Where Bolt and Bekele meet: the analytical basis of running performance estimates

Wim Westera

Abstract

This paper proposes a self-contained analytical model for the prediction of individual running performances. The model uses two personal bests for calibration, and then allows the prediction of the athlete's personal bests for any other distance. It is based on a simple, first order estimate of the way lap time increments with total distance. Also, the model accounts for delays that occur during start-up. It therefore covers a wide range of events including endurance and sprinting distances. The model is validated with empirical data of a variety of world class and sub top athletes. Outcomes display valid and reliable predictions with inaccuracies typically around 1%. It greatly outperforms existing models (typically 3% or higher). Importantly, the model is transparent, since it is based on theoretical principles rather than arbitrariness and negotiation. It is self-contained, easy to use and affordable, because it does not require any physiological or biomechanical tests to be carried out first. Also, the model displays a universal validity, as the results suggest its applicability for any speed and distance related sports event, including running, speed skating and swimming.

Keywords

Running, performance, sprinting, endurance, prediction

Introduction

Human performance is strongly dependent on the time span that the performance has to be delivered. Intense efforts that require high degrees of physical or mental power can only be made for a short period of time. Reversely, for tasks that require prolonged efforts, the internal economy of endurance reinforces effort levels well below maximum capacity.
In athletics it is well-established that average running speeds go down at longer distances. Indeed, the average speed of top athletes in a 10,000 m event is some 20% below the average speed in an 800 m race. This decay effect is obviously linked with the fundamental restrictions to human power and energy resources, as well as the underlying physiological processes and biomechanical constraints. Personal characteristics of the athlete (i.e. length, weight, muscle fibre type, lung capacity, anaerobic power) make a big difference in the amount of decay: sprinters tend to display severe decays of speed at longer distances, whereas for long distance runners the decays are less pronounced.

Various researchers have been working on the relationships between human performance and the duration of the efforts, in order to make predictions for individual athletes and devise personalised training schedules for these. For many decades it is known that endurance performance is directly related with maximal aerobic power (Costill, 1973). Measurement of maximum aerobic running speed, or speed at maximal oxygen uptake can be used for predicting performances across a wide range of running distances (Daniels &Daniels, 1992). At shorter distances, however, predictions are highly unreliable because unknown dependences of performance and maximal aerobic power. Bundle et al. (2003) proposed a model that combined the physiological limits of anaerobic and aerobic power for predicting performances in both the sprinting and mid-range distances (ranging from a few seconds to a few minutes). They suggest a simple exponential relationship between speed and run duration, and incorporate the different time scales of available anaerobic and aerobic power in relation to time. However, accuracies reported are poor: the predictions deviate on average well above 3 % from realised performances. An additional disadvantage is that accurate tests for assessing the athlete's speeds at maximal aerobic and anaerobic power are required. Likewise, progressive tables based on statistical processing of large numbers of performance data are known to be inaccurate and unreliable. The International Association of Athletics Federations (IAAF) provides and uses such scoring tables for several purposes: to evaluate scores in team competitions over multiple events, to determine result scores of performances for all event world rankings, or to produce national, school or club rankings (Spiriev, 2008). The official IAAF-committee that is watching over the validity of the table, regularly needs to make modifications to the scoring tables because of apparent irregularities in the relationship between results and assigned points. But even in the latest version of the scoring tables inconsistencies and suspect data can easily be tracked. For instance, the men's world records on 100 m, 200 m, 400 m, 800 m, 1,500 m, 3,000 m, 5,000 m, 10,000 m, and marathon, respectively, yield the following array of scores: 1374, 1356, 1300, 1321, 1302, 1299, 1294, 1295, 1293. Unfortunately, these scores fail to be equal, or even be close to each other; instead the scores suggest a unjustified bias toward shorter distances. Similarly, empirical scoring tables for decathlon and heptathlon are criticised for their inaccuracies, producing questionable rankings and records (Westera, 2006). Historical bias and frequent changes seem to demonstrate the arbitrariness and opportunism of performance alignments (Trkal, 2003). Harder (2001) takes a different approach by considering population fractions achieving a certain performance level. Calibration of the fractions between different events enables statistical mapping for inter-sport comparisons. Although Harder's tables deviate from the IAAF-tables, they correlate very well with these, and thus display the same inaccuracies. These statistical approaches have two things in common. First, they reflect a population-based average representing a mixture of many different human features and conditions, which may severely affect their applicability for individuals. Second, they are phenomenological in kind and do not rely on an underlying theory that would improve our understanding of the mechanisms involved.

Rather than an empirical or a statistical model, this paper presents an analytical model that allows accurate predictions without requiring any physiological or biomechanical test data. Instead, the model uses two personal bests for calibration, and then it allows predicting an athlete's hypothetical personal bests for any other distance. The model is based on a first order estimate of the way lap time (which is equivalent with reciprocal speed) increments with total distance. Also, the model accounts for delays that occur during start-up. This way the model covers the entire range including endurance and sprinting distances.
As first step, the basic model will be explained. Second, the basic model is validated using a variety of empirical data. Third, the basic model will be extended with a mechanism that accounts for start-up bias. Finally, the paper will present various practical applications.

Analytical model

Let t be the running time of an athlete required for a total distance s. What we are actually looking for is a mathematical formula which expresses the relationship between running time t and distance s. As an intermediate step we introduce the average lap time L (this corresponds with reciprocal running speed), which is given by:

$$L = \frac{t}{s} \quad . \tag{1}$$

The starting point of the model is that an infinitesimal increment *ds* of the running distance *s* would produce an increase *dL* of the average lap time, while accounting for the effect that *dL* gradually fades out at longer distances.
As a first order estimate the lap time increment *dL* is assumed to be reciprocally proportional to the distance s, yielding

$$dL = \alpha \cdot \frac{ds}{s} \quad , \tag{2}$$

where α is a constant.
Integration over *s* gives a simple logarithmic expression

$$L = \alpha \cdot \ln(\frac{s}{\beta}) \quad , \tag{3}$$

where β is a constant. The first order approximation reflected in equations (1) - (3) is not only theoretically grounded. Empirical evidence of its appropriateness can be found by using some existing data. This is done in figure 1 which displays a plot of lap time *L* against total distance *s* for men's track running world records. The data suggest the validity of a linear relationship in accordance with equation (3), even though data of different athletes were used.

