Performance comparison of long-distance running competitions in different meteorology and environment based influential conditions

Dávid Berke

Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics Budapest University of Technology and Economics Műegyetem rkp. 3., H-1111 Budapest, Hungary

Author E-mail: berkedavid89@gmail.com

Abstract—Long-distance running is one of the most popular daily sports, through which we can keep our health or take care of our body. Millions of people around the world run for different reasons like recreation, professionalism, mental balance, or fun. That is why long-distance running competitions are organized very often. The organizers have to know the expected meteorological conditions and several specific parameters and information about performance of the runners retroactively to support them effectively as well as adapting to the running attitude and main goals.

The purpose of this study was on one side to develop a data analyzing tool with flexible and multifunctional analytical capabilities to identify which specific – especially meteorological – attributes of long-distance running competitions have relevant impact on the performance of the competitors, while on the other side to compare competitions with each other, which have different environmental and social properties in terms of running pace and age.

Starting to design a tool named Running Competition Comparison Tool (RCCT), availability of existing data was discovered, focusing on Hungarian long-distance running competitions. Altogether results and environmental attributes of 28 Hungarian half-marathons and further 7 variable distance competitions were investigated. In all, 146,924 participants were involved and analyzed within this study between 1996 and 2017 within 5 km to 300 km. Specific and unique data classification methods were used to specify age and pace categories for realizing different kinds of statistical analyses. In a final phase, all classified data was represented and visualized by a two-dimensional colored matrix with automatic evaluation capability, combined with column and bar charts.

Numbers of recreational athletes in Budapest national half-marathon have been continuously increased during the last two decades, especially in case of middle pace (5:00-6:59 min/km) runners. Due to a trend prediction of the study, numbers of half-marathon finishers could be reach 20,000 for 2030, which will mean 80–85% additional growth. Extraordinary meteorological conditions like unexpected high air temperature or
humidity and several environmental parameters like ultra-distance and mountain course are able to impair recreational athletes’ pace in a similar way as tiredness of well-trained ironman triathletes.

According to this study, running distance and absolute elevation have maximum impact on long-distance running pace and separate age groups mostly. Air temperature and humidity together also have relevant impact on endurance running performance in cases of different distances. Especially high temperature (higher than 25°C) or high humidity (higher than 70%) could influence long-distance running speed and fluid loss more than lower temperature and humidity. That study highlighted well that the impact of meteorological attributes are relevant in terms of running performance of long-distance running pace, hence organizers have to take meteorological conditions into account during their plan to be done.

Numbers of half-marathon finishers have been continuously increasing since 1996, especially numbers of middle-aged runners. Orientation and main goal of the participants of Hungarian half-marathons have been completely changed from competitive (elite) to recreation and hobby. Because of the athlete separation of multi-distance running events, there are no significant differences between variable distanced (5 km, 10 km, 21 km, 42 km) competitions in terms of running pace.

Key-words: meteorological conditions, environmental dependence, long-distance running, recreational sport, statistical analysis, visual comparison tool

1. Introduction

There is an ongoing debate of the effects of biological (gender, age, body mass index, genetics), environmental (air temperature, humidity, air pollution, atmospheric pressure, absolute elevation, wind speed), intellectual (willpower, tendentiousness, competitive spirit, mental balance), physical fitness (endurance, velocity, strength), and real-time physiological (heart rate, fluid loss, thirst, body weight loss, blood sugar, body temperature, etc.) attributes on performance of long-distance running athletes. Although, it is commonly believed that most of these attributes are relevant but not at the same degrees.

Popularity of long-distance running events has been growing day by day, especially in cases of half-marathons and marathons. As a result, many physiological and sport science research have been done in the last couple years, not just focusing on elite athletes, but recreational and hobby runners as well, as a part of health maintenance or e-health.

Gaining deeper knowledge about drinking behaviors of long-distance runners is very important. Not a long ago, runners have not had deep understanding of the physiological consequences of hydration, to avoid conditions of dehydrations and over-hydration is not trivial for recreational competitors (Winger et al., 2010).

Another research was focusing on effects of fluid intake and thirst sensation on performance of trained half-marathon runners. In that case, rate of fluid intake
above that dictated by their thirst sensation is not able to improve running speed. Based on that, thirst is an efficient identifier to schedule fluid replacement during a race (Dion et al., 2013).

