Article

Marathon performance depends on pacing oscillations between non symmetric extreme values.

Jean-Renaud Pycke [1,‡], Véronique Billat [2,‡]

[1] LaMME CNRS (UMR 8071); University of Paris-Saclay (Evry)
[2] Department of STAPS ; University of Paris-Saclay (Evry)
* Correspondence: jeanrenaud.pycke@univ-evry.fr
‡ These authors contributed equally to this work.

Abstract: The marathon has been recently run in less than 2 hours by the man who ran the three fastest marathons ever recorded in the span of three years, by Eliud Kipchoge the Tokyo Olympic gold medals. Here we demonstrate that the best marathons were run according to a pace distribution that is statistically not constant and has a pace distribution with negative asymmetry. The concept of mirror race enables us to show that the sign of asymmetry is not due to sampling fluctuations. We show that marathon performance depends on pacing oscillations between extreme values and that even the best marathons ever run differ and can be improved upon. The utilization of extreme values and oscillations allows recovery and optimization of the complementary aerobic and anaerobic metabolisms. Our findings suggest new ways forward to approach the pacing for optimizing endurance performance.

Keywords: Extreme values ; symmetry breaking; pacing strategy ; optimization

Citation: Pycke, J.-R.; Billat, V. Marathon performance, oscillations between asymmetric extreme values. *Int. J. Environ. Res. Public Health* **2022**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

0. Introduction

For centuries, the limits of physiology and athletic records have fascinated scientists. In 2019, the sub-two-hour marathon barrier was broken with a time of 1:59:40.2. In 1925, Nobel laureate, Archibald Vivian (AV) Hill, published "The Physiological Basis of Athletic Records" [1].

Nowadays, one century later, Advances in wearable sensor technology have enabled real-time measurement of physiological data during exercise ([2]). Future directions in training are going to be about encouraging the marathon runner. To achieve this, we must provide ways of improving satisfaction regarding training progress and with the feeling that it has been optimized. The results show that according to our hypothesis We are testing the hypothesis that the ideal race, with the world's best marathon runners, can have several degrees of optimization. This can be observed in Kipchoge's last 5 marathon's, since Kipchoge ran the three fastest marathons at the time in the span of three years, and even winning Olympic gold medals in the process (2016, 2021)

Hill's and Kennelly's approach to running the fastest marathon was based on the constant pace paradigm.

In Hill's reference to Kennelly's 1906 paper, he stated that the ideal way to run a race was not necessarily to win, but to achieve a new athletic record for the distance, and to do it by running it at constant speed. Fifty years later, the physicist Joseph Keller modelled the predominating view that a runner should maintain constant pace to achieve the shortest time in his paper: "A theory of competitive running" [3]. To determine the optimal race strategy, he used simple physics and mathematics to correlate the physiological attributes of runners with world track records. He derived the optimal speed variation $t \mapsto v(t)$ by formulating and solving a problem in optimal control theory. For distances greater than 291 meters, his theory predicted a maximum acceleration for one or two seconds, then constant speed throughout the race, until the final one or two seconds, and finally, a slight slowing down. At the time, his results confirmed the recognized view that a runner should maintain constant speed to achieve the shortest time (Keller, 1973). Importantly, Keller admits his theory omitted several important variables, such as the up-and-down motion of the limbs, internal and external resistance, the depletion of fuel, and the accumulation and removal of waste products. These ideas were perhaps valid in 1925, when the fastest marathon time was only 2h29min (by Albert Michelson, an average speed of 16.99 km/h, Port Chester, USA).

In the same way, Arthur Kennelly desired to measure all points of a race. Delving deeper in Kennelly's paper: "It is to be noted that all these speeds are average speeds over the courses. There is no evidence among the records to show what the speed was at different points in the course. So far as concerns anything appearing in the data, the speed of a runner, for example, which averages 7.17 meters per second over a 1-kilometer course, might be 10 meters per second in the first part and 5 in the last part, or vice versa. Evidence is lacking to show what the facts are, and they are of great importance to the science of athletics. The speed of a world's record type of trained runner might be determined at any or all points of a course, either by securing a light recording chronograph on the back of his belt, with a thread paid out as he ran, or by pacing the runner with a light motor-car carrying an automatic speed " ([4], p. 328). Nowadays, thanks to microchip technology, it is possible to measure every point during a marathon.

Physiologic processes are inherently never constant. French physiologist Claude Bernard wrote, in his classic book [5] that: "The use of averages in physiology and medicine most often gives only a false precision to the results by destroying the biological character of the phenomena."

In the same way, the observed paces in a marathon are far from constant at the average speed. The best marathons are run with an asymmetric distribution of paces ([6]).

Nowadays, it is possible to test the hypothesis that the best marathon performance is run with variable pace and challenge the paradigm of constant pace. Statistically, a race is run with an oscillating pace between extreme values that are inter-played by optimizing fuel, recovery, and avoiding VO2 and heart rate drift. Using mathematical statistics, we aim to show that an asymmetric distribution is optimal for marathon running, to delve deeper into the pace distribution, and to find the exact outline of the optimal pacing signature by analysing the official best world marathon performances by Dennis Kimoto (Berlin 2014) and Eliud Kipchoge (Monza 2017, Berlin 2018, 2019, Vienna 2019, and the Tokyo Olympics Games, 2021). Furthermore previous marathon studies reported, also run in actual conditions, that physiological oscillations were the subjacent of real pace variations, even when the marathon pace is planned to be constant ([6]).

The Berlin marathons were official competitions and are official world records. While Monza 2017 and Vienna 2019 were exhibition marathons accomplished with rotating pacers, an electric pace vehicle, a laser beam projecting the ideal position on the road and ideal conditions. At Vienna, Kipchoge ran at a consistent average pace of 2:50 minutes per kilometre (4:33.5 minutes per mile) and was 11 seconds ahead of schedule halfway through the marathon. He even accelerated in the final kilometre. Given that the best marathon performances, including the Olympics, were run by a single man (Eliud Kipchoge), this shows the possibility of adding a new dimension of running performance beyond the chronometer and proposing a qualitative way of optimizing the physiological capacity for running a marathon at 21 km/h. Our approach represents a powerful possibility for the marathon runner to validate the optimal aspects of his/her performance. However, before exploring the possibilities of providing feedback to a marathon runner, we must test the consistency of such a model in Eliud Kipchoge, who is purported to have a robust " performance template." They have shown this to be primarily related to the increased confidence that the distance in question can be completed without unreasonable levels of exertion or injury ([7]). This can apply to real-world conditions with the aid of mathematical modelling and wearable technology as already proposed also in the journal Nature by the physicist Emig Thorsten, see [2].

1. Materials and Methods

Our data consists of six marathon races: Berlin 2014, Monza 2017, Berlin 2018, Berlin 2019, Vienna 2019, and Tokyo 2021. They were performed by Dennis Kimetto, D.K. (Berlin 2014), Eliud Kipchoge, E.K. (Monza 2017, Berlin 2018, Vienna 2019 and Tokyo 2021), and Kenenise Bekele, K.B. (Berlin 2019).

For a given marathon our data were collected from http://run.hwinter.de, accessed on 01/09/2021.

The original dataset includes marathons that provide timing data for 1km race segments (plus the final 41–42.195km segment); the requirement for 1km segments is based on the need to track changes in pacing during different stages of the marathon. The accuracy of the data is, then, higher than those collected by strava from the individual GPS measurement which depend on GPS models (between 0.47 and 1.65%, see [8]). Thus we will assume our data consist of a sequence, denoted by $(p) = (p_i)_{1\le i\le 42}$, where p_i is the pace, i.e. the duration measured in seconds, taken to cover the ith kilometre, all values being integers.

Following standard notations from the field of probability and statistics, if this sequence is arranged in order of magnitude and rewritten as

$$p_{(1)} \le \cdots \le p_{(42)}, \tag{1}$$

then $p_{(i)}$, for $1 \le i \le 42$, is called the ith order statistic (see §11.4 in [9], or §1.1 in [10]).

We call *average pace* the real number $\overline{p}$ computed from the sequence by the two equivalent relations

$$\overline{p} = \frac{1}{42}\sum_{i=1}^{42} p_i, \ \sum_{i=1}^{42}(p_i - \overline{p}) = 0 \tag{2}$$

It is important to note that this value may differ slightly from what can be called the *official average pace* obtained by dividing the official total time by the distance 42.195 km. On our data this only occurs for for Vienna 2019. The resulting difference may, as usual, come from the phenomenon of rounding error, with the result that a total of rounded numbers is not equal to the rounded version of the original total. Furthermore in our case, a rather precise value of the sum of the paces (minus the time taken to run the last 195 metres) is known as the total time. The difference may also, in the case of a marathon, be accounted for, at least partially, by a very fast, or slow pace during the last 195 metres, that are not included in our measurements.

Since our aim is to provide computational methods for big data that have more available, we will therefore describe the stages of a statistical study starting after the sequence $(p_i)_{1\le i\le 42}$ has been produced.

