

RESEARCH ARTICLE

Determining jumping performance from a single body-worn accelerometer using machine learning

Mark G. E. White^{1,2*}, Neil E. Bezodis¹, Jonathon Neville³, Huw Summers², Paul Rees²

1 Applied Sports, Technology, Exercise and Medicine Research Centre, Swansea University, Swansea, United Kingdom, **2** Department of Biomedical Engineering, Swansea University, Swansea, United Kingdom, **3** Sport Performance Research Institute New Zealand, Auckland University of Technology, Auckland, New Zealand

* markgwhite@gmail.com



OPEN ACCESS

Citation: White MGE, Bezodis NE, Neville J, Summers H, Rees P (2022) Determining jumping performance from a single body-worn accelerometer using machine learning. PLoS ONE 17(2): e0263846. <https://doi.org/10.1371/journal.pone.0263846>

Editor: Chris Connaboy, University of Pittsburgh, UNITED STATES

Received: September 22, 2021

Accepted: January 27, 2022

Published: February 10, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0263846>

Copyright: © 2022 White et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data files are available for download from here: DOI [10.17605/OSF.IO/K4WRB](https://doi.org/10.17605/OSF.IO/K4WRB) Matlab code is also available to process this data.

Abstract

External peak power in the countermovement jump is frequently used to monitor athlete training. The gold standard method uses force platforms, but they are unsuitable for field-based testing. However, alternatives based on jump flight time or Newtonian methods applied to inertial sensor data have not been sufficiently accurate for athlete monitoring. Instead, we developed a machine learning model based on characteristic features (functional principal components) extracted from a single body-worn accelerometer. Data were collected from 69 male and female athletes at recreational, club or national levels, who performed 696 jumps in total. We considered vertical countermovement jumps (with and without arm swing), sensor anatomical locations, machine learning models and whether to use resultant or triaxial signals. Using a novel surrogate model optimisation procedure, we obtained the lowest errors with a support vector machine when using the resultant signal from a lower back sensor in jumps without arm swing. This model had a peak power RMSE of 2.3 W·kg⁻¹ (5.1% of the mean), estimated using nested cross validation and supported by an independent holdout test (2.0 W·kg⁻¹). This error is lower than in previous studies, although it is not yet sufficiently accurate for a field-based method. Our results demonstrate that functional data representations work well in machine learning by reducing model complexity in applications where signals are aligned in time. Our optimisation procedure also was shown to be robust can be used in wider applications with low-cost, noisy objective functions.

Introduction

The ability to generate high levels of neuromuscular power is a critical aspect of sports performance [1–3]. It is strongly correlated with sprint acceleration [4–8] and serves as an indicator of overtraining or fatigue [9–12]. Accordingly, peak external power in the countermovement jump (CMJ) is monitored frequently in many professional athletes [13,14]. Reports of peak

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

power reductions typically range from 3.5% for a training protocol [10] to 12.4% for Australian rules football players 24 hours after a match [9]. The gains in peak power from training can be more substantial with average increases over three years of 13% and 46% reported for male Australian Rules football players and female collegiate gymnasts, respectively [15,16].

Jump testing has practical advantages as it can be administered quickly without residual fatigue, typically before a training session [13,14]. However, the gold-standard method relies on force platforms, which are cumbersome, expensive and unsuited to field-based testing [17]. Many coaches prefer to measure jump height instead of peak power as it requires minimal setup time and equipment [14]. However, although it has some factorial validity [18], jump height is a distinct measure from peak power ($r = 0.93$) [19]. Whereas jump height depends on the cumulative work done (impulse converts to take-off velocity through the conservation of momentum), power is an instantaneous measure reflecting the ability of the locomotor apparatus to perform rapid movements.

Formulae have been proposed for predicting peak power from jump height and body mass, with errors of 247–562 W, equivalent to 6.0–16.5% [20–25]. When those same equations were tested independently, their errors ranged more widely from 3.8–25.3% [24–27]. In recent years, researchers have investigated the use of body-worn inertial sensors to estimate instantaneous power from the body's vertical acceleration and velocity. The peak power estimates, however, were similarly inaccurate with errors of 10.7–21.2% [28–30]. These Newtonian approaches are highly sensitive to the small errors that arise from corrections needed to the sensor's changing orientation [31–33]. Moreover, even without such errors, the computed peak power would not be the same as the true external power because the sensor does not follow the trajectory of the body's centre of mass [34].

Rather than computing peak power directly from the signal, a machine learning approach may be more successful by relating patterns in the data to the performance outcome. Machine learning (ML) models and deep neural networks have been used to predict discrete performance measures derived from the ground reaction force (VGRF), such as peak force or loading rate [35–40]. The ML models in these studies performed at least as well as the neural networks, but without needing high data volumes that can be challenging to obtain, especially in the study of human movement. Different techniques for extracting characteristic features from the data have ranged from devising bespoke metrics to collecting generic statistical measures or employing dimensional reduction techniques such as Principal Component Analysis (PCA). Features based on functional principal components (FPCs) were considerably more accurate than expert-determined discrete measures when predicting jump height from VGRF data [41,42]. Indeed, Functional Principal Component Analysis (FPCA) has been applied to a diverse range of applications in biomechanics. Studies have reported strong associations between FPCs and various performance or injury risk measures in sports, including rowing, swimming, weightlifting, race walking and jumping [41,43–51]. These applications analysed the FPC scores using descriptive statistics, *t*-tests, ANOVA, discriminant analysis or a simple regression model to address their research questions. However, more sophisticated ML models in conjunction with FPCA have not yet been investigated.

When developing an ML model, its parameters need to be tuned through an optimisation procedure, typically using cross validation. However, if model selection is not made independently of model evaluation, then the model selection bias leads to an under-estimation of the model's generalised predictive error [52–55]. Nested cross validation (NCV), also known as double cross validation, overcomes this problem by enforcing the separation between model selection and evaluation, yielding unbiased error estimates [55–58]. It allows different K-fold cross validation (CV) designs to be used for model selection and evaluation, which have

distinct requirements [59]. However, despite its advantages NCV is rarely used in machine learning studies, not least in biomechanics.

This paper presents new models based on functional principal components for predicting peak power in the CMJ from body-worn sensor data. We focus on a single sensor solution for practical reasons as athletes often wear a single inertial measurement unit (IMU) in team sports. Our aim was to produce a model with a predictive error smaller than a typical athlete's inter-day variability. In order to obtain a threshold value for this inter-day variability *a priori*, we averaged the reported inter-day variability in trained athletes across three studies [10,60,61], obtaining a target error level of 3.4%. We modified existing techniques to develop a novel and rigorous optimisation procedure within a nested cross validation framework [62–64]. The optimisation concerned parameters for data preprocessing and the model itself, thereby encompassing the whole modelling procedure [59]. We used this procedure to answer the following research questions: (1) How accurately can peak external power be determined during a CMJ using an ML model based on body-worn accelerometer data? (2) Which of the anatomical locations considered is best for the sensor? (3) How should the signal data be processed?

Materials and methods

Data collection

We recruited 69 healthy participants (45 males, 24 females: body mass 73.1 ± 13.1 kg (mean \pm SD); height 1.74 ± 0.10 m; age 21.6 ± 1.5 years) who gave their written informed consent. The study was approved by the Research Ethics and Governance Committee of Swansea University's College of Engineering. All the participants played a sport, either at recreational (15), club (43) or national (11) level, except for four who trained regularly in the gym. The most frequent sports were football (10), volleyball (7), netball (5), rugby union (5) and rowing (5). The participants each performed either 8 or 16 maximal effort CMJs, divided equally between jumps with arm swing (CMJ_A) and those without (CMJ_{NA}), where hands were placed on hips. Most participants (55) completed 8 jumps as they also performed 8 broad jumps as part of a wider research project. The order of jumps was randomised to minimise potential learning and fatigue effects. The participants were given one minute's rest between each jump. All jumps were performed on two portable 400 × 600 mm force platforms (9260AA, Kistler, Winterthur, Switzerland), which recorded the vertical component of the ground reaction force at a sampling frequency of 1000 Hz. For convenience, all abbreviations used in this paper are listed in Table 1.

The unfiltered VGRF data, summed from both platforms, with body weight (BW) subtracted, gave the net force. The resulting acceleration (i.e. net force/mass) was integrated using the trapezoidal rule to obtain the vertical velocity. The product of velocity and VGRF gave the instantaneous power, from which the maximum value, normalised to body mass, gave the criterion value for peak power in the models below ($W \cdot kg^{-1}$). Jump initiation, the start point for the integration procedure, was identified using a two-step procedure adapted from [65]. The jump was detected initially where VGRF deviated by more than 8% BW, yet the movement must have begun earlier. Rather than using a fixed 30 ms backwards offset [65], the offset depended on where the VGRF deviation had exceeded 1% BW immediately before reaching the 8% threshold.

Delsys Trigno sensors (Delsys Inc., Natick, MA, USA) were attached over the L4 vertebra on the lower back (LB sensor), the C7 vertebra on the upper back (UB sensor), and the lower anterior medial aspect of the tibiae (LS/RS sensors), three anatomical positions commonly used in field-based testing [66] (Fig 1). They were attached directly to the skin using double-

Table 1. Summary of abbreviations.

Acronym	Definition
AM	Accelerometer Model
ANOVA	Analysis of Variance
BW	Body Weight
CMJ	Countermovement Jump
CV	Cross Validation
FPC	Functional Principal Component
FPCA	Functional Principal Component Analysis
GPR	Gaussian Process Regression
GPS	Global Positioning System
IMU	Inertial Measurement Unit.
LB	Lower Back
LR	Linear Regression
LS	Left Shank
ML	Machine Learning
NCV	Nested Cross Validation
PCA	Principal Component Analysis.
PSO	Particle Swarm Optimisation.
RMSE	Root Mean Squared Error.
RS	Right Shank
SM	Surrogate Model).
SVM	Support Vector Machine
UB	Upper Back
VGRF	Vertical Ground Reaction

<https://doi.org/10.1371/journal.pone.0263846.t001>

sided surgical tape and held firmly in place by an elastic adhesive bandage to minimise soft-tissue movement [67,68]. The sensors transmitted the analogue triaxial accelerations (± 9 g) for each jump to a receiving station connected to a computer. Vicon Nexus v2.5 software (Vicon, Oxford, UK) sampled the analogue accelerometer data at 250 Hz and synchronised it with the VGRF data. Although the sensors could digitally sample the measurements, the analogue form

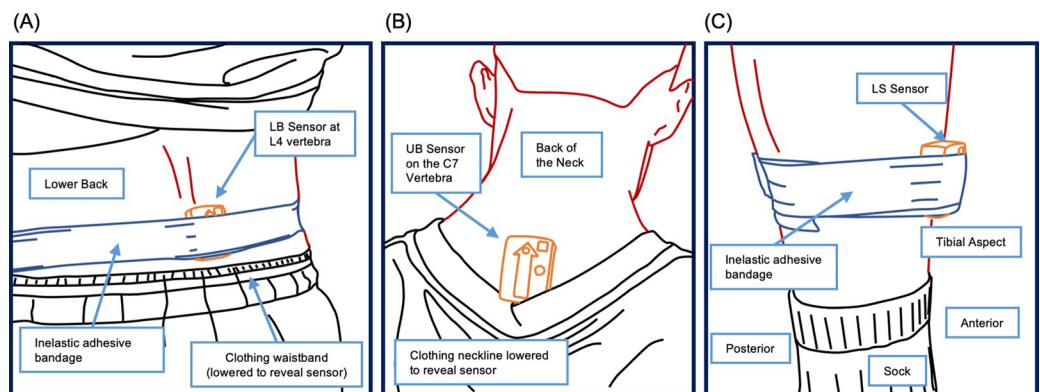


Fig 1. Illustration showing the anatomical position of the inertial sensors. (A) Lower back (LB) sensor attached with double-side tape and held in place with an inelastic adhesive bandage wrapped around the waist; (B) Upper back (UB) sensor attached only with double-sided tape; (C) Left shank (LS) sensor also attached with the same tape and held firmly in place by an adhesive bandage wrapped around the leg. The right shank (RS) sensor (not shown) was attached in the same way.

