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

Paul Swinton [1*], Ben Stephens Hemingway[1,2], Christian Rasche[3], Mark Pfeiffer[3], Ben Ogorek[4]

DOI (TBC)
SportRχiv hosted preprint version 1
Last updated 26/02/2021
PREPRINT - NOT PEER REVIEWED

Institutions

[1] School of Health Sciences, Robert Gordon University, Aberdeen, UK

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

[3] Faculty Theory and Practice of Sports, Johannes Gutenberg University of Mainz, Mainz, Germany

[4] Nousot, Chicago, IL, USA

*** Corresponding Author**

Dr. Paul Swinton
School of Health Sciences, Robert Gordon University
Garthdee Road
Aberdeen, UK,
AB10 7QG
p.swinton@rgu.ac.uk, +44 (0) 1224 262 3361

Disclosure of funding

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

Conflicts of interest

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

Supplementary material

Available at fitnessfatigue.com – Refer to section 1 for an introduction to these resources, and appendix for further detailed information. Resources associated with this part of the review series are focussed on illustrative code to guide the user via an interactive procedural flow. Maintained by baogorek@gmail.com and b.stephens-hemingway@rgu.ac.uk

Keywords

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

Abstract:

The standard fitness-fatigue model (FFM) is known to include several limitations described by the linearity assumption, the independence assumption, and the deterministic assumption. These limitations ensure that the modelled response to chronic training does not match the complexity observed in practice. The purpose of part II of this review series was to describe previous extensions to the standard FFM to address these limitations, providing key mathematical insights and resources to both explain technical elements and enable researchers and practitioners to fit these extended models to their own data. To address the linearity assumption of the standard FFM and the associated limitation that doubling the training load predicts twice the performance improvement, two distinct extensions are reviewed including the addition of a non-linear transform to training inputs and inclusion of non-linear terms within the system of differential equations. To address the independence assumption where the response to a training session is unaffected by previous sessions, a popular extension where fatigue is updated as an exponentially weighted moving average of previous training loads is reviewed. Finally, the review introduces the concept of state-space models where uncertainty in the estimates of fitness, fatigue and performance measurement can be directly modelled eliminating the unsuited deterministic assumption of the standard FFM. The review also highlights how state-space models can be further expanded to include features such as the Kalman filter where parameter estimates can be updated with incoming performance measurements to better predict and manipulate training to optimise performance. Collectively, the range of topics covered in this review series and the resources provided should enable researchers and practitioners to better investigate the extensive area of FFMs and determine in what contexts models can assist with training monitoring and prescription.

1.0 Introduction

Individualised exercise prescription in sport research and practice relies heavily on coach and athlete experience, prior knowledge of physiological mechanisms underpinning average response, and can be supplemented with quantitative monitoring approaches (Scarf et al., 2019). The process of individualised exercise prescription and the subsequent training regimes developed can be further enhanced through increased quantitative monitoring, modelling and prediction (Scarf et al., 2019). Each component of this process represents an active area of research where many open problems remain. Ideally, any performance model developed would possess specific qualities relating training to response with respect to causality, predictability, interdependence, non-linearity, relevance to specific training variables, and ease of identifying underlying physiological structures (Rasche & Pfeiffer, 2019). However, development of any performance model is an iterative process of refinement, and it is unlikely an ideal model will be derived in a first attempt. It was over forty years ago that Banister and colleagues (Banister et al., 1975) described the first dynamical systems model of training response referred to as the standard fitness-fatigue model (FFM). Despite consistent scientific attention over a long history, their model and associated extensions remain in the domain of exploratory research. It was identified in part I of this review series that there has been little uptake of FFMs across contemporary research and in real world practical environments (Stephens Hemingway et al., 2021). This is likely explained by a combination of factors including: 1) research that assumes readers are proficient with requisite mathematical concepts and have an awareness of historical conventions, nomenclature and terminology; 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. We attempted to address some of these factors in part I of this review series by providing a detailed overview of FFMs and the overarching general model (Stephens Hemingway et al., 2021). We also reviewed four key steps of fitness-fatigue modelling including training load quantification, criterion performance selection, parameter estimation and model evaluation, the latter two steps of which have received minimal discussion previously. From a technical point of view the modelling process can be conceptualized as illustrated in figure 1, regardless of the actual performance model being used. Initially, the given performance model is trained by fitting the model parameters. This procedure may be understood as learning the athlete-specific dynamics in terms of an individual fingerprint, which could help to understand adaptational characteristics. In practice the data usually consists of measured load and performance data in combination with an initial set of model parameters. Through an iterative parameter optimisation process aimed at minimising a loss function an optimised set of parameters is generated, which results in performance estimates in conjunction with a measure of in-sample model-fit. In a second step the performance model is tested to validate the optimised parameter set using either a fraction of the measured load, which was not included in the fitting process, or future training load of interest. The resulting performance estimates estimate the out-of-sample model-fit. In that way, besides predicting future performance, existing knowledge of researchers and practitioners regarding the individual effects of given load on performance may be extended.