The standard deviation $\sigma = \sigma(p)$ of the pace sequence is defined to be

$$\sigma^2 = \frac{1}{42}\sum_{i=1}^{42}(p_i - \overline{p})^2. \tag{3}$$

A coefficient of variation used extensively is Karl Pearson's dimensionless coefficient of variation given by

$$V = 100\,\frac{\sigma}{\overline{p}}, \tag{4}$$

and is one of the the standards for quantifying the relative pace variation.

In our study, a fundamental role will be played by the skewness coefficient

$$\gamma_1 = \frac{\sum_{i=1}^{42}(p_i - \overline{p})^3/42}{\{\sum_{i=1}^{42}(p_i - \overline{p})^2/42\}^{3/2}}, \tag{5}$$

see formula (3.89) in [9].

Computations from our data are summarized in Table 1. The races are not listed chronologically, but according to the performance and the sign of γ_1.

It might seem, in view of the small values of V that the two performances, Vienna 2019 and Monza 2017, approached the uniform pace strategy. We will see, in fact, that these two performances and neither the other four marathon were run uniformly. This is discussed in the next paragraph.

2. Results

2.1. *The pace is not uniform*

As in everyday language, uniformity of the pace will mean for us that the race is run at an approximately constant pace. Mathematically the sequence (p) is in this case, a constant, or a sample of measurements drawn from a distribution highly concentrated around the average pace, for example a normal distribution with mean $\overline{p}$ and very small variance.

Non-uniformity of the pace can be measured by the range of the observed pace distribution, i.e. the difference between the greatest and smallest values observed, given by

$$r_1 = p_{(42)} - p_{(1)}. \tag{6}$$

Non-uniformity of the pace may also be measured by the difference between the fastest and the slowest 10 kilometres,

$$r_{10} = \sum_{i=1}^{10} \{p_{(43-i)} - p_{(i)}\},$$

or by the difference between the fastest and the slowest half-marathon

$$r_{21} = \sum_{i=1}^{21} \{p_{(43-i)} - p_{(i)}\} = (p_{(22)} + p_{(23)} + \cdots p_{(42)}) - (p_{(1)} + p_{(2)} + \cdots + p_{(21)}),$$

Note that the well-known negative split is given by formula

$$r'_{21} = \sum_{i=1}^{21} \{p_{43-i} - p_i\} = (p_{22} + p_{23} + \cdots p_{42}) - (p_1 + p_2 + \cdots + p_{21}),$$

so that the sum is calculated from paces ordered chronologically, whereas our sums in r_{21} is performed after reordering the paces. Our ranges are known in the field of mathematical statistics as linear rank statistics.

The ratios

$$V_1 = \frac{r_1}{\overline{p}},\ V_{10} = \frac{r_{10}}{10\overline{p}}, V_{(21)} = \frac{r_{10}}{21\overline{p}} \tag{7}$$

are the corresponding variation coefficients for 1, 10 and 21 kilometres. The values are given in Table 1.

In view of this Table, it is apparent that none of the six races manifests a range attributable to random fluctuations about an average target speed. In other words, none was run at an approximately constant pace, in spite of the fact that two of them, Monza 2017 and Vienna 2019 were planned to be run at a constant pace. The smallest ranges are achieved (as expected at Monza) at races designed to be run at a constant pace. As a matter of fact, 8 seconds, even for a non elite runner, is a significant time-difference for a single kilometre, as well as 43 s for 10 kilometres.

The most spectacular variation occurs for Tokyo 2021's r_{21}, with more than 4 minutes between the fastest and the slowest half-marathon in the same race.

2.2. The pace is not symmetric

Once the spread of the pace has been shown to be significant, it is legitimate to raise the issue of its shape, in particular its symmetry or lack of symmetry.

A first insight into the general shape of our distribution is provided by a diagrammatic representation of the data, see the bar charts of grouped data in Figures 1-4 and Figure 5.

A striking feature shared by the six races is that they belong to the class of moderately asymmetric bell-curves (see [11], §4.20 and Figure 4.7 p. 83-87): the class frequencies present one maximum, and fall rapidly on one side of the maximum when compared to the other.

Note that for asymmetric distributions with negative skewness, the median value is usually greater than the mean, which means in our case that the runner will run more than half of the race with a pace above the average pace value, making the race more comfortable. We discussed this aspect in [6].

Statistically, we can assess that our sequences of paces do not behave as random samples from populations (in the sense used in mathematical statistics, see §1.1 and Chapter 9 in [9]) with symmetric, for example normal, distributions. In this case γ_1 would be close to zero, up to sampling fluctuations. Large positive or negative values of γ_1 indicate a departure from normality, or symmetry.

To this end, we make use of Zar's statistic denoted by z (see p. 115-116 in [12], or Equations I.23 p. 21, (4.2)-(4.8) p. 227-228 in[13], restated in relations $(10) - (16)$, to be applied with $n = 42$).

In the case of a normal distribution for (p), then z, computed from our γ_1, would approximately follow the distribution of a standard random normal variable, denoted by $\mathcal{N}(0, 1)$.

For the six races, z takes values fairly, or even significantly different from zero. A positive skewness of the pace is associated with the failed performance of Monza 2017.

A negative skewness of the space is associated with the successful performances.

Furthermore, when the skewness is negative, we observe the final times are better resulting in a higher (in absolute value) skewness (see Table1)

The skewness may also be measured in the way the range r_1 splits into a left range $\overline{p} - p_{(1)}$ (difference between the average pace and the fastest kilometre's pace) and a right range $p_{(42)} - \overline{p}$ (difference between average pace and the slowest kilometre's pace). The difference between these two ranges, say $\Delta 2R$, defined by

$$\Delta 2R = (p_{(42)} - \overline{p}) - (\overline{p} - p_{(1)}) = p_{(42)} + p_{(1)} - 2\overline{p}, \tag{8}$$

can be seen as a measure of skewness.

Table 1 confirms that negative asymmetry is associated with a successful performance.

2.3. The sign of asymmetry is not due to random sampling fluctuations

Even though most distributions from the real world manifest asymmetry, we can show that the negative asymmetry observed on the six registered races cannot be explained by mathematical randomness, as shown by the following thought experiment (in the sense of Einstein's *Gedankenexperiment*).

If it were the case, the observation of a race would be as frequent as that of what we call the *associated mirror-race*, say p^*, defined by one of the two equivalent equalities

$$p_i^* = 2\overline{p} - p_i, \ \frac{p_i^* + p_i}{2} = \overline{p} \quad (1 \le i \le 42). \tag{9}$$

For example for Berlin 2018, where the average pace was $\overline{p} \approx 173$ seconds, a kilometre run in 175 seconds would be replaced in the mirror race by a kilometre run in 171 seconds. The average of these two values is $\overline{p} = 173$.

From the equalities

$$p_i^* - \overline{p} = -(p_i - \overline{p}) \quad (1 \le i \le 42)$$

and in view of (2) − (5) we infer that a race and its mirror image have the same average pace (in other words they correspond to the same performance measured by the final time), and opposite asymmetries as measured by γ_1. In the 2018 Berlin marathon, 16, 6 and 20 kilometres were run at a pace smaller, equal, or greater than the average $\overline{p}$. In other words it, 16 kilometres were ran fast, 6 were medium, 20 were slow. In the mirror-race, 16 would be slow, 6 would remain medium, and 20 would be fast. Firstly, it is clear that the actual race is more comfortable to run than the mirror race. Secondly, physiologically, whereas 20 slow kilometres give the possibility to recover the energy for 16 fast kilometres, it is not certain that 16 slow kilometres would enable the runner to run 20 fast kilometres, as shall be discussed below.

Remarkably, Berlin 2019 manifests a similar split of 15, 6 and 21 kilometres (fast, medium and slow kilometres). Hence the mirror race would also be much more difficult to perform in this case.

For Tokyo 2021, Berlin 2014 and Monza 2017 the average pace is distinct enough from an integer value, so that each kilometre can be labelled as fast/difficult (red) or slow/easy (green), see Figures 1 and 5. For Tokyo 2021 and Berlin 2014 there were more easy kilometre than difficult kilometres, so that running the mirror race would again be unreasonable. For Monza 2017, the situation is opposite, there were more difficult than easy kilometres. Therefore, the mirror race, with positive skewness, would have been easier to run.

For Vienna 2019, a cursory look at the bar chart suggests that the distribution can be seen as the mixture (see §1.2.14 in [14]) of two independent distributions. The first one, represented by the right component of the graph, is rather symmetric and concentrated around the mean value $\overline{p} \approx 170$ s. The second one, is reduced to that of one value $p_{(1)} = p_{42} = 161$ seconds, a very fast pace during the last kilometre. The negative asymmetry is mainly due to this last kilometre, i.e. the fact that the runner had a excessive amount of energy at the end of the race. The mirror race would consist of the superposition of the same symmetric distribution centred about the average pace, and of a very high value corresponding to a very slow kilometre, which is unrealistic.

