

Middle- and long-distance races viewed from the perspective of complexity: Macroscopic analysis based on behaviour as a power law

 © by IAAF
21:1; 17-25, 2006

By Juan M. García-Manso, Juan M. Martín-González, Enrique Arriaza, Lucía Quintero

The term 'power law' describes the organising principle that very few nodes will maintain a large percentage of links in a network or system. In a continuation of an earlier work, the authors use this idea to characterise the different performance levels into which top-class male athletes in the middle- and long-distance races in athletics can be grouped. They assume that the total system has a critical behaviour and that the performances in these races should strictly follow a power law. Using the best times of the all-time top 550 ranked performers in the events from 1500m to marathon (excluding the steeplechase) on 30 October 2003 as basis for analysis, they attempt to detect those values (performances) that show clearly abnormal behaviour within their performance level and thus compare the level of one event to the others. A box-plot of the residuals from the regression model is used to analyse exceptional performers or outliers, who act as targets, barriers and/or powerful attractors that increase the level of performance in one event in comparison to the other distances analysed.

ABSTRACT

Dr. Juan M. García-Manso works in the Department of Physical Education at the University of Las Palmas of Grand Canaria in Spain. He is an athletics coach and was formerly responsible for the national junior team of Spain.

Dr. Juan M. Martín-González works in the Department of Physics at the University of Las Palmas of Grand Canaria in Spain. He holds a doctorate in Mathematics.

Prof. Enrique Arriaza works at the University of Valparaíso, Chile. He is a professor of Training Theory.

Lucía Quintero is a graduate student in Physical Activities and Sport in the Department of Physical Education at the University of Las Palmas of Grand Canaria in Spain.

AUTHORS

Introduction

In an earlier paper (GARCIA-MANSO et al., 2005), we verified that performances in the middle- and long-distance athletics races behave as power laws when time (or average velocity) and distance are related, regardless of their individual characteristics. This suggests the presence of critical phenomena. We also emphasised the importance of the *universe* (the

number of people involved in the activity) but we limited the study to the world's best athletes at each competition distance, referring particularly to the world records for each distance.

In the present study, we used the top 550 positions in the all-time world rankings for 1500m, 3000m, 5000m, 10,000m, half marathon and marathon events as samples. These give only the best time for each of the athletes running the different competition distances. The world records represent the times officially recognised as such as of 30 October 2003 and the times used for the analysis of the other positions correspond to those athletes occupying that position in the all-time world ranking as of the same date.

The data we utilised showed that, in spite of the performance level (world ranking position) of the individual concerned, when the

average velocity of a race is related to the distance covered, the same type of scaling law is always found:

$$V(r) = C(r) d^{-\alpha(r)} \quad (r = 1^{st}, 2^{nd}, \dots, 100^{th}, \dots) \quad (1)$$

In Equation 1, r is the athlete's position in the all-time world ranking, $C(r)$ and $\alpha(r)$ are constants at each level of performance r , while d represents the competition distances covered. The value $\log(C(r))$ represents the intersection point of the regression line at each level with the axis $\log(v)$ and, to some extent, is an indicator of how average race velocity decreases as we go down the world ranking. This fact led us to define $C(r)$ as "Performance Index" (PI), which also seems to roughly follow a scaling law.

A deeper analysis of the PI shows the existence of natural barriers, in the evolution of the times, which seem to correspond to per-

