University of Kent

Kent Academic Repository

Griffin, Jim E., Spyropoulou, Maria-Zafeiria and Hopker, James G. (2026) *Modelling between- and within-season trajectories in elite athletic performance data*. Journal of the Royal Statistical Society Series C: Applied Statistics . ISSN 0035-9254.

Downloaded from
https://kar.kent.ac.uk/112975/ The University of Kent's Academic Repository KAR

The version of record is available from
https://doi.org/10.1093/jrsssc/qlag002

This document version
Publisher pdf

DOI for this version

Licence for this version
CC BY (Attribution)

Additional information

Versions of research works

Versions of Record
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries
If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Journal of the Royal Statistical Society Series C: Applied Statistics, 2026, **00**, 1–15
https://doi.org/10.1093/jrsssc/qlag002

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society Series C Applied Statistics

Original Article

Modelling between- and within-season trajectories in elite athletic performance data

Jim E. Griffin[1], Maria-Zafeiria Spyropoulou[2] and James Hopker[2]

[1]Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK
[2]School of Sport and Exercise Sciences, University of Kent, Canterbury, Kent CT2 7NZ, UK

Address for correspondence: Jim E. Griffin, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK. Email: j.griffin@ucl.ac.uk

Abstract

Athletic performance follows a typical pattern of improvement and decline during a career. This pattern is also often observed within-seasons, as an athlete aims for their performance to peak at key events such as the Olympic Games or World Championships. A Bayesian hierarchical model is developed to analyse the evolution of athletic sporting performance throughout an athlete's career and separate these effects whilst allowing for confounding factors such as environmental conditions. Our model works in continuous time and estimates both $g(t)$, the average performance level of the population at age t, and $f_i(t)$, the difference of the ith athlete from this average. We further decompose $f_i(t)$ into a season-to-season trajectory and a within-season trajectory, which is modelled by a restricted Bernstein polynomial. The model is fitted using an adaptive Metropolis-within-Gibbs algorithm with a carefully chosen blocking scheme. The model allows us to understand seasonal patterns in athlete performance, how these differ between athletes, and provides individual fitted and trend performance trajectories. The properties of the model are illustrated using a simulation study and an application to 100 and 200 m freestyle swimming for both female and male athletes.

Keywords: Bayesian inference, global-local prior, longitudinal modelling, restricted Bernstein polynomial, splines

1 Introduction

The availability of large databases of elite athlete performance allows the modelling of performance levels over an athlete's career. Results from these models can be used for retrospective analysis (understanding how an athlete's performance level evolved over their career), short-term predictions (such as the results of future events), or long-term predictions (such as talent spotting through prediction of the evolution of performance levels at the start of an athlete's career). Understanding the variation across athletes and time is critical to effectively performing these analyses. We will concentrate on retrospective analysis in centimetre, gram, and seconds (CGS) sports, where performance is measured using one of these units.

Figure 1a shows the performances of an elite female 100 m swimmer from ages 16 to 27 and a curve fitted through the data (which we refer to as the athlete's *individual performance trajectory*). Performance is clearly improving over the athlete's career but there are also annual cyclical patterns. These follow from sports usually being organized into annual seasons with events of high importance (e.g. Olympics or World Championships) occurring at similar times in different years. Many athletes will tailor their training to peak for these events. Figure 1b and c decomposes the individual performance trajectory into two parts. An *individual career trend trajectory* (Figure 1b) which is a linear interpolation of the estimated performance level on the 1st January of each year and an effect for each year/season (Figure 1c), which we call *within-season performance trajectories*. Figure 1b shows the improvement in performance more clearly. Figure 1c shows that on average

Received: May 27, 2024. **Revised:** December 23, 2025. **Accepted:** January 5, 2026
© The Royal Statistical Society 2026.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

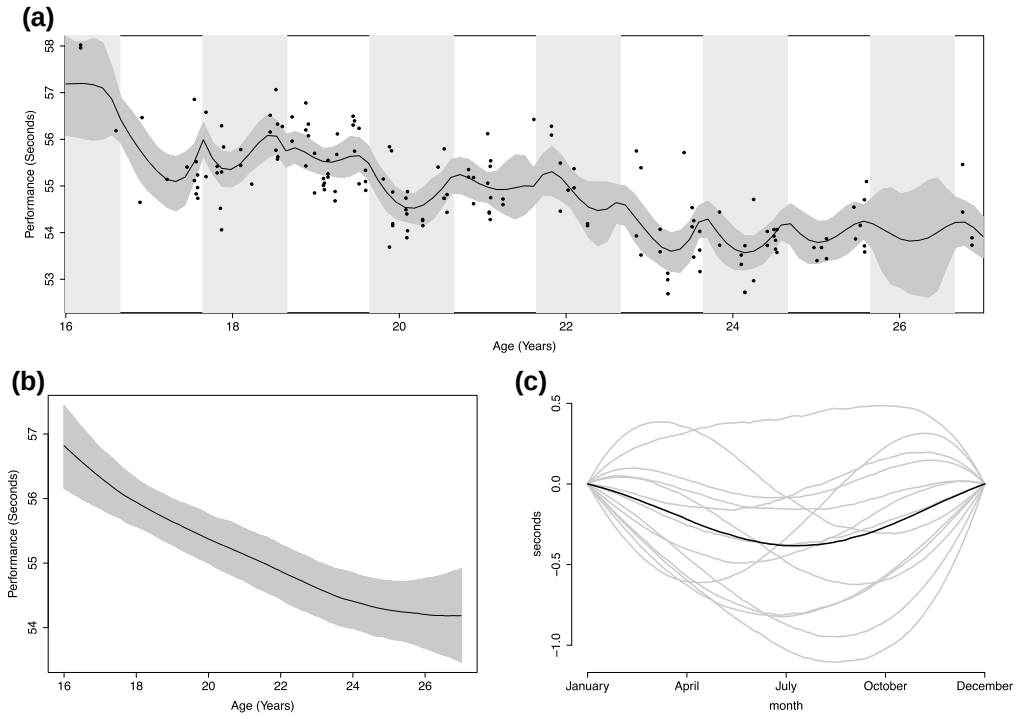

Figure 1. Analysis of a 100 m swimmer. Panel (a) shows the observed performances with individual performance trajectory (shown as posterior median (black line) and 95% credible interval) with each calendar year indicated by alternating grey and white bands. Panel (b) shows the individual career trend trajectory shown as posterior median (black line) and 95% credible interval. Panel (c) shows posterior median within-season performance trajectories (light grey lines) and the posterior median of the athlete's mean within-season performance trajectory (black line).

the swimmer peaks in August with a performance improvement of about 0.4 s compared to January. However, there are also substantial differences in the within-season performance trajectory from season-to-season ranging from 1 second improvement in some season and close to 0 s in others. Being able to estimate the variability of within-season performance trajectories is important for understanding whether an athlete is able to peak at the same time and by the same amount in different years.

Mixed models with scalar athlete and season effects have been used to understand variation within and between seasons in athletic performance (Bullock et al., 2009; Paton & Hopkins, 2005; Pyne et al., 2004; Trewin et al., 2004). However, this often leads to an overestimation of variability between consecutive performances because variability is calculated for all competitions within a season regardless of how the competitions were distributed within a given season (Malcata & Hopkins, 2014). Longitudinal models were originally proposed for sporting performance data by Berry et al. (1999), who build a linear mixed model with a common effect of age (usually called the aging function or population performance trajectory), in addition to scalar athlete and season effects. The population performance trajectory reflects the usual effect of age on athletic performance which typically follows a 'u' shape (Stival et al., 2023) with athlete's improving to a mid-career peak followed by deterioration. The age of peak performance often occurs between the ages 23 and 28 but can differ between sports, gender, and individuals (Griffin et al., 2022).

