



Kent Academic Repository

Montagna, Silvia and Hopker, James G. (2018) *A Bayesian approach to the use of athlete performance data within anti-doping*. *Frontiers in Physiology*, 9 (884). ISSN 1664-042X.

Downloaded from

<https://kar.kent.ac.uk/67380/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.3389/fphys.2018.00884>

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](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



A Bayesian Approach for the Use of Athlete Performance Data Within Anti-doping

Silvia Montagna¹ and James Hopker^{2*}

¹ ESOMAS Department, University of Torino, Torino, Italy, ² Endurance Research Group, School of Sport and Exercise Sciences, University of Kent, Medway, United Kingdom

OPEN ACCESS

Edited by:

Leonardo Alexandre Peyré-Tartaruga,
Universidade Federal do Rio Grande
do Sul (UFRGS), Brazil

Reviewed by:

Luca Paolo Ardigo,
University of Verona, Italy
Wolfgang Schobersberger,
Institut für Sport-, Alpinmedizin and
Gesundheitstourismus (ISAG), Austria
Fábio Juner Lanferdini,
URI Santo Ângelo, Brazil

*Correspondence:

James Hopker
j.g.hopker@kent.ac.uk

Specialty section:

This article was submitted to
Integrative Physiology,
a section of the journal
Frontiers in Physiology

Received: 31 January 2018

Accepted: 19 June 2018

Published: 19 July 2018

Citation:

Montagna S and Hopker J (2018) A
Bayesian Approach for the Use of
Athlete Performance Data Within
Anti-doping. *Front. Physiol.* 9:884.
doi: 10.3389/fphys.2018.00884

The World Anti-doping Agency currently collates the results of all doping tests for athletes involved in elite sporting competition with the aim of improving the fight against doping. Existing anti-doping strategies involve either the direct detection of use of banned substances, or abnormal variation in metabolites or biological markers related to their use. As the aim of any doping regime is to enhance athlete competitive performance, it is interesting to consider whether performance data could be used within the fight against doping. In this regard, the identification of unexpected increases in athlete performance could be used as a trigger for their closer scrutiny via a targeted anti-doping testing programme. This study proposes a Bayesian framework for the development of an “athlete performance passport” and documents some initial findings and limitations of such an approach. The Bayesian model was retrospectively applied to the competitive results of 1,115 shot put athletes from 1975 to 2016 in order to establish the interindividual variability of intraindividual performance in order to create individualized career performance trajectories for a large number of presumed clean athletes. Data from athletes convicted for doping violations (3.69% of the sample) was used to assess the predictive performance of the Bayesian framework with a probit model. Results demonstrate the ability to detect performance differences (~1 m) between doped and presumed clean athletes, and achieves good predictive performance of doping status (i.e., doped vs. non-doped) with a high area under the curve (AUC = 0.97). However, the model prediction of doping status was driven by the correct classification of presumed non-doped athletes, misclassifying doped athletes as non-doped. This lack of sensitivity is likely due to the need to accommodate additional longitudinal covariates (e.g., aging and seasonality effects) potentially affecting performance into the framework. Further research is needed in order to increase the framework structure and improve its accuracy and sensitivity.

Keywords: anti-doping, monitoring, target testing, data analytics, competition results

1. INTRODUCTION

Over the last decade the fight against doping in sports has evolved from purely biochemical testing for specific substances, to longitudinal profiling of specific biomarkers in the form of the athlete biological passport (ABP). The ABP uses longitudinal blood tests or steroid profiles, also known as modules, to indirectly demonstrate use of banned substances via a mathematical probabilistic

approach (Sottas et al., 2008). This longitudinal approach has improved the efficiency of targeted testing and resulted in an increased number of positive erythropoiesis stimulating agent doping cases in recent years (Zorzoli et al., 2014). Moreover, other anti-doping initiatives as outlined in the WADA technical document for sport specific analysis, signifies a move toward a more forensic intelligence led anti-doping system, which gathers broader sources of information to inform the planning of doping tests (WADA, 2016).

A major source of information that is not currently used within anti-doping practice is athlete performance. As the goal of most doping regimes is to improve athlete performance to gain an unfair advantage in competition, it seems logical to assume that the effect of doping might be identified through the evolution of performance of an athlete over time. For example, on a global perspective, (Schumacher and Pottgiesser, 2009) demonstrate marked improvements in world best performances in male 5,000 and 10,000 m running performances following the commercial introduction of recombinant erythropoietin in the 1990s. Conversely, they also highlight a down turn in female world best discus performances following the introduction of out-of-competition anti-doping tests for anabolic steroids in the late 1980s. Therefore, it is possible to question whether longitudinal tracking of athlete performance might provide additional information (alongside haematological and steroidal ABP modules) that can be used as part of the intelligence gathering process to inform anti-doping organization’s testing programmes.

The main objective of performance profiling in the form of an athlete performance passport (APP), would be to track individual performance over time in order to identify unexpected or disproportionate increases in performance that might be indicative of doping (Hopker et al., 2018). However, analysis of performance data of this nature is complex as it is difficult to differentiate between physiological increases in performance arising from normal training and/or maturation from unphysiological improvements caused by doping. Moreover, excellent athletic performance within a competition in itself is not proof of any wrong doing, or doping. Nevertheless, an APP may be useful in strengthening the sensitivity and applicability of the current ABP in the fight against doping by providing information to trigger targeted anti-doping tests for specific athletes (Iljukov et al., 2018).

