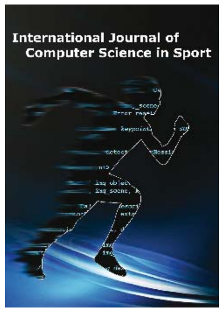
International Journal of Computer Science in Sport

Volume 18, Issue 1, 2019

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2019-0007



Including the Past: Performance Modeling Using a Preload Concept by Means of the Fitness-Fatigue Model

Melanie Ludwig[1], Alexander Asteroth[1], Christian Rasche[2] and Mark Pfeiffer[2]

[1]*Department of Computer Sciences, University o.a.S. Hochschule Bonn-Rhein-Sieg, St. Augustin, Germany;*

[2]*Department of Theory and Practical Performance in Sports, Johannes Gutenberg-University Mainz, Mainz, Germany*

Abstract

In mathematical modeling by means of performance models, the Fitness-Fatigue Model (FF-Model) is a common approach in sport and exercise science to study the training performance relationship. The FF-Model uses an initial basic level of performance and two antagonistic terms (for fitness and fatigue). By model calibration, parameters are adapted to the subject's individual physical response to training load. Although the simulation of the recorded training data in most cases shows useful results when the model is calibrated and all parameters are adjusted, this method has two major difficulties. First, a fitted value as basic performance will usually be too high. Second, without modification, the model cannot be simply used for prediction. By rewriting the FF-Model such that effects of former training history can be analyzed separately – we call those terms preload – it is possible to close the gap between a more realistic initial performance level and an athlete's actual performance level without distorting other model parameters and increase model accuracy substantially. Fitting error of the preload-extended FF-Model is less than 32% compared to the error of the FF-Model without preloads. Prediction error of the preload-extended FF-Model is around 54% of the error of the FF-Model without preloads.

KEYWORDS: FITNESS-FATIGUE MODEL, PERFORMANCE MODELING, PERFORMANCE PREDICTION, TRAINING PERFORMANCE RELATIONSHIP, MATHEMATICAL MODELING

Introduction

The purpose of tracking the progress of an athlete's training and performance has been a topic of great interest ever since. Banister, Calvert, Savage, and Bach (1975) presented a mathematical model that allows simulating impulse response of training stress, called Impulse Response Model or Fitness-Fatigue Model (FF-Model). Usage of mathematical models itself in sport science has become more relevant in the last years. Additionally, the increase in usage of wearables can improve and simplify analysis of several topics in the area of training (Ludwig, Hoffmann, Endler, Asteroth, & Wiemeyer, 2018; Passfield & Hopker, 2016).

Up to now, the FF-Model is still the most popular model in terms of simulating the relationship between training and performance (Kolossa et al., 2017; Passfield & Hopker, 2016) and many researchers presented promising and interesting extensions (see, e.g., Busso, 2003; Busso, Candau, & Lacour, 1994; Busso, Carasso, & Lacour, 1991; Busso, Denis, Bonnefoy, Geyssant & Lacour, 1997; Busso & Thomas, 2006; Hellard et al., 2006, 2005; Kolossa et al., 2017; Thomas, Mujika, & Busso, 2008; Turner, Mazzoleni, Little, Sequeira, & Mann, 2017, for more details). In the beginning of this century, Busso (2003) and Hellard et al. (2005) proposed nonlinear extensions for the classical FF-Model, which were further processed a few years later by Thomas et al. (2008) and Thomas, Mujika, and Busso (2009). In addition to the FF-Model, other models have been proposed, e.g., the PerPot-Metamodel (Perl, 2001; Perl & Pfeiffer, 2011), which has been proven to be a reasonable model for the task of simulating the training-performance relationship, too. Contrary to the FF-Model, which is static once the model parameters are optimized, the PerPot-Model takes internal potential levels during the simulation process into account, which results in a more dynamic nonlinear model behavior and allows for an incorporation of overtraining effects as well. In this paper, we will focus on the FF-Model despite its weaknesses, since it is widely discussed in the literature and well known to a broader readership. The preload concept which we present below can be applied to other models (like PerPot) and will be explained on the FF-Model as an example. Recently, Turner et al. (2017) stated again that extending the model based on the linear FF-Model might limit accuracy and applicability too much and therefore suggested a nonlinear development of this model, where the authors add exponential parameters to certain parts of the model. A comprehensive overview of performance modeling including the analysis of physiological assumptions motivating different types of models as well as various examples was recently published by Rasche and Pfeiffer (2018).

Figure 1 illustrates a typical modeling process applied to performance modeling. Therefore, input data is specified as training load, output as performance measures, and the model as (any) performance model. The process starts with a calibration of the model. In this part, a starting set of (free) model parameters are optimized within given parameter boundaries and constraints using both, input and output, to determine the goodness-of-fit of the model. After calibration, the model can be used to estimate the performance development in two different ways (simulation): On the one hand, if the same training load data is used as in calibration, simulation is only the simulation of the fitting itself to determine the accuracy of calibration. On the other hand, new or different training loads may be used for a performance prediction by using the optimized parameter set attained in the calibration phase. Of course it is possible to use a non calibrated, theoretical parameter set for any kind of simulation, too. We will not consider the latter in our analysis. Both kinds of simulation can be used to, e.g., compute values to describe model accuracy or to visually compare simulated and measured (empirical) output data.

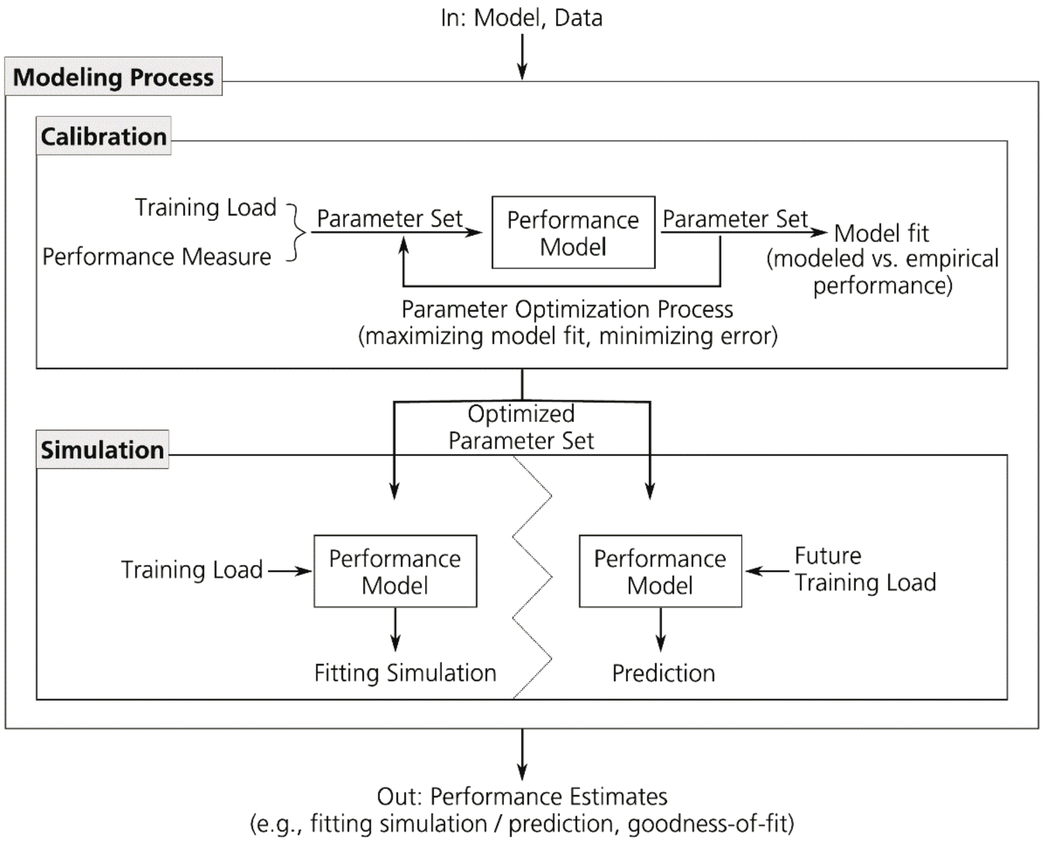


Figure 1. Methodical overview of the performance modeling process using sport and exercise science specific denominations used in this paper. The model input (training load) for calibration and fitting simulation is identical, while the training load used for performance prediction is different, either theoretically planned or a training load sequence following the afore used training load