<https://doi.org/10.1371/journal.pone.0263846.g001>

made the direct synchronisation of accelerometer and VGRF data possible. The sensors were calibrated following the manufacturer's instructions by placing them in six stationary, orthogonal orientations.

Data processing

Data from 696 CMJs were recorded, although four jumps had to be discarded owing to an issue with the accelerometer data. The data (VGRF and accelerometer time series) from 60 participants were assigned to a training/validation data set (548 jumps), while data from the remaining 9 participants (randomly chosen) were placed in an independent holdout test set (144 jumps). All bodyweight-normalised VGRF time series were padded to a standard length, the longest time series. A series of 1's was inserted (i.e. equal to body weight), as required, at the start of the time series to mimic quiet standing before the jump and at the end to reflect the standing position regained after the landing. A similar operation was performed on the accelerometer signal, inserting values equal to the mean acceleration recorded at the start in quiet standing. However, the optimal model may not require the full-length time series, so data were extracted from a time window beginning at a specified time before take-off (t_{pre}) and ending at a time after take-off (t_{post}). These two parameters were allowed to vary in steps of 0.1 s. The take-off time for the accelerometer data was identified where the VGRF dropped below 10 N for the first time [69]. The landing time was when VGRF subsequently rose above 10 N. The flight time was the difference between the take-off and landing times. All processing, unless otherwise stated, was performed in MATLAB R2021a (MathWorks, Natick, MA, USA). The code is available from GitHub: <https://github.com/markgwhite/accModel>.

The accelerometer signals were padded to the same duration as the VGRF data with the mean acceleration vector over the first 0.5 s when the participant stood still before the jump. The signals were then converted into smooth, continuous functions using b-splines [42]. The number of basis functions was defined indirectly as a density (ρ , b-splines per unit time) to make it independent of the time window parameters. The basis function specification (Φ) incorporated the basis order and the penalty order for the roughness penalty. The basis ranged from 4th order (cubic) to 6th order to offer more flexibility and greater b-spline overlap. The accelerometer signal was smoothed either by penalising high curvature (2nd order derivative) or by the rate of change of curvature (3rd order derivative), which would permit abrupt acceleration changes, such as preserving the amplitude of the high acceleration peak on landing. Using a single categorical parameter, Φ , rather than having two parameters (basis and penalty order) reduced the parameter space dimensionality and provided a list of valid combinations (see Search Range in Table 2). N_C specified the number of retained Functional Principal Components (FPCs) [42]. No varimax rotation was used in order to preserve the FPCs' independence and avoid multicollinearity. These procedures were applied to the accelerometer data from all four sensors, providing separate data sets for the ML models. All the parameters are summarised in Table 2.

Modelling procedures

Three common machine learning models were considered: regularised linear regression (LR), support vector machine (SVM) and Gaussian process regression (GPR). The regression models' predictor variables were the accelerometer FPC scores, and the outcome variable was the peak external power computed from the VGRF data. The model hyperparameters and the data processing parameters were determined through the optimisation procedure described below.

Nested cross validation. The optimisation procedure was run within an NCV framework to produce unbiased estimates of the model's generalised predictive error [54–57,70]. The data were first partitioned at the participant level with a 10-fold design for the outer loop. Jumps

Table 2. Data processing parameters with their respective ranges.

Model	Parameter	Description	Type †	Search Range ‡	Optimisation Bounds §
All	t_{pre}	Time before take-off	I	$[-35, \dots, 5] \times 100$ ms	[0.51, 30.49]
	t_{post}	Time after take-off	I	$[-5, \dots, 35] \times 100$ ms	[0.51, 30.49]
	ρ	Density of basis functions per second	R	[2 ... 22]	[4, 20]
	Φ	Basis function (order and penalty derivative)	C	{4-2, 5-2, 5-3, 6-2, 6-3, 6-4} *	[0.51, 6.49]
	λ	Roughness penalty	R	[-12 ... 12]	[-10, 10]
	N_C	No. retained FPCs	I	[1, ... , 35]	[3.51, 30.49]
	Z	Standardisation ¶	C	{No, Yes}	[0.51, 2.49]
LR	R	Regularisation Method	C	{Ridge, Lasso}	[0.51, 2.49]
	S	Solver Method	C	{SVM, Least Squares}	[0.51, 2.49]
	λ_{LR}	Regularisation Parameter	R	[-12 ... 12]	[-10, 10]
SVM	K_{SVM}	Kernel	C	{Gaussian, Linear, Polynomial}	[0.51, 3.49]
	BC	Box Constraint	R	[-7 ... 9]	[-6, 8]
	KS	Kernel Scale	R	[-7 ... 9]	[-6, 8]
	ϵ		R	[-5 ... 3]	[-4, 2]
GPR	B	Basis	C	{None, Constant, Linear, Pure Quadratic}	[0.51, 4.49]
	K_{GPR}	Kernel	C	{Exponential, Squared Exponential, Matérn 3/2, Matérn 5/2, Rational Quadratic}	[0.51, 5.49]
	σ	Noise	R	[-5 ... 3]	[-4, 2]

† Parameter type: C = Categorical; I = Integer; R = Real (continuous, \log_{10} transformed).

‡ Random search range has a broader range to gather data outside the optimisation bounds to prevent boundary effects—see text. Further constraints are imposed by the surrogate model.

§ Particle Swarm Optimisation uses real parameters, so the bounds extend 0.50 below and 0.49 above for categorical (indexed) and integer parameters so when rounded, there is no bias at lower and upper limits.

* Basis encoding = \langle basis order \rangle - \langle penalty order \rangle , defines valid combinations.

¶ Standardise the predictors and outcome variables as Z scores during fitting.

<https://doi.org/10.1371/journal.pone.0263846.t002>

from 54 of the 60 participants were assigned randomly to a training set for each iteration, while the remaining 6 participants were placed in a validation set (Fig 2A). The 10-fold partitioning was repeated (2×10 outer loop iterations) to reduce uncertainty in the predictive error estimate [57,71,72]. A 10-fold design was recommended for model evaluation as validation error estimates have low bias [59,73] and to provide a large proportion of the data for the model selection [55]. For each outer iteration, model selection was performed on the outer training set using 2-fold CV. The inner training and validation sets both comprised data from 26 participants (Fig 2B). Two-fold CV provides a large validation set to increase the likelihood of selecting the best regression model [74,75]. The best model that emerged from the inner loop was evaluated on the outer validation set. Since this data had been kept separate, the validation RMSE was an independent test of the whole modelling procedure.

Objective function. The accelerometer model function (AM) performed all aspects of the modelling process and served as the objective function, returning the 2-fold validation RMSE (loss) for the outer training data set. It carried out time series padding, functional smoothing, data partitioning and the CV inner loop, including FPCA and model fitting, prediction and validation error calculations. The AM defined FPCs based on the inner training partition alone and used them to compute the FPC scores for both the inner training and validation sets. It penalised invalid parameter combinations by returning a high loss ($10 \text{ W}\cdot\text{kg}^{-1}$). Invalid parameter combinations arose when there were insufficient basis functions for the number of

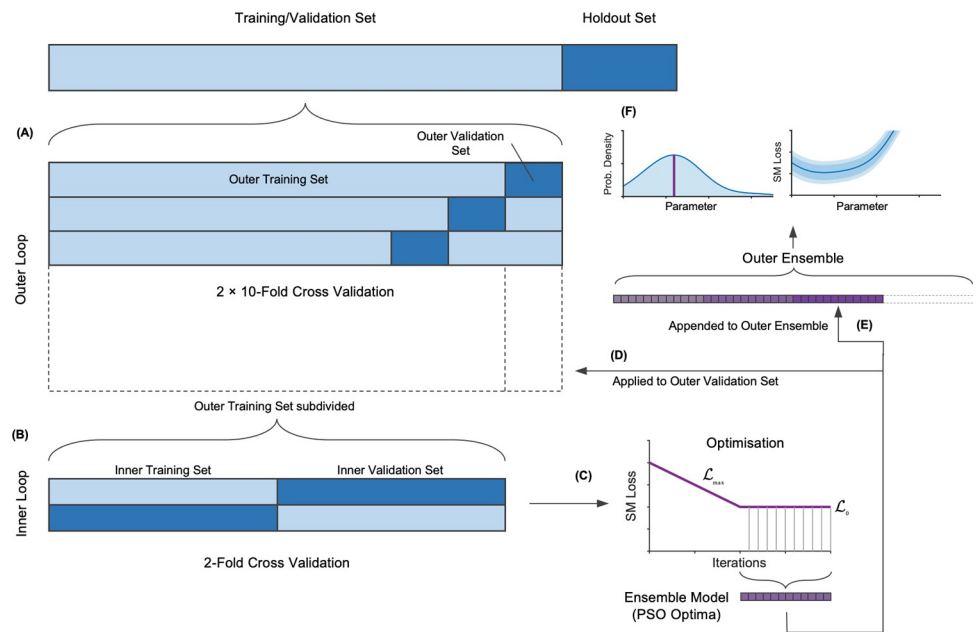


Fig 2. Schematic design of the nested cross validation and optimisation procedure. (A) The training/validation set is partitioned 10-fold for the outer loop, and then (B) each training outer set is re-partitioned 2-fold for the inner loop. (C) Optimisation works with the observations from the inner loop to determine an ensemble model based on the series of optimal parameters determined by Particle Swarm Optimisation. (D) The ensemble model is then evaluated on the outer training set. (E) The process repeats for each outer fold, adding to the series of optimal parameters used to determine the outer ensemble model. (F) This yields parameter distribution and partial plots. It also produces the final ensemble model that may be applied to the holdout data.

<https://doi.org/10.1371/journal.pone.0263846.g002>

FPCs required or if the resulting FPC-score training matrix did not have full column rank. Losses were also capped at $10 \text{ W} \cdot \text{kg}^{-1}$ to prevent very occasional extreme losses from destabilising the optimisation.

Surrogate model. A Bayesian approach is needed to accommodate the objective function’s stochastic behaviour arising from the CV subsampling variance. Many AM observations were required given the high-dimensional parameter space, but we found Bayesian optimisation became prohibitively expensive as more observations were added. Instead, we adapted a random search procedure [76] with a low overhead so it was tractable to make hundreds of observations. A surrogate model (SM) was fitted to the observations based on a Gaussian Process (GP), thus retaining the Bayesian approach [77]. The SM had the same specification used by MATLAB for its *bayesopt* optimiser: an anisotropic Matérn 5/2 kernel and a constant basis function with no predictor standardisation.

Constrained random search. The random search made 400 observations in each optimisation that were constrained to regions of the parameter space where the SM predicted a low AM loss. The constraint was imposed according to the following probability function that governed whether a randomly generated point was accepted:

$$p = \begin{cases} 1, & \mathcal{L}_i \leq \mathcal{L}_{max} \\ \exp\left[\frac{-(\mathcal{L}_i - \mathcal{L}_{max})^2}{2\delta^2}\right], & \mathcal{L}_i > \mathcal{L}_{max} \end{cases}$$

where

$$\mathcal{L}_{max} = \mathcal{L}_0 + \alpha_j \text{SD}(\mathcal{L}), \quad \delta = 1/2 \text{SD}(\mathcal{L})$$

\mathcal{L}_i is the SM prediction for the i -th observation using the j -th surrogate model, such that $i \in \{1 \dots 400\}$ and $j \in \{1 \dots 20\}$. Hence, the SM was retrained every 20 observations. \mathcal{L}_{\max} is a progressively declining upper limit to a baseline loss, \mathcal{L}_0 . δ is a measure of how likely points exceeding \mathcal{L}_{\max} will be accepted. α_j has a ramp profile to constrain the search such that it decreases linearly from 0.5 to 0 over the first half of the search and thereafter remains at zero (Fig 2C). Thus, the search initially surveys the parameter space when almost all points are accepted before tightly focusing its search in promising regions, relying almost entirely on the probability function. Often hundreds of candidate points could be rejected until one is accepted, but the overhead was minimal as the SM predictions were computed quickly (~ 0.0005 s vs 0.2 s for the AM).

