBELIN, A., VAN KLEUNEN, M., VON HOLLE, B., WINTER, M., ZENNI, R. D., MATTSSON, B. J. & ROURA-PASCUAL, N. 2020. Drivers of future alien species impacts: An expert-based assessment. *Glob Chang Biol,* 26**,** 4880-4893.

ESTES, J. A., TERBORGH, J., BRASHARES, J. S., POWER, M. E., BERGER, J., BOND, W. J., CARPENTER, S. R., ESSINGTON, T. E., HOLT, R. D., JACKSON, J. B., MARQUIS, R. J., OKSANEN, L., OKSANEN, T., PAINE, R. T., PIKITCH, E. K., RIPPLE, W. J., SANDIN, S. A., SCHEFFER, M., SCHOENER, T. W., SHURIN, J. B., SINCLAIR, A. R., SOULE, M. E., VIRTANEN, R. & WARDLE, D. A. 2011. Trophic downgrading of planet Earth. *Science,* 333**,** 301-6.

EUROPEAN COMMISSION 2020. EU Biodiversity Strategy for 2030. *In:* EUROPEAN COMISSION (ed.). Brussels.

EVANS, P. G. & BJØRGE, A. 2013. Impacts of climate change on marine mammals. *MCCIP Science Review***,** 134-148.

EVANS, P. G. H. & HAMMOND, P. S. 2004. Monitoring cetaceans in European waters. *Mammal Review,* 34**,** 131-156.

EVANS, R., LEA, M. A. & HINDELL, M. A. 2021. Predicting the distribution of foraging seabirds during a period of heightened environmental variability. *Ecol Appl,* 31**,** e02343.

FERRIER, S. & GUISAN, A. 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology,* 43**,** 393-404.

FINUCCI, B., DUFFY, C. A. J., BROUGH, T., FRANCIS, M. P., MILARDI, M., PINKERTON, M. H., PETERSEN, G. & STEPHENSON, F. 2021. Drivers of Spatial Distributions of Basking Shark (*Cetorhinus maximus*) in the Southwest Pacific. *Frontiers in Marine Science,* 8.

FREY, S., FISHER, J. T., BURTON, A. C., VOLPE, J. P. & ROWCLIFFE, M. 2017. Investigating animal activity patterns and temporal niche partitioning using camera-trap data: challenges and opportunities. *Remote Sensing in Ecology and Conservation,* 3**,** 123-132.

FROESE, R. & PAULY, D. 2000. FishBase 2000: concepts, design and data sources. Los Baños, Laguna, Philippines: ICLARM.

FULLARD, K. J., EARLY, G., HEIDE-JORGENSEN, M. P., BLOCH, D., ROSING-ASVID, A. & AMOS, W. 2000. Population structure of long-finned pilot whales in the North Atlantic: a correlation with sea surface temperature? *Mol Ecol,* 9**,** 949-58.

GAGNE, T. O., REYGONDEAU, G., JENKINS, C. N., SEXTON, J. O., BOGRAD, S. J., HAZEN, E. L. & VAN HOUTAN, K. S. 2020. Towards a global understanding of the drivers of marine and terrestrial biodiversity. *PLoS One,* 15**,** e0228065.

GARCÍA-BARÓN, I., SANTOS, M. B., SAAVEDRA, C., ASTARLOA, A., VALEIRAS, J., GARCÍA BARCELONA, S. & LOUZAO, M. 2020. Essential ocean variables and high value biodiversity areas: Targets for the conservation of marine megafauna. *Ecological Indicators,* 117.

GEBCO COMPILATION GROUP 2021. GEBCO 2021 Grid: a continuous terrain model of the global oceans and land. . *In:* BRITISH OCEANOGRAPHIC DATA CENTRE (ed.). National Oceanography Centre, NERC, UK.

GEIJZENDORFFER, I. R., REGAN, E. C., PEREIRA, H. M., BROTONS, L., BRUMMITT, N., GAVISH, Y., HAASE, P., MARTIN, C. S., MIHOUB, J.-B., SECADES, C., SCHMELLER, D. S., STOLL, S., WETZEL, F. T., WALTERS, M. & CADOTTE, M. 2016. Bridging the gap between biodiversity data and policy reporting needs: An Essential Biodiversity Variables perspective. *Journal of Applied Ecology,* 53**,** 1341-1350.

GILLES, A., HAMMOND, P., TAYLOR, N., GEELHOED, S. C. V., SVEEGAARD, S., CARLSTRÖM, J., AUTHIER, M., SVEEDRA, C. & ARAÚJO, M. B. 2023. *SCANS-IV survey*

[Online]. Available: https://www.tiho-hannover.de/en/clinics-institutes/institutes/institute-for-terrestrial-and-aquatic-wildlife-research-itaw/scans-iv-survey [Accessed 15th February 2023].

GISSI, E., FRASCHETTI, S. & MICHELI, F. 2019. Incorporating change in marine spatial planning: A review. *Environmental Science & Policy,* 92**,** 191-200.

GOTELLI, N. J. & COLWELL, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters,* 4**,** 379-391.

GRAY, J. S. 2001. Marine diversity: the paradigms in patterns of species richness examined. *Scientia Marina,* 65**,** 41-56.

GRECIAN, W. J., WITT, M. J., ATTRILL, M. J., BEARHOP, S., BECKER, P. H., EGEVANG, C., FURNESS, R. W., GODLEY, B. J., GONZALEZ-SOLIS, J., GREMILLET, D., KOPP, M., LESCROEL, A., MATTHIOPOULOS, J., PATRICK, S. C., PETER, H. U., PHILLIPS, R. A., STENHOUSE, I. J. & VOTIER, S. C. 2016. Seabird diversity hotspot linked to ocean productivity in the Canary Current Large Marine Ecosystem. *Biol Lett,* 12.

GRÉMILLET, D., LEWIS, S., DRAPEAU, L., VAN DER LINGEN, C. D., HUGGETT, J. A., COETZEE, J. C., VERHEYE, H. M., DAUNT, F., WANLESS, S. & RYAN, P. G. 2008. Spatial match-mismatch in the Benguela upwelling zone: should we expect chlorophyll and sea-surface temperature to predict marine predator distributions? *Journal of Applied Ecology,* 45**,** 610-621.

GUERRA, A. 1992. Mollusca, cephalopoda. *Fauna iberica,* 1.

GUINET, C., DUBROCA, L., LEA, M., GOLDSWORTHY, S., CHEREL, Y., DUHAMEL, G., BONADONNA, F. & DONNAY, J. 2001. Spatial distribution of foraging in female Antarctic fur seals *Arctocephalus gazella* in relation to oceanographic variables: a scale-dependent approach using geographic information systems. *Marine Ecology Progress Series,* 219**,** 251-264.

GUISAN, A. & RAHBEK, C. 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography,* 38**,** 1433-1444.

GUISAN, A. & THUILLER, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecol Lett,* 8**,** 993-1009.

GUISAN, A. & ZIMMERMANN, N. E. 2000. Predictive habitat distrubtion models in ecology. *Ecological Modelling,* 135**,** 147-186.

GULKA, J., RONCONI, R. A. & DAVOREN, G. K. 2019. Spatial segregation contrasting dietary overlap: niche partitioning of two sympatric alcids during shifting resource availability. *Marine Biology,* 166.

GUSMÃO, R. A. F., HERNANDES, F. A., VANCINE, M. H., NAKA, L. N., DOÑA, J., GONÇALVES-SOUZA, T. & CUNNINGHAM, S. 2020. Host diversity outperforms climate as a global driver of symbiont diversity in the bird-feather mite system. *Diversity and Distributions,* 27**,** 416-426.

HALPERN, B. S., FRAZIER, M., AFFLERBACH, J., LOWNDES, J. S., MICHELI, F., O'HARA, C., SCARBOROUGH, C. & SELKOE, K. A. 2019. Recent pace of change in human impact on the world's ocean. *Sci Rep,* 9**,** 11609.

HALPERN, B. S., FRAZIER, M., POTAPENKO, J., CASEY, K. S., KOENIG, K., LONGO, C., LOWNDES, J. S., ROCKWOOD, R. C., SELIG, E. R., SELKOE, K. A. & WALBRIDGE, S. 2015. Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nat Commun,* 6**,** 7615.

HAMILTON, C. D., LYDERSEN, C., AARS, J., ACQUARONE, M., ATWOOD, T., BAYLIS, A., BIUW, M., BOLTUNOV, A. N., BORN, E. W., BOVENG, P., BROWN, T. M., CAMERON, M., CITTA, J., CRAWFORD, J., DIETZ, R., ELIAS, J., FERGUSON, S. H., FISK, A., FOLKOW, L. P., FROST, K. J., GLAZOV, D. M., GRANQUIST, S. M., GRYBA, R., HARWOOD, L., HAUG, T., HEIDE-JØRGENSEN, M. P., HUSSEY, N. E., KALINEK, J., LAIDRE, K. L., LITOVKA, D. I., LONDON, J. M., LOSETO, L. L., MACPHEE, S., MARCOUX, M., MATTHEWS, C. J. D., NILSSEN, K., NORDØY, E. S., O’CORRY-CROWE, G., ØIEN, N., OLSEN, M. T., QUAKENBUSH, L., ROSING-ASVID, A., SEMENOVA, V., SHELDEN, K. E. W., SHPAK, O. V., STENSON, G., STORRIE, L., SVEEGAARD, S., TEILMANN, J., UGARTE, F., VON DUYKE, A. L., WATT, C., WIIG, Ø., WILSON, R. R., YURKOWSKI, D. J., KOVACS, K. M. & GUO, B. 2022. Marine mammal hotspots across the circumpolar Arctic. *Diversity and Distributions,* 28**,** 2729-2753.

HAMMOND, P., LACEY, C., GILLES, A., VIQUERAT, S., BÖRJESSON, P., HERR, H., MACLEOD, K., RIDOUX, V., SANTOS, M., SCHEIDAT, M., TEILMANN, J., VINGADA, J., ØIEN, N. 2017. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys.: Wageningen Marine Research.

HAMMOND, P. S., LACEY, C., GILLES, A., VIQUERAT, S., BÖRJESSON, P., HERR, H., MACLEOD, K., RIDOUX, V., SANTOS, M., SCHEIDAT, M. AND TEILMANN, J. 2021. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. *In:* SMRU (ed.). St Andrews, Scotland, UK: Sea Mammal Research Unit (SMRU).

HANSEN, J., SATO, M., RUEDY, R., LO, K., LEA, D. W. & MEDINA-ELIZADE, M. 2006. Global temperature change. *Proc Natl Acad Sci U S A,* 103**,** 14288-93.

HARLEY, C. D. 2011. Climate change, keystone predation, and biodiversity loss. *Science,* 334**,** 1124-7.

HARRIS, D. J., TAYLOR, S. D. & WHITE, E. P. 2018. Forecasting biodiversity in breeding birds using best practices. *PeerJ,* 6**,** e4278.

HAZEN, E. L., ABRAHMS, B., BRODIE, S., CARROLL, G., JACOX, M. G., SAVOCA, M. S., SCALES, K. L., SYDEMAN, W. J. & BOGRAD, S. J. 2019. Marine top predators as climate and ecosystem sentinels. *Frontiers in Ecology and the Environment,* 17**,** 565-574.

HAZEN, E. L., JORGENSEN, S., RYKACZEWSKI, R. R., BOGRAD, S. J., FOLEY, D. G., JONSEN, I. D., SHAFFER, S. A., DUNNE, J. P., COSTA, D. P., CROWDER, L. B. & BLOCK, B. A. 2012. Predicted habitat shifts of Pacific top predators in a changing climate. *Nature Climate Change,* 3**,** 234-238.

HELAOUET, P., BEAUGRAND, G. & EDWARDS, M. 2013. Understanding long-term changes in species abundance using a niche-based approach. *PLoS One,* 8**,** e79186.

HIGHAM, J. E. & LÜCK, M. 2007. *Marine wildlife and tourism management: Insights from the natural and social sciences,* Wallingford, UK, CABI Publishing.

HIJMANS, R. J. 2020. raster: Geographic Data Analysis and Modeling. *R package version 3.4-5.*

HILL, T. S. & HOOGENBOOM, M. O. 2022. The indirect effects of ocean acidification on corals and coral communities. *Coral Reefs,* 41**,** 1557-1583.

HILLEBRAND, H., BLASIUS, B., BORER, E. T., CHASE, J. M., DOWNING, J. A., ERIKSSON, B. K., FILSTRUP, C. T., HARPOLE, W. S., HODAPP, D., LARSEN, S., LEWANDOWSKA, A. M., SEABLOOM, E. W., VAN DE WAAL, D. B., RYABOV, A. B. & CADOTTE, M. 2018. Biodiversity change is uncoupled from species richness trends: Consequences for conservation and monitoring. *Journal of Applied Ecology,* 55**,** 169-184.

HINDELL, M. A., REISINGER, R. R., ROPERT-COUDERT, Y., HUCKSTADT, L. A., TRATHAN, P. N., BORNEMANN, H., CHARRASSIN, J. B., CHOWN, S. L., COSTA, D. P., DANIS, B., LEA, M. A., THOMPSON, D., TORRES, L. G., VAN DE PUTTE, A. P., ALDERMAN, R., ANDREWS-GOFF, V., ARTHUR, B., BALLARD, G., BENGTSON, J., BESTER, M. N., BLIX, A. S., BOEHME, L., BOST, C. A., BOVENG, P., CLEELAND, J., CONSTANTINE, R., CORNEY, S., CRAWFORD, R. J. M., DALLA ROSA, L., DE BRUYN, P. J. N., DELORD, K., DESCAMPS, S., DOUBLE, M., EMMERSON, L., FEDAK, M., FRIEDLAENDER, A., GALES, N., GOEBEL, M. E., GOETZ, K. T., GUINET, C., GOLDSWORTHY, S. D., HARCOURT, R., HINKE, J. T., JEROSCH, K., KATO, A., KERRY, K. R., KIRKWOOD, R., KOOYMAN, G. L., KOVACS, K. M., LAWTON, K., LOWTHER, A. D., LYDERSEN, C., LYVER, P. O., MAKHADO, A. B., MARQUEZ, M. E. I., MCDONALD, B. I., MCMAHON, C. R., MUELBERT, M., NACHTSHEIM, D., NICHOLLS, K. W., NORDOY, E. S., OLMASTRONI, S., PHILLIPS, R. A., PISTORIUS, P., PLOTZ, J., PUTZ, K., RATCLIFFE, N., RYAN, P. G., SANTOS, M., SOUTHWELL, C., STANILAND, I., TAKAHASHI, A., TARROUX, A., TRIVELPIECE, W., WAKEFIELD, E., WEIMERSKIRCH, H., WIENECKE, B., XAVIER, J. C., WOTHERSPOON, S., JONSEN, I. D. & RAYMOND, B. 2020. Tracking of marine predators to protect Southern Ocean ecosystems. *Nature,* 580**,** 87-92.

HOBSON, K. A. 1993. Trophic relationships among high Arctic seabirds: insights from tissue-dependent stable-isotope models. *Marine Ecology Progress Series,* 95**,** 7-18.

HOEGH-GULDBERG, O. & BRUNO, J. F. 2010. The impact of climate change on the world's marine ecosystems. *Science,* 328**,** 1523-8.

HUELSMANN, M., HECKER, N., SPRINGER, M. S., GATESY, J., SHARMA, V. & HILLER, M. 2019. Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. *Science advances,* 5**,** 1-12.

HUMPHRIES, N. E., SIMPSON, S. J., WEARMOUTH, V. J. & SIMS, D. W. 2016. Two’s company, three’s a crowd: fine-scale habitat partitioning by depth among sympatric species of marine mesopredator. *Marine Ecology Progress Series,* 561**,** 173-187.

HUNT, G. L. & MCKINNELL, S. 2006. Interplay between top-down, bottom- up, and wasp-waist control in control in marine ecosystems. *Progress in Oceanography,* 68**,** 115–124.

HYRENBACH, K. D., VEIT, R. R., WEIMERSKIRCH, H., METZL, N. & HUNT JR, G. L. 2007. Community structure across a large-scale ocean productivity gradient: Marine bird assemblages of the Southern Indian Ocean. *Deep Sea Research Part I: Oceanographic Research Papers,* 54**,** 1129-1145.

IUCN. 2023. *The IUCN Red List of Threatened Species* [Online]. Available: https://www.iucnredlist.org [Accessed 1st February 2023].

ISOJUNNO, S., MATTHIOPOULOS, J. & EVANS, P. G. H. 2012. Harbour porpoise habitat preferences: robust spatio-temporal inferences from opportunistic data. *Marine Ecology Progress Series,* 448**,** 155-170.

JESSOPP, M., ARNEILL, G. E., NYKÄNEN, M., BENNISON, A. & ROGAN, E. 2020. Central place foraging drives niche partitioning in seabirds. *Oikos,* 129**,** 1704-1713.

JORGENSEN, S. J., ANDERSON, S., FERRETTI, F., TIETZ, J. R., CHAPPLE, T., KANIVE, P., BRADLEY, R. W., MOXLEY, J. H. & BLOCK, B. A. 2019. Killer whales redistribute white shark foraging pressure on seals. *Scientific Reports,* 9.

JUNGBLUT, S., LIEBICH, V. & BODE-DALBY, M. YOUMARES 9 - The Oceans: Our Research, Our Future. *In:* JUNGBLUT, S., LIEBICH, V. & BODE-DALBY, M., eds. Young Marine Researcher (YOUMARES), 2018 Oldenburg, Germany. Springer Open.

KANE, A., PIROTTA, E., WISCHNEWSKI, S., CRITCHLEY, E. J., BENNISON, A., JESSOPP, M. & QUINN, J. L. 2020. Spatio-temporal patterns of foraging behaviour in a wide-ranging seabird reveal the role of primary productivity in locating prey. *Marine Ecology Progress Series,* 646**,** 175-188.

KARPOUZI, V. S. 2005. *Modelling and mapping trophic overlap between fisheries and the world's seabirds.* Zoology, University of British Columbia.

KASCHNER, K., TITTENSOR, D. P., READY, J., GERRODETTE, T. & WORM, B. 2011. Current and future patterns of global marine mammal biodiversity. *PLoS One,* 6**,** e19653.

KRAMER-SCHADT, S., NIEDBALLA, J., PILGRIM, J. D., SCHRÖDER, B., LINDENBORN, J., REINFELDER, V., STILLFRIED, M., HECKMANN, I., SCHARF, A. K., AUGERI, D. M., CHEYNE, S. M., HEARN, A. J., ROSS, J., MACDONALD, D. W., MATHAI, J., EATON, J., MARSHALL, A. J., SEMIADI, G., RUSTAM, R., BERNARD, H., ALFRED, R., SAMEJIMA, H., DUCKWORTH, J. W., BREITENMOSER-WUERSTEN, C., BELANT, J. L., HOFER, H., WILTING, A. & ROBERTSON, M. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions,* 19**,** 1366-1379.

KRONFELD-SCHOR, N. & DAYAN, T. 2003. Partitioning of Time as an Ecological Resource. *Annual Review of Ecology, Evolution, and Systematics,* 34**,** 153-181.

KRÜGER, L., RAMOS, J. A., XAVIER, J. C., GRÉMILLET, D., GONZÁLEZ-SOLÍS, J., PETRY, M. V., PHILLIPS, R. A., WANLESS, R. M. & PAIVA, V. H. 2018. Projected distributions of Southern Ocean albatrosses, petrels and fisheries as a consequence of climatic change. *Ecography,* 41**,** 195-208.

KULETZ, K. J., FERGUSON, M. C., HURLEY, B., GALL, A. E., LABUNSKI, E. A. & MORGAN, T. C. 2015. Seasonal spatial patterns in seabird and marine mammal distribution in the eastern Chukchi and western Beaufort seas: Identifying biologically important pelagic areas. *Progress in Oceanography,* 136**,** 175-200.

LARAN, S., AUTHIER, M., VAN CANNEYT, O., DORÉMUS, G., WATREMEZ, P. & RIDOUX, V. 2017. A Comprehensive Survey of Pelagic Megafauna: Their Distribution, Densities, and Taxonomic Richness in the Tropical Southwest Indian Ocean. *Frontiers in Marine Science,* 4.

LAZZARI, N., MARTÍN-LÓPEZ, B., SANABRIA-FERNANDEZ, J. A. & BECERRO, M. A. 2020. Alpha and beta diversity across coastal marine social-ecological systems: Implications for conservation. *Ecological Indicators,* 109.

LEA, D. J., MIROUZE, I., MARTIN, M. J., KING, R. R., HINES, A., WALTERS, D. AND THURLOW, M. 2015. Assessing a new data assimilation system based on the Met Office coupled atmosphere-land-ocean-sea ice model. *Monthly Weather Review,* 143**,** 4678-4694.

LEAR, K. O., WHITNEY, N. M., MORRIS, J. J. & GLEISS, A. C. 2021. Temporal niche partitioning as a novel mechanism promoting co-existence of sympatric predators in marine systems. *Proc Biol Sci,* 288**,** 20210816.

LEE-YAW, J. A., L. MCCUNE, J., PIRONON, S. & SHETH, N. S. 2021. Species distribution models rarely predict the biology of real populations. *Ecography,* 2022.

LEJEUSNE, C., CHEVALDONNE, P., PERGENT-MARTINI, C., BOUDOURESQUE, C. F. & PEREZ, T. 2010. Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea. *Trends Ecol Evol,* 25, 250-60.

LEMONS, G. E., LEWISON, R. L., SEMINOFF, J. A., COPPENRATH, C. M. & POPP, B. N. 2020. Nitrogen isotope fractionation of amino acids from a controlled study on the green turtle (*Chelonia mydas*): expanding beyond Glx/Phe for trophic position. *Marine Biology,* 167.

LOTZE, H. K., LENIHAN, H. S., BOURQUE, B. J., BRADBURY, R. H., COOKE, R. G., KAY, M. C., KIDWELL, S. M., KIRBY, M. X., PETERSON, C. H. & JACKSON, J. B. 2006. Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science,* 312, 1806-9.

LOUZAO, M., HYRENBACH, K. D., ARCOS, J. M., ABELLO, P., GIL DE SOLA, L. & ORO, D. 2006. Oceanographic habitat of an endangered Mediterranean procellariiform: implications for marine protected areas. *Ecol Appl,* 16, 1683-95.

LUCZAK, C., BEAUGRAND, G., JAFFRE, M. & LENOIR, S. 2011. Climate change impact on Balearic shearwater through a trophic cascade. *Biol Lett,* 7, 702-5.

LYONS, D. A., LOWEN, J. B., THERRIAULT, T. W., BRICKMAN, D., GUO, L., MOORE, A. M., PEÑA, M. A., WANG, Z. & DIBACCO, C. 2020. Identifying marine invasion hotspots using stacked species distribution models. *Biological Invasions,* 22, 3403-3423.

MACLEOD, C. D. 2009. Global climate change, range changes and potential implications for the conservation of marine cetaceans: a review and synthesis. *Endangered Species Research,* 7, 125-136.

MARTIN, C. S., TOLLEY, M. J., FARMER, E., MCOWEN, C. J., GEFFERT, J. L., SCHARLEMANN, J. P. W., THOMAS, H. L., VAN BOCHOVE, J. H., STANWELL-SMITH, D., HUTTON, J. M., LASCELLES, B., PILGRIM, J. D., EKSTROM, J. M. M. & TITTENSOR, D. P. 2015. A global map to aid the identification and screening of critical habitat for marine industries. *Marine Policy,* 53, 45-53.

MASELLO, J. F., MUNDRY, R., POISBLEAU, M., DEMONGIN, L., VOIGT, C. C., WIKELSKI, M. & QUILLFELDT, P. 2010. Diving seabirds share foraging space and time within and among species. *Ecosphere,* 1.

MATEAR, L., ROBBINS, J. R., HALE, M. & POTTS, J. 2019. Cetacean biodiversity in the Bay of Biscay: Suggestions for environmental protection derived from citizen science data. *Marine Policy,* 109, 103672.

MAURITZEN, C., MELSOM, A. & SUTTON, R. T. 2012. Importance of density-compensated temperature change for deep North Atlantic Ocean heat uptake. *Nature Geoscience,* 5, 905-910.

MCCAULEY, D. J., HOFFMANN, E., YOUNG, H. S. & MICHELI, F. 2012. Night shift: expansion of temporal niche use following reductions in predator density. *PLoS One,* 7, e38871.

MCCLELLAN, C. M., BRERETON, T., DELL'AMICO, F., JOHNS, D. G., CUCKNELL, A. C., PATRICK, S. C., PENROSE, R., RIDOUX, V., SOLANDT, J. L., STEPHAN, E., VOTIER, S. C., WILLIAMS, R. & GODLEY, B. J. 2014. Understanding the distribution of marine megafauna in the English channel region: identifying key habitats for conservation within the busiest seaway on earth. *PLoS One,* 9, e89720.

MENDONÇA, C. D. O., APTROOT, A., LÜCKING, R. & CÁCERES, M. E. D. S. 2020. Global species richness prediction for Pyrenulaceae (Ascomycota: Pyrenulales), the last of the “big three” most speciose tropical microlichen families. *Biodiversity and Conservation,* 29, 1059-1079.

MERCHANT, C. J., EMBURY, O., BULGIN, C. E., BLOCK, T., CORLETT, G. K., FIEDLER, E., GOOD, S. A., MITTAZ, J., RAYNER, N. A., BERRY, D., EASTWOOD, S., TAYLOR, M., TSUSHIMA, Y., WATERFALL, A., WILSON, R. & DONLON, C. 2019. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Sci Data,* 6, 223.

MEYER, C., KREFT, H., GURALNICK, R. & JETZ, W. 2015. Global priorities for an effective information basis of biodiversity distributions. *Nat Commun,* 6, 8221.

MEYER, X., MACINTOSH, A. J. J., CHIARADIA, A., KATO, A., MATTERN, T., SUEUR, C. & ROPERT-COUDERT, Y. 2017. Shallow divers, deep waters and the rise of behavioural stochasticity. *Marine Biology,* 164.

MEYERS, E. K. M., TUYA, F., BARKER, J., JIMÉNEZ ALVARADO, D., CASTRO-HERNÁNDEZ, J. J., HAROUN, R. & RÖDDER, D. 2017. Population structure, distribution and habitat use of the Critically Endangered Angelshark, *Squatina squatina*, in the Canary Islands. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 27, 1133-1144.

MILLER, P. 2009. Composite front maps for improved visibility of dynamic sea-surface features on cloudy SeaWiFS and AVHRR data. *Journal of Marine Systems,* 78, 327-336.

MILLER, P. I. & CHRISTODOULOU, S. 2014. Frequent locations of oceanic fronts as an indicator of pelagic diversity: Application to marine protected areas and renewables. *Marine Policy,* 45**,** 318-329.

MILLER, P. I., SCALES, K. L., INGRAM, S. N., SOUTHALL, E. J., SIMS, D. W. & COSTA, D. 2015. Basking sharks and oceanographic fronts: quantifying associations in the north-east Atlantic. *Functional Ecology,* 29**,** 1099-1109.

MONTERROSO, P., ALVES, P. C. & FERRERAS, P. 2014. Plasticity in circadian activity patterns of mesocarnivores in Southwestern Europe: implications for species coexistence. *Behavioral Ecology and Sociobiology,* 68**,** 1403-1417.

MONTEVECCHI, W. A., BENVENUTI, S., GARTHE, S., DAVOREN, G. K. & FIFIELD, D. 2009. Flexible foraging tactics by a large opportunistic seabird preying on forage- and large pelagic fishes. *Marine Ecology Progress Series,* 385**,** 295-306.

NAGELKERKEN, I. & CONNELL, S. D. 2022. Ocean acidification drives global reshuffling of ecological communities. *Glob Chang Biol,* 28**,** 7038-7048.

NAIMI, B., HAMM, N., GROEN, T., SKIDMORE, A. & AG, T. 2014. Where is positional uncertainty a problem for species distribution modelling. *Ecography,* 37**,** 191-203.

NAVARRO, J., CARDADOR, L., FERNÁNDEZ, Á. M., BELLIDO, J. M. & COLL, M. 2016. Differences in the relative roles of environment, prey availability and human activity in the spatial distribution of two marine mesopredators living in highly exploited ecosystems. *Journal of Biogeography,* 43**,** 440-450.

NEODAAS. 2022. *NEODAAS Access Data* [Online]. Plymouth Marine Laboratory: National Centre for Earth Observation (NCEO). Available: https://www.neodaas.ac.uk/Access_Data [Accessed 02/06/2021 2021].

NICOL, S., PAULY, T., BINDOFF, N. L., WRIGHT, S., THIELE, D., HOSIE, G. W., STRUTTON, P. G. & WOEHLER, E. 2000. Ocean circulation off east Antarctica affects ecosystem structure and sea-ice extent. *Nature,* 406**,** 504-507.

NICOLAS, D., LOBRY, J., LEPAGE, M., SAUTOUR, B., LE PAPE, O., CABRAL, H., URIARTE, A. & BOËT, P. 2010. Fish under influence: A macroecological analysis of relations between fish species richness and environmental gradients among European tidal estuaries. *Estuarine, Coastal and Shelf Science,* 86**,** 137-147.

NORD, M. & FORSLUND, P. 2015. Environmental gradients explain species richness and community composition of coastal breeding birds in the Baltic Sea. *PLoS One,* 10**,** e0118455.

NOTARBARTOLO DI SCIARA, G. 2002. Cetaceans of the Mediterranean and Black Seas: state of knowledge and conservation strategies. *A report to the ACCOBAMS Secretariat.* Monaco: ACCOBAMS.

NOTARBARTOLO DI SCIARA, G. & BIRKUN, A. J. 2010. Conserving whales, dolphins and porpoises in the Mediterranean and Black Seas: an ACCOBAMS status report. *In:* ACCOBAMS (ed.). Monaco.

NUR, N., JAHNCKE, J., HERZOG, M. P., HOWAR, J., HYRENBACH, K. D., ZAMON, J. E., AINLEY, D. G., WIENS, J. A., MORGAN, K., BALLANCE, L. T. & STRALBERG, D. 2011. Where the wild things are: predicting hotspots of seabird aggregations in the California Current System. *Ecol Appl,* 21**,** 2241-57.

O'HARA, C. C., FRAZIER, M. & HALPERN, B. S. 2021. At-risk marine biodiversity faces extensive, expanding, and intensifying human impacts. *Science,* 372**,** 84-87.

O'HARA, P. D., MORGAN, K. H. & SYDEMAN, W. J. 2006. Primary producer and seabird associations with AVHRR-derived sea surface temperatures and gradients in the southeastern Gulf of Alaska. *Deep Sea Research Part II: Topical Studies in Oceanography,* 53**,** 359-369.

ORGERET, F., THIEBAULT, A., KOVACS, K. M., LYDERSEN, C., HINDELL, M. A., THOMPSON, S. A., SYDEMAN, W. J. & PISTORIUS, P. A. 2022. Climate change impacts on seabirds and marine mammals: The importance of study duration, thermal tolerance and generation time. *Ecol Lett,* 25**,** 218-239.

OVERLAND, J. E., ALHEIT, J., BAKUN, A., HURRELL, J. W., MACKAS, D. L. & MILLER, A. J. 2010. Climate controls on marine ecosystems and fish populations. *Journal of Marine Systems,* 79**,** 305-315.

PACE, D. S., TIZZI, R. & MUSSI, B. 2013. Cetaceans Value and Conservation in the Mediterranean Sea. *Journal of Biodiversity & Endangered Species,* 01.

PALACIOS, D. M., BAUMGARTNER, M. F., LAIDRE, K. L. & GREGR, E. J. 2013. Beyond correlation: integrating environmentally and behaviourally mediated processes in models of marine mammal distributions. *Endangered Species Research,* 22**,** 191-203.

PAPASTAMATIOU, Y. P., WETHERBEE, B. M., LOWE, C. G. & CROW, G. L. 2006. Distribution and diet of four species of carcharhinid shark in the Hawaiian Islands: evidence for resource partitioning and competitivie exclusion. *Marine Ecology Progress Series,* 239–251**,** 239–251.

PATENAUDE-MONETTE, M., BELISLE, M. & GIROUX, J. F. 2014. Balancing energy budget in a central-place forager: which habitat to select in a heterogeneous environment? *PLoS One,* 9**,** e102162.

PENN, J. L. & DEUTSCH, C. 2022. Avoiding ocean mass extinction from climate warming. *Science,* 376**,** 524-526.

PEREIRA, H. M., FERRIER, S., WALTERS, M., GELLER, G. N., JONGMAN, R. H., SCHOLES, R. J., BRUFORD, M. W., BRUMMITT, N., BUTCHART, S. H., CARDOSO, A. C., COOPS, N. C., DULLOO, E., FAITH, D. P., FREYHOF, J., GREGORY, R. D., HEIP, C., HOFT, R., HURTT, G., JETZ, W., KARP, D. S., MCGEOCH, M. A., OBURA, D., ONODA, Y., PETTORELLI, N., REYERS, B., SAYRE, R., SCHARLEMANN, J. P., STUART, S. N., TURAK, E., WALPOLE, M. & WEGMANN, M. 2013. Essential biodiversity variables. *Science,* 339**,** 277-8.

PHILLIPS, S. J., ANDERSON, R. P. & SCHAPIRE, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling,* 190**,** 231-259.

PINSKY, M. L., COMTE, L. & SAX, D. F. 2022. Unifying climate change biology across realms and taxa. *Trends Ecol Evol,* 37**,** 672-682.

POLOCZANSKA, E. S., BROWN, C. J., SYDEMAN, W. J., KIESSLING, W., SCHOEMAN, D. S., MOORE, P. J., BRANDER, K., BRUNO, J. F., BUCKLEY, L. B., BURROWS, M. T., DUARTE, C. M., HALPERN, B. S., HOLDING, J., KAPPEL, C. V., O'CONNOR, M. I., PANDOLFI, J. M., PARMESAN, C., SCHWING, F., THOMPSON, S. A. & RICHARDSON, A. J. 2013. Global imprint of climate change on marine life. *Nature Climate Change,* 3**,** 919-925.

POLOCZANSKA, E. S., BURROWS, M. T., BROWN, C. J., GARCÍA MOLINOS, J., HALPERN, B. S., HOEGH-GULDBERG, O., KAPPEL, C. V., MOORE, P. J., RICHARDSON, A. J., SCHOEMAN, D. S. & SYDEMAN, W. J. 2016. Responses of Marine Organisms to Climate Change across Oceans. *Frontiers in Marine Science,* 3.

R CORE TEAM 2022. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

REEVES, R. & NOTARBARTOLO DI SCIARA, G. 2006. The status and distribution of cetaceans in the Black Sea and Mediterranean Sea. *IUCN Centre for Mediterranean Cooperation,.*

RETANA, M. V. & LEWIS, M. N. 2017. Suitable habitat for marine mammals during austral summer in San Jorge Gulf, Argentina. *Revista de biología marina y oceanografía,* 52**,** 275-288.

RIEDL, R. 1983. Fauna y flora del Mar Mediterráneo: Una Guía Sistemática para Biólogos y Naturalistas: Ediciones Omega. *SA, Barcelona*.

RIPPLE, W. J., WIRSING, A. J., WILMERS, C. C. & LETNIC, M. 2013. Widespread mesopredator effects after wolf extirpation. *Biological Conservation,* 160**,** 70-79.

RITCHIE, E. G. & JOHNSON, C. N. 2009. Predator interactions, mesopredator release and biodiversity conservation. *Ecol Lett,* 12**,** 982-98.

ROBINSON, K. P., TETLEY, M. J. & MITCHELSON-JACOB, E. G. 2009. The distribution and habitat preference of coastally occurring minke whales (*Balaenoptera acutorostrata*) in the outer southern Moray Firth, northeast Scotland. *Journal of Coastal Conservation,* 13**,** 39-48.

ROSWELL, M., DUSHOFF, J. & WINFREE, R. 2021. A conceptual guide to measuring species diversity. *Oikos,* 130**,** 321-338.

SABARROS, P. S., GRÉMILLET, D., DEMARCQ, H., MOSELEY, C., PICHEGRU, L., MULLERS, R. H. E., STENSETH, N. C. & MACHU, E. 2014. Fine-scale recognition and use of mesoscale fronts by foraging Cape gannets in the Benguela upwelling region. *Deep Sea Research Part II: Topical Studies in Oceanography,* 107**,** 77-84.

SADYKOVA, D., SCOTT, B. E., DE DOMINICIS, M., WAKELIN, S. L., WOLF, J. & SADYKOV, A. 2020. Ecological costs of climate change on marine predator-prey population distributions by 2050. *Ecol Evol,* 10**,** 1069-1086.

SALGADO KENT, C., BOUCHET, P., WELLARD, R., PARNUM, I., FOUDA, L. & ERBE, C. 2021. Seasonal productivity drives aggregations of killer whales and other cetaceans over submarine canyons of the Bremer Sub-Basin, south-western Australia. *Australian Mammalogy,* 43.

SANTORA, J. A. & VEIT, R. R. 2013. Spatio-temporal persistence of top predator hotspots near the Antarctic Peninsula. *Marine Ecology Progress Series,* 487**,** 287-304.

SCALES, K. L., MILLER, P. I., EMBLING, C. B., INGRAM, S. N., PIROTTA, E. & VOTIER, S. C. 2014. Mesoscale fronts as foraging habitats: composite front mapping reveals oceanographic drivers of habitat use for a pelagic seabird. *J R Soc Interface,* 11**,** 20140679.

SCALES, K. L., MILLER, P. I., INGRAM, S. N., HAZEN, E. L., BOGRAD, S. J. & PHILLIPS, R. A. 2016. Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. *Diversity and Distributions,* 22**,** 212-224.

SCALES, K. L., MILLER, P. I., VARO-CRUZ, N., HODGSON, D. J., HAWKES, L. A. & GODLEY, B. J. 2015. Oceanic loggerhead turtles *Caretta caretta* associate with thermal fronts: evidence from the Canary Current Large Marine Ecosystem. *Marine Ecology Progress Series,* 519**,** 195-207.

SCHERER, L., GÜRDAL, İ. & VAN BODEGOM, P. M. 2022. Characterization factors for ocean acidification impacts on marine biodiversity. *Journal of Industrial Ecology,* 26**,** 2069-2079.

SCHMITT, S., POUTEAU, R., JUSTEAU, D., DE BOISSIEU, F. & BIRNBAUM, P. 2017. SSDM: an R package to predict distribution of species richness and endemism based on stacked species distribution models. *Methods in Ecology and Evolution,* 8**,** 1795-1803.

SCOTT, B. E., SHARPLES, J., ROSS, O. N., WANG, J., PIERCE, G. J. & CAMPHUYSEN, C. J. 2010. Sub-surface hotspots in shallow seas: fine-scale limited locations of top predator foraging habitat indicated by tidal mixing and sub-surface chlorophyll. *Marine Ecology Progress Series,* 408**,** 207-226.

SEPULVEDA, M., QUINONES, R. A., ESPARZA, C., CARRASCO, P. & WINCKLER, P. 2020. Vulnerability of a top marine predator to coastal storms: a relationship between hydrodynamic drivers and stranding rates of newborn pinnipeds. *Sci Rep,* 10**,** 12807.

SERRATOSA, J., HYRENBACH, K. D., MIRANDA-URBINA, D., PORTFLITT-TORO, M., LUNA, N. & LUNA-JORQUERA, G. 2020. Environmental Drivers of Seabird At-Sea Distribution in the Eastern South Pacific Ocean: Assemblage Composition Across a Longitudinal Productivity Gradient. *Frontiers in Marine Science,* 6.

SHEPARD, E. L., AHMED, M. Z., SOUTHALL, E. J., WITT, M. J., METCALFE, J. D., & SIMS, D. W. 2006. Diel and tidal rhythms in diving behaviour of pelagic sharks identified by signal processing of archival tagging data. *Marine Ecology Progress Series,* 328**,** 205-213.

SHIPLEY, O. N., GALLAGHER, A. J., SHIFFMAN, D. S., KAUFMAN, L. & HAMMERSCHLAG, N. 2019. Diverse resource-use strategies in a large-bodied marine predator guild: evidence from differential use of resource subsidies and intraspecific isotopic variation. *Marine Ecology Progress Series,* 623**,** 71-83.

SHUMWAY, N., WATSON, J. E. M., SAUNDERS, M. I. & MARON, M. 2018. The Risks and Opportunities of Translating Terrestrial Biodiversity Offsets to the Marine Realm. *BioScience,* 68**,** 125-133.

SIMS, D. W., WITT, M. J., RICHARDSON, A. J., SOUTHALL, E. J. & METCALFE, J. D. 2006. Encounter success of free-ranging marine predator movements across a dynamic prey landscape. *Proc Biol Sci,* 273**,** 1195-201.

SOLDATINI, C., ALBORES-BARAJAS, Y. V., MASSA, B. & GIMENEZ, O. 2016. Forecasting ocean warming impacts on seabird demography: a case study on the European storm petrel. *Marine Ecology Progress Series,* 552**,** 255-269.

STANDISH, R. J., DAWS, M. I., GOVE, A. D., DIDHAM, R. K., GRIGG, A. H., KOCH, J. M., HOBBS, R. J. & GIBSON, D. 2015. Long-term data suggest jarrah-forest establishment at restored mine sites is resistant to climate variability. *Journal of Ecology,* 103**,** 78-89.

STOKES, C. R., ABRAM, N. J., BENTLEY, M. J., EDWARDS, T. L., ENGLAND, M. H., FOPPERT, A., JAMIESON, S. S. R., JONES, R. S., KING, M. A., LENAERTS, J. T. M., MEDLEY, B., MILES, B. W. J., PAXMAN, G. J. G., RITZ, C., VAN DE FLIERDT, T. & WHITEHOUSE, P. L. 2022. Response of the East Antarctic Ice Sheet to past and future climate change. *Nature,* 608**,** 275-286.

SURYAN, R. M., SANTORA, J. A. & SYDEMAN, W. J. 2012. New approach for using remotely sensed chlorophyll a to identify seabird hotspots. *Marine Ecology Progress Series,* 451**,** 213-225.

SUTCLIFFE, P. R., MELLIN, C., PITCHER, C. R., POSSINGHAM, H. P. & CALEY, M. J. 2014. Regional-scale patterns and predictors of species richness and abundance across twelve major tropical inter-reef taxa. *Ecography,* 37**,** 162-171.

SUTTON, G., PICHEGRU, L., BOTHA, J. A., KOUZANI, A. Z., ADAMS, S., BOST, C. A. & ARNOULD, J. P. Y. 2020. Multi-predator assemblages, dive type, bathymetry and sex influence foraging success and efficiency in African penguins. *PeerJ,* 8**,** e9380.

SYDEMAN, W. J., SCHOEMAN, D. S., THOMPSON, S. A., HOOVER, B. A., GARCIA-REYES, M., DAUNT, F., AGNEW, P., ANKER-NILSSEN, T., BARBRAUD, C., BARRETT, R., BECKER, P. H., BELL, E., BOERSMA, P. D., BOUWHUIS, S., CANNELL, B., CRAWFORD, R. J. M., DANN, P., DELORD, K., ELLIOTT, G., ERIKSTAD, K. E., FLINT, E., FURNESS, R. W., HARRIS, M. P., HATCH, S., HILWIG, K., HINKE, J. T., JAHNCKE, J., MILLS, J. A., REIERTSEN, T. K., RENNER, H., SHERLEY, R. B., SURMAN, C., TAYLOR, G., THAYER, J. A., TRATHAN, P. N., VELARDE, E., WALKER, K., WANLESS, S., WARZYBOK, P. & WATANUKI, Y. 2021. Hemispheric asymmetry in ocean change and the productivity of ecosystem sentinels. *Science,* 372**,** 980-983.

THOMPSON, P. M., MCCONNELL, B. J., TOLLIT, D. J., MACKAY, A., HUNTER, C. & RACEY, P. A. 1996. Comparative distribution, movements and diet of harbour and grey seals from Moray Firth, NE Scotland. *Journal of Applied Ecology,* 33**,** 1572-1584.

THORNE, L. H., FOLEY, H. J., BAIRD, R. W., WEBSTER, D. L., SWAIM, Z. T. & READ, A. J. 2017. Movement and foraging behavior of short-finned pilot whales in the Mid-Atlantic Bight: importance of bathymetric features and implications for management. *Marine Ecology Progress Series,* 584**,** 245-257.

THUILLER, W., LAFOURCADE, B., ENGLER, R. & ARAÚJO, M. B. 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography,* 32**,** 369-373.

TITTENSOR, D. P., MORA, C., JETZ, W., LOTZE, H. K., RICARD, D., BERGHE, E. V. & WORM, B. 2010. Global patterns and predictors of marine biodiversity across taxa. *Nature,* 466**,** 1098-101.

TOLOTTI, M., BAUER, R., FORGET, F., BACH, P., DAGORN, L. & TRAVASSOS, P. 2017. Fine-scale vertical movements of oceanic whitetip sharks (*Carcharhinus longimanus*). *Fishery Bulletin,* 115**,** 380-395.

TRITES, A. W. 2019. Marine Mammal Trophic Levels and Trophic Interactions. *Encyclopedia of Ocean Sciences.*

TRUDELLE, L., CERCHIO, S., ZERBINI, A. N., GEYER, Y., MAYER, F. X., JUNG, J. L., HERVE, M. R., POUS, S., SALLEE, J. B., ROSENBAUM, H. C., ADAM, O. & CHARRASSIN, J. B. 2016. Influence of environmental parameters on movements and habitat utilization of humpback whales (*Megaptera novaeangliae*) in the Madagascar breeding ground. *R Soc Open Sci,* 3**,** 160616.

TYNAN, C. T., AINLEY, D. G., BARTH, J. A., COWLES, T. J., PIERCE, S. D. & SPEAR, L. B. 2005. Cetacean distributions relative to ocean processes in the northern California Current System. *Deep Sea Research Part II: Topical Studies in Oceanography,* 52**,** 145-167.

UNEP 2015. Marine No Net Loss: A feasibility assessment of implementing no net loss of biodiversity in the sea. *In:* CENTRE, U. E. W. C. M. (ed.). Cambridge, UK: UNEP-WCMC.

URMY, S. S. & WARREN, J. D. 2018. Foraging hotspots of common and roseate terns: the influence of tidal currents, bathymetry, and prey density. *Marine Ecology Progress Series,* 590**,** 227-245.

VAN LLEEUWEN, P. I., RIJNSDORP, A. D. & VINGERHOED, B. 1994. Variations in abundance and distribution of demersal fish species in the coastal zone of the Southeastern North Sea between 1980 and 1993. *In:* COMMITTEE, D. F. (ed.) *ICES.* The Netherlands: Netherlands Institute for Fisheries Research.

VAUGHN, R. L., WÜRSIG, B., SHELTON, D. S., TIMM, L. L. & WATSON, L. A. 2008. Dusky dolphins influence prey accessibility for seabirds in Admiralty Bay, New Zealand. *Journal of Mammalogy,* 89**,** 1051-1058.

VILLALOBOS, F., DOBROVOLSKI, R., PROVETE, D. B. & GOUVEIA, S. F. 2013. Is rich and rare the common share? Describing biodiversity patterns to inform conservation practices for South American anurans. *PLoS One,* 8**,** e56073.

VIRGILI, A., AUTHIER, M., BOISSEAU, O., CAÑADAS, A., CLARIDGE, D., COLE, T., CORKERON, P., DORÉMUS, G., DAVID, L., DI-MÉGLIO, N., DUNN, C., DUNN, T. E., GARCÍA-BARÓN, I., LARAN, S., LAURIANO, G., LEWIS, M., LOUZAO, M., MANNOCCI, L., MARTÍNEZ-CEDEIRA, J., PALKA, D., PANIGADA, S., PETTEX, E., ROBERTS, J. J., RUIZ, L., SAAVEDRA, C., SANTOS, M. B., VAN CANNEYT, O., VÁZQUEZ BONALES, J. A., MONESTIEZ, P., RIDOUX, V. & KEITH, S. 2019. Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale. *Global Ecology and Biogeography,* 28**,** 300-314.

VOLIS, S. 2022. "Species Distribution Modeling - a conservation tool with a wide spectrum of applications in Central Asia". *Plant Diversity of Central Asia,* 1.

WAGGITT, J. J., CAZENAVE, P. W., HOWARTH, L. M., EVANS, P. G. H., VAN DER KOOIJ, J. & HIDDINK, J. G. 2018a. Combined measurements of prey availability explain habitat selection in foraging seabirds. *Biol Lett,* 14.

WAGGITT, J. J., DUNN, H. K., EVANS, P. G. H., HIDDINK, J. G., HOLMES, L. J., KEEN, E., MURCOTT, B. D., PIANO, M., ROBINS, P. E., SCOTT, B. E., WHITMORE, J., VENERUSO, G. & CAMPHUYSEN, K. 2018b. Regional-scale patterns in harbour porpoise occupancy of tidal stream environments. *ICES Journal of Marine Science,* 75**,** 701-710.

WAGGITT, J. J., EVANS, P. G. H., ANDRADE, J., BANKS, A. N., BOISSEAU, O., BOLTON, M., BRADBURY, G., BRERETON, T., CAMPHUYSEN, C. J., DURINCK, J., FELCE, T., FIJN, R. C., GARCIA-BARON, I., GARTHE, S., GEELHOED, S. C. V., GILLES, A., GOODALL, M., HAELTERS, J., HAMILTON, S., HARTNY-MILLS, L., HODGINS, N., JAMES, K., JESSOPP, M., KAVANAGH, A. S., LEOPOLD, M., LOHRENGEL, K., LOUZAO, M., MARKONES, N., MARTÍNEZ-CEDEIRA, J., Ó CADHLA, O., PERRY, S. L., PIERCE, G. J., RIDOUX, V., ROBINSON, K. P., SANTOS, M. B., SAAVEDRA, C., SKOV, H., STIENEN, E. W. M., SVEEGAARD, S., THOMPSON, P., VANERMEN, N., WALL, D., WEBB, A., WILSON, J., WANLESS, S., HIDDINK, J. G. & PUNT, A. 2019. Distribution maps of cetacean and seabird populations in the North-East Atlantic. *Journal of Applied Ecology,* 57**,** 253-269.

WAKEFIELD, E. D., MILLER, D. L., BOND, S. L., LE BOUARD, F., CARVALHO, P. C., CATRY, P., DILLEY, B. J., FIFIELD, D. A., GJERDRUM, C., GONZÁLEZ-SOLÍS, J., HOGAN, H., LAPTIKHOVSKY, V., MERKEL, B., MILLER, J. A. O., MILLER, P. I., PINDER, S. J., PIPA, T., RYAN, P. M., THOMPSON, L. A., THOMPSON, P. M. & MATTHIOPOULOS, J. 2021. The summer distribution, habitat associations and abundance of seabirds in the sub-polar frontal zone of the Northwest Atlantic. *Progress in Oceanography,* 198.

WAKEFIELD, E. D., OWEN, E., BAER, J., CARROLL, M. J., DAUNT, F., DODD, S. G., GREEN, J. A., GUILFORD, T., MAVOR, R. A., MILLER, P. I. & NEWELL, M. A. 2017. Breeding density, fine-scale tracking, and large-scale modeling reveal the regional distribution of four seabird species. *Ecological Applications,* 27**,** 2074-2091.

WHITE, E., MESSINA, F., MOSS, L. & MESELHE, E. 2018. Salinity and Marine Mammal Dynamics in Barataria Basin: Historic Patterns and Modeled Diversion Scenarios. *Water,* 10.

WHITEHEAD, H., MCGILL, B. & WORM, B. 2008. Diversity of deep-water cetaceans in relation to temperature: implications for ocean warming. *Ecol Lett,* 11**,** 1198-1207.

WILLIAMS, C. L. & PONGANIS, P. J. 2021. Diving physiology of marine mammals and birds: the development of biologging techniques. *Philos Trans R Soc Lond B Biol Sci,* 376**,** 20200211.

WINDSOR, F. M., VAN DEN HOOGEN, J., CROWTHER, T. W. & EVANS, D. M. 2022. Using ecological networks to answer questions in global biogeography and ecology. *Journal of Biogeography,* 50**,** 57-69.

WITMAN, J. D., CUSSON, M., ARCHAMBAULT, P., PERSHING, A. J. & MIESKOWSA, N. 2008. The relation between productivity and species diversity in temperate–arctic marine ecosystems. *Ecology,* 89**,** S66-S80.

YEN, P. P. W., SYDEMAN, W. J. & HYRENBACH, K. D. 2004. Marine bird and cetacean associations with bathymetric habitats and shallow-water topographies: implications for trophic transfer and conservation. *Journal of Marine Systems,* 50**,** 79-99.

ZERBINI, A. N., FRIDAY, N. A., PALACIOS, D. M., WAITE, J. M., RESSLER, P. H., RONE, B. K., MOORE, S. E. & CLAPHAM, P. J. 2016. Baleen whale abundance and distribution in relation to environmental variables and prey density in the Eastern Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography,* 134**,** 312-330.

ZHANG, C., CHEN, Y., XU, B., XUE, Y., REN, Y. & TREML, E. 2019. How to predict biodiversity in space? An evaluation of modelling approaches in marine ecosystems. *Diversity and Distributions,* 25**,** 1697-1708.

ZHANG, X. 2022. Predicting global seasonal distributions and population exchange routes of a Critically Endangered shark. *Biological Conservation,* 275.

ZHANG, Y., QIN, B., CHEN, W., GAO, G. & CHEN, Y. 2004. Chlorophyll a content and primary productivity of phytoplankton in Meiliang Bay of Taihu Lake. . *Ying Yong Sheng tai xue bao= The Journal of Applied Ecology,* 15**,** 2127-2131.

ZIKA, J. D., SKLIRIS, N., BLAKER, A. T., MARSH, R., NURSER, A. J. G. & JOSEY, S. A. 2018. Improved estimates of water cycle change from ocean salinity: the key role of ocean warming. *Environmental Research Letters,* 13.

Chapter 6 General discussion

The overall aim of my thesis was to utilise large available datasets to address the research gap in mapping the biodiversity of top marine predators across multiple taxa, around the UK. Many studies have highlighted the difficulty in mapping biodiversity at a large scale due to patchy data coverage, and this is further intensified using a particularly cryptic set of species that inhabit and spend much of their life beneath the surface. The importance of mapping biodiversity hotspots is critical for conservation and management, where mapping and protecting biodiversity hotspots represent a cost-effective use of resources to protect species (Pimm and Raven, 2000, Roberts *et al.*, 2002, Mittermeier *et al.*, 2011, Marchese, 2015).

In this concluding chapter, I synthesise the findings of each of my chapters in the context of their wider application to biodiversity mapping and the coexistence of top marine predators. I highlight limitations faced while undertaking this research, discuss the broader scope of these findings and present some direction for future work in this field.

6.1 Utilising existing historical data to maximise coverage

Chapter 2 demonstrated available data does exist to maximise temporal and spatial coverage of such a large-scale area. For marine predators utilising past historical datasets is critical to maximise survey effort and coverage (Waggitt *et al.*, 2019), as monitoring underwater or diving animals in the open ocean environment is difficult (Thomson *et al.*, 2013) and can lead to patchy spatially biased coverage. It is well known that occurrence datasets have a variety of sampling biases within them (Hijmans *et al.*, 2000, Reddy and Dávalos, 2003, Kadmon *et al.*, 2004, Aiello-Lammens *et al.*, 2015) due to factors such as area accessibility (Davis *et al.*, 1990, Reddy and Dávalos, 2003, Rondinini *et al.*, 2006, McBride-Kebert *et al.*, 2019), sampling perspective (Henkel *et al.*, 2007, Hodgson *et al.*, 2016, Phillips *et al.*, 2019), availability bias (Thomson *et al.*, 2013, Williams *et al.*, 2017), perception bias (Buckland *et al.*, 2001, Hobbs and Waite, 2010) and observer error (Nichols *et al.*, 2000) amongst others. Having repeated coverage of such large marine areas with consistent methodologies is highly unlikely to be achievable at the spatial scale of this study, and requires a huge amount of resource, financial cost and time (Bull *et al.*, 2013). Utilising years of past data over the same area can address this problem and provide the wide temporal and spatial coverage needed for marine predators (Pacifici *et al.*, 2016) which are often wide-ranging with migrations at various life stages (Valente *et al.*, 2017, Yamaura and Royle, 2017). This chapter highlighted that suitable data exists without paywalls that can be used in revealing spatial and temporal patterns in marine predators (Dickinson *et al.*, 2010), and using this existing diverse data may aid in meeting conservation and management needs, where cost and time of new individual surveys may be expensive. The Aichi Target 19 specifically encouraged the development of a shared advanced biodiversity knowledge base (Meyer *et al.*, 2015). This is critical in an era where the drive to conserve biodiversity is pressing, both in legislation and in research – we do not have the time to collect perfect standardised data, and instead must utilise the years of sighting data we have already available.

6.2 Quantifying biases within datasets

The availability of distribution data from online data portals and databases is ever-increasing (with a vast array of datasets available over the area of interest in this thesis evidenced in Chapter 2) due to increasing awareness of the importance of sharing data to tackle information gaps (Leonard *et al.*, 2006, Henry *et al.*, 2008, Grémillet *et al.*, 2022, Saran *et al.*, 2022), and mapping can be carried out through the mobilisation of existing historical data in combination with new data (Kot *et al.*, 2010).

However, it is well-known that biodiversity databases have inherent biases within them due to heterogeneous data sources and lack of standardisation (Jetz and Rahbek, 2001, Soberón *et al.*, 2007, Lobo, 2008, Boakes *et al.*, 2010, Rocchini *et al.*, 2011, Ladle and Hortal, 2013, Meyer *et al.*, 2015, Lobo *et al.*, 2018). Often databases lack quantification of the level of risk or bias in using these databases. Chapter 3 addressed this issue by utilising a well-recognised risk-assessment matrix approach to quantify biases within four example datasets. The study highlighted that the level of risk in using heterogeneous datasets is lower for assessing patterns of association in marine predators, rather than counts or abundances and therefore research questions should be carefully considered when utilising datasets that are not designed to a specific research hypothesis. The use of collated databases such as Global Biodiversity Information Facility (GBIF) 24 has been criticised due to inherent biases (Hortal *et al.*, 2008, Boakes *et al.*, 2010, Beck *et al.*, 2013, Beck *et al.*, 2014, Bartomeus *et al.*, 2019) however, importantly, most criticisms have been aimed at using occurrence data to reconstruct and model species' distribution ranges (Orr *et al.*, 2021). However, Zattara and Aizen (2021) suggested a binning approach in which simpler questions (e.g. “has a species been recorded anywhere in the world during a given period?”) that yield a yes/no answer can potentially be much more robust to sampling-effort heterogeneity and geographic uncertainty (Telfer *et al.*, 2002). This agrees with the findings of the bias matrix in Chapter 3, where risks were reduced when data were used for relative questions about species coexistence or trends rather than estimating absolute abundance. As discussed in Chapter 3, the matrix method has two important roles to advance this field of research. The first is to aid people to analyse existing datasets and provide a standardised approach to enumerating bias quantitatively as opposed to just describing bias assumptions. The second is to guide research to design better surveys by understanding which risk factors are most influential on their study. Even heavily biased datasets might still have value to answer relative questions (where they are not suitable to answer absolute questions) and help contribute to the study of species mapping and coexistence where maximising effort in space and time is critical.

6.3 Species richness mapping

To address the gap in maps of marine predator biodiversity hotspots around the UK, richness maps and species richness per unit effort (SRPUE) can be used to identify areas of shared space use by multiple marine predators, across taxa (Chapter 4). This study is the largest cross-taxa study of marine predator biodiversity around the UK to date filling in the gap where inter-taxa biodiversity mapping is lacking at large spatial scales (Cox *et al.*, 2018), and even more so in the marine realm.

When compared both spatially and over time, raw species richness had artefacts of survey effort leading to areas of high species richness being driven by higher survey effort over those cells. As such, using SRPUE provides a method of offsetting survey effort bias, and is recommended over raw richness when identifying areas of relatively high species richness. Successfully identifying areas of coexistence for top sentinel species will have significant top-down ecosystem implications if those areas are protected. As keystone umbrella species with important ecological roles and wide home ranges (Albert *et al.*, 2018), protecting hotspots of top predators will be advantageous for healthy ecosystems and functioning. Interestingly, when survey effort is factored in, the concept of carrying capacity becomes evident and this is worth further exploration. Although this thesis did not have the time to go into life history traits in much detail, splitting out species richness into monthly or annual periods by taxa and exploring the link between drivers and life history traits (such as in Chapter 5) could perhaps reveal if there are changes in species assemblages.

However, the methodology is not without limitations, particularly due to its grid-based diversity measure. It has been shown that richness values in quadrats or grid systems coupled at a set distance often show autocorrelation, positive or negative (Diniz-filho *et al.*, 2003), and in particular, for marine predators with wide-ranging migrations, geographical range size relative to chosen cell size means adjacent cells could be pseudo-replicated in space. Therefore, a study into aggregations or spread of the biodiversity patterns is recommended for the next step or future study.

6.4 What drives biodiversity hotspots?

Understanding the potential drivers of species richness patterns has been identified as important in the literature, particularly with anthropogenic climate change causing shifts in species coexistence. Chapter 5 used three modelling approaches to look at associations of high species richness with a suite of environmental variables. Outputs of maps demonstrated if wanting to simply quantify relative areas of species richness, maps from macro-ecological models (MEMs) using generalised additive models (GAMs) and stacked species distribution models (SSDMs) provided similar results. This chapter provided a comparison of successful mapping and modelling approaches that can be used to quantify biodiversity patterns around the UK.

However, the modelling of marine species is a challenge due to the physical and biological characteristics of marine habitats and species (Robinson *et al.*, 2011); we have limited knowledge of pelagic species ecology and behaviour (Bentlage *et al.*, 2013). Identifying true drivers of species richness in the three-dimensional dynamic environment is difficult as most available environmental data layers describe the uppermost layers or the ocean floor (Assis *et al.*, 2018). Furthermore, marine habitats are dynamic with constant changes over different temporal and spatial scales (Redfern *et al.*, 2006, Scales *et al.*, 2017) and marine predator species are highly mobile (Fernandez *et al.*, 2017, Fernandez *et al.*, 2018) therefore whilst compiling many sources of data to one occurrence dataset may aid with minimising spatial bias and improve coverage, it removes the temporal details of patterns and therefore may not reflect fine-scale true environmental drivers of richness. Therefore, though the methodology in Chapter 5 is proof of concept, future recommendations are to carry out similar modelling at smaller temporal and spatial scales – to identify seasonal or monthly patterns in richness or identify more fine-scale drivers of biodiversity.

The primary example, whilst a significant driver of richness, fronts did not rank in high importance in driving hotspots of richness. However, this is likely due to the methodology of the study which averaged environmental variables over 39 years of data, and this may not allow the required spatial and temporal resolution for such highly dynamic and active variables. Utilising the proven methodology but over smaller spatial scales or splitting into smaller periods may show different patterns.

SST was identified as a high driver of species richness hotspots, aligning with other studies (Whitehead *et al.*, 2008, Astudillo-Scalia *et al.*, 2020, Chaudhary *et al.*, 2021, Evans *et al.*, 2021), and this has critical importance in a warming climate. If biodiversity hotspots are tied to SST, it may lead to marine species shifting their distributions or adapting in response (Burrows *et al.*, 2011, Poloczanska *et al.*, 2013, Poloczanska *et al.*, 2016, Donelson *et al.*, 2019, Pinsky *et al.*, 2022) to avoid extinction. A vast array of literature which supported associations between chlorophyll *a* and high marine predator densities or diversity (Louzao *et al.*, 2006, Scott *et al.*, 2010, Embling *et al.*, 2012, Sabarros *et al.*, 2014, Bennison and Jessopp, 2015, Miller *et al.*, 2015, Grecian *et al.*, 2016, Trudelle *et al.*, 2016, Zerbini *et al.*, 2016, Chen *et al.*, 2020, Kane *et al.*, 2020, Salgado Kent *et al.*, 2021). However, as highlighted in Chapter 5, better inclusion of other layers to delineate food web spatial dynamics would improve this study, for example using fish biomass or density layers (see Grémillet *et al.* (2008)).

An interesting finding from Chapter 5 was that those critically endangered species showed high proportions of SST and salinity drivers, further indicating that increasing global warming and climate change is a specific threat to species already at extremely high risk of extinction. Those species that are sensitive to fluctuations in heat content and salinity may be under higher levels of stress than those species that are not, and therefore prioritising areas of high biodiversity may aid in reducing the extinction risk of critically endangered species.

6.5 Practical applications and implications

In summary, this thesis has very high practical value for use in real-world management and conservation.

- It utilises available historical data which is freely available and therefore can aid in answering large research questions at very little cost. The thesis provides an overview of available data around the UK, which can be reviewed quickly for a summary of the status of information available which is useful for a wide range of stakeholders who are often under time pressure, under-resourced and trying to be proactive in a fast-paced legislative environment.
- People can use the bias risk matrix approach to design better surveys, and they can also assess what types of research questions can be answered by a given database if utilising the bias risk matrix.
- The maps of species richness can be utilised by a wide range of people to meet different goals. For example, the push for offshore renewable energies to reach net zero by 2030 is needed to be balanced against the need for biodiversity net gain (with marine biodiversity net gain likely to be mandatory in upcoming years). These maps could crucially demonstrate areas of high species richness where it is better to avoid construction but also highlight lower areas of biodiversity where net gain could be more easily achievable as it begins at a lower species richness. In another example, it could be used in marine management to target conservation and protection methods such as MPAs to areas of higher marine predator biodiversity. For research, these maps could be utilised with other measures of biodiversity to compare how taxonomic richness compares with those other measures of biodiversity (e.g. phylogenetic, functional diversity).
- Chapter 5 compared methods for understanding species richness drivers and allows future scientists to decide which modelling approach best suits their research. It highlighted two methods for obtaining species richness maps, with both MEMs and SSDMs having benefits and drawbacks, and suggests research question and the size and scale of data should drive choice of model. It identifies the important environmental variables of species richness and provides insight into the drivers of hotspots.

6.6 Future avenues for research

This thesis inevitably has unanswered questions which would provide clear avenues for future study. Firstly, utilising the methodology to create species richness maps (in Chapter 4) and exploring drivers of these patterns (Chapter 5) to carry out analysis over smaller temporal periods would be useful for looking at how biodiversity may have changed over time, and whether there are any seasonal differences in biodiversity hotspots, such as fluctuating with species migration patterns. Particularly for more dynamic drivers such as fronts and mixed layer thickness, using annual or monthly maps of richness may show these to be important, as these drivers may not be revealed in the average value over 39 years of data. Using maps at smaller spatial scales with relevant environmental driver layers for those periods may show more fine-scale patterns and drivers of marine predator species richness. There is also a critical need to look at how marine predator richness is associated with areas of high prey biomass. It has been highlighted that for example that the environmental variable chlorophyll *a* is a proxy for primary production, and therefore may not truly reflect the drive on predator hotspots due to several intermediate food chain levels. It would be beneficial to look at the links with marine predator hotspots across intermediate trophic levels, particularly with the increased fishing pressures on prey species.

Secondly, the logical progression of this research study is to look at hotspots of biodiversity overlaid with spatial management measures such as MPAs or SACs, to see if these areas of high biodiversity are contained within these protected areas. The maps can also be compared with offshore wind developments both in construction and for future creation. For example, there are ongoing leasing rounds for seabed rights to plan and build windfarms (The Crown Estate, 2023), and it would be useful for developers and stakeholders alike to look at the potential overlap with high biodiversity areas.

Another avenue of study to explore is looking at the functional richness for hotspots and whether certain life history traits drive the areas of higher species richness. An original idea for this thesis aimed to cross-compare hotspots of taxonomic richness, phylogenetic and functional richness, but the scale of the study and computational complexity did not allow for this and would be a useful logical next step given the increasing literature on biodiversity being more than simple taxonomic richness.

6.7 Conclusion

In conclusion, this thesis is critically the first study to map marine predator biodiversity across multiple taxa at this spatial scale. It contributes to the gap in the literature surrounding quantifying predator species richness for more than one taxonomic group at a time, which is much needed given the drive for mitigation or slowing of biodiversity decline and the legislative push for monitoring and reporting on ecosystem biodiversity. Available maps of biodiversity have much wider implications for management and conservation, with accurate maps of species diversity facilitating the development of targeted conservation strategies. It provides a valuable resource for policymakers and stakeholders alike.

It provides several methodologies that can be utilised and repeated to create species richness maps and identify potential drivers of biodiversity. It highlights that factoring in survey effort is critical for true richness patterns and should be considered in advance of ascertaining species richness hotspots. It demonstrates that available data is out there to be utilised, and if the biases inherent in these datasets are acknowledged and quantified, it provides a vast cost-effect array of data to be used for answering large-scale ecosystem-level questions for which small individual surveys cannot. Thus, a novel method for using the well-known risk assessment matrix for biases can guide the use of large-scale observation databases to inform research, management, and policy. Users can both use this matrix approach to design better surveys (i.e. minimise the level of bias in their surveys) and can also assess the type of research question that can be answered by a given database, allowing the selection of data to be well evaluated in terms of the level of bias within them.

This thesis highlights the importance of evaluating and standardising heterogeneous presence-only data before using it in further analysis. It introduces the novel SRPUE measure, whereby species richness is adjusted by effort and can be used to identify relative areas of high and low species richness. Failure to factor in survey effort when determining species richness hotspots can be detrimental to marine planning, we need to make sure we are crucially targeting protection measures in the right areas, rather than those skewed by increased survey effort. Using SRPUE can allow areas of high biodiversity to be prioritised in an era where quickly and proactively protecting biodiversity is at the forefront of global policy. Particularly when the UK and much of the developed world are rapidly transitioning to renewable energy, the marine realm will face increasing spatial conflict between biodiversity conservation and development, and having robust quantification of biodiversity is critical to balance the increasing overlap in the future.

6.8 References

AIELLO-LAMMENS, M. E., BORIA, R. A., RADOSAVLJEVIC, A., VILELA, B. & ANDERSON, R. P. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography,* 38**,** 541-545.

ALBERT, C., LUQUE, G. M. & COURCHAMP, F. 2018. The twenty most charismatic species. *PLoS One,* 13**,** e0199149.

ASSIS, J., TYBERGHEIN, L., BOSCH, S., VERBRUGGEN, H., SERRÃO, E. A., DE CLERCK, O. & TITTENSOR, D. 2018. Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography,* 27**,** 277-284.

ASTUDILLO-SCALIA, Y., ALBUQUERQUE, F. S. & PERES-NETO, P. 2020. The geography of high-priority conservation areas for marine mammals. *Global Ecology and Biogeography,* 29**,** 2097-2106.

BARTOMEUS, I., STAVERT, J. R., WARD, D. & AGUADO, O. 2019. Historical collections as a tool for assessing the global pollination crisis. *Philosophical Transactions of the Royal Society B,* 374**,** 20170389.

BECK, J., BALLESTEROS-MEJIA, L., NAGEL, P. & KITCHING, I. J. 2013. Online solutions and the 'Wallacean shortfall': what does GBIF contribute to our knowledge of species' ranges? *Diversity and Distributions,* 19**,** 1043-1050.

BECK, J., BÖLLER, M., ERHARDT, A. & SCHWANGHART, W. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics,* 19**,** 10-15.

BENNISON, A. & JESSOPP, M. 2015. At-sea surveys confirm a North Atlantic biodiversity hotspot. *Bird Study,* 62**,** 262-266.

BENTLAGE, B., PETERSON, A. T., BARVE, N. & CARTWRIGHT, P. 2013. Plumbing the depths: extending ecological niche modelling and species distribution modelling in three dimensions. *Global Ecology and Biogeography,* 22**,** 952-961.

BOAKES, E. H., MCGOWAN, P. J., FULLER, R. A., CHANG-QING, D., CLARK, N. E., O'CONNOR, K. & MACE, G. M. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol,* 8**,** e1000385.

BUCKLAND, S., ANDERSON, D., BURNHAM, K., LAAKE, J., BORCHERS, D. & THOMAS, L. 2001. *Introduction to distance sampling: estimating abundance of biological populations.,* Oxford, Oxford University Press.

BULL, J. C., MASON, S., WOOD, C. & PRICE, A. R. G. 2013. Benthic marine biodiversity patterns across the United Kingdom and Ireland determined from recreational diver observations: A baseline for possible species range shifts induced by climate change. *Aquatic Ecosystem Health & Management,* 16**,** 20-30.

BURROWS, M. T., SCHOEMAN, D. S., BUCKLEY, L. B., MOORE, P., POLOCZANSKA, E. S., BRANDER, K. M., BROWN, C., BRUNO, J. F., DUARTE, C. M., HALPERN, B. S., HOLDING, J., KAPPEL, C. V., KIESSLING, W., O'CONNOR, M. I., PANDOLFI, J. M., PARMESAN, C., SCHWING, F. B., SYDEMAN, W. J. & RICHARDSON, A. J. 2011. The pace of shifting climate in marine and terrestrial ecosystems. *Science,* 334**,** 652-5.

CHAUDHARY, C., RICHARDSON, A. J., SCHOEMAN, D. S. & COSTELLO, M. J. 2021. Global warming is causing a more pronounced dip in marine species richness around the equator. *Proc Natl Acad Sci U S A,* 118.

CHEN, B., HONG, Z., HAO, X. & GAO, H. 2020. Environmental models for predicting habitat of the Indo-Pacific humpback dolphins in Fujian, China. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 30**,** 787-793.

COX, S. L., EMBLING, C. B., HOSEGOOD, P. J., VOTIER, S. C. & INGRAM, S. N. 2018. Oceanographic drivers of marine mammal and seabird habitat-use across shelf-seas: A guide to key features and recommendations for future research and conservation management. *Estuarine, Coastal and Shelf Science,* 212**,** 294-310.

DAVIS, F. W., STOMS, D. M., ESTES, J. E., SCEPAN, J. & SCOTT, J. M. 1990. An information systems approach to the preservation of biological diversity. *International Journal of Geographical Information System,* 4**,** 55-78.

DICKINSON, J. L., ZUCKERBERG, B. & BONTER, D. N. 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics,* 41**,** 149-172.

DINIZ-FILHO, J. A. F., BINI, L. M. & HAWKINS, B. A. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology & Biogeography,* 12**,** 53-64.

DONELSON, J. M., SUNDAY, J. M., FIGUEIRA, W. F., GAITAN-ESPITIA, J. D., HOBDAY, A. J., JOHNSON, C. R., LEIS, J. M., LING, S. D., MARSHALL, D., PANDOLFI, J. M., PECL, G., RODGERS, G. G., BOOTH, D. J. & MUNDAY, P. L. 2019. Understanding interactions between plasticity, adaptation and range shifts in response to marine environmental change. *Philos Trans R Soc Lond B Biol Sci,* 374**,** 20180186.

EMBLING, C. B., ILLIAN, J., ARMSTRONG, E., VAN DER KOOIJ, J., SHARPLES, J., CAMPHUYSEN, K. C. J. & SCOTT, B. E. 2012. Investigating fine-scale spatio-temporal predator-prey patterns in dynamic marine ecosystems: a functional data analysis approach. *Journal of Applied Ecology,* 49**,** 481-492.

EVANS, R., LEA, M. A. & HINDELL, M. A. 2021. Predicting the distribution of foraging seabirds during a period of heightened environmental variability. *Ecol Appl,* 31**,** e02343.

FERNANDEZ, M., YESSON, C., GANNIER, A., MILLER, P. I. & AZEVEDO, J. M. N. 2017. The importance of temporal resolution for niche modelling in dynamic marine environments. *Journal of Biogeography,* 44**,** 2816-2827.

FERNANDEZ, M., YESSON, C., GANNIER, A., MILLER, P. I. & AZEVEDO, J. M. N. 2018. A matter of timing: how temporal scale selection influences cetacean ecological niche modelling. *Marine Ecology Progress Series,* 595**,** 217-231.

GRECIAN, W. J., WITT, M. J., ATTRILL, M. J., BEARHOP, S., BECKER, P. H., EGEVANG, C., FURNESS, R. W., GODLEY, B. J., GONZALEZ-SOLIS, J., GREMILLET, D., KOPP, M., LESCROEL, A., MATTHIOPOULOS, J., PATRICK, S. C., PETER, H. U., PHILLIPS, R. A., STENHOUSE, I. J. & VOTIER, S. C. 2016. Seabird diversity hotspot linked to ocean productivity in the Canary Current Large Marine Ecosystem. *Biol Lett,* 12.

