IOSR Journal of Sports and Physical Education (IOSR-JSPE)
e-ISSN: 2347-6737, p-ISSN: 2347-6745, Volume 5, Issue 6, (Nov. – Dec. 2018), PP 06-22
www.iosrjournals.org

Modeling World Record Predictions in Track Events

G. Donald Allen
Department of Mathematics, Texas A&M University, USA
Corresponding Author: G. Donald Allen

Abstract: *We consider here the estimation of best possible running track events. A logistic type curve is applied with a least squares estimator and applying the gradient descent method. The facts seem to indicate a non-uniqueness of the best estimator and a rather slow convergence. The initial data must be reasonably accurate to obtain convergence, and convergence can require thousands of iterations. Since these are not production codes lengthy iterations are not significant, but there can be multiple convergence values depending on initial values. The implication is that unlike other literature which offers up best times, we demonstrate there can be substantial variance is best estimates. Predicting world records involves combining human mechanics, training, psychology, diet, and coaching. Our focus, mathematical models, groups these as four parameter models.*

Keywords: *Logistic, prediction, sigmoidal, track records*

Date of Submission: 07-12-2018 Date of acceptance: 22-12-2018

I. Introduction

The study of legged human locomotion is complex. It involves muscles, oxygen, neural coordination, centers of gravity, stresses, strains, diet, and an assortment of metabolic processes. The basic categories of physiology, mechanics, and neural commitment command a deep knowledge of a multiple of topics to devise the metrics of speed. In this paper, we take a fundamentalist viewpoint of modeling speed in terms of best possible and other data of best records for common races. Maximum running speeds have been estimated based on body mass (Hutchinson and Garcia, 2002), as well as rates at which energy can be consumed (Keller, 1973). Training methods have advanced into the realm of true science (McKlusky, 2014) in which few details are omitted.

There is little disagreement that for a given species, there is a maximal speed or time for a given distance, with the exception of stochastic effects to be discussed later. There is substantial controversy on how this maximum is approached. Other approaches are available from phenomenology to statistics (Gembris,Taylor, Suter, 2007; Gembris, et.al., 2002), from extreme value theory (Denny, 2008) to analytical approaches (Nevill and Whyte, 2005). Thus, one basic question is the nature of how best running speeds approach their asymptotes? In this paper, we take an analytic approach, applying several models of how this can occur. It is based on logistic-type differential equations which differ in how the logistic driver meets or approaches the asymptotes.

Extreme value theory examines the probability of a given extreme value is achieved during a given year. Based on the Fisher–Tippett–Gnedenko theorem which provides asymptotic estimates, it posits that extreme values in a specific event are independent identically distributed random variables (Gumbel, 2004, Gnedenko, 1943, Coles, 2001).For extreme events of weather, this is a reasonable assumption, but for race events, each new extreme value, such as the world or track record, does impact the corpus of participants in many ways, in the simplest notions of inspiring greater performance or discouraging future efforts. The statistical approach applies statistical methods to demonstrate that records can improve in time even if athletic performance remains constant. This implies the presence of stochastic effects and systematic trends. It also is impacted by sample size, importance, location, the number of competitions, track conditions, and other factors. (Gembris, 2007). Results are both significant and interesting.

In the present paper, we examine several analytical models based generalized logistical curves and their associated differential equations, discuss qualitatively how they differ near the asymptotes, and compare predicted records into the future. For example, the classical logistic curve converges much faster to the asymptotes than the arctangent method. We note that final values for the optimization codes are highly dependent on initial conditions. As to phenomenology, we simply look at the progression of records and measure by how much they successively change and applied a weighted average to predict the next record. Such a program, we will see, provides results consistent with the more sophisticated models and with extreme value theory.

The upshot of the current approach is to demonstrate that no matter what models are developed, there must be inherent assumptions that can at best be considered reasonable, at worst suspicious, but rarely true or false. In section 2, we consider factors affecting records. In Section 3, the nature of logistic models. In Section 4, we outline the optimization method applied; in Section 5, testing the method is discussed; in Section 6, the nature of pre-predictions, or the steps required prior to making predictions. Finally, the predictions themselves are given in Section 7.

II. Factors affecting world records

A natural question about the improvement of records is whether we are predicting records within the spectrum current conditions or against the future. For example, suppose it is discovered that athletes breathing an odd combination of oxygen, nitrogen, and hydrogen gasses (completely legal) for 30 minutes prior to a race dramatically improve their performance. Can you imagine future track stars appearing on the field wearing scuba-like equipment? How can this be factored into any model? So, we are confronted with the racing system as it is versus what it may be. Is it closed, or must it allow for new factors? For example, some models allow for the increase of population (Denny, 2008).

With this in mind, we consider the many factors impacting the world, national, Olympic, and most other track records. The makes it important to classify just what are the factors involved. We focus completely on the rate of change of records and the asymptotes. Concerning the rate, it is safe to say that some of the factors change the rate to the extent that though there is a governing equation or differential equation, the rate can change. Other affects the ultimate best possible records themselves. In this section, we review these simply with an overview, giving some examples to support them. Most have been well documented in the literature. We have only room for a brief sketch of nine of the twelve main topics. See (McKlusky, 2013) for more information. Sports science is a highly technical topic contributing much to our knowledge of athletic performance.

1. World population - A larger pool of athletes of athletes gives a greater number of persons trying to set new records and overall increasing the competition. This is not as true today as in years past, but it is a factor.
2. Genetic stock - When racers sampled from different geographical regions, special genetic proclivities appear. As we will see, the number of Africans entering long distance races not only have improved world records but increase the rate they are achieved.
3. Sports fashion - With records so difficult to achieve for so little gain, some of the most talented athletes will gravitate toward other venues, possibly more lucrative. This decreases the fraction of available athletes. Case-in-point: competitive walking, also called racewalking has a diminished fashion as an event. Relatively few compete.
4. Equipment - Of course, we've heard of friction reducing swimming attire that has helped lower record swim times. The same is true in most sports, including track. The shoe and its weight stand out. Most competitive track stars pay particular attention to their gear, particularly shoes. shoe for comfort and elasticity, traction, tunic for temperature control. New advanced design shoes seem promising to be a component of new records in the longer distance races https://www.nytimes.com/2017/03/08/sports/nikes-vivid-shoes-and-the-gray-area-of-performance-enhancement.html)
5. Training methods - There was a time when the prospective elite athlete would be told (Ericsson, Krampe, Tesch-Romer (1993)) that to reach the highest level performance, thousands of hours of training was required. Though not due to Ericsson, a number, at least 10,000 hours of training, became associated with this goal. This is time on task and became the prescription to achieve the elite level. However, with too many counter-examples about, the amount of training is now regarded as dependent on age, sport, motivation, commitment, and other factors.It has now been determined that athletes, suitably selected, can be brought to elite levels in less than 18 months — sometimes.
6. Focus on particular body types for the given event. The advent of the discoveries mentioned above, including focusing on body type can be used to short-circuit this rule. In particular, the English Olympic Committee developed a method of recruitment of women to their rowing team. Criteria were previous high level of active in other sports, but tall and strong. With intense coaching the rowing team thrived, producing elite rowers in only a few years, including at least one Olympic gold medal.
7. Coaches and coaching methods. Who gets the best coaches? What are the rewards for good coaches? As well, the perfect fit of coach and athlete is not an easy match to make. Under the aegis of coaching is the relatively modern use of videos for the athlete and coach to reflect on performance.
8. Race preparation. Just as an example, warm–up rooms are now routinely operated by individual countries for athletes to tune-up just prior to big races. This in addition to dietetic prescriptions clearly improve race performance.
9. Random effects, aka black swan events. Random effects are sometimes emphasized in lieu of training

methods as strongly affecting new records (Gembris, 2007). These may be more significant than we now understand. For suppose, the best time possible by the human in the 100-meter race has been achieved. This does not end the story. Indeed, it is clearly possible that on a particular day, with particular circumstances, an elite athlete may exceed only slightly a given world record. An example of this is with Florence Griffith-Joyner in the 100-meter women's race, who set the current record in 1988 of almost 0.3 seconds faster than the previous record. It has remained unchanged. The best time since Joyner is still 0.15 seconds slower than this now almost ancient record. It is certainly an indicator than any record unchanged in 20-30 years may be very near the best possible - except for black swan (Taleb, 2010) events.

In Table 1 we highlight how these events affect records and rates. In general, these same factors affect race performance at all levels.

