Journal of Sports Sciences, 1999, **17**, 467–476

Power laws and athletic performance

J. SYLVAN KATZ[1]* and LEON KATZ[2]

[1]*Science Policy Research Unit, University of Sussex, Brighton, East Sussex BN1 9RF, UK and* [2]*Department of Physics, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0, Canada*

Accepted 26 July 1998

In a previous study, we showed that the 1992 men's world record running times in the 100 m to 200 km could be represented accurately by the equation $T = cD^n$, where T is the calculated record time for distance D, and c and n are positive constants. Here, we extend that to cover the years 1925–65 at 10-year intervals and 1970–95 in 5-year intervals for distances of 100 m to 10 km. Values of n for all years lie along a straight line with a small negative slope. A regression analysis yields an equation for values of n covering the period 1925–95. Values of c from 1925 to 1995 were fitted by a quadratic equation. These two equations define a surface in three-dimensional space $\langle \log(T), \log(D), \text{date}\rangle$ for all men's world record runs over the 70-year period for distances of 100 m to 10 km. We also demonstrated previously that event times, t, do not scatter randomly with respect to the values of T but form a consistent pattern about the straight lines in $\log(T)$ versus $\log(D)$ plots. In this study, we show that the pattern of $(t - T)/t$ as a function of date has remained constant for the past 70 years.

Keywords: athletic limitation, chaos, energy, fractal, non-linear dynamics, power law.

Introduction

Since the early 1900s, many attempts have been made to develop an equation which represents world record running times. Kennelly (1906) was the first to use a power law to relate running time, T, to distance, D, for athletic events and horse races (equation 1). Equation (1) will be referred to as the power law and its log-log graph will be called a log-log plot. In equation (1), c and n are positive constants:

$$T = cD^n \tag{1}$$

Since then, others (Lietzke, 1954; Henry, 1955; Riegel, 1981; Blest, 1996) have used this equation because of its excellent fit to the data. The large range of event times and distances involved in these analyses (over two orders of magnitude) and the general monotonic character of the data have resulted in the power law equation fitting the data with R^2 values (coefficients of determination) very close to 1. However, even a linear regression of T on D has an R^2 of close to 1. Thus an important question is, 'which equation is most appropriate for this type of analysis?'

* Author to whom all correspondence should be addressed. e-mail: j.s.katz@sussex.ac.uk

0264–0414/99

Despite the excellent agreement between the power law or log-log linear representation of world running records, detailed analysis will show that, except for a small random variation, the event times are scattered about the regression lines in a pattern that has remained constant for more than seven decades. We show that a power law plot of the data values at 200 m and 400 m are consistently about 6% below the regression line, and at 1 km, 1.5 km, 2 km and 1 mile the time values are consistently about 3% above. In the linear model, the 100, 200 and 400 m points are 209%, 73% and 13% respectively above the regression line. The power law model appears to result in a better fit to the data. However, there is a more fundamental reason for choosing it, which we will explain shortly. The distance of the power law regression value from the actual data expressed as the percent change in the ordinate will be called the 'power law offset' or simply the 'offset'.

To our knowledge, the origin of these offsets has not previously been studied, and a possible explanation for their cause has not been proposed. This will form an important part of the present analysis. Henry (1955) considered offsets in his analysis of speed versus distance; however, speed is not an appropriate variable for such an analysis because it is an average quantity, making the offsets more difficult to interpret. Before presenting the details of this analysis, we explain why we chose the power law.

With the development of the computer, it became possible for scientists to explore the behaviour of non-linear dynamic dissipative systems. This exploration culminated in the publication of many research papers in the early 1970s followed by Gleick's (1987) popular book on chaos theory. A parallel independent but related development occurred about the same time, when Mandelbrot (1977) published his seminal book *The Fractal Geometry of Nature*. Both non-linear systems and fractal geometry reveal different facets of the structure and behaviour of real-world systems. As few non-linear equations are solvable by mathematical techniques, it was assumed in the past that many non-linear systems could be closely approximated by a linear system. With the advent of modern computers, the study of non-linear systems revealed a strange and interesting world filled with chaos, order, power laws and fractals. We do not intend to provide a detailed discussion of these developments here. For a mathematical description of the interdependent relationships between non-linear systems, power laws and fractals, see Schroeder (1991). Also, an excellent booklet by Lodge (1981) shows the human senses perceive the intensity of stimuli such as loudness, brightness, taste, smell, pain and pressure with power law characteristics. Buchanan (1997) reviewed many fields in which the power law applies.

From DNA (Oliver *et al.*, 1993) to human physiology (Goldberger *et al.*, 1990; Elbert *et al.*, 1994), the human body is composed of many non-linear systems and fractal structures that can be reasonably represented with power laws. The ubiquitous use of power laws to represent systems in the human body suggests that they may be the logical choice for the present analysis.

We initiated an analysis of world running and swimming records in 1990 to determine what can be learned about a dynamic system from a careful examination of its power laws. Sports data were chosen because they were the most accurate we could find for such a system. The analysis of 1992 world running and swimming records by men and women disclosed some interesting offsets from a strict power law dependence and this was the subject of a previous paper (Katz and Katz, 1994). Here, we extend that work and present new information.

