



Data-driven decadal climate forecasting using Wasserstein time-series generative adversarial networks

Ahmed Bouteska¹ · Marco Lavazza Seranto² · Petr Hajek³ ·
Mohammad Zoynul Abedin⁴

Received: 6 April 2023 / Accepted: 26 October 2023
© Crown 2023

Abstract

Recent trends in global climate modeling, coupled with the availability of more fine-scale datasets, have opened up opportunities for deep learning-based climate prediction to improve the accuracy of predictions over traditional physics-based models. For this, however, large ensembles of data are needed. Generative models have recently proven to be a suitable solution to this problem. For a sound generative model for time-series forecasting, it is essential that temporal dynamics are preserved in that the generated data obey the original data distributions over time. Existing forecasting methods aided by generative models are not adequate for capturing such temporal relationships. Recently, generative models have been proposed that generate realistic time-series data by exploiting the combinations of unsupervised and supervised learning. However, these models suffer from instable learning and mode collapse problems. To overcome these issues, here we propose Wasserstein Time-Series Generative Adversarial Network (WTGAN), a new forecasting model that effectively imitates the dynamics of the original data by generating realistic synthetic time-series data. To validate the proposed forecasting model, we evaluate it by backtesting the challenging decadal climate forecasting problem. We show that the proposed forecasting model outperforms state-of-the-art generative models. Another advantage of the proposed model is that once WTGAN is tuned, generating time-series data is very fast, whereas standard simulators

✉ Mohammad Zoynul Abedin
m.z.abedin@swansea.ac.uk

Ahmed Bouteska
ahmedcbouteska@gmail.com

Marco Lavazza Seranto
marcolaser@gmail.com

Petr Hajek
petr.hajek@upce.cz

¹ Faculty of Economics and Management of Tunis, University of Tunis El Manar, Tunis, Tunisia

² Ca' Foscari University of Venice, 3246 Dorsoduro, 30123 Venice, VE, Italy

³ Faculty of Economics and Administration, Science and Research Centre, University of Pardubice, Studentska 95, Pardubice 53210, Czech Republic

⁴ Department of Accounting and Finance, School of Management, Swansea University, Bay Campus, Fabian Way, Swansea SA1 8EN, Wales, UK

consume considerable computer time. Thus, a large amount of climate data can be generated, which can substantially improve existing data-driven climate forecasting models.

Keywords Forecasting · Climate · Deep learning · Time series · Generative adversarial learning

1 Introduction

Effective long-term climate forecasting has become a central issue in climate change detection. Indeed, there is an increasing need for accurate long-term climate predictions because climate forecasts have become a critical factor for strategic decisions of business, governments, and socio-economic sectors vulnerable to climate changes (Kumar et al., 2021; Lemos & Rood, 2010; Lomborg, 2020). For example, agriculture is highly sensitive to climate variations with serious impacts on agricultural production and water resources (Lu et al., 2019), and climate uncertainty typically results in less productive conservative agricultural strategies in an effort to reduce the climate risk (AitSahlia et al., 2011; Jones et al., 2000; Yerlikaya et al., 2020). Furthermore, effective climate forecasts play a key role in future energy demand estimates as energy use is exposed to climate changes (Franco & Sanstad, 2008; Sadefo Kamdem et al., 2023; Shahzad et al., 2023; Van Ruijven et al., 2019). That is why focus on forecasting climate change and its impact intensified during the last decade (Sarin et al., 2020).

The foremost focus of climate research is climate prediction over a timescale horizon of about ten years. Decadal forecasting includes both annual and multi-year (decadal) predictions (Boer et al., 2016). Decadal variations in climate span the gap between short-term climate forecasts, in which initial conditions play a key role, and long-term climate change forecasts, in which the predominant factor is external forcing. Hence, this requires taking into account all uncertainties that arise from the interaction between external climate impacts and the ongoing inherent variation in climate. Decadal climate forecasts have revived interest in ocean circulations and processes associated with ocean and atmosphere, as understanding these has been shown to be the keystone to achieving successful decadal predictions of climate. Long-term climate forecasting models have therefore received much attention in the past decade to deliver future climate scenarios. However, long-term climate forecasting is a challenging issue as it retains a high level of complexity and uncertainty in the mechanism behind the changes in the climate. In the long term, climate forecasting models also struggle with the problem of behavioural and technological uncertainty (Van Ruijven et al., 2019). Moreover, these problems vary across temporal and geographical scales (Lemos & Rood, 2010).

Decadal climate forecasts currently employ the very latest linked climate and Earth system models that address the underlying system of biogeochemical and physical equations that drive climate. The intrinsic complexity of such instruments poses an inherent challenge for decadal projections of climate, which is the trade-off between the vast amount of simulations needed to provide accurately probabilities related to different plausible decadal climate fluctuations, and the demanding computational needs of climate models. For example, increases in the horizontal resolution and the number of vertical levels of the Earth system model leads to an increased demand for computing capacity (Müller et al., 2018).

Existing research in climate forecasting tends to use either physics-based computational models or data-driven models (Meng et al., 2023; Scher, 2018). As climate-related databases grow in Earth system models, data-driven machine learning models have become particularly

relevant (Rolnick et al., 2022). Nevertheless, recent advances in long-term climate forecasting have shown that large ensembles of forecasting models are needed to achieve accurate forecasts (Smith et al., 2019). This brings with it requirements to acquire/generate sufficient data for these model ensembles. Indeed, when it comes to representing global climate trends, it is challenging to develop sound forecasting models due to greater complexity and data constraints. This study seeks to address this problem with the purpose of enabling the training of powerful and efficient deep learning-based models that can generate decadal climate forecasts in rapid time with a time-series distribution of data close to that supplied as training data to generative adversarial networks (GANs). Although GANs provides an effective solution to problems where synthetic data are required, it fails to take into account the temporal dynamics in the data. To overcome this problem, Yoon et al. (2019) introduced Time-series.

GANs (TimeGAN), a GAN model that can be explicitly trained to retain the temporal dynamics of time-series data. As a result, realistic time-series data for various forecasting problems can be generated. Both GANs and TimeGANs suffer from the same problems associated with the cross-entropy loss function, which can result in mode freezing (collapse) and the null gradient problem with convergence. Inspired by recent developments in cost functions for GANs, here we propose a forecasting model that we call Wasserstein Time-Series GAN (WTGAN), which not only effectively generates realistic time-series data by modelling the dynamics of the original climate data but it is also less susceptible to mode collapse and converges quickly to the desired solution. In summary, the contributions of this study are threefold:

- A novel WTGAN forecasting model is proposed that exploits the generated realistic time-series data, which allows us to accurately model the temporal features in complex dynamic systems while improving GAN's learning capacity.
- Backtesting of the decadal climate forecasting problem is used to validate the proposed model. We demonstrate that the state-of-the-art GAN-based forecasting models fail to adequately model the underlying dynamics in the climate data.
- We show that the trained WTGAN model can quickly generate large amounts of realistic climate time-series data, thus substantially outperforming existing Earth system simulation models in terms of computational time.