Category	**Affects record* (*a* and *b*)**	**Affects rate (c)**
Population	√	√
Genetic stock	√	
Equipment	√	√
Identification of physical types	√	√
Training	√	
Sports fashion	√	√
Diet, VO_2, etc	√	√
Coaching	√	√
Race preparation	√	
Random factors	√	
Longevity of records***	√	√
Resources (e.g. government)	√	√

Table 1.

Factors affecting world records
* Generally to improve the best record
** To accelerate or decelerate time to achieve a new record
*** Dependent on model weights

To supplement our discussion of population issues in the United States, consider Table 2 which shows American participation in the schools over the years 2003-2016. As is evident, US schools have not really increased their participation rates by much over the last dozen or so years.

Track and Field				
School Year	**Boys Participation**	**Girls Participation**	**Boys Participation PCT increase**	**Girls Participation PCT increase**
2002/2003	498,027	415,602	1.36%	0.65%
2003/2004	504,801	418,322	2.36%	2.36%
2004/2005	516,703	428,198	3.34%	2.57%
2005/2006	533,985	439,200	1.91%	1.13%
2006/2007	544,180	444,181	0.85%	0.75%
2007/2008	548,821	447,520	1.67%	2.28%
2008/2009	558,007	457,732	2.53%	2.50%
2009/2010	572,123	469,177	1.25%	1.30%
2010/2011	579,302	475,265	-0.63%	-1.37%
2011/2012	575,628	468,747	-0.20%	-0.18%
2012/2013	574,451	467,897	1.02%	2.35%
2013/2014	580,321	478,885	-0.29%	-0.03%
2014/2015	578,632	478,726	2.16%	1.51%
2015/2016	591,133	485,969		
Average	**554,008**	**455,387**	**1.33%**	**1.22%**

Table 2. Participation in Track and Field in the US

However, a deeper point needs mention. The very top performers in these groups are not necessarily drawn into the group by encouraging greater participation. They are already there. Additionally, Table 2 includes all sports events, with few of the top performers coming from track and field events. To support the genetic stock or fashion arguments or simply fashion, consider the following tables (Track and Field News, 2015).The first, Table 3, refers to the numbers of athletes ranked in the top ten by year since 1947 for the 100-meter dash. Overall, 40 countries placed athletes in the top ten over this period. There is every reason to accept the true contenders for setting a world record are among the top ten - whatever the event. It reveals the dominant athletes from the US dropped in percentage from about 46% to 32%. Note that Jamaica, a tiny country, ranked second has moved up considerably. Most other countries, aside from Great Britain and Trinidad are minor contributors.

Note also that East Germany is listed, but combining East Germany with Germany, the numbers are the same. For the Marathon race, the picture is different. In Table 4, we note the Kenyan contribution is first while the US, a major player prior to 1990 is now in the minor ranks. Kenya has become the dominant world leader in the Marathon race.

Similar rankings and changes thereof for the 10,000-meter race also obtain. Remarkably, Japan suffered the greatest percentage loss in the top ten rankings. Ethiopia nearly tripled its percentage in rankings. Overall, about 83% of the modern era top ten Marathon athletes are citizens of Kenya, Ethiopia, and Morocco. In both tables, there is overlap in the counts of top-ten athletes, 1947-2015 and 1990-2015. The goal here is to recognize recent changes in the national demography within the top tiers. Similar results hold for the other races — depending on the distance. Another factor clearly evident in these lists is that some governments expend large resources on sports, notably the former Soviet Union and East Germany, and can enhance competition and affect performance.

Country	Number since 1947	Number since 1990	PCT of all	PCT since 1990		Rank	Country	Number since 1947	Number since 1990	PCT of all	PCT since 1990
US	316	51	45.8%	31.9%		11	Italy	11	0	1.6%	0.0%
Jamaica	67	43	9.7%	26.9%		12	Namibia	10	2	1.4%	1.3%
Great Britain	42	12	6.1%	7.5%		13	Poland	10	0	1.4%	0.0%
Trinidad	30	11	4.3%	6.9%		14	Australia	9	0	1.3%	0.0%
Cuba	27	0	3.9%	0.0%		15	St. Kitts	8	7	1.2%	4.4%
Canada	26	1	3.8%	0.6%		16	East Germany	8	0	1.2%	0.0%
Nigeria	23	7	3.3%	4.4%		17	Portugal	5	5	0.7%	3.1%
France	18	8	2.6%	5.0%		18	Ghana	5	4	0.7%	2.5%
West Germany	16	0	2.3%	0.0%		19	Panama	4	0	0.6%	0.0%
Soviet Union	13	0	1.9%	0.0%		20	Antigua	3	3	0.4%	1.9%

Table 3 - International rankings for the 100-meter dash.

Rank	Country	Number since 1947	Number since 1990	PCT of all	PCT since 1990		Rank	Country	Number since 1947	Number since 1990	PCT of all	PCT since 1990
1	Kenya	118	86	17.1%	53.8%		11	Belgium	16	0	2.3%	0.0%
2	Japan	97	3	14.1%	1.9%		12	Mexico	14	0	2.0%	0.0%
3	Great Britain	60	0	8.7%	0.0%		13	Tanzania	12	1	1.7%	0.6%
4	Ethiopia	55	34	8.0%	21.3%		14	New Zealand	12	0	1.7%	0.0%
5	US	41	3	5.9%	1.9%		15	Sweden	12	0	1.7%	0.0%
6	Finland	37	0	5.4%	0.0%		16	South Africa	11	6	1.6%	3.8%
7	Soviet Union	27	0	3.9%	0.0%		17	Portugal	11	1	1.6%	0.6%
8	Australia	24	0	3.5%	0.0%		18	East Germany	11	0	1.6%	0.0%
9	Italy	17	4	2.5%	2.5%		19	Czechoslovakia	11	0	1.6%	0.0%
10	Morocco	16	13	2.3%	8.1%		20	Spain	10	4	1.4%	2.5%

Table 4 - International rankings for the Marathon.

There is not space in this short report to discuss all of the factors in Table 1. However, the reader can gain great insight from the book by McClusky, 2014 and related literature. Notably, performance in sprint races supports the US dominance.

III. The Logistic Model

Consider for a moment what kind of model, or function, is needed to model ultimate world records. Assuming we have a fixed population, possibly growing in size, the first thing needed is two asymptotes, one for the distant past and one for the distant future, which is the ultimate record. Obviously, the record time is decreasing; so, the function must be decreasing. For example, fitting a quadratic model to extant data will give a minimum but not an asymptote. Moreover, with a quadratic model, the distant past must be ignored completely. The term "fitting" is important because of native irregularities of data. Smooth is important, with the requirement at least two realms persist. The first is that up to some point records are decreasing at an increasing rate as training methods and populations increase, and afterward where the record decreases at a decreasing rate tapering toward the ultimate record. This indicates the existence of at least one inflection point. Numerous factors, some discussed above, can affect the record bound, and these serve to complexify the model. Such are discussed later in the paper. With our basic requirements in place, decreasing over time, with concavity negative to some point with concavity positive afterward, we posit general criteria.

Generalized logistic growth. The steps in making a model are routine: make a model with undetermined parameters, apply the data to determine the parameters. This, rather general description applies to most modeling, from mathematical to statistical. In this paper,the focus is on the mathematical models of the logistic type.

The notion of logistic growth or decay can be generalized with simple conditions: Given two values $b < a$, a logistic function is monotone with these values as asymptotes as the independent variable (usually time) tends to either $\pm\infty$. We call it strictlylogistic if it has but one identifiable inflection point, implying it has the classical sigmoidal shape. In Fig. 1a we show both increasing and decreasing strictly logistic forms. In Fig. 1b we show how the data fits to the logistic curve in the women's mile race. To model record race times, we will use the decreasing form, with the higher/larger asymptote referring to the distant past, and the lower/smaller asymptote referring to the distant future, i.e. the best possible record time. In most growth models, particularly biological growth, time begins at 0, and often the logistic function value is zero as well.