Optimisation. The SM was deterministic so global optimisers other than those using Bayesian methods could be employed. We chose Particle Swarm Optimisation (PSO) as it has been used to good effect in previous model optimisation problems [78–80]. It was set up to use 100 particles with an objective tolerance of 0.001. Since PSO only works with continuous variables, it was necessary to index the categorical parameters (Table 2) and use an intermediate objective function that rounded the categorical and integer parameters to the nearest whole number. PSO was run after the SM was retrained on these indexed parameters (Fig 2C). The random search was wider than PSO parameter bounds in order to populate a border region to ensure the SM was well-defined at the periphery (Table 2, last column).

Ensemble models. Once the random search and the final PSO were complete, an ensemble model was selected based on maximum likelihood. For categorical parameters, the most frequently occurring category was chosen, and for numeric parameters, the value with the highest probability density. The values were drawn from the series of PSO-determined optimal models, taken from the second half of the search when $\alpha_j = 0$, provided the constraints were satisfied (Fig 2C). The ensemble model was trained on the outer training set and evaluated on the outer validation set (Fig 2D). The NCV predictive error estimate was the average outer validation RMSE. Since model selection yielded a different model for each outer fold, the same maximum likelihood procedure was used to determine the final ensemble model parameters for the whole data set. The procedure ran on an aggregated list of PSO-determined optimal models taken from all outer folds (Fig 2E). The ensemble AM was trained on the entire training/validation set and then evaluated on the holdout test set to provide a final independent test of the model (Fig 2F).

Model analysis

Statistical comparisons. The whole modelling procedure above was run for each combination of model type (LR, SVM, GPR), sensor location (LB, UB, LS, RS) and jump type (CMJ_{NA} , CMJ_{A}). This analysis was performed on data sets based on the resultant or triaxial accelerometer signals to determine the best signal representation. The outer validation errors were compared between conditions (signal representation, model type, sensor location, jump type) using a two-way ANOVA with 960 observations (2 signal representations \times 3 model types \times 4 sensor locations \times 2 jump types \times 20 outer folds). The ANOVA model was another surrogate model predicting AM loss, which, although inferior to the GP model, allowed hypothesis testing. Effect sizes were based on semi-partial ω^2 , the proportion of the total variance (significance level 0.05) [81]. It was necessary to Winsorise all the data because a few outer validation errors for the SVM models were extremely large ($\gg 10 \text{ W}\cdot\text{kg}^{-1}$, including six $> 20 \text{ W}\cdot\text{kg}^{-1}$, three $> 40 \text{ W}\cdot\text{kg}^{-1}$), rendering otherwise significant effects undetectable. Accordingly, ten observations at opposite ends of the range were adjusted, equivalent to the 1st and 100th percentiles.

The statistical procedures were run in SAS Studio 3.8 (SAS Institute Inc., Cary, NC, USA) using *Proc GLM*. These procedures were bootstrapped (1000 replicas, stratified by condition) to obtain robust estimates because there was no homogeneity of variances at the model type level, according to Levene's test (no suitable transformations would suffice). The bootstrapped estimates are reported with 90% confidence intervals using the median for the central estimate and the 5th and 95th percentiles for the limits.

Model refinement. We selected the dataset with the lowest RMSE across the three model types for further refinement through repeated optimisations. For each model type, some parameter distributions indicated a strong preference for a certain optimal value. In addition, the associated SM partial plots generally showed an advantageous lower predicted loss. Where this was the case, the parameter was fixed at this value, removing it from the optimisation. We judged this subjectively as no satisfactory objective rules could be devised. In other cases, where there was no clear choice, specific values could not be excluded from the search range. Four rounds of optimisation were run for each model type, successively intensifying the search each time. The fourth and final optimal model was then applied to the holdout data set as an independent test.

Results

The peak power computed from the VGRF data (criterion measure) was similar between the training/validation and holdout groups, with higher peak powers recorded in jumps with arm swing (Table 3).

The bootstrapped ANOVA reported an overall effect of $F_{(9,959)} = 24.7$ [19.3, 31.1], $p < 0.0001$ with total $\omega^2 = 0.190$ [0.155, 0.228]. The strongest effects on the outer validation error were made by model type and jump type, respectively, explaining 10.6% and 4.6% of the variance (Table 4). These two factors, and sensor location, were the only ones that were significant across the 90% confidence interval. Signal representation did not always reach significance as the bootstrapped interval for the p -value extended beyond 0.05. It explained less than 1% of the variance, as did the interaction between model type and jump type, the only significant interaction.

The distributions of the Winsorised outer validation errors, grouped by condition, are shown in Fig 3 (top row), revealing which levels within each condition yield more accurate models. Predictions of peak power in the CMJ_{NA} are significantly more accurate in absolute terms than in the CMJ_A: 3.82 W·kg⁻¹ vs 4.62 W·kg⁻¹ (Fig 3A). However, relative to the data set's mean peak power the difference was less marked: 8.5% vs 9.0%. Using the LB sensor yielded more accurate models than when sensors were located elsewhere (Fig 3B), although this difference only reached significance compared to RS models. The errors of the UB, LS and RS sensor-based models were not significantly different from one another. Models based on the resultant accelerometer signal were marginally more accurate than those based on the tri-axial signal, but this difference was not significant (Fig 3C). The model types' errors were all significantly different from one another, with the GPR model being most accurate (Fig 3D).

Table 3. Peak power (W·kg⁻¹) computed from VGRF data.

		Mean ± SD	10th– 90th Percentile	Min, Max
Training / Validation data set	CMJ _{NA}	45.0 ± 7.3	35.2–54.1	27.2, 63.6
	CMJ _A	51.5 ± 8.6	39.7–62.1	28.1, 72.5
Holdout data set	CMJ _{NA}	47.6 ± 8.1	33.0–55.9	29.4, 59.0
	CMJ _A	53.4 ± 10.0	34.9–63.9	31.6, 67.0

<https://doi.org/10.1371/journal.pone.0263846.t003>

Table 4. ANOVA Type I effects for the optimised models' outer validation RMSE.

Effect	DF†	F	ω^2
Model	2	63.5 [46.4, 81.7] *** §	0.106 [0.080, 0.134]
Jump Type	1	55.1 [32.8, 84.9] *** §	0.046 [0.028, 0.069]
Sensor	3	7.7 [3.2, 14.0] *** §	0.017 [0.006, 0.033]
Signal	1	6.1 [0.9, 17.5] *	0.004 [0.000, 0.014]
Model × Jump Type	2	3.5 [0.4, 10.4] *	0.004 [-0.001, 0.015]

Significance for the central estimate indicated by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

§ indicates significance across the bootstrapped 90% CI shown in brackets.

† DF = Degrees of Freedom.

<https://doi.org/10.1371/journal.pone.0263846.t004>

(This general comparison between model types will be revised as the models are refined below.) Considering the models based on the resultant signal for the CMJ_{NA} (best combination for the jump type and signal representation conditions), the GPR models based on the LB sensor data yielded the lowest error ($2.67 \text{ W}\cdot\text{kg}^{-1}$, Fig 3E). As the LB-CMJ_{NA} resultant data set yielded the best models, it was carried forward for the further optimisation of the three model types below.

The distribution of optimal parameters, aggregated over all outer folds, is shown in Fig 4 (data parameters) and Fig 5 (model parameters) for each model type. The ensemble optimal value for each parameter is highlighted on each plot (peak probability density or peak frequency). Most distributions are spread widely across the range with only a modest peak (e.g. t_{pre} and t_{post}), but for some there are more prominent peaks (e.g. SVM model parameters; Fig 4B), none more so than the strong preference for no standardisation. Peaks in the optimisation parameter distributions reflect minima in the partial plots of the SM, as expected (Figs 5–7).

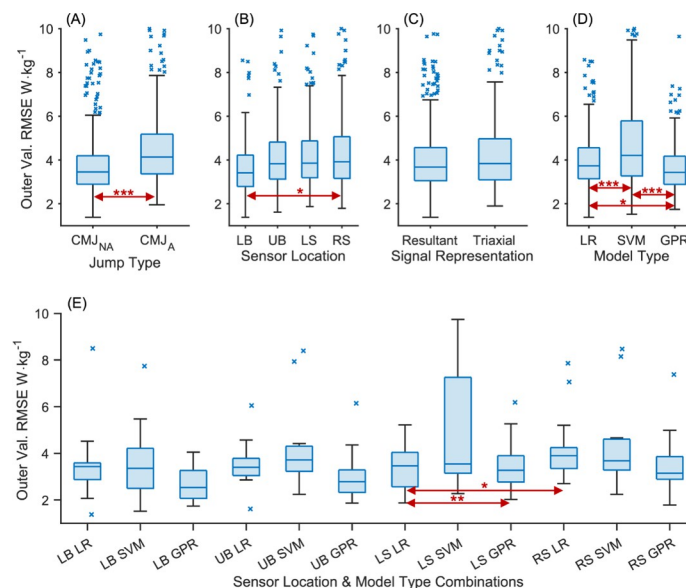


Fig 3. Outer validation RMSE distribution by level for each condition. Top row: single factors in the GLM, namely (A) Jump Type; (B) Sensor location; (C) Signal representation; (D) Model type. Bottom row: two factors (E) Model type and sensor location for the CMJ_{NA} using the resultant signal representation. Horizontal arrows indicate significant differences, where * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

<https://doi.org/10.1371/journal.pone.0263846.g003>

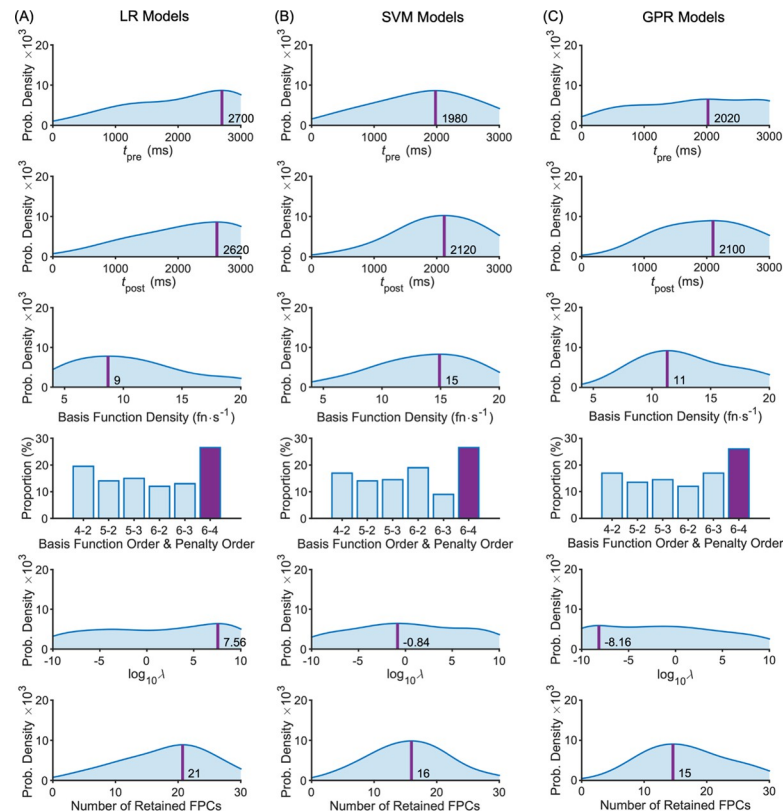


Fig 4. Data parameter distributions across the intermediate models for the LB-CMJ_{NA} data set with ensemble optimal values highlighted. (A) LR model type; (B) SVM model type; (C) GPR model type. Optimal values are shown by the darker shaded bar for categorical parameters and by a darker vertical line at the peak position for numeric parameters with that optimal value shown.