Finally, in part I of this review series we identified novel approaches that could be used to enhance the use of basic FFMs including a focus on predicting physical performance in simple biomechanical tasks or exercises, the generation of high frequency performance data through routinely collected training metrics, and development of unique training load weightings to account for multiple modalities (e.g. hypertrophy, strength or power training) and their specificity according to the performance outcome (Stephens Hemingway et al., 2021). Whilst these recommendations may provide productive approaches and avenues for future research, it is important to recognise that they fail to address some of the limitations inherent to the structure of basic FFMs. It has been known for an extended period that the standard FFM and subsequent first-order filter extensions suffer from several limitations that are at odds with the conceptual understanding of training response (Banister et al., 1975; Busso et al., 1991, 1994; Calvert et al., 1976; Morton et al., 1990; Rasche & Pfeiffer, 2019). Three key limitations include: 1) the *linearity assumption*, where performance can be arbitrarily increased by simply increasing the training dose and for example doubling the training dose leads to double the improvement (Hellard et al., 2006; Rasche & Pfeiffer, 2019); 2) the *independence assumption*, where there is no interaction between training sessions and therefore training performed on a given day does not influence the response generated from another session; and 3) the *deterministic assumption*, where uncertainty in model parameters and observed performance are not modelled directly and are not updated based on incoming information (Kolossa et al., 2017). Two general approaches have been proposed to address limitations inherent to basic FFMs: 1) altering the model input through constraining or saturating training loads (Hellard et al., 2005), and 2) altering the model formulation (Busso, 2003; Kolossa et al., 2017; Turner et al., 2017). Under the first of these approaches, Hellard et al. (2005) proposed the use of a threshold saturation function, termed the Hill function (Krzyzanski et al., 1999) to address the limitations of the linearity assumption and the result that indefinite increases in training loads result in continual increases in performance. Their use of a threshold saturation function to transform training loads to a non-linear input with asymptote worked outside of the model formulation and therefore avoided changing the structural relationship between the input and output specified (Rasche & Pfeiffer, 2019). In contrast, under the second approach the structural relationship between model components is changed with previous research focussing on modifications to account for each of the three key limitations previously identified. In part II of this review series, we introduce and detail examples of both approaches. For changes in the model input we examine the Hill function. For changes in model specification, we include the variable dose-response (VDR) model designed to create interactions between subsequent training sessions, the Kalman filter to explicitly model uncertainty in model parameters and update model results based on incoming information, and inclusion of non-linear power terms to capture non-linearities and saturation effects.

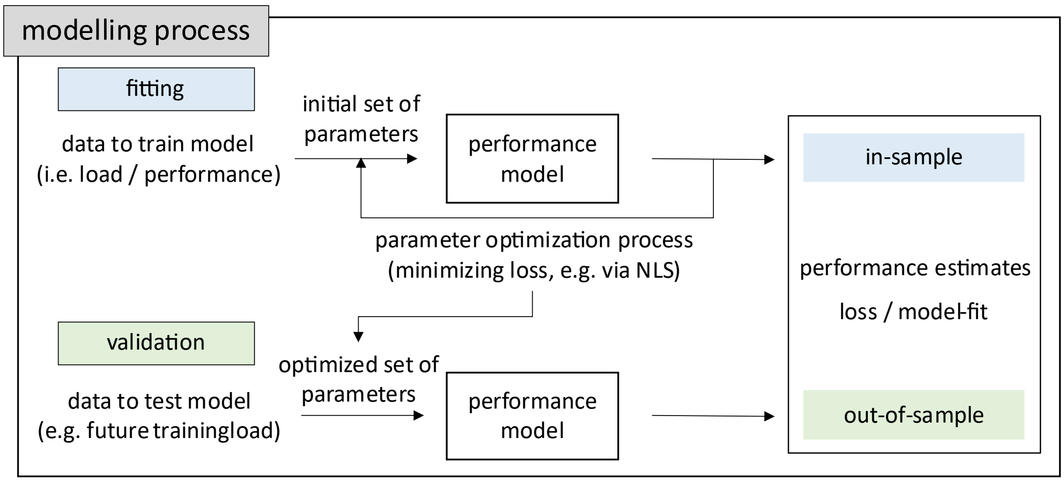

Figure 1: *The performance modelling process in terms of fitting and validating a given performance model by an estimation of in- and out-of-sample model-fit using corresponding performance estimates.*

Effective development of data-driven modelling approaches to assist with individualised prescription is likely to require greater numbers of researchers and practitioners to adopt and investigate more advanced models. However, it appears that as with a lack of uptake of even basic FFMs, reticence of researchers and practitioners to investigate and implement more advanced FFMs is influenced by limited exposure and availability of practical resources. Therefore, the purpose of part II of this review series is to detail advanced model developments and provide clarity to the underlying mathematics and associated strengths and limitations. Importantly, the review is accompanied by supplementary code resources written in the commonly used language R, provided in a dedicated online repository (fitnessfatigue.com, github.com/bsh2/fitness-fatigue-models). Included in this repository are didactic code resources that serve to provide clarity of methods and associated output described in this review, as well as the option to apply utility-driven functions offering user-friendly functionality to fit and assess models. Clear and detailed documentation is provided with each code resource (see appendix for further details). Additionally, interactive R notebooks are hosted on Kaggle's platform and these can be located through fitnessfatigue.com or links provided in the appendix.

2.0 Contemporary methods to address limitations

To enable this part of the review to stand alone and provide a comparator to the more advanced models that are the focus, we briefly outline the standard FFM developed by Banister and colleagues. The model takes as input a series of training load impulses that create positive ("fitness") and negative ("fatigue") responses, where modelled performance is a linear sum of these two antagonistic outputs (eq.1). The fitness $g(t)$ and fatigue $h(t)$ responses are scaled by multiplicative factors that depend upon the units used to measure the training load and performance, and typically follow a relationship $k_h \geq k_g > 0$ (Calvert et al., 1976),

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

where the parameter p^* represents a baseline performance level that an athlete is unlikely to fall chronically below. The underlying dynamical system was specified by a set of first-order differential equations (eq.2) with rate decay constants $\tau_g, \tau_h > 0$ for fitness and fatigue, respectively,

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

In the appendices of part I of this review series (Stephens Hemingway et al., 2021), we present the derivation of the solution to this system of differential equations and the discrete approximation (eq.3) that is commonly presented in the literature.

$$\begin{aligned} &\text{fitness response} \qquad g(n) = \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_g}} \\ &\text{fatigue response} \qquad h(n) = \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_h}} \end{aligned} \qquad (eq.3)$$

where continuous time t, has been discretised to day n. Additionally, eq.3 can be expressed in recursive form

$$\begin{aligned} g(n) &= [g(n-1) + \omega(n-1)]e^{-\frac{1}{\tau_g}} \\ h(n) &= [h(n-1) + \omega(n-1)]e^{-\frac{1}{\tau_h}}. \end{aligned} \qquad (eq.4)$$

