McNair Scholars Journal

Volume 19 | Issue 1 Article 9

2015

# Modeling the Development of World Records in Track and Field

Marie Griffith
*Grand Valley State University*

# Modeling the Development of World Records in Track and Field

**Marie Griffith**
*McNair Scholar*

**Paul Stephenson**
*Faculty Mentor*

## I. INTRODUCTION

Previous researchers have used various analytical techniques to develop a model of world records in track and field events. This can be a difficult task considering that there is not a set time frame in which world records are broken. Some records have been broken several times within one year, while others have taken multiple years only to move very slightly (as in one-hundredth of a second or one-fourth of an inch). It is assumed that there is a limit in human performance for track and field events since it seems unreasonable that a runner can run a race in zero seconds or that a high jumper reaches 100 feet. The purpose of this study is to assess the trends of world records in track and field events over time and to discover the possible threshold in which world records start to approach for the events of the 100 m, 200 m, 400 m, long jump, and shot put.

World records are those that have been officially ratified by the International Association of Athletics Federations (IAAF). If the ratified world record is wind-aided over an altitude of 1,000 meters in an event where air has a positive effect, the unofficial best time or measurement is sub-listed (World Records-Men, 2015). Specific criteria of submitted qualifying marks must be satisfied for ratification such as the dimensions of the track and equipment used must conform to IAAF standards, events must be accurately measured by a certified measurer, and running events up to 800 meters in distance are required to have a photo finish fully automatic timing. It is of interest to the researchers to explore the patterns, differences, and influential factors of records in numerous events over time. It is assumed that there will be a significant improvement in the world records from the starting years of the sport to the most current years. The eventual goal of our research is to find the limit values that these times and measurements appear to be approaching and witness if there seems to be asymptotic convergence along these thresholds.

## II. LITERATURE REVIEW

To comprehend the changes in speed of athletes based off of various distances with respect to time, "Models for Comparing Athletic Performances" was reviewed for modeling that is fitted to 1996 world records for various distances (Grubb, 1998). A developed three parameter model including speed at long distances, the maximum speed over distance obtained, and decrease in speed with respect to distance was used to compare performances by the same athlete at different distances to determine an athlete's strengths and training effects (Grubb, 1998). This model was constructed given that every athlete is built differently, has different strengths, and different endurance levels. After data was collected, the composed model showed the decrease in world records over several years and predicted lower bounds on these records using various parametric forms. Given the fact that it is believed that there is a threshold for every track and field world record and that there will be asymptotic convergence above these thresholds, this was alluring information.

Since it is favorable to develop a model that displays asymptotic convergence, an article titled "Are there limits to running world records?" was examined in which world record breaking for men and women was assessed for modeling techniques (Nevill and Whyte, 2005). Instead of using a linear model, it was decided to use the "S-shaped" logistic curve to provide a better fit for world record data overtime. In this study, it was identified that there was a "slow rise" in world record speeds at the beginning of the century. This was followed by an acceleration period, which showed a major increase in speeds and decrease in times due to the enhancement of the sport and new technology. At the end of the century, there appeared to be a reduction in record breaking performances due to the challenges of beating previous times now that the sport is more advanced (Nevill and Whyte, 2005). Their model also hinted towards men's world records reaching an asymptotic limit and that it is highly unlikely that women's world records will reach those achieved by men.

After reviewing the two previous sources, the idea of asymptotic convergence and curve fitting techniques seemed appealing. In another literature source, "Modelling the Development of World Records in Running," the authors demonstrated the progression of world records for metric running events from the 100 meter dash to the marathon for men and women using time-series curves (Kuper and Sterken, 2008). Multiple curve fittings were examined including the linear model, exponential curve, logistic curve, and the Gompertz curve to identify the "best fit" for modeling world record data. The Gompertz-curve was ultimately chosen for modeling. This curve was used for sixteen events and implied limit values were computed for each event. After thresholds were determined, a log-log model was used and determined that if the distance increased by 10%, the limit value increased by 11%, not necessarily taking gender into account (Kuper and Sterken, 2008). Based off of this modeling, the researchers chose the Gompertz curve for analysis due to validation of the curve choice and the high $R^2$ values that were produced from their computed threshold limits. This study influenced the researchers to develop a model that compares using similar methodology to see the differences in limit value predictions.