GRÉMILLET, D., CHEVALLIER, D., GUINET, C. & VOTIER, S. 2022. Big data approaches to the spatial ecology and conservation of marine megafauna. *ICES Journal of Marine Science,* 79**,** 975-986.

GRÉMILLET, D., LEWIS, S., DRAPEAU, L., VAN DER LINGEN, C. D., HUGGETT, J. A., COETZEE, J. C., VERHEYE, H. M., DAUNT, F., WANLESS, S. & RYAN, P. G. 2008. Spatial match-mismatch in the Benguela upwelling zone: should we expect chlorophyll and sea-surface temperature to predict marine predator distributions? *Journal of Applied Ecology,* 45**,** 610-621.

HENKEL, L. A., FORD, R. G., TYLER, W. B. & DAVIS, J. N. 2007. Comparison of aerial and boat-based survey methods for marbled murrelets *Brachyramphus marmoratus* and other marine birds. *Marine Ornithology,* 35**,** 145–151.

HENRY, P.-Y., LENGYEL, S., NOWICKI, P., JULLIARD, R., CLOBERT, J., ČELIK, T., GRUBER, B., SCHMELLER, D. S., BABIJ, V. & HENLE, K. 2008. Integrating ongoing biodiversity monitoring: potential benefits and methods. *Biodiversity and Conservation,* 17**,** 3357-3382.

HIJMANS, R. J., GARRETT, K. A., HUAMAN, Z., ZHANG, D. P., SCHREUDER, M. & BONIERBALE, M. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology,* 14**,** 1755–1765.

HOBBS, R. C. & WAITE, J. M. 2010. Abundance of harbor porpoise (*Phocoena phocoena*) in three Alaskan regions, corrected for observer errors due to perception bias and species misidentification, and corrected for animals submerged from view. *Fishery Bulletin,* 108**,** 251-267.

HODGSON, J. C., BAYLIS, S. M., MOTT, R., HERROD, A. & CLARKE, R. H. 2016. Precision wildlife monitoring using unmanned aerial vehicles. *Scientific Reports,* 6**,** 22574.

HORTAL, J., JIMÉNEZ-VALVERDE, A., GÓMEZ, J. F., LOBO, J. M. & BASELGA, A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos,* 117**,** 847-858.

JETZ, W. & RAHBEK, C. 2001. Geometric constraints explain much of the species richness pattern in African birds. *Proc Natl Acad Sci U S A,* 98**,** 5661-6.

KADMON, R., FARBER, O. & DANIN, A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications,* 14**,** 401-413.

KANE, A., PIROTTA, E., WISCHNEWSKI, S., CRITCHLEY, E. J., BENNISON, A., JESSOPP, M. & QUINN, J. L. 2020. Spatio-temporal patterns of foraging behaviour in a wide-ranging seabird reveal the role of primary productivity in locating prey. *Marine Ecology Progress Series,* 646**,** 175-188.

KOT, C. Y., FUJIOKA, E., HAZEN, L. J., BEST, B. D., READ, A. J. & HALPIN, P. N. 2010. Spatio-temporal gap analysis of OBIS-SEAMAP project data: assessment and way forward. *PLoS One,* 5**,** e12990.

LADLE, R. & HORTAL, J. 2013. Mapping species distributions: living with uncertainty. *Frontiers of Biogeography,* 5.

LEONARD, D. R., ROBERT CLARKE, K., SOMERFIELD, P. J. & WARWICK, R. M. 2006. The application of an indicator based on taxonomic distinctness for UK marine biodiversity assessments. *J Environ Manage,* 78**,** 52-62.

LOBO, J. M. 2008. Database records as a surrogate for sampling effort provide higher species richness estimations. *Biodiversity and Conservation,* 17**,** 873-881.

LOBO, J. M., HORTAL, J., YELA, J. L., MILLÁN, A., SÁNCHEZ-FERNÁNDEZ, D., GARCÍA-ROSELLÓ, E., GONZÁLEZ-DACOSTA, J., HEINE, J., GONZÁLEZ-VILAS, L. & GUISANDE, C. 2018. KnowBR: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. *Ecological Indicators,* 91**,** 241-248.

LOUZAO, M., HYRENBACH, K. D., ARCOS, J. M., ABELLO, P., GIL DE SOLA, L. & ORO, D. 2006. Oceanographic habitat of an endangered Mediterranean procellariiform: implications for marine protected areas. *Ecol Appl,* 16**,** 1683-95.

MARCHESE, C. 2015. Biodiversity hotspots: A shortcut for a more complicated concept. *Global Ecology and Conservation,* 3**,** 297-309.

MCBRIDE-KEBERT, S., TAYLOR, J. S., LYN, H., MOORE, F. R., SACCO, D. F., KAR, B. & KUCZAJ II, S. A. 2019. Controlling for survey effort is worth the effort: comparing Bottlenose dolphin (*Tursiops truncatus*) habitat use between standardized and opportunistic photographic-identification surveys. *Aquatic Mammals,* 45**,** 21-29.

MEYER, C., KREFT, H., GURALNICK, R. & JETZ, W. 2015. Global priorities for an effective information basis of biodiversity distributions. *Nat Commun,* 6**,** 8221.

MILLER, P. I., SCALES, K. L., INGRAM, S. N., SOUTHALL, E. J., SIMS, D. W. & COSTA, D. 2015. Basking sharks and oceanographic fronts: quantifying associations in the north-east Atlantic. *Functional Ecology,* 29**,** 1099-1109.

MITTERMEIER, R. A., TURNER, W. R., LARSEN, F. W., BROOKS, T. M. & GASCON, C. 2011. Global Biodiversity Conservation: The Critical Role of Hotspots. . *In:* HABEL, F. E. Z. A. J. C. (ed.) *Biodiversity hotspots: distribution and protection of conservation priority areas. .* Springer-Verlag, Berlin, Heidelberg, Germany.

NICHOLS, J. D., HINES, J. E., SAUER, J. R., FALLON, W. F., FALLON, J. E. & HEGLUND, P. J. 2000. A double-observer approach for estimating detection probability and abundance from point counts. *The Auk,* 117**,** 393-408.

ORR, M. C., HUGHES, A. C., CHESTERS, D., PICKERING, J., ZHU, C.-D. & ASCHER, J. S. 2021. Global Patterns and Drivers of Bee Distribution. *Current Biology,* 31**,** 451-458.e4.

PACIFICI, K., REICH, B. J., DORAZIO, R. M. & CONROY, M. J. 2016. Occupancy estimation for rare species using a spatially-adaptive sampling design. *Methods in Ecology and Evolution,* 7**,** 285-293.

PHILLIPS, E. M., HORNE, J. K., ZAMON, J. E., FELIS, J. J. & ADAMS, J. 2019. Does perspective matter? A case study comparing Eulerian and Lagrangian estimates of common murre (Uria aalge) distributions. *Ecol Evol,* 9**,** 4805-4819.

PIMM, S. L. & RAVEN, P. 2000. Biodiversity. Extinction by numbers. *Nature,* 403**,** 843-5.

PINSKY, M. L., COMTE, L. & SAX, D. F. 2022. Unifying climate change biology across realms and taxa. *Trends Ecol Evol,* 37**,** 672-682.

POLOCZANSKA, E. S., BROWN, C. J., SYDEMAN, W. J., KIESSLING, W., SCHOEMAN, D. S., MOORE, P. J., BRANDER, K., BRUNO, J. F., BUCKLEY, L. B., BURROWS, M. T., DUARTE, C. M., HALPERN, B. S., HOLDING, J., KAPPEL, C. V., O'CONNOR, M. I., PANDOLFI, J. M., PARMESAN, C., SCHWING, F., THOMPSON, S. A. & RICHARDSON,

A. J. 2013. Global imprint of climate change on marine life. *Nature Climate Change,* 3**,** 919-925.

POLOCZANSKA, E. S., BURROWS, M. T., BROWN, C. J., GARCÍA MOLINOS, J., HALPERN, B. S., HOEGH-GULDBERG, O., KAPPEL, C. V., MOORE, P. J., RICHARDSON, A. J., SCHOEMAN, D. S. & SYDEMAN, W. J. 2016. Responses of Marine Organisms to Climate Change across Oceans. *Frontiers in Marine Science,* 3.

REDDY, S. & DÁVALOS, L. M. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography,* 30**,** 1719-1727.

REDFERN, J. V., FERGUSON, M. C., BECKER, E. A., HYRENBACH, K. D., GOOD, C., BARLOW, J., KASCHNER, K., BAUMGARTNER, M. F., FORNEY, K. A. & BALLANCE, L. T. 2006. Techniques for cetacean–habitat modeling. *Marine Ecology Progress Series,* 310**,** 271-295.

ROBERTS, C. M., MCCLEAN, C. J., VERON, J. E., HAWKINS, J. P., ALLEN, G. R., MCALLISTER, D. E., MITTERMEIER, C. G., SCHUELER, F. W., SPALDING, M., WELLS, F., VYNNE, C. & WERNER, T. B. 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science,* 295**,** 1280-4.

ROBINSON, L. M., ELITH, J., HOBDAY, A. J., PEARSON, R. G., KENDALL, B. E., POSSINGHAM, H. P. & RICHARDSON, A. J. 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography,* 20**,** 789-802.

ROCCHINI, D., HORTAL, J., LENGYEL, S., LOBO, J. M., JIMÉNEZ-VALVERDE, A., RICOTTA, C., BACARO, G. & CHIARUCCI, A. 2011. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography: Earth and Environment,* 35**,** 211-226.

RONDININI, C., WILSON, K. A., BOITANI, L., GRANTHAM, H. & POSSINGHAM, H. P. 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters,* 9**,** 1136-1145.

SABARROS, P. S., GRÉMILLET, D., DEMARCQ, H., MOSELEY, C., PICHEGRU, L., MULLERS, R. H. E., STENSETH, N. C. & MACHU, E. 2014. Fine-scale recognition and use of mesoscale fronts by foraging Cape gannets in the Benguela upwelling region. *Deep Sea Research Part II: Topical Studies in Oceanography,* 107**,** 77-84.

SALGADO KENT, C., BOUCHET, P., WELLARD, R., PARNUM, I., FOUDA, L. & ERBE, C. 2021. Seasonal productivity drives aggregations of killer whales and other cetaceans over submarine canyons of the Bremer Sub-Basin, south-western Australia. *Australian Mammalogy,* 43.

SARAN, S., CHAUDHARY, S. K., SINGH, P., TIWARI, A. & KUMAR, V. 2022. A comprehensive review on biodiversity information portals. *Biodiversity and Conservation,* 31**,** 1445-1468.

SCALES, K. L., HAZEN, E. L., JACOX, M. G., EDWARDS, C. A., BOUSTANY, A. M., OLIVER, M. J. & BOGRAD, S. J. 2017. Scale of inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data. *Ecography,* 40**,** 210-220.

SCOTT, B. E., SHARPLES, J., ROSS, O. N., WANG, J., PIERCE, G. J. & CAMPHUYSEN, C. J. 2010. Sub-surface hotspots in shallow seas: fine-scale limited locations of top predator foraging habitat indicated by tidal mixing and sub-surface chlorophyll. *Marine Ecology Progress Series,* 408**,** 207-226.

SOBERÓN, J., JIMÉNEZ, R., GOLUBOV, J. & KOLEFF, P. 2007. Assessing completeness of biodiversity databases at different spatial scales. *Ecography,* 30**,** 152-160.

TELFER, M. G., PRESTON, C. D. & ROTHERY, P. 2002. A general method for measuring relative change in range size from biological atlas data. *Biological Conservation,* 107**,** 99-109.

THE CROWN ESTATE. 2023. *Leasing Round 4: driving a net zero clean energy future* [Online]. Online: The Crown Estate,. Available: https://www.thecrownestate.co.uk/round-4/#:~:text=Offshore%20Wind%20Leasing%20Round%204%20creates%20the%20opportunity%20for%20c,to%20net%20zero%20by%202050. [Accessed 11th March 2023].

THOMSON, J. A., COOPER, A. B., BURKHOLDER, D. A., HEITHAUS, M. R. & DILL, L. M. 2013. Correcting for heterogeneous availability bias in surveys of long-diving marine turtles. *Biological Conservation,* 165**,** 154-161.

TRUDELLE, L., CERCHIO, S., ZERBINI, A. N., GEYER, Y., MAYER, F. X., JUNG, J. L., HERVE, M. R., POUS, S., SALLEE, J. B., ROSENBAUM, H. C., ADAM, O. & CHARRASSIN, J. B. 2016. Influence of environmental parameters on movements and habitat utilization of humpback whales (*Megaptera novaeangliae*) in the Madagascar breeding ground. *R Soc Open Sci,* 3**,** 160616.

VALENTE, J. J., HUTCHINSON, R. A. & BETTS, M. G. 2017. Distinguishing distribution dynamics from temporary emigration using dynamic occupancy models. *Methods in Ecology and Evolution,* 8**,** 1707-1716.

WAGGITT, J. J., EVANS, P. G. H., ANDRADE, J., BANKS, A. N., BOISSEAU, O., BOLTON, M., BRADBURY, G., BRERETON, T., CAMPHUYSEN, C. J., DURINCK, J., FELCE, T., FIJN, R. C., GARCIA-BARON, I., GARTHE, S., GEELHOED, S. C. V., GILLES, A., GOODALL, M., HAELTERS, J., HAMILTON, S., HARTNY-MILLS, L., HODGINS, N., JAMES, K., JESSOPP, M., KAVANAGH, A. S., LEOPOLD, M., LOHRENGEL, K., LOUZAO, M., MARKONES, N., MARTÍNEZ-CEDEIRA, J., Ó CADHLA, O., PERRY, S. L., PIERCE, G. J., RIDOUX, V., ROBINSON, K. P., SANTOS, M. B., SAAVEDRA, C., SKOV, H., STIENEN, E. W. M., SVEEGAARD, S., THOMPSON, P., VANERMEN, N., WALL, D., WEBB, A., WILSON, J., WANLESS, S., HIDDINK, J. G. & PUNT, A. 2019. Distribution maps of cetacean and seabird populations in the North-East Atlantic. *Journal of Applied Ecology,* 57**,** 253-269.

WHITEHEAD, H., MCGILL, B. & WORM, B. 2008. Diversity of deep-water cetaceans in relation to temperature: implications for ocean warming. *Ecol Lett,* 11**,** 1198-1207.

WILLIAMS, R., ASHE, E., GAUT, K., GRYBA, R., MOORE, J. E., REXSTAD, E., SANDILANDS, D., STEVENTON, J. & REEVES, R. R. 2017. Animal Counting Toolkit: a practical guide to small-boat surveys for estimating abundance of coastal marine mammals. *Endangered Species Research,* 34**,** 149-165.

YAMAURA, Y. & ROYLE, J. A. 2017. Community distance sampling models allowing for im- perfect detection and temporary emigration. . *Ecosphere,* 8.

ZATTARA, E. E. & AIZEN, M. A. 2021. Worldwide occurrence records suggest a global decline in bee species richness. *One Earth,* 4**,** 114-123.

ZERBINI, A. N., FRIDAY, N. A., PALACIOS, D. M., WAITE, J. M., RESSLER, P. H., RONE, B. K., MOORE, S. E. & CLAPHAM, P. J. 2016. Baleen whale abundance and distribution in relation to environmental variables and prey density in the Eastern Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography,* 134**,** 312-330.

Appendices

Appendix A Supplementary information for Chapter 2

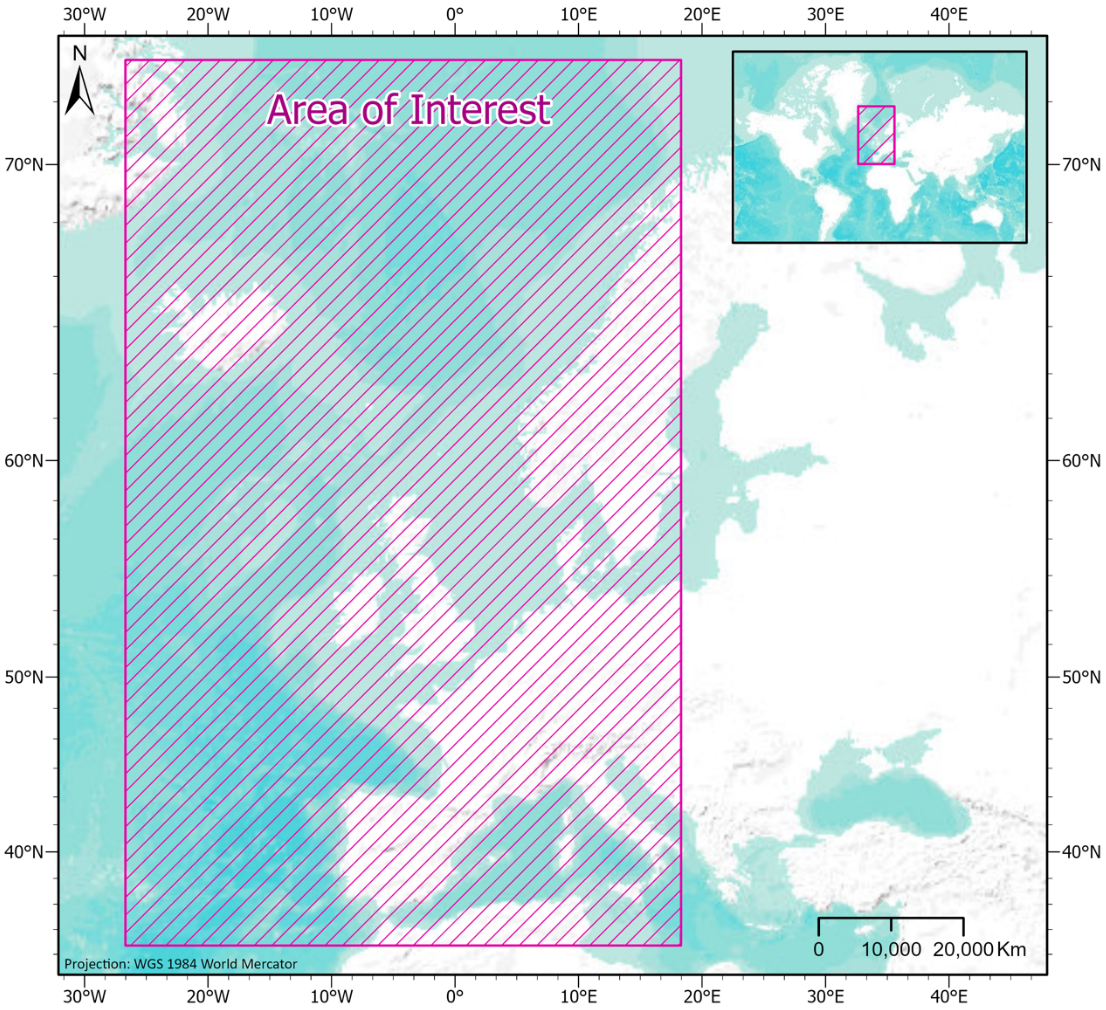


Figure A.1: Area of Interest for the chapter.

Table A-1 Identifier for datasets used in Figure A.2 to Figure A.5.

ID	Final name field	Link (as of June 2019)
1	All Wales Common Scoter (Melanitta nigra) Survey (1994-2002)	https://portal.medin.org.uk/portal/start.php#details?tpc=007_4f4c4942-4343-5764-6473-3§03436313333
2	Allied Humpback Whale Catalogue (1976 - 2003)	https://seamap.env.duke.edu/dataset/73
3	Amec, Islay Offshore Wind Farm, Marine Mammals (2010-2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_fb493fd2-268f-44e9-93a6-87788b813af3
4	AMEC, Kintyre Offshore Wind Farm, Bird and Mammal Surveys (2009 - 2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_61da586c129d4a609e8dd9c863ef856f
5	Amec, Scottish territorial Waters Offshore Wind Farms West Coast, Marine Mammals Cumulative Study (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_9aa5e24a-bede-10c7-9ae2-fe2e81a796e4
6	APEM, Blyth Narec Offshore Demonstration Site, Ornithology Assessment (2010 - 2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_eb81f7f2-e6d8-4bd9-9e40-c95030e1b020
7	APEM, London Array Offshore Wind Farm, Ornithological Aerial Survey (2012 - 2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_48513d049b76e7f45a61e7d40a90e66f
8	Argyll Array Offshore Wind Farm, Bird Survey (2009 - 2011)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_333cc8cc63a346f0947796686610ce19
9	Argyll Biological Records Dataset	https://registry.nbnatlas.org/public/show/dr819
10	Atlantic Array Underwater Noise Monitoring	https://portal.medin.org.uk/portal/start.php#details?tpc=010_d58098fe494c99c7c6c5896a30345407
11	Atlas of the Marine Mammals of Wales	https://portal.medin.org.uk/portal/start.php#details?tpc=007_4f4c4942-4343-5764-6473-313131383133
12	Baltic Porpoise Acoustic Surveys 01-02	https://seamap.env.duke.edu/dataset/343
13	Baltic Porpoise Sightings 01-02	https://seamap.env.duke.edu/dataset/344
14	BioConsult, Methodologies for measuring and assessing potential changes in MM behaviour, abundance or distribution arising from offshore windfarms (2008)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_8f2e4fc0-816e-4ec6-908d-c601190a5247
15	Biodiversity of the North Sea - Sylt	https://seamap.env.duke.edu/dataset/103152532
16	BLM CETAP OPP Sightings	https://seamap.env.duke.edu/dataset/284
17	Bottlenose Dolphin and Harbour Porpoise Monitoring in Cardigan Bay and Pen Llyn ar Sarnau SAC	https://portal.medin.org.uk/portal/start.php#details?tpc=007_4f4c4942-4343-5764-6473-313135313639
18	BOWind, Barrow Offshore Wind Farm, Construction Monitoring	https://portal.medin.org.uk/portal/start.php#details?tpc=015_1bbfb61f-811d-4651-b289-eb1f42937b81

19	Breeding Coastal Seabirds for oil spill contingency plans	https://portal.medin.org.uk/portal/start.php#details?tpc=012_Marine_Scotland_FishDAC_1777
20	BRERC species records from all years at full resolution excluding Notable Species within the last 10 years	https://registry.nbnatlas.org/public/show/dr938
21	British records in the osteology collections of World Museum, National Museums Liverpool	https://registry.nbnatlas.org/public/show/dr2614
22	BTO, A review of the Potential Impacts on Marine Aggregate Extraction on Seabirds (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_23e4ae3c885d570693fadd0377947bf9
23	BTO, Use of Aerial Surveys to Detect Bird Displacement by Offshore Windfarms (2006)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_2b77fdf2df465e6fb03b85d21ad5cb56
24	Caloo Ecological Services, Brough Head Wave Farm, Marine Wildlife Surveys (2010 - 2011)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_842a1e69-bb4a-4dac-b317-47381751b0f1
25	Caloo Ecological Services, Brough Head Wave Farm, Shore Based Monitoring of Seabirds and Cetaceans at the Outshore Point Development Area (2012 - 2014)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_01901a54-c796-414e-8012-814f23c8bc0b
26	Caroline Weir, Gunfleet Sands Offshore Wind Farm, Marine Mammals	https://portal.medin.org.uk/portal/start.php#details?tpc=015_8b796524352d57deabf3dd21f218c577
27	CEDaR Online Recording	https://registry.nbnatlas.org/public/show/dr1954
28	CEFAS Gillnet retrieval survey 2006/07 - Fisheries Science Partnership	https://portal.medin.org.uk/portal/start.php#details?tpc=009_CEFAS29a4f7ba-bf73-4fc0-a46d-fff918085a4b
29	CEFAS SHARK BYWATCH CATCH DATA	https://portal.medin.org.uk/portal/start.php#details?tpc=009_CEFASfea6bfb9-3b7f-4a6a-ad4d-3d7f2923f138
30	CEFAS Western Edge ghost nets 2005/06 - Fisheries Science Partnership	https://portal.medin.org.uk/portal/start.php#details?tpc=009_CEFASe049ffc5-38e9-43d4-8712-be8799007ec7
31	CMACS Burbo Bank Offshore Wind Farm, Post (2008 - 2009)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_b5dffdc9-d96e-4483-b74d-cc9e3599e21a
32	CMACS Greater Gabbard Offshore Wind Farm, Year 1 Post construction Benthic Ecology and Turbine colonisation survey (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_0db3ca7715ec52a3973569eb7114edeb
33	CMACS Kentish Flats Offshore Wind Farm, EMF Modelling and Interpretation for Electrosensitive Fish Species (2004)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_a265b90eed735c6996d291cd8d430912
34	CMACS West of Duddon Sands Offshore Wind Farm, Boat-based Ornithological Survey (2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_64c341cf-23e3-4f3d-87be-5c062f48e2ff
35	CMACS, Zone 9 Celtic Array, Autumn Fish Trawl Surveys (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_54e49e8cee064c41f64616ddd1deae7b

36	Coastal Sensitivity Atlas	https://portal.medin.org.uk/portal/start.php#details?tpc=007_4f4c4942-4343-5764-6473-313039393231
37	Coastwatch Marine Species Records from Northern Ireland	https://registry.nbnatlas.org/public/show/dr1972
38	CODA cetacean sightings on primary platform of vessel surveys 2007	https://seamap.env.duke.edu/dataset/1180
39	Commissioned surveys and staff surveys and reports for Scottish Wildlife Trust reserves	https://registry.nbnatlas.org/public/show/dr793
40	Copeland Island Bird Observatory SeaWatch Survey	https://registry.nbnatlas.org/public/show/dr1930
41	Costa Head, Natural Research Projects Ltd. (NRP) Marine Wildlife Survey (2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_1bdd066c559646a7b4840825b6c69546
42	COWRIE, Assessment and costs of potential engineering solutions for the mitigation of the impacts of underwater noise arising from the construction of offshore windfarms (2007)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_739a052fe37c54be81e183eae86e9bc3
43	COWRIE, Assessment of the potential acoustic deterrents to mitigate the impact on marine mammals of underwater noise arising from the construction of Offshore Wind Farms (2007)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_c8e3f67bdc085f7dbfc4df4c66a785fd
44	COWRIE, Comparison of design and model based estimates of seabird abundance derived from visual, digital still transects and digital video aerial surveys in Carmarthen Bay (2009)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_7b0fe22a-9c9a-4041-92ec-4a7f8d03fc4a
45	COWRIE, Effects of offshore wind farm noise on marine mammals and fish (2006)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_f99c7b2abb4253e69e5e886acfed7988
46	COWRIE, Measurement and interpretation of underwater noise during construction and operation of offshore windfarms in UK waters (2007)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_b6f9a377eed554658b9e92ce988f6b12
47	COWRIE, Trial High Definition Video Survey of Seabirds (2007)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_1c4cbcd04529520db9c7efa82882eaca
48	CREEM, Comparison of visual and digital aerial survey results of avian abundance for Round 3, Norfolk Region (2009)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_9a8fd593-259a-44e9-8d7e-8e9f707c556b
49	CRRU, Cetacean sighting in Scotland waters 1997-2015	https://seamap.env.duke.edu/dataset/647
50	DAERA Marine and Fisheries Division Marine Survey Data	https://registry.nbnatlas.org/public/show/dr1928
51	DASSH Data Archive Centre volunteer sightings records	https://registry.nbnatlas.org/public/show/dr1204
52	DASSH Data Archive Centre volunteer survey data	https://registry.nbnatlas.org/public/show/dr1200

53	Deepwater Elasmobranch Species Data From MSS Trawling Surveys (1996 - 2019)	https://portal.medin.org.uk/portal/start.php#details?tpc=012_Marine_Scotland_FishDAC_12326
54	Defra MB0102 2B Distribution of Mobile species half rectangle resolution layer (2009)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_97afe93c5da190ac68dde4aa35723e34
55	Devon Biodiversity Records Centre (DBRC) Strandings Database (1913)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_e6ca0bc5a3286b1b151b72cafd726139
56	DEVON Spawning and Nursery Grounds Layers for Selected Fish in UK Waters in 2010	https://portal.medin.org.uk/portal/start.php#details?tpc=009_CEFAS9690c5d5-dd4d-4f81-94ce-207cc6b58466
57	Devon Wildlife Trust (DWT), Sea Watch Foundation (SWF) Seaquest Sightings Database (1840)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_b555f22d7d464b90130166afbccdf5fc
58	DFO Maritimes Region Cetacean Sightings	https://seamap.env.duke.edu/dataset/103152397
59	DHI Water and Environment, Environmental Impact Assessment, Marine Mammals in the NW3 Area, Irish Sea (2006)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_7c79b44f781f582985610c3ce4682c8e
60	DHI, Offshore Renewables Joint Industry Programme (ORJIP), Bird Collision Avoidance Study (2014 - 2016)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_8153729f-f40c-435f-825a-67c79fe5ffa9
61	Diveboard - Scuba diving citizen science observations	https://seamap.env.duke.edu/dataset/103152427
62	DMP, Zone 3 Dogger Bank Teesside A and B, Statistical Analysis of Marine Mammal Data (2014)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_9dfd6d8b20414954acb623d73b9dc43b
63	DONG Energy Burbo Extension (UK) Ltd. Environmental Statement Series (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_ecb079a779693083ed770cc7e43e565b
64	DONG Energy, Walney Extension Preliminary Environmental Information (2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_ba76603e-a803-4984-b426-03bd6fbc9e9c
65	DONG Energy, Walney Extension Offshore Wind Farm, Environmental Statement & HRA (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_5447a39342655c9b633555d50c59e7ef
66	Ecologic, Lincs Offshore Wind Farm, Passive Acoustic Monitoring for Porpoises, Fieldwork and Preliminary Analysis (2004)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_2cf46001407959f5b5f605e634d44ca2
67	ECON Ecological Consultancy, Dudgeon Offshore Wind Farm, Boat based Ornithological Surveys (2007 - 2008)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_a86e599ade025cf5a742d64e4614f171
68	ECON, Lincs Offshore Wind Farm, Pre-construction Ornithological Baseline Surveys (2007 - 2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_7aae2b16-565e-4071-86dd-a5da71e3d3e7
69	ECON, Zone 8 Bristol Channel Atlantic Array, Ornithological Survey (2009 - 2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_98637447c9f8480ebc912efedafabdd7

70	ECON, Zone 9 Celtic Array, Bird Mammal Survey (2010 - 2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_f7876859-6865-4e8c-be1f-8a9011b118b1
71	Environment Agency Rare and Protected Species Records	https://registry.nbnatlas.org/public/show/dr815
72	ERIC NE Combined dataset to 2017	https://registry.nbnatlas.org/public/show/dr1574
73	FishDAC - 2013 West Scotland Deepwater Trawl Survey (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=012_Marine_Scotland_FishDAC_1312
74	FishDAC - Historic Fisheries - Sharks - Cetorhinidae	https://portal.medin.org.uk/portal/start.php#details?tpc=012_Marine_Scotland_FishDAC_956
75	Gardline Environmental Ltd, Teesside Offshore Wind Farm, Marine Mammal Observations and Passive Acoustic Monitoring (2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_aeb77834-01e8-4e91-9208-f737210aeaa9
76	Greater Wash Offshore Wind Farm Cumulative Effects: Scoping Study	https://portal.medin.org.uk/portal/start.php#details?tpc=015_4732b9fab6975bdf9b6a8b78f4552848
77	Happywhale - Atlantic White-sided Dolphin in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1937
78	Happywhale - Common Bottlenose Dolphin in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1947
79	Happywhale - Cuvier's Beaked Whale in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1733
80	Happywhale - Fin Whale in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1750
81	Happywhale - Harbour Porpoise in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1979
82	Happywhale - Humpback Whale in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1764
83	Happywhale - Killer Whale in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1717
84	Happywhale - Long-finned Pilot Whale in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1780
85	Happywhale - Minke Whale in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1741
86	Happywhale - Risso's Dolphin in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1706
87	Happywhale - Short-beaked Common Dolphin in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1697
88	Happywhale - Sperm Whale in North Atlantic Ocean	https://seamap.env.duke.edu/dataset/1726
89	Harbour porpoises, white-beaked dolphins and minke whales in North Sea - Boat surveys	https://seamap.env.duke.edu/dataset/425
90	Harbour porpoises, white-beaked dolphins and minke whales in North Sea - Land surveys	https://seamap.env.duke.edu/dataset/423

91	Hebridean Dolphin and Whale Trust killer whale sightings (1990-2006)	https://seamap.env.duke.edu/dataset/694
92	HiDef Aerial Surveying Ltd., Lincs Offshore Wind Farm, Aerial Ornithological Surveys (2013 - 2016)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_d8181392-58cf-444f-8a16-4b67cd0de50f
93	HiDef, Dogger Bank, Aerial Bird Surveys of Dogger Bank (Ultra High Resolution Video) (2012 - 2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_dbcc5a4e-2b70-450b-be16-d7c147f58c5f
94	High Resolution Video Survey of Seabirds and Mammals in the Norfolk Area	https://portal.medin.org.uk/portal/start.php#details?tpc=015_5da2ee1aaa5d5fc1a7c519b0e90bde1b
95	Historical distribution of whales shown by logbook records 1785-1913	https://seamap.env.duke.edu/dataset/885
96	HR Wallingford, Zone 6 Rampion, Underwater Noise Modelling (2016)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_c733e024a286b715ba90d6ab3bea5d4e
97	Hyder Consulting (UK) Ltd., Burbo Bank Offshore Wind Farm, Pre (2005 - 2006)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_5765faeb7e3b962b81dc59e65d633699
98	Identification of marine habitats relevant to Special Areas of Conservation (2004)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_8b68fcd5c9734859b1f4a3372a29e445
99	Inch Cape Offshore Limited, Environmental Statement (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_1fc61cfa3c45d3709295268296c4450c
100	Incidental sightings of marine mammals	https://seamap.env.duke.edu/dataset/103152572
101	Individual based data on Bottlenose dolphins in the Moray Firth Special Area of Conservation	https://portal.medin.org.uk/portal/start.php#details?tpc=010_85cbc7e5ca44f9586fdf11667b88794b
102	Institute of Estuarine and Coastal Studies, Humber Gateway Windfarm Seabird Survey Programme (2003 - 2008)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_bec9b2c2bd095ef593a48771f058ee23
103	Institute of Estuarine and Coastal Studies, Islay Offshore Wind Farm, Marine Mammals and Ornithology Survey (2010-2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_3eeb51572cb642b584cca02c5f709c85
104	Institute of Estuarine and Coastal Studies, Westermost Rough, Seabird Survey and Assessment Programme (2004 - 2006)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_a71dec812d53554f964cdf341682c30a
105	Isle of Man historical wildlife records 1995 to 1999	https://registry.nbnatlas.org/public/show/dr1614
106	Isle of Man wildlife records from 01/01/2000 to 13/02/2017	https://registry.nbnatlas.org/public/show/dr1148
107	Isle of Man wildlife records from 01/01/2000 to 13/02/2017	https://registry.nbnatlas.org/public/show/dr1148
108	Isle of Wight Notable Species	https://registry.nbnatlas.org/public/show/dr814
109	Isles of Scilly Seabird numbers and breeding success (2002 - 2004)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_1076670fe11ba6e27bd7781f1af919e3

110	JNCC European Seabirds at Sea	https://seamap.env.duke.edu/dataset/103152547
111	JNCC European Seabirds at Sea Seabird summer distribution / Seabird Winter distribution (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_968114002a1f4e612489c4c27ee79a59
112	JNCC seabird distribution and abundance data (all trips) from ESAS database	https://seamap.env.duke.edu/dataset/427
113	JNCC/MSS Solan Bank Reef Drop Down Video Survey (2014)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_04dc2dba1661e5f793936c7f0991ecb5
114	Machrihanish Seabird Observatory, Kintyre Offshore Wind Farm, ornithology survey (2009 - 2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_e32033f91fd053479c2af095e16cc732
115	Mammal Records (West Wales)	https://registry.nbnatlas.org/public/show/dr1486
116	Mammal records for Dumfries and Galloway	https://registry.nbnatlas.org/public/show/dr1398
117	Mammal records from Britain from the Atlas of Mammals (1993), with some subsequent records	https://registry.nbnatlas.org/public/show/dr743
118	Manx Basking Shark Watch 2017-2018	https://seamap.env.duke.edu/dataset/1512
119	Manx Biological Recording Partnership VERIFIED Isle of Man records between 14/02/2017 and 05/09/2019	https://registry.nbnatlas.org/public/show/dr2079
120	Marine Awareness North Wales, Wildlife Trust harbour porpoise baseline surveys on the north coast of Anglesey, Wales, UK	https://seamap.env.duke.edu/dataset/703
121	Marine Consents and Environment Unit, Burbo Bank Offshore Wind Farm, FEPA Licence (2003 - 2005)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_df1fcf7fbacf5b6b952d8903bda36b77
122	Marine Life Information Network (MarLIN) marine survey data (Professional)	https://seamap.env.duke.edu/dataset/103151986
123	Marine List of Ireland	https://obis.org/dataset/a97def2e-3a89-4a40-83c3-82ed18859ed1
124	Marine mammal monitoring from coastal sites in Cardigan Bay, UK, 2004-2009	https://seamap.env.duke.edu/dataset/716
125	Marine records from Pembrokeshire Marine Species Atlas	https://seamap.env.duke.edu/dataset/103152364
126	Marine Scotland Reports - Marine Environment - Scotland's Marine Atlas	https://marine.gov.scot/data-source-types/scotlands-marine-atlas
127	Marine Species Records from Skomer Marine Conservation Zone (MCZ) Marine Monitoring Programme	https://registry.nbnatlas.org/public/show/dr1253
128	Marinelife Cetacean sightings during survey work (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_cc700db4dd0776627a75e263fb3eb447

129	Ministry of Agriculture, Fisheries and Food (MAFF), Inner Dowsing Offshore Wind Farm, Food and Environment Protection Act (FEPA) 1985, Licence (2003)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_a08328a69aea56cfba1fa93f7cb8a087
130	Miscellaneous records held on the Cofnod database	https://registry.nbnatlas.org/public/show/dr1420
131	Monitoring of the effects of Belgian wind mill parks on benthic macro-invertebrates and the fish fauna of soft substrates	https://seamap.env.duke.edu/dataset/103152424
132	Monitoring the Consequences of the Northwestern North Sea Sandeel Fishery Closure	https://portal.medin.org.uk/portal/start.php#details?tpc=012_Marine_Scotland_FishDAC_1503
133	National Inventory of the Natural Heritage: Data from the air monitoring campaigns of marine megafauna (SAMM) in the French metropolitan area	https://seamap.env.duke.edu/dataset/103152508
134	National Mammal Atlas Project, online recording	https://registry.nbnatlas.org/public/show/dr671
135	National Trust for Scotland Species Records	https://registry.nbnatlas.org/public/show/dr681
136	Natural England (NE) Aerial Bird Surveys in the Outer Thames SPA (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_1029463f26a62ea9bf8ad0b5cae54aa6
137	Natural England (NE) Digital video aerial surveys of red throated diver in the Outer Thames Estuary Special Protection Area (2018)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_21a79c3c1de95de1a93456d3c9885a13
138	Natural England Marine Monitoring surveys	https://registry.nbnatlas.org/public/show/dr833
139	Natural Research Projects Limited, Lewis Wave Array, Baseline Bird Surveys (2010 - 2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_c6ba4c9e-7718-4dee-9cd5-c94b7f4fc7b6
140	Natural Research Projects Limited, Skerries Tidal Stream Array, Pre installation Seabird Pilot Study (2014)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_728a73b4-90af-4a25-982f-97c5485279ed
141	NatureScot Digital Aerial Survey of Moray Firth pSPA (2020)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_b0f9d873a3bbdf9f6382dcd1132e6b29
142	NatureScot Digitial Aerial Survey of Solway Firth SPA (2021)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_48a066f948a3e93eba3851ab9dd1b539
143	NBIS Records to December 2016	https://registry.nbnatlas.org/public/show/dr755
144	NE Scotland marine mammal records 1800-2010	https://registry.nbnatlas.org/public/show/dr1322
145	Non-avian taxa (BTO+partners)	https://registry.nbnatlas.org/public/show/dr529
146	Norfolk Offshore Wind Ltd., Cromer Offshore Wind Farm, Environmental Statement (2002)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_6e39e107-be82-4306-923b-4d7df366645c

147	Norfolk Offshore Windfarm Ltd., Cromer Offshore Wind Farm, FEPA Licence (2003)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_c79bbf6e807f524c85059bbffe88e612
148	Norman and Florence Hammond records. Seawatch and coastal survey records	https://seamap.env.duke.edu/dataset/103152569
149	North East Cetacean Project, Location of Harbour Porpoise Critical Habitat (2003 - 2009)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_23c6a897c7f163ec43a8416b5cb3b730
150	North East Cetacean Project, Location of Minke Whale / White Beaked dolphin critical habitat (2003 - 2009)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_f7a817d21c7ba442cb7607c3b177353d
151	Northern Ireland Marine Recorder Database - Collated Marine Species and Biotope Point Data	https://www.nmni.com/CEDaR/CEDaR-Information-service.aspx
152	Northern Seas Programme Data Set (2001-2007)	https://portal.medin.org.uk/portal/start.php#details?tpc=006_aa05e044fb015ff0acb36db74ce3edf6
153	NRW Regional Data: all taxa (sensitive species only), West Wales	https://registry.nbnatlas.org/public/show/dr1229
154	NRW Regional Data: North Wales	https://registry.nbnatlas.org/public/show/dr1248
155	NRW Regional Data: South East Wales Non-sensitive species	https://registry.nbnatlas.org/public/show/dr1935
156	Observatoire Pelagis - Reseau National Echouage (French stranding network) strandings 1934-2020	https://seamap.env.duke.edu/dataset/1406
157	Observatoire Pelagis aerial surveys 2002-2021	https://seamap.env.duke.edu/dataset/1404
158	Observatoire Pelagis boat surveys 2003-2021	https://seamap.env.duke.edu/dataset/1403
159	OCEAMM harbour porpoise sightings in the North Sea	https://seamap.env.duke.edu/dataset/779
160	Ongoing Marine Biological Association of the United Kingdom (MBA) UK volunteer collected Sealife Survey records (1998)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_331598372eda719709e2bf3aec25222a
161	OpenHydro and SSE, Brims Tidal Array, Environmental Statement (2016)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_3653d67f-37b0-4f56-8d97-beee2a3eaed8
162	Pembrokeshire Marine Species Atlas (PMSA)	https://portal.medin.org.uk/portal/start.php#details?tpc=007_4f4c4942-4343-5764-6473-303234333238
163	PIROP Northwest Atlantic 1965-1992	https://seamap.env.duke.edu/dataset/280
164	PMSS, North Hoyle Wind Farm, Annual FEPA Monitoring Report (2003 - 2004)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_4b880037910459b2a968b03fa3c9f148
165	Porpoise Feeding Areas	https://portal.medin.org.uk/portal/start.php#details?tpc=010_b9fb3949d22ec6cde9262c715951212e
166	present Whale and Dolphin Conservation (WDC) Shorewatch Programme records (2005)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_938b7aafb7f3d7bbab6e4d6f917be693

167	present, Robin Rigg, Monitoring Reports (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_90eb9f295456b6ea1dee3b92e52d1848
168	Purbeck Marine Wildlife Reserve, Kimmeridge Dorset, Marine Wildlife Recordings and Sightings (2005 - 2006)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_81e4c35004f4e33dff5570fd000a973f
169	RBA, Walney & West of Duddon Sands, Marine Mammals Impact Assessment (2004)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_6a5f3144-51d0-4e6f-ab2e-d489a3e687df
170	RECORD Mammal Data	https://registry.nbnatlas.org/public/show/dr1163
171	Robin Rigg Offshore Wind Farm, Bird and Marine Mammal Surveys (2001-2015)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_f54635d7cf20fa899b96d0e1ca51e46a
172	Robin Rigg Offshore Wind Farm, Environmental Statement & Aerial Survey	https://portal.medin.org.uk/portal/start.php#details?tpc=015_0670b6b08b674d839b9e3932b237d596
173	Royal Haskoning, Cumulative Ornithological Collision Risk Database (2017)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_f5117361-fadc-41ee-86e6-ea4a47382abd
174	Royal Society for the Protection of Birds, seabird colony locations within and adjacent to Net Gain regional project boundary (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_ec69178f7b5390000f240bd7a6eeac87
175	Royal Society for the Protection of Birds, Seabird mean foraging ranges within and adjacent to Net Gain regional project boundary (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_1227f480d1e1843cead2c73fbff4fc5f
176	RPS and ECON, Westermost Rough Offshore Wind Farm, Seabird Survey (2004 - 2006)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_5b1bd247fc5159fc9d72254f46fb03a7
177	RPS Argyll Array Offshore Wind Farm Basking Shark Survey (2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_753cc8cc-63a3-46f0-9477-96686610ce19
178	RPS, Bristol Channel Atlantic Array, Environmental Statement (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_717bfcc4c2b13dc67f5522b576960ee4
179	RPS, Bristol Channel, Atlantic Array Marine Mammal Survey (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_2782aaf8e09847308b94465a1ad62a3c
180	RPS, Humber Gateway Offshore Wind Farm, Marine Mammal Survey (2003 - 2005)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_9440c18aef4e54208bf68568115ac88d
181	RPS, Lincs Offshore Wind Farm, Boat based ornithological Monitoring (2011 - 2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_40377e1e-2736-464a-b28c-dc7bb6d695ff
182	RPS, Lincs Offshore Wind Farm, Boat based Ornithological Monitoring: Year 2 Construction Phase (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_1ed4e129-72e6-4345-bf01-69a52732aa16
183	RPS, Zone 8 Bristol Channel Atlantic Array, Underwater Noise Monitoring (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_ae44b489c6344980a5531791b54c7b28
184	SCANS I cetacean sightings 1994	https://seamap.env.duke.edu/dataset/1183

185	SCANS II cetacean sightings from aerial surveys 2005	https://seamap.env.duke.edu/dataset/1153
186	SCANS II cetacean sightings on primary platform of vessel surveys 2005	https://seamap.env.duke.edu/dataset/1150
187	SCANS II cetacean sightings on tracker platform of vessel surveys 2005	https://seamap.env.duke.edu/dataset/1152
188	SEA3 Technical report Marine mammals, North Sea (Addendum to SEA2 report) (2002)	https://portal.medin.org.uk/portal/start.php#details?tpc=004_aba64100-c119-4de3-e044-0003ba6f30bd
189	SEA4 seabird and marine mammal observations onboard R.V. Kommandor Jack, during a multibeam survey of the seabed in the Faroe (2002)	https://portal.medin.org.uk/portal/start.php#details?tpc=004_aba64100-c178-4de3-e044-0003ba6f30bd
190	SEA4 Technical report Marine mammals (north UKCS) (2003)	https://portal.medin.org.uk/portal/start.php#details?tpc=004_aba64100-c124-4de3-e044-0003ba6f30bd
191	SEA5 Technical Report Marine mammals North Sea (2004)	https://portal.medin.org.uk/portal/start.php#details?tpc=004_aba64100-c178-4de3-e044-0003ba6f30bd
192	SEA6 Technical Report Sea Mammal Research Background Information on Marine Mammals (Irish Sea) (2005)	https://portal.medin.org.uk/portal/start.php#details?tpc=004_aba64100-c14f-4de3-e044-0003ba6f30bd
193	SEA7 Marine Mammal Observations Kommandor Jack, July August 2005 (NE Atlantic west of Scotland) (2005)	https://portal.medin.org.uk/portal/start.php#details?tpc=004_aba64100-c17c-4de3-e044-0003ba6f30bd
194	SEA7 seabird and marine mammal observations onboard the RSS Charles Darwin during British Geological Survey cruise CD174 in the Rockall Trough (2005)	https://portal.medin.org.uk/portal/start.php#details?tpc=004_aba64100-c17d-4de3-e044-0003ba6f30bd
195	Seabird Colony Surveillance along Castlemartin Coastline, South Pembrokeshire	https://portal.medin.org.uk/portal/start.php#details?tpc=007_4f4c4942-4343-5764-6473-303436313336
196	Seabirds at Sea Evidence Base	https://portal.medin.org.uk/portal/start.php#details?tpc=007_c7ec3e4f-f697-46b9-a300-70db95b57bbf
197	SEAMARCO, Acoustic Mitigation Devices (AMDs) to Deter Marine Mammals from Pile Driving Areas at Sea (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_10301a75-5bbd-4c5a-bb05-d794ced6540d
198	SeaScape Energy Ltd., Burbo Bank Offshore Wind Farm, Environmental Statement	https://portal.medin.org.uk/portal/start.php#details?tpc=015_28f4db2c76f7531dbbbd90f02230543a
199	Seasearch Marine Surveys in England	https://registry.nbnatlas.org/public/show/dr690
200	Seasearch Marine Surveys in Ireland	https://registry.nbnatlas.org/public/show/dr693
201	Seasearch Marine Surveys in the Isle of Man	https://registry.nbnatlas.org/public/show/dr663
202	Seatrust Cetacean Records West Wales	https://registry.nbnatlas.org/public/show/dr1487

203	SEWBReC Mammals (South East Wales)	https://registry.nbnatlas.org/public/show/dr807
204	Shetland Oil Spill Sensitivity Maps_Black Guillemot (2016)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_1d7b9c928d952de467ca96772f4b1534
205	SMRU Ltd, Argyll Array Windfarm Basking Shark Draft (FINAL) Chapter for Environmental Statement (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_995cc5cc-63a3-46f0-9477-96686610ce19
206	SMRU Ltd., SeaGen Strangford Lough, Monitoring of Harbour Porpoise in Strangford Lough with focus on the Narrows (2006-2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_d97e350d-0995-4bc3-bbe3-6a2c298eb8c0
207	SMRU Ltd., SeaGen Strangford Lough, SeaGen Noise Biological Assessment (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_3eeb51572cb642b584cca02c5f709c85
208	SMRU Ltd., SeaGen Strangford Lough, Sonar Monitoring of Marine Mammals around SeaGen (2008-2009)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_3b6752fa-ab0a-4762-8533-e8d5589c1842
209	SMRU Ltdorigi, Islay Offshore Wind Farm, West Coast Wind Farms Marine Mammal Desk study (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_662cf2e4-9d73-45a5-965e-2e5aa8091ffc
210	SMRU Marine, Skerries Tidal Stream Array, Marine Mammal Monitoring: Land Based Vantage Point Survey Trial (2014)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_fb849137-1773-41d5-ab7d-7db16a4ae896
211	Stackpole National Nature Reserve Species Inventory and Ad-hoc Sightings from Across Pembrokeshire	https://registry.nbnatlas.org/public/show/dr1312
212	Subacoustech Ltd., Measurements of Underwater Noise Generated by Acoustic Mitigation Devices (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_011ccca2-a749-4c34-9ad3-97dfa4f57036
213	Suffolk Biodiversity Information Service (SBIS) Dataset	https://registry.nbnatlas.org/public/show/dr867
214	TCE, Robin Rigg, Aerial Bird Survey Data (2002 - 2004)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_0670b6b08b674d839b9e3932b237d596
215	TEL Ltd., Ramsey Sound, Marine Mammal Observations (2007-2015)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_acbc05f9-bee9-4ed3-93bd-5c9ba57dd645
216	The Belgian Marine Mammals Database (BMM)	https://seamap.env.duke.edu/dataset/103152589
217	The East Coast Marine Mammal Acoustic Study (ECOMASS)	https://portal.medin.org.uk/portal/start.php#details?tpc=012_Marine_Scotland_FishDAC_1969
218	The Food and Environment Research Agency, FERA, Skerries Tidal Site, Seabird Survey (2009)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_efa589c4-b45f-426b-946e-db414848ab28
219	The UK Archive for Marine Species and Habitats Data	https://seamap.env.duke.edu/dataset/103152350

220	UK Cetacean Strandings Investigation Programme (CSIP)- strandings around the UK coast (1990-present)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_815ecbf4af7e62e3fc2a507ae0bb848a
221	UK Royal Navy Marine Mammal Observations	https://seamap.env.duke.edu/dataset/64
222	United Kingdom Integrated Marine Observing Network (UK-IMON) initiative ocean Autumn 2013 glider deployments in the Celtic Sea.	https://portal.medin.org.uk/portal/start.php#details?tpc=006_1b063e666b93570bb97b3445db276383
223	United Kingdom National Whale Stranding Database 1913-2008	https://seamap.env.duke.edu/dataset/731
224	Vertebrates Outer Hebrides	https://registry.nbnatlas.org/public/show/dr797
225	Visual sightings data set 2003-2020	https://registry.nbnatlas.org/public/show/dr537
226	Visual sightings from Song of the Whale 1993-2013	https://seamap.env.duke.edu/dataset/1158
227	Walney Offshore Wind Farm, Year 1 Post-construction Monitoring (2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_ba76603e-a803-4984-b426-03bd6fbc9e9c
228	Walney Offshore Wind Farm, Year 2 & Year 3 Post-construction Monitoring (2013)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_73421273-30fc-4d3f-824d-6d71537fa223
229	WDC Lewis survey data	https://registry.nbnatlas.org/public/show/dr2027
230	WDC Shorewatch Sightings	https://registry.nbnatlas.org/public/show/dr911
231	WDC, Liverpool Bay UK, Whale and Dolphin Critical Habitats (2011)	https://portal.medin.org.uk/portal/start.php#details?tpc=010_6c7b9ae291dbf18aea19d04f0c617419
232	Westermost Rough Ornithological Technical EIA Report	https://portal.medin.org.uk/portal/start.php#details?tpc=015_6343090f0e635fe9a229ca33696bf4bb
233	Wintering Birds Oct - March_oil spill response	https://portal.medin.org.uk/portal/start.php#details?tpc=010_ecec15970f04441927e300fd78b3dbb3
234	WTSWW Data: All Taxa (West Wales)	https://registry.nbnatlas.org/public/show/dr1241
235	WWT Consulting, Lincs Offshore Wind Farm, Aerial Survey for Waterbirds and Seabirds (2007 - 2008)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_ce83c483-2198-4ab2-9f23-7909143e117f
236	WWT Consulting, Lincs Offshore Wind Farm, Aerial Surveys for Waterbirds and Seabirds (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_007e7ef8-c2d2-4efa-81e0-3e9ce3176cd9
237	WWT Consulting, Lynn and Inner Dowsing Offshore Wind Farm, Aerial Survey for Waterbirds and Seabirds in GW4 (2007 - 2008)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_d36e0748-b674-40c0-8739-35488408452a
238	WWT Consulting, Lynn and Inner Dowsing Offshore Wind Farms, Aerial Surveys for Waterbirds and Seabirds in GW4 (2010)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_5db2c7fbda994ab9876b803c03885f33

239	WWT, Bristol Channel Atlantic Array, Aerial Surveys for Waterbirds and Seabirds in the South West of England and Wales: 2007 Final Report (2007)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_3416a5f72f2580780dabd4266d6fb726
240	Yorkshire Naturalists Union Marine and Coastal Section Records	https://registry.nbnatlas.org/public/show/dr871
241	Zone 3 Dogger Bank Creyke Beck, DMP Statistical Soutions Ltd, Hi-Def Aerial Marine Mammal Survey (2009-2011)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_8be80558-8515-4920-a464-ca19f6f832e8
242	Zone 6 Rampion, NaturalPower, Bird and Mammal Survey (2010 - 2011)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_d71898d1-1165-4346-b98e-48f2470e5100
243	Zone 6 Rampion, SeaRoc, Bird and Mammal Survey (2010 - 2011)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_d1ee34f077861b18b148eacff208e219
244	Zone 8 Bristol Channel Atlantic Array, RPS Nocturnal Bird Surveys (2010 - 2011)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_315f06e675a13c52658abf516085a8da
245	Zone 9 Celtic Array, Hidef Aerial Bird Survey (2012)	https://portal.medin.org.uk/portal/start.php#details?tpc=015_8c3fbd3b-840d-442b-b196-f53322316eef
246	Zone 9 Celtic Array, Stage 2 PEI Chapter 11 (Marine Mammals, Basking Sharks and Turtles) and Chapter 12 Ornithology (2013)	https://https://portal.medin.org.uk/portal/start.php#details?tpc=009_CEFASfea6bfb9-3b7f-4a6a-ad4d-3d7f2923f138portal.medin.org.uk/portal/start.php#details?tpc=015_f2a4226c-90a8-4b4b-adc9-7820fd049f61

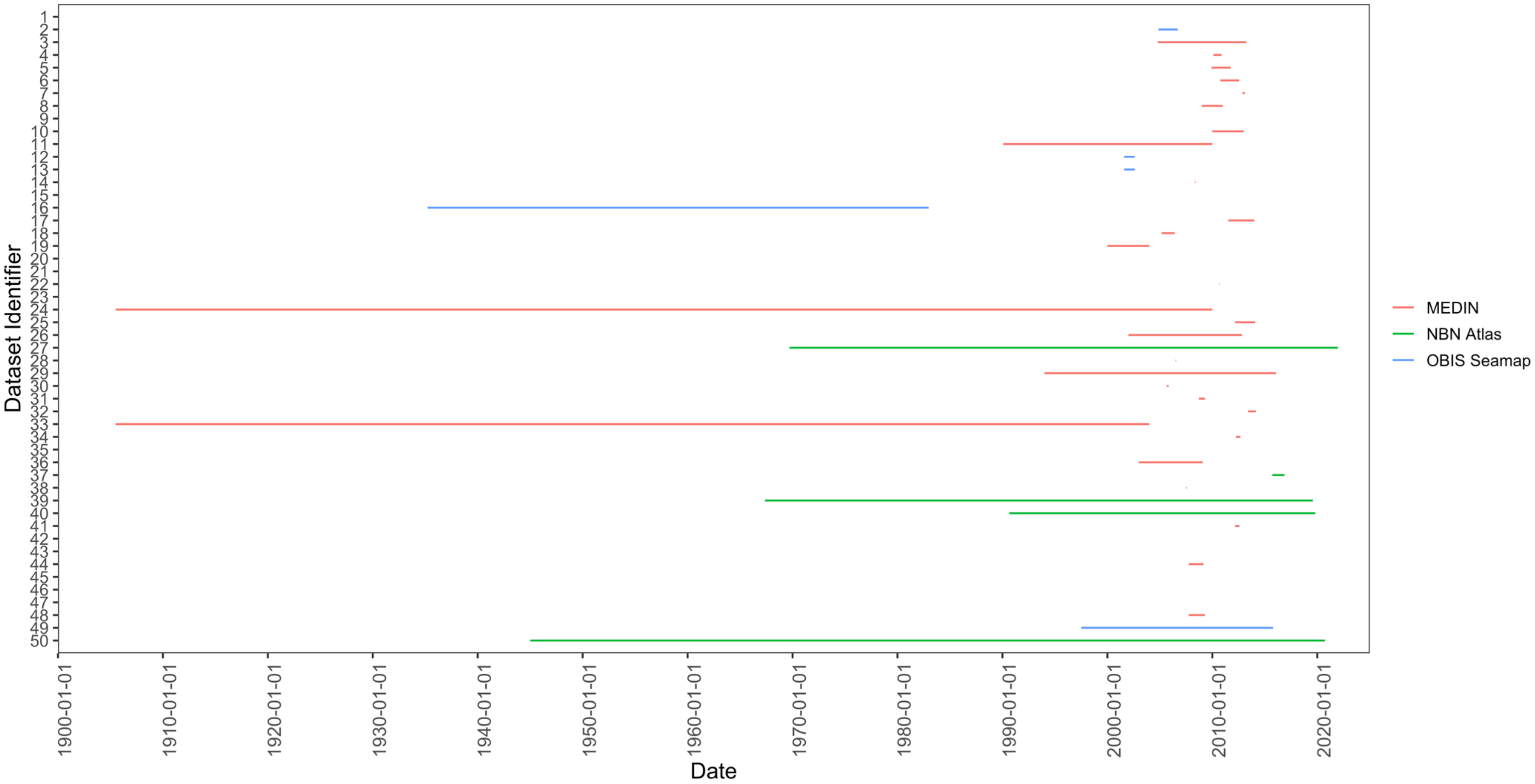


Figure A.2: Gantt chart of datasets 1 to 50 obtained from databases (Medin, NBN Atlas and OBIS Seamap), demonstrating temporal coverage for all datasets from 1900 onwards.

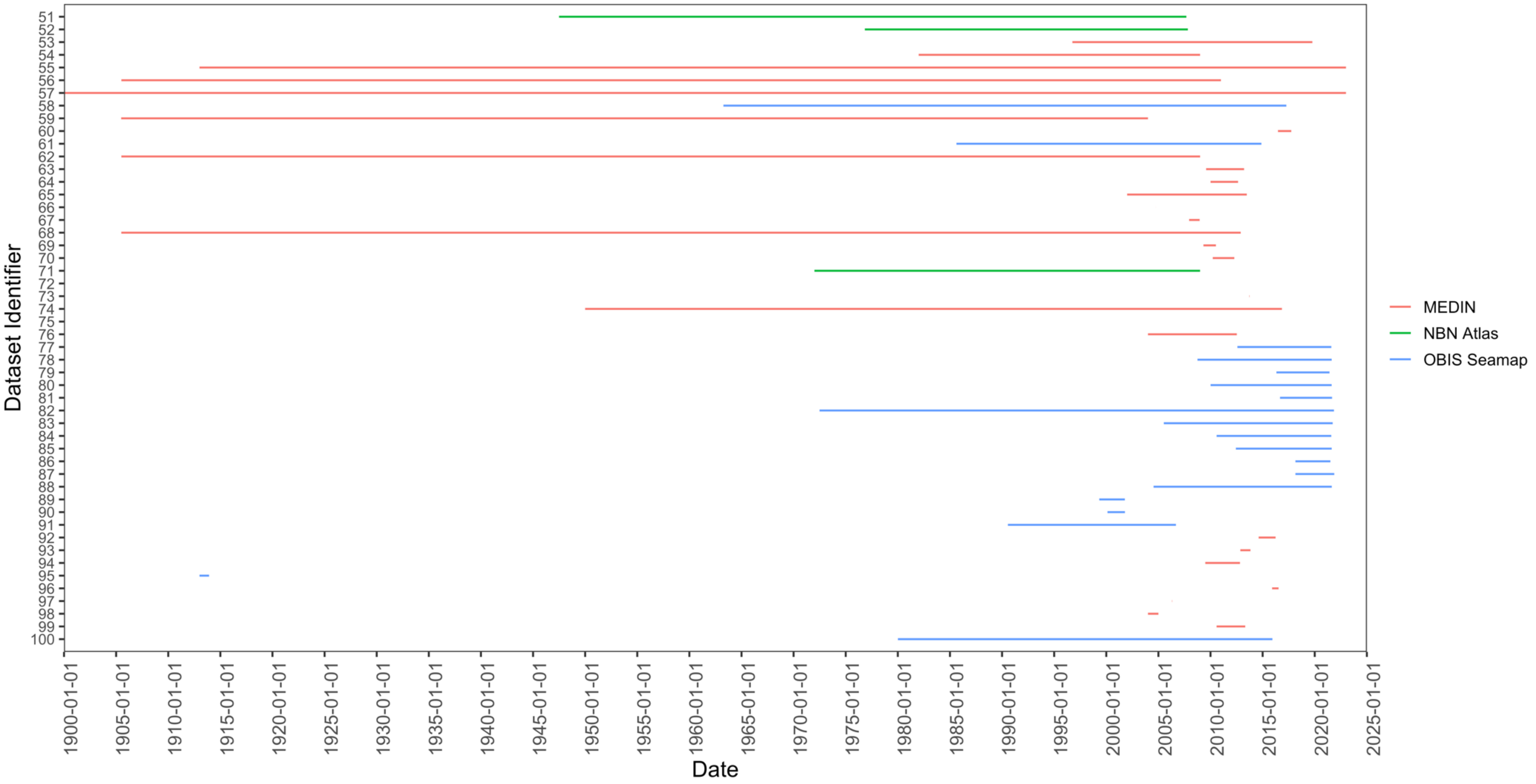


Figure A.3: Gantt chart of datasets 50 to 100 obtained from databases (Medin, NBN Atlas and OBIS Seamap), demonstrating temporal coverage for all datasets from 1400 onwards.

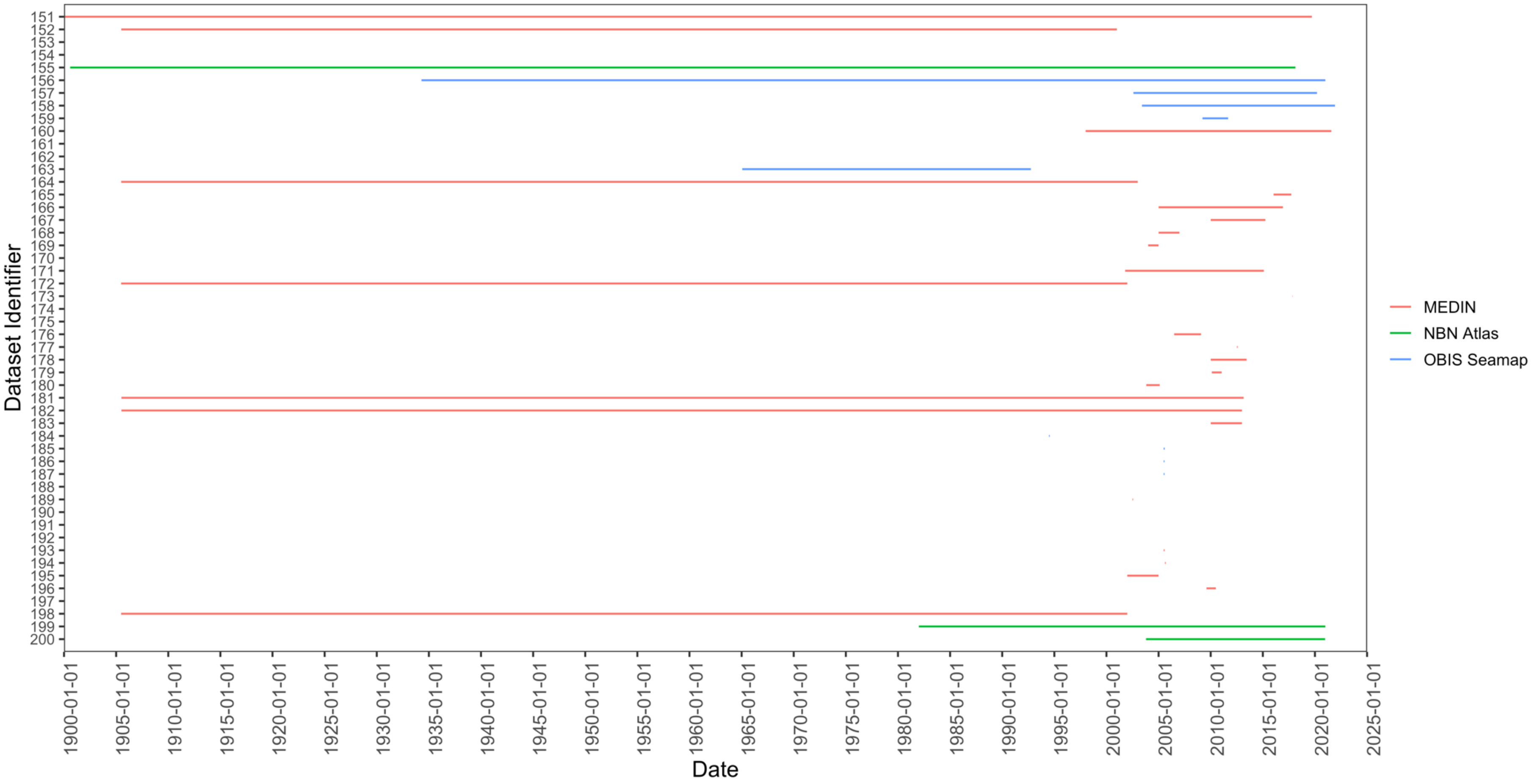


Figure A.4: Gantt chart of datasets 151 to 200 obtained from databases (Medin, NBN Atlas and OBIS Seamap), demonstrating temporal coverage for all datasets from 1400 onwards.

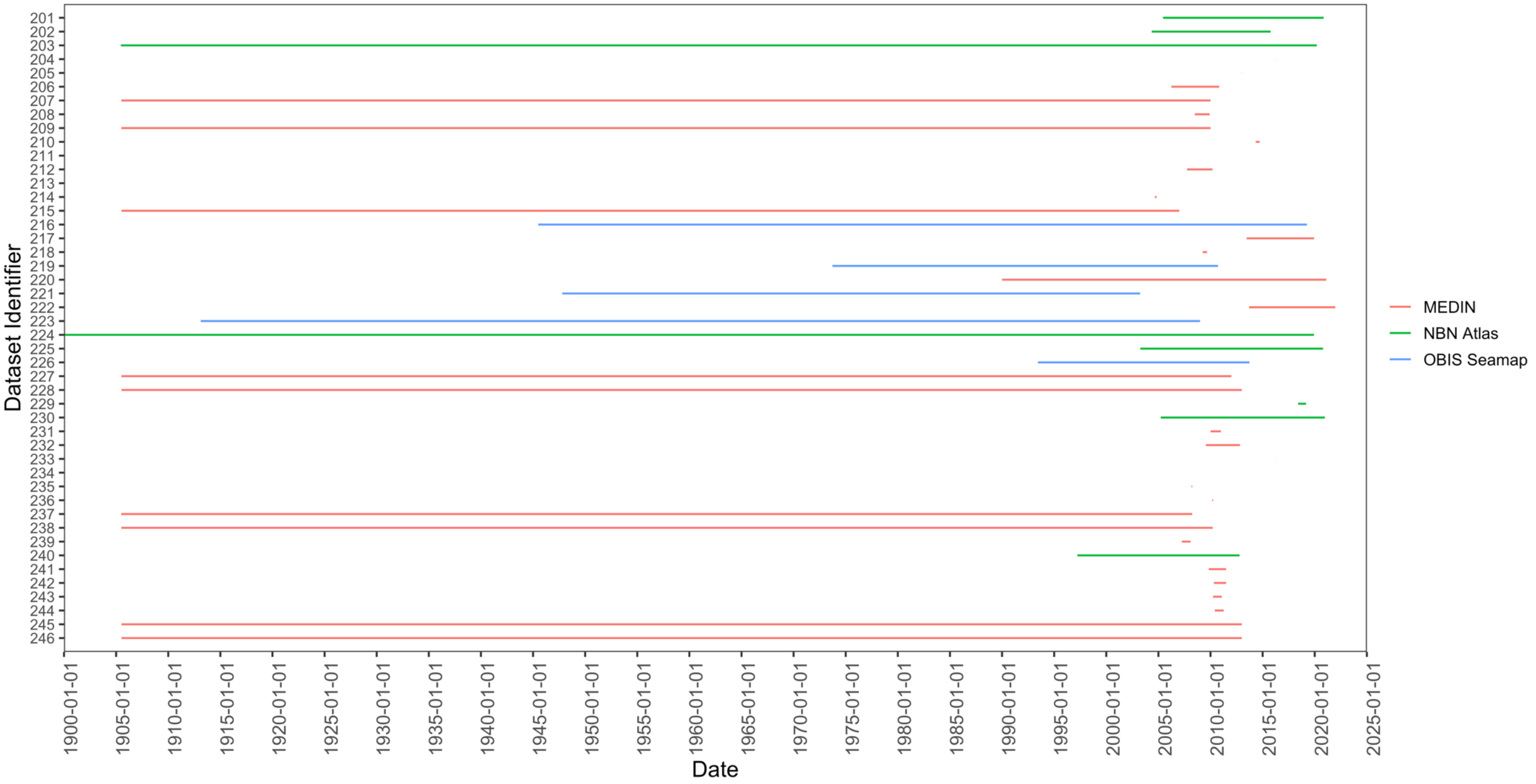


Figure A.5: Gantt chart of datasets 200 to 249 obtained from databases (Medin, NBN Atlas and OBIS Seamap), demonstrating temporal coverage for all datasets from 1900.

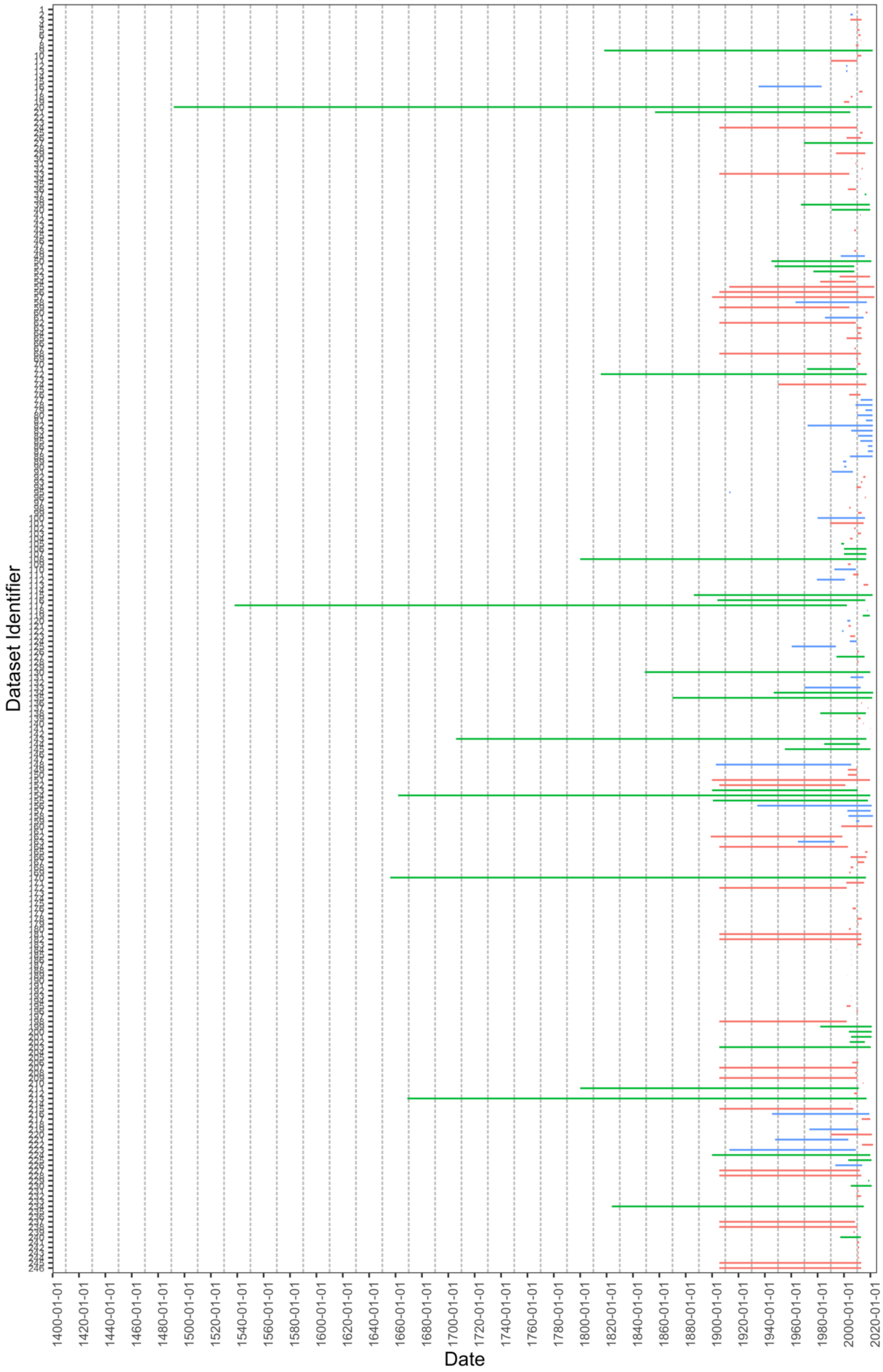


Figure A.6: Gantt of full datasets, from 1471 to 2022.