2.1 Modification to model input and inclusion of non-linearity

Prior to modifying the formulation of the standard FFM, an approach to address the limitation of arbitrarily large increases in performance is to apply restrictions to unconstrained training inputs (ω). An initial method to consider is to simply adopt an interpretable scale with a maximum value (e.g. 0 to 100). If for example a maximum value of 100 is set, then from the standard model (eq.2) the maximum amount of fitness (or fatigue) that can be achieved from a single training session is $100k_g$ (or $100k_h$) where the units of the multiplicative factors match that of the performance being modelled. Additionally, given the nature of general FFMs and their solutions involving exponential functions, these fitness and fatigue effects are at their maximum immediately after a training session. It follows that researchers and practitioners can identify maximum improvements that could reasonably occur over a short period (e.g. 1-week) and use this value to set upper bounds when estimating model parameters. Whilst this approach assists with creating more interpretable parameters and training inputs, it does not address limitations of linearity in training response. A potential solution to this limitation was proposed by Hellard et al. (2005). The authors proposed the use of the Hill function that mapped training inputs in a non-linear fashion to a threshold (κ), such that higher inputs have no additional effect on performance. Figure-2 demonstrates the behaviour of the Hill saturation function, as recreated from parameter values provided in Hellard et al. (2005). The Hill function is described by the following equation:

$$Hill(\omega) = \kappa\left(\frac{\omega^{\gamma}}{\delta^{\gamma} + \omega^{\gamma}}\right), \qquad (eq. 5)$$

where the parameter γ expresses the sensitivity to the training load and controls the time to reach the threshold κ (larger γ leads to shorter times to reach κ), and δ is the inertia of the function to the threshold value (low δ expresses a strong effect of dose on performance). In applying the Hill function to the standard FFM there are multiple options available. One option includes setting κ to an interpretable maximum value (e.g., 100) and then setting γ, δ to control the non-linearity desired. The scaling parameters k_g and k_h would then be used to transform training loads to change values in the performance measure. Alternatively, there is potential to develop more complex transforms for a given training input (ω) and map this to different non-linear fitness $\{\kappa_g, \gamma_g, \delta_g\}$ and fatigue $\{\kappa_h, \gamma_h, \delta_h\}$ forms. Here, κ_g and κ_h replace the scaling coefficients k_g and k_h and represent the maximum amount of fitness and fatigue expected from a single training session. The parameters γ_g, δ_g, and γ_h, δ_h tune the degree of non-linearity for fitness and fatigue, with performance estimated with

$$\hat{p}(n) = p^{*} + \sum_{i=1}^{n-1} \omega_g(i) \cdot e^{-\frac{(n-i)}{\tau_g}} - \sum_{i=1}^{n-1} \omega_h(i) \cdot e^{-\frac{(n-i)}{\tau_h}}, \qquad (eq. 6)$$

where the fitness component is defined as

$$\omega_g(i) = \kappa_g\left(\frac{\omega(i)^{\gamma}}{\delta_g^{\gamma} + \omega(i)^{\gamma}}\right), \qquad (eq. 7)$$

and the fatigue component is defined as

$$\omega_h(i) = \kappa_h\left(\frac{\omega(i)^{\gamma}}{\delta_h^{\gamma} + \omega(i)^{\gamma}}\right). \qquad (eq. 8)$$

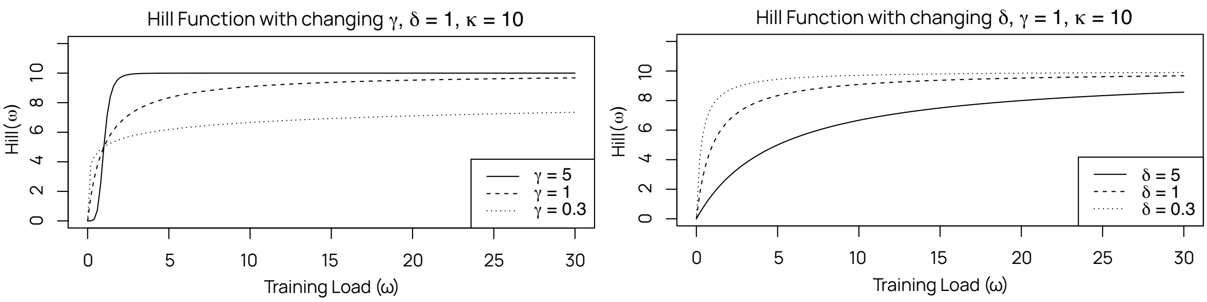

Figure 2: *Arbitrary behaviour of Hill training load saturation function (eq.5) when $\kappa = 10$, and γ, δ vary as shown. Plots recreated from parameter values provided in Hellard et al. (2005)*

Hellard et al. (2005) investigated whether modifying training input with a Hill function resulted in better model fit compared with standard linear input across 7 elite swimmers engaging in a long study period (100-200 weeks). After statistically controlling for the additional parameters, Hellard et al. (2005) identified slightly improved model fit with the addition of the Hill function (R^2_{adj} = 0.43 ± 0.1 vs. 0.36 ± 0.1). However, goodness-of-fit was poor across all athletes and both models. The authors included no out-of-sample assessment of model predictions, and so it is difficult to evaluate the extent of model overfit considering the additional flexibility the Hill function affords. It is possible that fitting the model over a long duration (100-200 weeks) contributed to the generally poor fit across both models, as parameters obtained may not remain stable for such long periods, either due to estimation issues or underlying change (discussed in part I of this review series). In the supplementary files we include an optimisation routine that is well suited to estimating the Hill function parameters for fitness and fatigue.