Analysis of world record runs by men

In this study, only runs of 10 km or less were analysed because they are held in stadiums with good tracks and are timed accurately. Timing techniques have improved immensely over the past 80 years. In 1912, the 100-m dash was measured to the nearest 0.1 s, a 1% accuracy; by 1968, timing methods had improved to 0.01 s, a 0.1% accuracy for the 100-m run and a 0.01% accuracy for the 800-m run. Distances have always been measured to an accuracy better than 0.01%. The main source of error in all athletic data arises from the erratic occurrence of record-breaking performances. The 'earlier than expected' appearance of an athlete with super abilities disrupts the even flow of the data. On the other hand, a record can last for a prolonged period. For example, the one-mile record remained unchanged from August 1985 to September 1993, resulting in a sudden improvement of 0.9%.

Men's world running records were explored using power law analysis for the years 1925–65 at 10-year intervals and thereafter at 5-year intervals until 1995 (Table 1). The units adopted in this analysis were minutes for time and kilometres for distance. The use of kilometres is helpful because, as can be seen from equation (1), the parameter c is equal to the regression time, T, in the 1-km event. Thus a graph of c versus the

Table 1 World record times (min) for men's running events

Distance (m)	1925	1935	1945	1955	1965	1970	1975	1980	1985	1990	1995
100	0.1733	0.1717	0.1700	0.1700	0.1667	0.1658	0.1658	0.1658	0.1655	0.1653	0.1642
200[a]	0.3467	0.3383	0.3383	0.3433	0.3367	0.3305	0.3305	0.3287	0.3287	0.3287	0.3287
400	0.7900	0.7700	0.7667	0.7567	0.7483	0.7310	0.7310	0.7310	0.7310	0.7215	0.7215
800	1.843	1.830	1.777	1.762	1.738	1.738	1.728	1.706	1.696	1.696	1.696
1 000	2.477	2.393	2.358	2.317	2.270	2.270	2.232	2.223	2.203	2.203	2.203
1 500	3.877	3.813	3.717	3.680	3.593	3.552	3.537	3.523	3.491	3.491	3.456
1 609[b]	4.173	4.113	4.023	3.967	3.893	3.852	3.823	3.813	3.772	3.772	3.740
2 000	5.433	5.363	5.197	5.037	5.020	4.937	4.937	4.859	4.857	4.847	4.798
3 000	8.460	8.307	8.020	7.927	7.660	7.660	7.585	7.535	7.535	7.491	7.419
5 000	14.470	14.283	13.970	13.677	13.403	13.277	13.217	13.140	13.007	12.973	12.740
10 000	30.103	30.103	29.590	28.903	27.657	27.657	27.513	27.372	27.230	27.138	26.726

[a] Before 1951, 200-m runs were held on a straight track and these values are used. [b] One-mile event.

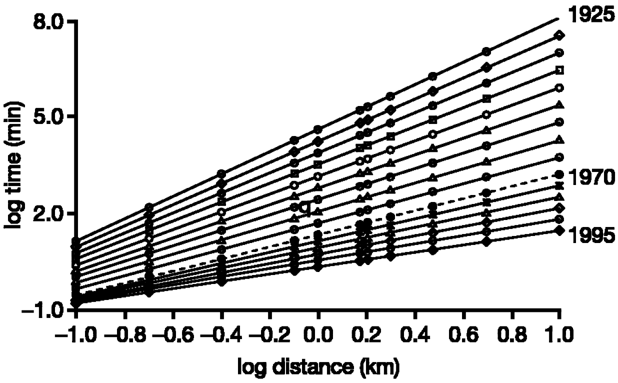

Figure 1 Schematic plot of time versus distance for men's running events in various years.

date shows how the 1-km regression value for T changed over the years. Blest (1996) performed a similar analysis using metres instead of kilometres, resulting in the loss of some useful information, which we discuss later.

Record run times for this analysis were taken from the *Progressive World Record Lists 1913–1970* (IAAF, 1972). The remaining values were taken from various World Almanacs and *Guinness Books of Records*; these were cross-checked against a world-wide web site, http://www.uta.fi/~csmipe/sport/eng/mrun.html, which has up-to-date running event times.

It is not possible to show all the regression lines derived from a log-log plot of the data for each year on a single graph, since adjacent lines would be displayed too close for viewing. Instead, a schematic diagram was constructed with the distances between lines enlarged 100-fold (Fig. 1). At this increased scale, only the calculated values from the regression equation are shown. Since the data for 1925–65 were analysed at 10-year intervals, values for 1930, 1940, 1950 and 1960 were obtained by interpolation. This was done to show the general progression in improvement over the whole period at 5-year intervals. Although the lines in Fig. 1 appear to show a change in the line spacing around 1970, this may in part be due to the way this schematic was constructed and in part due to the distribution of random errors. A more detailed analysis shows that a smooth curve can be passed through the data points with good accuracy.

The changing values over the years of slope, n_i, and parameter, c_i, are shown in Table 2. The slope (Fig. 2a) decreased linearly with time (about 0.024% per year), but the change in c_i required a quadratic equation for its representation (Fig. 2b). This equation should not be used to extrapolate to future years, as it has a minimum in 2037 and rises thereafter. A cubic equation fit to the data was rejected because it has a minimum in 2003. The regression lines through the data are displayed in Figs 2a and 2b. We analysed the world running records reported in Blest (1996) to determine if his data also exhibited the same general shape in the distribution of c_i values. We also tried to select data from Blest for years that we did not analyse. The 1980 data are common to both analyses. It is noteworthy that Blest included times for marathons, which are less reproducible than stadium events because of the different terrains they are run over. These times have been discarded and his distances expressed in kilometres. The resulting slope and c_i values of his data are also shown in Figs 2a and 2b. Lines of equations fitted to these values are lower by about 0.8% and 1.5%, respectively, as Blest did not include the one-mile, 1-km, 2-km and 3-km events in his data. For this reason, his values also show more scatter.