During an original marathon, the finishing time and the percentage of body weight loss of the runners are strongly correlated. Competitors, who can complete the whole distance faster, use to lose less percentage of their body weight than slower participants do in an inverse way. The nature of this body loss effect is linear, and statistically independent from gender or age (Zouhal et al., 2010).

Environmental parameters like air temperature, humidity, dew point, and atmospheric pressure also can influence running performance. According to several papers, it is not easy to prove the direct effects of the weather conditions even in an obvious area as transport (Vécsei et al., 2014). In our case, from all these environmental attributes, obviously, air temperature is the most important factor, because it is significantly correlated with running speed, which describes performance in this case (El-Helou et al., 2012). Seemingly, from environmental attributes, air temperature is the only factor, which is linked directly to running speed, but as we can see later in this paper, characterization of a running course, especially ratio of absolute elevation and running distance should be able to generate stronger impact on running speed.

The main goal and purpose to locate refreshment or aid station distributions along long-distance running courses is trying to decrease negative effects of dehydration, fluid loss, and thirst in a human body, to recover fluid, some kind of minerals, salt, and sugar as well. According to public information of long-distance running competitions, it would be possible to estimate and simulate impact of thirst on running performance for optimizing positions of aid stations (Berke, 2015, 2017).

Long-distance sports are not consisting of only running, cycling also a relevant and important part of it. Seemingly, dehydration can always influence the endurance sport participants’ performance and speed, especially in high temperature environment. According to an environment chamber research for testing well-trained cyclists, the previously mentioned statement is not necessarily true. In some special cases, dehydration does not impair performance in the heat automatically (Wall et al., 2013).

Through those research examples, importance and variability of long-distance running performance investigations can be highlighted. Because of diversity of these parameters, which can influence running performance, analyzing all of them has to be not so simple. That is why this research was focusing on modeling, to describe and present an information technological tool, through which additional impact of several hardly comparable factors can be revealed and identified, like athletic habits, course elevation, and running distance.
2. Data and methods

2.1. Data management

In Hungary, specific personal information of the long-distance finishers, like birth year, running time, and gender are mostly available on the official website of the competitions or the organizer company. For this reason, collecting and calculating relevant data of thousands of runners for supporting this research has not been complicated. After the data collection, manual data pre-processing had to be done, like missing data interpolation, false and incorrect data value handling, and different types of data conversion. Finally, all pre-processed data were integrated into an automatic, own designed information technological tool named RCCT, through which different types of statistical analyses could be realized and evaluated, as can be seen hereinafter.

2.2. Running Competition Comparison Tool (RCCT)

Current version of the Running Competition Comparison Tool (RCCT) is a well-structured table sheet, optimized for data storage, automatic data classification, cluster visualization, and specific diagram tracing. Specific parts of a previously referred research (Berke, 2017), which focusing on optimization of locations of refreshment stations on half-marathons, gave the basis idea of RCCT.

Competitor classification

This research was describing running performance not just using running pace but age of the competitors as well. Combining both parameters into a two-dimensional classification applied for identifying runner performance categories. Both attributes were separated into seven disjoint clusters independently from each other, considering practical separation techniques of athletics and sport rules and regulations of the Hungarian Athletics Association and IAAF (HAA, 2016; IAAF, 2016).

The first age category represents U20 (under 20), then further categories continuously describe a ten-year-long age interval, lastly the seventh cluster includes runners at least 70. In cases of half-marathon and marathon distance competitions, participating for specified age groups is prohibited for medical and health keeping reasons by several national and international laws. As a result, participating a marathon under 16 and participating a half-marathon under 13 are not allowed in Hungary.

On the other hand, the first pace category represents all those runners, who can reach the finish line performing < 4 min/km average pace. Further categories continuously describe a 1 min/km pace interval, the last one represents ≥9 min/km runners. This classification was optimized for recreational endurance running activities, not for elite or professional runners. Performance categories
combine age and pace categories in every possible way, and as a result, 49 performance clusters has been implemented.

Visual representation of running performance

Distribution of percentage of these performance categories is capable of representing importance of all those clusters, which is being analyzed. All of these 49 categories were structured into a 7×7-matrix form to be colored in a logarithmic way, according to the percentage values. The X-axis of that matrix represents pace categories form the slowest (≥ 9 min/km) to the fastest (< 4 min/km). The Y-axis represents age categories from the eldest (≥ 70) to the youngest (< 20). The origin cluster of the matrix can be found on the left bottom corner, which represents all those runners who are not younger than 70 and not faster than 9 min/km in average. Besides that, percentage values of each one-dimensional categories were summarized like border values and visualized into a column or a bar charts next to the original axes. All these visual elements can be seen in Fig. 1.