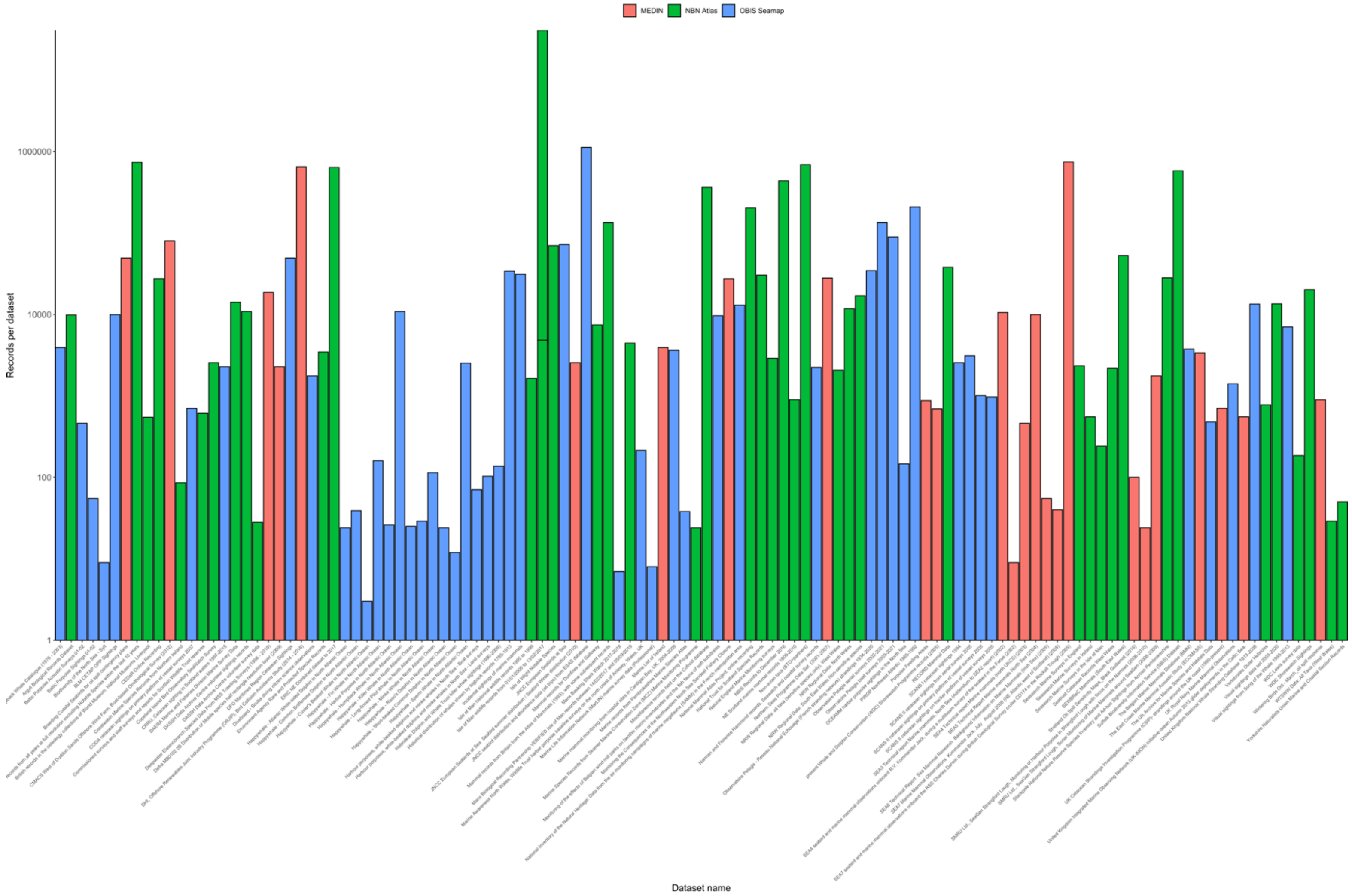


Figure A.7 Number of records per dataset, on log scale, for those datasets that were freely accessible online.

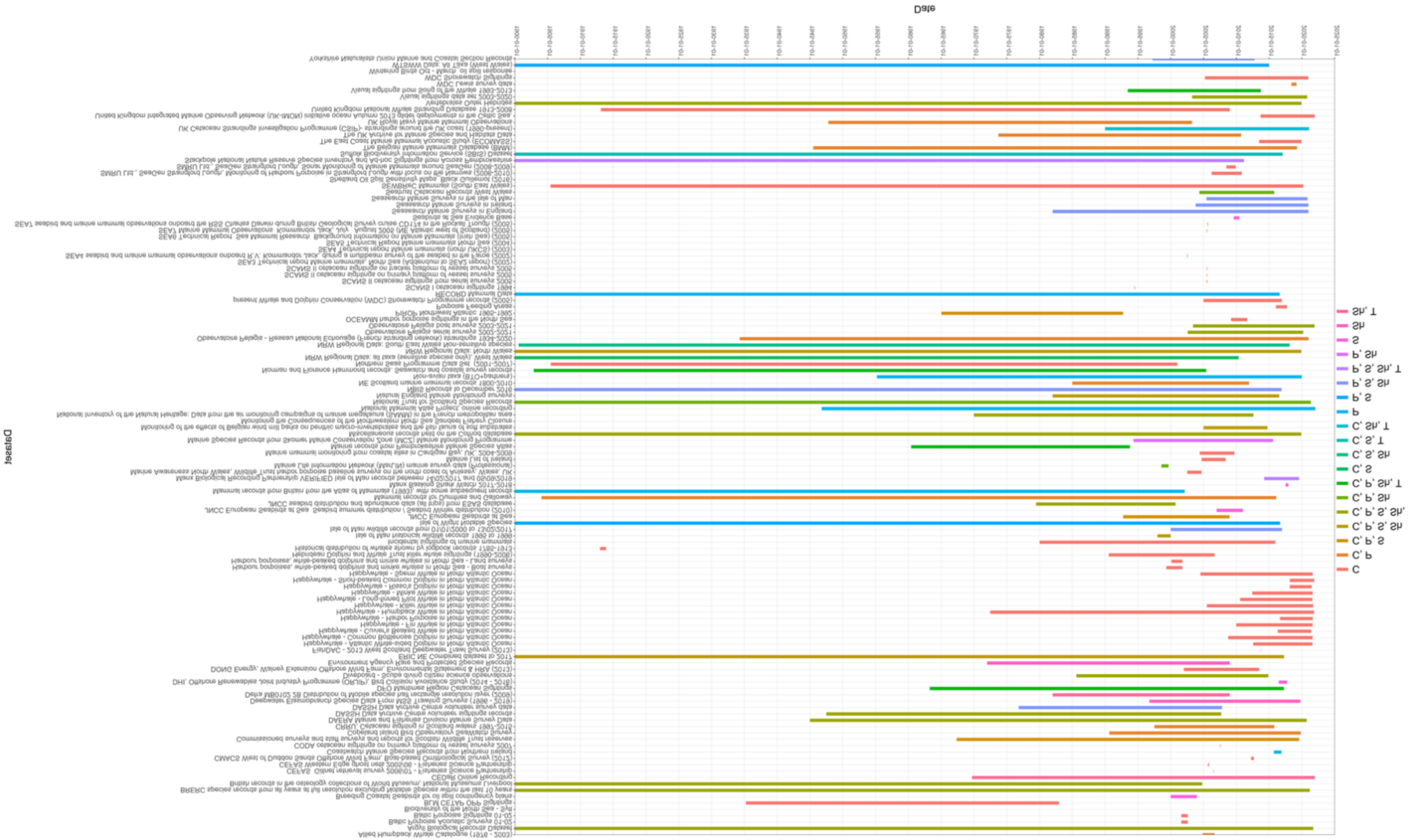


Figure A.8: Gannt of datasets from 1900 onwards with codes for taxa contained in dataset. C = Cetacean , P = Pinniped, S = Seabird, Sh = Shark, T = Turtle.

Appendix B Species list for thesis

Table B-1 – Species list for species in the ESAS and ORCA database

Taxa	Species name	Scientific Name	IUCN Status
Seabird	Arctic skua	*Stercorarius parasiticus*	Least concern
	Arctic tern	*Sterna paradisaea*	Least concern
	Atlantic puffin	*Fratercula arctica*	Vulnerable
	Black guillemot	*Cepphus grylle*	Least concern
	Black legged kittiwake	*Rissa tridactyla*	Vulnerable
	Black-headed gull	*Chroicocephalus ridibundus*	Least concern
	Common Guillemot	*Uria aalge*	Least concern
	Common gull	*Larus canus*	Least Concern
	Common tern	*Sterna hirundo*	Least concern
	European herring gull	*Larus argentatus*	Least concern
	European shag	*Phalacrocorax aristotelis*	Least concern
	European Storm-petrel	*Hydrobates pelagicus*	Least concern
	Great black-backed gull	*Larus marinus*	Least concern
	Great cormorant	*Phalacrocorax carbo*	Least concern
	Great shearwater	*Ardenna gravis*	Least concern
	Great skua	*Stercorarius skua / Catharacta skua*	Least concern
	Grey phalarope	*Phalaropus fulicarius*	Least concern
	Iceland gull	*Larus glaucoides*	Least concern
	Leach's Storm-petrel	*Oceanodroma leucorhoa*	Vulnerable
	Lesser black-backed gull	*Larus fuscus*	Least concern
	Little auk	*Alle alle*	Least concern
	Little gull	*Larus minutus / Hydrocoloeus minutus*	Least concern
	Little shearwater	*Puffinus assimilis*	Least concern
	Little tern	*Sternula albifrons*	Least concern
	Long-tailed skua	*Stercorarius longicaudus*	Least Concern
	Manx shearwater	*Puffinus puffinus*	Least Concern

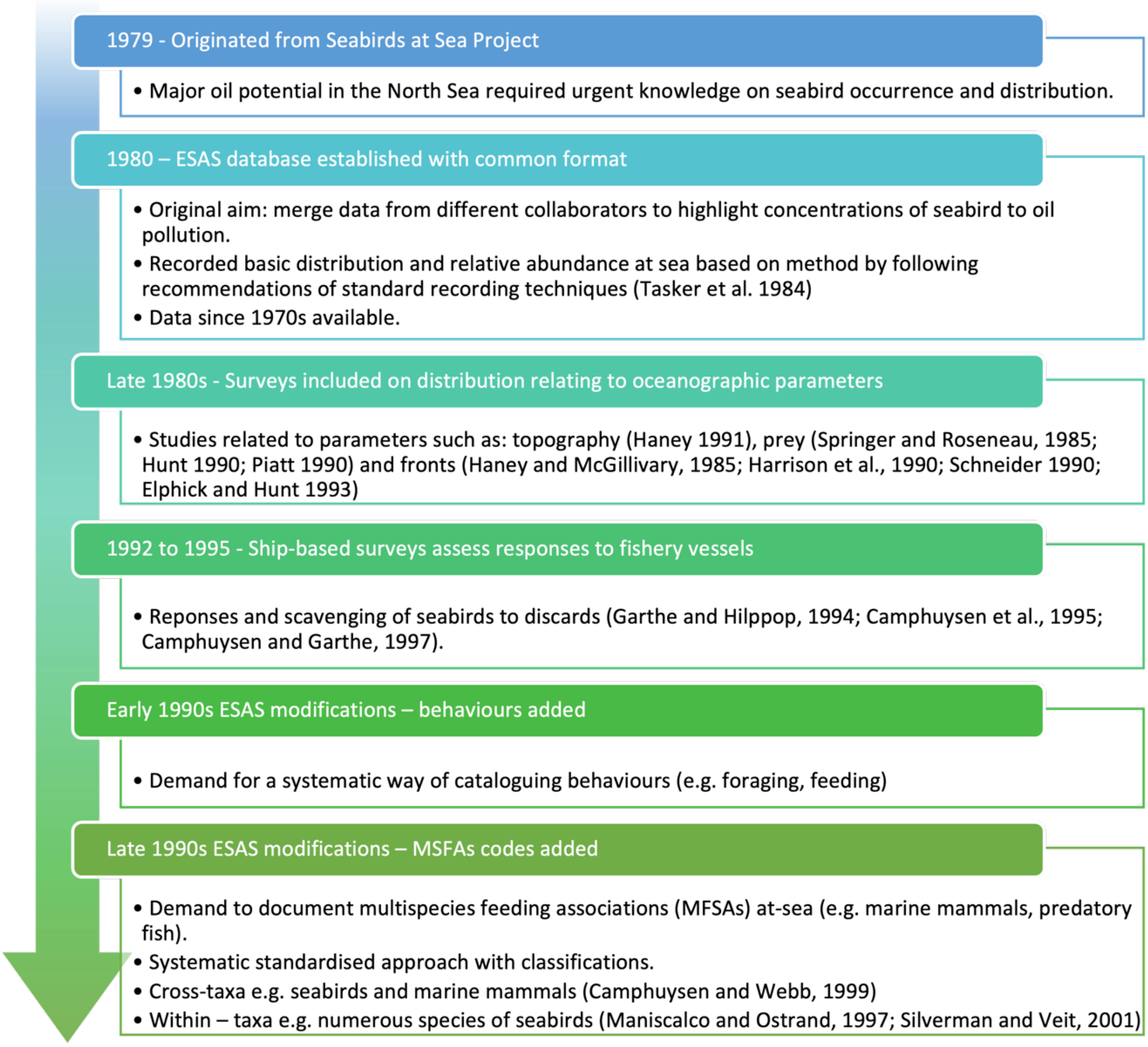
	Mediterranean gull	*Larus melanocephalus*	Least concern
	Northern Fulmar	*Fulmarus glacialis*	Least concern
	Northern gannet	*Morus bassanus*	Least concern
	Pomarine skua	*Stercorarius pomarinus*	Least concern
	Razorbill	*Alca torda*	Near Threatened
	Ring-billed gull	*Larus delawarensis*	Least concern
	Roseate tern	*Sterna dougallii*	Least concern
	Ross's gull	*Rhodostethia rosea*	Least concern
	Sabine's gull	*Xema sabini*	Least concern
	Sandwich tern	*Sterna sandvicensis*	Least concern
	Sooty Shearwater	*Ardenna grisea*	Near Threatened
	Wilson's storm petrel	*Oceanites oceanicus*	Least concern
	Yelkouan shearwater	*Puffinus yelkouan*	Vulnerable
	Yellow-legged gull	*Larus michahellis*	Least concern
Cetacean	Atlantic spotted dolphin	*Stenella frontalis*	Least concern
	Atlantic White-sided dolphin	*Lagenorhynchus acutus*	Least concern
	Blue whale	*Balaenoptera musculus*	Endangered
	Common bottlenose dolphin	*Tursiops truncatus*	Least concern
	Cuvier's beaked whale	*Ziphius cavirostris*	Least concern
	False killer whale	*Pseudorca crassidens*	Near Threatened
	Fin whale	*Balaenoptera physalus*	Vulnerable
	Harbour porpoise	*Phocoena phocoena*	Least concern
	Humpback whale	*Megaptera novaeangliae*	Least concern
	Killer whale	*Orcinus orca*	Data deficient
	Long-finned pilot whale	*Globicephala melas*	Least concern
	North Atlantic right whale	*Eubalaena glacialis*	Critically endangered
	Northern bottlenose whale	*Hyperoodon ampullatus*	Data deficient
	Northern minke whale	*Balaenoptera acutorostrata*	Least concern
	Risso's dolphin	*Grampus griseus*	Least concern
	Rough-toothed dolphin	*Steno bredanensis*	Least concern

	Sei whale	*Balaenoptera borealis*	Endangered
	Short-beaked common dolphin	*Delphinus delphis*	Least concern
	Sowerby's whale	*Mesoplodon bidens*	Least concern
	Sperm whale	*Physeter macrocephalus*	Vulnerable
	Striped dolphin	*Stenella coeruleoalba*	Least concern
	True's beaked whale	*Mesoplodon mirus*	Unknown
	White-beaked dolphin	*Lagenorhynchus albirostris*	Least concern
Fish	Ocean sunfish	*Mola mola*	Vulnerable
Pinniped	Grey seal	*Halichoerus grypus*	Least concern
	Harbour seal	*Phoca vitulina*	Least concern
	Harp seal	*Pagophilus groenlandicus*	Least concern
	Hooded seal	*Cystophora cristata*	Vulnerable
Shark	Basking shark	*Cetorhinus maximus*	Endangered
	Blue shark	*Prionace glauca*	Near Threatened
	Mako shark	*Isurus oxyrinchus*	Endangered
	Porbeagle	*Lamna nasus*	Vulnerable
Turtle	Green turtle	*Chelonia mydas*	Endangered
	Loggerhead turtle	*Caretta caretta*	Vulnerable

Appendix C Detailed methodologies for ESAS and ORCA

ESAS

The ESAS database is managed by JNCC, comprises data from boat and aerial surveys carried out from 1979 to 2011 (current database v5), and contains over three million records of seabirds, cetaceans, pinnipeds, and other marine megafauna from Northwest European and North Atlantic waters. Ship and aircraft surveys are a vital component of seabird (Hunt, 1990, Certain and Bretagnolle, 2008, Ainley *et al.*, 2009, Hunt Jr *et al.*, 2018, Phillips *et al.*, 2019), marine mammal (Berrow, 2001), pinniped (Southwell *et al.*, 2008) and other marine predator research. The ESAS database compiles data from numerous institutes for the entire North Sea, with data based on standardised survey methods (Tasker *et al.*, 1984) and merged in a common database administered by JNCC (Busch *et al.*, 2013). The original underpinning database remains intact, but the extent of information stored within it has evolved over time to accommodate supplementary information (such as foraging behaviour and multi-species feeding associations) and expand in scope (Figure C.1).



1979 - Originated from Seabirds at Sea Project

- Major oil potential in the North Sea required urgent knowledge on seabird occurrence and distribution.

1980 – ESAS database established with common format

- Original aim: merge data from different collaborators to highlight concentrations of seabird to oil pollution.
- Recorded basic distribution and relative abundance at sea based on method by following recommendations of standard recording techniques (Tasker et al. 1984)
- Data since 1970s available.

Late 1980s - Surveys included on distribution relating to oceanographic parameters

- Studies related to parameters such as: topography (Haney 1991), prey (Springer and Roseneau, 1985; Hunt 1990; Piatt 1990) and fronts (Haney and McGillivary, 1985; Harrison et al., 1990; Schneider 1990; Elphick and Hunt 1993)

1992 to 1995 - Ship-based surveys assess responses to fishery vessels

- Reponses and scavenging of seabirds to discards (Garthe and Hilppop, 1994; Camphuysen et al., 1995; Camphuysen and Garthe, 1997).

Early 1990s ESAS modifications – behaviours added

- Demand for a systematic way of cataloguing behaviours (e.g. foraging, feeding)

Late 1990s ESAS modifications – MSFAs codes added

- Demand to document multispecies feeding associations (MFSAs) at-sea (e.g. marine mammals, predatory fish).
- Systematic standardised approach with classifications.
- Cross-taxa e.g. seabirds and marine mammals (Camphuysen and Webb, 1999)
- Within – taxa e.g. numerous species of seabirds (Maniscalco and Ostrand, 1997; Silverman and Veit, 2001)

Figure C.1: Graphic showing data evolution over time of ESAS dataset.

The ESAS database incorporates many data sources and thus a standardised survey method is crucial (Tasker *et al.*, 1984, Camphuysen and Garthe, 2004) with a set ESAS survey method and coding system (JNCC, 2020). Only data collected by accredited ESAS surveyors are included in the ESAS database. It is designed to gather several different parameters simultaneously during seabird and cetacean sightings surveys, with four recording protocols - one for marine mammals, fish, and reptiles and three for birds (‘birds on the water’, ‘birds in flight’, ‘birds in snapshot’). The recommended set-up is a 90-degree angle of view to be selected from bow to beam (**Error! Reference source not found.**Figure C.2– Panel A), and then a 300-metre-wide strip transect of birds in the water is recorded, using a distance band method (**Error! Reference source not found.** Figure C.2– Panel B) at set time intervals. Birds in flight within the 90° area are recorded, but not placed into distance bands. The ‘birds in snapshot’ method uses a series of instantaneous counts that can be used to derive the total number of birds in flight in a given area (i.e. a density of birds in flight) at a given time (further detailed method in JNCC (2020)). For marine mammals and other species (e.g. basking sharks, sunfish etc.) the parameters described for collecting bird data are used (i.e. species, number, distance band, direction of travel etc.) but. A 180° or 90° forward scan operated at the same time is sometimes present to record scarcer seabirds, which whilst not viable for density calculations can provide information on multi-species associations or marine mammal presence. There have been several method improvements over time, such as new survey platforms and two observers and binoculars replacing one observer as in the original method. Shorter time intervals have also been introduced (1 and 5 minutes), where it was originally 10-minute intervals.

The ESAS database file contains base component information (such as base type, date, time, place, observers, and environmental conditions) and species information (species observed, count type). Additionally, association codes (A-codes) can be provided which gives certain surface phenomena details, such as whether the animal is heading for a fishing boat or offshore platform (wood, litter, seaweed) or associated with a colony or land. Foraging behaviour codes are also provided (such as actively searching, surface pecking) and general behaviour codes such as sleeping, preening and courtship. Marine mammal behaviour codes also included where relevant – such as swimming mode (fast / slow / escape behaviour), aerial displays (breaching), or other types (basking, spy hopping, tail / flipper flapping). Other data on death codes (misfortune, disease) and prey codes are also present.

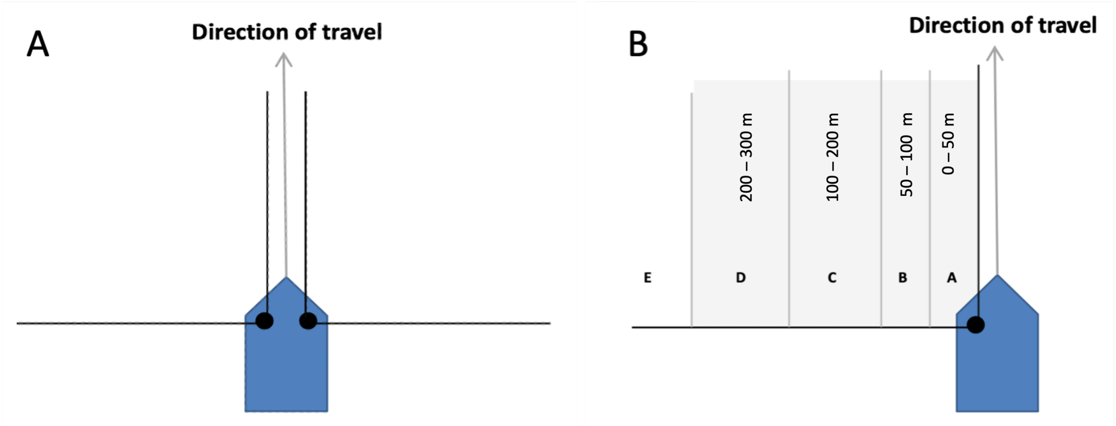


Figure C.2: Diagram from JNCC (2020) illustrating A) recommended viewing angles for ESAS survey methodology, where the black circles represent the observer and the black lines the recommended angle of view. B) demonstration of distance bands to the sea surface and delineation of the transect, with Band A – 0 to 50 m, Band B – 50 to 100 m, Band C – 100 to 200 m, Band D – 200 to 300 m, Band E – anything further than 300 m.

ORCA dataset

ORCA is a citizen science charity, and the ORCA dataset is a cetacean-only opportunistic dataset using research platforms on cruise and ferry routes around the UK and North-Eastern Atlantic since the mid 1990s (Matear *et al.*, 2019). Observation data is collected over extended periods of time and is a low-cost method of obtaining sighting data, as the cost of dedicated surveys on chartered research vessels is generally expensive if wanting regular repeat surveys. These platforms of opportunity operate all year round, broadening the temporal scale of surveying but is however limited by other factors such as survey design - e.g. routes cannot be adapted according to surveyors needs (Kiszka *et al.*, 2007, Robbins *et al.*, 2019) and areas of more coverage (ships passing more frequently, such as ferry routes) may give higher densities than traditional survey methods (Ban *et al.*, 2013, Harvey *et al.*, 2018). However, a recent study has suggested ORCA data has exhibited similar trends to data collected by professional scientists (Robbins *et al.*, 2019).

Like ESAS surveyor training, ORCA marine mammal surveyors complete a training course to learn ID skills, survey protocol and conduct on ships. Surveyors can range in technical backgrounds, careers, and ages but teams generally comprise three experienced surveyors and one less experienced to aid development. Trained surveyors are deployed on ferries or cruise ships. Ferry routes span the English Channel, Bay of

Biscay, Celtic Sea, Irish Sea and North Sea (Figure C.3**Error! Reference source not found.**), whilst cruise routes span a larger spatial range around Europe and across the wider Atlantic and Arctic waters (Figure C.4) (ORCA, 2020). From 2014, all ORCA cruise surveys applied a different survey protocol, the ORCA Observation Protocol (detailed below, ORCA (2017).

Teams consist of four members: two observers, one recorder and one resting, with individuals rotating roles every 30 minutes to avoid observer fatigue. Surveyors collect data from the forward-facing bridge of vessels according to standard distance sampling methodologies (Buckland *et al.*, 2001, Buckland *et al.*, 2015) (Figure C.5) The two observers scan an area of 100°, with the starboard observer scanning an area from 90° off the starboard side to 10° to port side, and port observer will scan an area from 90° off the port side to 10° to starboard. Both use reticle binoculars to scan distant waters for animals between periods of searching with the naked eye, arcing from side to side and near to distant. If sighting occurs, distance from the reticule and angle from angle board is taken to first sighting of single animal and best judged centre of school of animals. The recorder records the GPS position and time alongside. Weather conditions and changes in survey effort is also recorded.

Data recorded by ORCA surveys undergo a quality control procedure where it is checked by trained scientists, and then kept in both physical and digital formats using archives, log sheets, and geodatabases (Matear *et al.*, 2019). An annual additional check is completed to reduce inaccurate data being archived.

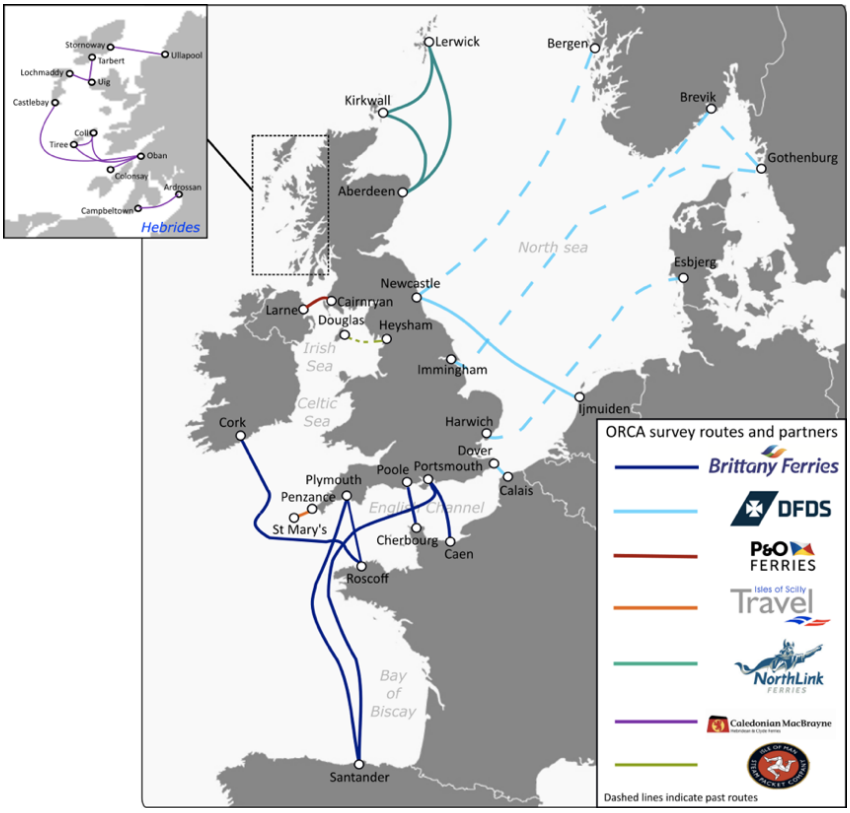


Figure C.3: Ferry routes of ORCA vessels (ORCA, 2020).

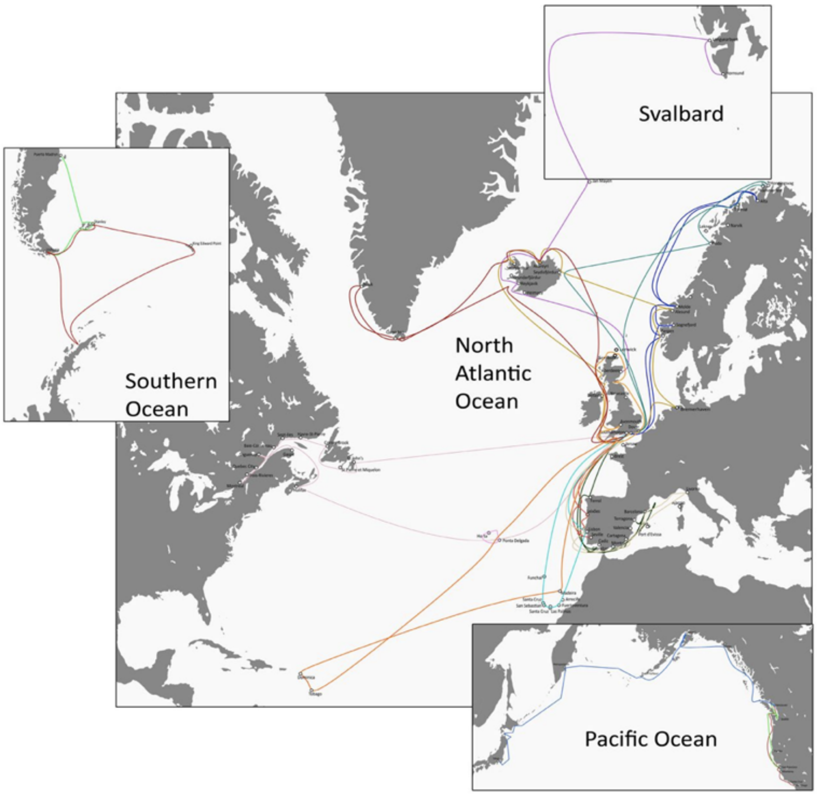


Figure C.4: Cruise routes of ORCA vessels (ORCA, 2020).

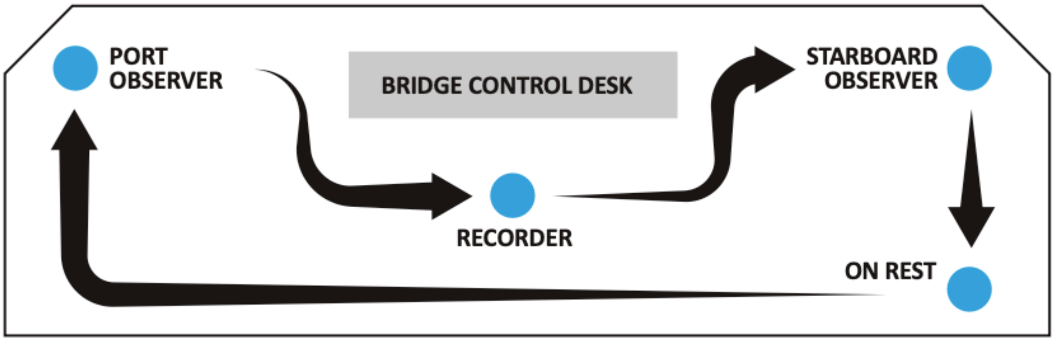


Figure C.5: Rotation diagram from ORCA survey protocol and survey guide (ORCA, 2017).

Two sets of data were provided by ORCA. The first is Distance Sampling data from the bridge of ferries during dedicated surveys with limited distractions and cruise ships, whilst the second is collected by Wildlife Officers on set ferry routes daily for up to six months of the year. The Wildlife Officers and Cruise teams operate with similar methods to Distance Sampling, but there may be some differences - such as the platform may not be forward facing, so they may be looking from the sides of the ship, and no bearing is taken of the sighting as such. Effort is also more variable as observers are often situated on the open deck of ships, not closed bridges, leading to reduced effort due to passenger distractions or increased due to more observations as a result of passengers helping.

Appendix D Pilot consultation

Method

The matrix method approach was also sent out as a pilot to three participants in diverse stakeholder groups that may utilise historical datasets to gather feedback: JNCC, Swansea University and Carmarthen County Council. Participants were identified as candidates that use large-scale heterogenous data in their work, and therefore likely to be susceptible to bias having an impact on their work.

The two questions asked of candidates were:

1. Whether this sort of bias quantification is carried out at all? Are there any preliminary judgements of bias or a risk assessment process when assessing the use of a dataset for research, legislation etc.? For example, if using a sightings dataset to work out abundance or distributions of species, are the various levels of bias considered in the data collection before using the dataset, or is it just used as deemed the best available data regardless of biases?
2. Whether quantification of the "risk of bias" in datasets would be useful in your area of expertise (legislation, marine management, research)? Would it be constructive and valuable to have a risk assessment of the dataset to include, before using it further?

Stakeholder opinions/email discussion

Feedback from candidates is given below on the matrix approach.

"Matrix would be useful, and could help researchers filter the data by way of data quality… Researchers could then choose the quality of data to use, but would this then result in conservative estimates of for example, abundance in this case"

Chiara Bertelli, Scientific Officer Swansea University

"As far as I know, there is nothing like your idea in place at the moment, so in principle, it's a really good idea. However, I have a few reservations... with your example of misidentification, this is pretty much impossible to quantify – there is no reason to assume that professional and experienced surveyors are doing any better a

job of identifying things than citizen scientists and less experienced people (at least not in the world of bird surveys, unfortunately), likewise with experience, what I find is that deviations from standard methodologies are just as likely to come from experienced surveyors (as they 'know better') than inexperienced surveyors who lack practice. So, I think you'd need to be very careful about what you considered to be risks/biases and how exactly you quantified them. That said, it would be an excellent way of flagging up issues with data such as an ageing dataset, how standardised data collection was (or even, how/if data collection was standardised at all), whether there was some sort of qualification observers needed to achieve to collect the data etc."

Mark Lewis, JNCC Data Officer

"I think there can be a general assumption that data from these sources have already been through some sort of verification and are 'endorsed' in some way, but this is not the case. You often have to dig into the metadata to get an idea of any biases that may affect the interpretation of the data. It is very difficult to do any analysis on data that has been collected by someone else without any prior knowledge of how that data has been collected or any assumptions that have been made in doing so. A big part of my role is presenting other people's data as evidence that underpins my assessment of the sustainability and environmental impact of development. We work to often-tight deadlines and so don't always have time to dig into the supporting documents of data to identify biases or flaws in the data. A simple and easy to understand quantification of data bias would be very useful. I think that the risk assessment method is a really interesting way of quantifying data biases. It is a straightforward format which is likely to be familiar to people who may have used it in other contexts. It would be a clear and quick way of understanding data biases, which is particularly useful when data is being used in contexts with often tight deadlines"

Rebecca Stone, Sustainability and Ecology Policy Officer, Carmarthen County Council

Conclusions

Key conclusions for the matrix approach are summarised below. Positives of the matrix approach include:

- Aid researchers to filter data by way of data quality.

- A way of flagging issues with data – such as aged datasets, how standardised data is, qualification level of observers.
- It provides a simple and easy to understand quantification of data bias, which is useful for people under time constraints.

Some further considerations of the approach:

Some biases are hard to quantify and therefore would require careful consideration.

Whilst I would have liked to explore a feasibility study more for my PhD thesis, it is at the pilot stage currently and did not escalate to a larger scale piece.

Appendix E Scaling by effort

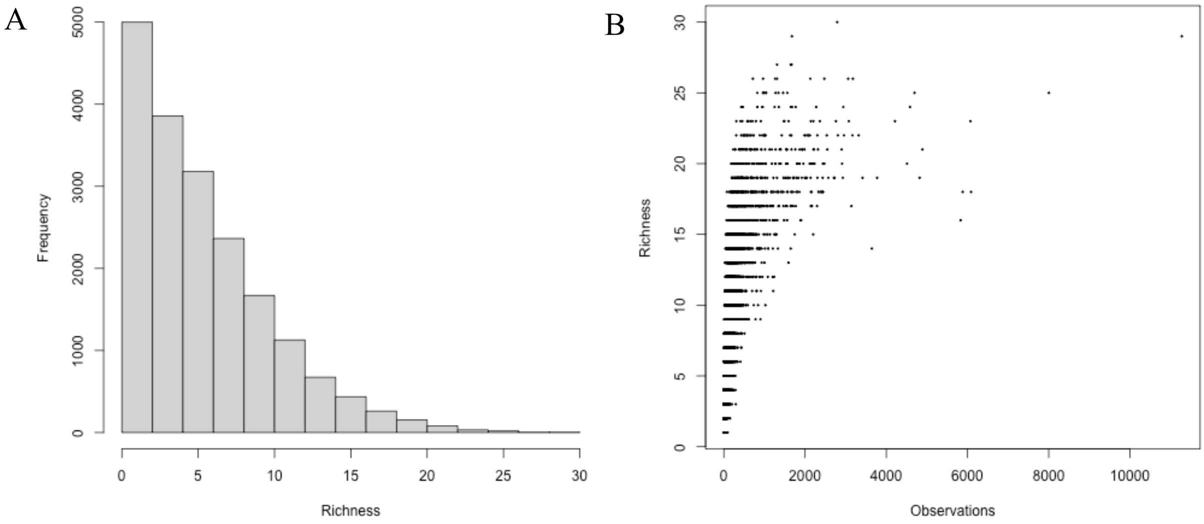


Figure E-1 A) Histogram of raw species richness values B) Scatterplot of species richness against observations for each grid cell at 100 km^2 resolution.

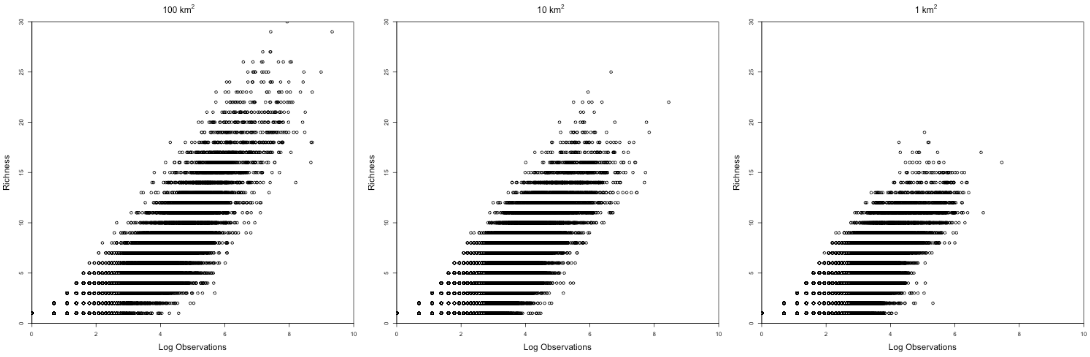


Figure E-2 Species richness per grid cell at different resolutions (100 km^2, 10 km^2, 1 km^2) against the logarithm of the number of observations in each grid cell. Demonstrates semi-log relationship.

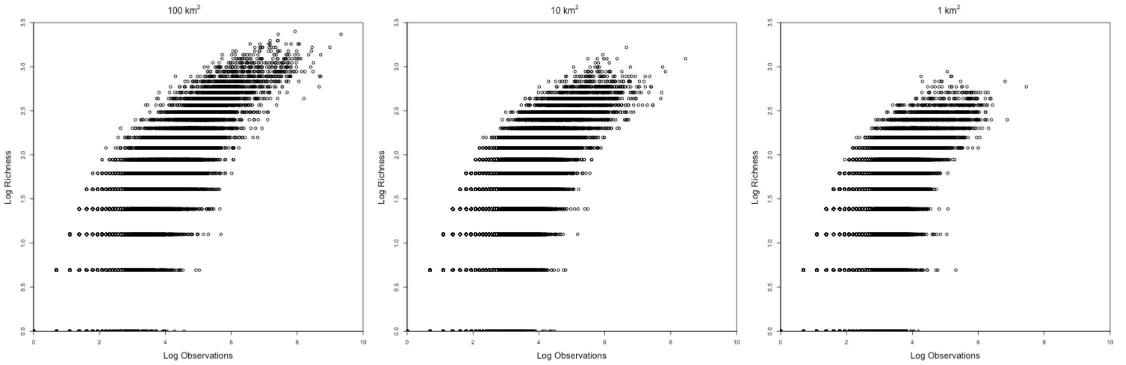


Figure E-3 Log species richness per grid cell at different resolutions (100 km^2, 10 km^2, 1 km^2) against the logarithm of the number of observations in each grid cell. Demonstrates log-log relationship.

Appendix F Time series decomposition of measures

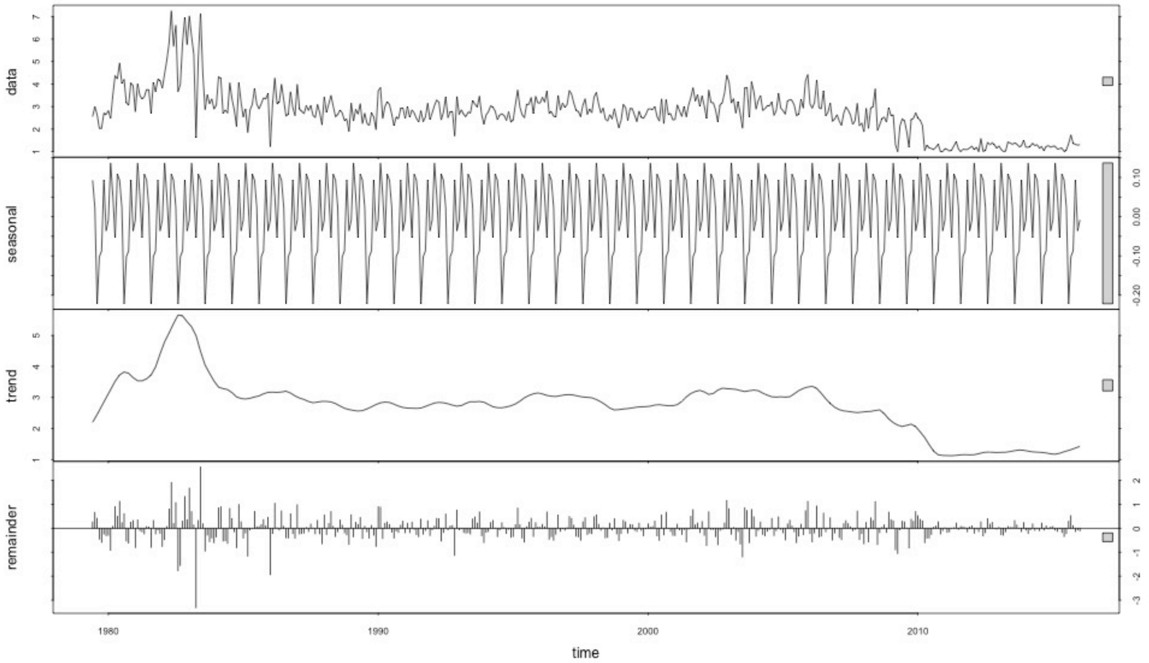


Figure F-1 Time-series decomposition of mean species richness over time using a 100 km^2 resolution grid from June 1979 to September 2018.

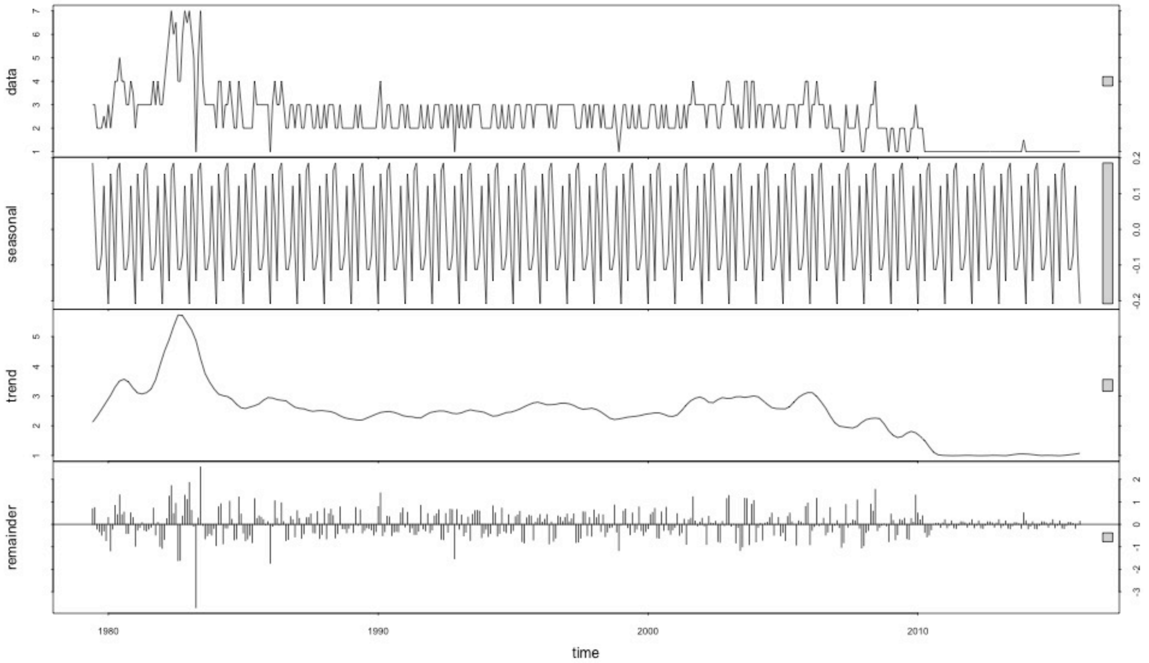


Figure F-2 Time-series decomposition of median species richness over time using a 100 km^2 resolution grid from June 1979 to September 2018.

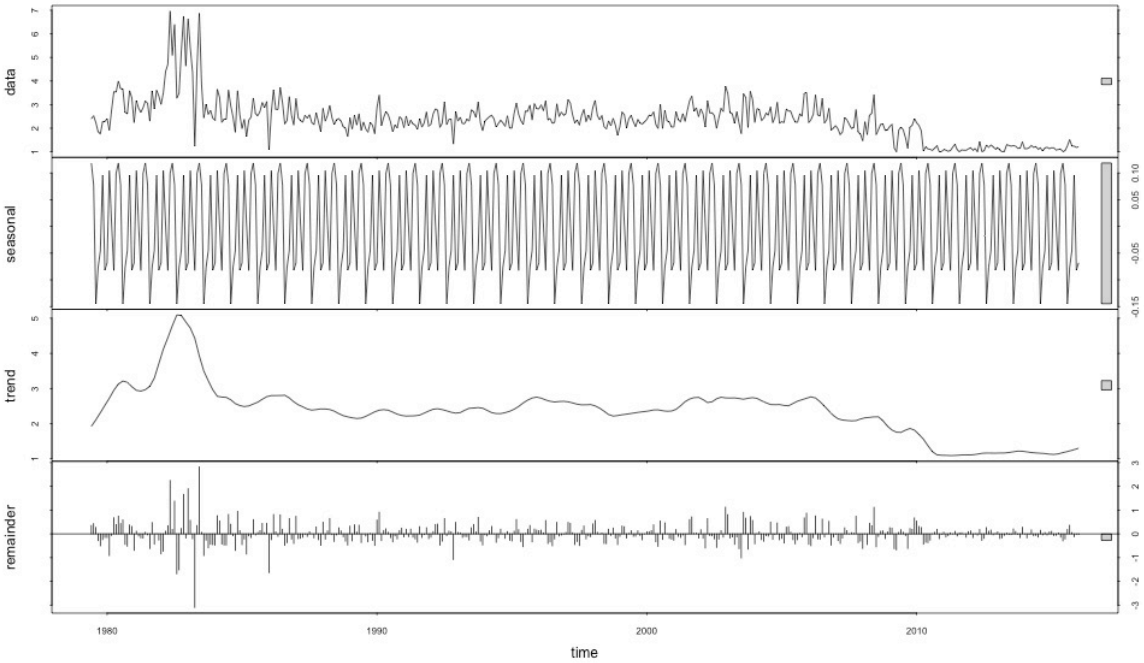


Figure F-3 Time-series decomposition of the geometric mean of species richness over time using a 100 km^2 resolution grid from June 1979 to September 2018.

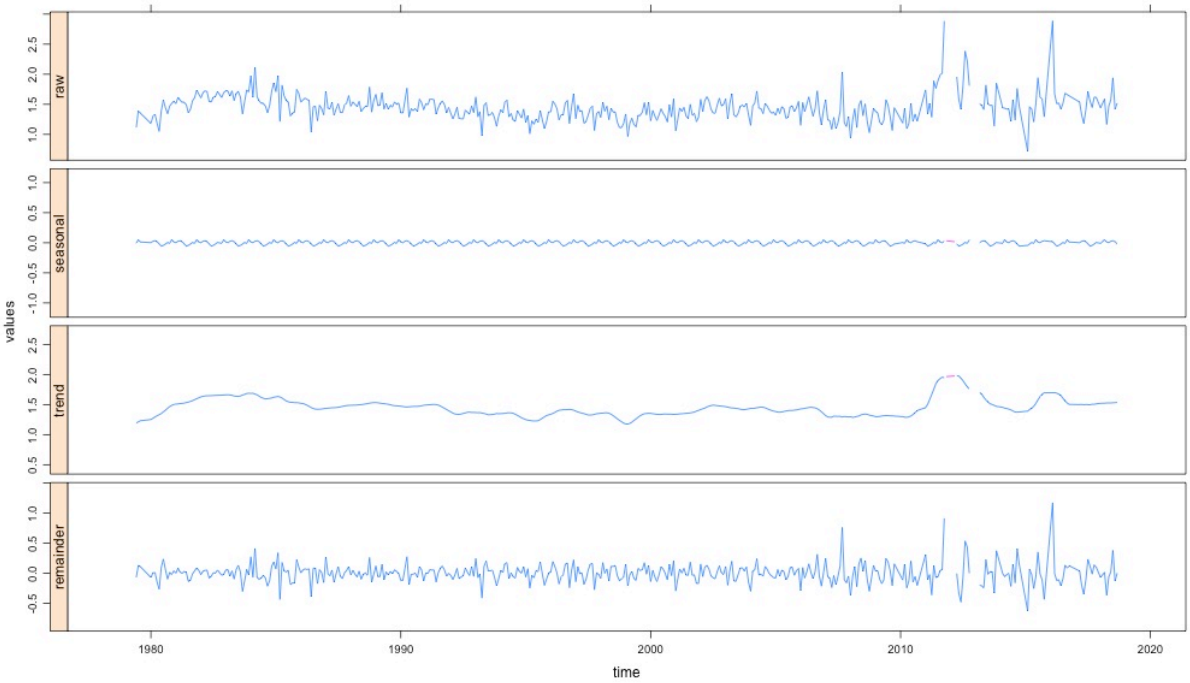


Figure F-4 Time-series decomposition of mean species richness per unit effort (SRPUE) over time using a 10 km^2 resolution grid from June 1979 to September 2018.

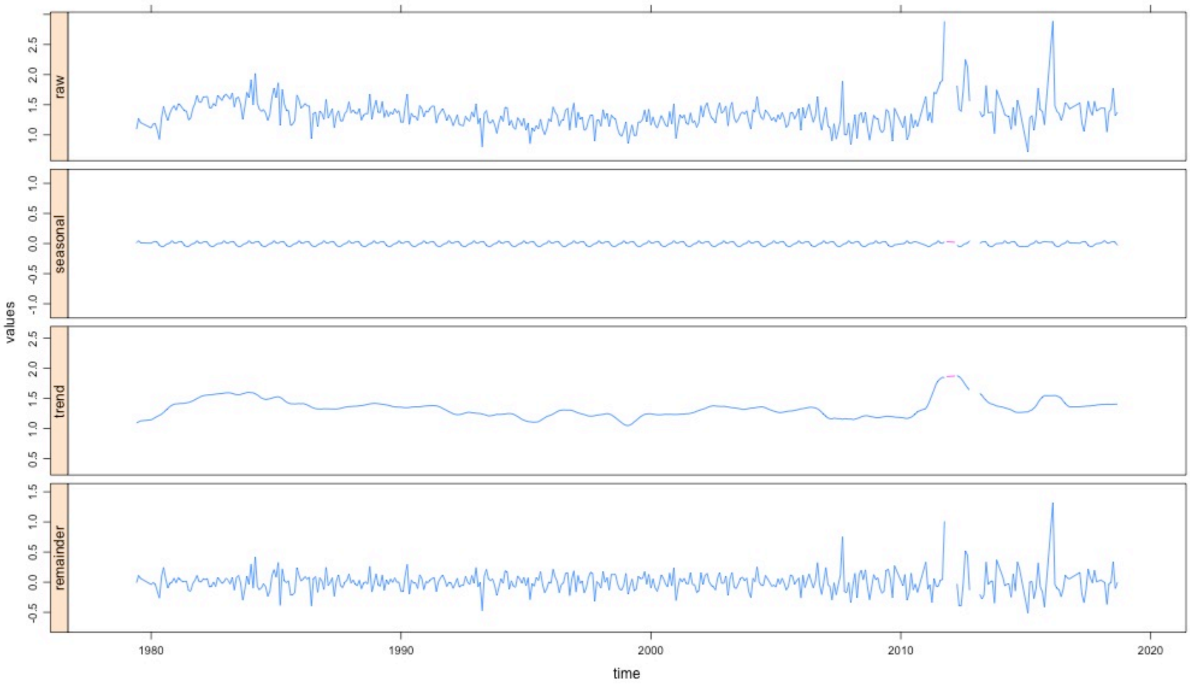


Figure F-5 Time-series decomposition of mean species richness per unit effort (SRPUE) over time using a 1 km^2 resolution grid from June 1979 to September 2018.

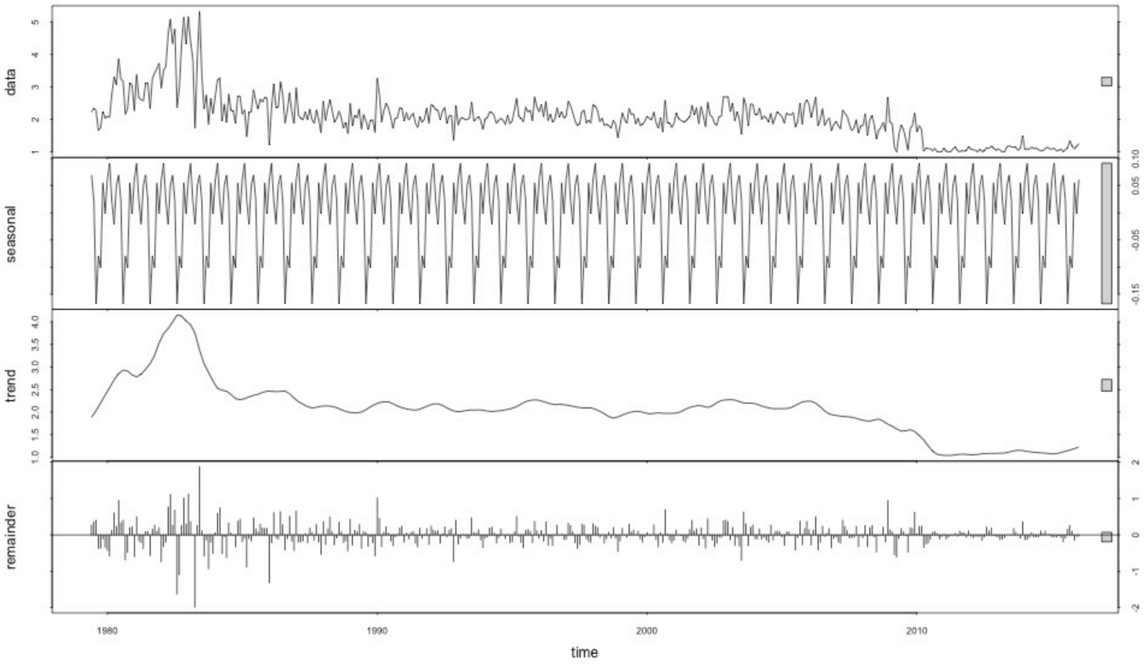


Figure F-6 Time-series decomposition of mean species richness over time using a 10 km^2 resolution grid from June 1979 to September 2018.

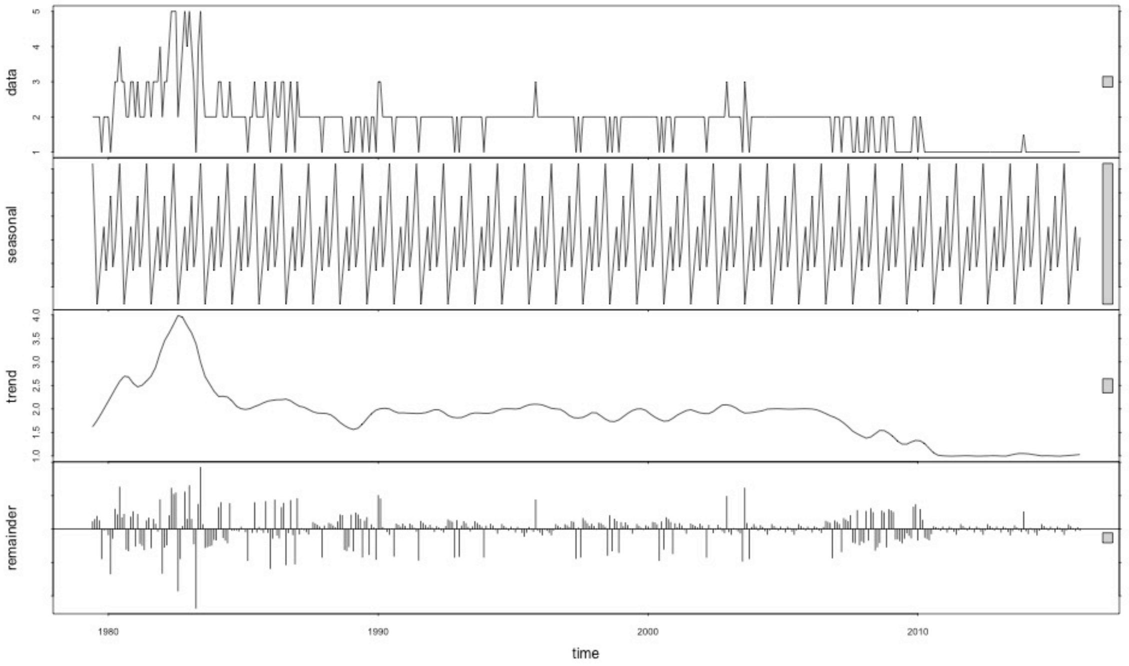


Figure F-7 Time-series decomposition of median species richness over time using a 10 km^2 resolution grid from June 1979 to September 2018.

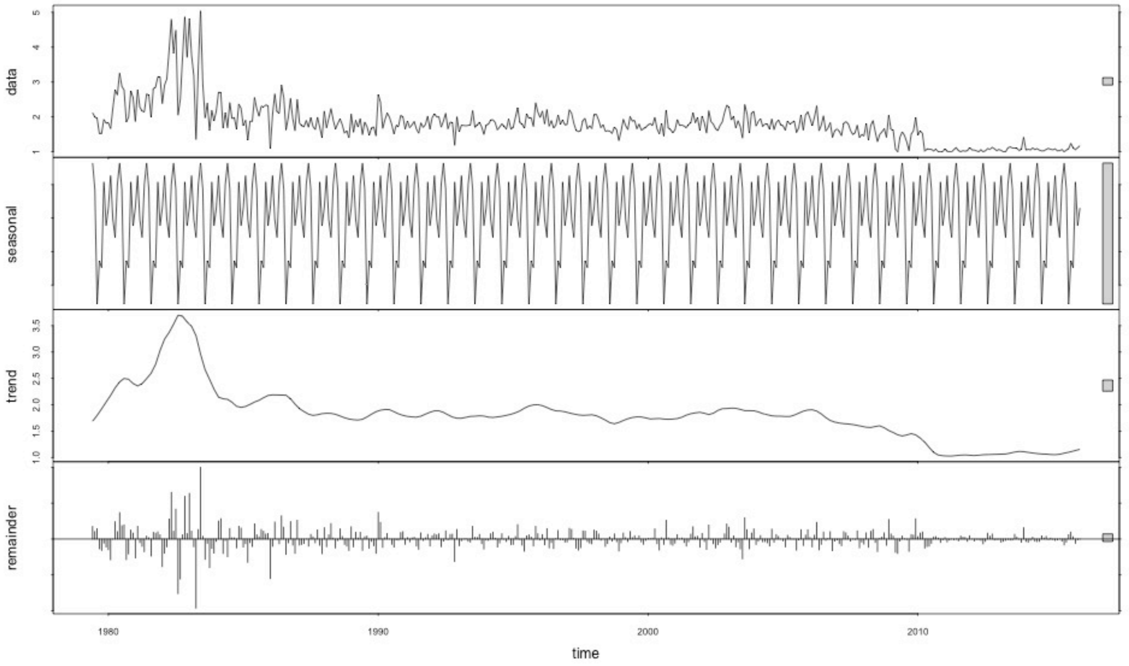


Figure F-8 Time-series decomposition of geometric mean of species richness over time using a 10 km^2 resolution grid from June 1979 to September 2018.

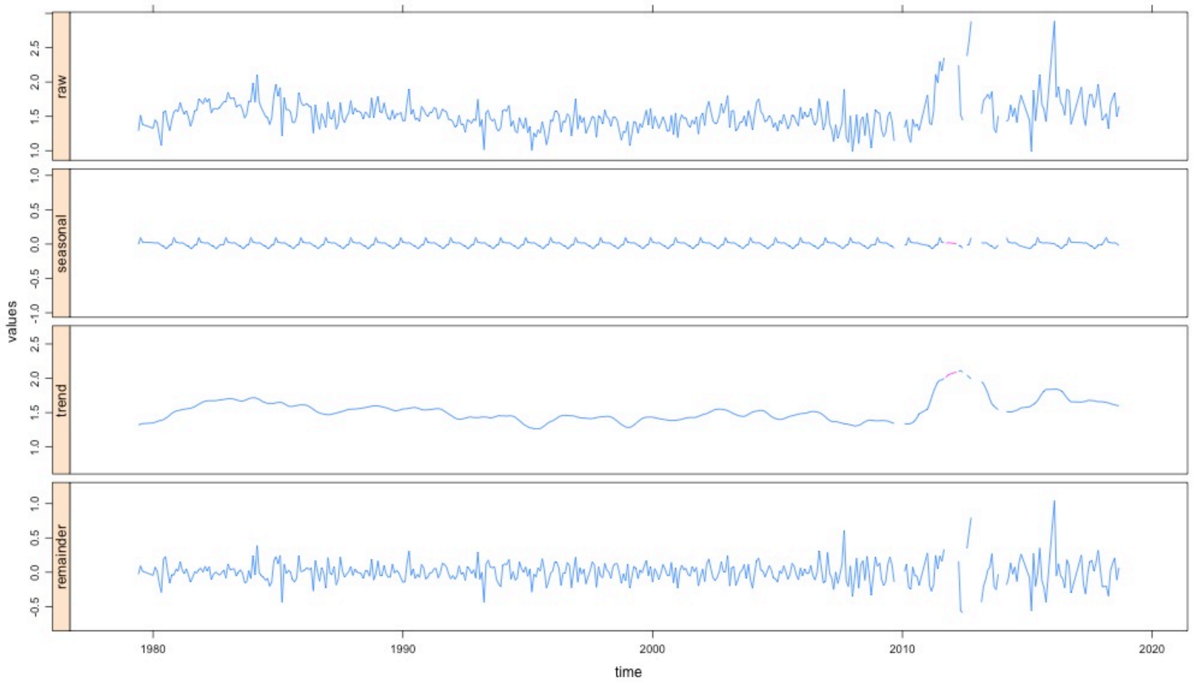


Figure F-9 Time-series decomposition of mean species richness per unit effort (SRPUE) over time using a 10 km^2 resolution grid from June 1979 to September 2018.

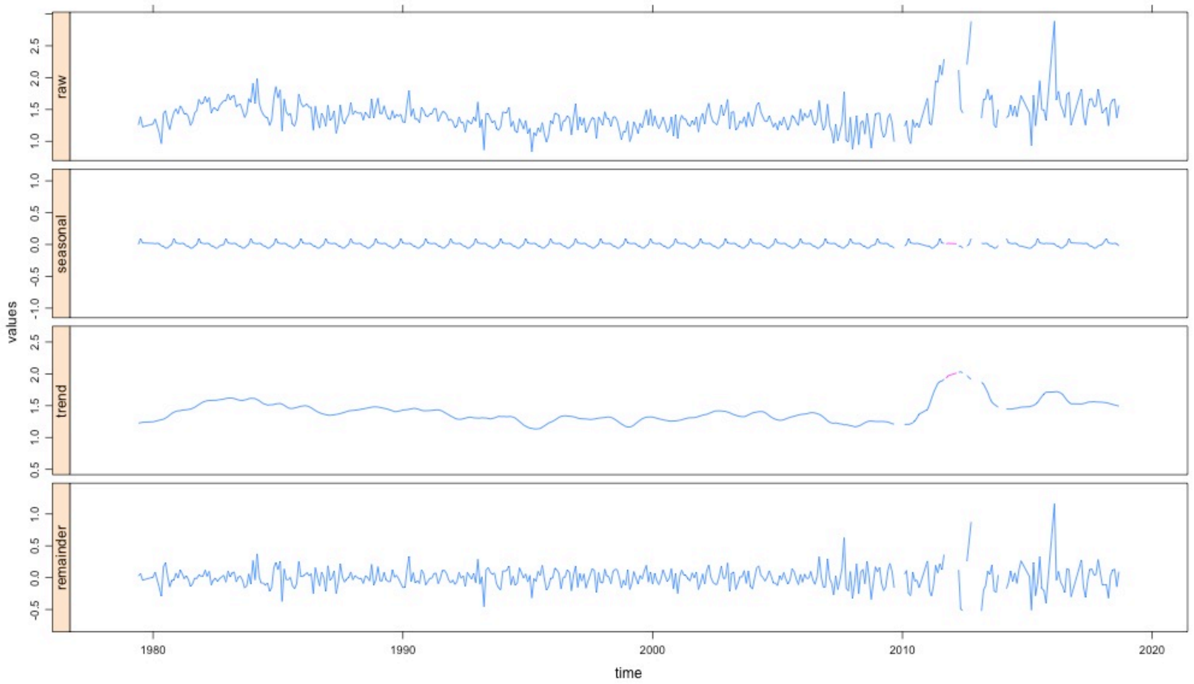


Figure F-10 Time-series decomposition of mean species richness per unit effort (SRPUE) over time using a 10 km^2 resolution grid from June 1979 to September 2018.

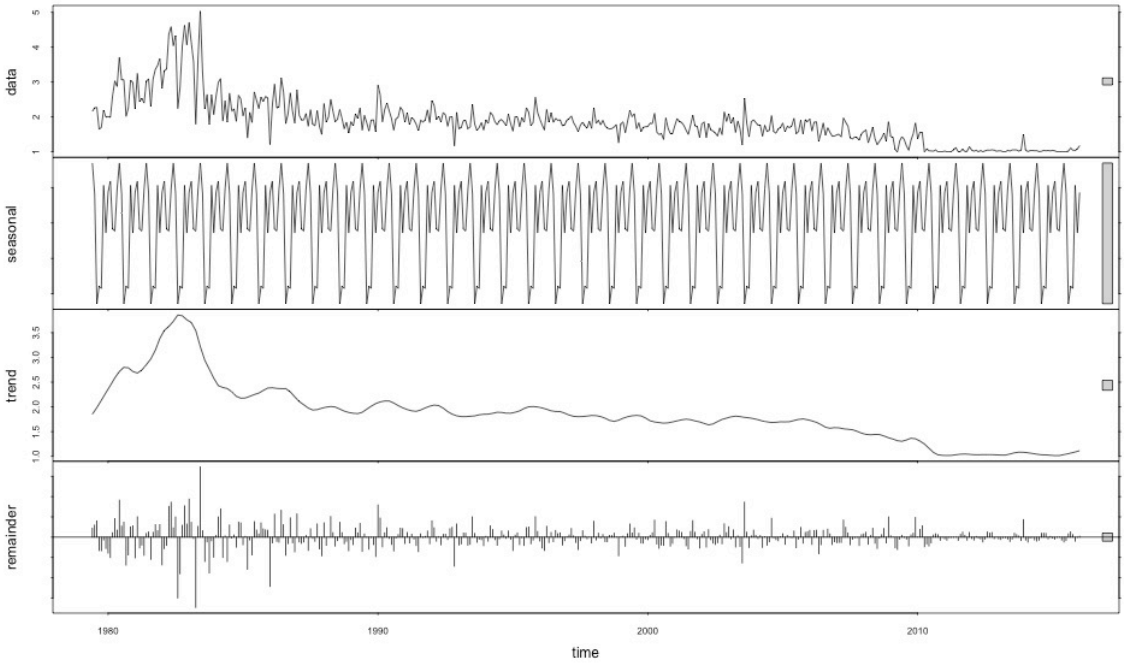


Figure F-11 Time-series decomposition of mean species richness over time using a 1 km^2 resolution grid from June 1979 to September 2018.

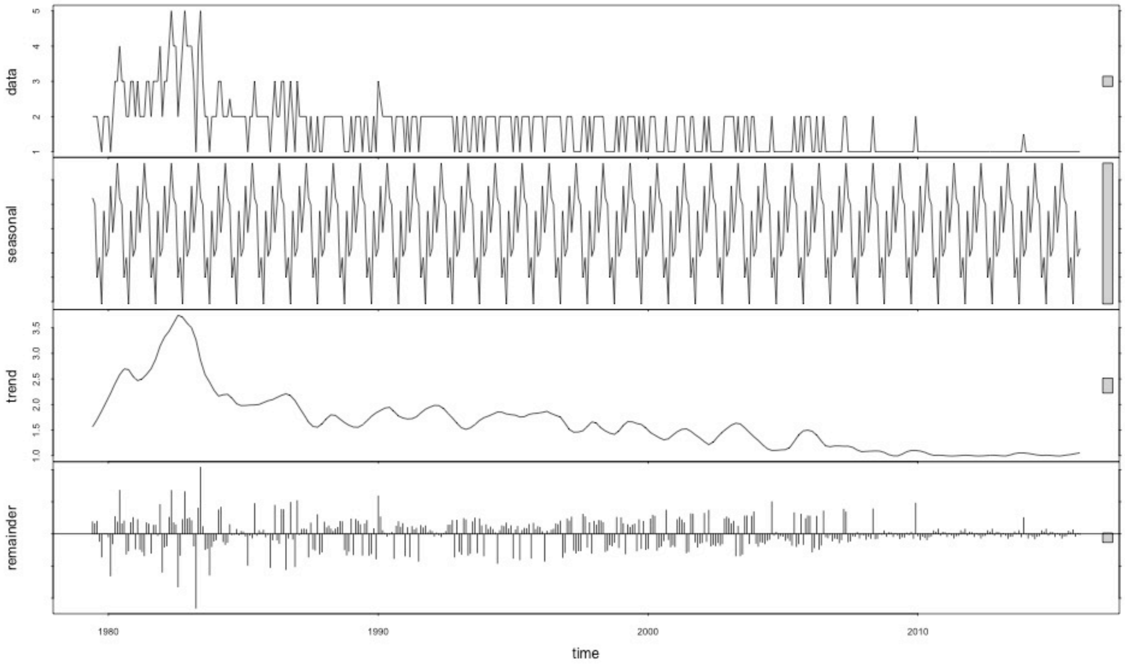


Figure F-12 Time-series decomposition of median species richness over time using a 1 km^2 resolution grid from June 1979 to September 2018.

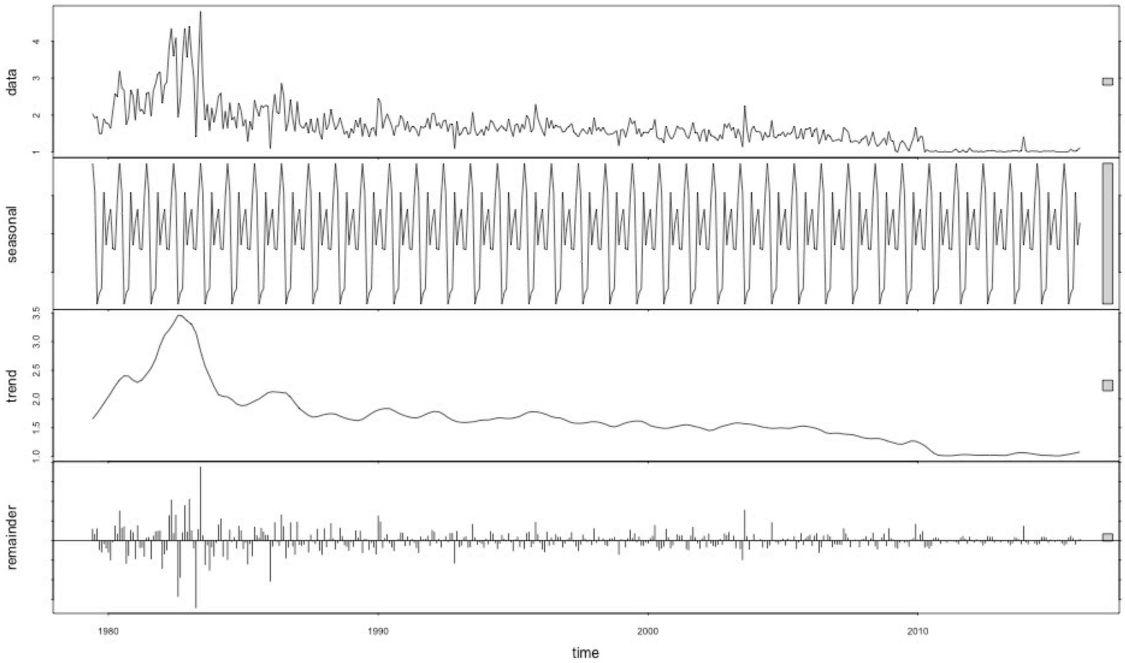


Figure F-13 Time-series decomposition of the geometric mean of species richness over time using a 1 km^2 resolution grid from June 1979 to September 2018.

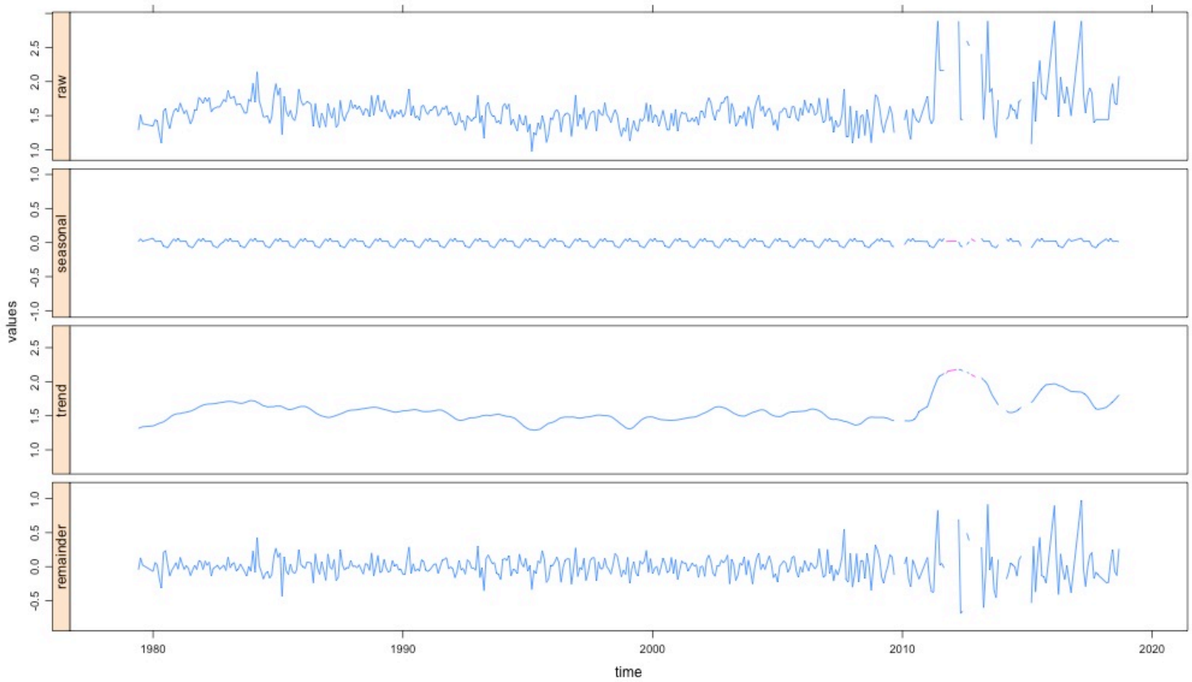


Figure F-14 Time-series decomposition of mean species richness per unit effort (SRPUE) over time using a 10 km^2 resolution grid from June 1979 to September 2018.