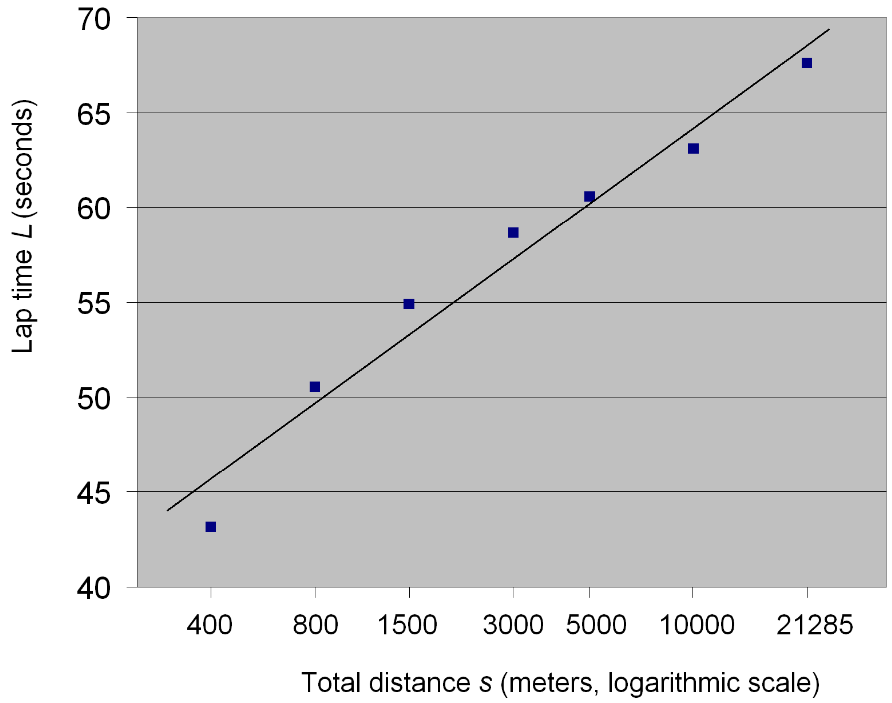

Figure 1. Growth of lap times with distance for men's track running world records.

Substituting equation (1) into equation (3) produces running time *t* as a function of total distance *s*:

$$t(s) = \alpha \cdot s \cdot \ln(\frac{s}{\beta}) \quad . \tag{4}$$

The unknown coefficients α and β can be resolved from equation (4) by applying a twofold calibration procedure. It requires the availability of two data pairs: [s_1, t_1] and [s_2, t_2], respectively. This would mean using two personal records on different distances. The outcomes of this procedure are:

$$\alpha = \frac{\frac{t_1}{s_1} - \frac{t_2}{s_2}}{\ln(\frac{s_1}{s_2})} \tag{5}$$

and

$$\ln(\beta) = \ln(s_1) - \frac{\ln(\frac{s_1}{s_2})}{(1 - \frac{t_2 \cdot s_1}{t_1 \cdot s_2})} \quad . \tag{6}$$

Inserting the values of α and $\ln(\beta)$ into equation (4) enables us to produce prognoses of running times for any distance. The two distances used for the calibration should produce a sufficiently wide interval: Using someone's personal records at 800 m and 3,000 m would probably make a sound basis for forecasting the person's achievable 1,500 m performance (reliable interpolation), but are likely not to produce reliable forecasts of marathon performance (unreliable extrapolation).

Model validation

For assessing the empirical validity of the method covered by equation (4), the following procedure is used:

1. Performance data of a 4 different groups of athletes are user:
   -World class male athletes
   -World class female athletes
   -Committed male sub top athletes
   -Committed female sub top athletes
2. An additional requirement is that each of the selected athletes cover a minimum of three distances. This is because two personal records are needed for calibration cf. equations (5) and (6), and the third one for checking the prediction based on these. This prediction is calculated via equation (4) and the outcome can easily be compared with the real personal record.
3. Only track events will be accounted for, because road events may be subject to inaccuracies of the traversed distance due to uncontrolled bends and corners.
4. Sprinting distances will be omitted for the moment, because of the disturbing effects of reaction time delays and the time needed for acceleration from the starting blocks. Later on we will go into corrections for these disturbances.

5. The validation is preferably based on interpolation rather than extrapolation, because of better accuracies.
6. For each personal record prediction of an athlete the outcome is compared with the athlete's real personal record. The accuracy of the prediction is expressed as a percentage of the deviation relative to the existing record.
7. As an alternative yardstick a simple linear interpolation procedure is used. Linear interpolation (and extrapolation) uses the following equation as a replacement for equation (4):

$$t = t_1 + \frac{(s - s_1)}{(s_2 - s_1)} \cdot (t_2 - t_1) \quad . \qquad (7)$$

This alternative procedure will be used as a reference for assessing the added value of the logarithmic model of equation (4).

The outcomes of this validation procedure are summarised in the tables below. Table 1 presents the calculations of a sample of 8 male world class athletes. The personal record data were collected from the all time performances lists of IAAF (2009).

Table 1. Logarithmic and linear performance predictions for male world class athletes.