A linear regression analysis of the slope data (Fig. 2a) yields the following equation:

$$\text{Slope} = n_{ic} = (1.1498 \pm 0.0018) - (0.000280 \pm 0.00003)\, Y_i \quad (2)$$

where Y = year − 1900. A more detailed discussion of this analysis will be presented below where we reveal and discuss an important property.

A quadratic equation was fitted to the c_i values by least squares (Fig. 2b) and is given by:

$$c_{ic} = 2.4868 \pm 0.0109 - (0.0058 \pm 0.0004) Y_i + (0.000021 \pm 0.000003) Y_i^2 \quad (3)$$

The goodness-of-fit of the quadratic will be discussed later. Equations (2) and (3) can now be introduced into equation (1) to give a three-dimensional model of the relationship between the variables $\log(T)$, $\log(D)$ and Y:

$$T_{ij} = c_i D_j^{n_i} \quad (4)$$

Table 2 Slope and parameter (c) values with standard errors of the regression lines fitted to the data in Table 1

	1925	1935	1945	1955	1965	1970	1975	1980	1985	1990	1995
Slope	1.141	1.143	1.138	1.134	1.128	1.130	1.129	1.128	1.126	1.126	1.123
σ_n	0.003	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
c-value	2.353	2.315	2.271	2.241	2.200	2.181	2.171	2.158	2.148	2.143	2.129
σ_c	0.009	0.006	0.003	0.004	0.004	0.003	0.003	0.003	0.004	0.004	0.004

Table 3 Values of offsets s_{ij} for all years and distances in the analysis[a]

Distance (m)	s_{ij} 1925	1935	1945	1955	1965	1970	1975	1980	1985	1990	1995	Average s_j (%)	Standard deviation[b] of s_j	Average time-offset $(t-T)_{ja}$
100	1.92	3.15	2.87	2.91	1.68	2.54	3.19	2.96	2.92	3.05	2.25	2.82	±0.35	0.00452
200	−8.15	−8.57	−7.44	−7.54	−6.38	−7.03	−8.14	−6.98	−6.71	−6.45	−6.32	−6.94	±0.67	−0.02412
400	−4.67	−5.39	−4.38	−4.03	−4.60	−5.92	−5.25	−5.09	−4.73	−5.85	−5.47	−5.38	±0.54	−0.03799
800	1.07	2.03	0.83	1.13	1.58	2.51	2.55	1.59	1.44	1.68	2.26	2.01	±0.55	0.02972
1 000	5.02	3.31	3.68	3.19	3.06	3.93	2.88	2.91	2.47	2.71	3.36	3.04	±0.55	0.07657
1 500	3.63	3.52	3.03	3.51	3.25	2.90	3.36	3.21	2.83	3.08	2.89	3.04	±0.25	0.11607
1 609	2.99	3.06	2.95	3.05	3.33	3.06	2.93	3.20	2.65	2.90	2.88	2.94	±0.18	0.11740
2 000	4.52	4.68	3.77	2.33	4.20	3.31	3.89	2.93	3.42	3.48	3.38	3.40	±0.36	0.18343
3 000	2.61	2.15	1.07	1.74	0.80	1.46	1.10	1.13	1.73	1.41	1.48	1.38	±0.25	0.11931
5 000	−1.99	−2.06	−1.60	−1.60	−0.87	−1.27	−1.13	−0.85	−1.20	−1.20	−1.80	−1.24	±0.33	−0.19202
10 000	−8.12	−6.99	−5.60	−5.44	−6.84	−6.41	−6.35	−5.78	−5.52	−5.59	−5.68	−5.89	±0.48	−1.75650

[a] The third last column is the average value, s_{ja}, for the years 1970–95 and the second last column its standard deviation. The last column lists the time-offsets averaged over the period 1925–95.

[b] Only calculated for 1970 and 1995 times.