2.2 Modification to model specificity and inclusion of interaction between training sessions

Identified in section 1.0, one of the key limitations of the standard FFM is the independence assumption such that the response to the current training session is not influenced by previous sessions. Busso (2003) was the first to propose a model to address this limitation that included the addition of a first-order filter on the fatigue component and is commonly referred to as the variable dose-response (VDR) model. To account for the interaction between training sessions Busso (2003) focused on the perspective that fatigue was most likely to be influenced by previous training (with higher previous training loads generating greater fatigue) and so introduced a further 'gain' term for fatigue with associated time constant τ_{h_2}. One specific implementation of the model can be presented as

$$\hat{p}(n) = p^* + k_g \sum_{i=1}^{n-1} \omega(i) \cdot e^{-\frac{(n-i)}{\tau_g}} - k_h \sum_{i=1}^{n-1} k_{h_2}(i) \cdot e^{-\frac{(n-i)}{\tau_h}}, \qquad (eq.9)$$

where the first-order filter $k_{h_2}(i)$ is calculated by a series of decaying exponentials with time-constant $\tau_{h_2} \geq 0$, such that

$$k_{h_2}(i) = \sum_{j=1}^{i} \omega(j) \cdot e^{-\frac{(i-j)}{\tau_{h_2}}}. \qquad (eq.10)$$

These dynamics are intuitive and appear to be a more reasonable description of the change in fatigue response with accumulation of previous training compared to the standard FFM (Chiu & Barnes, 2003). It can be seen from (eq.9) that both fitness and fatigue components are represented by a sum of exponentially decaying training loads. For the fitness component each decaying training load is set by the individual training input $\omega(i)$. In contrast, it can be seen from (eq.10) that each decaying training load for fatigue is set by a weighted average comprising $\omega(i)$ and previous training sessions $(\omega(1), \ldots, \omega(i))$. If τ_{h_2} is set to a low value, there is limited interaction between training sessions and the value of weighted average is set primarily by $\omega(i)$. However, for larger values of τ_{h_2} there is greater interaction, and the value of the weighted average is influenced to a greater degree by previous training sessions.

To assess whether the VDR model could better describe the response to training compared with the standard FFM, Busso (2003) fit both models to six healthy males undergoing a prospectively designed training schedule comprising progressive overload followed by training cessation. Whilst the findings demonstrated the VDR model produced significantly improved fit ($R^2_{adj} = 0.931 - 0.958$)($P < 0.001$) compared to the standard FFM ($R^2_{adj} = 0.917 - 0.943$), the magnitude of the improvement was small. Importantly, Busso (2003) did not include out-of-sample testing and so an assessment of predictive validity is not provided. As with the Hill function, the additional parameters afford increased flexibility which could improve in-sample fit at the cost of less accurate out-of-sample due to overfitting. Busso (2003) also demonstrated that the VDR model created an inverted-U-shape relationship between constant daily training thereby partly addressing issues of linearity with the standard FFM.

2. 3 Modification to model specification and inclusion of uncertainty and feedback

The VDR model and standard FFM do not attempt to model uncertainty in the processes that generate fitness, fatigue and performance estimates. Additionally, estimates typically remain fixed irrespective of new incoming data. There have been some limited examples where estimates are updated by applying constraints that link for example successive sets of parameter estimates via a least squares algorithm (Busso et al., 2002). However, how FFMs could be expressed as linear state-space models thereby integrating uncertainty. A state-space representation is a mathematical model of a physical system as a set of input (e.g. training load), output (e.g. performance) and state variables (e.g. fitness and fatigue), where the unobserved state evolves over time depending on current values (e.g. in a recursive format as described in eq.4) and the system's input. The linear descriptor refers to the ability to specify the models in terms of matrices and linear algebra calculations. The recursive form of the standard FFM (eq.2 and eq.4) can be specified as a linear state-space model using the following matrix calculations.

$$\mathbf{x}_{n+1} = \mathbf{A}_n \mathbf{x}_n + \mathbf{B}_n \omega_n + \mathbf{v}_n, \qquad (eq. 11)$$

where $\mathbf{x}_{n+1}$ is a vector comprising fitness and fatigue on day $n+1$, $\mathbf{A}_n$ is a diagonal 2×2 "transition matrix" of coefficients that multiply the current fitness and fatigue values, $\mathbf{B}_n$ is a 2×1 matrix of coefficients that multiplies the scalar training input, and $\mathbf{v}_n$ is the state noise quantified by a 2×2 covariance matrix that is generally denoted by $\boldsymbol{Q}$. The state noise describes the random changes in state (e.g. fitness and fatigue) above and beyond the deterministic component involving training and past fitness or fatigue. Kolossa et al. (2017) noted that these random changes can be considered unmodeled exertions of the athlete and additional features not captured by the fitness and fatigue relationship with training load. To match the standard FFM the matrices in eq.11 are set to the following:

$$\mathbf{x}_n = \begin{pmatrix} g(n) \\ h(n) \end{pmatrix}, \mathbf{A}_n = \mathbf{A} = \begin{pmatrix} e^{-\frac{1}{\tau_g}} & 0 \\ 0 & e^{-\frac{1}{\tau_h}} \end{pmatrix}, \mathbf{B}_n = \mathbf{B} = \begin{pmatrix} e^{-\frac{1}{\tau_g}} \\ e^{-\frac{1}{\tau_h}} \end{pmatrix}, \mathrm{Var}(\mathbf{v}_n) = \boldsymbol{Q} = \begin{pmatrix} \sigma_g^2 & \sigma_{g,h} \\ \sigma_{g,h} & \sigma_h^2 \end{pmatrix}, \qquad (eq. 12)$$

where the state of the system is not observed directly but is accessible by means of indirect measurements of performance p_n from the equation

$$p_n = p^* + \mathbf{C}_n \mathbf{x}_n + \boldsymbol{\eta}_n, \qquad (eq. 13)$$

where $\mathbf{C}_n = \mathbf{C} = (k_g, -k_h)$ a 1×2 matrix and $\boldsymbol{\eta}_n$ is the observed performance noise described by actual measurement errors with variance denoted by ξ^2.