Athlete	Distance (m)	Personal best (IAAF)	Logarithmic model		Linear interpolation	
			Predicted personal best	Deviation	Predicted personal best	Deviation
H.E.G.	1500	3:26.00	3:26.00		3:26.00	
	3000	*7:23.09*	*7:20.87*	*0.005*	*7:27.82*	*-0.011*
	5000	12:50.24	12:50.24		12:50.24	
B.L.	1500	3:26.34	3:26.34		3:26.34	
	3000	*7:33.15*	*7:24.26*	*0.020*	*7:31.86*	*0.003*
	5000	12:59.22	12:59.22		12:59.22	
K.B.	3000	7:25.79	7:25.79		7:25.79	
	5000	*12:37.35*	*12:42.41*	*-0.007*	*12:49.14*	*-0.016*
	10000	26:17.53	26:17.53		26:17.53	
H.G.	1500	3:33.73	3:33.73		3:33.73	
	3000	*7:25.09*	*7:24.77*	*0.001*	*7:35.32*	*-0.023*
	5000	*12:39.36*	*12:42.53*	*-0.004*	*12:57.44*	*-0.024*
	10000	26:22.75	26:22.75		26:22.75	
R.R.	800	1:44.05	1:44.05		1:44.05	
	1500	*3:29.14*	*3:32.61*	*-0.017*	*3:38.53*	*-0.045*
	3000	7:43.85	7:43.85		7:43.85	
D.K.	1500	3:29.46	3:29.46		3:29.46	
	3000	*7:20.67*	*7:20.18*	*0.001*	*7:25.29*	*-0.010*
	5000	12:39.74	12:39.74		12:39.74	
K.M.	3000	7:45.44	7:45.44		7:45.44	
	5000	*13:13.06*	*13:15.85*	*-0.004*	*13:22.83*	*-0.012*
	10000	27:26.29	27:26.29		27:26.29	
J.H.	3000	7:44.40	7:44.40		7:44.40	
	5000	*13:21.90*	*13:18.03*	*0.005*	*13:26.36*	*-0.006*
	10000	27:41.25	27:41.25		27:41.25	
K.L.	1500	3:38.83	3:38.83		3:38.83	
	3000	*7:52.50*	*7:44.61*	*0.017*	*8:01.04*	*-0.018*
	5000	*13:36.10*	*13:27.44*	*0.011*	*13:50.66*	*-0.018*
	10000	28:24.70	28:24.70		28:24.70	

Overall deviation				0.008		0.017

World class athletes provide an important sample because top athletes are usually well-prepared and perform near the limits of human capability. For each of the athletes 3 or occasionally 4 official personal bests are listed in the third column. Outer distances (shortest and longest) have been used for the calibration, viz. the calculation of α and β according to equations (5) and (6). These parameters were used to make predictions for the intermediate events, both by the logarithmic model and the linear model, covered by equation (4) and equation (7), respectively. The predictions of the logarithmic model are quite close to the official personal bests. The average deviation (minus signs neglected) is only 0.8% (0.008, cf. bottom row of table 1). The average deviation of the linear model is 1.7%. Likewise, table 2 displays the results for a sample of 6 female world class athletes, excelling at multiple distances.

Table 2. Logarithmic and linear performance predictions for female world class athletes.

Athlete	Distance (m)	Personal best (IAAF)	Logarithmic model		Linear interpolation	
			Predicted personal best	Deviation	Predicted personal best	Deviation
P.R.	3000	8:22.20	8:22.20		8:22.20	
	5000	*14:29.11*	*14:23.96*	*0.006*	*14:33.31*	*-0.005*
	10000	30:01.09	30:01.09		30:01.09	
V.C.	3000	8:28.66	8:28.66		8:28.66	
	5000	*14:12.88*	*14:45.20*	*-0.038*	*14:58.19*	*-0.053*
	10000	31:12.00	31:12.00		31:12.00	
M.D.	3000	8:24.51	8:24.51		8:24.51	
	5000	*14:12.88*	*14:25.78*	*-0.015*	*14:34.42*	*-0.025*
	10000	29:59.20	29:59.20		29:59.20	
T.D.	3000	8:29.55	8:29.55		8:29.55	
	5000	*14:11.15*	*14:29.65*	*-0.022*	*14:36.72*	*-0.030*
	10000	29:54.66	29:54.66		29:54.66	
M.Y.J.	800	1:57.80	1:57.80		1:57.80	
	1500	*3:56.18*	*3:56.87*	*-0.003*	*4:06.78*	*-0.045*
	3000	*8:28.87*	*8:29.01*	*0.000*	*8:43.17*	*-0.028*
	5000	14:51.68	14:51.68		14:51.68	
T.T.	800	2:02.49	2:02.49		2:02.49	
	1500	*3:56.91*	*4:01.36*	*-0.019*	*4:08.61*	*-0.049*
	3000	*8:25.56*	*8:28.52*	*-0.006*	*8:38.87*	*-0.026*
	5000	14:39.22	14:39.22		14:39.22	
Overall deviation				0.014		0.033

The calculations for female top athletes show similar results as those for top class males, be it that deviations are somewhat higher (1.4%). Again the logarithmic model clearly outperforms the linear model. A general comment on the sample of top athletes would be that there are only very few athletes that perform at world class level at multiple distances. Top athletes tend to specialise in one or two events. Although they might occasionally participate in a third or even fourth event, this is often not taken as serious as their specialism, nor is there training optimised for it. For this reason, the validation procedure is extended to a male group and a female group of amateur runners, under the condition that these amateur runners are well-trained and highly ambitious sportsmen rather than leisure joggers. Table 3 lists the calculation for a sample of male sub top runners.

Table 3. Logarithmic and linear performance predictions for male sub-top athletes.

Athlete	Distance (m)	Personal best	Logarithmic model		Linear interpolation	
			Predicted personal best	Deviation	Predicted personal best	Deviation
E.G.	800	1:56.82	1:56.82		1:56.82	
	1500	*3:54.80*	*3:53.83*	*0.004*	*4:09.20*	*-0.061*
	3000	*8:16.22*	*8:20.30*	*-0.008*	*8:52.86*	*-0.074*
	5000	*14:28.61*	*14:33.92*	*-0.006*	*15:11.08*	*-0.049*
	10000	30:56.62	30:56.62		30:56.62	
M.V.	800	1:57.20	1:57.20		1:57.20	
	1500	*3:53.90*	*3:54.12*	*-0.001*	*4:09.03*	*-0.065*
	3000	*8:19.50*	*8:19.91*	*-0.001*	*8:51.52*	*-0.064*
	5000	*14:22.28*	*14:32.10*	*-0.011*	*15:08.17*	*-0.053*
	10000	30:49.80	30:49.80		30:49.80	
J.V.	800	1:58.30	1:58.30		1:58.30	
	1500	*3:59.51*	*3:54.80*	*0.020*	*4:08.29*	*-0.037*
	3000	*8:25.60*	*8:18.25*	*0.015*	*8:46.83*	*-0.042*
	5000	*14:28.50*	*14:25.61*	*0.003*	*14:58.22*	*-0.034*
	10000	30:26.70	30:26.70		30:26.70	
H.K.	1500	4:00.51	4:00.51		4:00.51	
	3000	*8:48.02*	*8:32.92*	*0.029*	*8:52.37*	*-0.008*
	5000	*15:05.50*	*14:54.04*	*0.013*	*15:21.52*	*-0.018*
	10000	31:34.40	31:34.40		31:34.40	
H.D.	1500	4:00.70	4:00.70		4:00.70	
	3000	*8:26.70*	*8:27.96*	*-0.002*	*8:44.16*	*-0.034*
	5000	*14:45.50*	*14:39.23*	*0.007*	*15:02.12*	*-0.019*
	10000	30:47:00	30:47:00		30:47:00	
S.H.	800	1:46.40	1:46.40		1:46.40	
	1500	*3:46.10*	*3:45.62*	*0.002*	*4:01.80*	*-0.069*
	3000	*8:25.90*	*8:28.83*	*-0.006*	*8:51.94*	*-0.051*
	5000	15:18.80	15:18.80		15:18.80	
Overall deviation				0.009		0.046

