

Traditional and contemporary approaches to mathematical fitness-fatigue models in exercise science: A practical guide with resources. Part I.

Ben Stephens Hemingway^{1,2}, Leon Greig¹, Mladen Jovanovic³, Ben Ogorek⁴, Paul Swinton^{1*}

DOI 10.31236/OSF.IO/AP75J

SportRxiv hosted preprint version 4

Last updated 26/02/2021

PREPRINT - NOT PEER REVIEWED

Institutions

¹ School of Health Sciences, Robert Gordon University, Aberdeen, Scotland, UK

² School of Computing Science and Digital Media, Robert Gordon University, Aberdeen, Scotland, UK

³ Faculty of Sport and Physical Education, University of Belgrade, Belgrade, Serbia

⁴ Noursot, Chicago, Illinois, USA

* Corresponding author:

Dr Paul Swinton

School of Health Sciences

Robert Gordon University

Aberdeen, UK, AB10 7QE

p.swinton@rgu.ac.uk

Disclosure of funding

No sources of funding were used to assist in the preparation of this article.

Conflicts of interest

BSH, LG, MJ, BO, and PS declare they have no conflicts of interest relevant to the content of this review.

Acknowledgements

The authors would like to thank Christian Rasche and Mark Pfeiffer (Johannes Gutenberg University of Mainz, Mainz, Germany), Mark Connor (UCD Smurfit, Dublin, Ireland), and Thierry Busso (Université Jean Monnet, Saint-Etienne, France) for their insightful feedback during the early refinement of this article following initial publication in pre-print leading to this version 4 update.

Supplementary material

Available at fitnessfatigue.com - Refer to section 1 for an introduction to these resources, and appendix B for further information.

Maintained by b.stephens-hemingway@rgu.ac.uk and baogorek@gmail.com

Keywords

Banister's model, fitness-fatigue, TRIMP, performance modelling, impulse-response, dose-response

Author agreement statement

All authors are aware and agree that this work can be submitted to SportRxiv for publication in preprint form.

Preferred Citation

Stephens Hemingway, B., Greig, L., Jovanovic, M., Ogorek, B., & Swinton, P. (2021). Traditional and contemporary approaches to mathematical fitness-fatigue models in exercise science: A practical guide with resources. Part I. *SportRxiv (Preprint)*. <https://doi.org/10.31236/osf.io/ap75j>

Abstract

The standard fitness-fatigue model (FFM) was developed more than 40 years ago and remains a prominent conceptual model within exercise science. Translation from a conceptual form into a mathematical structure reveals a collection of models (FFMs) with common properties linking training inputs to performance outcomes. The greatest potential use of FFMs is to predict future performance of athletes with sufficient accuracy to assist with training program design and planning key tapering periods. However, despite a long history and consistent study, there has been limited uptake in practice. This is most likely due to a lack of accessible resources explaining key concepts and processes required to fit models, and uncertainty regarding overall predictive validity. Part I of this review series provides a comprehensive overview of FFMs and discusses three key aspects of fitting models including: 1) training load quantification; 2) criterion performance selection; and 3) parameter estimation. As the majority of athletes engage in sports where performance is complex and determined by a range of physical, psychological and technical factors, it is argued that FFMs may be best placed to quantify components of fitness (e.g., strength, power) targeted by an athlete's training, rather than competitive outcomes. Additionally, contemporary approaches to training monitoring (e.g., measurement of barbell velocity and repetitions in reserve) are recommended as tools to generate high frequency "performance" measures to better fit FFMs. Essential further developments require collaboration between researchers and practitioners with larger data sets to establish conditions where accurate predictions to future training could be obtained. A set of code-based resources in the R programming language are included with the review series to assist in fitting and evaluating models, and to enhance understanding of concepts described.

Contents

1.0 Introduction.....	3
2.0 Historical Development.....	4
2.1 The standard FFM and fitness-delay extension.....	4
2.2 The general model.....	7
2.3 The variable baseline performance model.....	8
2.4 Further developments to model structure.....	9
3.0 Application of fitness-fatigue models in research and practice	9
3.1 Training load quantification.....	9
3.2 Criterion performance selection	11
3.3 Parameter estimation approaches and limitations.....	12
3.4 Model evaluation: Integrating practitioner data into future research efforts	14
3.5 The utility of an FFM: Informing training program design.....	15
4.0 Conclusions and practical recommendations	16
References	19
Appendix A: Tables of nomenclature, abbreviations, and key terms	22
Appendix B: Supplementary code resources	23
Appendix C: Derivation of solutions to fitness-fatigue systems.....	24
Appendix D: Physiological correlates of model fitness and fatigue traces.....	27
Appendix E: Model formulae reference table.....	28

1.0 Introduction

Fitness-fatigue models (FFMs) underpin concepts guiding the theory and practice of physical training in exercise science¹. In general, FFMs posit that a single bout of training creates two antagonistic after-effects including a positive long-lasting and low magnitude fitness effect, and a negative short-lasting and high-magnitude fatigue effect². It is the combination of these antagonistic components that is said to describe an individual's current level of physical performance over time. In conceptual form, FFMs are frequently used to guide decisions around exercise prescription, recovery, and tapering strategies¹. However, FFMs can also be expressed as multi-parameter mathematical models, with five parameters in the standard formulation^{2,3}, and reaching upward of eight or more parameters in complex representations⁴⁻⁸. The initial mathematical framework was devised by Banister and colleagues over four decades ago², with the underlying mathematical structure revealing a collection of related models with similar properties^{2-4,6-13}. All FFMs take as input a series of quantified daily training doses and frequently measured performance values and attempt to model, and ultimately predict, performance across time². Individual tailoring is achieved by setting parameters in the model to match elements such as the magnitude and decay rate of the positive and negative after-effects experienced by the individual. This is achieved by conducting a period of *model fitting*, where frequent performance measurements are collected, and parameters retrospectively fit to best match input and output data. If a suitable model can be fit, FFM predictions may assist practitioners with training program design and athlete monitoring^{2,6,14-16}. Despite the central position of FFMs in training theory and their long history and study by researchers, there has been little uptake across contemporary research and in real world practical environments. This is likely explained by a combination of factors including: 1) inaccessibility of research that assumes readers are proficient with requisite mathematical concepts and have an awareness of historical conventions, nomenclature and terminology (tables provided in appendix A); 2) a lack of accessible resources featuring clear methods, and simple tools for model implementation and evaluation; and 3) limited empirical research evaluating key aspects such as prediction accuracy, model stability, and appropriate methods for parameter estimation across diverse scenarios. Previous research has summarised various aspects of FFMs and provided a useful entry point for those interested^{1,17-22}. However, no up to date resource exists that comprehensively summarises historical developments, methods, and limitations of FFMs. Similarly, few simple and flexible tools exist to assist researchers fit, evaluate and investigate FFMs using their own data. Researchers have typically developed their own implementations exclusive to their specific needs, limiting uptake of several potentially useful resources^{6,14,20}. Therefore, the aim of part I of this review series is to synthesise existing information regarding FFMs and their potential limitations, and balance this with practical considerations surrounding their prospective application in exercise science. Key mathematical content is included but accompanied by applied insight, considerations, and recommendations, to increase usefulness for both researchers and practitioners. A secondary aim of part I of this review series is to simplify and elucidate the process of building, modifying, and evaluating basic FFMs for use in research and teaching. Toward this end, supplementary code resources have been developed, and are available at <http://fitnessfatigue.com/>. Included are robust utilities written in the R²³ programming language for fitting and evaluating FFMs with modern optimisers, cross-validation, and input checking. In part II of this review series²⁴, more advanced contemporary methods are introduced to account for some of the limitations of basic FFMs. For the more advanced models didactic code has been developed to guide users through each of the methods (e.g., non-linear models and inclusion of the Kalman filter) via an interactive procedural flow. Thorough documentation, explanatory examples, and mock data are provided within the repository for the utilities and didactic code (see appendix B for further details). The nature of the open-source repository in which these resources are kept encourages others to contribute to new or existing models, and promotes open research practices²⁵. Given the vast amount of data that practitioners and researchers generate, increased usage of existing data and intensified dissemination of information is required to better develop and validate models. Training theory and design has developed considerably since Banister and colleagues first developed the standard FFM². This increased knowledge, combined with a rapid increase in volume of data collected by practitioners over the last decade presents unique opportunities to accelerate development of data-driven modelling approaches to assist with individualised

prescription. Collectively, the two parts of this review highlight several future developments that could be considered by collaborating groups of practitioners and researchers to better align FFM with current training theory.

2.0 Historical Development

2.1 The standard FFM and fitness-delay extension

In 1975, Banister and colleagues proposed the first systems model describing athletic performance in terms of training undertaken and fatigue accumulated ^{2,3}. The authors goal was to ultimately develop a predictive model of training that could improve understanding of relevant physiological processes, and in turn support selection of appropriate training programs ². The authors envisioned a complex systems model that could account for several basic determinants of performance such as psychology, skill, and physical capacity. As a first step toward a more complete systems model, the authors isolated training of physical capability, and then attempted to mathematically quantify the relationship with athletic performance. We refer to Banister and colleagues' seminal model as the standard FFM where physical training was defined in terms of training impulses (TRIMP ω) that exert an immediate positive and negative training effect that subsequently decay at different rates ². Under the model structure, each training session can be thought of as "dumping" a quantity of positive fitness and fatigue proportional to the training impulse ω ; and it is generally assumed that the initial inflow of fatigue is greater than fitness. Immediately, these quantities begin to decay exponentially with a slower rate of decay for fitness (Figure 1). At any time t , the total amount of fitness, denoted by the function $g(t)$, and fatigue, denoted by the function $h(t)$, are equal to the sum of the remaining quantities for each component. The difference between these positive functions represents the model prediction of performance $\hat{p}(t)$ which is hoped to reflect actual performance $p(t)$ throughout the training period (eq.1) (Figure 2).

$$\hat{p}(t) = p^* + k_g g(t) - k_h h(t), \quad g(t) \geq 0, h(t) \geq 0, k_h \geq k_g > 0 \quad (eq. 1)$$

Where parameter p^* acts as an additive term (i.e., an intercept) of the model and can be interpreted as an estimation of an individual's baseline performance level. Parameters k_g and k_h are the scaling factors that determine the magnitude of the fitness and fatigue realised after a single training session. The dynamical system at the core of the standard model is specified by the following system of first-order differential equations:

$$\begin{aligned} g'(t) &= \omega(t) - \frac{1}{\tau_g} g(t) \\ h'(t) &= \omega(t) - \frac{1}{\tau_h} h(t) \end{aligned} \quad (eq. 2)$$

Where $\omega(t) \geq 0$, and $\tau_g, \tau_h \geq 0$ are the parameters that describe the decay rate of fitness and fatigue, respectively. Initial conditions of the system are often set at $g(0) = h(0) = 0$ for convenience, and this approximation may be appropriate if the athlete has not engaged in strenuous training for an extended period. Larger values of $\omega(t)$ represent greater training doses and therefore a greater stimulus for adaptation. An explicit solution to this linear first-order independent system of differential equations is obtained by convolving the training impulse series $\omega(t)$ with exponential functions as follows:

$$\begin{aligned} g(t) &= \omega(t) * e^{\frac{-t}{\tau_g}} \\ &= \int_0^t e^{\left(\frac{-(t-u)}{\tau_g}\right)} \omega(u) du \end{aligned} \quad (eq. 3)$$

$$\begin{aligned} h(t) &= \omega(t) * e^{\frac{-t}{\tau_h}} \\ &= \int_0^t e^{\left(\frac{-(t-u)}{\tau_h}\right)} \omega(u) du \end{aligned} \quad (eq. 4)$$