## III. RESEARCH DESIGN AND METHODS

The purpose of this research is to model the development of world records in track and field using statistical analysis. Throughout this process, raw data was obtained by records listed on the official websites of the Olympic Movement and IAAF organizations (iaaf.org and Olympic.org). Only officially documented records from these two organizations were used for analysis. The analysis of this data was conducted using the JMP statistical software package that produced a graphical representation for various events using a nonlinear curve fitting technique. A model was also produced to predict the thresholds of world records in various track and field events.

Initially, raw data on world records was acquired from the first officially documented world record to the most current on the following events: 100 m 200 m, 400 m, long jump, and shot put. The data collected included the following variables: Name, Time or Measurement, Record Year, and *T*. Record Year represents the year in which a world record was broken and *T* represents the current year (that includes every consecutive year from the first documented world record for that event until the year 2014). The data was then imported into JMP for analysis.

Next, various curves were reviewed based off of previous literature sources in an effort to identify the most appropriate way to model the development of world records in track and field. While a linear model was first examined, it was evident that the data did not follow a linear form because the world records were not constantly being broken at a steady rate. Because of this, the linear trend was declared as not a good fit for this type of modeling. The exponential curve was the next model explored for fitting the data. This curve was also problematic because the events did not produce a trend that is constantly decreasing or increasing until it reaches an asymptote.

The next type of curve that was reviewed was the logistic curve. The shape of this curve was appealing to this particular type of modeling because it has the potential to form an "S" pattern when graphed. This shape is favorable for world record data because there appeared to be a slow change in times/measurements in the beginning stages of the sport, followed by a more rapid change during its development stages as the sport progressed, and leveling off of records in the later years once the sport started getting perfected. Although the logistic curve seemed reasonable for modeling, it was identified that the Gompertz curve was the model of "best fit" for predicting the threshold limits. The form of the Gompertz curve facilitates the prediction of thresholds because it asymptotically approaches the upper bound and unlike the logistic curve, it is not symmetric about the inflection point (Kuper and Sterken, 2008).

After the Gompertz curve was selected, a four parameter growth model was produced to represent the rate of change of world records with respect to time. This model was then utilized on the data for five events previously mentioned (100 m 200 m, and 400 m, long jump and shot put). The Gompertz curve growth model is given by:

$$\mathbf{Y}_T = a + (b - a)e^{-e^{-c(T-d)}}$$

The lower asymptote (slowest time or measurement) is represented by *a*. The upper asymptote (fastest time or measurement) is represented by *b*. The growth rate of the times and measurements is denoted by *c*. The point of inflection of the produced curve is represented by *d*. The last component *T* represents the current year.

The implied limit values are computed as such:

$$\lim_{T\to\infty} \mathrm{Y}_t = a + (b - a)e^{-e^{-c(T-d)}}$$
$$\lim_{T\to\infty} \mathrm{Y}_t = a + (b - a)e^{-\frac{1}{e^{c(T-d)}}}$$
$$\lim_{T\to\infty} \mathrm{Y}_t = a + (b - a)e^{-\frac{1}{\infty}}$$
$$\lim_{T\to\infty} \mathrm{Y}_t = a + (b - a)e^{-0}$$
$$\lim_{T\to\infty} \mathrm{Y}_t = a + (b - a)1$$
$$\lim_{T\to\infty} \mathrm{Y}_t = a + (b - a)$$
$$\lim_{T\to\infty} \mathrm{Y}_t = \mathrm{b}$$

Graphical output of the Gompertz curve and implied limit values were produced for the selected events and reviewed to ascertain the reasonableness of the fit. Produced $R^2$ values were used to assess how much variation is explained by the model and whether the limit values were reasonable estimates.