More recently, continuous-time longitudinal models have been used for the irregular observations in data sets of all athlete performances. Griffin et al. (2022) build a model which allows for time-varying athlete effects and confounders such as meteorological factors (e.g. wind speed or temperature) or geographical factors (e.g. altitude). The observational variation is modelled by a skew-*t* distribution since athletes have a higher chance of underperforming rather than overperforming to the same degree. This model was adequate for the disciplines considered in their

Downloaded from https://academic.oup.com/jrssc/advance-article/doi/10.1093/jrssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

Table 1. Summaries of the number of performances, number of athletes, and number of performances per athlete for the 100 and 200 m freestyle data sets

Event	Number of performances	Number of athletes	Number of performances per athlete		
			Median	Minimum	Maximum
100 m, female	23,669	500	37.5	5	267
100 m, male	23,440	500	38	5	191
200 m, female	21,112	500	33	5	274
200 m, male	19,696	500	32	5	162

paper (weightlifting and 100 m sprint), but does not model seasonal effects. In this paper, we develop this model to allow for changes in performance over a season using a restricted Bernstein polynomial (Wang & Ghosh, 2012). We also increase the flexibility of the model in two additional directions. Firstly, we investigate the use of a distribution which allows for different tail heaviness in each tail. Secondly, we address the sparsity of observations for some athletes using global–local shrinkage priors in a normal hierarchical model. This encourages parts of the hierarchy to be strongly shrunk towards the mean unless the data supports differences.

Similar to our approach is the model introduced by Dolmeta et al. (2023). The authors model performance with a non-linear function of time, a GARCH model for interseason changes and the effects of age, gender, and environmental covariates. Our model uses a more structured, yet flexible, model of the variation within seasons, which allows us to understand both individual and population average intra-seasonal effects.

The paper is organized as follows: *Section 2* discusses the available data for our application to 100 and 200 m freestyle performances for female and male swimmers. *Section 3* explains how the model is formed. Inference is discussed in *Section 4*. *Section 5* describes an application of our model to simulated data and a simulation study. *Section 6* discusses the application of our model to 100 and 200 m freestyle performances for both males and females. Lastly, in *Section 7*, a discussion is provided. The online supplementary material for this paper (Griffin et al., 2026) includes some derivations, the full Markov chain Monte Carlo (MCMC) for Bayesian inference, an exploratory data analysis, MCMC diagnostics for the real data applications, and some additional results. R code to fit the model can be downloaded from https://jimegriffin.github.io/website/.

2 Elite swimming data

We use a large database of performance data from 100 and 200 m freestyle swimming for both females and males to illustrate our methods. We downloaded performance data from 19 March 2008 to 08 September 2023 from the World Aquatics website[1] and selected the 500 swimmers with the fastest personal bests for each combination of distance and gender. Summaries of the number of performances and athletes are provided in Table 1. The data includes performances in both 25 and 50 m pools. Swimmers are able to achieve faster times in a 25 m pool and so pool length is included as a confounder. To make performances comparable when presenting data in the paper, we plot 25 m pool performances adjusted to a 50 m pool using the posterior mean of this pool effect. An exploratory data analysis of this data is presented in online supplementary material Section 3 (Griffin et al., 2026), which shows the effect of ages and seasonal effects within the data.

3 Model

3.1 Sampling model

We construct a continuous-time longitudinal model for athlete performance which allows for effects of age, the time of year, and confounders building on Griffin et al. (2022). Suppose that we have observations for M athletes where the ith athlete has n_i performances $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,n_i})$,

[1] https://www.worldaquatics.com.

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

which are observed at ages $\boldsymbol{a}_i = (a_{i,1}, \ldots, a_{i,n_i})$ and calendar time $\boldsymbol{t}_i = (t_{i,1}, \ldots, t_{i,n_i})$ (where 0 refers to the start of an athlete's first season) with confounders $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,n_i})$. We assume that the ith athlete competes in S_i seasons and define $\boldsymbol{s}_i = (s_{i,1}, \ldots, s_{i,n_i})$ where $s_{i,k} \in \{1, \ldots, S_i\}$ is the season of the kth performance. We assume the basic model

$$y_{i,k} = \mu_i(a_{i,k}, t_{i,k}) + x_{i,k}\,\zeta + \epsilon_{i,k}, \tag{1}$$

where $\mu_i(a, t)$ is the *individual performance trajectory* of the ith athlete at age a and calendar time t (which is defined to be 0 at the start of an athlete's first season), ζ are the effects of the confounders, and the $\epsilon_{i,k}$'s are i.i.d. observation errors. These errors will typically be both skewed and heavy tailed since athletes have larger probabilities of extreme underperformance compared to overperformance. We first describe the distribution of the errors before discussing the form of $\mu_i(a, t)$.

Griffin et al. (2022) assume that the errors follow a skew-t distribution but this assumes the same heaviness for both tails. We consider a distribution which allows differences in the heaviness of the two tails by generalizing the skew-t distribution. We define $T\mathcal{N}_R(0, \mu, \sigma^2)$ to denote the normal distribution with mean μ and variance σ^2 truncated to the region R and $\mathcal{IG}a(a, b)$ to denote an inverse gamma distribution with density $p(x) = \frac{b^a}{\Gamma(a)} x^{-a-1} \exp\{-b/x\}$. The observation error is defined by $\epsilon_{i,k} = \epsilon^{\star}_{i,k} + \frac{\alpha}{\sqrt{1+\alpha^2}}\,\kappa_{i,k}$, where $\epsilon^{\star}_{i,k} \sim \mathcal{N}(0, \omega_{i,k}\,\sigma_i^2)$, $\kappa_{i,k} \sim T\mathcal{N}_{[0,\infty]}(0, \phi_{i,k}\,\sigma_i^2)$, $\omega_{i,k} \sim \mathcal{IG}(\frac{\nu_1}{2}, \frac{\nu_1}{2})$, and $\phi_{i,k} \sim \mathcal{IG}(\frac{\nu_2}{2}, \frac{\nu_2}{2})$ for $k = 1, \ldots, n_i$, $i = 1, \ldots, M$, and all elements are independent. This reverts to a skew-t distribution if $\omega_{i,k} = \phi_{i,k}$ (and so $\nu_1 = \nu_2$). If $\alpha > 0$, the heaviness of the right-hand tail is controlled by the minimum of ν_1 and ν_2 and the heaviness of the left-hand tail by ν_1 with smaller values representing heavier tails (with the effects on left- and right-hand tails reversed if $\alpha < 0$).

The model of the individual performance trajectory $\mu_i(a, t)$ is initially decomposed into population and individual parts,

$$\mu_i(a, t) = g(a) + \theta_i(a, t), \tag{2}$$

where $g(a)$ is the *population performance trajectory* and $\theta_i(a, t)$ is the *individual excess performance*. We follow Griffin et al. (2022) by modelling $g(a)$ using a dth order polynomial function $g(a) = \sum_{k=0}^{d} \delta_k\,(a - \bar{a})^k$, where $\delta_1, \ldots, \delta_d$ are coefficients and $\bar{a}$ is the mean age of all observed performances (we find that $d = 4$ is sufficiently flexible in our examples).

The individual excess performance is further decomposed into two parts,

$$\theta_i(a, t) = f_i(a) + h_i(t), \tag{3}$$

where $f_i(a)$ is called the *trend excess performance trajectory* for the ith athlete and $h_i(t)$ is called the *seasonal performance trajectory*. We define $g(a) + f_i(a)$ to be the ith athlete's *individual trend performance trajectory*, which adjusts the individual performance trajectory $\mu_i(a, t)$ for seasonal effects.