In order to assess the feasibility of an APP approach, this study explores the important issue of modeling and analyzing the relationship between doping status and athlete performance accounting for some of the other covariates that impact on performance. Specifically, as a first step toward an APP, this study aims to characterize trajectories in athlete shot put performance results, while simultaneously accounting for sex and age. To address these aims, we used a Bayesian latent factor model for functional data (Montagna et al., 2012) characterizing the curve for each athlete as a linear combination of a high-dimensional set of basic functions, and placed a sparse latent factor regression model on the basis coefficients. Within this framework, it is possible to study the dependence of the curve shapes on covariates incorporated through the distribution of the latent

factors, and regress a scalar response (e.g., the doping status) on a functional covariate (e.g., an athlete shot put performance curve).

2. METHODOLOGY

Our goal is predicting an athlete’s doping status (a scalar response) given his/her shot put performance results and other non-functional covariates (e.g., age and sex). In section 2.1, we will investigate how the Bayesian latent factor regression model of (Montagna et al., 2012) can be used to represent an athlete’s shot put trajectory, and in section 2.2 we will explore the extension of this method to the joint modeling of the functional predictor and the scalar outcome. Finally, in section 2.3 we will briefly discuss the advantages of a Bayesian inferential approach over classical methods.

2.1. Functional Latent Factors Representation for Shot Put Performance

Let n denote the number of athletes in the study. We supposed that shot put performance results for athlete i are available as noisy measurements of an underlying smooth curve $f_i(t)$ at n_i time points $t_{ij}, j = 1, \dots, n_i$. Here t_{ij} denotes the time (in days) from the first measurement available for athlete i , thus $t_{i1} = 0$ for every i . We denote the shot put result at t_{ij} as y_{ij} and model:

$$y_{ij} = f_i(t_{ij}) + \epsilon_{ij}, \tag{1}$$

with $\epsilon_{ij} \sim N(0, \psi^2)$, independently across i and j . **Figure 1** shows the true shotput performance results for athletes who competed in the year of 2012. For the modeling of the random underlying smooth curves f_1, \dots, f_n , we follow (Montagna et al., 2012) and write $\{f_i\}_{i=1}^n$ in terms of a collection of fixed basis functions:

$$f_i(t_{ij}) = \sum_{m=1}^p \theta_{im} b_m(t_{ij}) = \mathbf{b}(t_{ij})^\top \boldsymbol{\theta}_i. \tag{2}$$

It is important to use a sufficiently large p and to choose locally concentrated basis elements so that a rich variety of shapes for f_i are entertained. Hereafter, $\{b_m(\cdot)\}_{m=1}^p$ are chosen to be 2D isotropic Gaussian kernels:

$$b_1(t) = 1, \quad \text{and} \quad b_m(t) = \exp(-b \|t - \phi_m\|^2), m = 2, \dots, p, \tag{3}$$

with kernel locations $\{\phi_m\}_{m=2}^p$ and bandwidth b to be specified according to prior knowledge.

The basis coefficients vectors $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$ capture all athlete-to-athlete variations in shot put performance curves. However, these vectors have a large dimension p and are non-sparse. To obtain a low dimensional representation of the shot put curves, we place a sparse latent factor model (Arminger, 1998) on the basis coefficients:

$$\boldsymbol{\theta}_i = \boldsymbol{\Lambda} \boldsymbol{\eta}_i + \boldsymbol{\zeta}_i, \quad \text{with} \quad \boldsymbol{\zeta}_i \sim N_p(0, \boldsymbol{\Sigma}), \tag{4}$$

where $\boldsymbol{\Lambda} = \{\lambda_{ml}\}$ is a $p \times k$ factor loading matrix with $k \ll p$, $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{ik})^\top$ is a vector of latent factors for athlete i