The personal record data in table 3 originate from the all time best lists of a local athletics association (Achilles-top, 2009). Average deviation of the logarithmic prediction is 0.9%, against 4.6% for the linear prediction. Similar results (0.9% versus 5.0 %) hold for sub top females, cf. table 4.

Table 4. Logarithmic and linear performance predictions for female sub-top athletes

Athlete	Distance (m)	Personal best	Logarithmic model		Linear interpolation	
			Predicted personal best	Deviation	Predicted personal best	Deviation
J.W.	800	2:08.49	2:08.49		2:08.49	
	1500	*4:28.71*	*4:27.74*	*0.004*	*4:44.36*	*-0.058*
	3000	*9:44.00*	*9:54.63*	*-0.018*	*10:18.36*	*-0.059*
	5000	17:43.70	17:43.70		17:43.70	
J.B.	800	2:16.00	2:16.00		2:16.00	
	1500	*4:37.47*	*4:29.50*	*0.029*	*4:44.55*	*-0.026*
	3000	*9:38.50*	*9:30.98*	*0.013*	*10:02.88*	*-0.042*
	5000	*16:38.60*	*16:30.90*	*0.008*	*17:7.31*	*-0.029*
	10000	34:48.40	34:48.40		34:48.40	
C.S.	800	2:13.50	2:13.50		2:13.50	
	1500	*4:30.60*	*4:29.82*	*0.003*	*4:50.08*	*-0.072*
	3000	*9:43.00*	*9:42.68*	*0.001*	*10:25.61*	*-0.073*

	10000	36:31.40	36:31.40		36:31.40	
J.M.	800	2:17.80	2:17.80		2:17.80	
	1500	*4:43.40*	*4:43.81*	*-0.001*	*4:59.57*	*-0.057*
	3000	*10:25.60*	*10:23.70*	*0.003*	*10:46.21*	*-0.033*
	5000	18:28.40	18:28.40		18:28.40	
Overall deviation				0.009		0.050

From the above it can be concluded that the logarithmic model produces far more accurate predictions than the linear model. The overall average deviation of the presented cases is found to be 0,9% for the logarithmic model, against 3.7% for the linear model. Compared with existing models and tables, the logarithmic model produces far better accuracies.

Compensating for start-up bias

For the purpose of validation, only distances of 800 m and longer were used to avoid the effect of time delays incurred during the early phase of the event, when the athlete has to accelerate from standstill to cruise speed. For shorter distances, however, start-up delays will strongly confound the outcomes of predictions. Various researches carried out biomechanical studies of the running start-up process (DiPrampero, 2005), but these are mostly concerned with the dynamics of body angle, required metabolic power and the techniques for start-up optimisation. For evaluating the impact of start-up empirical data of sprinting events should be analysed. Because of ongoing specialisation, however, only very few world class sprinters excel in three sprinting disciplines (100-200-400). Yet, some exceptions could be found and a number of these are presented in table 5 , together with some amateur examples (IAAF, 2009; Achilles-Top, 2009).

Table 5 Logarithmic and linear performance predictions for sprinters.

Athlete	Distance (m)	Personal best	Logarithmic model		Linear interpolation	
			Predicted personal best	Deviation	Predicted personal best	Deviation
I.S. (female)	100	0:11.10	0:11.10		0:11.10	
	200	*0:22.21*	*0:23.42*	*-0.055*	*0:23.83*	*-0.073*
	400	0:49.29	0:49.29		0:49.29	
M.K. (female)	100	0:10.83	0:10.83		0:10.83	
	200	*0:21.71*	*0:22.73*	*-0.047*	*0:23.09*	*-0.063*
	400	0:47.60	0:47.60		0:47.60	
M.J.P (female)	100	0:10.96	0:10.96		0:10.96	
	200	*0:21.99*	*0:23.02*	*-0.047*	*0:23.39*	*-0.064*
	400	0:48.25	0:48.25		0:48.25	
H.M. (male)	100	0:10.30	0:10.30		0:10.30	
	200	*0:20.40*	*0:21.78*	*-0.067*	*0:22.17*	*-0.087*
	400	0:45.90	0:45.90		0:45.90	
T.S. (male)	100	0:10.10	0:10.10		0:10.10	
	200	*0:19.83*	*0:21.23*	*-0.070*	*0:21.57*	*-0.088*
	400	0:44.50	0:44.50		0:44.50	
E.M. (male)	100	0:10.96	0:10.96		0:10.96	
	200	*0:22.30*	*0:23.11*	*-0.036*	*0:23.50*	*-0.054*
	400	0:48.58	0:48.58		0:48.58	
L.B. (male)	100	0:10.79	0:10.79		0:10.79	
	200	*0:21.41*	*0:22.57*	*-0.054*	*0:22.89*	*-0.069*
	400	0:47.10	0:47.10		0:47.10	

N.K. (female)	100 *200* 400	0:12.44 *0:26.04* 1:00.94	0:12.44 *0:27.68* 1:00.94	*-0.063*	0:12.44 *0:28.61* 1:00.94	*-0.099*
I.W. (female)	100 *200* 400	0:13.05 *0:26.46* 0:59.02	0:13.05 *0:27.14* 0:59.02	*-0.026*	0:13.05 *0:28.18* 0:59.02	*-0.065*
Overall deviation|				0.054		0.074

The resulting prediction have poor matches with the realised personal bests. On average the predictions deviates 5.4 % for the logarithmic model and 7.4 % for the linear model. These deviations are substantially higher than those of the previous tables with endurance athletes. Also, most deviations are negative which is in accordance with the neglect of acceleration losses. For preserving the predictive quality of our model at lower distances this start-up bias should be corrected for. Estimates for such correction can be given by analysing the sprinting start-up process in more detail.