The trend excess performance trajectory is modelled by a piecewise linear function where the knots occur at the start/end of each season,

$$f_i\big((s - 1 + r)\Delta\big) = \eta_{i,s}\,(1 - z) + \eta_{i,s+1}\,z, \quad z \in (0, 1),\ s = 1, \ldots, S_i.$$

where Δ is the length of each season and start on the same day of each year (although this could easily be changed in the model) and $\eta_{i,s}$ is the value at the start of the sth season.

The seasonal performance trajectory is modelled season-by-season. Let the *within-season performance trajectory* for the ith athlete in the sth season be $h^{\star}_{i,s}(z)$ for $0 < z < 1$. We define the seasonal performance trajectory to be

$$h_i((s - 1 + z)\Delta) = h^{\star}_{i,s}(z), \quad z \in (0, 1),\ s = 1, \ldots, S_i.$$

We want a flexible form for $h^{\star}_{i,s}$ which is constrained by $h^{\star}_{i,s}(0) = 0$ and $h^{\star}_{i,s}(1) = 0$. This allows us to identify the model in (3) since

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

$$\theta_i(a, s\Delta) = f_i(a) + h_i(s\Delta) = f_i(a) + h^\star_{i,s}(1) = f_i(a) + h^\star_{i,s+1}(0) = f_i(a) = \eta_{i,s+1}.$$

Therefore, the individual excess performance $\theta_i(a, s\Delta)$ is equal to the trend excess performance $f_i(a)$ at the start and end of each season.

3.2 Prior distributions for $f_i(a)$ and $h^\star_{i,s}(r)$

We wish to have flexible forms for both the individual trend excess performance $f_i(a)$ and the within-season performance trajectories $h^\star_{i,s}(r)$. The values of f_i at the start of each season are $\eta_{i,1}, \ldots, \eta_{i,S_i+1}$. These are modelled using a random walk with initial value $\frac{\eta_{i,1}}{\sigma^2_\mu} \sim t_{\nu^\mu}$ and increments $\frac{\eta_{i,s+1}-\eta_{i,s}}{\sigma^2_\eta} \overset{i.i.d.}{\sim} t_{\nu^\eta}$ for $s = 1, \ldots, S_i$. This specification allows for heavy tails in both the distribution of the initial level $\eta_{i,1}$ and possibly large changes from season to season.

We want the functions $h^\star_{i,s}$ to have some specific features. Firstly, for identifiability, the functions $h^\star_{i,s}$ are constrained by $h^\star_{i,s}(0) = 0$ and $h^\star_{i,s}(1) = 0$. Secondly, we want a hierarchical structure to allow sharing of information across different seasons for each athletes and between athletes. This allows the identification of those athletes who are better able to peak and those who can achieve this consistently. Thirdly, we assume that the seasonal variation is largely caused by athletes peaking for particular events or parts of the year (for example, the summer season in athletics) and that this is shared across athletes (i.e. athletes are peaking for events in the same point of the year).

A Bayesian hierarchical model built using Restricted Bernstein polynomials (RBPs) (Wang & Ghosh, 2012) is a convenient way to achieve our three goals. An nth order RBP $m(z)$ with $m_n(0) = m_n(1) = 0$ has the form

$$m_n(z) = \sum_{\nu=1}^{n-1} \beta_{n,\nu}\, b_{n,\nu}(z), \quad 0 < z < 1, \tag{4}$$

where

$$b_{n,\nu}(z) = \binom{n}{\nu} z^\nu (1-z)^{n-\nu}, \quad \nu = 1, \ldots, n-1.$$

To provide additional flexibility, we use the sum of RBPs of order 2 to N giving the model

$$h^\star_{i,s}(z) = \sum_{n=2}^{N} \sum_{\nu=1}^{n-1} \beta^{(i,s)}_{n,\nu}\, b_{n,\nu}(z), \quad 0 < z < 1. \tag{5}$$

It is convenient to group the coefficients of the RBP in $\boldsymbol{\beta}^{(i,s)} = (\beta^{(i,s)}_{2,1}, \beta^{(i,s)}_{3,1}, \beta^{(i,s)}_{3,2}, \ldots, \beta^{(i,s)}_{N,1}, \ldots, \beta^{(i,s)}_{N,N-1})$.

To allow sharing of information across seasons and athletes, we use independent hierarchical priors for each coefficient in the RBP, which nest seasons within athletes. This leads to, for $i = 1, \ldots, M$ and $s = 1, \ldots, S_i$,

$$\beta^{(i,s)}_{n,\nu} \overset{i.i.d.}{\sim} \mathcal{N}\left(\beta^{(i)}_{n,\nu}, \lambda_i^2\, c^2_{n,\nu}\right), \quad \beta^{(i)}_{n,\nu} \overset{i.i.d.}{\sim} \mathcal{N}\left(\beta_{n,\nu}, \tau_i^2\, d^2_{n,\nu}\right), \quad \nu = 1, \ldots, n-1, \;\; n = 2, \ldots, N.$$

The prior variances are a product of athlete-specific effects λ_i^2 and τ_i^2, and coefficient-specific effects $c^2_{n,\nu}$ and $d^2_{n,\nu}$. The use of coefficient-specific variance parameters allows for the variability of the functions $h^\star_i(z)$ and $h^\star(z)$ to change with z. The use of individual-specific variance parameters λ_i^2 and τ_i^2 allows for differences in the variability of the functions from athlete-to-athlete. The importance of allowing for differences is discussed in *Section 4*. In addition, since some athletes only have a few performances, we use global–local shrinkage priors (Bhadra et al., 2019) which avoids

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

overfitting and allows the effects to be shrunk towards the corresponding mean. Since the scale parameters $c^2_{n,v}$ and λ^2_i enter multiplicatively (similarly, $d^2_{n,v}$ and τ^2_i), we centre $c^2_{n,v}$'s and $d^2_{n,v}$'s around 1 to avoid problems of interpretation, and use gamma priors to encourage regularization of the coefficients (see Griffin & Brown, 2010),

$$c^2_{n,v} \sim \mathcal{G}a(5,5), \quad d^2_{n,v} \sim \mathcal{G}a(5,5), \quad v = 1, \ldots, n-1, \; n = 2, \ldots, N,$$
$$\lambda^2_i \sim \mathcal{G}a(\lambda_0, \lambda_0/\lambda_1), \quad \tau^2_i \sim \mathcal{G}a(\tau_0, \tau_0/\tau_1), \quad i = 1, \ldots, M$$

where $\mathcal{G}a(a, b)$ represents a gamma distribution with density $p(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp\{-b\,x\}$ and so λ_1 and τ_1 are the prior means of λ^2_i and τ^2_i, respectively.

The hierarchical structure introduces parameters $\beta^{(i)} = (\beta^{(i)}_{2,1}, \beta^{(i)}_{3,1}, \beta^{(i)}_{3,2}, \ldots, \beta^{(i)}_{N,1}, \ldots, \beta^{(i)}_{N,N-1})$ and $\beta = (\beta_{2,1}, \beta_{3,1}, \beta_{3,2}, \ldots, \beta_{N,1}, \ldots, \beta_{N,N-1})$, which can be interpreted as average coefficients over all seasons for the ith athlete and average coefficients over all seasons and all athletes, respectively. It is convenient to define functions formed by the coefficients β,

$$h^\star(z) = \sum_{n=2}^{N} \sum_{v=1}^{n-1} \beta_{n,v}\, b_{n,v}(z), \quad 0 < z < 1,$$

is called the *average within-season performance trajectory at the population level*, and by the coefficients $\beta^{(i)}$,

$$h^\star_i(z) = \sum_{n=2}^{N} \sum_{v=1}^{n-1} \beta^{(i)}_{n,v}\, b_{n,v}(z), \quad 0 < z < 1.$$

is called the *average within-season performance trajectory for the* i*th individual*.