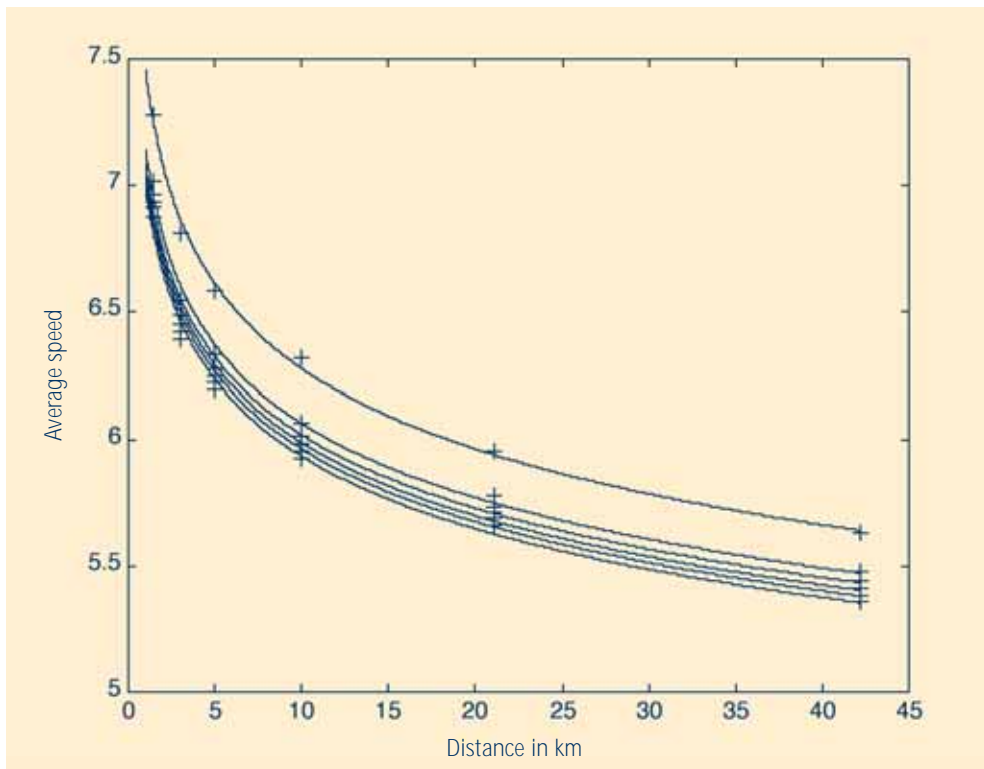


Figure 1: Graphic representation of equation (1) fitted for specific examples of athletic performance (1st, 100th, 200th, 300th, 400th and 550th ranked performers).

Table 1: Numerical results obtained for the differences $u^d(r) - \hat{u}^d(r)$, where $u^d(r)$ is the logarithm of the average velocity of subject r^{th} for competition distance d , and $\hat{u}^d(r)$ is the value given by the regression line. The table shows only some of the positions.

Ranking	1500m	3000m	5000m	10,000m	Half-marathon	Marathon
1st	0.0064	-0.0087	-0.0040	0.0063	0.0018	-0.0018
10th	0.0081	-0.0096	-0.0044	0.0039	0.0032	-0.0012
50th	0.0103	-0.0106	-0.0041	-0.0005	0.0047	0.0002
100th	0.0104	-0.0092	-0.0059	-0.0005	0.0044	0.0008
150th	0.0117	-0.0106	-0.0060	-0.0009	0.0045	0.0013
200th	0.0121	-0.0104	-0.0071	-0.0007	0.0048	0.0013
250th	0.0121	-0.0108	-0.0067	-0.0006	0.0051	0.0010
300th	0.0125	-0.0115	-0.0065	-0.0009	0.0057	0.0007
350th	0.0124	-0.0113	-0.0064	-0.0014	0.0059	0.0008
400th	0.0124	-0.0114	-0.0063	-0.0013	0.0061	0.0005
450th	0.0125	-0.0116	-0.0064	-0.0012	0.0059	0.0007
500th	0.0127	-0.0118	-0.0063	-0.0012	0.0056	0.0009

formance levels, or significant times, in an athlete's evolution towards better results and world records. Furthermore, the way in which PI times are distributed on different scales seems to show an underlying multifractal structure.

The sporadic appearance of subjects able to deliver times that are clearly better than those existing, would define new goals or targets that would act as attractors or reference points for other athletes. At the same time, a new elite group or area specific to the race in question would be established. As this happens, Equation 1 would be defined more clearly as a power law to which all the values would evolve in these resistance races.