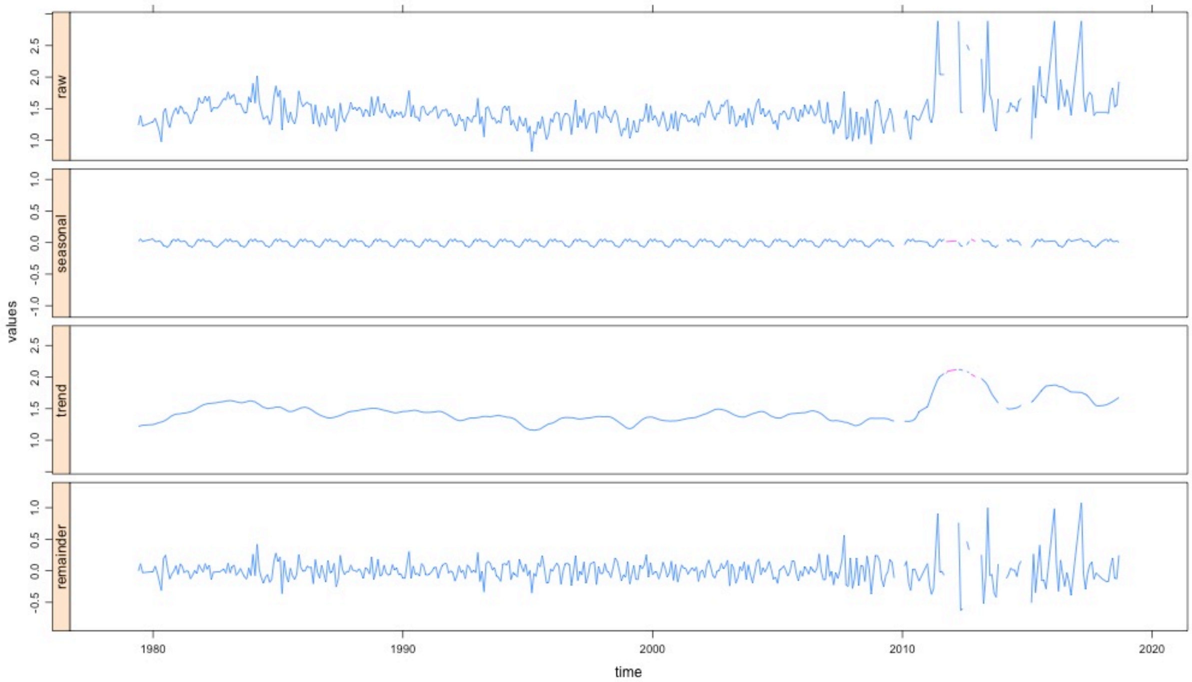


Figure F-15 Time-series decomposition of mean species richness per unit effort (SRPUE) over time using a 10 km^2 resolution grid from June 1979 to September 2018.

Appendix G Investigation into dispersion values from GLM versus GAM

Initially this chapter was to include an investigation into dispersion values to assess any aggregations or spread in the data. Dispersion values of SRPUE were calculated from quasi-poisson generalised linear models (GLMs) which excluded spatial elements using stats package in R (R Core Team, 2022) and quasi-poisson generalised additive models (GAMs) using the package mgcv (Wood, 2011) to include X and Y coordinates. The GAM was set to have a minimum of 5 data points within each subset by month to be included in the model, to give a fair estimation of dispersion. Species richness was the response variable, and the terms variable used was log(Occurrences) as an indicator of the sampling intensity of the grid cell.

Initial outputs of dispersion have been included in this appendix but was **not retained for the thesis chapter.**

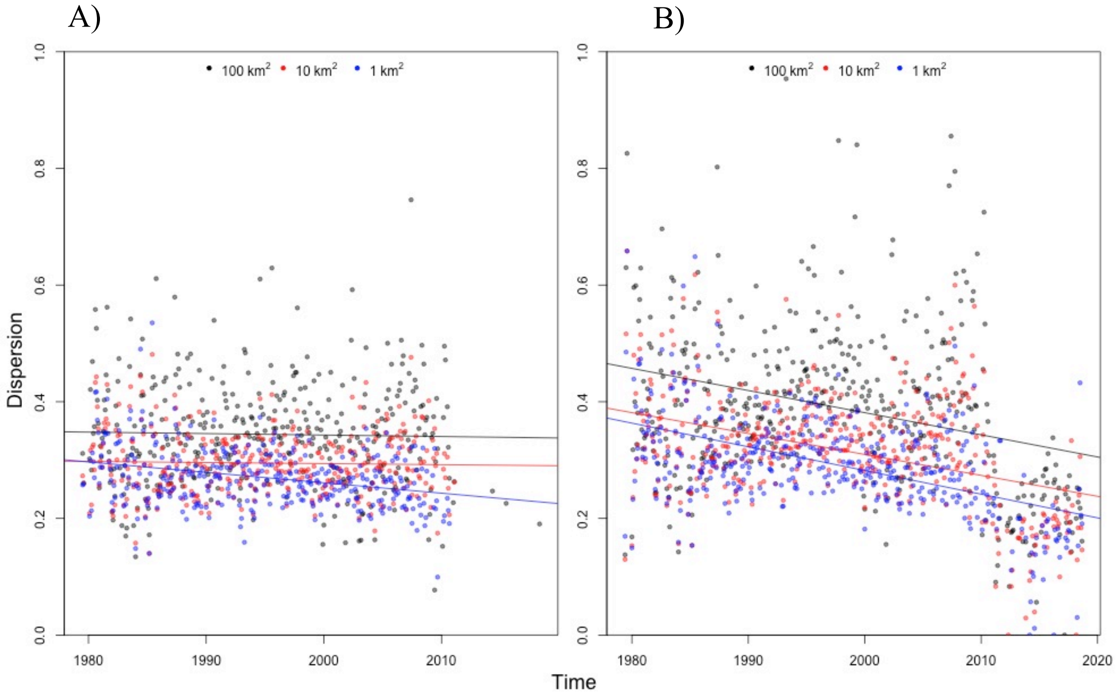


Figure G-1 Scatterplot of dispersion values over time for species richness as a function of log(effort) (SPUE) for GLM without coordinates (A) and GAM with coordinates (B), for different resolutions of grid size (100 km^2 , 10 km^2, 1 km^2).

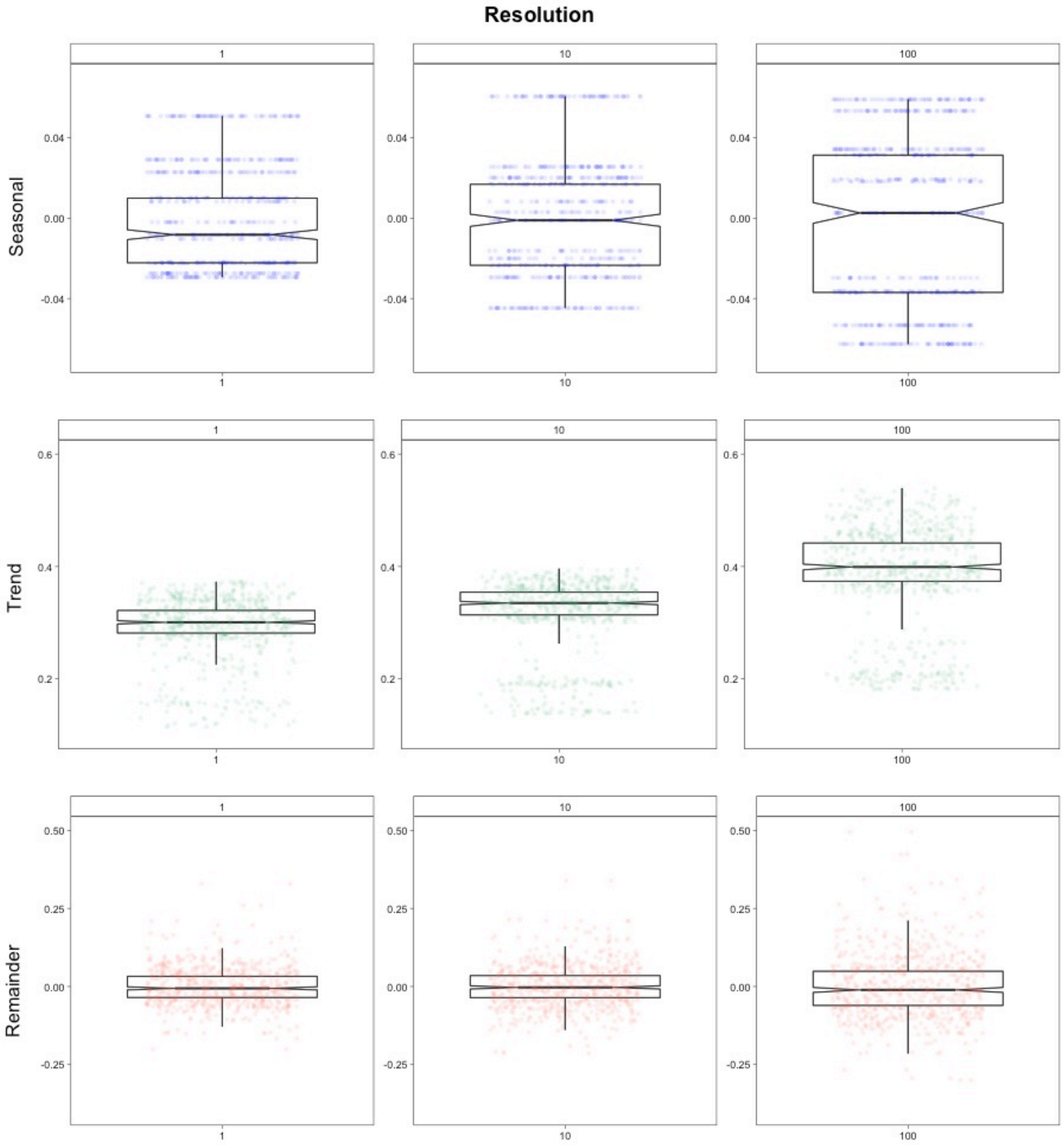


Figure G-2 Boxplot of different components of STL dispersion values of species richness as a function of log(effort) (SPUE) for GLM without coordinates for different resolutions of grid size (100 km^2, 10 km^2, 1 km^2). The raw data is made up of seasonal, trend and remainder components.

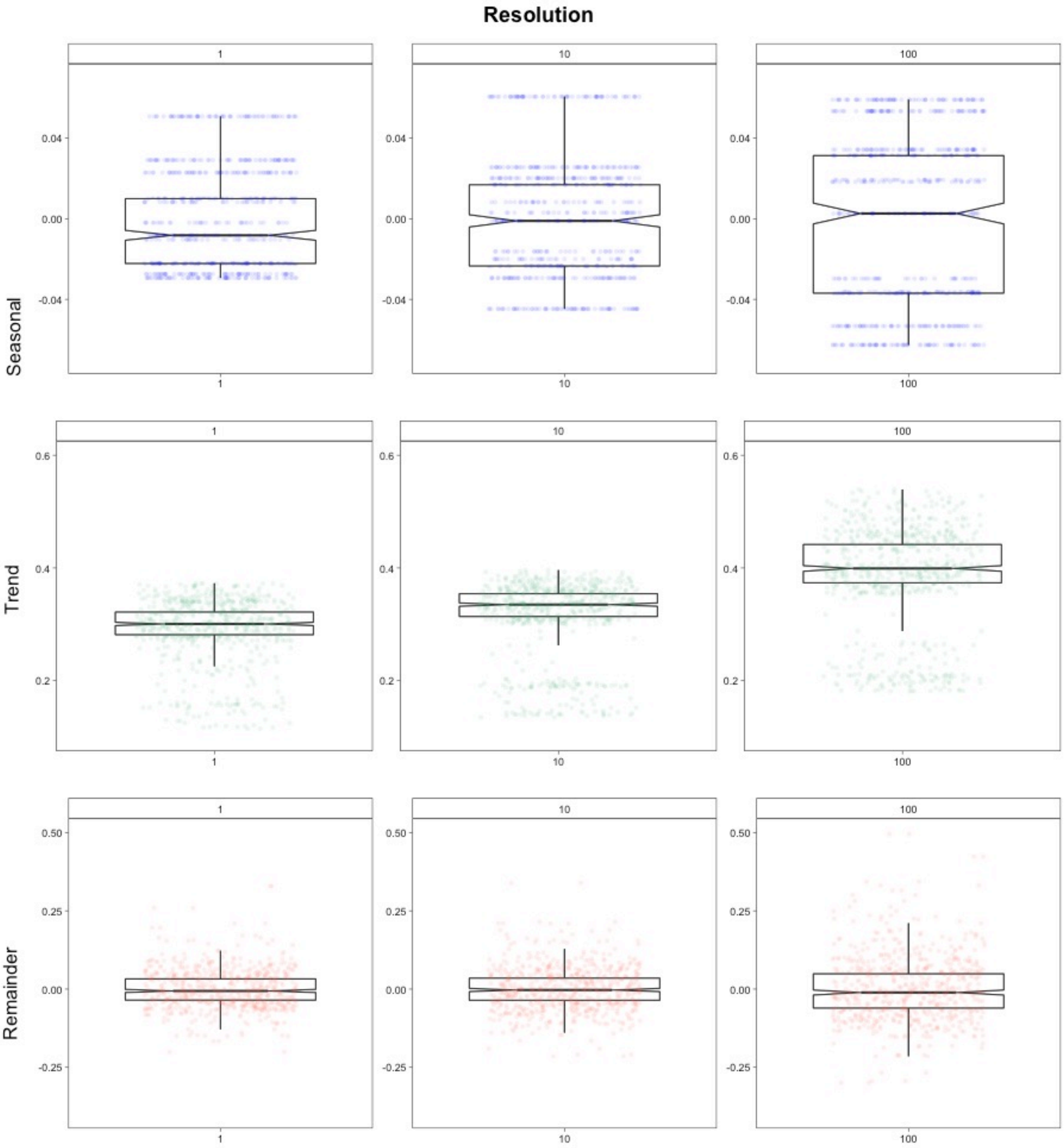


Figure G-3 Boxplot of different components of STL dispersion values of species richness as a function of log(effort) (SPUE) for GAM with coordinates for different resolutions of grid size (100 km^2, 10 km^2, 1 km^2). The raw data is made up of seasonal, trend and remainder components.

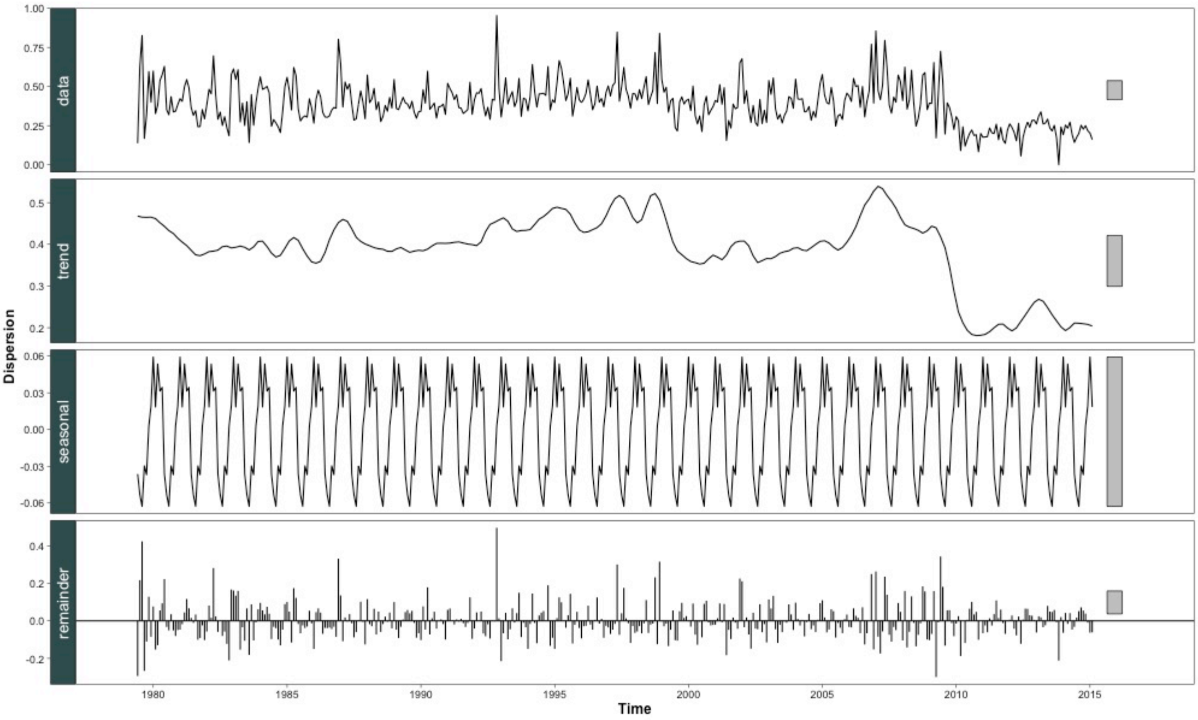


Figure G-4 Time-series decomposition of dispersion values over time from SRPUE using a 100 km^2 resolution grid from 1979 to 2015 from GLM, with no inclusion of coordinates.

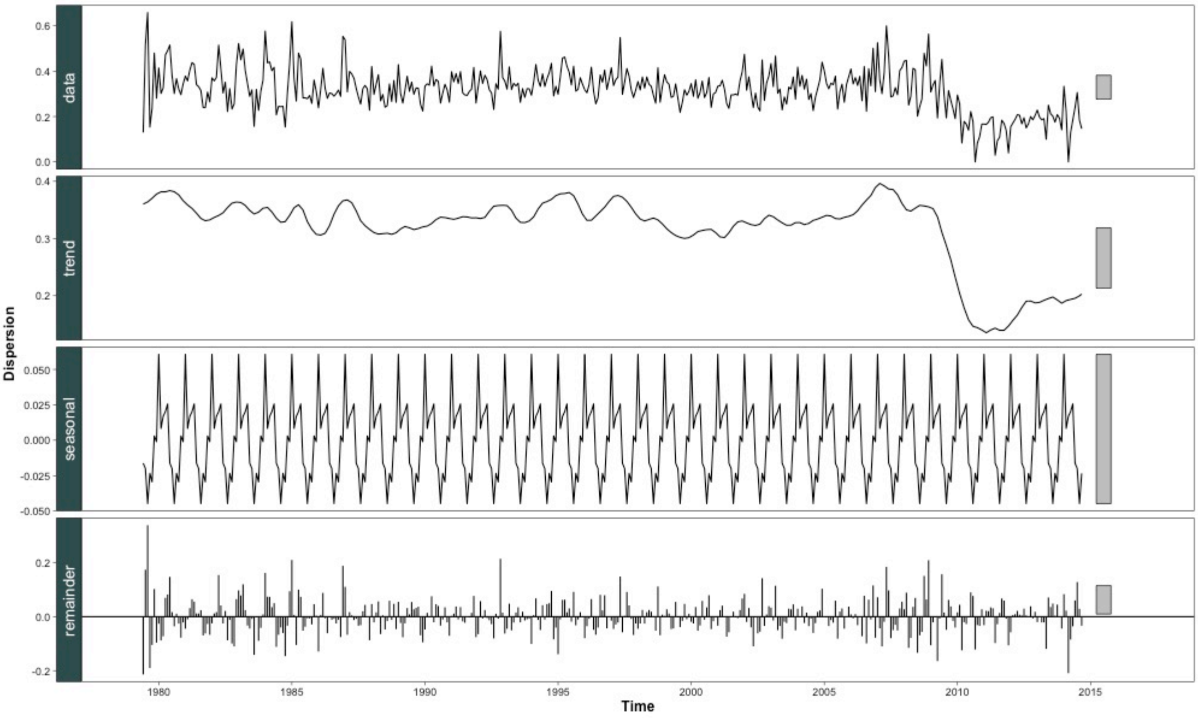


Figure G-5 Time-series decomposition of dispersion values from SRPUE over time using a 10 km^2 resolution grid from 1979 to 2015 from GLM, with no inclusion of coordinates.

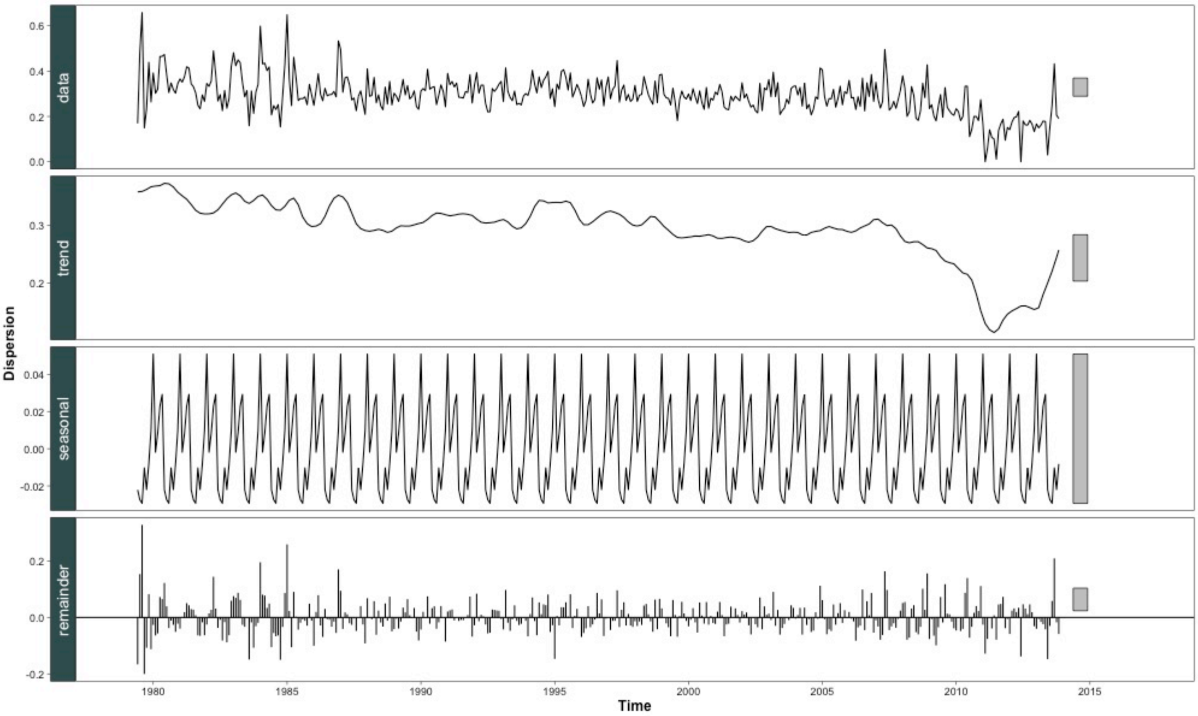


Figure G-6 Time-series decomposition of dispersion values over time from SRPUE using a 1 km^2 resolution grid from 1979 to 2015 from GLM, with no inclusion of coordinates.

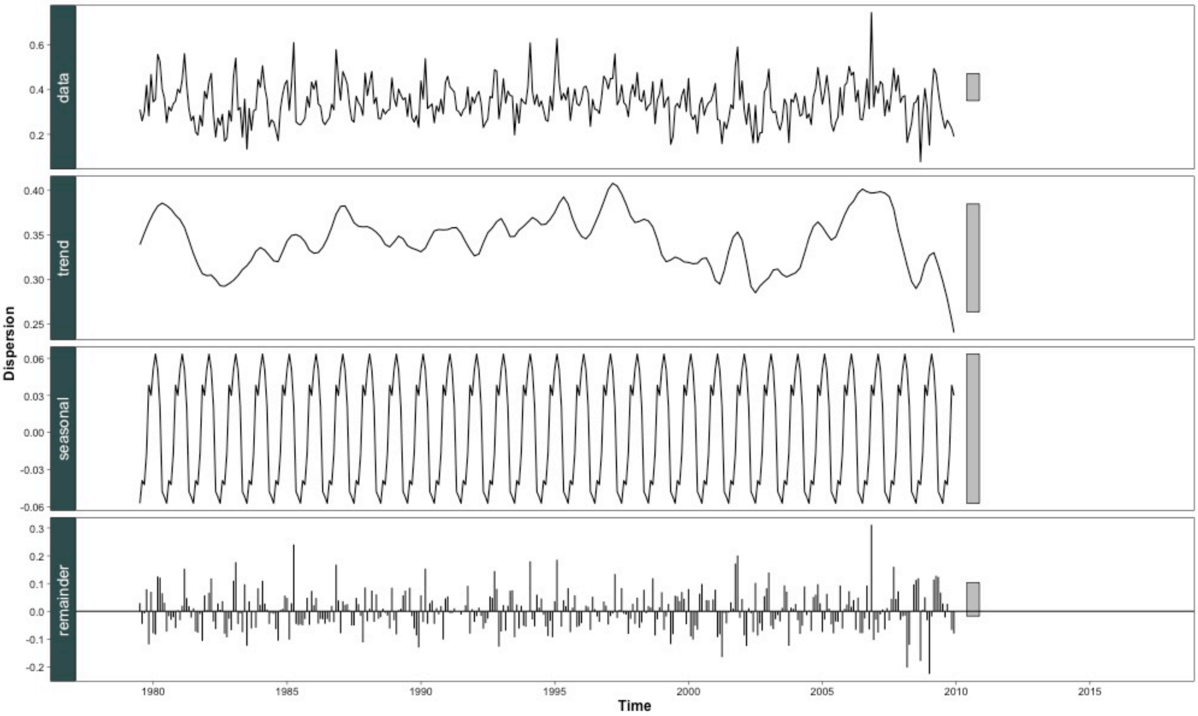


Figure G-7 Time-series decomposition of dispersion values from SRPUE over time using a 100 km^2 resolution grid from 1979 to 2015 from GAM, with inclusion of coordinates to account for spatial element.

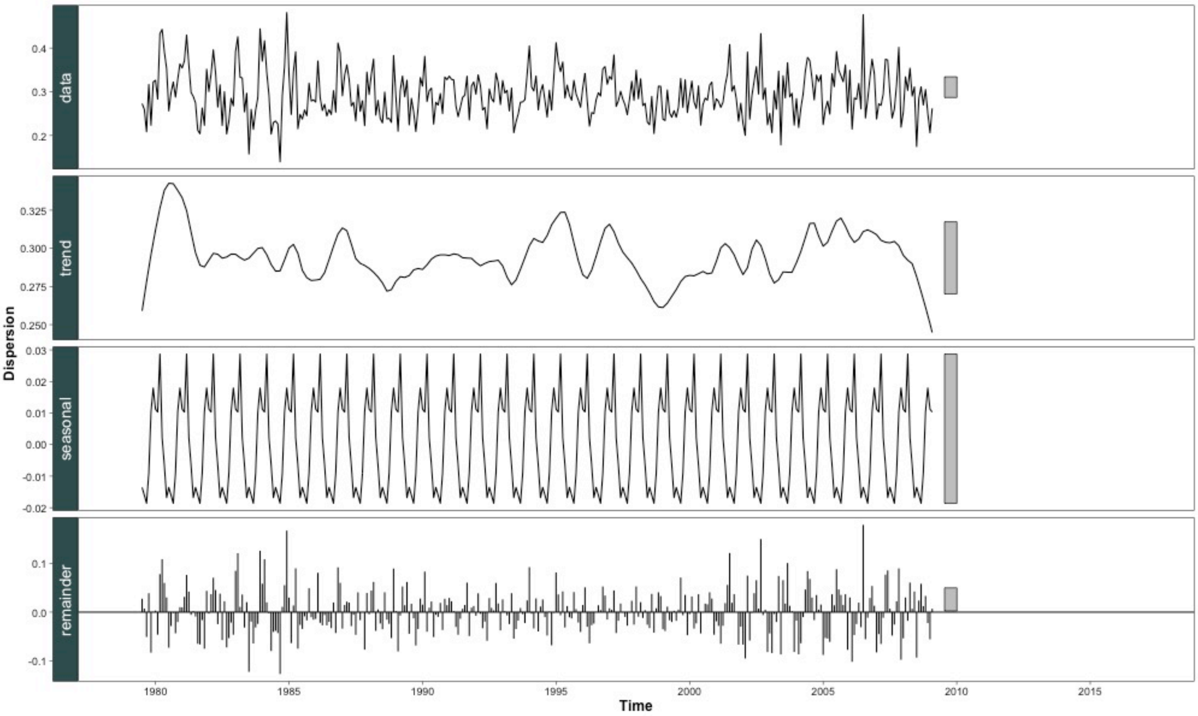


Figure G-8 Time-series decomposition of dispersion values from SRPUE over time using a 10 km^2 resolution grid from 1979 to 2015 from GAM, with inclusion of coordinates to account for spatial element.

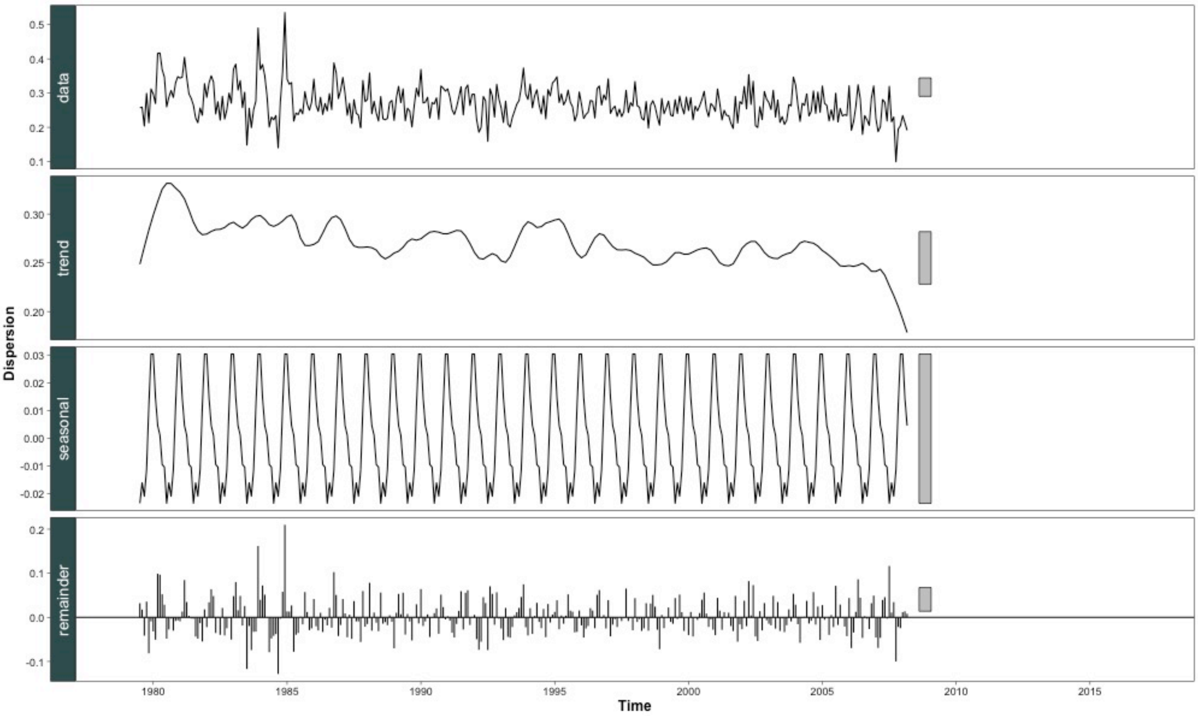


Figure G-9 Time-series decomposition of dispersion values from SRPUE over time using a 100 km^2 resolution grid from 1979 to 2015 from GAM, with inclusion of coordinates to account for spatial element.

Appendix H Maps of environmental variables

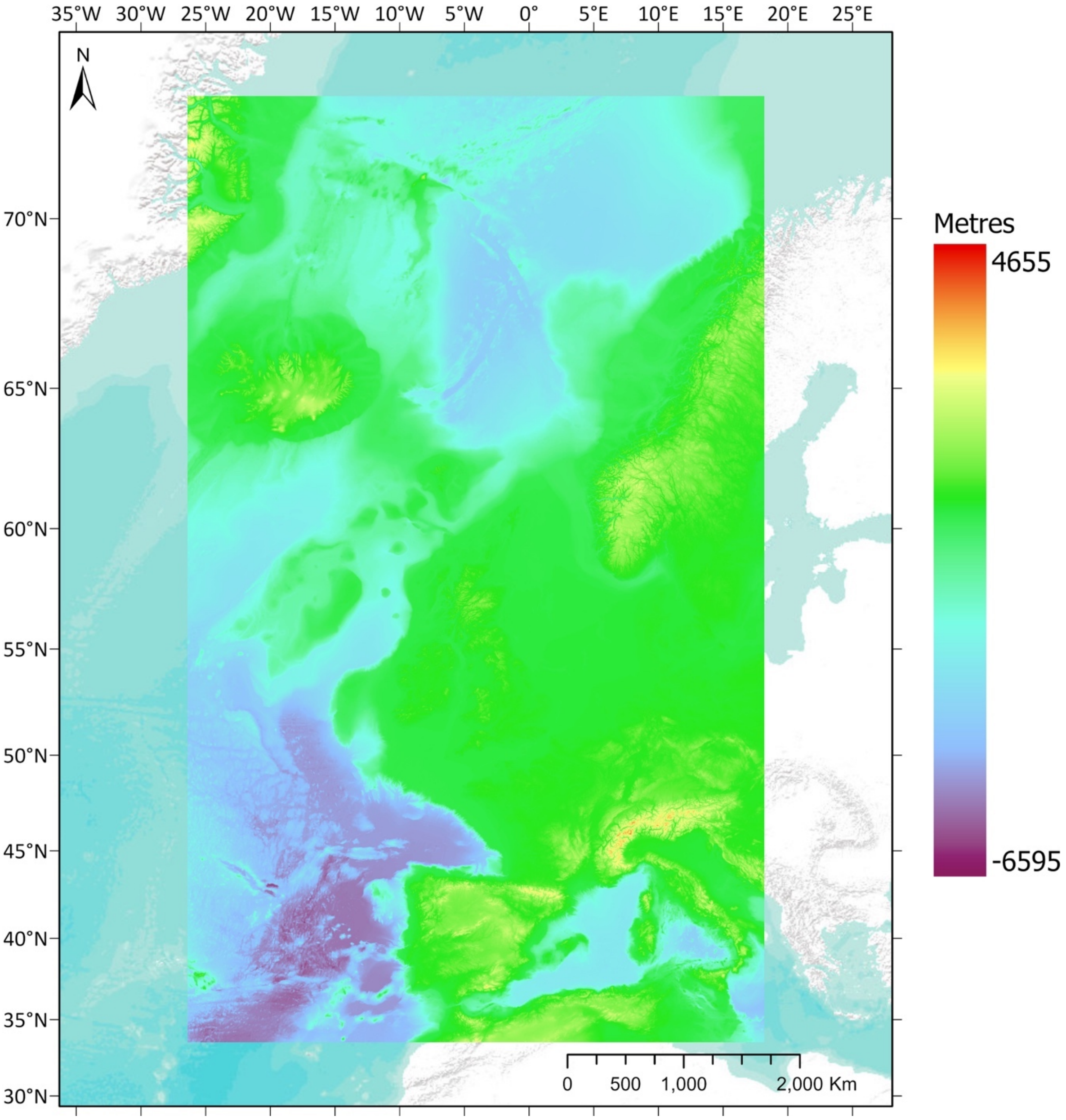


Figure H1 Map of bathymetry used in study, averaged from 1979 to 2018. Original data from GEBCO (GEBCO Compilation Group, 2021)

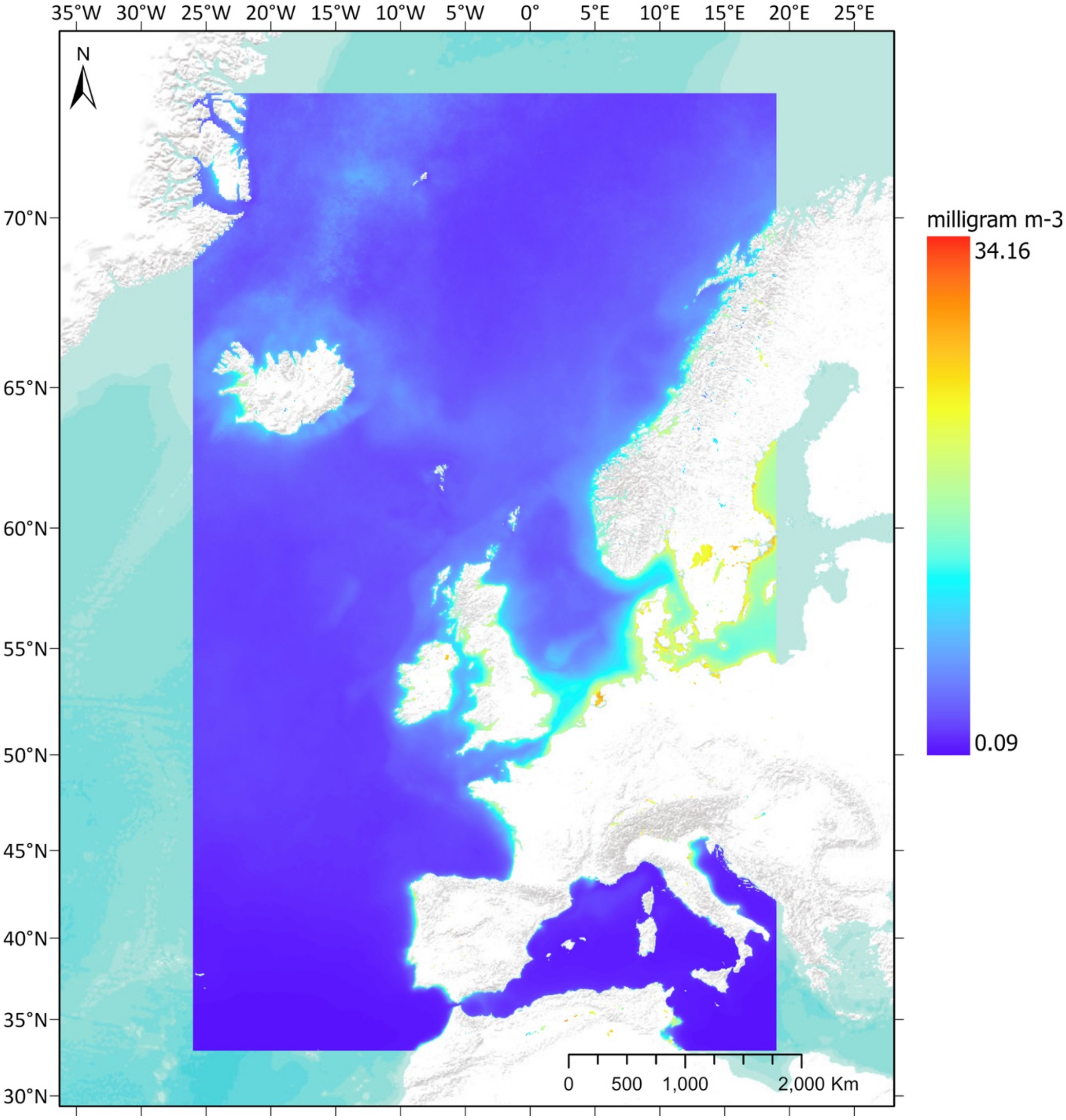


Figure H-2 Map of chlorophyll a used in study, averaged from 1979 to 2018. Original data from Copernicus (CMEMS, 2021).

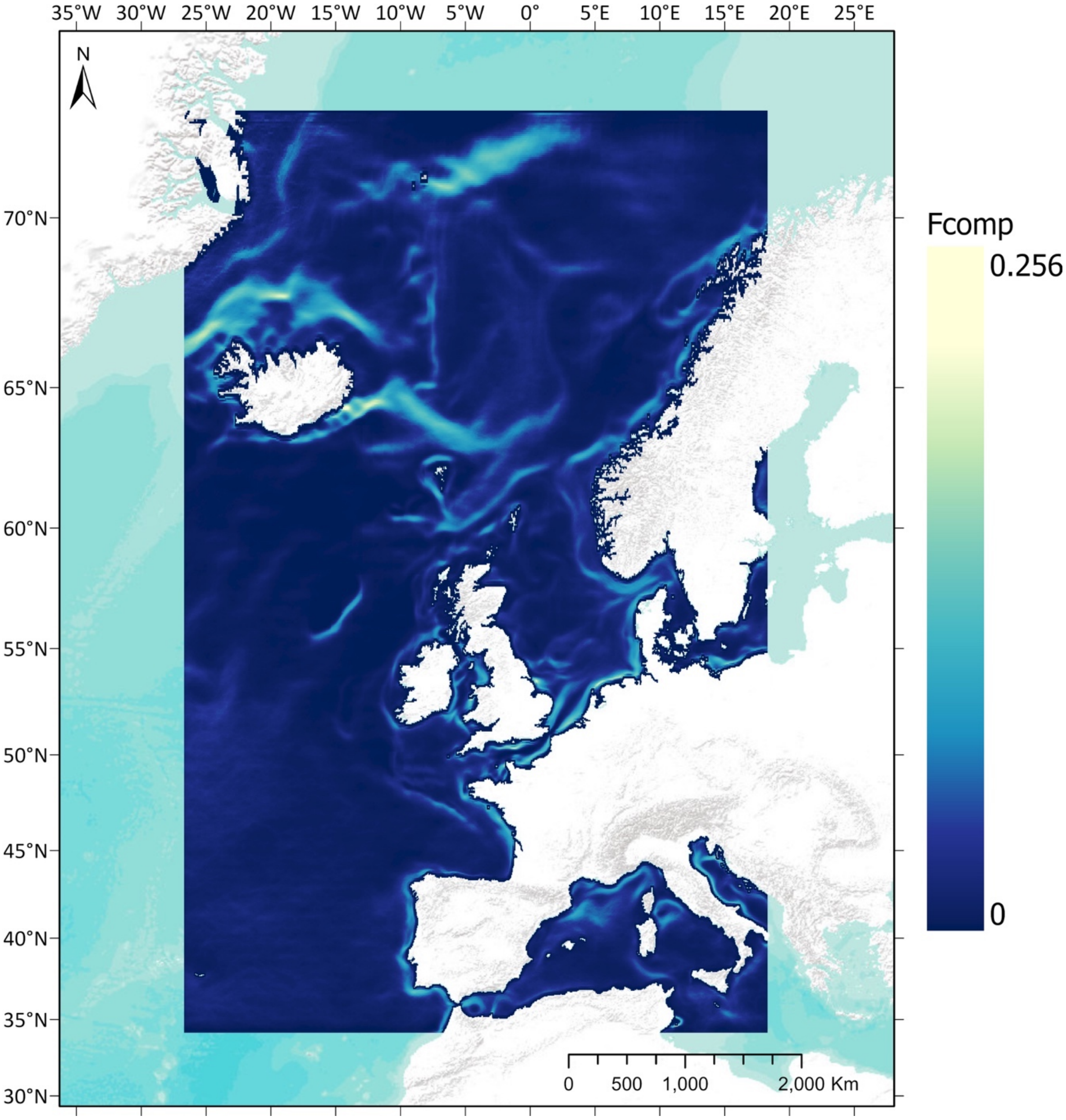


Figure H-3 Map of fronts used in study, averaged from 1979 to 2018. Original data from NEODAAS (NEODAAS, 2022)

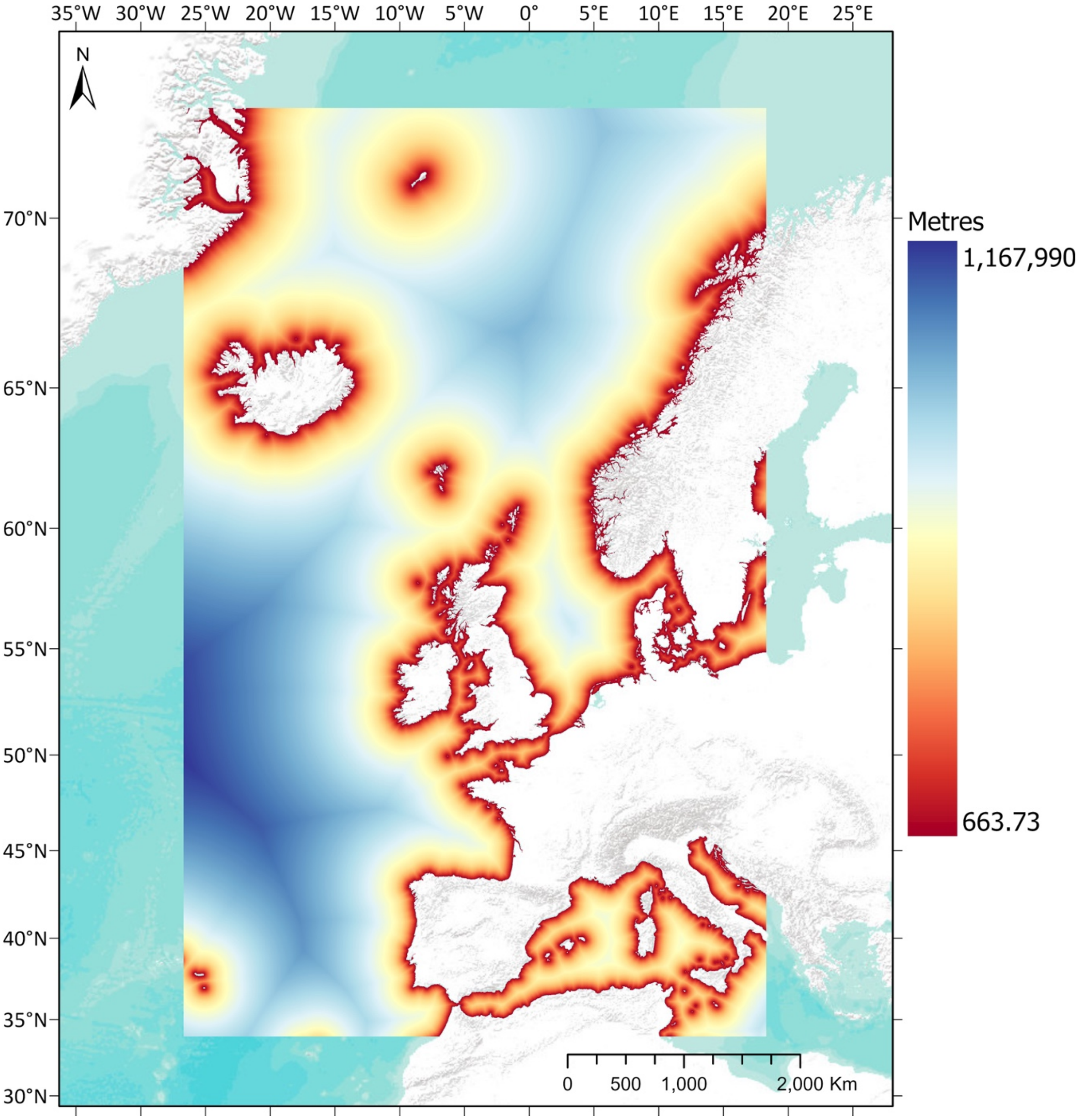


Figure H-4 Map of distance to land used in study, averaged from 1979 to 2018. Calculated using Distance to Land ARCGIS Tool.

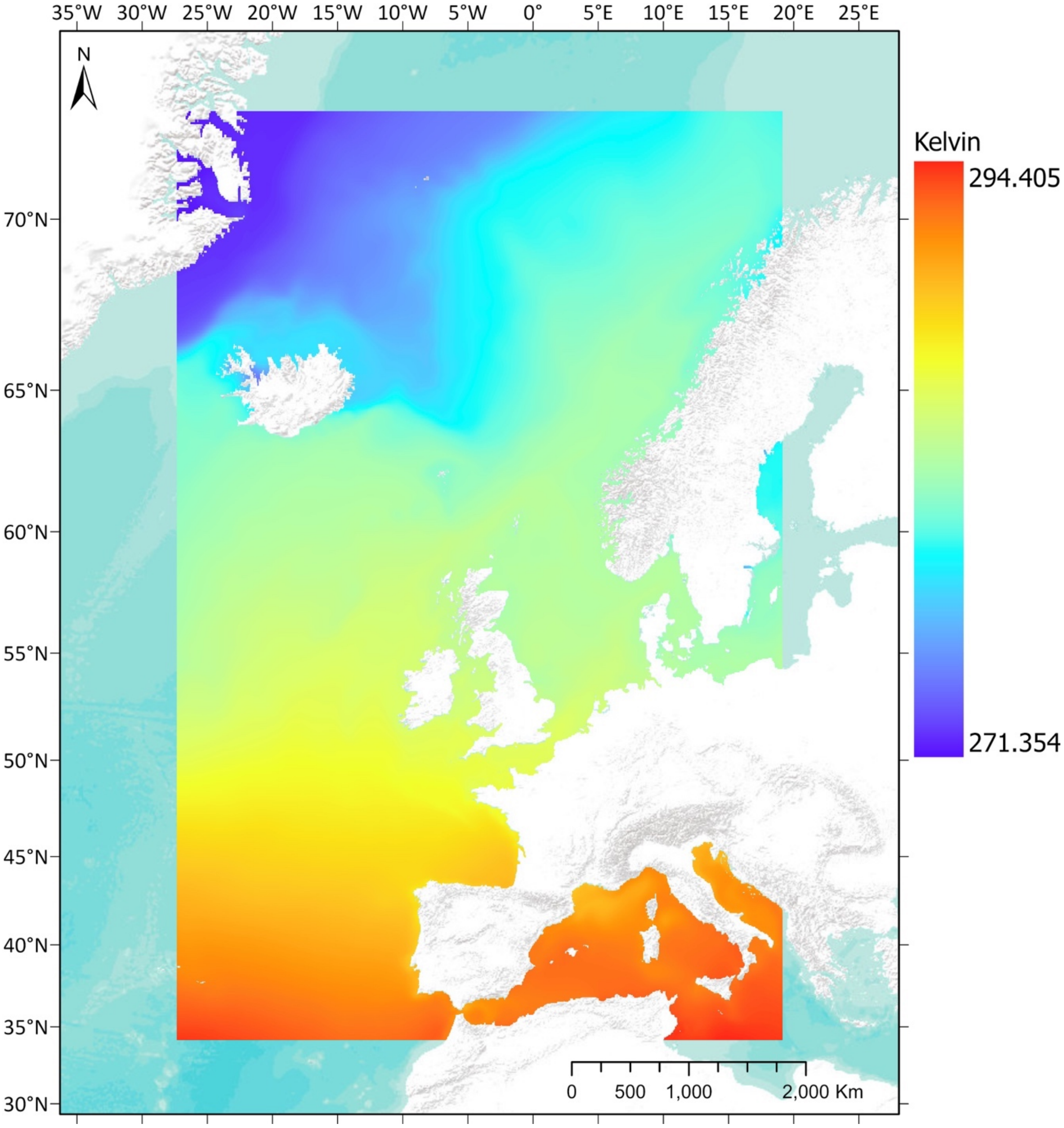


Figure H-5 Map of sea surface temperature used in study, averaged from 1979 to 2018. (NEODAAS, 2022)

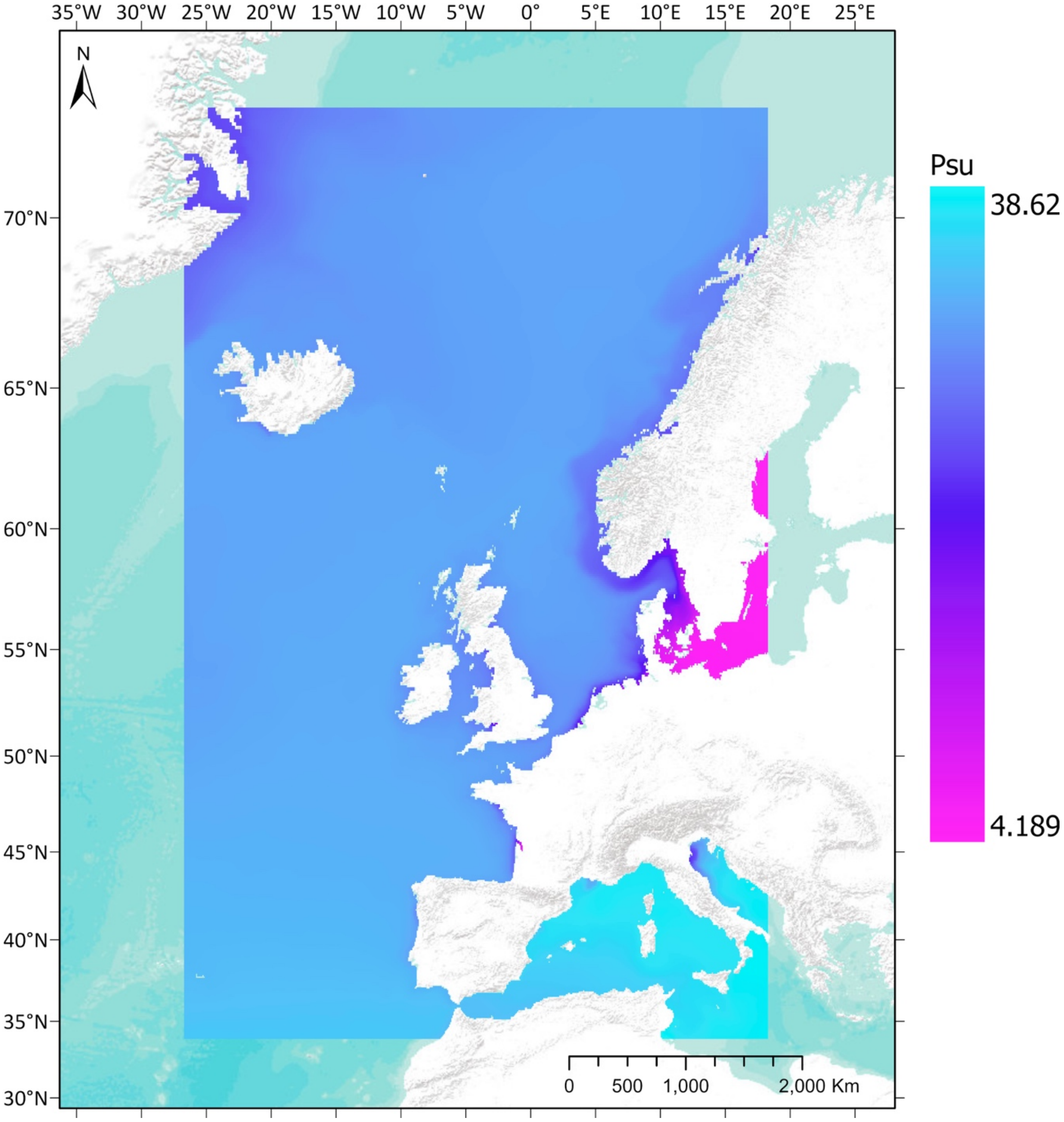


Figure H-6 Map of salinity used in study, averaged from 1979 to 2018. Original data from Copernicus (Lea, 2015).

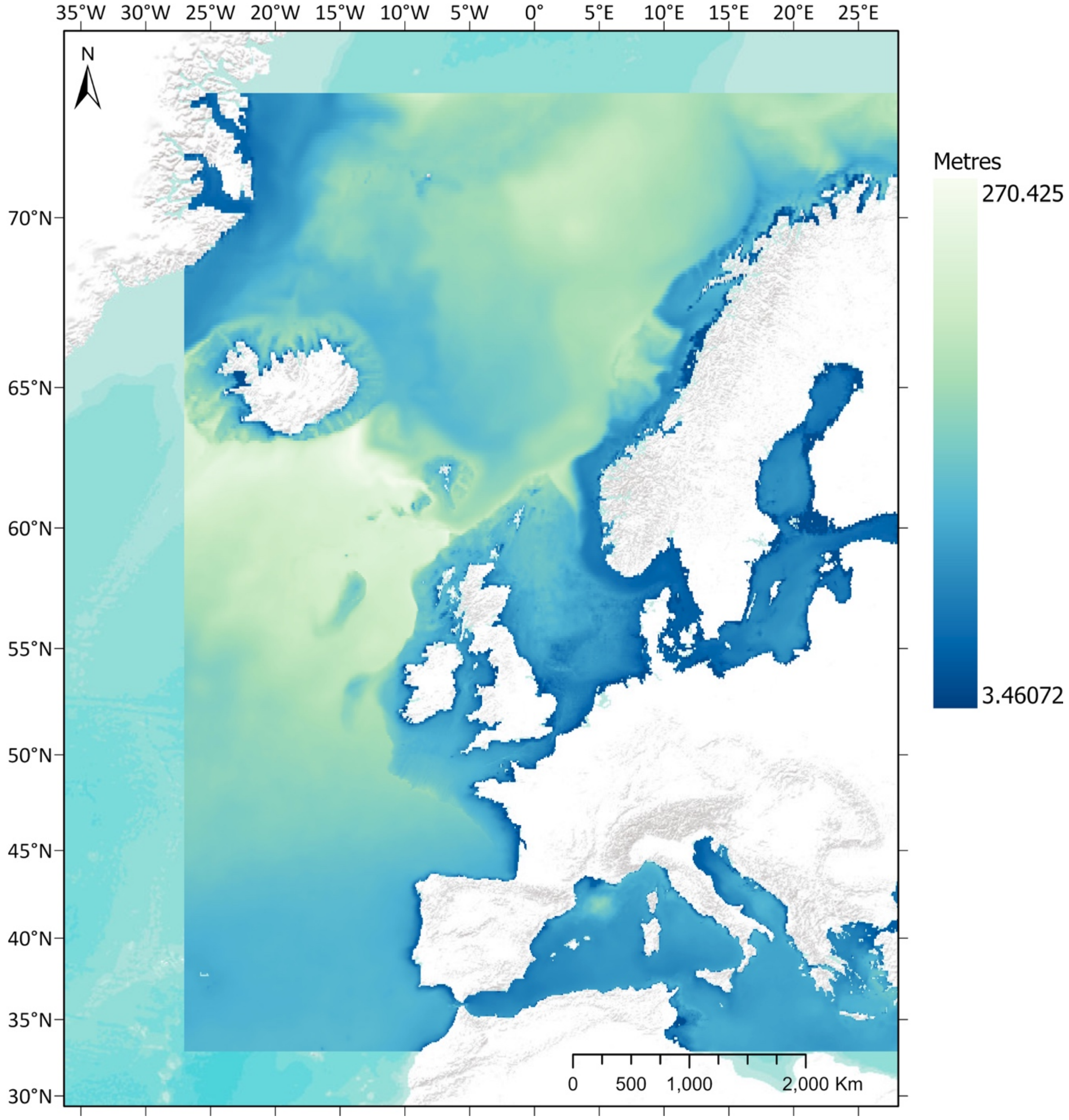


Figure H-7 Map of density ocean mixed layer thickness used in study, averaged from 1979 to 2018. Original data from Copernicus (Lea, 2015).

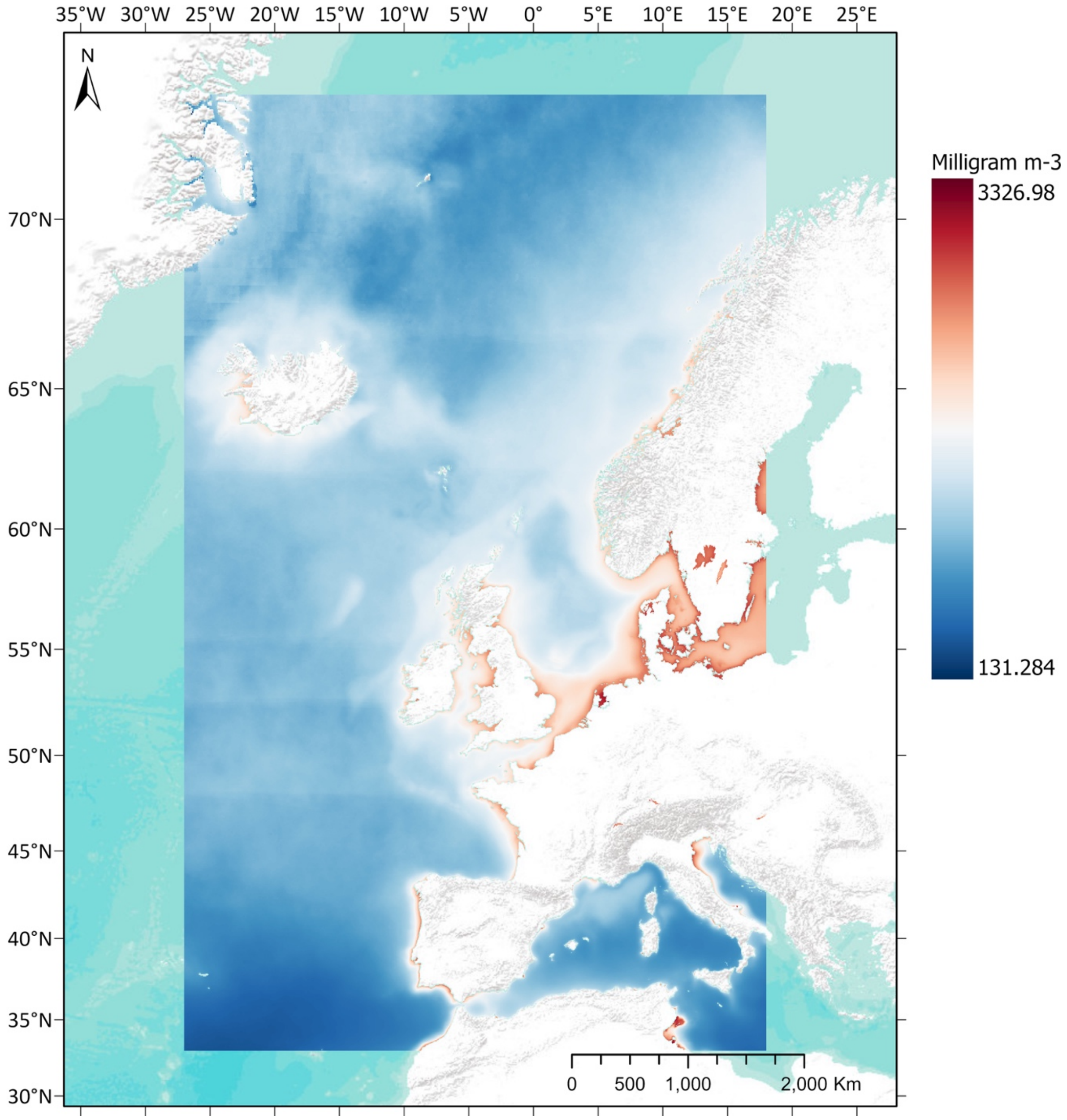


Figure H-8 Map of primary productivity, averaged from 1979 to 2018. Original data from Copernicus (Lea, 2015).

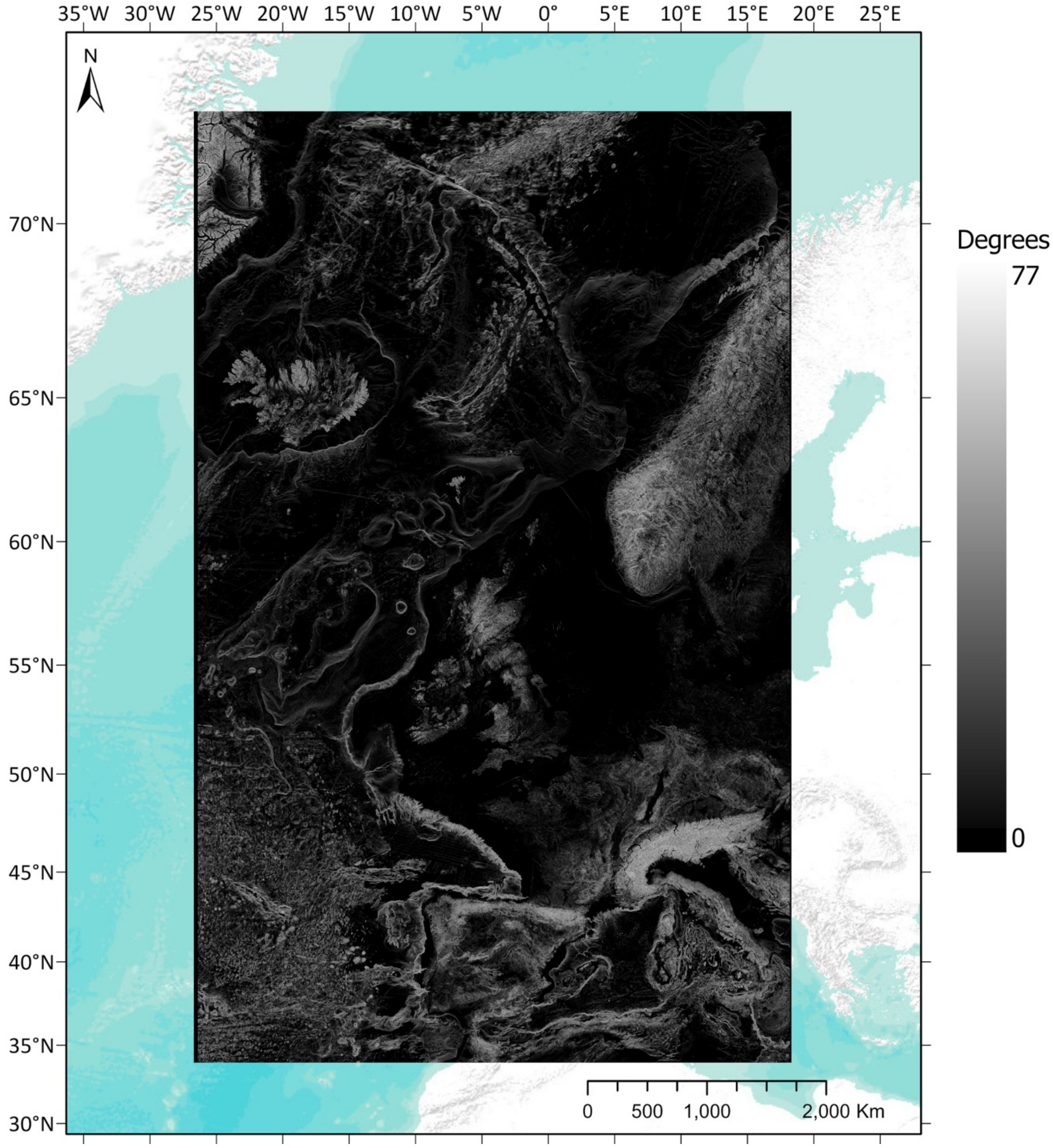
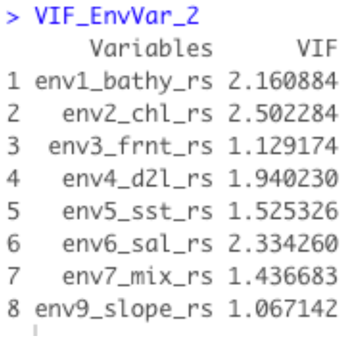


Figure H-9 Map of slope, averaged from 1979 to 2018. Calculated from GEBCO using Slope tool.

Appendix I VIF Outputs

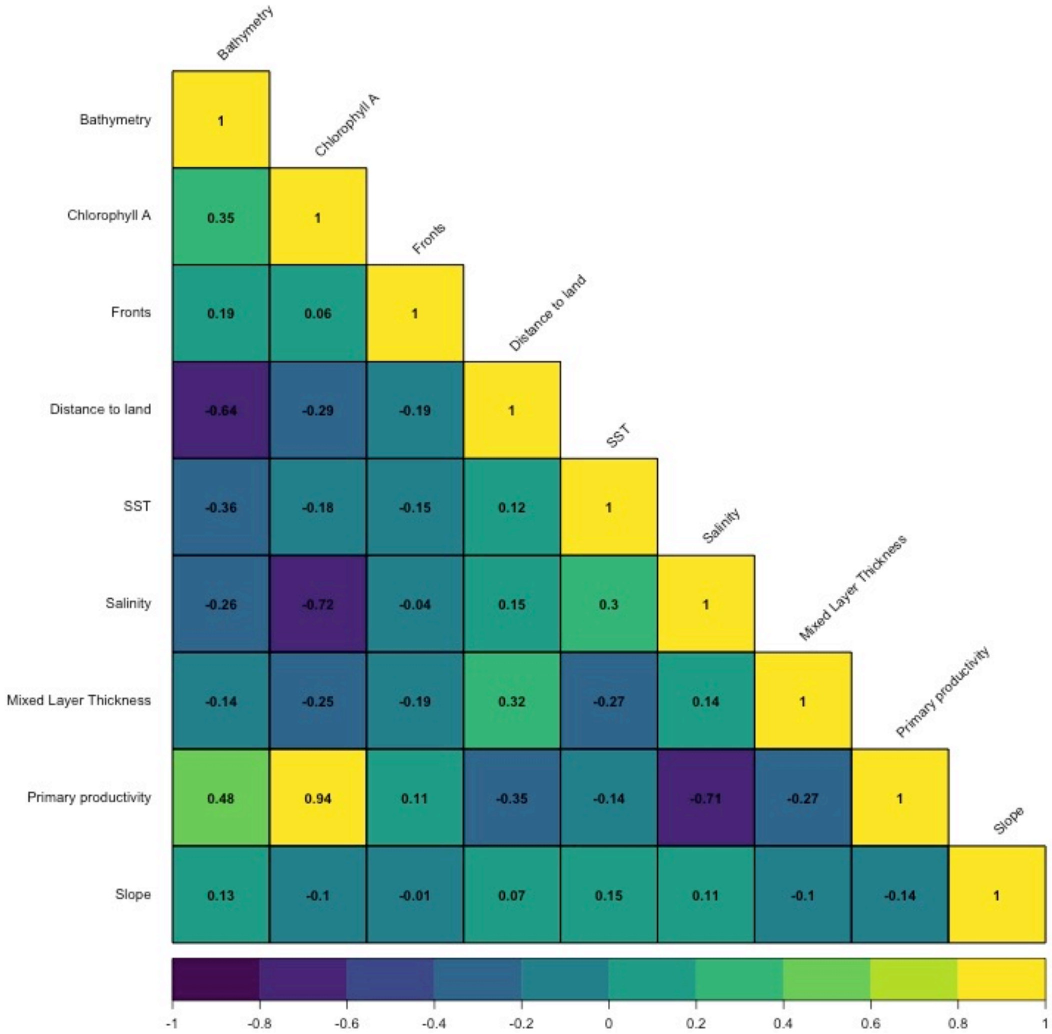
```
> VIF_EnvVar_1
      Variables       VIF
1 env1_bathy_rs  2.494523
2   env2_chl_rs 10.388065
3  env3_frnt_rs  1.140955
4   env4_d2l_rs  1.951260
5   env5_sst_rs  1.646633
6   env6_sal_rs  2.782843
7   env7_mix_rs  1.430508
8    env8_pp_rs 12.661083
9 env9_slope_rs  1.098944
```

Figure I-1 VIF output for all variables – bathymetry, chlorophyll a, fronts, distance to land, sea surface temperature, salinity, mixed layer thickness, primary productivity and slope.



```
> VIF_EnvVar_2
      Variables      VIF
1 env1_bathy_rs 2.160884
2   env2_chl_rs 2.502284
3  env3_frnt_rs 1.129174
4   env4_d2l_rs 1.940230
5   env5_sst_rs 1.525326
6   env6_sal_rs 2.334260
7   env7_mix_rs 1.436683
8 env9_slope_rs 1.067142
```

Figure I-2 VIF output for new set variables (no primary productivity) - bathymetry, chlorophyll *a*, fronts, distance to land, sea surface temperature, salinity, mixed layer thickness, primary productivity and slope.



	Bathymetry	Chlorophyll A	Fronts	Distance to land	SST	Salinity	Mixed Layer Thickness	Primary productivity	Slope
Bathymetry	1								
Chlorophyll A	0.35	1							
Fronts	0.19	0.06	1						
Distance to land	-0.64	-0.29	-0.19	1					
SST	-0.36	-0.18	-0.15	0.12	1				
Salinity	-0.26	-0.72	-0.04	0.15	0.3	1			
Mixed Layer Thickness	-0.14	-0.25	-0.19	0.32	-0.27	0.14	1		
Primary productivity	0.48	0.94	0.11	-0.35	-0.14	-0.71	-0.27	1	
Slope	0.13	-0.1	-0.01	0.07	0.15	0.11	-0.1	-0.14	1

Figure I-3 Correlation matrix for all variables. PP 0.94, very high.

Appendix J SSDM Map

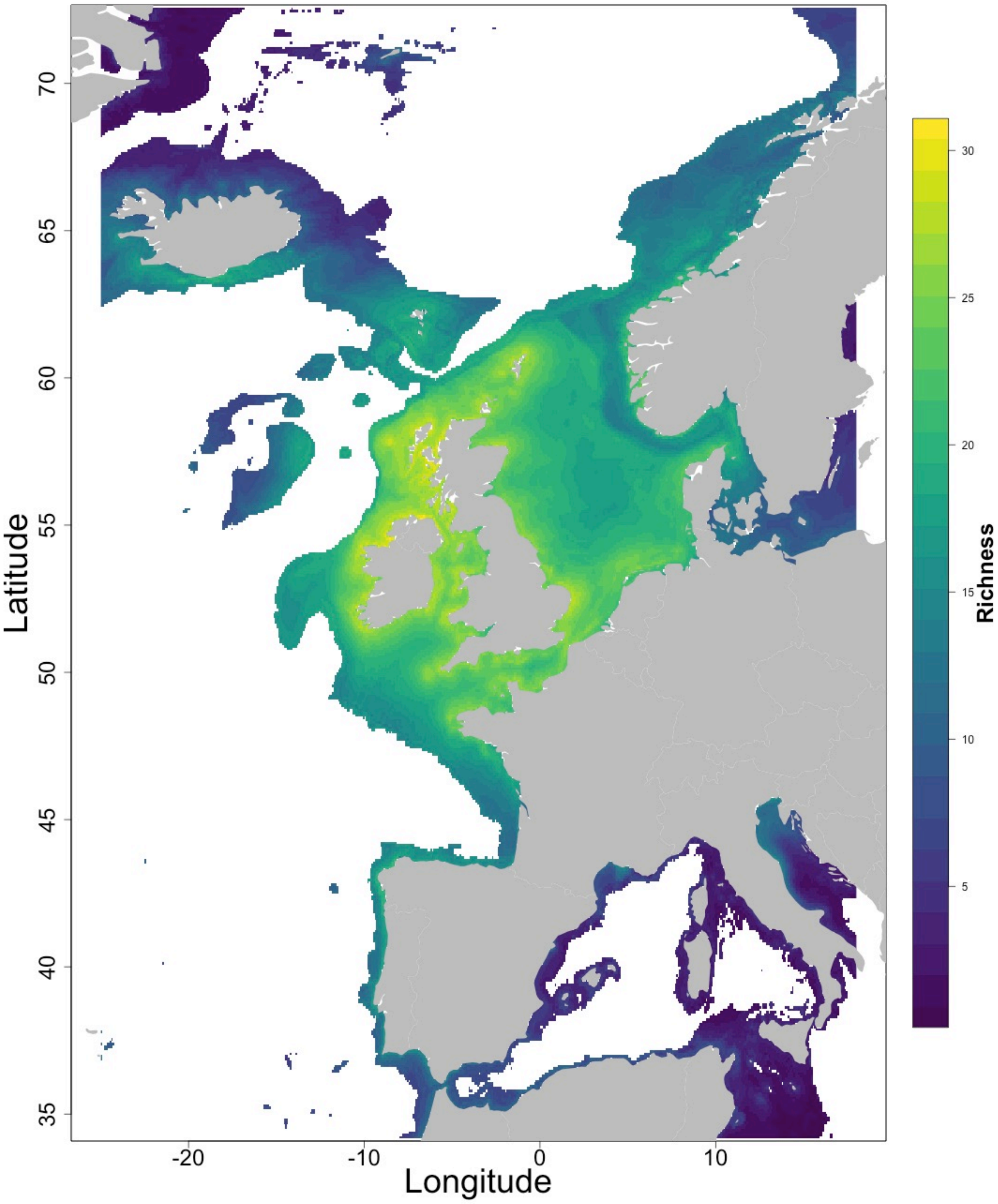
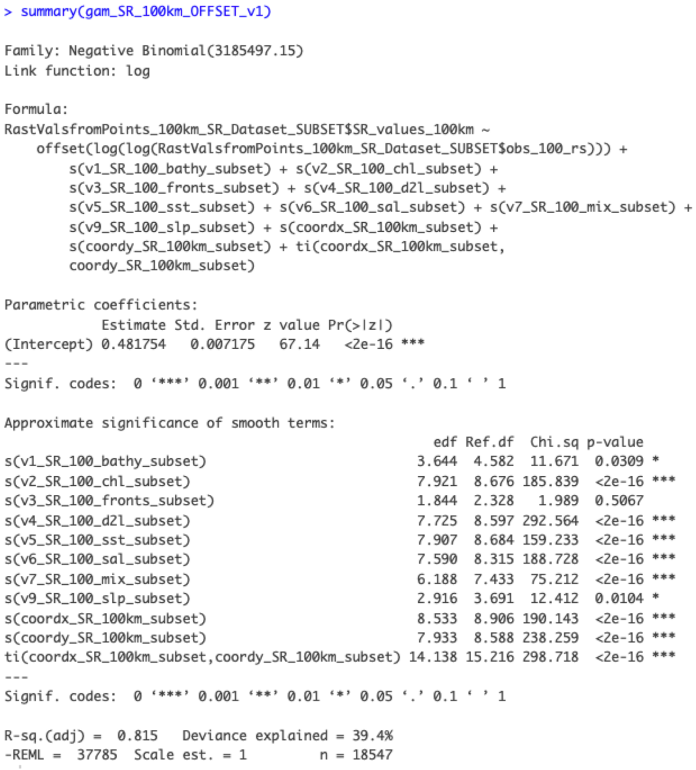


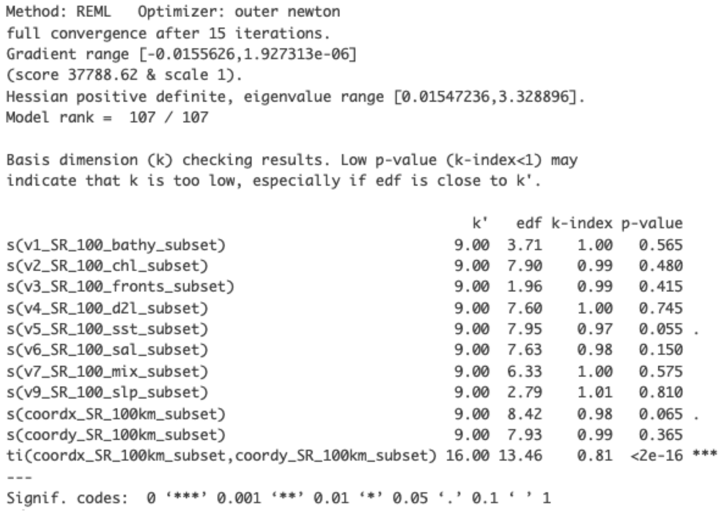
Figure J-1 Map of species richness from SSDM model.