2.4. Figures, Tables and Schemes

2.5. Mathematical Formulas

Formulas leading to Zar's statistic are as follows.

$$A = \gamma_1 \sqrt{\frac{(n+1)(n+3)}{6(n-2)}}, \tag{10}$$

$$B = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \tag{11}$$

$$C = \sqrt{2(B-1)} - 1, \tag{12}$$

$$D = \sqrt{C}, \tag{13}$$

$$E = \frac{1}{\ln \sqrt{D}}, \tag{14}$$

$$F = A\sqrt{\frac{C-1}{2}}, \tag{15}$$

$$z = E \ln(F + \sqrt{F^2+1}) \tag{16}$$

2.6. Tables and Figures

Race	min $= p_{(1)}$	max $= p_{(42)}$	r_1	r_{10}	r_{21}	V_1	V_{10}	V_{21}
Tokyo 2021	167	194	27	189	254	15%	10%	7%
Berlin 2014	165	181	16	106	137	9%	6%	4%
Berlin 2019	165	179	14	76	98	8%	4%	3%
Berlin 2018	164	178	14	73	124	8%	4%	3%
Vienna 2019	161	172	11	43	43	6%	2%	1%
Monza 2017	168	176	8	43	58	5%	3%	2%

Race	total time	$\overline{p}$	σ	V	γ_1
Tokyo 2021 (E. K.)	2 h 08 min 38 s	182.4 s	7.2	4%	- 0.31
Berlin 2014 (D.K)	2 h 2 min 57 s	174.8	4.1	2%	- 0.47
Berlin 2019 (K.B)	2 h 01 min 41 s	173.0	3.0	2%	- 0.55
Berlin 2018 (E.K)	2 h 01 min 39 s	173.1	3.0	2%	- 0.82
Vienna 2019 (E.K)	1 h 59 min 40 s	169.7	1.9	1%	- 2.32
Monza 2017 (E.K)	2 h 00 min 25 s	171.2	1.8	1%	+0.56

Race	$p_{(42)} - p_{(1)}$	$\overline{p} - p_{(1)}$	$p_{(42)} - \overline{p}$	$\Delta 2R$
Tokyo 2021	27	15.4	11.6	-3.8
Berlin 2014	16	9.8	6.2	-3.6
Berlin 2018	14	9.1	4.9	-4.2
Vienna 2019	11	8.7	2.3	-5.4
Monza 2017	8	3.3	4.7	1.2

Race	γ_1	z	p-value of z
Tokyo 2021	-0.31	-0.92	18%
Berlin 2014	-0.47	-1.36	9%
Berlin 2019	-0.55	-1.50	7%
Berlin 2018	-0.82	-2.25	1%
Vienna 2019	-2.32	-4.79	$< 10^{-6}\%$
Monza 2017	0.56	2.07	2%

Table 1. Zar statistic z. The p-value of z is $P(\mathcal{N}(0,1) > z)$ if $z > 0$, or $P(\mathcal{N}(0,1) < z)$ if $z < 0$.

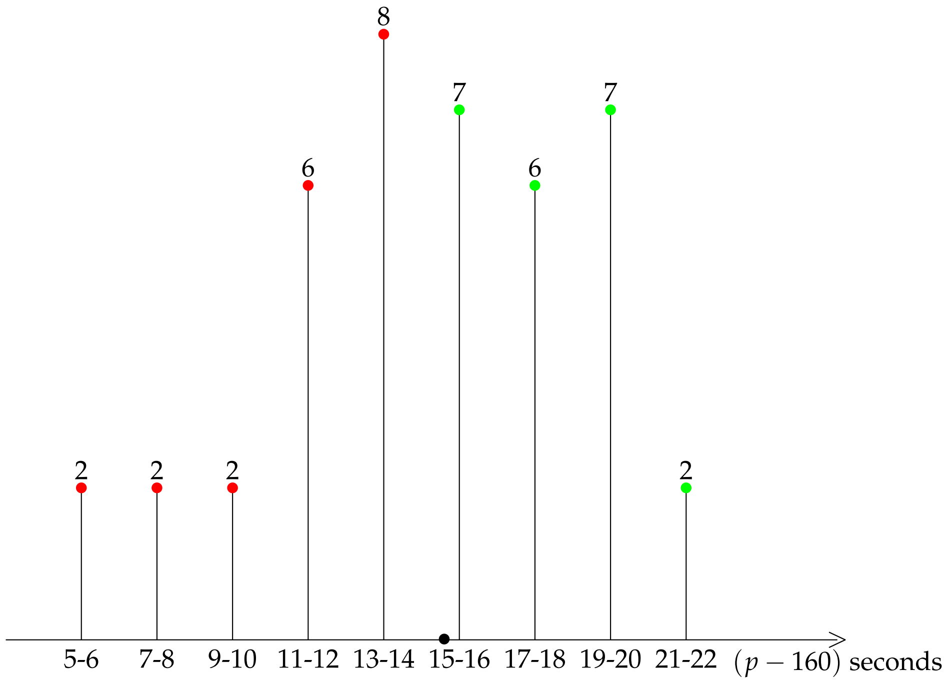

Figure 1. Frequency of paces of Berlin 2014. For instance 8 kilometres were run at the pace 173 or 174 seconds. Slow (resp. fast) kilometres in green (resp. red). The black bullet on the p axis marks the average pace.

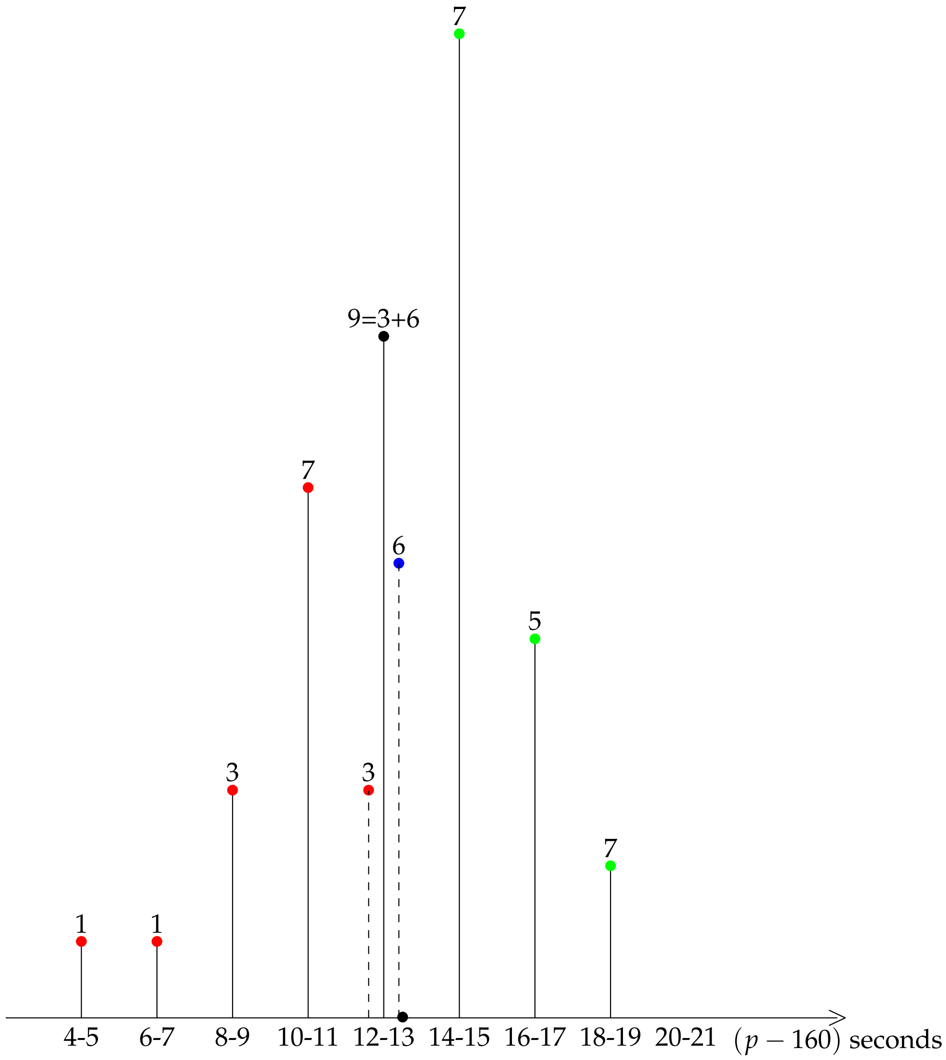

Figure 2. Frequency of paces of Berlin 2019. Slow (resp. medium, fast) kilometres in green (resp. blue, red). The 12-13 class is decomposed into two subclasses, 3 fast and and 6 medium kilometres.

