The suitability of a quasi-Newton algorithm for estimating fitness-fatigue models: sensitivity, troublesome local optima and implications for future research (an in silico experimental design).

STEPHENS HEMINGWAY, B., SWINTON, P. and OGOREK, B.

2021

The suitability of a quasi-Newton algorithm for estimating fitness-fatigue models: Sensitivity, troublesome local optima, and implications for future research (An *in silico* experimental design)

Ben Stephens Hemingway [1,2], Paul Swinton [1], Ben Ogorek [3]

PREPRINT – NOT PEER REVIEWED

Corresponding author:

Dr Paul Swinton
School of Health Sciences
Robert Gordon University
Aberdeen, UK, AB10 7QE

p.swinton@rgu.ac.uk

Institutions:

[1] School of Health Sciences, Robert Gordon University, Aberdeen, UK
[2] School of Computing, Robert Gordon University, Aberdeen, UK
[3] Nousot, Chicago, Illinois, USA

Conflicts of interest: BSH, PS, BO declare that they have no conflicts of interest relevant to the content of this review

Author agreement statement: All authors are aware and agree that this work can be submitted to SportRχiv for publication in preprint

Preferred citation:

Stephens Hemingway, B., Swinton, P., Ogorek, B. (2021) The suitability of a quasi-Newton algorithm for estimating fitness-fatigue models: Sensitivity, troublesome local optima, and implications for future research (An *in silico* experimental design). *SportRχiv* (Preprint) DOI: 10.31236/osf.io/dx7gm

Abstract

Fitting an FFM via NLS in practice assumes that a unique optimal solution exists and can be found by the algorithm applied. However, this idealistic scenario may not hold for two reasons: 1) the absolute minimum may not be unique; and 2) local minima, saddle points, and/or plateau features may exist that cause problems for certain algorithms. If there exist different parameter sets in the domain that share the same global minimum under standard NLS, then there is a situation where parameters aren't uniquely identified without additional constraints or regularisation terms. However, more likely is that problems with the typical FFM fitting process will stem from the existence of local minima, saddles, or plateau features that cause the algorithm to converge to a solution not equal to the global minimum. Local optima can provoke sensitivities in the fitting process for first and second-order algorithms that are by definition local optimisers. This manifests as sensitivity to initial parameter estimates (i.e., the starting point the algorithm initialises the search from). The extent of starting point sensitivity is largely unknown in the context of FFMs for common algorithms adopted and has not been studied directly. Given this concern, research reporting a single model solution derived from 'one shot' minimisation of NLS via typical first and second-order algorithms is fundamentally limited by possible uncertainty as to the suitability of fitted estimates as global minimisers. Therefore, the primary aim of this study was to investigate the sensitivity of a classical first-order search algorithm to selection of initial estimates when fitting a fitness-fatigue model (FFM) via nonlinear least-squares (NLS), and to subsequently assess the existence of local optima. A secondary aim of this study was to examine the implications of any findings in relation to previous research and provide considerations for future experimentation. The aims of the study were addressed through a computer experiment (*in silico*) approach that adopted a deterministic assumption the FFM completely specified athlete response. Under this assumption, two FFMs (standard, and fitness-delay) were simulated under a set of hypothetical model inputs and manually selected 'true' parameter values (for each FFM), generating a set of synthetic performance data. The two FFMs were refitted to the synthetic performance data without noise (and under the same model inputs) by the quasi-Newton L-BFGS-B algorithm in a repetitive fashion initiated from multiple starting points in the parameter space, attempting to at each search recover the true parameter values. Estimates obtained from this process were then further transformed into prediction errors quantifying in-sample model fit across the iterations and non-true solutions. Within the standard model scenarios, 69.1-70.3% of solutions found were the true parameters. In contrast, within the fitness-delay model scenarios, 17.6-17.9% of solutions found were the true parameters. A large number of unique non-true solutions were found for both the standard model (N=275-353) and the fitness-delay model (N=383-550) in this idealistic environment. Many of the non-true extrema found by the algorithm were local minima or saddles. Strong in-sample model fit was also observed across non-true solutions for both models. Collectively, these results indicate the typical NLS approach to fitting FFMs is harder for a hill-climbing algorithm to solve than previously recognised in the literature, particularly for models of higher complexity. The findings of this study add weight to the hypothesis that there exists substantial doubt in reported estimates across prior literature where local optimisers have been used or models more complex than the standard FFM applied, particularly when optimisation procedures reported have lacked the relevant detail to indicate that these issues have been considered. Future research should consider the use of global optimisation algorithms, hybrid approaches, or different perspectives (e.g., Bayesian optimisation).

1 Introduction

The standard FFM emerges from a linear system of first-order ordinary differential equations (Banister et al., 1975). When solved, this ODE system yields a nonlinear function in the unknown model parameters (Busso et al., 1990; Clarke & Skiba, 2013; Morton et al., 1990). Other FFMs arise from similar ODE systems, although some involve non-linear system dynamics (Turner et al., 2017), higher derivatives (Calvert et al., 1976) (see appendix A), and recursion (Matabuena & Rodríguez-López, 2016, 2019). Therefore, fitting an FFM constitutes a nonlinear optimisation problem in its model parameters. FFMs dependent on time-invariant parameters (Banister et al., 1975), or in special cases time-varying parameters (Busso et al., 1997; Kolossa et al., 2017) that cannot be inferred from observation and must instead be estimated from quantified training load and measured performance data (Stephens Hemingway et al., 2021). The fitting process takes as input a time-series of measured performances (denoted p) and training load values (denoted ω) and provides as output model parameter estimates ($\theta \in \mathbb{R}^+$) that give good or preferably the best possible agreement between iteratively computed model values (denoted $\hat{p}$) and measured data (p). Essentially, to fit an FFM, a researcher or practitioner requires a series of suitable training load and performance measurement data, and a method to alter the parameters to best match these through an optimisation perspective.

The most common optimisation approach for fitting FFMs has been NLS (eq. 1) (Clarke & Skiba, 2013; Connor & O'Neill, 2020; Hellard et al., 2006; Pfeiffer, 2008; Proshin & Solodyannikov, 2018), or a maximum likelihood perspective (Busso, 2017; Scarf et al., 2019). Least-squares and maximum likelihood estimation coincide under the assumption of independent and identically distributed Gaussian model errors. Of the two, NLS has represented the most accessible approach across prior research and involves minimising the sum of squared deviations (also called errors) between modelled and measured performance (eq. 1) (a twice differentiable function) using some iterative algorithm.

$$\min \sum_{i=1}^{m} (\hat{p}_i - p_i)^2 \qquad (1)$$

Where in (eq. 1) i is an index over a set of m of data points $\{(p_1, \hat{p}_1), (p_2, \hat{p}_2), \dots, (p_m, \hat{p}_m)\}$ that represent measured (p) and modelled ($\hat{p}$) criterion performance values at specific integer time points $t_i \in \mathbb{N}$. The term $\hat{p}_i$ is determined by the FFM function $f\left(t_i, \ \theta, \ \{\omega_1, \dots, \omega_{t_i}\}\right)$ that not only depends on the time-step input (i.e. $\Delta_t = 1$) up to time t_i (i.e. training load series) $\{\omega_1, \omega_2, \dots, \omega_{t_i}\}$ but also on n model parameters (θ) with $m \geq n$. For example, with the standard FFM (eq. 2) the parameters θ comprise the set $\{p^*, k_g, \tau_g, k_h, \tau_h\}$, where p^* is an additive term representing baseline performance, τ_g, τ_h are the decay time constants on fitness and fatigue, respectively, and k_g, k_h are the associated scaling factors. NLS regression problems are typically solved using general minimisation methods, where the algorithm evaluates the cost function (eq. 1) and uses specific update and stopping criteria to travel the available parameter space to search for the best possible set (i.e., the absolute minimum of the function).

Fitting an FFM via NLS in practice assumes that a unique optimal solution exists and can be found by the algorithm applied. However, this idealistic scenario may not hold for two reasons: 1) the absolute minimum may not be unique; and 2) local minima, saddle points, and/or plateau features may exist that cause problems for certain algorithms. The FFM in basic form is a model in five dimensions (Banister et al., 1975), or six if a delay on fitness is also included (Calvert et al., 1976). Therefore, the parameter surface cannot be plotted or visually inspected via standard techniques to assess convexity. If there

exist different parameter sets in the domain that share the same global minimum under standard NLS, then there is a situation where parameters aren't uniquely identified without additional constraints or regularisation terms. However, more likely is that problems with the typical FFM fitting process will stem from the existence of local minima, saddles, or plateau features that cause the algorithm to converge to a solution not equal to the global minimum, or become lost (Philippe et al., 2018). Local optima can provoke sensitivities in the fitting process for first and second-order algorithms that are by definition local optimisers. This manifests as sensitivity to initial parameter estimates (i.e., the starting point the algorithm initialises the search from). The extent of starting point sensitivity is largely unknown in the context of FFMs for common algorithms adopted and has not been studied directly. Given this concern, research reporting a single model solution derived from 'one shot' minimisation of NLS via typical first and second-order algorithms is fundamentally limited by possible uncertainty as to the suitability of fitted estimates as global minimisers. Therefore, the primary aim of the experiment was to study the sensitivity of a quasi-Newton algorithm to selection of initial estimates, and the existence of local optima, when fitting an FFM under an NLS perspective. A secondary aim was to examine the implications of any findings in relation to previous research as well as considerations for future investigations. The aims of the experiment were addressed through an *in silico* (computer experiment) approach that adopted a deterministic assumption that the FFM completely specified athlete response. Under this assumption, two FFMs (standard, and fitness-delay model) were simulated under a set of hypothetical model inputs and manually selected parameter values (for each model), generating in the process a set of synthetic performance data. The parameter values represented *true* values for the model under the deterministic assumption, associated with the synthetic performance data. The two FFMs were refitted to the synthetic performance data without any noise by a quasi-Newton algorithm, under the same training load inputs, in a repetitive fashion starting from multiple points in the parameter space. This allowed starting point sensitivity of the algorithm to be assessed under best case conditions (no noise), and identification of the presence of local optima in the search space.

2 Materials and methods

2.1 Experimental approach to the problem

An *in silico* approach was developed, employing a first-order search algorithm (with second-order approximation) to fit two FFMs (eq's. 2, 3) from multiple starting values to associated synthetic performance data, in an iterative fashion. The performance data (a set of) was generated for each model under pre-defined true parameters and training loads (model inputs) via model simulation. At each iteration in the experiment, the algorithm was initialised from a different starting point (selected sequentially from a large grid of pre-determined values) and the appropriate model fit to the associated performance data via successive minimisation of the NLS objective function (eq. 1). The goal of the optimisation algorithm at each iteration was to try and recover the true parameters (global minimum) in the case where no additional noise exists. A *scenario* was defined as the combination of the model involved, and the proportion of simulated data used in the fitting process, with this latter factor described shortly. The total number of iterations in each scenario was equal to the total number of starting sets in the grid, which was therefore also equivalent to the total number of fitted estimates obtained. The synthetic model input values (i.e., daily training loads) used in the experiment were manually constructed to exhibit a realistic distribution (with regard to pattern, shape, and relative magnitude). Additionally, the true parameter values were selected such that the simulated performance values represented realistic performance change and variation over time.