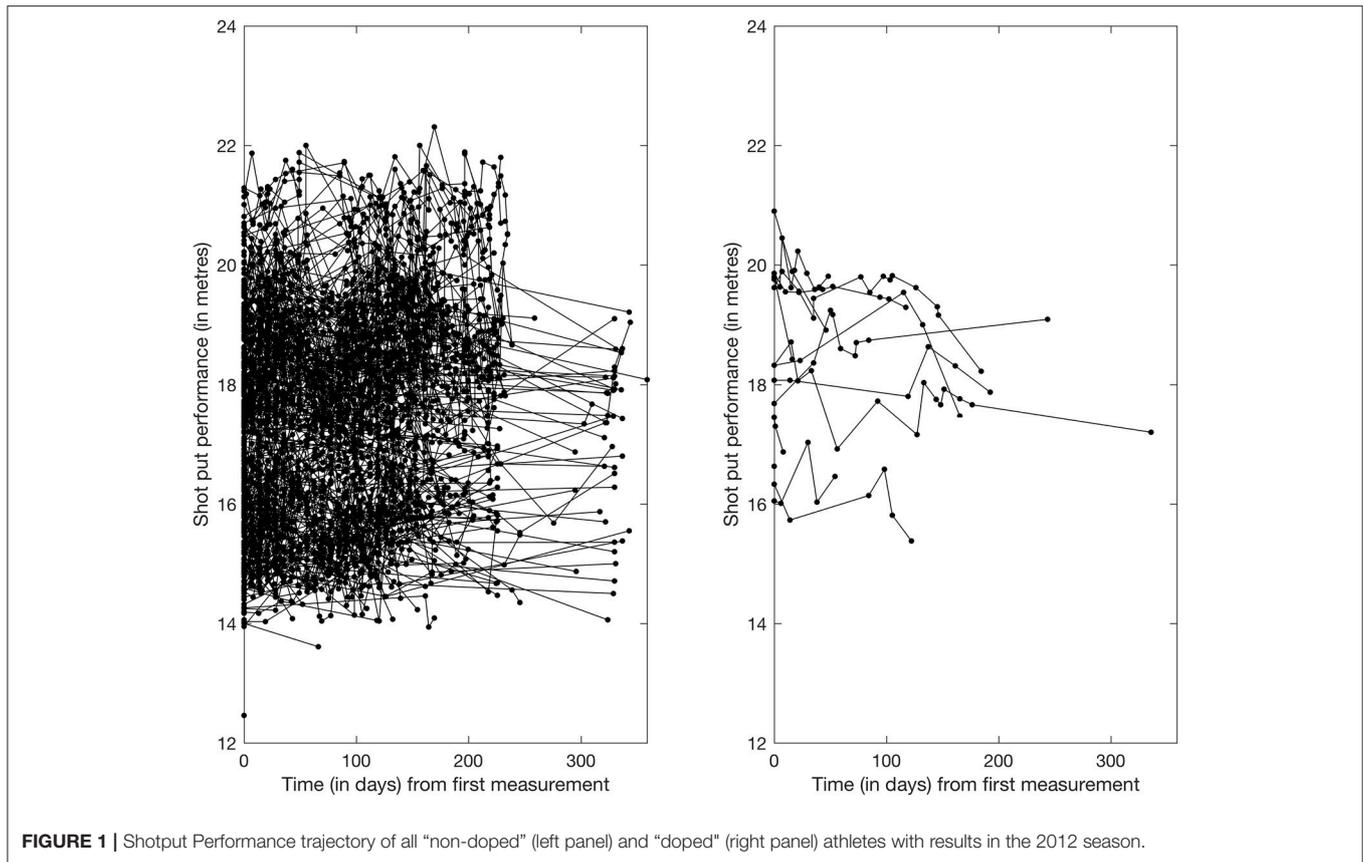


FIGURE 1 | Shotput Performance trajectory of all “non-doped” (left panel) and “doped” (right panel) athletes with results in the 2012 season.

and $\zeta_i = (\zeta_{i1}, \dots, \zeta_{ip})^\top$ is a residual vector that is independent with the other variables in the model and is normally distributed with mean zero and a diagonal covariance matrix $\Sigma = (\sigma_1^2, \dots, \sigma_p^2)$. The low-dimensional vectors η_1, \dots, η_n are used in all subsequent parts of the model where one seeks to link the shot put curves f_1, \dots, f_n with other variables of interest.

Information on covariates \mathbf{x}_i (e.g., sex and age) available for athlete i can be incorporated through a simple linear model on the latent factors:

$$\eta_i = \beta^\top \mathbf{x}_i + \Delta_i, \quad \Delta_i \sim N_k(0, I), \quad (5)$$

where β is a $r \times k$ matrix of covariates coefficients, with r denoting the number of covariates. We remark that the current formulation of model (1)–(5) cannot accommodate covariates changing over time, thus \mathbf{x}_i can only be a vector of static covariates. For a more detailed discussion on the properties of this model, we defer to (Montagna et al., 2012).

2.2. Prediction of Doping Status

Our goal is to use the shot put performance trajectory modeled through Equations (1)–(5) to predict the doping status of an athlete z_i . Let $z_i = 1$ if athlete i is convicted for doping offences at any point in his/her career, and $z_i = 0$ otherwise. From a statistical perspective, z_i is a Bernoulli random variable $z_i \sim \text{Bernoulli}(p_i)$, where p_i is the probability that athlete i will use

doping at any point in his/her career. We follow (Montagna et al., 2012) and model p_i as:

$$p_i = \mathbf{P}(z_i = 1 \mid \alpha, \boldsymbol{\gamma}, \eta_i) = \Phi(\alpha + \boldsymbol{\gamma}^\top \eta_i), \quad (6)$$

where $\Phi(\cdot)$ denotes the standard normal distribution function, α is an intercept, and $\boldsymbol{\gamma}$ is a vector of unknown regression coefficients. We remark that the same set of latent factors η_i impacts on the shot put performance curve via the basis coefficients θ_i and on the response variable (the doping status) via the probability of doping. We remark that Equation (6) defines a probit model for doping status.

2.3. The Bayesian Inferential Framework

To summarize our methodology, we provide a graphical representation of the model outlined in Equations (1)–(6) in **Figure 2**. The vector of latent factors plays the key role in linking the two component models for doping status and shot put performance curves, and the doping status z_i is conditionally independent of all nodes in the model given the latent factors η_i . All parameters located outside of the dashed rectangle $(\alpha, \boldsymbol{\gamma}, \beta, \Lambda, \Sigma)$ are shared by all athletes (parameters not indexed by i), and they are estimated by pulling information (*borrowing of information*) across all athletes. This is of crucial importance especially when data are sparse. If covariate information is