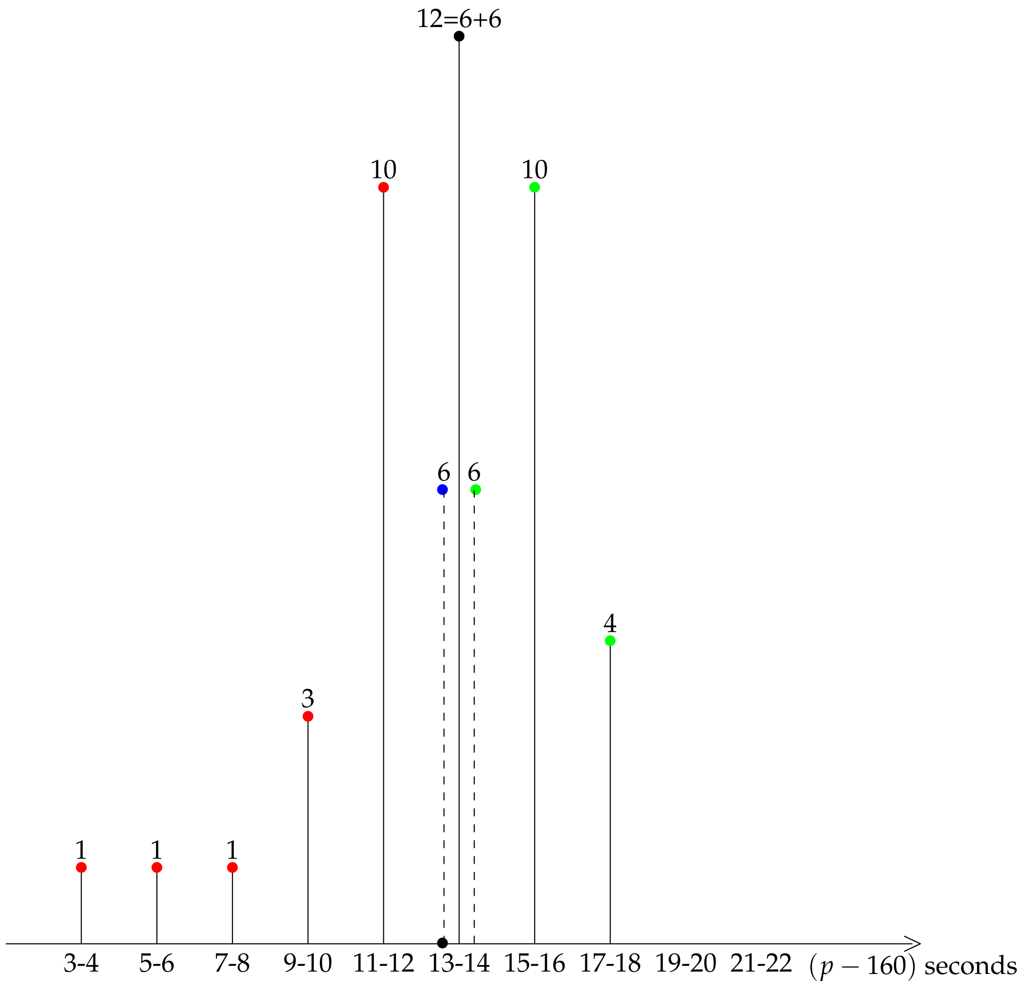

Figure 3. Frequency of paces of Berlin 2018. Slow (resp. medium, fast) kilometres in green (resp. blue, red). The 12-13 class is decomposed into two subclasses, 6 medium and 6 slow kilometres

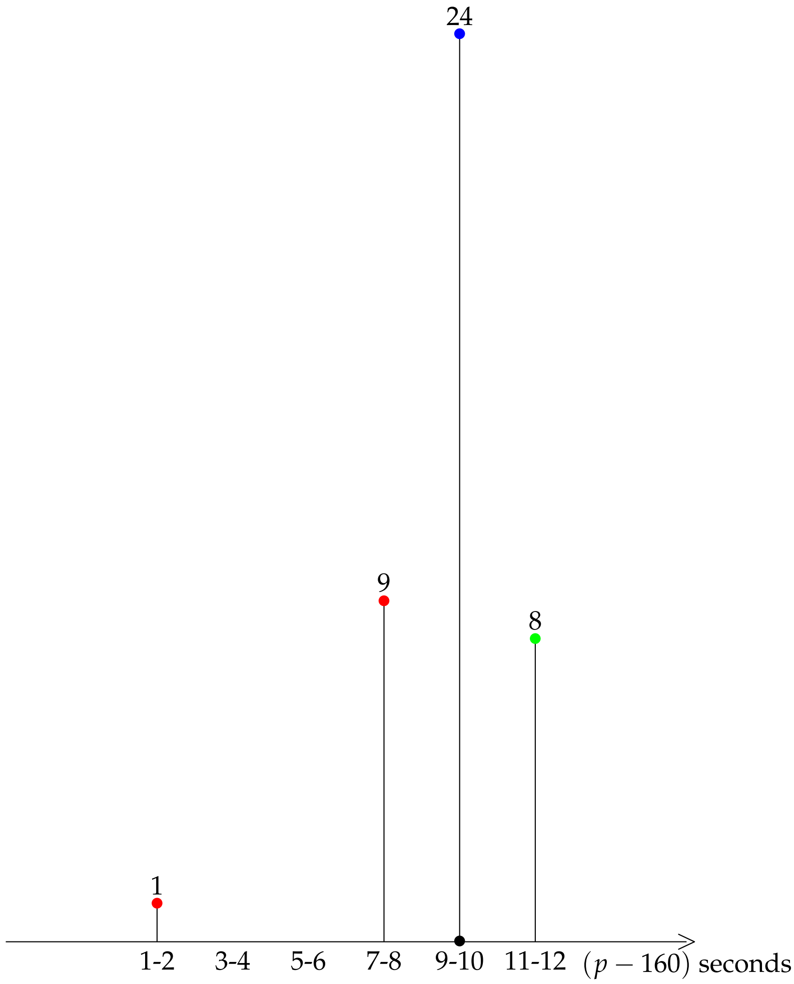

Figure 4. Frequency of paces of Vienna 2019.

. 217

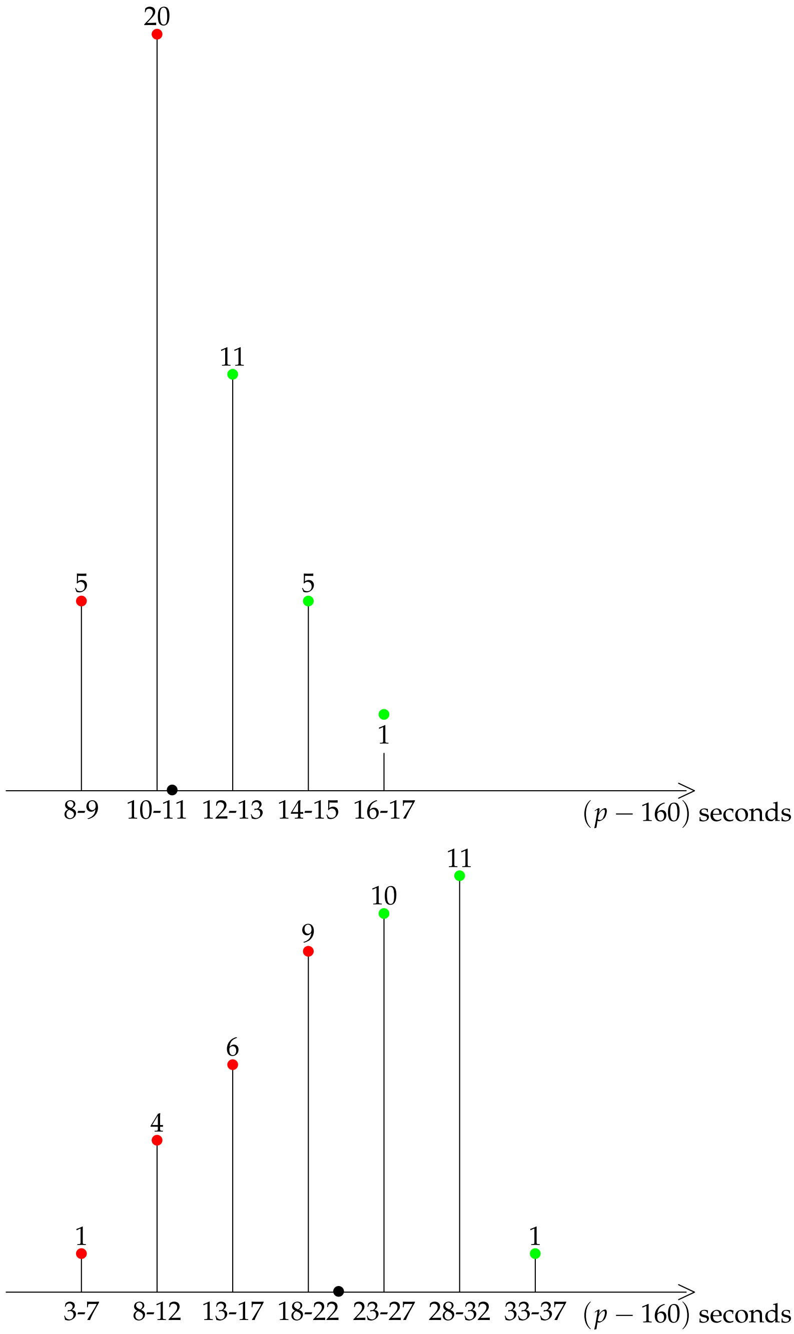

Figure 5. Frequency of paces of Monza 2017 and Tokyo 2021. Number of slow (resp. fast) kilometres in green (resp. red). The black bullet on the p axis marks the average pace.