Where the symbol $*$ denotes the convolution product. Readers are directed to appendix C for a complete mathematical derivation. In practice, eq.3 and eq.4 are generally approximated by the following functions with continuous time (t) replaced by discrete time (n) representative of days:

$$g(n) = \Delta_t \sum_{i=0}^{n-1} \omega(i) e^{\frac{-(n-i)}{\tau_g}}, \quad (eq.5)$$

$$h(n) = \Delta_t \sum_{i=0}^{n-1} \omega(i) e^{\frac{-(n-i)}{\tau_h}}, \quad (eq.6)$$

Where Δ_t is the discrete time step-size, typically set to one day. Fitness or fatigue at any given time point can be considered the sum of prior training convolved with the relevant decay function, and therefore the effect of a once-daily training impulse $\omega(n) \geq 0$ on performance is described by the following discrete multiparameter series approximation ^{8,10,11}:

$$\hat{p}(n) = p^* + \underbrace{k_g \sum_{i=1}^{n-1} \omega(i) \cdot e^{\frac{-(n-i)}{\tau_g}}}_{\text{fitness component}} - \underbrace{k_h \sum_{i=1}^{n-1} \omega(i) \cdot e^{\frac{-(n-i)}{\tau_h}}}_{\text{fatigue component}} \quad (eq.7)$$

Figure 1: Impulse-response dynamics of the standard model for a single arbitrary training dose

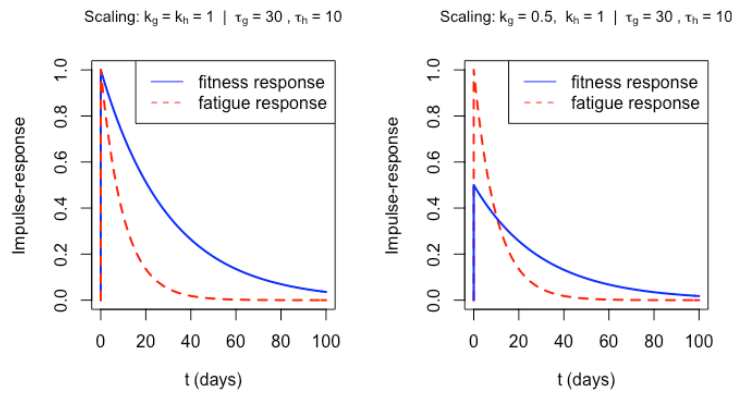
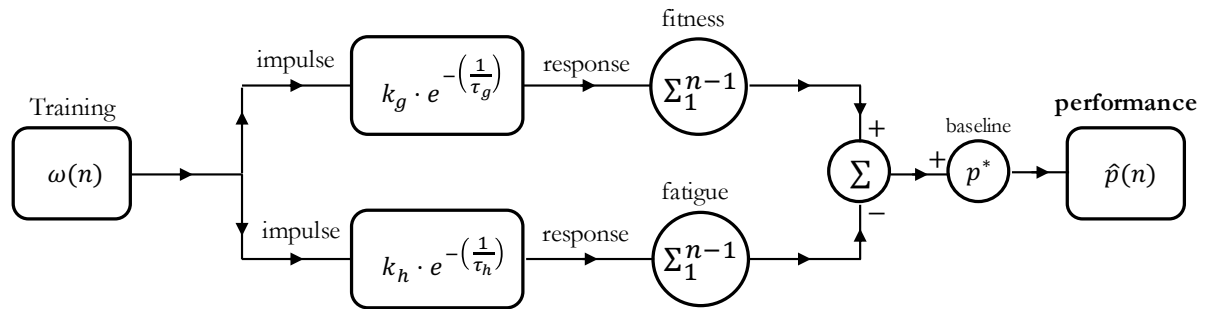


Figure 2: Diagrammatic illustration of the standard model function



The solution to the model system (eq.2) may also be stated in recursive form as follows:

$$\hat{p}(n) = p^* + k_g g(n) - k_h h(n) \quad (eq.8)$$

$$g(n) = [g(n-1) + \omega(n-1)] e^{\frac{-1}{\tau_g}} \quad g(0) \geq 0. \quad (eq.9)$$

$$h(n) = [h(n-1) + \omega(n-1)] e^{\frac{-1}{\tau_h}} \quad h(0) \geq 0 \quad (eq.10)$$

The relative magnitudes of the scaling factors (k_g, k_h) and the decay time constants (τ_g, τ_h) determine the primary dynamics of the model. If a relationship $\tau_g > \tau_h \geq 0$ holds, a net positive increase in performance will be obtained at points in the training cycle provided there is sufficient recovery time. The values of the scaling factors depend on the units used to measure the training load and performance, and frequently values are constrained so that $k_h > k_g > 0$ ¹⁰. Combining both sets of commonly applied constraints, modelled performance is suppressed during and immediately following periods of high training stress until fatigue dissipates sufficiently and fitness effects are revealed². This model behaviour (commonly referred to as supercompensation) is theoretically consistent with observed effects of high demand training on performance within a normal training cycle, or during short term periods of overreaching. However, these dynamics may not be representative for some groups of elite athletes where training is focussed on maintenance of existing physical performance levels rather than seeking or expecting large increases. Finally, the parameter p^* in eq.7 and eq.8 remains the common additive term that can be thought of conceptually as a base level of performance that the individual is not expected to fall chronically below, even after a prolonged absence of training.

A key limitation of the dynamics of the standard FFM is the unrealistic response of the fitness component with a maximum value obtained immediately following training^{3,5,13}. Noting this limitation, a year after their seminal model², Bannister and colleagues³ developed a simple extension introducing a delay term into the fitness component (eq.11). The extension created a period of non-linear growth to a delayed peak, before undergoing exponential decay (figure 3A). Despite the conceptual attractiveness of this extension^{3,10,26}, the model has rarely been investigated.

$$\hat{p}(n) = p^* + k_g \cdot \sum_{i=1}^{n-1} \omega(i) \cdot \left(e^{\frac{-(n-i)}{\tau_{g1}}}_{\text{effect}} - e^{\frac{-(n-i)}{\tau_{g2}}}_{\text{delay}} \right) - k_h \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{\frac{-(n-i)}{\tau_h}} \quad (\text{eq. 11})$$

Although the underlying system is not clearly stated in the authors' paper³, it can be determined from the transfer functions provided that the model arises from an independent system of differential equations comprising a second-order differential equation for fitness and a first-order differential equation for fatigue (see appendix C).

Time to reach maximum fitness response ($t_{g_{max}}$) to any single training load can be obtained by setting the derivative of the two fitness exponential terms to zero and solving for t to give:

$$t_{g_{max}} = \frac{\tau_{g1} \tau_{g2} \ln \frac{\tau_{g1}}{\tau_{g2}}}{\tau_{g1} - \tau_{g2}} \quad (\text{eq. 12})$$

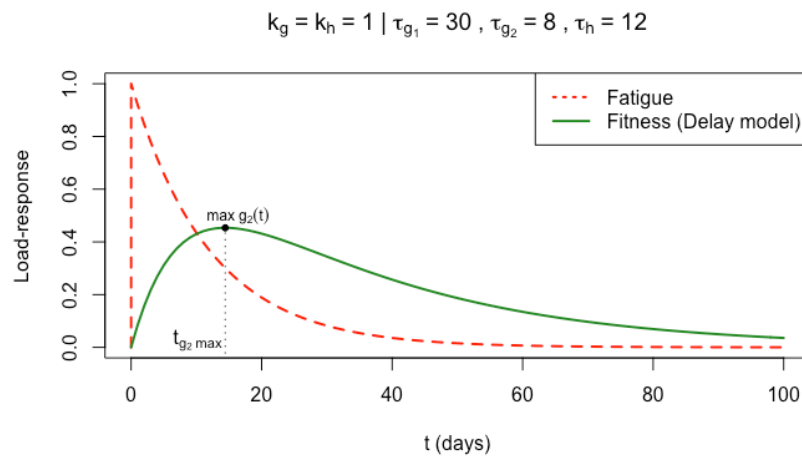


Figure 3: Load-response dynamics of the FFM with delay for a single arbitrary training dose

Several studies have attempted, with limited success, to correlate fitness and fatigue components described within FFM to observable physiological variables including heart rate⁷, maximal oxygen uptake (VO₂max)^{7,27,28}, respiratory gas parameters²⁸, iron markers^{29,30}, hormonal fluctuations^{8,11}, and enzyme markers³¹. In some cases, moderate to strong correlations have been observed (a summary table is provided in appendix D). However, there is noticeable inconsistency in these relationships. Whilst interesting, these results do not necessarily indicate that FFMs are ineffective. Fitness and fatigue are abstract concepts used to model the resultant effect of a range of physiological processes positively and negatively influencing the expression of physical performance. Therefore, perfect correlations with single observed physiological measures should not be expected. Additionally results may indicate the existence of factors not included in the model which is typical of the modelling process^{17,18,20}.

2.2 The general model

In view of a systems perspective and the potential of FFMs to model complex physiological phenomena via a limited number of resultant components, it follows that previous authors have attempted to investigate the optimum number of components^{7,13,32–38}. Toward this end, the standard FFM can be specified in general closed form such that the system is considered by N additive first-order components³³. In line with the process for solving the standard system (eq.4-6), the discrete solution of the general model with N components is obtained by convolving the variable representing the time-series of training loads $\omega(t)$ individually with N exponential functions of the form $e^{-\frac{t}{\tau_r}}$ ($r = 1, 2, \dots, N$), and then taking their sum. The resultant discrete function approximating the solution to the general model is then stated as:

$$\hat{p}(n) = p^* + \sum_{r=1}^N \left(k_r \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_r}} \right) \quad (\text{eq. 13})$$

Where the general scaling factor k_r is taken as positive or negative dependent upon whether the underlying system response represented by the component results in a performance increase or decrease. Several studies have applied the general model with experimental data to assess different levels of model complexity as a function of additional components, using laboratory study and/or retrospective analyses of field data^{7,30,32–34,36–39}. Several of these studies have found that additional first-order components did not significantly increase model fit in the majority of participants^{7,33–38}. However, a large proportion of these studies have been limited by sample size, the included population (i.e. limited study of elite performers) and inadequate number of data points. Each of these limitations would advantage models with a low number of components. In contrast, none of the above studies have assessed model prediction accuracy using out-of-sample data or in the context of external interventions purposively manipulating training loads to markedly change the distribution⁴⁰. Therefore, it remains difficult to identify the extent of overfitting and whether increasing model complexity provides a predictive advantage. A lack of out-of-sample assessment is not exclusive to studies assessing different levels of model complexity but is a consistent limitation throughout research investigating FFMs and will be discussed in section 3.4. Given the overarching philosophy of a systems model, it is recommended that future research attempts to assess several levels of model complexity and to evaluate predictions (particularly out of sample) across different scenarios. Nevertheless, some authors have recommended that a simple two-component model (i.e. the standard model, or the fitness-delay model) be applied in the meantime on the basis that it is likely to provide better conceptual and observed agreement with supercompensation phenomena that occurs during the recovery period in athletes that take part in regular training³³.