Predicting an athlete’s training progress can enhance information about the athlete's reaction to stress. It can support a trainer in planning or might be even useful for ambitious hobby athletes. With a suitable performance prediction, the model can be enhanced into a useful tool for a even more appropriate long-term training planning and to reduce trial-and-error adjustments in exercise prescription (Clarke & Skiba, 2013). Nevertheless, most research focused on fitting simulation of performance progress, while prediction of performance is still very sparsely covered (Kolossa et al., 2017). While Kolossa et al. (2017) presented a promising method for predicting performance progresses online during training, only Chalencon et al. (2015) seem to use the original FF-Model for fitting simulation and prediction, but do not describe the model calibration process in detail. According to Taha and Thomas (2003) the published studies employing the FF-Model for performance prediction reveal major inaccuracies especially for small data sample sizes. An overview of different settings and accuracies is given in Figure 2: In performance modeling with (e.g.) the FF-Model, data of training and performance is needed to calibrate the model. In general, if all parameters of the FF-Model including the starting performance level (p^*) are optimized simultaneously (upper gray boxes), fitting simulation accuracy is mostly decent, but the model tends to overfit with unreasonable values of p^*, whereas predefining the value of p^* (lower part) reduces the model accuracy. We will explain this in ssubsection 2.1 and subsection 2.2 in more detail.

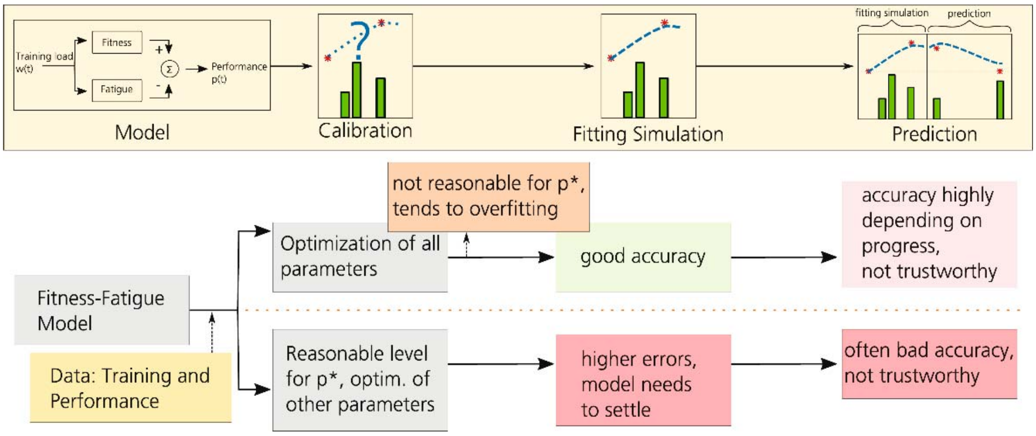


Figure 2. Overview of different settings and accuracy of results for FF-Model in terms of fitting simulation and prediction. The upper box gives an abstract overview of different states of modeling, while the lower part illustrates dependencies between the way of parameter optimization and corresponding outcomes in those different states of modeling

Therefore, we focus on the fitness and fatigue effects within performance models preceding the actual fitting simulation and prediction phases. Those phases strongly affect resulting performance estimation. We call the load leading to those former fitness and fatigue effects *preload*. We will show how preload terms can be extracted from the FF-Model without changing the model itself, and how preload can separately be analyzed. Ultimately, preload can be used to enhance fitting simulation and prediction accuracy in performance modeling.

Challenges of Performance Modeling

Depending on the point of view, modeling a person's performance is both, a problem of understanding dynamic processes in physiology as well as a curve fitting problem. Any model used for simulation needs to be calibrated (i.e., model parameters need to be *fitted*), and therefore, certain days of measurements are necessary. In terms of performance modeling, input data to a model usually is training load, while a performance measure is used as output.

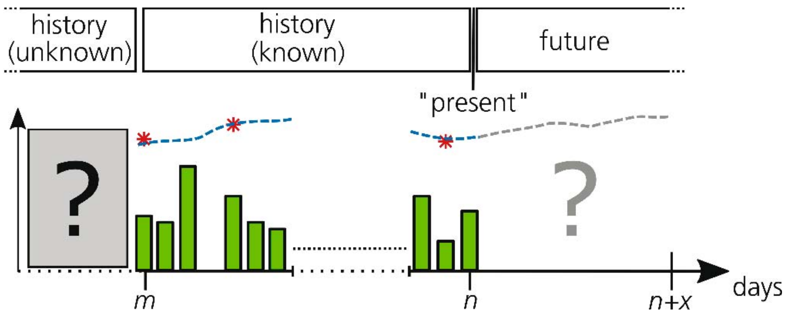


Figure 3. Independent on calibration or prediction, there is always an unknown history of training in the past

One of the big problems in modeling a person's performance progress is shown in Figure 3. Most of the time, only data from known history (m to n) is used in performance simulation. The starting point m is usually set to $m = 0$. The prediction (i.e., a forecast) starts after the last known day from training history (e.g., x days from $n + 1$ to $n + x$). While training data of known history is used for model calibration, every athlete has a training history which is often unknown up to a certain point (all days before day m), but can strongly affect the individual parameters of any model. Otherwise, the model would need to balance itself into some stable behavior, which might need many data points and of which the certain amount is not easy to figure out.

Technically, the starting point of each performance modeling process regarding the performance output needs to be determined as a basis of the modeled performance development. Furthermore, it is reasonable to assume this starting point as a baseline performance capability of a person, which can improve a model's accuracy if chosen properly. Notwithstanding the fact that we cannot measure a real baseline performance, we can assume that there is some very basic performance level for each individual, even if this level might change depending on, e.g., age and health condition. Additionally, it is often a helpful way in modeling to include a baseline, though using an unsuitable basic performance level as baseline can highly influence the whole modeling. We will explain the problem with a wrongly chosen or even fitted baseline performance with the *FF-Model* exemplary.

The Fitness-Fatigue model

The FF-Model by Banister et al. (1975) is still one of the most important and fundamental models (Rasche & Pfeiffer, 2018). It is used to estimate an athletes performance based on individual fitness values and a body's strain as reaction to stress which is called fatigue. Within this section, we will shortly outline the structure of this model and explain limitations regarding performance prediction with this model exemplary in subsection 2.2.

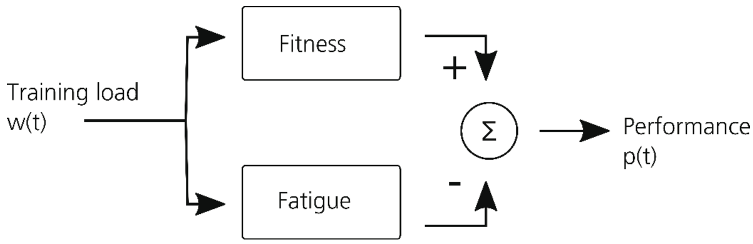


Figure 4. General concept of the FF-Model as first presented in Banister et al. (1975)

As shown in Figure 4, the basic idea of the FF-Model is that performance is made up of two contracting principles. On the one hand, training results in an improved performance, but on the other hand it induces fatigue which diminishes performance. In 1991, Busso et al. analyzed different versions of the FF-Model with differing amount of components. They concluded, that a model version "composed of two antagonistic first-order transfer functions" (Busso et al., 1991, p. 2048) sufficiently models the training response.

So Busso et al. (1994) derived a feasible two-component version:

$$\hat{p}(n) = p^* + \underbrace{k_1 \cdot \sum_{t=1}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}}}_{fitness} - \underbrace{k_2 \cdot \sum_{t=1}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}}}_{fatigue} \quad (1)$$

where $\hat{p}(n)$ describes performance at day n and p^* is the initial basic performance level. The input w (e.g., velocity or wattage) is considered for the past $n-1$ days of training. Furthermore τ_1 and τ_2 are time constants, and k_1 and k_2 are multiplicative scaling factors.