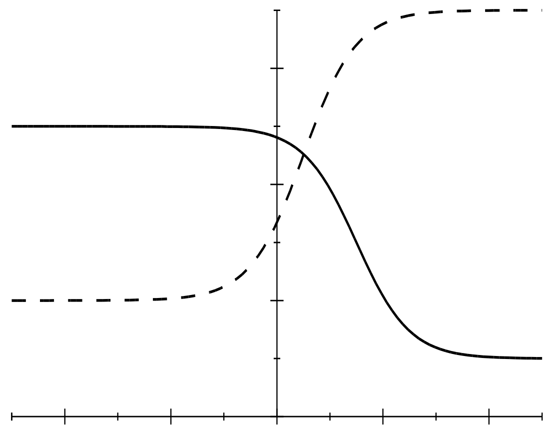

Figure 1a. Increasing and decreasing logistic forms

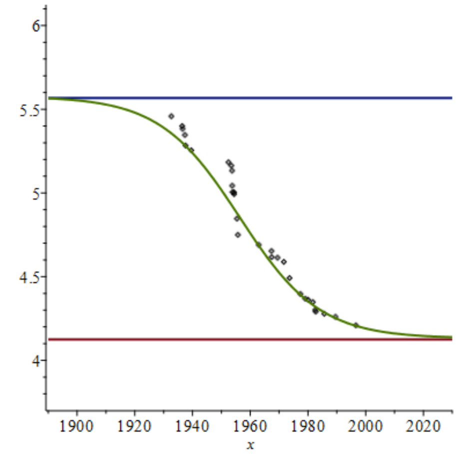

Figure 1b. Data and fitted curve for the Women's mile

To generate models for record race times with dates requires at least a four-parameter model, the two asymptotes referencing record times, past and future, and two other parameters, c and d, pertaining to record dates. In many cases d, is related to the mean or median of record dates and c is related to the slope of the curve at the inflection point which in turn is inversely related to the standard deviation. All dates are given in decimal years since 1900.

A note about the upper asymptote, b, is that in the distant past (prior to 1900 in most cases) there were limited official races and timing methods, and even more limited participation in them, and still more relatively constant training methods (practice a lot) and equipment. Such conditions persisted for decades or centuries. Only in the modern era have these factors been carefully examined and improved, together with a very large increase in the population of competitors. Thus, we conjecture b to be this past relative "best" time. Below we discuss other factors and how they related to record times. Similarly, the record best time is dependent on the model and is at it very best an estimate, which we will show is often in error. To the racing community, it is a

greater event to modelers to record a time exceeding best estimates than to merely record a world record. Any new record implies the models will be reapplied to re-compute a new best estimate.
While not necessary, logistic growth is often associated with a separable initial value problem of the form

$$\frac{dy}{dx} = \pm\kappa(x)L(y), \qquad y(x_0) = y_0$$

If $L(a) = L(b) = 0$, and if $b < y_0 < a$, if $L(y) > 0$ and is bounded on (b, a), if $\kappa(x)$, called the rate, is nonnegative and bounded for all x, and if the solution exists for all x, it exhibits the generalized logistic behavior. In this paper,we focus on autonomous or time-invariant types, namely those where the rate $\kappa(x)$ is a constant. The value of x where the $y''(x) = 0$ is the inflection point. The change point of concavity figures into most of the literature on the subject. The function $L$is called the logistic driver. If we are to consider population as a factor, then □ truly depends on time. Consider a few examples, some of which may be familiar.

Example 1. We can view the prediction problem as something of a boundary value problem about how the curve approaches the limits, top and bottom. Each presents its own advantages and problems. For example, with the standard logistic driver $L_1(y) = -c(a - y)(y - b)$, where $b < a$, the curve approaches both asymptotes with positive finite slopes. The solution to the differential equation is

$$y(x) = a - \frac{a - b}{1 + \exp(-(a - b)(cx - d))}$$

As is apparent $\lim_{x\to-\infty} y(x) = a$, and $\lim_{x\to\infty} y(x) = b$. The inflection point occurs at $x = c/d$.The examples shown in Fig. 1 are in fact logistic curves of this type. This is the logistic equation as originally discussed by (Verhultst, 1838).

Example 2. Using the logistic function $L_2(y) = -c\cos^2 y$, the slopes are zero at the asymptotes, $\pm\frac{\pi}{2}$.

Consider $L_2(y) = -c\cos^2(ry + s)$. Then $y = a$, we have $ra + s = \pi/2$, and when $y = b$ we have $rb + s = -\pi/2$. This gives $ra + s = \pi/2$ and $rb + s = -\pi/2$. So,

$$r = \frac{\pi}{a - b}$$

$$s = \frac{\pi}{2} - \frac{\pi}{a - b}a = -\frac{1}{2}\frac{\pi}{a - b}(a + b)$$

In any event, the slopes are zero at the asymptotes, implying the limiting behavior of the solution to $y' = L_2(y)$ converges somewhat slower at the asymptotes that do that for $y' = L_1(y)$, and this means the function is somewhat flatter in between. For L_2 the ultimate solution to the logistic ODE involves the arc tangent function, namely

$$y = \frac{a + b}{2} - \frac{a - b}{\pi}\arctan\frac{\pi}{a - b}(cx - d)$$

This example is similar to $L(y) = -c\cos\left(\frac{\pi}{2}y\right)$ on $[-1,1]$, which is entirely similar to $M(y) = (1 - y)(y - 1)$,and this in turn is essentially a beta distribution. One problem with using such a logistic driver is the difficulty with inverting the resulting integral $\int \sec y\,dy = \ln|\sec x + \tan x|$. This and the previous example suggest that finite probability distributions play a significant role in the description of generalized logistic functions. Many of these are discussed in the literature (Johnson, Kotz, Balakrishnan,1995).

Example 3. With the logistic-driver function $L_3(y) = -c\sqrt{1 - y^2}$note the slopes are vertical at the endpoints, ± 1. This translates with the usual asymptotes $b < a$ to

$$L_3(y) = -c\sqrt{\left(\frac{a - b}{2}\right)^2 - \left(y - \left(\frac{a + b}{2}\right)\right)^2}$$

This forces the trajectory to approach the asymptotes in a finite time. In fact, for $L_3(y)$, the resulting solution is a sinusoid and thus has meaning it's valid over only one half period and is constant (rsp. a and b) outside of $[b, a]$. To solve, compute

$$\int -\frac{1}{\sqrt{\left(\frac{a-b}{2}\right)^2-\left(y-\left(\frac{a+b}{2}\right)\right)^2}}dy = \arcsin\frac{1}{\sqrt{(a-b)^2}}(a+b-2y) = cx-d$$

Solve for y to get$\arcsin\frac{1}{\sqrt{(a-b)^2}}(a+b-2y) = cx-d$. This gives the solution

$$y = \frac{1}{2}(a+b) - \frac{\sqrt{(a-b)^2}}{2}\sin(cx-d)$$

but only for $\frac{-\pi}{2} \leq cx-d \leq \frac{\pi}{2}$. Beyond this range, the solution takes on the asymptotic values. This implies the asymptotes are reached in finite time. Both models, $L_2(y)$ and $L_3(y)$, have no previous appearance in the literature. Nor does the next.

Example 4. This is an example not directly related to a differential equation of the type reviewed above. Based on the normal density, we apply its cumulative distribution function $\frac{1}{2}(1+\operatorname{erf}(x))$, where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$ to achieve the functional form

$$y = a + \frac{b-a}{2}(1-\operatorname{erf}(-c(x-d)))$$

It is sigmoidal in the desired shape and the minus sign on $\operatorname{erf}(\cdot)$ makes y decreasing, as needed. Here c is related to $\frac{1}{2\sigma}$, where σ is the standard deviation of the record date, and d is related to their mean or median. In this and all cases, these provide starting points for the optimization problem. As can be easily checked, the appropriate differential equation that generates the normal density is

$$\sigma^2 y' + (x-\mu)y = 0, \qquad y(0) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

To achieve the error function, we need a second integration, which implies we can convert this to an appropriate second order differential equation for the erf. What is interesting about this model is that it is phenomenological but based on the (normal-type) nature of data. It is also interesting that this method can be fitted to the data with remarkably small residuals. However, we can construct the differential equation with a rate function

$$y' = -\frac{\frac{1}{\sqrt{\pi}}ce^{-c^2(d-x)^2}(a-b)}{a+\frac{b-a}{2}(1-\operatorname{erf}(-c(x-d)))}y$$

Of course, this $\kappa(x) = \frac{\frac{1}{\sqrt{\pi}}ce^{-c^2(d-x)^2}(a-b)}{a+\frac{b-a}{2}(1-\operatorname{erf}(-c(x-d)))}$ is bounded. Solving gives

$$\ln y(x) = -\int \frac{\frac{1}{\sqrt{\pi}}ce^{-c^2(d-x)^2}(a-b)}{a+\frac{b-a}{2}(1-\operatorname{erf}(-c(x-d)))}dx = \ln\left(a+\frac{b-a}{2}(1-\operatorname{erf}(-c(x-d)))\right)$$

$$y(x) = a + \frac{b-a}{2}(1-\operatorname{erf}(-c(x-d)))$$

Now we have the built-in initial condition, $y(d) = \frac{a+b}{2}$. It is easy to check for this $y(x)$ the two appropriate asymptotes are determined $\lim_{x\to\infty} y(x) = a$ and $\lim_{x\to-\infty} y(x) = b$.