For sprinting distances the effects of acceleration at start-up cannot be ignored. In contrast with longer distances, time losses during the acceleration phase of sprinting cause the average speed $<v>$ of the event to drop well below the athlete's cruise speed v_c, which is the steady state speed achieved after completing the acceleration. Since any acceleration was ignored in the model of equation (4), the model essentially assumes cruise speed v_c, while average speed $<v>$ has been inserted so far. For longer distances this doesn't make any difference, because v_c and $<v>$ are nearly the same, but for sprinting relative large differences will occur. To eliminate the start-up effects, we should be replacing average speed $<v>$ with cruise speed v_c in equation (4). This means that acceleration losses can be accounted for by replacing the event distance *s* with a extended value *s** that is given by:

$$s^*(t) = s(t) \cdot \frac{v_c}{<v>} = v_c \cdot t \qquad . \qquad (8)$$

Thus, using *s*(t)* in equation (4) rather than *s(t)* would compensate for acceleration losses since it virtually accounts for a flying start and the associated extra meters that have to be traversed within the same time span t.

As a next step we need to establish practical values for cruise speed v_c to be substituted in equation (8). For this we will use some split times available for various world top sprinters. Table 6 lists a sample of cumulative 100 m splits (Lee, 2008; Lee, 2009a):

Table 6 World class 100 m split times.

	Ben Johnson 1998	Carl Lewis 1988	Maurice Green 1999	Maurice Green 2001	Tim Montg. 2002	Asafa Powell 2005	Usain Bolt 2009	Usain Bolt 2009	Average
10	1.83	1.89	1.86	1.83	1.89	1.89	1.85	1.89	1.87
20	2.87	2.96	2.89	2.83	2.92	2.91	2.87	2.88	2.89
30	3.80	3.9	3.81	3.75	3.83	3.83	3.78	3.78	3.81
40	4.66	4.79	4.69	4.64	4.70	4.69	4.65	4.64	4.68
50	5.50	5.65	5.57	5.50	5.54	5.54	5.50	5.47	5.53
60	6.33	6.48	6.40	6.33	6.37	6.39	6.32	6.29	6.36
70	7.17	7.33	7.23	7.16	7.21	7.23	7.14	7.10	7.20
80	8.02	8.18	8.09	8.02	8.05	8.07	7.96	7.92	8.04
90	8.89	9.04	8.94	8.91	8.90	8.92	8.79	8.75	8.89

100	9.79	9.92	9.79	9.82	9.78	9.77	9.69	9.58	9.77
v_c	11.69	11.63	11.71	11.53	11.76	11.78	11.84	12.07	11.75
$v_c \cdot t$	114.4	115.3	114.6	113.2	115.1	115.1	114.8	115.6	114.8

From these splits it can be derived that steady state cruise speed v_c is readily achieved after some 30 meters. The table allows us to determine cruise speeds v_c and s* for each sprinter by using

$$v_c = \frac{100-30}{t(100)-t(30)} \quad . \qquad (9)$$

Resulting values for v_c and s^* ($=v_c.t$) are also given in table 6, as well as their averages. So, according to the column at the extreme right, running a 100 m event in 9.77 s (average speed 100/9.77=10.24 m/s) is technically equivalent with running 114.8 m at cruise speed 11.75 m/s.

For other distances, unfortunately, only few split recordings are available. There is some incidental data of Usain Bolts 2009 world record at 200 m available though, which yield accumulative 50 m splits of 5.60, 9.92, 14.44 and 19,19 seconds, respectively (Lee, 2009b). From these data it follows that for this 200 m event v_c=11.04 m/s. Before using this single outcome for our formulas, a slight correction should be carried out though, since it follows from table 6 that Bolt's 100 m performance deviates substantially from average performance. Accounting for the same (relative) deviation at 200 m reduces the 200 m reference value to v_c= 10,96 m/s. So, now we have two reference data of v_c (v_c= 11.75 m/s for 100 m, and v_c= 10.96 m/s for 200m, respectively) which can be used for compensating acceleration losses. Although our calibration value of v_c at 200 m may not be very accurate because it is only based on a single athlete's performance at one event, it will be used for the time being, since recording new split data is beyond the scope of this study. Note, however, that any modifications of the value v_c would only induce second order effects, since it would concern a correction of a correction.

For being able the apply the correction two things still have to be sorted out. First, having used split times of world class sprinters, raises the question if the outcomes also hold for amateurs. Second, having calibration points for v_c at 100m and 200 m leaves the question how to estimate the values of v_c at other distances.

With respect to the first question, it is important to note that we actually need the value of $v_c.t$. The available 100 m split data show no significant relationship of the product $v_c.t$ with time t (which acts as an indicator for performance). Note that v_c and t counterbalance each other: when cruise speed v_c drops, performance goes down, which means that time t goes up. As a first approximation we assume that the value of s^* ($=v_c.t$) doesn't change significantly with performance t.

With respect to the second question it is important to note that the disturbing effects of start-up will gradually disappear at longer distances. In fact, the difference between cruise speed v_c and average speed $\langle v \rangle$ will gradually approach zero at longer distances. Assuming exponential decay of this correction with distance s yields:

$$v_c - \langle v \rangle = \gamma \cdot e^{-\delta \cdot s} \quad , \qquad (10)$$

where γ and δ are constants. After substitution of the calibration values for v_c at 100 m and 200 m, we obtain γ=21.3 and δ=0.00365. Figure 2 displays the resulting graph for $v_c.t$, representing the extra distance *(s*-s)*, which compensates acceleration losses, versus distance s.