Limitations of the Fitness-Fatigue Model for Prediction

General limitations of the FF-Model such as parameter stability, parameter interpretability and ill-conditioning, and model accuracy especially for future performance prediction have already been discussed by, e.g., Chiu and Barnes (2003); Hellard et al. (2006); Pfeiffer (2008); Taha and Thomas (2003). While Hellard et al. (2006) criticized the parameter stability, they assumed the small sample size and inter-dependent parameters to be reasons for this instability among other reasons.

Taha and Thomas (2003) criticized the bad prediction accuracy of the model, if parameters are taken from calibration on former data points. In our approach, we will focus on handling the basic performance level, which also affects stability of parameters in general. In the FF-Model, past training effects are completely covered by an additive constant p^*, which is meant to be the baseline performance level of a person. While parameters k_1, k_2, τ_1, τ_2 can be optimized during model calibration, this is not a reasonable way to find the initial performance level p^*. In literature, the process of finding p^* is sparsely covered. We found three different methods published: First, p^* is fitted as any of the other four parameters (Chalencon et al., 2015; Clarke & Skiba, 2013). Second, p^* is set to 80% of the value of performance at the beginning of the experiment, since this would "correspond to the subjects' performance after a few months of detraining" (Busso et al., 1997, p. 1688) as suggested by Busso et al. (1997). Third, p^* can be estimated from prior training history data if possible or a performance level of the subject can be chosen if relatively detrained as suggested by Wood, Hayter, Rowbottom, and Stewart (2005). While the latter might be the most preferable option, it is also most difficult to realize. Independently on the method chosen for finding a suitable value of p^*, there are two possibilities: First, value of p^* is set near the actual performance, which often suits the remaining model calibration best. In this case, p^* will be too high if the person was not relatively detrained, resulting in a high *lower bound*. Due to prolonged absence of training load, performance will always converge against this lower bound over time. A lower bound set too high connotes that performance will stay that high even without training.

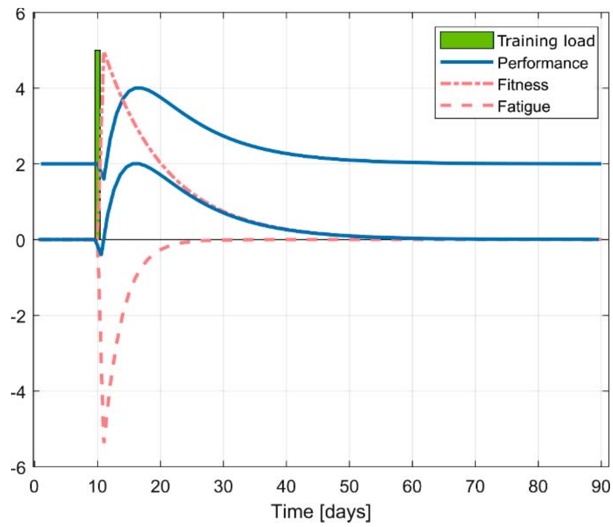


Figure 5. Convergence to the basic performance level $\boldsymbol{p^*}$ with two different example performance curves; parameters are set to $\boldsymbol{k_1 = 1.1, k_2 = 1.5, \tau_1 = 10, \tau_2 = 3}$ and $\boldsymbol{p^* = 0}$ (lower performance curve) resp. $\boldsymbol{p^* = 2}$ (upper performance curve). If there is no new training load, modeled performance will converge to the level of $\boldsymbol{p^*}$

Second, value of p^* is set to a more realistic basic performance level. This will lead to an inaccurate performance simulation for any person who is not accordingly detrained and whose actual fitness level is above her or his basic level. Both cases are illustrated in Figure 5 with two different exemplary values of p^*. When using the FF-Model to simulate or predict performance, the process always starts with this initial performance parameter p^*, despite the previous performed training. Any information about possible performance progress is missing. Furthermore, any choice of p^* affects the model calibration in terms of remaining parameters.

The Preload Concept

In performance modeling, simulation of any performance value strongly depends on stress of past training sessions and is based on formerly afforded performances. Therefore, it is reasonable to store information about past training effects within the model.

Since any training effect fades away over time, influence of the stored information should also disappear over time.

The basic idea of an explicitly used preload in the FF-Model to enhance performance prediction accuracy was first published by Ludwig and Asteroth (2016). The great advantage of formalizing and analyzing the preload separately is the possibility to even fit a preload value without knowing past training loads.

Since there is no clear "start" for training effects of fitness and fatigue, we assume an infinite period of time. Accordingly, we can define a preload in the FF-Model as follows.

Definition 1 (Preload) *The preload terms for fitness and fatigue are defined as:*

$$pr^{+}_{[-\infty,n]} = k_1 \cdot e^{\frac{-x}{\tau_1}} \cdot \sum_{t=-\infty}^{n} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} \tag{2}$$

$$and\ analogously\ pr^{-}_{[-\infty,n]} = k_2 \cdot e^{\frac{-x}{\tau_2}} \cdot \sum_{t=-\infty}^{n} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}}$$

with $x \in \mathbb{N}$ depending on the specific day simulated or predicted; x is set to the number of the day simulated minus the number of the first day taken into account for simulation (i.e., the day after computation of the preload terms ended) plus one.

These preload-terms are added to (resp. substracted from) p^* in the preload FF-Model. The exponential function multiplicated with the sum ($e^{\frac{-x}{\tau_i}}$) is used to model the vanishing effect of the preload with respect to the currently modeled day n. For example, if the last day included in preload is denoted with $m-1$, the simulation starts at day m and is performed for some day ℓ ($m \leq \ell \in \mathbb{N}$), then $x = \ell - m + 1$.

Definition 2 (FF-Model with Preload) *We can interpret the preload as additional component for the initial performance level p^*, which closes the gap between the initial performance level and the actual performance level reached by recent training stress. Performance on day n within an observed training time frame starting at day m is calculated as*

$$\hat{p}(n) = \left(p^* + pr^{+}_{[-\infty,m-1]} - pr^{-}_{[-\infty,m-1]}\right) + k_1 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} - k_2 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}} \tag{3}$$

Basically, Equation 3 is a mathematical re-formulation of a FF-Model including the complete history of a person. Please note that the actual fitting simulation or prediction of the FF-Model with preload starts at day m, and not at day 0 or 1 as usual. Additionally, we can define a more feasible version of the preload terms restricted to a certain time frame. This is especially usable to estimate the preload present in a certain training period and for most applications more realistic, since the effect long bygone training load disappears over time:

Definition 3 (Short Term Preload) *We define a short term preload for the fitness and fatigue component up to day n as*

$$pr_n^+ = k_1 \cdot e^{\frac{-x}{\tau_1}} \cdot \sum_{t=1}^{n} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} \tag{4}$$

$$pr_n^- = k_2 \cdot e^{\frac{-x}{\tau_2}} \cdot \sum_{t=1}^{n} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}} \tag{5}$$

with pr^+ the fitness component and pr^- the fatigue component. Parameters τ_i, k_i with $i \in \{1,2\}$ originate from the FF-Model as before. Again, $x \in \mathbb{N}$ is depending on the specific day simulated or predicted.

The FF-Model with short term preload is specified analogously to Equation 3. The (short term) preload concept should be deployed for performance prediction by using the given information of the calibration data. This can be performed either by including the training load used for calibration in prediction as well (as history only, i.e., without integrating this part in any statistical analysis of prediction accuracy), or by saving fitness and fatigue effects in a *short term preload*, for which the time frame is limited to the length of data used for calibration. This short term preload can always be used, if effects of a specific period of time are to be determined. Technically, performing a prediction including the (calculated/computed) short term preload terms instead of the data used for calibration (which is based on the calibration data in terms of a calculated preload) is equivalent to a simulation of the complete time of both, calibration phase and prediction phase. This relationship is illustrated in Example 7.1 (Appendix).

For illustration, the scenery of Equation 3 is shown in Figure 6**(a)**, while Figure 6**(b)** shows the scenario where the short term preload is necessary.