Researchers have also previously considered the case where an FFM is fit to an athlete experiencing residual fitness and fatigue effects that are expected to carry over into the next study period; i.e. $g(0) \neq 0, h(0) \neq 0$ ^{8,41}. To adjust

for this, the addition of P exponentially decaying initial components into the general model structure was suggested as a suitable approach by Busso and colleagues ⁸, as follows:

$$\hat{p}(n) = p^* + \sum_{r=1}^N \left(k_r \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_r}} \right) + \sum_{p=1}^P q_p^0 \cdot e^{-\frac{n}{\tau_p}} \quad (eq. 14)$$

Where the additional parameters (according to q_p^0) denote the initial level of each component, and τ_p is the decay time constant on the corresponding initial level. Conceptually, each additional parameter can be thought of as an additional ‘extra-large’ load-response at time $t = 0$, the sum of which will act to translate the function up or down on the vertical axis for a short initial period depending on the magnitudes with each associated component (i.e. fitness or fatigue). Note that not every component may require an initial decaying state ⁸. The use of a “grace period” (alternately “burn-in” period) where early observations are discarded is a technique used in Bayesian computation ⁴² and provided alongside Maximum Likelihood Estimation within Python’s popular statsmodels module ⁴³. Only a subset of these FFM variations have been implemented across the literature. In one experimental study, Busso and colleagues found that the number of statistically significant training components for each athlete ranged from 1-component ($n = 1$) to 2-component’s ($n = 5$) ⁸. The number of initial components (IC) in the final model equation were 0-IC ($n = 2$), 1-IC ($n = 2$), and 2-IC ($n = 3$). However, R^2 values in the study ranged substantially from 0.29 to 0.85 and statistical analyses were limited by low numbers of data points making it difficult to draw meaningful conclusions. The approach was further explored with regard to improvement in prediction accuracy by Ludwig and colleagues in 2016, and was referred to as an a posteriori ‘pre-load’ function ⁴¹.

A reference table for the various FFMs described so far is presented in appendix E to assist readers. An increase in model complexity (i.e., the addition of further filters increasing the number of free parameters) may be used to capture more known phenomena of the training response. However, it is worth noting that models with large numbers of free parameters are at a higher risk of overfitting and should be evaluated by a suitable method to assess out-of-sample prediction accuracy. Additionally, with large numbers of free parameters the requirements on the training data (i.e. the number of performance measures) may quickly become impractical. These issues are considered in more depth in sections 3.2 and 3.3.

2.3 The variable baseline performance model

FFMs generally include an intercept parameter p^* (baseline performance) which is conceptualised as the performance value obtained after a substantive period of recovery such that the residual effects of all previous training sessions have dissipated. However, baseline performance in many cases will change between training periods and the effect of this will depend on model use. For example, where an FFM is being used to predict short term change for tapering and peaking, change in baseline performance may not occur and any prior change can be accounted for within the model fitting process. However, where an FFM is being used to guide multiple training blocks over extended periods, baseline performance may require re-estimating, unless expected movement in the parameter itself can be predicted. Gouba and colleagues suggested an approach to address this limitation by introducing a recursive term p_{n_j} in place of p^* (eq.15) ¹². The basis of the authors extension is that performance on day n is influenced by both the previous level of measured performance, and the number of training doses between them ¹². Their model can be stated as follows:

$$\hat{p}_{n_j} = p_{n_{j-1}} + k_g \cdot \sum_{i=n_{j-1}}^{n_j-1} e^{-\frac{(n_j-i)}{\tau_g}} \cdot \omega(i) - k_h \cdot \sum_{i=n_{j-1}}^{n_j-1} e^{-\frac{(n_j-i)}{\tau_h}} \cdot \omega(i) \quad (eq. 15)$$

Where j is a measured performance, n_j is the day of the j -th measured performance, and p_{n_j} is the observed value. Under experimental conditions, the authors observed poor model fit for both the standard FFM ($R^2 = 0.28$) and their novel extension ($R^2 = 0.38$) when modelling the data of a high-level monofin swimmer. However, little can be drawn from these results due to confounding effects of low measurement frequency ($n = 6$ over a 24-week period), and a lack of cross-validation to assess any practical differences. Therefore, more research applying this model extension is required to assess whether there exists any substantive improvement in accuracy or conceptual validity.

2.4 Further developments to model structure

Three key limitations have been identified with the basic FFMs presented in previous sections including: 1) irrespective of training dose presented in any single session, performance is maximised a fixed number (η) of days later; 2) performance can be arbitrarily increased by simply increasing the training dose¹⁸; 3) there is no interaction between training sessions, and therefore training performed on a given day does not influence the response generated from another session. Collectively, these limitations indicate that the best training plan would be to complete all the planned training in one session, η days prior to competition. The general approaches proposed to address these limitations include either constraining and manipulating the model input (i.e. training dose)⁴⁴ or manipulating the model formulation^{4,5,7,9}. For example, the use of a threshold saturation function to restrict the effective load without changing the relationship between the input and output specified has been proposed⁴⁴, along with the introduction of specific structural modifications to account for non-linearities⁶ and interactions between training sessions^{4,7,9}. These approaches and other advanced topics including the use of the Kalman filter are presented in part II²⁴ of this review series. Further proposed extensions within the literature include exponential growth kinetics⁵ and secondary-signal models¹³.

3.0 Application of fitness-fatigue models in research and practice

FFMs depend on constant (or in special cases time-varying^{9,45}) parameter values that cannot be inferred through observation and must instead be estimated from training and performance data, for example, by the method of nonlinear least squares^{18,19,46,47}. FFMs are nonlinear in their parameters and therefore the model fitting process constitutes a nonlinear optimisation problem. The process takes as input a time-series of measured performance and training load values, and provides as output, model parameter estimates that give good (preferably the best possible in some sense) agreement between iteratively computed model performance values and the measured performance data. In summary, to fit an FFM a researcher or practitioner requires a series of suitable training quantification data, performance measurements, and a method to alter the parameters to best match these through an optimisation process. Addressing these three requirements is discussed in their respective order in the following sections.

3.1 Training load quantification

Existing FFMs require a time-step input in the form of a discrete time-series of training load values $\omega(n) \geq 0$, where a single number represents the training load for the time-step chosen. Typically, the time-step will be set to one and each numeric value intended to represent the resultant effect of the daily training session(s) on the physical response being modelled. Reducing athletic training sessions to numeric values is a highly complex and unresolved problem, particularly with respect to the broad spectrum of training modalities and performance measures used. Within laboratory settings, training load quantification might be easier due to tight controls and the use of single training modalities and performance measures that are expected to map directly to the intervention strategy⁴⁸. However, this is rarely practical in real world settings where training comprises different physical capabilities trained concurrently. Proposed solutions to address the problem of training load quantification can be described as predominantly exploratory, typically resulting from attempts to identify training variables that are believed to account for the largest variation of training response in a population. The result of these attempts are usually an amalgamation of

assumptions, coaching experience, and empirical knowledge to construct and implement one or more methods that seem reasonable but are not extensively validated. The full scope of the training load quantification problem is extensive and requires substantial development as both a standalone area and within the context of fitness-fatigue modelling. Interested readers are directed toward previous reviews of this topic ^{49–52}. In the following sections an overview of the area in the context of fitness-fatigue modelling is provided.

A general form for quantifying training load (ω) using a volume-by-intensity approach may be written mathematically as:

$$\omega = \mathcal{D} \cdot u(I) \quad (\text{eq. 16})$$

Where \mathcal{D} represents the density of the exercise (e.g. volume, duration, distance), I is the intensity (e.g. percent-of-maximum, perceived intensity) and u is a function (potentially non-linear) that can create relatively large changes in training dose from relatively small changes in intensity (illustrated in Figure 4). The interaction between density and intensity to provide appropriate training load values may vary between individuals and the type of training relative to the underlying performance measure. Different performance measures are likely to require different weighting functions based on the principle of specificity ⁵³. For example, an endurance athlete may perform both high-volume/low-intensity and low-volume/high-intensity training. At a basic level, these two types of training session will be structured to take advantage of improvements in different underlying energy systems and physiological processes. The effect of these adaptations on performance will therefore have different magnitudes depending on the properties of the performance measure selected (e.g. VO_2max , 5k TT or an aerobic threshold test). Athlete level and prior training history are also likely to mediate this difference in magnitude. Therefore, it is reasonable to suggest that further work is required to develop training load quantification methods that are sensitive to the differences described. Previous research has attempted to address this issues with other models such as the PerPot DoMo model which isolates intensity and volume into two concurrent load flows ⁵⁴. Similarly, within fitness-fatigue modelling, different inputs for each component (i.e. ω_{fitness} , ω_{fatigue}) may be required to account for the effects of certain exercises on systematic fatigue but not fitness in relation to the criterion performance measure. For example, in resistance training an exercise such as the squat may generate a fatigue response but not evoke a simultaneous fitness response when modelling upper-body physical performance.

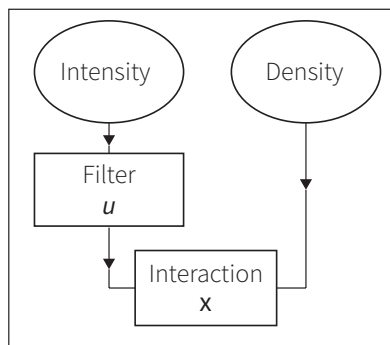


Figure 4: A common approach to quantifying training load for modelling physical performance change in sport and exercise applications.

Variations of the density-intensity approach have featured in the majority of previous FFM applications, in swimming ^{2,3,12,18,44,55–60}, cycling ^{7,33,45,61}, weightlifting ^{8,11}, judo ³², gymnastics ⁶², hammer throwing ⁶³, and running ^{28,64,65}. In 1985, Banister and colleagues introduced Banister’s TRIMPs, which included a ratio to quantify intensity via elevation in heart-rate relative to basal and maximum values ²⁹:

$$\mathcal{D} \times u \left(\frac{[HR_{\text{exercise}} - HR_{\text{basal}}]}{[HR_{\text{maximum}} - HR_{\text{basal}}]} \right) \quad (\text{eq. 17})$$

Where \mathcal{D} is the duration (in minutes), u is a non-linear function of the intensity $0 \leq I \leq 1$ which is the fractional elevation of the maximal HR, and $u(I) = 0.86 \cdot e^{1.67 \cdot I}$ (coefficients derived for female athletes in the original study). Therefore, u weights brief efforts at higher heart rates positively compared to sustained steady state efforts at lower heart rates. Banister's TRIMPS has been used in several applications of fitness-fatigue modelling in running^{10,29,31,65}, skiing³⁰, swimming¹², triathlon^{36,50,66}, and cardiac rehabilitation^{34,35,39}. Although Banister's TRIMPS and similar derived metrics for training load quantification (e.g. session RPE, summated heart rate zone score, Lucia's TRIMPS)^{49,67} comprise the majority of prior research applications, other less known methods may provide further avenues for future study. These include adapting models to incorporate multiple inputs as described earlier (e.g. an input specific to each model component, which may also reduce observed inter dependency between parameters k_g, k_h), and feature selection within a statistical-learning framework as investigated by Rozendaal²⁶. While there is no universal method to quantify training load series and work remains to be completed in this area, reasonable attempts to quantify training load can and should still be made. From the perspective of fitness-fatigue modelling, care should be taken to ensure that the process of quantifying a full training session to a single value incorporates elements of the training that are most likely to map to the performance outcome measure. Furthermore, researchers and practitioners should seek to record data at the highest possible frequency within-sessions and minimise instrumentation error⁶⁸. Subjective variables such as repetitions-in-reserve and RPE may be used to provide high-frequency mapping of intensity across a training-session. Similarly, objective measures such as heart rate^{59,69} and barbell velocity can be sampled with little obstruction at high frequencies and may provide insight into session intensity over time, and in particular the relative difficulty of the task compared to previous efforts.