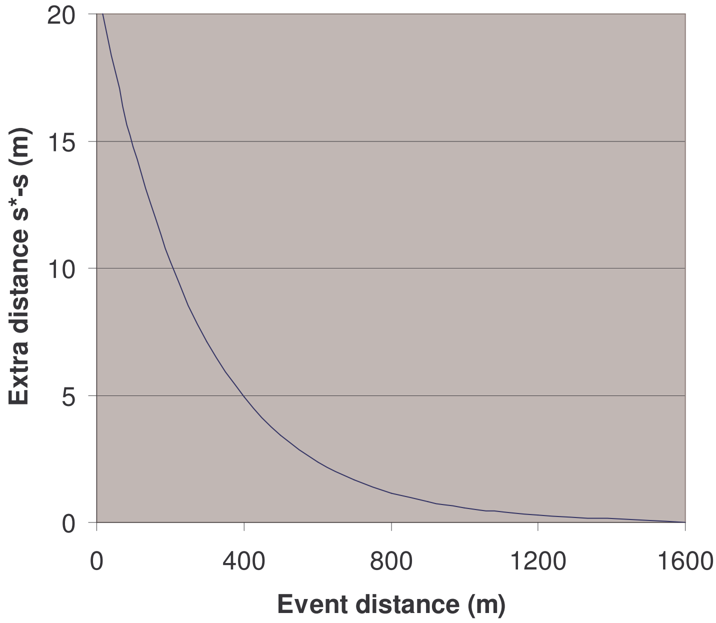

Figure 2. Fictitious extra distance for compensating acceleration losses, against event distance *s*.

Now that we are able to correct for start-up delays with the help of equation (10) or the curve in figure 2, it is interesting to see the effects of this on the sprinting predictions of table 5. Table 7 shows the outcomes of the recalculation of sprinting performances of table 5, while taking into account start-up effects.

Table 7 Performance predictions for sprinters, corrected for start-up losses.

Athlete	Distance (m)	Personal best (IAAF)	Logarithmic model		Linear interpolation	
			Predicted personal best	Deviation	Predicted personal best	Deviation
I.S. (female)	100	0:11.10	0:11.10		0:11.10	
	200	*0:22.21*	*0:22.86*	*-0.029*	*0:23.67*	*-0.066*
	400	0:49.29	0:49.29		0:49.29	
M.K. (female)	100	0:10.83	0:10.83		0:10.83	
	200	*0:21.71*	*0:22.18*	*-0.022*	*0:22.93*	*-0.056*
	400	0:47.60	0:47.60		0:47.60	
M.J.P (female)	100	0:10.96	0:10.96		0:10.96	
	200	*0:21.99*	*0:22.47*	*-0.022*	*0:23.23*	*-0.056*
	400	0:48.25	0:48.25		0:48.25	
H.M. (male)	100	0:10.30	0:10.30		0:10.30	
	200	*0:20.40*	*0:21.25*	*-0.042*	*0:22.01*	*-0.079*
	400	0:45.90	0:45.90		0:45.90	

T.S.	100	0:10.10	0:10.10		0:10.10	
(male)	*200*	*0:19.83*	*0:20.71*	*-0.045*	*0:21.42*	*-0.080*
	400	0:44.50	0:44.50		0:44.50	
E.M.	100	0:10.96	0:10.96		0:10.96	
(male)	*200*	*0:22.30*	*0:22.55*	*-0.011*	*0:23.34*	*-0.047*
	400	0:48.58	0:48.58		0:48.58	
L.B.	100	0:10.79	0:10.79		0:10.79	
(male)	*200*	*0:21.41*	*0:22.02*	*-0.028*	*0:22.74*	*-0.062*
	400	0:47.10	0:47.10		0:47.10	
N.K.	100	0:12.44	0:12.44		0:12.44	
(female)	*200*	*0:26.04*	*0:27.04*	*-0.038*	*0:28.40*	*-0.091*
	400	1:00.94	1:00.94		1:00.94	
I.W.	100	0:13.05	0:13.05		0:13.05	
(female)	*200*	*0:26.46*	*0:27.14*	*-0.026*	*0:28.18*	*-0.065*
	400	0:59.02	0:59.02		0:59.02	
Overall deviation				0.029		0.067

In all cases the deviations between predicted and realised performances go down. The average deviation for the whole sprinters' sample goes down from 5.4 to 2.9% for the logarithmic model, and from 7.4 to 6.7% for the linear model. The logarithmic model still greatly outperforms the linear interpolation model. So, by accounting for start-up losses, the accuracy of the approach has improved substantially.

In sum, the general prediction procedure now reads as follows:

1. Establish two sound personal bests: s_1, t_1, and s_2, t_2.
2. Replace all distances s_1 and s_2 with s_1* and s_2*, respectively by using

   $$s^*(t) = s(t) + \gamma \cdot t \cdot e^{-\delta \cdot s} \quad , \qquad (11)$$

   with γ=21.3 and δ=0.00365 (estimation of these parameters might be improved by extended recording of split times).
3. Calculate the personal coefficients α and β through

   $$\alpha = \frac{\frac{t_1}{s_1^*} - \frac{t_2}{s_2^*}}{\ln(\frac{s_1^*}{s_2^*})} \qquad (12)$$

   and

   $$\ln(\beta) = \ln(s_1^*) - \frac{\ln(\frac{s_1^*}{s_2^*})}{(1 - \frac{t_2 \cdot s_1^*}{t_1 \cdot s_2^*})} \quad . \qquad (13)$$

4. Choose a distance s for predicting time t.
5. Replace distance s with s^*, using equation (11).
6. Calculate predicted time t by using:

   $$t(s^*) = \alpha \cdot s^* \cdot \ln(\frac{s^*}{\beta}) \quad . \qquad (14)$$

Application

Now that the model compensates for acceleration losses it can be used across a wide range of distances. To do the calculations, a computer programme might be devised, a preliminary version of which is available on the web (Westera, 2009). Users enter their two personal bests required for calibration and enter one or more distances for which they receive their prophesised times.

The logarithmic linearity of equation (3) also offers the opportunity of a simple graphical representation of the model. This is displayed in figure (3). While the vertical axis denotes lap time and the horizontal axis covers the logarithmic scale of total distance of the event, the performances of an individual athlete are given by a unique straight line. For reasons of convenience, performance times (derived from the product of lap time and distance) for each event are projected at the appropriate co-ordinates.

