Assessing heterogeneity (and predictability ??) of runners' performance in Switzerland

Gianrocco Lazzari[1,*], Stefano Savaré[2], Antonio Iubatti[2], Maxime Peschard[2], Ondine Chanon[2], Michele Catasta[3], and Marcel Salathé[1]

[1]Global Health Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
[2]School of Computer Science, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
[3]Department of Computer Science, Stanford University, Stanford, USA.
[*]gianrocco.lazzari@epfl.ch

ABSTRACT

keywords: Aging, Distance running, Endurance performance, Sex difference

Introduction

as show in[1] blabalblab

Results

Demographics

In fig. 1(a) and 1(b) we show respectively how the number of runners increased in the last 15 years, by distances and gender. This raise was steeper for man than for women (fig. 1(b)), and steeper in the shorter distance (10 Km) than in the longer ones (fig. 1(a) - participants in full marathons seems to have decreased though) [1].

For a significant analysis of runners' performance, it is important to check the amount of data present in the various events, and therefore first assess the heterogeneity of events popularity. In fig. 2(a) we show the distribution of number of editions each event was hosted, across all history. Counting all editions of all races as independent, we recorded 222 events. More interestingly in fig. 2(b) one can see the broad distribution of number of participants in the different events. In particular, a power-law fit ($f(x) \sim x^{-\alpha}$) provides an exponent of $\alpha = 1.69 \pm 0.05$[2].

Inversely, one can measure how participative runners have been across Switzerland. In fig. 3(a) we show the distribution[3] of the number of events to which each runners participated. We collected data from a total number of 531426 runners.

The case of Lausanne Marathon

We consider the case of 2016 Lausanne Marathon event to present some relevant statistics on age and performance distribution, due to his popularity across all Switzerland (see indeed also fig. 9 for the distribution of runners' origin towns). Apart from the 812 kids (younger than 14 years) running a shorter race, the 2016 edition had more than 10 thousands participants, divided per race and gender, as shown in table 1.

[1]For simplicity we only include the most popular distances. There are many events that include shorter distances, like 3 Km, 5 Km, usually attended by a small fraction of young runners.
[2]The power-law starts from a lower-bound, whose value results as well from the fitting procedure: $x_{min} = 688$ runners/race
[3] One can see that a log-normal would fit better in this case than a power-law model. In particular the fitted parameters are $\mu = -0.70, \sigma = 1.55$

Category	Number of participants
10 km	5515
half-marathon	4414
marathon	1318

Gender	Number of participants
M	6905
F	4655

Table 1. Distribution of number of participants in 2016 Lausanne Marathon, by distance and gender.