The present study will analyse the current situation of three groups of athletics running events (middle-distance, long-distance and marathon) in terms of the corresponding power laws that characterise the different performance levels that group together the athletes occupying the top 550 positions of the all-time world ranking for the respective

distances. We will thus try to detect those values that show clearly abnormal behaviour within their performance level. With this intention, we shall assume that the total system has a critical behaviour and that the races should strictly follow the power-law shown in the Equation 1. Or, in other words, the differences in each average velocity value (or time taken) to the curve (or logarithms to the regression line) should be or tend towards zero. To this end, we have taken Equation 1 to a logarithmic form:

$$u(r) = \log(v(r)) = \log(C(r)) - a(_r) \log(d) \quad (2)$$

From this point onwards, we will use this equation as a reference base for each position in the ranking. We calculated the differences (residuals) from the actual data (time or velocity) to the value suggested by the regression model. This allows us to compare the position or state of each race to the rest, as well as to organise the times for each competition distance. We will thus have $D^d(r) = u^d(r) - \hat{u}^d(r)$ where $u^d(r)$ is the logarithm of the average velocity of subject r^{th} for competition

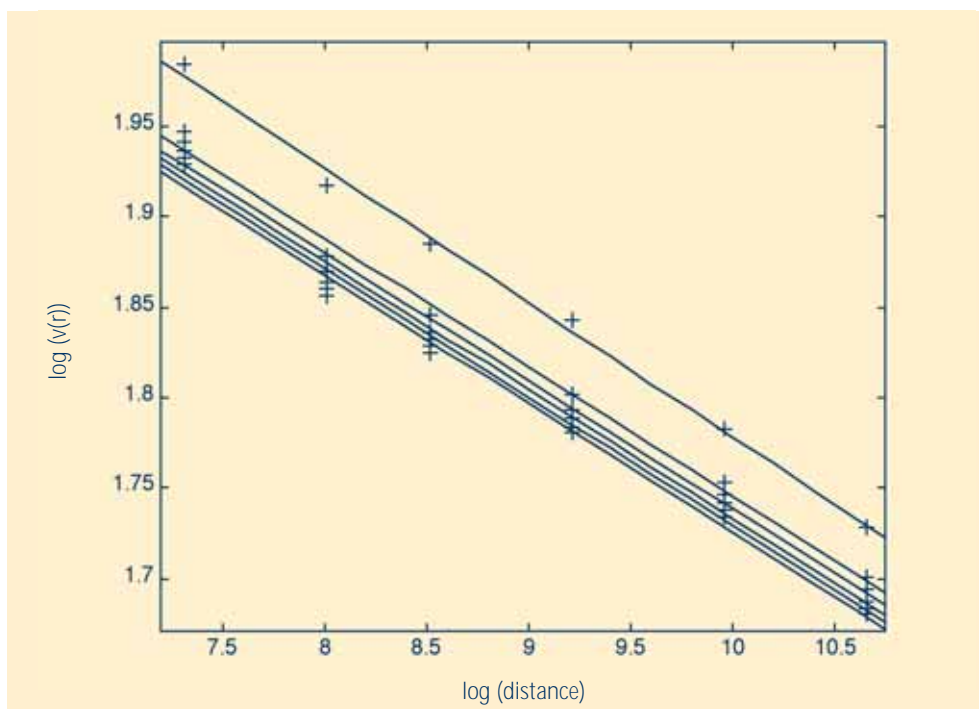


Figure 2: Log-log plot of equation (1) for the same levels (1st, 100th, 200th, 300th, 400th and 550th) in the men's all-time world ranking, together with their regression lines.

distance d , and $\hat{u}^d(r)$ is the value given by the regression line. The numerical values, for some of the positions, are given in Table 1.