Although the assumption of a completely deterministic model of athlete response is unrealistic due to simplification by design within the modelling process, this experimental approach is believed to be reasonable in a research context to enable lower-bound study of the fitting process in a manner not possible with real data. Furthermore, the simulated performance data were not unreasonable with regard to change in the performance profile, and true parameters (such as the decay constants) were chosen as to be interpretable with regard to model dynamics (Stephens Hemingway et al., 2021). In the real world, fitting FFMs involves non-zero (possibly large) residual solutions that make it impossible to be sure that the fitted estimates represent a unique global minimum. In contrast, the approach developed in this experiment allows the convergence to the true parameters representing a global minimum to be reliably assessed for different initial estimates (starting values) used by the algorithm, and when fitting to different proportions of the data (i.e., a lower measurement frequency). In the experiment, this second factor in each scenario (proportion of fitting data used) was contextually referred to as the *measurement frequency* based on its correspondence with the availability of data in practice. A reduction in measurement frequency was reflected by repeating the process described above for each model whilst fitting to a reduced subset of the simulated data. Three frequencies were studied: 1) *Every day* (ED) equivalent to 100% of the data; 2) *Every 2 days* (E2D) equivalent to approximately 50% of the data; and 3) *Every 3 days* (E3D) equivalent to approximately 33% of the data. At the heart of the research is to determine whether the fitting algorithm adopted is suitable for use in practice based on its ability (or lack of) to consistently recover the true parameters regardless of starting point or a decrease in volume of data supplied (given that no additional noise is incorporated). A flowchart detailing the computational process is presented in Figure 1, and the algorithm for the experimental flow stated in Appendix A, part A-1.
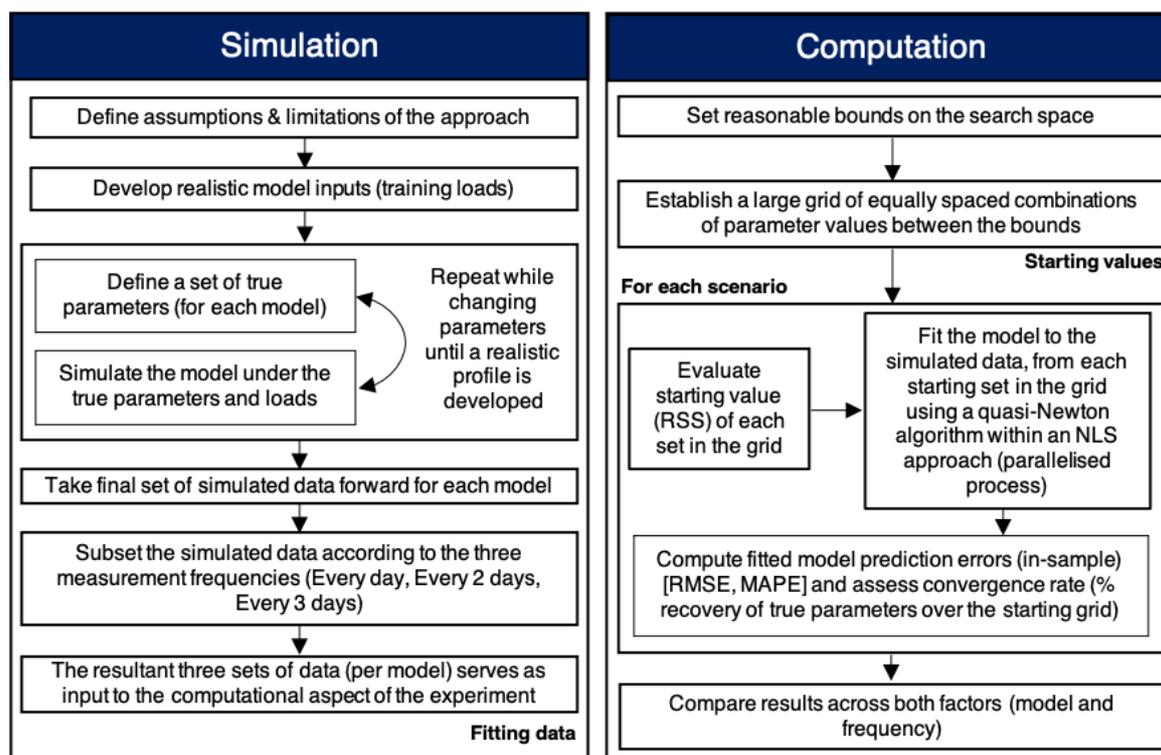
The standard model (Banister et al., 1975)

$$\hat{p}(t) = p^* + k_g \underbrace{\sum_{i=1}^{t-1} \omega_i \cdot e^{\frac{-(t-i)}{\tau_{g_1}}}}_{\text{fitness component}} - k_h \underbrace{\sum_{i=1}^{t-1} \omega_i \cdot e^{\frac{-(t-i)}{\tau_h}}}_{\text{fatigue component}} \qquad (2)$$

The fitness-delay model (Calvert et al., 1976)

$$\hat{p}(t) = p^* + k_g \underbrace{\sum_{i=1}^{t-1} \omega_i \cdot (e^{\overbrace{\frac{-(t-i)}{\tau_{g_1}}}^{effect}} - e^{\overbrace{\frac{-(t-i)}{\tau_{g_2}}}^{delay}})}_{\text{fitness component}} - k_h \underbrace{\sum_{i=1}^{t-1} \omega_i \cdot e^{\frac{-(t-i)}{\tau_h}}}_{\text{fatigue component}} \qquad (3)$$

Where for both models, $g(0) = h(0) = \omega_0 = 0$.

Figure 1: Flowchart describing the experimental approach to the problem (for both models)

## 2.2 Development of the synthetic model inputs (training loads)

The series of model input (training loads) used in the experiment were constructed based on concepts of daily variation and a subtle wave-like profile (Baker, 1998). Pre-simulation scaling of the load series was set to a maximum of 5% of the true baseline performance (additive term p^*), to avoid excessively small values of the scaling parameters (k_g, k_h), but the values were otherwise arbitrary. Primary emphasis was placed on the shape of the distribution (Baker, 1998) and reasonable relative differences in magnitude (i.e., between session), rather than absolute scale. Figure 2 provides a visual plot of the training loads developed over a 150-day time-series.
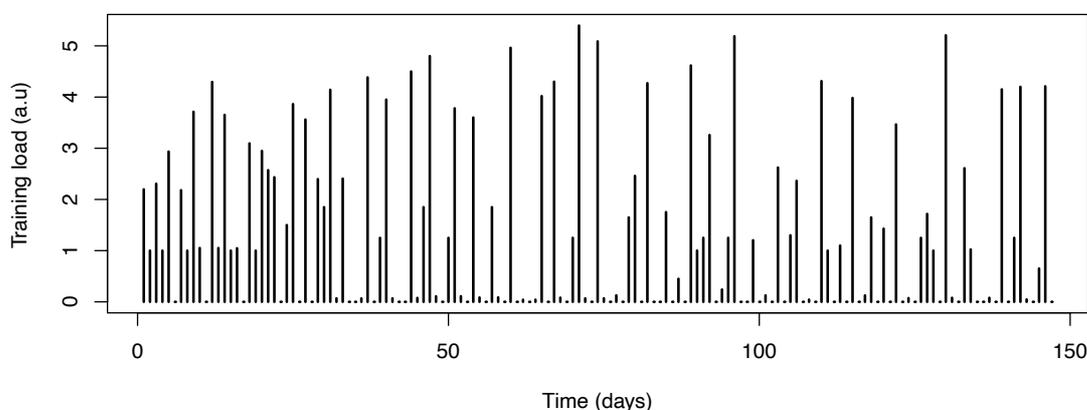

Figure 2: Hypothetical training load values for the experiment with realistic variation and wave-like profile

2.3 Simulated performance data

As described, the set of simulated daily performance data for each model were developed by first selecting a set of 'true parameters' associated with each model (Table 1) that produced a realistic profile with regard to magnitude and shape of performance change under the training load series. To assess magnitude, improvement in the maximum bench press (kg) for a moderately trained athlete over a 150-day period was used as a guide. Although the introduction of context is useful, it is also recognised that is it not necessary here and that these values could be left undefined without affecting validity of the experiment. The data used to fit (train) the model were developed by isolating a proportion of the model-generated performance values according to the measurement frequency condition in the scenario. In contrast to the experiment in Stephens Hemingway *et al*. (2019), no noise was introduced to the training data as this would disrupt the presence and purpose of establishing a known global minimum. The total proportion of values isolated from the simulated data and used to fit the model was defined by three measurement frequency conditions described previously (ED, E2D, E3D) and reflected by subsets of 100%, ~50% and ~33% respectively. Figure 3 shows the set of simulated performance data for each model, and figure 4 illustrates the underlying component profile.
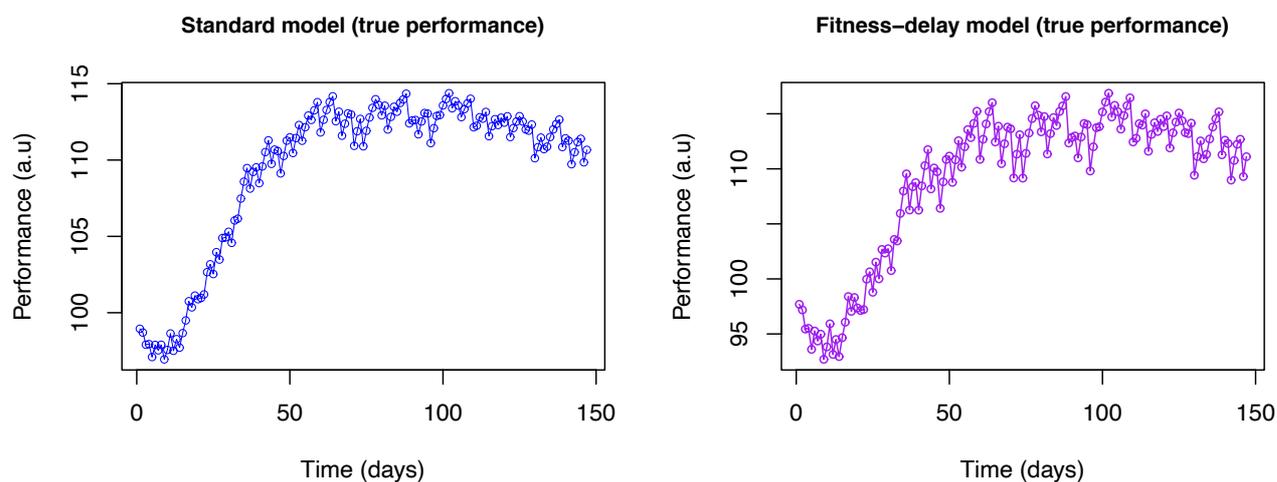
Figure 3: Simulated performance data generated for each model: Standard model (left) and fitness-delay model (right)
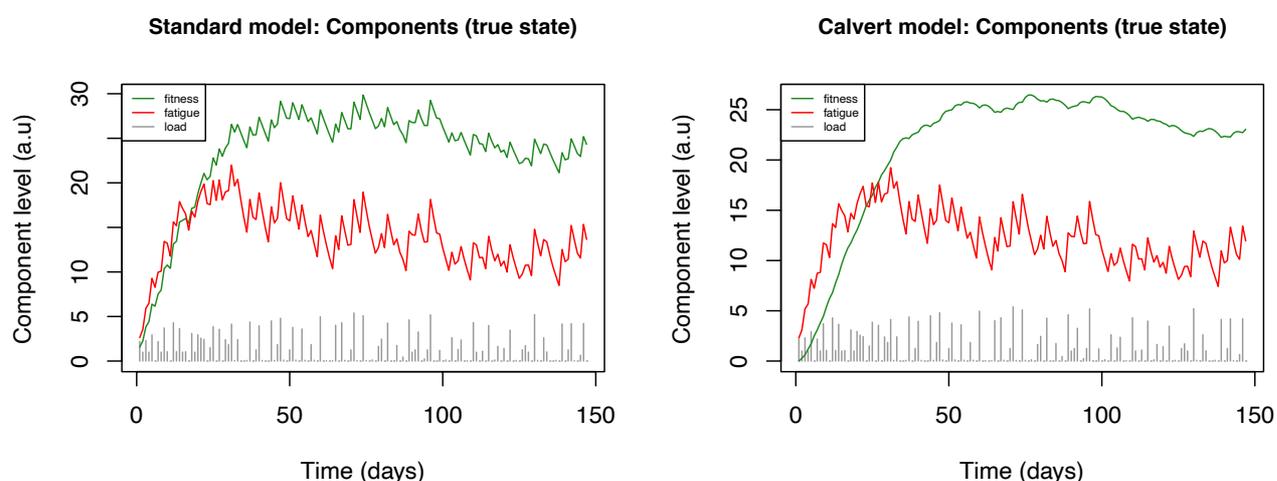
Figure 4: Simulated model component states: standard model (left) and fitness-delay model (right); fitness component (green), fatigue component (red), training load (grey)

Table 1: True parameters used in the experimental (simulated data) for each model

|	p^*	k_g	k_h	τ_h	τ_{g_1}	τ_{g_2}
Standard model	100	0.72	1.2	8.6	28.5	-
Fitness-delay model	100	0.72	1.05	8.6	32.5	4.3

2.4 Computation framework

A total of 6 scenarios were investigated comprising conditions of two factors (model [2] × measurement frequency [3]), and the following process was employed for each: 1) reasonable bounds on the parameter space were established on each parameter (see Table 2, equivalent across scenarios); 2) between the bounds, a discrete grid of feasible parameter combinations was constructed, with equally spaced step changes in each parameter (See Figure 5); 3) the value of each parameter set in the discrete grid was evaluated via the objective function (eq. 1) representing the 'fitness' (cost) of each starting point prior to fitting; 4) in an iterative and parallelised model fitting process, each parameter set in the grid was applied as the starting point of the quasi-Newton algorithm; that takes as input the training load series and simulated criterion performance values (fitting data, see section 2.3), and returns as output a set of fitted parameter estimates via minimisation of NLS; 5) fitted parameter estimates were combined with the training load series to generate fitted model predictions (project performance) and these were transformed into in-sample goodness-of-fit statistics: the root mean percentage error (RMSE), and mean absolute percentage error (MAPE).