So far, the slopes at the asymptotes have been anti-symmetric. That is $L'(a) = -L'(b)$. This implies the same rate and functional mechanisms apply both in forward and backward in time. This may not be the case, though it is not discussed in the literature. For example, in some unusual races such as the 50K walk, there may be so few racers that not only are races difficult to find but racers to compete with are so very few. This can retard the lowering of records, in contrast to when there are multiple racers and events. We consider a few examples of such logistic drivers. Simply put, the race population can result in different asymptotic rates at $\pm\infty$.

Example 5. A fourth logistic function, which is applied only in the interval $[b, a]$, and $c > a$ is
$L_4(y) = -(a-y)(y-b)(c-y)$.
This function is negative on $[b, a]$ and has different slopes at a and b.
$L_4'(a) = (a-b)(a-c)$
$L_4'(b) = -(a-b)(b-c)$
This gives different slopes at the asymptotes $b < a$. Assuming the initial condition places $y(0) \in (b, a)$, the trajectories remain in (b, a) Yet, the trajectories approach the asymptotes at different rates. To solve the ODE in his case we integrate (from the ODE
$y' = L_4(y)$)

$$\int \frac{dy}{L_4(y)} = \frac{1}{(a-b)(a-c)(b-c)}(a\ln(y-b) - b\ln(y-a) - a\ln(y-c) + c\ln(y-a) + b\ln(y-c) - c\ln(y-b))$$

$$= -\frac{1}{(a-b)(a-c)(b-c)}\ln\left((y-a)^{c-b}(y-b)^{a-c}(y-c)^{b-a}\right)$$

This is difficult to invert, and we don't use it here. However, we could still use the optimization, done implicitly. A typical graph looks like for L_4 (e.g. $L_4(y) = -(5-y)(y-2)(7-y)$ is shown in Fig. 2.

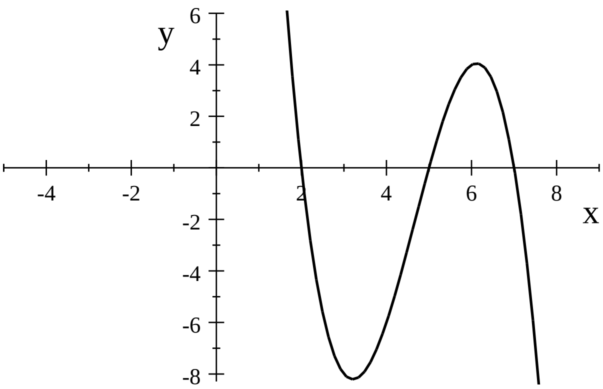

Figure 2. A non-symmetric logistic driver

So, the solution of the ODE remains in $[2,5]$ and is decreasing. A special case of L_4 is a specific form of the beta distribution (online at https://en.wikipedia.org/wiki/Beta_distribution)

$$L_4 = ly - my^p = ly\left(1 - \frac{m}{l}y^{p-1}\right)$$

The asymptotes are 0 and $\left(\frac{l}{m}\right)^{1/(p-1)}$. If the initial condition is between these values, it remains so. If $l < 0$, and n is even, the solution remains between them. If $l > 0$, the solution increases and if $l < 0$, the solution decreases. This model has been used with success in the case by (Richards, 1959) in modeling somatic growth.

Recasting for decreasing measures, or race records for example, the model takes the form

$$y' = -c(a-y)\left(1 - \frac{(a-y)^{p-1}}{(a-b)^{p-1}}\right)$$

with solution

$$y = a - \left(\frac{1}{(a-b)^{-(p-1)} + \exp(-c(p-1)(x-d))} \right)^{\frac{1}{p-1}}$$

with asymptotes $b < a$ and rate $c > 0$. Usually $p > 1$.In general, when $p > 1$, the inflection point, i.e. the point where $y'' = 0$, occurs at

$$x = -\frac{1}{c(p-1)} \left(\ln (a-b)^{-p+1} (p-1) - cd(p-1) \right)$$

This somatic growth model, converted to race record decrease, may be a reasonable model, because when $1 < p < 2$, the growth actually stops, and likewise the race records decrease. This means the best record is actually achieved with the exception of stochastic and other special effects, as discussed above. Of course, this introduces a five-parameter model, providing we optimize on all.

Many other sigmoidal curves are available. The Gompertz (Gompertz, 1825). curve is among them. Also called the double exponential (Allen, 2017), it has the form

$y_G(x) = ae^{-re^{-sx}}$, with a, r, s are constants

Normally, r and s are positive constants and usually, c is called the growth rate. Under these conditions, $\lim_{x\to\infty} y_G(x) = a$. It satisfies the differential equation of the form $y_G' = \pm k y_G \ln\left(\frac{K}{y_G}\right)$. We can reform it as a decreasing function of the type we need (with horizontal asymptotes $b < a$) by recasting it as

$$y(x) = a - (a-b)e^{-re^{-(sx-d)}}, \qquad \text{for all } x$$

The general logistic driver. Suppose $f(x)$ is a generalized logistic function, continuously differentiable with $f'(x)$ never zero, and with asymptotes $b < a$. Thus $f(x)$ has a differentiable inverse with domain (b, a). Define

$$L(y) = \left(\frac{d}{dy} f^{-1}(y) \right)^{-1}$$

Then the differential equation

$$y' = L(y)$$

has solution $y(x) = f(x)$. Of course, an initial condition needs to be specified. For the error function case discussed above, it is required to find the inverse error function, $\text{erf}^{-1}(y)$, which is unknown in closed form. However, the process is to invert $y = a + \frac{b-a}{2}(1 - \text{erf}(-c(x-d)))$ to obtain

$$f^{-1}(y) = -\frac{1}{c} \text{erf}^{-1}\left(1 - (y-a)\frac{2}{b-a}\right) - d$$

with domain $[b, a]$. It can be shown that the resulting $L(y)$ has finite (nonzero) slopes at the endpoints, implying this logistic function resembles that of Example 1 insofar at endpoints are concerned. However, the best least squares fit for the quadratic version rises more slowly toward the asymptote to the right $(+\infty)$, while both behave similarly in a neighborhood of the common inflection point. (See MathworldWolfram or Wikipedia, online.)

In physical terms, the unique inflection point implies that prior to it, the record times are decreasing at an increasing rate, i.e. the record speeds are increasing, and from thence on the record speeds increase at a decreasing rate. Many, many other models of growth or decay, some of which are above may be found in a comprehensive paper (Sakanoue, 2007).

IV. The basic method for optimization

The method employed here is to use extant data and then fit it to the equation by adjusting the parameters using the method of least squares. The most elementary method is to make a large sample of possible parameters and determine which set fits the best. This brings us to the notion of "fit, or precisely "best fit." Having four parameters gives this nonlinear problem its first challenge.