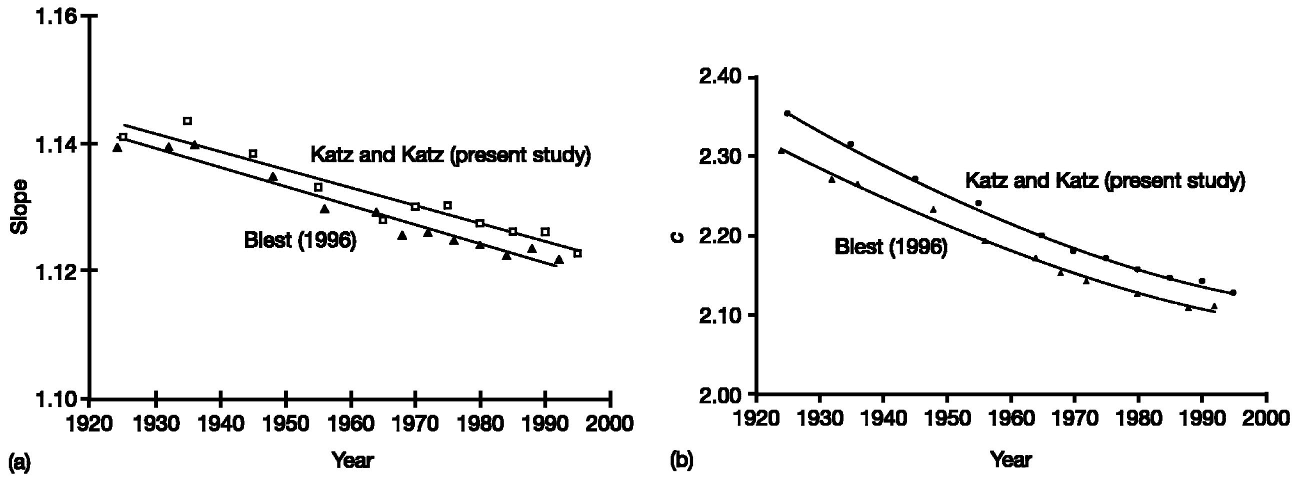

Figure 2 (a) Regression slope versus year for men's running events. (b) Regression intercept (c) versus year for men's running events.

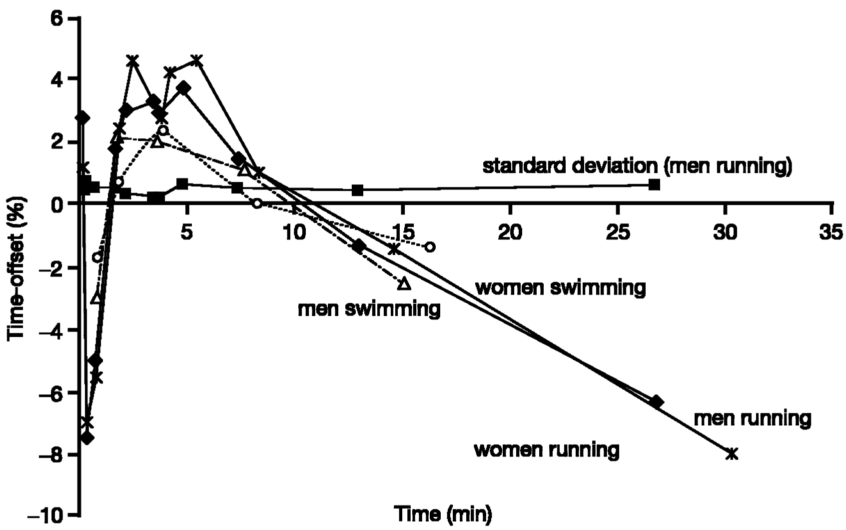

Figure 3 Percent offset versus time for men's and women's running and swimming events.

where j is the event and i the year. This equation defines a ruled surface in three-dimensional space of $\langle \log(T)$, $\log(D)$ and $Y \rangle$ for the period 1925–95.

The offsets of the data from the regression lines have been mentioned several times. The offset values, s_{ij}, of the world record events during the past 70 years are listed in Table 3, where:

$$s_{ij} = 100(t_{ij} - T_{ij})/t_{ij} \quad (5)$$

Again, subscript i represents the year, j the event, t_{ij} the measured event times and T_{ij} the calculated regression times. The average value of the offset of an event j will be designated by s_{ja}. The second and third columns from the end in Table 3 give the offsets averaged over the years 1970–95 (the most accurate values) and their standard deviation calculated from the scatter of these values. Offsets were also calculated for women's running events in 1990 (Katz and Katz, 1994) and men's and women's swimming events in 1995. All these data, including those in the last two columns, are shown in Fig. 3.

Because swims over this range of distances take about five times as long, the fairly good coincidence of all these curves (Fig. 3) shows that the offsets are related to the athlete's *time of exertion* and not the distance covered. Keller (1973, 1974) divided short-distance events into start time and full speed time. Such a division is not relevant in this case, because the start times are quite different in running and swimming yet the offsets are the same when compared on the basis of time. These considerations, combined with the reproducibility of the

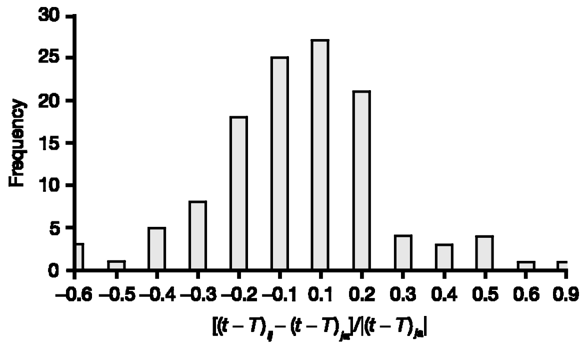

Figure 4 Cumulative frequency distribution of normalized residuals.

offsets for men's running events over the past 70 years, leads to the conclusion that the offsets are mainly the result of a systematic cause related to the exertion time of the athlete, and only 14.7% of the offset values can be attributed to random effects. Within the accuracy of the data, the offsets have not changed over the past 70 years.

Based on this conclusion, it is now possible to analyse the data in greater detail, to comment on the least square analysis and make a goodness-of-fit test of the regression lines to the slope n_i and intercept c_i values shown in Figs 2a and 2b. As the offsets at each event have remained relatively constant over the 70-year period of this analysis, it is appropriate to subtract average (1925–95) 'time-offsets' $(t-T)_{ja}$ from each event j. This ensures any random variations (residuals) remain in the remaining time value; we label these corrected event times $t_{ij}*$. This procedure does not take into account minor corrections because the event times changed over the years (this is why values of s_{ij} were expressed in terms of percent change).

