

Training Load Responses Modelling in Elite Sports: How to Deal with Generalisation?

Frank Imbach (✉ frank.imbach@umontpellier.fr)

University of Montpellier

Stephane Perrey

University of Montpellier

Romain Chailan

University of Montpellier

Thibaut Meline

University of Perpignan

Robin Candau

University of Montpellier

Research Article

Keywords: sport performance modelling, elastic net (ENET), principal component regression (PCR), random forest (RF)

Posted Date: December 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-128940/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Training load responses modelling in elite sports: how to deal with generalisation?

Frank Imbach^{1,2,3,*}, Stephane Perrey², Romain Chailan¹, Thibaut Meline⁴, and Robin Candau³

¹Seenovate, Univ. Montpellier, Montpellier, 34000, France

²EuroMov Digital Health in Motion, Univ. Montpellier, IMT Mines Ales, Montpellier, 34090, France

³DMeM, Univ. Montpellier, INRAe, Montpellier, 34000, France

⁴Laboratoire Européen Performance Santé Altitude, Univ. Perpignan Via Domitia, Font-Romeu, 66120, France

*frank.imbach@umontpellier.fr

ABSTRACT

This study aims to provide a transferable methodology in the context of sport performance modelling, with a special focus to the generalisation of models. Data were collected from seven elite Short track speed skaters over a three months training period. In order to account for training load accumulation over sessions, cumulative responses to training were modelled by impulse, serial and bi-exponential responses functions. The variable dose-response (DR) model was compared to elastic net (ENET), principal component regression (PCR) and random forest (RF) models, while using cross-validation within a time-series framework. ENET, PCR and RF models were fitted either individually (M_I) or on the whole group of athletes (M_G). Root mean square error criterion was used to assess performances of models. ENET and PCR models provided a significant greater generalisation ability than the DR model ($p = 0.012$, $p < 0.001$, $p = 0.005$ and $p < 0.001$ for $ENET_I$, $ENET_G$, PCR_I and PCR_G , respectively). Only $ENET_I$, $ENET_G$ and RF_I were significantly more accurate in prediction than DR ($p = 0.020$, $p < 0.001$ and $p = 0.043$, respectively). In conclusion, ENET achieved greater generalisation and predictive accuracy performances. Thus, building and evaluating models within a generalisation enhancing procedure is a prerequisite for any predictive modelling.

Introduction

The relationship between training load and performance in sports has been studied since decades. A key point of the performance optimisation is the training prescription delivered by coaches, physical trainers or the athlete himself. Such a programming involves both various modalities of exercise (i.e. the type of training regarding to the physical quality required to perform) and adjusted training load. Training load is usually dissociated into i) an external load defined by the work completed by the athlete, independently of his internal characteristics¹ and ii) an internal load corresponding to the psycho-physiological stresses imposed on the athlete in response to the external load².

Models of training load responses emerged with the impulse response model promoted by Banister *et al.*³ in order to describe human adaptations to training loads. Afterwards, a simplified version of the original model built on a two-way antagonistic first order transfer function (fitness and fatigue components, so called Fitness - Fatigue model) has showed a large interest to describe the training process⁴⁻⁸. However, several limitations regarding to the model stability, parameter interpretability, ill-conditioning and predictive accuracy were reported^{9,10}. Such models may be considered as non-linear models according to their component structure¹¹ and therefore, require a sufficient number of observations (i.e. performances) to correctly estimate relationships between training load and performance^{9,12}. To overcome some of the limits, refinements of the former impulse response model were proposed by using a recursive algorithm in order to estimate parameters according to each model input (i.e. the training load)¹¹ and by introducing variations in the fatigue response to a single training bout¹³. Further adaptations to the Fitness - Fatigue model were also developed with the aim of improving both goodness-of-fit and prediction accuracy^{14,15}. Nonetheless, because impulse-response models aiming to predict training effects in both endurance (running, cycling, skiing and swimming)^{6,11,16-21} and more complex (hammer throw, gymnastic and judo)^{8,22,23} activities sought to mitigate the underpinning physiological processes involved by exercise into a small number of entities, accuracy of predictions might greatly suffer from²⁴. Moreover, these models assume that the training effect is maximal by the end of the training session. This assumption is reasonable only for the negative component of the model (i.e. "Fatigue"), where its maximal value is taken immediately after the session. Regarding to the positive effects induced by training (i.e. "Fitness"), such a response is quite questionable since physiological adaptations are continuing from the end of the exercise session. For instance, skeletal muscle adaptations to training described by increases in muscle mass, fiber shortening velocity and myosin

ATPase activity modifications are known to be progressive (i.e. short to long term after-effects) rather than instantaneous²⁵⁻²⁷. Consequently, serial and bi-exponential functions were proposed to counteract these limitations and better describe training adaptations through exponential growth and decay functions, according to physiological responses in rats²⁸.

A more statistical approach was used to investigate the effects of training load on performance by using principal component analysis and linear mixed models on different time frames¹². Such models infer parameters from all available data (i.e. combining subjects instead of by-subject model) but allow parameters to vary in respect of heterogeneity between athletes. The model being multivariate, the multi-faceted nature of the performance could be conserved by including psychological, nutritional and technical information as predictors^{12,16,18}. However, authors did not consider the cumulative facet of daily training loads, where exponential and decay cumulative functions such as proposed by Candau *et al.*¹⁷ may be suitable for performance modelling.

Alternatives from computer sciences field were also used to clarify the training load - performance relationship in a predictive aim. Most notably, machine learning approaches are usually focused on the generalisation of models (i.e. a model ability to make accurate predictions on unknown data). Various approaches tend to maximise such a criterion. For instance, one can perform cross-validation (CV) procedures, where data are separated into training sets for parameters estimation and testing sets for prediction²⁹. Such a procedure fosters the determination of optimal models, relatively to the family of models considered and regarding to their ability for generalisation. In the same time, CV procedures allow to diagnose under- and over-fitting of the model. Underfitting commonly describes an inflexible model unable of capturing noteworthy regularities in a set of exemplary observations³⁰. In contrast, overfitting represents an over-trained model, which tends to memorise each particular observation thus leading to high error rates when predicting on unknown data³¹. While aforementioned studies aimed to describe the training load - performance relationships by estimating model parameters and by testing the model on a single data set, generalisation of models cannot be ensured. This challenges their usefulness in a predictive application. On the other hand, modelling methodologies using CV procedures are the standard in a predictive aim rather than only being descriptive. To our knowledge, only a few recent studies modelled performances with Fitness-Fatigue models using a CV procedure^{10,32,33} and one separated data into two equals training and testing data sets respectively³⁴. Two of them reported an overfitting of the model at the expense of accurate predictions^{10,33}. Consequently, interpretations drawn from predictions as well as model parameters may be incorrect.

The physiological adaptations involved by exercise being complex, some authors investigated the relationship between training and performance by using Artificial Neural Networks (ANN), non-linear machine learning models^{35,36}. Despite low prediction errors reported (e.g. 0.05 seconds error over a 200m swimming performance³⁵), the methodological consideration in their study mostly influenced by a small sample size and the "black-box" nature of ANN question their use in sport performance modelling^{9,37}. Computer sciences offer plenty of machine learning models although being often resumed to ANN for physical performance prediction. Some algorithms might be preferred for sport performance modelling by preventing high dimensionality and multicollinearity issues. To cite a few, non-linear approaches such as Random Forest (RF) models³⁸ and regularised linear regressions^{39,40} proved their efficiency and could be profitable in a sport performance context.

To date, not any consensus is claimed about which model family (i.e. impulse response and physiological based, statistical and machine learning models) should be preferred for physical performance prediction based on a data set, mainly due to a lack of evidence and confidence in training effect modelling and performance prediction accuracy. In addition, because generalisation ability is not systemically appraised, practical and physiological interpretations drawn from models may be incorrect and at least should be taken with caution.

Beyond the model employed to elucidate the relationships between training load and performance in a predictive application, we hypothesised that modelling approaches including CV procedures, regularisation and dimension reduction methods would lead to a greater generalisation than former impulse response models.

In order to prescribe an optimal training programming, sport practitioners need to understand the physiological effects involved by each training session and its after-effects on physical performance. Hence, this study aimed to provide a reliable and transferable methodology relying on model generalisation in a context of sport performance modelling. To do so, the variable dose-response (DR) model¹³ was considered as a baseline framework and compared to two regularisation methods and one machine learning regression model: a principal component regression (PCR), an Elastic net (ENET) regression and a RF regression model, respectively.

Materials and methods

Participants

Seven national elite Short-track speed skaters (mean age 22.7 ± 3.4 years old; 3 males, body mass of 71.4 ± 9.4 kg, and 4 females, body mass of 55.9 ± 3.9 kg) voluntarily participated to the study. Each athletes experienced the 2018 Olympic Winter Games in PyeongChang, South Korea ($n = 2$) or were preparing the Olympics Games of Pekin, China ($n = 7$). The whole team was trained by the same coach, responsible for training programming and data collection. Mean weekly volume of training was

16.6 ± 2.5 hours. Data were collected over a three months training period, beginning one month after resuming training for a new season. Participants were fully informed about data collection and written consent was obtained from them. The study was performed in agreement with the standards set by the declaration of Helsinki (2013) involving human subjects. The protocol was reviewed and approved by the local research Ethics Committee (EuroMov, University of Montpellier, France). The present retrospective study relied on the collected data without causing any changes in the training programming of athletes.