Determining the parameters. Begin by looking at some data (x_i, y_i), $i = 1, \ldots, n$, where x_i is the year the record occurred, and y_i is the record time. Naturally, the record times decrease, with corresponding record speeds increasing. With the logistic equation and the data, we compute the residuals

$$R(a,b,c,d) = \sum_{j=1}^{n} \quad w_i(f(x_i) - y_i)^2$$

$$= \sum_{j=1}^{n} \quad w_i(a - \frac{b}{1 + e^{-(cx_j - d)b}} - y_j)^2$$

where $w_i, i = 1 \dots n$ are weights, and with the problem to minimize

$$\min_{(a,b,c,d)} R\,(a,b,c,d)$$

This is accomplished using the method of steepest descent or other optimization techniques. To review, for a given set of parameters (a_n, b_n, c_n, d_n) compute the next set by the equation

$$(a_{n+1}, b_{n+1}, c_{n+1}, d_{n+1}) = (a_n, b_n, c_n, d_n) - \nabla R(a_n, b_n, c_n, d_n)\delta$$

When using data of given world records $(x_i, y_i), i = 1 \dots N$, it is apparent the logistic curve is only an approximation. So, our task is to estimate the parameters. This is achieved using a least squares estimator. Compute

$$R(a,b,c,d,q) = \sum_{j=1}^{N} \quad w_i(f(x_i) - y_i)^2$$

$$= \sum_{j=1}^{N} \quad w_i(a - \frac{b}{1 + e^{-cq(bx_j - d)}} - y_j)^2$$

where $w_i, i = 1, \dots, N$ are weights, and with the problem to minimize

$$\min_{(a,b,c,d,q)} R\,(a,b,c,d,q)$$

This is accomplished using the method of steepest descent. For a given set of parameters $(a_n, b_n, c_n, d_n, q_n)$ compute the next set by the equation

$$(a_{n+1}, b_{n+1}, c_{n+1}, d_{n+1}) = (a_n, b_n, c_n, d_n) - \delta_n \nabla R(a_n, b_n, c_n, d_n)$$

where ∇R is the gradient of $R(x)$

$$\nabla R = \left(\frac{\partial R}{\partial a}, \frac{\partial R}{\partial b}, \frac{\partial R}{\partial c}, \frac{\partial R}{\partial d}\right)$$

evaluated at (a_n, b_n, c_n, d_n) and δ_n is a small parameter multiplier of the gradient indicating how much distance to travel. This is the most basic version of the gradient descent method. Other versions seek to minimize the residuals in the given direction using a variety of methods. For several reasons, such enhanced method seems not to work well. What we do here is modify this method using an adaptive scheme. That is, if after several iterations the residuals increase, we then decrease δ ($\delta \to \delta/2$) and revert to the previous iteration. In this way, we continue the iteration until we achieve a stopping point defined by

$$\delta_n < \epsilon_1$$

where ϵ_1 is preassigned, usually about 10^{-7} or until the relative residual error

$$E(a_{n+1}, b_{n+1}, c_{n+1}, d_{n+1}) = \left|\frac{R(a_{n+1}, b_{n+1}, c_{n+1}, d_{n+1}) - R(a_n, b_n, c_n, d_n)}{R(a_n, b_n, c_n, d_n, q_n)}\right| \le \epsilon_2$$

where ϵ_2 is also preassigned, usually about 10^{-6}. Still another criteria for stopping is that

$$\|\nabla R\| < \epsilon_3$$

When the starting values are close to the optimal values, we use the Barzilai-Borwein criterion, namely

$$\delta_n = \frac{(z_{n+1} - z_n)\big(\nabla R(z_{n+1}) - \nabla R(z_n)\big)^T}{\|\nabla R(z_{n+1}) - \nabla R(z_n)\|^2}$$

where $z_n = (a_n, b_n, c_n, d_n)$. Usually, this converges (or diverges) rapidly.

In addition, we stop all iterations after about $N = 75{,}000$ iterations, indicating the convergence is very, very slow. What is the case is that the iterations are very sensitive to changes in δ. One form of our adaptive methods is to gradually increase δ throughout the iterations to maintain maximal changes.

Weighting the Records. The first question about weights is how data should be regarded as important. We might agree that early records say those dating prior to 1920 should not count as much as more recent records. That is, for example, the weights should be increasing by date. However, there seems no clear method for assigning such weights. In fact, doing so effectively allows almost any desired outcomes in determining best possible records. In our calculations, we take the possible weights to be identical, all ones. This implies each record counts equally over the time span of all records — with one exception.

By this, we first note all records have a time duration. We might also agree that the duration of a record should have an effect. This means that a recent record that has sustained for several years should have a proportionally greater effect. The longer the most recent record has been in place, the stronger should be its effect on the ultimate record. This means that a given record that survives for say 10 years should have a weight of 10. Call these longevity weights. Pictorially, this looks as in Fig. 3, where the horizontal values are dates past 1900.

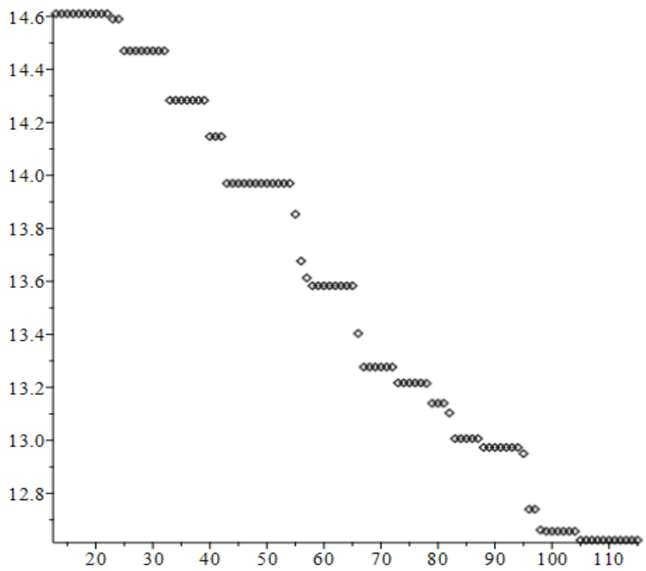

Figure 3. Men's 5000 data with weights

Another issue is the "now." That is, do we extrapolate the data as given or extrapolate from now, as in today. The consequence is the current record, which may have been in place for a number of years, should be so weighted. Such factors are built into our calculations, ultimately applying there can be really no unique fastest time.

Numerically, it does not really matter what acceleration method is used. What matters is to minimize in the context the residuals. The smaller they are, the better. We use two acceleration methods. First, after a number of steps, we increase the step size, δ, by a small amount. This permits a faster convergence rate. If it is too great, the main criteria requires it to decrease by a factor, usually of two. Second, we adjust the iterates according to the gradient rates relative to the iterate values. A few more notes about results are appropriate.1. The iterations normally converge slowly unless starting values are reasonably close to the optimal values. 2. The inflection point engenders possibly several optimal solutions. Moreover, it is not always clear where the inflection point is. This is data dependent. 3. High precision computations seem to be needed, normally about 16 digits.

In Fig. 4 we show this curve for the Men's 5000-meter race. Normally residuals are very small. However, note that a linear regression also generates small residuals. Naturally, the usual question is whether small residuals imply the correctness of the model. The answer is usually no. Small residuals provide indicators only. For this reason, we test the method in a couple of ways.

Phenomenology. Another way to predict is to use the predictions over multiple years to predict or to make a secondary prediction. That is, we can make predictions based upon predictions. Another method is more statistical. Roughly, what is done is to collect all the times over some period, compute their mean and standard deviation. Simply take two standard deviations from this mean. This could be a reasonable approximation to the best possible time. Although this seems reasonable, it does not seem to work well, partly because the data, a tail of all top times, is not normally distributed.

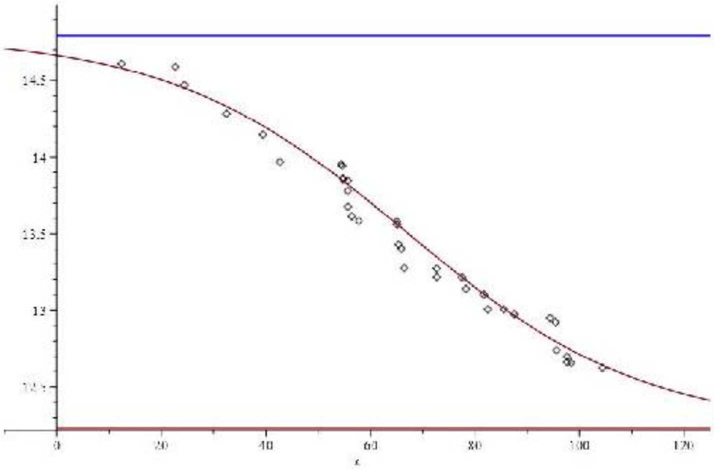

Figure 4. Men's 5000-meter race

V. Testing the Predictions

Suppose the race data were logistic of type L_1. How can we test for this? One test is to examine the data from the onset of records, year-by-year, up to the date of current record. If the best predicted time remains more-or-less constant for each successive period, this gives assurance of the logistic nature of the fit. Such a test was executed on true generated logistic data and the results were affirmative. Less than half the actual number of values used (with the same dates as real data), in fact, regenerated the four parameters of the logistic curve. However, using actual data, this was usually not the case. We show for a typically well-behaved race, the 10,000-meter women's race, calculations are not exact, but estimates are reasonably consistent over time. For other races, variation is wide, indicating the logistic premise may be inaccurate. In Table 5 we see how the estimated final race times are computed using only partial information beginning from the onset of recorded records to the first year of estimation in about 1989. The serial number, s, gives the number of data points used in the calculation, (x_i, y_i), $i = 1..s$. The highest value, 37, is the total number recorded records. All times are given in minutes and dates in years. The year of inflection is basically when at the modeled race date has value (i.e. time) $\frac{a+b}{2}$.