An analysis of the residuals showed that they were randomly distributed; however, because the values in Table 1 vary by more than two orders of magnitude, the residuals were first normalized by the following equation to allow comparison:

$$[(t-T)_{ij}-(t-T)_{ja}]/\,|\,(t-T)_{ja}\,| \qquad (6)$$

When ordered, these values yield the curve shown in Fig. 4. A comparison of the cumulative frequency under this curve with that of a normal distribution (Ogive plot) results in a straight line ($R^2 = 0.99$), showing that the distribution of the residuals is normal. This showed that the regression lines calculated by least square analysis of $\log(t^*) - \log(D)$ have minimum variance. The standard deviation of the $\log(t^*) - \log(D)$ regression lines is also reduced. From equation (1), the equation for these lines is:

$$\log(T_{ij}^*) = \log(c_i) + n_i \log(D_j) = a_i + b_i x_j \qquad (7)$$

where $T_{ij}*$ are the new regression line values of event j in year i with $(t-T)_{ja}$ subtracted. The standard deviations of $\log(c_i)$ values, σ_{ia}, are on average 14.5% of those without subtracting $(t-T)_{ja}$; the standard deviations of n_i, σ_{ib}, are on average 15.2% of those calculated without this subtraction. The recalculated values of the slope, n_i, and parameter, c_i, differ by less than one part in 10^4 from those calculated without the time-offset subtraction.

The standard deviations associated with the constants of the regression lines $T_{ij}*$ are σ_{ia} and σ_{ib}, respectively, and the variances of the lines for the 1-km event in different years are given by the following expression (see Sveshnikov, 1968, for an example):

$$\sigma_i^2(x) = \sigma_{ia}^{\ 2} + \sigma_{ib}^{\ 2}x^2 + 2\sigma_{iab}^{\ 2}x \qquad (8)$$

Because the value of x at 1 km is zero (i.e. $\log(1) = 0$), the last two terms drop out giving a variance $\sigma_i^2(0) = \sigma_{ja}^{\ 2}$. These are the values of the standard deviations of c_i listed in Table 2. Calculation of the goodness-of-fit of the quadratic equation to the c_i values gave $F_{3,7} = 733$ and $P(F_{3,7} > 8.45) = 0.01$, indicating an excellent fit.

We realize that the columns in Table 1 are highly correlated and that the resulting n_i and c_i values obtained by regression analysis of the columns are not independent quantities. The aim of this analysis was to determine how this correlation changes with time in the dynamics of competitive running. Equations (2), (3) and (4) are the result.

Energy expended in running

There are a few published studies on aerobic and anaerobic energy expenditure, which, with proper analysis, show that the combined aerobic and anaerobic energies of a runner do not give a smooth energy–time curve in the first 5 min. We will try to show that the power law time-offsets probably reflect these energy fluctuations modulated by the efficiency of their utilization in running.

Ward-Smith (1985) presented an analysis of the energetics of running based on the first law of thermodynamics. He derived the aerobic and anaerobic energies expended by an athlete from world record running data. In a similar study, Peronnet and Thibault (1989) used an 'empirical model relating human running performance to some characteristics of metabolic energy yielding processes'. In 1954, estimates were collected from several sources of the power developed in running up a hill. In each case, the average horsepower

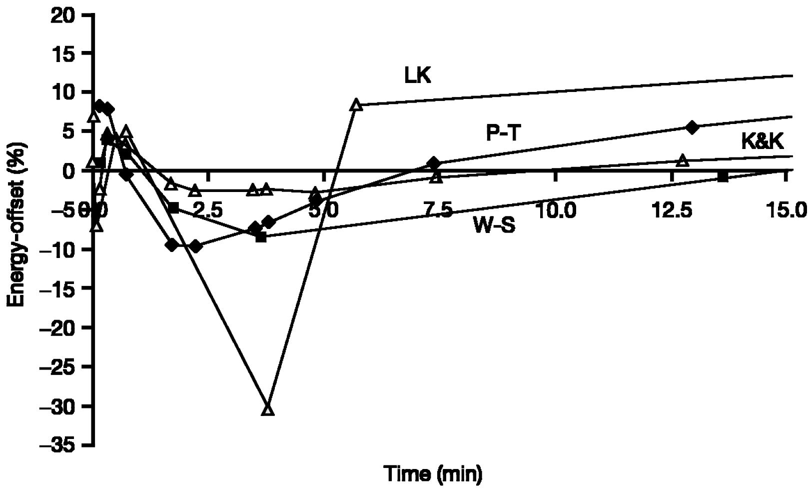

Figure 5 Energy-offset derived from various sources (LK = Katz, 1954; W-S = Ward-Smith, 1985; P-T = Peronnet and Thibault, 1989; K&K = present study).