<https://doi.org/10.1371/journal.pone.0263846.g004>

The NCV predictive errors declined progressively with less variance between outer folds when the optimal parameters were refined (Table 5). The ranking between the three model types changed, resulting in the SVM model achieving the lowest predictive error of $2.27 \text{ W}\cdot\text{kg}^{-1}$. In the final round there was no significant difference between the models' predictive error ($p > 0.860$). The LR model achieved marginally the lowest error, but all three were within $0.1 \text{ W}\cdot\text{kg}^{-1}$ of one another (Table 6). In many cases, but not all, excluding specific parameters from the optimisation resulted in more peaked distributions, as can be seen in supplementary material, S1–S3 Figs.

Discussion

This study developed ML models for estimating peak power in the CMJ from accelerometer data from a body-worn inertial sensor. We aimed to produce a model with a predictive error smaller than a typical athlete's inter-day variability. If that level of accuracy were achieved, such a field-based system could be used reliably for monitoring athletes' neuromuscular power. To this end, robust procedures were implemented to obtain unbiased estimates of how the models would perform on independent data. The best model achieved a generalised predictive error of $2.3 \text{ W}\cdot\text{kg}^{-1}$ according to NCV and an independent error of $2.0 \text{ W}\cdot\text{kg}^{-1}$ with the holdout data set. In percentage terms, these errors amount to 5.1% and 4.2% of the mean peak power. These errors are higher than the 3.4% target level for inter-day variability, determined *a priori* from three studies [10,60,61], as presented in the introduction. The 3.4% level is

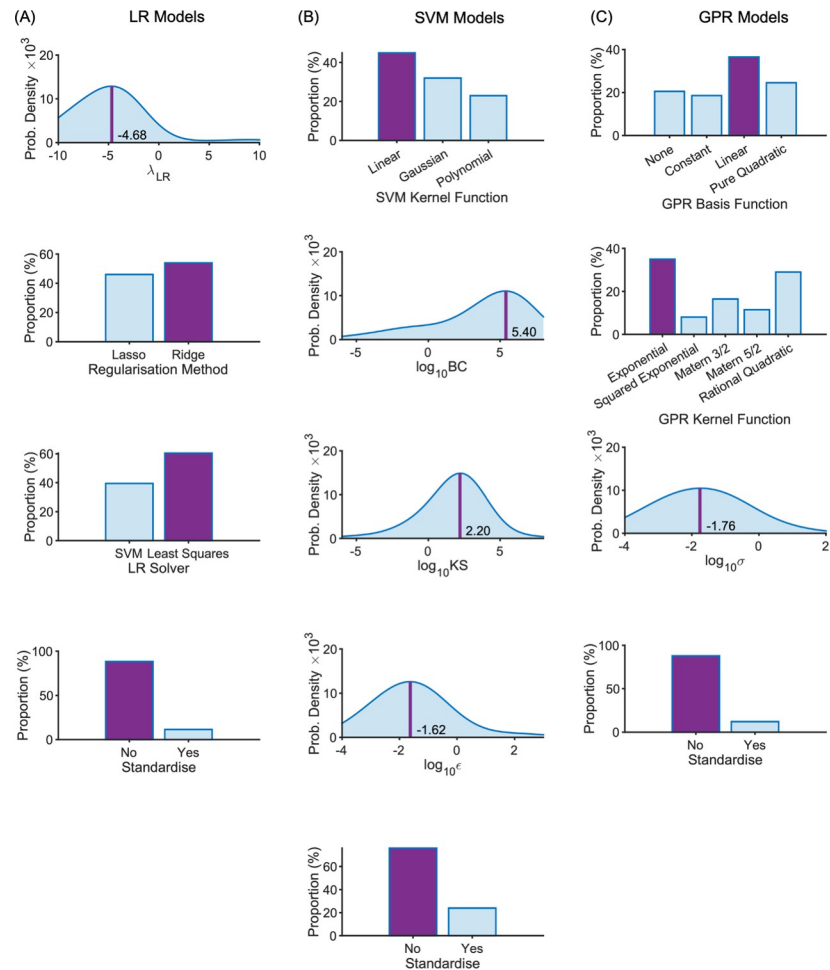


Fig 5. Model parameter distributions across the intermediate models for the LB-CMJ_{NA} data set with ensemble optimal values highlighted. (A) LR model type; (B) SVM model type; (C) GPR model type. Optimal values are shown by the darker shaded bar for categorical parameters and by a darker vertical line at the peak position for numeric parameters with that optimal value shown.

<https://doi.org/10.1371/journal.pone.0263846.g005>

equivalent to 1.55 W·kg⁻¹ with this data set. Thus, our sensor-based system and model does not meet the level of accuracy needed for practical day-to-day use.

Although our approach did not produce a sufficiently accurate model, the results are a considerable improvement over previous attempts in the literature. Estimates of peak power based on jump height had errors of 6.0–16.5% [20–25] while the Newtonian sensor-based calculations resulted in errors of 10.7–21.2% [28–30]. The lowest error reported in those studies was for the Canavan-Vescovi equation [21], but it was based on data from only 20 participants. In subsequent larger studies using the same equation, errors of 2.0%, 25.3% and 27.6% were reported [23,25,27]. The Sayers equation was the most consistent with errors of 5.3 ± 1.2 W·kg⁻¹ (10.5 ± 4.3%) across six studies [20–25]. These studies did not use similarly robust methods to estimate the expected error on independent data, as we did in our study, so their true generalised errors may in fact be higher.

It should also be noted that the performance levels achieved in our study are representative of those reported in the literature. For example, the CMJ_{NA} mean power output of 48.4 W·kg⁻¹ for men in our study compares with 53.6 W·kg⁻¹ for professional rugby players [82], 54 W·kg⁻¹

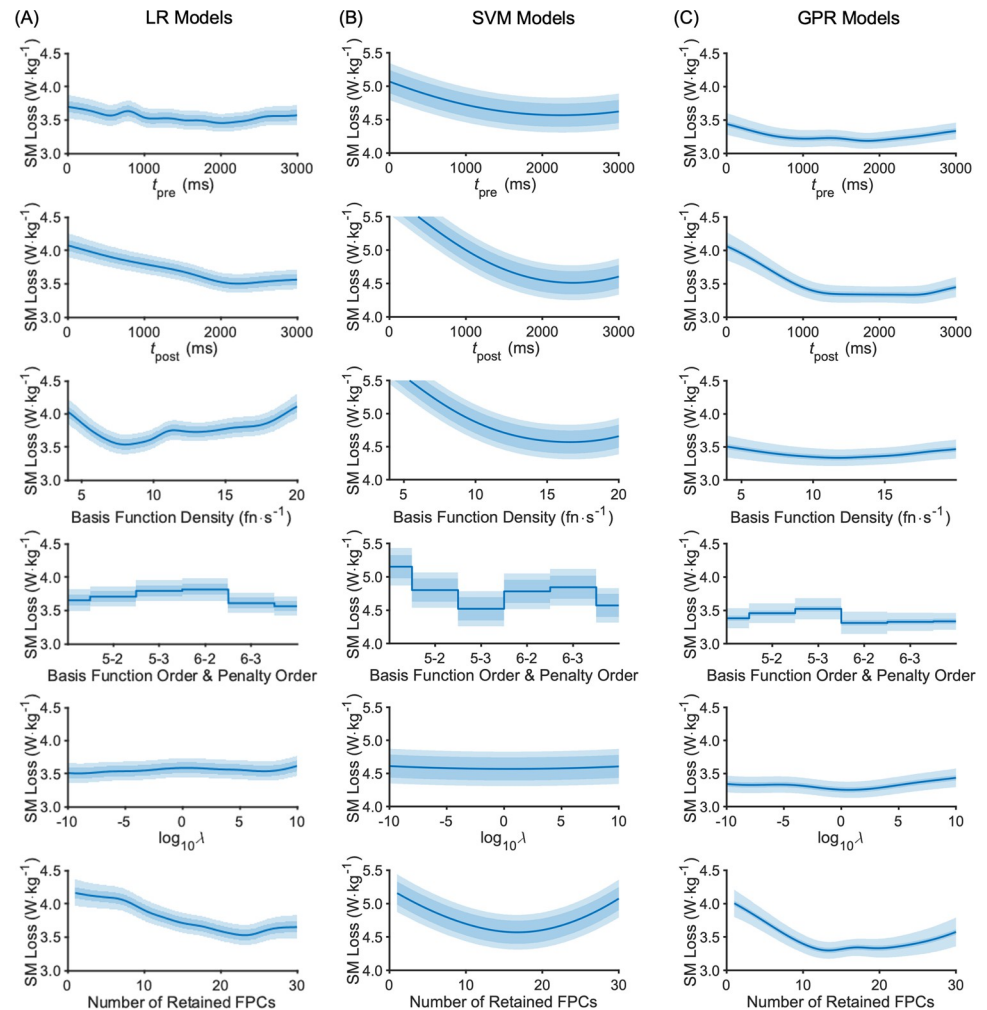


Fig 6. Aggregated surrogate model partial plots for the data parameters from the LB-CMJ_{NA} data set showing the predicted AM loss at the global minimum. (A) LR model type; (B) SVM model type; (C) GPR model type. The central blue line is the central SM estimate. The darker shaded area about this line is the SM fitted noise level. The lighter shaded area covers the standard deviation. Note that for SVM (middle column), the SM range (y-axis) is higher.

<https://doi.org/10.1371/journal.pone.0263846.g006>

for Australian rules football players [9], and $65.1 \text{ W}\cdot\text{kg}^{-1}$ for college-level team-sport athletes [10]. Our female participants' mean performance of $38.2 \text{ W}\cdot\text{kg}^{-1}$ places them in between the $34.8 \text{ W}\cdot\text{kg}^{-1}$ reported for college students who played sports recreationally [83] and the $43.4 \text{ W}\cdot\text{kg}^{-1}$ achieved by NCAA volleyball players [84].

Conditions influencing the model

Jump type. The errors for the CMJ_{NA} were significantly lower than for the CMJ_A in absolute terms ($0.8 \text{ W}\cdot\text{kg}^{-1}$), but in relative terms they were much closer (0.5%) (Fig 3A). The additional degrees of freedom associated with arm swing makes peak power harder to predict, but only moderately so. It should be noted that these are comparisons between different (optimised) models on different data sets, not comparisons of how the same model performs on different data sets. This reveals the adaptability of model fitting and optimisation, as is reflected by jump type having only a weak effect on the model and data processing parameters. Despite the arm swing introducing more degrees of freedom with the possibility of different swing