Thirdly, the observation that athletes often peak at a particular time is included by constraining the within-season performance trajectory at the population level to have a single peak (rather than at the individual or seasonal level). This allows for differences in peak time (or the number of peaks) for some athletes and some seasons, which allows athletes to have different goals in some seasons (for example, in Olympic and non-Olympic years).

The shape of the population-level within-season performance trajectory is determined by the direction of improvement in the sport. In timed events such as swimming or running, the direction of improvement is negative since a better performances leads to a faster time whereas, in events such as weightlifting or throwing, the direction of improvement is positive since a better performance leads to a heavier lift or a longer throw. Therefore, we constrain the trajectory to be concave if the direction of improvement is positive and convex if the direction of improvement is negative. As discussed by Wang and Ghosh (2012), the second derivative of the RBP in (4) can be expressed as

$$\frac{m''_n(z)}{n(n-1)} = \left(\beta_{n,2} - 2\beta_{n,1}\right) b_{n-2,0}(z) + \sum_{v=1}^{n-2} \left(\beta_{n,v+2} - 2\beta_{n,v+1} + \beta_{n,v}\right) b_{n-2,v}(z) + \left(\beta_{n,n-2} - 2\beta_{n,n-1}\right) b_{n-2,n-2}(z).$$

Let $\beta_n = (\beta_{n,1}, \ldots, \beta_{n,n-1})$ then $m_n(z)$ is convex if $\boldsymbol{D}_n\boldsymbol{\beta}_n \geq 0_{n-1}$ and concave if $\boldsymbol{D}_n\boldsymbol{\beta}_n \leq 0_{n-1}$ where

$$\boldsymbol{D}_n = \begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 \end{pmatrix}_{(n-1)\times(n-1)},$$

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

and $\mathbf{0}_{n-1}$ represents an $(n-1)$-dimensional vector of 0's. Therefore, if the direction of improvement is positive, the prior is a normal distribution with mean $\mathbf{0}_{n-1}$ and covariance matrix $\text{diag}(\sigma^2_{\boldsymbol{\beta},n})$ truncated to the region $\{\boldsymbol{\beta}_n \mid \boldsymbol{D}_n \boldsymbol{\beta}_n \leq \mathbf{0}_{n-1}\}$, where $\text{diag}(\boldsymbol{a})$ represents a diagonal matrix with the elements of $\boldsymbol{a}$ on the diagonal, and $\sigma^2_{\beta,n}$ is an $(n-1)$-dimensional vector of variance parameters. If the direction of improvement is negative, the prior is truncated to the region $\{\boldsymbol{\beta}_n \mid \boldsymbol{D}_n \boldsymbol{\beta}_n \geq \mathbf{0}_{n-1}\}$. We apply the constraint for each value of n in the sum of RBPs prior for $h^{\star}_{i,s}(z)$ to ensure that the sum is either convex or concave.

3.3 Priors for other parameters

The Bayesian model is completed by specifying priors for all other parameters. Following Griffin and Brown (2010), we centre the shape parameters λ_0 and τ_0 on 1 and give vague priors to the means λ_1 and τ_1 giving $\lambda_0, \tau_0 \overset{i.i.d.}{\sim} \mathcal{E}x(1)$, where $\mathcal{E}x(\theta)$ represents an exponential distribution with mean $\frac{1}{\theta}$ and $\lambda_1, \tau_1 \overset{i.i.d.}{\sim} \mathcal{IG}(0.001, 0.001)$. The scale of the observation error for each athlete is given the population distribution $\sigma^2_i \sim \mathcal{IG}(\sigma^2_a, \sigma^2_a/\sigma^2_m)$, and the skewness parameter is given the prior $\alpha \sim \mathcal{N}(0, 3^2)$, which provides support for reasonable values of the skewness. The degrees of freedom parameters are given a prior mean of 20 and wide spread of possible values $\nu^{\mu} \sim \mathcal{G}a(2, 0.1)$, $\nu^{\eta} \sim \mathcal{G}a(2, 0.1)$, $\nu_1 \sim \mathcal{G}a(2, 0.1)$, and $\nu_2 \sim \mathcal{G}a(2, 0.1)$. All other parameters are given vague priors $p(\zeta, \delta) \propto 1$, $\sigma^2_{\eta} \sim \mathcal{IG}(0.001, 0.001)$, $\sigma^2_a \sim \mathcal{IG}(0.001, 0.001)$, $\sigma^2_m \sim \mathcal{IG}(0.001, 0.001)$, and $\sigma^2_{\mu} \sim \mathcal{IG}(0.001, 0.001)$.

4 Inference

We use MCMC to fit the model and the full algorithm is described in online supplementary material Section 2 (Griffin et al., 2026). Blocking (Knorr-Held & Rue, 2002; Roberts & Sahu, 1997) and interweaving (Yu & Meng, 2011) are used to avoid slow mixing of some parameters.

The hierarchical model for the within-season performance trajectory allows us to estimate functional effects at the season, athlete, and population level. Figure 1c plots the within-season performance trajectory at the individual and seasonal levels. This shows substantial variation in the within-season performance trajectories for this individual. A univariate summary of the amount of variation is useful to compare athletes. Similarly, we define a univariate summary of the difference between an athlete's within-season performance trajectory and the population within-season performance trajectory. To establish these summaries, we use the following results for the RBP in (5) (a proof is given in the online supplementary material Section 1 Griffin et al., 2026). Suppose that $\epsilon(z) = \sum_{n=2}^{N} \sum_{v=1}^{n-1} a_{n,v}\, b_{n,v}(z)$ then

(a) $\int_0^1 \epsilon(z)\, dz = \sum_{n=2}^{N} \frac{1}{n+1} \sum_{v=1}^{n-1} a_{n,v}$,
(b)

$$\int_0^1 \epsilon(z)^2\, dz = \sum_{n_1=2}^{N} \sum_{v_1=1}^{n_1-1} \sum_{n_2=2}^{N} \sum_{v_2=1}^{n_2-1} a_{n_1,v_1}\, a_{n_2,v_2}\, B_{n_1,n_2,v_1,v_2},$$

where

$$B_{n_1,n_2,v_1,v_2} = \binom{n_1}{v_1}\binom{n_2}{v_2} \frac{(v_1+v_2)!(n_1+n_2-v_1-v_2)!}{(n_1+n_2+1)!}.$$

The variability over different seasons for the ith athlete can be measured by considering the differences between the within-season performance trajectory for the sth season and the average within-season performance trajectory for the ith athlete which is

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

$$\psi_i(z) = h^{\star}_{i,s}(z) - h^{\star}_i(z) = \sum_{n=2}^{N}\sum_{v=1}^{n-1}\left(\beta^{(i,s)}_{n,v} - \beta^{(i)}_{n,v}\right) b_{n,v}(z).$$