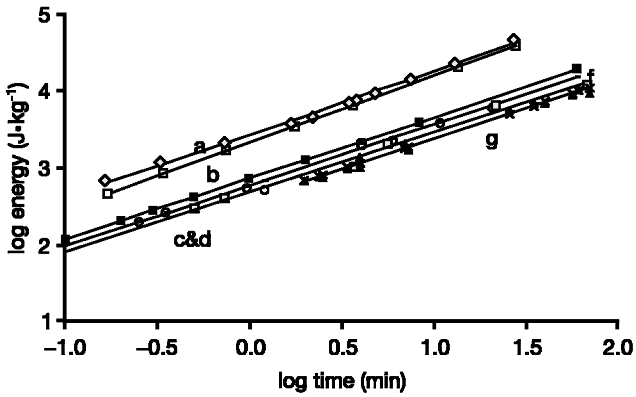

Figure 6 Energy expenditure versus time. Sources of data: a, Peronnet and Thibault (1989); b, Ward-Smith (1985); c, Harris (1970); d, Pons and Vaughan (1989); e, Wilkie (1960); f, Johnson (1991); g, Katz (1954).

generated was found from the weight of the subject, the time and the height of the hill climbed (Katz, 1954). The data covered a run up a 7.5-foot hill in 1.2 s by an athlete to an ascent of a 23,000 foot mountain in 22 h.

For reasons previously outlined, power laws were also used to analyse these three sets of data. Calculations of the energy-offset (fine structure) values from their respective regression lines are shown in Fig. 5. One would not expect the three graphs to coincide, but they all show the same general structure. In particular, we would like to draw attention to how well the general shapes of these curves reflect the shapes of the time-offset curves in Fig. 3 (except for an inversion about the x-axis). It is our contention that this variation in total energy as a function of time is reflected in similar variations in the time–distance running records.

We realize that, in many respects, this amounts to a circular argument in terms of the calculations of Ward-Smith (1985) and Peronnet and Thibault (1989). The running data, which include time-offsets, were used by these authors to calculate the energy, which was then used to show that there is a correspondence between the time-offsets and energy-offsets. However, for the data of Katz (1954), energy was calculated from the time it took to climb a hill. Although the accuracy of the hill-climbing data leaves much to be desired, and hill climbing is quite different from running on a horizontal surface, the general shape of the energy curve is the same as for the other calculations.

To support the above assumption, it is necessary to convert the time-offsets of Table 2 presented in Fig. 3 into equivalent energy-offsets that the athlete uses in running. Several laboratory experiments have measured the energy a male subject can develop at maximum sustained exertion for a given time (Wilkie, 1960; Harris, 1970; Pons and Vaughan, 1989; Johnson, 1991). As running requires maximum exertion, these energy values are directly comparable to the three energies discussed above. All seven sets of data are shown as log-log plots in Fig. 6 (note that the data of Harris, 1970, and of Pons and Vaughan, 1989, fall on the same line).

The lines in Fig. 6 fall into two groups. First, the upper two lines are based on the calculations of Ward-Smith (1985) and Peronnet and Thibaut (1989) and represent outputs of energy by highly trained runners. The athletes represented by the lower graphs were not as intensely trained and had lower energy outputs. The bottom line is from the scattered results of various individuals running up hills (Katz, 1954). The ratio of the averages of these two sets of data is 4.5 in energy output. All the lines have similar slopes. The data of Ward-Smith (1985) gave the largest slope of 0.858, whereas the overall average slope was 0.798 ± 0.020. The data in this figure may be represented by a power law:

$$E(T) = bT^{m} \tag{9}$$

where E is the total energy developed by the subject in $J \cdot kg^{-1}$ during maximum exertion time, T. It is assumed there is a one-to-one correspondence between the time T of the regression lines in running events and the T values of equation (7) (at least to the first order). Both involve energy production during periods of maximum exertion and, in both cases, the data seem to have the same relative distribution about the regression time lines (Figs 3 and 5), except that one is a reflection of the other.

Let $\Delta T = t - T$ and let the energy developed by an athlete in time t be:

$$E(t) = E(T + \Delta T) = E(T) + \Delta E(T)$$

$$= E(T)[1 + \Delta E(T)/E(T)] \tag{10}$$

where $\Delta E(T)$ is the energy corresponding to the deviation ΔT (Fig. 3). From equation (9) we have:

$$\Delta E(T)/E(T) = -m\Delta T/T \tag{11}$$

A negative sign was introduced into this equation because, all things being equal, an athlete who has more energy available presumably can maintain his or her top speed longer; thus a positive $\Delta E(T)$ would result in a negative ΔT.

Introducing equation (11) into equation (10) gives:

$$E(t)/E(T) = (1 - m\Delta T/T) \tag{12}$$

where m is the average value derived above. From equation (5) we have:

$$\Delta T/T = (s_{ja}/100)(t/T) \tag{13}$$

where s_{ja} is the average percent time-offset corresponding to the time-offsets listed in Table 2. Thus, the energy-offset $\Delta E(T)$ is given by:

$$\Delta E(T) = 100[E(t) - E(T)]/E(t)$$

$$= -ms_{ja}(t/T)[1 - 1/(1 - ms_{ja}/100T)] \tag{14}$$

The average power law offsets, s_{ja}, from the second last column of Table 2, are converted to the corresponding energy-offsets and are shown in Fig. 5.

In summary, there is good agreement in the general shape of the four data sets, although this is not the case for the magnitudes of the offsets. This is not surprising, as the data of Ward-Smith (1985) and of Peronnet and Thibault (1989) relate to the total rate of energy production in an athlete's body, while the values calculated from the offsets are these energies multiplied by the efficiency by which they are converted into running.