## IV. RESULTS

The Gompertz curve was used to compute limit values for the following events: 100 m 200 m, 400 m, long jump, and shot put. The computed thresholds and their associated $R^2$ values for each event tested are listed in Table 1.

**Table 1: Computed Limit Values**

| *Event* | *Men* | $R^2$ *(men)* | *Women* | $R^2$ *(women)* |
|---|---|---|---|---|
| 100 m | 9.36 | .9668 | 10.06 | .9513 |
| 200 m | 18.69 | .9325 | 21.32 | .9894 |
| 400 m | 40.58 | .9516 | 47.54 | .9549 |
| Long Jump | 29 ft 5.15 in | .9511 | 27 ft 0.41 in | .9552 |
| Shot Put | 77 ft 8.28 in | .9840 | 76 ft 9.65 in | .9636 |

World record data on the five tested events for men and women were fitted through the Gompertz curve for modeling. The figures below show the data for each event that is fitted through the Gompertz curve from the first world record until the year 2014.

**Figure 1: Men's 100 m**

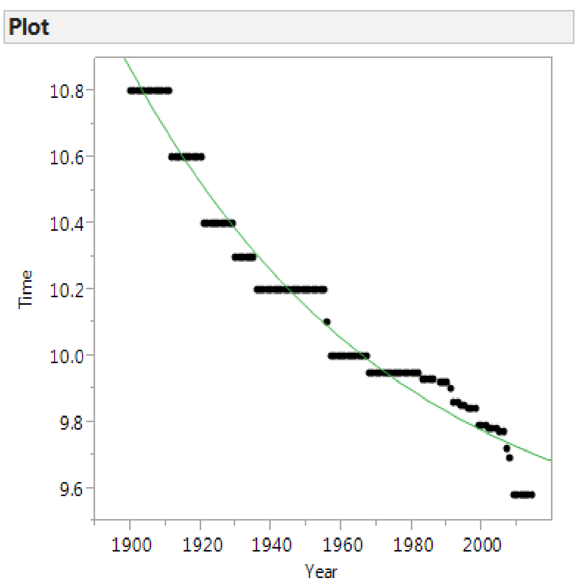



**Figure 2: Women's 100 m**

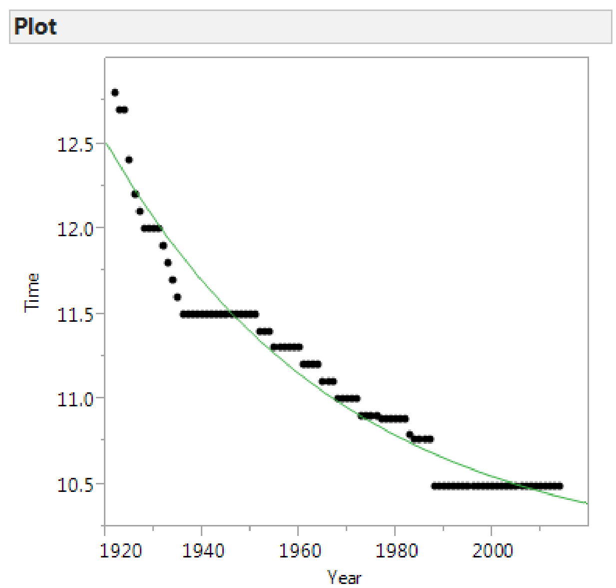