For the most part, the predicted best time for all was about 30-40 seconds from the projected best using all data. For example, using just the first 30 data point, the best projected time is 26.89 minutes, about $(6.89 - 6.12)60 = 46.2$ seconds. Currently, at this writing, the projected best time is about 10 seconds less than the current record. It is interesting to note that visually, the inflection point was hardly noticeable throughout. Note also, the computations are rather consistent in its determination, and note as well the distant past asymptote is reasonably steady.

Processing this data a bit more, we observe the average decrease in projected best time over the past eleven new records, since 1989, is 5.2 seconds while the average decrease in the actual record is 5.1 seconds. This appears to confirm the model is "trying" to predict what is actually happening.

Serial	Date	Projected Best Time	Current record	Year of Inflection	Residuals	Upper Asymptote
27	1989.6	26.99	27.14	1954.5	0.651	30.76
28	1993.5	27.00	27.13	1954.4	0.652	30.76
29	1993.5	26.95	26.97	1954.6	0.665	30.78
30	1994.6	26.89	26.87	1954.9	0.694	30.8
31	1995.4	26.80	26.73	1955.2	0.759	30.85
32	1996.7	26.71	26.63	1955.6	0.836	30.89
33	1997.5	26.60	26.52	1955.9	0.938	30.95
34	1997.6	26.49	26.46	1956.3	1.040	31.02
35	1998.4	26.37	26.38	1956.7	1.153	31.1
36	2004.4	26.24	26.34	1956.9	1.204	31.2
37	2005.7	26.12	26.29	1957.1	1.243	31.3

Table 5. Women's 10,000-meter race projected best time

Another method for testing the logistic nature of the data is to apply least squares to model the efficacy of actual differential equation. Call this the differential method. As above, we consider the original differential equation, $y' = -c(a-y)(y-b)$. We have the data (x_i, y_i), $i = 1..n$. Using the data, estimate the derivatives, called $\tilde{y}_i'$, and make the fit to minimize the residual for

$$\sum_{i=1}^{n} \left(\tilde{y}_i' + c(a-y_i)(y_i-b)\right)^2$$

overall variables, $a, b,$ and $c.$ This should generate approximate asymptotes, $a,$ $b,$ and the rate factor, c. In fact, when using true logistic data, this scheme does find the variables. But in general, are the residuals small? Are the variables $a, b,$ and c consistent with the full nonlinear optimization? Again, if we use true logistic data, this method accurately, almost exactly reproduces these parameters. For actual, race data, the results are not spectacular, but when this method works, the full nonlinear optimization works well. Applying this to the data for the 10,000 meter women's race yields the quadratic

$$-0.01(y-26.22)(35.72-y)$$

which approximates the values to the output from the general nonlinear model given above. Another method, call it the integral method, is similar in its nature. We compute the residual

$$\sum_{i=1}^{n-1} \left(\int_{x_i}^{x_{i+1}} y'(x)dx - \int_{x_i}^{x_{i+1}} c\,(a-y)(y-b)dx\right)^2$$

where the right-side integral is computed numerically using the data, and the left side integral simply takes a difference, $y_{i+1} - y_i$. Simply minimize the residual over the constants $a, b,$ and c, as above. For the 3000 meter men's race, this direct integration of the differential equation gives

$$-0.0086(y-7.06)(9.18-y)$$

with residual 0.04. These asymptotes compare favorably with 8.9 and 7.25 which come from optimizing the logistic fit model. The approximation is not that good, but it does reveal a basic underlying logistic structure. One problem with the differential and integral methods is that derivatives or integrals of must be approximated, and this can generate a substantial error. Two methods for computing the right-side were used, a Taylor expansion approximation and the quadratic interpolation of each successive triple of data values. Different asymptotes resulted, though both determined potential asymptotes. For other data samples, the "asymptotes" became complex, indicating little or no logistic behavior.

Normally, we use the first method over either the differential or integrals method, as it better smooths data, avoiding the difficulties of computing accurately derivatives over rather large x-axis spans, and integral approximation errors. As well, note none of the methods are affected by normalizing the data as we are computing with high precision

The principle point of these "tests" is that the data often does not reveal even a trace of a logistic nature. Indeed, the data fits well in many cases to a linear model, from which no asymptotes are possible. Fitting the data to a quadratic model can reveal a lower asymptote, but there seems little more than a phenomenalistic motivation for so doing.

VI. The Predictions.

We consider several races and give predictions based on the four models, plus the difference method. One important note is that times are given in seconds and minutes. For these, most improvements are given in seconds even when the data is given in minutes. Only for the Marathon are basic units in hours, and for these improvements are given in minutes. In all other cases, the record times are in minutes or seconds, but the gain is always reported in seconds. In some cases, note the predicted value is negative. This means the predicted record was greater than the current record. Though rare, it does happen, and usually, implies data with a variance against the imposed logistic nature. Overall, inflection points (the year) are relatively distant in the past.

\NOTE: To obtain this data, go to Wikipedia and search xxx meter progressions or go to https://www.iaaf.org/records/toplists/sprints/100-metres/outdoor/men/senior

A few other points require note. For the integral test, it is mentioned only whether the roots are real, indicating a logistic nature to the raw data, or complex, indicating a non-logistic nature to the data. It is important to note, the logistic models, here and in other papers, use four parameters, and there is the best fit even for non-logistic data. Note also two rows pertain to the logistic method. The first is the more pure form, where the optimization seeks all four parameters simultaneously. Here the results can appear wrong. Notice also for some methods the time to gain is negative. This indicates the model predicts a "best" time greater than the current record time. It happens. For the optimizations, the same starting values were used for all methods.

The first set of records, Table 6, is about the most renown of all races, the Marathon. For the men's data, since the record dropped so precipitously in the early 20th century, we have deleted the first ten records.

Method	Race	*n*	Predicted Record	Current Record	Minutes to gain	Speed - Current (m/s)	Speed - Predicted (m/s)	Upper Asymptote	Inflection	Residuals
Logistic	Men's Marathon	37	2.064	2.049	-0.897	5.720	5.678	2.587	1953.0	0.000
Logit_alt		37	1.952	2.049	5.847	5.720	6.005	2.674	1955.0	0.001
Arctangent		37	1.973	2.049	4.572	5.720	5.941	2.645	1955.3	0.016
Sinusoid		37	2.043	2.049	0.360	5.720	5.737	2.561	1956.6	0.034
Errror fcn		37	2.039	2.049	0.587	5.720	5.747	2.566	1956.5	0.024
Men's Marathon --- Difference method prediction 2.0285. Minutes to gain 1.239. Integral test - real roots										
Logistic	Womens Marathon	37	2.087	2.257	10.23	5.193	5.617	22.530	1795.9	0.064
Logit_alt		37	2.178	2.257	4.74	5.193	5.382	3.803	1971.4	0.006
Arctangent		37	2.175	2.257	4.92	5.193	5.389	3.778	1971.3	0.110
Sinusoid		37	2.171	2.257	5.14	5.193	5.398	3.752	1971.7	0.239
Errror fcn		37	2.270	2.257	-0.75	5.193	5.165	3.638	1971.9	0.146
Women's Marathon --- Difference method prediction 2.2160. Minutes to gain 2.45. Integral test - real roots										

Table 6 – Predictions for the Marathon

For the 100-meter race, results in Table 7 are mostly consistent, but for the 100-meter women's sprint, the integral test fails. Indeed, the data is terrible, almost linear, from a logistic viewpoint. However, almost all the predictions reveal there is very little to gain for all the methods. Clearly, the logistic method produces an outlier.

As in all the tables, important features to note are the residuals and the consistency of the inflection points. It would be possible to average the gains, but difficult to interpret what it might mean. For the 200-meter race, Table 8 note the arctangent method simply gives completely unrealistic results for the men. The results for the women seem much more consistent among the methods. For the 400-meter race, results seem to sustain a pattern, Table 9. When projecting with the full four parameter logistic model, odd results obtain. However, the two-stage optimization appears to give more consistent results. This is sustained in Table 10 for the 800-meter race. Table 11 shows the results for the 1500-meter race, which is akin to the mile race. These results are the most consistent across all methods, but note relatively wide variations in the "second to gain" column. In Table 12, for the 10,000-meter race, reveals there is much time for improvement. Finally, in Table 13, in the 5,000-meter race, the results show that for men the integral test fails but passes for the women. In most cases, the date of inflection for men precedes that for the women, indicating men have been competitively somewhat longer than women.