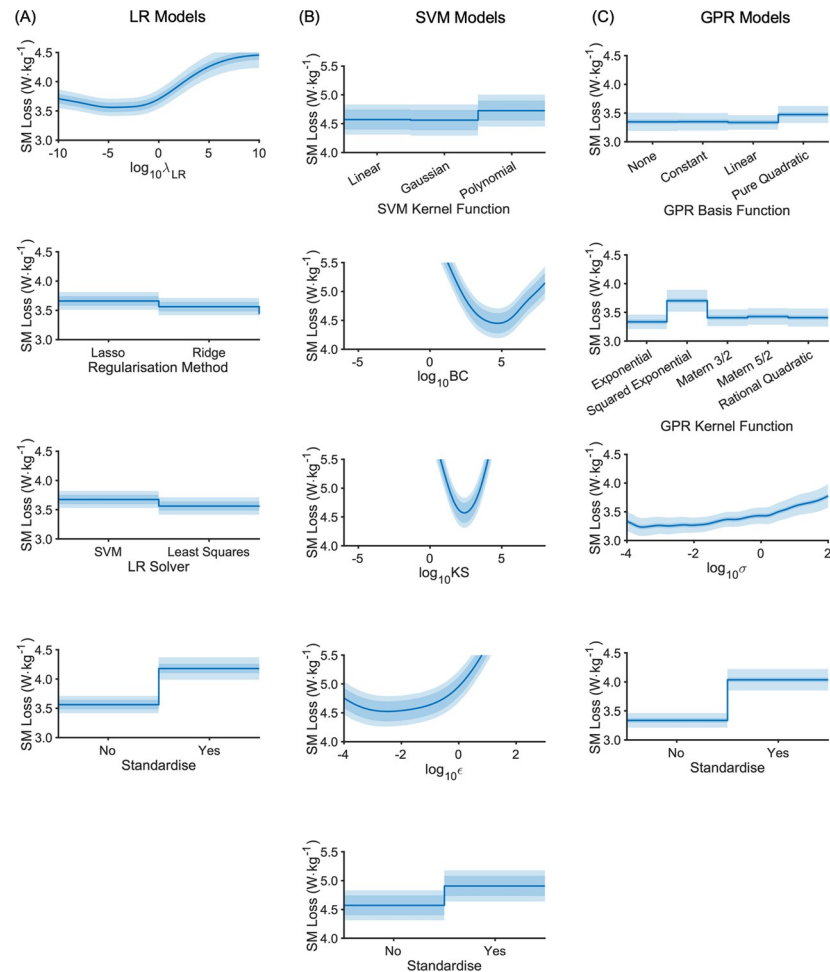


Fig 7. Aggregated surrogate model partial plots for the model parameters from the LB-CMJ_{NA} data set showing the predicted AM loss at the global minimum. (A) LR model type; (B) SVM model type; (C) GPR model type. The central blue line is the central SM estimate. The darker shaded area about this line is the SM fitted noise level. The lighter shaded area covers the standard deviation. Note that for SVM (middle column), the SM range (y-axis) is higher.

<https://doi.org/10.1371/journal.pone.0263846.g007>

movement patterns, the models could accommodate the greater complexity. This finding suggests that such a modelling approach may be suitable for estimating performance metrics in more complex movements.

Table 5. Predictive error estimates over progressive optimisations for each model type using the resultant LB sensor for the CMJ_{NA}, based on nested cross validation and the independent holdout test.

RMSE (W.kg ⁻¹)	LR	SVM	GPR
NCV- 1 st round †	3.50 ± 1.37	3.53 ± 1.44	2.67 ± 0.68
NCV- 2 nd round †	3.11 ± 0.89	3.38 ± 2.11	2.59 ± 0.69
NCV- 3 rd round †	2.93 ± 0.89	2.44 ± 0.47	2.47 ± 0.49
NCV- 4 th round †	2.82 ± 0.87	2.27 ± 0.51	2.38 ± 0.54
Holdout ‡	1.91	2.02	2.02

† For NCV (Nested Cross Validation) estimates, the mean loss is shown ± standard deviation over 20 outer folds. The standard errors in the final round estimates are 0.19 W.kg⁻¹, 0.12 W.kg⁻¹ and 0.11 W.kg⁻¹, respectively for LR, SVM and GPR.

‡The holdout test has a single error specific to that data set.

<https://doi.org/10.1371/journal.pone.0263846.t005>

Table 6. Ensemble optimal parameters over successive optimisations (1st, 2nd, 3rd, 4th) for each model type using the resultant LB sensor for the CMJ_{NA}.

Optimisation Round	LR				SVM				GPR			
	1 st	2 nd	3 rd	4 th	1 st	2 nd	3 rd	4 th	1 st	2 nd	3 rd	4 th
t_{pre} (s)	2.8	2.9	1.9	<u>2.8</u>	1.9	2.7	2.8	<u>1.2</u>	2.9	2.8	2.9	<u>1.2</u>
t_{post} (s)	2.9	2.8	2.8	<u>2.7</u>	1.8	1.7	1.5	<u>1.2</u>	2.2	2.4	1.2	<u>2.8</u>
ρ (fn·s ⁻¹)	9	<u>7</u>	16	<u>8</u>	12	10	<u>8</u>	...
F	6–4	6–4	<u>6–4</u>	...	6–4	6–4	<u>6–4</u>	...	6–4	4–2	<u>6–4</u>	...
$\log_{10} \lambda$	8.8	-9.0	<u>-9.5</u>	...	-2.4	-3.0	0.8	<u>-9.0</u>	-9.5	-1.7	-9.1	<u>-8.4</u>
N_C	21	20	<u>24</u>	...	16	20	<u>24</u>	...	16	12	<u>23</u>	...
Z	<u>No</u>	<u>No</u>	<u>No</u>
R	<u>Ridge</u>								
S	<u>LSq</u>								
$\log_{10} \lambda_{LR}$	-4.0	-1.6	...	<u>-1.0</u>								
K_{SVM}					Linear	<u>Gaussian</u>				
$\log_{10} BC$					5.8	3.1	<u>3.3</u>	...				
$\log_{10} KS$					2.4	<u>2.2</u>				
$\log_{10} \epsilon$					-1.5	<u>-0.7</u>				
B									Linear	<u>None</u>
K_{GPR}									Exp	<u>Exp</u>
$\log_{10} \sigma$									-1.3	<u>-0.5</u>

Successive optimisations progressively involve fewer parameters, narrowing the search and intensifying observations in promising regions. When a parameter’s ‘true’ value has been determined, is it underlined and then held fixed in subsequent optimisations, indicated by the ellipsis. The ‘true’ optimal value is accepted when the parameter distribution shows a narrowly defined, unambiguous peak.

<https://doi.org/10.1371/journal.pone.0263846.t006>

Sensor location. Placing a sensor on the lower back provided the most accurate estimates of peak power of the four anatomical locations considered. The LB models’ mean errors were consistently lower than those based on sensor data from other locations. In biomechanics, the lower back tends to be used more often for sensor attachment, but it will depend on the application in question [66]. In the case of predicting peak power in vertical jumping, having a sensor close to the body’s CM appears to be advantageous, as seen in our results, even though the CM does not have a fixed anatomical location. In comparison, the Newtonian approaches of previous investigations using inertial sensors [31–34] rely on the assumption that the sensor’s movements match those of the body’s CM. Even if those algorithms could perfectly correct for the sensor’s changing orientation, the resulting peak power estimate would pertain to the motion of a sensor rather than the body as a whole. The sensor would have a fixed anatomical location while the body’s CM would move dynamically relative to such a reference point. Hence, the differences in the trajectories of the body’s CM and the sensor will be a source of error in Newtonian methods.

A machine learning approach, in contrast, compensates for different sensor positions in the fitting procedure, determining the best (linear) combination of features to approximate the outcome variable. Hence, models based on sensor data from other locations, seemingly less advantageous, were only slightly less accurate. Both the LB and UB sensors detect trunk movement, which makes the largest segmental contribution to the work done and take-off velocity in a vertical jump [85–87]. However, the same cannot be said for the shank sensors, although the LS/RS models were as accurate as the UB models. The LS/RS sensors tracked the shanks’ changing inclination, a movement with fewer degrees of freedom. As with the comparison

between jump types, this is further evidence of the adaptability of a modelling procedure based on extracting patterns from the data.

In professional team sports, players often wear an inertial measurement unit on the upper back, which usually includes a GPS tracker. In principle, such a sensor could be re-purposed for peak power measurements, which would make it convenient for players and coaches as no additional setup would be required to attach a second sensor for a jump test. The UB model is less accurate than its LB equivalent, but the difference is only marginal. If further improvements could be made to feature extraction methods or the modelling procedure, then using such a sensor-based system for peak power measurement could become a realistic proposition provided the sensor is well-coupled to the player. Field-based testing of peak power could then be incorporated into training programmes, provided other limitations can be overcome, allowing many more tests to be conducted, which may in certain applications partly compensate for the lower level of accuracy compared to the force platform gold standard.

Signal representation. The models based on the resultant signal had marginally lower errors than their triaxial counterparts. The inertial accelerations would have been primarily vertical, making the resultant signal a reasonable first approximation. In principle, the triaxial models had more information, but with many more predictors the model was more prone to overfitting. Furthermore, the sensor's changing inclination in the sagittal plane would bias the accelerations measured along each axis. The baseline gravity vector would shift proportionally between the sensor's X- and Z-axes while the body's inertial acceleration moves in and out of alignment with those axes. However, orientation correction is not a requirement when using a pattern-based machine learning approach in our case, in contrast to the Newtonian approaches discussed above. Our models will have found the best weighting for the FPCs, thus implicitly compensating for the effects of changing sensor orientation, albeit imperfectly. Since the CMJ is a well-controlled movement, making it a valid and reliable test [18,88,89], the changing bias in the inertial accelerations would generally be consistent across jumps. However, differences in strength, coordination, fatigue, limb lengths and muscle morphology will account for variations in the movement pattern, limiting the accuracy of the models [90–93]. Whilst IMUs could correct for orientation, they have an inherent lag in responding to changes of orientation [31], which may limit their suitability for explosive movements. Further research would be needed to determine whether using IMUs rather than simple accelerometers could improve the model predictions, as the additional gyroscope data would permit a correction for sensor orientation [e.g. [94]].

Model type. The final condition was model type where common algorithms were considered, including parametric (LR) and non-parametric methods (SVM and GPR). After the first round of optimisation, the GPR appeared to be the best for this application, but further refinements revealed SVM achieved the lowest errors. This indicates that the global optimum for SVM was harder to find, as may be expected with non-parametric models, which are generally sensitive to the values of the kernel parameters. SVM had three strong parameters (BC, KS and ϵ), all with a continuous range. If one of those parameters was slightly adrift from the true optimum, the errors could be substantially higher. On the other hand, LR and GPR had only had one strong, real parameter each (λ_{LR} and σ , respectively), but more categorical parameters that were easier to optimise. GPR was less prone to overfitting with its Bayesian approach, choosing the most likely solution from the distribution of possible fits. In contrast, SVM had a propensity to produce wildly inaccurate predictions if its hyperparameters were chosen poorly, hence the need to Winsorise the estimates. Furthermore, SVM fitting could occasionally be time-consuming due to its kernel-type design, as indicated by AM execution times: median 0.187 s, 90% CI [0.095 s, 5.080 s]. The times for LR and GPR were more consistently shorter overall: 0.133 s, [0.079, 0.221] s and 0.185 s, [0.122, 0.283] s for GPR. In summary, SVM optimisation

was time-consuming and at times unreliable, but it produced the best estimates in the end. GPR models were more forgiving, less prone to overfitting and easier to work with in practical terms. Ultimately, if the exploration of parameter space is thorough and properly directed, which was achieved by narrowing the parameter search ranges, then the challenges of optimisation with these models can be met.

Optimal parameter values

The optimal parameter values provide answers for our third research question on the best data processing setup. In optimising the time window, the model needs to define a period that includes all the relevant information for the prediction, but there is a trade-off. Extending the time window provides more information, but it comes with the risk of overfitting. A longer period will increasingly encompass periods outside of the jumping movement, especially for jumps that are performed more quickly than others. In these periods, the only inertial accelerations should be due to body sway in standing. In all cases, the optimal time window extended beyond take-off to include flight and landing, indicating valuable information in this latter portion of the signal relating to mechanical power generation and dissipation. The final SVM model may have outperformed the others because its window [-1.2 s, 1.2 s] was limited to these more substantial inertial accelerations, making it less prone to overfitting.

The flight time itself may be useful as the first FPC, which had the highest correlation with peak power, mainly described variations in the timing of the landing impact spike (580 ± 65 ms after take-off compared to an actual flight time according to the VGRF data of 480 ± 66 ms). The second FPC primarily described variations in the impact spike amplitude, indicating an association with peak power via jump height. The models in our study made more accurate predictions than the peak power formulae from previous research because using several FPCs as predictors provides more information than flight time alone. To verify this, we fitted a simple regression model based on flight time and body mass, the same as those previous peak power formulae, and obtained a cross-validated RMSE of $3.49 \text{ W}\cdot\text{kg}^{-1}$.

