Exploring Effect of Vaporfly Shoes in Marathon Performance Using Semi-parametric Hierarchical Models

Hunter Jiang (hjiang24@ncsu.edu)

Abstract: In this report, we compare and then select a semi-parametric model to predict an athlete's performance on marathons through cross-validation. Wearing Vaporfly shoes will improve their performance, and such improvement will vary across the course(greatest), runner, and gender(mildest). Except for other effects, our model shows wearing Vaporfly shoes will improve 2.16 to 4.32 minutes' marathon performance on average.

1 Introduction

Vaporfly, as Nike Corporation's new racing shoes, has made a huge slash in the marathon running community. Guinness et al. [2] have reported speed improvements for runners wearing these shoes in marathon contests. The objective of this study is to determine the scale of the improvement and whether the effect varies across (1) gender, (2) runner and/or (3) course.

Viewing our task as a data exploitation process, we extract runner's age form the "name_age" column. As for missing ages, we fixed 408 records based on the given data. E.g. if a runner's $age = 25$ in 2008, we plug $age = 26$ into her/his 2009's course record. Combining age with others, we have 5 explanatory variables: Age (X_a, numeric), Runner (factor), Course (factor), Gender (factor), and Vaporfly(binary). The response variable is time in minutes (Y, numeric).

On average, from 1502 non-missing observations, we have 16 records for each course, 3 records for each athlete, and 13 percents of observations that wearing Vaporfly shoes. Consider the scale and availability of the data, we implement a random effect model [5] to factor variables, because it can catch the effect of each factor with acceptable degrees of freedom. Wimmer et al. [6] reported a clear non-linear association between age and performance in the decathlon, a combination of track and field events. Interested in whether age has a similar property in marathon performance, we applied semi-parametric basis functions to test it.

In general, we construct different models and simulate their performance, and select the best one through the MSE and DIC criteria. Factors' effect on wearing Vaporfly shoes will be analyzed by comparing the difference of posterior distributions between "full" and "reduced" models via Kolmogorov-Smirnov test [1]. Finally, we will report our findings in the final chapter.

2 Methods

We construct our "full" hierarchical models with the following assumptions. (1). All explanatory variables affect marathon performance. (2). Age has a non-linear effect, and factor variables have random effects, to marathon performance. (3). All effects are additive.

Following (1) $\sim$ (3), our model could be constructed as:

$$Y_{ijk} = \sum_{p=1}^{s} c_p f_p(X_a) + \alpha_i + \beta_j + \gamma_k + \tau + \epsilon,$$

Where $f_p(\cdot)$ is the p^{th} basis function; α_i, β_j and γ_k indicate the random effects of Gender i, Runner j, and Course k; Besides, τ serve as a linear effect of wearing Vaporfly shoes.

We assign Gaussian priors to gender's random effect because there are only 2 levels exist. However, we apply Gaussian and double-exponential priors separately when modeling the effects of runner and course, intending to check whether there are only some runners (courses) that will affect marathon performance. Meanwhile, Gamma and half-Cauchy priors are tested as standard deviation. As a result, we have 8 different models, within the same hierarchical structure, as candidates.

Using a Directed Acyclic Graph (DAG) to illuminate, our hierarchical models could be represented as Fig.1. Note that we need to choose one candidate from line-dot rectangles within the process and prior layer.