3.2 Criterion performance selection

The second major consideration for researchers and practitioners wishing to apply FFM with their own data is the selection of an appropriate outcome measure (i.e. model target). This outcome measure is commonly referred to within the literature as the *criterion performance* measure. The use of the term *performance* in sport is broad and for the purposes of fitness-fatigue modelling benefits from being conceptualised as either physical or competitive performance. Physical performance is defined here as the state of any dependent variable that describes physical capability in a sporting or biomechanical task, and typically reflects dimensions of fitness (e.g., strength, stamina, speed). In contrast, competitive performance is composed of the interaction of multiple dynamic, stochastic, and chaotic processes (e.g., physical performance, psychological state, team and opposition behaviour, environment). These processes eventually converge to a conclusive outcome variable (win, loss, draw), which we separately term the *competitive outcome*. Viewed in isolation, competitive outcomes may fail to provide useful information, and any discernible patterns over time are very unlikely to be explained by concepts of fitness and fatigue alone. Therefore, future use of performance models such as FFM, particularly within team sport environments, may require a shift in emphasis where training data is used to predict response in terms of dimensions of physical capability measured through common exercises (e.g. one-repetition maximum squat, vertical jump height, or peak power produced during an explosive movement) rather than sporting outcomes which are likely to demonstrate complex and often confounded relationships with training loads. Exceptions to this may exist in endurance sports or power/strength-based sports such as weightlifting where outcomes in competitive events relate closely to physical capability (e.g., 100m swim sprint, 5k time-trial run or cycle, maximum weight lifted by a powerlifter in the squat). However, this data is only likely to supplement a modelling approach to increase the number of data points, rather than be the focus of it, due to introduction of other factors such as those highlighted above. Previous examples of criterion performance measures used in FFM applications include performance in mock or actual sporting events particularly within endurance sports^{18,59,66}, maximum strength or peak power output produced during specific exercises^{7,63}, and outcomes obtained from clinical exercise capacity tests^{34,35,39}.

Measurement frequency and underlying measurement error should also be considered when selecting an appropriate performance measure and how these may influence the ability to obtain stable and accurate parameter estimates^{22,70}. It has been highlighted that missing data may lead to non-interpretable model behaviour in iteratively updated (i.e.

continuous) modelling applications²². For multiple linear regression, a minimum of 15 observations per-parameter is recommended⁷¹. However, the FFM comprises a non-linear function (in its parameters) with subsequent statistical analyses based on asymptotic theory⁷²⁻⁷⁴. Therefore, more data points are required per model parameter¹⁸ with previous recommendations of 60-200 performance tests over a period of model fitting. As a result, the ability to accurately model and predict performance is expected to require performance tests that can be completed at a high weekly frequency (e.g. at least every 1-3 days)⁷⁰. Consequently, it is likely that only non-fatiguing performance tests (for example, the counter-movement jump) or those without persistent learning effects may be able to meet this requirement. An alternative strategy that has clear potential but has received limited investigation in the context of fitness-fatigue modelling is the extrapolation of performance measures from standard training session data⁷⁵. Such procedures could incorporate either objective (e.g., barbell velocity or measurements of power) or subjective variables (e.g. repetitions in reserve) to predict maximum performance with high frequencies. Combined with the previous recommendation of a shift to emphasise physical performance in simple biomechanical tasks or exercise movements, effective use of training monitoring data within fitness-fatigue modelling may generate increased collaboration between researchers and practitioners. The use of extrapolation methods to generate high frequency measures may reveal a novel approach to studying FFMs in certain sports, such as powerlifting, where performance can be considered simultaneously a largely isolated dimension of physical capability and competitive outcome.

When using FFMs in practice, it is important to acknowledge that all measurement of physiological systems comprises error (instrumentation error and biological variability) which often makes it difficult to accurately quantify underlying physical capability^{68,76}. Thus, true performance can be defined as a random variable where measured (observed) performance is only ever an estimate^{76,77}. FFMs acknowledge measurement error through introduction of a model error term, often assumed to be independent and identically distributed as Gaussian. Even where these assumptions are reasonable, higher error variances generally lead to lower precision in estimation. Several basic steps can be taken to reduce measurement error and its influence on any model's accuracy. These steps include selection of tests that comprise low measurement error (preferably less than 3-5% of coefficient of variation), standardisation of testing procedures and where possible taking the average of repeated independent tests^{68,70}.

3.3 Parameter estimation approaches and limitations

Fitting an FFM (i.e., estimating the unknown parameters in the model) constitutes a nonlinear optimisation problem. The objective is to determine values for the parameters that maximise a goodness-of-fit function known as a loss, cost, or objective function⁴⁷. Estimation of model parameters in FFMs has been historically approached from a least-squares^{18,20,21} or maximum likelihood perspective^{13,21,76,77}. The nonlinear least-squares approach involves minimising the sum of squared deviations between modelled and measured performance via the following cost function:

$$\min \sum_{i=1}^m (\hat{p}_i - p_i)^2 \quad (eq. 18)$$

Where i is an index over a set of m of data points $\{(p_1, \hat{p}_1), (p_2, \hat{p}_2), \dots, (p_m, \hat{p}_m)\}$ that represent measured (p) and modelled (\hat{p}) criterion performance values at specific integer time points. \hat{p}_i is the fitness-fatigue model function $f(t_i, \theta, x(t_i))$ that depends on the step input up to time t_i (i.e. training load series) $\{x(1), x(2), \dots, x(t_i)\}$ and n model parameters θ (i.e. $\theta = \{p^*, k_g, \tau_g, k_h, \tau_h\}$ for the standard model). An optimisation algorithm wraps around the objective function such as eq.18 and uses specific update and stopping criteria to methodically or stochastically travel over the available parameter space to search for the best possible set (i.e., the absolute or global minimum). First and second order algorithms (including those that approximate the Hessian by the outer product of the gradient) have predominantly been used to direct the iterative search to minimise least-squares or maximise likelihood.

Finding suitable parameter values in this manner has been described as “notoriously difficult” in the study of similar modelling problems ⁴⁶. Fitting an FFM assumes that an optimum solution exists, is unique, and can be easily found. The FFM even in its most basic form is a model in five dimensions, and therefore cannot be visualised using standard plotting techniques. Furthermore, if there are truly different parameter sets that have the same global minimum under standard nonlinear least squares, there exists a situation where parameters aren’t uniquely identified without additional constraints or regularisation terms. However, it is more likely that local minima and plateau features present the primary challenge in a typical FFM fitting process. First and second order algorithms can sometimes mistakenly converge to local optima, or become stuck on plateau features, provoking sensitivities in the fitting process and to the starting point or region on the cost surface (i.e., jumping-off point for the algorithm). Identification of suitable model parameters from training and performance data is therefore an outstanding concern in fitness-fatigue modelling ^{18,20}, and is an area that has received minimal attention.

In most complex models of real-world systems, parameters have compensatory (dependent) effects relative to the systems collective behaviour ^{18,46}. For example, with the FFM it is possible to modify a single decay constant without changing the resultant behaviour of the system, provided values in other parameters adjust to compensate ⁴⁶. Additionally, FFMs have too been shown to be ill-conditioned, such that a small change in parameter values can reflect large changes in model behaviour ^{18,20,70}. This ill conditioning, also termed ‘sloppiness’ ⁴⁶ in systems modelling has been said to affect the accuracy and precision of parameter estimates ⁷². It has been suggested that the fundamental value of FFMs is that they seek to link underlying physiological concepts to observed data, and therefore interpretability of parameter estimates in the real-world is viewed as a valuable property. However, model sloppiness has not precluded good predictive power in application of biological, physical and chemical systems models ⁴⁶. Prediction accuracy (particularly out-of-sample) is an area where FFMs have never been rigorously assessed, particularly under well-designed experimental conditions or using large retrospective datasets from practice. In studies where poor model fit has been reported, there have typically been issues such as low sample size, selection of inappropriate performance measures, or limited information has been reported on the fitting process. It is hypothesised that accurate prediction of future response may still be possible under an FFM, even in the absence of stable, characteristic model parameters (i.e., those which can be carried over between training programs). Historically, researchers have fitted FFMs under one set of ‘best guess’ initial candidates for parameters under first and second order algorithms ¹⁹. Given the concerns outlined above, research reporting a single model solution derived from a ‘one shot’ run of gradient-based optimisation creates uncertainty surrounding parameter estimates. As a basic starting point for future use, practitioners and researchers fitting FFMs via least-squares using traditional derivative-based algorithms should re-fit the model under multiple starting points. In-sample model stability can then be assessed through repeated refitting to subsets of the data with data points removed ¹⁸, and out-of-sample prediction accuracy should be tested where possible under different intervention distributions (i.e. future training load programs, see section 3.4).

Researchers and practitioners interested in fitting FFMs may also investigate whether derivative-free approaches achieve better results. Such approaches often perform better with non-convex functions comprising multiple local optima. While linear least squares is guaranteed to be convex, there are no such guarantees for non-linear least squares problems. The class of evolutionary algorithms offers an alternative approach and may also be effective in cases where constraints or penalty functions are applied using regularisation terms or custom cost functions. However, evolutionary algorithms including differential evolution, genetic algorithms, and particle swarm can require significant tuning of control parameters to obtain appropriate results. These algorithms are also slow compared to gradient-based methods that can exploit a smooth surface. However, hybrid approaches may go some way toward offering a resolution to both the problems of finding suitable initial candidate estimates under first/second order methods to reduce the chances of becoming stuck at local optima, and the lack of guarantee on optimal convergence using evolutionary algorithms. A hybrid algorithm typically comprises a stochastic tunnelling method to identify convex

basins of attraction surrounding optimum parameter estimates, and then subsequently passes these as initial values for gradient-based local search. More advanced fitting approaches have rarely been applied to FFM⁶, however, previous research has employed evolutionary algorithms to solve training program design problems using pre-defined FFM parameter estimates^{14–16}.

Lastly, the problem of over-fitting can occur when the cost-function is fit too closely to a limited set of data points reflecting noise in the data⁷⁸. If a model comprises more free parameters than required to explain most of the variance in observed performance, this may permit explanation of idiosyncrasies within the data. In summary, resolving these issues often requires a larger number of data points, lower relative error within the measured data, re-formulation of the optimisation problem to reduce fitting sensitivity to noise in the data^{13,76,77}, or in some cases re-specification of the model to reduce complexity^{18,20}. Researchers and practitioners in future collaborations may consider setting relatively tight a priori constraints on likely parameter values (or use of informative priors using a Bayesian approach). Alternatively, ensemble or hierarchical modelling approaches may be considered, aggregating parameters over multiple training and performance sets, or across groups of similar athletes. The introduction of regularisation terms and associated penalty functions that enforce constraints on the cost function may also assist in obtaining parameter values that provide stable predictions or limit overfitting.

3.4 Model evaluation: Integrating practitioner data into future research efforts

The fundamental stages of mathematical modelling comprise: 1) problem formulation; 2) determination and examination of solutions; and 3) model validation across diverse cases applicable in the real world⁷⁹. Of the three phases, FFM research is most limited in model validation. As an appropriate starting point, simulation-based approaches can be used to provide an efficient and inexpensive method to study potential limitations of a model and early indications of predictive performance both in and out of sample^{48,80,81}. For example, fitting-based simulation approaches under realistic mock data can be used to identify lower bounds on properties such as prediction error via systematic and iterative deviation of controlled ‘small world’ conditions (e.g., various levels of measurement error and different testing frequencies). In these instances, the model can only be assumed to perform worse in the real world⁷⁰. Simulation can also be used to explore elements such as fitting sensitivity and explore the optimisation approach. If lower bounds on model properties identified via simulation appear unreasonable or indicate poor performance in real world applications, then further investment in studying the model under empirical/experimental conditions should be carefully considered or avoided.

