The Fitness–Fatigue Model: What’s in the Numbers?

Kobe Vermeire,1 Michael Ghijs,2 Jan G. Bourgois,1,3 and Jan Boone1,3

1Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium; 2BIOMATH, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium; 3Center of Sports Medicine, Ghent University Hospital, Ghent, Belgium

Purpose: The purpose of this commentary is to outline some of the pitfalls when using the fitness–fatigue model to unravel the interaction between training load and performance. By doing so, we encourage sport scientists and coaches to interpret the parameters from the model with some extra caution. Conclusions: Caution is needed when interpreting the fitness–fatigue model since the parameter values are influenced by the starting parameter values, the modeling technique, and the input of the model. Also, the use of general constants should be avoided since they do not account for interindividual differences and differences between training-load methods. Therefore, we advise sport scientists and coaches to use the model as a way to work more data-informed rather than working data-driven.

Keywords: impulse-response model, performance modeling, periodization, training data, training load

Quantification of the relationship between training load (TL) and changes in performance can be considered the Holy Grail in sport science, as this would allow coaches to select the most adequate training stimulus to maximize performance of an athlete at a given moment in time. Therefore, already in the 70s, the impulse-response model, more commonly known as the fitness–fatigue model, was developed.1 In this fitness–fatigue model, performance is considered the resultant of both a fitness and a fatigue component that are both induced to a different extent from a training session. Although the model has been further developed throughout the decades, as such obtaining several versions, the most used mathematical form is

\[ \hat{p}_n = p^* + k_1 \sum_{i=1}^{n-1} w_i e^{-(n-i)/\tau_1} - k_2 \sum_{i=1}^{n-1} w_i e^{-(n-i)/\tau_2}. \]

The modeled performance at day n (\( \hat{p}_n \)) is estimated from successive TL’s (\( w_i \)) with i varying from 1 to \( n-1 \). \( p^* \) is an additive term that represents the initial performance level of the subject, \( \tau_1 \) and \( \tau_2 \) are the exponential time constants, expressed in days, for respectively the fitness and the fatigue term, and magnitude factors have been added to both fitness (\( k_1 \)) and fatigue (\( k_2 \)). In other words, the \( \tau \) parameter represents the number of days the positive and the negative effects of a training session are apparent in the body, while the \( k \) parameter represents the magnitude, or the size, of that effect. In general, the value of \( \tau \) is higher for the fitness term than for the fatigue term, meaning that fitness will last longer than fatigue. The opposite is true for the \( k \) parameter, here the value will be larger for the fatigue term than for the fitness term, indicating that fatigue is larger than fitness immediately after a training session. Together, this leads to the fact that fatigue will be more apparent directly after training sessions, but that this fatigue will dissipate faster than fitness. After a certain period, fitness will surpass the fatigue, and performance improvement will be the result. Moreover, the \( k \) parameter also works as a conversion factor. This makes it possible to convert arbitrary TL units to the required unit of the performance measure (often a power metric or finish time). Over time, adaptations were made to the model, resulting in different mathematical forms that incorporate 1, 2, or sometimes even 3 \( k \) parameters. This means that the exact value and meaning of the \( k \) parameter will depend on the model used. However, it is not the goal of this commentary to outline all models, for an overview of the models, we refer to the manuscript of Clarke and Skiba.2

Use of the Model

The fitness–fatigue model has been used in order to relate TL to changes in performance in different sports such as swimming,6–8 cycling,6–8 running,9,10 and triathlon.11 In these studies, a good to excellent model fit has often been found, suggesting that the model is able to relate TL with performance in an adequate manner. Based on these results, the model and derivatives of the model were integrated in several coaching platforms, making it accessible to a large group of athletes and coaches. This integration was made to support coaches in assessing and predicting the performance of their athletes without having to do extensive testing. It has also enabled coaches to make the right selection of TL in the planning and periodization phase, and to understand how the training program, with the corresponding TL, influences performance over time. However, in a lot of platforms, a simplification of the original model is used where the \( k \) parameter is removed, which then leads to a more simple (exponentially) weighted average. This means that the conversion to a real performance measure is no longer possible, and arbitrary units are used. This also implies that conclusions based on one model, or the other are not per se directly transferable to each other.