To interpret the results we shall use a box-plot of the matrix $D^d(r)$, with $r = 1\text{st}, 2\text{nd}, \dots, 550\text{th}$ for each value of d . The box-plot produces a box and whisker plot for each d value. The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. In our box-plot analysis, an outlier corresponds to a time whose value is more than 1.5 times the inter-quartile range away from the top or the bottom for the specific box for each race distance.

Detection of atypical behaviour in the races

If we analyse the competition distances in Figure 2, we can see that in each event there are extraordinary runners whose performance

levels do not conform to the norm of the compiled data. The incidence of this situation is not the same, either in the way it occurs or in proportion, in the different distances studied.

Thus, we can see (Figure 3) that this type of runner (known as an *outlier*), capable of recording times that are significantly different to those of most of the athletes analysed in the series used for each distance, is more frequent in the 1500m (34 athletes) and 5000m (27 athletes). By contrast, the tendency is something different for the 3000m (11 athletes), 10,000m (16 athletes), half marathon (7 athletes) and marathon (15 athletes) races.

We understand that outliers act as a kind of target, barrier and/or powerful attractor leading the event towards a greater position in comparison with the other distances analysed. The importance of the outliers and their effect on the evolution of such a system depends on many factors. However,

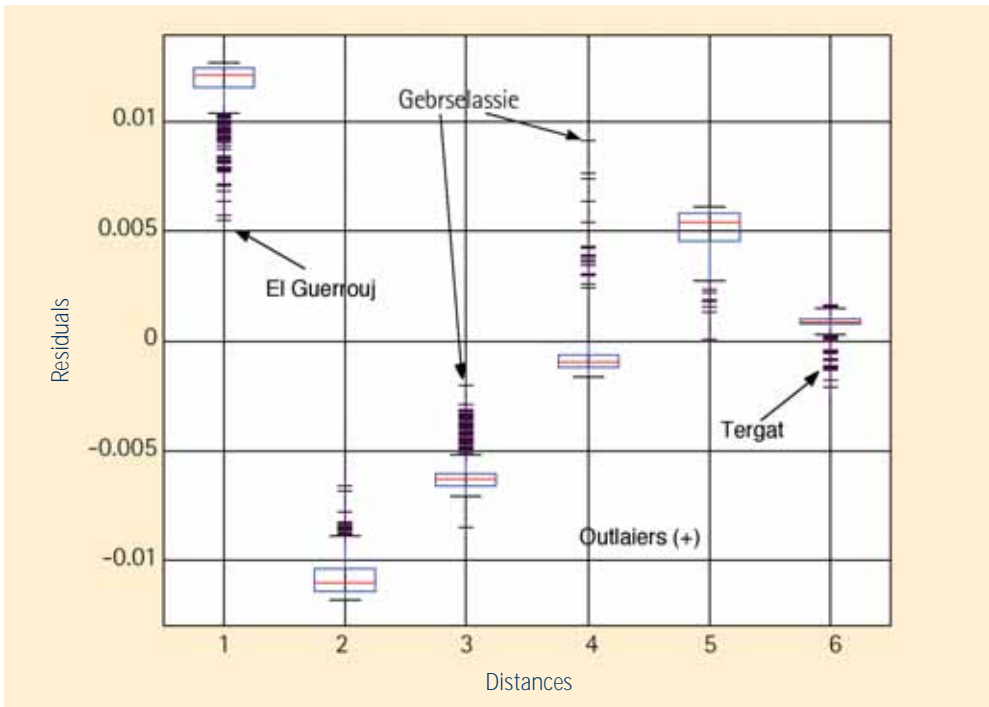


Figure 3: Box-plot of the matrix $Dd(r)$, with $r = 1st, 2nd, \dots, 550th$ for each of the distances considered d_i , ($i = 1, \dots, 6$), equivalent to 1500m, 3000m, 5000m, 10,000m, half marathon and marathon, in this order.

given the practical experience that characterises these races, we can mention the most relevant factors: the potential value of the time, the number of outliers in the event and the circumstances in which the exceptional times are run. The quality of the behaviour in this case is relevant given that, according to our initial hypothesis, these attractors are responsible for leading the system (composed of competition distances) to a critical state.