3. Discussion

The future of marathon training will focus on the qualitative aspects versus only focusing on the quantitative data. This is certainly a step in the right direction as it was Kipchoge's confidence allowing him to break sub-2-hour marathon barrier. Advances in wearable sensor technology have enabled real-time measurement of physiological data during exercise ([2]). Future directions in training are going to be about encouraging the marathon runner. To achieve this, we must provide ways of improving satisfaction regarding training progress and with the feeling that it has been optimized. We are testing the hypothesis that the ideal race, with the world's best marathon runners, can have several degrees of optimization. This can be observed in Kipchoge's last 5 marathon's, including

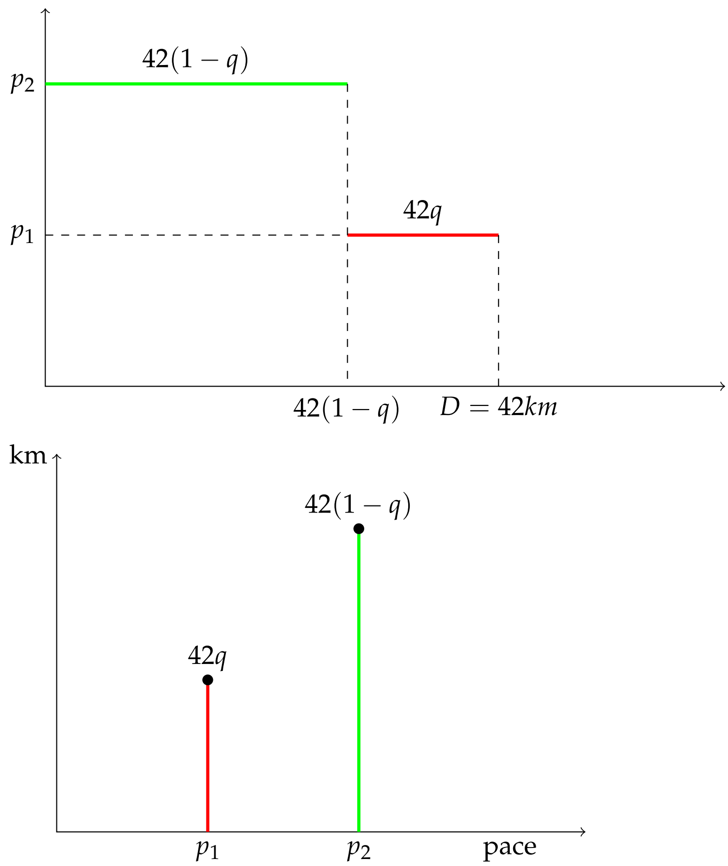

Figure 6. Frequencies of a theoretical race with two paces and negative asymmetry

the exhibition marathons, where the paradigm of a perfect constant pace was applied. In the official marathons (Berlin), the pace was not constant and run with other competitors; these races are considered being the best performances ever ran.

The best races are run with a negative asymmetry of pace distribution. This means that more time is spent running below Vmarathon and that the pace is variable, never constant. Our recent publication about the asymmetric characteristics of marathon pace in elite marathon runners, were not random, but well founded on mathematically demonstrated principles. Estimating the optimal marathon pace should be done as a consequence of optimal physiological homeostasis. Our results showed how to obtain a true individual distribution shape, for instance concave or bathtub, by using the negative asymmetry of pace distribution. We reported an optimal value for the negative asymmetry, showing how this asymmetric distribution results from an oscillating pattern and the interaction between extreme values.

3.1. The pace is variable, not uniform

Publications showing that pace is non-uniform are in agreement with prior studies of short-distance competitions. Indeed, variability in pacing has been studied in short and middle-distance running (e.g., 3000 m to 10 km). See [15], [16], [17], [18], [19], [20].

These studies have focused on the influence of variability of pacing on metabolic and performance parameters. However, [21], reported that elite marathon runners had few changes in pace, suggesting low speed variability. More recently, the sub two-hour marathon attempts in Monza (2017) and Vienna (2019) were based on the belief that constant speed is the best way of running. These ideas were derived from the notion that an optimal pace is non variable, according to the seminal model of [3].

Recreational marathon runners adopt the same paradigm of constant speed by running with a pace-group leader as provided by the marathon organizers. This blindly reinforces

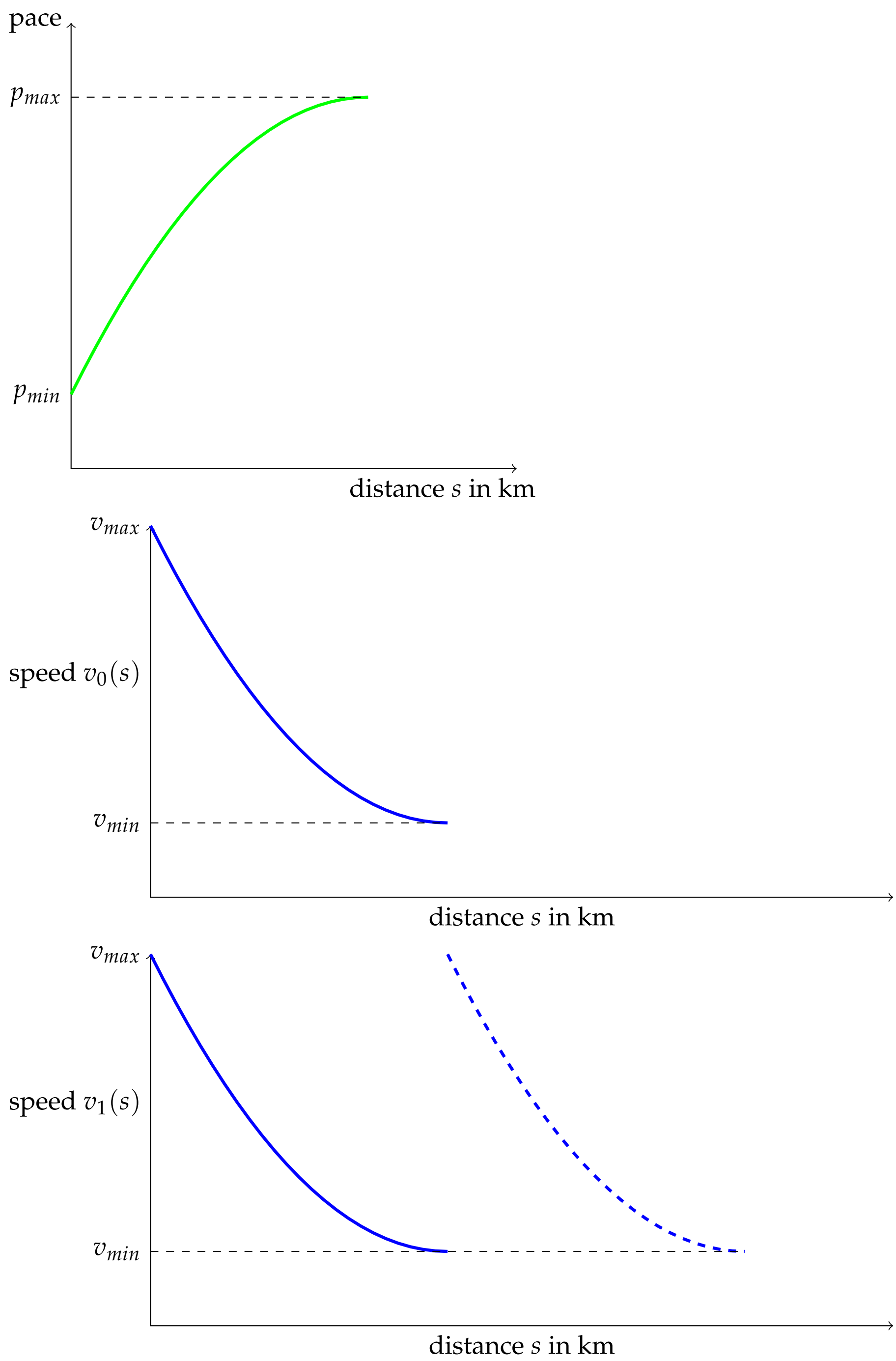

Figure 7. A concave pace function on an elementary time stretch, a convex speed function v_0, for instance (22) on an elementary time stretch, and a speed function v_1 with the same frequencies and asymmetry as v_0.