![Fig. 1. Two-dimensional matrix representation of running performance according to clusters of age and pace.](image)

2.3. Experimental design

The main concept to identify hidden effects of environmental on running performance (age and pace) was managed to design experimental analyses through which it could be possible to take control over set of attributes. Thus, three independent groups of influencing parameters were specified: weather conditions, running course, and event date. As can be seen, the following
experimental analysis will be able to take control over some of these parameters to reveal impact of the remaining ones on running performance.

**Trend analysis**

On the first experimental analysis, weather conditions and running course were fastened. To achieve those conditions, such competitions were involved which has been organized on the same place and almost at the same date year by year. This solution was the best to ensure nearly equal weather conditions and equal running environments as well.

The first national Budapest half-marathon was organized more than 30 years ago, however results have been available in public since 1996. Every event was on the first week of September, the running course although was changed sometimes, but the main characteristics of that remained. As a result, the main conditions to fasten weather and course parameters would be guaranteed. Together 22 competitions from 1996 to 2017 through 115,972 finishers’ age and pace values were involved into this research. Via these analyses, it has been possible to identify how the nature of running attitude has been changed during these two decades. Besides, calculating and estimating trend functions are also realizable for describing this nature deeper as well as taking predictions for the future.

**Half-marathon comparison**

This experimental analysis is responsible for fastening event date value and distance of competitions. Alltogether seven Hungarian half-marathon competitions in 2017 through 19,433 finishers were analyzed within this study. All environmental values have been given by the Hungarian Meteorological Service (OMSZ) and reported in Table 1.

<table>
<thead>
<tr>
<th>Environment characteristics</th>
<th>Competitions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Siófok</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>7.0</td>
</tr>
<tr>
<td>Humidity (%)</td>
<td>90</td>
</tr>
<tr>
<td>Wind force (m/s)</td>
<td>2.9</td>
</tr>
<tr>
<td>Average elevation (m)</td>
<td>10</td>
</tr>
</tbody>
</table>

*Table 1. Environment description of examined competitions*
**Distance modification**

In this part of the study, a long-distance running event was chosen within which independent competitions has been organized with variable distances. Alltogether 24,550 finishers of five running competitions in 2017 were involved within this study, which were going on in parallel. The available distances were 5 km, 7 km, 10 km, 21 km, and 42 km, from which only one has to be chosen by participants. Based on that, athletes of multi-distance running events always use to be separated and distributed, that nature can be analyzed and characterized through this part of the study.

**Extraordinary factors**

Extraordinary environmental attributes are able to influence running performance in unexpected ways. In this experimental analysis, three competitions through 2,173 finishers were involved into that part of the study: a mountain race, a marathon phase of an Ironman, and an ultra-marathon running event. Relevant attributes are reported in Table 2.

**Table 2.** Extraordinary environments

<table>
<thead>
<tr>
<th>Environment characteristics</th>
<th>Competitions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MT*</td>
</tr>
<tr>
<td>Distance (km)</td>
<td>11,6</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>22</td>
</tr>
<tr>
<td>Humidity (%)</td>
<td>81</td>
</tr>
<tr>
<td>Average elevation (m/km)</td>
<td>57,84483</td>
</tr>
</tbody>
</table>

* Uphill running  
† Marathon part of an Ironman  
‡ Ultra-marathon running

**3. Results and discussion**

**3.1. Trend description of half-marathon running**

During the last two decades, nature and performance of runners have been significantly changed in Budapest national half-marathon. In Fig. 2 differences between 1997 and 2017 can be clearly seen. More than twenty years ago, faster
and younger runners dominated that competition, but now number of middle aged, slower competitors have been indisputably increasing.

Fig. 2. Visual representation of running performance clusters in the Budapest half-marathon between 1996 and 2017.