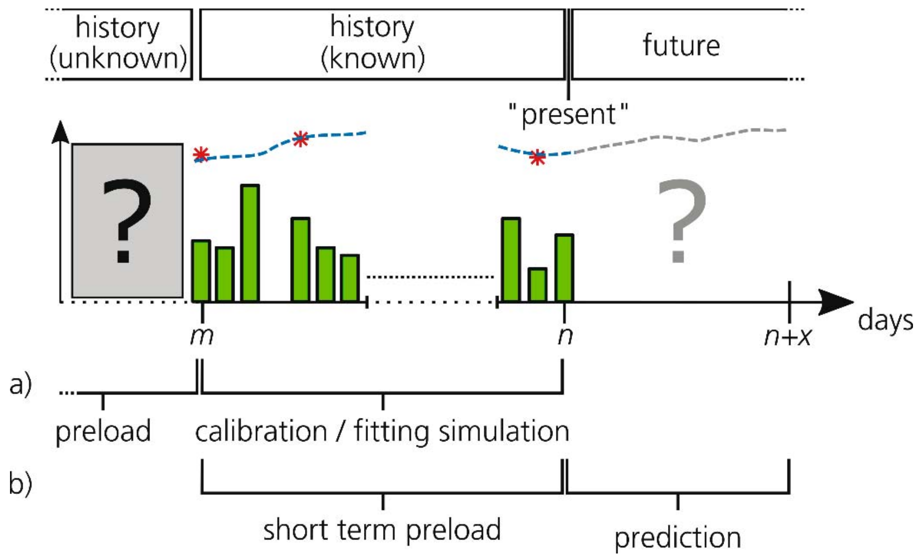


Figure 6. Scenarios for (a) preloads and (b) short term preloads. In both cases, preloads are used to memorize past training effects. In case of a unknown training history, preloads can also be estimated during the parameter optimization process. In case of the short term preload: $\boldsymbol{m = 0}$. Bars: training load at the corresponding day; stars: performance measures; line: simulated performance estimation

Materials and Methods

Participants and Data

All experiments are based on the performance and training load data of four male cyclists of the German national squad for road cycling (Table 1), collected for 31 (P1) resp. 35 weeks (P2, P3, P4).

The season was divided into four different training periods: A preparation period of seven (P1) resp. nine weeks for basic endurance training (including strength endurance training), a second preparation period of seven weeks for basic endurance training (including maximum power training), eight weeks preparatory competitions including intensity training, and seven (P1) resp. nine weeks with competitions. Days with no cycling training at the end of the last period were excluded.

Cycling performance was quantified as the relative accumulated energy exposure (kJ/kg) during a weekly all-out incremental step test on a cycling ergometer. Starting resistance was set to 150 W followed by increments of 10 W every 10 seconds. To calculate the daily training load, training time spent cycling was divided into five levels based on the intensity (individual wattage) according to the *Bavarian cycling federation* and accumulated at last by using weighting coefficients: 1 × active recovery, 2 × endurance, 4 × tempo, 7 × lactate threshold and 10 × VO_2max / anaerobic threshold. The additional power training is not included in this analysis.

Table 1. Characteristics of the cyclists who participated in this study

	P1	P2	P3	P4
Age (at start of the study)	17	17	18	18
Height [cm]	174	180	182	179
Weight [kg]	61.2 ± 0.6	68.3 ± 0.9	66.1 ± 0.6	60.8 ± 0.6
Training load per year [km]	ca. 18'000	ca. 17'000	ca. 18'000	ca. 18'000

General Study Design

The goal of our study is to investigate whether the additional preload terms (pr^+, pr^-) can enhance the accuracy of individual performance developments modeled by the preload-extended FF-Model. To minimize the effects of varying training patterns and competition phases, which represent specific adaptation and regeneration phases, we cut a varying amount of days (0, 15, 30, 45 or 60) off at the end of each data set before continuing with the analysis. In doing so, the stability of the new model can be validated and the number of data sets is increased. The resulting data sets were analyzed concisely according to the following partitioning.

The analysis in this paper focuses on short-term preloads and, therefore, the data set of each cyclist is divided into three different parts: the "unknown" training history, which would usually not be available (part A, 105 days or more), calibration data to fit the model for each individual (part B, 60 days) and a prediction period (part C, 30 days). Since fitting simulations as well as predictions are typically evaluated to assess a model's accuracy, this paper analyzes both: model-fits for the fitting simulation of part B and the prediction of part C.

Hence, we need to specify basic performance levels (p^*) as well as the utilized preload calculation methods on whose basis we calibrate the model using data of part B. According to the goal of our study, both may be set variously and are partially dependent on each other, which is explained in the following paragraph. Additionally, a comprehensive overview of the different parts of analysis and their purpose as well as calibration settings can be found in Figure 7.

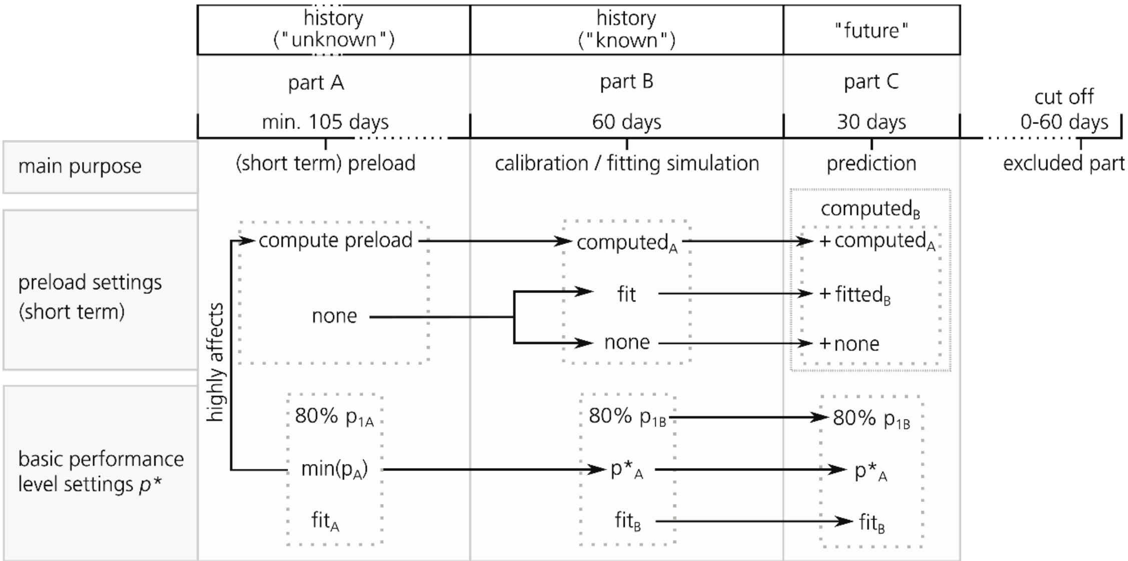


Figure 7. Partitioning of training data before analysis and different settings for short term preload and basic performance level

Short term preload (pr $^+$, pr $^-$)

To assess the athlete-specific preload, part A is first used to calibrate the model's parameters followed by a fitting simulation to compute the accumulated fitness and fatigue states at the end of part A. These states are the preloads and are dependent on the choice of a basic performance level p^* (see next paragraph) and labeled as *computed preload* values for further modeling. We contrast these to *fitted preloads* stemming from the calibration process of part B (optimized parameters: τ_1, τ_2, k_1, k_2 plus pr $^+$ and pr $^-$) on the one hand and to no preload (*none*) on the other hand representing the classic FF-Model.

Basic performance level ($\boldsymbol{p}^*$)

Generally, the basic performance level needs to be defined mainly prior to calibration and fitting simulation of part B, but also for the calculation of the preload based on part A. Regarding the former, we analyzed three options for the choice of p^*: First, p^* is set as 80 % of the part B's first performance measurement ($0.8 \cdot p_{1B}$) or, second, p^* is fitted besides the other parameters of the FF-Model for part B (optimized parameters: p^*, τ_1, τ_2, k_1, k_2 and pr $^+$/pr $^-$ if applicable). Third, p^* is set to the same value which is used to calculate the preload based on part A referred to as p_A^*, which may be either 80% of the first performance measurement ($0.8 \cdot p_{1A}$), the minimal performance, or a fitted value based on part A's data.