For example, the record and/or times for 10,000m and marathon races behave like outliers in comparison with the rest of the times that appear in the ranking used, although they do so in different ways. The outlier for the 10,000m (difference 0.0063, in Table 1) shows the extraordinary merit of the world record (26:22.75) recognised at the time this table was produced, not to mention the subsequent performances by Bekele (ETH). Other athletes at this distance will find the record extremely difficult to beat. Sometimes,

although it is not this case, experience shows us that extraordinary times can be attributed to the effect of variables that are not connected with the athlete. These may include particularly favourable competition conditions (excellent pace-marker, ideal weather conditions, significant competition incentives, sea level, etc.). Alternatively, factors that are difficult to control may come into play (these might include doping). Besides these considerations, proximity to the fit and the presence of an outlier, the 10,000m shows a relatively compact behaviour, with a normal distribution of recorded times and low dispersion of times (Figure 3).

The overall behaviour of times in the marathon is very similar to that observed in the 10,000m, with concentrated times and a low level of dispersion, showing the internal normality of this race, particularly if we consider the top positions. The previous world record (2:05:38) by Khannouchi (USA), seemed, at the time, to differ significantly

from the average behaviour of other values analysed for this distance. However, at that time, its value could be considered as low in comparison with the rest of the times recorded for other races and for the same distance and, therefore, open to improvement in the short term by a fair number of specialists. This was confirmed by the times recorded by different athletes at the end of 2003, particularly by Tergat (KEN) and Korir (KEN), who broke the existing world record during the Berlin marathon with their respective times of 2:04:55 and 2:04:56. Despite this latest qualitative change in the all-time records, an overall analysis of the event seems to indicate that the ranking list will undergo significant changes over coming seasons.

This hypothesis of the future can also be accepted from a physiological point of view. We understand that a runner with a low BMI (Body Mass Index), a VO_2max near to $80 \text{ ml/kg}^{-1}/\text{m}^{-1}$, and the ability to run the distance to an intensity close to 90% of this value with a high economy ($\sim 0.185\text{-}0.190 \text{ ml.kg}^{-1}.\text{m}^{-1}$), will be able to beat these marks or others of greater level.

Analysis by race groups

For this analysis, we organised the competition distances into three categories (middle-distance, long-distance and marathon), which represent similar distances in terms of the athletes who are successful in each category. Obviously, each of these categories could be reorganised based on the individual profile of its athletes, but we do not think this step is either necessary or helpful in this overall race analysis. The middle-distance group includes the 1500m, 3000m and 5000m; the long-distance group covers the 10,000m and the half marathon while the third category includes only the marathon.

Middle-distances

The current situation of races traditionally known as middle-distance events constitutes a very interesting case (Figure 3). From a metabolic point of view, runners of 3000m

and 5000m races have the maximum requirements for aerobic power, which correspond to times of about 420sec (PERONNET and THIBAUT, 1989), 450-550sec (ALVAREZ-RAMIREZ, 2002) or 600sec (MORTON and BILLAT, 1999). By contrast, the 1500m is the frontier between the resistance and speed events (power and/or speed endurance). When we analyse the position of these events on the box-plot, we realise the low relative value of the times for the 3000m and 5000m. This tendency increases as the performance of the athletes decreases. This type of behaviour, although it clearly exists, is less acute in the longer distance (5000m), though we have to bear in mind that this distance has been included in the official programme of the Olympic Games, the World Championships in Athletics and the continental championships since these competitions began. On the other hand, the 3000m has only been included in indoor competitions or some outdoor invitational meetings over the last few decades. This leads us to think that the all-time ranking of these two distances, particularly the 3000m, could undergo a significant change and improvement if the following variables were changed: an increase in the universe of athletes with the current training characteristics and/or incorporation of the 3000m in the programmes of the major outdoor championships. We believe that the current world records for these two distances could be improved, and that the new records would drag the other competing athletes with them to a higher level of performance. The potential record times could lie around 7:16.55 (3000m) and 12:36.55 (5000m), similar to the 5000m world record of 12:39.36 used in our analysis and Bekele's subsequent mark of 12:37.35.