The number of runners who have been be able to complete a half-marathon is five times higher this year than twenty years ago. In 1997 the number of finishers could not reach two thousand (1,749), until now this value has been increased around ten thousand (9,054 in 2015). Different trend line estimations have been done for predicting number of finishers in 2020, 2025, and 2030. Estimated numbers, equations, and proper R-square values are reported in table 3 and visualized in Fig. 8. Accordingly, the number of participants might be doubled in 2030.
Using another, re-categorized version of the results between 1996 and 2017 was able to highlight how the performance changed during this time interval. According to the left diagram in Fig. 3, the number of elderly runners (50+) has been increasing slowly, number of middle age participants (30–49) has been increasing faster, but number of younger runners (U 30) has been stagnating within the examination time interval. Due to the other diagram on the right, similar observations can be done to examine pace: number of slower runners (7:00+) has been increasing in a slowly way, middle speed runners’ (5:00–6:59) number has been increasing faster, but number of fast participants (U5:00) has not been statistically changed. A strong correlation between categories of age and pace also can be detected in Fig. 3.
3.2. Effect of variable environments on performance

According to Table 1, effect of variable environment on age and pace can be examined. As we can see on the left diagram in Fig. 4, there is no relevant differences between the analyzed competitions in terms of age categories. On the other hand, distributions of pace categories are much more different due to variable temperature, humidity, and absolute elevation. According to the chart on the right in Fig. 4, the average pace of the runners significantly decreased in case of two competitions Hortobágy(Jun) and Szekszárd(May). On the first event the temperature was quite high (27.2 °C), in turn, on the second event the absolute elevation (177 m) was higher than in case of any other competitions. Consequently, extremely high temperature or steep running course can impair running pace almost similarly.

As we know, optimal temperature for elite marathon runners is around 5–7 °C, naturally this value should be higher in case of recreational half-marathon competitors. An optimal environment might be included by flat course, tiny wind, and advantageous combination of temperature and humidity. According to the analyzed half-marathon events, another observations, like the best environment including not too high and not too low temperature with standard humidity is justifiable. Siófok(Nov) event has got 7 °C with 90% humidity, so temperature was not far from the marathon’s optimal, but humidity was very high. The other faster competition was Budapest(Apr), which had 19 °C with 52% humidity, so the temperature was still not too high and humidity was on a normal level in Hungary.

Fig. 4. Age (left) and pace (right) category distributions of individual finishers in Hungarian half-marathons in 2017.
3.3. Effect of variable distance on performance

According to Fig. 5, the most popular distance is the half-marathon (21 km), while the popularity of marathon (42 km) and 10 km is almost the same. Better-trained recreational athletes run on as long distance as they can, that is why the number of faster runners (U5:00) is the greatest at half-marathon. Distributions of age groups are able to highlight an expected nature of recreational runners, i.e., that number of younger athletes were higher, who completed shorter distance competitions like 5, 7, 10 km than numbers of marathon and half-marathon runners. As previously mentioned, there are different age limitations to participate a half-marathon and a marathon, this is one reason why children use to attend on shorter distances. Those assumptions can be seen on the left chart of Fig. 5.

![Fig. 5. Distributions of age (left) and pace (right) categories of individual finishers in variable distance competitions in 2017 in Budapest, Hungary without normalization.](image)

In case of long-distance running when runnable distance increases, running pace decreases in a linear way, in general (Takács et al., 2004). Variable distance competitions always separate competitors; hence, the aforementioned statement would not be true in that special case. According to Fig. 6, it can be clearly seen, that running pace does not follow the hypothesis, as the fastest distance was the half-marathon and the slowest one was the 7 km competition in Budapest, in 2017. Therefore, significant differences between these competitions cannot be identified in terms of pace.
3.4. Effect of extraordinary conditions on performance

Effects of extraordinary environment conditions on running performance can hardly to identify without any investigations. In Fig. 7, visual matrix representation of three extraordinary and one standard long-distance competition can be seen. In the left top chart, a fundamental marathon was represented, which disposes standard environment. In the right top matrix, evaluated results of a marathon part of an ironman distance triathlon competition can be seen. There were not any participants under 20 who could be faster than 4:00 min/km in that event. In addition, significant differences can be found between pace distributions, but seemingly, that is not as many as it would be expected. Nevertheless, Ironman is not a recreational sport event; just well-trained athletes can manage to complete it, that might explain the previous observation.