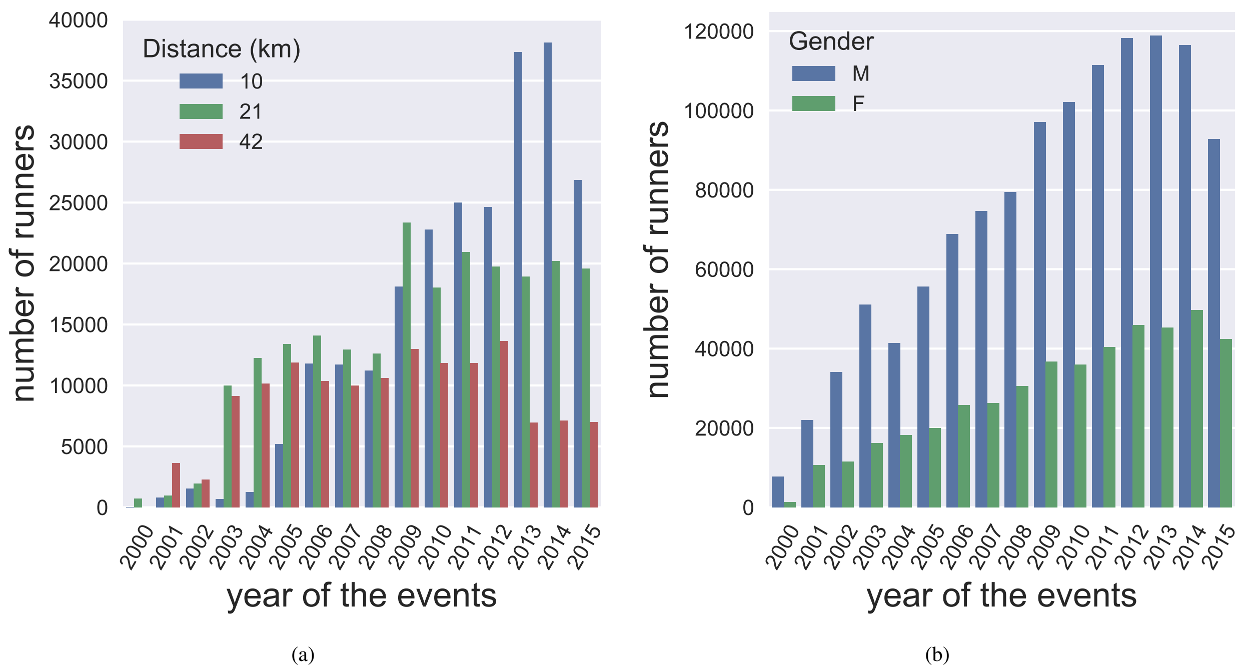


Figure 1. Number of participants in running competition in Switzerland, across time, by distance (a) and gender (b).

In fig. 4(a) and 4(b) we show the distribution of runners' ages in the three main races and divided by gender. For women, even if the median age increases with the distance, begin respectively 34 y.o. (10 Km), 36 y.o. (21 Km) and 38 y.o. (42 Km), only the women running the 10 Km were significantly[4] younger. For men, median ages also slightly increases with distance 39 y.o. (10 Km), 39 y.o. (21 Km) and 42 y.o. (42 Km), but only the men running the full marathon are significantly older. Furthermore, women are significantly younger than men in all three races.

In fig. 5(a), 5(b) and 5(c) we compare the distribution of performances by gender, for each race. We found the men were significantly faster in all races.

It is also worthy to compare the *pace*-π (min/Km) across different races. One of course would expect that longer races require higher pace ($\pi_{10} < \pi_{21} < \pi_{42}$). These differences were found to be significant for men 6(b) for all races and for women running 6(a) the marathon (difference $\langle \pi_{10} \rangle - \langle \pi_{21} \rangle$ for women was not significant)

Overall performance analysis

Age-performance relation

We wanted to check whether the previously found[1–4] U-relation between age and performance holds as well in the case of swiss races. In fig. 7 (**plot to be improved!**) we show the dependence of runners' performance on age, for four of the most popular swiss marathons. The above-mentioned U-shaped although still slightly appearing in the longest distance (42Km), it does emerge more clearly in the half-marathons, as shown indeed in fig. 8 (**plot to be improved!**)

Temperature-performance relation

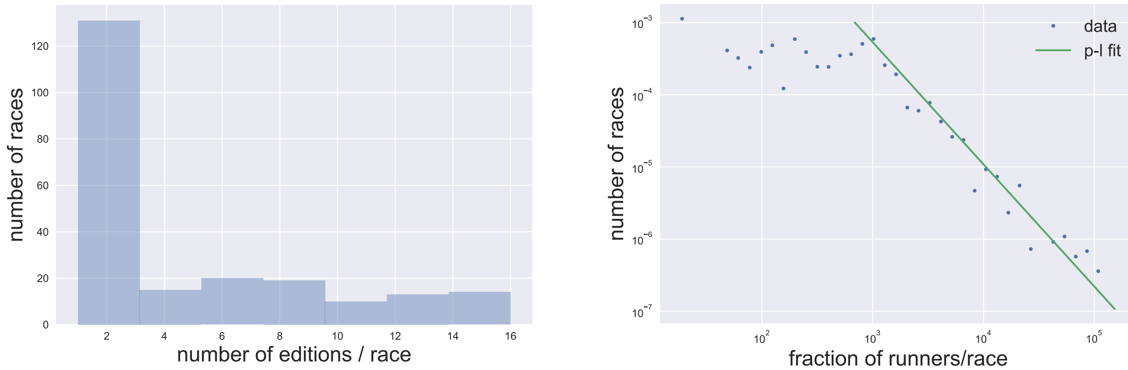
we don't have enough data (can be re-checked)...