The roughness penalty, λ , controls how much the signal is smoothed, but its final optimal value ($< 10^{-8}$) was very low (Table 6). In comparison, generalised cross validation, the standard method of determining the roughness penalty, yielded 10^2 [42]. Light smoothing preserves the amplitude of sharp peaks, particularly the impact acceleration spike on landing. It appears the modelling procedure relied partly on reducing the basis function density, ρ , to control complexity. It was helped by using 6th order b-splines, which made up for the low density with considerable flexibility, not just from the quintic polynomials but from their high degree of overlap. The low densities reduced the FPCA computational cost considerably, which is roughly proportional to the square of the number of basis functions. In summary, functional smoothing had quite a limited role in controlling complexity. Indeed, it was optimal to retain a large number of components from FPCA, many of which described very small signal variations.

Having a long list of potentially complex features appeared to be tolerable because in part the models had their own ways of regulating complexity. The LR model favoured ridge regression, which reduced coefficients through the regularisation parameter λ_{LR} , diminishing the influence of some features. The final SVM model had a narrow support vector margin (ϵ), facilitated by a high box constraint (BC) or soft margin, making the model more flexible and less prone to overfitting. At one level, the GPR model used the fitted noise level (σ) of $10^{-0.5}$ ($\sim 0.3 \text{ W}\cdot\text{kg}^{-1}$), but overfitting was controlled mainly through its Bayesian approach. The other part of the explanation can be attributed to the unrotated FPCs having an inherent reduction in amplitude with each successive component. Finally, the optimiser favoured no feature

standardisation because the influence of higher-order FPCs diminished, thus providing a natural form of regularisation.

Modelling procedures

Cross validation is widely regarded as an essential element of machine learning, yet there are comparatively few examples of nested cross validation in the literature. In our study, twice-repeated 10-fold CV (20 outer folds) produced reasonable estimates of the expected generalised error with a standard error of $\sim 0.15 \text{ W}\cdot\text{kg}^{-1}$. More iterations could refine this estimate, but it is already small enough to make no meaningful difference in practice. However, the expected value should not obscure the fact that there was considerable variation in error between folds, indicating a high degree of model sensitivity to the data. It follows that the error for any given jump is somewhat uncertain. Only in aggregate with the large samples can model performance be assessed with the precision reported above.

The statistical model comparing different conditions could only account for 19% of the outer validation errors. The unexplained variance can be attributed to the subsampling variation of the CV inner loop and to differences in the distributions between the inner training set and the outer validation set. The variance could be reduced by averaging over more CV repetitions [62,57,95,96], but that would come with a higher computational cost. For example, the AM loss with two-fold CV without repetition had a noise level of $0.577 \text{ W}\cdot\text{kg}^{-1}$, while five repeats reduced noise to $0.258 \text{ W}\cdot\text{kg}^{-1}$ but the execution time rose by a factor of 4.1.

Optimising a noisy objective function would typically be the task of a Bayesian optimiser. However, although the search directed by its expected-improvement algorithm (or similar) is highly efficient, it comes with a high overhead that rises steeply as more observations are added, as others have reported [97]. We found MATLAB's *bayesopt* optimiser exceeded the AM cost by a factor of 10 after just 50 iterations. Researchers have previously investigated more efficient Bayesian alternatives, but the overhead remains significant [98–100]. The overhead with our method, including SM fitting and PSO, was only 3.5% of the total execution time, allowing a high proportion of computing resources to be devoted to the search.

Limitations

The models depended on accelerometer signals being aligned perfectly with take-off, which had been achieved by referring to the synchronised VGRF data. If an accelerometer-based system were to be implemented, it would have to be self-sufficient by detecting take-off from the accelerometer data alone. That would introduce an alignment error, which could potentially reduce the effectiveness of FPCA, depending on the algorithm's accuracy [42]. Algorithms for estimating CMJ flight time from body-worn inertial sensors have errors of 21–37 ms [34,101,102]. Assuming the take-off and landing detection errors have identical normal distributions, the take-off errors would be 15–26 ms. Further research is needed to develop a suitable algorithm and quantify its effect on the AM validation error.

FPCA as a feature extraction method is based on a linear decomposition that requires more components to represent a pattern than would otherwise be the case with nonlinear representations, such as those obtained using autoencoders [103]. Using such feature encodings may improve the models, although it may be more appropriate to use a second neural network to make the performance predictions. Such an approach may work well in more complex situations where athletic movements have more degrees of freedom. What our study has shown is that reasonably accurate estimates can be obtained using linear feature representations provided the movement is carefully controlled in a test environment.

Finally, the NCV error estimates assumed independent, identically distributed data, an assumption that is common in machine learning. Were the model applied to a new cohort with a different peak power distribution to the one used here, the errors would have a different spread making the RMSE perhaps higher or lower. This can be seen in the holdout errors where the LR model outperformed the other two and its NCV estimate. Recruiting participants from a range of sports partly addressed this as it created a heterogeneous data set without being specific to a single cohort. A replication study evaluating the same methodology with different sensors, researchers and participants would contribute greatly to the ecological validity of the research.

Conclusions

The final models developed in this study using accelerometer data from body-worn sensors predicted peak power in the CMJ_{NA} more accurately than has hitherto been achieved by a field-based system. The error estimates reported above can be considered realistic owing to the robust procedures implemented. However, with errors of $2.3 \text{ W}\cdot\text{kg}^{-1}$ or 5.1%, they do not reach the level of accuracy desired for practical use. Nevertheless, with further developments, this gap may be bridged such that a valuable single-sensor system could be applied for certain practical applications. The models themselves were based on FPCA, which has been successful in biomechanics, with optimisation of data processing parameters, as well as the model's hyperparameters. We believe this is the first biomechanics study to take this comprehensive approach to optimisation. In yielding a small number of features characterising time series data, FPCA allows classical machine learning models to be employed. It would be suitable where there is a natural point of alignment, such as jump take-off, so the modes of variation become apparent without further data manipulation. It is a modelling approach that has potentially wider applications in biomechanics as it has been shown to be adaptable to different data sets.

Supporting information

S1 Fig. LR model optimal parameter distributions over four successive optimisations.

Parameters may be eliminated in successive rounds if there is a clear preference for an optimal value. Alternatively, the range of possible values may be reduced. In doing so, subsequent distributions tend to have more prominent peaks, but not always, as with the time window parameters. Abbreviations. Vertical axes: Proportion = Proportion (%); Density = Probability Density Function $\times 10^3$. Standardise Axis: N = No; Y = Yes. Regularisation Axis: L = Lasso; R = Ridge. LR Solver Axis: S = SVM; L = Least Squares. (TIF)

S2 Fig. SVM model optimal parameter distributions over four successive optimisations.

Parameters may be eliminated in successive rounds if there is a clear preference for an optimal value. Alternatively, the range of possible values may be reduced. Abbreviations. Vertical axes: Proportion = Proportion (%); Density = Probability Density Function $\times 10^3$. Standardise Axis: N = No; Y = Yes. SVM Kernel Axis: L = Linear; G = Gaussian; P = Polynomial. (TIF)

S3 Fig. GPR model optimal parameter distributions over four successive optimisations.

Parameters may be eliminated in successive rounds if there is a clear preference for an optimal value. Alternatively, the range of possible values may be reduced. Abbreviations. Vertical axes: Proportion = Proportion (%); Density = Probability Density Function $\times 10^3$. GPR Basis Axis: N = None; C = Constant; L = Linear; RQ = Rational Quadratic. GPR Kernel Axis:

E = Exponential; SE = Squared Exponential; M3 = Matérn 3/2; M5 = Matérn 5/2;
RQ = Rational Quadratic. Standardise Axis: N = No; Y = Yes.
(TIF)

S4 Fig.
(TIF)

Author Contributions

Conceptualization: Mark G. E. White, Neil E. Bezodis, Huw Summers.

Data curation: Mark G. E. White.

Formal analysis: Mark G. E. White.

Funding acquisition: Huw Summers.

Investigation: Mark G. E. White.

Methodology: Mark G. E. White, Jonathon Neville, Paul Rees.

Project administration: Neil E. Bezodis.

Software: Mark G. E. White.

Supervision: Neil E. Bezodis, Jonathon Neville, Paul Rees.

Validation: Mark G. E. White.

Writing – original draft: Mark G. E. White.

Writing – review & editing: Neil E. Bezodis, Jonathon Neville, Paul Rees.

References

1. Cronin J, Sleivert G. Challenges in Understanding the Influence of Maximal Power Training on Improving Athletic Performance: *Sports Med.* 2005; 35: 213–234. <https://doi.org/10.2165/00007256-200535030-00003> PMID: 15730337
2. Cormie P, McGuigan MR, Newton RU. Developing Maximal Neuromuscular Power. *Sports Med.* 2011; 22.
3. Claudino JG, Cronin J, Mezêncio B, McMaster DT, McGuigan M, Tricoli V, et al. The countermovement jump to monitor neuromuscular status: A meta-analysis. *J Sci Med Sport.* 2017; 20: 397–402. <https://doi.org/10.1016/j.jsams.2016.08.011> PMID: 27663764
4. Baker D. Differences in strength and power among junior-high, senior-high, college-aged, and elite professional rugby league players. *J Strength Cond Res.* 2002; 16: 581–585. PMID: 12423189
5. Baker D, Nance S. The Relation Between Running Speed and Measures of Strength and Power in Professional Rugby League Players. *J Strength Cond Res.* 1999; 13: 230–235.
6. Baker D, Nance S. The Relation Between Strength and Power in Professional Rugby League Players. *J Strength Cond Res.* 1999; 13: 224–229.
7. Chelly SM, Denis C. Leg power and hopping stiffness: relationship with sprint running performance: *Med Sci Sports Exerc.* 2001; 326–333. <https://doi.org/10.1097/00005768-200102000-00024> PMID: 11224825
8. Young W, Newton R, Doyle T, Chapman D, Cormack S, Stewart C, et al. Physiological and anthropometric characteristics of starters and non-starters and playing positions in elite Australian Rules football: a case study. *J Sci Med Sport.* 2005; 8: 333–345. [https://doi.org/10.1016/s1440-2440\(05\)80044-1](https://doi.org/10.1016/s1440-2440(05)80044-1) PMID: 16248474
9. Cormack SJ, Newton RU, McGuigan MR. Neuromuscular and Endocrine Responses of Elite Players to an Australian Rules Football Match. *Int J Sports Physiol Perform.* 2008; 3: 359–374. <https://doi.org/10.1123/ijsp.3.3.359> PMID: 19211947