Data set

Dependent variable: Performance

Participants performed each week standing start time trials ($distance = 166.68$ meters equal 1.5 lap) after a standardised warm-up. At the finish line, photocell timing system (Brower timing system, USA) recorded individual time trial performance. A total of $n = 248$ performances were recorded for an average of 35.4 ± 2.23 individual performances. The performance test being a gold standard for the assessment of acceleration ability⁴¹, athletes were all familiar with it prior to the study.

In the sequel, let $\mathcal{Y} \subset \mathbb{R}$ be the domain of definition of such a performance and $Y \in \mathcal{Y}$ the random variable. In this context, each observation $y_j \in Y$ can be referenced by both its athlete i and its day of realisation t as $y_{i,t}$.

Independent variables

Independent variables stem from various sources, which are summarised in Table 1. In the sequel, let $\mathcal{X} \subset \mathbb{R}^d$ with $d \in \mathbb{N}$ be the domain of definition of the random variable $X = [X_1, \dots, X_d] \in \mathcal{X}$. The variable X is thus defined as a vector composed of the independent variables detailed hereafter. First, $\{X_1\}$ refers to the raw training loads (TL, Figure 1c), calculated from on-ice and off-ice training sessions (see details on Supplementary material Appendix). Then, $\{X_2, X_3\}$ represent two aggregations of daily TL. Those aggregations come from the daily training loads $w(t)$ –also known here as X_1 – convoluted to two transfer functions adapted from Philippe *et al.*²⁸, which are denoted $g_{\text{imp}}(t)$ and $g_{\text{ser}}(t)$.

The associated impulse response $G_{\text{imp}}(t)$ reflects the acute response to exercise (e.g. fatigue). It is defined as

$$G_{\text{imp}}(t) = e^{-\frac{t}{\tau_I}}, \quad (1)$$

where τ_I is a short time constant equals to 3 days in this study (Figure 1a). Respectively, the response $G_{\text{ser}}(t)$ describes a serial and bi-exponential function reflecting training adaptations over time. It is defined as

$$G_{\text{ser}}(t) = \left(1 - e^{-\frac{t}{\tau_G}}\right)U + e^{-\frac{-(t-TD)}{\tau_D}} |U - 1|, \quad \text{with } U = \begin{cases} 1 & \text{if } t < TD \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The time delay for the decay phase to begin only after the growth phase is given by TD . Here, $TD = 4\tau_G$. Both τ_G and τ_D are the time constants of respectively the growth phase and the decline phase with $\tau_G = 1$ day and $\tau_D = 7$ days (Figure 1b). Note that the time constants τ_I , τ_G , τ_D were averaged and based on empirical knowledge and previous findings¹³. Hence, for a given athlete,

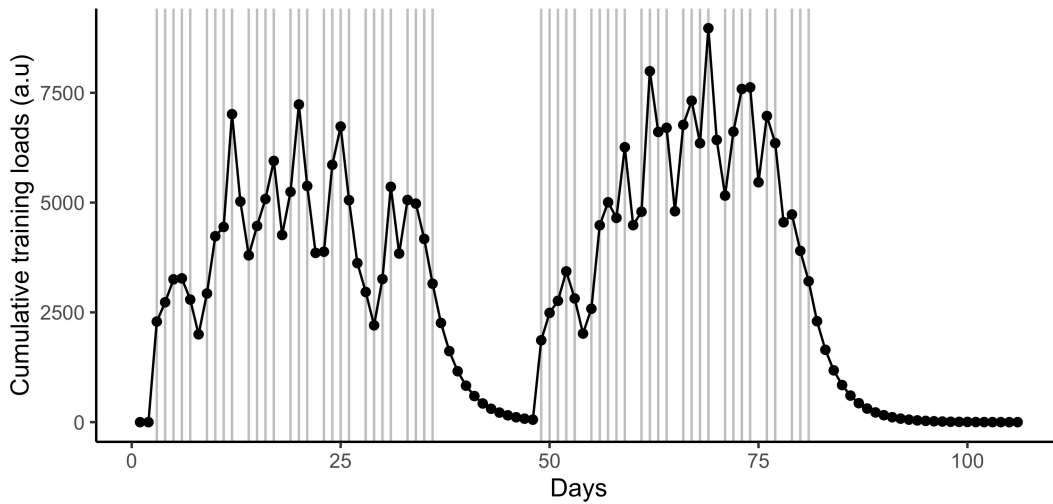
$$X_2(t) = (w * g_{\text{imp}})(t) = \sum_{l=1}^t w(l) \left(e^{-\frac{-(t-l)}{\tau_I}} \right), \quad \text{and}$$

$$X_3(t) = (w * g_{\text{ser}})(t) = \sum_{l=1}^t w(l) \left(\left(1 - e^{-\frac{-(t-l)}{\tau_G}}\right)U + e^{-\frac{-(t-TD-l)}{\tau_D}} |U - 1| \right), \quad \text{with } U = \begin{cases} 1 & \text{if } t < TD \\ 0 & \text{otherwise.} \end{cases}$$

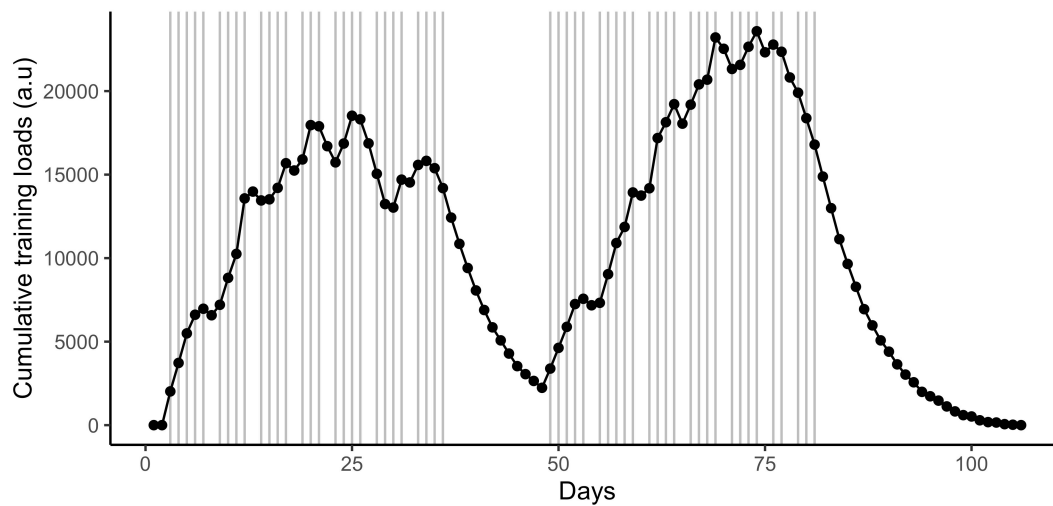
Note that the symbol $*$ denotes the convolution product.

Similarly, some characteristics components of each session were aggregated. This encompasses Rate of Perceived Exertion (RPE) $\{X_4, X_5\}$, averaged power $\{X_6, X_7\}$, maximal power output $\{X_8, X_9\}$, relative intensity $\{X_{10}, X_{11}\}$, session duration $\{X_{12}, X_{13}\}$ and session density $\{X_{14}, X_{15}\}$. Also, for each session ice quality $\{X_{16}\}$ and rest between two consecutive sessions $\{X_{17}\}$ were considered. Since some models may benefit from time through autocorrelated performances $y_{i,t}$, the preceding performance $y_{i,t-k}$ with $k = 1$ was included as predictor, denoted $\{X_{18}\}$. Finally, athlete $\{X_{19}\}$ was considered excepted for individually built models.

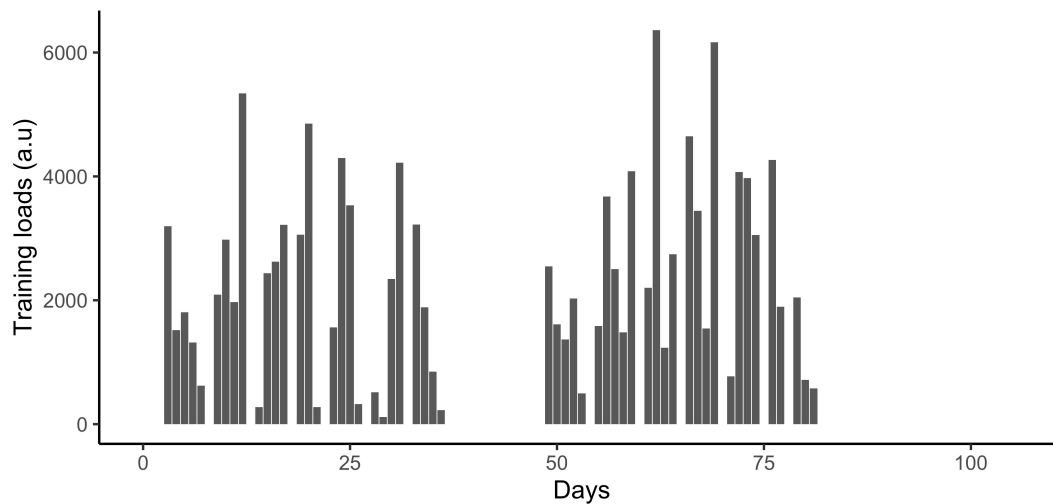
Applied to the observed data of this study a data set of $n = 248$ observations of performances associated with 19 independent variables was obtained (see Table 1). To formalise, allowing that $X \times Y \sim f$ with f a function of density, the built data set is a sample $S = \{(x_j, y_j)\}_{j \leq n} \sim f^n$.