Bounds on the parameter space were imposed in the form of box-constraints and these were chosen to not be too tight nor to close to the true parameters, but also not too large as to be physiologically non-interpretable (in particular for decay constants and additive term) or more than five times the magnitude of the difference between the performance values and the training load values (for scaling factors). For the scenarios comprising the standard model (eq. 2), each grid of starting values comprised a total of 10^5 parameter sets, and for the fitness-delay model scenarios (eq. 3) each grid comprised a total of 7^6 sets. A slightly larger size of grid for the fitness-delay model allowed for a reasonable step-size between changes in parameter values over the bounds to be maintained. A toy example of this method used to construct the grid is illustrated in Figure 5.

Table 2: Bounds on the parameter space and starting grid for each model

Parameter	Standard model (eq. 2)		Fitness-delay model (eq. 3)	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
p^*	60	140	60	140
k_g	0.01	5	0.01	5
k_h	0.01	5	0.01	5
τ_h	1	50	1	50
τ_{g_1}	1	50	1	50
τ_{g_2}	-	-	1	50
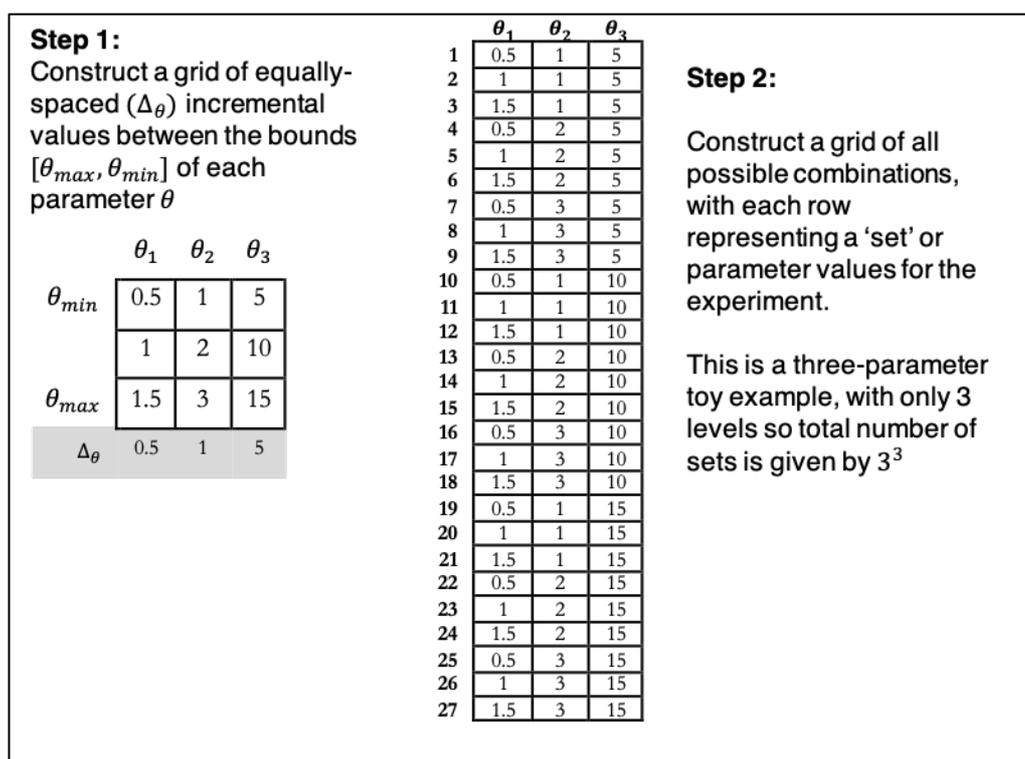
Figure 5: An illustration of the method used to construct the grids (generalised example)

2.5 The quasi-Newton algorithm

Most iterative algorithms to find minimisers of smooth NLS problems require computation and storage of first or second-order derivatives of the objective function (Mohammad & Waziri, 2019). For example, Newton-type methods require the exact Hessian, Gauss-Newton (GN) and Levenberg-Marquardt (LM) methods make use of the first-derivative information and ignore the second-order part of the Hessian, and quasi-Newton (QN) methods approximate the Hessian rather than having to iteratively compute and store it at each step (Mohammad & Waziri, 2019). However, each of these methods have associated limitations. For example, the exact second-order derivatives of the objective function are not normally available at a reasonable cost, and the analytic expression is often intractable for non-linear problems (Dennis Jr & Schnabel, 1996; Sun & Yuan, 2006). Thus, exact Newton-type methods that use on the exact Hessian (reflecting a description of the curvature of the function) are not typically suitable. GN and LM methods are expected to perform well with zero-residual problems, however when solving large-residual problems these methods can perform poorly and may not be suitable (Dennis Jr & Schnabel, 1996; Mohammad & Waziri, 2019). Quasi-Newton methods are a class of methods similar to the full Newton method but instead these approximate the Hessian, with approximations generally improving at each step. Quasi-Newton methods are typically computationally cheap, for example the Broydon-Fletcher-Goldfarb-Shannon (BFGS) algorithm (Byrd et al., 1995) requires $\mathcal{O}(n^2)$ operations per iteration, compared to the full Newton method that requires $\mathcal{O}(n^3)$ (Henao, 2014). Quasi-Newton methods have represented a popular choice for NLS optimisation in data fitting problems and are often available across a multiple of programming languages and mathematical suites (e.g., Mathematica, MATLAB, GNU Octave, R, SciPy). In particular, the BFGS algorithm is a standard tool for the optimisation of smooth functions (Wright & Nocedal, 1999) and includes an exact or inexact line search method to determine step size (Henao, 2014). The algorithm used in this experiment to solve the least-

squares problem at each iteration was an implementation of the limited memory modification of the BFGS method (L-BFGS) in R, with the inclusion of a further modification to incorporate box constraints (L-BFGS-B). The limited memory modification variant of the BFGS method uses less computer memory to update the approximation to the inverse of the Hessian, by only storing a record of the last m iterations rather than an $n \times n$ matrix where m is a small number and n is the number of parameters (Henao, 2014). As such, L-BFGS only requires $\mathcal{O}(mn)$ operations per iteration so is well suited to problems where the number of free parameters n is large. The inclusion of bounding however increases the cost of the line search slightly due to extra necessary steps to ensure the algorithm remains in the defined box with each step (Henao, 2014). The algorithm is available as part of the *optim* function included as part of the *stats* library included in the standard R environment (R Core Team, 2020). Analytic gradients were not supplied to the function, for reasons described above, and therefore the algorithm attempts to approximate the gradient using finite differencing, and this increases the possibility that in some instances abnormal (unsuccessful) termination in the line search may occur. Algorithm convergence is reported by the *optim* function as part of the convergence code and message returned following the search (Henao, 2014). Although supplying precise analytic gradient functions may improve the success of the algorithm, this is a challenging to intractable task and unrealistic approach for a sport science researcher when fitting FFMs Readers are referred to the works of Henao (2014) and Wright and Nocedal (1999) for more in-depth analysis of the behaviour of this algorithm. The parallelised searches were run on an 8-Core Intel® Xeon® Gold 6230 CPU @ 2.10GHz, with 8.0GB available RAM (80% average usage). Note, successful termination (convergence) refers to achieving an iterative reduction of the objective function that is within a factor (1e7) of the machine tolerance (2.2e-16), giving an approximate tolerance of 2e-9.

2.6 Analyses

Given that the procedure was evaluated on a large deterministic grid, it is sufficient to treat the results as a "complete population" given that no stochastic element was introduced and therefore there is no intention to perform inference about a superpopulation. Findings were best communicated by descriptive statistics and visualisations to summarise spread, shape, and centrality of fitted parameter estimates, prediction errors, and the rate of convergence to the true parameters (and other local optima) in each scenario. Local optima were appraised based on the definiteness of the Hessian matrix. For solutions where the associated Hessian was positive semi-definite and objective value (RSS) not equivalent to the known global extremum (i.e., 0 at the true parameters), the critical point was indicated to be a local minimum. Similarly, if the Hessian at a given solution was indefinite this indicated the point was a saddle. Appendix A provides further distributional summary tables of solutions at a resolution of each parameter.

Experimental code and analysis files can be found at: github.com/bsh2/experiments/qNewtonNLS

3 Results

## 3.1 Parameter estimates (convergence)

Within each scenario (i.e., model [standard, fitness-delay] × proportion of data [100%, ~50%, ~33%]), the L-BFGS-B algorithm terminated successfully during 99.71-99.97% of the iterations. The estimates from searches that terminated successfully are referred to as "solutions", although this term does not imply whether the estimates reached the true parameters. For all models, a reduction in the amount of data did not appear to influence the number of solutions that reached the true parameters (i.e., the number of fitting iterations that recovered the true global minimum). Within the standard model scenarios, 69.1-70.3% of solutions found were the true parameters. In contrast, within the fitness-delay model scenarios, 17.6-17.9% of solutions found were the true parameters. Within the standard model scenarios, the remaining non-true solutions resolved to other critical points including saddles (27.6-28.8%) and a small number of local minima (2-2.1%). Within the fitness-delay model scenarios, the remaining non-true solutions resolved to predominantly local minima (76.1-78.3%) and a small number of saddle points (4.0-5.9%). Table 3 provides comparison between the scenarios with respect to the results outlined so far. The parameter distributions of the solutions that did not find the true parameters were similar between the three standard model scenarios (i.e., 100%, 50%, 33% of fitting data) (see Figure 6). This was also the case for the fitness-delay model scenarios (figure 6). Tables of summary statistics describing the fitted parameter estimate distributions are provided in Appendix A parts A-2 and A-3. Tables of the highest frequency (non-true) solutions found for each scenario (model × data proportion) are given in Appendix A parts A-4 and A-5.

Table 3: Convergence rates of the solutions found by the L-BFGS-B algorithm (to critical points in the parameter space)