Appendix K GAM outputs: Raw species richness



```
> summary(gam_SR_100km_OFFSET_v1)

Family: Negative Binomial(3185497.15)
Link function: log

Formula:
RastValsfromPoints_100km_SR_Dataset_SUBSET$SR_values_100km ~
    offset(log(log(RastValsfromPoints_100km_SR_Dataset_SUBSET$obs_100_rs))) +
        s(v1_SR_100_bathy_subset) + s(v2_SR_100_chl_subset) +
        s(v3_SR_100_fronts_subset) + s(v4_SR_100_d2l_subset) +
        s(v5_SR_100_sst_subset) + s(v6_SR_100_sal_subset) + s(v7_SR_100_mix_subset) +
        s(v9_SR_100_slp_subset) + s(coordx_SR_100km_subset) +
        s(coordy_SR_100km_subset) + ti(coordx_SR_100km_subset,
        coordy_SR_100km_subset)

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.481754   0.007175   67.14   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                                                  edf Ref.df  Chi.sq p-value
s(v1_SR_100_bathy_subset)                       3.644  4.582  11.671  0.0309 *
s(v2_SR_100_chl_subset)                         7.921  8.676 185.839  <2e-16 ***
s(v3_SR_100_fronts_subset)                      1.844  2.328   1.989  0.5067
s(v4_SR_100_d2l_subset)                         7.725  8.597 292.564  <2e-16 ***
s(v5_SR_100_sst_subset)                         7.907  8.684 159.233  <2e-16 ***
s(v6_SR_100_sal_subset)                         7.590  8.315 188.728  <2e-16 ***
s(v7_SR_100_mix_subset)                         6.188  7.433  75.212  <2e-16 ***
s(v9_SR_100_slp_subset)                         2.916  3.691  12.412  0.0104 *
s(coordx_SR_100km_subset)                       8.533  8.906 190.143  <2e-16 ***
s(coordy_SR_100km_subset)                       7.933  8.588 238.259  <2e-16 ***
ti(coordx_SR_100km_subset,coordy_SR_100km_subset) 14.138 15.216 298.718  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.815   Deviance explained = 39.4%
-REML =  37785  Scale est. = 1         n = 18547
```

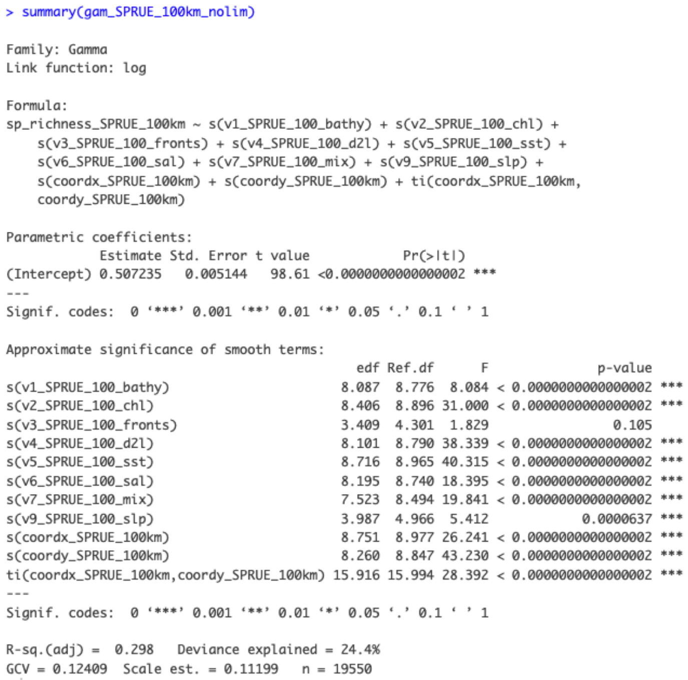
Figure K-1 Output from summary (GAM) of raw species richness with variables.



```
Method: REML   Optimizer: outer newton
full convergence after 15 iterations.
Gradient range [-0.0155626,1.927313e-06]
(score 37788.62 & scale 1).
Hessian positive definite, eigenvalue range [0.01547236,3.328896].
Model rank =  107 / 107

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                                                     k'   edf k-index p-value
s(v1_SR_100_bathy_subset)                          9.00  3.71    1.00   0.565
s(v2_SR_100_chl_subset)                            9.00  7.90    0.99   0.480
s(v3_SR_100_fronts_subset)                         9.00  1.96    0.99   0.415
s(v4_SR_100_d2l_subset)                            9.00  7.60    1.00   0.745
s(v5_SR_100_sst_subset)                            9.00  7.95    0.97   0.055 .
s(v6_SR_100_sal_subset)                            9.00  7.63    0.98   0.150
s(v7_SR_100_mix_subset)                            9.00  6.33    1.00   0.575
s(v9_SR_100_slp_subset)                            9.00  2.79    1.01   0.810
s(coordx_SR_100km_subset)                          9.00  8.42    0.98   0.065 .
s(coordy_SR_100km_subset)                          9.00  7.93    0.99   0.365
ti(coordx_SR_100km_subset,coordy_SR_100km_subset) 16.00 13.46    0.81  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

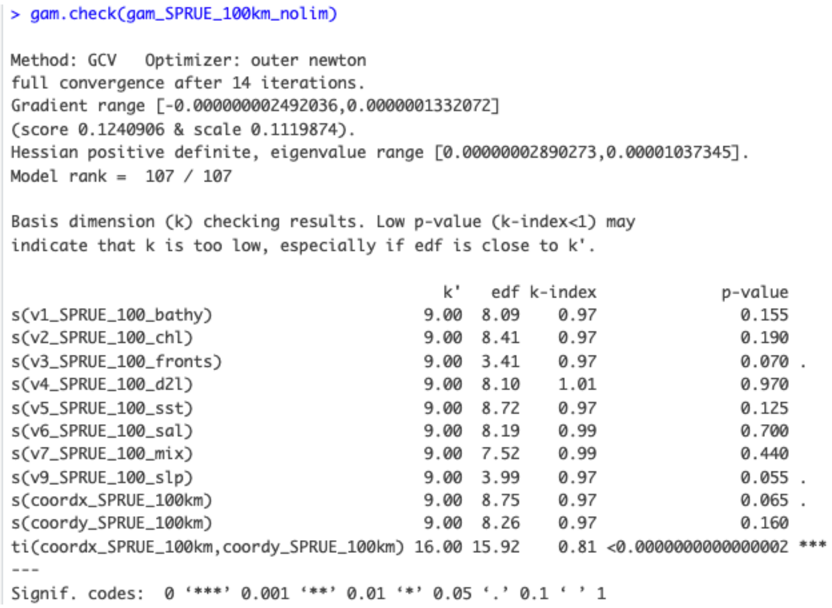
Figure K-2 Output from gam. check of raw species richness with variables.

Appendix L GAM Outputs: SRPUE



```
> summary(gam_SPRUE_100km_nolim)

Family: Gamma
Link function: log

Formula:
sp_richness_SPRUE_100km ~ s(v1_SPRUE_100_bathy) + s(v2_SPRUE_100_chl) +
    s(v3_SPRUE_100_fronts) + s(v4_SPRUE_100_d2l) + s(v5_SPRUE_100_sst) +
    s(v6_SPRUE_100_sal) + s(v7_SPRUE_100_mix) + s(v9_SPRUE_100_slp) +
    s(coordx_SPRUE_100km) + s(coordy_SPRUE_100km) + ti(coordx_SPRUE_100km,
    coordy_SPRUE_100km)

Parametric coefficients:
            Estimate Std. Error t value            Pr(>|t|)
(Intercept) 0.507235   0.005144   98.61 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                                          edf Ref.df      F             p-value
s(v1_SPRUE_100_bathy)                   8.087  8.776  8.084 < 0.0000000000000002 ***
s(v2_SPRUE_100_chl)                     8.406  8.896 31.000 < 0.0000000000000002 ***
s(v3_SPRUE_100_fronts)                  3.409  4.301  1.829                0.105
s(v4_SPRUE_100_d2l)                     8.101  8.790 38.339 < 0.0000000000000002 ***
s(v5_SPRUE_100_sst)                     8.716  8.965 40.315 < 0.0000000000000002 ***
s(v6_SPRUE_100_sal)                     8.195  8.740 18.395 < 0.0000000000000002 ***
s(v7_SPRUE_100_mix)                     7.523  8.494 19.841 < 0.0000000000000002 ***
s(v9_SPRUE_100_slp)                     3.987  4.966  5.412            0.0000637 ***
s(coordx_SPRUE_100km)                   8.751  8.977 26.241 < 0.0000000000000002 ***
s(coordy_SPRUE_100km)                   8.260  8.847 43.230 < 0.0000000000000002 ***
ti(coordx_SPRUE_100km,coordy_SPRUE_100km) 15.916 15.994 28.392 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.298   Deviance explained = 24.4%
GCV = 0.12409  Scale est. = 0.11199   n = 19550
```

Figure L-1 Output from summary(GAM) of SRPUE with variables.



```
> gam.check(gam_SPRUE_100km_nolim)

Method: GCV   Optimizer: outer newton
full convergence after 14 iterations.
Gradient range [-0.000000002492036,0.0000001332072]
(score 0.1240906 & scale 0.1119874).
Hessian positive definite, eigenvalue range [0.00000002890273,0.00001037345].
Model rank =  107 / 107

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                                             k'   edf k-index             p-value
s(v1_SPRUE_100_bathy)                      9.00  8.09    0.97               0.155
s(v2_SPRUE_100_chl)                        9.00  8.41    0.97               0.190
s(v3_SPRUE_100_fronts)                     9.00  3.41    0.97               0.070 .
s(v4_SPRUE_100_d2l)                        9.00  8.10    1.01               0.970
s(v5_SPRUE_100_sst)                        9.00  8.72    0.97               0.125
s(v6_SPRUE_100_sal)                        9.00  8.19    0.99               0.700
s(v7_SPRUE_100_mix)                        9.00  7.52    0.99               0.440
s(v9_SPRUE_100_slp)                        9.00  3.99    0.97               0.055 .
s(coordx_SPRUE_100km)                      9.00  8.75    0.97               0.065 .
s(coordy_SPRUE_100km)                      9.00  8.26    0.97               0.160
ti(coordx_SPRUE_100km,coordy_SPRUE_100km) 16.00 15.92    0.81 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure L-2 Output from gamcheck of GAM of SRPUE with variables.

Appendix M R Code

Overview of scripts

Script	Description of script	Chapter of relevance
Script 1: Data set up	Setting up data for into R for working out species richness and SPRUE.	4
Script 2: Species effort calculations	Plotting out raw, log-log and semi log plots.	4
Script 3: Scaling by effort	Species richness and SRPUE calculations from combined ESAS/ORCA dataset using grid cells.	4
Script 4: Density plots	Density plots for species richness and SRPUE.	4
Script 5: Splitting out by month	Splitting out species richness and SRPUE per month (e.g. Jan – Dec combined for 39 years) into rasters.	4
Script 6: Splitting out by years	Splitting out species richness and SRPUE per year (e.g. 1979 – 2018) into rasters.	4
Script 7: Splitting out by month and year	Splitting out species richness and SRPUE per month and year (e.g. Jan 1979, Feb 1979…) into rasters.	4
Script 8: Loading data from rasters for analysis	Quick method of re-loading data rather than running Script 5-7 each time wanting to load data into the environment.	4
Script 9: Levelplot maps	Maps of rasters.	4
Script 10: Convert data to dataframes for subsequent analysis	Converting rasters from Script 5-7 into dataframes for use in analysis.	4
Script 11: Seasonal Decomposition of Time Series (STL) analysis	Seasonal Decomposition of Time Series (STL) analysis.	4
Script 12: Richness on SD scale	Putting species richness and SRPUE on a standard deviation unit scale for comparison.	4

Script 13: Setting up for GAMs	Data set up for GAMS for use in chapter 5, exploring environmental drivers.	5
Script 14: GAMS	GAMS of Species richness and SRPUE. Variable importance for each. Predicted maps using GAMS.	5
Script 15: Histograms of rasters	Histograms from predicted raster maps.	5
Script 16: SSDM	Stacked species distribution model approach using SSDM package.	5
Script 17: Variable importance from SSDM	Determining variable importance from SSDM.	5

Script 1: Data set up

```
### 1. Setup ###--------------------------------------------------------------------------------
# Load packages
library(sp)
library(raster)
library(rgdal)
library(viridis)
options(scipen = 999)
# Make sure max memory is available given number of datapoints
# For Mac
# Open Terminal
# Paste below:
cd ~
touch .Renviron
open .Renviron
# Save following into that file
R_MAX_VSIZE=100Gb
# Save
# For PC
memory.limit(size = 100000000)
### 2. Set colour / pixel schemes for plots ### --------------------------------------------------------------------------------
maxpalette <- viridis(n = 30) # N = Maximum value from 100km rasters.
# Set pixel max for calculations from 1km resolutions - do after running code.
r1
# Dimensions are 15480288 for n-cells, therefore max pixels
pixmax <- 15480288
### 3. Load Data into R ### --------------------------------------------------------------------------------
# To note: make sure pointDropZ = TRUE - else adds extra dimension
# For mac:
points <-readOGR("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/data/ALL_Data_FINAL_LAEA.shp", "ALL_Data_FINAL_LAEA", pointDropZ = TRUE)
mapproj <- crs(points) # check correct projection - LAEA
# bring in outline data for countries
outlines <-readOGR("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/data/worldoutline.shp", "worldoutline", pointDropZ = TRUE)
crs(outlines)
# For Windows #
points <-readOGR(dsn = "data/ALL_Data_FINAL_LAEA.shp", "ALL_Data_FINAL_LAEA", pointDropZ = TRUE)
mapproj <- crs(points)
outlines <-readOGR(dsn = "data/worldoutline.shp", "worldoutline", pointDropZ = TRUE)
crs(outlines)
# To save figure of base points
pngfile <- "Figs/Points_Overall.png"
png(pngfile, width=729, height=729)   # open the file
plot(points, pch = 20, cex = 0.5, col = "darkslategrey", xlab = "X Coord", ylab = "Y Coord")
plot(outlines, add=TRUE, col = "darkgrey", lwd = 0.1, border = "gray50")
box(which = "plot", lty = "solid")
dev.off() # close the file
### 4. Species Richness & Observation Rasters - FOR ALL DATA ### -------------------------------------------------------------------
# a) For raster extent - at 100km2 #--------------------------------------------------------------------------------
# Create rasters of the points extent for each resolution
# Create blank raster of extent
r100 <- raster(points)
# Set resolution 100km2 – To note: Set resolution of the raster - set by length of sides, not area
res(r100) <- 10000
# Generate Species Richness
rich_100 <- rasterize(points, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
plot(rich_100)
# Count observations - number of observations - per grid cell
obs_100 <- rasterize(points, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )

### b) For raster extent - 10km2 --------------------------------------------------------------------------------
r10 <- raster(points)
res(r10) <- 3160  # 10km2 --> 3.16x3.16 = 3160 m
# Set resolution of the raster - set by length of sides, not area
# Species Richness
rich_10 <- rasterize(points, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
```

```
# Observations per cell
obs_10 <- rasterize(points, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
### c) For raster extent - 1 km2 ### -----------------------------------------------------------------------------------
r1 <- raster(points)
res(r1) <- 1000 # 1km2 --> 1000 x 1000 = 10,000
# Note: set resolution of the raster - set by length of sides, not area
# Species Richness RICHNESS
rich_1 <- rasterize(points, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
# Observations per cell
obs_1 <- rasterize(points, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
### 5. Save out main map richness plots ### ----------------------------------------------------------------------------
# Specify cutpoints for legend breaks
cuts_100 = seq(0,30,1)
pngfile <- "Figs/All/Richness_100km2.png"
# open the file
png(pngfile, width=729, height=729)
plot(rich_100, col = maxpalette, maxpixels = pixmax, breaks = cuts_100, xlab = "X Coord", ylab = "Y Coord")
plot(outlines, add=TRUE, col = "darkgrey", lwd = 0.1, border = "gray50")
dev.off()
# close the file
cuts_10 = seq(0,25,1)
pngfile <- "Figs/All/Richness_10km2.png"
png(pngfile, width=729, height=729)   # open the file
plot(rich_10, col = maxpalette[0:25], maxpixels = pixmax, breaks = cuts_10, xlab = "X Coord",  ylab = "Y Coord")
plot(outlines, add=TRUE, col = "darkgrey", lwd = 0.1, border = "gray50")
dev.off()
# close the file
cuts_1 = seq(0,19,1)
pngfile <- "Figs/All/Richness_1km2.png"
png(pngfile, width=729, height=729)   # open the file
plot(rich_1, col = maxpalette[0:19], maxpixels = pixmax, breaks = cuts_1, xlab = "X Coord", ylab = "Y Coord")
plot(outlines, add=TRUE, col = "darkgrey", lwd = 0.1, border = "gray50")
dev.off()
# close the file
### 6) Save out rasters ###-------------------------------------------------------------------------------
writeRaster(rich_100, "Tiffs/All/rich_100.tif", format="GTiff", overwrite=TRUE)
writeRaster(rich_10, "Tiffs/All/rich_10.tif", format="GTiff", overwrite=TRUE)
writeRaster(rich_1, "Tiffs/All/rich_1.tif", format="GTiff", overwrite=TRUE)
writeRaster(obs_100, "Tiffs/All/obs_100.tif", format="GTiff", overwrite=TRUE)
writeRaster(obs_10, "Tiffs/All/obs_10.tif", format="GTiff", overwrite=TRUE)
writeRaster(obs_1, "Tiffs/All/obs_1.tif", format="GTiff", overwrite=TRUE)
# Save extent rasters
pngfile <- "Figs/Grid100km2.png"
png(pngfile, width=729, height=729)   # open the file
plot(r100, col = viridis ,xlab = "X Coord", lty = 5,
     ylab = "Y Coord")
r100_vals <- setValues(r100, 1:ncell())
vals <- 1:ncell(r100)
r100_vals <- setValues(r100, vals)
vals <- 1:ncell(r10)
r10_vals <- setValues(r10, vals)
vals <- 1:ncell(r1)
r1_vals <- setValues(r1, vals)
writeRaster(r100_vals, "Tiffs/raster100", format="GTiff", overwrite=TRUE)
writeRaster(r10_vals, "Tiffs/raster10", format="GTiff", overwrite=TRUE)
writeRaster(r1_vals, "Tiffs/raster1", format="GTiff", overwrite=TRUE)
### End of script ###-------------------------------------------------------------------------------
```

Script 2: Species effort calculations

```
### 1. Setup -###---------------------------------------
# Make sure packages installed
library(sp)
library(raster)
library(rgdal)
library(viridis)
options(scipen = 999)
pixmax <- 15480288
#Using data (check if it is previously loaded into session)
obs_100
obs_10
obs_1
rich_100
rich_10
rich_1
# Can import rasters via:
obs_100 <- raster(“tiffs/All/obs_100.tif”)
obs_10 <- raster(“tiffs/All/obs_10.tif”)
obs_1 <- raster(“tiffs/All/obs_1.tif”)
rich_100 <- raster(“tiffs/All/rich_100.tif”)
rich_10 <- raster(“tiffs/All/rich_10.tif”)
rich_1 <- raster(“tiffs/All/rich_1.tif”)
# For info: Following Ugland et al 2003 reading – semilog scale versus log-log scale. Checking which is best fit.
# Rate of change = S / ln©
# Convert data from raster, need to make variables  matrix or list, so that can carry out model
richness100 <- getValues(rich_100)
observations100 <- getValues(obs_100)
richness10 <- getValues(rich_10)
observations10 <- getValues(obs_10)
richness1 <- getValues(rich_1)
observations1 <- getValues(obs_1)
# Plot raw data
par(mar = c(5,5, 5, 5))
pngfile <- “Figs/Survey_Effort_Calcs/Final/100_HistScatterOriginalData.png”
png(pngfile, width=1000, height=500)
par(mfrow = c(1,2))
hist(rich_100, main = “”, xlab = “Richness”)
plot(obs_100, rich_100, maxpixels = pixmax, ylab = “Richness”, xlab = “Observations”)
mtext(“Hist of Richness + Richness x Observations”, side = 3, line = -3 , outer = TRUE)
dev.off()
### 2) Get logs for each for further analysis ###------------------------------------------------------------------------------
logrichness100 <- log(richness100)
logrichness10 <- log(richness10)
logrichness1 <- log(richness1)
logobservations100 <- log(observations100)
logobservations10 <- log(observations10)
logobservations1 <- log(observations1)

### 3) On a semi-log scale ###-------------------------------------------------------------------------------
# 100km2
plot(logobservations100, richness100)
lm100_semilog <- lm(richness100 ~ logobservations100)
plot(lm100_semilog)
abline(lm100_semilog, col = "red")
# Check diagnostics plot
par( mfrow = c(2, 2) )
plot(lm100_semilog)
# 10km2
par(mfrow = c(1, 1))
plot(logobservations10, richness10)
lm10_semilog <- lm(richness10 ~ logobservations10)
abline(lm10_semilog, col = "red")
# Check diagnostics plot
par(mfrow = c(2, 2))
plot(lm10_semilog)
# 1km2
```

```
par(mfrow = c(1, 1))
plot(logobservations1, richness1)
lm1_semilog <- lm(richness1 ~ logobservations1)
abline(lm1_semilog, col = "red")
# Check diagnostics plot
par(mfrow = c(2, 2))
plot(lm1_loglog)
# Get R2 values for plots:
summary(lm100_semilog) # 0.7183
summary(lm10_semilog) # 0.6341
summary(lm1_semilog) # 0.5799
# Neat figure for all 3: --------------
# Labels for plots
lab100 <- expression('100 km' ^2)
lab10 <- expression('10 km' ^2)
lab1 <- expression('1 km' ^2)
par(mar = c(5,5, 5, 5))
pngfile <- "Figs/Survey_Effort_Calcs/Final/3semiloglm.png"
png(pngfile, width=1500, height=500)
par(mfrow = c(1,3))
plot(logobservations100, richness100, ylab = "Richness", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,30), xaxs = "i", yaxs = "i",cex.lab = 1.5)
abline(lm100_semilog, col = "red")
mtext(lab100, side = 3, line = -3 , outer = TRUE, adj = 0.15)
plot(logobservations10, richness10, ylab = "Richness", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,30), xaxs = "i", yaxs = "i",cex.lab = 1.5)
abline(lm10_semilog, col = "red")
mtext(lab10, side = 3, line = -3, outer = TRUE)
plot(logobservations1, richness1, ylab = "Richness", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,30), xaxs = "i", yaxs = "i",cex.lab = 1.5)
abline(lm1_semilog, col = "red")
mtext(lab1, side = 3, line = -3, outer = TRUE, adj = 0.85)
dev.off()
### 4. On a log-log scale:### -----------------------------------------------------------------------------
# 100km2
plot(logobservations100, logrichness100)
lm100_loglog <- lm(logrichness100 ~ logobservations100)
abline(lm100_loglog, col = "red")
# Check diagnostics plot
par( mfrow = c(2, 2) )
plot(lm100_loglog)
# 10km2
par(mfrow = c(1, 1))
plot(logobservations10, logrichness10)
lm10_loglog <- lm(logrichness10 ~ logobservations10)
abline(lm10_loglog, col = "red")
# Check diagnostics plot
par(mfrow = c(2, 2))
plot(lm10_loglog)
# 1km2
par(mfrow = c(1, 1))
plot(logobservations1, logrichness1)
lm1_loglog <- lm(logrichness1 ~ logobservations1)
abline(lm1_loglog, col = "red")
# Check diagnostics plot
par(mfrow = c(2, 2))
plot(lm1_loglog)
# Get R2 values for plots:
summary(lm100_loglog) # 0.8072
summary(lm10_loglog) # 0.6937
summary(lm1_loglog) # 0.6273

### 5. Figures ###------------------------------------------------------------------------------------------
# Labels for plots
lab100 <- expression('100 km' ^2)
lab10 <- expression('10 km' ^2)
lab1 <- expression('1 km' ^2)
par(mar = c(5,5, 5, 5))
pngfile <- "Figs/Survey_Effort_Calcs/Final/3logloglm.png"
png(pngfile, width=1500, height=500)
```

```
par(mfrow = c(1,3))

plot(logobservations100, logrichness100, ylab = "Log Richness", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,3.5), xaxs = "i", yaxs = "i",cex.lab = 1.5, col = "black")
abline(lm100_loglog, col = "red")
mtext(lab100, side = 3, line = -3 , outer = TRUE, adj = 0.15)

plot(logobservations10, logrichness10, ylab = "Log Richness", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,3.5), xaxs = "i", yaxs = "i",cex.lab = 1.5)
abline(lm10_loglog, col = "red")
mtext(lab10, side = 3, line = -3, outer = TRUE)

plot(logobservations1, logrichness1, ylab = "Log Richness", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,3.5), xaxs = "i", yaxs = "i",cex.lab = 1.5)
abline(lm1_loglog, col = "red")
mtext(lab1, side = 3, line = -3, outer = TRUE, adj = 0.85)
dev.off()

### 3 x 3 plot for corrections ###-------------------------------------------------------------------------------------------------------
viridis(3)
# Labels for plots
lab100 <- expression('100 km' ^2)
lab10 <- expression('10 km' ^2)
lab1 <- expression('1 km' ^2)

pngfile <- "Figs/Survey_Effort_Calcs/Final/3x3_CorrectionsPlot.png"
png(pngfile, width=2000, height=1500)
par(mfrow = c(3,3)) # 3 rows x 3 columns

# Row 1) Raw data
par(mar = c(5,6,4,1.5)) #bltr # left and top is higher due to label
plot(observations100, richness100, ylab = "Raw Richness", xlab = "Observations",ylim = c(0,30), xaxs = "i", yaxs = "i", cex.lab = 3, cex.axis = 2.2, pch = 20, cex = 0.5, col = "#440154FF")
par(mar = c(5,3,4,1.5))#bltr
plot(observations10, richness10, ylab = "", yaxt="t", xlab = "Observations", ylim = c(0,30), xaxs = "i", yaxs = "i",cex.lab = 3, cex.axis = 2.2, pch = 20, cex = 0.5, col = "#440154FF")
plot(observations1, richness1, ylab = "", yaxt="t",xlab = "Observations", ylim = c(0,30), xaxs = "i", yaxs = "i",cex.lab = 3,cex.axis = 2.2, pch = 20, cex = 0.5, col = "#440154FF")

# Row 2) semi-log
par(mar = c(5,6,1,1.5)) #bltr # left is higher due to label
plot(logobservations100, richness100, ylab = "Richness", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,30), xaxs = "i", yaxs = "i",cex.lab = 3, cex.axis = 2.2, pch = 20, cex = 0.5, col = "#3B528BFF")
#mtext(lab100, side = 3, line = -3 , outer = TRUE, adj = 0.15)
par(mar = c(5,3,1,1.5))#bltr
plot(logobservations10, richness10, ylab = "", yaxt="t", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,30), xaxs = "i", yaxs = "i",cex.lab = 3, cex.axis = 2.2, pch = 20, cex = 0.5, col = "#3B528BFF")
#mtext(lab10, side = 3, line = -3, outer = TRUE)
plot(logobservations1, richness1, ylab = "", yaxt="t", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,30), xaxs = "i", yaxs = "i",cex.lab = 3, cex.axis = 2.2, pch = 20, cex = 0.5, col = "#3B528BFF")
#mtext(lab1, side = 3, line = -3, outer = TRUE, adj = 0.85)

# Row 3) log log
par(mar = c(5,6,1,1.5)) #bltr # left is higher due to label
plot(logobservations100, logrichness100, ylab = "Log Richness", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,3.5), xaxs = "i", yaxs = "i",cex.lab = 3, cex.axis = 2.2, pch = 20, cex = 0.5, col = "#21908CFF")
mtext(lab100, side = 3, line = -3 , outer = TRUE, adj = 0.15, cex = 2)
par(mar = c(5,3,1,1.5))#bltr
plot(logobservations10, logrichness10, ylab = "", yaxt="t",, xlab = "Log Observations", xlim = c(0,10), ylim = c(0,3.5), xaxs = "i", yaxs = "i",cex.lab = 3, cex.axis = 2.2, pch = 20, cex = 0.5, col = "#21908CFF")
mtext(lab10, side = 3, line = -3, outer = TRUE, cex = 2)
plot(logobservations1, logrichness1, ylab = "",yaxt="t", xlab = "Log Observations", xlim = c(0,10), ylim = c(0,3.5), xaxs = "i", yaxs = "i",cex.lab = 3, cex.axis = 2.2, pch = 20, cex = 0.5, col = "#21908CFF")
mtext(lab1, side = 3, line = -3, outer = TRUE, adj = 0.85, cex = 2)
dev.off()

### -End of Script ###-------------------------------------------------------------------------------
```

Script 3: Scaling by effort

```
### 1. Setup -###---------------------------------------
# Make sure packages installed
library(raster)
pixmax <- 15480288
# Load rasters if needed:
rich_100 <- raster("Tiffs/All/rich_100.tif")
rich_10 <- raster("Tiffs/All/rich_10.tif")
rich_1 <- raster("Tiffs/All/rich_1.tif")

obs_100 <- raster("Tiffs/All/obs_100.tif")
obs_10 <- raster("Tiffs/All/obs_10.tif")
obs_1 <- raster("Tiffs/All/obs_1.tif")

# Write as function for reuse
# A custom function consists of a defined set of commands performed on a input object.
# Custom functions are particularly useful for tasks that need to be repeated over and over in the code.
# A simplified syntax for writing a custom function in R is:
# functionName <- function(variable1, variable2){WhatYouWantDone, WhatToReturn}
# My effort function = divide raw species richness values by the LOG of effort.
# Raster 1 must be richness, raster 2 observations.
# Observations is always 1 or more, therefore log will never be negative.
effortfactor <- function(r1, r2){
  fun1 <- r1 / (log(r2))
  return(fun1)
}
# New Measure (SPRUE) Rasters ------------------------------------------
rast_measure_100 <- overlay(rich_100, obs_100, fun=effortfactor, maxpixels = pixmax)
plot(rast_measure_100)
rast_measure_10 <- overlay(rich_10, obs_10, fun=effortfactor, maxpixels = pixmax)
plot(rast_measure_10)
rast_measure_1 <- overlay(rich_1, obs_1, fun=effortfactor, maxpixels = pixmax)
plot(rast_measure_1)
dev.off()
# Write out rasters to file
writeRaster(rast_measure_100, "Tiffs/All/rast_measure_100.tif", format="GTiff", overwrite=TRUE)
writeRaster(rast_measure_10, "Tiffs/All/rast_measure_10.tif", format="GTiff", overwrite=TRUE)
writeRaster(rast_measure_1, "Tiffs/All/rast_measure_1.tif", format="GTiff", overwrite=TRUE)
# Plot new histograms of "effort-scaled richness'
par(mar = c(5,5, 5, 5))
pngfile <- "Figs/Survey_Effort_Calcs/Hists_EffortScaled.png"
png(pngfile, width=1500, height=500)
par(mfrow = c(1,3))
hist(rast_measure_100, maxpix = pixmax, main = lab100, ylab = "Frequency", xlab = "Effort-scaled Species Richness", xlim = c(0,5), ylim = c(0,3000), xaxs = "i", yaxs = "i", cex.lab = 1.5)
hist(rast_measure_10, maxpix = pixmax, main = lab10, ylab = "Frequency", xlab = "Effort-scaled Species Richness", xlim = c(0,5), ylim = c(0,20000), xaxs = "i", yaxs = "i",cex.lab = 1.5)
hist(rast_measure_1, maxpix = pixmax, main = lab1, ylab = "Frequency", xlab = "Effort-scaled Species Richness", xlim = c(0,5), ylim = c(0,50000), xaxs = "i", yaxs = "i",cex.lab = 1.5)
dev.off()
### Remove any unused files
rm(rich_100, rich_10, rich_1, obs_100, obs_10, obs_1)
gc()
# clear memory
removeTmpFiles(h=0)
### -End of Script ###--------------------------------------------------------------------------------
```

Script 4: Density plots

```
# Setup Packages
library(ggplot2)
library(ggridges)
library(lattice)
library(latticeExtra)
library(hrbrthemes)  #imports library hrbrthemes
library(dplyr)       #imports the dplyr function
Library (tidyr)      #imports the tidyverse package
library(viridis)     #imports the library viridis
library(scales)

show_col(viridis_pal()(20))
# Load rasters if needed:
rast_measure_100 <- raster("Tiffs/All/rast_measure_100.tif")
rast_measure_10 <- raster("Tiffs/All/rast_measure_10.tif")
rast_measure_1 <- raster("Tiffs/All/rast_measure_1.tif")
# Density plots for rasters - from rasters, using base R-------------------------------------
dens100 <- density(rast_measure_100, maxpix = pixmax, plot = FALSE)
dens10 <- density(rast_measure_10, maxpix = pixmax, plot = FALSE)
dens1 <- density(rast_measure_1, maxpix = pixmax, plot = FALSE)
par(mar = c(5,5, 5, 5))
pngfile <- "Figs/Survey_Effort_Calcs/Densityplots.png"
png(pngfile, width=1500, height=500)
par(mfrow = c(1,3))

plot(dens100)
plot(dens10)
plot(dens1)
dev.off()
# Using GGPlots ---------------------------------------------------------------------------
# Preperation
# Need to get values from the raster - GGplot needs dataframes
vals_measure_100 <- getValues(rast_measure_100)
vals_measure_10 <- getValues(rast_measure_10)
vals_measure_1 <- getValues(rast_measure_1)
# Add Resolution to column for splitting by group
vals_measure_100_lab <- cbind(vals_measure_100, "100")
vals_measure_10_lab <- cbind(vals_measure_10, "10")
vals_measure_1_lab <- cbind(vals_measure_1, "1")
# Name column names
colnames(vals_measure_100_lab) <- c("Measure","Res")
colnames(vals_measure_10_lab) <- c("Measure","Res")
colnames(vals_measure_1_lab) <- c("Measure","Res")
# Combined to one matrix
vals_allcombo <- rbind(vals_measure_100_lab, vals_measure_10_lab, vals_measure_1_lab)
# Remove NA's from the data
library(IDPmisc)
mydata <- NaRV.omit(vals_allcombo)
# Check data type and levels - needs to be dataframe
str(mydata)
# Convert to df if needed.
mydata <- as.data.frame(mydata, row.names = NULL, optional = FALSE,
                        make.names = TRUE,
                        stringsAsFactors = default.stringsAsFactors())
is.data.frame(mydata)
# Make data numeric for ggplot success
as.numeric(mydata$Measure)
Legend_1 <- expression('Resolution (km'^2~')')
# Plots
# GGplot code for Density Plot - overlapped
p1 <-
  ggplot(data = mydata, aes(
    x = as.numeric(Measure),
    group = Res,
    fill = Res
  )) +
```

```
  geom_density(adjust = 1.5, alpha = 0.3) +
  theme_classic() +
  scale_fill_manual(values = c("#32648EFF", "#1F968BFF", "#74D055FF")) +
  scale_x_continuous(name = "Richness Measure",
                     limits = c(0, 5),
                     expand = c(0, 0)) +
  scale_y_continuous(name = "Density",
                     limits = c(0, 1.5),
                     expand = c(0, 0)) +
  labs(fill = Legend_1) +
  theme(
    legend.position = c(0.8, 0.8),
    legend.key.size = unit(1.5, "lines"),
    legend.text = element_text(colour = "black", size = 16),
    legend.title = element_text(colour = "black", size = 16),
    panel.spacing = unit(0.1, "lines"),
    axis.ticks.x = element_blank(),
    axis.title.x = element_text(
      color = "Black",
      size = 14,
      face = "bold",
      hjust = 0.5,
      vjust = 0.4
    ),
    axis.title.y = element_text(
      color = "Black",
      size = 14,
      face = "bold",
      hjust = 0.5,
      vjust = 0.1
    ),
    axis.text.x = element_text(size = 16),
    axis.text.y = element_text(size = 16),
    plot.title = element_text(
      color = "Black",
      size = 14,
      face = "bold.italic",
      hjust = 0.5
    ),
    strip.text.x = element_text(
      color = "Black",
      size = 20,
      hjust = 0.5
    ),
    plot.margin = unit(c(1, 1, 1, 1), "cm")
  ) #top, right, bottom, left

# GGplot code for Density Plot All seperate
p2 <-
  ggplot(data = mydata, aes(
    x = as.numeric(Measure),
    group = Res,
    fill = Res)) +
  geom_density(adjust = 1.5, alpha = 0.5) +
  scale_fill_manual(values=c("#32648EFF", "#1F968BFF", "#74D055FF"))+
  scale_x_continuous(name = "Richness Measure", limits = c(0, 5)) +
  scale_y_continuous(name = "Density", limits = c(0, 1.5), expand = c(0,0)) +
  theme_classic() +
  facet_wrap(~ Res) +
  theme(
    legend.position = "none",
    panel.spacing = unit(2, "lines"),
    axis.ticks.x = element_blank(),
    axis.title.x = element_text(
      color = "Black",
      size = 14,
      face = "bold",
      hjust = 0.5,
      vjust = 0.5),
    axis.title.y = element_text(
      color = "Black",
      size = 14,
      face = "bold",
      hjust = 0.5,
      vjust = 0.1),
    axis.text.x = element_text(size = 12),
```

```
    axis.text.y = element_text(size = 16),
    plot.title = element_text(
      color = "Black",
      size = 14,
      face = "bold.italic",
      hjust = 0.5),
    strip.text.x = element_text(
      color = "Black",
      size = 20,
      hjust = 0.5))

p3 <- ggplot(data = mydata, aes(
  x = as.numeric(Measure),
  y = Res,
  fill = factor(stat(quantile)))) +
  stat_density_ridges(
    geom = "density_ridges_gradient",
    calc_ecdf = TRUE,
    quantiles = c(0.025, 0.975),
    scale = 0.5) +
  scale_fill_manual(
    name = "Probability",
    values = c("#FF0000A0", "#A0A0A0A0", "#0000FFA0"),
    labels = c("(0, 0.025]", "(0.025, 0.975]", "(0.975, 1]")) +
  labs(x = "Richness Measure", y = "Resolution")+
  theme_minimal() +
  theme(axis.text.x = element_text(size = 12),
    axis.text.y = element_text(size = 16),
    legend.position = "right",
    axis.title.x = element_text(
      color = "Black",
      size = 14,
      face = "bold",
      hjust = 0.5,
      vjust = 0.5),
    axis.title.y = element_text(
      color = "Black",
      size = 14,
      face = "bold",
      hjust = 0.5,
      vjust = 0.1))+
  scale_y_discrete(expand = expansion(mult = c(0.01, 0.4)))

# Plot out plots:
# P1 - Overlap, transparency
par(mar = c(5,5, 5, 5))
par(mfrow = c(1,1))
pngfile <- "Figs/Survey_Effort_Calcs/Density_plots/Density_All3_Overlap.png"
png(pngfile, width=1000, height=800)
p1
dev.off()


# P2 - Overlap, transparency
par(mar = c(5,5, 5, 5))
pngfile <- "Figs/Survey_Effort_Calcs/Density_plots/Density_All3_Sep.png"
png(pngfile, width=1000, height=500)
par(mfrow = c(1,1))
p2
dev.off()

# P3 Density ridges plot - 3 layers
par(mar = c(5,5, 5, 5))
pngfile <- "Figs/Survey_Effort_Calcs/Density_plots/Density_All3_RidgesPanel.png"
png(pngfile, width=500, height=500)
par(mfrow = c(1,1))
p3
dev.off()

### Remove any unused files:
rm(pngfile)
rm(dens100, dens10, dens1)
rm(vals_measure_100, vals_measure_10, vals_measure_1)
### -End of Script ###--------------------------------------------------------------------------------
```

Script 5: Splitting out by month

```
# 1) Setup ----------------------------------------
# Load packages
library(sp)
library(raster)
library(rgdal)
library(viridis)
options(scipen = 999)
### Data load ------------------------------------------------------------------------------------------------------------------------------------
### Reload from rasters if not in environment
### Richness measure (SRPUE)
measure100_months_1<- raster("tiffs/Month/measure100_months__1.tif")
measure100_months_2<- raster("tiffs/Month/measure100_months__2.tif")
measure100_months_3<- raster("tiffs/Month/measure100_months__3.tif")
measure100_months_4<- raster("tiffs/Month/measure100_months__4.tif")
measure100_months_5<- raster("tiffs/Month/measure100_months__5.tif")
measure100_months_6<- raster("tiffs/Month/measure100_months__6.tif")
measure100_months_7<- raster("tiffs/Month/measure100_months__7.tif")
measure100_months_8<- raster("tiffs/Month/measure100_months__8.tif")
measure100_months_9<- raster("tiffs/Month/measure100_months__9.tif")
measure100_months_10<- raster("tiffs/Month/measure100_months__10.tif")
measure100_months_11<- raster("tiffs/Month/measure100_months__11.tif")
measure100_months_12<- raster("tiffs/Month/measure100_months__12.tif")

stack_richmeasure_months_100 <- stack(measure100_months_1, measure100_months_2, measure100_months_3,
measure100_months_4, measure100_months_5, measure100_months_6,measure100_months_7, measure100_months_8,
measure100_months_9, measure100_months_10, measure100_months_11, measure100_months_12)
rm(measure100_months_1, measure100_months_2, measure100_months_3, measure100_months_4, measure100_months_5,
measure100_months_6,measure100_months_7, measure100_months_8, measure100_months_9, measure100_months_10,
measure100_months_11, measure100_months_12)

measure10_months_1<- raster("tiffs/Month/measure10_months__1.tif")
measure10_months_2<- raster("tiffs/Month/measure10_months__2.tif")
measure10_months_3<- raster("tiffs/Month/measure10_months__3.tif")
measure10_months_4<- raster("tiffs/Month/measure10_months__4.tif")
measure10_months_5<- raster("tiffs/Month/measure10_months__5.tif")
measure10_months_6<- raster("tiffs/Month/measure10_months__6.tif")
measure10_months_7<- raster("tiffs/Month/measure10_months__7.tif")
measure10_months_8<- raster("tiffs/Month/measure10_months__8.tif")
measure10_months_9<- raster("tiffs/Month/measure10_months__9.tif")
measure10_months_10<- raster("tiffs/Month/measure10_months__10.tif")
measure10_months_11<- raster("tiffs/Month/measure10_months__11.tif")
measure10_months_12<- raster("tiffs/Month/measure10_months__12.tif")

stack_richmeasure_months_10 <- stack(measure10_months_1, measure10_months_2, measure10_months_3,
measure10_months_4, measure10_months_5, measure10_months_6,measure10_months_7, measure10_months_8,
measure10_months_9, measure10_months_10, measure10_months_11, measure10_months_12)
rm(measure10_months_1, measure10_months_2, measure10_months_3, measure10_months_4, measure10_months_5,
measure10_months_6,measure10_months_7, measure10_months_8, measure10_months_9, measure10_months_10,
measure10_months_11, measure10_months_12)

measure1_months_1<- raster("tiffs/Month/measure1_months__1.tif")
measure1_months_2<- raster("tiffs/Month/measure1_months__2.tif")
measure1_months_3<- raster("tiffs/Month/measure1_months__3.tif")
measure1_months_4<- raster("tiffs/Month/measure1_months__4.tif")
measure1_months_5<- raster("tiffs/Month/measure1_months__5.tif")
measure1_months_6<- raster("tiffs/Month/measure1_months__6.tif")
measure1_months_7<- raster("tiffs/Month/measure1_months__7.tif")
measure1_months_8<- raster("tiffs/Month/measure1_months__8.tif")
measure1_months_9<- raster("tiffs/Month/measure1_months__9.tif")
measure1_months_10<- raster("tiffs/Month/measure1_months__10.tif")
measure1_months_11<- raster("tiffs/Month/measure1_months__11.tif")
measure1_months_12<- raster("tiffs/Month/measure1_months__12.tif")

stack_richmeasure_months_1 <- stack(measure1_months_1, measure1_months_2, measure1_months_3, measure1_months_4,
measure1_months_5, measure1_months_6,measure1_months_7, measure1_months_8, measure1_months_9,
measure1_months_10, measure1_months_11, measure1_months_12)
rm(measure1_months_1, measure1_months_2, measure1_months_3, measure1_months_4, measure1_months_5,
measure1_months_6,measure1_months_7, measure1_months_8, measure1_months_9, measure1_months_10,
measure1_months_11, measure1_months_12)
```

```
### 2) Months Data analysis ### --------------------------------------------------------------------------------------------------------------
#Create subsets of points per month ----------------------------------------
M01 = subset(points, points$MONTH == 1)
M02 = subset(points, points$MONTH == 2)
M03 = subset(points, points$MONTH == 3)
M04 = subset(points, points$MONTH == 4)
M05 = subset(points, points$MONTH == 5)
M06 = subset(points, points$MONTH == 6)
M07 = subset(points, points$MONTH == 7)
M08 = subset(points, points$MONTH == 8)
M09 = subset(points, points$MONTH == 9)
M10 = subset(points, points$MONTH == 10)
M11 = subset(points, points$MONTH == 11)
M12 = subset(points, points$MONTH == 12)
# Species richness all months - 100km resolution ----------------------------------------
rich100_M01 <- rasterize(M01, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M02 <- rasterize(M02, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M03 <- rasterize(M03, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M04 <- rasterize(M04, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M05 <- rasterize(M05, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M06 <- rasterize(M06, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M07 <- rasterize(M07, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M08 <- rasterize(M08, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M09 <- rasterize(M09, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M10 <- rasterize(M10, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M11 <- rasterize(M11, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_M12 <- rasterize(M12, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
#Stack, remove originals, save out raster
stack_rich100_months <- stack(rich100_M01, rich100_M02, rich100_M03, rich100_M04, rich100_M05, rich100_M06,
rich100_M07,rich100_M08,rich100_M09,rich100_M10,rich100_M11,rich100_M12)
rm(rich100_M01, rich100_M02, rich100_M03, rich100_M04, rich100_M05, rich100_M06, rich100_M07, rich100_M08,
rich100_M09, rich100_M10, rich100_M11, rich100_M12)
writeRaster(stack_rich100_months, "tiffs/Month/rich100_months_", format = 'GTiff', bylayer = TRUE, suffix='numbers',
overwrite = TRUE)
# Species richness all months - 10km resolution ----------------------------------------
rich10_M01 <- rasterize(M01, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M02 <- rasterize(M02, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M03 <- rasterize(M03, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M04 <- rasterize(M04, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M05 <- rasterize(M05, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M06 <- rasterize(M06, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M07 <- rasterize(M07, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M08 <- rasterize(M08, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M09 <- rasterize(M09, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M10 <- rasterize(M10, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M11 <- rasterize(M11, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_M12 <- rasterize(M12, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
#stack
stack_rich10_months <- stack(rich10_M01, rich10_M02, rich10_M03, rich10_M04, rich10_M05, rich10_M06,
rich10_M07,rich10_M08,rich10_M09,rich10_M10,rich10_M11,rich10_M12)
rm(rich10_M01, rich10_M02, rich10_M03, rich10_M04, rich10_M05, rich10_M06, rich10_M07, rich10_M08, rich10_M09,
rich10_M10, rich10_M11, rich10_M12)
writeRaster(stack_rich10_months, "tiffs/Month/rich10_months_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite
= TRUE)
# Species richness all months - 1km resolution ----------------------------------------
rich1_M01 <- rasterize(M01, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M02 <- rasterize(M02, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M03 <- rasterize(M03, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M04 <- rasterize(M04, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M05 <- rasterize(M05, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M06 <- rasterize(M06, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M07 <- rasterize(M07, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M08 <- rasterize(M08, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M09 <- rasterize(M09, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M10 <- rasterize(M10, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M11 <- rasterize(M11, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_M12 <- rasterize(M12, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
# stack
stack_rich1_months <- stack(rich1_M01, rich1_M02, rich1_M03, rich1_M04, rich1_M05, rich1_M06,
rich1_M07,rich1_M08,rich1_M09,rich1_M10,rich1_M11,rich1_M12)
rm(rich1_M01, rich1_M02, rich1_M03, rich1_M04, rich1_M05, rich1_M06, rich1_M07, rich1_M08, rich1_M09, rich1_M10,
rich1_M11, rich1_M12)
```

```
writeRaster(stack_rich1_months, "tiffs/Month/rich1_months_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite =
TRUE)
# Count of observations
# Raster of the number of observations - per grid cell
# Observations all months - 100km resolution ----------------------------------------
obs100_M01 <- rasterize(M01, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M02 <- rasterize(M02, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M03 <- rasterize(M03, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M04 <- rasterize(M04, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M05 <- rasterize(M05, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M06 <- rasterize(M06, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M07 <- rasterize(M07, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M08 <- rasterize(M08, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M09 <- rasterize(M09, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M10 <- rasterize(M10, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M11 <- rasterize(M11, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_M12 <- rasterize(M12, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
# Stack
stack_obs100_months <- stack(obs100_M01, obs100_M02, obs100_M03, obs100_M04, obs100_M05, obs100_M06,
obs100_M07,obs100_M08,obs100_M09,obs100_M10,obs100_M11,obs100_M12)
rm(obs100_M01, obs100_M02, obs100_M03, obs100_M04, obs100_M05, obs100_M06, obs100_M07, obs100_M08,
obs100_M09, obs100_M10, obs100_M11, obs100_M12)
writeRaster(stack_obs100_months, "tiffs/Month/obs100_months_", format = 'GTiff', bylayer = TRUE, suffix='numbers',
overwrite = TRUE)
# OBS all months - 10km resolution ----------------------------------------
obs10_M01 <- rasterize(M01, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M02 <- rasterize(M02, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M03 <- rasterize(M03, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M04 <- rasterize(M04, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M05 <- rasterize(M05, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M06 <- rasterize(M06, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M07 <- rasterize(M07, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M08 <- rasterize(M08, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M09 <- rasterize(M09, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M10 <- rasterize(M10, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M11 <- rasterize(M11, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_M12 <- rasterize(M12, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
#Stack
stack_obs10_months <- stack(obs10_M01, obs10_M02, obs10_M03, obs10_M04, obs10_M05, obs10_M06, obs10_M07,
obs10_M08 ,obs10_M09,obs10_M10,obs10_M11,obs10_M12)
rm(obs10_M01, obs10_M02, obs10_M03, obs10_M04, obs10_M05, obs10_M06, obs10_M07, obs10_M08, obs10_M09,
obs10_M10, obs10_M11, obs10_M12)
writeRaster(stack_obs10_months, "tiffs/Month/obs10_months_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite
= TRUE)
# OBS all months - 1km resolution ----------------------------------------
obs1_M01 <- rasterize(M01, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M02 <- rasterize(M02, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M03 <- rasterize(M03, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M04 <- rasterize(M04, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M05 <- rasterize(M05, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M06 <- rasterize(M06, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M07 <- rasterize(M07, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M08 <- rasterize(M08, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M09 <- rasterize(M09, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M10 <- rasterize(M10, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M11 <- rasterize(M11, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_M12 <- rasterize(M12, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
# Stack
stack_obs1_months <- stack(obs1_M01, obs1_M02, obs1_M03, obs1_M04, obs1_M05, obs1_M06,
obs1_M07,obs1_M08,obs1_M09,obs1_M10,obs1_M11,obs1_M12)
rm(obs1_M01, obs1_M02, obs1_M03, obs1_M04, obs1_M05, obs1_M06, obs1_M07, obs1_M08, obs1_M09, obs1_M10,
obs1_M11, obs1_M12)
writeRaster(stack_obs1_months, "tiffs/Month/obs1_months_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite =
TRUE)
### Effort calculations ### ------------------------------------------------------------------------------------
# Ready made function from #3_ScalingbyEffort
effortfactor
stack_richmeasure_months_100 <- overlay(stack_rich100_months, stack_obs100_months, fun=effortfactor, maxpixels =
pixmax)
plot(stack_richmeasure_months_100)
writeRaster(stack_richmeasure_months_100, "tiffs/Month/measure100_months_", format = 'GTiff', bylayer = TRUE,
suffix='numbers', overwrite = TRUE)
```

```
# If needed: to remove inf numbers.
stack_richmeasure_months_100[!is.finite(stack_richmeasure_months_100)] <- NA
stack_richmeasure_months_10 <- overlay(stack_rich10_months, stack_obs10_months, fun=effortfactor, maxpixels = pixmax)
writeRaster(stack_richmeasure_months_10, "tiffs/Month/measure10_months_", format = 'GTiff', bylayer = TRUE,
suffix='numbers', overwrite = TRUE)
# If needed: to remove inf numbers.
# stack_richmeasure_months_10[!is.finite(stack_richmeasure_months_10)] <- NA
stack_richmeasure_months_1 <- overlay(stack_rich1_months, stack_obs1_months, fun=effortfactor, maxpixels = pixmax)
writeRaster(stack_richmeasure_months_1, "tiffs/Month/measure1_months_", format = 'GTiff', bylayer = TRUE,
suffix='numbers')
stack_richmeasure_months_1[!is.finite(stack_richmeasure_months_1)] <- NA
### -End of Script ###--------------------------------------------------------------------------------
```