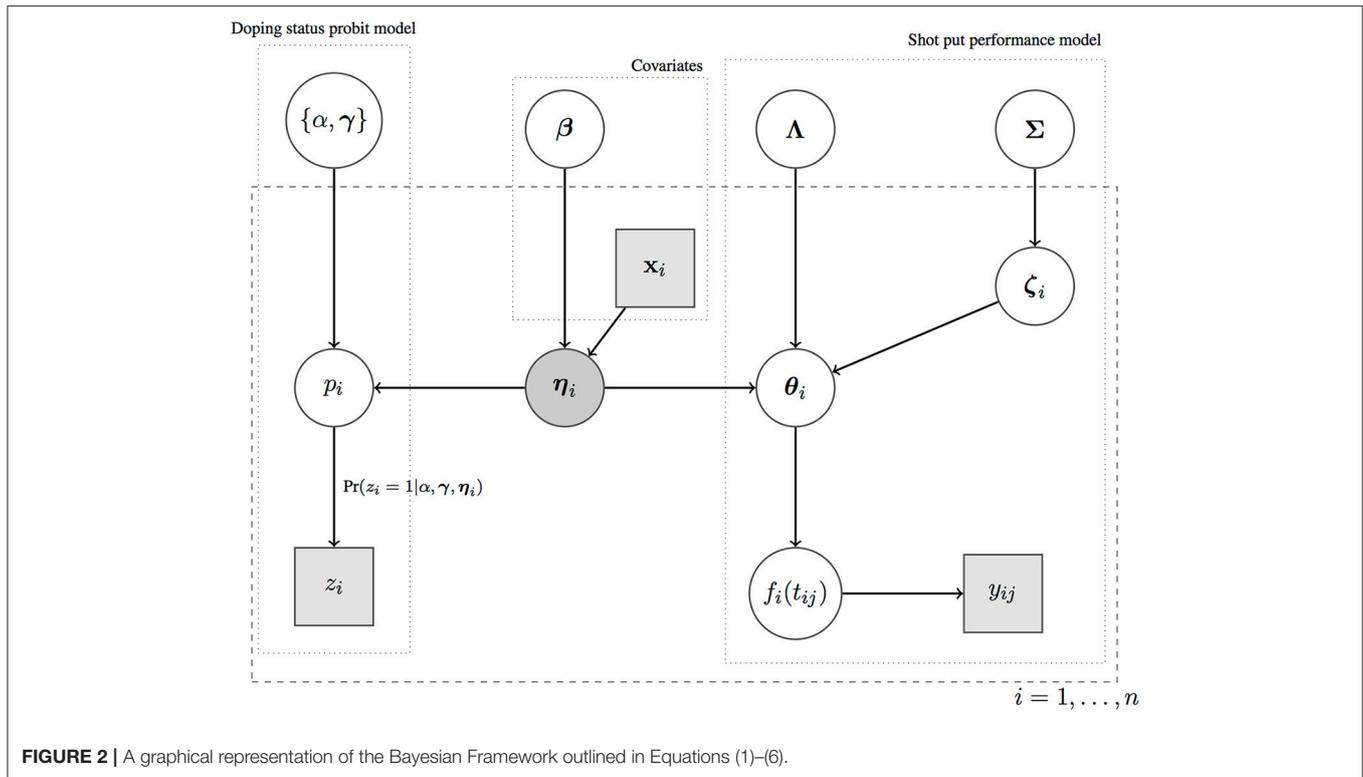


FIGURE 2 | A graphical representation of the Bayesian Framework outlined in Equations (1)–(6).

available, covariates impact on the η_i 's via a linear regression model.

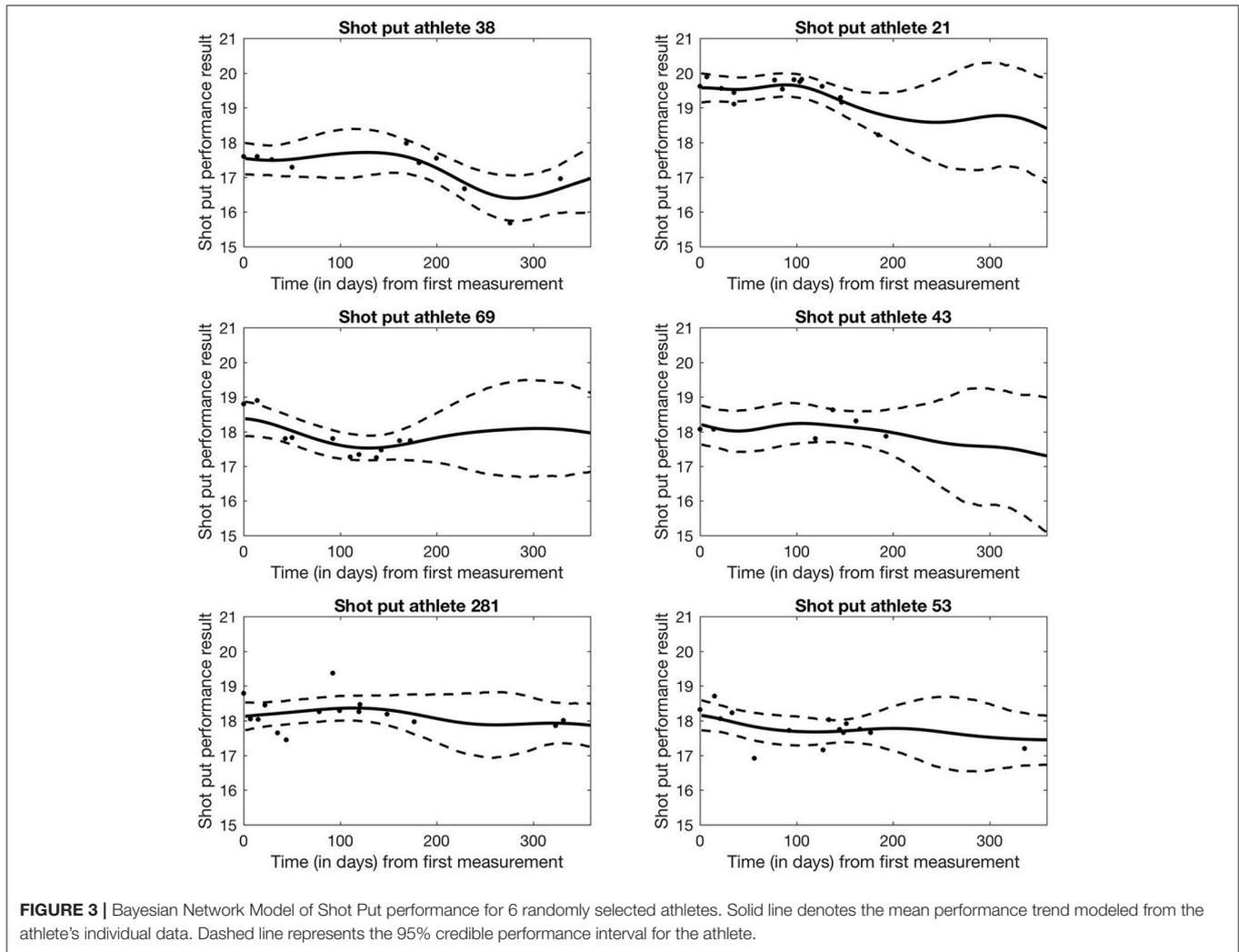
All model parameters, which are represented by circles in **Figure 2**, need to be estimated from the data. To estimate these parameters, one can either proceed via classical methods or by embedding model (1)–(6) in a Bayesian framework. In classical methods, the model parameters are unknown, but fixed constants. In Bayesian statistics, the model parameters are considered to be random variables and are therefore assigned *prior* probability distributions. The priors are then combined with the likelihood to obtain the *posterior* distributions of the model parameters.