(a)



(b)



(c)

Figure 1. Cumulative daily training loads of a representative athlete following (a) the impulse response function (X_2 , Equation 1) and (b) the serial bi-exponential response function (X_3 , Equation 2). (c) illustrates the raw daily training loads X_1 , expressed by $w(t)$. In (a) and (b), dots represent daily values of the cumulative training load and vertical solid lines indicate occurrence of training sessions. Values are represented in arbitrary units (a.u.).

Table 1. Summary of independent variables

Independent variables	X_i	Description	Aggregation
Raw training load	X_1	Daily training load computed from $TL_{ice}, TL_{RT}, TL_{aer}, TL_{RS}, TL_{aer}$ (see Supplementary material Appendix)	Daily recorded
Cumulative Training load	X_2, X_3	Daily computed from X_1 values	Impulse and serial cumulative responses
Rate of Perceived Exertion (RPE)	X_4, X_5	Borg category ratio (CR) 0-10 scale	Impulse and serial cumulative responses
Averaged power	X_6, X_7	On-ice sessions	Impulse and serial cumulative responses
Maximal power	X_8, X_9	On-ice sessions	Impulse and serial cumulative responses
Relative intensity	X_{10}, X_{11}	On-ice sessions (see Supplementary material Appendix , Equation S1)	Impulse and serial cumulative responses
Session duration	X_{12}, X_{13}	All sessions, overall session duration	Impulse and serial cumulative responses
Session density	X_{14}, X_{15}	All sessions, effective work only	Impulse and serial cumulative responses
Ice quality	X_{16}	Subjective information quoted on a Borg 0-10 CR scale	Recorded the day of performance
Rest	X_{17}	Rest between two consecutive sessions (days)	Sum of rest days preceding the performance
Past performance	X_{18}	Significantly correlated past performance $_{t-k}$ with performance $_t$	Performance at day $_{t-k}$
Athlete	X_{19}	Athlete's id	

Modelling Methodology

Formally, the goal is to find a function $h : X \rightarrow Y$ which minimises the generalisation error

$$R(h) = \mathbb{P}(h(X) \neq Y) = \mathbb{E}[\mathbb{1}[h(X) \neq Y]].$$

In practice the minimisation of R is unreachable. Instead, we get a sample set $S = (x_i, y_i)_{i \leq n} \in X \times Y$ and note the empirical error as

$$Re(h) = \frac{1}{n} \sum_i^n [\mathbb{1}[h(x_i) \neq y_i]].$$

The objective becomes to find the best estimate $\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} Re(h)$ with \mathcal{H} the class of function that we accept to consider.

Here, four family of models are evaluated in this context. With the exception of the DR, all models were individually and collectively computed (h_I and h_G , respectively).

Reference: Variable dose-response

The non-linear mathematical model developed by Busso¹³ was considered as the model of reference. Formally and according to the previously introduced notation, this model is a function $h^{(busso)} : X_1 \rightarrow Y$. It describes the training effects on performance over time, $y(t)$, from the raw training loads X_1 . TL are convoluted to a set of transfer functions $g_{\text{apt}}(t)$ and $g_{\text{fat}}(t)$, relating respectively to the aptitude and to the fatigue impulse responses as

$$g_{\text{apt}}(t) = e^{-\frac{t}{\tau_1}}$$

$$g_{\text{fat}}(t) = e^{-\frac{t}{\tau_2}},$$

with τ_1 and τ_2 two time constants. Combined with the basic level of performance y^* of the athlete, the modelled performance is

$$\hat{y}^{(busso)}(t) = y^* + k_1 (w * g_{\text{apt}})(t) - ((k_2 w) * g_{\text{fat}})(t),$$

with k_1 and $k_2(t)$ being gain terms. The later is related to the training doses by a second convolution to the transfer function

$$g_{\text{fat}'}(t) = e^{-\frac{t}{\tau_3}},$$

with τ_3 a time constant. Since is defined as $k_2(t) = k_3 (w * g_{\text{fat}'}) (t)$ where k_3 is a gain term, one may note that $k_2(t)$ increases proportionally to the training load and decay decreases exponentially from this new value. From discrete convolutions, the modelled performance can be rewritten as

$$\hat{y}^{(busso)}(t) = y^* + k_1 \sum_{l=1}^{t-1} w(l) e^{-\frac{-(t-l)}{\tau_1}} - \sum_{l=1}^{t-1} k_2(l) w(l) e^{-\frac{-(t-l)}{\tau_2}},$$

with $k_2(l) = k_3 \sum_{m=1}^l w(m) e^{-\frac{-(l-m)}{\tau_3}}$.

The five parameters of the model (i.e. k_1, k_3, τ_1, τ_2 and τ_3) are fitted by minimizing the residual sum of squares (RSS) between modelled and observed performances, such as

$$RSS = \sum_{t=1}^T (\hat{y}^{(\text{busso})}(t) - y(t))^2,$$

where $t \in T$ being the day in which the performance is measured. A non-linear minimisation was employed according to a Newton-type algorithm⁴².

Unlike this model of reference, the next presented models take benefit from the augmented data space $X^* = X \setminus X_1$.

Regularisation procedures

Elastic net. In highly dimensional contexts, multivariate linear regressions may lead to unsteady models by being excessively sensitive to the expanded space of solutions. To tackle this issue, cost functions penalising some parameters on account of correlated variables exist. On one side, Ridge penalisation reduces the space of possible functions by assigning a constraint to the parameters, thus minimising their amplitude to almost null values. On the other side, Least Absolute Shrinkage and Selection Operator (LASSO) penalisation has the capacity to fix parameters coefficient to null. The ENET regularisation combines both Ridge and LASSO penalisation³⁹. Hence, the multivariate linear model $h^{(\text{enet})} : X^* \rightarrow Y$ is

$$y_t^{(\text{enet})} = \mathbf{x}_t^t \beta + \varepsilon_t,$$

with $\mathbf{x} \in X^*$ the predictors, $\beta \in \mathbb{R}^d$ the parameters of the model and ε_t the error term. The regularisation stems from the optimisation of the objective

$$\min_{\beta \in \mathbb{R}^d} \frac{1}{2} \|y_t^{(\text{enet})} - y_t\|_2^2 + \lambda ((1 - \alpha) \|\beta\|_2^2 + \alpha \|\beta\|_1),$$

where $\alpha \in [0, 1]$ denotes the mixing parameter which defines the balance between the Ridge regularisation and the LASSO regularisation. λ denotes the impact of the penalty with $\lambda \rightarrow \infty$. For $\alpha = 0$ and $\alpha = 1$, the model will use a ridge and a lasso penalisation, respectively. Thus, for $\alpha \rightarrow 1$ and a fixed value of λ , the number of removed variables (null coefficients) increases with monotony from 0 to the LASSO most reduced model. The model was adjusted by hyper-parameters α and λ during the model selection, being part of the CV process (as described below).

Principal component regression. In this multivariate context with potential multicollinearity issues, principal component analysis aims to project the original data set from X^* into a new space \tilde{X}^* of orthogonal dimensions called principal components. These dimensions are built from linear combinations of the initial variables. One may use the principal components to regress the dependent variable: also known as Principal Components Regression (PCR). The regularisation is performed by using as regressors only the first principal components which retain the maximum of variance of the original data, by construction. In our study and according to the Kaiser's rule⁴³, p principal components with an eigenvalue higher than 1 were retained and further used in linear regression.

Such a model, $h^{(\text{pcr})} : \tilde{X}^* \rightarrow Y$, can be defined as a linear multivariate regression over principal components as

$$y_t^{(\text{pcr})} = \tilde{\mathbf{x}}_t^t \beta + \varepsilon_t,$$

with $\mathbf{x} \in \tilde{X}^* \setminus \{\tilde{X}_{p+1}^*, \dots, \tilde{X}_d^*\}$ the predictors, $\beta \in \mathbb{R}^p$ the parameters of the model and ε_t the error term. In addition to being a regularisation technique by using a subset of principal components only, PCR also exerts a discrete shrinkage effect on the low variance components (the lower eigenvalue components), nullifying their contribution in the original regression model.

Random Forest

Random Forest model consists of a large number of regression trees that operate as an ensemble. RF is random in two ways, (i) each tree is based on a random subset of observations and (ii) each split within each tree is created based on a random subset of candidate variables. The overall performance of the forest is defined by the average of predictions from the individual trees⁴⁴. In this study, random subset of variables and number of trees were the two hyper-parameters for adjusting the model within the model selection. The model is a function $h^{(\text{rf})} : X^* \rightarrow Y$.

Time series cross-validation and prediction

Data were separated –respectively to the time index– into one training data set for time series CV (80 % of the total data) and the remaining data for an unbiased model evaluation (evaluation data set). In this procedure, a model selection occurs first with the search of hyper-parameters values that minimise the predictive model error over validation subsets. The procedure is detailed in Algorithm 1.