Script 6: Splitting out by years

```
# 1) Setup -----------------------------------------------------------------------------------------------------------
# Load packages
library(sp)
library(raster)
library(rgdal)
library(viridis)
options(scipen = 999)
### Note: Load from previous rasters
### 2) Years ### -----------------------------------------------------------------------------------------------------------
# Create subsets of points per year
Y1979 = subset(points, points$YEAR == 1979)
Y1980 = subset(points, points$YEAR == 1980)
Y1981 = subset(points, points$YEAR == 1981)
Y1982 = subset(points, points$YEAR == 1982)
Y1983 = subset(points, points$YEAR == 1983)
Y1984 = subset(points, points$YEAR == 1984)
Y1985 = subset(points, points$YEAR == 1985)
Y1986 = subset(points, points$YEAR == 1986)
Y1987 = subset(points, points$YEAR == 1987)
Y1988 = subset(points, points$YEAR == 1988)
Y1989 = subset(points, points$YEAR == 1989)
Y1990 = subset(points, points$YEAR == 1990)
Y1991 = subset(points, points$YEAR == 1991)
Y1992 = subset(points, points$YEAR == 1992)
Y1993 = subset(points, points$YEAR == 1993)
Y1994 = subset(points, points$YEAR == 1994)
Y1995 = subset(points, points$YEAR == 1995)
Y1996 = subset(points, points$YEAR == 1996)
Y1997 = subset(points, points$YEAR == 1997)
Y1998 = subset(points, points$YEAR == 1998)
Y1999 = subset(points, points$YEAR == 1999)
Y2000 = subset(points, points$YEAR == 2000)
Y2001 = subset(points, points$YEAR == 2001)
Y2002 = subset(points, points$YEAR == 2002)
Y2003 = subset(points, points$YEAR == 2003)
Y2004 = subset(points, points$YEAR == 2004)
Y2005 = subset(points, points$YEAR == 2005)
Y2006 = subset(points, points$YEAR == 2006)
Y2007 = subset(points, points$YEAR == 2007)
Y2008 = subset(points, points$YEAR == 2008)
Y2009 = subset(points, points$YEAR == 2009)
Y2010 = subset(points, points$YEAR == 2010)
Y2011 = subset(points, points$YEAR == 2011)
Y2012 = subset(points, points$YEAR == 2012)
Y2013 = subset(points, points$YEAR == 2013)
Y2014 = subset(points, points$YEAR == 2014)
Y2015 = subset(points, points$YEAR == 2015)
Y2016 = subset(points, points$YEAR == 2016)
Y2017 = subset(points, points$YEAR == 2017)
Y2018 = subset(points, points$YEAR == 2018)
# Species richness for all years - 100km resolution  ----------------------------------------
rich100_Y1979 <- rasterize(Y1979, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1980 <- rasterize(Y1980, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1981 <- rasterize(Y1981, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1982 <- rasterize(Y1982, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1983 <- rasterize(Y1983, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1984 <- rasterize(Y1984, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1985 <- rasterize(Y1985, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1986 <- rasterize(Y1986, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1987 <- rasterize(Y1987, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1988 <- rasterize(Y1988, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1989 <- rasterize(Y1989, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1990 <- rasterize(Y1990, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1991 <- rasterize(Y1991, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1992 <- rasterize(Y1992, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1993 <- rasterize(Y1993, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1994 <- rasterize(Y1994, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1995 <- rasterize(Y1995, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1996 <- rasterize(Y1996, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1997 <- rasterize(Y1997, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y1998 <- rasterize(Y1998, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
```

```
rich100_Y1999 <- rasterize(Y1999, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2000 <- rasterize(Y2000, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2001 <- rasterize(Y2001, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2002 <- rasterize(Y2002, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2003 <- rasterize(Y2003, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2004 <- rasterize(Y2004, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2005 <- rasterize(Y2005, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2006 <- rasterize(Y2006, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2007 <- rasterize(Y2007, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2008 <- rasterize(Y2008, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2009 <- rasterize(Y2009, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2010 <- rasterize(Y2010, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2011 <- rasterize(Y2011, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2012 <- rasterize(Y2012, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2013 <- rasterize(Y2013, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2014 <- rasterize(Y2014, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2015 <- rasterize(Y2015, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2016 <- rasterize(Y2016, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2017 <- rasterize(Y2017, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich100_Y2018 <- rasterize(Y2018, r100, 'NAME', function(x, ...) length(unique(na.omit(x))))
# Stack up years to one rasterstack for ease
stack_rich100_years <-
  stack(rich100_Y1979,
    rich100_Y1980,
    rich100_Y1981,
    rich100_Y1982,
    rich100_Y1983,
    rich100_Y1984,
    rich100_Y1985,
    rich100_Y1986,
    rich100_Y1987,
    rich100_Y1988,
    rich100_Y1989,
    rich100_Y1990,
    rich100_Y1991,
    rich100_Y1992,
    rich100_Y1993,
    rich100_Y1994,
    rich100_Y1995,
    rich100_Y1996,
    rich100_Y1997,
    rich100_Y1998,
    rich100_Y1999,
    rich100_Y2000,
    rich100_Y2001,
    rich100_Y2002,
    rich100_Y2003,
    rich100_Y2004,
    rich100_Y2005,
    rich100_Y2006,
    rich100_Y2007,
    rich100_Y2008,
    rich100_Y2009,
    rich100_Y2010,
    rich100_Y2011,
    rich100_Y2012,
    rich100_Y2013,
    rich100_Y2014,
    rich100_Y2015,
    rich100_Y2016,
    rich100_Y2017,
    rich100_Y2018)
# write out rasters
writeRaster(stack_rich100_years, "tiffs/Year/rich100_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite =
TRUE)
#removeformemory
rm(rich100_Y1979,rich100_Y1980,                                                    rich100_Y1981,
rich100_Y1982,rich100_Y1983,rich100_Y1984,rich100_Y1985,rich100_Y1986,
rich100_Y1987,rich100_Y1988,rich100_Y1989,rich100_Y1990,rich100_Y1991,rich100_Y1992,          rich100_Y1993,
rich100_Y1994,
rich100_Y1995,rich100_Y1996,rich100_Y1997,rich100_Y1998,rich100_Y1999,rich100_Y2000,rich100_Y2001,rich100_Y20
02,rich100_Y2003,rich100_Y2004,rich100_Y2005,rich100_Y2006,rich100_Y2007,rich100_Y2008,rich100_Y2009,rich100_Y
2010,rich100_Y2011,rich100_Y2012,
rich100_Y2013,rich100_Y2014,rich100_Y2015,rich100_Y2016,rich100_Y2017,rich100_Y2018)
```

```
# SPR for all years - 10km resolution  ----------------------------------------
rich10_Y1979 <- rasterize(Y1979, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1980 <- rasterize(Y1980, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1981 <- rasterize(Y1981, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1982 <- rasterize(Y1982, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1983 <- rasterize(Y1983, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1984 <- rasterize(Y1984, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1985 <- rasterize(Y1985, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1986 <- rasterize(Y1986, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1987 <- rasterize(Y1987, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1988 <- rasterize(Y1988, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1989 <- rasterize(Y1989, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1990 <- rasterize(Y1990, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1991 <- rasterize(Y1991, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1992 <- rasterize(Y1992, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1993 <- rasterize(Y1993, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1994 <- rasterize(Y1994, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1995 <- rasterize(Y1995, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1996 <- rasterize(Y1996, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1997 <- rasterize(Y1997, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1998 <- rasterize(Y1998, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y1999 <- rasterize(Y1999, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2000 <- rasterize(Y2000, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2001 <- rasterize(Y2001, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2002 <- rasterize(Y2002, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2003 <- rasterize(Y2003, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2004 <- rasterize(Y2004, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2005 <- rasterize(Y2005, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2006 <- rasterize(Y2006, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2007 <- rasterize(Y2007, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2008 <- rasterize(Y2008, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2009 <- rasterize(Y2009, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2010 <- rasterize(Y2010, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2011 <- rasterize(Y2011, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2012 <- rasterize(Y2012, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2013 <- rasterize(Y2013, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2014 <- rasterize(Y2014, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2015 <- rasterize(Y2015, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2016 <- rasterize(Y2016, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2017 <- rasterize(Y2017, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich10_Y2018 <- rasterize(Y2018, r10, 'NAME', function(x, ...) length(unique(na.omit(x))))
# Stack up years to one rasterstack for ease
stack_rich10_years <-
stack(rich10_Y1979,rich10_Y1980,rich10_Y1981,rich10_Y1982,rich10_Y1983,rich10_Y1984,rich10_Y1985,rich10_Y1986,ri
ch10_Y1987, rich10_Y1988,rich10_Y1989,
rich10_Y1990,rich10_Y1991,rich10_Y1992,rich10_Y1993,rich10_Y1994,rich10_Y1995,rich10_Y1996,rich10_Y1997,rich10_
Y1998,rich10_Y1999,
rich10_Y2000,rich10_Y2001,rich10_Y2002,rich10_Y2003,rich10_Y2004,rich10_Y2005,rich10_Y2006,rich10_Y2007,rich10_
Y2008,rich10_Y2009,
rich10_Y2010,rich10_Y2011,rich10_Y2012,rich10_Y2013,rich10_Y2014,rich10_Y2015,rich10_Y2016,rich10_Y2017,rich10_
Y2018)
# write out rasters
writeRaster(stack_rich10_years, "tiffs/Year/rich10_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite =
TRUE)
#removeformemory
rm(rich10_Y1979,rich10_Y1980, rich10_Y1981, rich10_Y1982,rich10_Y1983,rich10_Y1984,rich10_Y1985,rich10_Y1986,
rich10_Y1987,rich10_Y1988,rich10_Y1989,rich10_Y1990,rich10_Y1991,rich10_Y1992, rich10_Y1993, rich10_Y1994,
rich10_Y1995,rich10_Y1996,rich10_Y1997,rich10_Y1998,rich10_Y1999,rich10_Y2000,rich10_Y2001,rich10_Y2002,rich10_
Y2003,rich10_Y2004,rich10_Y2005,rich10_Y2006,rich10_Y2007,rich10_Y2008,rich10_Y2009,rich10_Y2010,rich10_Y2011,
rich10_Y2012, rich10_Y2013,rich10_Y2014,rich10_Y2015,rich10_Y2016,rich10_Y2017,rich10_Y2018)
# SPR for all years - 1km resolution  ----------------------------------------
rich1_Y1979 <- rasterize(Y1979, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1980 <- rasterize(Y1980, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1981 <- rasterize(Y1981, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1982 <- rasterize(Y1982, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1983 <- rasterize(Y1983, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1984 <- rasterize(Y1984, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1985 <- rasterize(Y1985, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1986 <- rasterize(Y1986, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1987 <- rasterize(Y1987, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1988 <- rasterize(Y1988, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
```

```
rich1_Y1989 <- rasterize(Y1989, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1990 <- rasterize(Y1990, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1991 <- rasterize(Y1991, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1992 <- rasterize(Y1992, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1993 <- rasterize(Y1993, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1994 <- rasterize(Y1994, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1995 <- rasterize(Y1995, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1996 <- rasterize(Y1996, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1997 <- rasterize(Y1997, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1998 <- rasterize(Y1998, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y1999 <- rasterize(Y1999, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2000 <- rasterize(Y2000, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2001 <- rasterize(Y2001, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2002 <- rasterize(Y2002, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2003 <- rasterize(Y2003, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2004 <- rasterize(Y2004, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2005 <- rasterize(Y2005, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2006 <- rasterize(Y2006, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2007 <- rasterize(Y2007, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2008 <- rasterize(Y2008, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2009 <- rasterize(Y2009, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2010 <- rasterize(Y2010, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2011 <- rasterize(Y2011, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2012 <- rasterize(Y2012, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2013 <- rasterize(Y2013, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2014 <- rasterize(Y2014, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2015 <- rasterize(Y2015, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2016 <- rasterize(Y2016, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2017 <- rasterize(Y2017, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
rich1_Y2018 <- rasterize(Y2018, r1, 'NAME', function(x, ...) length(unique(na.omit(x))))
# Stack up years to one rasterstack for ease
stack_rich1_years <-
stack(rich1_Y1979,rich1_Y1980,rich1_Y1981,rich1_Y1982,rich1_Y1983,rich1_Y1984,rich1_Y1985,rich1_Y1986,rich1_Y19
87, rich1_Y1988,rich1_Y1989,
rich1_Y1990,rich1_Y1991,rich1_Y1992,rich1_Y1993,rich1_Y1994,rich1_Y1995,rich1_Y1996,rich1_Y1997,rich1_Y1998,rich
1_Y1999,
rich1_Y2000,rich1_Y2001,rich1_Y2002,rich1_Y2003,rich1_Y2004,rich1_Y2005,rich1_Y2006,rich1_Y2007,rich1_Y2008,rich
1_Y2009,
rich1_Y2010,rich1_Y2011,rich1_Y2012,rich1_Y2013,rich1_Y2014,rich1_Y2015,rich1_Y2016,rich1_Y2017,rich1_Y2018)
# write out rasters
writeRaster(stack_rich1_years, "tiffs/Year/rich1_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite =
TRUE)
#removeformemory
rm(rich1_Y1979,rich1_Y1980, rich1_Y1981, rich1_Y1982,rich1_Y1983,rich1_Y1984,rich1_Y1985,rich1_Y1986,
rich1_Y1987,rich1_Y1988,rich1_Y1989,rich1_Y1990,rich1_Y1991,rich1_Y1992, rich1_Y1993, rich1_Y1994,
rich1_Y1995,rich1_Y1996,rich1_Y1997,rich1_Y1998,rich1_Y1999,rich1_Y2000,rich1_Y2001,rich1_Y2002,rich1_Y2003,rich
1_Y2004,rich1_Y2005,rich1_Y2006,rich1_Y2007,rich1_Y2008,rich1_Y2009,rich1_Y2010,rich1_Y2011,rich1_Y2012,
rich1_Y2013,rich1_Y2014,rich1_Y2015,rich1_Y2016,rich1_Y2017,rich1_Y2018)
writeRaster(richnessmeasure_100_Y1979, "tiffs/Year/Measure/richnessmeasure100_Y1979", format = 'GTiff', bylayer = TRUE,
suffix='numbers', overwrite = TRUE)
writeRaster(obs100_Y1979, "tiffs/Year/Obs/obs100_Y1979", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite =
TRUE)
# Observations for all years - 100km resolution ----------------------------------------
obs100_Y1979 <- rasterize(Y1979, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1980 <- rasterize(Y1980, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1981 <- rasterize(Y1981, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1982 <- rasterize(Y1982, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1983 <- rasterize(Y1983, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1984 <- rasterize(Y1984, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1985 <- rasterize(Y1985, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1986 <- rasterize(Y1986, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1987 <- rasterize(Y1987, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1988 <- rasterize(Y1988, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1989 <- rasterize(Y1989, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1990 <- rasterize(Y1990, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1991 <- rasterize(Y1991, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1992 <- rasterize(Y1992, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1993 <- rasterize(Y1993, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1994 <- rasterize(Y1994, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1995 <- rasterize(Y1995, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1996 <- rasterize(Y1996, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1997 <- rasterize(Y1997, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y1998 <- rasterize(Y1998, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
```

```
obs100_Y1999 <- rasterize(Y1999, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2000 <- rasterize(Y2000, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2001 <- rasterize(Y2001, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2002 <- rasterize(Y2002, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2003 <- rasterize(Y2003, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2004 <- rasterize(Y2004, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2005 <- rasterize(Y2005, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2006 <- rasterize(Y2006, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2007 <- rasterize(Y2007, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2008 <- rasterize(Y2008, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2009 <- rasterize(Y2009, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2010 <- rasterize(Y2010, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2011 <- rasterize(Y2011, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2012 <- rasterize(Y2012, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2013 <- rasterize(Y2013, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2014 <- rasterize(Y2014, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2015 <- rasterize(Y2015, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2016 <- rasterize(Y2016, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2017 <- rasterize(Y2017, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs100_Y2018 <- rasterize(Y2018, r100, field='NAME', fun=function(x, ...)length((na.omit(x))) )
#stack
stack_obs100_years <- stack(obs100_Y1979,
obs100_Y1980,obs100_Y1981,obs100_Y1982,obs100_Y1983,obs100_Y1984,obs100_Y1985,obs100_Y1986,obs100_Y1987,
obs100_Y1988,obs100_Y1989,obs100_Y1990,obs100_Y1991,obs100_Y1992,obs100_Y1993,obs100_Y1994,obs100_Y1995,
obs100_Y1996,obs100_Y1997,obs100_Y1998,obs100_Y1999,obs100_Y2000,obs100_Y2001,obs100_Y2002,obs100_Y2003,
obs100_Y2004,obs100_Y2005,obs100_Y2006,obs100_Y2007,obs100_Y2008,obs100_Y2009,obs100_Y2010,obs100_Y2011,
obs100_Y2012,obs100_Y2013,obs100_Y2014,obs100_Y2015,obs100_Y2016,obs100_Y2017,obs100_Y2018)
# write out rasters
writeRaster(stack_obs100_years, "tiffs/Year/obs100_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite =
TRUE)
#removeformemory
rm(obs100_Y1979,
obs100_Y1980,obs100_Y1981,obs100_Y1982,obs100_Y1983,obs100_Y1984,obs100_Y1985,obs100_Y1986,obs100_Y1987,
obs100_Y1988,obs100_Y1989,obs100_Y1990,obs100_Y1991,obs100_Y1992,obs100_Y1993,obs100_Y1994,obs100_Y1995,
obs100_Y1996,obs100_Y1997,obs100_Y1998,obs100_Y1999,obs100_Y2000,obs100_Y2001,obs100_Y2002,obs100_Y2003,
obs100_Y2004,obs100_Y2005,obs100_Y2006,obs100_Y2007,obs100_Y2008,obs100_Y2009,obs100_Y2010,obs100_Y2011,
obs100_Y2012,obs100_Y2013,obs100_Y2014,obs100_Y2015,obs100_Y2016,obs100_Y2017,obs100_Y2018)
# OBS for all years - 10km resolution  ----------------------------------------
obs10_Y1979 <- rasterize(Y1979, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1980 <- rasterize(Y1980, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1981 <- rasterize(Y1981, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1982 <- rasterize(Y1982, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1983 <- rasterize(Y1983, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1984 <- rasterize(Y1984, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1985 <- rasterize(Y1985, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1986 <- rasterize(Y1986, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1987 <- rasterize(Y1987, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1988 <- rasterize(Y1988, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1989 <- rasterize(Y1989, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1990 <- rasterize(Y1990, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1991 <- rasterize(Y1991, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1992 <- rasterize(Y1992, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1993 <- rasterize(Y1993, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1994 <- rasterize(Y1994, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1995 <- rasterize(Y1995, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1996 <- rasterize(Y1996, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1997 <- rasterize(Y1997, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1998 <- rasterize(Y1998, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y1999 <- rasterize(Y1999, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2000 <- rasterize(Y2000, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2001 <- rasterize(Y2001, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2002 <- rasterize(Y2002, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2003 <- rasterize(Y2003, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2004 <- rasterize(Y2004, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2005 <- rasterize(Y2005, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2006 <- rasterize(Y2006, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2007 <- rasterize(Y2007, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2008 <- rasterize(Y2008, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2009 <- rasterize(Y2009, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2010 <- rasterize(Y2010, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2011 <- rasterize(Y2011, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2012 <- rasterize(Y2012, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2013 <- rasterize(Y2013, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
```

```
obs10_Y2014 <- rasterize(Y2014, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2015 <- rasterize(Y2015, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2016 <- rasterize(Y2016, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2017 <- rasterize(Y2017, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs10_Y2018 <- rasterize(Y2018, r10, field='NAME', fun=function(x, ...)length((na.omit(x))) )
#stack
stack_obs10_years <- stack(obs10_Y1979,
obs10_Y1980,obs10_Y1981,obs10_Y1982,obs10_Y1983,obs10_Y1984,obs10_Y1985,obs10_Y1986,obs10_Y1987,obs10_Y1
988,obs10_Y1989,obs10_Y1990,obs10_Y1991,obs10_Y1992,obs10_Y1993,obs10_Y1994,obs10_Y1995,obs10_Y1996,obs10
_Y1997,obs10_Y1998,obs10_Y1999,obs10_Y2000,obs10_Y2001,obs10_Y2002,obs10_Y2003,obs10_Y2004,obs10_Y2005,ob
s10_Y2006,obs10_Y2007,obs10_Y2008,obs10_Y2009,obs10_Y2010,obs10_Y2011,obs10_Y2012,obs10_Y2013,obs10_Y201
4,obs10_Y2015,obs10_Y2016,obs10_Y2017,obs10_Y2018)
# write out rasters
writeRaster(stack_obs10_years, "tiffs/Year/obs10_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite =
TRUE)
#removeformemory
rm(obs10_Y1979,
obs10_Y1980,obs10_Y1981,obs10_Y1982,obs10_Y1983,obs10_Y1984,obs10_Y1985,obs10_Y1986,obs10_Y1987,obs10_Y1
988,obs10_Y1989,obs10_Y1990,obs10_Y1991,obs10_Y1992,obs10_Y1993,obs10_Y1994,obs10_Y1995,obs10_Y1996,obs10
_Y1997,obs10_Y1998,obs10_Y1999,obs10_Y2000,obs10_Y2001,obs10_Y2002,obs10_Y2003,obs10_Y2004,obs10_Y2005,ob
s10_Y2006,obs10_Y2007,obs10_Y2008,obs10_Y2009,obs10_Y2010,obs10_Y2011,obs10_Y2012,obs10_Y2013,obs10_Y201
4,obs10_Y2015,obs10_Y2016,obs10_Y2017,obs10_Y2018)
# Observations for all years - 1km resolution  ----------------------------------------
obs1_Y1979 <- rasterize(Y1979, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1980 <- rasterize(Y1980, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1981 <- rasterize(Y1981, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1982 <- rasterize(Y1982, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1983 <- rasterize(Y1983, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1984 <- rasterize(Y1984, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1985 <- rasterize(Y1985, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1986 <- rasterize(Y1986, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1987 <- rasterize(Y1987, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1988 <- rasterize(Y1988, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1989 <- rasterize(Y1989, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1990 <- rasterize(Y1990, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1991 <- rasterize(Y1991, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1992 <- rasterize(Y1992, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1993 <- rasterize(Y1993, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1994 <- rasterize(Y1994, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1995 <- rasterize(Y1995, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1996 <- rasterize(Y1996, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1997 <- rasterize(Y1997, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1998 <- rasterize(Y1998, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y1999 <- rasterize(Y1999, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2000 <- rasterize(Y2000, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2001 <- rasterize(Y2001, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2002 <- rasterize(Y2002, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2003 <- rasterize(Y2003, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2004 <- rasterize(Y2004, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2005 <- rasterize(Y2005, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2006 <- rasterize(Y2006, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2007 <- rasterize(Y2007, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2008 <- rasterize(Y2008, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2009 <- rasterize(Y2009, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2010 <- rasterize(Y2010, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2011 <- rasterize(Y2011, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2012 <- rasterize(Y2012, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2013 <- rasterize(Y2013, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2014 <- rasterize(Y2014, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2015 <- rasterize(Y2015, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2016 <- rasterize(Y2016, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2017 <- rasterize(Y2017, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
obs1_Y2018 <- rasterize(Y2018, r1, field='NAME', fun=function(x, ...)length((na.omit(x))) )
#stack
stack_obs1_years <- stack(obs1_Y1979,
obs1_Y1980,obs1_Y1981,obs1_Y1982,obs1_Y1983,obs1_Y1984,obs1_Y1985,obs1_Y1986,obs1_Y1987,obs1_Y1988,obs1_Y
1989,obs1_Y1990,obs1_Y1991,obs1_Y1992,obs1_Y1993,obs1_Y1994,obs1_Y1995,obs1_Y1996,obs1_Y1997,obs1_Y1998,o
bs1_Y1999,obs1_Y2000,obs1_Y2001,obs1_Y2002,obs1_Y2003,obs1_Y2004,obs1_Y2005,obs1_Y2006,obs1_Y2007,obs1_Y2
008,obs1_Y2009,obs1_Y2010,obs1_Y2011,obs1_Y2012,obs1_Y2013,obs1_Y2014,obs1_Y2015,obs1_Y2016,obs1_Y2017,ob
s1_Y2018)
# write out rasters
writeRaster(stack_obs1_years, "tiffs/Year/obs1_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite = TRUE)
#removeformemory
```

```
rm(obs1_Y1979,
obs1_Y1980,obs1_Y1981,obs1_Y1982,obs1_Y1983,obs1_Y1984,obs1_Y1985,obs1_Y1986,obs1_Y1987,obs1_Y1988,obs1_Y1989,obs1_Y1990,obs1_Y1991,obs1_Y1992,obs1_Y1993,obs1_Y1994,obs1_Y1995,obs1_Y1996,obs1_Y1997,obs1_Y1998,obs1_Y1999,obs1_Y2000,obs1_Y2001,obs1_Y2002,obs1_Y2003,obs1_Y2004,obs1_Y2005,obs1_Y2006,obs1_Y2007,obs1_Y2008,obs1_Y2009,obs1_Y2010,obs1_Y2011,obs1_Y2012,obs1_Y2013,obs1_Y2014,obs1_Y2015,obs1_Y2016,obs1_Y2017,obs1_Y2018)
#### Effort calculations ### ------------------------------------------------------------------------------------
# Ready made function from Script 3_ScalingbyEffort
effortfactor
pixmax <- 15480288
# Load in from previous script
stack_rich100_years
stack_obs100_years
stack_richmeasure_years_100 <- overlay(stack_rich100_years, stack_obs100_years, fun=effortfactor, maxpixels = pixmax)
# plot(stack_richmeasure_years_100)
writeRaster(stack_richmeasure_years_100, "tiffs/Year/measure100_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite = TRUE)
# Rasters Saved (y)
stack_richmeasure_years_10 <- overlay(stack_rich10_years, stack_obs10_years, fun=effortfactor, maxpixels = pixmax)
plot(stack_richmeasure_years_10)
writeRaster(stack_richmeasure_years_10, "tiffs/Year/measure10_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite = TRUE)
# Rasters Saved (y)
# 1 km2
stack_richmeasure_years_1 <- overlay(stack_rich1_years, stack_obs1_years, fun=effortfactor, maxpixels = pixmax)
# plot(stack_richmeasure_years_1)
writeRaster(stack_richmeasure_years_1, "tiffs/Year/measure100_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite = TRUE)
# Rasters Saved (y)
### Remove Years to remove data excess
rm(Y1979, Y1980, Y1981, Y1982, Y1983, Y1984, Y1985, Y1986, Y1987, Y1988, Y1989, Y1990, Y1991, Y1992, Y1993, Y1994, Y1995, Y1996, Y1997, Y1998, Y1999, Y2000, Y2001, Y2002, Y2003, Y2004, Y2005, Y2006, Y2007, Y2008, Y2009, Y2010, Y2011, Y2012, Y2013, Y2014, Y2015, Y2016, Y2017,Y2018)
rm(stack_rich100_years, stack_rich10_years, stack_rich1_years)
rm(stack_obs100_years, stack_obs10_years, stack_obs1_years)
### Long code if required needed.
# Yearly Measure Rasters:
# Years 100 100km2
richnessmeasure_100_Y1980 <- overlay(rich100_Y1980, obs100_Y1980, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1981 <- overlay(rich100_Y1981, obs100_Y1981, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1982 <- overlay(rich100_Y1982, obs100_Y1982, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1983 <- overlay(rich100_Y1983, obs100_Y1983, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1984 <- overlay(rich100_Y1984, obs100_Y1984, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1985 <- overlay(rich100_Y1985, obs100_Y1985, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1986 <- overlay(rich100_Y1986, obs100_Y1986, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1987 <- overlay(rich100_Y1987, obs100_Y1987, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1988 <- overlay(rich100_Y1988, obs100_Y1988, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1989 <- overlay(rich100_Y1979, obs100_Y1979, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1990 <- overlay(rich100_Y1990, obs100_Y1990, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1991 <- overlay(rich100_Y1991, obs100_Y1991, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1992 <- overlay(rich100_Y1992, obs100_Y1992, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1993 <- overlay(rich100_Y1993, obs100_Y1993, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1994 <- overlay(rich100_Y1994, obs100_Y1994, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1995 <- overlay(rich100_Y1995, obs100_Y1995, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1996 <- overlay(rich100_Y1996, obs100_Y1996, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1997 <- overlay(rich100_Y1997, obs100_Y1997, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1998 <- overlay(rich100_Y1998, obs100_Y1998, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y1999 <- overlay(rich100_Y1979, obs100_Y1979, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2000 <- overlay(rich100_Y2000, obs100_Y2000, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2001 <- overlay(rich100_Y2001, obs100_Y2001, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2002 <- overlay(rich100_Y2002, obs100_Y2002, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2003 <- overlay(rich100_Y2003, obs100_Y2003, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2004 <- overlay(rich100_Y2004, obs100_Y2004, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2005 <- overlay(rich100_Y2005, obs100_Y2005, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2006 <- overlay(rich100_Y2006, obs100_Y2006, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2007 <- overlay(rich100_Y2007, obs100_Y2007, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2008 <- overlay(rich100_Y2008, obs100_Y2008, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2009 <- overlay(rich100_Y2009, obs100_Y2009, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2010 <- overlay(rich100_Y2010, obs100_Y2010, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2011 <- overlay(rich100_Y2011, obs100_Y2011, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2012 <- overlay(rich100_Y2012, obs100_Y2012, fun=effortfactor, maxpixels = pixmax)
```

```
richnessmeasure_100_Y2013 <- overlay(rich100_Y2013, obs100_Y2013, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2014 <- overlay(rich100_Y2014, obs100_Y2014, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2015 <- overlay(rich100_Y2015, obs100_Y2015, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2016 <- overlay(rich100_Y2016, obs100_Y2016, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2017 <- overlay(rich100_Y2017, obs100_Y2017, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_100_Y2018 <- overlay(rich100_Y2018, obs100_Y2018, fun=effortfactor, maxpixels = pixmax)
# Years 10km2
richnessmeasure_10_Y1979 <- overlay(rich100_Y1979, obs100_Y1979, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1980 <- overlay(rich100_Y1980, obs100_Y1980, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1981 <- overlay(rich100_Y1981, obs100_Y1981, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1982 <- overlay(rich100_Y1982, obs100_Y1982, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1983 <- overlay(rich100_Y1983, obs100_Y1983, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1984 <- overlay(rich100_Y1984, obs100_Y1984, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1985 <- overlay(rich100_Y1985, obs100_Y1985, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1986 <- overlay(rich100_Y1986, obs100_Y1986, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1987 <- overlay(rich100_Y1987, obs100_Y1987, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1988 <- overlay(rich100_Y1988, obs100_Y1988, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1989 <- overlay(rich100_Y1979, obs100_Y1979, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1990 <- overlay(rich100_Y1990, obs100_Y1990, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1991 <- overlay(rich100_Y1991, obs100_Y1991, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1992 <- overlay(rich100_Y1992, obs100_Y1992, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1993 <- overlay(rich100_Y1993, obs100_Y1993, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1994 <- overlay(rich100_Y1994, obs100_Y1994, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1995 <- overlay(rich100_Y1995, obs100_Y1995, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1996 <- overlay(rich100_Y1996, obs100_Y1996, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1997 <- overlay(rich100_Y1997, obs100_Y1997, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1998 <- overlay(rich100_Y1998, obs100_Y1998, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y1999 <- overlay(rich100_Y1979, obs100_Y1979, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2000 <- overlay(rich100_Y2000, obs100_Y2000, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2001 <- overlay(rich100_Y2001, obs100_Y2001, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2002 <- overlay(rich100_Y2002, obs100_Y2002, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2003 <- overlay(rich100_Y2003, obs100_Y2003, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2004 <- overlay(rich100_Y2004, obs100_Y2004, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2005 <- overlay(rich100_Y2005, obs100_Y2005, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2006 <- overlay(rich100_Y2006, obs100_Y2006, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2007 <- overlay(rich100_Y2007, obs100_Y2007, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2008 <- overlay(rich100_Y2008, obs100_Y2008, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2009 <- overlay(rich100_Y2009, obs100_Y2009, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2010 <- overlay(rich100_Y2010, obs100_Y2010, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2011 <- overlay(rich100_Y2011, obs100_Y2011, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2012 <- overlay(rich100_Y2012, obs100_Y2012, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2013 <- overlay(rich100_Y2013, obs100_Y2013, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2014 <- overlay(rich100_Y2014, obs100_Y2014, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2015 <- overlay(rich100_Y2015, obs100_Y2015, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2016 <- overlay(rich100_Y2016, obs100_Y2016, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2017 <- overlay(rich100_Y2017, obs100_Y2017, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_10_Y2018 <- overlay(rich100_Y2018, obs100_Y2018, fun=effortfactor, maxpixels = pixmax)
# Years 1km2
richnessmeasure_1_Y1979 <- overlay(rich100_Y1979, obs100_Y1979, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1980 <- overlay(rich100_Y1980, obs100_Y1980, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1981 <- overlay(rich100_Y1981, obs100_Y1981, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1982 <- overlay(rich100_Y1982, obs100_Y1982, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1983 <- overlay(rich100_Y1983, obs100_Y1983, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1984 <- overlay(rich100_Y1984, obs100_Y1984, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1985 <- overlay(rich100_Y1985, obs100_Y1985, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1986 <- overlay(rich100_Y1986, obs100_Y1986, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1987 <- overlay(rich100_Y1987, obs100_Y1987, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1988 <- overlay(rich100_Y1988, obs100_Y1988, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1989 <- overlay(rich100_Y1979, obs100_Y1979, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1990 <- overlay(rich100_Y1990, obs100_Y1990, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1991 <- overlay(rich100_Y1991, obs100_Y1991, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1992 <- overlay(rich100_Y1992, obs100_Y1992, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1993 <- overlay(rich100_Y1993, obs100_Y1993, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1994 <- overlay(rich100_Y1994, obs100_Y1994, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1995 <- overlay(rich100_Y1995, obs100_Y1995, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1996 <- overlay(rich100_Y1996, obs100_Y1996, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1997 <- overlay(rich100_Y1997, obs100_Y1997, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1998 <- overlay(rich100_Y1998, obs100_Y1998, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y1999 <- overlay(rich100_Y1979, obs100_Y1979, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2000 <- overlay(rich100_Y2000, obs100_Y2000, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2001 <- overlay(rich100_Y2001, obs100_Y2001, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2002 <- overlay(rich100_Y2002, obs100_Y2002, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2003 <- overlay(rich100_Y2003, obs100_Y2003, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2004 <- overlay(rich100_Y2004, obs100_Y2004, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2005 <- overlay(rich100_Y2005, obs100_Y2005, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2006 <- overlay(rich100_Y2006, obs100_Y2006, fun=effortfactor, maxpixels = pixmax)
```

```
richnessmeasure_1_Y2007 <- overlay(rich100_Y2007, obs100_Y2007, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2008 <- overlay(rich100_Y2008, obs100_Y2008, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2009 <- overlay(rich100_Y2009, obs100_Y2009, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2010 <- overlay(rich100_Y2010, obs100_Y2010, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2011 <- overlay(rich100_Y2011, obs100_Y2011, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2012 <- overlay(rich100_Y2012, obs100_Y2012, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2013 <- overlay(rich100_Y2013, obs100_Y2013, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2014 <- overlay(rich100_Y2014, obs100_Y2014, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2015 <- overlay(rich100_Y2015, obs100_Y2015, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2016 <- overlay(rich100_Y2016, obs100_Y2016, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2017 <- overlay(rich100_Y2017, obs100_Y2017, fun=effortfactor, maxpixels = pixmax)
richnessmeasure_1_Y2018 <- overlay(rich100_Y2018, obs100_Y2018, fun=effortfactor, maxpixels = pixmax)
# Stack up these measures for ease in future scripts
stack_years_richmeasure_100 <- stack(
  richnessmeasure_100_Y1979,
  richnessmeasure_100_Y1980,
  richnessmeasure_100_Y1981,
  richnessmeasure_100_Y1982,
  richnessmeasure_100_Y1983,
  richnessmeasure_100_Y1984,
  richnessmeasure_100_Y1985,
  richnessmeasure_100_Y1986,
  richnessmeasure_100_Y1987,
  richnessmeasure_100_Y1988,
  richnessmeasure_100_Y1989,
  richnessmeasure_100_Y1990,
  richnessmeasure_100_Y1991,
  richnessmeasure_100_Y1992,
  richnessmeasure_100_Y1993,
  richnessmeasure_100_Y1994,
  richnessmeasure_100_Y1995,
  richnessmeasure_100_Y1996,
  richnessmeasure_100_Y1997,
  richnessmeasure_100_Y1998,
  richnessmeasure_100_Y1999,
  richnessmeasure_100_Y2000,
  richnessmeasure_100_Y2001,
  richnessmeasure_100_Y2002,
  richnessmeasure_100_Y2003,
  richnessmeasure_100_Y2004,
  richnessmeasure_100_Y2005,
  richnessmeasure_100_Y2006,
  richnessmeasure_100_Y2007,
  richnessmeasure_100_Y2008,
  richnessmeasure_100_Y2009,
  richnessmeasure_100_Y2010,
  richnessmeasure_100_Y2011,
  richnessmeasure_100_Y2012,
  richnessmeasure_100_Y2013,
  richnessmeasure_100_Y2014,
  richnessmeasure_100_Y2015,
  richnessmeasure_100_Y2016,
  richnessmeasure_100_Y2017,
  richnessmeasure_100_Y2018)

stack_years_richmeasure_10 <- stack(
  richnessmeasure_10_Y1979,
  richnessmeasure_10_Y1980,
  richnessmeasure_10_Y1981,
  richnessmeasure_10_Y1982,
  richnessmeasure_10_Y1983,
  richnessmeasure_10_Y1984,
  richnessmeasure_10_Y1985,
  richnessmeasure_10_Y1986,
  richnessmeasure_10_Y1987,
  richnessmeasure_10_Y1988,
  richnessmeasure_10_Y1989,
  richnessmeasure_10_Y1990,
  richnessmeasure_10_Y1991,
  richnessmeasure_10_Y1992,
  richnessmeasure_10_Y1993,
  richnessmeasure_10_Y1994,
  richnessmeasure_10_Y1995,
  richnessmeasure_10_Y1996,
  richnessmeasure_10_Y1997,
  richnessmeasure_10_Y1998,
```

```
  richnessmeasure_10_Y1999,
  richnessmeasure_10_Y2000,
  richnessmeasure_10_Y2001,
  richnessmeasure_10_Y2002,
  richnessmeasure_10_Y2003,
  richnessmeasure_10_Y2004,
  richnessmeasure_10_Y2005,
  richnessmeasure_10_Y2006,
  richnessmeasure_10_Y2007,
  richnessmeasure_10_Y2008,
  richnessmeasure_10_Y2009,
  richnessmeasure_10_Y2010,
  richnessmeasure_10_Y2011,
  richnessmeasure_10_Y2012,
  richnessmeasure_10_Y2013,
  richnessmeasure_10_Y2014,
  richnessmeasure_10_Y2015,
  richnessmeasure_10_Y2016,
  richnessmeasure_10_Y2017,
  richnessmeasure_10_Y2018)

stack_years_richmeasure_1 <- stack(
  richnessmeasure_1_Y1979,
  richnessmeasure_1_Y1980,
  richnessmeasure_1_Y1981,
  richnessmeasure_1_Y1982,
  richnessmeasure_1_Y1983,
  richnessmeasure_1_Y1984,
  richnessmeasure_1_Y1985,
  richnessmeasure_1_Y1986,
  richnessmeasure_1_Y1987,
  richnessmeasure_1_Y1988,
  richnessmeasure_1_Y1989,
  richnessmeasure_1_Y1990,
  richnessmeasure_1_Y1991,
  richnessmeasure_1_Y1992,
  richnessmeasure_1_Y1993,
  richnessmeasure_1_Y1994,
  richnessmeasure_1_Y1995,
  richnessmeasure_1_Y1996,
  richnessmeasure_1_Y1997,
  richnessmeasure_1_Y1998,
  richnessmeasure_1_Y1999,
  richnessmeasure_1_Y2000,
  richnessmeasure_1_Y2001,
  richnessmeasure_1_Y2002,
  richnessmeasure_1_Y2003,
  richnessmeasure_1_Y2004,
  richnessmeasure_1_Y2005,
  richnessmeasure_1_Y2006,
  richnessmeasure_1_Y2007,
  richnessmeasure_1_Y2008,
  richnessmeasure_1_Y2009,
  richnessmeasure_1_Y2010,
  richnessmeasure_1_Y2011,
  richnessmeasure_1_Y2012,
  richnessmeasure_1_Y2013,
  richnessmeasure_1_Y2014,
  richnessmeasure_1_Y2015,
  richnessmeasure_1_Y2016,
  richnessmeasure_1_Y2017,
  richnessmeasure_1_Y2018)
# Remove infinte values (inf)
stack_years_richmeasure_100[!is.finite(stack_years_richmeasure_100)] <- NA
stack_years_richmeasure_10[!is.finite(stack_years_richmeasure_10)] <- NA
stack_years_richmeasure_1[!is.finite(stack_years_richmeasure_1)] <- NA

### -End of Script ###-------------------------------------------------------------------------------
```

Script 7: Splitting out by month and year

```
# 1) Setup ---------------------------------------------------------------------------------------------------------------
library(raster)
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
# 2) Data subset ---------------------------------------------------------------------------------------------------------------
Y1979_M01 = subset(points, points$MONTH == 1 & points$YEAR == 1979)
Y1979_M02 = subset(points, points$MONTH == 2 & points$YEAR == 1979)
Y1979_M03 = subset(points, points$MONTH == 3 & points$YEAR == 1979)
Y1979_M04 = subset(points, points$MONTH == 4 & points$YEAR == 1979)
Y1979_M05 = subset(points, points$MONTH == 5 & points$YEAR == 1979)
Y1979_M06 = subset(points, points$MONTH == 6 & points$YEAR == 1979)
Y1979_M07 = subset(points, points$MONTH == 7 & points$YEAR == 1979)
Y1979_M08 = subset(points, points$MONTH == 8 & points$YEAR == 1979)
Y1979_M09 = subset(points, points$MONTH == 9 & points$YEAR == 1979)
Y1979_M10 = subset(points, points$MONTH == 10 & points$YEAR == 1979)
Y1979_M11 = subset(points, points$MONTH == 11 & points$YEAR == 1979)
Y1979_M12 = subset(points, points$MONTH == 12 & points$YEAR == 1979)
### Repeat this code for all months of each year, not included here in appendix due to space ###

### 3) Richness / Obs for 100km subsets ### ------------------------------------------------------------------------------
# SPR 100km
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
# Code runs so that if 'blank' aka no rows, it will run out the basic r100 raster to fill in gaps. If 1+ rows, will do the species richness calcs and save out raster.
#1979
if(length(Y1979_M01) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_01.tif", overwrite = TRUE)} else {rasterize(Y1979_M01, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_01.tif", overwrite = TRUE)}
if(length(Y1979_M02) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_02.tif", overwrite = TRUE)} else {rasterize(Y1979_M02, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_02.tif", overwrite = TRUE)}
if(length(Y1979_M03) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_03.tif", overwrite = TRUE)} else {rasterize(Y1979_M03, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_03.tif", overwrite = TRUE)}
if(length(Y1979_M04) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_04.tif", overwrite = TRUE)} else {rasterize(Y1979_M04, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_04.tif", overwrite = TRUE)}
if(length(Y1979_M05) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_05.tif", overwrite = TRUE)} else {rasterize(Y1979_M05, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_05.tif", overwrite = TRUE)}
if(length(Y1979_M06) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_06.tif", overwrite = TRUE)} else {rasterize(Y1979_M06, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_06.tif", overwrite = TRUE)}
if(length(Y1979_M07) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_07.tif", overwrite = TRUE)} else {rasterize(Y1979_M07, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_07.tif", overwrite = TRUE)}
if(length(Y1979_M08) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_08.tif", overwrite = TRUE)} else {rasterize(Y1979_M08, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_08.tif", overwrite = TRUE)}
if(length(Y1979_M09) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_09.tif", overwrite = TRUE)} else {rasterize(Y1979_M09, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_09.tif", overwrite = TRUE)}
if(length(Y1979_M10) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_10.tif", overwrite = TRUE)} else {rasterize(Y1979_M10, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_10.tif", overwrite = TRUE)}
if(length(Y1979_M11) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_11.tif", overwrite = TRUE)} else {rasterize(Y1979_M11, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_11.tif", overwrite = TRUE)}
if(length(Y1979_M12) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_12.tif", overwrite = TRUE)} else {rasterize(Y1979_M12, r100, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Richness/rich100_Y1979_12.tif", overwrite = TRUE)}
### Repeat this code for all months of each year, not included here in appendix due to space ###
# Stack for use:
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Richness")
files <- list.files(pattern = ".tif")
stack_rich100_subyear <- stack(files)
stack_rich100_subyear

# 4) Observations 100km ### ---------------------------------------
```

```
# Code runs so that if 'blank' aka no rows, it will run out the basic r100 raster to fill in gaps. If 1+ rows, will do the species Obs calcs and save out raster.
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
#1979
if(length(Y1979_M01) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_01.tif", overwrite = TRUE)} else {rasterize(Y1979_M01, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_01.tif", overwrite = TRUE)}
if(length(Y1979_M02) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_02.tif", overwrite = TRUE)} else {rasterize(Y1979_M02, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_02.tif", overwrite = TRUE)}
if(length(Y1979_M03) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_03.tif", overwrite = TRUE)} else {rasterize(Y1979_M03, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_03.tif", overwrite = TRUE)}
if(length(Y1979_M04) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_04.tif", overwrite = TRUE)} else {rasterize(Y1979_M04, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_04.tif", overwrite = TRUE)}
if(length(Y1979_M05) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_05.tif", overwrite = TRUE)} else {rasterize(Y1979_M05, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_05.tif", overwrite = TRUE)}
if(length(Y1979_M06) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_06.tif", overwrite = TRUE)} else {rasterize(Y1979_M06, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_06.tif", overwrite = TRUE)}
if(length(Y1979_M07) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_07.tif", overwrite = TRUE)} else {rasterize(Y1979_M07, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_07.tif", overwrite = TRUE)}
if(length(Y1979_M08) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_08.tif", overwrite = TRUE)} else {rasterize(Y1979_M08, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_08.tif", overwrite = TRUE)}
if(length(Y1979_M09) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_09.tif", overwrite = TRUE)} else {rasterize(Y1979_M09, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_09.tif", overwrite = TRUE)}
if(length(Y1979_M10) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_10.tif", overwrite = TRUE)} else {rasterize(Y1979_M10, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_10.tif", overwrite = TRUE)}
if(length(Y1979_M11) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_11.tif", overwrite = TRUE)} else {rasterize(Y1979_M11, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_11.tif", overwrite = TRUE)}
if(length(Y1979_M12) == 0){writeRaster(r100, filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_12.tif", overwrite = TRUE)} else {rasterize(Y1979_M12, r100, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/100/Obs/obs100_Y1979_12.tif", overwrite = TRUE)}
### Repeat this code for all months of each year, not included here in appendix due to space ###
# Stack for ease in futue scripts:
setwd(“D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Obs”)
files <- list.files(pattern = “.tif”)
stack_obs100_subyear <- stack(files)
stack_obs100_subyear

### 5) Effort 100 km calculations ###--------------------------------------------------------------------------------------------------------
# setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2") PC
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
# Ready made function from #3_ScalingbyEffort
effortfactor
# Using:
stack_rich100_subyear
stack_obs100_subyear
stack_richmeasure_subyear_100 <- overlay(stack_rich100_subyear, stack_obs100_subyear, fun=effortfactor, maxpixels = pixmax)
# plot(stack_richmeasure_subyear_100)
writeRaster(stack_richmeasure_subyear_100, "tiffs/Month_Year_Sub/100/Measure/Measure100_years_", format = 'GTiff', bylayer = TRUE, suffix='numbers', overwrite = TRUE)
# Rasters Saved

### Richness / Obs for 10km subsets -###----------------------------------------
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
# SPR 10km   ----------------------------------------------------------------------------------------------------------------------
# Code runs so that if 'blank' aka no rows, it will run out the basic r10 raster to fill in gaps. If 1+ rows, will do the species richness calcs and save out raster.
#1979
if(length(Y1979_M01) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_01.tif", overwrite = TRUE)} else {rasterize(Y1979_M01, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_01.tif", overwrite = TRUE)}
```

```
if(length(Y1979_M02) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_02.tif", overwrite
= TRUE)} else {rasterize(Y1979_M02, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_02.tif", overwrite = TRUE)}
if(length(Y1979_M03) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_03.tif", overwrite
= TRUE)} else {rasterize(Y1979_M03, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_03.tif", overwrite = TRUE)}
if(length(Y1979_M04) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_04.tif", overwrite
= TRUE)} else {rasterize(Y1979_M04, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_04.tif", overwrite = TRUE)}
if(length(Y1979_M05) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_05.tif", overwrite
= TRUE)} else {rasterize(Y1979_M05, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_05.tif", overwrite = TRUE)}
if(length(Y1979_M06) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_06.tif", overwrite
= TRUE)} else {rasterize(Y1979_M06, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_06.tif", overwrite = TRUE)}
if(length(Y1979_M07) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_07.tif", overwrite
= TRUE)} else {rasterize(Y1979_M07, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_07.tif", overwrite = TRUE)}
if(length(Y1979_M08) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_08.tif", overwrite
= TRUE)} else {rasterize(Y1979_M08, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_08.tif", overwrite = TRUE)}
if(length(Y1979_M09) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_09.tif", overwrite
= TRUE)} else {rasterize(Y1979_M09, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_09.tif", overwrite = TRUE)}
if(length(Y1979_M10) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_10.tif", overwrite
= TRUE)} else {rasterize(Y1979_M10, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_10.tif", overwrite = TRUE)}
if(length(Y1979_M11) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_11.tif", overwrite
= TRUE)} else {rasterize(Y1979_M11, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_11.tif", overwrite = TRUE)}
if(length(Y1979_M12) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_12.tif", overwrite
= TRUE)} else {rasterize(Y1979_M12, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_12.tif", overwrite = TRUE)}
#1980
if(length(Y1979_M01) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_01.tif", overwrite
= TRUE)} else {rasterize(Y1979_M01, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_01.tif", overwrite = TRUE)}
if(length(Y1979_M02) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_02.tif", overwrite
= TRUE)} else {rasterize(Y1979_M02, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_02.tif", overwrite = TRUE)}
if(length(Y1979_M03) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_03.tif", overwrite
= TRUE)} else {rasterize(Y1979_M03, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_03.tif", overwrite = TRUE)}
if(length(Y1979_M04) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_04.tif", overwrite
= TRUE)} else {rasterize(Y1979_M04, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_04.tif", overwrite = TRUE)}
if(length(Y1979_M05) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_05.tif", overwrite
= TRUE)} else {rasterize(Y1979_M05, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_05.tif", overwrite = TRUE)}
if(length(Y1979_M06) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_06.tif", overwrite
= TRUE)} else {rasterize(Y1979_M06, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_06.tif", overwrite = TRUE)}
if(length(Y1979_M07) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_07.tif", overwrite
= TRUE)} else {rasterize(Y1979_M07, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_07.tif", overwrite = TRUE)}
if(length(Y1979_M08) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_08.tif", overwrite
= TRUE)} else {rasterize(Y1979_M08, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_08.tif", overwrite = TRUE)}
if(length(Y1979_M09) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_09.tif", overwrite
= TRUE)} else {rasterize(Y1979_M09, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_09.tif", overwrite = TRUE)}
if(length(Y1979_M10) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_10.tif", overwrite
= TRUE)} else {rasterize(Y1979_M10, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_10.tif", overwrite = TRUE)}
if(length(Y1979_M11) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_11.tif", overwrite
= TRUE)} else {rasterize(Y1979_M11, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_11.tif", overwrite = TRUE)}
if(length(Y1979_M12) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_12.tif", overwrite
= TRUE)} else {rasterize(Y1979_M12, r10, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Richness/rich10_Y1979_12.tif", overwrite = TRUE)}
### Repeat this code for all months of each year, not included here in appendix due to space ###
# Stack for use:
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Richness")
files <- list.files(pattern = ".tif")
stack_rich10_subyear <- stack(files)
```

```
stack_rich10_subyear

# OBS 10km   ----------------------------------------
# Code runs so that if 'blank' aka no rows, it will run out the basic r10 raster to fill in gaps. If 1+ rows, will do the species Obs calcs and save out raster.
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
#1979
if(length(Y1979_M01) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_01.tif", overwrite = TRUE)} else {rasterize(Y1979_M01, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_01.tif", overwrite = TRUE)}
if(length(Y1979_M02) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_02.tif", overwrite = TRUE)} else {rasterize(Y1979_M02, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_02.tif", overwrite = TRUE)}
if(length(Y1979_M03) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_03.tif", overwrite = TRUE)} else {rasterize(Y1979_M03, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_03.tif", overwrite = TRUE)}
if(length(Y1979_M04) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_04.tif", overwrite = TRUE)} else {rasterize(Y1979_M04, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_04.tif", overwrite = TRUE)}
if(length(Y1979_M05) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_05.tif", overwrite = TRUE)} else {rasterize(Y1979_M05, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_05.tif", overwrite = TRUE)}
if(length(Y1979_M06) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_06.tif", overwrite = TRUE)} else {rasterize(Y1979_M06, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_06.tif", overwrite = TRUE)}
if(length(Y1979_M07) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_07.tif", overwrite = TRUE)} else {rasterize(Y1979_M07, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_07.tif", overwrite = TRUE)}
if(length(Y1979_M08) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_08.tif", overwrite = TRUE)} else {rasterize(Y1979_M08, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_08.tif", overwrite = TRUE)}
if(length(Y1979_M09) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_09.tif", overwrite = TRUE)} else {rasterize(Y1979_M09, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_09.tif", overwrite = TRUE)}
if(length(Y1979_M10) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_10.tif", overwrite = TRUE)} else {rasterize(Y1979_M10, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_10.tif", overwrite = TRUE)}
if(length(Y1979_M11) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_11.tif", overwrite = TRUE)} else {rasterize(Y1979_M11, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_11.tif", overwrite = TRUE)}
if(length(Y1979_M12) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_12.tif", overwrite = TRUE)} else {rasterize(Y1979_M12, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_12.tif", overwrite = TRUE)}
#1980
if(length(Y1979_M01) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_01.tif", overwrite = TRUE)} else {rasterize(Y1979_M01, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_01.tif", overwrite = TRUE)}
if(length(Y1979_M02) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_02.tif", overwrite = TRUE)} else {rasterize(Y1979_M02, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_02.tif", overwrite = TRUE)}
if(length(Y1979_M03) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_03.tif", overwrite = TRUE)} else {rasterize(Y1979_M03, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_03.tif", overwrite = TRUE)}
if(length(Y1979_M04) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_04.tif", overwrite = TRUE)} else {rasterize(Y1979_M04, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_04.tif", overwrite = TRUE)}
if(length(Y1979_M05) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_05.tif", overwrite = TRUE)} else {rasterize(Y1979_M05, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_05.tif", overwrite = TRUE)}
if(length(Y1979_M06) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_06.tif", overwrite = TRUE)} else {rasterize(Y1979_M06, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_06.tif", overwrite = TRUE)}
if(length(Y1979_M07) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_07.tif", overwrite = TRUE)} else {rasterize(Y1979_M07, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_07.tif", overwrite = TRUE)}
if(length(Y1979_M08) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_08.tif", overwrite = TRUE)} else {rasterize(Y1979_M08, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_08.tif", overwrite = TRUE)}
if(length(Y1979_M09) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_09.tif", overwrite = TRUE)} else {rasterize(Y1979_M09, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_09.tif", overwrite = TRUE)}
if(length(Y1979_M10) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_10.tif", overwrite = TRUE)} else {rasterize(Y1979_M10, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_10.tif", overwrite = TRUE)}
```

```
if(length(Y1979_M11) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_11.tif", overwrite =
TRUE)} else {rasterize(Y1979_M11, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_11.tif", overwrite = TRUE)}
if(length(Y1979_M12) == 0){writeRaster(r10, filename = "tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_12.tif", overwrite =
TRUE)} else {rasterize(Y1979_M12, r10, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/10/Obs/obs10_Y1979_12.tif", overwrite = TRUE)}
### Repeat this code for all months of each year, not included here in appendix due to space ###
# Stack for use:
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Obs")
files <- list.files(pattern = ".tif")
stack_obs10_subyear <- stack(files)
stack_obs10_subyear
#### 10km Effort calculations-----------------------------
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
# eady made function from #3_ScalingbyEffort
effortfactor
# Using:
stack_rich10_subyear
stack_obs10_subyear
stack_richmeasure_subyear_10 <- overlay(stack_rich10_subyear, stack_obs10_subyear, fun=effortfactor, maxpixels = pixmax)
# plot(stack_richmeasure_subyear_10)
writeRaster(stack_richmeasure_subyear_10, "tiffs/Month_Year_Sub/10/Measure/Measure10_years_", format = 'GTiff', bylayer
= TRUE, suffix='numbers', overwrite = TRUE)
# Rasters Saved
### Richness / Obs for 1km subsets ###-----------------------------------------
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
# SPR 1km resolution ---------------------------------------
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
# Code runs so that if 'blank' aka no rows, it will run out the basic r1 raster to fill in gaps. If 1+ rows, will do the species richness
calcs and save out raster.
#1979
if(length(Y1979_M01) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_01.tif", overwrite =
TRUE)} else {rasterize(Y1979_M01, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_01.tif", overwrite = TRUE)}
if(length(Y1979_M02) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_02.tif", overwrite =
TRUE)} else {rasterize(Y1979_M02, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_02.tif", overwrite = TRUE)}
if(length(Y1979_M03) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_03.tif", overwrite =
TRUE)} else {rasterize(Y1979_M03, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_03.tif", overwrite = TRUE)}
if(length(Y1979_M04) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_04.tif", overwrite =
TRUE)} else {rasterize(Y1979_M04, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_04.tif", overwrite = TRUE)}
if(length(Y1979_M05) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_05.tif", overwrite =
TRUE)} else {rasterize(Y1979_M05, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_05.tif", overwrite = TRUE)}
if(length(Y1979_M06) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_06.tif", overwrite =
TRUE)} else {rasterize(Y1979_M06, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_06.tif", overwrite = TRUE)}
if(length(Y1979_M07) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_07.tif", overwrite =
TRUE)} else {rasterize(Y1979_M07, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_07.tif", overwrite = TRUE)}
if(length(Y1979_M08) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_08.tif", overwrite =
TRUE)} else {rasterize(Y1979_M08, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_08.tif", overwrite = TRUE)}
if(length(Y1979_M09) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_09.tif", overwrite =
TRUE)} else {rasterize(Y1979_M09, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_09.tif", overwrite = TRUE)}
if(length(Y1979_M10) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_10.tif", overwrite =
TRUE)} else {rasterize(Y1979_M10, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_10.tif", overwrite = TRUE)}
if(length(Y1979_M11) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_11.tif", overwrite =
TRUE)} else {rasterize(Y1979_M11, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_11.tif", overwrite = TRUE)}
if(length(Y1979_M12) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_12.tif", overwrite =
TRUE)} else {rasterize(Y1979_M12, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1979_12.tif", overwrite = TRUE)}
TRUE)} else {rasterize(Y1983_M02, r1, 'NAME', function(x, ...) length(unique(na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Richness/rich1_Y1983_02.tif", overwrite = TRUE)}
### Repeat this code for all months of each year, not included here in appendix due to space ###
# Stack for use:
```

```
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Richness")
files <- list.files(pattern = ".tif")
stack_rich1_subyear <- stack(files)
stack_rich1_subyear
# Observations 1km resolution ----------------------------------------
# Code runs so that if 'blank' aka no rows, it will run out the basic r1 raster to fill in gaps. If 1+ rows, will do the species Obs calcs
and save out raster.
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
#1979
if(length(Y1979_M01) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_01.tif", overwrite =
TRUE)} else {rasterize(Y1979_M01, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_01.tif", overwrite = TRUE)}
if(length(Y1979_M02) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_02.tif", overwrite =
TRUE)} else {rasterize(Y1979_M02, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_02.tif", overwrite = TRUE)}
if(length(Y1979_M03) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_03.tif", overwrite =
TRUE)} else {rasterize(Y1979_M03, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_03.tif", overwrite = TRUE)}
if(length(Y1979_M04) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_04.tif", overwrite =
TRUE)} else {rasterize(Y1979_M04, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_04.tif", overwrite = TRUE)}
if(length(Y1979_M05) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_05.tif", overwrite =
TRUE)} else {rasterize(Y1979_M05, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_05.tif", overwrite = TRUE)}
if(length(Y1979_M06) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_06.tif", overwrite =
TRUE)} else {rasterize(Y1979_M06, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_06.tif", overwrite = TRUE)}
if(length(Y1979_M07) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_07.tif", overwrite =
TRUE)} else {rasterize(Y1979_M07, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_07.tif", overwrite = TRUE)}
if(length(Y1979_M08) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_08.tif", overwrite =
TRUE)} else {rasterize(Y1979_M08, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_08.tif", overwrite = TRUE)}
if(length(Y1979_M09) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_09.tif", overwrite =
TRUE)} else {rasterize(Y1979_M09, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_09.tif", overwrite = TRUE)}
if(length(Y1979_M10) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_10.tif", overwrite =
TRUE)} else {rasterize(Y1979_M10, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_10.tif", overwrite = TRUE)}
if(length(Y1979_M11) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_11.tif", overwrite =
TRUE)} else {rasterize(Y1979_M11, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_11.tif", overwrite = TRUE)}
if(length(Y1979_M12) == 0){writeRaster(r1, filename = "tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_12.tif", overwrite =
TRUE)} else {rasterize(Y1979_M12, r1, 'NAME', fun=function(x, ...)length((na.omit(x))), filename =
"tiffs/Month_Year_Sub/1/Obs/obs1_Y1979_12.tif", overwrite = TRUE)}
### Repeat this code for all months of each year, not included here in appendix due to space ###
# Stack for use:
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Obs")
files <- list.files(pattern = ".tif")
stack_obs1_subyear <- stack(files)
stack_obs1_subyear
#### 1km Effort calculations-###-----------------------------
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")
# Using:
stack_rich1_subyear
stack_obs1_subyear
stack_richmeasure_subyear_1 <- overlay(stack_rich1_subyear, stack_obs1_subyear, fun=effortfactor, maxpixels = pixmax)
writeRaster(stack_richmeasure_subyear_1, "tiffs/Month_Year_Sub/1/Measure/Measure1_years_", format = 'GTiff', bylayer =
TRUE, suffix='numbers', overwrite = TRUE)
df_rich1_subyear<- as.data.frame(stack_rich1_subyear)
colnames(df_rich1_months) <- Months
col_rich1_months <- gather(df_rich1_months) #Tidyr to Reshaping wide format to long format
# View(col_rich1_months)
rm(df_rich1_months)
### -End of Script ###-----------------------------------------------------------------------------
```

Script 8: Loading data from rasters for analysis

```
# To save running long code (Scripts 5-7) with processing each time wanting to use data, load via this script instead to save processing and memory.
# Setup Packages
library(ggplot2)
library(ggridges)
library(lattice)
library(latticeExtra)
library(hrbrthemes)  #imports library hrbrthemes
library(dplyr)       #imports the dplyr function
library (tidyr)      #imports the tidyverse package
library(viridis)     #imports the library viridis
library(scales)
library(raster)
### 1) All data ###------------------------------------------------------------------------------------------------------------------------
# a) Richness Tiffs
rich_100 <- raster("tiffs/All/rich_100.tif")
rich_10 <- raster("tiffs/All/rich_10.tif")
rich_1 <- raster("tiffs/All/rich_1.tif")
# b) Obs Tiffs
obs_100 <- raster("tiffs/All/obs_100.tif")
obs_10 <- raster("tiffs/All/obs_10.tif")
obs_1 <- raster("tiffs/All/obs_1.tif")
# c) Measure Tiffs
measure100 <- raster("tiffs/All/rast_measure_100.tif")
measure10 <- raster("tiffs/All/rast_measure_10.tif")
measure1 <- raster("tiffs/All/rast_measure_1.tif")
### 2) Months ###------------------------------------------------------------------------------------------------------------------------
# a) Richness Tiffs
# 100km2
rich100_months_1<- raster("tiffs/Month/rich100_months__1.tif")
rich100_months_2<- raster("tiffs/Month/rich100_months__2.tif")
rich100_months_3<- raster("tiffs/Month/rich100_months__3.tif")
rich100_months_4<- raster("tiffs/Month/rich100_months__4.tif")
rich100_months_5<- raster("tiffs/Month/rich100_months__5.tif")
rich100_months_6<- raster("tiffs/Month/rich100_months__6.tif")
rich100_months_7<- raster("tiffs/Month/rich100_months__7.tif")
rich100_months_8<- raster("tiffs/Month/rich100_months__8.tif")
rich100_months_9<- raster("tiffs/Month/rich100_months__9.tif")
rich100_months_10<- raster("tiffs/Month/rich100_months__10.tif")
rich100_months_11<- raster("tiffs/Month/rich100_months__11.tif")
rich100_months_12<- raster("tiffs/Month/rich100_months__12.tif")

stack_rich100_months <- stack(rich100_months_1, rich100_months_2, rich100_months_3, rich100_months_4, rich100_months_5, rich100_months_6,rich100_months_7, rich100_months_8, rich100_months_9, rich100_months_10, rich100_months_11, rich100_months_12)
rm(rich100_months_1, rich100_months_2, rich100_months_3, rich100_months_4, rich100_months_5, rich100_months_6,rich100_months_7, rich100_months_8, rich100_months_9, rich100_months_10, rich100_months_11, rich100_months_12)

# 10km2
rich10_months_1<- raster("tiffs/Month/rich10_months__1.tif")
rich10_months_2<- raster("tiffs/Month/rich10_months__2.tif")
rich10_months_3<- raster("tiffs/Month/rich10_months__3.tif")
rich10_months_4<- raster("tiffs/Month/rich10_months__4.tif")
rich10_months_5<- raster("tiffs/Month/rich10_months__5.tif")
rich10_months_6<- raster("tiffs/Month/rich10_months__6.tif")
rich10_months_7<- raster("tiffs/Month/rich10_months__7.tif")
rich10_months_8<- raster("tiffs/Month/rich10_months__8.tif")
rich10_months_9<- raster("tiffs/Month/rich10_months__9.tif")
rich10_months_10<- raster("tiffs/Month/rich10_months__10.tif")
rich10_months_11<- raster("tiffs/Month/rich10_months__11.tif")
rich10_months_12<- raster("tiffs/Month/rich10_months__12.tif")

stack_rich10_months <- stack(rich10_months_1, rich10_months_2, rich10_months_3, rich10_months_4, rich10_months_5, rich10_months_6,rich10_months_7, rich10_months_8, rich10_months_9, rich10_months_10, rich10_months_11, rich10_months_12)
rm(rich10_months_1, rich10_months_2, rich10_months_3, rich10_months_4, rich10_months_5, rich10_months_6,rich10_months_7, rich10_months_8, rich10_months_9, rich10_months_10, rich10_months_11, rich10_months_12)
```

```
# 1 km2
rich1_months_1<- raster("tiffs/Month/rich1_months__1.tif")
rich1_months_2<- raster("tiffs/Month/rich1_months__2.tif")
rich1_months_3<- raster("tiffs/Month/rich1_months__3.tif")
rich1_months_4<- raster("tiffs/Month/rich1_months__4.tif")
rich1_months_5<- raster("tiffs/Month/rich1_months__5.tif")
rich1_months_6<- raster("tiffs/Month/rich1_months__6.tif")
rich1_months_7<- raster("tiffs/Month/rich1_months__7.tif")
rich1_months_8<- raster("tiffs/Month/rich1_months__8.tif")
rich1_months_9<- raster("tiffs/Month/rich1_months__9.tif")
rich1_months_10<- raster("tiffs/Month/rich1_months__10.tif")
rich1_months_11<- raster("tiffs/Month/rich1_months__11.tif")
rich1_months_12<- raster("tiffs/Month/rich1_months__12.tif")

stack_rich1_months <- stack(rich1_months_1, rich1_months_2, rich1_months_3, rich1_months_4, rich1_months_5,
rich1_months_6,rich1_months_7, rich1_months_8, rich1_months_9, rich1_months_10, rich1_months_11, rich1_months_12)
rm(rich1_months_1, rich1_months_2, rich1_months_3, rich1_months_4, rich1_months_5, rich1_months_6,rich1_months_7,
rich1_months_8, rich1_months_9, rich1_months_10, rich1_months_11, rich1_months_12)


### b) Obs Tiffs ###-----------------------------------------------------------------------------------------------------------------------------
# 100km2
obs100_months_1<- raster("tiffs/Month/obs100_months__1.tif")
obs100_months_2<- raster("tiffs/Month/obs100_months__2.tif")
obs100_months_3<- raster("tiffs/Month/obs100_months__3.tif")
obs100_months_4<- raster("tiffs/Month/obs100_months__4.tif")
obs100_months_5<- raster("tiffs/Month/obs100_months__5.tif")
obs100_months_6<- raster("tiffs/Month/obs100_months__6.tif")
obs100_months_7<- raster("tiffs/Month/obs100_months__7.tif")
obs100_months_8<- raster("tiffs/Month/obs100_months__8.tif")
obs100_months_9<- raster("tiffs/Month/obs100_months__9.tif")
obs100_months_10<- raster("tiffs/Month/obs100_months__10.tif")
obs100_months_11<- raster("tiffs/Month/obs100_months__11.tif")
obs100_months_12<- raster("tiffs/Month/obs100_months__12.tif")
stack_obs100_months <- stack(obs100_months_1, obs100_months_2, obs100_months_3, obs100_months_4, obs100_months_5,
obs100_months_6,obs100_months_7, obs100_months_8, obs100_months_9, obs100_months_10, obs100_months_11,
obs100_months_12)
rm(obs100_months_1, obs100_months_2, obs100_months_3, obs100_months_4, obs100_months_5,
obs100_months_6,obs100_months_7, obs100_months_8, obs100_months_9, obs100_months_10, obs100_months_11,
obs100_months_12)
# 10km2
obs10_months_1<- raster("tiffs/Month/obs10_months__1.tif")
obs10_months_2<- raster("tiffs/Month/obs10_months__2.tif")
obs10_months_3<- raster("tiffs/Month/obs10_months__3.tif")
obs10_months_4<- raster("tiffs/Month/obs10_months__4.tif")
obs10_months_5<- raster("tiffs/Month/obs10_months__5.tif")
obs10_months_6<- raster("tiffs/Month/obs10_months__6.tif")
obs10_months_7<- raster("tiffs/Month/obs10_months__7.tif")
obs10_months_8<- raster("tiffs/Month/obs10_months__8.tif")
obs10_months_9<- raster("tiffs/Month/obs10_months__9.tif")
obs10_months_10<- raster("tiffs/Month/obs10_months__10.tif")
obs10_months_11<- raster("tiffs/Month/obs10_months__11.tif")
obs10_months_12<- raster("tiffs/Month/obs10_months__12.tif")

stack_obs10_months <- stack(obs10_months_1, obs10_months_2, obs10_months_3, obs10_months_4, obs10_months_5,
obs10_months_6,obs10_months_7, obs10_months_8, obs10_months_9, obs10_months_10, obs10_months_11,
obs10_months_12)
rm(obs10_months_1, obs10_months_2, obs10_months_3, obs10_months_4, obs10_months_5,
obs10_months_6,obs10_months_7, obs10_months_8, obs10_months_9, obs10_months_10, obs10_months_11,
obs10_months_12)

# 10km2
obs1_months_1<- raster("tiffs/Month/obs1_months__1.tif")
obs1_months_2<- raster("tiffs/Month/obs1_months__2.tif")
obs1_months_3<- raster("tiffs/Month/obs1_months__3.tif")
obs1_months_4<- raster("tiffs/Month/obs1_months__4.tif")
obs1_months_5<- raster("tiffs/Month/obs1_months__5.tif")
obs1_months_6<- raster("tiffs/Month/obs1_months__6.tif")
obs1_months_7<- raster("tiffs/Month/obs1_months__7.tif")
obs1_months_8<- raster("tiffs/Month/obs1_months__8.tif")
obs1_months_9<- raster("tiffs/Month/obs1_months__9.tif")
obs1_months_10<- raster("tiffs/Month/obs1_months__10.tif")
obs1_months_11<- raster("tiffs/Month/obs1_months__11.tif")
obs1_months_12<- raster("tiffs/Month/obs1_months__12.tif")
```

```
stack_obs1_months <- stack(obs1_months_1, obs1_months_2, obs1_months_3, obs1_months_4, obs1_months_5,
obs1_months_6,obs1_months_7, obs1_months_8, obs1_months_9, obs1_months_10, obs1_months_11, obs1_months_12)
rm(obs1_months_1, obs1_months_2, obs1_months_3, obs1_months_4, obs1_months_5, obs1_months_6,obs1_months_7,
obs1_months_8, obs1_months_9, obs1_months_10, obs1_months_11, obs1_months_12)

### c) Effort Tiffs ###---------------------------------------------------------------------------------------------------
# 100km2
# Effort Tiffs
measure100_months_1<- raster("tiffs/Month/measure100_months__1.tif")
measure100_months_2<- raster("tiffs/Month/measure100_months__2.tif")
measure100_months_3<- raster("tiffs/Month/measure100_months__3.tif")
measure100_months_4<- raster("tiffs/Month/measure100_months__4.tif")
measure100_months_5<- raster("tiffs/Month/measure100_months__5.tif")
measure100_months_6<- raster("tiffs/Month/measure100_months__6.tif")
measure100_months_7<- raster("tiffs/Month/measure100_months__7.tif")
measure100_months_8<- raster("tiffs/Month/measure100_months__8.tif")
measure100_months_9<- raster("tiffs/Month/measure100_months__9.tif")
measure100_months_10<- raster("tiffs/Month/measure100_months__10.tif")
measure100_months_11<- raster("tiffs/Month/measure100_months__11.tif")
measure100_months_12<- raster("tiffs/Month/measure100_months__12.tif")

stack_richmeasure_months_100 <- stack(measure100_months_1, measure100_months_2, measure100_months_3,
measure100_months_4, measure100_months_5, measure100_months_6,measure100_months_7, measure100_months_8,
measure100_months_9, measure100_months_10, measure100_months_11, measure100_months_12)
rm(measure100_months_1, measure100_months_2, measure100_months_3, measure100_months_4, measure100_months_5,
measure100_months_6,measure100_months_7, measure100_months_8, measure100_months_9, measure100_months_10,
measure100_months_11, measure100_months_12)

# 10km2
measure10_months_1<- raster("tiffs/Month/measure10_months__1.tif")
measure10_months_2<- raster("tiffs/Month/measure10_months__2.tif")
measure10_months_3<- raster("tiffs/Month/measure10_months__3.tif")
measure10_months_4<- raster("tiffs/Month/measure10_months__4.tif")
measure10_months_5<- raster("tiffs/Month/measure10_months__5.tif")
measure10_months_6<- raster("tiffs/Month/measure10_months__6.tif")
measure10_months_7<- raster("tiffs/Month/measure10_months__7.tif")
measure10_months_8<- raster("tiffs/Month/measure10_months__8.tif")
measure10_months_9<- raster("tiffs/Month/measure10_months__9.tif")
measure10_months_10<- raster("tiffs/Month/measure10_months__10.tif")
measure10_months_11<- raster("tiffs/Month/measure10_months__11.tif")
measure10_months_12<- raster("tiffs/Month/measure10_months__12.tif")

stack_richmeasure_months_10 <- stack(measure10_months_1, measure10_months_2, measure10_months_3,
measure10_months_4, measure10_months_5, measure10_months_6,measure10_months_7, measure10_months_8,
measure10_months_9, measure10_months_10, measure10_months_11, measure10_months_12)
rm(measure10_months_1, measure10_months_2, measure10_months_3, measure10_months_4, measure10_months_5,
measure10_months_6,measure10_months_7, measure10_months_8, measure10_months_9, measure10_months_10,
measure10_months_11, measure10_months_12)

# 1km2
measure1_months_1<- raster("tiffs/Month/measure1_months__1.tif")
measure1_months_2<- raster("tiffs/Month/measure1_months__2.tif")
measure1_months_3<- raster("tiffs/Month/measure1_months__3.tif")
measure1_months_4<- raster("tiffs/Month/measure1_months__4.tif")
measure1_months_5<- raster("tiffs/Month/measure1_months__5.tif")
measure1_months_6<- raster("tiffs/Month/measure1_months__6.tif")
measure1_months_7<- raster("tiffs/Month/measure1_months__7.tif")
measure1_months_8<- raster("tiffs/Month/measure1_months__8.tif")
measure1_months_9<- raster("tiffs/Month/measure1_months__9.tif")
measure1_months_10<- raster("tiffs/Month/measure1_months__10.tif")
measure1_months_11<- raster("tiffs/Month/measure1_months__11.tif")
measure1_months_12<- raster("tiffs/Month/measure1_months__12.tif")

stack_richmeasure_months_1 <- stack(measure1_months_1, measure1_months_2, measure1_months_3, measure1_months_4,
measure1_months_5, measure1_months_6,measure1_months_7, measure1_months_8, measure1_months_9,
measure1_months_10, measure1_months_11, measure1_months_12)
rm(measure1_months_1, measure1_months_2, measure1_months_3, measure1_months_4, measure1_months_5,
measure1_months_6,measure1_months_7, measure1_months_8, measure1_months_9, measure1_months_10,
measure1_months_11, measure1_months_12)

### 3) Years ###---------------------------------------------------------------------------------------------------
# a) Richness Tiffs
rich100_years_1 <- raster("tiffs/Year/Richness/rich100_years__1.tif")
```

```
rich100_years_2 <- raster("tiffs/Year/rich100_years__2.tif")
rich100_years_3 <- raster("tiffs/Year/rich100_years__3.tif")
rich100_years_4 <- raster("tiffs/Year/rich100_years__4.tif")
rich100_years_5 <- raster("tiffs/Year/rich100_years__5.tif")
rich100_years_6 <- raster("tiffs/Year/rich100_years__6.tif")
rich100_years_7 <- raster("tiffs/Year/rich100_years__7.tif")
rich100_years_8 <- raster("tiffs/Year/rich100_years__8.tif")
rich100_years_9 <- raster("tiffs/Year/rich100_years__9.tif")
rich100_years_10 <- raster("tiffs/Year/rich100_years__10.tif")
rich100_years_11 <- raster("tiffs/Year/rich100_years__11.tif")
rich100_years_12 <- raster("tiffs/Year/rich100_years__12.tif")
rich100_years_13 <- raster("tiffs/Year/rich100_years__13.tif")
rich100_years_14 <- raster("tiffs/Year/rich100_years__14.tif")
rich100_years_15 <- raster("tiffs/Year/rich100_years__15.tif")
rich100_years_16 <- raster("tiffs/Year/rich100_years__16.tif")
rich100_years_17 <- raster("tiffs/Year/rich100_years__17.tif")
rich100_years_18 <- raster("tiffs/Year/rich100_years__18.tif")
rich100_years_19 <- raster("tiffs/Year/rich100_years__19.tif")
rich100_years_20 <- raster("tiffs/Year/rich100_years__20.tif")
rich100_years_21 <- raster("tiffs/Year/rich100_years__21.tif")
rich100_years_22 <- raster("tiffs/Year/rich100_years__22.tif")
rich100_years_23 <- raster("tiffs/Year/rich100_years__23.tif")
rich100_years_24 <- raster("tiffs/Year/rich100_years__24.tif")
rich100_years_25 <- raster("tiffs/Year/rich100_years__25.tif")
rich100_years_26 <- raster("tiffs/Year/rich100_years__26.tif")
rich100_years_27 <- raster("tiffs/Year/rich100_years__27.tif")
rich100_years_28 <- raster("tiffs/Year/rich100_years__28.tif")
rich100_years_29 <- raster("tiffs/Year/rich100_years__29.tif")
rich100_years_30 <- raster("tiffs/Year/rich100_years__30.tif")
rich100_years_31 <- raster("tiffs/Year/rich100_years__31.tif")
rich100_years_32 <- raster("tiffs/Year/rich100_years__32.tif")
rich100_years_33 <- raster("tiffs/Year/rich100_years__33.tif")
rich100_years_34 <- raster("tiffs/Year/rich100_years__34.tif")
rich100_years_35 <- raster("tiffs/Year/rich100_years__35.tif")
rich100_years_36 <- raster("tiffs/Year/rich100_years__36.tif")
rich100_years_37 <- raster("tiffs/Year/rich100_years__37.tif")
rich100_years_38 <- raster("tiffs/Year/rich100_years__38.tif")
rich100_years_39 <- raster("tiffs/Year/rich100_years__39.tif")
rich100_years_40 <- raster("tiffs/Year/rich100_years__40.tif")

stack_rich100_years <-
  stack(rich100_years_1,
        rich100_years_2,
        rich100_years_3,
        rich100_years_4,
        rich100_years_5,
        rich100_years_6,
        rich100_years_7,
        rich100_years_8,
        rich100_years_9,
        rich100_years_10,
        rich100_years_11,
        rich100_years_12,
        rich100_years_13,
        rich100_years_14,
        rich100_years_15,
        rich100_years_16,
        rich100_years_17,
        rich100_years_18,
        rich100_years_19,
        rich100_years_20,
        rich100_years_21,
        rich100_years_22,
        rich100_years_23,
        rich100_years_24,
        rich100_years_25,
        rich100_years_26,
        rich100_years_27,
        rich100_years_28,
        rich100_years_29,
        rich100_years_30,
        rich100_years_31,
        rich100_years_32,
        rich100_years_33,
        rich100_years_34,
        rich100_years_35,
```

```
        rich100_years_36,
        rich100_years_37,
        rich100_years_38,
        rich100_years_39,
        rich100_years_40
  )

# 10km2
rich10_years_1 <- raster("tiffs/Year/rich10_years__1.tif")
rich10_years_2 <- raster("tiffs/Year/rich10_years__2.tif")
rich10_years_3 <- raster("tiffs/Year/rich10_years__3.tif")
rich10_years_4 <- raster("tiffs/Year/rich10_years__4.tif")
rich10_years_5 <- raster("tiffs/Year/rich10_years__5.tif")
rich10_years_6 <- raster("tiffs/Year/rich10_years__6.tif")
rich10_years_7 <- raster("tiffs/Year/rich10_years__7.tif")
rich10_years_8 <- raster("tiffs/Year/rich10_years__8.tif")
rich10_years_9 <- raster("tiffs/Year/rich10_years__9.tif")
rich10_years_10 <- raster("tiffs/Year/rich10_years__10.tif")
rich10_years_11 <- raster("tiffs/Year/rich10_years__11.tif")
rich10_years_12 <- raster("tiffs/Year/rich10_years__12.tif")
rich10_years_13 <- raster("tiffs/Year/rich10_years__13.tif")
rich10_years_14 <- raster("tiffs/Year/rich10_years__14.tif")
rich10_years_15 <- raster("tiffs/Year/rich10_years__15.tif")
rich10_years_16 <- raster("tiffs/Year/rich10_years__16.tif")
rich10_years_17 <- raster("tiffs/Year/rich10_years__17.tif")
rich10_years_18 <- raster("tiffs/Year/rich10_years__18.tif")
rich10_years_19 <- raster("tiffs/Year/rich10_years__19.tif")
rich10_years_20 <- raster("tiffs/Year/rich10_years__20.tif")
rich10_years_21 <- raster("tiffs/Year/rich10_years__21.tif")
rich10_years_22 <- raster("tiffs/Year/rich10_years__22.tif")
rich10_years_23 <- raster("tiffs/Year/rich10_years__23.tif")
rich10_years_24 <- raster("tiffs/Year/rich10_years__24.tif")
rich10_years_25 <- raster("tiffs/Year/rich10_years__25.tif")
rich10_years_26 <- raster("tiffs/Year/rich10_years__26.tif")
rich10_years_27 <- raster("tiffs/Year/rich10_years__27.tif")
rich10_years_28 <- raster("tiffs/Year/rich10_years__28.tif")
rich10_years_29 <- raster("tiffs/Year/rich10_years__29.tif")
rich10_years_30 <- raster("tiffs/Year/rich10_years__30.tif")
rich10_years_31 <- raster("tiffs/Year/rich10_years__31.tif")
rich10_years_32 <- raster("tiffs/Year/rich10_years__32.tif")
rich10_years_33 <- raster("tiffs/Year/rich10_years__33.tif")
rich10_years_34 <- raster("tiffs/Year/rich10_years__34.tif")
rich10_years_35 <- raster("tiffs/Year/rich10_years__35.tif")
rich10_years_36 <- raster("tiffs/Year/rich10_years__36.tif")
rich10_years_37 <- raster("tiffs/Year/rich10_years__37.tif")
rich10_years_38 <- raster("tiffs/Year/rich10_years__38.tif")
rich10_years_39 <- raster("tiffs/Year/rich10_years__39.tif")
rich10_years_40 <- raster("tiffs/Year/rich10_years__40.tif")

stack_rich10_years <-
  stack(rich10_years_1,
        rich10_years_2,
        rich10_years_3,
        rich10_years_4,
        rich10_years_5,
        rich10_years_6,
        rich10_years_7,
        rich10_years_8,
        rich10_years_9,
        rich10_years_10,
        rich10_years_11,
        rich10_years_12,
        rich10_years_13,
        rich10_years_14,
        rich10_years_15,
        rich10_years_16,
        rich10_years_17,
        rich10_years_18,
        rich10_years_19,
        rich10_years_20,
        rich10_years_21,
        rich10_years_22,
        rich10_years_23,
        rich10_years_24,
        rich10_years_25,
```

```
        rich10_years_26,
        rich10_years_27,
        rich10_years_28,
        rich10_years_29,
        rich10_years_30,
        rich10_years_31,
        rich10_years_32,
        rich10_years_33,
        rich10_years_34,
        rich10_years_35,
        rich10_years_36,
        rich10_years_37,
        rich10_years_38,
        rich10_years_39,
        rich10_years_40
  )

# 1km2
rich1_years_1 <- raster("tiffs/Year/rich1_years__1.tif")
rich1_years_2 <- raster("tiffs/Year/rich1_years__2.tif")
rich1_years_3 <- raster("tiffs/Year/rich1_years__3.tif")
rich1_years_4 <- raster("tiffs/Year/rich1_years__4.tif")
rich1_years_5 <- raster("tiffs/Year/rich1_years__5.tif")
rich1_years_6 <- raster("tiffs/Year/rich1_years__6.tif")
rich1_years_7 <- raster("tiffs/Year/rich1_years__7.tif")
rich1_years_8 <- raster("tiffs/Year/rich1_years__8.tif")
rich1_years_9 <- raster("tiffs/Year/rich1_years__9.tif")
rich1_years_10 <- raster("tiffs/Year/rich1_years__10.tif")
rich1_years_11 <- raster("tiffs/Year/rich1_years__11.tif")
rich1_years_12 <- raster("tiffs/Year/rich1_years__12.tif")
rich1_years_13 <- raster("tiffs/Year/rich1_years__13.tif")
rich1_years_14 <- raster("tiffs/Year/rich1_years__14.tif")
rich1_years_15 <- raster("tiffs/Year/rich1_years__15.tif")
rich1_years_16 <- raster("tiffs/Year/rich1_years__16.tif")
rich1_years_17 <- raster("tiffs/Year/rich1_years__17.tif")
rich1_years_18 <- raster("tiffs/Year/rich1_years__18.tif")
rich1_years_19 <- raster("tiffs/Year/rich1_years__19.tif")
rich1_years_20 <- raster("tiffs/Year/rich1_years__20.tif")
rich1_years_21 <- raster("tiffs/Year/rich1_years__21.tif")
rich1_years_22 <- raster("tiffs/Year/rich1_years__22.tif")
rich1_years_23 <- raster("tiffs/Year/rich1_years__23.tif")
rich1_years_24 <- raster("tiffs/Year/rich1_years__24.tif")
rich1_years_25 <- raster("tiffs/Year/rich1_years__25.tif")
rich1_years_26 <- raster("tiffs/Year/rich1_years__26.tif")
rich1_years_27 <- raster("tiffs/Year/rich1_years__27.tif")
rich1_years_28 <- raster("tiffs/Year/rich1_years__28.tif")
rich1_years_29 <- raster("tiffs/Year/rich1_years__29.tif")
rich1_years_30 <- raster("tiffs/Year/rich1_years__30.tif")
rich1_years_31 <- raster("tiffs/Year/rich1_years__31.tif")
rich1_years_32 <- raster("tiffs/Year/rich1_years__32.tif")
rich1_years_33 <- raster("tiffs/Year/rich1_years__33.tif")
rich1_years_34 <- raster("tiffs/Year/rich1_years__34.tif")
rich1_years_35 <- raster("tiffs/Year/rich1_years__35.tif")
rich1_years_36 <- raster("tiffs/Year/rich1_years__36.tif")
rich1_years_37 <- raster("tiffs/Year/rich1_years__37.tif")
rich1_years_38 <- raster("tiffs/Year/rich1_years__38.tif")
rich1_years_39 <- raster("tiffs/Year/rich1_years__39.tif")
rich1_years_40 <- raster("tiffs/Year/rich1_years__40.tif")

stack_rich1_years <-
  stack(rich1_years_1,
        rich1_years_2,
        rich1_years_3,
        rich1_years_4,
        rich1_years_5,
        rich1_years_6,
        rich1_years_7,
        rich1_years_8,
        rich1_years_9,
        rich1_years_10,
        rich1_years_11,
        rich1_years_12,
        rich1_years_13,
        rich1_years_14,
        rich1_years_15,
```

```
      rich1_years_16,
      rich1_years_17,
      rich1_years_18,
      rich1_years_19,
      rich1_years_20,
      rich1_years_21,
      rich1_years_22,
      rich1_years_23,
      rich1_years_24,
      rich1_years_25,
      rich1_years_26,
      rich1_years_27,
      rich1_years_28,
      rich1_years_29,
      rich1_years_30,
      rich1_years_31,
      rich1_years_32,
      rich1_years_33,
      rich1_years_34,
      rich1_years_35,
      rich1_years_36,
      rich1_years_37,
      rich1_years_38,
      rich1_years_39,
      rich1_years_40
  )

### a) Obs Tiffs
# 100km2
obs100_years_1 <- raster("tiffs/Year/obs100_years__1.tif")
obs100_years_2 <- raster("tiffs/Year/obs100_years__2.tif")
obs100_years_3 <- raster("tiffs/Year/obs100_years__3.tif")
obs100_years_4 <- raster("tiffs/Year/obs100_years__4.tif")
obs100_years_5 <- raster("tiffs/Year/obs100_years__5.tif")
obs100_years_6 <- raster("tiffs/Year/obs100_years__6.tif")
obs100_years_7 <- raster("tiffs/Year/obs100_years__7.tif")
obs100_years_8 <- raster("tiffs/Year/obs100_years__8.tif")
obs100_years_9 <- raster("tiffs/Year/obs100_years__9.tif")
obs100_years_10 <- raster("tiffs/Year/obs100_years__10.tif")
obs100_years_11 <- raster("tiffs/Year/obs100_years__11.tif")
obs100_years_12 <- raster("tiffs/Year/obs100_years__12.tif")
obs100_years_13 <- raster("tiffs/Year/obs100_years__13.tif")
obs100_years_14 <- raster("tiffs/Year/obs100_years__14.tif")
obs100_years_15 <- raster("tiffs/Year/obs100_years__15.tif")
obs100_years_16 <- raster("tiffs/Year/obs100_years__16.tif")
obs100_years_17 <- raster("tiffs/Year/obs100_years__17.tif")
obs100_years_18 <- raster("tiffs/Year/obs100_years__18.tif")
obs100_years_19 <- raster("tiffs/Year/obs100_years__19.tif")
obs100_years_20 <- raster("tiffs/Year/obs100_years__20.tif")
obs100_years_21 <- raster("tiffs/Year/obs100_years__21.tif")
obs100_years_22 <- raster("tiffs/Year/obs100_years__22.tif")
obs100_years_23 <- raster("tiffs/Year/obs100_years__23.tif")
obs100_years_24 <- raster("tiffs/Year/obs100_years__24.tif")
obs100_years_25 <- raster("tiffs/Year/obs100_years__25.tif")
obs100_years_26 <- raster("tiffs/Year/obs100_years__26.tif")
obs100_years_27 <- raster("tiffs/Year/obs100_years__27.tif")
obs100_years_28 <- raster("tiffs/Year/obs100_years__28.tif")
obs100_years_29 <- raster("tiffs/Year/obs100_years__29.tif")
obs100_years_30 <- raster("tiffs/Year/obs100_years__30.tif")
obs100_years_31 <- raster("tiffs/Year/obs100_years__31.tif")
obs100_years_32 <- raster("tiffs/Year/obs100_years__32.tif")
obs100_years_33 <- raster("tiffs/Year/obs100_years__33.tif")
obs100_years_34 <- raster("tiffs/Year/obs100_years__34.tif")
obs100_years_35 <- raster("tiffs/Year/obs100_years__35.tif")
obs100_years_36 <- raster("tiffs/Year/obs100_years__36.tif")
obs100_years_37 <- raster("tiffs/Year/obs100_years__37.tif")
obs100_years_38 <- raster("tiffs/Year/obs100_years__38.tif")
obs100_years_39 <- raster("tiffs/Year/obs100_years__39.tif")
obs100_years_40 <- raster("tiffs/Year/obs100_years__40.tif")

stack_obs100_years <-
  stack(obs100_years_1,
      obs100_years_2,
      obs100_years_3,
      obs100_years_4,
```

```
        obs100_years_5,
        obs100_years_6,
        obs100_years_7,
        obs100_years_8,
        obs100_years_9,
        obs100_years_10,
        obs100_years_11,
        obs100_years_12,
        obs100_years_13,
        obs100_years_14,
        obs100_years_15,
        obs100_years_16,
        obs100_years_17,
        obs100_years_18,
        obs100_years_19,
        obs100_years_20,
        obs100_years_21,
        obs100_years_22,
        obs100_years_23,
        obs100_years_24,
        obs100_years_25,
        obs100_years_26,
        obs100_years_27,
        obs100_years_28,
        obs100_years_29,
        obs100_years_30,
        obs100_years_31,
        obs100_years_32,
        obs100_years_33,
        obs100_years_34,
        obs100_years_35,
        obs100_years_36,
        obs100_years_37,
        obs100_years_38,
        obs100_years_39,
        obs100_years_40
  )

# 10km2
obs10_years_1 <- raster("tiffs/Year/obs10_years__1.tif")
obs10_years_2 <- raster("tiffs/Year/obs10_years__2.tif")
obs10_years_3 <- raster("tiffs/Year/obs10_years__3.tif")
obs10_years_4 <- raster("tiffs/Year/obs10_years__4.tif")
obs10_years_5 <- raster("tiffs/Year/obs10_years__5.tif")
obs10_years_6 <- raster("tiffs/Year/obs10_years__6.tif")
obs10_years_7 <- raster("tiffs/Year/obs10_years__7.tif")
obs10_years_8 <- raster("tiffs/Year/obs10_years__8.tif")
obs10_years_9 <- raster("tiffs/Year/obs10_years__9.tif")
obs10_years_10 <- raster("tiffs/Year/obs10_years__10.tif")
obs10_years_11 <- raster("tiffs/Year/obs10_years__11.tif")
obs10_years_12 <- raster("tiffs/Year/obs10_years__12.tif")
obs10_years_13 <- raster("tiffs/Year/obs10_years__13.tif")
obs10_years_14 <- raster("tiffs/Year/obs10_years__14.tif")
obs10_years_15 <- raster("tiffs/Year/obs10_years__15.tif")
obs10_years_16 <- raster("tiffs/Year/obs10_years__16.tif")
obs10_years_17 <- raster("tiffs/Year/obs10_years__17.tif")
obs10_years_18 <- raster("tiffs/Year/obs10_years__18.tif")
obs10_years_19 <- raster("tiffs/Year/obs10_years__19.tif")
obs10_years_20 <- raster("tiffs/Year/obs10_years__20.tif")
obs10_years_21 <- raster("tiffs/Year/obs10_years__21.tif")
obs10_years_22 <- raster("tiffs/Year/obs10_years__22.tif")
obs10_years_23 <- raster("tiffs/Year/obs10_years__23.tif")
obs10_years_24 <- raster("tiffs/Year/obs10_years__24.tif")
obs10_years_25 <- raster("tiffs/Year/obs10_years__25.tif")
obs10_years_26 <- raster("tiffs/Year/obs10_years__26.tif")
obs10_years_27 <- raster("tiffs/Year/obs10_years__27.tif")
obs10_years_28 <- raster("tiffs/Year/obs10_years__28.tif")
obs10_years_29 <- raster("tiffs/Year/obs10_years__29.tif")
obs10_years_30 <- raster("tiffs/Year/obs10_years__30.tif")
obs10_years_31 <- raster("tiffs/Year/obs10_years__31.tif")
obs10_years_32 <- raster("tiffs/Year/obs10_years__32.tif")
obs10_years_33 <- raster("tiffs/Year/obs10_years__33.tif")
obs10_years_34 <- raster("tiffs/Year/obs10_years__34.tif")
obs10_years_35 <- raster("tiffs/Year/obs10_years__35.tif")
obs10_years_36 <- raster("tiffs/Year/obs10_years__36.tif")
```

```
obs10_years_37 <- raster("tiffs/Year/obs10_years__37.tif")
obs10_years_38 <- raster("tiffs/Year/obs10_years__38.tif")
obs10_years_39 <- raster("tiffs/Year/obs10_years__39.tif")
obs10_years_40 <- raster("tiffs/Year/obs10_years__40.tif")

stack_obs10_years <-
  stack(obs10_years_1,
        obs10_years_2,
        obs10_years_3,
        obs10_years_4,
        obs10_years_5,
        obs10_years_6,
        obs10_years_7,
        obs10_years_8,
        obs10_years_9,
        obs10_years_10,
        obs10_years_11,
        obs10_years_12,
        obs10_years_13,
        obs10_years_14,
        obs10_years_15,
        obs10_years_16,
        obs10_years_17,
        obs10_years_18,
        obs10_years_19,
        obs10_years_20,
        obs10_years_21,
        obs10_years_22,
        obs10_years_23,
        obs10_years_24,
        obs10_years_25,
        obs10_years_26,
        obs10_years_27,
        obs10_years_28,
        obs10_years_29,
        obs10_years_30,
        obs10_years_31,
        obs10_years_32,
        obs10_years_33,
        obs10_years_34,
        obs10_years_35,
        obs10_years_36,
        obs10_years_37,
        obs10_years_38,
        obs10_years_39,
        obs10_years_40
  )
# 1km2
obs1_years_1 <- raster("tiffs/Year/obs1_years__1.tif")
obs1_years_2 <- raster("tiffs/Year/obs1_years__2.tif")
obs1_years_3 <- raster("tiffs/Year/obs1_years__3.tif")
obs1_years_4 <- raster("tiffs/Year/obs1_years__4.tif")
obs1_years_5 <- raster("tiffs/Year/obs1_years__5.tif")
obs1_years_6 <- raster("tiffs/Year/obs1_years__6.tif")
obs1_years_7 <- raster("tiffs/Year/obs1_years__7.tif")
obs1_years_8 <- raster("tiffs/Year/obs1_years__8.tif")
obs1_years_9 <- raster("tiffs/Year/obs1_years__9.tif")
obs1_years_10 <- raster("tiffs/Year/obs1_years__10.tif")
obs1_years_11 <- raster("tiffs/Year/obs1_years__11.tif")
obs1_years_12 <- raster("tiffs/Year/obs1_years__12.tif")
obs1_years_13 <- raster("tiffs/Year/obs1_years__13.tif")
obs1_years_14 <- raster("tiffs/Year/obs1_years__14.tif")
obs1_years_15 <- raster("tiffs/Year/obs1_years__15.tif")
obs1_years_16 <- raster("tiffs/Year/obs1_years__16.tif")
obs1_years_17 <- raster("tiffs/Year/obs1_years__17.tif")
obs1_years_18 <- raster("tiffs/Year/obs1_years__18.tif")
obs1_years_19 <- raster("tiffs/Year/obs1_years__19.tif")
obs1_years_20 <- raster("tiffs/Year/obs1_years__20.tif")
obs1_years_21 <- raster("tiffs/Year/obs1_years__21.tif")
obs1_years_22 <- raster("tiffs/Year/obs1_years__22.tif")
obs1_years_23 <- raster("tiffs/Year/obs1_years__23.tif")
obs1_years_24 <- raster("tiffs/Year/obs1_years__24.tif")
obs1_years_25 <- raster("tiffs/Year/obs1_years__25.tif")
obs1_years_26 <- raster("tiffs/Year/obs1_years__26.tif")
obs1_years_27 <- raster("tiffs/Year/obs1_years__27.tif")
```

```
obs1_years_28 <- raster("tiffs/Year/obs1_years__28.tif")
obs1_years_29 <- raster("tiffs/Year/obs1_years__29.tif")
obs1_years_30 <- raster("tiffs/Year/obs1_years__30.tif")
obs1_years_31 <- raster("tiffs/Year/obs1_years__31.tif")
obs1_years_32 <- raster("tiffs/Year/obs1_years__32.tif")
obs1_years_33 <- raster("tiffs/Year/obs1_years__33.tif")
obs1_years_34 <- raster("tiffs/Year/obs1_years__34.tif")
obs1_years_35 <- raster("tiffs/Year/obs1_years__35.tif")
obs1_years_36 <- raster("tiffs/Year/obs1_years__36.tif")
obs1_years_37 <- raster("tiffs/Year/obs1_years__37.tif")
obs1_years_38 <- raster("tiffs/Year/obs1_years__38.tif")
obs1_years_39 <- raster("tiffs/Year/obs1_years__39.tif")
obs1_years_40 <- raster("tiffs/Year/obs1_years__40.tif")

stack_obs1_years <-
  stack(obs1_years_1,
        obs1_years_2,
        obs1_years_3,
        obs1_years_4,
        obs1_years_5,
        obs1_years_6,
        obs1_years_7,
        obs1_years_8,
        obs1_years_9,
        obs1_years_10,
        obs1_years_11,
        obs1_years_12,
        obs1_years_13,
        obs1_years_14,
        obs1_years_15,
        obs1_years_16,
        obs1_years_17,
        obs1_years_18,
        obs1_years_19,
        obs1_years_20,
        obs1_years_21,
        obs1_years_22,
        obs1_years_23,
        obs1_years_24,
        obs1_years_25,
        obs1_years_26,
        obs1_years_27,
        obs1_years_28,
        obs1_years_29,
        obs1_years_30,
        obs1_years_31,
        obs1_years_32,
        obs1_years_33,
        obs1_years_34,
        obs1_years_35,
        obs1_years_36,
        obs1_years_37,
        obs1_years_38,
        obs1_years_39,
        obs1_years_40
  )

rm(obs1_years_1,
   obs1_years_2,
   obs1_years_3,
   obs1_years_4,
   obs1_years_5,
   obs1_years_6,
   obs1_years_7,
   obs1_years_8,
   obs1_years_9,
   obs1_years_10,
   obs1_years_11,
   obs1_years_12,
   obs1_years_13,
   obs1_years_14,
   obs1_years_15,
   obs1_years_16,
   obs1_years_17,
   obs1_years_18,
   obs1_years_19,
```

```
  obs1_years_20,
  obs1_years_21,
  obs1_years_22,
  obs1_years_23,
  obs1_years_24,
  obs1_years_25,
  obs1_years_26,
  obs1_years_27,
  obs1_years_28,
  obs1_years_29,
  obs1_years_30,
  obs1_years_31,
  obs1_years_32,
  obs1_years_33,
  obs1_years_34,
  obs1_years_35,
  obs1_years_36,
  obs1_years_37,
  obs1_years_38,
  obs1_years_39,
  obs1_years_40
)



### c) Effort Tiffs ------------------------------------------------------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------------------------------------------------------------------------
-------
# 100km2
measure100_years_1 <- raster("tiffs/Year/measure100_years__1.tif")
measure100_years_2 <- raster("tiffs/Year/measure100_years__2.tif")
measure100_years_3 <- raster("tiffs/Year/measure100_years__3.tif")
measure100_years_4 <- raster("tiffs/Year/measure100_years__4.tif")
measure100_years_5 <- raster("tiffs/Year/measure100_years__5.tif")
measure100_years_6 <- raster("tiffs/Year/measure100_years__6.tif")
measure100_years_7 <- raster("tiffs/Year/measure100_years__7.tif")
measure100_years_8 <- raster("tiffs/Year/measure100_years__8.tif")
measure100_years_9 <- raster("tiffs/Year/measure100_years__9.tif")
measure100_years_10 <- raster("tiffs/Year/measure100_years__10.tif")
measure100_years_11 <- raster("tiffs/Year/measure100_years__11.tif")
measure100_years_12 <- raster("tiffs/Year/measure100_years__12.tif")
measure100_years_13 <- raster("tiffs/Year/measure100_years__13.tif")
measure100_years_14 <- raster("tiffs/Year/measure100_years__14.tif")
measure100_years_15 <- raster("tiffs/Year/measure100_years__15.tif")
measure100_years_16 <- raster("tiffs/Year/measure100_years__16.tif")
measure100_years_17 <- raster("tiffs/Year/measure100_years__17.tif")
measure100_years_18 <- raster("tiffs/Year/measure100_years__18.tif")
measure100_years_19 <- raster("tiffs/Year/measure100_years__19.tif")
measure100_years_20 <- raster("tiffs/Year/measure100_years__20.tif")
measure100_years_21 <- raster("tiffs/Year/measure100_years__21.tif")
measure100_years_22 <- raster("tiffs/Year/measure100_years__22.tif")
measure100_years_23 <- raster("tiffs/Year/measure100_years__23.tif")
measure100_years_24 <- raster("tiffs/Year/measure100_years__24.tif")
measure100_years_25 <- raster("tiffs/Year/measure100_years__25.tif")
measure100_years_26 <- raster("tiffs/Year/measure100_years__26.tif")
measure100_years_27 <- raster("tiffs/Year/measure100_years__27.tif")
measure100_years_28 <- raster("tiffs/Year/measure100_years__28.tif")
measure100_years_29 <- raster("tiffs/Year/measure100_years__29.tif")
measure100_years_30 <- raster("tiffs/Year/measure100_years__30.tif")
measure100_years_31 <- raster("tiffs/Year/measure100_years__31.tif")
measure100_years_32 <- raster("tiffs/Year/measure100_years__32.tif")
measure100_years_33 <- raster("tiffs/Year/measure100_years__33.tif")
measure100_years_34 <- raster("tiffs/Year/measure100_years__34.tif")
measure100_years_35 <- raster("tiffs/Year/measure100_years__35.tif")
measure100_years_36 <- raster("tiffs/Year/measure100_years__36.tif")
measure100_years_37 <- raster("tiffs/Year/measure100_years__37.tif")
measure100_years_38 <- raster("tiffs/Year/measure100_years__38.tif")
measure100_years_39 <- raster("tiffs/Year/measure100_years__39.tif")
measure100_years_40 <- raster("tiffs/Year/measure100_years__40.tif")

stack_richmeasure_years_100 <-
  stack(measure100_years_1,
    measure100_years_2,
    measure100_years_3,
    measure100_years_4,
```

```
    measure100_years_5,
    measure100_years_6,
    measure100_years_7,
    measure100_years_8,
    measure100_years_9,
    measure100_years_10,
    measure100_years_11,
    measure100_years_12,
    measure100_years_13,
    measure100_years_14,
    measure100_years_15,
    measure100_years_16,
    measure100_years_17,
    measure100_years_18,
    measure100_years_19,
    measure100_years_20,
    measure100_years_21,
    measure100_years_22,
    measure100_years_23,
    measure100_years_24,
    measure100_years_25,
    measure100_years_26,
    measure100_years_27,
    measure100_years_28,
    measure100_years_29,
    measure100_years_30,
    measure100_years_31,
    measure100_years_32,
    measure100_years_33,
    measure100_years_34,
    measure100_years_35,
    measure100_years_36,
    measure100_years_37,
    measure100_years_38,
    measure100_years_39,
    measure100_years_40
  )
# 10km2
measure10_years_1 <- raster("tiffs/Year/measure10_years__1.tif")
measure10_years_2 <- raster("tiffs/Year/measure10_years__2.tif")
measure10_years_3 <- raster("tiffs/Year/measure10_years__3.tif")
measure10_years_4 <- raster("tiffs/Year/measure10_years__4.tif")
measure10_years_5 <- raster("tiffs/Year/measure10_years__5.tif")
measure10_years_6 <- raster("tiffs/Year/measure10_years__6.tif")
measure10_years_7 <- raster("tiffs/Year/measure10_years__7.tif")
measure10_years_8 <- raster("tiffs/Year/measure10_years__8.tif")
measure10_years_9 <- raster("tiffs/Year/measure10_years__9.tif")
measure10_years_10 <- raster("tiffs/Year/measure10_years__10.tif")
measure10_years_11 <- raster("tiffs/Year/measure10_years__11.tif")
measure10_years_12 <- raster("tiffs/Year/measure10_years__12.tif")
measure10_years_13 <- raster("tiffs/Year/measure10_years__13.tif")
measure10_years_14 <- raster("tiffs/Year/measure10_years__14.tif")
measure10_years_15 <- raster("tiffs/Year/measure10_years__15.tif")
measure10_years_16 <- raster("tiffs/Year/measure10_years__16.tif")
measure10_years_17 <- raster("tiffs/Year/measure10_years__17.tif")
measure10_years_18 <- raster("tiffs/Year/measure10_years__18.tif")
measure10_years_19 <- raster("tiffs/Year/measure10_years__19.tif")
measure10_years_20 <- raster("tiffs/Year/measure10_years__20.tif")
measure10_years_21 <- raster("tiffs/Year/measure10_years__21.tif")
measure10_years_22 <- raster("tiffs/Year/measure10_years__22.tif")
measure10_years_23 <- raster("tiffs/Year/measure10_years__23.tif")
measure10_years_24 <- raster("tiffs/Year/measure10_years__24.tif")
measure10_years_25 <- raster("tiffs/Year/measure10_years__25.tif")
measure10_years_26 <- raster("tiffs/Year/measure10_years__26.tif")
measure10_years_27 <- raster("tiffs/Year/measure10_years__27.tif")
measure10_years_28 <- raster("tiffs/Year/measure10_years__28.tif")
measure10_years_29 <- raster("tiffs/Year/measure10_years__29.tif")
measure10_years_30 <- raster("tiffs/Year/measure10_years__30.tif")
measure10_years_31 <- raster("tiffs/Year/measure10_years__31.tif")
measure10_years_32 <- raster("tiffs/Year/measure10_years__32.tif")
measure10_years_33 <- raster("tiffs/Year/measure10_years__33.tif")
measure10_years_34 <- raster("tiffs/Year/measure10_years__34.tif")
measure10_years_35 <- raster("tiffs/Year/measure10_years__35.tif")
measure10_years_36 <- raster("tiffs/Year/measure10_years__36.tif")
measure10_years_37 <- raster("tiffs/Year/measure10_years__37.tif")
```

```
measure10_years_38 <- raster("tiffs/Year/measure10_years__38.tif")
measure10_years_39 <- raster("tiffs/Year/measure10_years__39.tif")
measure10_years_40 <- raster("tiffs/Year/measure10_years__40.tif")

stack_richmeasure_years_10 <-
  stack(measure10_years_1,
        measure10_years_2,
        measure10_years_3,
        measure10_years_4,
        measure10_years_5,
        measure10_years_6,
        measure10_years_7,
        measure10_years_8,
        measure10_years_9,
        measure10_years_10,
        measure10_years_11,
        measure10_years_12,
        measure10_years_13,
        measure10_years_14,
        measure10_years_15,
        measure10_years_16,
        measure10_years_17,
        measure10_years_18,
        measure10_years_19,
        measure10_years_20,
        measure10_years_21,
        measure10_years_22,
        measure10_years_23,
        measure10_years_24,
        measure10_years_25,
        measure10_years_26,
        measure10_years_27,
        measure10_years_28,
        measure10_years_29,
        measure10_years_30,
        measure10_years_31,
        measure10_years_32,
        measure10_years_33,
        measure10_years_34,
        measure10_years_35,
        measure10_years_36,
        measure10_years_37,
        measure10_years_38,
        measure10_years_39,
        measure10_years_40
  )

rm(measure10_years_1,
        measure10_years_2,
        measure10_years_3,
        measure10_years_4,
        measure10_years_5,
        measure10_years_6,
        measure10_years_7,
        measure10_years_8,
        measure10_years_9,
        measure10_years_10,
        measure10_years_11,
        measure10_years_12,
        measure10_years_13,
        measure10_years_14,
        measure10_years_15,
        measure10_years_16,
        measure10_years_17,
        measure10_years_18,
        measure10_years_19,
        measure10_years_20,
        measure10_years_21,
        measure10_years_22,
        measure10_years_23,
        measure10_years_24,
        measure10_years_25,
        measure10_years_26,
        measure10_years_27,
        measure10_years_28,
        measure10_years_29,
```

```
        measure10_years_30,
        measure10_years_31,
        measure10_years_32,
        measure10_years_33,
        measure10_years_34,
        measure10_years_35,
        measure10_years_36,
        measure10_years_37,
        measure10_years_38,
        measure10_years_39,
        measure10_years_40
  )

# 1km2

measure1_years_1 <- raster("tiffs/Year/measure1_years__1.tif")
measure1_years_2 <- raster("tiffs/Year/measure1_years__2.tif")
measure1_years_3 <- raster("tiffs/Year/measure1_years__3.tif")
measure1_years_4 <- raster("tiffs/Year/measure1_years__4.tif")
measure1_years_5 <- raster("tiffs/Year/measure1_years__5.tif")
measure1_years_6 <- raster("tiffs/Year/measure1_years__6.tif")
measure1_years_7 <- raster("tiffs/Year/measure1_years__7.tif")
measure1_years_8 <- raster("tiffs/Year/measure1_years__8.tif")
measure1_years_9 <- raster("tiffs/Year/measure1_years__9.tif")
measure1_years_10 <- raster("tiffs/Year/measure1_years__10.tif")
measure1_years_11 <- raster("tiffs/Year/measure1_years__11.tif")
measure1_years_12 <- raster("tiffs/Year/measure1_years__12.tif")
measure1_years_13 <- raster("tiffs/Year/measure1_years__13.tif")
measure1_years_14 <- raster("tiffs/Year/measure1_years__14.tif")
measure1_years_15 <- raster("tiffs/Year/measure1_years__15.tif")
measure1_years_16 <- raster("tiffs/Year/measure1_years__16.tif")
measure1_years_17 <- raster("tiffs/Year/measure1_years__17.tif")
measure1_years_18 <- raster("tiffs/Year/measure1_years__18.tif")
measure1_years_19 <- raster("tiffs/Year/measure1_years__19.tif")
measure1_years_20 <- raster("tiffs/Year/measure1_years__20.tif")
measure1_years_21 <- raster("tiffs/Year/measure1_years__21.tif")
measure1_years_22 <- raster("tiffs/Year/measure1_years__22.tif")
measure1_years_23 <- raster("tiffs/Year/measure1_years__23.tif")
measure1_years_24 <- raster("tiffs/Year/measure1_years__24.tif")
measure1_years_25 <- raster("tiffs/Year/measure1_years__25.tif")
measure1_years_26 <- raster("tiffs/Year/measure1_years__26.tif")
measure1_years_27 <- raster("tiffs/Year/measure1_years__27.tif")
measure1_years_28 <- raster("tiffs/Year/measure1_years__28.tif")
measure1_years_29 <- raster("tiffs/Year/measure1_years__29.tif")
measure1_years_30 <- raster("tiffs/Year/measure1_years__30.tif")
measure1_years_31 <- raster("tiffs/Year/measure1_years__31.tif")
measure1_years_32 <- raster("tiffs/Year/measure1_years__32.tif")
measure1_years_33 <- raster("tiffs/Year/measure1_years__33.tif")
measure1_years_34 <- raster("tiffs/Year/measure1_years__34.tif")
measure1_years_35 <- raster("tiffs/Year/measure1_years__35.tif")
measure1_years_36 <- raster("tiffs/Year/measure1_years__36.tif")
measure1_years_37 <- raster("tiffs/Year/measure1_years__37.tif")
measure1_years_38 <- raster("tiffs/Year/measure1_years__38.tif")
measure1_years_39 <- raster("tiffs/Year/measure1_years__39.tif")
measure1_years_40 <- raster("tiffs/Year/measure1_years__40.tif")

stack_richmeasure_years_1 <-
  stack(measure1_years_1,
        measure1_years_2,
        measure1_years_3,
        measure1_years_4,
        measure1_years_5,
        measure1_years_6,
        measure1_years_7,
        measure1_years_8,
        measure1_years_9,
        measure1_years_10,
        measure1_years_11,
        measure1_years_12,
        measure1_years_13,
        measure1_years_14,
        measure1_years_15,
        measure1_years_16,
        measure1_years_17,
        measure1_years_18,
        measure1_years_19,
```

```
    measure1_years_20,
    measure1_years_21,
    measure1_years_22,
    measure1_years_23,
    measure1_years_24,
    measure1_years_25,
    measure1_years_26,
    measure1_years_27,
    measure1_years_28,
    measure1_years_29,
    measure1_years_30,
    measure1_years_31,
    measure1_years_32,
    measure1_years_33,
    measure1_years_34,
    measure1_years_35,
    measure1_years_36,
    measure1_years_37,
    measure1_years_38,
    measure1_years_39,
    measure1_years_40
 )

### 3) SubYears -###-------------------------------------------------------------------------------------------------------------------------------
# a) Richness Tiffs
stack_rich100_subyear
#setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Richness")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Richness")
files <- list.files(pattern = ".tif")
stack_rich100_subyear <- stack(files)

# setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Richness")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Richness")
files <- list.files(pattern = ".tif")
stack_rich10_subyear <- stack(files)

setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Richness")
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Richness")
files <- list.files(pattern = ".tif")
stack_rich1_subyear <- stack(files)

df_rich1_subyear<- as.data.frame(stack_rich1_subyear) # Convert to dataframe for use with GGplot
colnames(df_rich100_months) <- Months
col_rich100_months <- gather(df_rich100_months) #Tidyr to Reshaping wide format to long format
# View(col_rich100_months)
key <- unique(col_rich100_months$key) # Check for all months present
rm(df_rich100_months)

setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Richness")
# setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Richness")
files <- list.files(pattern = ".tif")
stack_rich1_subyear <- stack(files)

files <- list.files(pattern = ".tif")
results <- lapply(stack_rich1_subyear, as.data.frame)


# b) Obs Tiffs
stack_obs100_subyear
#setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Obs")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Obs")
files <- list.files(pattern = ".tif")
stack_obs100_subyear <- stack(files)

stack_obs10_subyear
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Obs")
#setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Obs")
files <- list.files(pattern = ".tif")
stack_obs10_subyear <- stack(files)
```

```
stack_obs1_subyear
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Obs")
#setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Obs")
files <- list.files(pattern = ".tif")
stack_obs1_subyear <- stack(files)

# c) Effort Tiffs
stack_richmeasure_subyear_100
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Measure")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Measure")
files <- list.files(pattern = ".tif")
stack_richmeasure_subyear_100 <- stack(files)

stack_richmeasure_subyear_10
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Measure")
files <- list.files(pattern = ".tif")
stack_richmeasure_subyear_10 <- stack(files)

stack_richmeasure_subyear_1
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Measure")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Richness")
files <- list.files(pattern = ".tif")
stack_richmeasure_subyear_1 <- stack(files)
dfrichmeasure_subyear_1<- as.data.frame(stack_richmeasure_subyear_1) # Convert to dataframe for use with GGplot

### -End of Script ###------------------------------------------------------------------------------
```