The rest of the manuscript is structured as follows. Section 2 briefly presents the existing deep learning approaches to data-driven climate forecasting. Section 3 outlines the research methodology, introduces the proposed generative model for time-series data, and describes the data used for climate forecasting. Section 4 presents the experimental setup and forecasting results. The obtained forecasts are discussed in Sect. 5, and Sect. 6 concludes with future research suggestions.

2 Related literature

To provide guidance to government decisions and estimate the risks associated with climate change, physics-based climate models are widely used, such as Earth system models and General circulation models (Rolnick et al., 2022). To address complex climate problems, these models rely on data simulations grounded in the laws of physics. However, the simulation of the dynamics of climate data is a complex task with a wide range of uncertain environmental factors, which in turn limits the accuracy of predictions based on these numerical models.

Unlike physics-based models, data-driven models fully rely on observational data with the aim of learning their underlying distribution (Meng et al., 2023). Even though the physical

interpretation of the data-driven model is problematic, data-driven models such as deep learning-based models advance the state-of-the-art in many climate forecasting problems. This is made possible by the ever-increasing data from climate observations. A major obstacle of existing simulation-based climate forecasting models is their computationally intensity. In contrast, data-driven climate forecasting models are not only fast to train but more importantly quickly provide predictions in real deployments.

Climate time-series forecasting presents several unique challenges and characteristics that distinguish it from other types of time-series data. Climate systems exhibit multivariate, spatial and long-term dependencies, requiring forecasting models to capture and maintain temporal coherence over extended durations (Ardabili et al., 2019). Climate systems are also susceptible to extreme events and influenced by various forcings, such as solar radiation and greenhouse gas emissions, which can lead to feedback loops and amplify or dampen changes (AitSahlia et al., 2011). Moreover, climate data can be sparse, especially in historical records (Ham et al., 2019). Forecasting models need to handle missing data and fill in gaps appropriately. Additionally, historical climate data might not fully capture the true range of climate variability due to limitations in measurement and recording. These peculiarities arise from the complex and interconnected nature of the Earth's climate system.

Deep learning-based forecasting models were applied successfully to various fields, including technology forecasting (San Kim and So, 2020), macroeconomic activity (Jabeur et al., 2022), and Earth system science (Reichstein et al., 2019). Indeed, the state-of-the-art performance was improved in many domains due to the capacity of multiple layer architectures for learning data representations at different levels of abstraction. Recent applications of deep learning also cover climate forecasting. For example, Rasp et al. (2018) proposed a deep learning-based model to exploit the performance of cloud simulations for climate modelling. Their data-driven model was fast and highly accurate, including the predictions of extreme events such as tropical waves and precipitation extremes. The potential of the deep learning-based model for forecasting global weather was examined by Dueben and Bauer (2018). This was justified by the enormous availability of data, the computational efficiency of deep learning-based models, and the shortcomings of existing climate models, especially with regard to model complexity and resolution. The deep learning-based machine learning methods have also been successfully used to predict extreme climate events in historical observations (Racah et al., 2017). Exascale-class deep learning architectures were also developed to efficiently scale deep learning in detecting extreme climate events in multivariate data with high resolution (Kurth et al., 2018). The main advantages of machine learning and deep learning in modelling climate change were identified in a systematic review of Ardabili et al. (2019): high accuracy, low computational cost, and robustness. Hybrid forecasting models that combine deep learning with traditional machine learning methods have also been recently used to enhance the performance in climate prediction tasks (de Mattos Neto et al., 2022). Deep learning and artificial intelligence in general are also making an important contribution to adaptation to climate change (Leal Filho et al., 2022).

Despite the above advantages of data-driven models, many climate forecasting problems struggle with limited data because the Earth generates only annual climate observations in each year, regardless of the number of sensors deployed (Rolnick et al., 2022). As a result, the climate data are only available for several decades. Another problem arises when we focus on the quality of the available data. Moreover, biases may exist between different observation systems and the data-driven forecasting models should also be robust against faulty and missing data (Dueben & Bauer, 2018). In this regard, GAN-based forecasting models were identified as a promising direction toward producing a stochastic deep learning-based parameterization that can effectively capture the data variability (Rasp et al., 2018). Indeed,

GAN models have gained prominence in the field of operations research due to their versatile capabilities and potential to address important operations research challenges, such as data augmentation and scenario generation. In operations research, where data availability can be limited, GAN-generated data can help augment datasets, leading to better model training and more accurate analyses (Gokasar et al., 2023). For example, GANs can simulate and generate data related to supply chain dynamics, aiding in inventory management, demand forecasting, and optimization of distribution networks (Wang et al., 2023). In addition, GANs can generate multiple scenarios or variations of data, which can be particularly useful for risk assessment and optimization under uncertainty (Zhao et al., 2022). GANs can model complex interactions between resources and constraints, offering insights into optimal resource allocation strategies that might not be readily apparent through traditional methods (Hua et al., 2019; Wang et al., 2021). GANs can also be used to learn the normal behavior of complex systems and then identify anomalies or deviations from this behavior (Shao et al., 2019; Yan, 2021). This is crucial for detecting credit risk, fraud, faults, or irregularities in various operations research contexts (Bui, 2023; Fiore et al., 2019; Gülmez, 2023; Zhang et al., 2022). In this study, we leverage these advantages of GANs and propose a forecasting model that is specifically designed for time-series data.

3 Methodology

This section first introduces the proposed WSGAN model, which aim is to build new realistic time-series instances for climate forecasting. Then, the data from the decadal climate prediction project are described.

3.1 Wasserstein time-series generative adversarial network

A GAN comprises mainly of two components: (1) a generator learning to produce credible data, and (2) a discriminator learning to discriminate between the fake generator data and the real original data. At the start of training, the generator creates data that is random and thus clearly fake. With training, the generator increasingly moves towards providing outputs that may fool the discriminator. The generator and discriminator are both represented by neural networks, with the generator being connected straight to the input of the discriminator. The classification of the discriminator yields an output that the generator applies to adjust the neural networks' weights via backpropagation. The flowchart is depicted in Fig. 1.

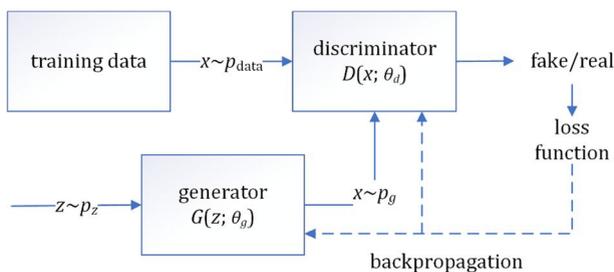


Fig. 1 Flowchart of GAN

For training, the positive instances are represented by real data, while the negative instances are those from the generator. While training the discriminator, the generator is kept unaltered, i.e., its neural network's weights remain unchanged, whilst producing the data that the discriminator is trained on. The loss function of the discriminator penalizes the discriminator for the errors produced.

The generator gets trained to produce false instances that can mislead the discriminator. The generator consists of a random noise generator and a generator neural network that converts the random noise taken at the input into data instances. It has been experimentally found that the distribution of the noise source does not matter much, so a uniform distribution is typically assumed. As training continues, the generator becomes increasingly capable of creating data instances similar to the real ones, and the discriminator starts making more errors.