Algorithm 1 Time series cross-validation

Require:

- A data set of n time ordered elements, $S = \{(x_j, y_j)\}_{j \leq m}$
- The number of partitions to split training data to, $K < m - s_{\min} - s_{\text{val}} + 1 \in \mathbb{N}$
- The minimum size of training set within a partition, $s_{\min} \geq 1$
- The size of validation set within a partition, $s_{\text{val}} \geq 1$
- A class of functions, \mathcal{H}

Ensure: An optimal model h^* fitted on S **for** $h^{(i)} \in \mathcal{H}$ **do****for** $k \in \{1, \dots, K\}$ **do**

- $t_{\text{val}} \leftarrow s_{\min} + k$
- $S_{\text{train}} \leftarrow (x_t, y_t)$ with $t \in [1, t_{\text{val}} - 1]$
- $S_{\text{valid}} \leftarrow (x_{t'}, y_{t'})$ with $t' \in [t_{\text{val}}, t_{\text{val}} + s_{\text{val}} - 1]$
- $h_{\text{trained}}^{(i)} \leftarrow \text{Train } h^{(i)} \text{ on } S_{\text{train}}$
- $E[i, k] \leftarrow \text{Evaluate RMSE of } h_{\text{trained}}^{(i)} \text{ on } S_{\text{valid}}$

end for**end for**

return $h^* = h^{(i^*)}$ with $i^* = \text{argmin}_i \{ \frac{1}{K} \sum_{k=1}^K E[i, k] \}$

Algorithm 1 iteratively evaluates a class of functions \mathcal{H} , in which each function $h^{(i)}$ differs from its hyper-parameters values. A time ordered data set S is partitioned into training and validation subsets (S_{train} and S_{valid} , respectively). For each partition k with $k \in \{1, \dots, K\}$, $h^{(i)}$ functions are fitted on the incremental S_{train} and evaluated on the fixed S_{valid} subset that occurs after the last element of S_{train} . Once $h^{(i)}$ functions are evaluated on K partitions of S , a function $h^{(i^*)}$ that provides the lowest and averaged root mean square error (*RMSE*) among validation subsets defines an optimal model denoted h^* .

Afterwards and for each partition of S , h^* is adjusted on new time ordered training subsets S'_{train} which combines both S_{train} and S_{valid} . Hence, the generalisation capability of h^* is evaluated on fixed length subsets of evaluation data, saved for that purpose.

Statistical analysis

For any model, the goodness of fit according to linear relationships and to performance were described by the coefficient of determination (R^2) and the *RMSE* criterion respectively. Their generalisation ability is described by the difference between *RMSE* computed on each training and testing data. The prediction error was reported through the Mean Absolute Error (*MAE*) between observations and predictions. After checking normality and variance homogeneity of the dependant variable by a Shapiro-wilk and a Levene test respectively, linear models were performed to assess differences in generalisation of models and differences of predictive performances for each group computed models compared to the DR model. For individual models comparisons, linear mixed models were used to account for subject variability in performances modelling by including athletes as random effect. Significance threshold was fixed to $p < 0.05$. Unstandardised regression coefficients β are reported along with 95% confidence interval (*CI*) as a measure of effect size. Models computation and statistical analysis were conducted with R statistical software (version 4.0.2). The DR model was computed with personal custom-built R package (version 1.0)⁴⁵.

Results

Through the times series CV, models provided heterogeneous generalisation and performance prediction. Distributions of *RMSE* per model are illustrated in Figure 2.

Models generalisation

Comparisons to the DR_I models showed significant differences and a greater generalisation capability for both ENET ($p = 0.012$, $\beta_{\text{diff}} = 0.027$ [0.007, 0.50] 95% *CI* and $p < 0.001$, $\beta_{\text{diff}} = 0.057$ [0.045, 0.069] 95% *CI*) and PCR models ($p = 0.005$, $\beta_{\text{diff}} = 0.031$ [0.010, 0.053] 95% *CI* and $p < 0.001$, $\beta_{\text{diff}} = 0.032$ [0.020, 0.044] 95% *CI* for individual and group

computed models respectively). In contrast, RF models reported a lower generalisation than the DR_I ($p = 0.032$, $\beta_{diff} = -0.023$ $[-0.045, -0.001]$ 95% CI and $p < 0.001$, $\beta_{diff} = -0.026$ $[-0.038, -0.014]$ 95% CI, for RF_I and RF_G respectively). Therefore, ENET and PCR models computed on overall data were the most generalisable models, followed by individual ENET and PCR models. A summary of model pairwise comparisons is provided in Table 2.

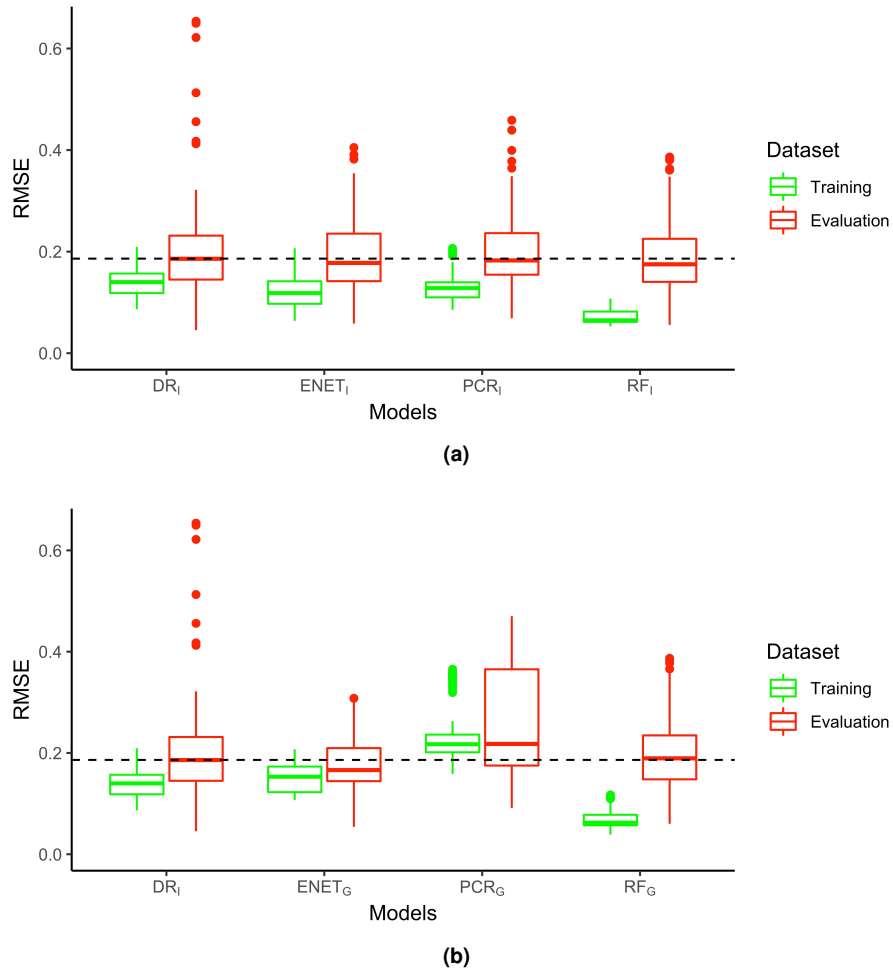


Figure 2. Distributions of models' performance. (a) shows RMSE distributions of each individual models and (b) the models computed on the whole group. Within boxplot midline represents the median of the distribution. All of them are compared to the dose-response (DR) model.

Table 2. Summary of models pairwise comparisons for generalisation and prediction abilities. β_{diff} represents the difference in parameter estimates between the DR model and the compared model.

Comparison	β_{diff}	CI_{lower}	CI_{upper}	$Pr(> t)$	Criterion
DR _I – ENET _G	0.057	0.045	0.069	< 0.001	Generalisation
DR _I – PCR _G	0.032	0.020	0.044	< 0.001	Generalisation
DR _I – PCR _I	0.031	0.010	0.053	0.005	Generalisation
DR _I – ENET _I	0.027	0.007	0.050	0.012	Generalisation
DR _I – RF _I	-0.023	-0.045	-0.001	0.032	Generalisation
DR _I – RF _G	-0.026	-0.038	-0.014	< 0.001	Generalisation
DR _I – ENET _G	0.032	0.016	0.048	< 0.001	Prediction
DR _I – ENET _I	0.025	0.004	0.047	0.020	Prediction
DR _I – RF _I	0.022	0.001	0.044	0.043	Prediction
DR _I – PCR _I	0.020	-0.002	0.041	0.072	Prediction
DR _I – RF _G	0.014	-0.002	0.029	0.088	Prediction
DR _I – PCR _G	-0.051	-0.067	-0.035	< 0.001	Prediction

Prediction performances

For any model, the observed RMSE on the evaluation data set showed significant differences to the reference DR_I, with the exception of PCR_I and RF_G ($p = 0.072$ and $p = 0.088$ respectively, see Table 2). The ENET_G model provided the greatest performances ($RMSE = 0.176 \pm 0.012$, $MAE = 0.15 \pm 0.01$ and $R^2 = 0.179 \pm 0.063$), closely followed by RF_G. Only PCR_G showed lower performances in prediction than DR_I. Distributions of $RMSE$ on data used for evaluation have shown heterogeneous variance between models. The greatest standard deviations were found for DR_I and PCR_G with $\sigma = 0.053$ and $\sigma = 0.062$ respectively. The ENET, PCR_I and RF models provided more consistent performances with lower standard deviations comprised within [0.023;0.027] and [0.012;0.017] intervals for individual and group computed models, respectively. Mean values of R^2 indicated that weak linear relationships between performance and predictors were identified by models ($R^2 \in [0.150;0.206]$). The highest averaged R^2 value but also the greatest standard deviations were reported for DR_I models ($R^2 = 0.206 \pm 0.093$). However, significant differences of averaged R^2 were only found for ENET_I, RF_G and PCR_G ($\beta = -0.056 [-0.10; -0.01]$ 95% CI, $p = 0.02$; $\beta = -0.041 [-0.08; -0.01]$ 95% CI, $p = 0.02$ and $\beta = -0.036 [-0.07; -0.01]$ 95% CI, $p = 0.04$ respectively). A summary of model performances is provided in Table 3. Predictions made from the two most generalisable models – ENET_G and PCR_G – and the reference DR_I illustrate the sensitivity of models for a representative athlete (Figure 3). Performances modelled from DR_I model were relatively steady and less sensitive to real performance variations. Standard deviation calculated on data used for model evaluation supported such a smooth prediction with $\sigma = 0.015$, $\sigma = 0.071$ and $\sigma = 0.062$ for DR_I, PCR_G and ENET_G, respectively.