Taking the expectation of $\int_0^1 \psi_i^2(z)\,dz$ with respect to the prior distribution of $\beta^{(i,j)} - \beta^{(i)}$ leads to the expression

$$\Psi_i = \mathrm{E}\left[\int_0^1 \psi_i^2(z)\,dz\right] = \lambda_i^2 \sum_{n=2}^{N}\sum_{v=1}^{n-1} c^2_{n,v}\, B_{n,n,v,v},$$

which we call the within-season variability for the ith athlete. In a similar way, we can summarize the size of the difference in effect between an athlete's average within-season performance trajectory and the average within-season performance trajectory across all athletes using the measure

$$\gamma_i(z) = h^{\star}_i(z) - h^{\star}(z) = \sum_{n=2}^{N}\sum_{v=1}^{n-1} (\beta^{(i)}_{n,v} - \beta_{n,v})\, b_{n,v}(z).$$

Taking the expectation of $\int_0^1 \gamma_i^2(z)\,dz$ with respect to the prior distribution of $\beta^{(i)} - \beta$ leads to the expression

$$\Gamma_i = \mathrm{E}\left[\int_0^1 \gamma^2(z)\,dz\right] = \tau_i^2 \sum_{n=2}^{N}\sum_{v=1}^{n-1} d^2_{k,v}\, B_{n,n,v,v}$$

which we call the average effect size for the ith athlete. We are often interested in the ordering of each of these measures across athletes to find athletes which have small or large variability in their season effects or athletes are far from the population average. This can be achieved by just report λ_i^2 (in place of Ψ_i) and τ_i^2 (in place of Γ_i).

5 Simulated data

We use a simulated example and a simulation study to show how the model can capture individual career trend trajectories, within-season performance trajectories at the seasonal, individual and population levels, and differences in variation and the seasonal and individual levels. We generated data from the full model in *Section 3* without confounders. The ith individual has observation for S_i seasons were

$$S_i \sim \begin{cases} \mathcal{P}o(4) & \text{with probability } p_1, \\ \mathcal{P}o(8) & \text{with probability } 1 - p_1. \end{cases}$$

The parameter p_1 controls the proportion of athletes with more or less observed seasons. A larger value of p_1 leads to more athletes with a lower number of observed seasons. In the jth season, there were $n_{i,j} \sim \mathcal{U}(3, 11)$ performances. The age at the start of the first season was assumed to be distributed $\mathcal{U}(18, 22)$ and the population performance trajectory was $g(t) = 40 + 0.1(t - 26)^2$. The excess performances $\eta_{i,1}, \ldots, \eta_{i,S_i+1}$ were generated as $\eta_{i,1} \sim \mathcal{N}(0, 4)$ and $\eta_{i,j+1} - \eta_{i,j} \sim \mathcal{N}(0, 0.09)$ for $j = 1, \ldots, S_i$. The population-level within-season performance trajectory was $h^{\star}(t) = 4(t - \frac{1}{2})^2 - 1$ for $0 < t < 1$, which has a minimum at $t = \frac{1}{2}$. The individual and seasonal performance trajectories were parabolas with a maximum or minimum at different values for different individuals or seasons. The turning point for the individual trajectory was $p_i \overset{i.i.d.}{\sim} \mathcal{U}(0.5, 0.7)$ and the value at that point was $a_i \sim \mathcal{N}(0, \sigma_a^2)$. The trajectory was

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

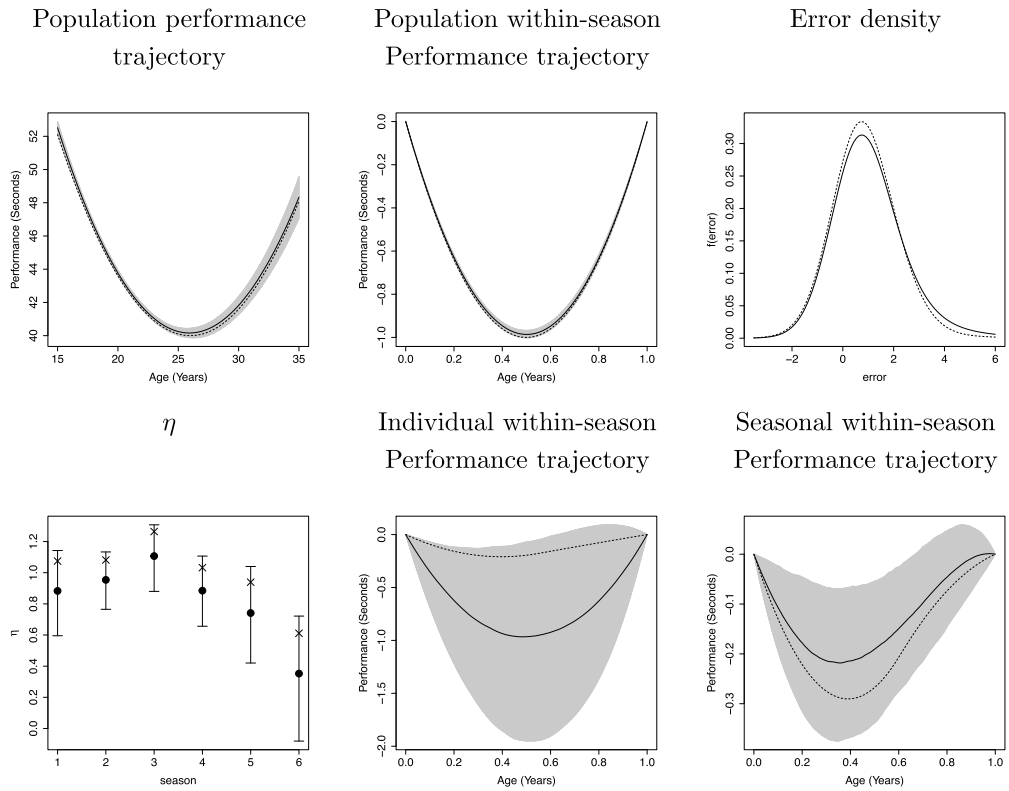

Figure 2. Results for the simulated example. Results for Population performance trajectory, Population, Individual and Seasonal within-season performance trajectories are shown as true value (dotted line), posterior median (solid line), and 95% credible interval. Results for the error density are shown as true value (dotted line) and posterior mean (solid line). The results for η represent $\eta_{i,1}, \ldots, \eta_{i,6}$ for one athlete and are shown as true value (block dot), posterior median (cross) with 95% credible interval.

$$h_i^\star(t) - h^\star(t) = \begin{cases} a_i\left(1 - \frac{(t-p_i)^2}{p_i^2}\right), & 0 < t \le p_i, \\ a_i\left(1 - \frac{(t-p_i)^2}{1-p_i^2}\right), & p_i < t < 1. \end{cases} \tag{6}$$

The turning point for the individual trajectory was $r_{i,j} \overset{i.i.d.}{\sim} \mathcal{U}(0.5, 0.7)$ and the value at that point was $b_{i,j} \sim \mathcal{N}(0, \sigma_b^2)$. The trajectory was

$$h_{i,j}^\star(t) - h_i^\star(t) = \begin{cases} b_{i,j}\left(1 - \frac{(t-r_{i,j})^2}{r_{i,j}^2}\right), & 0 < t \le r_{i,j}, \\ b_{i,j}\left(1 - \frac{(t-r_{i,j})^2}{1-r_{i,j}^2}\right), & r_{i,j} < t < 1. \end{cases} \tag{7}$$

The errors were generated with $\alpha = 3$, $\nu_1 = 30$, and $\nu_2 = 7$.

We simulated a data set of 500 individuals with $p_1 = 0.2$, $\sigma_a^2 = 0.5$, and $\sigma_b^2 = 0.25$. The results are shown in Figure 2. The population performance trajectory and the population within-season performance trajectory and error distribution are well-estimated. Estimates at the individual level are unsurprisingly less well-estimated. The true values of the individual trend performance trajectory (η) and individual and seasonal within-season performance trajectories are all within the corresponding 95% credible intervals.