Further analysis of the running data

The T_{ij}* values, calculated from equations (4) and (7), define a smooth surface in three-dimensional space covering event times minus the offsets in the 100-m to 10-km events over a 70-year period. To reconstruct the best estimate of event times (t_{ijc}), one must add the average offset $(t - T)_{ja}$ to all events j (alternatively one could use the average s_{ja} values). The values of T_{ij}* and the calculated event times t_{ijc} are plotted for four representative events (100 m, 1 mile, 1 km and 10 km). The progressive world record times are shown in Fig. 7. In addition to the rather small errors in measurement of event time and distance, a few other sources of error exist in calculating the values of t_{ijc}: these are the errors in the T_{ij}* regression values, in the equations fitted to the slope n_i and parameter c_i values, and in the average values of $(t - T)_{ja}$, the total being estimated at $\pm 0.5\%$. Figure 7 shows how t_{ijc} (solid lines) and T_{ij}* (dashed-and-dotted lines) changed over the years for the different events. Values of $t_{ij} \pm 0.5\%$ are represented by the dotted lines.

Actual world record event times from Table 1 are shown in Fig. 7 for comparison, but are plotted to the nearest fraction of the year during which it was made. Some values from the recent Olympic Games in Atlanta and a few records established in the years intermediate to those of Table 1 are included. It is interesting to note that the predicted rate at which event times change at present is by about −0.0025 s per year for the 100-m event and −2.6 s per year for the 10-km event.

Conclusion

Although power law analysis of running competitions has been used since the early 1900s, only recent improvements in our understanding of the behaviour of non-linear dynamic systems has made the power law a natural choice for such analyses. Furthermore, we believe that deviations from a power law might indicate an interaction with another part of the system. In the

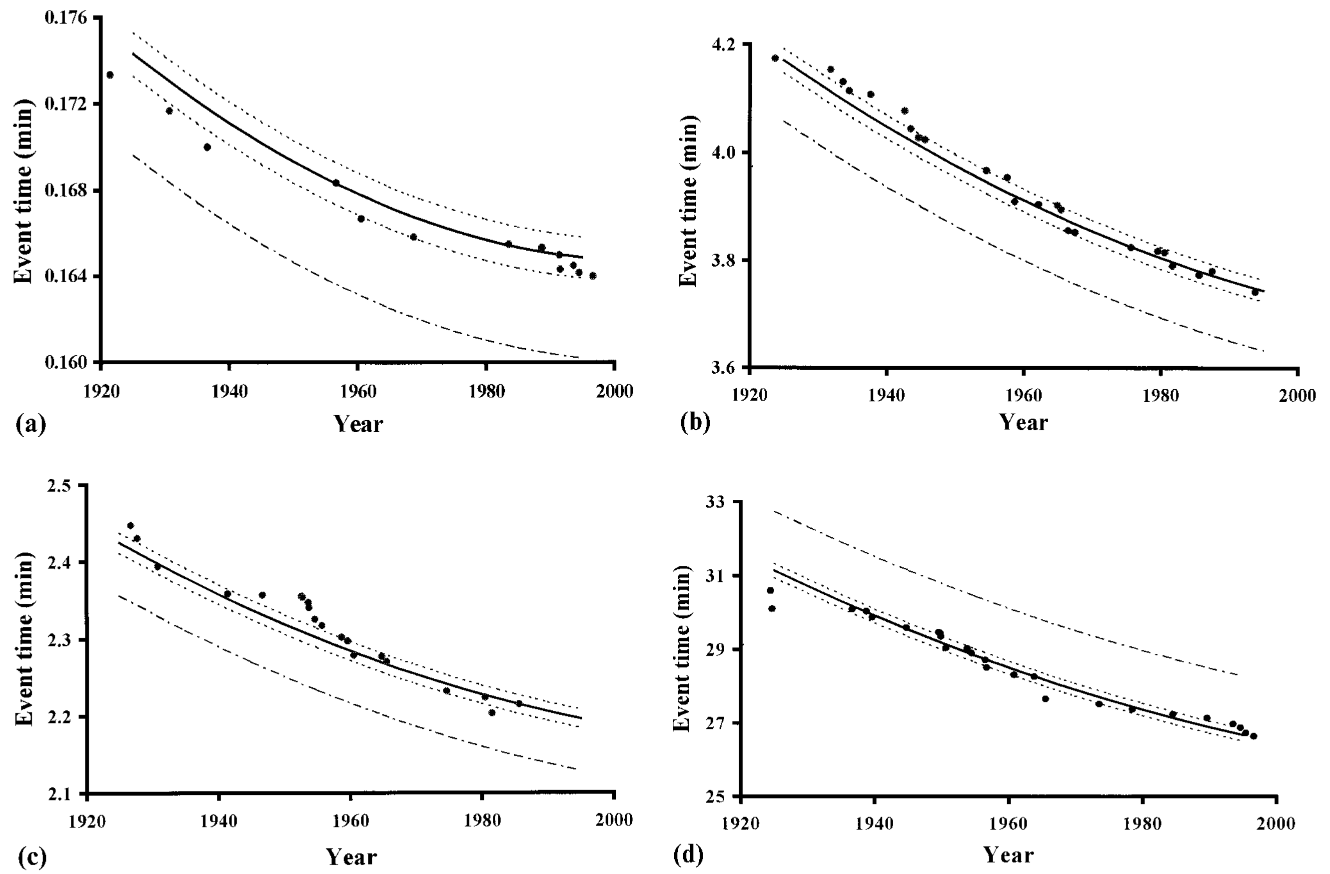

Figure 7 Event time versus year for: (a) the men's 100-m run; (b) the men's one-mile run; (c) the men's 1-km run; (d) the men's 10-km run.

present study, we used world record running data because they are the most accurate and historically most extensive that we could find.