Bayesian methods and classical methods both have advantages and disadvantages, and there are some similarities. When the sample size is large, Bayesian inference often provides results for parametric models that are very similar to the results produced by classical methods. A Bayesian analysis provides a convenient setting to a wide range of models, such as the hierarchical model outlined in Equations (1)–(6). Existing numerical methods make computations tractable for virtually all parametric models. Most importantly, Bayesian inference provides a natural and principled way of combining prior experts information with the data, within a solid decision theoretical framework. For example, let us look at the intercept parameter α in Equation (6). This parameter represents the baseline risk for an athlete to use doping at some point in his/her career, regardless of athlete-specific factors (e.g., sex, age, performance). While it is impossible to know the exact proportion of doped athletes in a given sport, experts may have a sense for what this proportion (or a plausible range for this proportion) may be based on their knowledge.

Suppose that experts believe that about 1% of athletes use doping. Such belief can be embedded in the prior for α , which is generally chosen to be Normal for computational convenience. Thus, α will be given a $\alpha \sim N(\Phi^{-1}(0.01), 1)$ prior so that the hyperprior mean is chosen to correspond to the experts knowledge (the variance can also be suitably adjusted). It is naturally possible that different people will produce different priors, and by trying different prior choices we can investigate how sensitive conclusions are to these choices. Unfortunately, it is not possible to make use of prior knowledge within a classical inferential framework.

For all the reasons above, we decide to follow the lead in (Montagna et al., 2012) and embed model (1)–(6) in a Bayesian framework by choosing prior distributions for all model parameters. Given the high dimensionality of the problem (large n , the number of subjects, and large p , the number of basis functions in Equation 2) it is practically important to choose conditionally *conjugate* prior distributions for parameters $\psi^2, \alpha, \Lambda, \Sigma, \beta, \gamma$. Conjugacy ensures that the posterior distribution of the parameters is of the same type of the prior (e.g., a Normal prior combined with a Normal likelihood as in Equation (1) returns a Normal posterior distribution), and this leads to efficient posterior computation via Markov Chain Monte Carlo. We adopt the same conjugate priors of (Montagna et al., 2012) and details on posterior computation, together with the MATLAB code to run the analysis, are available in (Montagna et al., 2012).

Despite the simplicity of this hierarchical model, the resulting model on the smooth shot put trajectories f_1, \dots, f_n allows a flexible accommodation of covariate information. In particular, these curves are independent



Gaussian processes with covariate dependent mean functions $\mathbb{E}[f_i(t)] = \sum_{m=1}^k \beta_m^T \mathbf{x}_i \tilde{\phi}_m(t)$ and a common covariance function $\text{Cov}\{f_i(t), f_i(s)\} = \sum_{m=1}^k \tilde{\phi}_m(t) \tilde{\phi}_m(s) + \sum_{l=1}^p \sigma_l^2 b_l(t) b_l(s)$, where β_m denotes the m th column of β and $\tilde{\phi}_m(t) = \sum_{l=1}^p \lambda_{lm} b_l(t)$. In summary, our model accommodates scalar (doping)-on-function (shot put trajectories) regression and the distribution of the curves is allowed to change flexibly with (static) predictors.

3. DATA

Following Institutional ethical approval (Prop_72_2017_18) athletes data to be included within the model was obtained with permission from an open results database (www.tilastopaja.eu). Specifically, 56,000 results of elite shot put competitions of 1,115 athletes from 1976 to 2016, inclusive, were collected. Sampled data included the athlete name, IAAF ID number, date of birth, sex, country of birth, event, result in meters, finishing position, and any doping violation during the athlete's career.

Hereafter, we focus on the analysis of data collected from 2012. This enables us to use both sex and age as predictors of shot put performance. We remark that the model could be applied to all the available data from 1976 to 2016, but we would be no longer able to use age (now a longitudinal predictor) as covariate. Elite shot put performance data were collected on 352 athletes between 5th January and 29th December, 2012, for a total of 3290 measurements. The average number of measurements per athlete was 9.34, ranging from a minimum of 1 to a maximum of 27 measurements. The average age of the athletes was 23 years, with the youngest athlete being 15 years old and the oldest athlete being 43 years old, and 175 athletes were males. Only 13 athletes in this dataset were convicted for doping offences ($z_i = 1$), representing 3.69% of the sample. We applied model (1)–(6) to the data, and results are presented in the next section.