However, mostly average values for the \( \tau \) and \( k \) parameters are used to establish these models since an individual model fit is hard to obtain in a real-world setting. For example, the \( \tau \) parameter is often estimated to be 42 days for fitness and 7 days for fatigue. It is known, however, that an individual model fit is necessary when trying to relate TL and performance.12 In literature, values are found ranging from 4 to 51, and from 4 to 74 for \( \tau_1 \) and \( \tau_2 \), respectively.2,4,12–15 Although some coaching platforms enable coaches to adjust these values to the user’s preference, it is futile to estimate what parameter values will best suit the individual athlete without fitting the model.

To fit the model, TL is needed as an input. However, many different methods to quantify TL exist nowadays with the external

Boone (Jan.Boone@UGent.be) is corresponding author, https://orcid.org/0000-0002-8485-6169
TL methods that quantify the work completed (power, distance, speed, etc) and internal TL methods, which represent the psychophysiological response to training (perceived exertion, heart rate, lactate, etc). This leads to different TL scores for the same training sessions, depending on the method used. Moreover, studies have shown that different input variables (ie, TL calculations) can lead to substantially different results in the parameter values. The study of Mitchell et al and the study of Vermeire et al have both shown that the input of different TL quantification methods will lead to different \( \tau \) values on the same data set. The parameter values of these studies differ considerably from the original papers, which gives rise to question the validity of the commonly used parameter values. Therefore, we want to point out 3 possible pitfalls when using performance models: (1) the interpretation of the model parameters, (2) the selection of the modeling technique, and (3) the input of the model.

**Interpretation of the Model Parameters**

When the model was originally constructed, it was stated that \( \tau_1 \) and \( \tau_2 \) are time constants which represent the number of days that the fitness and the fatigue, generated by a training session, will be apparent in the body. However, interpretation of this parameter is more complex and should be done with caution. To demonstrate this, a result is shown from a global optimization (see below) of the model to data of one of the subjects in an earlier study in Figure 1. As this optimization scanned a series of values for each combination of parameters in the fitness–fatigue model (a range of 0–3 for \( k \) and 0–60 for \( \tau \)), a minimum absolute model error (residual sum of squares) could be plotted for each parameter value combination in Figure 1. From this figure, we can deduct that different subsets of parameter values can lead to similar model errors, which makes it hard to select the right parameter values. As previously shown by Hellard et al, \( k_1 \) and \( k_2 \) but also \( \tau_1 \) and \( \tau_2 \) are highly correlated. This ill-conditioning of the model could therefore explain the plethora of possible parameter values that give similar model errors.

In other words, this means that we cannot interpret the \( \tau \) parameter as being a specific number of days, and that the \( k \) parameter does not solely serve as a magnitude and conversion factor, but rather that the fitness–fatigue model should be treated as one entity. Several authors have also previously indicated that the practical interpretation of the parameters might be difficult. For example, a positive correlation between the fatigue function and testosterone concentration was found, where a negative relation was expected. The model parameters were also criticized based on the fact that most of the models were poorly supported by physiological mechanisms.

**Selection of the Modeling Technique**

The model deficiencies as explained above are noticeable since global optimization techniques were used to fit the model. However, global optimization techniques are technically advanced; and therefore, coaches in the field do not always have the possibility, or the means, to infer such a model fit. Nevertheless, there are more accessible ways to fit a model by means of local optimization. As
proposed by Clarke and Skiba,\textsuperscript{2} one can use an Excel spreadsheet and fit the model to the individual by means of the Solver function in Microsoft Excel. The pitfall here is that local optimization values will depend greatly on the starting values. In Table 1, the locally optimized parameter values are presented, using the Solver function (generalized reduced gradient [GRG] nonlinear) in Excel (MS Excel, version 2111) without constraints, for 4 different sets of starting parameter values, taken from literature,\textsuperscript{2,14,15} using the same training and performance data. The weakness of using local optimization techniques is reflected in this table. The 4 sets of starting values lead to 3 different local optima (sets 2 and 4 are similar), which are determined to a great extent by the starting values.

In short, local optimization techniques search for the smallest model error in a smaller parameter space. The optimization will run as long as the model error keeps getting smaller. As soon as the model error starts rising again, the optimization will stop running at this local optimum (Figure 2). Initiating the optimization from different parameter values could lead to a smaller error and thus the starting set will determine the result. Global optimization techniques work around this problem in different ways, resulting in a true minimum error and thus more correct parameter values. Therefore, we advise, as Clarke and Skiba\textsuperscript{2} already suggested, when choosing for a local optimization method, different sets of starting parameter values should be used so to minimize the chance of missing the true optimum.