The 3000m shows relatively homogeneous times with a small inter-quartile range and low dispersion, and the times show a certain tendency to approach the outliers.

The times for the 5000m are similar, but their behaviour is different, given that in this case they form a compact block around a

concentrated average, with the obvious influence of a group of specialists who powerfully drag the race towards its natural position. This suggests to us the existence of a specific specialist profile (with times of <13:00) clearly differentiated from other world-class runners (whose times range between 13:15 and 13:30). Athletes from North Africa (Morocco and Algeria) and East Africa (Kenya and Ethiopia) have played an important part in shaping the dynamics of the 5000m.

The 1500m behaves in a very different way to the two races we have just discussed. This race has two significant characteristics: first, it is a key event in athletics (its popularity giving it a greater *universe*), and second, it is located close to the border between the endurance running (>1000m), determined mainly by aerobic metabolism, and the speed events (<1000m), which rely on anaerobic metabolism.

The results obtained in the series of times used demonstrates the high average value of the times for the 1500m in relation to the tendencies found in the other events analysed. There are three aspects that might affect the position occupied by this distance: energy dependence, it's *universe* and the profile of the athlete currently running the best times in the world.

Aerobic metabolism plays a very important role at this distance, although the metabolic contribution is significantly different in the case of each athlete according to his functional profile, muscular structure and physical fitness. In this race, aerobic metabolism would appear to act as an important attractor in the organism of current world-class athletes. The anaerobic metabolism is also a determining factor, although in a different proportion. According to WARD-SMITH (1999), the full potential of the anaerobic capacity is available for conversion during extended periods of running, but other authors say that the anaerobic energy contribution declines with race duration (GOLLNICK and HERMANSEN, 1973); PERONNET and THIBAUT, 1989).

According to most studies, aerobic metabolism is attributed as contributing between 60% and 90% of the total energy contribution (SPENCER et al., 1996; WEYAND et al., 1993). These figures include a wide degree of variability between individual athletes, which gives cause for thought. Clearly, part of this difference could be explained by the characteristics of the samples used in the different studies undertaken (whether they were physically very trained, quite trained or sedentary), but we need to consider other parameters as well. Experience has shown us that there are three prototypes for athletes running the 1500m: those that run this distance as well as shorter distances (800m and 1000m); those that run the 1500m well, but also perform well at longer distances (3000m and 5000m) and genuine specialists in the 1500m. This practical reality could explain some of the causes underlying this very wide range of energetic behaviour, although we should also bear in mind the possible effect of the procedures used in metabolic assessment.

Generally speaking, the main functional factors restricting performance in the 1500m could be taken as the elevated depletion of the muscular CrP, the high metabolic acidosis produced by a significant activation of the glycolytic metabolism and the insufficient capacity of the aerobic metabolism to produce enough energy. The importance of each of these aspects to the result of a race varies in each athlete, his functional profile and the type of training employed. The current world record holder at 1500m, El Guerrouj (MAR) is a particularly interesting case, because of his exceptionally good time (3:26.89). His times appear closer to the regression line (distance: 0.0064, in Table 1) than the rest of the times used in the study (up to the all-time 550th position). This led us to think that this athlete could be included in the group of runners who also run well at longer distances (2000m, 3000m and even 5000m). This opinion was confirmed when we checked his best times over 3000m (7:23.09) and 5000m (12:50.24).

If we add to the mixed metabolic dependence, the enormous historical importance of this event and the high number of athletes who have competed at this distance over the last century (*universe*), we should expect a different behaviour to that shown in the other events used to determine the scaling law proposed in this paper. If we look at Figure 2 can see that at all levels (from position 1st to 550th), this distance shows a much higher level of behaviour than the other distances, particularly in comparison with the next two events (3000m and 5000m). This could correspond to the two points analysed above.