A distinct lack of out-of-sample testing represents the biggest limitation of FFM research to date. FFM^s are flexible and authors who simply fit an FFM to all available data without assessing prediction accuracy via a cross-validation approach limit the potential to assess both the extent of overfitting in the model, and accuracy of predictions under future training programs. Cross-validation (in its purest form) is a method of estimating out-of-sample loss (i.e. model accuracy). In the context of fitness-fatigue modelling this involves partitioning data into two complementary subsets: 1) a training set; and 2) a testing or validation set. The data within the training set is used to fit the FFM (and is thus classified as “seen” by the model), and the testing set comprises data used to assess prediction accuracy either within or out-with the modelled period (and is therefore classed as “not seen” by the model). It is the comparison between the testing set data and fitted model predictions that provides a measure of model loss. Under most cross-validation techniques, this process is repeated several times using different choices of partitioning to reduce variability, and the results are averaged over the rounds to provide a summary estimate of out-of-sample accuracy and therefore provides some insight into the model’s potential predictive capability. There are several possible techniques that can be used to achieve out-of-sample loss estimation, the simplest of which is the *hold-out* method involving a single run of model fitting and testing of two (often evenly split) consecutive sets (figure 5-A). As the FFM is a time-series model, cross-validation techniques that mix previous and future data require special care as they do not respect the order between time-series data nor probable autocorrelation and thus might introduce look-ahead bias. Therefore, methods such as

random hold-out, and *k-fold* are not recommended, but *Combinatorial Purged Cross Validation*⁸² offers a more principled approach to mixing past and future. In general, walk-forward approaches with multiple splits provide a more robust assessment of out-of-sample model loss. A potential walk-forward approach suitable for FFM applications is the *expanding window* method where the proportion of data used to train the model progressively increases (Figure 5-B). As the primary purpose of cross-validation is out-of-sample loss estimation (and not parameter estimation), it is reasonable to return to the full dataset to obtain final point estimates of the FFM parameters. While it is tempting to tune parameters based on the cross-validation estimated loss, tuning using a validation set will result in overoptimistic loss estimates. In contrast, averaging the estimates from the expanding windows may reduce estimator variability in the same way as “bagging” in Machine Learning⁸³. How to best use estimates from cross-validation windows is another topic and avenue for future research.

With five parameters in the standard model (eq.7) and upwards of eight in more complex formulations, FFMs would be expected to fit most patterns of data, even those generated from a process very different to that underlying the FFM. If an athlete’s training loads are consistent and average performance is trending upwards linearly, for example, then both the FFM and a straight line are sufficiently expressive. Provided this pattern continues from training into the hold-out periods, the predictions, even those out of sample, say little about the FFM’s ability to meaningfully predict outcomes. In contrast, if the athlete is presented with a new training plan with substantial variation in the training loads, a more effective framework is available to determine if the model can make nontrivial, testable predictions. In statistical Design of Experiments (DoE), variation in the explanatory factors is purposefully created to establish uncertainty in parameter estimation. Optimal designs for evaluating FFMs including configurations of the training impulses best suited for learning represent another potential direction for future research. In the meantime, researchers can draw from essential DoE design concepts. Since the external “factors” in FFM are convolutions of training inputs, training loads are required that vary for enough time to modulate these convolutional quantities. For fatigue, this may mean prescribed rest periods of days to weeks. For fitness, creating variation is easiest when the individual has low prior levels, as it is a feature of the model that fitness is relatively stable. For maximal learning, longer rest periods or prolonged lighter training after the athlete has attained high levels of fitness may best facilitate this variation, potentially creating issues for participation and adherence to studies.

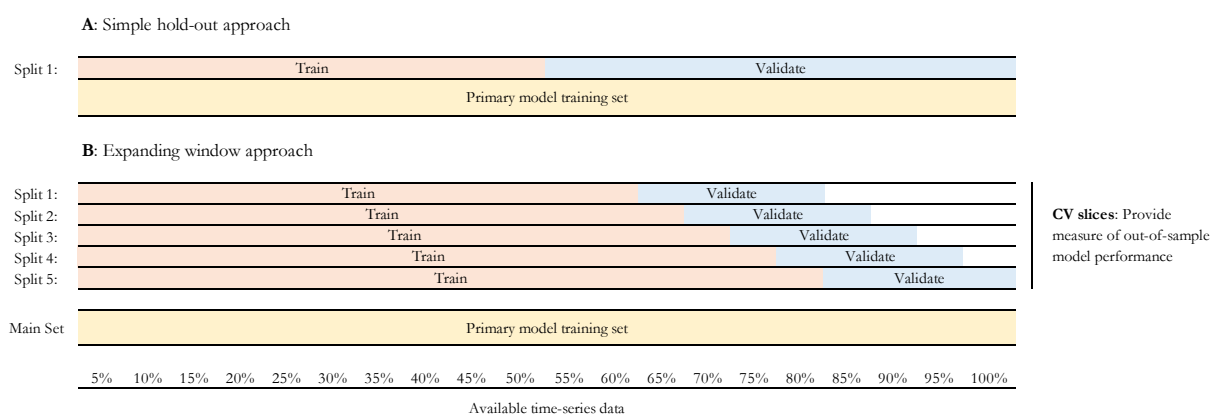


Figure 5: Example of non-exhaustive cross-validation methods: simple hold-out **(A)** and expanding-window **(B)**. These approaches may be used to assess model performance in existing FFMs

3.5 The utility of an FFM: Informing training program design

Currently, individual exercise prescription in research and practice relies heavily on qualitative reasoning, coach and athlete experience, and what is known about underpinning physiological mechanisms from the scientific literature⁷⁷. Conversely, individualisation of physical training via quantitative approaches is an area where many open problems

remain ⁷⁷. It is recognised that Banister and colleagues were pioneers in advocating for sizeable tapering periods prior to competition to achieve optimal performance ⁸³. For coaches and practitioners, the potential value in applying the quantitative representation of the FFM is to better inform training design to maximise performance at specific times ⁸⁴. Although the ultimate utility of existing FFMs (i.e. out-of-sample predictions) has not been rigorously tested, prospective methods for FFMs to generate future training programs under constraints are available and have been previously assessed ^{6,14–16}. If FFMs or alternative models are determined to be sufficiently accurate, they may be applied within constrained optimisation frameworks to assist in identifying particular programs that are likely to provide highest rates of adaptation or best performance over a planned training period ^{14,48}. To summarise briefly some of the previous work conducted so far, Schaefer and colleagues ¹⁴ treated plan generation as a constraint-satisfaction problem (CSP) and assessed several optimisation frameworks for solving them to derive training plans. Model parameter estimates were pre-defined from a previous study, and their work solely focused on assessing whether reasonable training plans could be generated using non-linear solvers ¹⁴. Similarly, Connor et al. used mathematical programming to solve the training plan generation problem under an FFM framework using grammatical evolution, obtaining reasonable training programs for an elite team of Gaelic football players ¹⁵. Kumyaito et al. ¹⁶ attempted to create practical cycling plans that satisfied physiological constraints (with regard to monotony, chronic training load ramp rate, and daily training impulse), but which were reasonable for improving athletic performance. The authors applied adaptive particle swarm optimisation using ϵ constraint methods to formulate the plans and simulate likely performance outcomes ¹⁶. Collectively, these studies represent viable approaches to deriving future training programs with constraints, following identification of useful performance models (perhaps FFMs) that attempt to describe the relationship between training and performance.

4.0 Conclusions and practical recommendations

Modelling complex human processes is scarcely straightforward. The scientific field of systems biology is at the forefront of modern inquiry and leading the way in several aspects of systems modelling. Even so, systems biologists are frequently confronted by and still make use of heavily parameterised, incomplete, or partially inaccurate mathematical models of complex processes ⁴⁶. Moreover, with any mathematical model of the real world there almost always exists an accuracy vs. complexity trade off. With the use of modelling to describe complex real-world systems now viewed as an important component of modern scientific investigation, it appears hasty to dismiss one of the only existing systems models in sport science without rigorous study, at the very least to inform and guide future modelling attempts ⁴⁶. Sport science has very few models to describe the interaction of complex human processes underpinning training and performance ⁸⁵, and this art exists at the fringe of our existing literature body and outside the purview of most practitioners. Part I of this two-part review series has attempted to present a methodical history and refreshed perspective of the study of these models and bring this area of systems research back into focus. While FFMs are not exempt from the common challenges associated with complex systems modelling (e.g., mis-specification and ill-conditioning), their potential utility and notional value reside in being theory-driven mathematical models of athletic performance from cumulative training loads, with counterfactual properties and the ability to generate testable predictions ⁴⁰. Furthermore, there have been several valuable and informative attempts within the literature to address existing limitations and common criticisms echoed in the literature body. It is noticeable that these model extensions have been largely overlooked with respect to further experimental investigation most likely due to historical criticisms of the standard model, and a lack of awareness amongst sport scientists of several recent and innovative advancements in this area in the last 5 to 10 years. Therefore, these model extensions must be studied and tested on their own merit and should not be grouped carelessly with prior criticisms of the standard model structure. Examples of these models are presented in part II ²⁴ of this review series, along with resources to fit and assess models with users own data.

To explain the lack of out-of-sample assessment of model predictions across fitted models, one possibility is a prior insufficiency in the availability and access to high-quality and high-frequency long term training and performance data. In the past, some authors have even turned to conducting interventions on themselves as sole participants to generate

sufficient data to conduct FFM research ³¹. It has only been in the last decade or so that data collection on training input and performance has substantially increased, making these types of investigations far more feasible across many sporting domains. Notably, some recent studies have begun to take advantage of the opportunity to utilise convenient field data with FFMs ^{9,26,58,68,76,87}. However, the majority of historical experimental research has focussed on quantifying relationships between underlying physiological processes of training, through iterative refinement of model structure and comparison with observed physiological phenomena ⁴⁸. While there may still be value in this approach, there is an overwhelming volume of this type of work compared to the study of FFM predictions, and attempts should now be made to restore a balance between these two modelling objectives. To improve the practical application of future research, more effective collaboration between researchers and practitioners is needed and development required in each of the three practical areas highlighted in this literature review (training load quantification, criterion performance selection, parameter estimation). With regards to training load quantification, greater discussion of how to selectively weight different training modalities, exercises and intensities on the basis of the target performance is needed. Additionally, researchers may consider including training load quantification parameters as part of the overall fitting process, potentially with a priori constraints suggested by practitioners. With regards to criterion performance measures, the biggest challenge with fitness-fatigue modelling appears to be the frequency of measurement. It is hypothesised that a method that seems most suited to generating useful high frequency measurements, particularly with regards to fitness aspects, is individually predicted maximum performances derived from actual training data. Given the large increase in applied research in this area (e.g. reps-in-reserve, barbell-velocity profiling) there are a range of candidate measurements that could be used and would benefit from the collaboration between researchers and practitioners. Finally, with regards to parameter estimation there is a clear requirement for extensive further development in methods and understanding. Of all areas of the modelling process, parameter estimation and reporting are the least well described aspects. In addition to the systematic study of which algorithms perform the best for obtaining suitable parameters under a range of conditions, there is a need to explore in greater depth issues such as parameter stability and estimation sensitivity, and methods to obtain parameters that are best able to predict future training such as the potential of tuning-based cross-validation and the introduction of regularisation terms. Future collaborations between researchers and practitioners may wish to consider setting a priori constraints on likely parameter values given the specific athlete (alternatively using informative priors using a Bayesian approach) or matching athletes and fitting multiple models simultaneously sharing information to obtain parameter values that are representative of athletes with similar adaptive responses. As with other practical areas of fitting FFMs, there is considerable scope for ingenuity and interdisciplinary work that will eventually require and could utilise large amounts of data routinely collected by practitioners working with athletes.