### Input of the Model

A last point to consider when interpreting the model parameters is the input (TL) of the model. Different methods to quantify the TL exist nowadays, either based on internal measures (eg, heart rate, rate of perceived exertion, lactate), or external measures (eg, power, speed). At first glance, it seems that all these methods result in reasonably small model errors, suggesting it does not matter which TL method is used. However, research has shown that different TL quantification methods lead to different $k$ and $\tau$ values.\textsuperscript{12,13} In an ideal situation, the $\tau$ values would not be different, depending on the input, and only the $k$ values would be influenced since this is the factor that makes it possible to convert the input to the required output. Again, this implies that the absolute parameter values are to be interpreted with caution. Often in practice, different methods are used to quantify TL due to data registration restrictions or data capture failure. However, these data clearly show that fitting the fitness–fatigue model requires that only one quantification method is used throughout the entire period.

Also, since most of the existing TL methods are a combination of volume and intensity, the specificity of training is not represented in this score. For instance, a training session of 2 hours at low intensity would result in a TL score of 120 (using Lucia TL),\textsuperscript{18} but a high intensity bout of 40 minutes can result in an identical score. Since the training adaptations performing such different training sessions with a similar TL are totally different,\textsuperscript{19,20} the relationship with performance improvement will always be distorted.

### Table 1 Local Optimization Values for Different Sets of Starting Values

<table>
<thead>
<tr>
<th>Set no.</th>
<th>Starting values</th>
<th>Local optima</th>
<th>Model error (RSS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>$k_1$ 1</td>
<td>0.6407</td>
<td>378.47</td>
</tr>
<tr>
<td></td>
<td>$k_2$ 2</td>
<td>0.7146</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_1$ 42</td>
<td>10.5313</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_2$ 7</td>
<td>8.0583</td>
<td></td>
</tr>
<tr>
<td>Set 2</td>
<td>$k_1$ 0.18</td>
<td>0.0649</td>
<td>323.06</td>
</tr>
<tr>
<td></td>
<td>$k_2$ 0.23</td>
<td>0.1057</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_1$ 36</td>
<td>24.1714</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_2$ 21</td>
<td>3.5542</td>
<td></td>
</tr>
<tr>
<td>Set 3</td>
<td>$k_1$ 0.0048</td>
<td>0.037352</td>
<td>367.26</td>
</tr>
<tr>
<td></td>
<td>$k_2$ 0.3860</td>
<td>0.055929</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_1$ 49</td>
<td>39.94221</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_2$ 4.3</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Set 4</td>
<td>$k_1$ 0.0193</td>
<td>0.065958</td>
<td>323.05</td>
</tr>
<tr>
<td></td>
<td>$k_2$ 0.0148</td>
<td>0.106808</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_1$ 40.8</td>
<td>23.89291</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_2$ 9.0</td>
<td>3.619263</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: RSS, residual sum of squared errors. Note: The different starting values are selected from literature.

### Figure 2 — Representation of local optima and a global optimum in a minimization problem. The circles represent local optima, while the “x” represents the global optimum.
improvements in performance, which may in turn lead to insights for individual periodization strategies. However, since the model parameters cannot be interpreted separately and since they will differ strongly between athletes and over different training periods,14 the conclusions taken from these models can only be used to support coaches. Nonetheless, if the pitfalls are accounted for, individualization of the parameters could aid in individualizing training programs.21

**Practical Applications**

For both coaches and sport scientists, we give some short recommendations when using the fitness–fatigue model:

- Use only one TL metric to fit the fitness–fatigue model. However, apart from the model, different metrics should be collected and compared to each other. When the external internal TL ratio dissociates, this could indicate training (mal-)adaptations.22

- When using local optimization techniques, use different sets of starting parameter values to infer the smallest error.

- Only compare parameter values intrasubject, and track changes over different training periods.

- Always interpret the parameter values using physiological knowledge and training experience.

**Conclusion**

Caution is needed when interpreting the fitness–fatigue model since the parameter values are influenced by the starting parameter values, the modeling technique, and the input of the model. The use of general parameter values should be avoided since they do not account for interindividually different differences and differences between TL methods. Therefore, the authors advise coaches and sport scientists to use the model as a way to work more data-informed rather than working data-driven.

**References**