Long-distances

The 10,000m provides the stable model around which the other distances analysed in this study vary. However, there are some details relating to the best times for this distance that we feel should be mentioned. The world record at this distance when the figures were produced, as well as the times recorded by the other top athletes over this distance, were all recorded at the end of the last century, when they broke the barrier of 27 minutes quite easily. Initially, these times produced a significant effect on the event. A subsequent period of stagnation appears to have been overcome now. This leads us to think that, all things being equal, this event is unlikely to see any spectacular surprises over the next few years.

The half marathon is a particularly interesting event. From here on, the scaling law, and its possible explanation, is somewhat more complex, but not for that reason less interesting. If we look at Figure 2, we see that, regardless of the level we look at (positions 1st,...,500th) the results of the half marathon are always clearly situated above the fit and, therefore, above those of the adjacent events. The box-plot (Figure 3) shows that half marathon times do not show many outliers, and that all levels present analogue distances (0.0030 – 0.0060, in Table 1). This behaviour is interesting if we bear in mind that this distance is not included in the Olympic Games,

World Championships in Athletics or the continental championships. It has only recently become widely popular (in the last three decades), although the level of participation is very significant. Given the results obtained, we could posit that performances by top class athletes in the half marathon benefit from the fact that this distance falls within the range of two types of specialists (10,000m and marathon). This produces better performance and bodes well for the event's evolution.

Marathon

The marathon is clearly different from the long-distance races considered above, as top-level success calls for a highly specialised type of runner whose energy needs (and therefore training needs) are clearly differentiated from the runners of shorter distances. There is no other sufficiently developed race with similar characteristics, in terms of length, functional or physical dependence, that can be considered in this category.

From the metabolic point of view we know that races over two hours in duration significantly increase the participation of fats in the aerobic metabolism: through β -oxidation replacing the reserves of glycogen, which is the major source of energy in races lasting around an hour (for example the half marathon) (BERGSTRÖM and HULTMAN, 1967; SALTIN and LARLSSON, 1971; COSTILL et al, 1971 and 1973; SHERMAN et al, 1981; MADSEN et al, 1990; SAHLIN et al, 1990; WELTMAN, 1995; TSINTZAS et al, 1996^a; HAWLEY et al, 1997). In the marathon, the runner's energy dependence on fats is around 20%, although the importance of this substrate increases with the length or duration of the race, and can reach 60-70% in 100km races (NEWSHOLME et al, 1992; LEIBA and TERRADOS, 1996).

The change in energy dependence in races lasting 90-120 minutes could possibly be expressed in a new scaling law, which would allow us to find the decisive point between long-distance and marathon. This cut-off point could define the limits between the two

aerobic metabolisms described above (carbohydrates and fats). However, in order to do this we would need enough information about other long-distance races (for example, 50km or 100km races). Unfortunately, this type of race is not sufficiently developed for our purposes, and potential top athletes for these races normally opt for those events that are currently more popular, such as the

triathlon. The only distance in athletics that has an established competitive tradition is the 100 kilometres, but the number of participating athletes is low and there are very few genuine specialists at this distance. This fact limits the effective *universe*, gives a biased behaviour pattern over this distance and prevents us from broadening the scope of our work.