Scenario			Totals			Convergence rates (Critical points)		
Model	Data (%)	Data points	Iterations (total sets)	Successful termination	Abnormal termination	True parameters	Other local minima	Saddle points
Standard	100 %	147	10^5	99967 (99.97%)	33 (0.03%)	69204 (69.2%)	2047 (2.1%)	28716 (28.7%)
Standard	50 %	74	10^5	99968 (99.97%)	32 (0.03%)	69145 (69.1%)	1995 (2.0%)	28828 (28.8%)
Standard	33 %	49	10^5	99960 (99.96%)	40 (0.04%)	70284 (70.3%)	2056 (2.1%)	27620 (27.6%)
Fitness-delay	100 %	147	7^6	117305 (99.71%)	344 (0.29%)	20588 (17.6%)	91909 (78.1%)	4808 (4.1%)
Fitness-delay	50 %	74	7^6	117492 (99.87%)	157 (0.13%)	20651 (17.6%)	92127 (78.3%)	4714 (4.0%)
Fitness-delay	33 %	49	7^6	117551 (99.92%)	98 (0.08%)	21065 (17.9%)	89518 (76.1%)	6968 (5.9%)
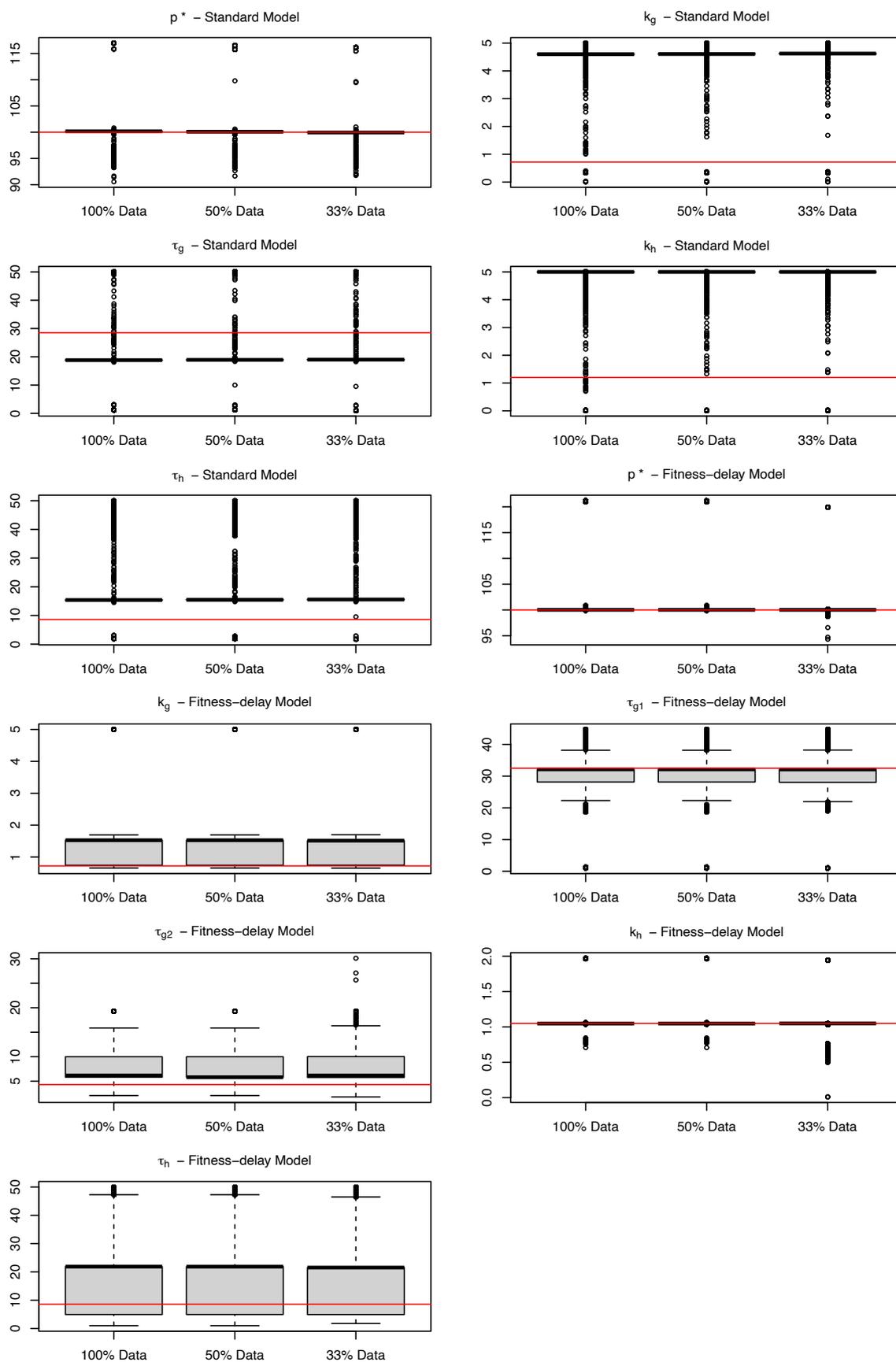
Figure 6: Parameter estimate distributions from the solutions that did not reach the true values (i.e., global minimum), for the standard and fitness-delay models. The red line indicates the true value.
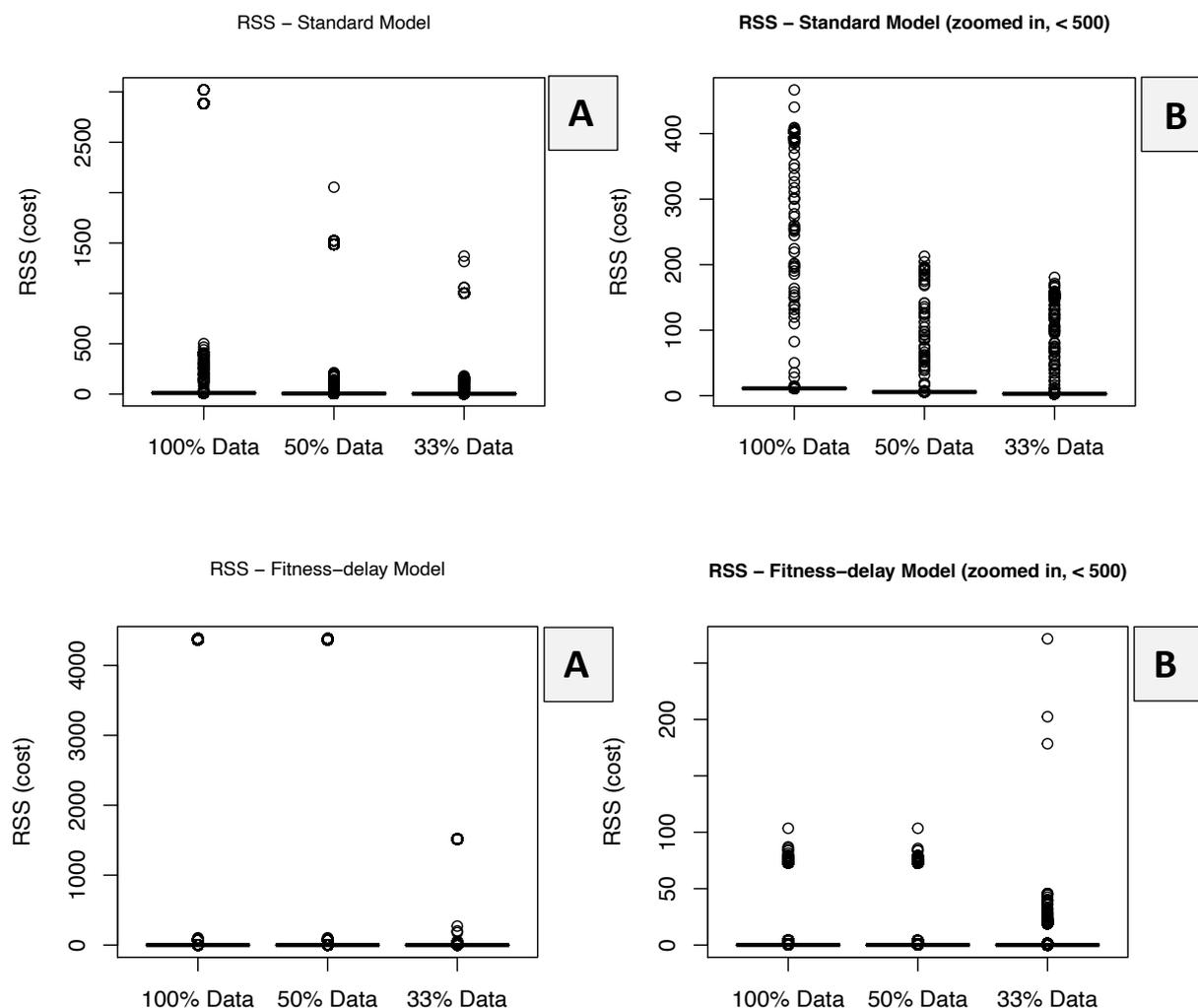
Figure 7: Objective function values (RSS) associated with solutions that did not reach the true values (i.e., global minimum), for the standard and fitness-delay models (A). The (B) plots offer a 'zoomed in' picture of the distribution following removal of large values (outliers, RSS < 500).

3.2 Prediction errors (model fit)

In-sample median model fit across solutions that did not reach the true parameters were similar and strong in all scenarios and metrics including root-mean-squared error (RMSE) and mean average percentage error (MAPE). For example, RMSE$_{(median)}$ for the standard FFM ranged from 0.24-0.28 and RMSE$_{(median)}$ for the fitness-delay model was 0.03-0.06 across all proportions of fitting data. Median absolute deviation RMSE$_{(m.a.d)}$ was ~0 for the standard model and 0.01- 0.03 for the fitness-delay model searches. Although there were a small number of solutions in each scenario that resolved to poor model fit (i.e., RMSE = 5.27 and correspondingly high RSS values) (seen visually as the outliers in the fitted model traces plotted in Figure 8 and 10). Figures 8 and 10 also illustrate visually that the range of daily prediction errors (performance profiles) across fitted parameter sets that did not reach the true parameter values was low, and figures 9 and 11 provide the associated distributional plots for the errors. In each scenario, parameter sets returned by the algorithm from searches that resulted in abnormal termination also resolved to good model fit (RMSE = 0.01 - 1.77, MAPE = 0.01 - 1.30%, across all

scenarios for both models) (Table 4) despite differences in the parameter values in these sets relative to the global minimum point (Tables A-2, A-3, appendix A).

Table 4: Model fit (in-sample) summary statistics for the fitted solutions

Data	Converged	Descriptive statistic	Standard model		Fitness-delay model	
			RMSE	MAPE	RMSE	MAPE
100%	Non-true solutions	minimum	0.28	0.21	0.01	0.01
		maximum	4.53	3.18	5.46	3.87
		median	0.28	0.22	0.03	0.02
		m.a.d	0.00	0.00	0.02	0.02
100%	Abnormal termination	minimum	0.28	0.22	0.01	0.01
		maximum	0.28	0.22	0.06	0.05
		median	0.28	0.22	0.06	0.04
		m.a.d	0.00	0.00	0.01	0.01
50%	Non-true solutions	minimum	0.28	0.21	0.00	0.00
		maximum	5.27	3.98	5.46	3.87
		median	0.28	0.21	0.03	0.02
		m.a.d	0.00	0.00	0.03	0.02
50%	Abnormal termination	minimum	0.28	0.21	0.01	0.01
		maximum	0.28	0.21	0.71	0.54
		median	0.28	0.21	0.05	0.04
		m.a.d	0.00	0.00	0.02	0.02
33%	Non-true solutions	minimum	0.24	0.19	0.01	0.01
		maximum	5.29	4.06	5.56	3.94
		median	0.24	0.19	0.03	0.02
		m.a.d	0.00	0.00	0.03	0.02
33%	Abnormal termination	minimum	0.24	0.19	0.01	0.01
		maximum	1.77	1.30	0.07	0.05
		median	0.24	0.19	0.06	0.04
		m.a.d	0.00	0.00	0.01	0.01