Table 3. Summary of the predictive models. According to model families, criteria were averaged among folders and displayed with their standard deviation. For individual models, averaged values of hyper parameters are displayed along with lower and upper recorded values. The greatest performance among criteria is listed in bold type.* indicates the DR_I as the reference model and specification of its averaged parameters.

Model	R^2	MAE	RMSE	Hyper parameters*
DR _I *	0.206 ± 0.093	0.189 ± 0.055	0.225 ± 0.053	$k_1 = -3.95e-05, k_1 \in [-4.85e-05; -3.19e-05]$ $k_3 = -7.75e-09, k_3 \in [-4.01e-09; -1.71e-08]$ $\bar{\tau}_1 = 36.02, \tau_1 \in [25.82; 42.28], \bar{\tau}_2 = 22.57, \tau_2 \in [14.58; 26], \bar{\tau}_3 = 5.23, \tau_3 \in [4.33; 6.67]$ $\bar{\alpha} = 0.176, \alpha \in [0; 0.6], \bar{\lambda} = 0.273, \lambda \in [0; 1]$ $\bar{n}_{comp} = 1.918, n_{comp} \in [1; 3]$ $\bar{m}_{ry} = 8.90, m_{ry} \in [1; 17]$ $\alpha = 0.28, \lambda = 0.02$ $n_{comp} = 3$ $m_{ry} = 16$
ENET _I	0.150 ± 0.010	0.169 ± 0.020	0.197 ± 0.023	
PCR _I	0.164 ± 0.068	0.173 ± 0.025	0.201 ± 0.027	
RF _I	0.193 ± 0.074	0.170 ± 0.023	0.199 ± 0.024	
ENET _G	0.179 ± 0.063	0.150 ± 0.010	0.176 ± 0.012	
PCR _G	0.17 ± 0.053	0.22 ± 0.044	0.259 ± 0.062	
RF _G	0.164 ± 0.069	0.163 ± 0.017	0.195 ± 0.017	

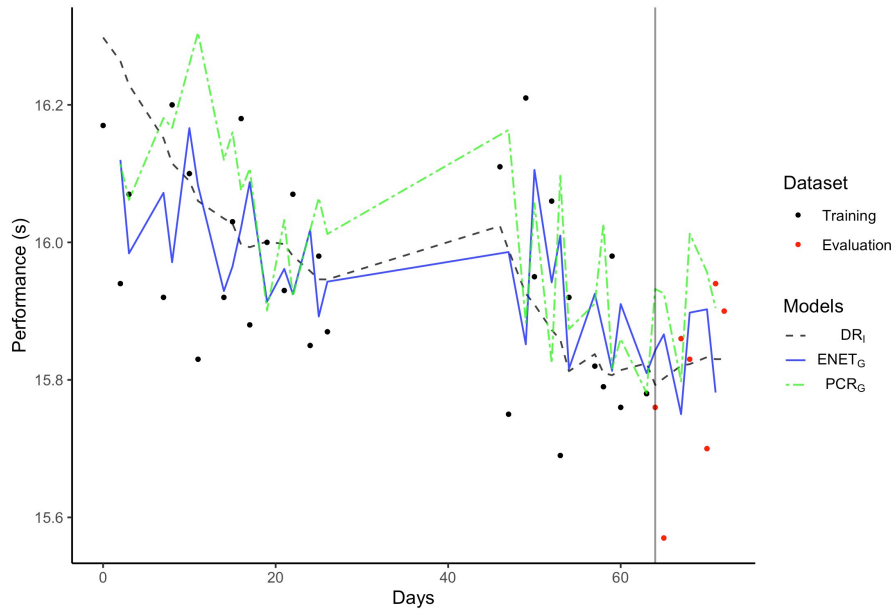


Figure 3. Modelled performance of a representative subject. Solid and dashed lines represent the DR model and the two most generalisable models on both training and evaluation data set, separated by the vertical solid line (80% and 20% of the total data respectively). Fitted parameters of the DR model were $k_1 = -2.45e-05$, $k_3 = -2.58e-09$, $\tau_1 = 39$, $\tau_2 = 26$, $\tau_3 = 5$. Hyper-parameters of the PCR and ENET models were $n_{comp} = 3$ and $\alpha = 0.28$, $\lambda = 0.02$.

Discussion

In the present study, we provided a modelling methodology that encompasses data aggregation relying on physiological assumptions and model validation for future predictions. Data were obtained from elite athletes, able of improving their performance by training being very sensitive to physical, psychological and emotional states. The variable dose-response model¹³ was fitted on individual data. It was compared to statistical and machine-learning models fitted on individual and on overall data: ENET, PCR and RF models.

Cross validation outcomes revealed significant heterogeneity in performances of models. The main criterion of interest, generalisation, was significantly greater for both ENET and PCR models than DR_I model. One can explain this result by the capabilities of the statistical models to better catch the underlying skating performance process using up to 19 independent variables when associated with regularisation methods. Conversely, the DR_I model relies on two antagonistic components strictly based on the training load dynamics. It does not deal with any other factors that may greatly impact the performance (e.g. psychological, nutritional, environmental, training-specific factors)^{12,18,46}. Thus, such a conceptual limit can be overtaken by employing multivariate modelling that may result in a greater comprehension of the training load - performance relationship, for the purpose of future predictions^{9,12}.

Distributions of RMSE from training and evaluation data sets allow us to establish a generalisation model ranking (Table 2). Linear models computed on overall data offer a better generalisation. This finding is essential because by handling the bias-variance trade-off, models are more suited for capturing a proper underlying function that maps inputs to the target even on unknown data. Hence, it allows further physiological and practical interpretations from the models. Sample size might be one source of explanation for such behaviour. It is well known that statistical inference on small samples leads to bad estimates and consequently to bad performances in prediction^{47,48}. A greater sample size obtained by combining individual data led to more accurate parameter estimates, being more suitable for sport performance modelling¹². This is particularly important to consider when we aim to predict a very few discipline specific performances throughout a season. However, predicting non-invasive physical quality assessments which can be daily performed (e.g. squat jumps and its variations for an indirect assessment of neuromuscular readiness⁴⁹, short sprints) may be an alternative for small sample size issues. Also, regularisation tends to stabilise parameters estimators and favour generalisation of the models. For instance, multicollinearity may occur in high-dimensional problems. Stochastic models generally suffer from such a conditioning. One would note that the ENET and PCR models attempt to overcome these issues in their own way by (i) penalising or removing features – or both – that are mostly linearly correlated and (ii) by projecting the initial data space onto a reduced space, which is optimised to keep the maximum of variance of the data from linear combinations of the initial features. Both approaches limit the number of unnecessary – or

noisy – dimensions. In contrast, in this study non-linear machine learning models expressed a lower generalisation capability than linear models even when models combine data from several athletes. We believe that such models may be powerful in multidimensional modelling but require an adequate data set with, in particular, ones with a sufficient sample size. Otherwise, model overfitting may occur at the expense of inaccurate predictions on unknown data.

As reported previously and with the exception of PCR_G , models were more accurate in prediction than DR_I (Table 3). The large averaged RMSE as well as large standard deviations provided by the DR_I among performance criteria tend to agree with the literature, since the model is prone to suffer from a weak stability and ill-conditioning raised by noisy data that impact its predictive accuracy^{9,10}. This evokes that linear relationships between the two components "Aptitude" – "Fatigue" and the performance are not clear. However, because of a lack of cross-validation procedures on impulse models and particularly the DR employed in our study, our results cannot be validly compared with the literature. Despite lower standard deviations of R^2 reported by ENET and PCR models, the weak averaged R^2 values suggest that linear models can only explain a few part of the total variance. Therefore and if the data allow it (i.e. a sufficient sample size and robustly collected data), non-linear models may still be effective and should be considered during the modelling process.