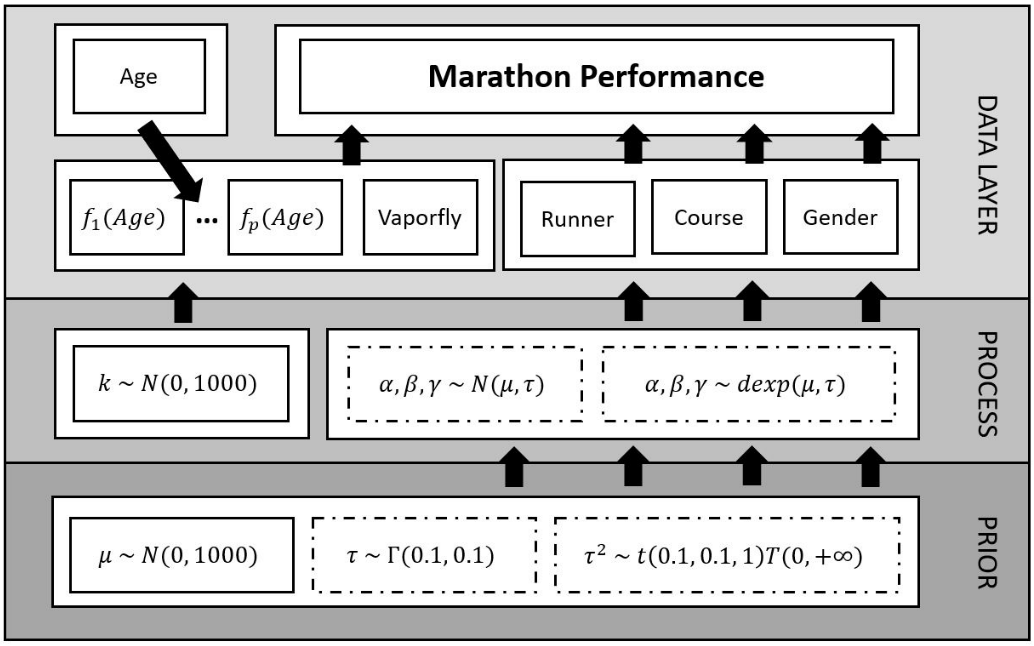


Figure 1: DAG for our hierarchical models

3 Computation

We name 8 different models as **Model_a_b_c**, where **a** and **b** are notation of prior for runner and course effect, which is selected in process layer, and **c** indicates the prior chosen from the prior layer. E.g. "Model_g_d_h" means the model has Gaussian prior for runner effect, double exponential priors to course effect, and both standard deviation will follow half-Cauchy prior.

To select the best model among them, we randomly separate our data into training (90%) and validation set (10%), and evaluate the prediction bias (through MSE), the Deviance Information Criterion (DIC), and Geweke's convergence results.

As for parameters in prior, all normal distribution have mean of 0 and precision of 0.001. $\alpha = \beta = 0.1$ are assigned to Gamma distributions; half-Cauchy distribution will have $x = 0$ and $\gamma = 0.01$. All simulation tests are running through R [4] and JAGS [3], on a Windows 10 PC (I5-9600K/Z370). Codes used in this report are available at: `https://github.com/HunterJiang97/VaporflyEffect/`

4 Model comparisons

We run models with pre-allocated settings shown in section 3, and the best performances within each index are highlighted. Given the results in table 1, we decide to choose model_g_g_h, which utilizes Gaussian prior to random effects and half-Cauchy prior to standard deviations.

	MSE	DIC	Con.	ES		MSE	DIC	Con.	ES
Model_g_g_i	29.26	**8424**	8.4%	22.3%	**Model_g_g_h**	**29.19**	**8424**	5.8%	17%
Model_g_d_i	30.86	8501	7.9%	8.4%	Model_g_d_h	30.87	8500	**3.0%**	5%
Model_d_g_i	30.15	8440	30.4%	3.4%	Model_d_g_h	30.05	8440	13.8%	3%
Model_d_d_i	30.15	8432	12.6%	1.1%	Model_d_d_h	30.30	8432	5.7%	**1%**

Table 1: Simulated Model Performance

Then we applied model_g_g_h to the whole data set, in order to fully exploit everything we have, with $20K$ more iterations. Self-validation shows such a model has MSE of 15.33, with DIC = 9297. Within all parameters, only 2 of them has effective samples less than 1000, and 1 parameter doesn't pass Geweke's convergence test.

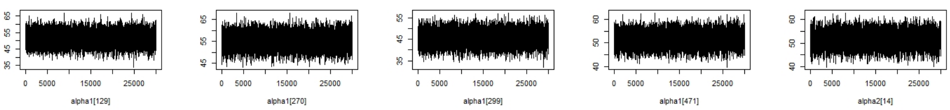
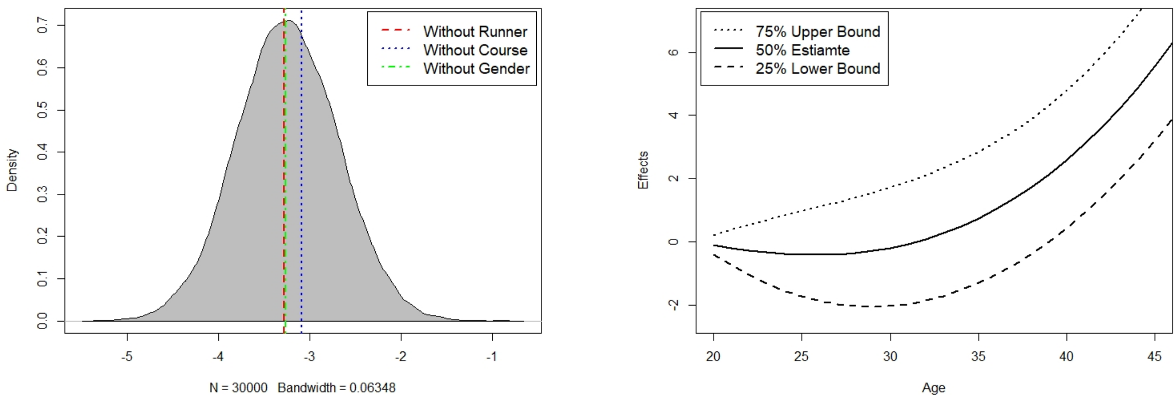


Figure 2: Five randomly selected sample traces

5 Results

We visualize the effect of Vaporfly and Age to the marathon performance in Fig.3. Three means of estimated τ in reduced models with one factor less than our best model are shown in (a).