However, the task of a forecasting model is not only to capture the feature distribution at each time point, but it also needs to capture the complex dynamics of these features over time. To address this issue, Yoon et al. (2019) introduced TimeGAN, which seeks to precisely represent the underlying conditional distribution of temporal transitions. More precisely, TimeGAN makes use of the autoregressive decomposition of the joint distribution to target the conditionals, leading to the complementary and more easily achieved goal of learning the original data density. This decision implies two different objectives. The first objective is the distance between both distributions and it is considered to be global. The other objective is a local measure of the distance between the original and the generated sequence at time t . From there, we set the objective that combines traditional GAN (linked to the first objective, which is configured as the Jensen-Shannon divergence) and a supervised learning of the second objective through the maximum-likelihood (configured as the Kullback-Leiber divergence).

The TimeGAN structure consists of the autoencoding components (represented by embedding and recovery functions) trained jointly with the adversarial components (sequence generator and sequence discriminator). In other words, TimeGAN concurrently performs feature encoding, generation of representations, and iterates over time. Thus, the embedding network supplies a latent space, while the adversarial network is operating in this space. A supervised loss is used to synchronize the latent dynamics of synthetic and data. Figure 2 presents the block scheme of TimeGAN.

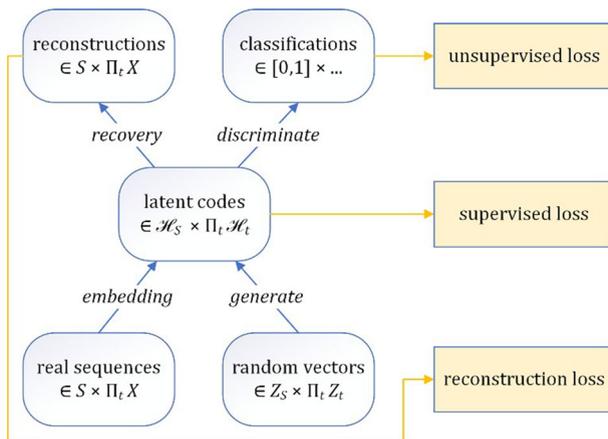


Fig. 2 Block scheme of TimeGAN

Let us briefly introduce the four parts of TimeGAN. The embedding and recovery components are represented by autoencoders, which are used to facilitate the mapping of latent spaces \mathcal{H}_S and \mathcal{H}_X representing static features S and temporal features X , respectively, to allow the adversarial network to learn temporal data patterns by reducing dimensionality. The embedding function e transforms the features S and X to their latent codes $h_s = e_S(S)$ and $h_t = e_X(h_s, h_{t-1}, x_t)$, where s and x represent specific values of S and X . The recovery function works in the opposite direction, that is $\tilde{s} = r_s(h_s)$ and $\tilde{x} = r_x(h_t)$.

The generator and discriminator work on both static and dynamic data. The generating function g transforms static and temporal data to synthetic latent codes $\tilde{h}_s = g_S(z_s)$ and $\tilde{h}_t = e_X(\tilde{h}_s, \tilde{h}_{t-1}, z_t)$, respectively. Finally, the discriminator performs the classification on both real and synthetic data with the following outputs: $\tilde{y}_s, \tilde{y}_t = (\tilde{h}_s, \tilde{h}_t)$.

First, the reconstruction loss is used to train the embedding and reconstruction networks:

$$\mathcal{L}_R = \mathbb{E}_{S, X_t \sim p} \left[\|s - \tilde{s}\|_2 + \sum_t \|x - \tilde{x}\|_2 \right] \tag{1}$$

where p is a joint distribution.

Second, the unsupervised loss is defined, which maximizes the classification performance of the discriminator:

$$\mathcal{L}_U = \mathbb{E}_{S, X_t \sim p} \left[\log y_s + \sum_t \log y_t \right] + \mathbb{E}_{S, X_t \sim \hat{p}} \left[\log(1 - \hat{y}_s) + \sum_t \log(1 - \hat{y}_t) \right] \tag{2}$$

To achieve greater consistency with conditional data distributions, an additional supervised loss is implemented to train GAN in supervised mode, where the generator is given sequences drawn from the embedded real data:

$$\mathcal{L}_U = \mathbb{E}_{S, X_t \sim p} \left[\left\| \sum_t h_t - g_x(h_s, h_{t-1}, z_t) \right\|_2 \right] \tag{3}$$

In the course of training, the difference is evaluated between the latent vector of the embedding network and the latent vector synthetically produced by the generator. Whilst \mathcal{L}_U forces the generator to produce realistic time-series data, \mathcal{L}_S makes sure that similar stepwise transitions are created. The parameters of the embedding and recovery networks θ_e and θ_r are trained by minimizing the reconstruction and supervised losses:

$$\min_{\theta_e, \theta_r} (\lambda \mathcal{L}_S + \mathcal{L}_R) \tag{4}$$

where $\lambda \geq 0$ denotes a hyperparameter used to balance the losses. Adversarial training is applied to optimize the parameters θ_g and θ_d of the generator and discriminator networks:

$$\min_{\theta_g} \left(\eta \mathcal{L}_S + \max_{\theta_d} \mathcal{L}_U \right) \tag{5}$$

where η represents another hyperparameter balancing supervised and unsupervised losses. Practically speaking, TimeGAN does not seem to be particularly hyperparameter-sensitive; therefore, in agreement with Yoon et al. (2019), the values for the hyperparameters were set as follows: $\lambda = 1$ and $\eta = 10$.

Considering the classification problem performed within the TimeGAN (fake/real data classification), the cross-entropy loss function is typically used. To solve the problem of low or zero gradients, here we propose to use the Wasserstein loss function in the TimeGAN model. The Wasserstein loss function is grounded on the distance of ground motion between

two probability distributions. Therefore, WTGANs are enhanced TimeGANs with the loss function.

\mathcal{L}_U (Eq. (2)) represented by the Wasserstein loss function defined as follows:

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in (P_r, P_g)} \mathbb{E}_{(x,y) \sim \gamma} \|x - y\| \quad (6)$$

where \mathbb{P}_r denotes the real data distribution, \mathbb{P}_g is the generated data distribution, both are marginals of joint distributions $\gamma(x, y)$. If we transform Eq. (6) using the Kantorovich-Rubinstein duality and consider x and y to be K-Lipschitz functions f_w , then the following objective function is obtained:

$$\max_{w \in W} \mathbb{E}_{x \sim \mathbb{P}_r} [f_w(x)] - \mathbb{E}_{z \sim p(z)} [g_\theta(z)] \quad (7)$$

Compared to GANs and TimeGANs, WTGANs are less prone to freezing than competing models while avoiding zero gradient problems. Unlike cross-entropy, the Wasserstein loss function also offers the virtue of providing a true metric of distance within the probability distribution space. The Wasserstein loss function may be realized through the calculation of the mean by multiplying the predicted score for each instance by the corresponding true label (element-wise).