The sensitivity of models according to gains and losses of performances differed between the two most generalisable models – $ENET_G$ and PCR_G – and the reference DR_I . Such differences can be explained by the influence of variables that may affect performance, other than training loads dynamic (e.g. ice quality the day of performance, cumulative training loads following a serial and bi-exponential function, the last known performance) or a DR_I model failure in parameter estimates regarding to the variability of the data. However, this applied example does not inform us about neither the generalisation ability of models nor accuracy of predictions because it concerns only a particular set of data, where the selected models are trained on the first 80 % of data and evaluated on the 20 % remaining data.

This study presents some limits. The first one concerns the data we used and particularly the criterion of performance: standing start time trials few times a week during an approximately 3-months period. Even though being a very discipline specific test in which athletes are familiar and being conducted in standardised conditions, each test requires high levels of arousal, buy-in and motivation. Therefore, psychological states and cognitive functions monitoring such as motivation and attentional focus^{50,51} should have been done prior performing each trial. Secondly, the time series cross-validation presented here has a certain cost, most notably when only few data are available (e.g. when models are individually computed). The rolling origin re-calibration evaluation performed as described by Beirgmer *et al.*⁵² implies a model training only on a incremental sub-sequence of training data. Hence, the downsized sample size of the first training sub-sequences may cause model failure in parameter estimates and consequently, an increase of prediction errors. Then, training and evaluation data sets present some dependencies. In order to evaluate models on fully independent data, some modifications of the current CV framework exist at the expense of withdrawing even more data in the learning procedure. According to Racine⁵³, the so-called *hv - block cross-validation* is one of the least costly alternative to the CV used in our study, requiring a certain gap between each training and validation subsets. However, due to a limited sample size, we voluntary chose to not adapt the original CV framework described in Algorithm 1. Nonetheless, we recommend researchers and practitioners to consider such alternatives in case of significant dependencies and when sample size is sufficient.

Finally, *backtesting* was performed in order to evaluate model performances on historical data. From a practical point of view, models are able to predict the coming performance following a given feature of data known until day t . However, the contribution of training load responses modelling also concerns training after-effects simulations over a longer time frame. Having identified a suitable model, simulations of independent variables within their own distributions would allow practitioners and coaches to simulate changes in performance following objective and subjective measures of training loads, and any performance factors that are monitored. Conditional simulations that consider known relationships between independent variables (e.g. relationships between training load parameters)^{54,55} may improve the credibility of simulations.

Conclusion

In this study, we provided a transferable modelling methodology which relies on the evaluation of models generalisation ability in a context of sport performance modelling. The mathematical variable dose-response model along Elastic net, principal component regression and random forest models were cross-validated within a time series framework. Generalisation of the DR model was outperformed by ENET and PCR models. The ENET model provided the greatest performances both in terms of generalisation and accuracy in prediction. Increasing sample size by computing models on the whole group of athletes led to more performing models than the individually computed ones. The methodology highlighted in our study can be reemployed whatever the data, with the aim of optimising elite sport performance through training protocols simulations. Moreover, we believe that model validation is a requisite for any physiological interpretation for the purpose of making future predictions. Further researches that involve training session simulations and model evaluations in forecasting would highlight the relevance of some model families for training programming optimisation.

Author contributions statement

Conceptualisation, F.I., S.P., R.C. (Romain Chailan), R.C. (Robin Candau); methodology and investigation F.I., R.C. (Robin Candau), R.C. (Romain Chailan); recruitment T.M.; formal analysis and data curation F.I., R.C. (Robin Candau), T.M.; resources R.C. (Romain Chailan), T.M.; writing original draft preparation, F.I.; writing—review and editing, F.I., R.C. (Robin Candau), R.C. (Romain Chailan), S.P.; visualisation, F.I.; supervision, R.C. (Robin Candau), S.P.; project administration, R.C. (Robin Candau), S.P.; funding acquisition, F.I. All authors have read and agreed to the published version of the manuscript.

Additional information

Competing interests.

The authors declare no conflict of interest.

Funding.

This research was funded by the Association Nationale de la Recherche et de la Technologie (ANRT) grant number 2018/0653.

Acknowledgements We are grateful to the Fédération Française des Sports de Glace, the Institut National du Sport, de l'Expertise et de la Performance and to Dr. Anthony MJ Sanchez and Robert Solsona (Laboratoire Européen Performance Santé Altitude, University of Perpignan Via Domitia) for collaboration and sharing data sets.

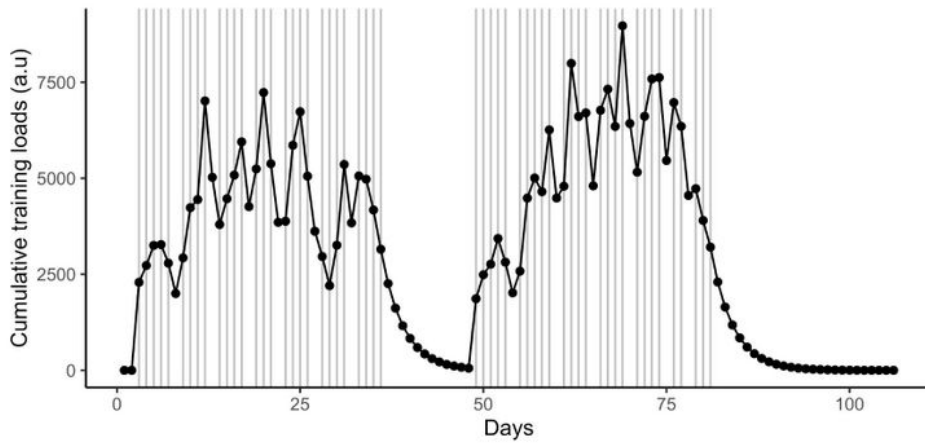
References

1. Wallace, L. K., Slattery, K. M. & Coutts, A. J. The ecological validity and application of the session-rpe method for quantifying training loads in swimming. *J. Strength Cond. Res.* **23**, 33–38 (2009).
2. Impellizzeri, F. M., Rampinini, E. & Marcora, S. M. Physiological assessment of aerobic training in soccer. *J. Sports Sci.* **23**, 583–592 (2005).
3. Banister, E., Calvert, T., Savage, M. & Bach, T. A systems model of training for athletic performance. *Aust. J. Sports Med.* **7**, 57–61 (1975).
4. Calvert, T. W., Banister, E. W., Savage, M. V. & Bach, T. A systems model of the effects of training on physical performance. *IEEE Trans. Syst. Man. Cybern. SMC* **6**, 94–102 (1976).
5. Banister, E. & Calvert, T. Planning for future performance: implications for long term training. *Can. J. Appl. Sport Sci.* **5**, 170–176 (1980).
6. Banister, E. W., Good, P., Holman, G. & Hamilton, C. L. *Modeling the training response in athletes*, vol. 3, 7–23 (Human Kinetics, 1986).
7. Busso, T., Carasso, C. & Lacour, J. Adequacy of a systems structure in the modeling of training effects on performance. *J Appl Physiol* **71**, 2044–2049 (1991).
8. Busso, T., Candau, R. & Lacour, J.-R. Fatigue and fitness modelled from the effects of training on performance. *Eur. J. Appl. Physiol. Occup. Physiol.* **69**, 50–54 (1994).
9. Hellard, P. *et al.* Assessing the limitations of the banister model in monitoring training. *J. Sports Sci.* **24**, 509–520 (2006).
10. Ludwig, M., Asteroth, A., Rasche, C. & Pfeiffer, M. Including the past: Performance modeling using a preload concept by means of the fitness-fatigue model. *Int. J. Comput. Sci. Sport* **18**, 115–134 (2019).
11. Busso, T., Denis, C., Bonnefoy, R., Geysant, A. & Lacour, J.-R. Modeling of adaptations to physical training by using a recursive least squares algorithm. *J. Appl. Physiol.* **82**, 1685–1693 (1997).
12. Avalos, M., Hellard, P. & Chatard, J.-C. Modeling the training-performance relationship using a mixed model in elite swimmers. *Med. Sci. Sports Exerc.* **35**, 838 (2003).
13. Busso, T. Variable dose-response relationship between exercise training and performance. *Med. Sci. Sports Exerc.* **35**, 1188–1195 (2003).
14. Kolossa, D. *et al.* Performance estimation using the fitness-fatigue model with kalman filter feedback. *Int. J. Comput. Sci. Sport* **16**, 117–129 (2017).
15. Matabuena, M. & Rodríguez-López, R. An improved version of the classical banister model to predict changes in physical condition. *Bull. Math. Biol.* **81**, 1867–1884 (2019).
16. Morton, R., Fitz-Clarke, J. & Banister, E. Modeling human performance in running. *J. Appl. Physiol.* **69**, 1171–1177 (1990).
17. Candau, R., Busso, T. & Lacour, J. Effects of training on iron status in cross-country skiers. *Eur. J. Appl. Physiol. Occup. Physiol.* **64**, 497–502 (1992).
18. Mujika, I. *et al.* Modeled responses to training and taper in competitive swimmers. *Med. Sci. Sports Exerc.* **28**, 251–258 (1996).
19. Millet, G. *et al.* Modelling the transfers of training effects on performance in elite triathletes. *Int. J. Sports Med.* **23**, 55–63 (2002).
20. Millet, G., Gros Lambert, A., Barbier, B., Rouillon, J. & Candau, R. Modelling the relationships between training, anxiety, and fatigue in elite athletes. *Int. J. Sports Med.* **26**, 492–498 (2005).
21. Thomas, L., Mujika, I. & Busso, T. A model study of optimal training reduction during pre-event taper in elite swimmers. *J. Sports Sci.* **26**, 643–652 (2008).
22. Sanchez, A. M. *et al.* Modelling training response in elite female gymnasts and optimal strategies of overload training and taper. *J. Sports Sci.* **31**, 1510–1519 (2013).
23. Agostinho, M. F. *et al.* Perceived training intensity and performance changes quantification in judo. *J. Strength Cond. Res.* **29**, 1570–1577 (2015).
24. Busso, T. & Thomas, L. Using mathematical modeling in training planning. *Int. J. Sports Physiol. Perform.* **1**, 400–405 (2006).