m.a.d : median absolute deviation
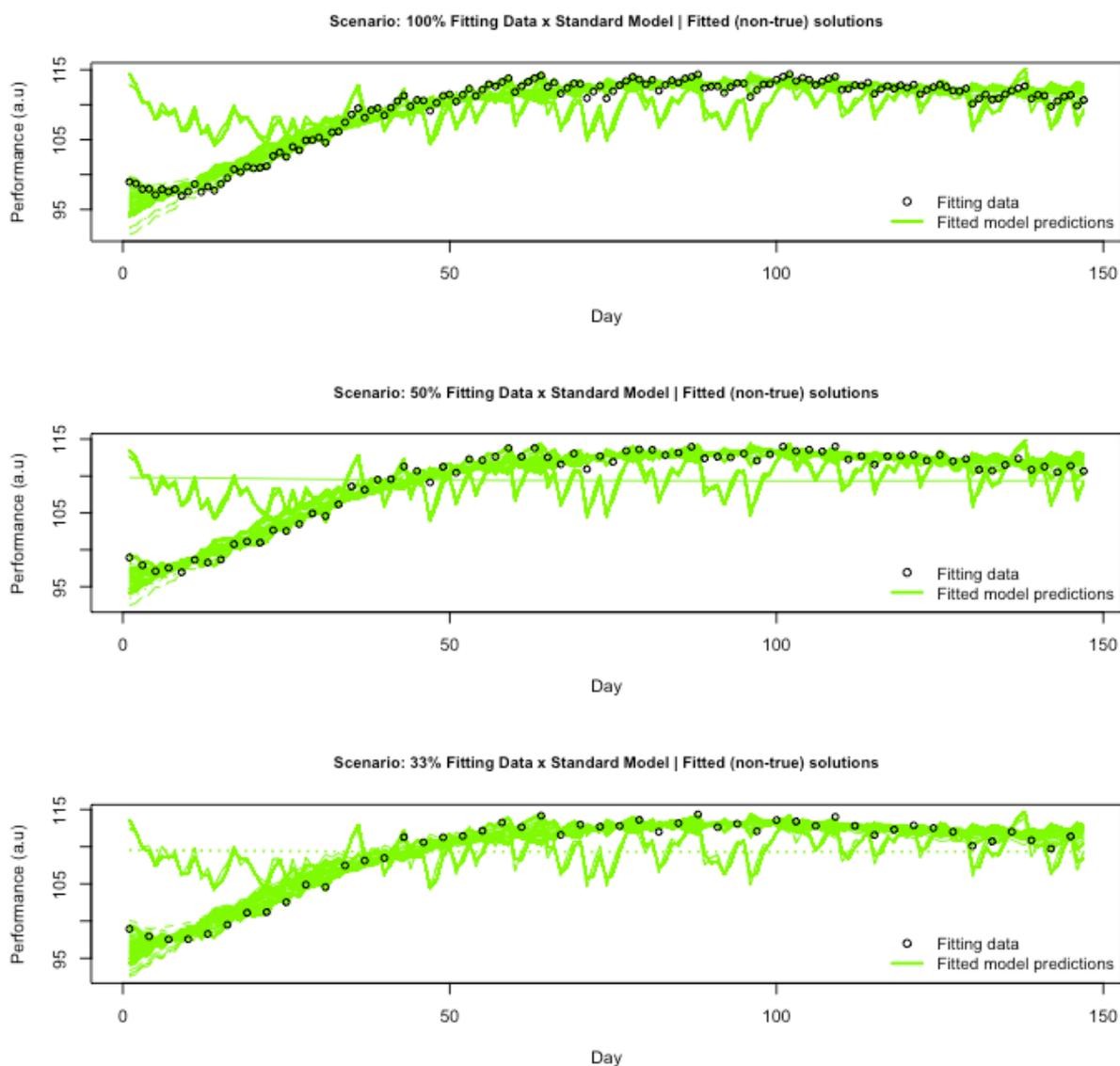
Figure 8: Fitted model predictions (in-sample) reflecting the range of performance profiles generated by the non-true solutions – Standard model scenarios (green)
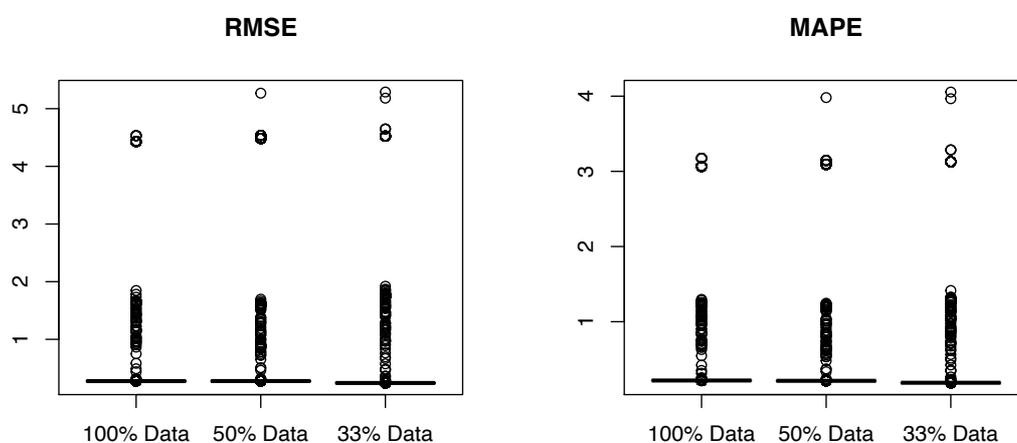
Standard Model

Figure 9: Comparison of in-sample goodness-of-fit (RMSE, MAPE) for non-true solutions, obtained for the standard model searches across the three proportions of fitting data
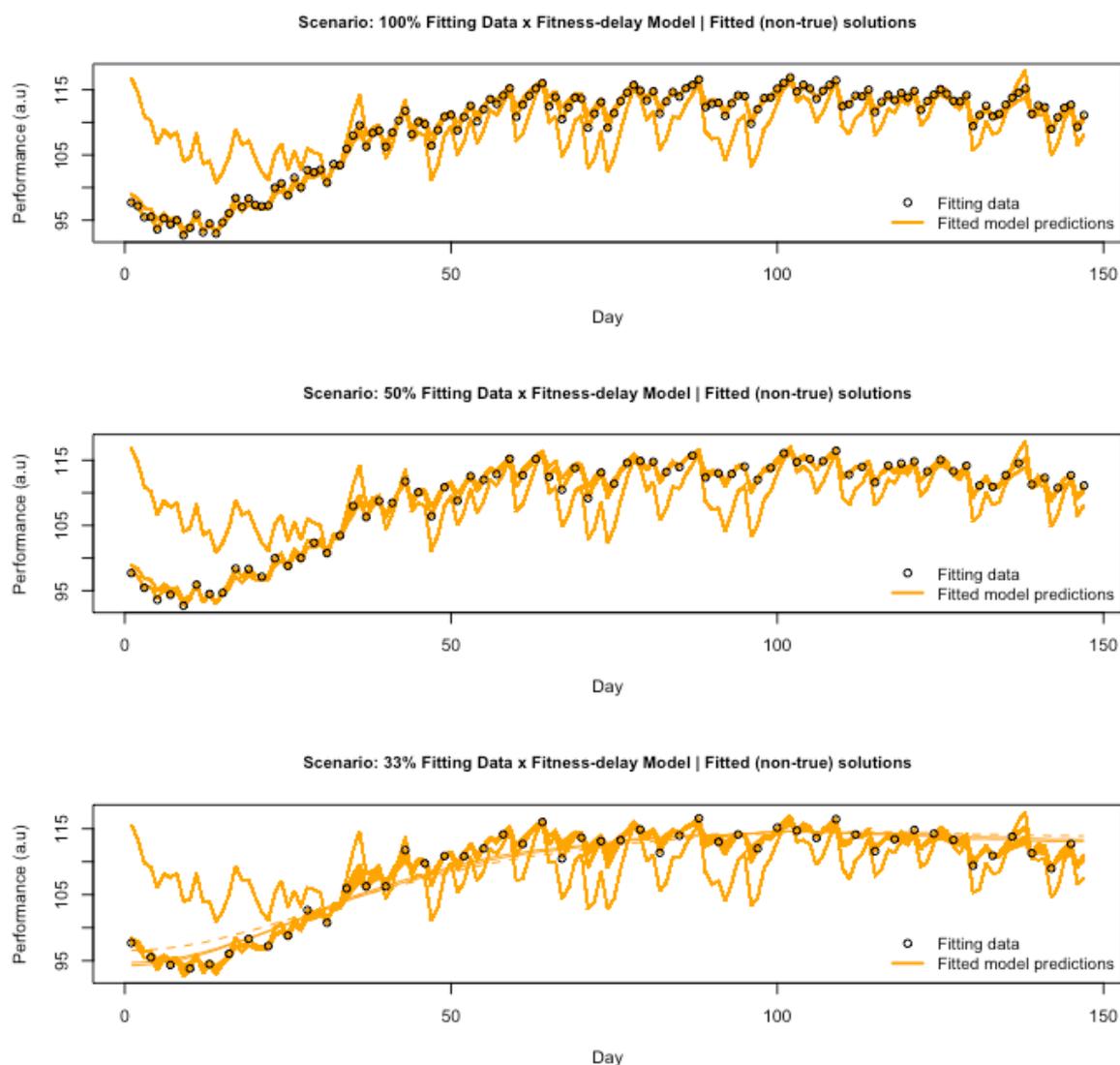
Figure 9: Fitted model predictions (in-sample) reflecting the range of performance profiles generated by the non-true solutions – Fitness-delay model scenarios (blue)
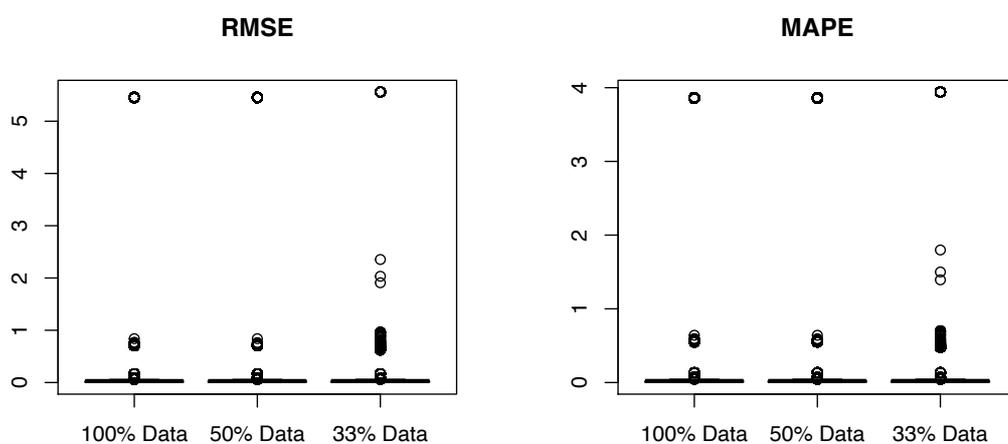

Fitness-delay model

Figure 5.11: Comparison of in-sample goodness-of-fit (RMSE, MAPE) for non-true solutions, obtained for the fitness-delay model searches across the three proportions of fitting data trialled.

3.3 Runtime

There were large differences in the fitting time between the two models, and within each model when fitting to additional data. An increase in data for both models (%) resulted in an approximately directly proportional increase (%) in fitting time per search. The fitness-delay model was also substantially slower to fit per search (approximately 2.8-2.96 times slower) in comparison to the standard model, likely due to the presence of the additional free parameter. These results are summarised in table 5.

Table 5: Fitting runtime across all scenarios for L-BFGS-B algorithm

Model	Data points	Iterations (N)	Runtime (Total)	Mean runtime (per search)
Standard	100% (N=147)	100,000	74 Hours	2.7 Seconds
Standard	50% (N=74)	100,000	38 Hours	1.4 Seconds
Standard	33% (N=49)	100,000	28 Hours	1.0 Seconds
Calvert	100% (N=147)	117,649	263 Hours	8.0 Seconds
Calvert	50% (N=74)	117,649	133 Hours	4.1 Seconds
Calvert	33% (N=49)	117,649	90 Hours	2.8 Seconds

Runtime (total) given to the approximate hour

4 Discussion

This study investigated the typical model fitting process for two common fitness-fatigue models under a well-known quasi-Newton algorithm (BFGS), with limited memory modification and bounding (L-BFGS-B). To enable direct study of the effectiveness of the optimisation algorithm, each model was assumed to fully specify the training response and simulated under pre-selected true parameters and synthetic load inputs to derive performance data without additional error. It was argued that for this common algorithm to be robust for FFM fitting problems, it should as a basic capability be able to find the true minimum (zero residual) known (only) to exist in this simulation framework. It was known prior to the study that the convexity of the objective function over its domain would influence the effectiveness of a local optimisation algorithm such as BFGS, but the implications of local optima had previously never been recognised in prior work and ignored as a possible issue with reported solutions.

When initiated from a wide array of initial points spanning the parameter space in a uniform fashion, the L-BFGS-B algorithm was successful at finding the true solution in approximately 69-70% of the searches in each of the standard model scenarios but was only successful in between 15-18% of searches in the fitness-delay model scenarios (Table 5.3). These results demonstrate concerns with the algorithms ability to obtain suitable fitted estimates for both models even under the idealistic condition of no model error. These concerns were further exacerbated for the fitness-delay model, where the optimiser was only able to find the true values in less than 18% of the searches, with a high frequency of local optima across the search space demonstrated. It is likely that a substantial reduction in the rate of successful convergence to the true solution in this experiment in the fitness-delay model scenarios is due to the added complexity in the search space as a result of the additional fitness parameter (τ_{g_2}) and its relationships with other parameters in the model creating a higher incidence of local optima. There were no discernible patterns in the distribution of initial estimates with respect to starting error (RSS) for the standard (appendix A-3, figure A-3.1, 3.2) and fitness-delay model scenarios (appendix A-4, figure A-4.1, 4.2).

Collectively, these results indicate that the typical NLS fitting approach is a harder optimisation problem for a typical Hill-climbing algorithm to solve than previously recognised in the literature, and in particular when fitting the fitness-delay model. However, these results would benefit from further confirmatory experimentation under different input distributions to determine average rates across varying inputs, and alternative standard algorithms (first and second-order methods). It is likely however, that more advanced global optimisation algorithms such as differential evolution, and genetic algorithms, may be required to adequately solve the NLS data fitting problem within fitness-fatigue modelling.