To quantify the performance of the model, we conducted a simulation study. We use 100 replications for eight different combinations generated by varying four factors over two levels:

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

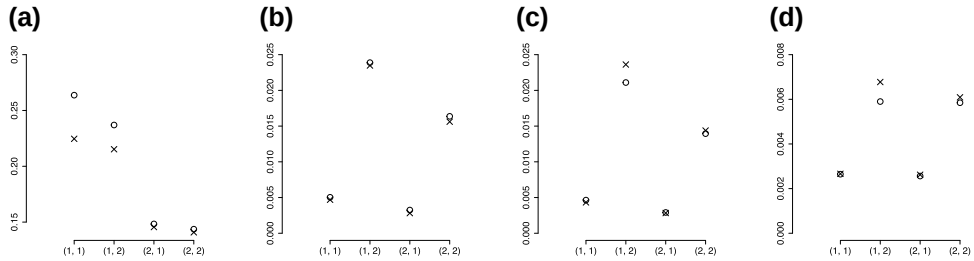

Figure 3. Simulation Study: The RMISE for (a) the Population Performance Trajectory $g(c)$, (b) Population Within-Season Performance Trajectories $h^\star(\cdot)$, (c) Individual Within-Season Performance Trajectories $h^\star(\cdot)_i$, and (d) Seasonal Within-Season Performance Trajectories $h^\star_{i,s}(t)$. For all plots, the x-axis labels are (i, j) where i is the level of M and j represents the level of (σ_a^2, σ_b^2), cross represents $p_1 = 0.2$ and circle represents $p_1 = 0.5$.

$M = (200, 500)$, $p_1 = (0.2, 0.5)$, $(\sigma_a^2, \sigma_b^2) = (0.1, 0.03), (0.5, 0.25)$. We use performance measures for seven parameters of interest: the Population Performance Trajectory $g(\cdot)$, the Population Within-Season Performance Trajectory $h^\star$, Individual Within-Season Performance Trajectory $h_i^\star$, the Within-Season Performance Trajectory $h_{i,s}^\star$, the excess performance at the start of each season $\eta_1, \ldots, \eta_M$, Ψ_i and Γ_i. For the first four functional parameters, we define the root mean squared integrated error to be

$$\text{RMISE} = \int \left(\hat{f}(z) - f(z) \right)^2 dz,$$

where f is the true value of the function and $\hat{f}$ is the posterior mean of f and the limit of integration are the possible range of z. The RMISE is averaged over all athletes for $h_i^\star(\cdot)$ and over all seasons and athletes for $h_{i,s}^\star(\cdot)$. For η_i, we calculate the average root mean squared error (AMRSE),

$$\text{AMRSE} = \frac{1}{\sum_{i=1}^{M} S_i + 1} \sum_{i=1}^{M} \sum_{s=1}^{S_i+1} \left(\hat{\eta}_{i,s} - \eta_{i,s} \right)^2,$$

where $\hat{\eta}_{i,s}$ is the posterior mean of $\eta_{i,s}$. We would like to use τ_i^2 as a proxy for the squared difference between the individual and population within-season trajectory ($(h_i^\star - h^\star)^2$). To understand the strength of this relationship, we calculate Spearman's rank correlation coefficients between the posterior median of $\tau_1^2, \ldots, \tau_M^2$ and $|a_1|, \ldots, |a_M|$ as defined in (6) since these are not directly comparable. Similarly, to understand the relationship between λ_i^2 and the squared difference between the seasonal and individual within-season trajectory, we calculate Spearman's correlation between $\bar{b}_i = \frac{1}{S_i} \sum_{s=1}^{S_i} |b_{i,s}|$ where $b_{i,s}$ is given in (7) and λ_i^2.

The results for the functional parameters are shown in Figure 3. A larger number of athlete, M, leads to lower RMISE's for each of these functions. The values of σ_a^2 and σ_b^2 have only a small effect on the estimation accuracy of the population performance trajectory. Conversely, σ_a^2 and σ_b^2 have a large effect on the estimation accuracy for all three within-season trajectories.

The results for the other three parameters of interest are shown in Figure 4. The estimation of η is most strongly affected by the number of athletes M. The two correlation measures increase with level of variation in the within-season trajectories but the other factors do not have a strong effect. We find that there is a stronger correlation between a_i and τ_i^2 than $\bar{b}_i$ and λ_i^2. In fact, for the within-season correlation, the correlation is close to zero with the lower level of variation but is around 0.7 for the higher level of variation. Therefore, these are useful summaries of the variation in these trajectories if there is sufficient variation in the data.

6 Application to elite swimming

We applied our model to performances for both female and male swimmers in the 100 and 200 m freestyle. We fitted the model separately to each combination of gender and distance, and included

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

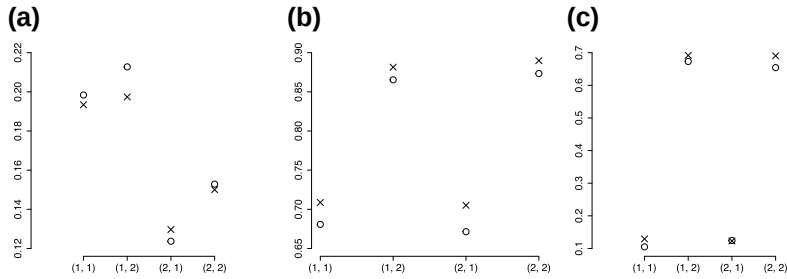

Figure 4. Simulation Study: (a) The ARMSE of $\eta_{i,j}$, (b) Spearman's rank correlation coefficient between a_i and τ_i^2, and (c) Spearman's rank correlation coefficient between $\bar{b}_i = \frac{1}{S_i}\sum_{j=1}^{S_i} |b_{i,j}|$ and λ_i^2. For all plots, the x-axis labels are (i, j) where i is the level of M and j represents the level of (σ_a^2, σ_b^2), cross represents $p_1 = 0.2$ and circle represents $p_1 = 0.5$.

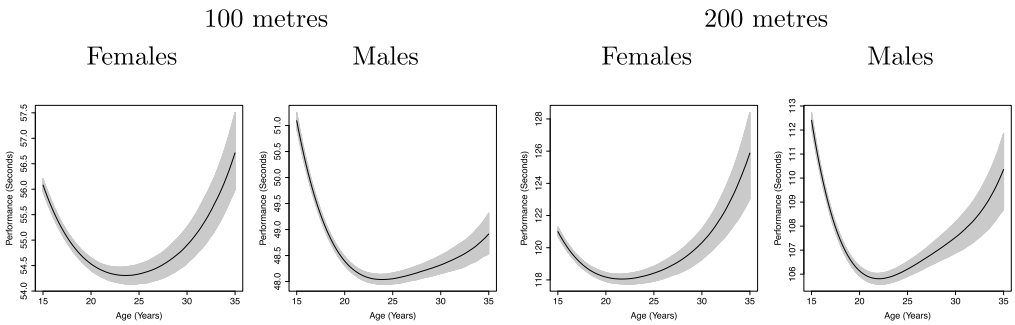

Figure 5. Estimated population performance trajectory $g(\cdot)$ for both females and males in the 100 and 200 m freestyle. The trajectories are shown as posterior median (black line) and 95% credible interval (grey shading).

pool length as a confounder. Some additional results are provided in online supplementary material Section 5 (Griffin et al., 2026). Two samplers were run for a total 50,000 iterations. The first 30,000 iterations were used as a burn-in and the subsequent 20,000 iterations were thinned every 20th value to 1,000 posterior samples. This took about ten hours to run with R using an Apple Mac with M4 chip. Some MCMC diagnostics for each data set are given in online supplementary material Section 4 (Griffin et al., 2026).