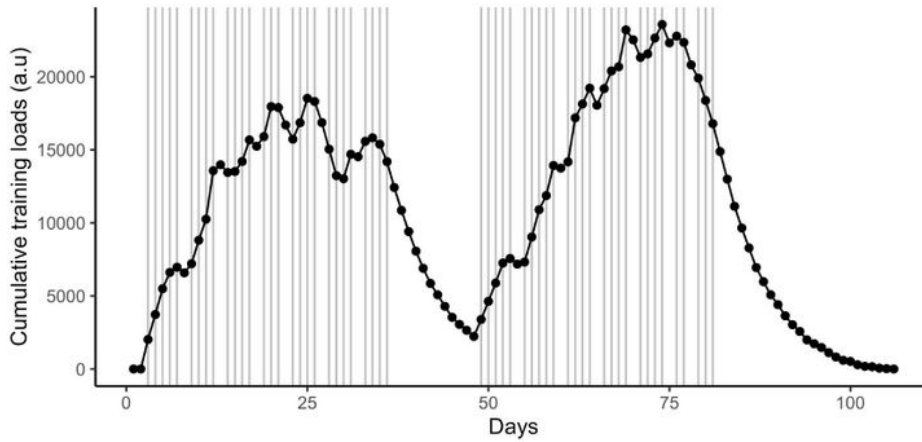
25. Begue, G. *et al.* Early activation of rat skeletal muscle il-6/stat1/stat3 dependent gene expression in resistance exercise linked to hypertrophy. *PLoS one* **8**, e57141 (2013).
26. D'Antona, G. *et al.* Skeletal muscle hypertrophy and structure and function of skeletal muscle fibres in male body builders. *J. Physiol.* **570**, 611–627 (2006).
27. Roels, B. *et al.* Paradoxical effects of endurance training and chronic hypoxia on myofibrillar atpase activity. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294**, R1911–R1918 (2008).
28. Philippe, A. G., Borrani, F., Sanchez, A. M., Py, G. & Candau, R. Modelling performance and skeletal muscle adaptations with exponential growth functions during resistance training. *J. Sports Sci.* **37**, 254–261 (2019).
29. Arlot, S., Celisse, A. *et al.* A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79 (2010).
30. Kouvaris, K., Clune, J., Kounios, L., Brede, M. & Watson, R. A. How evolution learns to generalise: Using the principles of learning theory to understand the evolution of developmental organisation. *PLoS Comput. Biol.* **13**, e1005358 (2017).
31. Lever, J., Krzywinski, M. & Altman, N. Points of significance: model selection and overfitting. *Nat. Methods* **13**, 703–704 (2016).
32. Mitchell, L. J., Rattray, B., Fowlie, J., Saunders, P. U. & Pyne, D. B. The impact of different training load quantification and modelling methodologies on performance predictions in elite swimmers. *Eur. J. Sport Sci.* 1–10 (2020).
33. Stephens Hemingway, B. H., Burgess, K. E., Elyan, E. & Swinton, P. A. The effects of measurement error and testing frequency on the fitness-fatigue model applied to resistance training: A simulation approach. *Int. J. Sports Sci. Coach.* **15**, 60–71 (2020).
34. Chalencon, S. *et al.* Modeling of performance and ans activity for predicting future responses to training. *Eur. J. Appl. Physiol.* **115**, 589–596 (2015).
35. Edelmann-Nusser, J., Hohmann, A. & Henneberg, B. Modeling and prediction of competitive performance in swimming upon neural networks. *Eur. J. Sport. Sci.* **2**, 1–10 (2002).
36. Carrard, J., Kloucek, P. & Gojanovic, B. Modelling training adaptation in swimming using artificial neural network geometric optimisation. *Sports* **8**, 8 (2020).
37. Lek, S. & Guégan, J.-F. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol. Model.* **120**, 65–73 (1999).
38. Qi, Y. Random forest for bioinformatics. In *Ensemble machine learning*, 307–323 (Springer, 2012).
39. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
40. Kosmidis, I. & Passfield, L. Linking the performance of endurance runners to training and physiological effects via multi-resolution elastic net (2015). Preprint at <https://arxiv.org/abs/1506.01388>.
41. Felser, S. *et al.* Relationship between strength qualities and short track speed skating performance in young athletes. *Scand. J. Med. Sci. Sports* **26**, 165–171 (2016).
42. Dennis Jr, J. E. & Schnabel, R. B. *Numerical methods for unconstrained optimization and nonlinear equations* (SIAM, 1996).
43. Kaiser, H. F. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **20**, 141–151 (1960).
44. Grömping, U. Variable importance assessment in regression: linear regression versus random forest. *Am. Stat.* **63**, 308–319 (2009).
45. Imbach, F. sysmod : An R package for dose-response modelling in sports. <https://github.com/fimbach/sysmod> (2020).
46. Stone, M. H., Stone, M. & Sands, W. A. *Principles and practice of resistance training* (Human Kinetics, 2007).
47. Kelley, K. & Maxwell, S. E. Sample size for multiple regression: obtaining regression coefficients that are accurate, not simply significant. *Psychol. Methods* **8**, 305 (2003).
48. Cui, Z. & Gong, G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage* **178**, 622–637 (2018).
49. Watkins, C. M. *et al.* Determination of vertical jump as a measure of neuromuscular readiness and fatigue. *J. Strength Cond. Res.* **31**, 3305–3310 (2017).
50. Gillet, N. *et al.* Examining the motivation-performance relationship in competitive sport: a cluster-analytic approach. *Int. J. Sport Exerc. Psychol.* **43**, 79 (2012).

51. Ille, A., Selin, I., Do, M.-C. & Thon, B. Attentional focus effects on sprint start performance as a function of skill level. *J. Sports Sci.* **31**, 1705–1712 (2013).
52. Bergmeir, C. & Benítez, J. M. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **191**, 192–213 (2012).
53. Racine, J. Consistent cross-validated model-selection for dependent data: hv-block cross-validation. *J. Econ.* **99**, 39–61 (2000).
54. Noble, B. J., Borg, G. A., Jacobs, I., Ceci, R. & Kaiser, P. A category-ratio perceived exertion scale: relationship to blood and muscle lactates and heart rate. *Med. Sci. Sports Exerc.* **15**, 523 (1983).
55. Casamichana, D., Castellano, J., Calleja-Gonzalez, J., San Román, J. & Castagna, C. Relationship between indicators of training load in soccer players. *J. Strength Cond. Res.* **27**, 369–374 (2013).
56. Van Ingen Schenau, G. J. The influence of air friction in speed skating. *J. Biomech.* **15**, 449–458 (1982).
57. De Koning, J. J., De Groot, G. & van Ingen Schenau, G. J. Ice friction during speed skating. *J. Biomech.* **25**, 565–571 (1992).

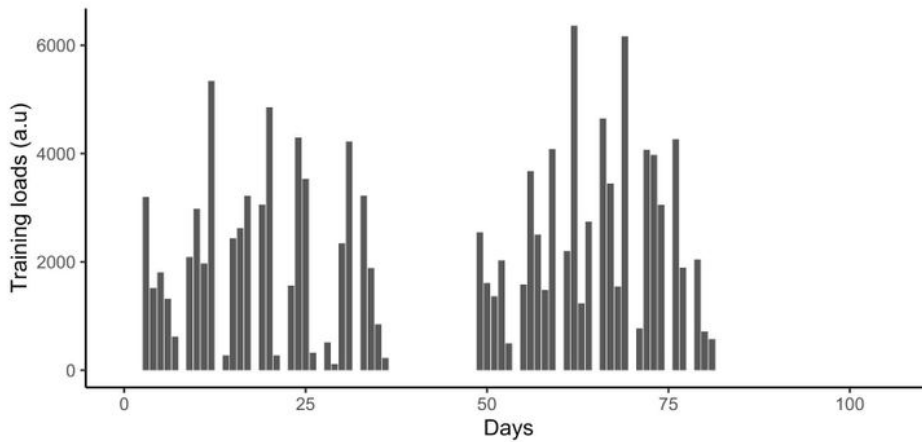
Figures



(a)



(b)