It seems the longer distances yield more stable and consistent results. However, note that some races seem complete, in the sense that predictions are only slightly smaller than extant records, while the longer races give much more latitude, even considering percentages. As already indicated, when the inflection points vary, this is a signal that results may be unreliable. In sum, the model is the model. Listening to what it informs us is important to reflect on the model itself. All models, viz-viz, how the logistic driver approaches the asymptotes have been considered, the classical logistic method with finite slopes, the arctangent model with zero slopes, and the sinusoid model with infinite slopes.

Method	Race	*n*	Predicted Record	Current Record	Seconds to gain	Speed - Current (m/s)	Speed - Predicted (m/s)	Upper Asymptote	Inflection	Residuals
100 meter men --- Difference method prediction 9.4462 Seconds to gain 0.1338										
logistic	100 meter men	78	9.5495	9.58	0.0305	10.438	10.472	12.5529	1881.7	0.0038
logit_alt		78	9.1023	9.58	0.4777	10.438	10.986	11.1318	1956.0	0.005
arctangent		78	9.3246	9.58	0.2554	10.438	10.724	10.9049	1955.8	0.3625
Sinusoid		78	9.4874	9.58	0.0926	10.438	10.540	10.7071	1958.4	0.3895
Errror fcn		78	9.4886	9.58	0.0914	10.438	10.539	10.7079	1958.3	0.3811
100 meter men --- Difference method prediction 9.4462 Seconds to gain 0.1338. Integral test - real roots										
logistic	100 meter women	42	5.6603	10.49	4.8297	9.533	17.667	32.0942	1124.4	0.2569
logit_alt		42	9.975	10.49	0.515	9.533	10.025	14.2891	1929.5	0.0749
arctangent		42	10.2774	10.49	0.2126	9.533	9.730	13.9399	1930.0	2.594
Sinusoid		42	10.4068	10.49	0.0832	9.533	9.609	13.7222	1931.4	3.2221
Errror fcn		42	10.4179	10.49	0.0721	9.533	9.599	13.717	1931.5	3.0703
100 meter women --- Difference method prediction 10.2663 Seconds to gain 0.2237. Integral test - complex roots										

Table 7 – Predictions for the 100-meter race

Method	Race	*n*	Predicted Record	Current Record	Seconds to gain	Speed - Current (m/s)	Speed - Predicted (m/s)	Upper Asymptote	Inflection	Residuals
Logistic	200 meter women	19	20.4939	21.83	1.3361	9.162	9.759	22.8722	1899.8	0.051
Logit_alt		19	20.7385	21.83	1.0915	9.162	9.644	22.407	2067.5	0.0096
Arctangent		19	2230.759	21.83	-2208.929	9.162	0.090	-2187.6183	2066.8	0.1816
Sinusoid		19	21.7035	21.83	0.1265	9.162	9.215	-15603.026	-13509.8	0.3261
Error fcn		19	15.7509	21.83	6.0791	9.162	12.698	40.1405	-1184.8	0.1816
200 meter women --- Difference method prediction 21.7004. Seconds to gain 0.1296. Integral test: real roots.										
Logistic	200 meter men	25	19.4512	19.19	-0.2612	10.422	10.282	20.728	1967.2	0.0188
Logit_alt		25	18.2501	19.19	0.9399	10.422	10.959	21.786	1975.3	0.0403
Arctangent		25	18.6231	19.19	0.5669	10.422	10.739	21.4278	1974.1	0.9143
Sinusoid		25	19.0644	19.19	0.1256	10.422	10.491	20.9414	1975.2	0.9727
Error fcn		25	19.0983	19.19	0.0917	10.422	10.472	20.9341	1974.0	0.8844
200 meter men --- Difference method prediction 18.9492. Seconds to gain 0.2408. Integral test: real roots.										

Table 8 – Predictions for the 200-meter race

Method	Race	*n*	Predicted Record	Current Record	Seconds to gain	Speed - Current (m/s)	Speed - Predicted (m/s)	Upper Asymptote	Inflection	Residuals
Logistic	400 meter men	24	42.6117	43.03	0.4183	9.296	9.387	48.498	1950.7	0.096
Logit_alt		24	40.9115	43.03	2.1185	9.296	9.777	50.180	1951.1	0.121
Arctangent		24	41.1719	43.03	1.8581	9.296	9.715	49.907	1951.1	2.410
Sinusoid		24	42.6882	43.03	0.3418	9.296	9.370	48.224	1952.5	2.400
Error fcn		24	42.6335	43.03	0.3965	9.296	9.382	48.268	1952.6	2.301
400 meter men --- Difference method prediction 42.5385. Seconds to gain 0.4915. Integral test: real roots										
Logistic	400 meter women	26	25.9144	47.6	21.6856	8.403	15.435	154.586	1829.7	1.320
Logit_alt		26	45.2233	47.6	2.3767	8.403	8.845	59.875	1966.2	0.605
Arctangent		26	43.9827	47.6	3.6173	8.403	9.095	61.242	1965.9	16.346
Sinusoid		26	47.1143	47.6	0.4857	8.403	8.490	57.612	1966.8	15.864
Error fcn		26	47.0837	47.6	0.5163	8.403	8.496	57.646	1966.8	16.253
400 meter women --- Difference method prediction 46.9541. Seconds to gain 0.6459. Integral test: complex roots										

Table 9 – Predictions for the 400-meter race

Method	Race	*n*	Predicted Record	Current Record	Seconds to gain	Speed - Current (m/s)	Speed - Predicted (m/s)	Upper Asymptote	Inflection	Residuals
Logistic	800 meter women	24	1.1861	1.9583	46.3361	6.809	11.242	7.173	1534.6	0.0048
Logit_alt		24	1.8611	1.9583	5.8345	6.809	7.164	2.6349	1950.0	0.0013
Arctangent		24	1.7787	1.9583	10.7797	6.809	7.496	2.7287	1949.0	0.0326
Sinusoid		24	1.9392	1.9583	1.1468	6.809	6.876	2.5351	1952.3	0.0311
Error fcn		24	1.9364	1.9583	1.3176	6.809	6.886	2.5387	1951.9	0.0321
800-meter-women --- Difference method prediction 1.9261. Seconds to gain 1.9350. Integral test: Complex roots										
Logistic	800 meter men	24	1.6718	1.6817	0.5942	7.929	7.976	1.9185	1944.9	0.0001
Logit_alt		24	1.5988	1.6817	4.9709	7.929	8.340	1.9588	1953.5	0.0002
Arctangent		24	1.6142	1.6817	4.0456	7.929	8.260	1.9421	1952.8	0.0031
Sinusoid		24	1.7064	1.6817	-1.4844	7.929	7.814	1.7896	1941.4	0.1560
Error fcn		24	1.6673	1.6817	0.8633	7.929	7.997	1.8838	1954.3	0.0031
800-meter-men --- Difference method prediction 1.6681. Seconds to gain 0.8124. Integral test: real roots										

Table 10 – Predictions for 800-meter race

Method	Race	*n*	Predicted Record	Current Record	Seconds to gain	Speed - Current (m/s)	Speed - Predicted (m/s)	Upper Asymptote	Inflection	Residuals
Logistic	1500 meter men	38	3.3730	3.4333	3.6225	7.282	7.412	4.0419	1950.2	0.0003
Logit_alt		38	3.2623	3.4333	10.2625	7.282	7.663	4.1267	1952.6	0.0004
Arctangent		38	3.2579	3.4333	10.5274	7.282	7.674	4.1312	1952.4	0.013
Sinusoid		38	3.4000	3.4333	1.9971	7.282	7.353	3.9697	1954.2	0.014
Error fcn		38	3.3995	3.4333	2.0325	7.282	7.354	3.9706	1954.0	0.0136
1500 meter race - men --- Difference method prediction 3.3894. Seconds to gain 2.6377. Integral test: real roots										
Logistic	1500 meter women	27	3.6347	3.8345	11.9873	6.520	6.878	6.2907	1936.0	0.0088
Logit_alt		27	3.6456	3.8345	11.3347	6.520	6.858	5.5705	1950.5	0.0091
Arctangent		27	3.4526	3.8345	22.9143	6.520	7.241	5.7826	1950.2	0.2757
Sinusoid		27	3.7830	3.8345	3.0899	6.520	6.609	5.3723	1951.9	0.2214
Error fcn		27	3.7861	3.8345	2.9021	6.520	6.603	5.3692	1951.9	0.2422
1500 meter women --- Difference method prediction 3.7860. Seconds to gain 2.9115. Integral test: real roots										