Direct comparisons with different choices of preload and different choices of p^* are not analyzed since the goal is to analyze the effects of preloads.

Statistical Measures

Mean absolute percentage error (MAPE) is computed to compare simulated and empirical (measured) performance values of different methods used, given as:

$$MAPE = \frac{100}{T} \sum_{t=1}^{T} \left| \frac{y(t) - \tilde{y}(t)}{y(t)} \right| \qquad (6)$$

with T being the total number of data points, $y(t)$ the measured performance value at time t and $\tilde{y}(t)$ the simulated performance value at time t. All MAPE values are expressed as median values, p.r.n. supplemented with interquartile range (IQR) to describe variances.

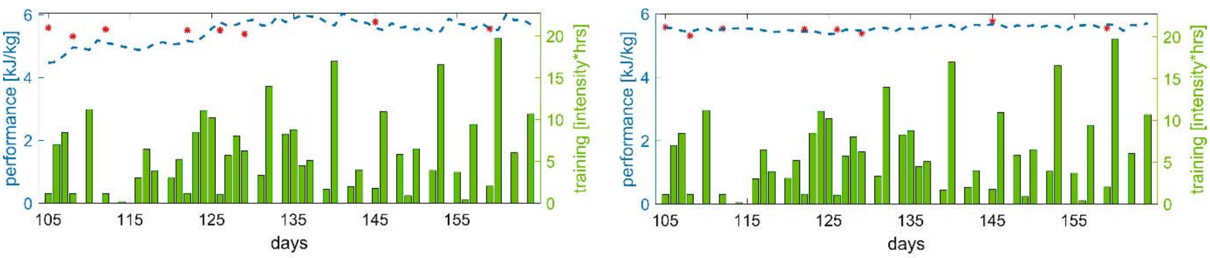
The null hypotheses that the underlying data is not normally distributed could not be discarded with level of significance at $\alpha = 0.05$ (Jarque-Bera test). Since we are dealing with small data sets in terms of statistics, differences in the error results between different experiments were tested with Wilcoxon rank-sum test at $\alpha = 0.05$. The amount of data is sufficient for this test, because it could be increased by the partitioning performed. With Wilcoxon rank-sum test, the null hypothesis indicates that results of the experiments to be compared are based on data with the same median. Therefore, the compared setups differ significantly from each other with level of significance at $\alpha = 0.05$, if the null hypothesis can be rejected.

Results

No significant differences dependent on the cuts were determined with Wilcoxon rank-sum test at $\alpha = 0.05$. Therefore, different cutoffs of the data sets are not differentiated but used all when taking the median.

Fitting Simulation Accuracy

An example of fitting accuracy is given in Figure 8. Subfigure **(a)** shows a typical fitting simulation, where $p^* = 0.8 \cdot p_{1A}$ and no preload is used. The simulation starts at a lower level and it takes multiple days (in this case around 20) for the FF-Model to reach the right performance level. The simulation is close to several empirical performances after day 125, but there is no trend corresponding to the underlying performance values detectable. Subfigure Figure 8**(b)** shows a typical fitting simulation, where again $p^* = 0.8 \cdot p_{1A}$ and a fitted preload is used. The simulated performance starts at a reasonable high. In this case, neither simulation nor measured performance fluctuate much.



(a) No preload, $p^* = 0.8 \cdot p_{1A}$ (b) Fitted preload, $p^* = 0.8 \cdot p_{1A}$

Figure 8. Fitting simulation examples (a) without and (b) with preload (athlete 3). Bars: training load at the corresponding day; stars: measures performance values; line: simulated performance estimation

Accuracy of fitting simulation results is given in Figure 9. The boxplot is shown exemplary for the case where p^* (for computation of preload) is set to 80% of the first measured performance value in the data of part A denoted as $p^* = 0.8\ p_{1A}$. Highest errors were produced in experiments without preload and where p^* is not fitted (i.e., both outer subboxes).

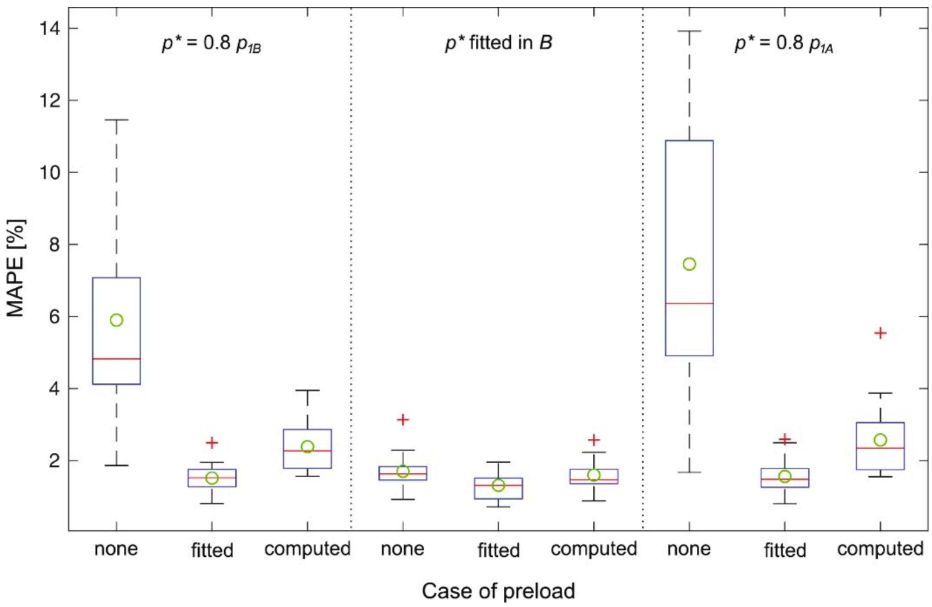


Figure 9. Boxplots of MAPE values in fitting simulation for different experiment setups. Median values are marked by a red horizontal line; mean values are marked by a green circle; outliers are marked with a red cross where existing. Different settings for $\boldsymbol{p}^*$ are separated by dashed lines and marked above

Adding a preload reduces the error values for setting $p^* = 0.8\ p_{1B}$ from a median MAPE of 4.82% to 1.52% in case of fitted preloads, respectively to 2.27% for computed preloads. If $p^* = 0.8\ p_{1A}$, error values reduce from a median MAPE of 6.36% to 1.48%, respectively to 2.35%. If p^* is fitted during calibration of part B, errors seem to have similar ranges with median values of 1.63% (no preload), 1.31% (fitted preloads), and 1.47% (computed preloads).

Table 2: p-values of Wilcoxon rank-sum test for all fitting simulation analysis cases. Significant results indicate a difference in the compared (preload) settings; Non-significant results are bracketed

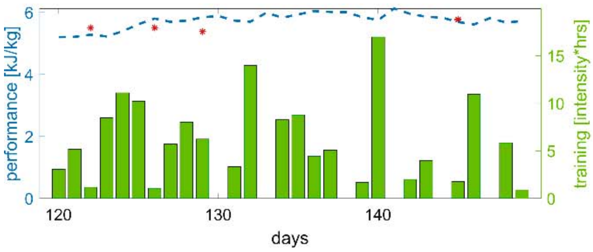
		options for $\boldsymbol{p_A^*}$		
$\boldsymbol{p^*}$	**preloads**	**80%**	**min**	**fitted**
$0.8 \cdot p_{1B}$	none vs. fitted	< 0.0001	< 0.0001	< 0.0001
	none vs. computed	< 0.0001	< 0.0001	< 0.0001
	fitted vs. computed	0.0002	0.0001	0.0003
fitted in B	none vs. fitted	0.0237	0.0237	0.0237
	none vs. computed	(0.3346)	(0.5798)	(0.2482)
	fitted vs. computed	(0.0847)	0.0354	(0.1329)
set to p_A^*	none vs. fitted	< 0.0001	< 0.0001	< 0.0001
	none vs. computed	< 0.0001	0.002	0.0445
	fitted vs. computed	0.0004	0.013	(0.1032)

The other cases got comparable results; detailed results of Wilcoxon rank-sum test for all cases with different values for p^* and different preloads is given in Table 2: If p^* was fitted during calibration of part B, there were no significant differences between the accuracy results of the experiments without preloads and those with computed preloads. Those with fitted preload reached a slightly better accuracy than those with computed preload. Those differences were significant in four out of nine settings. If p^* is set to $80\% \cdot p_{1B}$, comparing experiments without preloads to those with preloads (computed or fitted) results in significant rank-sum test for all nine settings, similar to the same comparison if p^* is set to the value previously determined in part A, where only one case got non-significant results.