**Figure 3: Men's 200 m**

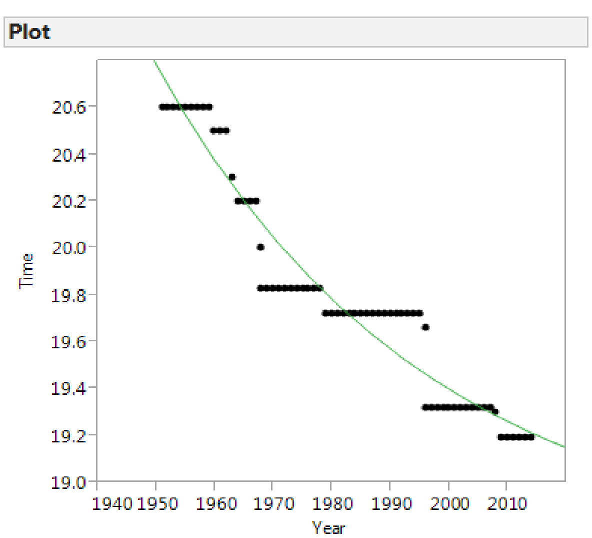



**Figure 4: Women's 200 m**

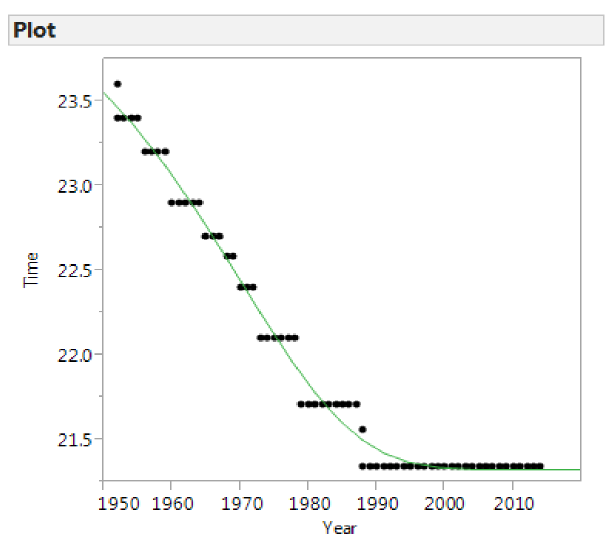


**Figure 5: Men's 400 m**

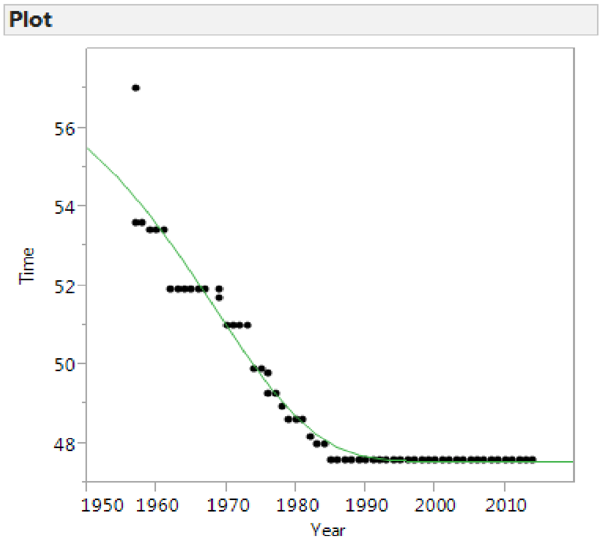



**Figure 6: Women's 400 m**

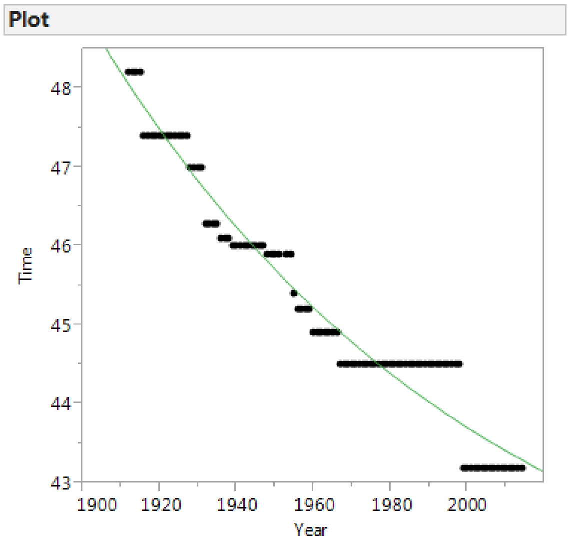



**Figure 7: Men's Long Jump**

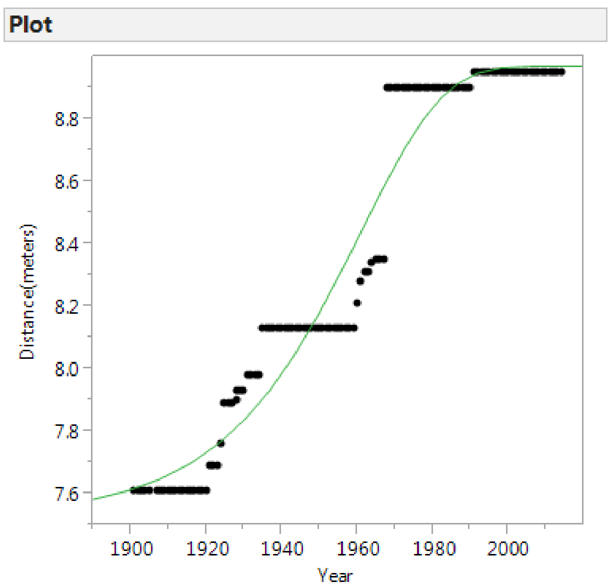



**Figure 8: Women's Long Jump**

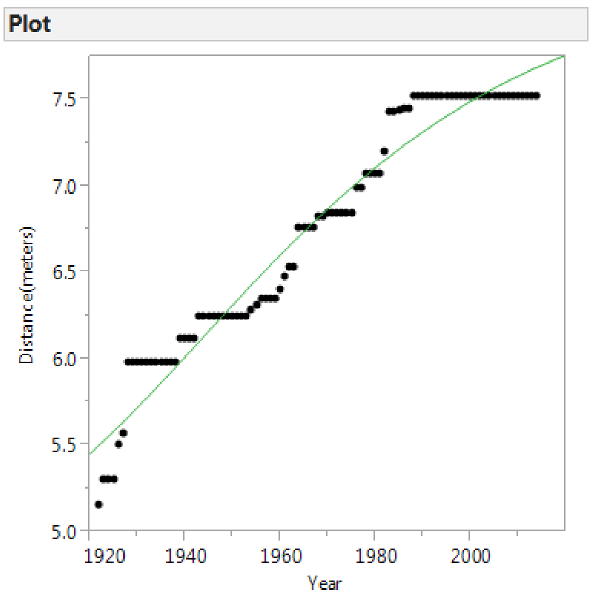


**Figure 9: Men's Shot Put**

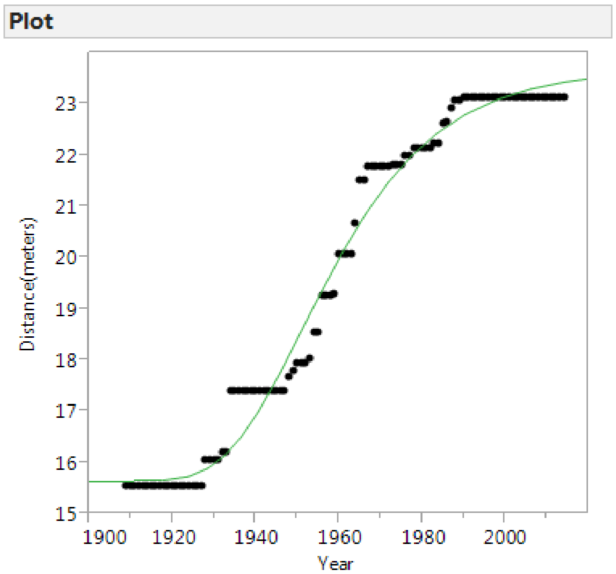
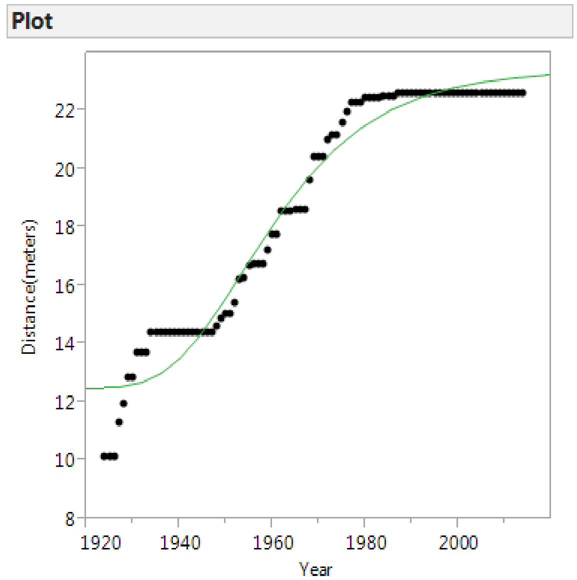