Expressing the FFM as a state-space model has the advantage that uncertainty in the state and measured performance can be modelled addressing the reality that all models are limited and contain uncertainty. Kolossa et al. (2017) also described how the Kalman filter could be combined with a state-space representation of an FFM to obtain better fitness and fatigue estimates with incoming data. The ability of the Kalman filter to operate on quantities exhibiting statistical noise over time and iteratively update may provide practitioners with an effective resource to optimise training prior to important events (an example of this process is provided in the supplementary resources). The goal of the Kalman filter is to generate an "a posteriori" state estimate $\hat{\mathbf{x}}_n$ based on "a priori" estimate $\mathbf{z}_n$ from the deterministic mechanics and the observed performance p_n. The extent to which the a priori estimate is updated depends on the relative extent of the uncertainty in the state to the uncertainty in the measurement (the ratio of $\boldsymbol{Q}$ to ξ^2). When the uncertainty in the state is large relative to the uncertainty in the measurement, the "Kalman gain" will be high and the filter will place more weight on the incoming performance data and relatively large corrections can be made in the a posteriori estimate. In contrast, when measurement uncertainty is large relative to uncertainty in the state, little weight will be placed on incoming data and there will be minimal correction to the initial a priori estimate. More formally, the stages of the Kalman filter are expressed in the following procedural flow.

1) Calculate the a priori state estimate $\mathbf{z}_n$

$$\mathbf{z}_n = \boldsymbol{A}\hat{\mathbf{x}}_{n-1} + \boldsymbol{B}\omega_{n-1}, \qquad (eq.14)$$

where $\hat{\mathbf{x}}_{n-1}$ is the previous a posteriori state estimate, or for the case where $n = 1$, $\hat{\mathbf{x}}_0$ is the initial state which may be set to $g(0) = h(0) = 0$ or estimated as unknown parameters. The second stage of the Kalman filter is

2) Calculate the Kalman gain $\mathbf{K}_n$ as a 2×1 matrix, defined as

$$\mathbf{K}_n = \boldsymbol{M}_n\boldsymbol{C}^{\boldsymbol{T}}(\xi^2 + \boldsymbol{C}\boldsymbol{M}_n\boldsymbol{C}^{\boldsymbol{T}})^{-1}, \qquad (eq.15)$$

where $\boldsymbol{M}_n$ is the covariance matrix of $\hat{\mathbf{x}}_n$ which is iteratively updated (eq.17). The Kalman gain scales the transformed matrix $\boldsymbol{M}_n\boldsymbol{C}^{\boldsymbol{T}}$ by the scalar value ($\xi^2 + \boldsymbol{C}\boldsymbol{M}_n\boldsymbol{C}^{\boldsymbol{T}}$), which describes the total variance in the state plus the variance in the measurement.

3) Calculate the a posteriori state estimate $\hat{\mathbf{x}}_n$ as

$$\hat{\mathbf{x}}_n = \mathbf{z}_n + \mathbf{K}_n(p_n - \mathbf{C}\mathbf{z}_n). \qquad (eq.16)$$

The third stage of the Kalman filter is the correction stage where the a priori estimate $\mathbf{z}_n$ is updated after observing the performance p_n. The Kalman gain $\mathbf{K}_n$ is expressed as a 2×1 matrix such that the correction to fitness and fatigue can be distinct and influenced by the state noise of each component and the error covariance specified. The Kalman gain for each component is multiplied by a scalar which is equal to the difference between the observed performance and the initial estimated performance ($\mathbf{C}\mathbf{z}_n$). Therefore, the larger the difference between the initial estimated performance and the observed performance the larger the correction. The fourth and final stage of the Kalman filter is

4) Update the estimation a posteriori error covariance matrix $\boldsymbol{M}_n$ with

$$\mathbf{M}_{n+1} = \boldsymbol{Q} + \boldsymbol{A}\boldsymbol{M}_n\boldsymbol{A}^{\boldsymbol{T}} - \boldsymbol{A}\boldsymbol{M}_n\boldsymbol{C}^{\boldsymbol{T}}(\xi^2 + \boldsymbol{C}\boldsymbol{M}_n\boldsymbol{C}^{\boldsymbol{T}})^{-1}\boldsymbol{C}\boldsymbol{M}_n\boldsymbol{A}^{\boldsymbol{T}}. \qquad (eq.17)$$

After initialisation, the updating of $\boldsymbol{M}_n$ governs how the Kalman gain evolves over time and the strength of the filtering effect of the model (for further details on fitting the Kalman filter in practice see appendix). In a non-stationary case like the FFM (due to non-stationary training driving the process), initialisation of $\boldsymbol{M}_0$ falls into three categories: 1) "known" initialisation, where values are set; 2) "approximate diffuse" initialisation, where $\boldsymbol{M}_0 = \kappa\boldsymbol{I}$ for large κ; 3) and "exact diffuse" initialisation which relies on limits as variances approach infinity (Fulton, C., & Opg 2017). The four stages of the Kalman filter can then be repeated with the estimated state and performance updated based on training input and corrections applied when performance is measured.