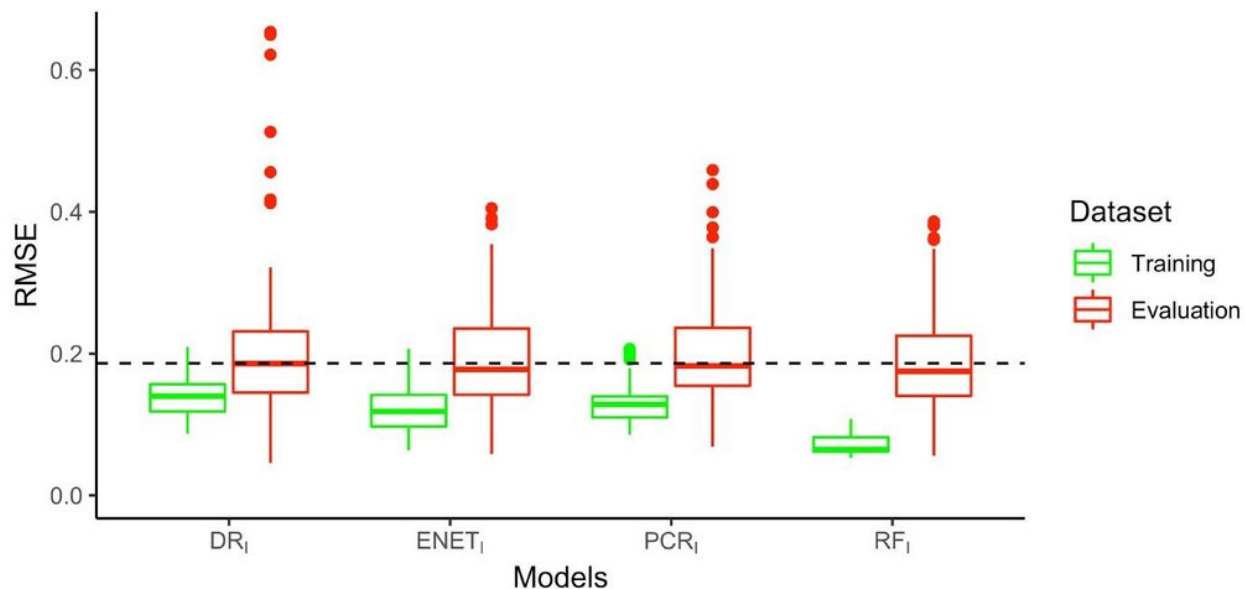


(c)

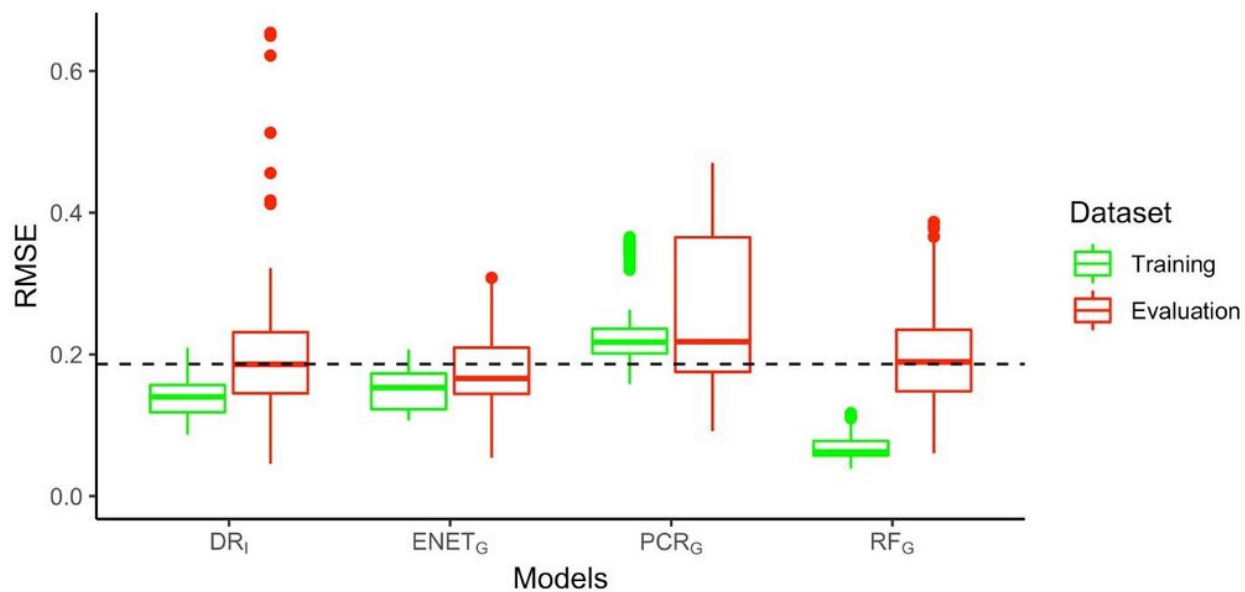
Figure 1

Cumulative daily training loads of a representative athlete following (a) the impulse response function (X2, Equation 1) and (b) the serial bi-exponential response function (X3, Equation 2). (c) illustrates the raw daily training loads X1, expressed by $w(t)$. In (a) and (b), dots represent daily values of the cumulative

training load and vertical solid lines indicate occurrence of training sessions. Values are represented in arbitrary units (a.u).



(a)



(b)

Figure 2

Distributions of models' performance. (a) shows RMSE distributions of each individual models and (b) the models computed on the whole group. Within boxplot midline represents the median of the distribution. All of them are compared to the dose-response (DR) model.

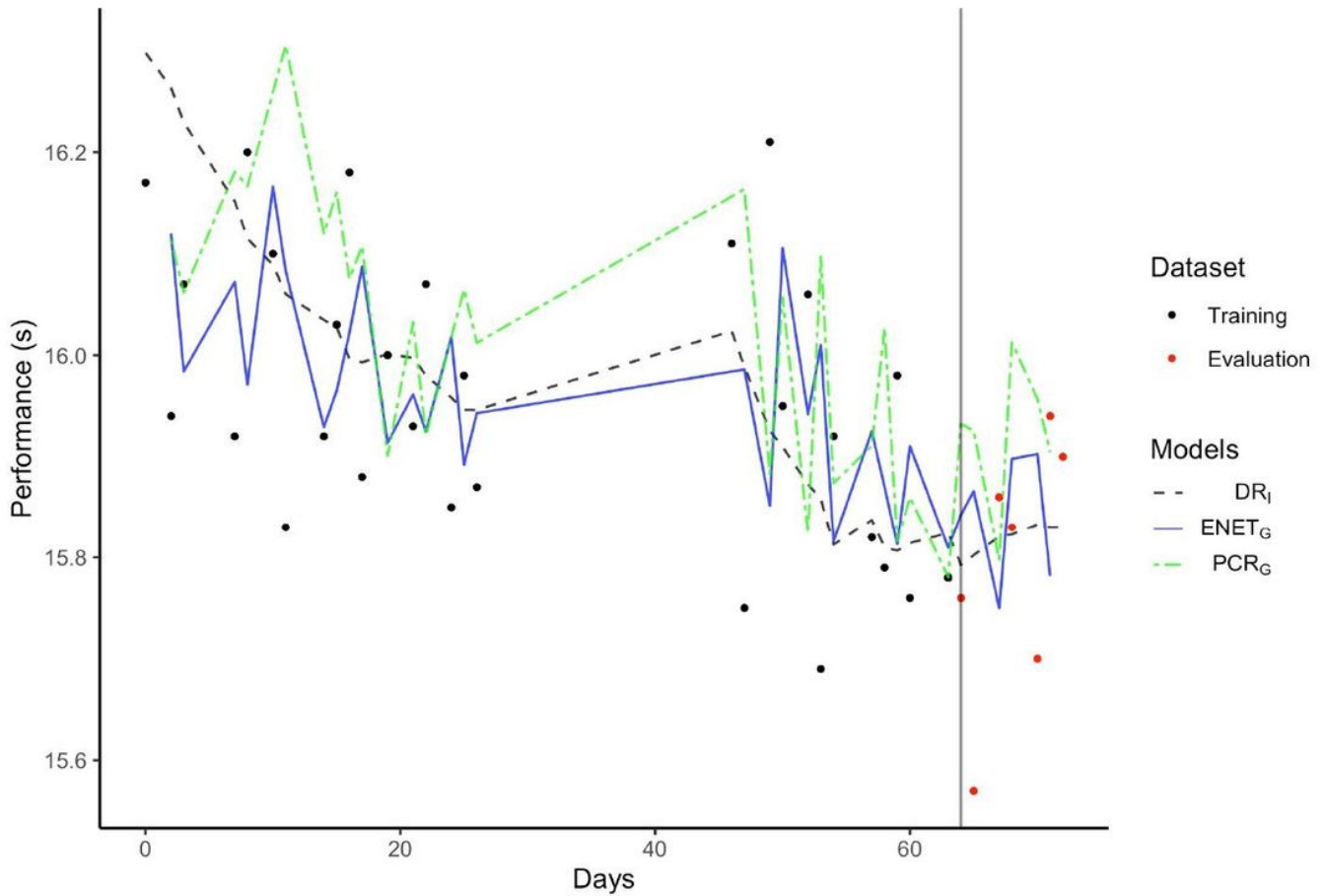


Figure 3

Modelled performance of a representative subject. Solid and dashed lines represent the DR model and the two most generalisable models on both training and evaluation data set, separated by the vertical solid line (80% and 20% of the total data respectively). Fitted parameters of the DR model were $k_1 = -2.45e-05$, $k_3 = -2.58e-09$, $\tau_1 = 39$, $\tau_2 = 26$, $\tau_3 = 5$. Hyper-parameters of the PCR and ENET models were $n \text{ comp} = 3$ and $\alpha = 0.28$, $\lambda = 0.02$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [appendix.pdf](#)