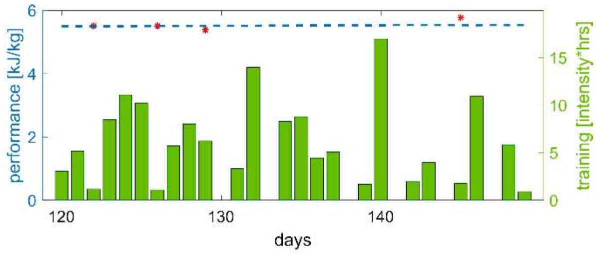
All in all, the null hypotheses that results from experiments without preloads and those with fitted preloads have equal median values can be rejected for all cases; for the comparison of experiments without preloads to those with computed preloads, the equivalent null hypotheses can be rejected for no case if p^* is fitted in B, but for every other setting. It can clearly be seen that experiments without preload and with a p^* value which was not fitted results in higher errors than a corresponding experimental setup with preload.

Prediction Accuracy

A qualitative example of prediction is given in Figure 10. Subfigure **(a)** shows a typical prediction, where $p^* = 0.8 \cdot p_{1A}$ and no preload is used. Based on the result of the fitting, where performance simulation fluctuated much as in Figure 10, prediction starts at a more or less suitable performance level but fluctuates around empirical measurements. But even regarding the general trend, predicted performance does not behave like the empirical performance. Subfigure **(b)** shows a typical prediction, where again $p^* = 0.8 \cdot p_{1A}$ and a fitted preload is used. The prediction has hardly any variances and seems to follow more or less a mean performance line of the measured performance values. The usage of preload in the former fitting process ensured that predicted performances starts at a suitable performance level.



(a) No preload, $p^* = 0.8 \cdot p_{1A}$



(b) Fitted preload, $p^* = 0.8 \cdot p_{1A}$

Figure 10. Prediction examples (athelte 3). Bars: training load at the corresponding day; stars: measures performance values; line: simulated performance estimation

Accuracy of prediction results is given in Figure 11. The boxplot is shown exemplary for the same case as for fitting simulation, i.e., where p^* is set to $80\% \cdot p_{1A}$.

Again, slightly higher errors were produced in experiments without preload and where p^* is not fitted, and where both are fitted, p^* and preload. Best accuracy results were produced where p^* is fitted and no preload is used with a median MAPE of 2.55%, and where p^* is fitted and the pre-computed preload is used with a median MAPE of 2.69%. While for the other cases of p^*, experiments including preload results in slightly better accuracy than those without preload, a clear trend cannot be determined.

The null hypotheses that results from experiments without preload and those with fitted preload have equal median values cannot be rejected for any case.

The other cases got comparable results, but none were significant with Wilcoxon rank-sum test at $\alpha = 0.05$. It has to be mentioned that the amount of data points for evaluation in the prediction case consists of only 1-6 performance values for evaluation of prediction accuracy.

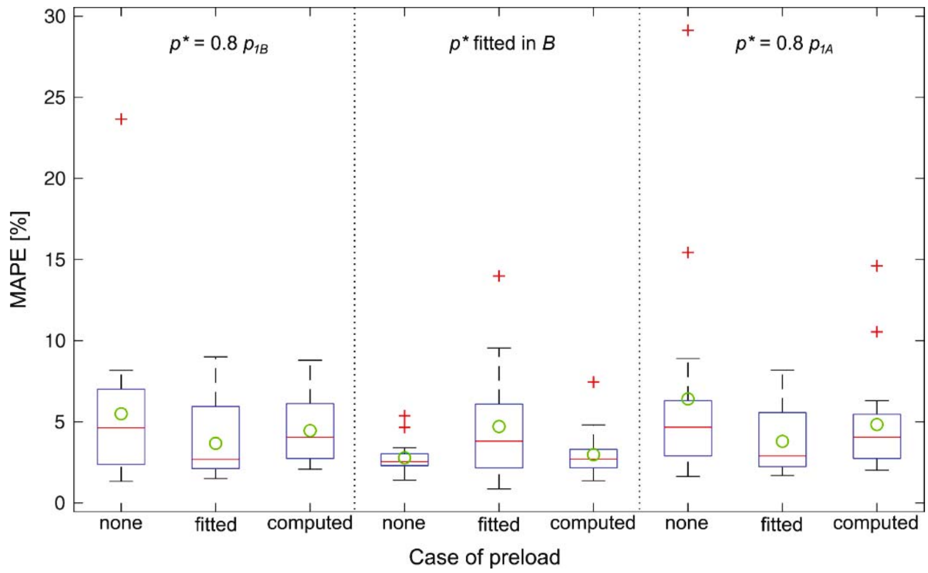


Figure 11. Boxplots of MAPE values in prediction for different experiment setups. Median values are marked by a red horizontal line; mean values are marked by a green circle; outliers are marked with a red cross where existing. Different settings for $\boldsymbol{p}^*$ are separated by dashed lines and marked above

Discussion

In the preceding sections we introduced a way to re-formulate the FF-Model. The new representation allows us to analyze effects of a former, usually unknown training history, which we called *preloads*. Furthermore, adding calculated preloads to the FF-Model or fitting the preloads during model calibration closes the gap between a low initial performance level and the actual performance level of an athlete. For four athletes with training data of half a year including different types of training, it was shown that the additional preloads significantly improve fitting simulation accuracies (in median) independent on the chosen initial performance value. In particular, the high errors which can occur by using a non-fitted, low initial performance value can be completely compensated by adding preloads to the FF-Model. In prediction, accuracies can not be improved significantly and seem to be comparable to results without preloads.

An first overview of the findings which will be explained in the following is given in Figure 12.

Without the additional preloads, optimizing all parameters (including p^*) resulted in unreasonable values of p^*, tending to overfitting simulation results, with which no trustworthy predictions were possible (upper path way). Predefining a reasonable value of p^* resulted in high errors (lower path way). The inclusion of preloads added more information about past training progresses, even if there is no documentation which can be used.

Both middle pathways are illustrations to highlight the gain in accuracy when using the preload concept: independent on the calibration of p^*, accuracy results in lower errors compared to the corresponding setting without preload in both kinds of simulation. Therefore, the preload enables the possibility to use some reasonable values for the basic performance level (p^*) of an athlete, without loosing accuracy in modeling results.

A more detailed discussion regarding parameter optimization and performance accuracy is given in the following.