4. RESULTS

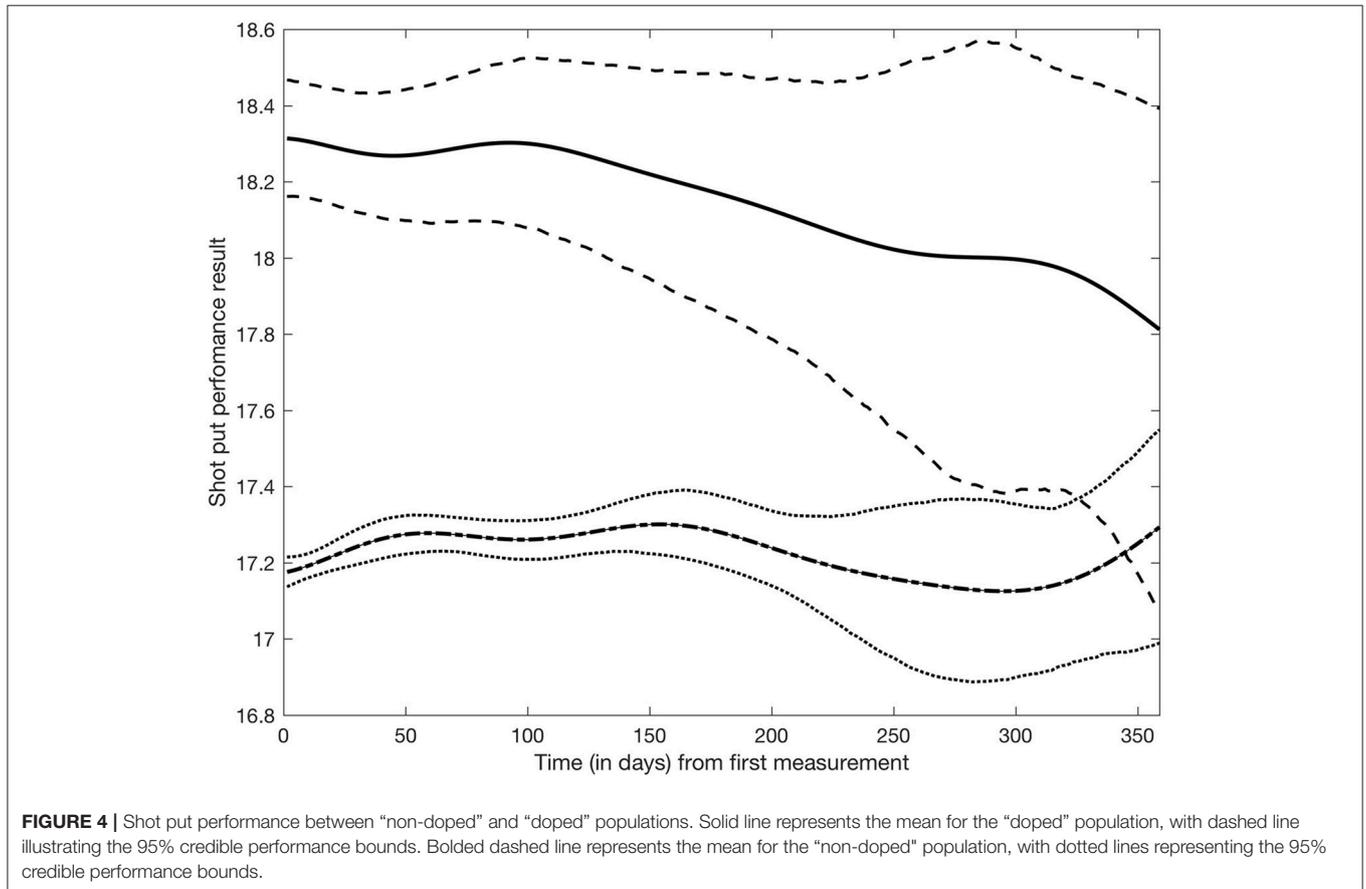
Figure 3 shows the estimated trajectories \hat{f}_i for six randomly selected athletes. As expected, we note that 95% credible intervals

expand when no measurements are available, and this effect is particularly evident toward the end of the year for athletes 21, 69, and 43. Wider bands denotes higher uncertainty in the estimates due to lack of data. Left panels show three randomly selected non-doped athletes, and right panels show three randomly selected doped athletes. No significant differences emerge in the shape of the estimated trajectories at this level of analysis between doped and non-doped athletes.

Figure 4 shows the average shot put performance function estimates for doping and non-doping athlete groups. Specifically, the solid and dash-dot lines in **Figure 4** were obtained by averaging the basis function coefficients for athletes belonging to either group, $\hat{f}_g(t) = \mathbf{b}(t)^\top \hat{\theta}_g$, where $\hat{\theta}_g = \frac{1}{\text{Card}(g)} \sum_{i \in g} \hat{\theta}_i$ and $\text{Card}(g)$ is the cardinality of group g . It appears that as a group, doped athletes achieve higher shot put results than non-doped athletes, though credible intervals expand and overlap at the end of the observation period due to lack of data in the later part of the year. Within this model, it is also possible to study the effect of covariates on shot put performance. **Figure 5** shows the estimated trajectories \hat{f}_i for a 30 years old male athlete (dash line), a 23 years old male athlete (solid line), and a 15 years old male athlete (dash-dot line), along with 95% credible intervals. It appears age has an impact on shot put performance, with more experienced athletes outperforming their younger

counterparts. No significant conclusions can be made at the end of the year (after day 200) given that credible bands are overlapping. This is again due to the sparsity of the data, thus increased uncertainty.