When using the Kalman filter, as is the case when fitting general FFMs, the model parameters must be estimated from training and performance data. This is achieved through algorithmically minimising some loss criterion, for example, the residual sum of squares between modelled and measured performance data, which, in the case of gaussian errors, coincides with the likelihood function (Mannakee et al 2016). With even the standard FFM, optimisation is analytically intractable and numerical procedures are used that require starting values for parameters $\{p^*, k_g, k_h, \tau_g, \tau_h\}$. While the available algorithms differ, they all are iterative in nature, stopping when the loss function falls below some prespecified threshold. When using the Kalman filter, the likelihood is available as a by-product of filtering operations (Fultcn, 2017), and thus the Kalman filter can be fit with the same optimisation routines as the standard FFM with additional starting values for the extra parameters. Given that an entire run of the filtering algorithm is required to obtain the likelihood of the sample, a "double loop" results when paired with a numerical optimisation procedure, and time to convergence may be slow for long series. Additionally, given the additional parameters that must be estimated in the Kalman filter model given the same setup as the standard FFM, some "sloppiness" in parameter estimation is likely. We found the full $\boldsymbol{Q}$ matrix difficult to recover, though the model's predictions were still quite good (see appendix).

Kolossa et al. (2017) outlined that more complex FFMs could be expressed as linear state-space models. The authors demonstrated that the VDR model could be easily expressed as a linear state space model and by including the Kalman filter could provide a means of addressing both the limitations of independence between training sessions and failure to take advantage of incoming data (Kolossa et al., 2017). To express the VDR model as a linear-state space model a simple change is required from eq.12 where the control matrix $\boldsymbol{B}$ is updated each iteration according to the following equation:

$$\mathbf{B}_{\boldsymbol{n}} = \begin{pmatrix} e^{-\frac{1}{\tau_g}} \\ k_{h_2}(i) \cdot e^{-\frac{1}{\tau_h}} \end{pmatrix}, \qquad (eq.18)$$

where $k_{h_2}(i)$ is given in eq.10. Further extensions combining FFMs and the Kalman filter are conceivable and include use of time-varying coefficients (p^*, k_g, k_h), cyclical variation, and external regressors (Commandeur et al., 2011).

2.4 Modification to model specification and inclusion of non-linearity

A common criticism of the standard FFM is that it is based on linear systems theory (i.e., control systems constructed of linear differential equations) that limits accuracy given the observed non-linearity of most human phenomena (Turner et al., 2017). Turner et al. (2017) presented a novel refinement of the standard FFM, introducing a generic mathematical framework to capture the non-linear effects of training (i.e., the problem of increased performance with arbitrary increases in training load and diminishing rates of return) enabling the search for optimum training programs in theory. The authors suggested that the standard FFM could be updated and specified as a system of non-linear differential equations for the fitness and fatigue components, as follows:

$$g'(t) = k_g \cdot \omega(t) - \frac{1}{\tau_g} g(t)^{\alpha} \qquad (eq.19)$$

$$h'(t) = k_h \cdot \omega(t) - \frac{1}{\tau_h} h(t)^{\beta}, \qquad (eq.20)$$

where α, β are power terms that represent the model's non-linearities, and where the original linear systems model can be recovered by simply setting $\alpha = \beta = 1$.

To explore the features of the non-linear model, Turner et al. (2017) first investigated the simplistic case of constant daily training ω. The authors demonstrated that under this simple analysis model a steady-state performance would be reached equal to

$$\hat{p}(t) = p^* + \left(k_g \tau_g \omega\right)^{\frac{1}{\alpha}} - (k_h \tau_h \omega)^{\frac{1}{\beta}}, \qquad (eq.21)$$

and that the constant training load that caused maximum steady-state performance was equal to

$$\omega_{max} = \left(\left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\beta}} \frac{\left(k_g \tau_g\right)^{\beta}}{\left(k_h \tau_h\right)^{\alpha}} \right)^{\frac{1}{\beta-\alpha}} . \qquad (eq. 22)$$

A limitation of the non-linear model was identified when investigating uniform weekly schedules where the daily training load could fluctuate but the same weekly schedule was repeated. Turner et al. (2017) demonstrated that repeated application of this weekly schedule created a periodic steady-state solution with the maximum obtained when the average training load was equal to that presented in eq.22. The analysis demonstrated under this constraint the equivalence of different patterns of training loads within a week, if the average training load was the same. Therefore, the model addressed the limitations of linearity but not independence of subsequent training sessions.

Turner et al. (2017) also investigated the predictive validity of the non-linear model on a data set collected from a single cyclist. Training load data were collected over 532 days with performance measured across 18 sessions throughout the period. The non-linear model comprised seven parameters $\{p^*, \tau_g, \tau_h, k_g, k_h, \alpha, \beta,\}$ and was fit to 9 of the performance measures with cross-validation performed on the remaining 9 performances. The cross-validation process was repeated many times for 9 randomly selected performances with parameters estimated each time with a genetic algorithm. Despite the small number of performance sessions and the large number of training days over which the model was fit, good predictive agreement was obtained. Additionally, the repetition of obtaining different parameters across the randomly selected performances was used to assess sensitivity in parameter estimation. Relatively tight variation was identified across all parameters with values generally ranging from 0.5 to 1.5 times the average estimated value. Turner et al. (2017) also used the variation in parameter sets to generate uncertainty in performance predictions and presented this graphically using a normalised 2-D histogram. In the accompanying code in the supplementary files we demonstrate how the nonlinear model can be fit to data and apply similar boot-strapping processes to generate multiple parameter estimates and plot predicted uncertainty to future training loads.