some reviews on the topic:
`http://runningstrong.com/temperature.html`
`http://believeperform.com/performance/the-effects-of-heat-on-sport-performance/`

Geographical analysis

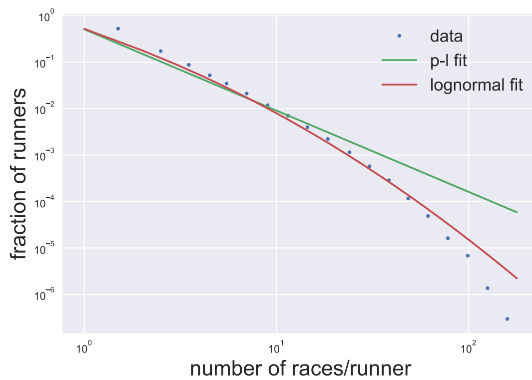
(to be included ??)(by Antonio & Gr)

[4] All reported tests are done using two-tailed p-value on Kolmogorov-Smirnov test $p \leq 10^{-3}$.



(a) (b)

Figure 2. Assess popularity of running competitions, in Switzerland.



(a)

Figure 3. Assess how participative runners are across competitions, in Switzerland.

It is interesting to assess from which cities runners come, and how many of them, from the different locations. In fig. 9 we report as an example the quite broad[5] distribution of number of runners, coming from the 2005 municipalities reported for the 2016 Lausanne Marathon.

Network of runners

(to be included ??) (by Gr)

Forecast of career advancement (??)

(not done yet)
nice article on fivethertyeight, pointing to one of the best/latest model[5]

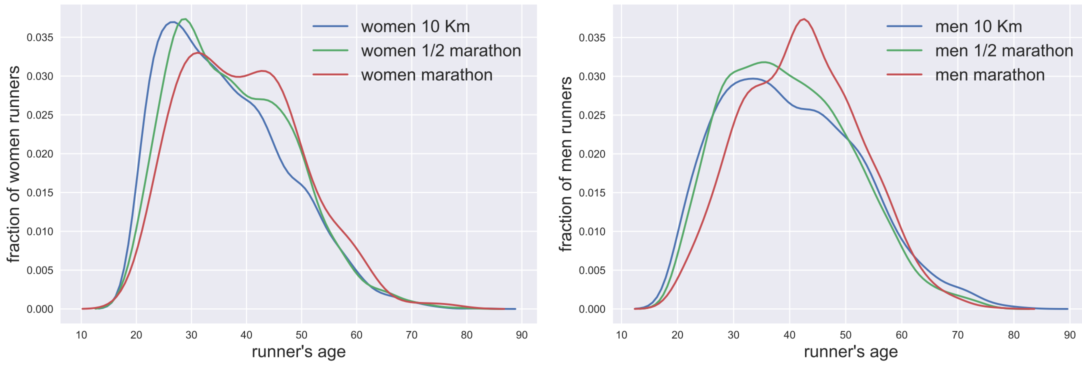
Discussion

Methods

Data parsing

@stefano (remember to add the *definition of runner*)

[5]Fitted exponent: $\alpha = 1.90 \pm 0.03$



(a) (b)

Figure 4. Runners' age distribution in 2016 Lausanne Marathon, for women (a) and men (b) and divided by race (color-coded). For clarity we only plot the Gaussian kernel estimate of the distribution.

Data analysis

All analysis were performed on python notebooks (available on the related repository), using standard python packages for data analysis and plotting, such as `pandas`, `seaborn`, `scipy`, `powerlaw`[6] and `networkx`.