3.2 Data

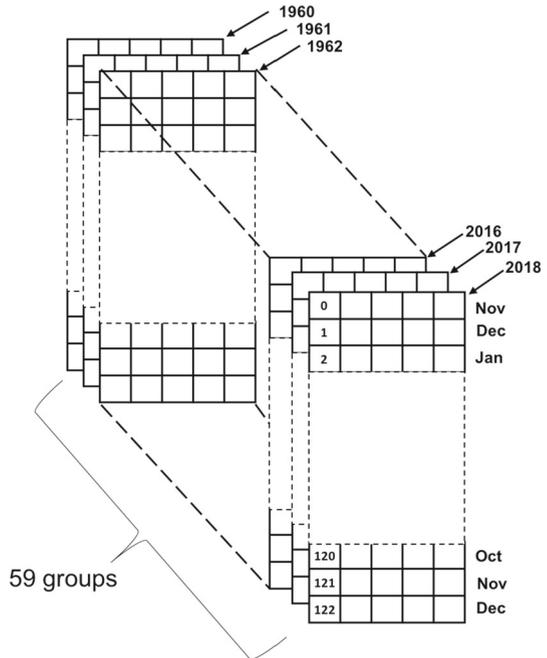
Various ocean measurement systems have been used to collect data on its temperature, oxygen content, salinity, and so on at the surface and at depth. Additional and more advanced sensors enable us to gather data on wave height, traces of oil or chlorophyll concentrations. This plethora of data adds up to build the dataset needed to simulate models of the system that are helpful in producing long-term forecasts.

The data for this study were obtained from the MiKlip preoperational system for decadal climate projections (Marotzke et al., 2016). The MiKlip system builds on the high-resolution Earth system model of the Max Planck Institute (MPI-ESM1.2-HR) (Müller et al., 2018), a conglomerate of coupled sub-models for soil and vegetation and a general circulation model. In the MPI-ESM1.2-HR, the general circulation model of the atmosphere exploits a horizontal resolution of T127/100 km and 95 hybrid sigma pressure levels that reach up to 0.01 hPa, and the ocean and sea ice model is equipped with a tripolar grid with a global resolution of 0.4° allowing eddies with 40 z-levels. MPI-ESM-HR and other versions of MPI-ESM were extensively validated in investigations of climate dynamics (Müller et al., 2018; Pohlmann et al., 2019).

We used the MiKlip simulations contributing to the decadal climate prediction project (DCPP) (Boer et al., 2016). The simulations incorporate historical assimilation data spanning 1958/11-2018/11 and hindcasts initially performed on November 1 for every year between 1960 and 2018 employing historical forcing. One major benefit of decadal climate projections is that their validity can be evaluated by looking at retrospective forecasts (the so-called hindcasts) and comparing them with later observations (Smith et al., 2019). Each hindcast in our data consists of five ensemble elements (r_1, r_2, \dots, r_5) with different initial states. The data can be found in the Earth System Grid Federation.¹ The AMV index was computed from the monthly Max Planck Institute Ocean Model Output as spatial averages of North Atlantic sea-surface temperature in the region covering 0–60°N latitude and 0°–80°W longitude. Prior to

¹ <https://confluence.ecmwf.int/display/COPSRV/CMIP6%3A+Decadal+climate+predictions>.

Fig. 3 Input data structure



the index being calculated, biases in the model, such as drift, were systematically eliminated from the grid-point data following DCPD recommendations.

The data supplied are illustrated as fifty-nine 122×5 matrices, as shown in Fig. 3. Each matrix is composed of five simulations of sea temperature in the North Atlantic. Each column represents the monthly average temperatures over 10 years starting from November of that year. That is, the first matrix carries five simulations ranging from November 1960 to December 1970, the second one covers November 1961 to December 1971 and so on until the last one, which covers 2018 to 2028.

Temperatures ranged from $17.86 \text{ }^\circ\text{C}$ to $24.27 \text{ }^\circ\text{C}$, with the average and standard deviation of $20.88 \pm 1.62 \text{ }^\circ\text{C}$. The original data were normalized and standardized as detailed in the following section.

4 Experiments

4.1 Experimental setup

The experiments were conducted using a MacBook pro laptop running macOS High Sierra with 8 GB of RAM. To train the WTGAN model, 1000 iterations were executed. Initially, tests were also run with 2000 and 5000 epochs, without obtaining particular improvements in the results. Each trained model was used to generate 100 instances in a $100 \times 59 \times 122 \times 5$ array. Each model took just over an hour to train, while creating 100 instances from the trained WTGAN model took under a minute.

The original code of the TimeGAN model can be found at the following link: <https://github.com/jsoyon0823/TimeGAN>. The code of the Wasserstein loss function is available at:

<https://github.com/kzkadc/wasserstein-gan>. The proposed model was implemented in Python 3.7 with the settings adopted from Yoon et al. (2019): hidden dimensions = 24, number of layers = 3, activation function = sigmoid, module = GRU (gated recurrent unit). The GRU module was chosen because, unlike LSTM (long short-term memory), it solves the vanishing gradient problem. It is worth noting that we also experimented with different numbers of hidden dimensions, but no improvement in forecasting was made.

4.2 Experimental results

In this section, the experimental results show the effectiveness of the proposed forecasting model compared with existing GAN-based forecasting models.

In the first run of experiments, we demonstrate both (1) the effect of the proposed Wasserstein loss function in the TimeGAN model and (2) the importance of data pre-treatment (normalization/standardization) for the quality of the generated time-series data. To this end, the original data were first used with the same experimental settings as presented in Sect. 4.1 using the cross-entropy loss as in the original TimeGAN model. Figure 4 shows the synthetic data produced with the increasing training time of TimeGAN. The patterns of the real data were captured in about seventy epochs. It is interesting to observe how the model accommodates these patterns gradually.

Figure 5 shows the result of the data processing. As can be seen from Fig. 5a, the generated time-series data do not follow the trend of the real data: the generated data averages are flat and the dynamics of growth are not reproduced. The result suggests that the TimeGAN generator did not converge and was trapped in a local minimum. To overcome this problem, we used the Wasserstein distance as the loss function. To this end, the TimeGAN was enhanced to allow us the use the Wasserstein loss instead of the standard cross-entropy loss function. However, Fig. 5b shows that growth dynamics were not captured in this case either.

This result indicated that the original data (without pre-processing) were not helpful for representing the temporal dynamics of the generated data. Therefore, we further examined two global (on all data columns) pre-processing schemes: (1) unity-based normalization $(x - x_{\min})/(x_{\max} - x_{\min})$ and (2) standardization $(x - \mu_x)/\sigma_x$. Figure 6 shows the result of the experiment for the unity-based normalization. However, the generated data exhibited the same unsatisfactory trend. It is worth noting that similar results were also obtained for the data standardization even with the Wasserstein loss function.