3.0 Conclusions

A central goal of applied sport and exercise science is to appropriately design training interventions with the aim of improving physical capabilities and ultimately sporting performance. The design, monitoring and systematic application of physical stimuli are, however, complex and there is a need to develop more appropriate frameworks to better understand the processes and achieve improved outcomes (Jeffries et al., 2020). Even in the most current frameworks the standard FFM developed by Banister and colleagues over 40 years ago plays a central role (Jeffries et al. 2020). Whilst this role is primarily a conceptual one, the mathematics of FFMs have the potential to contribute much more, especially given the primacy that is now placed on quantitative monitoring of athletes. Indeed, the overarching purpose of this review series has been to highlight the richness and extensive possibilities that exist to experiment with FFMs and support the physical development of athletes. In part I of this review series it was recommended that there was a shift in emphasis and FFMs be built and evaluated from training and measurements obtained during performance in simple physical tasks or exercise movements. From there, the scope of what can be attempted is extensive, from the design of innovative methods to quantify physical training, the development of high frequency performance data from training-based measurements, to the estimation of model parameters and associated uncertainty used to highlight the most likely outcomes rather than unrealistic point predictions (each discussed in part I). The scope can then be extended even further when as highlighted in part II of this review series, the underlying mathematics are developed creating more advanced models featuring for example interactions between training sessions, estimates of uncertainty and the ability to update predictions with incoming data, and facilities to account for likely non-linearities in the input to output relationships. The large data sets that practitioners are now routinely collecting provide an important source of information to build, experiment and evaluate models using a variety of cross-validation approaches (discussed in part I). The most promising approaches can then be assessed in fully prospective environments with continued dialogue and collaboration between researchers and practitioners used to enhance rigour and practical application. However, to facilitate this development, there is a need to overcome the initial barrier to entry which can be sizeable given the limited resources that exist to explain and enable experimentation with FFMs. It is hoped that this review series and the mix of resources providing both utilities to perform many of the operations with a user's own data, and didactic code to provide guidance encouraging users to be develop their own processes, can go some way to lowering the barrier to entry.

References

Banister, E. W., Calvert, T. W., Savage, M. V, & Bach, T. M. (1975). A Systems Model of Training for Athletic Performance. *Australian Journal of Sports Medicine*, *7*(3), 57–61.

Busso, T. (2003). Variable dose-response relationship between exercise training and performance. *Medicine and Science in Sports and Exercise*, *35*(7), 1188–1195. https://doi.org/10.1249/01.MSS.0000074465.13621.37

Busso, T., Benoit, H., Bonnefoy, R., Feasson, L., & Lacour, J. R. (2002). Effects of training frequency on the dynamics of performance response to a single training bout. *Journal of Applied Physiology*, *92*(2), 572–580. https://doi.org/10.1152/japplphysiol.00429.2001

Busso, T., Candau, R., & Lacour, J. R. (1994). Fatigue and fitness modelled from the effects of training on performance. *European Journal of Applied Physiology and Occupational Physiology*, *69*(1), 50–54. https://doi.org/10.1007/BF00867927

Busso, T., Carasso, C., & Lacour, J. R. (1991). Adequacy of a systems structure in the modeling of training effects on performance. *Journal of Applied Physiology*, *71*(5), 2044–2049.

Calvert, T. W., Banister, E. W., Savage, M. V, & Bach, T. M. (1976). A Systems Model of the Effects of Training on Physical Perfoffnance. *IEEE Transactions on Systems, Man, and Cybernetics*, *6*(2), 94–102. https://www.math.fsu.edu/~dgalvis/journalclub/papers/11_28_2016.pdf

Chiu, L. Z. F., & Barnes, J. L. (2003). The Fitness-Fatigue Model Revisited: Implications for Planning Short- and Long-Term Training. *Strength and Conditioning Journal*, *25*(6), 42–51. https://doi.org/10.1519/1533-4295(2003)025<0042

Cole, S. R., Chu, H., & Greenland, S. (2014). Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *American Journal of Epidemiology*, *179*(2), 252–260.

Commandeur, J. J. F., Koopman, S. J., Ooms, M., & others. (2011). Statistical software for state space methods. *Journal of Statistical Software*, *41*(1), 1–18.

Fultcn, C. (2017). *Estimating time series models by state space methods in Python: Statsmodels.* http://www.chadfulton.com/files/fulton_statsmodels_2017_v1.pdf

Hellard, P., Avalos, M., Lacoste, L., Barale, F., Chatard, J. C., & Millet, G. P. (2006). Assessing the limitations of the Banister model in monitoring training. *Journal of Sports Sciences*, *24*(5), 509–520. https://doi.org/10.1080/02640410500244697

Hellard, P., Avalos, M., Millet, G. P., Lacoste, L., Barale, F., & Chatard, J. C. (2005). Modeling the Residual Effects and Threshold Saturation of Training: A Case Study of Olympic Swimmers. *Journal of Strength and Conditioning Research*, *19*(1), 67–75.

Jeffries, A., Marcora, S., Coutts, A., Wallace, L., McCall, A., & Impellizzeri, F. (2020). Development of a revised conceptual framework of physical training for measurement validation and other applications. *Preprint*. https://doi.org/10.31236/osf.io/wpvek

Kolossa, D., Bin Azhar, M. A., Rasche, C., Endler, S., Hanakam, F., Ferrauti, A., & Pfeiffer, M. (2017). Performance estimation using the fitness-fatigue model with Kalman filter feedback. *International Journal of Computer Science in Sport*, *16*(2), 117–129. https://doi.org/10.1515/ijcss-2017-0010

Krzyzanski, W., Perez-Ruixo, J. J., & Vermeulen, A. (1999). Basic pharmacodynamic models for agents that alter the lifespan distribution of natural cells. *Journal of Pharmacokinetics and Biopharmaceutics*, *27*(5), 467–489. https://doi.org/10.1007/s10928-008-9092-6

Mannakee, B. K., Ragsdale, A. P., Transtrum, M. K., & Gutenkunst, R. N. (2016). Sloppiness and the geometry of parameter space. In *Uncertainty in Biology* (pp. 271–299). Springer.

Morton, R. H., Fitz-clarke, J. R., & Banister, E. W. (1990). Modeling Human Performance in Running. *The American Physiological Society: Modeling Methodology Forum*, *69*(3), 1171–1177.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (3.6.3). R Foundation for Statistical Computing. R-project.org

Rasche, C., & Pfeiffer, M. (2019). Training. In A. Baca & J. Perl (Eds.), *Modelling and Simulation in Sport and Exercise* (pp. 187–207). Routledge.