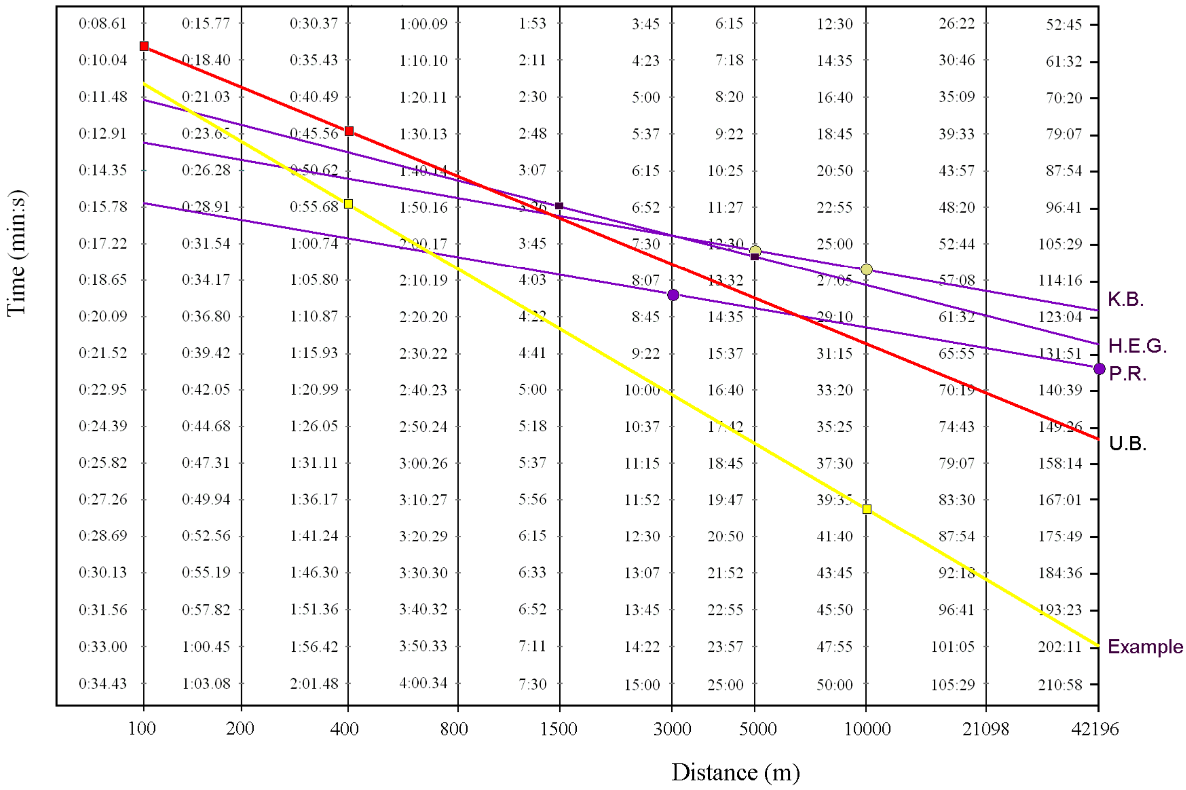

Figure 3. Graphical representation of the logarithmic model.

The yellow line (labelled "Example") is the performance curve of a fictitious athlete with personal bests of 0:55 at 400 m and 40:00 at 10,000 m. These two calibration points are indicated with yellow squares. The intersections with the ordinates provide predicted outcomes at the various distances: these are 0:10.89 at 100 m, 0:24.08 at 200 m, 2:06.57 at 800 m, 4:27.57 at 1,500 m, 10:02.65 at 3,000 m, 18:07,40 at 5,000 m, 1:32:55 at the half marathon and 3:21:41 at the marathon. The graphic representation also explains that the accuracy will go up when the two calibration events are chosen on the opposite ends of the horizontal scale. When these calibration events are too close to each other, the required extrapolation rather than interpolation will increase the inaccuracies. The dark lines in figure 3 are the performance curves of selected world class athletes who cover multiple distances: Kenenisa Bekele (K.B.), Hicham El Guerrouj (H.E.G.), Paula Radcliffe (P.R.), respectively. The calibration values used are marked with dots. According to the model Kenenisa Bekele is supposed to break the marathon world record (2:00:07). Also, he will be able to run 400 m in

52.20 s, which is only slightly below his 10,000 finish. Paula Radcliffe seems to have a "weak" 10,000 m record: the prognosis is 29:49 (against 30:01.09 personal best). Note that the curves of these two athletes display about the same slope, indicating the same type of decay that is probably distinctive for long distance runners. The curve of Hicham El Guerrouj shows a steeper slope, which can be attributed to the higher cruise speeds that he is capable of in the mid-range distances. One of the things that turns out is that Hicham El Guerrouj would be capable of running 10,000 m in 27:17. The performance curve of Usain Bolt (U.B.) is also displayed in figure 3. Here, his personal bests at 100 m and 400 m are used, although the latter (45.28) dates from 2007 which is well before his breakthrough as a world class sprinter; indeed the predicted 200 m time of 20.41 sec is far behind his actual record. Nevertheless, using the 100 m and 200 m events for calibration doesn't make much sense exactly because of the inaccuracies due to reinforced extrapolation. Still, the predictions at the longer distances don't make much sense here because of the extreme extrapolations they would require.

An interesting question would be at what distance world class sprinter Usain Bolt would compare with long distance runner Kenenisa Bekele. With the model it can be calculated that Bolt would meet Bekele at *1431 m*. This corresponds with the horizontal co-ordinate of the intersection of the two performance curves in figure 3 (although graphical estimation is less accurate). A pre-assumption of this prediction is that both calibration records are of equal quality. This is probably not the case. When Bolt would be able to update his 400 m best time, his curve will go up at the right hand side, so he might be able to make it until 1,500 m or more.

In principle, the approach is not limited to running events. It may also be applicable for similar events in other disciplines, like speed skating and swimming. Table 8 presents the outcomes for a small sample of world class speed skaters (KNSB, 2009; ISU, 2009).

Table 8. Logarithmic predictions for speed skating.