Script 9: Levelplot maps

```
# Setup ----------------------------------------
# Load packages
library(sp)
library(raster)
library(viridis)
library(rasterVis)
options(scipen = 999)

# 1) Plots of SPR's for all data  ----------------------------------------
# I 1_Data_Setup Script
# 2) Levelplots of Richness For Months ----------------------------------------
# requires from Script 8 Data loading:
stack_rich100_months
stack_rich10_months
stack_rich10_months
levelpalette <- viridis(22) # Set new palette
# 100km2  ----------------------------------------
names(stack_rich100_months)
months <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")
names(stack_rich100_months) <- months
# Lattice-style levelplot - 100km2
pngfile_1 <- "Figs/by_month/LevelplotMap_Months_100km2.png"
png(pngfile_1, width=729, height=729)   # open the file
cutpts_1 <- c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22)            # Define legend breaks
plt <- levelplot(stack_rich100_months,  # Raster of choice
               at=cutpts_1, # Cutpoints as defined above
               cuts=22, #Break points - n of cutpoints
               pretty=T,
               col.regions=levelpalette[0:22],
               xlab = 'X Coordinate',
               ylab = 'Y Coordinate',
               cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off() # close the file
# 10km2  ----------------------------------------
names(stack_rich10_months) <- months
# lattice-style levelplot - 10km2
pngfile_2 <- "Figs/by_month/LevelplotMap_Months_10km2.png"
png(pngfile_2, width=729, height=729)   # open the file
cutpts <- c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17)            # Define legend breaks
plt <- levelplot(stack_rich10_months, # Raster of choice
               at=cutpts, # Cutpoints as defined above
               cuts=17, #Break points - n of cutpoints
               pretty=T,
               col.regions=levelpalette[0:17],
               xlab = 'X Coordinate',
               ylab = 'Y Coordinate',
               cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off() # close the file
# 1km2  ----------------------------------------
names(stack_rich1_months) <- months
# lattice-style levelplot - 1km2
pngfile_3 <- "Figs/by_month/LevelplotMap_Months_1km2.png"
png(pngfile_3, width=729, height=729)   # open the file
cutpts <- c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16)            # Define legend breaks
plt <- levelplot(stack_rich1_months, # Raster of choice
               at=cutpts, # Cutpoints as defined above
               cuts=16, #Break points - n of cutpoints
               pretty=T,
               col.regions=levelpalette[0:16],
               xlab = 'X Coordinate',
               ylab = 'Y Coordinate')
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off()
# close the file
# Remove if needed:
rm(stack_rich100_months, stack_rich10_months, stack_rich1_months)
```

```
# 3) Levelplots of MEASURE For Months ----------------------------------------
# requires from # 8 Data loading:
stack_richmeasure_months_100
stack_richmeasure_months_10
stack_richmeasure_months_1
levelpalette <- viridis(22) # Set new palette
# 100km2  ----------------------------------------
names(stack_richmeasure_months_100)
months <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")
names(stack_richmeasure_months_100) <- months
stack_richmeasure_months_100
# Check max and min vals to set cutpoints
# Lattice-style levelplot - 100km2
pngfile_1 <- "Figs/by_month/LevelplotMap_MEASURE_Months_100km2.png"
png(pngfile_1, width=729, height=729)   # open the file
cutpts_1 <- c(0, 0.5, 1, 1.5,2,2.5,3,3.5,4,4.5)            # Define legend breaks
plt <- levelplot(stack_richmeasure_months_100,  # Raster of choice
           at=cutpts_1, # Cutpoints as defined above
           cuts=10, #Break points - n of cutpoints
           pretty=T,
           col.regions=levelpalette[0:10],
           xlab = 'X Coordinate',
           ylab = 'Y Coordinate',
           cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off()
# close the file
# 10km2  ----------------------------------------
names(stack_richmeasure_months_10) <- months
# lattice-style levelplot - 10km2
pngfile_2 <- "Figs/by_month/LevelplotMap_MEASURE_Months_10km2.png"
png(pngfile_2, width=729, height=729)   # open the file
plt <- levelplot(stack_richmeasure_months_10, # Raster of choice
           at=cutpts_1, # Cutpoints as defined above
           cuts=10, #Break points - n of cutpoints
           pretty=T,
           col.regions=levelpalette[0:10],
           xlab = 'X Coordinate',
           ylab = 'Y Coordinate',
           cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off() # close the file
# 1km2  ----------------------------------------
names(stack_richmeasure_months_1) <- months
# lattice-style levelplot - 1km2
pngfile_3 <- "Figs/by_month/LevelplotMap_MEASURE_Months_1km2.png"
png(pngfile_3, width=729, height=729)   # open the file
plt <- levelplot(stack_richmeasure_months_1, # Raster of choice
           at=cutpts_1, # Cutpoints as defined above
           cuts=10, #Break points - n of cutpoints
           pretty=T,
           col.regions=levelpalette[0:10],
           xlab = 'X Coordinate',
           ylab = 'Y Coordinate')
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off() # close the file
# Remove if needed:
rm(stack_rich100_months, stack_rich10_months, stack_rich1_months)

### 3) Levelplots of SPR's For Years ### ----------------------------------------
# 100 km
# Rename layers to years
names(stack_rich100_years)
years_labels                                                                          <-
c("Y1979","Y1980","Y1981","Y1982","Y1983","Y1984","Y1985","Y1986","Y1987","Y1988","Y1989","Y1990","Y1991","Y
1992","Y1993","Y1994","Y1995","Y1996","Y1997","Y1998","Y1999","Y2000","Y2001","Y2002","Y2003","Y2004","Y2005
","Y2006","Y2007","Y2008","Y2009","Y2010","Y2011","Y2012","Y2013","Y2014","Y2015","Y2016","Y2017","Y2018")
names(stack_rich100_years) <- years_labels
# 100km2
maxValue(stack_rich100_years)
pngfile <- "Figs/by_year/Levelplot_Richness_Years_100km2.png"
png(pngfile, width=1400, height=729)   # open the file
```

```
cutpts <- c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22)          # Define legend breaks
plt <- levelplot(stack_rich100_years,  # Raster of choice
                 at=cutpts, # Cutpoints as defined above
                 cuts=22, #Break points - n of cutpoints
                 pretty=T,
                 col.regions=levelpalette,
                 xlab = 'X Coordinate',
                 ylab = 'Y Coordinate',
                 cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off() # close the file
# 10km2
names(stack_rich10_years) <- years_labels
maxValue(stack_rich10_years)
pngfile <- "Figs/by_year/Levelplot__Richness_Years_10km2.png"
png(pngfile, width=1400, height=729)   # open the file
cutpts <- c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19)          # Define legend breaks
plt <- levelplot(stack_rich10_years,  # Raster of choice
                 at=cutpts, # Cutpoints as defined above
                 cuts=19, #Break points - n of cutpoints
                 pretty=T,
                 col.regions=levelpalette[0:19],
                 xlab = 'X Coordinate',
                 ylab = 'Y Coordinate',
                 cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off() # close the file
# 1km2
maxValue(stack_rich1_years)
names(stack_rich1_years) <- years_labels
# Max = 16
pngfile <- "Figs/by_year/Levelplot_Richness_Years_1km2.png"
png(pngfile, width=1400, height=729)   # open the file
cutpts <- c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16)          # Define legend breaks
plt <- levelplot(stack_rich1_years,  # Raster of choice
                 at=cutpts, # Cutpoints as defined above
                 cuts=16, #Break points - n of cutpoints
                 pretty=T,
                 col.regions=levelpalette[0:16],
                 xlab = 'X Coordinate',
                 ylab = 'Y Coordinate',
                 cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off() # close the file
dev.off() #
### 3) Levelplots of measure (SRPUE) for years ### ---------------------------------------
# Rename layers to years
names(stack_richmeasure_years_100)
years_labels                                                                                          <-
c("Y1979","Y1980","Y1981","Y1982","Y1983","Y1984","Y1985","Y1986","Y1987","Y1988","Y1989","Y1990","Y1991","Y
1992","Y1993","Y1994","Y1995","Y1996","Y1997","Y1998","Y1999","Y2000","Y2001","Y2002","Y2003","Y2004","Y2005
","Y2006","Y2007","Y2008","Y2009","Y2010","Y2011","Y2012","Y2013","Y2014","Y2015","Y2016","Y2017","Y2018")
names(stack_richmeasure_years_100) <- years_labels
names(stack_richmeasure_years_10) <- years_labels
names(stack_richmeasure_years_1) <- years_labels
# 100km2
max(maxValue(stack_richmeasure_years_100))
# Max = 22 again
pngfile <- "Figs/by_year/Levelplot_Measure_Years_100km2.png"
png(pngfile, width=1400, height=729)   # open the file
cutpts_1   # Define legend breaks
plt <- levelplot(stack_richmeasure_years_100,  # Raster of choice
                 at=cutpts_1, # Cutpoints as defined above
                 cuts=10, #Break points - n of cutpoints
                 pretty=T,
                 col.regions=levelpalette,
                 xlab = 'X Coordinate',
                 ylab = 'Y Coordinate',
                 cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off() # close the file
```

```
pngfile <- "Figs/by_year/Levelplot_Measure_Years_10km2.png"
png(pngfile, width=1400, height=729)   # open the file
cutpts_1   # Define legend breaks
plt <- levelplot(stack_richmeasure_years_10,  # Raster of choice
                at=cutpts_1, # Cutpoints as defined above
                cuts=10, #Break points - n of cutpoints
                pretty=T,
                col.regions=levelpalette,
                xlab = 'X Coordinate',
                ylab = 'Y Coordinate',
                cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off()

pngfile <- "Figs/by_year/Levelplot_Measure_Years_1km2.png"
png(pngfile, width=1400, height=729)   # open the file
cutpts_1   # Define legend breaks
plt <- levelplot(stack_richmeasure_years_1,  # Raster of choice
                at=cutpts_1, # Cutpoints as defined above
                cuts=10, #Break points - n of cutpoints
                pretty=T,
                col.regions=levelpalette,
                xlab = 'X Coordinate',
                ylab = 'Y Coordinate',
                cex.axis = 0.5)
plt + latticeExtra::layer(sp.lines(outlines, col="black", fill = "gray", lwd=0.01))
dev.off()
### -End of Script ###-----------------------------------------------------------------------
```

Script 10: Convert data to dataframes for subsequent analysis

```
# Load packages
library(raster)
library(terra)
library(tidyverse)
library(dplyr)
# Aim: to get data from rasters into useable dataframes with richness and observation per cell for each
# 100km
rich100_all <- terra::rast(rich_100)        # Make SpatRaster (Terra)
terra::values(rich100_all, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
view(rich100_all)
df_rich100_all <- terra::as.data.frame(rich100_all, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY
coordinates, keeps cell number
colnames(df_rich100_all) <- c("Cell", "X", "Y", "Richness") # Rename columns
head(df_rich100_all)

obs100_all <- terra::rast(obs_100)        # Make SpatRaster (Terra)
terra::values(obs100_all, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
view(obs100_all)
df_obs100_all <- terra::as.data.frame(obs100_all, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY
coordinates, keeps cell number
colnames(df_obs100_all) <- c("Cell", "X", "Y", "Observations") # Rename columns
head(df_obs100_all)

# Combine richness and obs to one df for use
df_100km_all <- cbind(df_rich100_all,df_obs100_all[4])
rm(rich100_all, obs100_all)

# 10km
rich10_all <- terra::rast(rich_10)        # Make SpatRaster (Terra)
terra::values(rich10_all, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
view(rich10_all)
df_rich10_all <- terra::as.data.frame(rich10_all, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY
coordinates, keeps cell number
colnames(df_rich10_all) <- c("Cell", "X", "Y", "Richness") # Rename columns
head(df_rich10_all)

obs10_all <- terra::rast(obs_10)        # Make SpatRaster (Terra)
terra::values(obs10_all, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
view(obs10_all)
df_obs10_all <- terra::as.data.frame(obs10_all, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY
coordinates, keeps cell number
colnames(df_obs10_all) <- c("Cell", "X", "Y", "Observations") # Rename columns
head(df_obs10_all)

# Combine richness and obs to one df for use
df_10km_all <- cbind(df_rich10_all,df_obs10_all[4])
rm(rich_10, obs_10, rich10_all, obs10_all)
# 1km
rich1_all <- terra::rast(rich_1)
# Make SpatRaster (Terra)
terra::values(rich1_all, mat = TRUE)
# Returns all cell values of a SpatRaster (a matrix),
df_rich1_all <- terra::as.data.frame(rich1_all, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY
coordinates, keeps cell number
colnames(df_rich1_all) <- c("Cell", "X", "Y", "Richness") # Rename columns
head(df_rich1_all)

obs1_all <- terra::rast(obs_1)
# Make SpatRaster (Terra)
terra::values(obs1_all, mat = TRUE)
# Returns all cell values of a SpatRaster (a matrix),
df_obs1_all <- terra::as.data.frame(obs1_all, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY
coordinates, keeps cell number
colnames(df_obs1_all) <- c("Cell", "X", "Y", "Observations") # Rename columns
head(df_obs1_all)

# Combine richness and obs to one df for use
```

```
df_1km_all <- cbind(df_rich1_all,df_obs1_all[4])
rm(rich_1, obs_1, rich1_all, obs1_all)
----------------------------------------------------------------------------------------------------------------------
### Months ###
----------------------------------------------------------------------------------------------------------------------
# 100 km Richness Values Months
rich100_months <- terra::rast(stack_rich100_months)
 # Make SpatRaster (Terra)
terra::values(rich100_months, mat = TRUE)
# Returns all cell values of a SpatRaster (a matrix),
view(rich100_months)
df_rich100_months <- terra::as.data.frame(rich100_months, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY coordinates, keeps cell number
labels <- c('Cell', 'x', 'y', 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec') # Col headings
colnames(df_rich100_months) <- labels # Rename columns
head(df_rich100_months)

rm(stack_rich100_months, rich100_months)

# 100 km OBS Values Months
obs100_months <- terra::rast(stack_obs100_months)         # Make SpatRaster (Terra)
terra::values(obs100_months, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
df_obs100_months <- terra::as.data.frame(obs100_months, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY coordinates, keeps cell number
labels <- c('Cell', 'x', 'y', 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec') # Col headings
colnames(df_obs100_months) <- labels # Rename columns
head(df_obs100_months)
rm(stack_obs100_months, obs100_months)

# 10 km Richness Values Months
rich10_months <- terra::rast(stack_rich10_months)
# Make SpatRaster (Terra)
terra::values(rich10_months, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
df_rich10_months <- terra::as.data.frame(rich10_months, xy = TRUE, cells = TRUE, na.rm = FALSE)
# Makes a dataframe - keeps XY coordinates, keeps cell number
labels <- c('Cell', 'x', 'y', 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec') # Col headings
colnames(df_rich10_months) <- labels # Rename columns
head(df_rich10_months)
rm(stack_rich10_months, rich10_months)

# 10 km OBS Values Months
obs10_months <- terra::rast(stack_obs10_months)         # Make SpatRaster (Terra)
terra::values(obs10_months, mat = TRUE)
# Returns all cell values of a SpatRaster (a matrix)
df_obs10_months <- terra::as.data.frame(obs10_months, xy = TRUE, cells = TRUE, na.rm = FALSE)
# Makes a dataframe - keeps XY coordinates, keeps cell number
head(df_obs10_months)
labels <- c('Cell', 'x', 'y', 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec') # Col headings
colnames(df_obs10_months) <- labels # Rename columns
head(df_obs10_months)

rm(stack_obs10_months, obs10_months)

# 1 km Richness Values Months
rich1_months <- terra::rast(stack_rich1_months)         # Make SpatRaster (Terra)
terra::values(rich1_months, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
df_rich1_months <- terra::as.data.frame(rich1_months, xy = TRUE, cells = TRUE, na.rm = FALSE)
# Makes a dataframe - keeps XY coordinates, keeps cell number
labels <- c('Cell', 'x', 'y', 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec') # Col headings
colnames(df_rich1_months) <- labels # Rename columns
head(df_rich1_months)
rm(stack_rich1_months, rich1_months)

# 1 km OBS Values Months
obs1_months <- terra::rast(stack_obs1_months)         # Make SpatRaster (Terra)
terra::values(obs1_months, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
df_obs1_months <- terra::as.data.frame(obs1_months, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps XY coordinates, keeps cell number
head(df_obs1_months)
```

```
labels <- c('Cell', 'x', 'y', 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec') # Col headings
colnames(df_obs1_months) <- labels # Rename columns
head(df_obs1_months)

rm(stack_obs1_months, obs1_months)

### 1 Matrix per Res - by Months ----------------------
# 100km
head(df_rich100_months)
df_1 <- cbind(df_rich100_months, "100")
rich_month_labels <- c("Cell", "X", "Y", "Richness_Jan", "Richness_Feb", "Richness_Mar", "Richness_Apr", "Richness_May",
"Richness_Jun", "Richness_Jul", "Richness_Aug", "Richness_Sep", "Richness_Oct", "Richness_Nov", "Richness_Dec",
"Resolution")
colnames(df_1) <- rich_month_labels
head(df_1)
df_2 <- cbind(df_obs100_months, "100")
obs_month_labels <- c("Cell", "X", "Y", "Observations__Jan", "Observations__Feb", "Observations__Mar",
"Observations__Apr", "Observations__May", "Observations__Jun", "Observations__Jul", "Observations__Aug",
"Observations__Sep", "Observations__Oct", "Observations__Nov", "Observations__Dec", "Resolution")
colnames(df_2) <- obs_month_labels
df_months_100 <- cbind(df_1, df_2[4:15])
# 10km
df_3 <- cbind(df_rich10_months, "10")
colnames(df_3) <- rich_month_labels

df_4 <- cbind(df_obs10_months, "10")
colnames(df_4) <- obs_month_labels

df_months_10 <- cbind(df_3, df_4[4:15])
head(df_months_10)

rm(df_1, df_2, df_3, df_4)

# -----------------------------------------------------------------------------------------------------------------
### Subyear ###
#----------------------------------------------------------------------------------------------------------------
# Richness:
# 100 ###

# Required:
stack_rich100_subyear
# PC:
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Richness")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Richness")
files <- list.files(pattern = ".tif")
stack_rich100_subyear <- stack(files)

setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")

rich100_subyear <- terra::rast(stack_rich100_subyear)        # Make SpatRaster (Terra)
rm(stack_rich100_subyear)
terra::values(rich100_subyear, mat = TRUE)
# Returns all cell values of a SpatRaster (a matrix),
df_rich100_subyear <- terra::as.data.frame(rich100_subyear, xy = TRUE, cells = TRUE, na.rm = FALSE)
# Makes a dataframe - keeps XY coordinates, keeps cell number
head(df_rich100_subyear)
colnames(df_rich100_subyear)
rm(rich100_subyear)

save(df_rich100_subyear, file="df_rich_100_subyear.Rda")

### 10 ###
# Required:
stack_rich10_subyear
# PC:
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Richness")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Richness")
files <- list.files(pattern = ".tif")
stack_rich10_subyear <- stack(files)
```

```
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")

rich10_subyear <- terra::rast(stack_rich10_subyear)        # Make SpatRaster (Terra)
rm(stack_rich10_subyear)
terra::values(rich10_subyear, mat = TRUE)
# Returns all cell values of a SpatRaster (a matrix),
df_rich10_subyear <- terra::as.data.frame(rich10_subyear, xy = TRUE, cells = TRUE, na.rm = FALSE)
# Makes a dataframe - keeps XY coordinates, keeps cell number
head(df_rich10_subyear)
colnames(df_rich10_subyear)
rm(rich10_subyear)

save(df_rich10_subyear, file="df_rich_10_subyear.Rda")


### 1 ###
# Required:
stack_rich1_subyear
# PC:
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Richness")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Richness")
files <- list.files(pattern = ".tif")
stack_rich1_subyear <- stack(files)

setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")

rich1_subyear <- terra::rast(stack_rich1_subyear)        # Make SpatRaster (Terra)
rm(stack_rich1_subyear)
terra::values(rich1_subyear, mat = TRUE)
# Returns all cell values of a SpatRaster (a matrix),
df_rich1_subyear <- terra::as.data.frame(rich1_subyear, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe -
keeps XY coordinates, keeps cell number
head(df_rich1_subyear)
colnames(df_rich1_subyear)
rm(rich1_subyear)

save(df_rich1_subyear, file="df_rich_1_subyear.Rda")


# Observations: ###
### 100 ###
# Required
stack_obs100_subyear
setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Obs")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/Obs")
files <- list.files(pattern = ".tif")
stack_obs100_subyear <- stack(files)

setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")

obs100_subyear <- terra::rast(stack_obs100_subyear)        # Make SpatRaster (Terra)
rm(stack_obs100_subyear)
# terra::values(obs100_subyear, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
df_obs100_subyear <- terra::as.data.frame(obs100_subyear, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe -
keeps XY coordinates, keeps cell number
head(df_obs100_subyear)
colnames(df_obs100_subyear)
rm(obs100_subyear)

save(df_obs100_subyear, file="df_obs_100_subyear.Rda")


### 10 km2 ###
# Required
stack_obs10_subyear
#setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Obs")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/Obs")
files <- list.files(pattern = ".tif")
stack_obs10_subyear <- stack(files)
```

```
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")

obs10_subyear <- terra::rast(stack_obs10_subyear)        # Make SpatRaster (Terra)
rm(stack_obs10_subyear)
# terra::values(obs10_subyear, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
df_obs10_subyear <- terra::as.data.frame(obs10_subyear, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe -
keeps XY coordinates, keeps cell number
head(df_obs10_subyear)
colnames(df_obs10_subyear)
rm(obs10_subyear)

save(df_obs10_subyear, file="df_obs_10_subyear.Rda")
### 1 km2 ###
# Required
stack_obs1_subyear
#setwd("D:/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Obs")
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Obs")
files <- list.files(pattern = ".tif")
stack_obs1_subyear <- stack(files)

setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")

obs1_subyear <- terra::rast(stack_obs1_subyear)        # Make SpatRaster (Terra)
rm(stack_obs1_subyear)
# terra::values(obs1_subyear, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
df_obs1_subyear <- terra::as.data.frame(obs1_subyear, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps
XY coordinates, keeps cell number
head(df_obs1_subyear)
colnames(df_obs1_subyear)
rm(obs1_subyear)

save(df_obs1_subyear, file="df_obs_1_subyear.Rda")


### Breaking it down
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/Obs")
files <- list.files(pattern = ".tif")
stack_obs1_subyear <- stack(files)
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")

obs1_subyear_1 <- terra::rast(stack_obs1_subyear[1])        # Make SpatRaster (Terra)
rm(stack_obs1_subyear)
# terra::values(obs1_subyear, mat = TRUE) # Returns all cell values of a SpatRaster (a matrix),
df_obs1_subyear <- terra::as.data.frame(obs1_subyear, xy = TRUE, cells = TRUE, na.rm = FALSE) # Makes a dataframe - keeps
XY coordinates, keeps cell number
head(df_obs1_subyear)
colnames(df_obs1_subyear)
rm(obs1_subyear)

save(df_obs1_subyear, file="df_obs_1_subyear.Rda")
###Data to df - subyears ###-------------------------------------------------------------------------------

library(raster)
library(terra)
library(tidyverse)
library(tidyr)
library(dplyr)
library(plyr)
library(lubridate)



# Getting one dataframe of all data for subyears (480 months) -----------------------------------------------------------------------------
# Load Richness Files for each of 480 months
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/txt files/Rich")
files_rich <- list.files(pattern = ".txt")
list_rich <- list()
for(i in 1:480){
  list_rich [[i]] <- read.csv(files_rich[[i]], sep="")
  print(i)
}
```

```
# Load Obs Files for each of 480 months
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/Month_Year_Sub/1/txt files/Obs")
files_obs <- list.files(pattern = ".txt")
list_obs <- list()
for(i in 1:480){
  list_obs [[i]] <- read.csv(files_obs[[i]], sep="")
  print(i)
}
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")

# Make it "longer" - so put  480 variables into one column
list_rich_longer <- list()
for(i in 1:480){
  list_rich_longer[[i]] <- pivot_longer(list_rich[[i]], cols = 4:4)
  colnames(list_rich_longer[[i]]) <- c("Cell", "X", "Y", "Name", "Richness")
  print(i)
}
list_rich_longer

list_obs_longer <- list()
for(i in 1:480){
  list_obs_longer[[i]] <- pivot_longer(list_obs[[i]], cols = 4:4)
  colnames(list_obs_longer[[i]]) <- c("Cell", "X", "Y", "Name", "Observations")
  print(i)
}

# Check what it looks like for DF with vals
list_rich[6]
list_rich_longer[6]
list_obs_longer[6]

# Combine richness list and observation list
list_all_combined <- list()
for(i in 1:480){
  list_all_combined[[i]] <- cbind(list_rich_longer[[i]], list_obs_longer[[i]])
  print(i)
}

# Combine all list into one dataframe
df <- ldply(list_all_combined, data.frame)
head(df)
# Data tidy
# Rename Columns
colnames(df)
final_df <- df %>% select(-Cell.1, -X.1, -Y.1, Name.1)
colnames(final_df) <- c("Cell", "X", "Y", "Rich_Raster", "Richness", "Observation_Raster", "Observations")
head(final_df)
# Add new labels column for dates / months
colnames(final_df)
final_df$Label <- final_df$Rich_Raster #Add new column for label splitting
final_df <- final_df %>% separate(Label,
            c("Res", "Year", "Month")) # Seperate out labels into Res / Year / Month

# Remove prefix characters
library(stringr)
final_df <- final_df %>% transform(Year=str_replace(Year,"Y","")) # Remove Y from YEar column
final_df <- final_df %>% transform(Res=str_replace(Res,"rich","")) # Remove Rich from res columns
head(final_df)
# Check class of data
sapply(final_df, class)
format(final_df$Year, format="%Y")
format(final_df$Month, format="%M")
sapply(final_df, class)
final_df <- unite(final_df, Fulldate, c(Year, Month), remove=FALSE) # Combine Year and Date into 1 column
final_df$Fulldate <- paste(final_df$Fulldate, "_01", sep = "") # Add "day" to give full format
colnames(final_df)
final_df <- final_df[, c(1,2,3,4,5,6,7,8,10,11,9)] # Rearrange columns to logical order

library(lubridate)
final_df$Fulldate <- ymd(final_df$Fulldate) # Make date format
```

```
sapply(final_df, class)
### Save out "Final Dataframe"
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/Final_dfs")
write.table(final_df,
            file = "Final_df_Res1")


### 10 km ###
# Load Richness Files for each of 480 months
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/txt files/Richness")
files_rich <- list.files(pattern = ".txt")
list_rich <- list()
for(i in 1:480){
  list_rich [[i]] <- read.csv(files_rich[[i]], sep="")
  print(i)
}


# Load Obs Files for each of 480 months
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/10/txt files/Observations")
files_obs <- list.files(pattern = ".txt")
list_obs <- list()
for(i in 1:480){
  list_obs [[i]] <- read.csv(files_obs[[i]], sep="")
  print(i)
}
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")



# Make it "longer" - so put  480 variables into one column
list_rich_longer <- list()
for(i in 1:480){
  list_rich_longer[[i]] <- pivot_longer(list_rich[[i]], cols = 4:4)
  colnames(list_rich_longer[[i]]) <- c("Cell", "X", "Y", "Name", "Richness")
  print(i)
}
head(list_rich_longer)

list_obs_longer <- list()
for(i in 1:480){
  list_obs_longer[[i]] <- pivot_longer(list_obs[[i]], cols = 4:4)
  colnames(list_obs_longer[[i]]) <- c("Cell", "X", "Y", "Name", "Observations")
  print(i)
}
# Check what it looks like for DF with vals
list_rich[6]
list_rich_longer[6]
list_obs_longer[6]
# Combine richness list and observation list
list_all_combined <- list()
for(i in 1:480){
  list_all_combined[[i]] <- cbind(list_rich_longer[[i]], list_obs_longer[[i]])
  print(i)
}
# Combine all list into one dataframe
df <- ldply(list_all_combined, data.frame)
head(df)
# Data tidy
# Rename Columns
colnames(df)
final_df <- df %>% select(-Cell.1, -X.1, -Y.1, Name.1)
colnames(final_df) <- c("Cell", "X", "Y", "Rich_Raster", "Richness", "Observation_Raster", "Observations")
head(final_df)
# Add new labels column for dates / months
colnames(final_df)
final_df$Label <- final_df$Rich_Raster #Add new column for label splitting
final_df <- final_df %>% separate(Label,
                                  c("Res", "Year", "Month")) # Seperate out labels into Res / Year / Month
# Remove prefix characters
```

```r
library(stringr)
final_df <- final_df %>% transform(Year=str_replace(Year,"Y","")) # Remove Y from YEar column
final_df <- final_df %>% transform(Res=str_replace(Res,"rich","")) # Remove Rich from res columns
head(final_df)

# Check class of data
sapply(final_df, class)

format(final_df$Year, format="%Y")
format(final_df$Month, format="%M")
sapply(final_df, class)

final_df <- unite(final_df, Fulldate, c(Year, Month), remove=FALSE) # Combine Year and Date into 1 column
final_df$Fulldate <- paste(final_df$Fulldate, "_01", sep = "") # Add "day" to give full format
colnames(final_df)
final_df <- final_df[, c(1,2,3,4,5,6,7,8,10,11,9)] # Rearrange columns to logical order

library(lubridate)
final_df$Fulldate <- ymd(final_df$Fulldate) # Make date format
sapply(final_df, class)
# Save out "Final Dataframe"
head(final_df)
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/Final_dfs")
write.table(final_df,
      file = "Final_df_Res10")

### 100 km2 ###
# Load Richness Files for each of 480 months
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/txt files/Richness")
files_rich <- list.files(pattern = ".txt")
list_rich <- list()
for(i in 1:480){
  list_rich [[i]] <- read.csv(files_rich[[i]], sep="")
  print(i)
}

# Load Obs Files for each of 480 months
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/100/txt files/Observations")
files_obs <- list.files(pattern = ".txt")
list_obs <- list()
for(i in 1:480){
  list_obs [[i]] <- read.csv(files_obs[[i]], sep="")
  print(i)
}

setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2")


# Make it "longer" - so put  480 variables into one column
list_rich_longer <- list()
for(i in 1:480){
  list_rich_longer[[i]] <- pivot_longer(list_rich[[i]], cols = 4:4)
  colnames(list_rich_longer[[i]]) <- c("Cell", "X", "Y", "Name", "Richness")
  print(i)
}
head(list_rich_longer)

list_obs_longer <- list()
for(i in 1:480){
  list_obs_longer[[i]] <- pivot_longer(list_obs[[i]], cols = 4:4)
  colnames(list_obs_longer[[i]]) <- c("Cell", "X", "Y", "Name", "Observations")
  print(i)
}

# Check what it looks like for DF with vals
list_rich[6]
list_rich_longer[6]
list_obs_longer[6]
```

```
# Combine richness list and observation list t
list_all_combined <- list()
for(i in 1:480){
  list_all_combined[[i]] <- cbind(list_rich_longer[[i]], list_obs_longer[[i]])
  print(i)
}

# Combine all list into one dataframe
df <- ldply(list_all_combined, data.frame)
head(df)

# Data tidy
# Rename Columns
colnames(df)
final_df <- df %>% select(-Cell.1, -X.1, -Y.1, Name.1)
colnames(final_df) <- c("Cell", "X", "Y", "Rich_Raster", "Richness", "Observation_Raster", "Observations")
head(final_df)
# Add new labels column for dates / months
colnames(final_df)
final_df$Label <- final_df$Rich_Raster #Add new column for label splitting
final_df <- final_df %>% separate(Label,
                                  c("Res", "Year", "Month")) # Seperate out labels into Res / Year / Month

# Remove prefix characters
library(stringr)
final_df <- final_df %>% transform(Year=str_replace(Year,"Y","")) # Remove Y from YEar column
final_df <- final_df %>% transform(Res=str_replace(Res,"rich","")) # Remove Rich from res columns
head(final_df)

# Check class of data
sapply(final_df, class)

format(final_df$Year, format="%Y")
format(final_df$Month, format="%M")
sapply(final_df, class)

final_df <- unite(final_df, Fulldate, c(Year, Month), remove=FALSE) # Combine Year and Date into 1 column
final_df$Fulldate <- paste(final_df$Fulldate, "_01", sep = "") # Add "day" to give full format
colnames(final_df)
final_df <- final_df[, c(1,2,3,4,5,6,7,8,10,11,9)] # Rearrange columns to logical order

library(lubridate)
final_df$Fulldate <- ymd(final_df$Fulldate) # Make date format
sapply(final_df, class)
# Save out "Final Dataframe"
head(final_df)
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/Final_dfs")
write.table(final_df,
        file = "Final_df_Res100")
### All 3 resolutions to load ### -------------------------------------------------------
Final_df_Res1 <- read.csv("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/Final_dfs/Final_df_Res1", sep="", stringsAsFactors=TRUE)
Final_df_Res10 <- read.csv("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/Final_dfs/Final_df_Res10", sep="", stringsAsFactors=TRUE)
Final_df_Res100 <- read.csv("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species
Richness/Species_Richness_v2/tiffs/Month_Year_Sub/Final_dfs/Final_df_Res100", sep="", stringsAsFactors=TRUE)
colnames(Final_df_Res100)

### End of Script ###-----------------------------------------------------------------------------
```