A reduction in the amount of data used to fit the models did not appear to affect the algorithms convergence rates to the true solution in this experiment, or the distributions of the non-true solutions (parameter values) and associated model fit (Table 5, Figure 6). As would be expected, a reduction in fitting data did improve fitting time. In isolation, this reduction in fitting time is of little practical interest, as fitting an individual model to data via the method used in this experiment is relatively short (i.e., seconds), and researchers would always be expected to maximise available data even at the expense of runtime. However, fitting time may become important when incorporating robust observational or tuning-based cross-validation approaches to evaluate model validity, and/or when using complex optimisation algorithms (e.g., genetic algorithms). In these instances, researchers are advised to allocate reasonable compute resources and sufficient time to the fitting process and consider the use of parallelisation.

Many unique (non-true) solutions (to 1.d.p $\{p^*, \tau_g, \tau_h\}$ and 2.d.p $\{k_g, k_h\}$) were found in each of the scenarios involving the standard model (range = 275-353 per scenario) (Appendix A-4). The set of unique solutions collectively spanned most of the search space in all parameters, and the most were saddle points. The frequency of each unique solution was highly variable (for example, a given unique solution appeared between 1 and 25838 times in one standard model scenario). Appendix D-3 demonstrates the top 10 most frequent solutions across each standard model scenario, and supplementary file 1 (SF-1)[1] contains the entire set of unique solutions for the standard model and fitness-delay model scenarios. In contrast to the standard model scenarios, most of the unique solutions found in the fitness-delay model scenarios (range = 383-550 per scenario) (Appendix A-5) were local minima, rather than saddle points. The frequency at which a given unique solution appeared also demonstrated high variation (for example between 1 and 39401 times in one fitness-delay model scenario). Collectively, these results demonstrate that there appear to be many points in the search space at which the algorithm can become stuck, but also that local solutions can be found across the search space.

The implications of these results on prior and future research are that unless efforts were (or are) made to improve the likelihood that a solution found is the absolute minimiser of the NLS problem, then subsequent results relating to prediction accuracy under solutions carried forward may not be interpretable or robust generalisations of model error. Although, quite obviously, there is never a known comparator solution (as is the case in a simulation approach) within real world experiments due to model misspecification and noise, an appropriate recommendation arising from the results of this experiment is the re-running of first and second-order algorithms that require starting points from a large grid of stochastically generated points covering most of the parameter space between the bounds. The second recommendation arising from this study is further scientific investigation of global optimisation

---

[1] SF-1 can be found at the following repository link:
  github.com/bsh2/experiments/qNewtonNLS/SF-1.xlsx

methods, such as evolutionary strategies for fitting FFMs via NLS, and in particular the prospective use of global methods with integrated random local search (via first and second-order methods).

Another notable result of this study was that strong model fit was observed across most solutions that did not converge to the true solutions for each model (and the associated scenarios) (Table 4, Figures 5.8-5.11). This finding would benefit from further confirmatory experimentation under different input distributions. However, it is highly conceivable that in prior applications of FFMs, researchers may have ignored the possibility that solution obtained is not the absolute minimiser, specifically when measures of in-sample model fit are particularly strong (e.g., R^2, RMSE, MAPE), and where no estimate of fitted estimate uncertainty or starting point sensitivity has been determined. Therefore, the findings in this experiment add weight to the hypothesis that there exists substantial doubt in reported estimates across prior research; particularly where optimisation procedures have not been stated clearly or have lacked the relevant procedural detail to indicate that these issues have been considered or addressed. The negative implications of this are primarily placed on the interpretation of prior model validity work, and subsequent decisions made by researchers with regard to the collective optimism (or lack thereof) toward further study of FFMs.

The main limitation of this study is that there exists a possibility that results may differ under different input (training load) distributions, and that the exact assumptions required to enable study of the optimisation process via computation are too unrealistic in the real world, due to extensive model misspecification. Specifically, fitting to measured performance data is unlikely to ever be zero residual optimisation problem due to the presence of noise and inherent simplification within the modelling process resulting in model misspecification. Further, it is unclear the role that model misspecification will play in these results, and it is possible that algorithm performance may be considerable worse such that changes in local minima have substantive differences in predictions and model fit. However, it seems unlikely they would be better. It appears clear that further work in the application of optimisation approaches for FFM problems is required, and that alterative algorithmic approaches (e.g., evolutionary or genetic algorithms) should be evaluated, and different perspectives (e.g., the use of priors under a Bayesian approach) represent sensible pathways for future work.

5 Conclusions

Collectively, this experiment highlights that significant care must be taken in future research and practice to ensure that the optimisation problem is appropriately posed, and that the algorithmic approach selected to fit the selected FFM is sufficient, due to the high likelihood of local optima. In particular, solutions may not be all that they appear following one-shot minimisation using a hill-climbing algorithm such as L-BFGS-B, even in the presence of very good in-sample fit. At a minimum, multiple runs of optimisation should be performed under this approach, starting from many points spanning the breadth of the search space (sensitivity), and include some form of observational cross-validation as discussed to estimate uncertainty. Notably, both fitness-fatigue models demonstrated that different non-true solutions may exhibit the same model behaviour and achieve strong model fit. This creates a series of challenges for researchers in obtaining solutions in real-world experiments via similar approaches, as absolute minimisation cannot be confirmed, and uncertainty only estimated. It also casts doubt across correctness of solutions reported in prior FFM literature. This experiment has highlighted that the search space of the standard model, and in particular the fitness-delay model are more complex

and challenging for standard algorithms than previously recognised; and it is likely that we require better algorithmic approaches to solve FFM data fitting problems. Bayesian methods and evolutionary algorithms may offer two possible routes toward improved fitting of FFMs. However, the role of cross-validation (out-of-sample testing) in the model evaluation process can also not be ignored going forward. Out-of-sample testing may also offer a qualitative approach for flagging solutions that are clearly incorrect, or it may also be used in a more modern sense within tuning-based cross-validation frameworks. This work, although extensive, is an n-of-1 with respect to the input distribution, and so would benefit from further replication under different training load distributions and performance profiles.

Bibliography

Baker, D. (1998). Applying the in-season periodization of strength and power training to football. *Strength and Conditioning Journal*. https://doi.org/10.1519/1073-6840(1998)020<0018:atispo>2.3.co;2

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. (2017). From an indirect response pharmacodynamic model towards a secondary signal model of dose-response relationship between exercise training and physical performance. *Scientific Reports*, *7*(November 2016), 1–11. https://doi.org/10.1038/srep40422

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: Modeling in Physiology*, *82*(7), 1685–1693.

Busso, T., Häkkinen, K., Pakarinen, A., Carasso, C., Lacour, J. R., Komi, P. V, & Kauhanen, H. (1990). A systems model of training responses and its relationship to hormonal responses in elite weight-lifters. *European Journal of Applied Physiology and Occupational Physiology*, *61*(1), 48–54. https://doi.org/10.1001/jama.1937.02780230056030

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208. https://doi.org/10.1137/0916069

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

Clarke, D. C., & Skiba, P. F. (2013). Rationale and resources for teaching the mathematical modeling of athletic training and performance. *American Journal of Physiology - Advances in Physiology Education*, *37*(2), 134–152. https://doi.org/10.1152/advan.00078.2011

Connor, M., & O'Neill, M. (2020). Optimizing the Parameters of A Physical Exercise Dose-Response Model: An Algorithmic Comparison. *ArXiv Preprint ArXiv:2012.09287*.

Dennis Jr, J. E., & Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations*. SIAM.

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

Henao, W. (2014). An L-BFGS-B-NS optimizer for non-smooth functions. *Master's Thesis*.

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

Matabuena, M., & Rodríguez-López, R. (2016). A new approach to predict changes in physical condition: A new extension of the classical Banister model. *ArXiv Preprint*. http://arxiv.org/abs/1612.08591

Matabuena, M., & Rodríguez-López, R. (2019). An Improved Version of the Classical Banister Model to Predict Changes in Physical Condition. *Bulletin of Mathematical Biology*, *81*(6), 1867–1884. https://doi.org/10.1007/s11538-019-00588-y

Mohammad, H., & Waziri, M. Y. (2019). Structured two-point stepsize gradient methods for nonlinear least squares. *Journal of Optimization Theory and Applications*, *181*(1), 298–317.

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.

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.

Philippe, A. G., Borrani, F., Sanchez, A. M. J., Py, G., & Candau, R. (2018). Modelling performance and skeletal muscle adaptations with exponential growth functions during resistance training. *Journal of Sports Sciences*, *37*(3), 254–261. https://doi.org/10.1080/02640414.2018.1494909

Proshin, A. P., & Solodyannikov, Y. V. (2018). Physiological Avatar Technology with Optimal Planning of the Training Process in Cyclic Sports. *Automation and Remote Control*, *79*(5), 870–883. https://doi.org/10.1134/S0005117918050089

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

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., Burgess, K., Elyan, E., & Swinton, P. (2019). The effects of measurement error and testing frequency in applying the Fitness Fatigue Model to resistance training : A simulation study. *International Journal of Sports Science and Coaching*, *0*(0), 1–12. https://doi.org/10.13140/RG.2.2.19730.56005

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

Sun, W., & Yuan, Y.-X. (2006). *Optimization theory and methods: nonlinear programming* (Vol. 1). Springer Science & Business Media.

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

Wright, S., & Nocedal, J. (1999). Numerical optimization. *Springer Science*, *35*(67–68), 7.

Appendix A

A-1 Experiment algorithm

Requirements:

- Let $f(\theta, \omega)$ denote the fitness-fatigue model function
- Let θ be the parameter space $\theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_m) \in \mathbb{R}^+$ of f
- Let ω denote the training load series $\omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ of length n and time-step 1 day.

```
BEGIN MAINPROGRAM
   SET lower bound on θ: l = (l₁,l₂,l₃,…,lₘ),lᵢ ∈ ℝ⁺
   SET upper bound on θ: u = (u₁,u₂,u₃,…,uₘ), uᵢ ∈ ℝ⁺,uᵢ > lᵢ
   SET η where ηᵐ is the total size of the grid (η,m ∈ ℕ)
   SET γ where nᵐ mod γ ≠ 0, γ ∈ ℕ
   COMPUTE Sample η equally spaced points from each open interval (lᵢ,uᵢ),i ∈ [1,m]
   COMPUTE Make grid_θ of size ηᵐ by taking all combinations
   COMPUTE Partition grid_θ into γ equal segments grid_θ ≡ {grid_θ⁽¹⁾,grid_θ⁽²⁾,….,grid_θ⁽ᵞ⁾}
   SET θ_TRUE ∈ (l,u)
   INPUT ω
   COMPUTE vector p ∈ ℝ of length n by computing f(θ_TRUE,ω)
   FOR ν = 1 to 3
         INITALISE empty array RESULT_ν of dim (nᵐ × (m + 3))
         COMPUTE the sequence s, where s goes from 1 to n by ν
         COMPUTE p_s by sampling p at each point s
         FOR j = 1 to γ
           INPUT grid_θ⁽ʲ⁾ into working memory
           DO IN PARALLEL FOR k = 1 to nᵐ/γ
                 COMPUTE p̂_k from f(θ_k,ω), θ_k ∈ grid_θ⁽ʲ⁾
                 COMPUTE p̂_kₛ by sampling p̂_k at each point s
                 COMPUTE RSS_INIT = ∑(p̂_kₛ − p_s)²
                 OUTPUT Store RSS_INIT to array RESULT_ν
                   INITIALISE The quasi-Newton method at starting point θ_k
                   SOLVE min∑(f(θ_FIT,ω) − p_s)² s.t. θ_FIT ∈ [u,l]
                   OUTPUT Row-bind θ_FIT and RSS_FIT to array RESULT_ν
                   COMPUTE p̂_θ_FIT from the model for {θ_FIT,ω}
                   COMPUTE Fit statistic ε = √(∑(p̂_θ_FIT − p_s)²)  (example RMSE)
                   OUTPUT Store ε to array RESULT_ν
           END PARALLEL FOR
           INPUT Pull grid segments {grid_θ⁽¹⁾,…,grid_θ⁽ᵞ⁾} into working memory
           Recombine grid_θ and remove segments
           Columnar-bind grid_θ with RESULT_ν: yields array of dim (nᵐ × (2m + 3))
         END FOR
   END FOR
END MAINPROGRAM
```

Algorithm Notes

- The algorithm was written in pseudocode to make it language agnostic and therefore most understandable.
- R stores and manipulates all objects in the physical memory and therefore to conserve working memory during the implementation we opted to split and save the grid into smaller segments that would be loaded in sequential order. In our experiment we only had 8GB of RAM available on the machine. The constant γ is the number of smaller grids used to conserve memory.
- `DO IN PARALLEL` indicates that the operations were distributed to available nodes (via multi-core) and executed in parallel. In our experiment we used 8 available nodes within a single machine.
- Dimensions of arrays are given in rows $\times$ column format.