There is also scope for researchers and practitioners to more creatively apply the standard FFM. In the context of resistance training and the development of fitness aspects, the FFM has been used in its conceptual form to highlight that different forms of training (e.g. strength, power, and hypertrophy) appear to generate distinct fitness and fatigue profiles ¹. The general FFM with N additive first-order components and other properties such as fitness-delays and threshold saturation, may provide a suitable mathematical framework to model these empirical observations with for example strength, power and hypertrophy training sessions inputting their effects through specific first-order components with scaling coefficients and decay time constants dependent on the fitness attribute being modelled. With all such innovative procedures, it is critical that emphasis is placed on determining the conditions under which suitable predictions to future training can be obtained.

Finally, it is clear from this review that historical and future research would benefit from increased transparency and reproducibility around data, assumptions, and methods/implementation. It is therefore recommended that where possible future experimental study follow basic tenants of reproducible research when disseminating work in peer-reviewed journals and preprint ²⁵. These include: 1) extensive documentation of code or software used to implement the models and perform the analysis; 2) presentation of computational output with explanatory text; 3) inclusion of

raw data alongside code and analysis to facilitate verification and reproduction of findings and methods. Where public release of raw or anonymised data is unrealistic due to reasonable concerns for privacy or commercial sensitivity, authors should attempt to make this available to reviewers within the peer-review process where possible. Code notebooks (e.g. markdown) hosted on public or private repositories provide convenient and simple methods for achieving the recommendations outlined above.

Concluding statements: While acknowledging several limitations in the standard model structure and criticisms across prior research, it is our position that sport science should not yet give up on FFMs, particularly given several promising approaches by researchers to address prior limitations in recent times. To achieve progress, this small field of sport science must pool ideas, concepts, and develop communal tools. Establishment of a coherent framework for future study would be a further helpful step, particularly one that prioritises tackling specific factors currently limiting model utility, and that outlines steps for rigorous assessment of new or existing models and methods. It is hoped that this review provides a detailed insight into the historical development of and practical considerations for fitness-fatigue modelling, and that the accompanying code tools and recommendations outlined offer a foundational basis for the development of a research framework for future FFM study.

References

1. Chiu, L. Z. F. & Barnes, J. L. The Fitness-Fatigue Model Revisited: Implications for Planning Short- and Long-Term Training. *Strength Cond. J.* **25**, 42–51 (2003).
2. Banister, E. W., Calvert, T. W., Savage, M. V & Bach, T. M. A Systems Model of Training for Athletic Performance. *Aust. J. Sport. Med.* **7**, 57–61 (1975).
3. Calvert, T. W., Banister, E. W., Savage, M. V & Bach, T. M. A Systems Model of the Effects of Training on Physical Performance. *IEEE Trans. Syst. Man. Cybern.* **6**, 94–102 (1976).
4. 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).
5. Philippe, A. G., Borrani, F., Sanchez, A. M. J., Py, G. & Candau, R. Modelling performance and skeletal muscle adaptations with exponential growth functions during resistance training. *J. Sports Sci.* **37**, 254–261 (2018).
6. Turner, J. D., Mazzoleni, M. J., Little, J. A., Sequeira, D. & Mann, B. P. A nonlinear model for the characterization and optimization of athletic training and performance. *Biomed. Hum. Kinet.* **9**, 82–93 (2017).
7. Busso, T. Variable dose-response relationship between exercise training and performance. *Med. Sci. Sports Exerc.* **35**, 1188–1195 (2003).
8. Busso, T. *et al.* Hormonal adaptations and modelled responses in elite weightlifters during 6 weeks of training. *Eur. J. Appl. Physiol. Occup. Physiol.* **64**, 381–386 (1992).
9. Kolossa, D. *et al.* Performance estimation using the fitness-fatigue model with Kalman filter feedback. *Int. J. Comput. Sci. Sport* **16**, 117–129 (2017).
10. Morton, R. H., Fitz-clarke, J. R. & Banister, E. W. Modeling Human Performance in Running. *Am. Physiol. Soc. Model. Methodol. forum* **69**, 1171–1177 (1990).
11. Busso, T. *et al.* A systems model of training responses and its relationship to hormonal responses in elite weight-lifters. *Eur. J. Appl. Physiol. Occup. Physiol.* **61**, 48–54 (1990).
12. Gouba, E., Konfe, B. O., Nakoulima, O., Some, B. & Hue, O. Applying a Mathematical Model to the Performance of a Female Monofin Swimmer. *Appl. Math.* **04**, 1673–1681 (2013).
13. Busso, T. From an indirect response pharmacodynamic model towards a secondary signal model of dose-response relationship between exercise training and physical performance. *Sci. Rep.* **7**, 1–11 (2017).
14. Schaefer, D., Asteroth, A. & Ludwig, M. Training plan evolution based on training models. *IEEE Symp.* (2015) doi:10.1109/INISTA.2015.7276739.
15. Connor, M., Fagan, D. & O'Neill, M. Optimising Team Sport Training Plans With Grammatical Evolution. 2474–2481 (2019) doi:10.1109/cec.2019.8790369.
16. Kumyaito, N., Yupapin, P. & Tamee, K. Planning a sports training program using Adaptive Particle Swarm Optimization with emphasis on physiological constraints. *BMC Res. Notes* **11**, 1–6 (2018).
17. Taha, T. & Thomas, S. G. Systems Modelling of the Relationship between Training and Performance. *Sport. Med.* **33**, 1061–1073 (2003).
18. Hellard, P. *et al.* Assessing the limitations of the Banister model in monitoring training. *J. Sports Sci.* **24**, 509–520 (2006).
19. Clarke, D. C. & Skiba, P. F. Rationale and resources for teaching the mathematical modeling of athletic training and performance. *Am. J. Physiol. - Adv. Physiol. Educ.* **37**, 134–152 (2013).
20. Pfeiffer, M. Modeling the Relationship between Training and Performance - A Comparison of Two Antagonistic Concepts. *Int. J. Comput. Sci. Sport* **7**, 13–32 (2008).
21. Proshin, A. P. & Solodyannikov, Y. V. Physiological Avatar Technology with Optimal Planning of the Training Process in Cyclic Sports. *Autom. Remote Control* **79**, 870–883 (2018).
22. Rasche, C. & Pfeiffer, M. Training. in *Modelling and Simulation in Sport and Exercise* (eds. Baca, A. & Perl, J.) 187–207 (Routledge, 2019).
23. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
24. Swinton, P., Stephens Hemingway, B., Rasche, C., Pfeiffer, M. & Ogorek, B. Traditional and contemporary approaches to mathematical fitness-fatigue models in exercise science: A practical guide with resources. Part II. *Sport.* (2021).
25. Peng, R. D. Reproducible research in computational science. *Science* (2011) doi:10.1126/science.1213847.
26. Rozendaal, R. Modeling performance of elite cyclists: The Effect of Training on Performance. (Delft University of Technology, 2017).
27. Corlett, J. T., Calvert, T. W. & Banister, E. W. Cybernetics of Human Physical Performance. *Curr. Top. Cybern. Syst.* 180–182 (1978).
28. Wood, R. E., Hayter, S., Rowbottom, D. & Stewart, I. Applying a mathematical model to training adaptation in a distance runner. *Eur. J. Appl. Physiol.* **94**, 310–316 (2005).
29. Banister, E. W. & Hamilton, C. L. Variations in iron status with fatigue modelled from training in female distance runners. *Eur. J. Appl. Physiol. Occup. Physiol.* **54**, 16–23 (1985).
30. Candau, R., Busso, T. & Lacour, J. R. Effects of training on iron status in cross-country skiers. *Eur. J. Appl. Physiol. Occup. Physiol.* **64**, 497–502 (1992).
31. Banister, E. W., Morton, R. H. & Fitz-Clarke, J. R. Dose/Response Effects of Exercise Modeled from Training: Physical and Biochemical Measures. *Ann. Physiol. Anthropol.* **11**, 345–356 (1992).
32. Agostinho, M. F. *et al.* Perceived training intensity and performance changes quantification in judo. *J. Strength Cond. Res.* **29**, 1570–1577 (2015).
33. Busso, T., Carasso, C. & Lacour, J. R. Adequacy of a systems structure in the modeling of training effects on performance. *J. Appl. Physiol.* **71**, 2044–2049 (1991).
34. le Bris, S., Ledermann, B., Topin, N., Messner-Pellenc, P. & Le Gallais, D. High versus low training frequency in cardiac rehabilitation using a systems model of training. *Eur. J. Appl. Physiol.* **96**, 217–224 (2006).
35. le Bris, S., Ledermann, B., Topin, N., Messner-

- Pellenc, P. & Le Gallais, D. A systems model of training for patients in phase 2 cardiac rehabilitation. *Int. J. Cardiol.* **109**, 257–263 (2006).
36. Millet, G. P. *et al.* Modelling the transfers of training effects on performance in elite triathletes. *Int. J. Sports Med.* **23**, 55–63 (2002).
37. Millet, G. P., Gros Lambert, A., Barbier, B., Rouillon, J. D. & Candau, R. Modelling the relationships between training, anxiety, and fatigue in elite athletes. *Int. J. Sports Med.* **26**, 492–498 (2005).
38. Philippe, A. G. *et al.* Modeling the responses to resistance training in an animal experiment study. *Biomed. Res. Int.* **1**–7 (2015) doi:10.1155/2015/914860.
39. le Bris, S. *et al.* Applying a systems model of training to a patient with coronary artery disease. *Med. Sci. Sports Exerc.* **36**, 942–948 (2004).
40. Pearl, J. An introduction to causal inference. *Int. J. Biostat.* **6**, 7 (2010).
41. Ludwig, M. & Asteroth, A. Predicting performance from outdoor cycling training with the fitness-fatigue model. *Work. Model. Endur. Sport.* **3**–6 (2016).
42. Hamra, G., MacLehose, R. & Richardson, D. Markov chain Monte Carlo: an introduction for epidemiologists. *Int. J. Epidemiol.* **42**, 627–634 (2013).
43. Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python. in *Proceedings of the 9th Python in Science Conference* vol. 57 61 (2010).
44. Hellard, P. *et al.* Modeling the Residual Effects and Threshold Saturation of Training: A Case Study of Olympic Swimmers. *J. Strength Cond. Res.* **19**, 67–75 (2005).
45. Busso, T., Denis, C., Bonnefoy, R., Geyssant, A. & Lacour, J. R. Modeling of adaptations to physical training by using a recursive least squares algorithm. *J. Appl. Physiol. Model. Physiol.* **82**, 1685–1693 (1997).
46. Transtrum, M. K. *et al.* Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *J. Chem. Phys.* (2015) doi:10.1063/1.4923066.
47. Soetaert, K. & Petzoldt, T. Inverse modelling, sensitivity and monte carlo analysis in R using package FME. *J. Stat. Softw.* (2010) doi:10.18637/jss.v033.i03.
48. Busso, T. & Thomas, L. Using Mathematical Modeling in Training Planning. *Int. J. Sport. Physiol. Perform. Invt. Comment.* **1**, 400–405 (2006).
49. Borresen, J. & Lambert, M. I. The Quantification of Training Load, the Training Response and the Effect on Performance. *Sport. Med.* **39**, 779–95 (2009).
50. Banister, E. W., Carter, J. B. & Zarkadas, P. C. Training theory and taper: validation in triathlon athletes. *Eur. J. Appl. Physiol. Occup. Physiol.* **79**, 182–191 (1999).
51. Hayes, P. R. & Quinn, M. D. A mathematical model for quantifying training. *Eur. J. Appl. Physiol.* **106**, 839–847 (2009).
52. Jobson, S. A., Passfield, L., Atkinson, G., Barton, G. & Scarf, P. The analysis and utilization of cycling training data. *Sport. Med.* **39**, 833–844 (2009).
53. Moxnes, J. F. & Hausken, K. The dynamics of athletic performance, fitness and fatigue. *Math. Comput. Model. Dyn. Syst.* **14**, 515–533 (2008).
54. Perl, J. & Pfeiffer, M. PerPot DoMo: Antagonistic Meta-Model Processing two Concurrent Load Flows. *Int. J. Comput. Sci. Sport* **10**(2011).
55. 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).
56. Chalencon, S. *et al.* A Model for the Training Effects in Swimming Demonstrates a Strong Relationship between Parasympathetic Activity, Performance and Index of Fatigue. *PLoS One* **7**, 1–10 (2012).
57. Chalencon, S. Prediction of performance in swimming by Autonomic Nervous System activity assessment : mathematical modeling. (University of Lyon, 2015).
58. Ishii, H., Akahashi, S. T., Hiba, T. C., Aeda, A. M. & Akahashi, Y. T. Prediction of swim performance in junior female swimmers by dynamic system model. *Hum. Perform. Meas.* **5**, 1–8 (2008).
59. Mitchell, L. J. G., 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.* (2020) doi:10.1080/17461391.2020.1719211.
60. Mujika, I. *et al.* Modeled responses to training and taper in competitive swimmers. *Med. Sci. Sports Exerc.* **28**, 251–258 (1996).
61. Busso, T., Benoit, H., Bonnefoy, R., Feasson, L. & Lacour, J. R. Effects of training frequency on the dynamics of performance response to a single training bout. *J. Appl. Physiol.* **92**, 572–580 (2002).
62. Sanchez, A. M. J. *et al.* Modelling training response in elite female gymnasts and optimal strategies of overload training and taper. *J. Sports Sci.* **31**, 1510–1519 (2013).
63. 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).
64. Suzuki, S., Sato, T., Maeda, A. & Takahashi, Y. Program design based on a mathematical model using rating of perceived exertion for an elite Japanese sprinter: A case study. *J. Strength Cond. Res.* **20**, 36–42 (2006).
65. Wallace, L. K., Slattery, K. M. & Coutts, A. J. A comparison of methods for quantifying training load: Relationships between modelled and actual training responses. *Eur. J. Appl. Physiol.* **114**, 11–20 (2014).
66. Zarkadas, P. C., Carter, J. B. & Banister, E. W. Modelling the effect of taper on performance, maximal oxygen uptake, and the anaerobic threshold in endurance triathletes. *Adv. Exp. Med. Biol.* **393**, 179–186 (1995).
67. Lucia, A., Hoyos, J., Santalla, A., Earnest, C. & Chicharro, J. L. Tour de France versus Vuelta a España: Which is harder? *Med. Sci. Sports Exerc.* **35**, 872–878 (2003).
68. Swinton, P. A., Stephens Hemingway, B., Saunders, B., Gualano, B. & Dolan, E. A Statistical Framework to Interpret Individual Response to Intervention: Paving the Way for Personalized Nutrition and Exercise Prescription. *Front. Nutr.* (2018) doi:10.3389/fnut.2018.00041.
69. Williams, S. *et al.* Modelling the HRV response to