10. Gathercole R, Sporer B, Stellingwerff T, Sleivert G. Alternative Countermovement-Jump Analysis to Quantify Acute Neuromuscular Fatigue. *Int J Sports Physiol Perform*. 2015; 10: 84–92. <https://doi.org/10.1123/ijsp.2013-0413> PMID: 24912201
11. McLean BD, Coutts AJ, Kelly V, McGuigan MR, Cormack SJ. Neuromuscular, Endocrine, and Perceptual Fatigue Responses During Different Length Between-Match Microcycles in Professional Rugby League Players. *Int J Sports Physiol Perform*. 2010; 5: 367–383. <https://doi.org/10.1123/ijsp.5.3.367> PMID: 20861526
12. Twist C, Highton J. Monitoring Fatigue and Recovery in Rugby League Players. *Int J Sports Physiol Perform*. 2013; 8: 467–474. <https://doi.org/10.1123/ijsp.8.5.467> PMID: 23319463
13. Jones T, Smith A, Macnaughton LS, French DN. Strength and Conditioning and Concurrent Training Practices in Elite Rugby Union: J Strength Cond Res. 2016; 30: 3354–3366. <https://doi.org/10.1519/JSC.000000000001445> PMID: 27870696
14. Taylor K-L, Chapman DW, Cronin JB, Newton MJ, Gill N. Fatigue monitoring in High Performance Sport: A Survey of Current Trends. *J Aust Strength Cond*. 2012; 20: 12.
15. McGuigan MR, Cormack S, Newton RU. Long-Term Power Performance of Elite Australian Rules Football Players. *J Strength Cond Res*. 2009; 23: 26–32. <https://doi.org/10.1519/jsc.0b013e31818753ca> PMID: 19130637
16. French DN, Gómez AL, Volek JS, Rubin MR, Ratamess NA, Sharman MJ, et al. Longitudinal tracking of muscular power changes of NCAA Division I collegiate women gymnasts. *J Strength Cond Res*. 2004; 18: 101–107. [https://doi.org/10.1519/1533-4287\(2004\)018<0101:ltompc>2.0.co;2](https://doi.org/10.1519/1533-4287(2004)018<0101:ltompc>2.0.co;2) PMID: 14971975
17. McMahon J, Lake JP, Suchomel TJ. Vertical jump testing. Performance Assessment in Strength and Conditioning. Abingdon, Oxon.: Routledge, Taylor & Francis Group; 2019.
18. Markovic G, Dizdar D, Jukic I, Cardinale M. Reliability and factorial validity of squat and countermovement jump tests. *J Strength Cond Res*. 2004; 18: 551–555. [https://doi.org/10.1519/1533-4287\(2004\)18<551:RAFVOS>2.0.CO;2](https://doi.org/10.1519/1533-4287(2004)18<551:RAFVOS>2.0.CO;2) PMID: 15320660
19. Dowling JJ, Vamos L. Identification of Kinetic and Temporal Factors Related to Vertical Jump Performance. *J Appl Biomech*. 1993; 9: 95–110. <https://doi.org/10.1123/jab.9.2.95>
20. Amonette WE, Brown LE, De Witt JK, Dupler TL, Tran TT, Tufano JJ, et al. Peak Vertical Jump Power Estimations in Youths and Young Adults: J Strength Cond Res. 2012; 26: 1749–1755. <https://doi.org/10.1519/JSC.0b013e3182576f1e> PMID: 22505126
21. Canavan PK, Vescovi JD. Evaluation of Power Prediction Equations: Peak Vertical Jumping Power in Women: *Med Sci Sports Exerc*. 2004; 36: 1589–1593. <https://doi.org/10.1249/01.mss.0000139802.96395.ac> PMID: 15354042
22. Sayers SP, Harackiewicz DV, Harman EA, Frykman PN, Rosenstein MT. Cross-validation of three jump power equations. *Med Sci Sports Exerc*. 1999; 31: 572–577. <https://doi.org/10.1097/00005768-199904000-00013> PMID: 10211854
23. Lara AJ, Abián J, Alegre LM, Jiménez L, Aguado X. Assessment of power output in jump tests for applicants to a sports sciences degree. *J Sports Med Phys Fitness*. 2006; 46: 419–424. PMID: 16998446
24. Quagliarella L, Sasanelli N, Belgiovine G, Moretti L, Moretti B. Evaluation of Standing Vertical Jump by Ankles Acceleration Measurement: *J Strength Cond Res*. 2010; 24: 1229–1236. <https://doi.org/10.1519/JSC.0b013e3181cb281a> PMID: 20386133
25. Lara-Sánchez AJ, Zagalaz ML, Berdejo-del-Fresno D, Martínez-López EJ. Jump Peak Power Assessment Through Power Prediction Equations in Different Samples: *J Strength Cond Res*. 2011; 25: 1957–1962. <https://doi.org/10.1519/JSC.0b013e3181e06ef8> PMID: 21399537
26. Tessier J-F, Basset F-A, Simoneau M, Teasdale N. Lower-Limb Power cannot be Estimated Accurately from Vertical Jump Tests. *J Hum Kinet*. 2013; 38: 5–13. <https://doi.org/10.2478/hukin-2013-0040> PMID: 24235979
27. Amonette WE, Brown LE, De Witt JK, Dupler TL, Tran TT, Tufano JJ, et al. Peak Vertical Jump Power Estimations in Youths and Young Adults: *J Strength Cond Res*. 2012; 26: 1749–1755. <https://doi.org/10.1519/JSC.0b013e3182576f1e> PMID: 22505126
28. Giroux C, Rabita G, Chollet D, Guilhem G. What is the Best Method for Assessing Lower Limb Force-Velocity Relationship? *Int J Sports Med*. 2014; 36: 143–149. <https://doi.org/10.1055/s-0034-1385886> PMID: 25259590
29. Hojka V, Tufano JJ, Maly T, Stastny P, Jebavy R, Feher J, et al. Concurrent validity of Myotest for assessing explosive strength indicators in countermovement jump. *Acta Gymnica*. 2018; 48: 95–102. <https://doi.org/10.5507/ag.2018.013>

30. Mauch M, Rist H-J, Kaelin X. Reliability and Validity of Two Measurement Systems in the Quantification of Jump Performance. *Swiss Sports Exerc Med*. 2014; 62. <https://doi.org/10.34045/SSEM/2014/6>
31. Godwin A, Agnew M, Stevenson J. Accuracy of Inertial Motion Sensors in Static, Quasistatic, and Complex Dynamic Motion. *J Biomech Eng*. 2009; 131: 114501. <https://doi.org/10.1115/1.4000109> PMID: 20353265
32. Shull PB, Xu J, Yu B, Zhu X. Magneto-Gyro Wearable Sensor Algorithm for Trunk Sway Estimation During Walking and Running Gait. *IEEE Sens J*. 2017; 17: 480–486. <https://doi.org/10.1109/JSEN.2016.2630938>
33. Tan T, Chiasson DP, Hu H, Shull PB. Influence of IMU position and orientation placement errors on ground reaction force estimation. *J Biomech*. 2019; 97: 109416. <https://doi.org/10.1016/j.jbiomech.2019.109416> PMID: 31630774
34. Picerno P, Camomilla V, Capranica L. Countermovement jump performance assessment using a wearable 3D inertial measurement unit. *J Sports Sci*. 2011; 29: 139–146. <https://doi.org/10.1080/02640414.2010.523089> PMID: 21120742
35. Derie R, Robberechts P, Van den Berghe P, Gerlo J, De Clercq D, Segers V, et al. Tibial Acceleration-Based Prediction of Maximal Vertical Loading Rate During Overground Running: A Machine Learning Approach. *Front Bioeng Biotechnol*. 2020; 8: 33. <https://doi.org/10.3389/fbioe.2020.00033> PMID: 32117918
36. Guo Y, Storm F, Zhao Y, Billings S, Pavic A, Mazzà C, et al. A New Proxy Measurement Algorithm with Application to the Estimation of Vertical Ground Reaction Forces Using Wearable Sensors. *Sensors*. 2017; 17: 2181. <https://doi.org/10.3390/s17102181> PMID: 28937593
37. Gurchiek RD, Rupasinghe Arachchige Don HS, Pelawa Watagoda LCR, McGinnis RS, van Werkhoven H, Needle AR, et al. Sprint Assessment Using Machine Learning and a Wearable Accelerometer. *J Appl Biomech*. 2019; 35: 164–169. <https://doi.org/10.1123/jab.2018-0107> PMID: 30676153
38. Neugebauer JM, Collins KH, Hawkins DA. Ground Reaction Force Estimates from ActiGraph GT3X+ Hip Accelerations. Hug F, editor. *PLoS ONE*. 2014; 9: e99023. <https://doi.org/10.1371/journal.pone.0099023> PMID: 24914946
39. Pogson M, Verheul J, Robinson MA, Vanrenterghem J, Lisboa P. A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities. *Med Eng Phys*. 2020; 78: 82–89. <https://doi.org/10.1016/j.medengphys.2020.02.002> PMID: 32115354
40. Wouda FJ, Giuberti M, Bellusci G, Maartens E, Reenalda J, van Beijnum B-JF, et al. Estimation of Vertical Ground Reaction Forces and Sagittal Knee Kinematics During Running Using Three Inertial Sensors. *Front Physiol*. 2018; 9: 218. <https://doi.org/10.3389/fphys.2018.00218> PMID: 29623042
41. Richter C, O'Connor NE, Marshall B, Moran K. Comparison of discrete-point vs. dimensionality-reduction techniques for describing performance-related aspects of maximal vertical jumping. *J Biomech*. 2014; 47: 3012–3017. <https://doi.org/10.1016/j.jbiomech.2014.07.001> PMID: 25059895
42. Ramsay JO, Silverman BW. *Functional data analysis*. 2nd ed. New York: Springer; 2005.
43. Donà G, Preatoni E, Cobelli C, Rodano R, Harrison AJ. Application of functional principal component analysis in race walking: An emerging methodology. *Sports Biomech*. 2009; 8: 284–301. <https://doi.org/10.1080/14763140903414425> PMID: 20169759
44. Donoghue OA, Harrison AJ, Coffey N, Hayes K. Functional Data Analysis of Running Kinematics in Chronic Achilles Tendon Injury. *Med Sci Sports Exerc*. 2008; 40: 1323–1335. <https://doi.org/10.1249/MSS.0b013e31816c4807> PMID: 18580414
45. Floría P, Sánchez-Sixto A, Harrison AJ. Application of the principal component waveform analysis to identify improvements in vertical jump performance. *J Sports Sci*. 2019; 37: 370–377. <https://doi.org/10.1080/02640414.2018.1504602> PMID: 30058950
46. Harrison AJ, Ryan W, Hayes K. Functional data analysis of joint coordination in the development of vertical jump performance. *Sports Biomech*. 2007; 6: 199–214. <https://doi.org/10.1080/14763140701323042> PMID: 17892096
47. Kipp K, Redden J, Sabick MB, Harris C. Weightlifting Performance Is Related to Kinematic and Kinetic Patterns of the Hip and Knee Joints: *J Strength Cond Res*. 2012; 26: 1838–1844. <https://doi.org/10.1519/JSC.0b013e318239c1d2> PMID: 21986692
48. Ryan W, Harrison AJ, Hayes K. Functional data analysis of knee joint kinematics in the vertical jump. *Sports Biomech*. 2006; 5: 121–138. <https://doi.org/10.1080/14763141.2006.9628228> PMID: 16521626
49. Sacilotto GBD, Warmenhoven JS, Mason BR, Ball N, Clothier J. Investigation of ATM propulsion Force-time Profiles using Functional Data Analysis on Front Crawl Sprint Swimmers. 33rd International Conference of Biomechanics in Sports. 2015.