To assess the predictive performance of our model, we split the data into a training set of 300 athletes and a test set of 52 athletes. Training and test sets were generated to maintain the same proportion of doped/non-doped athletes in the sample, thus 2 athletes in the test set (3.69% of 52) were randomly chosen from the doped athletes in the sample. All remaining athletes in the test set were randomly chosen from non-doped athletes in the sample. The complete data (shot put performance and doping status) were retained for the training set when running the analysis, whereas the doping status of athletes in test set was held out and predicted. **Figure 6** shows the ROC curve for correct classification of doping status for subjects in the test set. We also report the Area Under the Curve (AUC), with the closer this value to 1, the better the predictive performance. The model shows very good predictive performance, but we underline this result is mostly driven by the correct classification of supposed non-doped athletes as “non-doped,” rather than the classification of the two doped athletes in the test set as “doped” athletes. A discussion is provided in the next section.



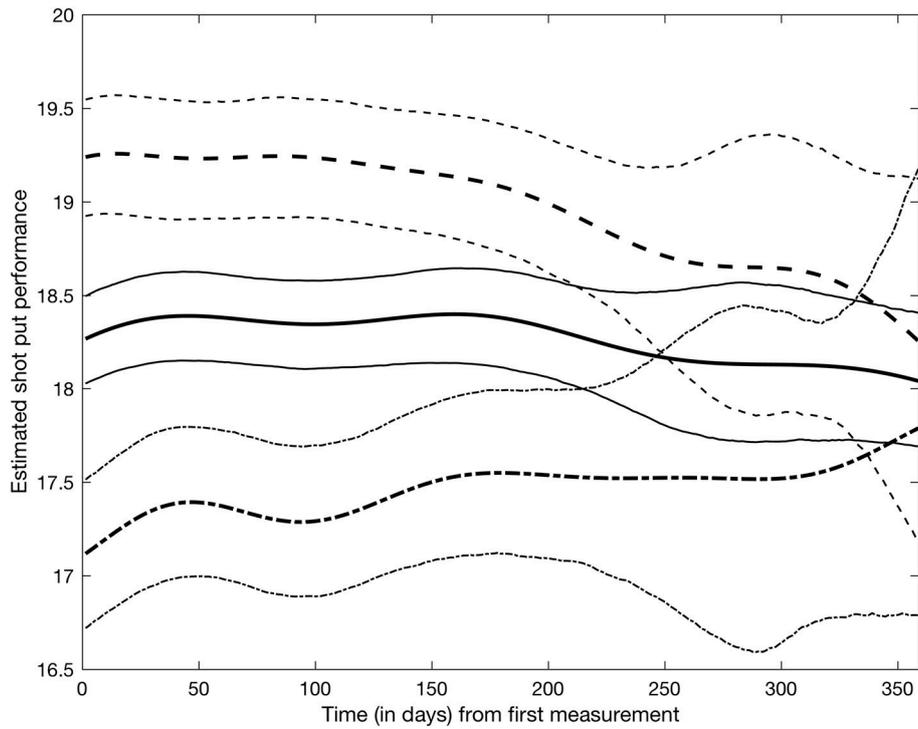


FIGURE 5 | Shotput performance trajectories for a 30 years old male athlete (dash line), a 23 years old male athlete (solid line), and a 15 years old male athlete (dash-dot) along with their 95% credible intervals.

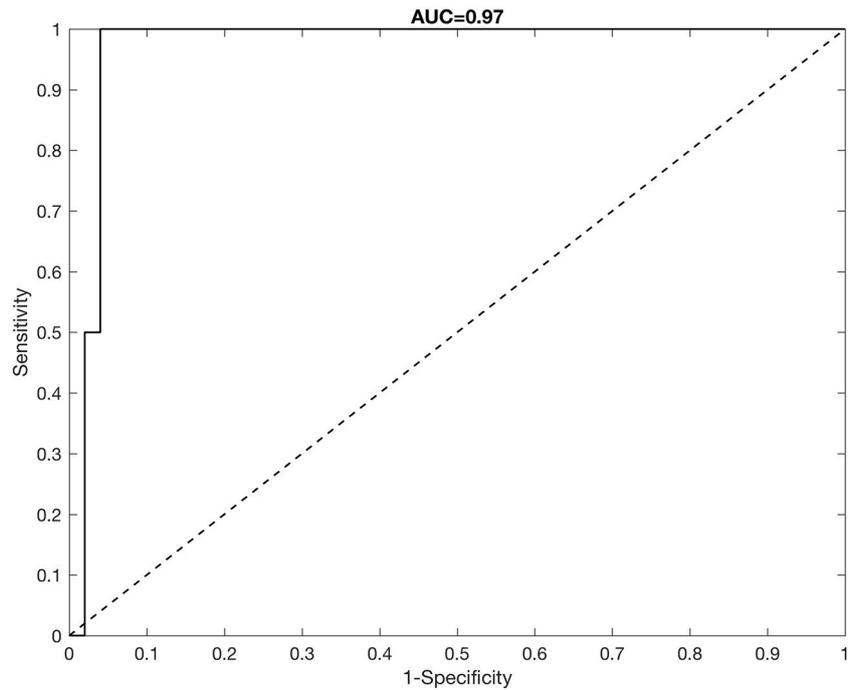


FIGURE 6 | ROC curve analysis for correct classification of doping status for athletes in the test set.

5. DISCUSSION

This study has proposed a Bayesian latent factor regression model for the analysis of performance data used in an anti-doping context. Following (Montagna et al., 2012), we include a high-dimensional set of pre-specified basis functions, while allowing automatic shrinkage and effective removal of basis coefficients not needed to characterize any of the athlete-specific functions. Further, we accommodate joint modeling of a functional predictor, the elite shot put performance trajectory, with a binary response, the doping status, within a framework of scalar-on-function regression. While the motivation of our work comes from the analysis of shot put performance data, the methodology presented in this paper is widely applicable to the analysis of performance data collected in all so-called centimeter-gram-second sports.