Figure 5 shows the estimated population performance trajectories, which all have a reverse J-shape, as in Griffin et al. (2022), with performance rapidly improving between 15 and 20, peaking around 22 (in both men and women), and subsequently decreasing. There is a clear difference between men and women. Men show a much greater improvement in performance between 15 and 20 (for example, in the 100 m, females improve by 1.5 s between 15 and 20 whereas males improve by 2.6 s. Men are also better able to maintain their performance in their late twenties (for example, female performance decreases by 2 s by 25 and 30 whereas men's performance decreases by 0.7 s).

The population-level within-season trajectories are shown in Figure 6. The trajectories have very similar shapes across gender and distance with the within-season performance improvement peaking around July. The posterior median improvement is around 0.17 s in the 100 m and around 0.39 s in the 200 m for both males and females. The error distributions were found to be positively skewed with evidence of a large difference in the heaviness of the left-hand and right-hand tails. Results are presented in online supplementary material Section 5 (Griffin et al., 2026).

Figure 7 shows within-season performance trajectories for six individuals in the 100 metres (three female and three male) with their average within-season performance trajectory. The athletes were chosen to show a range of behaviours where some individuals have a substantial difference between their best and worst performance level within-season (Swimmer 1, 4, and 6) and

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

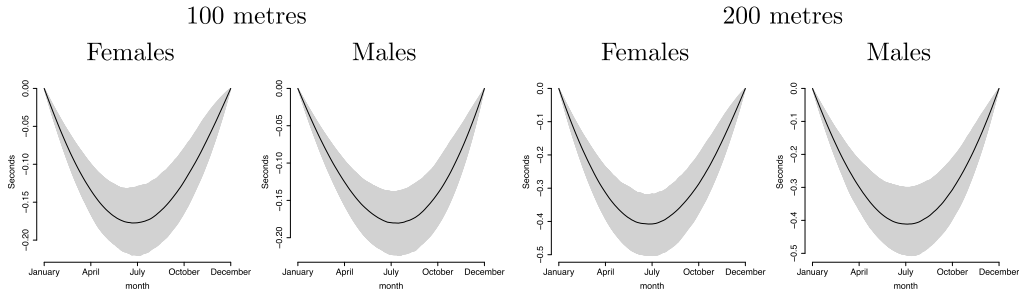

Figure 6. Estimated population within-season performance trajectory $h^{\star}(\cdot)$ for both females and males in the 100 and 200 m freestyle. The trajectories are shown as posterior median (black line) and 95% credible interval (grey shading).

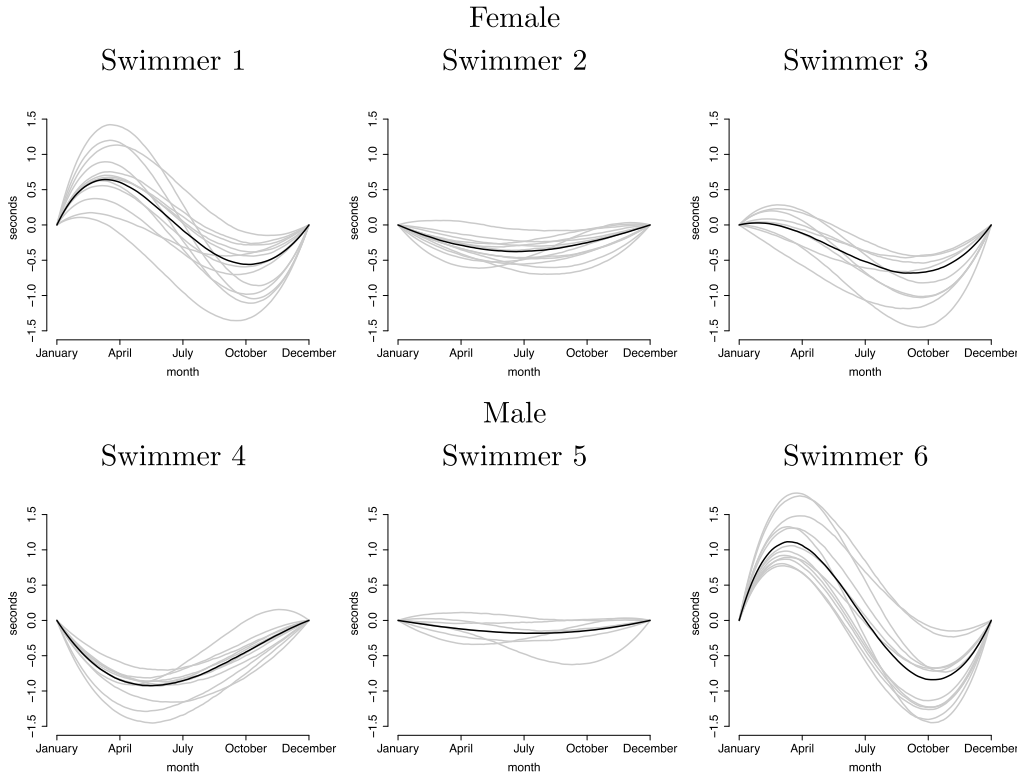

Figure 7. Estimated within-season trajectories. Posterior median athlete-level (black line) and within-season performance trajectories for each career season (grey lines).

others show much less variation over the season (Swimmers 2 and 5). There are also some clear differences in the shape of the curve. Swimmers 1, 3, and 6 peak in October, whereas Swimmers 2 and 5 peak in July and Swimmer 4 in May. This reflects differences in the aims of athletes. Some athletes will target events such as the Olympics whereas other athletes will target the winter rather than summer season. We can use the posterior median of τ_i^2 to understand how these individual trajectories relate to the population trajectory. Swimmers 5 ($\tau_i^2 = 0.029$), 2 ($\tau_i^2 = 0.089$), and 3 ($\tau_i^2 = 0.175$) are closest to the population trajectory with a similar shape and level of performance improvement. There are larger differences for Swimmers 1 ($\tau_i^2 = 0.288$), who shows a

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

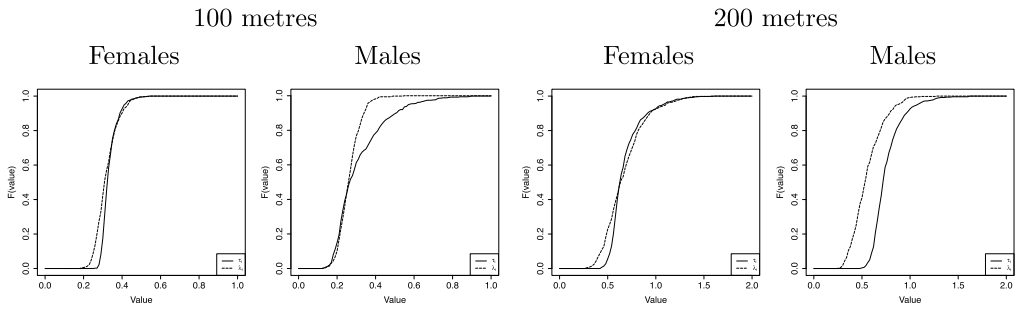

Figure 8. Empirical distribution function of the posterior median values of the within-season variability (Δ_i) and the average effect size (Γ_i) for all swimmers (divided by gender) in the 100 and 200 metre freestyle.

different shape, and 4 ($\tau_i^2 = 0.274$), who shows a larger level of improvement. Swimmer 6 ($\tau_i^2 = 0.803$) shows differences in both the shape of trajectory and the level of performance improvement. There are also clear differences in the consistency of the trajectories across seasons. Swimmers 2 ($\lambda_i^2 = 0.064$), 5 ($\lambda_i^2 = 0.077$), and 4 ($\lambda_i^2 = 0.086$) show the lowest level of variation in the seasonal trajectory. Swimmers 3 ($\lambda_i^2 = 0.1243$) and 6 show ($\lambda_i^2 = 0.154$) show more variation and Swimmer 1 ($\lambda_i^2 = 1.083$) shows the most variation in trajectory. This may reflect several factors including changing priorities over an swimmer's career or injury problems.