A-2 Parameter estimate distributions (standard model)

Table A-2: Parameter estimate and RSS distributions of solutions obtained for the standard model

Scenario	Convergence	Summary statistics	p^*	k_g	τ_{g_1}	k_h	τ_h	RSS$_{solutions}$	RSS$_{initial}$
Standard model **100%** **data**	**True parameters** **(N = 69204)**	Min	100.00	0.72	28.47	1.20	8.58	0.00	288
		Max	100.00	0.72	28.54	1.20	8.61	0.00	12893022
		Median	100.00	0.72	28.50	1.20	8.60	0.00	529882
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	710954
	Other solutions **(N = 30763)**	Min	90.55	0.01	1.19	0.01	1.93	11.17	457
		Max	116.97	5.00	50.00	5.00	50.00	3019.36	12771646
		Median	100.16	4.60	18.82	5.00	15.37	11.17	616682
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	845510
	Abnormal termination **(N = 33)**	Min	100.16	4.60	18.82	5.00	15.37	11.17	13268
		Max	100.16	4.60	18.82	5.00	15.37	11.17	7401559
		Median	100.16	4.60	18.82	5.00	15.37	11.17	655974
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	860297
Standard model **(50%** **data)**	**True parameters** **(N = 69145)**	Min	100.00	0.72	28.47	1.20	8.57	0.00	135
		Max	100.00	0.72	28.56	1.20	8.62	0.00	6470310
		Median	100.00	0.72	28.50	1.20	8.60	0.00	267974
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	359878
	Other solutions **(N = 30823)**	Min	90.55	0.01	1.19	0.01	1.93	0.00	457
		Max	116.97	5.00	50.00	5.00	50.00	3019.36	12858865
		Median	100.16	3.81	18.82	5.00	15.37	11.17	609208
		M.A.D.	0.24	1.18	14.35	0.00	10.04	16.56	830985
	Abnormal termination **(N = 32)**	Min	100.09	4.61	18.91	5.00	15.47	5.66	7159
		Max	100.09	4.61	18.91	5.00	15.47	5.66	4484160
		Median	100.09	4.61	18.91	5.00	15.47	5.66	480631
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	555066
Standard model **(33%** **data)**	**True parameters** **(N = 70284)**	Min	100.00	0.72	28.43	1.20	8.57	0.00	101
		Max	100.00	0.72	28.57	1.20	8.63	0.00	4266927
		Median	100.00	0.72	28.50	1.20	8.60	0.00	178693
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	239610
	Other solutions **(N = 29676)**	Min	91.75	0.01	1.00	0.01	1.68	2.89	306
		Max	116.14	5.00	50.00	5.00	50.00	1370.32	4018048
		Median	99.94	4.62	18.98	5.00	15.57	2.89	197519
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	269993
	Abnormal termination **(N = 40)**	Min	93.74	0.33	18.98	0.01	15.57	2.89	1920
		Max	99.94	5.00	50.00	5.00	49.64	154.04	2239363
		Median	99.94	4.62	18.98	5.00	15.57	2.89	415359
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	477229

M.A.D refers to the median absolute deviation. Data % refers to the proportion of data used in the fitting process (i.e., 100% corresponds to a measurement frequency of every day, 50% to every second day, 33% to every 3rd day). Other solutions include all non-true critical points (i.e., saddle and local minima). All parameter estimates rounded to 2.d.p, fitted RSS values to 3.d.p.

A-3 Parameter estimate distributions (fitness-delay model)

Table A-3: Parameter estimate and RSS distributions of solutions obtained for the fitness-delay model

Scenario	Convergence	Summary statistics	p^*	k_g	τ_{g_1}	τ_{g_2}	k_h	τ_h	RSS$_{solutions}$	RSS$_{initial}$
Fitness-delay model **100% data**	**True parameters** **(N = 20588)**	Min	100.00	0.72	32.45	4.29	1.05	8.55	0.000	544
		Max	100.00	0.72	32.56	4.32	1.05	8.65	0.001	35463988
		Median	100.00	0.72	32.50	4.30	1.05	8.60	0.000	897537
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	0.000	1180419
	Other solutions **(N = 96717)**	Min	99.94	0.65	1.04	2.06	0.71	1.00	0.010	143
		Max	121.18	5.00	44.70	19.29	1.97	50.00	4380.976	41612914
		Median	100.04	1.53	32.16	6.17	1.05	21.87	0.121	1060962
		M.A.D.	0.04	0.17	5.94	0.53	0.00	25.09	0.164	1454207
	Abnormal termination **(N = 344)**	Min	100.01	0.74	27.86	5.65	1.04	4.84	0.011	2648
		Max	100.10	1.69	44.69	10.45	1.05	50.00	0.598	15290067
		Median	100.09	1.68	32.23	6.36	1.04	31.11	0.457	1090636
		M.A.D.	0.01	0.02	5.72	0.06	0.00	11.20	0.130	1479867
Fitness-delay model **(50% data)**	**True parameters** **(N = 20651)**	Min	100.00	0.72	32.43	4.28	1.05	8.54	0.000	273
		Max	100.00	0.72	32.57	4.32	1.05	8.66	0.000	18983885
		Median	100.00	0.72	32.50	4.30	1.05	8.60	0.000	451439
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	0.000	592054
	Other solutions **(N = 96184)**	Min	99.94	0.65	1.04	2.06	0.71	1.00	0.000	143
		Max	121.18	5.00	44.70	19.29	1.97	50.00	4380.976	41612914
		Median	100.04	1.53	32.16	5.81	1.05	21.86	0.121	1060179
		M.A.D.	0.04	0.25	5.94	0.90	0.00	19.65	0.164	1453494
	Abnormal termination **(N = 157)**	Min	99.96	0.73	19.34	5.64	0.80	4.88	0.006	539
		Max	100.48	5.00	41.40	14.23	1.06	45.27	37.537	8804668
		Median	100.06	1.65	30.65	6.29	1.04	26.70	0.181	540733
		M.A.D.	0.02	0.06	2.32	0.14	0.00	5.42	0.144	743087
Fitness-delay model **(33% data)**	**True parameters** **(N = 21065)**	Min	100.00	0.72	32.41	4.28	1.05	8.53	0.000	181
		Max	100.00	0.73	32.58	4.32	1.05	8.68	0.001	10781682
		Median	100.00	0.72	32.50	4.30	1.05	8.60	0.000	307321
		M.A.D.	0.00	0.00	0.00	0.00	0.00	0.00	0.000	403900
	Other solutions **(N = 96486)**	Min	94.31	0.65	1.00	1.78	0.01	1.78	0.004	46
		Max	119.93	5.00	44.71	30.11	1.94	50.00	1514.567	13746207
		Median	100.04	1.51	32.12	6.16	1.05	21.53	0.042	348259
		M.A.D.	0.04	0.26	6.02	0.59	0.00	24.62	0.058	476901
	Abnormal termination **(N = 98)**	Min	100.01	0.74	27.83	5.63	1.05	4.87	0.004	2255
		Max	100.11	1.69	38.48	10.33	1.05	40.99	0.227	7949920
		Median	100.09	1.66	30.37	6.24	1.05	27.20	0.154	429570
		M.A.D.	0.02	0.05	1.97	0.15	0.00	4.25	0.083	593110

M.A.D refers to the median absolute deviation. Data % refers to the proportion of data used in the fitting process (i.e., 100% corresponds to a measurement frequency of every day, 50% to every second day, 33% to every 3rd day). Other solutions include all non-true critical points (i.e., saddle and local minima). All parameter estimates rounded to 2.d.p, fitted RSS values to 3.d.p.

A-4 Unique solutions (standard model)

Below are tables of the top 10 highest frequency solutions (to 1.d.p for p^*, τ; 2.d.p for k) found across the searches applying the standard model. The complete set of unique solutions for each scenario are available in spreadsheet form at the following repository link, but to conserve space are not copied in entirety here.

Proportion of fitting data	Total unique solutions (N)	Link (to all solutions)
100%	353	
50%	275	github.com/bsh2/experiments/qNewtonNLS/SF-1.xlsx
33%	275	

100% Fitting Data

Table A-4A: Top 10 highest frequency solutions (standard model, 100% fitting data)

| p^*	k_g	τ_{g_1}	k_h	τ_h	Type	Frequency	RSS
100.2	4.6	18.8	5	15.4	saddle	25838	21.42
117	3.8	1.2	5	1.9	minimum	1933	2896.74
94	5	50	4.7	49.7	saddle	164	395.04
93.4	0.34	50	0.01	50	saddle	88	406.15
93.4	0.34	50	0.01	48.7	saddle	77	405.69
93.4	0.34	50	0.01	48.6	saddle	76	405.65
93.4	0.34	50	0.01	48.9	saddle	75	405.76
93.4	0.34	50	0.01	49.9	saddle	64	406.12
93.4	0.34	50	0.01	49.8	saddle	59	406.08
93.4	0.34	50	0.01	48.8	saddle	58	405.72

50% Fitting Data

Table A-4B: Top 10 highest frequency solutions (standard model, 50% fitting data)

| p^*	k_g	τ_{g_1}	k_h	τ_h	Type	Frequency	RSS
100.1	4.61	18.9	5.00	15.5	saddle	26359	8.52
116.5	3.65	1.3	5.00	1.9	minimum	1922	1487.52
94.1	5.00	50.0	4.70	49.6	saddle	177	191.12
93.4	0.34	50.0	0.01	50.0	saddle	94	194.69
93.4	0.34	50.0	0.01	48.7	saddle	73	194.41
93.4	0.34	50.0	0.01	49.8	saddle	71	194.65
93.4	0.34	50.0	0.01	48.8	saddle	69	194.43
93.4	0.34	50.0	0.01	48.9	saddle	65	194.45
93.4	0.34	50.0	0.01	48.5	saddle	56	194.37
93.4	0.34	50.0	0.01	49.9	saddle	55	194.67

33% Fitting Data

Table A-4C: Top 10 highest frequency solutions (standard model, 33% fitting data)

p^*	k_g	τ_{g_1}	k_h	τ_h	Type	Frequency	RSS
100.2	4.6	18.8	5	15.4	saddle	14394	6.24
100	0.72	28.5	1.2	8.6	minimum	12798	0.00
117	3.8	1.2	5	1.9	minimum	1834	1018.39
94	5	50	4.7	49.7	saddle	115	150.38
93.4	0.34	50	0.01	50	saddle	48	154.93
117	3.81	1.2	5	1.9	minimum	44	1019.40
93.4	0.34	50	0.01	48.9	saddle	43	154.83
93.4	0.34	50	0.01	48.6	saddle	43	154.81
93.4	0.34	50	0.01	48.7	saddle	41	154.82
93.4	0.34	50	0.01	49.9	saddle	40	154.92