As noted previously, although time and distance measurements have been made with great accuracy, records are bettered at irregular intervals and by irregular amounts so that it is difficult to establish trend lines in individual events. However, with power law analysis, these individual variations are smoothed out and the basic dynamics of competitive running are revealed.

The presence of offsets disclosed by this analysis reveal how the aerobic and anaerobic energy components of the athlete's body multiplied by utilization efficiency in running affects event times. When variations in event times are corrected for this effect, the degree to which the ideal power law model applies is excellent. The precision with which event times in any year, corrected for energy variations and utilization efficiency, can be represented by this simple equation, signal a very important property of human locomotion – super athletes run as if they had identical bodies, ignoring energy considerations. In other words, a *single* super athlete trained to run all these events could in principle produce all the record times.

The time-offsets of individual events did not show any consistent trend over the 70-year period of this analysis; some increased over this period whereas others decreased, and the mean change for all 11 events was −6.5% per event between 1925 and 1995. We conclude that, on average, the aerobic plus anaerobic energies and the efficiency of utilization in running have remained relatively constant.

Hill (1926, p. 98) made the following observation: 'Some of the most consistent physiological data available are contained, not in books on physiology, not even in books on medicine, but in the world's records for running different horizontal distances'. This analysis shows that it may indeed be possible to extract physiological information from world running and swimming records, as the human body's bioenergetics system – the aerobic and anaerobic energy cycle – modulates the ideal power law model to produce a measurable and consistent offset pattern. With this modulation removed, world record runners behave in a similar way.

Acknowledgements

We would like to thank Professor Ken Paulson, Department of Physics, University of Saskatchewan, for his advice on statistical matters.

References

Blest, D.C. (1996). Lower bounds for athletic performance. *The Statistician*, **45**, 243–253.

Buchanan, M. (1997). One law to rule them all. *New Scientist*, **2107**, 30–35.

Elbert, T., Ray, W.J., Kowalik, Z.J., Skinner, J.E., Graf, K.E. and Birbaumer, N. (1994). Chaos and physiology: Deterministic chaos in excitable cell assemblies. *Physiological Reviews*, **74**, 1–47.

Gleick, J. (1987). *Chaos, the Making of a New Science*. New York: Viking Penguin.

Goldberger, A.L., Rigney, D.R. and West, B.J. (1990). Chaos and fractals in human physiology. *Scientific American*, **262**, 43–49.

Harris, J.Y. (1970). Maximal human power output by suitable selection of motion cycle and load. *Human Factors*, **12**, 315–329.

Henry, F.M. (1955). Prediction of world records in running sixty yards to twenty-six miles. *Research Quarterly*, **26**, 147–158.

Hill, A.V. (1926). *Muscular Activity*. Baltimore, MD: Williams & Wilkins.

International Amateur Athletic Federation (1972). *Progressive World Record Lists 1913–1970*. London: IAAF.

Johnson, A.T. (1991). *Biomechanics and Exercise Physiology*. New York: John Wiley.

Katz, L. (1954). Letter to the editor. *American Scientist*, **42**, 392 and 532.

Katz, L. and Katz, J.S. (1994). Fractal (power law) analysis of athletic performance. *Sport Medicine, Training and Rehabilitation*, **5**, 95–105.

Keller, J.B. (1973). A theory of competitive running. *Physics Today*, September, pp. 43–47.

Keller, J.B. (1974). Optimal velocity in a race. *American Mathematics Monthly*, **81**, 474–480.

Kennelly, A.E. (1906). An approximate law of fatigue in the speeds of racing animals. *Proceedings of the American Academy of Arts and Sciences*, **42**, 273–331.

Lietzke, M.H. (1954). An analytical study of world and Olympic racing records. *Science*, **119**, 333–336.

Lodge, M. (1981). *Magnitude Scaling*. London: Sage.

Mandelbrot, B.B. (1977). *The Fractal Geometry of Nature*. New York: W.H. Freeman.

Oliver, J.L., Barnaola-Galvan, P., Guerrero-Garcia, J. and Roman-Roldan, R. (1993). Entropic profiles of DNA sequences through chaos-game-derived images. *Journal of Theoretical Biology*, **160**, 457–470.

Peronnet, F. and Thibault, G. (1989). Mathematical analysis of running performance and world running records. *Journal of Applied Physiology*, **67**, 453–465.

Pons, D.J. and Vaughan, C.L. (1989). Mechanics of cycling. In *Biomechanics of Sport* (edited by C.L. Vaughan), pp. 305–306. Boca Raton, FL: CRC Press.

Riegel, P.S. (1981). Athletic records and human endurance. *American Scientist*, **69**, 985–990.

Schroeder, M. (1991). *Fractals, Chaos and Power Laws*. New York: W.H. Freeman.

Sveshnikov, A.A. (1968). *Problems in Probability Theory, Mathematical Statistics and Theory of Random Functions*. London: W.B. Saunders.

Ward-Smith, A.J. (1985). A mathematical theory of running, based on the first law of thermodynamics, and its application to the performance of world class athletes. *Journal of Biomechanics*, **18**, 337–349.

Wilkie, D.R. (1960). Man as a source of mechanical power. *Ergonomics*, **3**, 1–8.