Scarf, P., Shrahili, M., Alotaibi, N., Jobson, S. A., & Passfield, L. (2019). Modelling the effect of training on performance in road cycling: estimation of the Banister model parameters using field data. *ArXiv Preprint*, 1–14.

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

Turner, J. D., Mazzoleni, M. J., Little, J. A., Sequeira, D., & Mann, B. P. (2017). A nonlinear model for the characterization and optimization of athletic training and performance. *Biomedical Human Kinetics*, *9*(1), 82–93. https://doi.org/10.1515/bhk-2017-0013

Appendix: Supplementary code resources

Project website: fitnessfatigue.com
Repository URL (direct): www.github.com/bsh2/fitness-fatigue-models

The fitness-fatigue model code 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, and with particular attention to part II of this review series, illustrative code in a didactic-style has been constructed that guides the user through the application of several advanced methods (e.g. Hill saturation, VDR model, Kalman feedback) via an interactive procedural flow. The former of these resources, the utility scripts, form the focus of part I in this review series, and the latter illustrative notebooks form the focus of this review. Prior to this project, there has not been a centralised repository of code suitable for the study of FFMs and related models.

The goal of this project is to expedite research and development of FFMs 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 (R Core Team, 2020), 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.

Code notebooks:

These code notebooks act as 'documentation' for the illustrative code files, and provide a run through of the materials provided and underlying methods.

Notebook	Link
Fitness-fatigue models: Standard, Hill Saturation, VDR model	https://www.kaggle.com/baogorek/fitness-fatigue-models-illustrative-code
Fitness-fatigue model with Kalman feedback	https://www.kaggle.com/baogorek/kalman-filter-illustrative-code
Non-linear variant of standard model system	https://www.kaggle.com/bsh2020/turnermodel

Fitting the standard model, hill function, VDR model, and Kalman filter

Maximum Likelihood Estimation

Maximum likelihood estimation is a general-purpose method for estimating unknown parameters in models. For the basic FFM and its variants, it coincides with nonlinear least squares under the assumption of independent and identically distributed Gaussian errors (Cole et al., 2014). Within the state-space formulation of the Kalman Filter, the likelihood is available as a by-product of filtering operations. Thus, maximum likelihood can be employed for all models in this review.

The general applicability of maximum likelihood does not guarantee easy estimation; there may be pathological curvature, local optima, and "sloppiness" in the geometry of the parameter space (Mannakee et al., 2016). We found the L-BFGS-G optimisation algorithm (offered in most statistical software) to be convenient in addressing these issues. The L-BFGS-G algorithm affords the use lower and upper search bounds, which allows incorporation of content knowledge to guide the search. It also removes the need for parameter transformations to ensure parameters do not leave their natural range (e.g., negative variance estimates); working on the original scale is simpler computationally and conceptually.

Here we outline the decisions made while using the L-BFGS-B algorithm in our "illustrative" code, including the starting values, the upper and lower limits, and the scaling and tolerance parameters. The starting values are chosen with the help of linear regression, whereby a course grid of time constants $\tau_g, \tau_h, \tau_{h2}$ (if VDR) as well as δ, γ values (Hill transformation) are needed to seed the procedure. While the grid can be course-grained, the procedure is fast enough to run through thousands of combinations in seconds each linear regression has a closed-form solution. The specific methodology for choosing starting values is outlined below:

- For time constants $\tau_g, \tau_h,$ and τ_{h2} (VDR), and hill transformation parameters δ, γ: the values from the course grid corresponding to the linear regression with the smallest RMSE.
- For p^*, k_g, k_h: the intercept and regression coefficients corresponding to the linear regression with the smallest RMSE.
- For σ: the smallest RMSE.

With the best time constants in hand, convolutional estimates of fitness and fatigue can be computed directly. These are useful for computing other starting values.

- For initial fitness and fatigue parameters q_g, q_h in the FFMs and the Kalman Filter state vector, a linear extrapolation to time 0 of the fitness and fatigue convolutions is used. We discarded the initial 10% of convolution values to remove the starting effects.
- For **M**, the unconditional variance of the initial state in the Kalman Filter, we use 1 / 2 the magnitudes of the initial state vector for the standard deviations. This carries the interpretation that 2 standard deviations is 100% of the initial state guess. We used 0 as the covariance.
- The initial state update of the Kalman Filter depends on an initial state value (fitness and fatigue) but also on a training load for "time 0." Rather than using 0 or creating yet another parameter, we choose to use the average training intensity in the data set.
- For σ_g, σ_h, we obtain rough state error predictions by $f(t) - \exp\left(-\frac{1}{\tau}\right) f(t-1)$, for both fitness and fatigue, and take the standard deviations of those.
- For ρ, the state correlation, we set it equal to 0.

For upper and lower constraints in L-BFGS-B:

- For p^*, the bounds chosen are 1 / 2 and 2 times the intercept of the best starting values regression.
- For k_g, k_h, the lower bound chosen is 0. The upper bound is set at 10 times the corresponding coefficient of the best starting values regression.
- For τ_g, τ_h, the lower bound is chosen to be 1 (unit of time) and the upper bound is chosen as the number of time points in the series (i.e., n units of time).
- For σ_g, σ_h in the state error covariance, we use a lower bound of 1E-6 to avoid covariance matrix singularity and 10 times the RMSE estimate of σ as an upper bound.
- For ρ in the state error covariance, the bounds chosen are -.999, .999, again to avoid covariance matrix singularity.

For parameter scaling, the parscale argument in R's optim routine allows optimisation to be performed on parameter / parscale (R Core Team, 2020). We set it to be the parameter upper bounds described above with good results. The tolerance parameter controlling the convergence, called "factr" (as in "factor of machine precision"), we chose 10, which is considerably smaller than the default of 1e7. The documentation for Python's SciPy states that typical values for factr are: 1E12 for low accuracy; 1E7 for moderate accuracy; 10.0 for extremely high accuracy. The maximum number of iterations is limited to 10,000.