100% Fitting Data
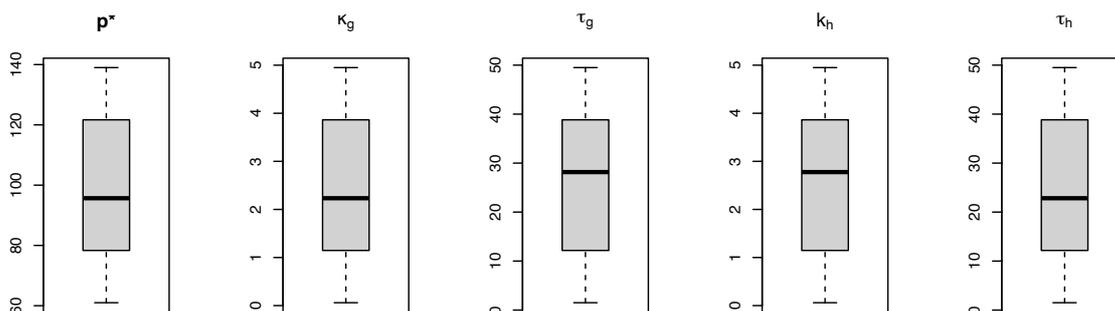
50% Fitting Data
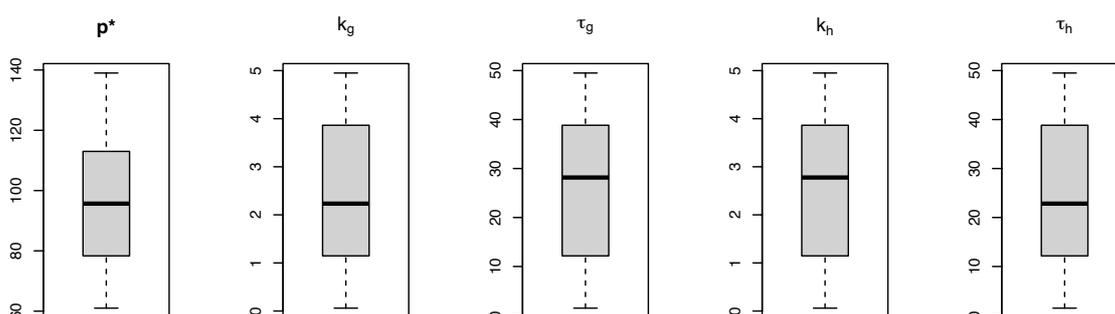
33% Fitting Data
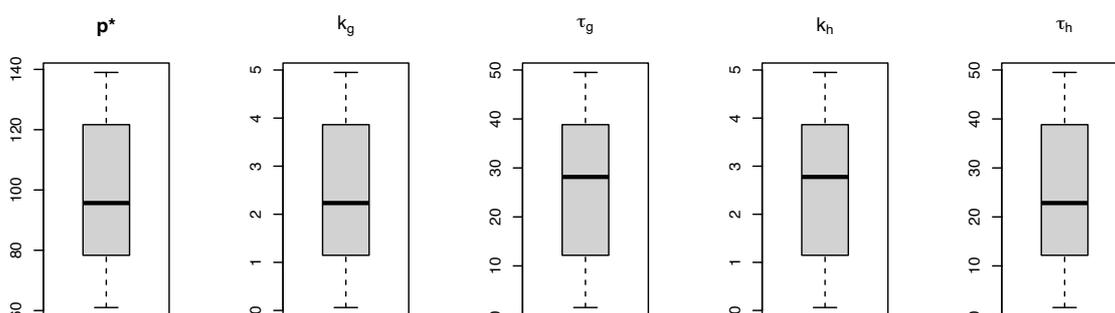
Figure A-4.1: Starting value (initial solution) distributions across iterations that successfully reached the true parameters in the standard model scenarios
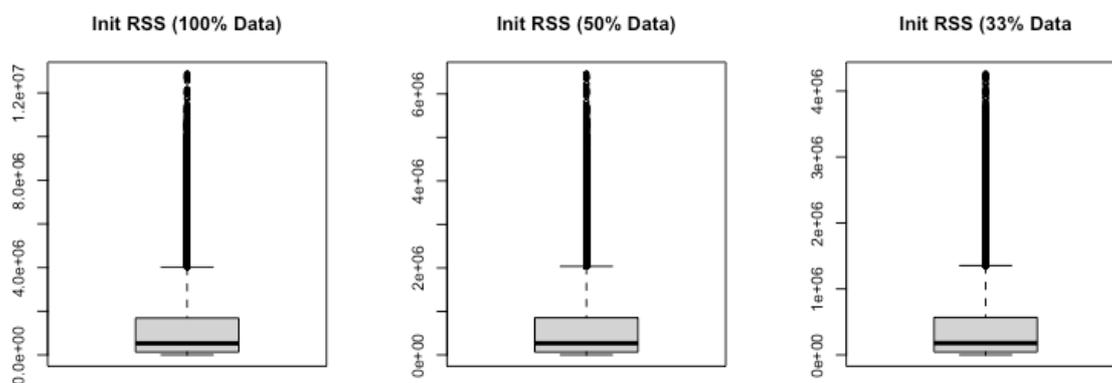
Figure A-4.2: RSS values associated with initial solution distributions across iterations that successfully reached the true parameters in the standard model scenarios.

A-5 Unique solutions (fitness-delay model)

Below are tables of the top 10 highest frequency solutions (to 1.d.p for p^*, τ; 2.d.p for k) found across the searches applying the fitness-delay model. The complete set of unique solutions for each scenario are available in spreadsheet form at the following repository link, but to conserve space are not copied in entirety here.

Proportion of fitting data	Total unique solutions (N)	Link (repository)
100%	383	github.com/bsh2/experiments/qNewtonNLS/SF-1.xlsx
50%	504	
33%	550	

100% Fitting Data

Table A-5A: Top 10 highest frequency solutions (fitness-delay model, 100% fitting data)

| p^*	k_g	τ_{g_1}	τ_{g_2}	k_h	τ_h	Type	Frequency	RSS
100	0.74	32.2	10	1.05	4.9	minimum	39401	0.10
100	1.53	28.1	5.8	1.05	21.9	minimum	33266	1.03
100	1.53	28.2	5.8	1.05	21.9	minimum	5811	0.94
100.1	1.64	44	6.2	1.04	49	minimum	4380	2.03
121.2	5	1.2	2.1	1.97	2.1	minimum	1519	4417.29
100	1.52	28.1	5.8	1.05	21.8	minimum	1505	7.01
100	1.52	28.1	5.8	1.05	21.9	minimum	953	15.93
100.1	1.64	44.7	6.1	1.04	50	minimum	940	30.56
100	5	22.4	19.3	1.06	4.1	saddle	621	8.27
100.7	5	19.2	14	0.82	19.2	saddle	613	98.30

50% Fitting Data

Table A-5B: Top 10 highest frequency solutions (fitness-delay model, 50% fitting data)

| p^*	k_g	τ_{g_1}	τ_{g_2}	k_h	τ_h	Type	Frequency	RSS
100	0.74	32.2	10	1.05	4.9	minimum	34291	0.05
100	1.53	28.1	5.8	1.05	21.9	minimum	31140	0.52
100	0.72	32.5	4.3	1.05	8.6	minimum	8089	0.00
100	1.53	28.2	5.8	1.05	21.9	minimum	5434	0.47
100.1	1.64	44	6.2	1.04	49	minimum	4320	1.02
121.2	5	1.2	2.1	1.97	2.1	minimum	1510	2298.04
100	1.52	28.1	5.8	1.05	21.8	minimum	1411	3.53
100.1	1.64	44.7	6.1	1.04	50	minimum	924	15.31
100	1.52	28.1	5.8	1.05	21.9	minimum	894	8.01
100	5	22.4	19.3	1.06	4.1	saddle	604	4.16

33% Fitting Data

Table A-5C: Top 10 highest frequency solutions (fitness-delay model, 33% fitting data)

p^*	k_g	τ_{g_1}	τ_{g_2}	k_h	τ_h	Type	Frequency	RSS
100	0.74	32.1	10	1.05	4.9	minimum	38922	0.12
100	1.51	28.1	5.8	1.05	21.6	minimum	17384	2.95
100	1.51	28.1	5.8	1.05	21.5	minimum	12086	0.98
100	1.52	28.1	5.8	1.05	21.6	minimum	8634	0.11
100.1	1.64	44.7	6.1	1.05	50	minimum	4099	0.25
119.9	5	1	1.8	1.94	1.8	minimum	1636	1515.64
100	1.51	28	5.8	1.05	21.5	minimum	1332	3.75
100.1	5	19.4	14.4	0.76	19.4	saddle	913	19.93
99.9	5	22.3	19.2	1.03	4.2	saddle	587	2.33
100	1.52	28.1	5.8	1.05	21.7	minimum	409	0.59

100% Fitting Data
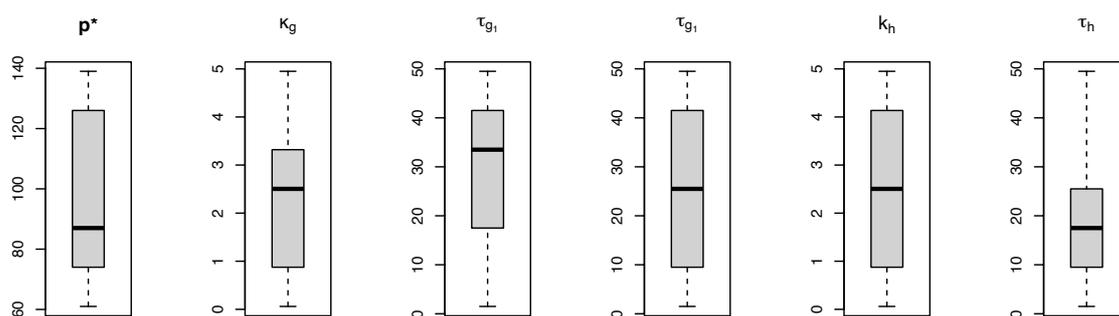
50% Fitting Data
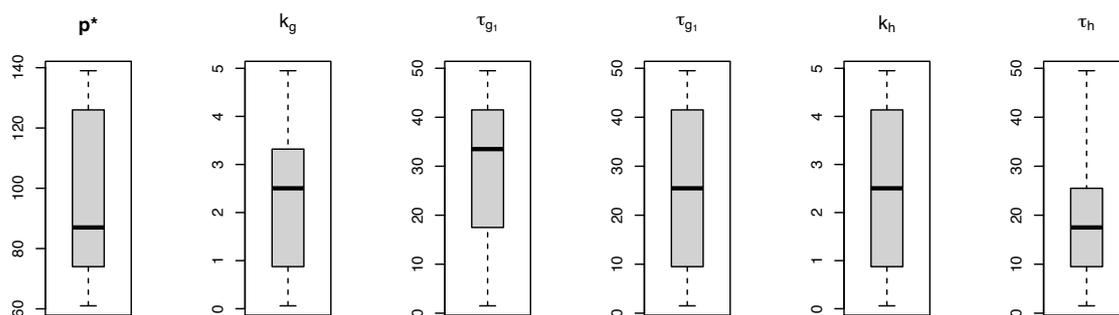
33% Fitting Data
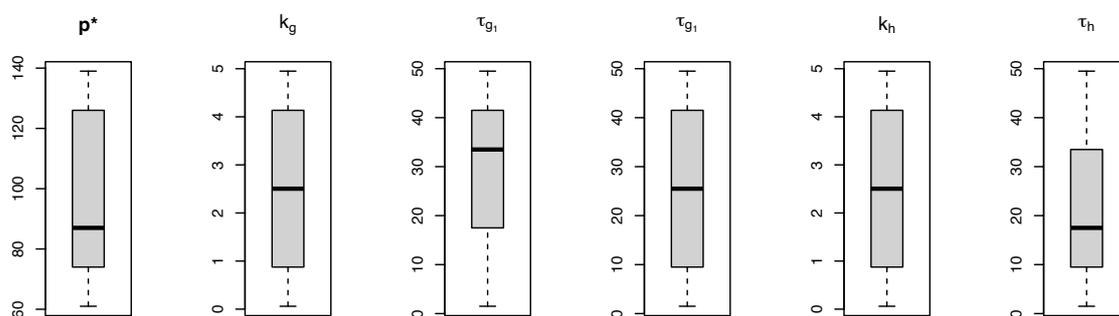
Figure A-5.1: Starting value (initial solution) distributions across iterations that successfully reached the true parameters in the fitness-delay model scenarios
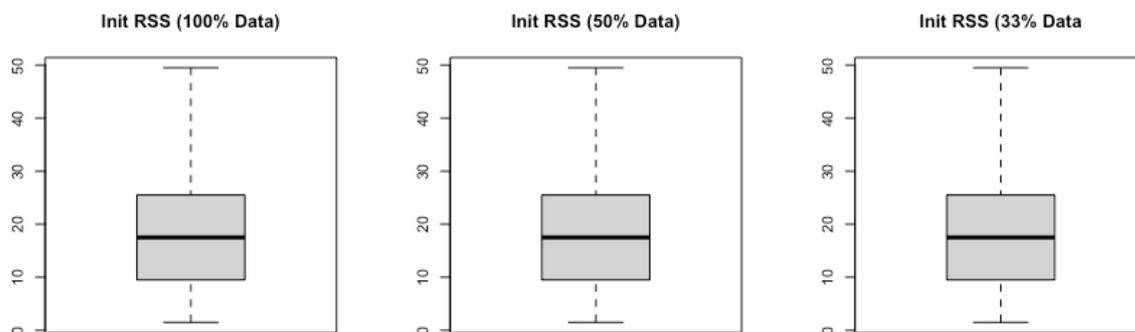
Figure A-5.2: RSS values associated with initial solution distributions across iterations that successfully reached the true parameters in the fitness-delay model scenarios.