Data visualization

We implemented interactive visualizations of some of our results and collected them in the Hop Suisse[7] website. After exporting the data needed for the plot in `.json` dumps, we used C3.js for the interactive plotting. More details on how datasets queries and plots were built can be found on the dedicated GitHub repository[8]. We also build an animated infographics[9], inspired by Hans Rosling's work. With such video we wanted to show in a more powerful and clear way the relations among runners' mean pace, experience and age, providing as well information on gender and race length (the python code used to construct the video frames can be found in the related folder[10] of our GitHub repository).

Author contributions statement

G.L. and A.I. performed the data analysis. S.S., O.C. and M.P. performed the data parsing. G.L. and S.S. wrote the manuscript. M.C. and M.S. review the manuscript.

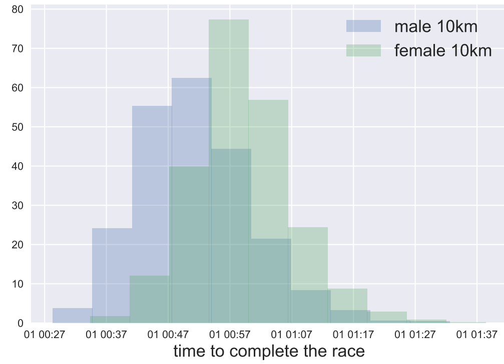
Additional information

All the code used to parse the data from https://www.datasport.com/en/, for data analysis and visualization can be found in our open GitHub repository: https://github.com/ggrrll/hop_suisse_ada_project_public.

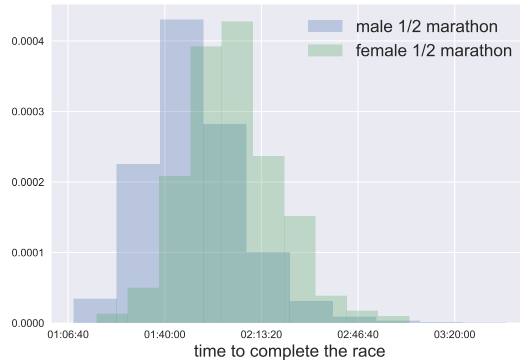
Competing financial interests

The authors declare no conflict of interests.

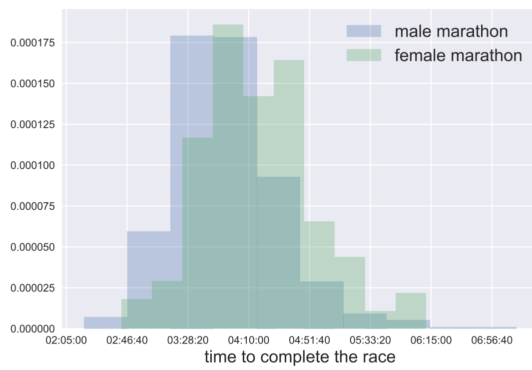
[6] https://pypi.python.org/pypi/powerlaw
[7] https://hopsuisse.github.io
[8] https://github.com/hopsuisse/hopsuisse.github.io
[9] https://www.youtube.com/watch?v=MyvbnOXHShw
[10] https://github.com/ggrrll/hop_suisse_ada_project_public/tree/master/8-video



(a)



(b)