A 11 km long mountain running without any downhill phase represented on the left bottom. Age group distributions are almost similar to those of standard marathons, but relevant differences can be found in terms of pace. Although runnable distance of the mountain event is around four times shorter than on a marathon, values of slower pace clusters are still higher. As a result, extraordinary elevation can impair recreational runners’ pace in a similar way as tiredness of ironman triathletes after the cycling phase.
The last chart on the right bottom of Fig. 7 represents an ultra-distance running competition. That running event is a four-day race, where around 50 km can be done by the participants on every day. Although the final distance is extremely high (193 km), runners have resting and sleeping time between the daily stages. Without extremely endurance and mental stability, there is no chance to manage to complete this competition, that fact can be highlighted in Fig. 7. There were not any participants from the fastest and slowest pace cluster, that firstly means distance is too long to complete it with high speed, secondly high training level is essential, that’s why extremely slow pace was also not so typical. Besides, clusters of under 30 and above 70 age groups are almost empty, that refers to an expected nature of ultra-distance running regarding keeping health condition.
4. Conclusions

Starting point of this study was to investigate the influence of endurance running performance (age and pace of runners) by taking meteorological and environmental attributes, running distance and course elevation into account. This study was focusing on how those parameters could influence recreational athletes’ distribution by age and running pace.

According to the first analysis, age parameter strongly correlated with running pace, so there is a statistically optimal age for every runner when recreational endurance sport activities can be done in the most efficient way (Osváth, 2010). Naturally, age is a fixed, unmodified, so-called independent attribute that cannot be influenced by other parameters, like pace (Ketskeméty et al., 2011). This research was also focusing on those results, through which environmental and other influencing parameters might be identified.

According to the Budapest half-marathon trend analysis, numbers of recreational runners has been continuously increased. Nowadays, the number of participants (Fig. 8) almost reached the attendance limit (10,000) which many Hungarian competitions have. Thus, organizers will have to consider how this situation can be handled to find a balance between willingness of athletes and organizing long-distance running events in a proper, securely way.

![Fig. 8. Numbers of individual finishers in Budapest half-marathons from 1996–2017.](image)

In Hungary, half-marathon competitions are the most famous running events for recreational athletes. The age groups distributed into the same way, regardless of locations and event dates. In this case, only pace attribute is carrying relevant information about running performance.
Those competitions, which have extraordinary environmental components like extremely long distance or steep running course, are managed to complete by trained athletes having well physical condition, rather than recreational participants. For that reason, such sport events are hardly to complete for young (under 20) and elderly (above 60) persons, that is why range of age categories could be shrunk.

A main result of the study is that the distance and elevation attributes of running environment impair running speed mostly, otherwise outside temperature and humidity have less effect on recreational long-distance running pace. The direct effect of distance can be seen in Fig. 7, greater pace than 4 min/km could not be reached, but in all other analyzed distances it could be possible. Uniform elevation without uphill and downhill pats can be run easier than mountain competitions. That can be perfectly highlighted by Figs. 4 and 7. At a half-marathon at Szekszárd, distribution of pace categories has been shifted: the number of faster runners (pace below than 5 min/km) was one third, while the number of slow runners (pace greater than 7 min/km) was three times greater than at a flat competition in Keszthely with similar air temperature and humidity.

4.1. Future direction

As can be seen, wide range analyses can be done by using two-dimensional visual matrix representation without any deeper primary data for investigating the influencing parameters. Comparing multi-distance competitions is possible through this model, but different sport events like running, swimming, kayaking, and cycling can be also compared. This tool would be support different types of statistical analyses in variable scientific areas like information technology, sport science, telecommunications, education, and society. In general, age and time is a very important attribute in many human and technological procedures as well, which could be analyzed by some kind of modified version of this Running Competition Comparison Tool model.

Another possible research direction is taking further investigation in terms of effects on meteorological conditions and physiological attributes of endurance running athletes by analyzing primary data collected in an environment chamber in 2017 at the Auckland University of Technology, New Zealand.

Acknowledgements: The authors would like to express their gratitude towards Dr. László Bozó for his valuable input on the mechanics of front type transitions. The authors also owe gratitude to Dr. Kálmán Kovács and the Data Supply Department of the Hungarian Meteorological Service for sharing the extensive local and medical meteorological and accident databases for the purpose of completion of the authors’ research. The research reported in this paper was supported by the BME- Artificial Intelligence FIKP grant of EMMI (BME FIKP-MI/SC), and by the EFOP 3.6.1. -16-2016-00014 program of the Ministry of Human Capacities (Disruptive Technologies).
References