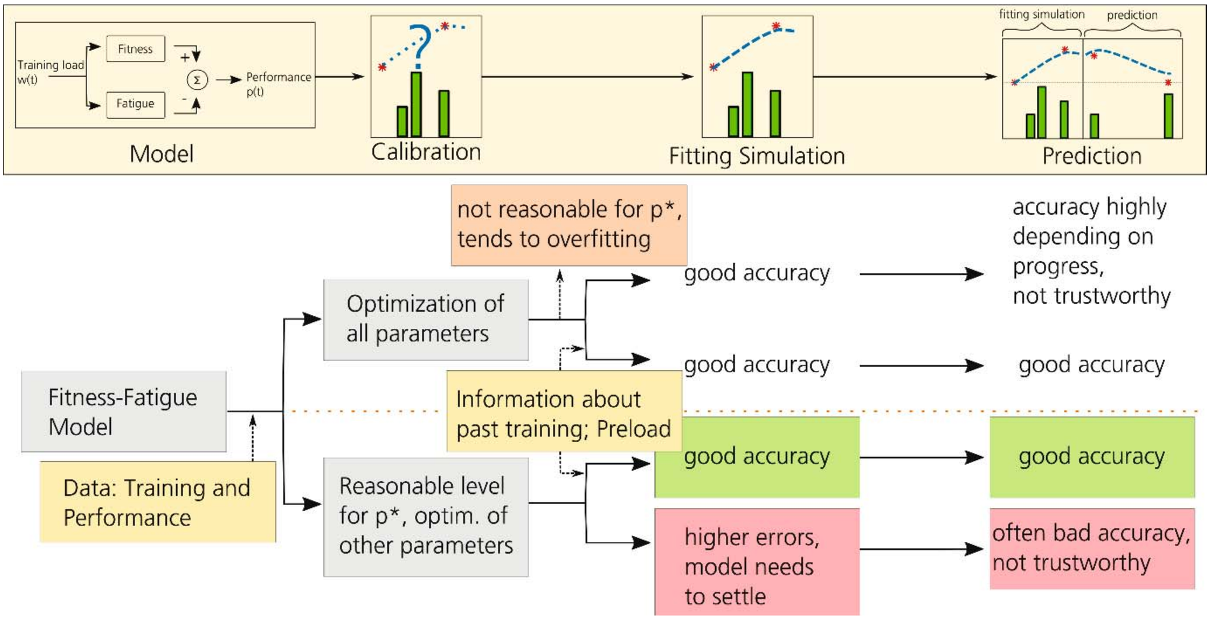


Figure 12. Overview of different settings and accuracy of results for FF-Model

Initial performance value and model parameters

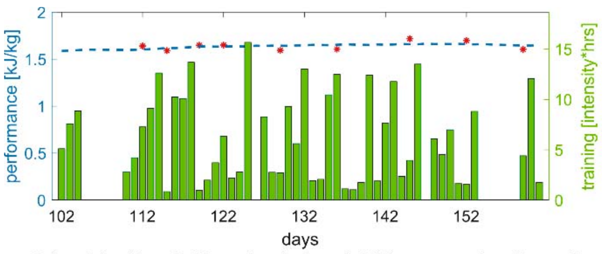
Calibration of the initial performance parameter value p^* might lead to seemingly suitable results, especially in elite athletes or athletes without loss in performance. Lowering the initial performance value therefore does not work out directly. Our results of this setting had smallest errors, too, whether preloads were included or not. A performance value of p^* clearly lower than the first *measured* performance value forces the model to settle first. The time needed for the model to settle to the right level of performance will result in either higher errors, or – at least in calibration – to different parameters of k_i and $\tau_i, i \in \{1,2\}$.

This finding might suggest that fitting p^* might be a good idea. But as we mentioned, p^* is a lower bound of performance in the FF-Model and any performance will converge to p^* if no further training is performed. We think that parameters fitted in combination with a wrong value for p^* cannot be interpreted. Errors can accumulate over time. Hence, even a prediction will have no physiological meaning: based on a wrong starting point and with distorted parameters, a prediction might be just a meaningless curve based on some inputs.

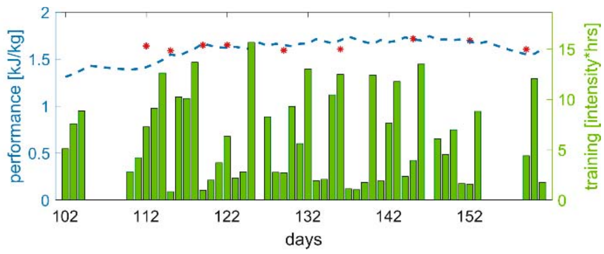
A qualitative example of those cases for fitting simulation of performance data of one cyclist (after cutting the data by 30 days) is shown in Figure 13. Part **(a)** illustrates the simulated performance while p^* is fitted like the other parameters, while part **(b)** illustrates the same for p^* set to 80% of the first performance value. In part (c) a pre-computed preload is added to the setting from part **(b)**.

Mean absolute percentage errors are 2.20% for case **(a)**, 3.65% for case **(b)**, and 1.89% for case **(c)**. These three cases give a first impression that a preload can compensate for a more reasonable initial performance value without losing accuracy.

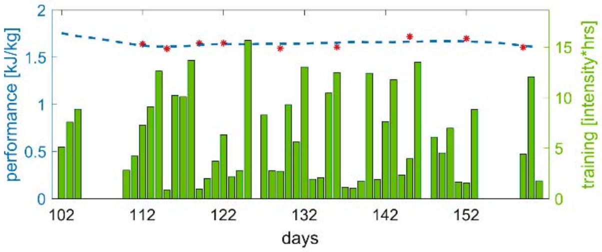
Consequently, since changes in the basic performance level p^* affect all other parameters, those changes strongly affect calculation of the preload. This has to be considered in interpretation of the results, where fitting simulation and prediction accuracies of the model using a calculated preload are included.



(a) $p*$ is fitted directly ($p*$ = 1.59), no preload used. MAPE = 2.20%



(b) $p*$ is set to 80% of the first performance value of part B ($p*$ = 1.31), no preload used. MAPE = 3.65%



(c) $p*$ is set to 80% of the first performance value of part B ($p*$ = 1.31), preload was computed beforehand. MAPE = 1.89%

Figure 13. Comparison of different settings of p^* and preload exemplary for one cyclist (cutoff of 30 days exemplary, athlete 1). Bars: training load at the corresponding day; stars: measures performance values; line: simulated performance estimation

Unfortunately, the true "initial / basic performance level" of a person is not known. Up to now, there are no tests or methods known to determine this value. While Wood et al. (2005) suggested to estimate p^* from former training sessions, this task might still be challenging and a concrete procedure is still missing. Additionally, estimate p^* if the athlete is relatively detrained as suggested might not be always possible depending on training goals, time constraints, or other reasons. Therefore we think that the approach to set p^* to about 80% of the first measured performance value as suggested by Busso et al. (1997) is the most suitable by now. According to them, this value corresponds approximately to a detraining of multiple weeks. Yet, it should not be forgotten that the first measured performance value might be on a day where fatigue decreased the performance capabilities of the athlete, or reverse that the athlete has a peak form day.

A problem with this approach for p^* is the resulting gap between this value p^* and the measured performances. The remaining model parameters can again be strongly distorted and hence not be well interpreted. We saw that preload can close this gap, even if not estimated but fitted. It has yet to be analyzed if parameters are more meaningful according to the athlete's physiology or not.

Prediction Accuracy

In our experiments, prediction accuracies did not profit from any additional preloads used during calibration. One reason might be, that prediction accuracies could only be evaluated on 1-6 data points.

Additionally, model calibration over 60 days seem to be sufficient enough for the model to settle independent on the choices for preloads and p^*. With given parameter sets and information about the average training load and performance of an athlete, it is possible to estimate the amount of days before the effects of the preload vanish by solving an optimization problem. With parameter settings of our athletes, preload effects in median (and inter quartile range) vanishes completely after 30 (31) days (athlete 1), 30 (52) days (athlete 2), 37.5 (104) days (athlete 3), or 25.5 (74) days (athlete 4).

The high inter quartile ranges indicate that in preload effects were still present in some cases and completely vanished in others. Especially for athletes 1 and 2, all preload effects might have vanished after calibration for most experimental settings. Due to a small amount of performance tests, we were not able to perform our experiments with smaller calibration periods.

In general, any evaluation based on that few data points is quite insufficient. It might give a first indication, but cannot be accepted to be a fact.

Conclusion

Fitting the initial performance value p^* might work in some cases, especially if the athlete is detrained at the beginning or has a stable performance without high variation. Nevertheless, it is methodically problematic, since p^* works as lower bound for performance if no new training input is given. Furthermore, any choice of p^* has influence on the remaining parameters which cannot have any reliable meaning if they are distorted by a wrong p^*.

On the other hand, a lower value of p^* leads to other artifacts. The model has to settle and the remaining parameters k_i, τ_i will compensate the offset and are accordingly distorted. Using preloads for effects in both, fitness and fatigue, closes the gap between a lower value of p^* and the actual performance level of the athlete. This way, model calibration with a fitting of parameters k_i and τ_i does not result in high amplitudes for compensation of the performance high. Nevertheless, determining a suitable choice of p^* based on physiological reasons and dependent on the type of sport is an important topic for future investigations of sport scientists.