Because even the global normalization failed to provide reasonable results, it was hypothesized that choosing parameters related to the entire dataset might obscure the dynamics

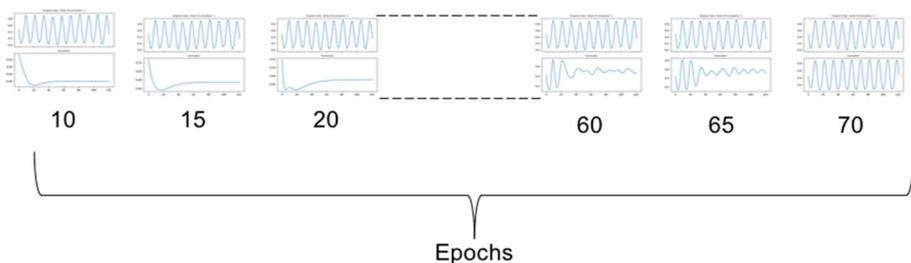


Fig. 4 The effect of the number of epochs on the data generated using TimeGAN. At the top is the trend of the simulation (the first one from 1960), at the bottom is the evolution of the curve generated

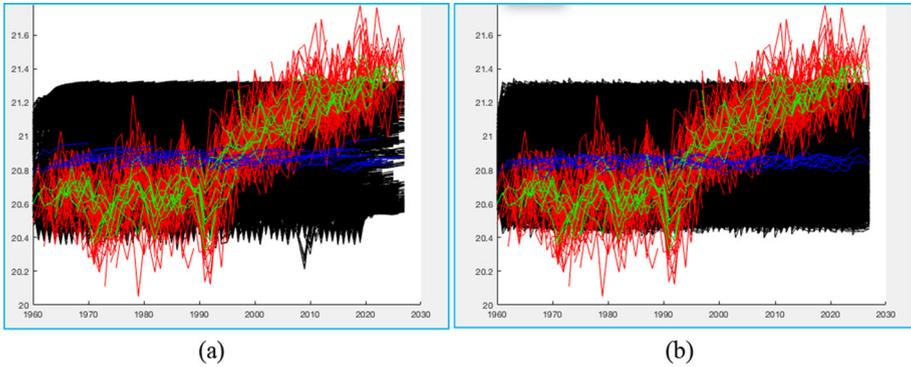


Fig. 5 Results of TimeGAN with original data for (a) cross-entropy loss and (b) Wasserstein loss. The red (black) curves are the average annual values of the real (generated) data, and the green (blue) curves show the average column values of the real (generated). (Color figure online)

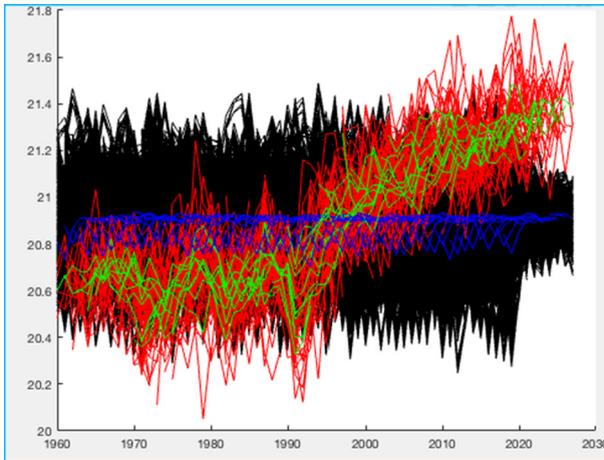


Fig. 6 Results of WTGAN with globally normalized data. The red (black) curves are the average annual values of the real (generated) data, and the green (blue) curves show the average column values of the real (generated). (Color figure online)

of the real data by flattening the results around the normalized values. Two additional data pre-processing schemes were then utilized to normalize and standardize the data locally (in each column). In contrast to the previous results, the locally normalized data fit the dynamics of the real data, with much larger deviations of the mean values (Fig. 7). Here, it can be seen that the TimeGAN with cross-entropy loss produces the data with significantly lower average values than for the real data, while the WTGAN produces a larger amplitude of data averages generated.

In Fig. 8, a different data representation can be seen to demonstrate the decadal forecasting capacity of generated data throughout the hindcasting period. Again, the TimeGAN with Wasserstein loss better simulates the original data, whereas the use of cross-entropy loss functions does not represent the original data distribution.

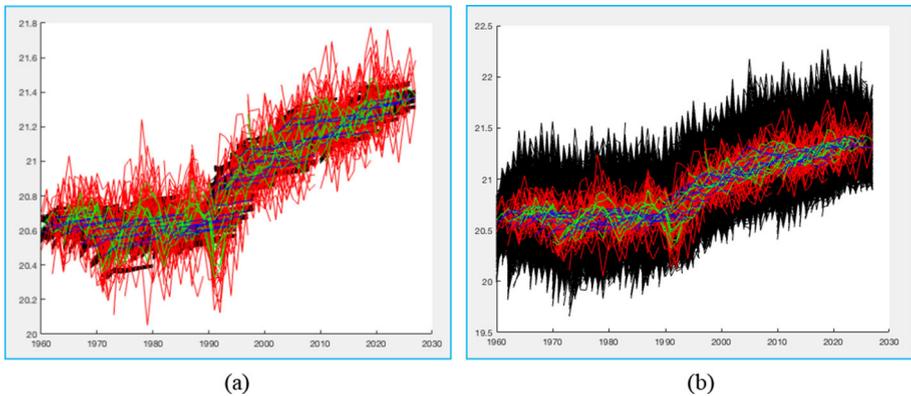


Fig. 7 Results of TimeGAN with locally normalized data for **(a)** cross-entropy loss and **(b)** Wasserstein loss. The red (black) curves are the average annual values of the real (generated) data, and the green (blue) curves show the average column values of the real (generated). (Color figure online)

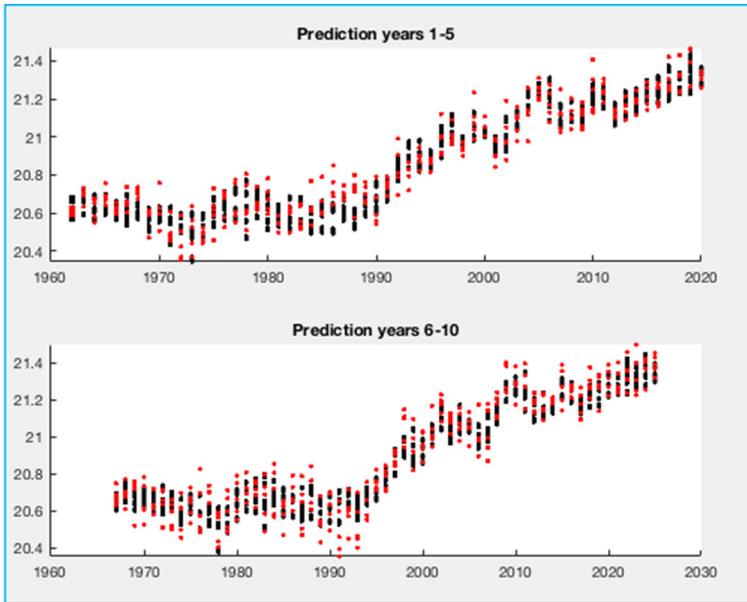
Figures 9 and 10 show the results of WTGAN when the data were standardized locally (by data column). Compared to that obtained for locally normalized data, this result suggests a more reliable data generation process, although in some cases the averages generated are still low.