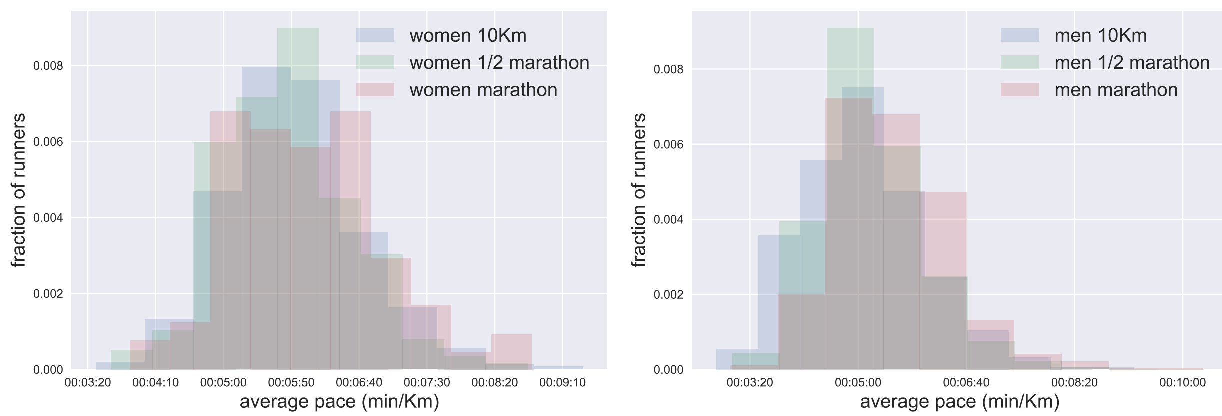


(c)

Figure 5. Runners' performance distribution in 2016 Lausanne Marathon, for 10 Km (a), half marathon (b) and marathon (c), divided by gender (color-coded).

References

1. Connick, M. J., Beckman, E. M. & Tweedy, S. M. Relative age affects marathon performance in male and female athletes. *J. sports science & medicine* **14**, 669 (2015).
2. Knechtle, B., Assadi, H., Lepers, R., Rosemann, T. & Rüst, C. A. Relationship between age and elite marathon race time in world single age records from 5 to 93 years. *BMC sports science, medicine rehabilitation* **6**, 31 (2014).
3. Lara, B., Salinero, J. J. & Del Coso, J. The relationship between age and running time in elite marathoners is u-shaped. *Age* **36**, 1003–1008 (2014).
4. Lehto, N. Effects of age on marathon finishing time among male amateur runners in stockholm marathon 1979–2014. *J. Sport Heal. Sci.* **5**, 349–354 (2016).
5. Vickers, A. J. & Vertosick, E. A. An empirical study of race times in recreational endurance runners. *BMC Sports Sci. Medicine Rehabil.* **8**, 26 (2016).



(a) (b)

Figure 6. Runners' pace distribution in 2016 Lausanne Marathon, for women (a) and men (b), divided by distance (color-coded).

Figure 7. Relation between runners' performance (time in minutes to complete the race) and age, for the most popular marathons, color coded by gender.