REFERENCES

1. ÁLVAREZ-RAMÍREZ, J.: An improved Peronnet-Thibault mathematical model of human running performance. *European Journal Applied Physiology*, 86, (2002) 517-525.
2. BERGSTROM, J.; HULTMAN, E.: Muscle glycogen synthesis after exercise: an enhancing factor localized to the muscle cells in man. *Nature*, 210, (1967). 309-310.
3. COSTILL, D.L.; BOWERS, R.; BRANAN, G.; SPARKS, K.: Muscle glycogen utilization during prolonged exercise on successive days. *Journal Applied Physiology*, 31, (1971). 834-838.
4. COSTILL, D.L.; GOLLNICK, P.D.; JANSSON, E.D.; SALTIN, B.; STEI, E.M.: Glycogen depletion pattern in human muscle fibres during distance running. *Acta Physiological Scandinavica*, 89, (1973). 374-383.
5. G^a-MANSO, J.M.; MARTÍN-GONZÁLEZ, J.M.; DÁVILA, N.; ARRIAZA, E.: Middle and long distance athletics races viewed from the perspective of complexity. *Journal Theoretical Biology*, 233: (2005). 191-198.
6. GOLLNICK, P.D.; HERMANSEN, L.: Biomechanical adaptations to exercise: anaerobic metabolism. In: Wilmore, J.H. (Ed.), *Exercise and Sports Science Reviews*. Vol. 1, Academic Press, New York, (1973). 1-43.
7. HAWLEY, J.A.; PALMER, G.S.; NOAKES, T.D.: Effects of 3 days of carbohydrate supplementation on muscle glycogen content and utilisation during a 1-h cycling performance. *European Journal of Applied Physiology*, 75: (1997). 407-412.
8. LEIBAR, X.; TERRADOS, N.: Un approccio biomédico alla corsa di maratona. *Rivista di Cultura Sportiva. Scuola dello Sport*, 34, (1996). 46-54.
9. MADSEN, K.; PEDERSEN, P.K.; ROSE, P.; RICHTER, E.A.: Carbohydrate supercompensation and muscle glycogen utilization during exhaustive running in highly trained athletes. *European Journal Applied Physiology*, 61, (1990). 467-472.
10. MORTON, R.H.; BILLAT, V.: Maximal endurance time at v_{O2max} . *Medicine & Science in Sport and Exercise*, 31, (1999). 1496-1504.
11. NEWSHOLME, E.A.; BLOMSTRAND, E.; MCANDREW, N.; PARRY-BILLINGS, M.: Biochemical causes of fatigue and overtraining, in *Endurance in Sport. The encyclopaedia of sports medicine*, (ed) Shephard, R.J., and Åstrand, P.O., Blackwell Scientific Publications, London, (1992). 351-364.
12. PERONNET, F.; THIBAUT, G.: Mathematical analysis of running performance and world running records. *Journal Applied Physiology*, 67, (1989). 453-465.
13. SAHLIN, K.; KATZ, A.; BROBERG, S.: Tricarboxylic acid cycle intermediates in human muscle during prolonged exercise. *American Journal Physiology*, 259, (1990). C834-C841.
14. SALTIN, B.; KARLSSON, J.: Muscle glycogen utilization during work of different intensities, in *Muscle metabolism during exercise*, (ed) Pernow, B and Saltin, B. Plenum Press, New York, (1971). 289-300.
15. SHERMAN, V.N.; COSTILL, D.L.; FINK, W.J.; MILLER, J.M.: Effect of exercise-diet manipulation on muscle glycogen and its subsequent utilization during performance. *International Journal in Sports Medicine*, 2, (1981). 114-118.
16. SPENCER, M.R.; GASTIN, P.B.; PAYNE, W.R.: Energy system contribution during 400 to 1500 metres running. *IAAF. New Studies in Athletics*, 11(1996)., 59-65.
17. TSINTZAS, O.K.; WILLIAMS, C.; BOOBIS, L.; GREENHAFF, P.L.: Carbohydrate ingestion and single muscle fiber glycogen metabolism during prolonged running in men. *Journal of Applied Physiology*, 81, (1996). 801-809.
18. WARD-SMITH, A.J.: The bioenergetics of optimal performances in middle-distance and long-distance track running. *Journal of Biomechanics*, 32: (1999). 461-465.
19. WELTMAN, A.: The blood lactate response to exercise. *Champaign (IL). Human Kinetics*, (1995). 1-115.
20. WEYAND, P.; CURETON, K.; CONLEY, D.; SLONIGER, M.: Percentage anaerobic energy utilized during track running events. *Medicine & Science in Sport and Exercise*, 25, (1993). S105.