50. Warmenhoven J, Cogley S, Draper C, Harrison AJ, Bargary N, Smith R. Bivariate functional principal components analysis: considerations for use with multivariate movement signatures in sports biomechanics. *Sports Biomech*. 2017; 18: 10–27. <https://doi.org/10.1080/14763141.2017.1384050> PMID: [29125036](https://pubmed.ncbi.nlm.nih.gov/29125036/)
51. Warmenhoven J, Smith R, Draper C, Harrison AJ, Bargary N, Cogley S. Force coordination strategies in on-water single sculling: Are asymmetries related to better rowing performance? *Scand J Med Sci Sports*. 2018; 28: 1379–1388. <https://doi.org/10.1111/sms.13031> PMID: [29222948](https://pubmed.ncbi.nlm.nih.gov/29222948/)
52. Zucchini W. An Introduction to Model Selection. *J Math Psychol*. 2000; 44: 41–61. <https://doi.org/10.1006/jmps.1999.1276> PMID: [10733857](https://pubmed.ncbi.nlm.nih.gov/10733857/)
53. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci*. 2002; 99: 6562–6566. <https://doi.org/10.1073/pnas.102102699> PMID: [11983868](https://pubmed.ncbi.nlm.nih.gov/11983868/)
54. Cawley GC, Talbot NLC. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J Mach Learn Res*. 2010; 11: 2079–2107.
55. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. 2006; 7: 91. <https://doi.org/10.1186/1471-2105-7-91> PMID: [16504092](https://pubmed.ncbi.nlm.nih.gov/16504092/)
56. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J R Stat Soc Ser B Methodol*. 1974; 36: 111–133. <https://doi.org/10.1111/j.2517-6161.1974.tb00994.x>
57. Filzmoser P, Liebmann B, Varmuza K. Repeated double cross validation. *J Chemom*. 2009; 23: 160–171. <https://doi.org/10.1002/cem.1225>
58. Baumann D, Baumann K. Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. *J Cheminformatics*. 2014; 6: 47. <https://doi.org/10.1186/s13321-014-0047-1> PMID: [25506400](https://pubmed.ncbi.nlm.nih.gov/25506400/)
59. Zhang Y, Yang Y. Cross-validation for selecting a model selection procedure. *J Econom*. 2015; 187: 95–112. <https://doi.org/10.1016/j.jeconom.2015.02.006>
60. Cormack SJ, Newton RU, McGuigan MR, Doyle TLA. Reliability of Measures Obtained During Single and Repeated Countermovement Jumps. *Int J Sports Physiol Perform*. 2008; 3: 131–144. <https://doi.org/10.1123/ijssp.3.2.131> PMID: [19208922](https://pubmed.ncbi.nlm.nih.gov/19208922/)
61. Hori N, Newton RU, Andrews WA, Kawamori N, McGuigan MR, Nosaka K. Comparison of four different methods to measure power during hang power clean and weighted jump squat. *J Strength Cond Res*. 2007; 21: 314–320. <https://doi.org/10.1519/R-22896.1> PMID: [17530989](https://pubmed.ncbi.nlm.nih.gov/17530989/)
62. Braga-Neto U. Small-sample error estimation: mythology versus mathematics. In: Astola JT, Tabus I, Barrera J, editors. San Diego, California, USA; 2005. p. 59160V. doi:10.1117/12.619331.
63. Brochu E, Cora VM, de Freitas N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. [arXiv.org](https://arxiv.org/abs/2010.04979). 2010; 49.
64. Eggenberger K, Lindauer M, Hoos HH, Hutter F, Leyton-Brown K. Efficient benchmarking of algorithm configurators via model-based surrogates. *Mach Learn*. 2018; 107: 15–41. <https://doi.org/10.1007/s10994-017-5683-z>
65. Owen NJ, Watkins J, Kilduff LP, Bevan HR, Bennett MA. Development of a Criterion Method to Determine Peak Mechanical Power Output in a Countermovement Jump. *J Strength Cond Res*. 2014; 28: 1552–1558. <https://doi.org/10.1519/JSC.0000000000000311> PMID: [24276298](https://pubmed.ncbi.nlm.nih.gov/24276298/)
66. Camomilla V, Bergamini E, Fantozzi S, Vannozzi G. Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review. *Sensors*. 2018; 18: 873. <https://doi.org/10.3390/s18030873> PMID: [29543747](https://pubmed.ncbi.nlm.nih.gov/29543747/)
67. Forner-Cordero A, Mateu-Arce M, Forner-Cordero I, Alcántara E, Moreno JC, Pons JL. Study of the motion artefacts of skin-mounted inertial sensors under different attachment conditions. *Physiol Meas*. 2008; 29: N21–N31. <https://doi.org/10.1088/0967-3334/29/4/N01> PMID: [18401071](https://pubmed.ncbi.nlm.nih.gov/18401071/)
68. Ziegert JC, Lewis JL. The Effect of Soft Tissue on Measurements of Vibrational Bone Motion by Skin-Mounted Accelerometers. *J Biomech Eng*. 1979; 101: 218–220. <https://doi.org/10.1115/1.3426248>
69. Street G, McMillan S, Board W, Rasmussen M, Heneghan JM. Sources of Error in Determining Countermovement Jump Height with the Impulse Method. *J Appl Biomech*. 2001; 17: 43–54. <https://doi.org/10.1123/jab.17.1.43>
70. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J Cheminformatics*. 2014; 6: 10. <https://doi.org/10.1186/1758-2946-6-10> PMID: [24678909](https://pubmed.ncbi.nlm.nih.gov/24678909/)
71. Burman P. A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods. *Biometrika*. 1989; 76: 503. <https://doi.org/10.2307/2336116>

72. Zhang P. Model Selection Via Multifold Cross Validation. *Ann Stat.* 1993; 21. <https://doi.org/10.1214/aos/1176349027>
73. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence.* Montreal, Quebec, Canada; 1995. pp. 1137–1143.
74. Shao J. Linear Model Selection by Cross-validation. *J Am Stat Assoc.* 1993; 88: 486–494. <https://doi.org/10.1080/01621459.1993.10476299>
75. Shao J. An asymptotic theory for linear model selection. *Stat Sin.* 1997; 7: 221–242.
76. Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. *J Mach Learn Res.* 2012; 13: 281–305.
77. Snoek J, Larochelle H, Adams RP. Practical Bayesian Optimization of Machine Learning Algorithms. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. *Advances in Neural Information Processing Systems 25.* Curran Associates, Inc.; 2012. pp. 2951–2959. Available: <http://papers.nips.cc/paper/4522-practical-bayesian-optimization-of-machine-learning-algorithms.pdf>
78. Kennedy J, Eberhart R. Particle swarm optimization. *Proceedings of ICNN'95—International Conference on Neural Networks.* Perth, WA, Australia: IEEE; 1995. pp. 1942–1948. doi:10.1109/ICNN.1995.488968.
79. Escalante HJ, Montes M, Sucar E. Ensemble particle swarm model selection. *The 2010 International Joint Conference on Neural Networks (IJCNN).* Barcelona, Spain: IEEE; 2010. pp. 1–8. doi:10.1109/IJCNN.2010.5596915.
80. Unler A, Murat A. A discrete particle swarm optimization method for feature selection in binary classification problems. *Eur J Oper Res.* 2010; 206: 528–539. <https://doi.org/10.1016/j.ejor.2010.02.032>
81. Field AP. *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock “n” roll.* 4th edition. Los Angeles: Sage; 2013.
82. McMahon J, Murphy S, Rej SJE, Comfort P. Countermovement-Jump-Phase Characteristics of Senior and Academy Rugby League Players. *Int J Sports Physiol Perform.* 2017; 12: 803–811. <https://doi.org/10.1123/ijsp.2016-0467> PMID: 27918658
83. Makaruk H, Winchester JB, Sadowski J, Czaplicki A, Sacewicz T. Effects of Unilateral and Bilateral Plyometric Training on Power and Jumping Ability in Women: *J Strength Cond Res.* 2011; 25: 3311–3318. <https://doi.org/10.1519/JSC.0b013e318215fa33> PMID: 22076090
84. Newton RU, Rogers RA, Volek JS, Häkkinen K, Kraemer WJ. Four Weeks of Optimal Load Ballistic Resistance Training at the End of Season Attenuates Declining Jump Performance of Women Volleyball Players. *J Strength Cond Res.* 2006; 20: 955. <https://doi.org/10.1519/R-5050502x.1> PMID: 17194257
85. Luhtanen P, Komi PV. Segmental contribution to forces in vertical jump. *Eur J Appl Physiol.* 1978; 38: 181–188. <https://doi.org/10.1007/BF00430076> PMID: 648508
86. Miller DI, East DJ. Kinematic and kinetic correlates of vertical jumping in woman. In: Komi PV, editor. *Biomechanics V-B.* Baltimore, University Park Press; 1976. pp. 65–72.
87. Pandy MG, Zajac FE, Sim E, Levine WS. An optimal control model for maximum-height human jumping. *J Biomech.* 1990; 23: 1185–1198. [https://doi.org/10.1016/0021-9290\(90\)90376-e](https://doi.org/10.1016/0021-9290(90)90376-e) PMID: 2292598
88. Aragón LF. Evaluation of Four Vertical Jump Tests: Methodology, Reliability, Validity, and Accuracy. *Meas Phys Educ Exerc Sci.* 2000; 4: 215–228. https://doi.org/10.1207/S15327841MPEE0404_2
89. Hatze H. Validity and Reliability of Methods for Testing Vertical Jumping Performance. *J Appl Biomech.* 1998; 14: 127–140. <https://doi.org/10.1123/jab.14.2.127>
90. Nuzzo JL, McBride JM, Cormie P, McCaulley GO. Relationship Between Countermovement Jump Performance and Multijoint Isometric and Dynamic Tests of Strength: *J Strength Cond Res.* 2008; 22: 699–707. <https://doi.org/10.1519/JSC.0b013e31816d5eda> PMID: 18438251
91. Rodacki ALF, Fowler NE, Bennett SJ. Vertical jump coordination: fatigue effects: *Med Sci Sports Exerc.* 2002; 34: 105–116. <https://doi.org/10.1097/00005768-200201000-00017> PMID: 11782655
92. Methenitis SK, Zaras ND, Spengos KM, Stasinaki A-NE, Karampatsos GP, Georgiadis GV, et al. Role of Muscle Morphology in Jumping, Sprinting, and Throwing Performance in Participants With Different Power Training Duration Experience. *J Strength Cond Res.* 2016; 30: 807–817. <https://doi.org/10.1519/JSC.0000000000001147> PMID: 26907845
93. Aouadi R, Jlid MC, Khalifa R, Hermassi S, Chelly MS, Van Den Tillaar R, et al. Association of anthropometric qualities with vertical jump performance in elite male volleyball players. *J Sports Med Phys Fitness.* 2012; 52: 11–17. PMID: 22327081

94. Madgwick SOH, Harrison AJL, Vaidyanathan R. Estimation of IMU and MARG orientation using a gradient descent algorithm. 2011 IEEE International Conference on Rehabilitation Robotics. Zurich: IEEE; 2011. pp. 1–7. doi:10.1109/ICORR.2011.5975346.
95. Kim J-H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal.* 2009; 53: 3735–3745. <https://doi.org/10.1016/j.csda.2009.04.009>
96. Rodríguez JD, Pérez A, Lozano JA. A general framework for the statistical analysis of the sources of variance for classification error estimators. *Pattern Recognit.* 2013; 46: 855–864. <https://doi.org/10.1016/j.patcog.2012.09.007>
97. Snoek J, Rippel O, Swersky K, Kiros R, Satish N, Sundaram N, et al. Scalable Bayesian Optimization Using Deep Neural Networks. ArXiv150205700 Stat. 2015 [cited 22 Aug 2021]. Available: <http://arxiv.org/abs/1502.05700>.
98. Balandat M, Karrer B, Jiang DR, Daulton S, Letham B, Wilson AG, et al. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. ArXiv191006403 Cs Math Stat. 2020 [cited 19 May 2021]. Available: <http://arxiv.org/abs/1910.06403>.
99. Letham B, Karrer B, Ottoni G, Bakshy E. Constrained Bayesian Optimization with Noisy Experiments. *Bayesian Anal.* 2019; 14: 495–519. <https://doi.org/10.1214/18-BA1110>
100. Li C, Gupta S, Rana Sm, Nguyen V, Venkatesh S, Shilton A. High Dimensional Bayesian Optimization using Dropout. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization; 2017. pp. 2096–2102. doi:10.24963/ijcai.2017/291.
101. Castagna C, Ganzetti M, Ditroilo M, Giovannelli M, Rocchetti A, Manzi V. Concurrent Validity of Vertical Jump Performance Assessment Systems: *J Strength Cond Res.* 2013; 27: 761–768. <https://doi.org/10.1519/JSC.0b013e31825dbcc5> PMID: 22648140
102. Monnet T, Decatoire A, Lacouture P. Comparison of algorithms to determine jump height and flight time from body mounted accelerometers. *Sports Eng.* 2014; 17: 249–259. <https://doi.org/10.1007/s12283-014-0155-1>
103. Hinton GE, Osindero S, Teh Y-W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* 2006; 18: 1527–1554. <https://doi.org/10.1162/neco.2006.18.7.1527> PMID: 16764513