The model achieves good predictive performance of doping status, with a high AUC. However, the model prediction is driven by the correct classification of non-doped athletes, thus the error in sensitivity we observe is due to the misclassification of the doped athletes as non-doped. For example, if we classified as “doped” all athletes having $\hat{p}_i > 0.50$ (the random classification rate), none of the truly doped athletes in the test set would be correctly classified as doped. We observed this misclassification problem not only when analysing data limited to 1 year, but also when analysing all available shot put performance data. There could be several reasons that lead to the misclassification. The number of doped athletes is very small compared to size of the sample (13 doped athletes out of 352 subjects), thus it is possible that the signal is too low for the model to be able to detect the doping-status of an athlete. In particular, (Montagna et al., 2012, 2018) show the model has good classification performance of any group in simulation studies with higher signal-to-noise ratio and more balanced representation of different groups in the sample. It might also be possible that some athletes denoted as “clean” in the sample are in fact “doped,” but were never caught through traditional testing methods.

Regardless of potential issues with the data, the model in its current formulation suffers from some limitations. The Bayesian latent factor regression methodology was originally developed for very sparse longitudinal data (Montagna et al., 2012) with the purpose of capturing a global trend in subject-specific trajectories. Instead, the shot put dataset has measurements often collected just a few days apart from each other on each athlete, and for a potentially long number of years. Thus, our dataset

shows more local, short-range variability, which the current version of the model cannot adequately represent. Further, the model does not currently accommodate longitudinal covariates potentially affecting performance, for example, as shown in **Figure 5**, athlete aging has a clear demonstrable effect on shot put throwing distance. Moreover, factors such as individual athlete seasonal training and competition patterns will affect their individual competition results, independently of any doping related effect and so must be accounted for in any longitudinal model of performance. Nonetheless, our current model appears to be able to detect differences between the two populations in the early phase of the competitive year, with about 1 meter difference in performance between doped and non-doped athletes (see **Figure 4**). Therefore, this retrospective data modeling provides an indication that the effects of doping can be identified from longitudinal athlete performance profiles, and that it could potentially be used as a tool to estimate theoretical future credible performances for a given athlete. The ultimate application of this type of modeling approach would be that when the projected credible level of performance is exceeded by an athlete (e.g., **Figure 3**—athlete 53 and 281 data points above the 95% credible interval), they would be identified for target testing via the ABP system. Thus, an athlete performance passport will potentially improve the effectiveness of in-competition anti-doping testing, where currently only placed athletes and a number of randomly selected others, are subjected to scrutiny.

While the above has to be considered as a preliminary analysis affected by some limitations, the model is still credible and might represent a useful tool for longitudinal tracking of athlete performance in order to discriminate expected from unexpected or disproportionate increases. Further research is required in order to increase the structure of the model by incorporating more covariates (e.g., training, maturation, seasonal variation) in order to increase its accuracy and sensitivity, and thus better fit the shot put data.

AUTHOR CONTRIBUTIONS

JH conceptualized the idea; SM developed the methods and conducted the analysis; JH and SM drafted, read and approved the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank Sergei Iljukov and Mirko Jalava for providing the data used in this manuscript.

REFERENCES

- Arminger, G. (1998). A bayesian approach to nonlinear latent variable models using the gibbs sampler and the metropolis-hastings algorithm. *Psychometrika* 63, 271–300. doi: 10.1007/BF02294856
- Hopker, J., Schumacher, Y. O., Fedoruk, M., Mørkeberg, J., Berman, S., Iljukov, S., et al. (2018). Athlete performance monitoring in anti-doping. *Front. Physiol.* 9:232. doi: 10.3389/fphys.2018.00232
- Iljukov, S., Berman, S., and Schumacher, Y. O. (2018). Application of the athlete's performance passport for doping control: a case report. *Front. Physiol.* 9:1102. doi: 10.3389/fphys.2018.00280
- Montagna, S., Tokdar, S. T., Neelon, B., and Dunson, D. B. (2012). Bayesian latent factor regression for functional and longitudinal data. *Biometrics* 68, 1064–1073. doi: 10.1111/j.1541-0420.2012.01788.x
- Montagna, S., Wager, T., Feldman Barrett, L., Johnson, T. D., and Nichols, T. E. (2018). Spatial Bayesian latent factor regression modeling of coordinate-based meta-analysis data. *Biometrics* 74, 342–353. doi: 10.1111/biom.12713

- Schumacher, Y. O., and Pottgiesser, T. (2009). Performance profiling: a role for sport science in the fight against doping? *Int. J. Sports Physiol. Perform.* 4, 129–133. doi: 10.1123/ijspp.4.1.129
- Sottas, P. E., Robinson, N., Saugy, M., and Niggli, O. (2008). A forensic approach to the interpretation of blood doping markers. *Law Prob. Risk* 7, 91–210. doi: 10.1093/lpr/mgm042
- WADA (2016). *Technical Document for Sport Specific Analysis: Version 3.0*. Montreal, QC: Technical Report, World Anti-doping Agency.
- Zorzoli, M., Pipe, A., Garnier, P. Y., Vouillamoz, M., and Dvorak, J. (2014). Practical experience with the implementation of an athlete's biological profile in athletics, cycling, football and swimming.

Br. J. Sports Med. 48, 862–866. doi: 10.1136/bjsports-2014-093567

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Montagna and Hopker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.