Usually, a preload cannot be computed, since information about further training history is not available. Otherwise, the model could have been fitted on those data. Our experimental results demonstrate that fitted preload values in average lead to as good accuracies as computed preload values, or reach even an even better accuracy. In fitting simulation, a low value of p^* can be compensated by preloads without disturbing the other parameters. In prediction, preloads enhance accuracies independent on the chosen p^* as well, but errors of 3-5.3% of performance in median are still quite high and might therefore not be helpful in elite sports. Nevertheless and supported by the effect that the simulation with preloads do not need to be balanced out during calibration, even usage of fitted preload values within the FF-Model might help amateur athletes to get at least an idea of their future performance progress, where errors around 5% are acceptable.

Evaluations on training data from elite swimmers validate these findings and suggest that this concept is not only valid for training data from elite cyclists. In future work, it will be necessary to evaluate the model with preloads for amateur athletes, too, and on data sets with more performance measurements, especially for the prediction part. Furthermore, stability and interpretation of parameters in the FF-Model with and without preload will be interesting to compare. The model's behavior on more specific performance changes due to, e.g., competitions, resting periods, injuries or similar would also be interesting to evaluate in this context.

Since the preload does not reverse other weaknesses of a model, the FF-Model with preload will still not be able to, e.g., predict overtraining effects. From a methodological standpoint, it should be avoided to fit the initial performance value. The preloads then enable a reliable calibration and give the opportunity for further analyzes of unknown training history effects.

Funding

We are very grateful for financial support given by the Ministry of Culture and Science (MKW) of the North Rhine-Westphalia state within the program FH-STRUKTUR 2017 (AZ: 322-8.03.04.02-FH-STRUKTUR 2017/07), and by the Equal Opportunity Commission at Bonn-Rhein-Sieg University o.a.S.

References

Banister, E., Calvert, T., Savage, M., & Bach, T. (1975). A systems model of training for athletic performance. *Aust J Sports Med*, 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.

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.

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), 2044–9.

Busso, T., Denis, C., Bonnefoy, R., Geyssant, A., & Lacour, J.-R. (1997). Modeling of adaptations to physical training by using a recursive least squares algorithm. *Journal of applied physiology*, 82(5), 1685–1693.

Busso, T., & Thomas, L. (2006). Using mathematical modeling in training planning. *International journal of sports physiology and performance*, 1(4), 400–405.

Chalencon, S., Pichot, V., Roche, F., Lacour, J.-R., Garet, M., Connes, P., . . . Busso, T. (2015). Modeling of performance and ans activity for predicting future responses to training. *European journal of applied physiology*, 115(3), 589–596.

Chiu, L. Z., & Barnes, J. L. (2003). The fitness-fatigue model revisited: Implications for planning short-and long-term training. *Strength & Conditioning Journal*, 25(6), 42–51.

Clarke, D. C., & Skiba, P. F. (2013). Rationale and resources for teaching the mathematical modeling of athletic training and performance. *Advances in physiology education*, 37(2), 134–152.

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(05), 509–520.

Hellard, P., Avalos, M., Millet, G., 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.

Kolossa, D., Bin Azhar, M., 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.

Ludwig, M., & Asteroth, A. (2016). Predicting performance from outdoor cycling training with the fitness-fatigue model. In A. Artiga Gonzalez, R. Bertschinger, T. Dahmen, M. Gratkowski, S. Wolf, & D. Saupe (Eds.), *Proceedings dvs-workshop modelling in* *endurance sports,* University of Konstanz, Sept. 9–11, 2016 (pp. 3–6). KOPS – The Institutional Repository of the University of Konstanz.

Ludwig, M., Hoffmann, K., Endler, S., Asteroth, A., & Wiemeyer, J. (2018). Measurement, prediction, and control of individual heart rate responses to exercise - basics and options for wearable devices. *Front. Physiol.*, 9(778).

Passfield, L., & Hopker, J. G. (2016). A mine of information: can sports analytics provide wisdom from your data? *International journal of sports physiology and performance*, 1–17.

Perl, J. (2001). Perpot: A metamodel for simulation of load performance interaction. *European Journal of Sport Science*, 1(2), 1–13.

Perl, J., & Pfeiffer, M. (2011). Perpot domo: Antagonistic meta-model processing two concurrent load flows. *International Journal of Computer Science in Sport (International Association of Computer Science in Sport)*, 10(2).

Pfeiffer, M. (2008). Modeling the relationship between training and performance-a comparison of two antagonistic concepts. *International journal of computer science in sport*, 7(2), 13–32.

Rasche, C., & Pfeiffer, M. (2018). Training. *In A. Baca & J. Perl (Eds.), Modelling and simulation in sport and exercise* (1st ed., p. 187—207). Routledge.

Taha, T., & Thomas, S. G. (2003). Systems modelling of the relationship between training and performance. *Sports Medicine*, 33(14), 1061–1073.

Thomas, L., Mujika, I., & Busso, T. (2008). A model study of optimal training reduction during pre-event taper in elite swimmers. *Journal of sports sciences*, 26(6), 643–652.

Thomas, L., Mujika, I., & Busso, T. (2009). Computer simulations assessing the potential performance benefit of a final increase in training during pre-event taper. *The Journal of Strength & Conditioning Research*, 23(6), 1729–1736.

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.

Wood, R. E., Hayter, S., Rowbottom, D., & Stewart, I. (2005). Applying a mathematical model to training adaptation in a distance runner. *European journal of applied physiology*, 94(3), 310–316.

Appendix

Example 7.1 *Assume a prediction of day* n *based on days* m *to* $n-1$ *(*$m<n$*). Calibration was based on days* 1 *to* $m-1$ *and accordingly, preload terms are the appropriate fitness and fatigue sums of the calibration phase. Predicting day* n *with the FF-Model including short term preloads is given as:*

$$
\begin{aligned}
\hat{p}(n) &= (p^* + \overbrace{pr^{+}_{m-1} - pr^{-}_{m-1}}^{Preloads}) + \overbrace{k_1 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}}}^{Fitness} - \overbrace{k_2 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}}}^{Fatigue} \\
&= (p^* + k_1 \cdot e^{\frac{-(n-m+1)}{\tau_1}} \cdot \sum_{t=1}^{m-1} w(t) \cdot e^{\frac{-(m-1-t)}{\tau_1}} - k_2 \cdot e^{\frac{-(n-m+1)}{\tau_2}} \cdot \sum_{t=1}^{m-1} w(t) \cdot e^{\frac{-(m-1-t)}{\tau_2}}) \\
&\quad + k_1 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} - k_2 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}} \\
&= (p^* + k_1 \cdot \sum_{t=1}^{m-1} w(t) \cdot e^{\frac{-(m-1-t)}{\tau_1}} \cdot e^{\frac{-(n-m+1)}{\tau_1}} - k_2 \cdot \sum_{t=1}^{m-1} w(t) \cdot e^{\frac{-(m-1-t)}{\tau_2}} \cdot e^{\frac{-(n-m+1)}{\tau_2}}) \\
&\quad + k_1 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} - k_2 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}} \\
&= (p^* + k_1 \cdot \sum_{t=1}^{m-1} w(t) \cdot e^{\frac{-(m-1-t)-(n-m+1)}{\tau_1}} - k_2 \cdot \sum_{t=1}^{m-1} w(t) \cdot e^{\frac{-(m-1-t)-(n-m+1)}{\tau_2}}) \\
&\quad + k_1 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} - k_2 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}} \\
&= (p^* + k_1 \cdot \sum_{t=1}^{m-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} - k_2 \cdot \sum_{t=1}^{m-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}}) \\
&\quad + k_1 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} - k_2 \cdot \sum_{t=m}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}} \\
&= p^* + k_1 \cdot \sum_{t=1}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_1}} - k_2 \cdot \sum_{t=1}^{n-1} w(t) \cdot e^{\frac{-(n-t)}{\tau_2}}
\end{aligned}
\tag{7}
$$

This way it can simply be seen that using the short term preload starting at day m is equivalent to a whole simulation starting at day 1.