To better evaluate these results, a violin plot was made. The central year of each treatment is shown on the abscissa; for illustration, the first column in the first chart (years 1960–1964) shows 1962. The results show that the synthetic data are consistent with the dynamics of the real data and retain average values that are close to the real ones.

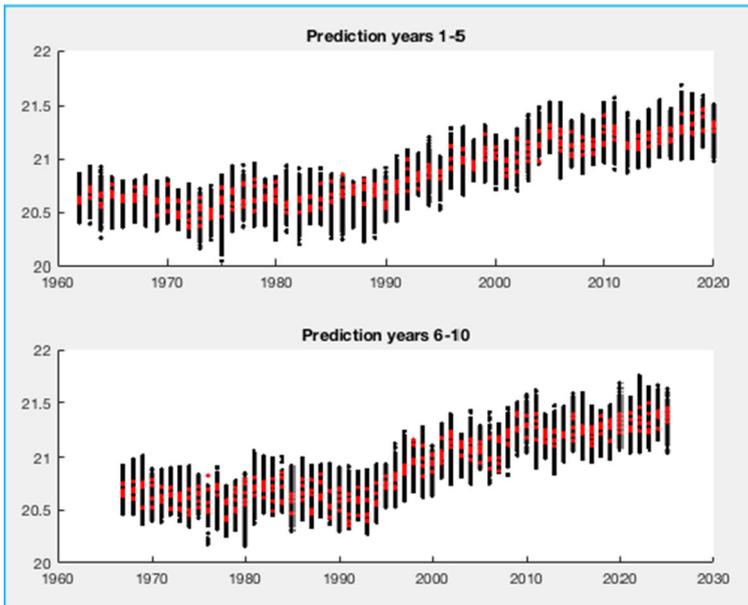
Given that the goal of this study was to generate synthetic data with the same trend and variance as their real counterparts in a short period of time, Table 1 contrasts the computing power used to create the simulations between the proposed WTGAN model and MPI-ESM. The reduction in computing resources is significant. On a conventional computer, approximately 500 simulations can be produced over 30 years in just over 1 h of machine time.

5 Discussion

Prior research has found that the limited amounts of valid observational data is the main limitation of successful application of deep learning-based models in climate forecasting (Ham et al., 2019). In the study conducted by Ham et al. (2019), it was demonstrated that a deep-learning-based statistical forecast model exhibits a noteworthy capability in generating accurate forecasts of the El Niño-Southern Oscillation. To address the inherent constraint of limited observational data, the researchers implemented transfer learning as a strategic approach. This involved the preliminary training of a convolutional neural network on historical simulation data. However, the CMIP5 models have limitations in spatial and temporal resolutions and climate models used in CMIP5 can have biases in simulating certain aspects of the climate. To overcome these problems, GAN models have recently been used to generate synthetic climate data that can be tailored to specific needs, allowing researchers to explore scenarios beyond historical data, especially in regions or time periods with sparse observations, while producing data at higher spatial and temporal resolutions (Kumar et al.,



(a)



(b)

Fig. 8 10-year average results of TimeGAN with locally normalized data for (a) cross-entropy loss and (b) Wasserstein loss. Average values calculated from the real (generated) data columns are shown in red (black). Each column represents ten years. (Color figure online)

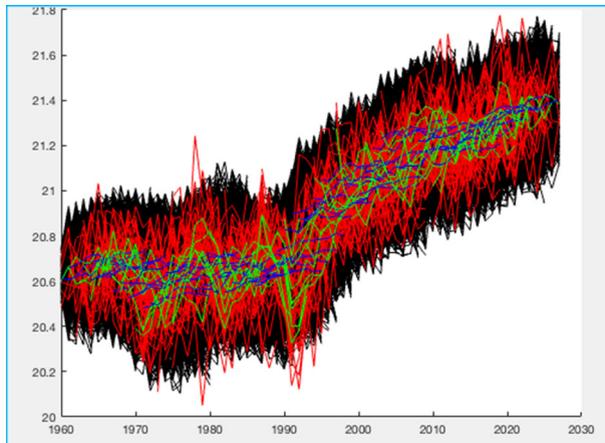


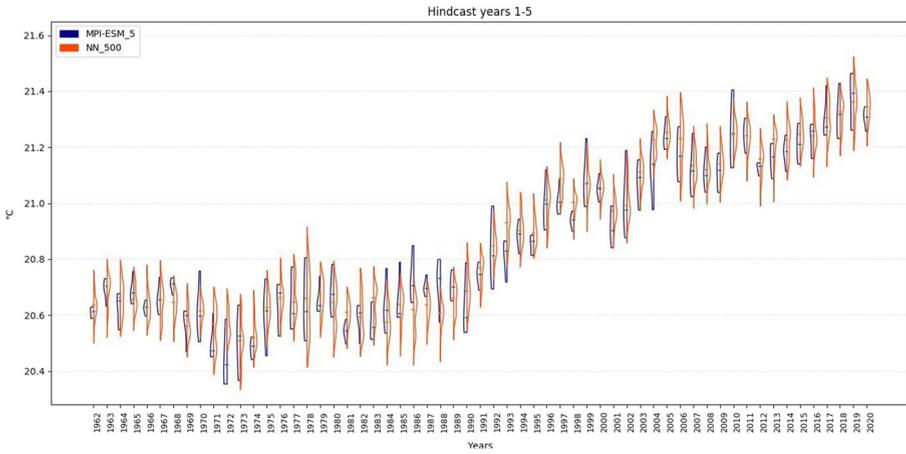
Fig. 9 Results of WTGAN with locally standardized data. The red (black) curves are the average annual values of the real (generated) data, and the green (blue) curves show the average column values of the real (generated). (Color figure online)

2023; Meng et al., 2023). Notably, Meng et al. (2023) demonstrated that the performance of the traditional assimilation models can be significantly improved by GAN-based forecasting models. Specifically, their GAN-based forecasting model was superior to the competing deep learning-based models, namely recurrent neural networks (Jia et al., 2019) and multilayer convolutional LSTM neural networks (Zhang et al., 2020), in the prediction of sea subsurface temperature. Inspired by these findings, this study introduced a novel WTGAN forecasting model, which leverages authentically generated time-series data, allowing us to adeptly characterize the temporal attributes within intricate dynamic systems.

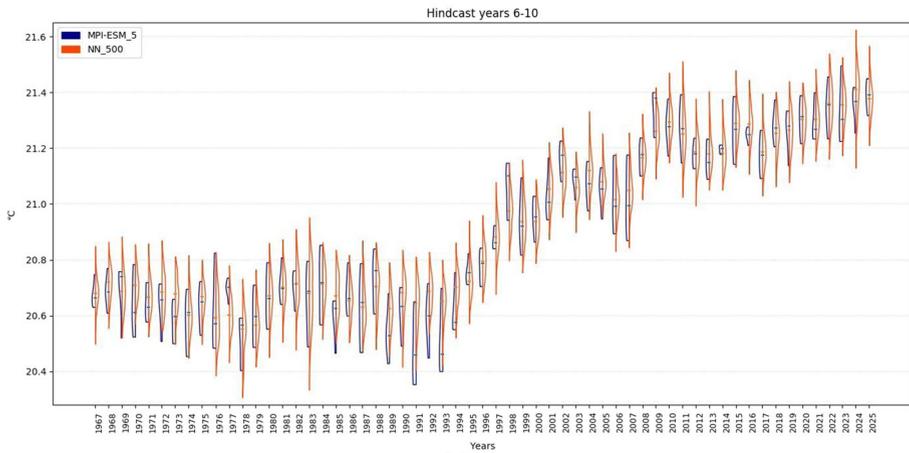
Similar to the above-mentioned deep learning-based data generation approaches, and compared to existing simulation models (de Mattos Neto et al., 2022; Leal Filho et al., 2022; Rasp et al., 2018), the model proposed in this study has the advantage of efficiently generating realistic time-series data, which is particularly convenient in many climate prediction problems with a very limited amount of valid observations (Rolnick et al., 2022).