- training loads in elite rugby sevens players. *J. Sport. Sci. Med.* **17**, 402–408 (2018).
70. Stephens Hemingway, B., Burgess, K., Elyan, E. & Swinton, P. The effects of measurement error and testing frequency in applying the Fitness Fatigue Model to resistance training: A simulation study. *Int. J. Sport. Sci. Coach.* **0**, 1–12 (2019).
 71. Stevens, J. *Applied Multivariate Statistics for the Social Sciences*. (Lawrence Erlbaum Associates, 1986). doi:10.2307/2685203.
 72. Bates, D. M. & Watts, D. G. *Nonlinear Regression Analysis and Its Applications*. (Wiley, 1988). doi:10.2307/1268866.
 73. Davidian, M. & Giltinan, D. M. Nonlinear models for repeated measurement data: An overview and update. *J. Agric. Biol. Environ. Stat.* (2003) doi:10.1198/1085711032697.
 74. Sen, A. & Srivastava, M. *Regression Analysis: Theory, Methods and Applications*. (Springer, 1990). doi:10.2307/3620306.
 75. Al-Otaibi, N. M. Statistical modelling of training and performance using power output and heart rate data collected in the field. (University of Salford, 2017).
 76. Shrahili, M. Modelling and optimising the sport and exercise training process. (University of Salford, 2014).
 77. Scarf, P., Shrahili, M., Alotaibi, N., Jobson, S. A. & Passfield, L. Modelling the effect of training on performance in road cycling: estimation of the Banister model parameters using field data. *arXiv Prepr.* 1–14 (2019).
 78. Everitt, B. & Skrondal, A. *The Cambridge dictionary of statistics*. vol. 106 (Cambridge University Press Cambridge, 2002).
 79. Domotor, Z. Philosophy of Science, Mathematical Models in. in *Mathematics of Complexity and Dynamical Systems* (ed. Meyers, R. A.) 1407–1422 (Springer New York, 2011). doi:10.1007/978-1-4614-1806-1_89.
 80. Morton, R. H. The Quantitative Periodization of Athletic Training: A Model Study. *Sport. Med. Train. Rehab* **3**, 19–28 (1991).
 81. Thomas, L., Mujika, I. & Busso, T. Computer simulations assessing the potential performance benefit of a final increase in training during pre-event taper. *J. Strength Cond. Res.* **23**, 1729–1736 (2009).
 82. Prado, M. L. De. *Advances in Financial Machine Learning*. (John Wiley & Sons Inc, 2018).
 83. Breiman, L. Bagging predictors. *Mach. Learn.* (1996) doi:10.1007/bf00058655.
 84. Morton, R. H. & Fukuba, Y. Professor Eric W. Banister, 1933–2010: Obituary. *Res. Sport. Med.* **19**, 144–144 (2011).
 85. Fitz-Clarke, J. R., Morton, R. H. & Banister, E. W. Optimizing athletic performance by influence curves. *J. Appl. Physiol.* **71**, 1151–1158 (1991).
 86. Jeffries, A. *et al.* Development of a revised conceptual framework of physical training for measurement validation and other applications. *Preprint* (2020) doi:10.31236/osf.io/wpvek.
 87. Ludwig, M., Schaefer, D. & Asteroth, A. Training simulation with nothing but training data simulating performance based on training data without the help of performance diagnostics in a laboratory. *icSPORTS 2016 - Proc. 4th Int. Congr. Sport Sci. Res. Technol. Support* 75–82 (2016) doi:10.5220/0006042900750082.
 88. Ramsay, J. O., Hooker, G., Campbell, D. & Cao, J. Parameter estimation for differential equations: A generalized smoothing approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69**, 741–796 (2007).
 89. Lamarra, N. Variables constants and parameters: clarifying the system structure. *Med. Sci. Sports Exerc.* **22**, 88–95 (1990).

Appendix A: Tables of nomenclature, abbreviations, and key terms

A1 Table of nomenclature

Variable	Description
$p(n)$	Measured performance representing an estimate of latent (true) performance, used to fit the model
$\hat{p}(n)$	Modelled performance (FFM)
$g(n)$	FFM fitness component: Sum of decayed impulse-response functions from days 1 to t
$h(n)$	FFM fatigue component: Sum of decayed impulse-response functions from days 1 to t
$\omega(n)$	Discrete time-series of training load values (model input)
Parameter	Description
Δ_t	Discrete time-step size (typically set to 1-day)
p^*	Baseline performance (i.e. model intercept): typically estimated during the optimisation process
k_r	Model component scaling factor ($r \in \mathbb{N}$)
τ_r	Model component (decay) time constant ($r \in \mathbb{N}$)

A2 Table of abbreviations

Abbreviation	Term
FFM	Fitness-Fatigue Model
IR	Impulse-Response
TRIMP	Training Impulse
RSS	Residual Sum of Squares
NLS	Non-linear least-squares

A3 Glossary of key terms

Term	Definition
Criterion performance	An observed (measured) performance for a dependent variable that acts as the fitness-fatigue model target. It is considered an estimate of true (latent) performance ⁷⁷ .
Cross-validation	A class of model prediction evaluation and tuning techniques (broadly categorised as exhaustive and non-exhaustive) that facilitate observational assessment of prediction accuracy or can be used to further tune models.
Dynamical systems modelling	A process that transforms a set of m input functions, with values as functions of time $t \in [0, T]$ indicated by vector $\mathbf{u}(t)$, into a set of d output functions with values $\mathbf{x}(t)$ ⁸⁸ .
Model assumptions	Abstractions of observable structure in real-world phenomena that define a ‘model world’ as part of the process of mathematical model development.
Model fit	A statistic or visual plot describing correspondence between measured and modelled performance within the training dataset.
Model predictions	The time-series values computed by the model under a set of input values (training load) and parameter estimates (obtained from fitting to observed performance data).
Model prediction accuracy	A statistic or visual plot describing correspondence between measured data not seen by the model (i.e. out-of-sample) and the model predicted values from a set of parameter estimates fitted to the ‘training’ dataset.
System	An entity which can be characterised by at least one input and one output variable, whose internal behaviour determines the relationship between input and output ⁸⁹ .
Training load / impulse series	A discrete time-series of a single numerical quantity that summarises training stress of a given day relative and comparable to previous and future training and which maps to the underlying performance measure.
True performance	Latent ‘unobservable’ performance. It is defined as a random variable with the state of the athlete (preparedness) as its expectation ⁷⁷

Appendix B: Supplementary code resources

www.fitnessfatigue.com or github.com/bsh2/fitness-fatigue-models (direct to repository)

The fitness-fatigue model project is focussed on developing open-source, robust, and flexible utilities for fitting and evaluating fitness-fatigue models (FFMs); with modern optimisers, out-of-sample assessment, and input checking. In addition, illustrative code in a didactic-style has been constructed and which guides the user through the application of several advanced methods via an interactive procedural flow. The latter resources forming the focus of part II ²⁴ in this review series, and the former relating closer to this review article representing a basic introduction to fitting and evaluating FFMs. Thorough documentation, explanatory code examples, and mock data are provided for each utility (fitting) function. Before this project, there has never been a centralised repository code suitable for the study of fitness-fatigue and related models.

The goal of this project is to expedite research and development of fitness-fatigue models and methods in the sport and exercise sciences, and to reduce initial barriers to study. It is hoped the resources contained in the repository will be useful to researchers, educators, students, and practitioners alike. The open-source nature of the repository encourages others to contribute code, offer improvements to existing methods, and assess the processes going on under the hood. The repository is not an endorsement of these models as “ready to use” in practice, but rather should be thought of as a set of scientific tools to aid understanding and future development.

The tools were written in the statistical programming language R ²³, given the large increase in uptake and interest amongst sport scientists for analysing experimental data over the last five years. It was therefore hoped that this medium would reach the largest audience in its initial form. There is scope in the future to explore the use of interactive dashboards if suitable models can be identified, and several planned future developments are ongoing for the repository to expand the range of available tools.

Appendix C: Derivation of solutions to fitness-fatigue systems

The systems of differential equations that define FFM's can be solved using the method of Laplace transform. The Laplace transform \mathcal{L} is an integral transform that can be used to convert a differential equation into an algebraic equation. The algebraic equation can then be solved and then the inverse Laplace transform used to find the solution to the original differential equation. The Laplace transform converts a function of a real variable t , to a function of a complex variable s , where

$$\mathcal{L}\{f(t)\} = F(s) = \int_0^{\infty} f(t)e^{-st} dt.$$

To use the Laplace transform to solve differential equations, the Laplace transform of a derivative is required, where

$$\mathcal{L}\{f'(t)\} = sF(s) - f(0); \mathcal{L}\{f''(t)\} = s^2F(s) - sf(0) - f'(0).$$

Frequently the final step of solving the differential equation will involve the inverse Laplace transform of a product $F(s) = M(s)N(s)$ such that:

$$f(t) = (m * n)(t) = \int_0^t m(u)n(t-u)du$$

Where $m(t) * n(t)$ is the convolution of the functions m and n .