Table 11 – Predictions for the 1500-meter race

Method	Race	*n*	Predicted Record	Current Record	Seconds to gain	Speed - Current (m/s)	Speed - Predicted (m/s)	Upper Asymptote	Inflection	Residuals
Logistic	10000 meter men	37	26.1348	26.2922	9.4421	6.339	6.377	31.2337	1957.5	0.0321
Logit_alt		37	24.9823	26.2922	78.591	6.339	6.671	32.5246	1956.5	0.0378
Arctangent		37	25.0404	26.2922	75.1063	6.339	6.656	32.4514	1956.3	1.1396
Sinusoid		37	26.0413	26.2922	15.0521	6.339	6.400	31.2814	1958.3	1.3714
Error fcn		37	26.0387	26.2922	15.2063	6.339	6.401	31.2856	1958.1	1.2472
10000 meter men --- Difference method prediction 26.0542. Seconds to gain 14.2782. Integral test: real roots.										
Logistic	10000 meter women	37	26.1224	26.2922	10.1846	6.339	6.380	31.3007	1957.1	0.0336
Logit_alt		37	24.9836	26.2922	78.5145	6.339	6.671	32.5289	1956.7	0.0383
Arctangent		37	25.0278	26.2922	75.8628	6.339	6.659	32.4718	1956.4	1.1854
Sinusoid		37	26.0416	26.2922	15.0349	6.339	6.400	31.2867	1958.4	1.4152
Error fcn		37	26.0395	26.2922	15.1586	6.339	6.401	31.2904	1958.2	1.3024
10000 meter women --- Difference method prediction 26.0542 Seconds to gain 14.2782. Integral test: real roots										

Table 12 – Predictions for the 10,000-meter race

Method	Race	*n*	Predicted Record	Current Record	Seconds to gain	Speed - Current (m/s)	Speed - Predicted (m/s)	Upper Asymptote	Inflection	Residuals
Logistic	5000 meter women	22	14.380	14.186	-11.652	5.874	5.795	16.144	1981.9	0.0124
Logit_alt		22	13.492	14.186	41.623	5.874	6.176	17.109	1982.4	0.0299
Arctangent		22	13.768	14.186	25.082	5.874	6.053	16.810	1982.1	0.4522
Sinusoid		22	14.091	14.186	5.689	5.874	5.914	16.433	1983.1	0.6253
Error fcn		22	14.102	14.186	5.049	5.874	5.910	16.431	1982.8	0.5218
5000 meter women --- Difference method prediction 14.0007. Seconds to gain 11.1092. Integral test: real roots										
Logistic	5000 meter men	26	11.147	12.623	88.526	6.602	7.476	17.502	1917.6	0.0157
Logit_alt		26	11.992	12.623	37.818	6.602	6.949	14.548	1974.2	0.0068
Arctangent		26	11.835	12.623	47.249	6.602	7.041	14.712	1974.0	0.1823
Sinusoid		26	12.495	12.623	7.629	6.602	6.669	13.997	1975.2	0.1830
Error fcn		26	12.493	12.623	7.767	6.602	6.670	14.001	1975.0	0.1896
5000 meter men --- Difference method prediction 12.5025. Seconds to gain 7.1988. Integral test: complex roots										

Table 13 – Predictions for the 5000-meter race

VII. Conclusions.

In total, we have used seven methods to predict ultimate world records, the logistic, arctangent, sinusoid, and error function, plus the differential (not reported) and integrals methods, and finally, the simplistic different method. From the results and previous discussions, we see the prediction problem is rather difficult, and certainly not uniform. Not only as to which model to use, the race and therefore predictions themselves have evolved. In addition, the point plots of the data often do not even appear logistic, and many fail the logistic integral test, as well. Absent here is a collection of tables of record predictions using various weight schemes – to be reported later.

If a new event was created, say the 300-meter sprint, we would see a very rapid decline in records until a near optimal level was reached and most subsequent results would range only slightly smaller in race time.

The might be an exponential decline, ala Bertanffly curve adjusted to decrease, and thus having no distant past asymptote (Bertanffly, 1957). This is because training methods and most other factors already mentioned have stabilized in the past decade or so. The methods above may not work at all for predictions of best possible as there is no possibility of a logistic evolution of the records. This suggests the difference method or even statistical methods may be better.

Currently, the best we can give is a range of possible ultimate records are all viable, depending on methods used and of course the myriad of possible future-based effects. Predicting a future with numerous unknowns is at best risky.

References

[1]. Hutchison, J. R. and Garcia, M. (2002). Tyrannosaurus was not a fast runner. Nature, 415, 1018-1021.
[2]. Keller, J. B. (1973). A theory of competitive running. Phys. Today 26, 42-46.
[3]. McClusky, Mark. (2014) Faster, Higher, and Stronger, Faster, Higher, Stronger - How Sports Science is Creating a New Generation of Superathletes - and What We Can Learn from Them, Hudson Street Press.
[4]. Denny, Mark W., (2008) Limits to running speed in dogs, horses and humans,The Journal of Experimental Biology 211, 3836-3849 , doi:10.1242/jeb.024968.
[5]. Nevill, A. M. and Whyte, G. (2005). Are there limits to running world records? Med. Sci. Sports Exerc. 37, 1785-1788.).
[6]. Gembris, Daniel, Taylor, John G., Suter, Dieter, (2007), Evolution of Athletic Records: Statistical Effects versus Real Improvements, Journal of Applied Statistics, Vol. 34, No. 5, 529–545.
[7]. Gembris, D., Taylor, J. G. & Suter, D. (2002) Sports statistics-trends and random fluctuations in athletics, Nature, 417, p. 506.
[8]. Gumbel, E. J. (2004) [1958], Statistics of Extremes, Mineola, NY: Dover, ISBN 0-486-43604-7
[9]. Gnedenko, B.V. (1943), "Sur la distribution limiteduterme maximum d'uneseriealeatoire", Annals of Mathematics, 44: 423–453, doi:10.2307/1968974.
[10]. Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. London: Springer-Verlag.
[11]. Ericsson, K. Anders, Krampe, Ralf Th., Tesch-Romer, Clemens, (1993) The Role of Deliberate Practice in the Acquisition of Expert Performance Psychological Review,Vol. 100. No. 3, 363-406.
[12]. Taleb, Nassim Nicholas (2010) [2007], The Black Swan: the impact of the highly improbable (2nd ed.), London: Penguin, ISBN 978-0-14103459-1
[13]. Track and Field News (2015), online at https://www.trackandfieldnews.com/index.php/special-articles/817
[14]. Verhulst P. F. (1838) Notice sur la loi que la population suit dans son accroissement. Correspondance Mathématique et Physique Publiee par A. Quetelet, Brussels 10: 113-121.
[15]. Johnson, Norman L, Kotz, Balakrishnan, N. (1995), Continuous Univariate Distributions, volume 2, Wiley.
[16]. https://en.wikipedia.org/wiki/Beta_distribution
[17]. Richards F. J. 1959. A flexible growth function for empirical use. Journal of Experimental Botany 10: 290-300.
[18]. Gompertz B. (1825). On the nature of the function expressive of the law of human mortality, and on a new method of determining the value of life contingencies. Philosophical Transactions of the Royal Society 27: 513-585.
[19]. Sakanoue, Seiichi, (2007), Extended logistic model for growth of single-species populations, Ecological Modelling, 205, 159-168.
[20]. Allen, G. Donald, (2017), Hierarchy of Knowledge – from Data to Wisdom, International Journal of Current Research in Multidisciplinary (IJCRM), Vol 2, 1, 2017, 15-23.
[21]. The Inverse Error Function, online at http://mathworld.wolfram.com/InverseErf.html.
[22]. The Error Function, online at https://en.wikipedia.org/wiki/Error_function.
[23]. World track progression records, online at https://www.iaaf.org/records/toplists/sprints/100-metres/outdoor/men/senior
[24]. Bertalanffy L. V., 1957. Quantitative laws in metabolism and growth. The Quarterly Review of Biology 32: 217-231.

G. Donald Allen. "Modeling World Record Predictions in Track Events." IOSR Journal of Sports and Physical Education (IOSR-JSPE) 5.6 (2018): 06-22.