Script 11: Seasonal Decomposition of Time Series (STL) analysis

```
library(mgcv)
library(tidyr)
library(lubridate)  # if needed to change time formats
library(zoo)
library(forecast)
library(ggplot2)
library(scales)
library(ggpubr)
library(cowplot)
library(gridExtra)
library(tibble)
library(stlplus)
library(reshape2)
getwd()

### STL for TREND - for both geometric mean raw richness and SPUE, for all resolutions
# Data from df in script 10
# Data of mean, median etc values
load("Final_Vals_100.Rda")
load("Final_Vals_10.Rda")
load("Final_Vals_1.Rda")

par(mfrow=c(1,1))
options(scipen = 999)

# For FINAL Dataset - All
Final_Vals_100 <- as.data.frame(Final_Vals_100)
Final_Vals_100 <- rownames_to_column(Final_Vals_100, var = "Date")
as_date(Final_Vals_100$Date)
Final_Vals_10 <- as.data.frame(Final_Vals_10)
Final_Vals_10 <- rownames_to_column(Final_Vals_10, var = "Date")
as_date(Final_Vals_10$Date)
Final_Vals_1 <- as.data.frame(Final_Vals_1)
Final_Vals_1 <- rownames_to_column(Final_Vals_1, var = "Date")
as_date(Final_Vals_1$Date)

str(Final_Vals_100) # Check formats of columns
Final_Vals_100$Date <- (ymd(Final_Vals_100$Date))
str(Final_Vals_100) # Col now date format

str(Final_Vals_10) # Check formats of columns
Final_Vals_10$Date <- (ymd(Final_Vals_10$Date))
str(Final_Vals_10) # Col now date format

str(Final_Vals_1) # Check formats of columns
Final_Vals_1$Date <- (ymd(Final_Vals_1$Date))
str(Final_Vals_1) # Col now date format

### 1) STL of geometric richness for species richness ---------------------------------------------------------------------------

### RAW RICHNESS ### ----------
# fig <- "Figs/Analysis/stl/stl_100_GeomeansRich.jpeg"
#jpeg(fig, width=1000, height=600)   # open the file
plot(stl(ts(Final_Vals_100$Res100_GeoMeans, start=c(1979,6), frequency = 12), "periodic"))
# dev.off()

geomeanrichness_100 <- stlplus(Final_Vals_100$Res100_GeoMeans, t = Final_Vals_100$Date, n.p = 12, s.window = "periodic")
geomeanrichness_10 <- stlplus(Final_Vals_10$Res10_GeoMeans, t = Final_Vals_10$Date, n.p = 12, s.window = "periodic")
geomeanrichness_1 <- stlplus(Final_Vals_1$Res1_GeoMeans, t = Final_Vals_1$Date, n.p = 12, s.window = "periodic")

# 3 x 1 plot
fig <- "Figs/Analysis/stl/Figures for Chapter/3GeomeansRichness_Trend.jpeg"
jpeg(fig, width=1500, height=1000)   # open the file
par(mfrow = c(3,1), oma = c(3,5,2,2), mar = c(2,1,1,1)) # bottom, left, top, and right
plot(trend(geomeanrichness_100), x = geomeanrichness_100$time, type = "l", xlab = "", ylab = "", ylim = c(0,6), xaxs = "r", yaxs = "i", cex.axis = 1.5, cex.lab = 1.5)
plot(trend(geomeanrichness_10), x = geomeanrichness_10$time, type = "l", xlab = "", ylab = "", cex.lab = 1.5, ylim = c(0,6), xaxs = "r", yaxs = "i", cex.axis = 1.5)
```

```
plot(trend(geomeanrichness_1), x = geomeanrichness_1$time, type = "l", xlab = "", cex.lab = 1.5, ylab = "", ylim = c(0,6), xaxs = "r", yaxs = "i", )
mtext(text="Geometric mean Species Richness",side=2,line=2.5,outer=TRUE, cex = 1.5)
mtext(text="Time",side=1, line=1, adj = 0.53, outer=TRUE, cex = 1.2)
dev.off()
# All on one plot
fig <- "Figs/Analysis/stl/Figures for Chapter/1GeomeansRichness_Trend.jpeg"
jpeg(fig, width=1500, height=1000)   # open the file
par(mfrow = c(1,1), oma = c(3,5,2,2), mar = c(2,1,1,1)) # bottom, left, top, and right
plot(trend(geomeanrichness_100), x = geomeanrichness_100$time, type = "l", xlab = "", ylab = "", ylim = c(0,6), xaxs = "r", yaxs = "i", cex.axis = 1.5, cex.lab = 1.5, col = "black")
lines(trend(geomeanrichness_10), x = geomeanrichness_10$time, type = "l", xlab = "", ylab = "", cex.lab = 1.5, ylim = c(0,6), xaxs = "r", yaxs = "i", cex.axis = 1.5, col = "red")
lines(trend(geomeanrichness_1), x = geomeanrichness_1$time, type = "l", xlab = "", cex.lab = 1.5, ylab = "", ylim = c(0,6), xaxs = "r", yaxs = "i", col = "blue")
mtext(text="Geometric mean Species Richness",side=2,line=2.5,outer=TRUE, cex = 1.5)
mtext(text="Time",side=1, line=1, adj = 0.53, outer=TRUE, cex = 1.5)
dev.off()

### SRPUE RICHNESS ### ------------------------------------------------------------------------------------------------------------------
fig <- "Figs/Analysis/stl/stl_100_GeomeansRich.jpeg"
jpeg(fig, width=1000, height=600)   # open the file
plot(stlplus(Final_Vals_100$Res100_GeoMeanMeasure, t = Final_Vals_100$Date, n.p = 12, s.window = "periodic"))
dev.off()
geomeanSPUE_100 <- stlplus(Final_Vals_100$Res100_GeoMeanMeasure, t = Final_Vals_100$Date, n.p = 12, s.window = "periodic")
geomeanSPUE_10 <- stlplus(Final_Vals_10$Res10_GeoMeanMeasure, t = Final_Vals_10$Date, n.p = 12, s.window = "periodic")
geomeanSPUE_1 <- stlplus(Final_Vals_1$Res1_GeoMeanMeasure, t = Final_Vals_1$Date, n.p = 12, s.window = "periodic")
# 3 x 1 plot
fig <- "Figs/Analysis/stl/Figures for Chapter/3GeomeanSPUE_Trend.jpeg"
jpeg(fig, width=1500, height=1000)   # open the file
par(mfrow = c(3,1), oma = c(3,5,2,2), mar = c(2,1,1,1)) # bottom, left, top, and right
plot(trend(geomeanSPUE_100), x = geomeanSPUE_100$time, type = "l", xlab = "", ylab = "", ylim = c(1,2.2), xaxs = "r", yaxs = "i", cex.axis = 1.5, cex.lab = 1.5)
plot(trend(geomeanSPUE_10), x = geomeanSPUE_10$time, type = "l", xlab = "", ylab = "", cex.lab = 1.5, ylim = c(1,2.2), xaxs = "r", yaxs = "i", cex.axis = 1.5)
plot(trend(geomeanSPUE_1), x = geomeanSPUE_1$time, type = "l", xlab = "", cex.lab = 1.5,cex.axis = 1.5, ylab = "", ylim = c(1,2.2), xaxs = "r", yaxs = "i")
mtext(text="Geometric mean SPUE",side=2,line=2.5,outer=TRUE, cex = 1.5)
mtext(text="Time",side=1, line=1, adj = 0.53, outer=TRUE, cex = 1.2)
dev.off()
# All on one plot
fig <- "Figs/Analysis/stl/Figures for Chapter/1GeomeansSPUE_Trend.jpeg"
jpeg(fig, width=1500, height=1000)   # open the file
par(mfrow = c(1,1), oma = c(3,5,2,2), mar = c(2,1,1,1)) # bottom, left, top, and right
plot(trend(geomeanSPUE_100), x = geomeanSPUE_100$time, type = "l", xlab = "", ylab = "", ylim = c(0,3), xaxs = "r", yaxs = "i", cex.axis = 1.5, col = "black")
lines(trend(geomeanSPUE_10), x = geomeanSPUE_10$time, type = "l", xlab = "", ylab = "", cex.lab = 1.5, ylim = c(0,3), xaxs = "r", yaxs = "i", cex.axis = 1.5, col = "red")
lines(trend(geomeanSPUE_1), x = geomeanSPUE_1$time, type = "l", xlab = "", cex.lab = 1.5, ylab = "", ylim = c(0,3), xaxs = "r", yaxs = "i", col = "blue")
mtext(text="Geometric mean Species Richness",side=2,line=2.5,outer=TRUE, cex = 1.5)
mtext(text="Time",side=1, line=1, adj = 0.53, outer=TRUE, cex = 1.5)
dev.off()
# All on one plot
fig <- "Figs/Analysis/stl/Figures for Chapter/BOTH_TREND_v1.jpeg"
jpeg(fig, width=1500, height=1000)   # open the file
par(mfrow = c(2,1), oma = c(3,5,2,2), mar = c(2,1,1,1)) # bottom, left, top, and right
plot(trend(geomeanrichness_100), x = geomeanrichness_100$time, type = "l", xlab = "", ylab = "", ylim = c(0,6), xaxs = "r", yaxs = "i", cex.axis = 2, cex.lab = 2, col = "black")
lines(trend(geomeanrichness_10), x = geomeanrichness_10$time, type = "l", xlab = "", ylab = "", cex.lab = 2, ylim = c(0,6), xaxs = "r", yaxs = "i", col = "red")
lines(trend(geomeanrichness_1), x = geomeanrichness_1$time, type = "l", xlab = "", cex.lab = 2, ylab = "", ylim = c(0,6), xaxs = "r", yaxs = "i", col = "blue")
mtext(text="Geometric mean Species Richness",side=2,line=2.5,outer=TRUE, cex = 2.5)
legend("top", inset = 0.02, horiz=TRUE,legend=c(as.expression(bquote("100" ~ km^2)),
                                    as.expression(bquote("10" ~ km^2)),
                                    as.expression(bquote("1" ~ km^2))),
      col=c("black", "red", "blue"), lty = 1, cex=2)

plot(trend(geomeanSPUE_100), x = geomeanSPUE_100$time, type = "l", xlab = "", ylab = "", ylim = c(0,3), xaxs = "r", yaxs = "i", cex.axis = 2, col = "black")
```

```
lines(trend(geomeanSPUE_10), x = geomeanSPUE_10$time, type = "l", xlab = "", ylab = "", cex.lab = 2, ylim = c(0,3), xaxs = "r", yaxs = "i", cex.axis = 2, col = "red")
lines(trend(geomeanSPUE_1), x = geomeanSPUE_1$time, type = "l", xlab = "", cex.lab = 2, ylab = "", ylim = c(0,3), xaxs = "r", yaxs = "i", col = "blue")
mtext(text="Time",side=1, line=1.5, adj = 0.53, outer=TRUE, cex = 2.5)
legend("top", inset = 0.02, horiz=TRUE,legend=c(as.expression(bquote("100" ~ km^2)),
                                as.expression(bquote("10" ~ km^2)),
                                as.expression(bquote("1" ~ km^2))),
      col=c("black", "red", "blue"), lty = 1, cex=2)
dev.off()


## Install required packages for this Script
library(YesSiR)
library("ggpubr")
library(flextable)
library(stlplus)

### STL of geometric richness for species richness - cutting at 2010 - not doing, just trialling -----------------------------------------
# raw SP richness
plot(stl(ts(Final_Vals_100$Res100_GeoMeans, start=c(1979,6), end =c(2010,12), frequency = 12), "periodic"))
geomeanrichness_100 <- stlplus(Final_Vals_100$Res100_GeoMeans, t = Final_Vals_100$Date, n.p = 12, s.window = "periodic")
geomeanrichness_10 <- stlplus(Final_Vals_10$Res10_GeoMeans, t = Final_Vals_10$Date, n.p = 12, s.window = "periodic")
geomeanrichness_1 <- stlplus(Final_Vals_1$Res1_GeoMeans, t = Final_Vals_1$Date, n.p = 12, s.window = "periodic")

# SPRUE
plot(stlplus(Final_Vals_100$Res100_GeoMeanMeasure, t = Final_Vals_100$Date, start = c(1979,6), end = c(2010,12), n.p = 12, s.window = "periodic"))
geomeanSPUE_100 <- stlplus(Final_Vals_100$Res100_GeoMeanMeasure, t = Final_Vals_100$Date, n.p = 12, s.window = "periodic")
geomeanSPUE_10 <- stlplus(Final_Vals_10$Res10_GeoMeanMeasure, t = Final_Vals_10$Date, n.p = 12, s.window = "periodic")
geomeanSPUE_1 <- stlplus(Final_Vals_1$Res1_GeoMeanMeasure, t = Final_Vals_1$Date, n.p = 12, s.window = "periodic")
max(geomeanrichness_100$time)
# All on one plot
# fig <- "Figs/Analysis/stl/Figures for Chapter/Corrections_STL_SR_2010.jpeg"
# jpeg(fig, width=1500, height=1000)   # open the file
par(mfrow = c(1,1), oma = c(3,5,2,2), mar = c(2,1,1,1)) # bottom, left, top, and right
plot(trend(geomeanrichness_100), x = geomeanrichness_100$time, type = "l", xlab = "", ylab = "", ylim = c(0,6), xaxs = "i",  xlim =c(as.Date("1979-01-01"),as.Date("2011-01-01")), yaxs = "i", cex.axis = 1.5, cex.lab = 1.5, col = "black")
lines(trend(geomeanrichness_10), x = geomeanrichness_10$time, type = "l", xlab = "", ylab = "", cex.lab = 1.5, ylim = c(0,6), xaxs = "r", yaxs = "i", cex.axis = 1.5, col = "red")
lines(trend(geomeanrichness_1), x = geomeanrichness_1$time, type = "l", xlab = "", cex.lab = 1.5, ylab = "", ylim = c(0,6), xaxs = "r", yaxs = "i", col = "blue")
mtext(text="Geometric mean Species Richness",side=2,line=2.5,outer=TRUE, cex = 1.5)
mtext(text="Time",side=1, line=1, adj = 0.53, outer=TRUE, cex = 1.5)
dev.off()


# All on one plot for both SR and SPRUE
fig <- "Figs/Analysis/stl/Figures for Chapter/Corrections_BOTH_TREND_v1.jpeg"
jpeg(fig, width=1500, height=1000)   # open the file
par(mfrow = c(2,1), oma = c(3,5,2,2), mar = c(2,1,1,1)) # bottom, left, top, and right
plot(trend(geomeanrichness_100), x = geomeanrichness_100$time, type = "l", xlab = "", ylab = "", ylim = c(0,6), xaxs = "r", xlim =c(as.Date("1979-01-01"),as.Date("2018-09-01")), yaxs = "i", cex.axis = 2, cex.lab = 2, col = "black")
lines(trend(geomeanrichness_10), x = geomeanrichness_10$time, type = "l", xlab = "", ylab = "", cex.lab = 2, ylim = c(0,6), xaxs = "r", yaxs = "i", col = "red")
lines(trend(geomeanrichness_1), x = geomeanrichness_1$time, type = "l", xlab = "", cex.lab = 2, ylab = "", ylim = c(0,6), xaxs = "r", yaxs = "i", col = "blue")
mtext(text="Geometric mean",side=2,line=2.5,outer=TRUE, cex = 2.5)
legend("top", inset = 0.02, horiz=TRUE,legend=c(as.expression(bquote("100" ~ km^2)),
                                as.expression(bquote("10" ~ km^2)),
                                as.expression(bquote("1" ~ km^2))),
      col=c("black", "red", "blue"), lty = 1, cex=2)
plot(trend(geomeanSPUE_100), x = geomeanSPUE_100$time, type = "l", xlab = "", ylab = "", ylim = c(0,3), xaxs = "r", xlim =c(as.Date("1979-01-01"),as.Date("2018-09-01")), yaxs = "i", cex.axis = 2, col = "black")
lines(trend(geomeanSPUE_10), x = geomeanSPUE_10$time, type = "l", xlab = "", ylab = "", cex.lab = 2, ylim = c(0,3), xaxs = "r", yaxs = "i", cex.axis = 2, col = "red")
lines(trend(geomeanSPUE_1), x = geomeanSPUE_1$time, type = "l", xlab = "", cex.lab = 2, ylab = "", ylim = c(0,3), xaxs = "r", yaxs = "i", col = "blue")
mtext(text="Time",side=1, line=1.5, adj = 0.53, outer=TRUE, cex = 2.5)
legend("top", inset = 0.02, horiz=TRUE,legend=c(as.expression(bquote("100" ~ km^2)),
                                as.expression(bquote("10" ~ km^2)),
```

```
                                        as.expression(bquote("1" ~ km^2))),
       col=c("black", "red", "blue"), lty = 1, cex=2)
dev.off()

### Boxplots removing data past 2012 ###-------------------------------------------------------------------
### a) Boxplot of geometric richness for species richness RAW -------------------------------------------
# Cut dataset down to <2010
BoxplotData_Pre2012_100 <- subset(Final_Vals_100, Date < "2012-01-01")
BoxplotData_Pre2012_10 <- subset(Final_Vals_10, Date < "2012-01-01")
BoxplotData_Pre2012_1 <- subset(Final_Vals_1, Date < "2012-01-01")

head(BoxplotData_Pre2012_100)
head(BoxplotData_Pre2012_10)
head(BoxplotData_Pre2012_1)

BoxplotData_SR_All <- cbind(BoxplotData_Pre2012_100$Date, BoxplotData_Pre2012_100$Res100_GeoMeans,
BoxplotData_Pre2012_10$Res10_GeoMeans, BoxplotData_Pre2012_1$Res1_GeoMeans)
head(BoxplotData_SR_All)
BoxplotData_SR_All <- as.data.frame(BoxplotData_SR_All)
names(BoxplotData_SR_All)[1] <- "Date"
names(BoxplotData_SR_All)[2] <- "100 km"
names(BoxplotData_SR_All)[3] <- "10 km"
names(BoxplotData_SR_All)[4] <- "1 km"
BoxplotData_SR_All <- gather(BoxplotData_SR_All[2:4])

Corrections_bp_geomeans_rawrichness <- ggplot(data = BoxplotData_SR_All, aes(x = key, y = value)) +
  geom_boxplot(
    colour = "black",
    notch = TRUE,
    outlier.color = "black",
    outlier.shape = NA,
    outlier.alpha = 0.1,) +
  geom_jitter(width = 0.2,
              shape = 16,
              alpha = 0.4,
              aes(colour = key)) +
  scale_color_manual(values = c('red', 'blue', 'black')) + # for scatter plot dots
  facet_wrap( ~ key, scales = "free") +
  theme_linedraw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        strip.background = element_rect(fill = "white"),
        strip.text = element_text(colour = "black", size = rel(2)),
        axis.title = element_text(colour = "black", size = rel(2)),
        axis.text.x = element_blank(),
        axis.text = element_text(size = rel(2))) +
  ylim(0, 8)+
  labs(x = "", y = "")
Corrections_bp_geomeans_rawrichness

### b) Boxplot of geometric richness for Measure ---------------------------------------------------------------
head(BoxplotData_Pre2012_100)
BoxplotData_SPRUE_All <- cbind(BoxplotData_Pre2012_100$Date, BoxplotData_Pre2012_100$Res100_GeoMeanMeasure,
BoxplotData_Pre2012_10$Res10_GeoMeanMeasure, BoxplotData_Pre2012_1$Res1_GeoMeanMeasure)
head(BoxplotData_SPRUE_All)
BoxplotData_SPRUE_All <- as.data.frame(BoxplotData_SPRUE_All)
names(BoxplotData_SPRUE_All)[1] <- "Date"
names(BoxplotData_SPRUE_All)[2] <- "100 km"
names(BoxplotData_SPRUE_All)[3] <- "10 km"
names(BoxplotData_SPRUE_All)[4] <- "1 km"
BoxplotData_SPRUE_All <- gather(BoxplotData_SPRUE_All[2:4])
head(BoxplotData_SPRUE_All)
# test <- na.omit(data_geomeanSPUE$value) # for getting boxplot limits
# min(test)
Corrections_bp_geomeans_SPRUE <- ggplot(data = BoxplotData_SPRUE_All, aes(x = key, y = value)) +
  geom_boxplot(
    colour = "black",
    notch = TRUE,
    outlier.color = "black",
    outlier.shape = NA,
    outlier.alpha = 0.1) +
  geom_jitter(width = 0.2,
              shape = 16,
              alpha = 0.4,
              aes(colour = key)) +
```

```
  scale_color_manual(values = c('red', 'blue', 'black')) + # for scatter plot dots
  facet_wrap( ~ key, scales = "free") +
  theme_linedraw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        strip.background = element_rect(fill = "white"),
        strip.text = element_text(colour = "black", size = rel(2)),
        axis.title = element_text(colour = "black", size = rel(2)),
        axis.text.x = element_blank(),
        axis.text.y = element_text(size = rel(2.2))) +
  ylim(0, 3)+
  labs(x = "", y = "")
Corrections_bp_geomeans_SPRUE

# c) Final Figure  #####--------------------------------------------------------------------------------------------------------------------------
Corrections_boxplot_all_geomeans <- ggarrange(nrow = 2,
                                  ncol = 1,
                                  Corrections_bp_geomeans_rawrichness,
                                  Corrections_bp_geomeans_SPRUE,
                                  legend = FALSE)+
  theme(plot.margin = margin(1, 1, 0, 1, "cm"))
Corrections_boxplot_all_geomeans <- annotate_figure(Corrections_boxplot_all_geomeans, bottom = text_grob("Grid size
resolution", color = "black", rot = 0, hjust = 0.2, vjust = 0, size = 20))
Corrections_boxplot_all_geomeans<- annotate_figure(Corrections_boxplot_all_geomeans, left = text_grob("Geometric mean of
Species Richness", color = "black", rot = 90, hjust = -0.2, vjust = 2, size = 20))
Corrections_boxplot_all_geomeans <- annotate_figure(Corrections_boxplot_all_geomeans, left = text_grob("Geometric mean of
SRPUE", color = "black", rot = 90, hjust = 1.6, vjust = 4, size = 20))
Corrections_boxplot_all_geomeans


fig <- "Figs/Analysis/stl/Figures for Chapter/00_Corrections/CompareGeomeanRichness_v1.jpeg"
jpeg(fig, width=1200, height=1000)
Corrections_boxplot_all_geomeans
dev.off()


### 2) T test to compare if actually different when taking out post 2010 ###------------------------------------------------------------
shapiro.test(Final_Vals_100$Res100_GeoMeans) # not normal
shapiro.test(Final_Vals_10$Res10_GeoMeans) # not normal
shapiro.test(Final_Vals_1$Res1_GeoMeans) # not normal
shapiro.test(Final_Vals_100$Res100_GeoMeanMeasure) # not normal
shapiro.test(Final_Vals_10$Res10_GeoMeanMeasure) # not normal
shapiro.test(Final_Vals_1$Res1_GeoMeanMeasure) # not normal

library("ggpubr")
ggdensity(Final_Vals_100$Res100_GeoMeans)
ggdensity(Final_Vals_100$Res100_GeoMeanMeasure)
ggdensity(Final_Vals_10$Res10_GeoMeans)
ggdensity(Final_Vals_10$Res10_GeoMeanMeasure)
ggdensity(Final_Vals_1$Res1_GeoMeans)
ggdensity(Final_Vals_1$Res1_GeoMeanMeasure)

# T test SR Mean against SR <2012
TT_SR_100 <- t.test(Final_Vals_100$Res100_GeoMeans, BoxplotData_Pre2012_100$Res100_GeoMeans, alternative =
"two.sided", var.equal = FALSE, paired = FALSE)
TT_SR_10 <- t.test(Final_Vals_10$Res10_GeoMeans, BoxplotData_Pre2012_10$Res10_GeoMeans, alternative = "two.sided",
var.equal = FALSE, paired = FALSE)
TT_SR_1 <- t.test(Final_Vals_1$Res1_GeoMeans, BoxplotData_Pre2012_1$Res1_GeoMeans, alternative = "two.sided",
var.equal = FALSE, paired = FALSE)

# T test SPRUE Mean against SR <2012
TT_SPRUE_100 <- t.test(Final_Vals_100$Res100_GeoMeanMeasure, BoxplotData_Pre2012_100$Res100_GeoMeanMeasure,
alternative = "two.sided", var.equal = FALSE, paired = FALSE)
TT_SPRUE_10 <- t.test(Final_Vals_10$Res10_GeoMeanMeasure, BoxplotData_Pre2012_10$Res10_GeoMeanMeasure,
alternative = "two.sided", var.equal = FALSE, paired = FALSE)
TT_SPRUE_1 <- t.test(Final_Vals_1$Res1_GeoMeanMeasure, BoxplotData_Pre2012_1$Res1_GeoMeanMeasure, alternative =
"two.sided", var.equal = FALSE, paired = FALSE)

flextable::as_flextable(TT_SR_100)
flextable::as_flextable(TT_SR_10)
flextable::as_flextable(TT_SR_1)
exportxlsx(flextable::as_flextable(TT_SR_100), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/TT_SR_100.xlsx")
```

```
exportxlsx(flextable::as_flextable(TT_SR_100), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/TT_SR_10.xlsx")
exportxlsx(flextable::as_flextable(TT_SR_100), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/TT_SR_1.xlsx")
flextable::as_flextable(TT_SPRUE_100)
flextable::as_flextable(TT_SPRUE_10)
flextable::as_flextable(TT_SPRUE_1)
exportxlsx(flextable::as_flextable(TT_SPRUE_100), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/TT_SRPUE_100.xlsx")
exportxlsx(flextable::as_flextable(TT_SPRUE_10), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/TT_SRPUE_10.xlsx")
exportxlsx(flextable::as_flextable(TT_SPRUE_1), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/TT_SRPUE_1.xlsx")


library(flextable)
flextable::as_flextable(TT_SR_100)
install.packages("remotes")
remotes::install_github("Sebastien-Le/YesSiR")
library(YesSiR) # to export a flextable into MS Excel: exportxlsx() function
exportxlsx(flextable::as_flextable(TT_SR_100), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/TT_SR_100.xlsx")
getwd()

# WT (non para) Ntest SR Mean against SR <2012-------
WT_SR_100 <- wilcox.test(Final_Vals_100$Res100_GeoMeans, BoxplotData_Pre2012_100$Res100_GeoMeans, paired =
FALSE)
WT_SR_10 <- wilcox.test(Final_Vals_10$Res10_GeoMeans, BoxplotData_Pre2012_10$Res10_GeoMeans)
WT_SR_1 <- wilcox.test(Final_Vals_1$Res1_GeoMeans, BoxplotData_Pre2012_1$Res1_GeoMeans)
WT_SPRUE_100 <- wilcox.test(Final_Vals_100$Res100_GeoMeanMeasure,
BoxplotData_Pre2012_100$Res100_GeoMeanMeasure)
WT_SPRUE_10 <- wilcox.test(Final_Vals_10$Res10_GeoMeanMeasure,
BoxplotData_Pre2012_10$Res10_GeoMeanMeasure)
WT_SPRUE_1 <- wilcox.test(Final_Vals_1$Res1_GeoMeanMeasure, BoxplotData_Pre2012_1$Res1_GeoMeanMeasure)

flextable::as_flextable(WT_SR_100)
flextable::as_flextable(WT_SR_10)
flextable::as_flextable(WT_SR_1)
flextable::as_flextable(WT_SPRUE_100)
flextable::as_flextable(WT_SPRUE_10)
flextable::as_flextable(WT_SPRUE_1)

exportxlsx(flextable::as_flextable(WT_SR_100), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/WT_SR_100.xlsx")
exportxlsx(flextable::as_flextable(WT_SR_10), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/WT_SR_10.xlsx")
exportxlsx(flextable::as_flextable(WT_SR_1), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1.
Species Richness/Species_Richness_v2/Tables/WT_SR_1.xlsx")

### a) Combined boxplot ###--------------------------------
boxplot_all_geomeans
Corrections_boxplot_all_geomeans

Fig4.6 <- ggarrange(nrow = 1,
          ncol = 2,
          boxplot_all_geomeans,
          Corrections_boxplot_all_geomeans,
          legend = FALSE)+
  theme(plot.margin = margin(1, 1, 0, 1, "cm"))
Fig4.6 <- annotate_figure(Fig4.6, bottom = text_grob("Grid size resolution", color = "black", rot = 0, hjust = 0.2, vjust = 0, size
= 20))
Fig4.6<- annotate_figure(Fig4.6, left = text_grob("Geometric mean of Species Richness", color = "black", rot = 90, hjust = -0.1,
vjust = 4, size = 20))
Fig4.6 <- annotate_figure(Fig4.6, left = text_grob("Geometric mean of SRPUE", color = "black", rot = 90, hjust = 1.6, vjust = 6,
size = 20))
Fig4.6 <- annotate_figure(Fig4.6, top = text_grob("All data", color = "black", rot = 0, hjust = 5, vjust = 4, size = 20))
Fig4.6 <- annotate_figure(Fig4.6, top = text_grob("Post 2012 removed", color = "black", rot = 0, hjust = -2, vjust = 5.5, size =
20))

Fig4.6

fig <- "Figs/Analysis/stl/Figures for Chapter/00_Corrections/Correction_Fig4.6.jpeg"
jpeg(fig, width=1500, height=1000)
Fig4.6
dev.off()
```

```
### 3) STL boxplots without post 2012 data ###------------------------------------------------------------------------------------------------
# a) Analysis preparing for STL boxplots#-------------------------------------------------------------------------------------------------
# Data to use
head(BoxplotData_Pre2012_100)
head(BoxplotData_Pre2012_10)
head(BoxplotData_Pre2012_1)

# Setup stl data for 100km #------------------------------------------------------------------------------------------------------------
Pre2012_stl_geomeans_100 <- stlplus(BoxplotData_Pre2012_100$Res100_GeoMeans, t = BoxplotData_Pre2012_100$Date,
n.p = 12, s.window = "periodic")
# plot(seasonal(Pre2012_stl_geomeans_100), x = Pre2012_stl_geomeans_100$time, type = "l", xlab = "", ylab = "", ylim =
NULL, xaxs = "r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_artmeans_100 <- stlplus(BoxplotData_Pre2012_100$Res100_means, t = BoxplotData_Pre2012_100$Date, n.p =
12, s.window = "periodic")
plot(seasonal(Pre2012_stl_artmeans_100), x = Pre2012_stl_artmeans_100$time, type = "l", xlab = "", ylab = "", ylim = NULL,
xaxs = "r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_median_100 <- stlplus(BoxplotData_Pre2012_100$Res100_median, t = BoxplotData_Pre2012_100$Date, n.p = 12,
s.window = "periodic")
plot(seasonal(Pre2012_stl_median_100), x = Pre2012_stl_median_100$time, type = "l", xlab = "", ylab = "", ylim = NULL, xaxs
= "r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_SPUEmean_100 <- stlplus(BoxplotData_Pre2012_100$Res100_meanMeasure_Remove1s, t =
BoxplotData_Pre2012_100$Date, n.p = 12, s.window = "periodic")
Pre2012_stl_SPUEgeomean_100 <- stlplus(BoxplotData_Pre2012_100$Res100_GeoMeanMeasure, t =
BoxplotData_Pre2012_100$Date, n.p = 12, s.window = "periodic")

Pre2012_sum_geomeans_100 <- summary(Pre2012_stl_geomeans_100$data$seasonal)
Pre2012_sum_artmeans_100 <- summary(Pre2012_stl_artmeans_100$data$seasonal)
Pre2012_sum_medians_100 <- summary(Pre2012_stl_median_100$data$seasonal)
Pre2012_sum_SPUEmeans_100 <- summary(Pre2012_stl_SPUEmean_100$data$seasonal)
Pre2012_sum_SPUEgeomeans_100 <- summary(Pre2012_stl_SPUEgeomean_100$data$seasonal)
Pre2012_seasonal_all_100 <- cbind(Pre2012_sum_artmeans_100,Pre2012_sum_medians_100, Pre2012_sum_geomeans_100,
Pre2012_sum_SPUEmeans_100, Pre2012_sum_SPUEgeomeans_100)
colnames(Pre2012_seasonal_all_100) <- c("Mean Richness", "Median Richness", "Geo-Mean Richness", "Mean SPRUE", "Geo-
Mean SPRUE")
Pre2012_seasonal_all_100
boxplot(Pre2012_seasonal_all_100)

# Setup stl data for 10km #-------------------------------------------------------------------------------------------------------------
Pre2012_stl_geomeans_10 <- stlplus(BoxplotData_Pre2012_10$Res10_GeoMeans, t = BoxplotData_Pre2012_10$Date, n.p =
12, s.window = "periodic")
# plot(seasonal(Pre2012_stl_geomeans_10), x = Pre2012_stl_geomeans_10$time, type = "l", xlab = "", ylab = "", ylim = NULL,
xaxs = "r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_artmeans_10 <- stlplus(BoxplotData_Pre2012_10$Res10_means, t = BoxplotData_Pre2012_10$Date, n.p = 12,
s.window = "periodic")
# plot(seasonal(Pre2012_stl_artmeans_10), x = Pre2012_stl_artmeans_10$time, type = "l", xlab = "", ylab = "", ylim = NULL,
xaxs = "r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_median_10 <- stlplus(BoxplotData_Pre2012_10$Res10_median, t = BoxplotData_Pre2012_10$Date, n.p = 12,
s.window = "periodic")
# plot(seasonal(Pre2012_stl_median_10), x = Pre2012_stl_median_10$time, type = "l", xlab = "", ylab = "", ylim = NULL, xaxs
= "r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_SPUEmean_10 <- stlplus(BoxplotData_Pre2012_10$Res10_meanMeasure_Remove1s, t =
BoxplotData_Pre2012_10$Date, n.p = 12, s.window = "periodic")
Pre2012_stl_SPUEgeomean_10 <- stlplus(BoxplotData_Pre2012_10$Res10_GeoMeanMeasure, t =
BoxplotData_Pre2012_10$Date, n.p = 12, s.window = "periodic")

Pre2012_sum_geomeans_10 <- summary(Pre2012_stl_geomeans_10$data$seasonal)
Pre2012_sum_artmeans_10 <- summary(Pre2012_stl_artmeans_10$data$seasonal)
Pre2012_sum_medians_10 <- summary(Pre2012_stl_median_10$data$seasonal)
Pre2012_sum_SPUEmeans_10 <- summary(Pre2012_stl_SPUEmean_10$data$seasonal)
Pre2012_sum_SPUEgeomeans_10 <- summary(Pre2012_stl_SPUEgeomean_10$data$seasonal)
Pre2012_seasonal_all_10 <- cbind(Pre2012_sum_artmeans_10,Pre2012_sum_medians_10, Pre2012_sum_geomeans_10,
Pre2012_sum_SPUEmeans_10, Pre2012_sum_SPUEgeomeans_10)
colnames(Pre2012_seasonal_all_10) <- c("Mean Richness", "Median Richness", "Geo-Mean Richness", "Mean SPRUE", "Geo-
Mean SPRUE")
Pre2012_seasonal_all_10
boxplot(Pre2012_seasonal_all_10)

# setup stl data for 1km #--------------------------------------------------------------------------------------------------------------
Pre2012_stl_geomeans_1 <- stlplus(BoxplotData_Pre2012_1$Res1_GeoMeans, t = BoxplotData_Pre2012_1$Date, n.p = 12,
s.window = "periodic")
```

```
# plot(seasonal(Pre2012_stl_geomeans_1), x = Pre2012_stl_geomeans_1$time, type = "l", xlab = "", ylab = "", ylim = NULL,
xaxs = "r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_artmeans_1 <- stlplus(BoxplotData_Pre2012_1$Res1_means, t = BoxplotData_Pre2012_1$Date, n.p = 12,
s.window = "periodic")
# plot(seasonal(Pre2012_stl_artmeans_1), x = Pre2012_stl_artmeans_1$time, type = "l", xlab = "", ylab = "", ylim = NULL, xaxs
= "r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_median_1 <- stlplus(BoxplotData_Pre2012_1$Res1_median, t = BoxplotData_Pre2012_1$Date, n.p = 12, s.window
= "periodic")
# plot(seasonal(Pre2012_stl_median_1), x = Pre2012_stl_median_1$time, type = "l", xlab = "", ylab = "", ylim = NULL, xaxs =
"r", yaxs = "r", cex.axis = 1, cex.lab = 1)
Pre2012_stl_SPUEmean_1 <- stlplus(BoxplotData_Pre2012_1$Res1_meanMeasure_Remove1s, t =
BoxplotData_Pre2012_1$Date, n.p = 12, s.window = "periodic")
Pre2012_stl_SPUEgeomean_1 <- stlplus(BoxplotData_Pre2012_1$Res1_GeoMeanMeasure, t = BoxplotData_Pre2012_1$Date,
n.p = 12, s.window = "periodic")

Pre2012_sum_geomeans_1 <- summary(Pre2012_stl_geomeans_1$data$seasonal)
Pre2012_sum_artmeans_1 <- summary(Pre2012_stl_artmeans_1$data$seasonal)
Pre2012_sum_medians_1 <- summary(Pre2012_stl_median_1$data$seasonal)
Pre2012_sum_SPUEmeans_1 <- summary(Pre2012_stl_SPUEmean_1$data$seasonal)
Pre2012_sum_SPUEgeomeans_1 <- summary(Pre2012_stl_SPUEgeomean_1$data$seasonal)
Pre2012_seasonal_all_1 <- cbind(Pre2012_sum_artmeans_1,Pre2012_sum_medians_1, Pre2012_sum_geomeans_1,
Pre2012_sum_SPUEmeans_1, Pre2012_sum_SPUEgeomeans_1)
colnames(Pre2012_seasonal_all_1) <- c("Mean Richness", "Median Richness", "Geo-Mean Richness", "Mean SPRUE", "Geo-
Mean SPRUE")
Pre2012_seasonal_all_1
boxplot(Pre2012_seasonal_all_1)


# b) Get trend out from stl #-----------------------------------------------------------------------------------------------------------------

a <- summary(Pre2012_stl_geomeans_100$data$trend)
b <- summary(Pre2012_stl_artmeans_100$data$trend)
c <- summary(Pre2012_stl_median_100$data$trend)
d <- summary(Pre2012_stl_SPUEmean_100$data$trend)
e <- summary(Pre2012_stl_SPUEgeomean_100$data$trend)
Pre2012_trend_all_100 <- cbind(b,c,a, d, e)

a <- summary(Pre2012_stl_geomeans_10$data$trend)
b <- summary(Pre2012_stl_artmeans_10$data$trend)
c <- summary(Pre2012_stl_median_10$data$trend)
d <- summary(Pre2012_stl_SPUEmean_10$data$trend)
e <- summary(Pre2012_stl_SPUEgeomean_10$data$trend)
Pre2012_trend_all_10 <- cbind(b,c,a, d, e)

a <- summary(Pre2012_stl_geomeans_1$data$trend)
b <- summary(Pre2012_stl_artmeans_1$data$trend)
c <- summary(Pre2012_stl_median_1$data$trend)
d <- summary(Pre2012_stl_SPUEmean_1$data$trend)
e <- summary(Pre2012_stl_SPUEgeomean_1$data$trend)
Pre2012_trend_all_1 <- cbind(b,c,a, d, e)

colnames(Pre2012_trend_all_100) <- c("Mean Richness", "Median Richness", "Geo-Mean Richness", "Mean SPRUE", "Geo-
Mean SPRUE")
colnames(Pre2012_trend_all_10) <- c("Mean Richness", "Median Richness", "Geo-Mean Richness", "Mean SPRUE", "Geo-
Mean SPRUE")
colnames(Pre2012_trend_all_1) <- c("Mean Richness", "Median Richness", "Geo-Mean Richness", "Mean SPRUE", "Geo-Mean
SPRUE")

# c) Figure #--------------------------------------------------------------------------------------------------------------------------
fig <- "Figs/Analysis/stl/Figures for Chapter/00_Corrections/Boxplot_all_trendandseason.jpeg"
jpeg(fig, width=2500, height=1200)
par(mfrow=c(2,3), mar = c(1,2,2,1), oma = c(26,8,4,1)) #c(bottom, left, top, right))
boxplot(Pre2012_seasonal_all_100, xaxt = "n", yaxt = "t", las = 2, ylim = c(-0.25, 0.2), col = "white", border = "black", cex.axis
= 3)
boxplot(Pre2012_seasonal_all_10, xaxt = "n", yaxt = "n", las = 2, ylim = c(-0.25, 0.2), col = "white", border = "red")
boxplot(Pre2012_seasonal_all_1, xaxt = "n", yaxt = "n", las = 2, ylim = c(-0.25, 0.2), col = "white", border = "blue")
boxplot(Pre2012_trend_all_100, xaxt = "t", yaxt = "t", las = 2, ylim = c(0,7), col = "white", border = "black", cex.axis = 3)
boxplot(Pre2012_trend_all_10, xaxt = "t", yaxt = "n",las = 2, ylim = c(0,7), col = "white", border = "red", cex.axis = 3)
boxplot(Pre2012_trend_all_1, xaxt = "t",yaxt = "n", las = 2, ylim = c(0,7), col = "white", border = "blue", cex.axis = 3)
mtext("Value", side = 2, line = 1.5, outer = TRUE, cex = 2)
mtext("Measure", side = 1, line =24, outer = TRUE, cex = 2)
mtext("Seasonal", side = 2, line =3.5, adj = 0.8, outer = TRUE, cex = 2)
mtext("Trend", side = 2, line =3.5, adj = 0.2, outer = TRUE, cex = 2)
par(fig = c(0, 1, 0, 1), oma = c(0, 0, 0, 0), mar = c(0, 0, 0, 0), new = TRUE)
plot(0, 0, type = "n", bty = "n", xaxt = "n", yaxt = "n")
```

```
# make a new "overarching plot"to put over the panel, for legend
legend(
  "top",
  c(as.expression(bquote("100" ~ km^2)),
    as.expression(bquote("10" ~ km^2)),
    as.expression(bquote("1" ~ km^2))),
  xpd = TRUE,
  horiz = TRUE,
  bty = "n",
  lty = 1,
  col = c("black", "red", "blue"),
  cex = 3)
dev.off()
### -End of Script ###-----------------------------------------------------------------------------
```

Script 12: Richness on SD scale

```
### Load packages required ### -----------------------------------------------------------------------------------------------------------------
library(raster)
library(sp)
library(rgdal)
library(terra)
library(RColorBrewer)
library(ggplot2)
library(scales)
library(YesSiR)
library(sf)
library(ggpubr)
par(mfrow=c(1,1))
options(scipen = 999)

# Load required rasters
Raster_SP_100 <- raster("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/All/Rich_100.tif")
Raster_SP_10 <- raster("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/All/Rich_10.tif")
Raster_SP_1 <- raster("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/All/Rich_1.tif")

Raster_SRPUE_100 <- raster("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/All/rast_measure_100.tif")
Raster_SRPUE_10 <- raster("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/All/rast_measure_10.tif")
Raster_SRPUE_1 <- raster("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/All/rast_measure_1.tif")
plot(Raster_SRPUE_100)

# Get onto SD scale  ### -------------------------------------------------------------------------------------------------------------------------
# SR
a <- cellStats(Raster_SP_100, 'mean')
# Find mean of raster layer
b <- Raster_SP_100 - a
# Raster of SR minus mean
c <- cellStats(b, 'sd')
# Get the SD of this new raster layer
Raster_SD_SR_100 <- b / c
# Raster of SR - mean divided by SD = data range scale becomes unit of standard deviation

a <- cellStats(Raster_SP_10, ‘mean’)
b <- Raster_SP_10 – a
c <- cellStats(b, ‘sd’)
Raster_SD_SR_10 <- b / c # Raster of SR – mean divided by SD = data range scale becomes unit of standard deviation

a <- cellStats(Raster_SP_1, ‘mean’) # Find mean of raster layer
b <- Raster_SP_1 – a # Raster of SR minus mean
c <- cellStats(b, ‘sd’) # Get the SD of this new raster layer
Raster_SD_SR_1 <- b / c # Raster of SR – mean divided by SD = data range scale becomes unit of standard deviation

par(mfrow = c(3,1))
plot(Raster_SD_SR_100, zlim = c(-1.5,9))
plot(Raster_SD_SR_10, zlim = c(-1.5,9))
plot(Raster_SD_SR_1, zlim = c(-1.5,9))

# SRPUE
A <- cellStats(Raster_SRPUE_100, ‘mean’)
b <- Raster_SRPUE_100 - a
c <- cellStats(b, 'sd')
Raster_SD_SRPUE_100 <- b / c

a <- cellStats(Raster_SRPUE_10, 'mean')
b <- Raster_SRPUE_10 - a # Raster of SR minus mean
c <- cellStats(b, 'sd')
Raster_SD_SRPUE_10 <- b / c

a <- cellStats(Raster_SRPUE_1, 'mean')
```

```
b <- Raster_SRPUE_1 - a
c <- cellStats(b, 'sd')
Raster_SD_SRPUE_1 <- b / c

# Put on diverging colour scale ###-------
# load outlines
library(sf)
world <- read_sf('/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/data/worldoutline.shp')
cols <- brewer.pal((12), "RdYlBu")
# set colour scheme


# Plot all
fig <- "Figs/Analysis/Corrections/SD_Maps_Onelegend.jpeg"
jpeg(fig, width=1800, height=1200)   # open the file
par(mfrow = c(2,3), mai = c(0,0.5,0,0)) # bottom, left, top and right margins
plot(Raster_SD_SR_100, col = rev(cols), legend = FALSE, horizontal = FALSE, cex.axis = 2)+
  plot(world, col = "lightgrey", add = TRUE)
  plot(Raster_SD_SR_10,  col = rev(cols), zlim = c(-2.5,10), legend = FALSE, horizontal = TRUE, cex.axis = 2,  #main = "Raw species Richness", cex.main=2
     )+
  plot(world, col = "lightgrey", add = TRUE)
plot(Raster_SD_SR_1,  col = rev(cols), zlim = c(-2.5,10), legend = FALSE, cex.axis = 2)+
  plot(world, col = "lightgrey", add = TRUE)
plot(Raster_SD_SRPUE_100, col = rev(cols), zlim = c(-2.5,10), legend = FALSE, horizontal = TRUE, cex.axis = 2)+
  plot(world, col = "lightgrey", add = TRUE)
plot(Raster_SD_SRPUE_10,  col = rev(cols), zlim = c(-2.5,10), legend = FALSE, horizontal = TRUE, cex.axis = 2, #main = "SRPUE", cex.main=2
     )+
  plot(world, col = "lightgrey", add = TRUE)
plot(Raster_SD_SRPUE_1,  col = rev(cols), zlim = c(-2.5,10), legend = FALSE, horizontal = TRUE, cex.axis = 2)+
  plot(world, col = "lightgrey", add = TRUE)
dev.off()

### Correlation Plot between two measures ### -----------------------------------------

Stack_100 <- stack(Raster_SD_SR_100, Raster_SD_SRPUE_100)
Stack_10 <- stack(Raster_SD_SR_10, Raster_SD_SRPUE_10)
Stack_1 <- stack(Raster_SD_SR_1, Raster_SD_SRPUE_1)

install.packages('spatialEco')
library(raster)
jnk=layerStats(Stack_10, 'pearson', na.rm=T)
corr_matrix=jnk$'pearson correlation coefficient'

overlay_100 <- data.frame(na.omit(values(Stack_100)))
cor_100 <- cor.test(overlay_100[,1], overlay_100[,2], )
overlay_10 <- data.frame(na.omit(values(Stack_10)))
cor_10 <- cor.test(overlay_10[,1], overlay_10[,2])
overlay_1 <- data.frame(na.omit(values(Stack_1)))
cor_1 <- cor.test(overlay_1[,1], overlay_1[,2])

### Corr plot ###-------------------------------------------------------------------------------------------------------------
library(corrplot)
mydata.cor = cor(overlay_100)
corrplot(mydata.cor, method="circle", is.corr=FALSE)
raster.cor(Stack_100$rich_100, Stack_100$rast_measure_100, method = "spearman")

### Compare rasters - difference between two rasters ###
# Get differences between two SD rasters
overlay_100$diff <- overlay_100$rich_100 - overlay_100$rast_measure_100
head(overlay_100)
overlay_10$diff <- overlay_10$rich_10 - overlay_10$rast_measure_10
head(overlay_10)
overlay_1$diff <- overlay_1$rich_1 - overlay_1$rast_measure_1
head(overlay_1)

# Plot differences as a map


# Stacks are SR raster (SD) and SRPUE Raster (SD)
Diff_100 <- Stack_100$rich_100 - Stack_100$rast_measure_100
```

```
cellStats(Diff_100, mean)
Diff_10 <- Stack_10$rich_10 - Stack_10$rast_measure_10
cellStats(Diff_10, mean)
Diff_1 <- Stack_1$rich_1 - Stack_1$rast_measure_1
cellStats(Diff_1, mean)

# looking at areas of difference
dif100 <- Diff_100
dif100[dif100 != 0] <- 1
plot(dif100)

dif10 <- Diff_10
dif10[dif10 != 0] <- 1
plot(dif10)

dif1 <- Diff_1
dif1[dif1 != 0] <- 1
plot(dif1)

plot(dif100, col = "blue", legend = TRUE, horizontal = FALSE, cex.axis = 1)+
  plot(world, col = "lightgrey", add = TRUE)
dev.off()
# fig <- "Figs/Analysis/Corrections/SD_Differences_100.jpeg"
# jpeg(fig, width=1500, height=1500)   # open the file
plot(Diff_100, col = pal(21), legend = TRUE, horizontal = FALSE, cex.axis = 1)+
  plot(world, col = "lightgrey", add = TRUE)
dev.off()

# Get range for palette
scale_range_100 <- c(cellStats(Diff_100, min), cellStats(Diff_100, max))
library(leaflet)
pal <- colorNumeric("RdBu", domain = scale_range_100)
scale_range_10 <- c(cellStats(Diff_10, min), cellStats(Diff_10, max))
scale_range_1 <- c(cellStats(Diff_1, min), cellStats(Diff_1, max))

diff_100_df <- as.data.frame(Diff_100, xy = TRUE)
diff_10_df <- as.data.frame(Diff_10, xy = TRUE)
diff_1_df <- as.data.frame(Diff_1, xy = TRUE)
lower <- min(test_df$layer, na.rm = TRUE)
midway <- median(test_df$layer, na.rm = TRUE)
upper <- max(test_df$layer, na.rm = TRUE)
# get coordinate ref system to match first figure set
selectedCRS <- st_crs(Raster_SD_SR_100)
# plot difference in measures
# 100km2
plot_diff_100 <- ggplot() +
  geom_raster(data = diff_100_df, aes(x = x, y = y, fill = layer)) +
  scale_fill_gradient2(low = "#54bebe", mid = "ghostwhite", high ="#c80064",
                       midpoint = 0,
                       space = "Lab",
                       guide = "colourbar",
                       na.value="transparent",
                       limits = c((min(diff_1_df$layer, na.rm = TRUE)), (max(diff_1_df$layer, na.rm = TRUE))),
                       breaks= pretty_breaks(n = 7),
                       # breaks = waiver()
                       )+
  geom_sf(data = world, color = "black", fill = "lightgray") +
  coord_sf(xlim = c(min(test_df$x+10), max(test_df$y)-10),
           ylim  = c(min(test_df$y)+10, max(test_df$y)-10),
           datum = st_crs(selectedCRS))+ # set CRS to match previous figure (not degrees)
  theme_linedraw()+
  theme(axis.title = element_blank(),
        axis.text.y = element_text(angle = 90, hjust = 0.5),
        axis.text = element_text(size = 12),
        plot.margin = margin(t = 0,  # Top margin
                             r = 0.1,  # Right margin
                             b = 0,  # Bottom margin
                             l = 0.1,
                             unit = "cm"),
        panel.grid.major = element_blank(),
        legend.position = "right",
        legend.text = element_text(size = 14))+ # change this to none to remove legend for plot
  labs(fill='Diff')+
```

```
  guides(fill = guide_colorbar(barwidth = 1, barheight = 20))
plot_diff_100

# Plot difference
plot_diff_10 <- ggplot() +
  geom_raster(data = diff_10_df, aes(x = x, y = y, fill = layer)) +
  scale_fill_gradient2(low = "#54bebe", mid = "ghostwhite", high ="#c80064",
                    midpoint = 0,
                    space = "Lab",
                    guide = "colourbar",
                    na.value="transparent",
                    limits = c((min(diff_1_df$layer, na.rm = TRUE)), (max(diff_1_df$layer, na.rm = TRUE))),
                    breaks= pretty_breaks(n = 5)
                    # breaks = waiver()
                    )+
  geom_sf(data = world, color = "black", fill = "lightgrey") +
  coord_sf(xlim = c(min(diff_10_df$x)+10, max(diff_10_df$y)-10),
         ylim  = c(min(diff_10_df$y)+10, max(diff_10_df$y)-10),
         datum = st_crs(selectedCRS))+ # set CRS to match previous figure (not degrees)
  theme_linedraw()+
  theme(axis.title = element_blank(),
      axis.text.y = element_text(angle = 90, hjust = 0.5),
      axis.text = element_text(size = 12),
      plot.margin = margin(t = 0,  # Top margin
                         r = 0.1,  # Right margin
                         b = 0,  # Bottom margin
                         l = 0.1,
                         unit = "cm"),
      panel.grid.major = element_blank(),
      legend.position = "none")+ # change this to none to remove legend for plot
  labs(fill='Diff')+
  guides(fill = guide_colorbar(barwidth = 1, barheight = 20, horizontal = TRUE))

plot_diff_10

# Plot 1 diff
plot_diff_1 <- ggplot() +
  geom_raster(data = diff_1_df, aes(x = x, y = y, fill = layer)) +
  scale_fill_gradient2(low = "#54bebe", mid = "ghostwhite", high ="#c80064",
                    midpoint = 0,
                    space = "Lab",
                    guide = "colourbar",
                    na.value="transparent",
                    limits = c((min(diff_1_df$layer, na.rm = TRUE)), (max(diff_1_df$layer, na.rm = TRUE))),
                    breaks= pretty_breaks(n = 5)
                    # breaks = waiver()
                    )+
  geom_sf(data = world, color = "black", fill = "lightgrey") +
  coord_sf(xlim = c(min(diff_1_df$x)+10, max(diff_1_df$y)-10),
         ylim  = c(min(diff_1_df$y)+10, max(diff_1_df$y)-10),
         datum = st_crs(selectedCRS))+ # set CRS to match previous figure (not degrees)
  theme_linedraw()+
  theme(axis.title = element_blank(),
      axis.text.y = element_text(angle = 90, hjust = 0.5),
      axis.text = element_text(size = 12),
      plot.margin = margin(t = 0,  # Top margin
                         r = 0.1,  # Right margin
                         b = 0,  # Bottom margin
                         l = 0.1,
                         unit = "cm"),
      panel.grid.major = element_blank(),
      legend.position = "none")+ # change this to none to remove legend for plot)
  labs(fill='Diff')+
  guides(fill = guide_colorbar(barwidth = 1, barheight = 20, horizontal = TRUE))
plot_diff_1


fig <- "Figs/Analysis/Corrections/Differences_SD_3scales_Rev2.jpeg"
jpeg(fig, width=1800, height=1200, res = 100)   # open the file
ggarrange(plot_diff_100, plot_diff_10, plot_diff_1, ncol = 3, nrow = 1)
dev.off()
print("finished")

#### Scaling 0-1 ###------------------------------------------------------------------------------
mnv <- cellStats(Diff_100,'min')
```

```
mxv <- cellStats(Diff_100,'max')
NormalDiff_100 <- (Diff_100 - mnv) / (mxv - mnv)
plot(NormalDiff_100)

mnv <- cellStats(Diff_10,'min')
mxv <- cellStats(Diff_10,'max')
NormalDiff_10 <- (Diff_10 - mnv) / (mxv - mnv)
plot(NormalDiff_10)

mnv <- cellStats(Diff_1,'min')
mxv <- cellStats(Diff_1,'max')
NormalDiff_1 <- (Diff_1 - mnv) / (mxv - mnv)
plot(NormalDiff_1)

NormalDiff_100_df <- as.data.frame(NormalDiff_100, xy = TRUE)
NormalDiff_10_df <- as.data.frame(NormalDiff_10, xy = TRUE)
NormalDiff_1_df <- as.data.frame(NormalDiff_1, xy = TRUE)

hcl_palettes(plot = TRUE)


#Figure plot
plot_normal_100 <- ggplot() +
  geom_raster(data = NormalDiff_100_df, aes(x = x, y = y, fill = layer)) +
  scale_fill_gradientn(colours = rev(sequential_hcl(10,"PurpleYellow")),
                       guide = "colourbar",
                       na.value="transparent",
                       breaks= pretty_breaks(n = 5),
                       # breaks = waiver()
  )+
  geom_sf(data = world, color = "black", fill = "lightgray") +
  coord_sf(xlim = c(min(test_df$x+10), max(test_df$y)-10),
           ylim  = c(min(test_df$y)+10, max(test_df$y)-10))+
  theme_bw()+
  theme(axis.title = element_blank(),
        plot.margin = margin(t = 0,  # Top margin
                             r = 0.1,  # Right margin
                             b = 0,  # Bottom margin
                             l = 0.1,
                             unit = "cm"),
        legend.position = "right")+
  labs(fill='Diff.')+
  guides(fill = guide_colorbar(barwidth = 1, barheight = 20))
plot_normal_100

plot_normal_10 <- ggplot() +
  geom_raster(data = NormalDiff_10_df, aes(x = x, y = y, fill = layer)) +
  scale_fill_gradientn(colours = rev(sequential_hcl(10,"PurpleYellow")),
                       guide = "colourbar",
                       na.value="transparent",
                       breaks= pretty_breaks(n = 5),
                       # breaks = waiver()
  )+
  geom_sf(data = world, color = "black", fill = "lightgray") +
  coord_sf(xlim = c(min(test_df$x+10), max(test_df$y)-10),
           ylim  = c(min(test_df$y)+10, max(test_df$y)-10))+
  theme_bw()+
  theme(axis.title = element_blank(),
        plot.margin = margin(t = 0,  # Top margin
                             r = 0.1,  # Right margin
                             b = 0,  # Bottom margin
                             l = 0.1,
                             unit = "cm"),
        legend.position = "right")+
  labs(fill='Diff.')+
  guides(fill = guide_colorbar(barwidth = 1, barheight = 20))
plot_normal_10

plot_normal_1 <- ggplot() +
  geom_raster(data = NormalDiff_1_df, aes(x = x, y = y, fill = layer)) +
  scale_fill_gradientn(colours = rev(sequential_hcl(10,"PurpleYellow")),
                       guide = "colourbar",
                       na.value="transparent",
                       breaks= pretty_breaks(n = 5),
                       # breaks = waiver()
  )+
```

```
  geom_sf(data = world, color = "black", fill = "lightgray") +
  coord_sf(xlim = c(min(test_df$x+10), max(test_df$y)-10),
           ylim  = c(min(test_df$y)+10, max(test_df$y)-10))+
  theme_bw()+
  theme(axis.title = element_blank(),
        plot.margin = margin(t = 0,  # Top margin
                             r = 0.1,  # Right margin
                             b = 0,  # Bottom margin
                             l = 0.1,
                             unit = "cm"),
        legend.position = "right")+
  labs(fill='Diff.')+
  guides(fill = guide_colorbar(barwidth = 1, barheight = 20))
plot_normal_1


fig <- "Figs/Analysis/Corrections/Differences_SD_Normalised.jpeg"
jpeg(fig, width=3000, height=1500, res = 100)   # open the file
ggarrange(plot_normal_100, plot_normal_10, plot_normal_1, ncol = 3, nrow = 1)
dev.off()
print("finished")

## Comparison stats ###----------------------------------------------------------------------------------------------------------------

# SR stats
stats100 <- cbind(cellStats(Raster_SP_100, 'min'),
      cellStats(Raster_SP_100, 'mean'),
      cellStats(Raster_SP_100, 'max'),
      cellStats(Raster_SP_100, 'sd'))
stats10 <- cbind(cellStats(Raster_SP_10, 'min'),
                 cellStats(Raster_SP_10, 'mean'),
                 cellStats(Raster_SP_10, 'max'),
                 cellStats(Raster_SP_10, 'sd'))
stats1 <- cbind(cellStats(Raster_SP_1, 'min'),
                 cellStats(Raster_SP_1, 'mean'),
                 cellStats(Raster_SP_1, 'max'),
                 cellStats(Raster_SP_1, 'sd'))
stats_SR <- rbind(stats100,stats10,stats1)
colnames(stats_SR) <- c("Min","Mean","Max","SD")
rownames(stats_SR) <- c("100km2","10km2","1km2")
stats_SR <- as.data.frame(stats_SR)
class(stats_SR)
#get variance
Var_SP_100 <- as.data.frame(Raster_SP_100, xy = TRUE)
Var_SP_100 <- var(Var_SP_100$rich_100, na.rm = TRUE)
Var_SP_10 <- as.data.frame(Raster_SP_10, xy = TRUE)
Var_SP_10 <- var(Var_SP_10$rich_10, na.rm = TRUE)
Var_SP_1 <- as.data.frame(Raster_SP_1, xy = TRUE)
Var_SP_1 <- var(Var_SP_1$rich_1, na.rm = TRUE)
Variance_SP <- rbind(Var_SP_100,Var_SP_10,Var_SP_1)
stats_SR$Var <- as.numeric(Variance_SP)
# get coefficient of varation
# CV = SD / Mean
CV_SP <- stats_SR$SD/stats_SR$Mean
stats_SR$CV <- CV_SP

# Final table SR-----------------------------------
flextable::as_flextable(stats_SR)

stats_SRPUE_100 <- cbind(cellStats(Raster_SRPUE_100, 'min'),
                 cellStats(Raster_SRPUE_100, 'mean'),
                 cellStats(Raster_SRPUE_100, 'max'),
                 cellStats(Raster_SRPUE_100, 'sd'))
stats_SRPUE_10 <- cbind(cellStats(Raster_SRPUE_10, 'min'),
                cellStats(Raster_SRPUE_10, 'mean'),
                cellStats(Raster_SRPUE_10, 'max'),
                cellStats(Raster_SRPUE_10, 'sd'))
stats_SRPUE_1 <- cbind(cellStats(Raster_SRPUE_1, 'min'),
               cellStats(Raster_SRPUE_1, 'mean'),
               cellStats(Raster_SRPUE_1, 'max'),
               cellStats(Raster_SRPUE_1, 'sd'))
stats_SRPUE <- rbind(stats_SRPUE_100,stats_SRPUE_10,stats_SRPUE_1)
colnames(stats_SRPUE) <- c("Min","Mean","Max","SD")
rownames(stats_SRPUE) <- c("100km2","10km2","1km2")
```

```
stats_SRPUE <- as.data.frame(stats_SRPUE)
class(stats_SRPUE)
#get variance
Var_SRPUE_100 <- as.data.frame(Raster_SRPUE_100, xy = TRUE)
Var_SRPUE_100 <- var(Var_SRPUE_100$rast_measure_100, na.rm = TRUE)
Var_SRPUE_10 <- as.data.frame(Raster_SRPUE_10, xy = TRUE)
Var_SRPUE_10 <- var(Var_SRPUE_10$rast_measure_10, na.rm = TRUE)
Var_SRPUE_1 <- as.data.frame(Raster_SRPUE_1, xy = TRUE)
Var_SRPUE_1 <- var(Var_SRPUE_1$rast_measure_1, na.rm = TRUE)
Variance_SRPUE <- rbind(Var_SRPUE_100,Var_SRPUE_10,Var_SRPUE_1)
stats_SRPUE$Var <- as.numeric(Variance_SRPUE)
# get coefficient of varation
# CV = SD / Mean
CV_SRPUE <- stats_SRPUE$SD/stats_SRPUE$Mean
stats_SRPUE$CV <- CV_SRPUE

# pop vairance
# df_SR_100 <- as.data.frame(Raster_SP_100)
# n <- length(df_SR_100$rich_100)
# popSR100 <- var(df_SR_100$rich_100, na.rm = TRUE) * (n-1)/n
# df_SR_10 <- as.data.frame(Raster_SP_10)
# n <- length(df_SR_10$rich_10)
# popSR10 <- var(df_SR_10$rich_10, na.rm = TRUE) * (n-1)/n

# Final table SR
flextable::as_flextable(stats_SRPUE)

SummaryTableAll <- rbind(stats_SR,stats_SRPUE)
flextable::as_flextable(SummaryTableAll)
exportxlsx(flextable::as_flextable(SummaryTableAll), path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/Tables/SummaryTableAll.xlsx")

### geometric mean
library('psych')
df_SR_100 <- as.data.frame(Raster_SP_100)
df_SR_10 <- as.data.frame(Raster_SP_10)
df_SR_1 <- as.data.frame(Raster_SP_1)
df_SRPUE_100 <- as.data.frame(Raster_SRPUE_100)
df_SRPUE_10 <- as.data.frame(Raster_SRPUE_10)
df_SRPUE_1 <- as.data.frame(Raster_SRPUE_1)

amean_100 <- mean(na.omit(df_SR_100$rich_100))
amean_10 <- mean(na.omit(df_SR_10$rich_10))
amean_1 <- mean(na.omit(df_SR_1$rich_1))

geometric.mean(df_SR_100$rich_100)
geometric.mean(df_SR_10$rich_10)
geometric.mean(df_SR_1$rich_1)
geometric.mean(df_SRPUE_100$rast_measure_100)
geometric.mean(df_SRPUE_10$rast_measure_10)
geometric.mean(df_SRPUE_1$rast_measure_1)

# Data from rasters
data100  #100km df from raster
data10  #10km df from raster
data1  #1km df from raster

### End of Script ###------------------------------------------------------------------------------
```

Script 13: Setting up for GAMs

```
### GAMS for Environmental Drivers  ###------------------------------------------------------------------------------------------------------
### Species Richness (raw) ### ----------------------------------------------------------------------------------------------------------------
# Load packages
library(mgcv)
library(sp)
library(raster)
library(ggplot2)
library(dplyr)
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2. Environmental_drivers/GAMs/Env_Drivers_GAMS")
# A) 100 km2 GRID CELLS # -------------------------------------------------------------------------------------------------------------
#  1) Get coords for centre of grid cells # ----------------------------------------------------------------------------------------------
rich_100.tif <- raster("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/1. Species Richness/Species_Richness_v2/tiffs/All/rich_100.tif")
plot(rich_100.tif)
crs(rich_100.tif)
# check they are integers when imported
test <- na.omit(as.data.frame(rich_100.tif))
unique(unlist(test))              # Return unique values
# Need to reproject to lat long for matching env variables (needs loading below) and SSDM
# Important: when change to new projection, it changes the richness values as it is RESAMPLING.
# You need to set the method depending on whether this is wanted or not
# https://gis.stackexchange.com/questions/326180/values-change-after-re-projecting-the-netcdf-files-to-tiff-file
# Works by changing METHOD ("ngb" or "bilinear") below:
rich_100_reproj <- projectRaster(rich_100.tif, crs = crs(stack_env_OBS), method = "ngb") # method ngb is nearest neighbour, so keeps closest value
# Syntax is projectRaster(RasterObject, crs = CRSToReprojectTo)
plot(rich_100_reproj)
# check still integers in species richness
test <- na.omit(as.data.frame(rich_100_reproj))
unique(unlist(test))
# get coordinates of center coordinates of each cell and RICHNESS
SR_points_100km <- coordinates(rich_100_reproj)
SR_values_100km <- values(rich_100_reproj)
### 2) Extract raster values from points - including OBS ### -------------------------------------------------------------------------------
# NOTE Can re-load stack from raster, rather than reprocessing to same resolution via steps below:
setwd("/Volumes/#1 PhD Working Seagate/PhD/R Stats/PhD Analysis/2. Environmental_drivers/GAMs/Env_Drivers_GAMS/Data/Env_Stack")
rastlist_OBS <- list.files(path = "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2. Environmental_drivers/GAMs/Env_Drivers_GAMS/Data/Env_Stack", pattern='.tif$', all.files=TRUE, full.names=FALSE)
stack_env_OBS <- stack(rastlist_OBS)
stack_env_OBS
# Processing steps to Resample - NOT NEEDED NOW DONE AND SAVED
# If only need to resample OBSERVATIONS (rest have been done and saved) -----
# Step 1: Create a Raster stack or Raster brick of your raster files using “raster” package in R.
# Set WD to ones with all the files in
# This one INCLUDES observations (aka effort)
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2. Environmental_drivers/GAMs/Env_Drivers_GAMS/Data/Env_Stack")
env1_bathy_rs <- raster("env1_bathy_rs.tif")
obs_100 <- raster("obs_100.tif") # OBSERVATIONS not RICHNESS
# Reproject to lat long
obs_100 <- projectRaster(obs_100,= crs(stack_env), method = "ngb")
# Resample to match env resolutions
# ngb, nearest neighbour - to keep values
obs_100_rs <- resample(obs_100, env1_bathy_rs, method="ngb")
plot(obs_100_rs)
# Use AOI from Script 2
AOI <- raster(ncol=1, nrow=1, xmn=-26.6875, xmx=18.30417, ymn=33.92917, ymx=72.70417)
# Extend all to be same extent
obs_100_rs = extend(obs_100_rs, AOI)
# Crop to extent of AOI if over
obs_100_rs = crop(obs_100_rs, AOI)
# Save resampled tiff - OBS only
```

```
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/GAMs/Env_Drivers_GAMS/Data/Env_Stack")
writeRaster(obs_100_rs, "obs_100_rs", format = 'GTiff', overwrite = TRUE)
### 3) Extract raster values from points ###) ### ---------------------------------------------------------------------------------------------
# Raster stack of env variables + obs
stack_env_OBS
# Point data - get centre points of the rich_100 grid as "points" using coordinates
# From first steps: points, values from rich_100_proj
head(SR_points_100km)
head(SR_values_100km)
# Extract raster value by points (ONLY need spatial POINTS bit here)
rasValue_SR_100km <- extract(stack_env_OBS, SR_points_100km)
head(rasValue_SR_100km)
# check has values
test <- na.omit(as.data.frame(rasValue_SR_100km))
head(test)
# Combine raster values with point and save as a CSV file.
combinePointValue_SR_100km2 <- cbind(SR_points_100km,SR_values_100km,rasValue_SR_100km)
head(combinePointValue_SR_100km2)
test <- na.omit(combinePointValue_SR_100km2)
head(test)
write.table(combinePointValue_SR_100km2, file= "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/GAMs/Env_Drivers_GAMS/Data/Outputs/RastValsfromPoints_100km_SR_Dataset.csv",
append=FALSE, sep= ",", row.names = FALSE, col.names=TRUE)
### End of Script ###--------------------------------------------------------------------------------
```

Script 14: GAMS

```
# Load Subset from SCRIPT A following steps until below is in environment:
head(RastValsfromPoints_100km_SR_Dataset_SUBSET)

# Step 1) Amended GAM to remove any unneccesary Env Variables ###-------------------------------------------------------------------
# a) Gam for species richness, with offset, and lat and long REMOVED - select = TRUE ###----------------------------------------
gam_SR_100km_OFFSET_Corrections_v4 <-
  gam(
    RastValsfromPoints_100km_SR_Dataset_SUBSET$SR_values_100km ~ offset(log(
      log(RastValsfromPoints_100km_SR_Dataset_SUBSET$obs_100_rs)
    )) +
      s(v1_SR_100_bathy_subset, k = 4) + s(v2_SR_100_chl_subset, k = 4) + s(v3_SR_100_fronts_subset, k = 4) +
s(v4_SR_100_d2l_subset, k = 4) + s(v5_SR_100_sst_subset, k = 4) + s(v6_SR_100_sal_subset, k = 4) +
s(v7_SR_100_mix_subset, k = 4) + s(v9_SR_100_slp_subset, k = 4),
    family = nb(theta = NULL, link =  "log"),
    select = TRUE
  )
# Interpet - https://noamross.github.io/gams-in-r-course/chapter2
summary(gam_SR_100km_OFFSET_Corrections_v4)
par(mfrow = c(2, 2)) # Create a 2 x 2 plotting matrix
gam.check(gam_SR_100km_OFFSET_Corrections_v4)
flextable::as_flextable(gam_SR_100km_OFFSET_Corrections_v4)
plot(gam_SR_100km_OFFSET_Corrections_v4, pages = 1, all.terms = TRUE, rug = TRUE, residuals = FALSE, pch = 1, cex =
1, seWithMean = TRUE, shade = TRUE, shade.col = "lightblue")
## b) GAMS of SPRUE ### -------------------------------------------------------------------------------------------------------------
# SPRUE - K limited for all env variables, lat/long removed,select = TRUE #--------------
gam_SPRUE_100km_OFFSET_Corrections_v4 <-
  gam(
    sp_richness_SPRUE_100km ~ s(v1_SPRUE_100_bathy, k = 4) + s(v2_SPRUE_100_chl, k = 4) + s(v3_SPRUE_100_fronts,
k = 4) + s(v4_SPRUE_100_d2l, k = 4) + s(v5_SPRUE_100_sst, k = 4) + s(v6_SPRUE_100_sal, k = 4) + s(v7_SPRUE_100_mix,
k = 4) + s(v9_SPRUE_100_slp, k = 4),
    family = Gamma(link =  "log"),
    select = TRUE)
# Interpet - https://noamross.github.io/gams-in-r-course/chapter2
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)
par(mfrow = c(2, 2)) # Create a 2 x 2 plotting matrix
gam.check(gam_SPRUE_100km_OFFSET_Corrections_v4)
flextable::as_flextable(gam_SPRUE_100km_OFFSET_Corrections_v4)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, pages = 1, all.terms = TRUE, rug = TRUE, residuals = FALSE, pch = 1,
cex = 1, seWithMean = TRUE, shade = TRUE, shade.col = "lavender")
# check for shrunk terms
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$s.pv
gam_SPRUE_100km_OFFSET_Corrections_v3, gam_SPRUE_100km_OFFSET_Corrections_v4)
#  Figures – maps of GAMS ### ----------------------------------------------------------------------------------------------------------
library(dplyr)
library(ggmap)
library(ggplot2)
library(viridis)
library(maps)
library(raster)
### 1) Setup GAMs  ### ------------------------------------------------------------------------------------------------------------------
# Requires GAMS:
gam_SR_100km_OFFSET_Corrections_v4 # SR selected from above
gam_SPRUE_100km_OFFSET_Corrections_v4 # SRPUE selected selected from above
# Get predicted probabilities for each value using predict on gam
gampred_SR <- predict(gam_SR_100km_OFFSET_Corrections_v4, type = "response")
head(gampred_SR)
class(gampred_SR)
test <- as.data.frame.table(gampred_SR)
gampred_SPRUE <- predict(gam_SPRUE_100km_OFFSET_Corrections_v4, type = "response")
### Attach predict values to the spatial coordinates - using original dataframe
# original dataframes from Script 1 & Script 2
head(RastValsfromPoints_100km_SR_Dataset) # Species richness dataset
head(RastValsfromPoints_100km_SPRUE_Dataset) # SPRUE dataset
# Species Richness
# Remove NA's to match up with outcome of predict()
```

```
prediction_SR <- na.omit(RastValsfromPoints_100km_SR_Dataset_SUBSET) # Used subset
RastValsfromPoints_100km_SR_Dataset_SUBSET - no 1's SR
# trying to remove df with 1 SR's
head(RastValsfromPoints_100km_SR_Dataset)
test <- prediction_SR %>%  filter(obs_100_rs>'1')
test$predict <- gampred_SR # attach output of predict
head(test) #check
class(test)

# SPRUE
prediction_SPRUE <- na.omit(RastValsfromPoints_100km_SPRUE_Dataset) # note/check that it matches point #8509
head(prediction_SPRUE)
prediction_SPRUE$predict <- gampred_SPRUE # attach output of predict
head(prediction_SPRUE)
plot(prediction_SPRUE$predict)
class(prediction_SPRUE)

### 2) Reformatting to match SSDM format  ### ------------------------------------------------------------------------------------------
library(dplyr)
# SR
# select columns for converting to raster
prediction_SR_clean <- dplyr::select(prediction_SR, c('x','y','predict'))
head(prediction_SR_clean)
plot(prediction_SR_clean)
prediction_SR_raster <- rasterFromXYZ(prediction_SR_clean, digits = 1)
#Convert first two columns as lon-lat and third as value
library(terra)
test <- rast(prediction_SR_clean, type="xyz", crs="", digits=NA, extent=NULL)
plot(test)
jpeg("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/GAMs/Env_Drivers_GAMS/Exports/Predicted maps/SR_predict.jpeg", width = 1200, height = 1500)
par(bg = NA, mar = c(5,5,5,5), oma = c(5, 5, 5, 5))
plot(
  prediction_SR_raster,
  xaxs = "i",
  yaxs = "i",
  ylim = c(30, 73),
  xlim = c(-25, 18),
  cex.axis = 2,
  xlab = "Longitude",
  ylab = "Latitude",
  cex.lab = 3,
  legend = FALSE,
  col = viridis(131199),
  legend.width = 2,
  legend.shrink = 0.8,
  legend.args = list(
    text = 'Richness',
    side = 4,
    font = 2,
    line = 2.5,
    cex = 2))
map('world', add = T, col="grey", fill=TRUE, bg="white", lwd=0.05, border=0)
plot(
  prediction_SR_raster,
  legend.only = TRUE,
  cex = 2,
  col = viridis(131199),
  legend.width = 4,
  legend.shrink = 0.8,
  text.width = 1,
  legend.args = list(
    text = 'Richness',
    side = 4,
    font = 2,
    line = 2.5,
    cex = 2))
dev.off()
#SPRUE
# Select columns for converting to raster
prediction_SPRUE_clean <- dplyr::select(prediction_SPRUE, c('x','y','predict'))
head(prediction_SPRUE_clean)
```

```
prediction_SPRUE_raster <- rasterFromXYZ(prediction_SPRUE_clean)  #Convert first two columns as lon-lat and third as value
plot(prediction_SPRUE_raster)
test1 <- dplyr::select(test, c('x','y','predict'))
prediction_SR_raster <- rasterFromXYZ(test1)  #Convert first two columns as lon-lat and third as value
plot(prediction_SPRUE_raster)
### Figures of GAMS ###-----------------------------------------------------------
# required GAMS
gam_SR_100km_OFFSET_Corrections_v4
gam_SPRUE_100km_OFFSET_Corrections_v4
options(scipen=999) # disable scientific notation
### Line up responses from MEM for SP versus SSDM
jpeg(file="/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/GAMs/Env_Drivers_GAMS/Exports/Corrections/GAM
Outputs/GAM_SR_100km_SRversusSRPUE_Part1.jpeg", width=1500, height = 2000)
par(mai=c(1,0.8,0.1,0.1), mfrow = c(4, 2)) # # B,L,T,R #  Create a 2 x 2 plotting matrix
# 1) bathymetry
plot(gam_SR_100km_OFFSET_Corrections_v4, select = 1, trans = exp, shade = TRUE, shade.col = "lightblue", ylab = "Sp
Richness", xlab = "Bathymetry (m)", ylim = c(0,2), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, select = 1, shift =
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$p.coeff, trans = exp, shade = TRUE, shade.col = "lavender", ylab =
"\n\nSRPUE", xlab = "\nBathymetry (m)", ylim = c(1,3), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
# 2) Chla
plot(gam_SR_100km_OFFSET_Corrections_v4, select = 2, trans = exp, shade = TRUE, shade.col = "lightblue", ylab = "Sp
Richness", xlab = "Chl A (milligram m-3)", ylim = c(0,2), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, select = 2, shift =
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$p.coeff, trans = exp, shade = TRUE, shade.col = "lavender", ylab =
"\n\nSRPUE", xlab = "\nChl A (mg-3)", ylim = c(1,3.5), cex.lab = 3, cex.axis = 2.5,rug = TRUE )
# 3) Fronts
plot(gam_SR_100km_OFFSET_Corrections_v4, select = 3, trans = exp, shade = TRUE, shade.col = "lightblue", ylab = "Sp
Richness", xlab = "Fronts (Fcomp)", ylim = c(0,2), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, select = 3, shift =
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$p.coeff, trans = exp, shade = TRUE, shade.col = "lavender", ylab =
"\n\nSRPUE", xlab = "\nFronts", ylim = c(1,2.5), cex.lab = 3, cex.axis = 2.5, rug = TRUE )
# 4) D2L
plot(gam_SR_100km_OFFSET_Corrections_v4, select = 4, trans = exp, shade = TRUE, shade.col = "lightblue", ylab = "Sp
Richness", xlab = "Distance to land (m)", ylim = c(0,2), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, select = 4, shift =
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$p.coeff, trans = exp, shade = TRUE, shade.col = "lavender", ylab =
"\n\nSRPUE", xlab = "\nDistance to land (m)", ylim = c(0,2.5), cex.lab = 3, cex.axis = 2.5,rug = TRUE )
dev.off()
jpeg(file="/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/GAMs/Env_Drivers_GAMS/Exports/Corrections/GAM
Outputs/GAM_SR_100km_SRversusSRPUE_Part2.jpeg", width=1500, height = 2000)
par(mai=c(1,0.8,0.1,0.1), mfrow = c(4, 2)) # # B,L,T,R #  Create a 2 x 2 plotting matrix
# 5) SST
plot(gam_SR_100km_OFFSET_Corrections_v4, select = 5, trans = exp, shade = TRUE, shade.col = "lightblue", ylab = "Sp
Richness", xlab = "SST (K)", ylim = c(0,6), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, select = 5, shift =
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$p.coeff, trans = exp, shade = TRUE, shade.col = "lavender", ylab =
"\n\nSRPUE", xlab = "\nSST (K)", cex.lab = 3, cex.lab = 3, cex.axis = 3, rug = TRUE )
#6) Sal
plot(gam_SR_100km_OFFSET_Corrections_v4, select = 6, trans = exp, shade = TRUE, shade.col = "lightblue", ylab = "Sp
Richness", xlab = "Salinity (psu)", ylim = c(0,2), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, select = 6, shift =
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$p.coeff, trans = exp, shade = TRUE, shade.col = "lavender", ylab =
"\n\nSRPUE", xlab = "\nSalinity (psu)", ylim = c(0,3.5), cex.lab = 3, cex.axis = 2.5,rug = TRUE )
#7) MLT
plot(gam_SR_100km_OFFSET_Corrections_v4, select = 7, trans = exp, shade = TRUE, shade.col = "lightblue", ylab = "Sp
Richness", xlab = "Mixed layer thickness (m)", ylim = c(0,2), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, select = 7, shift =
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$p.coeff, trans = exp, shade = TRUE, shade.col = "lavender", ylab =
"\n\nSRPUE", xlab = "\nMixed Layer thickness (m)", ylim = c(1,2.5), cex.lab = 3, cex.axis = 2.5,rug = TRUE )
# 8) Slope
plot(gam_SR_100km_OFFSET_Corrections_v4, select = 8, trans = exp, shade = TRUE, shade.col = "lightblue", ylab = "Sp
Richness", xlab = "Slope (degrees)", ylim = c(0,2), cex.lab = 3, cex.axis = 2.5, rug = TRUE)
plot(gam_SPRUE_100km_OFFSET_Corrections_v4, select = 8, shift =
summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$p.coeff, trans = exp, shade = TRUE, shade.col = "lavender", ylab =
"\n\nSRPUE", xlab = "\nSlope (degrees)", ylim = c(0,4), cex.lab = 3, cex.axis = 2.5,rug = TRUE )
dev.off()
```

```
### Variable importance new GAMS #-----------------------------------------------------------------------------------------------------
# Sp Richness
VI_SPR <- 100 * summary(gam_SR_100km_OFFSET_Corrections_v4)$chi.sq /
sum(summary(gam_SR_100km_OFFSET_Corrections_v4)$chi.sq)
sd(VI_SPR)
as.data.frame(VI_SPR)
# Sp Richness
VI_SRPUE <- 100 * summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$chi.sq /
sum(summary(gam_SPRUE_100km_OFFSET_Corrections_v4)$chi.sq)
as.data.frame(VI_SRPUE)
sd(VI_SRPUE)
barplot(VI_SPR,main="SPR",ylab="Importance",xlab="Variable",las=2)
barplot(VI_SRPUE,main="SPR",ylab="Importance",xlab="Variable",las=2)
### -End of Script ###-----------------------------------------------------------------------------
```

Script 15: Histograms of rasters

```
### Histogtam of raster data for chapter ### -----------------------
library(viridis)
library(raster)
library(tidyverse)
library(sf)
library(stars)
# get correct rasters that are normalised
prediction_SR_raster_normalized <- RASTER_SR_Normalised # SR
prediction_SPRUE_raster_normalized #SRPUE
prediction_SSDM_raster_normalized <- rasterFromXYZ(cbind(ssdm_clean$x, ssdm_clean$y,
ssdm_clean$normalized))# SSDM
prediction_SR_raster_normalized$normalized
prediction_SR_raster_normalized <- rasterFromXYZ(SP_OFFSET_df_Normalized, )
#Convert first two columns as lon-lat and third as value
prediction_SPRUE_raster_normalized
prediction_SSDM_raster_normalized
dev.off()
hist(prediction_SR_raster_normalized$normalized, col = 'red')
hist(prediction_SPRUE_raster_normalized, col='green', add=TRUE)
hist(prediction_SSDM_raster_normalized, col='blue', add=TRUE)
[1] "#CB627B" "#D17287" "#D68193" "#DB909F" "#E09FAB" "#E4ADB7" "#E7BCC4" "#EBCBD1" "#EDDADD"
"#F0E9EA" "#E5ECEC"
[12] "#CFE3E3" "#B7D9DA" "#9FD0D2" "#84C7C9" "#65BEC0" "#3AB5B8" "#00ACB0" "#00A4A7" "#009B9F"
viridis(25)
[1] "#440154FF" "#471164FF" "#481F70FF" "#472D7BFF" "#443A83FF" "#404688FF" "#3B528BFF" "#365D8DFF"
[9] "#31688EFF" "#2C728EFF" "#287C8EFF" "#24868EFF" "#21908CFF" "#1F9A8AFF" "#20A486FF" "#27AD81FF"
[17] "#35B779FF" "#47C16EFF" "#5DC863FF" "#75D054FF" "#8FD744FF" "#AADC32FF" "#C7E020FF" "#E3E418FF"
[25] "#FDE725FF"
library(scales)
# set bar colours
col1 <- "lightblue"
col2 <- "lavender"
col3 <- "#65BEC0"
layout(mat = matrix(c(1,2),2,1, byrow=TRUE),  height = c(2,8))
par(mar=c(0, 5, 1, 2.1)) #bottom, left, top, and right
boxplot(prediction_SR_raster_normalized$normalized, horizontal=TRUE , xaxt="n" , yaxs="i", xaxs = "i", frame=F, col = col1,
outcol = alpha("black", 0.1), pch=16)
par(mar=c(5, 5, 1, 2.1)) #bottom, left, top, and right
hist(prediction_SR_raster_normalized$normalized,
    main = "Species Richness",
    xlab = "Normalized Richness",
    ylab = "Frequency",
    border=T,
    col = alpha(col1), pch=16,
    yaxs ="i",
    xaxs = "i",
    ylim = c(0,5000))
dev.off()
#SRPUE
layout(mat = matrix(c(1,2),2,1, byrow=TRUE),  height = c(2,8))
par(mar=c(0, 5, 1, 2.1))
boxplot(prediction_SPRUE_raster_normalized , horizontal=TRUE , xaxt="n" , yaxs="i", xaxs = "i", frame=F, col = col2, outcol
= alpha("black", 0.1), pch=16)
par(mar=c(5, 5, 1, 2.1)) #bottom, left, top, and right
hist(prediction_SPRUE_raster_normalized,
    main = "SRPUE",
    xlab = "Normalized Richness",
    ylab = "Frequency",
    border=T,
    col = alpha(col2), pch=16,
    yaxs ="i",
    xaxs = "i")
#SSDM
layout(mat = matrix(c(1,2),2,1, byrow=TRUE),  height = c(2,8))
par(mar=c(0, 5, 1, 2.1))
boxplot(prediction_SSDM_raster_normalized , horizontal=TRUE , xaxt="n" , yaxs="i", xaxs = "i", frame=F, col = col3, outcol
= alpha("black", 0.1), pch=16)
par(mar=c(5, 5, 1, 2.1)) #bottom, left, top, and right
hist(prediction_SSDM_raster_normalized,
```

```
     main = "SSDM",
     xlab = "Normalized Richness",
     ylab = "Frequency",
     border=T,
     col = alpha(col3), pch=16,
     yaxs ="i",
     xaxs = "i")
# Non normalised
#Rasters:
prediction_SR_raster_normalized$layer
prediction_SR_raster
prediction_SPRUE_raster
ssdm_richness
layout(mat = matrix(c(1,2,3), nrow = 3, ncol = 1, byrow=TRUE),  height = c(1,1,10))
par(mar=c(0, 5, 0, 2.1)) #bottom, left, top, and right
boxplot(prediction_SR_raster , horizontal=TRUE , xaxt="n" , yaxs="i", xaxs = "i", frame=F, col = col1, outcol = alpha("black", 0.7), pch=16, ylim = c(0,42))
boxplot(ssdm_richness , horizontal=TRUE , xaxt="n" , yaxs="i", xaxs = "i", frame=F, col = col3, outcol = alpha("black", 0.7), pch=16, ylim = c(0,42))

par(mar=c(5, 5, 0, 2.1)) #bottom, left, top, and right
hist(ssdm_richness,
     main = "",
     xlab = "Species Richness",
     ylab = "Frequency",
     border=T,
     col = alpha(col3, 0.7),
     yaxs ="i",
     xaxs = "i",
     ylim = c(0, 3500),
     xlim = c(0,42),
     add = F,
     breaks = 30,
     cex.axis = 1.5,
     cex.lab = 2)
hist(prediction_SR_raster,
     main = "Species Richness",
     xlab = "",
     ylab = "",
     border=T,
     col = alpha(col1, 0.8),
     yaxs ="i",
     xaxs = "i",
     add = T, breaks = 30)
dev.off()
### End of Script ###------------------------------------------------------------------------------
```

Script 16: SSDM

```
#load required packages
library(SSDM)
library(raster)
library(rgdal)
library(sf)
library(dplyr)

setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/SSDM/SCW_SSDM/Test_1")

path <- getwd()

Env_Vars_minusPP <-
  load_var(
    path = path,
# set in WD code above
    files = c(
      "1_Bathymetry.tif",
      "2_Mean_CHL.tif",
      "3_Mean_Fronts.tif",
      "4_Distancetoland.tif",
      "5_SST_Mean.tif",
      "6_MeanSalinity.tif",
      "7_Mean_MixedLayerThickness.tif",
      "9_Slope.tif"
    ),
    # If set Null - brings in all from folder.
    format = ".tif",
    categorical = NULL,
    # Change if any are Categorical
    tmp = FALSE,
    # temp file to avoid memory overload
    verbose = TRUE,
    Norm = FALSE,
    # Do not want to normalise between 0 and 1
  )

Occ_full <- load_occ(
  path = path,
  Env_Vars,
  # This is from Environmental Variables above, so must be done before this stage
  Xcol = 'LONGITUDE',
  Ycol = 'LATITUDE',
  Spcol = "NAME",
  file = 'occ_full.csv',
  GeoRes = FALSE,
  sep = ',',
  verbose = FALSE
)
# Create list of species names
list_of_species <- split(Occ_full, Occ_full$NAME)
# NOTE - Test with small list IF needed before running full species list e.g.
temp <- subset(Occ_full, subset = Occ_full$NAME %in% c("BASKING SHARK", "BLUE SHARK", "BOTTLENOSE
DOLPHIN"))
# Put this into output bit below instead of 'list_of_species'
SSDM_minusPP <- stack_modelling(
  c('GAM'),
  Occ_full,
  Env_Vars_minusPP,
  rep = 1,
  Xcol = 'LONGITUDE',
  Ycol = 'LATITUDE',
  Spcol = 'NAME',
  method = "pSSDM",
  verbose = TRUE,
  uncertainty = FALSE,
```

```
  cores = 7,
  save = TRUE,
  ensemble.thresh = 0
)
# save.stack(SSDM, name = "SSDM", path = getwd(), verbose = FALSE, GUI = FALSE)

load_stack(name = "SSDM_minusPP", path = getwd(), GUI = FALSE)

plot(SSDM_minusPP)
### -End of Script ###--------------------------------------------------------------------------------
```

Script 17: Variable importance from SSDM

```
### Var Imp by splits ### ----------------------------------------------------------------
# Used to look at variable importance by splits such as taxa, feeding strategy etc

# Load required packages
library(dplyr)
library(cowplot)

# Bring in data ### ----------------------------------------------------------------
setwd("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/SSDM/SCW_SSDM/Test_1/Data/Var Importances")
#Get a list with all csv files from the directory that is set as 'working directory'
filelist = list.files(pattern="*.csv$")
#reading in csv files can also be done using the base R function read.csv(), without needing to load package "data.table":
df_input_list <- lapply(filelist, read.csv)
#get the filenames, remove extension for use as "id"
names(df_input_list) <- gsub(filelist, pattern="\\..*", replacement="")
# Merge all the dataframes and use the filenames as id
varImp_All <- bind_rows(df_input_list, .id = "SPECIES")
varImp_All <- varImp_All[-c(2)] # Remove the column with axes evaluation in it
head(varImp_All)
write.csv(varImp_All, "/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/SSDM/SCW_SSDM/Test_1/Data/varImp_All.csv")
# Fill in taxa in excel

### Load file with taxa now filled in ### ----------------------------------------------------------------
varImp_All <- read.csv("/Volumes/#1 PhD - Working Seagate/PhD/R Stats/PhD Analysis/2.
Environmental_drivers/SSDM/SCW_SSDM/Test_1/Data/varImp_All.csv", stringsAsFactors=TRUE)
head(varImp_All)
varImp_All <- varImp_All[-c(1)] # Remove number column
head(varImp_All)

library(viridis)
library(RColorBrewer)
display.brewer.all(colorblindFriendly = TRUE)
# https://stackoverflow.com/questions/17721126/simplest-way-to-do-grouped-barplot
library(ggplot2)
ggplot(varImp_All, aes(factor(Taxa), X1_Bathymetry, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()
# 9 x 9 plot
library(ggpubr)
var1 <- ggplot(varImp_All, aes(factor(Taxa), X1_Bathymetry, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = "", y = "Bathymetry (m)")
var2 <- ggplot(varImp_All, aes(factor(Taxa), X2_Mean_CHL, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = "", y = "Chl A")
var3 <- ggplot(varImp_All, aes(factor(Taxa), X3_Mean_Fronts, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = "", y = "Fronts")
var4 <- ggplot(varImp_All, aes(factor(Taxa), X4_Distancetoland, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
```

```
        panel.grid.minor = element_blank())+
  labs(x = "", y = "Distance to Land")
var5 <- ggplot(varImp_All, aes(factor(Taxa), X5_SST_Mean, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = "", y = "SST")
var6 <- ggplot(varImp_All, aes(factor(Taxa), X6_MeanSalinity, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = "", y = "Mean Salinity)")
var7 <- ggplot(varImp_All, aes(factor(Taxa), X7_Mean_MixedLayerThickness, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = "", y = "Mixed Layer thickness")
var8 <- ggplot(varImp_All, aes(factor(Taxa), X8_Mean_PP, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = "", y = "Mean PP")
var9 <- ggplot(varImp_All, aes(factor(Taxa), X9_Slope, fill = Taxa)) +
  geom_bar(stat="identity", position = "dodge") +
  scale_fill_brewer(palette = "Paired")+
  theme_linedraw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = "", y = "Slope")

plot_grid(var1, var2, var3, var4, var5, var6, var7, var8, var9 + rremove("x.text"),
          ncol = 3, nrow = 3)
### End of Script ###-----------------------------------------------------------------------------
```