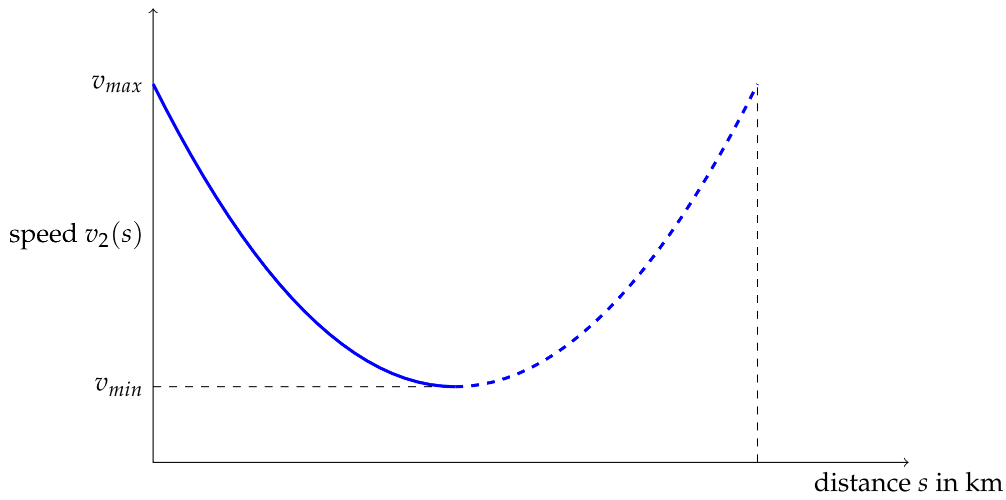

Figure 8. bathtub curve function v_2 with the same frequencies ans asymmetry as v_0

the idealistic paradigm of constant speed, distorting the practice of running. However, considering the physiological limitations (glycogen availability), means that runners must choose the ideal pace allowing them to achieve the optimal performance without famously "hitting the wall" or slowing down from extreme fatigue 10 kilometres or less before the finish line (see [22]).

We must underline that negative pacing can help runners to achieve better results (both recreational and professional). However, a great majority of runners just want to complete a marathon or enjoy running, without focusing too much on optimizing their pacing. Therefore, for them, even pacing is easier to follow and control. Indeed, for both elite and recreational runners, planning variable pacing is not practical. The pacing is mainly controlled by CNS and it is highly automated. Therefore, any change in pacing can be hard to achieve, resulting from runners either slowing or speeding up too much. In particular, if they are less experienced runners. Please elaborate some more, and possibly add this as a limitation of this study. In addition, speeding and slowing are requesting an additional muscle force to be exerted, thus spending extra energy even if it allow to elevate the average pace. We still have to check the difference of energy spent by continuous energy cost measurement between variable and even pace race, at least on half marathon real race.

The non-arbitrary sign of asymmetry, which cannot be accounted for by sampling fluctuations in speed but by physiology, is explained by the "mirror race". It is important to understand that a runner cannot recover lost time in the first part of a marathon (as in Monza) using their speed reserve (see [23], [24]). Recall that speed reserve is defined as the difference between the velocity at VO2max and the maximal speed during a sprint or a 1000m distance (see [25]).

Variable pace running focuses the optimization of energy. These concepts were developed by AV Hill over 100 years ago with his papers on muscular exercise, lactic acid, the supply and utilization of oxygen, all helping to establish the concept of "an-aerobic" energy production during exercise, with oxidative restoration in recovery (see [26]); it is unlikely that he thought humans should run at a constant pace to achieve athletic records. Ironically, ever since Hill's time, schools often still teach students to run at a strict constant pace as in, rather than to self-pace according to the rate of perception of exhaustion. It has been proven that humans are natural runners and capable of accurately self-pacing our accelerations and speed variations at three levels of intensity (soft, medium and hard accelerations) (see [27]). Even if control of speed, and thus power output, are voluntary (see [28]), the physiological signals that athletes receive estimating their abilities to sustain any instantaneously chosen speed have yet to be to be elucidated. Computing the critical speed from a personal best has shown, in the same way as for VO2max, that all the elite marathon runners could run very close to their critical speed (90%-98%) (J[29], [30].

Furthermore, it has been shown that critical speed can be accurately estimated from a non-exhaustive self-pace run ([31]). Even if these two paradigms of exercise (self-pace or constant-load model) have different limitations (see [32]), the self-pace model has allowed new insights into the optimization of energy transformation (see [33]). Energy transformation during exercise is thought to be variable; this allows for self-regulation so that feed-forward or anticipatory regulation remains a critical signal to preserve homeostasis and to avoid cellular catastrophe (see [34]). [33] showed that it is possible to consider the time required for task completion, which can explain why power output (or speed) and skeletal muscle recruitment (and hence VO2) in self-pace trials are almost immediately reduced in exercise, only to return to near-optimal values during the "end spurt" (see [35]). Examining the pacing signatures of the two best marathon performances and, given that the same human has achieved these results, offers us the unique possibility to compare these performances and further understand how the limits of human physiology were elevated. In addition, we could also pattern this novel research to the over 9 million of marathon runners crossing the finish line in 2019.

3.2. The pace is not symmetric, and the sign of asymmetry is explained via a "mirror race", by physiology, and not by sampling fluctuations.

Independent of the race or the runner, the pace is never symmetric, implying that the central tendency and deviation do not entirely characterize the pacing signature. A negative asymmetry of pace distribution signifies a faster race and the higher the absolute value of the pacing asymmetry is, the better the performance is. Indeed, according to our hypothesis, our results show that for each of these high-level races, we could give the general shape of the data distribution. Independent of the marathons studied, we always obtained an asymmetric shape. Interestingly, we showed that the asymmetries for the two exhibition marathons, were different: negative for Vienna (the first 2-sub-hour marathon race (1h59min40s and $\gamma = -2.50$) and positive in Monza (2h00min25s and $\gamma = +0.56$). Furthermore, we showed Vienna was comprised of two independent distributions, therefore it was really two different races: one was constant for 40 km, and the other, also constant, but at a much higher speed. Remarkably, in the Vienna attempt, Kipchoge had a remarkable "end spurt" in the final kilometer, helping him to break the 2-hour barrier and thus creating an artificial asymmetry. Similarly in Tokyo, after running a very slow first half marathon while attempting to help his compatriot, Kipchoge finished the ten last kilometres extremely fast. The marathon pace (or speed) frequency-distribution manifests two remarkable features: non-uniformity, and skewness of the pace which has the particularity of being negative. This is a condition of comprehending a good performance i.e., the shortest time possible without excessive suffering. A U-shaped curve allows the complete distribution of pacing. By now, pacing profiles and tactical behaviors of elite runners have been shown to be a U-shape from the 1500 to the 10,000m (Casado et al., 2020). This U-shape is characterized by a fast start, a slowing down during the middle part of the race, and a fast finish characterizes world record performances (see [36]).

Although a parabolic J-shaped pacing profile (in which the "fast-start" is faster than the middle part of the race but is slower than the end spurt) is observed in many championship races ([36]), we showed that in the marathon, the best race is in U shape, which induced a negative speed asymmetric distribution measured by $\Delta 2R$. Here, we reported the shape distribution of the entire marathon, and we have formalized the role of the extreme values and oscillations between them. In these world record performances, we show that non-uniformity of a sequence of measurement means that they do not arise from a population highly concentrated around a sample value as a normal distribution with a very slight variance.

We also showed how the difference in extreme values distributions allows a characterization of the optimal marathon as negatively asymmetric in the sense that the fast-kilometres pace frequency decay occurs more progressively than the slower kilometres pace. Analysing the "big data" of marathons reveals that foremost, the individual experience of training and running a marathon offers personal gratification, whether or not they achieve their target of finishing less than 4h21min03s or 4h48min45s (the average male and female marathon performances in 2018, respectively). Endurance running has never reached this level of popularity, especially among women. Marathon running is truly human by nature (see [37]). In this fundamental article, the authors showed how evolution has not only shaped the human for long-distance running, but has also conditioned the human brain to enjoy this type of physical activity through the development of endorphin release and mood-elevating neuroendocrine mechanisms. When humans run, we are balanced between a quest for performance and "the minimization of effort" which is defined as the process that aims to achieve the most cost-effective behaviour based on our perceptions (see[38]). Energy is the ability to produce physical action and effort is the cortical brain activity associated with the initiation or maintenance of a behaviour. The brain constructs our perceptions based not only on the current physical effort but also on previous similar experiences, motivation, awareness, and affects (see [39]). Therefore, our theory of a mirror race that, may or may not be possible to run, shows that an optimal marathon performance must be performed without suffering and hence with variable pace.

3.3. Some mathematical speed functions as the bathtub with negatively skewed pace distributions

Let us show that the foregoing results are, obtained via calculus, in agreement with some already identified running strategies, especially faster start and finish that are observed even in the traditional (even-pacing) approach to record-setting results.