Skater	Distance (m)	Personal best (ISU)	Logarithmic model	
			Predicted personal best	Deviation
Cindy Klassen (female)	1500 *3000* 5000	1:51.79 *3:53.34* 6:48.79	1:51.79 *3:55.59* 6:48.79	*-0.010*
Martine Sablikova (female)	1500 *3000* 5000	1:51.79 *3:55.83* 6:45.61	1:51.79 *3:54.49* 6:45.61	*0.006*
Shani Davis (male)	1500 *5000* 10000	1:41.80 *6:10.49* 13:05.94	1:41.80 *6:08.16* 13:5.94	*0.006*
Sven Kramer (male)	1500 *5000* 10000	1:43.99 *6:03.32* 12:41.69	1:43.99 *6:03.19* 12:41.69	*0.000*
Overall deviation				0.005

For easy estimation of start-up corrections for each speed skating event, the split times of the associated world records have been used. Outcomes are promising, showing average inaccuracy of 0.5% for the sample.

Likewise, table 9 shows the results for a sample of world class swimmers (Swimnews, 2009).

Table 9. Logarithmic predictions for swimming.

Swimmer	Distance (m)	Personal best (ISU)	Logarithmic model	
			Predicted personal best	Deviation
Laure Manaudou (female)	200	1:55.52	1:55.52	
	400	*4:02.13*	*3:59.90*	*0.009*
	800	*8:18.80*	*8:17.53*	*0.003*
	1500	16:03.01	16:3.01	
Ian Thorpe (male)	200	1:44.06	1:44.06	
	400	*3:40.08*	*3:38.85*	*0.006*
	800	7:39.16	7:39.16	
Grant Hackett (male)	200	1:45.61	1:45.61	
	400	*3:42.51*	*3:38.79*	*0.017*
	800	*7:38.65*	*7:32.71*	*0.013*
	1500	14:34.56	14:34.56	
Overall deviation				0.009

Because of the low speeds in swimming and for reasons of simplicity start-up corrections have been neglected here (γ =0). Nevertheless, predictive power for this sample of swimmers is around 1 %. Note that the eventual boosting effects of turning points (every 50 m) have not been taken into account.

Conclusions

The calculations presented in this paper provide strong evidence that the proposed model produces valid and reliable predictions. Inaccuracies are typically around 1%. Therefore the model greatly outperforms existing models (typically 3% or higher). Specific accuracies of the calculations, however, are quite dependent on the conditions for interpolation or extrapolation. Besides its unchallenged accuracy, the model has some additional advantages. Importantly, the model is transparent, since it is based on theoretical principles rather than arbitrariness and negotiation. Furthermore it is self-contained, easy to use and affordable, because it does not require any physiological or biomechanical tests to be carried out first: it just uses two personal bests for individual self–calibration. Since the model compensates for start-up delays it is valid across a wide range of events, including sprinting, mid range and long distance running. Finally, the model displays a universal validity: it seems to be highly applicable for any speed and distance related sports event, including running, speed skating and swimming.

References

Achilles-Top (2009) Clubrecords Athletiekvereniging Achilles-Top Kerkrade, retrieved September 17, 2009 from http://www.achilles-top.nl/

Di Prampero, P.E., Fusi, S. Sepulcri, L., Morin, J.B., Belli, A., and Antonutto, G. (2005) Sprint running: a new energetic approach, *Journal of Experimental Biology*, 208, 2809-2816.

Costill, D.L., Thomason, H., and Roberts, E. (1973) Fractional utilization of the aerobic capacity during distance running, *Medicine & Science in Sports & Exercise*, 5, 248–252.

Bundle, M.W., Hoyt, R.W., and Weyand, P.G. (2003) High-speed running performance: a new approach to assessment and prediction, *Journal of Applied Physiology*, 95, 1955-1962.

Daniels, J., and Daniels, N. (1993) Running economy in elite male and elite female runners, *Medicine & Science in Sports & Exercise*, 24, 483–489.

Harder, D. (2001) Apples to Oranges, *Education Plus*, 1.

IAAF (2009) Top lists, International Association of Athletics Federations, retrieved September 17, 2009 from http://www.iaaf.org/statistics/recbycat/index.html

ISU (2009) World all time top 100 best, International Skating Union, retrieved September 17, 2009 from http://www.isu.org/vsite/vfile/page/fileurl/0,11040,4844-187597-204819-115404-0-file,00.pdf

KNSB (2009) Speed skating world records, Koninklijke Nederlandse Schaatsbond, retrieved September 17, 2009 from http://www.knsb.nl/content/langebaan/records/lbwr.asp

Lee, J. (2008) Usain Bolt 100m 10 meter Splits and Speed Endurance, August 22, Speedendurance.com, retrieved September 17, 2009 from http://speedendurance.com/2008/08/22/usain-bolt-100m-10-meter-splits-and-speed-endurance/

Lee, J. (2009a) Usain Bolt 10 meter splits, Fastest Top Speed, 2008 vs 2009, August 19, Speedendurance.com, retrieved September 17, 2009 from http://speedendurance.com/2009/08/19/usain-bolt-10-meter-splits-fastest-top-speed-2008-vs-2009/

Lee, J. (2009b) Usain Bolt 200 meter splits, Speed Reserve and Speed Endurance, August 21, Speedendurance.com, retrieved September 17, 2009 from http://speedendurance.com/2009/08/21/usain-bolt-200-meter-splits-speed-reserve-and-speed-endurance/

Spiriev, B. (2008) IAAF Scoring Tables of Athletics, 2008 revised edition, retrieved September 17, 2009 from http://www.iaaf.org/mm/Document/Competitions/TechnicalArea/ScoringOutdoor2008_742.pdf

Swimnews (2009) Best performers lists, retrieved September 17, 2009 from http://www.swimnews.com/rank/index

Trkal, V (2003) The development of combined'events scoring tables and implications for the training of decathletes, *New studies in Athletics*, 18(4), 7-17

Westera, W. (2006). Decathlon, towards a balanced and sustainable performance assessment method. *New Studies in Athletics*, IAAF, 21(1), 39-51.

Westera, W. (2009) Personal record predictor, preliminary version, JAVA applet retrieved September 17, 2009 from http://www.open.ou.nl/wim/calculate