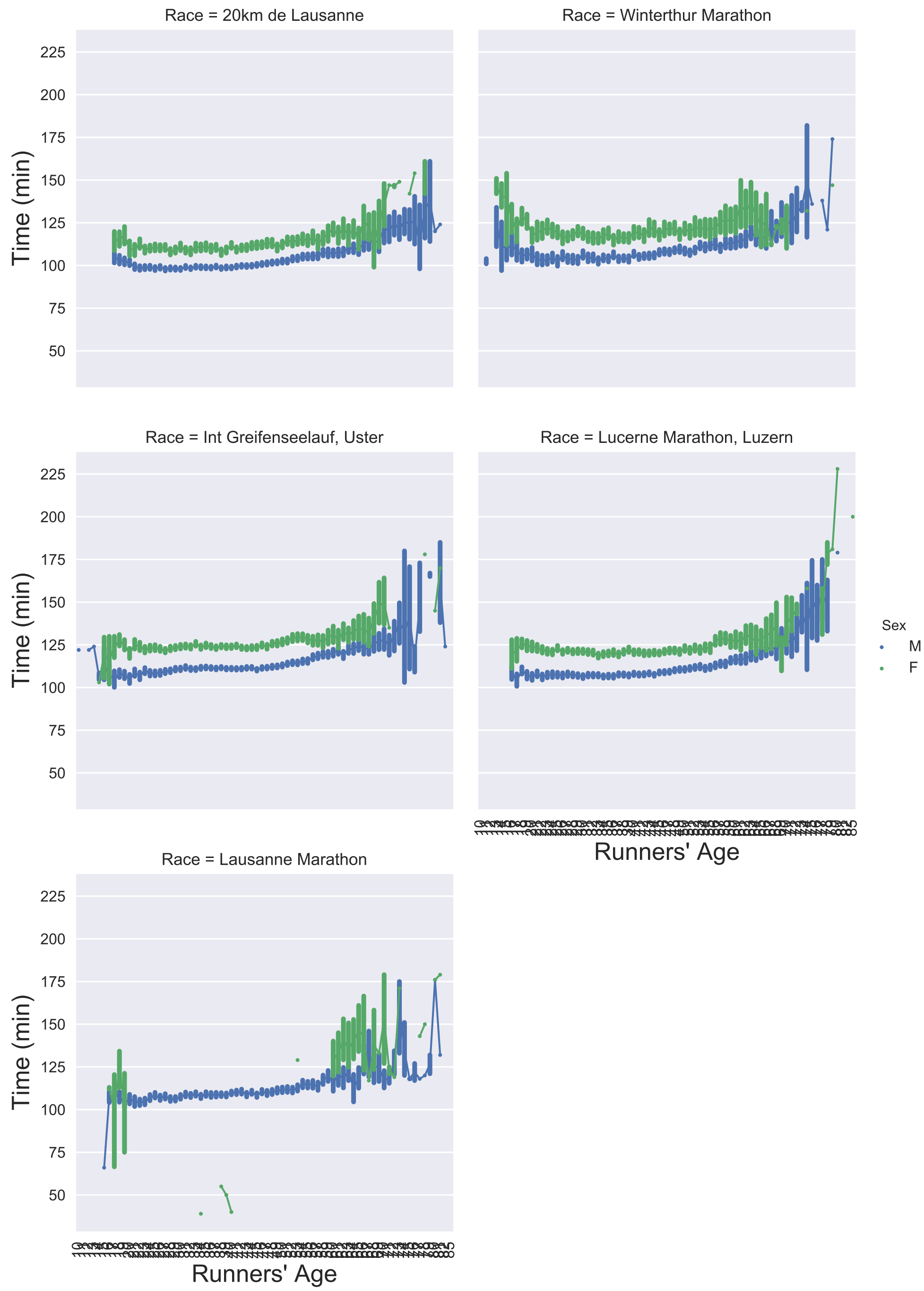
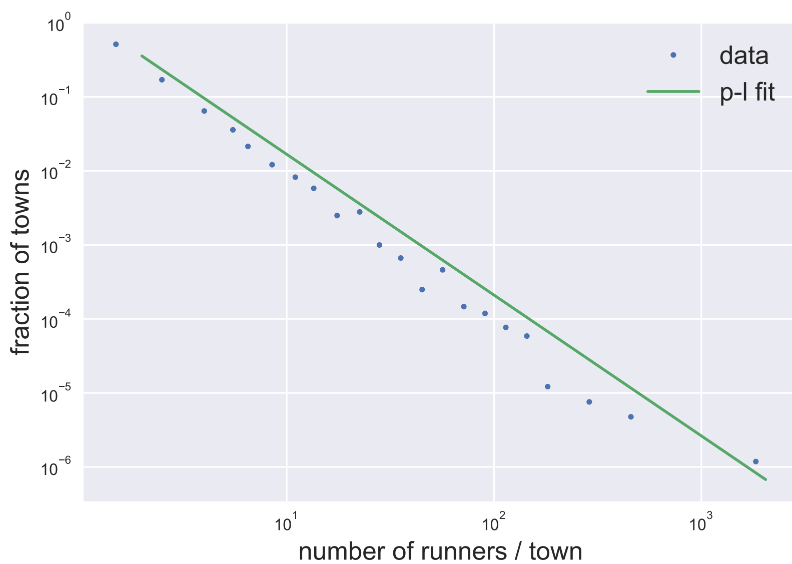


Figure 8. Relation between runners' performance (time in minutes to complete the race) and age, for the most popular half-marathons (20 Km and 21 Km), color coded by gender.



data
p-l fit
fraction of towns
number of runners / town
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{1}
10^{2}
10^{3}

Figure 9