Assume that $q \in (0, 1/2)$. Typically, q will represent the proportion of distance run fast during a marathon. For instance the runner runs $42q$ kilometres fast at pace p_1, and $42(1-q)$ kilometres slowly at a pace $p_2 > p_1$. The bar chart of the pace sequence corresponds to the increasing function, say ϕ, defined by

$$42q = \phi(p_1) < \phi(p_2) = 42(1-q) \quad (p_1 < p_2).$$

The bar chart of the pace distribution ϕ, given by Figure 6, is that of a Bernoulli distribution with skewness coefficient

$$\gamma_1 = \frac{2q-1}{\sqrt{q(1-q)}} < 0.$$

Let us generalize this fundamental example. Let $s \in [0, 42]$ denote the curvilinear abcissa, i.e. the distance from the start line. The instantaneous pace at s is denoted by $p(s)$.

The bar charts of the pace sequence considered up to now will be replaced by a continuous curve, the density function $p \mapsto \phi(p)$, defined as follows. The infinitesimal distance $ds = \phi(p)dp$ represents the distance spent by the runner at the pace ranging from p to $p + dp$. In accordance with the preceding example we assume that ϕ is increasing. The derivative of $s \mapsto p(s)$ is $p'(s) = dp/ds$. The preceding equalities lead to

$$\frac{dp}{ds} = \frac{1}{\phi(p)}, \; \phi(p)dp = ds, \; \Phi(p) = s, \tag{17}$$

where $\Phi(p)$ is such that $\Phi'(p) = \phi(p)$. We obtain

$$p(s) = \Phi^{-1}(s) \tag{18}$$

as the pace function. Note that equation $p'(s) = \phi(p)$ and the fact that ϕ is assumed to be an increasing function of s, which is itself increasing with p, imply that p' is decreasing, thus p is concave. Up to a constant factor the speed function is given by $v(s) = 1/p(s)$, so that

$$v''(s) = -\frac{p''}{p^2} + 2\frac{(p')^2}{p^3} > 0, \tag{19}$$

and v is an increasing convex function.

Conversely, every increasing speed function $s \mapsto v(s)$ gives rise to a concave decreasing function $s \mapsto p(s)$ (invert the roles of v and p in (19)) and to a density function $\phi(p) = 1/p'(s))$ that is increasing, as the inverse of a decreasing function.

Let us make these formulas explicit in the case

$$\phi(p) = p^{\alpha}, \quad (0 < \alpha < 1) \tag{20}$$

so that ϕ is a beta distribution, with negative skewness [14] Chapter. Then up to multiplicative factors, $\Phi(p) = p^{\alpha+1}$,

$$p = \Phi^{(-1)}(s) = s^{\frac{1}{\alpha+1}} = s^{\beta} \quad (0 < \beta < 1) \tag{21}$$

Since $\gamma_1(p)$ is invariant under affine transformations, we can set

$$p(s) = C_1 + C_2 s^{\beta},\ v_0(s) = \frac{1}{C_1 + C_2 s^{\beta}} \tag{22}$$

These computations lead to rather unrealistic pace strategies. But keep in mind that our aim was only to produce some pace distribution with negative skewness.

From the speed function in (22), we can build two speed functions that will lead to the same density function ϕ:

$$v_1(s) = \begin{cases} v_0(s), & \text{for } 0 < s < T, \\ v_2(s) = v_0(s - T), & \text{for } T < s < 2T \end{cases}, \tag{23}$$

and

$$v_2(s) = \begin{cases} v(s), & \text{for } 0 < s < T, \\ v_2(s) = v(2T - s), & \text{for } T < s < 2T \end{cases} \tag{24}$$

The graphs of v_1 and v_2 are drawn in Figures 7 and 8, the dashed parts being those which were added. Periodicity and symmetry, used to build v_1 and v_2 respectively, can be used an arbitrary number of times.

From the mathematical point of view, a running strategy can be outlined as follows: First divide the race into sub-races, called "pace-cells", and for each of these cells, select a speed function as v_0, v_1 or v_2. Provided these functions are built from the same elementary function as v_0, the global skewness will be that of v_0.

Figure 7 can be associated with an oscillator as a clock with frictions, receiving regular impulses, see [40], §III.4.2.

The theory of oscillators seems well-suited for providing models for oscillating functions. Our analogy is not based on purely mathematical similarities, but is a consequence of our results. The interplay between extreme (small and large) values that we have unveiled, corresponds, for oscillators, to the well-known restoring forces. Indeed, Claude Bernard already pointed out that all biological systems are oscillators, in the sense that the value of a variable remaining constant at its mean value is never observed, that is to say biological systems are oscillators. Our statistical analysis illustrates this fact concerning the world's best marathon runners.

3.4. Oscillations between extremes values allow the interplay between aerobic and anaerobic metabolisms, optimizing the recovery of energy.

We proposed to focus on a time series generated by the oscillating pattern of the damped clock discussed in [40]. Remarkably, when human and other mammals run, the body's complex system of muscles, tendons and ligament springs behaves like a single linear spring ("leg spring"). A simple spring-mass model, comprising a single linear leg spring and a mass equivalent to the animal, has been shown to describe the mechanisms of running remarkably well (see [41]). Hence, the stride can be considered as a damped clock. We have shown that the first sub-2-hour attempt (Vienna) was not optimal.

Therefore, by considering the pacing oscillations between extremes values according to the athlete's physiology, these world record performances could be run faster, even the 1h50min marathon as predicted by [42]; but only when the method for predicting the record is based on physiology. Indeed, currently it is thought that running a marathon at the fastest speed possible appears to be regulated by the rate of aerobic metabolism (i.e., marathon oxygen uptake) of a limited amount of carbohydrate energy (i.e., muscle glycogen and blood glucose) and the pace that can be maintained without developing hyperthermia.

However, to oscillate between extreme pace values, there is a need to have a high range of speed incompatible with monotonous training ([43], [44]). Indeed, our results confirmed that the best performance was achieved by an oscillating pattern between extreme paces.

Extreme value theory is used to predict the occurrence of rare events such as extreme floods, large insurance losses, stock market crashes, and human life expectancy. Indeed, some authors have already applied the extreme value theory to athletic events, using estimation methods involving moment method and maximum likelihood methods (MLE). Accurate estimations of future athletic records were accomplished using both methods. Hence, an interesting question is: under the present knowledge of training, materials (shoes, suits and equipment), and anti-doping regulations, by how much further could athletes possibly exceed current world records in the near future? We used the extreme value theory dealing with the issues of extremes without considering them to be independent because we took into account (as stated by Hill), the "fatigue factor".

Running a marathon between extreme values, training must use the so-called "polarized training" emphasizing low and high intensity, rather than medium intensity (see [43] , [45]). A recent study based on around 14,000 individuals with more than 1.6 million exercise sessions containing duration and distance, and with a total distance of 20 million km showed that the analysis of individual long-term training protocols leads to a wide spectrum of physiological responses ([2]). This study on big data confirms the concept of "polarized training", which involves running a wide range of speeds; these techniques are currently used by elite athletes and have been shown to be the most efficient training methods, resulting in the greatest improvements in the key variables of endurance performance in well-trained endurance athletes.

4. Conclusions

We analysed the three best real-world marathon performances ever ran to the hypothesis that Kipchoge optimizes his pacing in relation to his aerobic and anaerobic power and endurance. The conclusions from this could be an inspiration source for the over 9 million of marathon runners who crossed the finish line in 2018 (Esther Fleming, https://www.sidmartinbio.org/how-many-runners-are-there/, accessed on 01/09/2021). Using mathematics and physiology, we opened new perspectives on how optimize endurance and power in the marathon. The pacing strategy of elite marathon runners is to start fast and then to recover by running just below their average pace for the 2/3 racing distance, generating a negative asymmetry, i.e., a median speed below the average. This "lazy" race confirms AV Hills' discoveries about the concept of "an-aerobic" energy production during exercise, with oxidative restoration in recovery are factors determining the variation of speed with distance.

However, our aim was not to replace the coach advice, but only to show that the future of best performance as world record, is perhaps to look forward the variable pace shape.