**Figure 10: Women's Shot Put**




## V. DISCUSSION

Data was analyzed by the Gompertz model and computed limit values were produced indicated by the upper asymptote. These computed limit values were checked visually by the produced curve to see how well it fit the data. $R^2$ values were then evaluated to check whether there is a large variation of the data that is explained by the model, which could indicate that these computed limit values were "reasonable." All of the produced $R^2$ values were between 0.9 and 1 which proved a very strong amount of variation that was explained by the Gompertz model. These $R^2$ values and curve fittings demonstrated that the Gompertz curve was a sufficient choice for modeling world record data and that the computed limit values were reasonable estimates.

Sensitivity analysis of three scenarios were then conducted on the men's and women's 100 M to demonstrate if these implied limit values will hold over time. These scenarios included extending the current world record by one hundred years, deleting the current world record holder's time and using the previously documented world record for computation of limits, and examining the steroid effect for the men's data by adding Ben Johnson's previous world record time of 9.79 seconds done in 1988. The computed limit values from the sensitivity analysis are represented in Table 2.

**Table 2: Computed Limit Values Based on Sensitivity Analysis**

| *Scenario* | *Men* | *$R^2$ (men)* | *Women* | *$R^2$ (women)* |
|---|---|---|---|---|
| **Scenario 1:** 100 M Current world record extended 100 years | 9.55 | .9791 | 10.47 | .9648 |
| **Scenario 2:** Current world record deleted | 9.57 | .9767 | 10.65 | .9559 |
| **Scenario 3:** Steroid use | 9.28 | .9719 | *** | *** |

In scenario 1, the researchers extended the current world record for the 100 M as of the year 2015 for one hundred years for both men and women to see its effects on the computed limit values. The current world record for men is 9.58 seconds by Usain Bolt and 10.49 seconds by Florence Griffith-Joyner for the women.

As stated in Table 2, if Usain Bolt's record of 9.58 seconds for the 100 M (set in 2009) stood for one hundred years, the predicted threshold would change from 9.36 to 9.55 seconds. For the women, if Florence Griffith-Joyner's record of 10.49 seconds (set in 1988) lasted for one hundred years as well, the predicted threshold would change from 10.06 to 10.47 seconds. This scenario demonstrates that the longer a current world record holds, the more the threshold limit gravitates toward that record. The model will keep readjusting itself over time to get closer to that time/measurement unless that record is broken again.