Also, the inverse Laplace transform of $1/(s-a)$ where a is a constant is e^{at} .

With the above we can solve the independent system of differential equations that define the standard FFM

$$\begin{aligned} g'(t) &= \omega(t) - \frac{1}{\tau_g}g(t) \\ h'(t) &= \omega(t) - \frac{1}{\tau_h}h(t). \end{aligned}$$

We simply solve each differential equation in isolation. We show how this would be achieved for fitness $g(t)$. Taking the Laplace transform of each term gives:

$$sG(s) - g(0) = \Omega(s) - \frac{1}{\tau_g}G(s)$$

Where $G(s)$ is the Laplace transform of $g(t)$ and $\Omega(s)$ is the Laplace transform of $\omega(t)$. Under the assumption $g(0) = 0$, we can rearrange the equation above to give the following

$$sG(s) + \frac{1}{\tau_g}G(s) = \Omega(s)$$

$$G(s) \left(s + \frac{1}{\tau_g} \right) = \Omega(s)$$

$$G(s) = \Omega(s) \left(\frac{1}{\left(s + \frac{1}{\tau_g} \right)} \right)$$

On the right hand side we have a product and therefore taking the inverse Laplace transform gives

$$g(t) = \omega(t) * e^{-\frac{t}{\tau_g}} = \int_0^t e^{-\frac{(t-u)}{\tau_g}} \omega(u) du.$$

This is the solution that was presented in eq.3. Given FFM's are generally conceptualised as impulses performed on single days, the integrals expressed in the solution of the system are discretised and approximated with the area of rectangles. The discretisation is highlighted by changing from continuous time (t) to discrete time (n). In the exercise science literature any training impulse on day n is omitted such that the index of the series terminates at $n - 1$ to give the following approximation:

$$g(n) = \Delta t \sum_{i=0}^{n-1} \omega(i) e^{-\frac{(n-i)}{\tau_g}}$$

Setting $\Delta t = 1$ and $w(0) = 0$, we obtain

$$g(n) = \sum_{i=1}^{n-1} \omega(i) e^{-\frac{(n-i)}{\tau_g}}$$

Repeating the same process for fatigue $h(n)$ and then combining with the scaling constants and baseline performance gives the familiar form presented for the standard FFM

$$\hat{p}(n) = p^* + \underbrace{k_g \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_g}}}_{\text{fitness component}} - \underbrace{k_h \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_h}}}_{\text{fatigue component}}$$

We will now show that the system of differential equations that defines the FFM with a delay term introduced by Calvert et al. ³ is the following

$$\left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right)^{-1} g''(t) = \omega(t) - \left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right)^{-1} \left(\frac{1}{\tau_{g_1}} + \frac{1}{\tau_{g_2}}\right) g'(t) - \left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right)^{-1} \frac{1}{\tau_{g_1} \tau_{g_2}} g(t)$$

$$h'(t) = \omega(t) - \frac{1}{\tau_h} h(t)$$

As identified at the beginning of this appendix, the same routine can be performed on fatigue $h(n)$ to obtain

$$h(n) = \sum_{i=1}^{n-1} \omega(i) e^{-\frac{(n-i)}{\tau_h}}$$

More work is required for $g(t)$. Taking the Laplace transform of each term gives

$$\left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right)^{-1} (s^2 G(s) - s g(0) - g'(0)) = \Omega(s) - \left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right)^{-1} \left(\frac{1}{\tau_{g_1}} + \frac{1}{\tau_{g_2}}\right) (s G(s) - g(0)) - \left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right)^{-1} \frac{1}{\tau_{g_1} \tau_{g_2}} G(s)$$

Setting $g(0) = g'(0) = 0$ and rearranging gives

$$\begin{aligned} \left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right)^{-1} \left(s^2 G(s) + \left(\frac{1}{\tau_{g_1}} + \frac{1}{\tau_{g_2}}\right) s G(s) + \frac{1}{\tau_{g_1} \tau_{g_2}} G(s) \right) &= \Omega(s) \\ G(s) \left(s^2 + \left(\frac{1}{\tau_{g_1}} + \frac{1}{\tau_{g_2}}\right) s + \frac{1}{\tau_{g_1} \tau_{g_2}} \right) &= \left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right) \Omega(s) \\ G(s) \left(\left(s + \frac{1}{\tau_{g_1}}\right) \left(s + \frac{1}{\tau_{g_2}}\right) \right) &= \left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right) \Omega(s) \\ G(s) &= \frac{\left(\frac{1}{\tau_{g_2}} - \frac{1}{\tau_{g_1}}\right)}{\left(s + \frac{1}{\tau_{g_1}}\right) \left(s + \frac{1}{\tau_{g_2}}\right)} \Omega(s) \end{aligned}$$

Partial fraction decomposition gives

$$G(s) = \Omega(s) \left(\frac{1}{\left(s + \frac{1}{\tau_{g_1}}\right)} \right) - \Omega(s) \left(\frac{1}{\left(s + \frac{1}{\tau_{g_2}}\right)} \right)$$

Application of the inverse Laplace transform gives

$$g(t) = \omega(t) * e^{-\frac{t}{\tau_{g_1}}} - \omega(t) * e^{-\frac{t}{\tau_{g_2}}} = \int_0^t e^{-\frac{(t-u)}{\tau_{g_1}}} \omega(u) du - \int_0^t e^{-\frac{(t-u)}{\tau_{g_2}}} \omega(u) du$$

Approximation and combining the fitness and fatigue components with the associated scaling coefficients and baseline performance gives Calvert's model presented in eq.11

$$\hat{p}(n) = p^* + k_g \cdot \sum_{i=1}^{n-1} \omega(i) \cdot \left(\underbrace{e^{-\frac{(n-i)}{\tau_{g_1}}}}_{\text{effect}} - \underbrace{e^{-\frac{(n-i)}{\tau_{g_2}}}}_{\text{delay}} \right) - k_h \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_h}}$$

Appendix D: Physiological correlates of model fitness and fatigue traces

Ref	<i>n</i>	Domain	Duration	Physiological Measures	Fitness Component	Fatigue Component
27	1	Lab-based general exercise	4 months	Serum Bilirubin & Maximal Oxygen Uptake	Maximal oxygen uptake demonstrated moderate correlations with traces of the fitness component (visual plot only, no quantitative data provided)	Bilirubin demonstrated moderate correlations with traces of the fatigue component (visual plot only, no quantitative data provided)
29	5	Distance Running (Amateur)	300 Days	Serum Iron	No data provided	All markers of iron status (Ferritin, RBC and Haemoglobin) appear to demonstrate low-moderate correlations with fatigue trace. Inter-individual variation evident, with correlations likely stronger for some individuals than others
11	6	Weight-lifting (Elite)	1 year	Serum testosterone (T), cortisol (C), SHBG as well as ratios between T:C and T:SHBG	[Testosterone, $r = 0.61 - 0.92$](significant for $n=2$); [Cortisol, $r = -0.75 - 0.76$]; [SHBG, $r = -0.64 - 0.75$]; [Testosterone:Cortisol, $r = 0.11 - 0.94$](significant for $n=2$); [Testosterone:SHBG, $r = -0.27 - 0.89$](significant for $n=1$)	[Testosterone, $r = 0.34 - 0.88$](significant for $n=1$); [Cortisol, $r = -0.75 - 0.75$]; [SHBG, $r = -0.75 - 0.76$]; [Testosterone:Cortisol, $r = 0.0 - 0.97$](significant for $n=1$); [Testosterone:SHBG, $r = -0.80 - 0.92$](significant for $n=1$)
31	2	Cycling (Amateur)	28 days	Serum biomarkers and enzymes including: Lactate dehydrogenase, creatine kinase, and aspartate transaminase	No graphs or data reported for correlations between any measured physiological marker and fitness parameter	Fatigue and LDH showed a degree of correlation across all subjects, however these would likely be poor-moderate Temporal trends out of phase with LDH generally lagging behind observed changes in fatigue. Similar trend apparent from fatigue and creatine kinase. Greatest correlation appears to be between fatigue and aspartate transaminase (GOT) with much less variation between subjects and data points fitting much closer. Likely moderate-strong.
8	6	Weight-lifting (Elite)	6 weeks	Serum testosterone, luteinising hormone, SHBG and T:SHBG ratio	Mean Serum Luteinising Hormone ($r=0.97$) during training period 1 [appears to be for all subjects]	Nothing reported for fatigue parameter in either table, graph or written format. Same data set utilised in Busso 1990, however serum LH results not reported in earlier work.
28	1	Running (Amateur)	12 weeks	Velocity at ventilatory threshold (VTRS), Profile-of-mood-states (POMS) and VO ₂ Max, Running economy	[VTRS, $r=0.94$]; [VO ₂ Max, $r=0.42$]; [Running Economy, $r=-0.61$]	[POMS, $r=0.75$]

Appendix E: Model formulae reference table

Model	Formula
	<p>One component (i.e. fitness only) – The basic model</p> $p(t) = p^* + Kp(t)$ $p'(t) = \omega(t) - \frac{1}{\tau}p(t)$
Model systems of differential equations	<p>Two components (i.e. fitness and fatigue) – The standard model</p> $p(t) = p^* + k_1g(t) - k_2h(t)$ $g'(t) = \omega(t) - \frac{1}{\tau}g(t)$ $h'(t) = \omega(t) - \frac{1}{\tau}h(t)$
	$\hat{p}(n) = p^* + k_g \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_g}} - k_h \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_h}}$ <p>Alternatively, this can be written with a single summation as follows:</p> $\hat{p}(n) = p^* + \sum_{i=1}^{n-1} \left(k_g \cdot e^{-\frac{(n-i)}{\tau_g}} - k_h \cdot e^{-\frac{(n-i)}{\tau_h}} \right) \cdot \omega(i)$
The standard model (2 components)	<p>This could also be adjusted to incorporate a single scaling factor to address criticisms^{18,20} of parameter interdependence:</p> $\hat{p}(n) = p^* + K \cdot \sum_{i=1}^{n-1} \left(e^{-\frac{(n-i)}{\tau_g}} - e^{-\frac{(n-i)}{\tau_h}} \right) \cdot \omega(i)$ <p>And include additional delay terms on the fitness component³</p> $\hat{p}(n) = p^* + k_g \cdot \sum_{i=1}^{n-1} \omega(i) \cdot \left(e^{-\frac{(n-i)}{\tau_{g1}}} - e^{-\frac{(n-i)}{\tau_{g2}}} \right) - k_h \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_h}}$
The general model (N components)	$\hat{p}(n) = p^* + \sum_{r=1}^N \left(k_r \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_r}} \right)$ <p>$N \in \mathbb{N}; (r = 1, \dots, N)$</p>
The general model with initial components (N components with P initial components)	$\hat{p}(n) = p^* + \sum_{r=1}^N \left(k_r \cdot \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_r}} \right) + \sum_{p=1}^P q_p^0 \cdot e^{-\frac{n}{\tau_p}}$ <p>Where q_p^0 denotes the initial level of the component, and τ_p is the decay time constant on the corresponding initial level.</p>
Baseline performance model	$\hat{p}_{n_j} = p_{n_{j-1}} + k_g \cdot \sum_{i=n_{j-1}}^{n_j-1} e^{-\frac{(n_j-i)}{\tau_g}} \cdot \omega(i) - k_h \cdot \sum_{i=n_{j-1}}^{n_j-1} e^{-\frac{(n_j-i)}{\tau_h}} \cdot \omega(i)$ <p>Where j is a measured performance, n_j is the day of the j-th measured performance, and p_{n_j} is the observed value.</p>