Figure 8 shows the empirical distribution function of the posterior median values of the individual within-season variability (λ_i) and the individual between-season variability (τ_i) for both females and males and the two distances. The distribution is similar for both females and males and so we concentrate on the results for female. Both distributions have a very heavy right-tail so that a few effects have much larger values of τ_i and λ_i than others. The distribution of τ_i is shifted to the left of the distribution of λ_i indicating that the within-season variability tends to be smaller than the average effect size. In other words, the season-on-season variability within a swimmer tends to be smaller than the variability between swimmers. This indicates that swimmers have some ability to control how their performance levels evolve over a season through the training process and are able to replicate that improvement in different seasons.

7 Discussion

In this work, we have developed a Bayesian longitudinal model which can account for the variation in performance change within a season across population-level, athlete-level, and within-season (i.e. within athletes). An application to freestyle swimming data shows that there is substantial variation between swimmers and between seasons with some having a clear pattern of peaking for major events (e.g. Olympics Games and World Championships) which usually occur during the summer months (July–August). We use a dth degree of polynomial to model the population-level effect of age and an error distribution which allows for skewness and different heaviness of the left and right tail. We find that the population-level effect of age follows the expected reverse J shape in freestyle swimming with a difference between the improvement in performance of females and males in years 15–23. We find that the error distribution has a much lighter left tail than right tail. The result suggests that swimmers are much less likely to have a performance that is substantially worse than expected, rather than one that is substantially better than expected. One explanation is that elite athletes generally perform close to their optimal level and so improvements are much harder to achieve than poor performances (which can be due to many factors including things such as poor race execution, illness, and injury).

The model provides some interesting insights about athlete performance but there are some limitations. An individual's trend excess performance trajectory is assumed to follow a random walk, which is appropriate for retrospective analysis. This approach may also be able to provide short-term prediction but the lack of structure will struggle to provide good long-term prediction performance. This would need additional structure to explain the evolution of career trajectories, which could include additional covariates. The model is currently restricted to a single discipline

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

but athletes compete in multiple disciplines in some sports (across different distances in track running or distances and styles in swimming). It would be interesting to extend this model to better understand differences in performance trajectories across disciplines for a single athlete. The model also assumes a single season but the results show that athletes may follow different seasonal patterns. For example, Australian swimmers have different seasons to European swimmers. Our model is able to capture these differences but a more complicated model allowing for different seasonal patterns could provide better estimates. The model also assumes independence across athletes conditional on the population performance trajectory and population within-season performance trajectory. However, there could be further sharing of information across athletes at other levels of the hierarchy, such as the individual average within-season performance trajectories. Colombi et al. (2025) present an interesting Bayesian nonparametric approach to achieve this goal, which they use to stratify athletes by performance whilst allowing for difference between athletes and seasons.

Conflicts of interest: The authors have no conflicts of interest.

Funding

This research was supported by a Partnership for Clean Competition research grant awarded to James Hopker (Grant: 514).

Data availability

The data used in this paper were collected from the World Aquatics website.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series C*.

References

Berry S. M., Reese C. S., & Larkey P. D. (1999). Bridging different eras in sports. *Journal of the American Statistical Association*, *94*(447), 661–676. https://doi.org/10.1080/01621459.1999.10474163

Bhadra A., Datta J., Polson N. G., & Willard B. (2019). LASSO meets horseshoe: A survey. *Statistical Science*, *34*(3), 405–427. https://doi.org/10.1214/19-STS700

Bullock N., Gulbin J. P., Martin D. T., Ross A., Holland T., & Marino F. (2009). Talent identification and deliberate programming in skeleton: Ice novice to Winter Olympian in 14 months. *Journal of Sports Science*, *27*(4), 397–404. https://doi.org/10.1080/02640410802549751

Colombi A., Argiento R., Camerlenghi F., & Paci L. (2025). Hierarchical mixture of finite mixtures. *Bayesian Analysis*, *20*(4), 1231–1259. http://doi.org/10.1214/24-BA1501

Dolmeta P., Argiento R., & Montagna S. (2023). Bayesian GARCH modeling of functional sports data. *Statistical Methods & Applications*, *32*(2), 401–423. https://doi.org/10.1007/s10260-022-00656-z

Griffin J. E., & Brown P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, *5*(1), 171–188. https://doi.org/10.1214/10-BA502

Griffin J. E., Hinoveanu L. C., & Hopker J. G. (2022). Bayesian modelling of elite sporting performance with large databases. *Journal of Quantitative Analysis in Sports*, *18*(4), 253–268. https://doi.org/10.1515/jqas-2021-0112

Griffin J. E., Spyropoulou M.-Z., & Hopker J. G. (2026). Supplementary Material for "Modelling between- and within-season trajectories in elite athletic performance data". *Journal of the Royal Statistical Society: Series C, Applied Statistics*.

Knorr-Held L., & Rue H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, *29*(4), 597–614. https://doi.org/10.1111/sjos.2002.29.issue-4

Malcata R. M., & Hopkins W. G. (2014). Variability of competitive performance of elite athletes: A systematic review. *Sports Medicine*, *44*(12), 1763–1774. https://doi.org/10.1007/s40279-014-0239-x

Paton C. D., & Hopkins W. G. (2005). Combining explosive and high-resistance training improves performance in competitive cyclists. *Journal of Strength & Conditioning Research*, *19*, 826–830. https://doi.org/10.1519/R-16334.1

Pyne D., Trewin C., & Hopkins W. (2004). Progression and variability of competitive performance of Olympic swimmers. *Journal of Sports Science*, *22*(7), 613–620. https://doi.org/10.1080/02640410310001655822

Downloaded from https://academic.oup.com/jrssc/advance-article/doi/10.1093/jrssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026

Roberts G. O., & Sahu S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *59*(2), 291–317. https://doi.org/10.1111/1467-9868.00070

Stival M., Bernardi M., Cattelan M., & Dellaportas P. (2023). Missing data patterns in runners' careers: Do they matter? *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *72*(1), 213–230. https://doi.org/10.1093/jrsssc/qlad009

Trewin C. B., Hopkins W. G., & Pyne D. B. (2004). Relationship between world-ranking and Olympic performance of swimmers. *Journal of Sports Science*, *22*(4), 339–345. https://doi.org/10.1080/02640410310001641610

Wang J., & Ghosh S. K. (2012). Shape restricted nonparametric regression with Bernstein polynomials. *Computational Statistics & Data Analysis*, *56*(9), 2729–2741. https://doi.org/10.1016/j.csda.2012.02.018

Yu Y., & Meng X.-L. (2011). To center or not to center: That is not the question—An ancillarity–sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, *20*(3), 531–570. https://doi.org/10.1198/jcgs.2011.203main

Downloaded from https://academic.oup.com/jrsssc/advance-article/doi/10.1093/jrsssc/qlag002/8456346 by Centre for Health Services Studies user on 03 February 2026