(a) Posterior density plot of Vaporfly effects (b) Trust bounds of Age effects

Figure 3: Results of Semi-parametric Hierarchical Models

Three two-sided Kolmogorov-Smirnov tests are conducted to the effect of Vaporfly between the full model and reduced model, which shows all 4 samples, statistically saying, comes from a different distribution. Thus, we believe all three effects will affect the improvement when wearing Vaporfly shoes, and the ”scale” of effect could be sorted from smallest to largest as Course, Runner, then Gender.

	Reduced	D-value	p-value
1	Runner	0.05	9.6e-10
2	Course	0.11	2.2e-16
3	Gender	0.03	5.3e-4

Table 2: Test results between full and reduced models

Conclusion: Wearing Vaporfly shoes will improve athlete’s performance in marathons, and such improvement will vary across course, runner, and gender. Among these factors, the course will have the greatest impact, which might indicate Vaporfly shoes have more advantages in some weather/field conditions. Except for other effects, our model shows wearing Vaporfly shoes will improve 2.16 to 4.32 minutes’ marathon performance on average.

References

[1] W. Conover. *Practical nonparametric statistics*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1980.

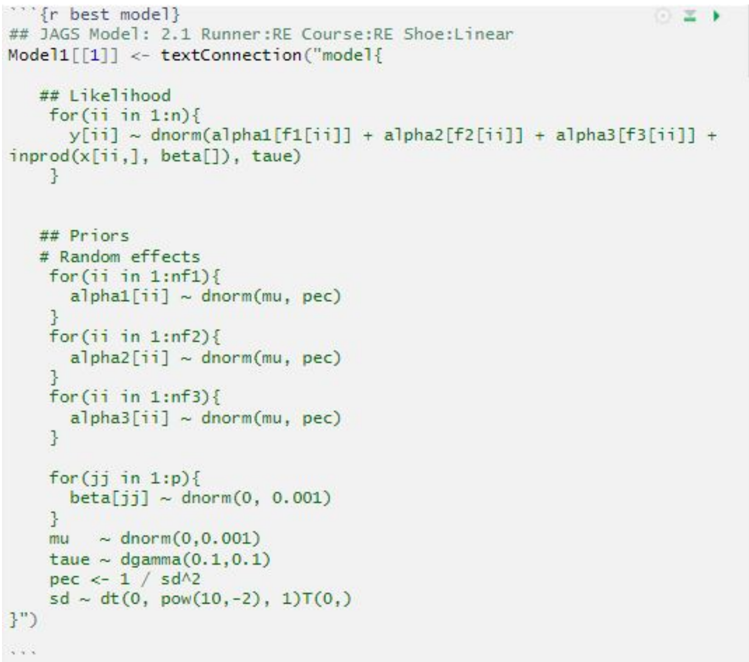
[2] J. Guinness, D. Bhattacharya, J. Chen, M. Chen, and A. Loh. An observational study of the effect of nike vaporfly shoes on marathon performance, 2020.

[3] M. Plummer. *JAGS Version 4.3.0 user manual*. World Health Organization, Lyon, Fance, 2017.

[4] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[5] B. Reich and S. Ghosh. *Bayesian Statistical Methods*. ASA-CRC series on statistical reasoning in science and society. CRC Press, Taylor & Francis Group, 2019.

[6] V. Wimmer, N. Fenske, P. Pyrka, and L. Fahrmeir. Exploring competition performance in decathlon using semi-parametric latent variable models. *Journal of Quantitative Analysis in Sports*, 7(4), 2011.



```
```{r best model}
## JAGS Model: 2.1 Runner:RE Course:RE Shoe:Linear
Model1[[1]] <- textConnection("model{

   ## Likelihood
    for(ii in 1:n){
      y[ii] ~ dnorm(alpha1[f1[ii]] + alpha2[f2[ii]] + alpha3[f3[ii]] +
inprod(x[ii,], beta[]), taue)
    }


   ## Priors
   # Random effects
    for(ii in 1:nf1){
      alpha1[ii] ~ dnorm(mu, pec)
    }
    for(ii in 1:nf2){
      alpha2[ii] ~ dnorm(mu, pec)
    }
    for(ii in 1:nf3){
      alpha3[ii] ~ dnorm(mu, pec)
    }

    for(jj in 1:p){
      beta[jj] ~ dnorm(0, 0.001)
    }
    mu    ~ dnorm(0,0.001)
    taue ~ dgamma(0.1,0.1)
    pec <- 1 / sd^2
    sd ~ dt(0, pow(10,-2), 1)T(0,)
}")
```
```

Figure 4: JAGs Codes of our best model