Furthermore, our results show that the standard GANs, as used in related studies (Meng et al., 2023), are not sufficient to produce realistic time-series climate data because the transition dynamics in the data is difficult to be captured efficiently. Our evidence shows that WTGAN can learn and replicate complex temporal patterns that are inherent in climate time-series data, thereby exhibiting realistic transitions in long-term sequences and maintaining temporal consistency. Our study not only confirms the superiority of the GANs designed specifically for time-series data, but we also demonstrate that the traditional loss function easily leads to the collapse mode and failure of the training algorithm. This is consistent with recent studies (Mi et al., 2023), which demonstrate that the inclusion of the Wasserstein loss function provides both robustness to the model training process and an increased range of output diversity.

Finally, some limitations of the WTGAN-based model should be noted. First, the model might lack the extensive validation and verification that historical simulation data have undergone. This raises questions about the accuracy and reliability of the synthetic data, especially when sufficient amounts of high-quality observational data are not available. Moreover, developing a reliable GAN model for different climate data and forecasting tasks requires careful



(a)



(b)

Fig. 10 Violin plots of the results of WTGAN with locally standardized data. For each year, the distribution of the average values of each simulation are shown, the distribution of the five elements (500 elements) of the real (generated) data are in blue (red). (Color figure online)

Table 1 Machine resources needed to perform climate simulations

Model	Machine	Cores/nodes	Years/computing time
MPI-ESM	Mistral	36	50/day
WTGAN	MacBook pro	2	30 × 5 × 100/h

training, hyperparameter tuning, and validation to ensure that the generated data aligns with the underlying physical processes. Although this study can provide important clues for conducting such experiments, the computational complexity of the models and the expertise required to validate them should also be mentioned, which may hinder their applicability by other researchers and practitioners.

6 Conclusion

This study focused on the rapid generation of synthetic data to improve the forecasting performance of a GAN-based model. To this end, the WTGAN model was proposed, which attempts to replicate the temporal dynamics in the real time-series data. The previously developed TimeGAN model comprises four basic modules, while adopting the cross-entropy loss function. It was considered worthwhile to circumvent the collapse mode by applying the Wasserstein loss function in the discriminator module of TimeGAN rather than the cross-entropy loss function. Additionally, we decided to transform the training data locally on each individual simulation as opposed to globally on the entire dataset. It was precisely the models that used the local standardized data in the WTGAN model.

This work has raised several issues that need further investigation. Continuing this work calls for the generation of multivariate synthetic data covering various physical variables such as salinity and temperature. To this, geolocation can be used to create maps illustrating the climate situation over time. Another issue that needs to be addressed is assimilation with respect to difficult-to-predict extreme events, such as volcanic eruptions that impose significant variations in the data. To this end, the introduction of fixed points corresponding to these events is needed. The WTGAN model should account for these events by forcing the curves to pass through the fixed points and then incorporate them in the generation of synthetic data. Even though these modifications will demand much higher computational resources than those used here, these are still within the capabilities of a computer with an appropriate GPU. We believe that the presented modifications of the GAN model will enable its further use in various prediction problems in operations research. Potential areas of application of the WTGAN model in operations research are multivariate, spatial and non-stationary prediction problems with extreme events that may suffer from sparse data, such as inventory management (Shajalal et al., 2023) and demand forecasting (Efat et al., 2022).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- AitSahlia, F., Wang, C. J., Cabrera, V. E., Uryasev, S., & Fraise, C. W. (2011). Optimal crop planting schedules and financial hedging strategies under ENSO-based climate forecasts. *Annals of Operations Research*, 190, 201–220.

- Ardabili, S., Mosavi, A., Dehghani, M., Várkonyi-Kóczy, A. R. (2019). Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In: International Conference on Global Research and Education, Springer, Cham, pp. 52–62.
- Arjovsky, M., Chintala, S., Bottou, L. (2017). Wasserstein generative adversarial networks. In: International Conference on Machine Learning (ICML), pp. 214–223.
- Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., & Eade, R. (2016). The decadal climate prediction project (dcpp) contribution to cmip6. *Geoscientific Model Development*, 9(10), 3751–3777.
- Bui, A. T. (2023). Root cause analysis of manufacturing variation from optical scanning data. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-05077-5>
- de Neto, M. P. S., Cavalcanti, G. D., de Santos Júnior, O. D. S., & Silva, E. G. (2022). Hybrid systems using residual modeling for sea surface temperature forecasting. *Scientific Reports*, 12(1), 1–16.
- Dueben, P. D., & Bauer, P. (2018). Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10), 3999–4009.
- Efat, M. I. A., Hajek, P., Abedin, M. Z., Azad, R. U., Jaber, M. A., Aditya, S., & Hassan, M. K. (2022). Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04838-6>
- Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, 479, 448–455.
- Franco, G., & Sanstad, A. H. (2008). Climate change and electricity demand in California. *Climatic Change*, 87(1), 139–151.
- Gokasar, I., Timurogullari, A., Ozkan, S. S., & Deveci, M. (2023). IDILIM: Incident detection included linear management using connected autonomous vehicles. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-023-05280-y>
- Gülmez, B. (2023). A novel deep neural network model based Xception and genetic algorithm for detection of COVID-19 from X-ray images. *Annals of Operations Research*, 328, 617–641.
- Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, 573(7775), 568–572.
- Hua, Y., Li, R., Zhao, Z., Chen, X., & Zhang, H. (2019). GAN-powered deep distributional reinforcement learning for resource management in network slicing. *IEEE Journal on Selected Areas in Communications*, 38(2), 334–349.
- Jabeur, S. B., Ballouk, H., Mefteh-Wali, S., & Omri, A. (2022). Forecasting the macrolevel determinants of entrepreneurial opportunities using artificial intelligence models. *Technological Forecasting and Social Change*, 175, 121353.
- Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., Kumar, V. (2019). Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles. In: Proceedings of the 2019 SIAM International Conference on Data Mining, pp. 558–566.
- Jones, J. W., Hansen, J. W., Royce, F. S., & Messina, C. D. (2000). Potential benefits of climate forecasting to agriculture. *Agriculture, Ecosystems & Environment*, 82(1–3), 169–184.
- Ko, Y. C., Zigan, K., & Liu, Y. L. (2021). Carbon capture and storage in South Africa: A technological innovation system with a political economy focus. *Technological Forecasting and Social Change*, 166(2021), 120633.
- Kumar, B., Atey, K., Singh, B. B., Chattopadhyay, R., Acharya, N., Singh, M., & Rao, S. A. (2023). On the modern deep learning approaches for precipitation downscaling. *Earth Science Informatics*, 16(2), 1459–1472.
- Kumar, N., Poonia, V., Gupta, B. B., & Goyal, M. K. (2021). A novel framework for risk assessment and resilience of critical infrastructure towards climate change. *Technological Forecasting and Social Change*, 165, 120532.
- Kurth, T., Treichler, S., Romero, J., Mudigonda, M., Luehr, N., Phillips, E., Mahesh, A., Matheson, M., Deslippe, J., Fatica, M., Prabhat, P., Houston, M. (2018). Exascale deep learning for climate analytics. In: SC18: International Conference for High Performance Computing, Networking, Storage and Analysis, IEEE, pp. 649–660.
- Leal Filho, W., Wall, T., Mucova, S. A. R., Nagy, G. J., Balogun, A. L., Luetz, J. M., & Gandhi, O. (2022). Deploying artificial intelligence for climate change adaptation. *Technological Forecasting and Social Change*, 180, 121662.
- Lemos, M. C., & Rood, R. B. (2010). Climate projections and their impact on policy and practice. *Wiley Interdisciplinary Reviews: Climate Change*, 1(5), 670–682.