Author Contributions: Conceptualization, J-R P. and V.B; methodology, J-R P. and V.B; software, J-R P. and V.B.; validation, J-R P. .; formal analysis, J-R P. ; investigation, J-R P. and V.B; resources, J-R P. and V.B; data curation, J-R P. and V.B; writing—original draft preparation, J-R P. and V.B; writing—review and editing, J-R P. and V.B; visualization, J-R P. and V.B; supervision, J-R P. and V.B; project administration, J-R P. and V.B; funding acquisition, this research received no external funding. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: https://labs.strava.com/

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hill, A. The Physiological Basis of Athletic Records. *Nature* **1925**, *116*, 544–548.
2. Emig, T.; Peltonen, J. Human running performance from real-world big data. *Nature communications* **2020**, *11*, 1–9.
3. Keller, J.B. A theory of competitive running. *Physics today* **1973**, *26*, 42–47.
4. Kennelly, A.E. An approximate law of fatigue in the speeds of racing animals. Proceedings of the American Academy of Arts and Sciences. JSTOR, 1906, Vol. 42, pp. 275–331.
5. Bernard, C. *Introduction à la médecine expérimentale*; Ligaran, France, 2015.
6. Billat, V.; Carbillet, T.; Correa, M.; Pycke, J.R. Detecting the marathon asymmetry with a statistical signature. *Physica A: Statistical Mechanics and its Applications* **2019**, *515*, 240–247.
7. Foster, C.; Hendrickson, K.J.; Peyer, K.; Reiner, B.; deKoning, J.J.; Lucia, A.; Battista, R.A.; Hettinga, F.J.; Porcari, J.P.; Wright, G. Pattern of developing the performance template. *British journal of sports medicine* **2009**, *43*, 765–769.
8. Lluch, J.; Rebollo, M.; Calduch-Losa, Á.; Mollá, R. Precision of wearable GPS in marathon races. *IEEE Consumer Electronics Magazine* **2020**, *10*, 32–38.
9. Stuart, A. *Kendall's advanced theory of statistics*; Vol. 1, John Wiley & Sons, 1994.
10. David, H.A.; Nagaraja, H.N. *Order statistics*; John Wiley & Sons, 2004.
11. Udny Yule, G.; Kendall. *An introduction to the theory of statistics.*; Number 14th ed, Charles Griffin & Co. Ltd., London, 1950.
12. Zar, J. Biostatistical Analysis., 4th edn.(Prentice-Hall: Englewood Cliffs, NJ.) **1999**.
13. Sheskin, D.J. *Handbook of parametric and nonparametric statistical procedures*; Chapman and Hall/CRC, London, 2003.
14. Johnson, N.L.; Kemp, A.W.; Kotz, S. *Univariate discrete distributions*; Vol. 444, John Wiley & Sons, 2005.
15. Léger, L.A.; Ferguson, R.J. Effect of pacing on oxygen uptake and peak lactate for a mile run. *European journal of applied physiology and occupational physiology* **1974**, *32*, 251–257.
16. Ariyoshi, M.; Yamaji, K.; Shephard, R. Influence of running pace upon performance: effects upon treadmill endurance time and oxygen cost. *European journal of applied physiology and occupational physiology* **1979**, *41*, 83–91.
17. Billat, V.L.; Slawinski, J.; Danel, M.; Koralsztein, J.P. Effect of free versus constant pace on performance and oxygen kinetics in running. *Medicine and Science in Sports and Exercise* **2001**, *33*, 2082–2088.
18. Cottin, F.; Papelier, Y.; Durbin, F.; Koralsztein, J.P.; Billat, V.L. Effect of fatigue on spontaneous velocity variations in human middle-distance running: use of short-term Fourier transformation. *European journal of applied physiology* **2002**, *87*, 17–27.
19. Sandals, L.E.; Wood, D.; Draper, S.B.; James, D. Influence of pacing strategy on oxygen uptake during treadmill middle-distance running. *International journal of sports medicine* **2006**, *27*, 37–42.
20. Garcin, M.; Danel, M.; Billat, V. Perceptual responses in free vs. constant pace exercise. *International journal of sports medicine* **2008**, *29*, 453–459.
21. Ely, M.R.; Martin, D.E.; Cheuvront, S.N.; Montain, S.J. Effect of ambient temperature on marathon pacing is dependent on runner ability. *Med Sci Sports Exerc* **2008**, *40*, 1675–1680.
22. Jones, A.M.; Vanhatalo, A.; Burnley, M.; Morton, R.H.; Poole, D.C. Critical power: implications for determination of V O2max and exercise tolerance. *Medicine & Science in Sports & Exercise* **2010**, *42*, 1876–1890.
23. Fukuba, Y.; Whipp, B.J. A metabolic limit on the ability to make up for lost time in endurance events. *Journal of Applied Physiology* **1999**, *87*, 853–861.
24. Morton, R.H. A new modelling approach demonstrating the inability to make up for lost time in endurance running events. *IMA Journal of Management Mathematics* **2009**, *20*, 109–120.
25. Blondel, N.; Berthoin, S.; Billat, V.; Lensel, G. Relationship between run times to exhaustion at 90, 100, 120, and 140% of vV O2max and velocity expressed relatively to critical velocity and maximal velocity. *International journal of sports medicine* **2001**, *22*, 27–33.
26. Bassett Jr, D.R. Scientific contributions of AV Hill: exercise physiology pioneer. *Journal of applied physiology* **2002**, *93*, 1567–1582.
27. Billat, V.; Brunel, N.J.; Carbillet, T.; Labbé, S.; Samson, A. Humans are able to self-paced constant running accelerations until exhaustion. *Physica A: Statistical Mechanics and its Applications* **2018**, *506*, 290–304.
28. BILLAT, V.L.; DEMARLE, A.; SLAWINSKI, J.; PAIVA, M.; KORALSZTEIN, J.P. Physical and training characteristics of top-class marathon runners. *Medicine & Science in Sports & Exercise* **2001**, *33*, 2089–2097.
29. Jones, A.M.; Kirby, B.S.; Clark, I.E.; Rice, H.M.; Fulkerson, E.; Wylie, L.J.; Wilkerson, D.P.; Vanhatalo, A.; Wilkins, B.W. Physiological demands of running at 2-hour marathon race pace. *Journal of Applied Physiology* **2021**, *130*, 369–379.

30. Billat, V.; Vitiello, D.; Palacin, F.; Correa, M.; Pycke, J.R. Race analysis of the world's best female and male marathon runners. *International journal of environmental research and public health* **2020**, *17*, 1177.
31. Nakamura, F.Y.; Okuno, N.M.; Perandini, L.A.; Caldeira, L.F.S.; Simões, H.G.; Cardoso, J.R.; Bishop, D.J. Critical power can be estimated from nonexhaustive tests based on rating of perceived exertion responses. *The Journal of Strength & Conditioning Research* **2008**, *22*, 937–943.
32. Marino, F.E. The limitations of the constant load and self-paced exercise models of exercise physiology. *Comparative Exercise Physiology* **2010**, *7*, 173–178.
33. Ulmer, H.V. Concept of an extracellular regulation of muscular metabolic rate during heavy exercise in humans by psychophysiological feedback. *Experientia* **1996**, *52*, 416–420.
34. Billman, G.E. Homeostasis: the underappreciated and far too often ignored central organizing principle of physiology. *Frontiers in physiology* **2020**, *11*, 200.
35. Amann, M.; Eldridge, M.W.; Lovering, A.T.; Stickland, M.K.; Pegelow, D.F.; Dempsey, J.A. Arterial oxygenation influences central motor output and exercise performance via effects on peripheral locomotor muscle fatigue in humans. *The Journal of physiology* **2006**, *575*, 937–952.
36. Casado, A.; Hanley, B.; Jiménez-Reyes, P.; Renfree, A. Pacing profiles and tactical behaviors of elite runners. *Journal of Sport and Health Science* **2021**, *10*, 537–549.
37. Bramble, D.M.; Lieberman, D.E. Endurance running and the evolution of Homo. *nature* **2004**, *432*, 345–352.
38. Cheval, B.; Boisgontier, M.P. The Theory of Effort Minimization in Physical Activity. *Exercise and Sport Sciences Reviews* **2021**, *49*, 168.
39. Abbiss, C.R.; Peiffer, J.J.; Meeusen, R.; Skorski, S. Role of ratings of perceived exertion during self-paced exercise: what are we actually measuring? *Sports medicine* **2015**, *45*, 1235–1243.
40. Andronov, A.; Vitt, A.; Khaikin, S. *Theory of oscillators, Translation published by Pergamon Press*; Oxford, 1966.
41. Farley, C.T.; Gonzalez, O. Leg stiffness and stride frequency in human running. *Journal of biomechanics* **1996**, *29*, 181–186.
42. Joyner, M.J.; Coyle, E.F. Endurance exercise performance: the physiology of champions. *The Journal of physiology* **2008**, *586*, 35–44.
43. Boullosa, D.; Medeiros, A.R.; Flatt, A.A.; Esco, M.R.; Nakamura, F.Y.; Foster, C. Relationships between Workload, Heart Rate Variability, and Performance in a Recreational Endurance Runner. *Journal of Functional Morphology and Kinesiology* **2021**, *6*, 30.
44. Foster, C.; Schrager, M.; Snyder, A.C.; Thompson, N.N. Pacing strategy and athletic performance. *Sports Medicine* **1994**, *17*, 77–85.
45. Myrkos, A.; Smilios, I.; Zafeiridis, A.; Iliopoulos, S.; Kokkinou, E.M.; Douda, H.; Tokmakidis, S.P. Effects of Work and Recovery Duration and Their Ratio on Cardiorespiratory and Metabolic Responses During Aerobic Interval Exercise. *The Journal of Strength & Conditioning Research* **2021**.