**Figure 11: Men's 100 M Current World Record Extended 100 Years**

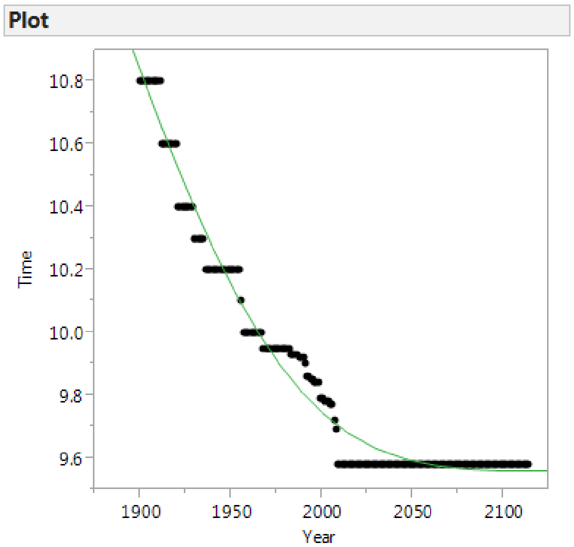
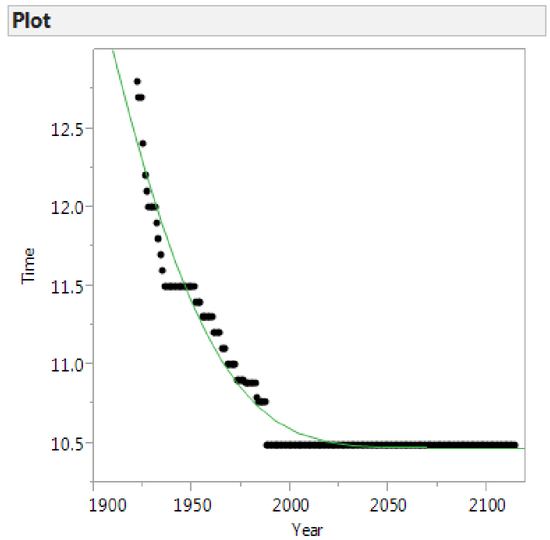



**Figure 12: Women's 100 M Current World Record Extended 100 Years**




In scenario 2, there is a special case for the men's 100 meters since Usain Bolt has both the current and previous world records with the times of 9.69 and 9.58 seconds. For this analysis, it was decided to test the effects if Usain Bolt did not exist at all as a world record holder. The last documented world record holder before Usain Bolt is Asafa Powell with a record of 9.72 seconds (set in 2007). For the women, if Florence Griffith-Joyner had not been a world record holder, the previously documented record holder would be Evelyn Ashford with a time of 10.76 seconds (set in 1984).

**Figure 13: Men's 100 m without Current World Record Usain Bolt**

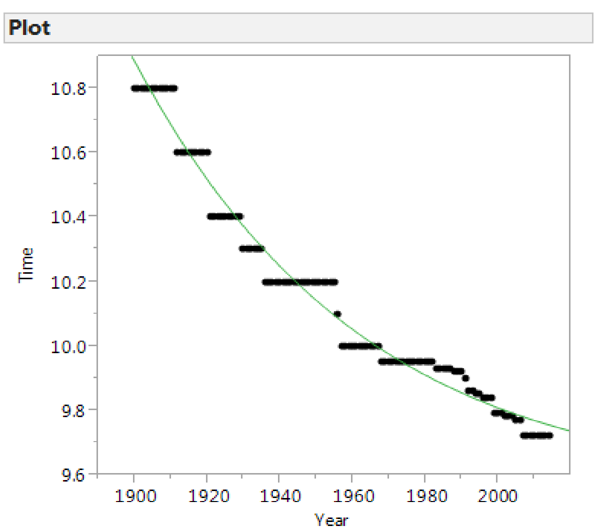
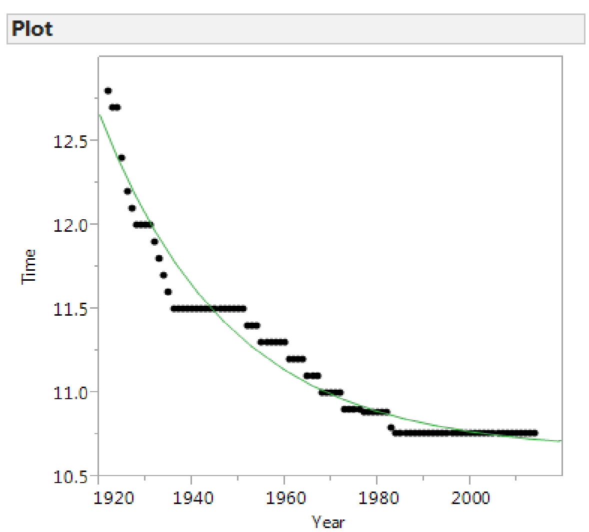