- Lomborg, B. (2020). Welfare in the 21st century: Increasing development, reducing inequality, the impact of climate change, and the cost of climate policies. *Technological Forecasting and Social Change*, *156*, 119981.
- Lu, S., Bai, X., Li, W., & Wang, N. (2019). Impacts of climate change on water resources and grain production. *Technological Forecasting and Social Change*, *143*, 76–84.
- Marotzke, J., Müller, W. A., Vamborg, F. S., Becker, P., Cubasch, U., Feldmann, H., Kaspar, F., Kottmeier, Ch., Marini, C., Polkova, I., Prammel, K., Rust, H. W., Stammer, D., Ulbrich, U., Kadow, Ch., Kahl, A., Krager, J., Kruschke, T., Pinto, J. G., ... Ziese, M. (2016). MiKlip: A national research project on decadal climate prediction. *Bulletin of the American Meteorological Society*, *97*(12), 2379–2394.
- Meng, Y., Rigall, E., Chen, X., Gao, F., Dong, J., & Chen, S. (2023). Physics-guided generative adversarial networks for sea subsurface temperature prediction. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(7), 3357–3370.
- Mi, J., Ma, C., Zheng, L., Zhang, M., Li, M., & Wang, M. (2023). WGAN-CL: A Wasserstein GAN with confidence loss for small-sample augmentation. *Expert Systems with Applications*, *233*, 120943.
- Müller, W. A., Jungclauss, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., ... Marotzke, J. (2018). A higher-resolution version of the max planck institute earth system model (mpi-esm1.2-hr). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413.
- Pohlmann, H., Müller, W. A., Bittner, M., Hettrich, S., Modali, K., Pankatz, K., & Marotzke, J. (2019). Realistic quasi-biennial oscillation variability in historical and decadal hindcast simulations using CMIP6 forcing. *Geophysical Research Letters*, *46*(23), 14118–14125.
- Racah, E., Beckham, C., Maharaj, T., Ebrahimi Kahou, S., Prabhat, M., & Pal, C. (2017). Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Advances in Neural Information Processing Systems*, *30*, 1–12.
- Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., & Bengio, Y. (2022). Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)*, *55*(2), 1–96.
- Sadefo Kamdem, J., Miano Mukami, P., & Njong, J. B. (2023). Time-frequency analysis and machine learning models for carbon market forecasting. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-023-05443-x>
- San Kim, T., & Sohn, S. Y. (2020). Machine-learning-based deep semantic analysis approach for forecasting new technology convergence. *Technological Forecasting and Social Change*, *157*, 120095.
- Sarin, S., Haon, C., Belkhouja, M., Mas-Tur, A., Roig-Tierno, N., Segó, T., & Carley, S. (2020). Uncovering the knowledge flows and intellectual structures of research in technological forecasting and social change: A journey through history. *Technological Forecasting and Social Change*, *160*, 120210.
- Scher, S. (2018). Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophysical Research Letters*, *45*(22), 12–616.
- Shahzad, U., Sengupta, T., Rao, A., & Cui, L. (2023). Forecasting carbon emissions future prices using the machine learning methods. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-023-05188-7>
- Shajalal, M., Hajek, P., & Abedin, M. Z. (2023). Product backorder prediction using deep neural network on imbalanced data. *International Journal of Production Research*, *61*(1), 302–319.
- Shao, S., Wang, P., & Yan, R. (2019). Generative adversarial networks for data augmentation in machine fault diagnosis. *Computers in Industry*, *106*, 85–93.
- Smith, D. M., Eade, R., Scaife, A. A., Caron, L. P., Danabasoglu, G., DelSole, T. M., & Yang, X. (2019). Robust skill of decadal climate predictions. *Npj Climate and Atmospheric Science*, *2*(1), 1–10.
- Van Ruijven, B. J., De Cian, E., & Sue Wing, I. (2019). Amplification of future energy demand growth due to climate change. *Nature Communications*, *10*(1), 1–12.
- Wang, J., Zhang, Y., Xing, X., Zhan, Y., Chan, W. K. V., & Tiwari, S. (2023). A data-driven system for cooperative-bus route planning based on generative adversarial network and metric learning. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04842-w>
- Wang, P., Wang, Z., Ye, F., & Chen, X. (2021). Bytesgan: A semi-supervised generative adversarial network for encrypted traffic classification in SDN edge gateway. *Computer Networks*, *200*, 108535.
- Yan, K. (2021). Chiller fault detection and diagnosis with anomaly detective generative adversarial network. *Building and Environment*, *201*, 107982.
- Yerlikaya, B. A., Ömezli, S., & Aydoğan, N. (2020). Climate change forecasting and modeling for the year of 2050. *Environment Climate, Plant and Vegetation Growth* (pp. 109–122). Springer.

- Yoon, J., Jarrett, D., Van der Schaar, M. (2019). Time-series generative adversarial networks. *Advances in Neural Information Processing Systems*, 32. In: Annual Conference on Neural Information Processing Systems 2019 (NeurIPS 2019), pp. 5509–5519.
- Zhang, K., Geng, X., & Yan, X. H. (2020). Prediction of 3-D ocean temperature by multilayer convolutional LSTM. *IEEE Geoscience and Remote Sensing Letters*, 17(8), 1303–1307.
- Zhang, X., Yu, L., Yin, H., & Lai, K. K. (2022). Integrating data augmentation and hybrid feature selection for small sample credit risk assessment with high dimensionality. *Computers & Operations Research*, 146, 105937.
- Zhao, F., Lu, Y., Li, X., Wang, L., Song, Y., Fan, D., & Chen, X. (2022). Multiple imputation method of missing credit risk assessment data based on generative adversarial networks. *Applied Soft Computing*, 126, 109273.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.