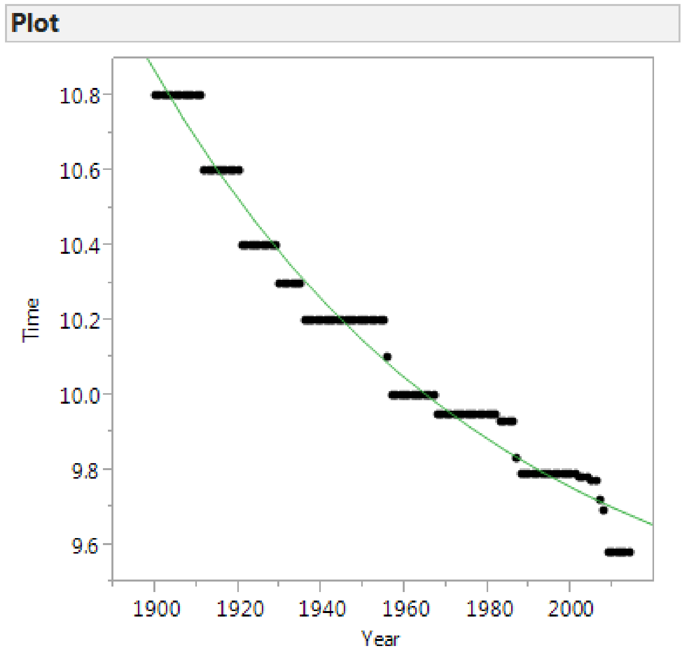

**Figure 14: Women's 100 m without Current World Record Florence Griffith - Joyner**




Without the times of Usain Bolt and Florence Griffith-Joyner in the data used to predict the limit values, the limits would readjust and the new computed thresholds for the 100 M would be 9.57 seconds for the men and 10.65 seconds for the women (Table 2).

In the third and final scenario conducted using sensitivity analysis, steroid use was infused into the data to see its effects on the thresholds. Since this research deals strictly with world record data, the women were not used for this scenario. The only clear indication of steroid use in track and field is whether an athlete's world record is removed due to positive detection of steroids. There has not been any documented indication of steroid use for women involving world records. However, for the men's data Canadian runner Ben Johnson's world record of 9.79 seconds for the 100 M (set in 1988) that was stripped away due to the positive testing for anabolic steroids, was added to the data for new limit value computation (Cashmore, 2010).

**Figure 15: Men's 100 m with Ben Johnson's Record**




The inclusion of Ben Johnson's record appeared to have an effect on the threshold limit but not the curve itself. The Gompertz curve readjusted itself allowing for more performance improvement of future athletes. The limit valuedecreased from the originally predicted 9.36 to 9.28 seconds (Table 2).

After computing these limit values using both the original and altered data, it was found that there can be a wide range of computed limit values for track and field events. While in theory a definite limit value for track and field events exists, there is a level of uncertainty embedded in the Gompertz model since human performance fluctuates over time. Due to this uncertainty in human performance, thresholds for track and field events can change over time as well. This uncertainty in performance can also affect the prediction of when threshold limits will be met. Given the high $R^2$ values and curve fitting from the acquired data, it is believed that the computed limit values listed in Table 1 are reasonable estimates. However, due to this unpredictability in human performance, which was proven through sensitivity analysis in Table 2, it is shown that the original thresholds predicted by the researchers are not definite.

In summary, assuming that there is a limit in human performance, the researchers used the Gompertz curve as the chosen analytical technique to develop a model of world records over time and predict thresholds. Thresholds were computed using a four parameter growth model for the five selected events (100 M, 200 M, 400 M, long jump, and shot put) for both men and women. Given that the curve appeared to fit the data for these events quite well and all of the produced $R^2$ values were over 0.9, the researchers assumed that the predicted limit values were reasonable (Table 1). However, after sensitivity analysis was completed on the performance of the model, fluctuation in human performance proved a level of uncertainty is embedded in the model process (Table 2). Therefore, predicted threshold limits for track and field events are not definite and can vary over time.

## References


Cashmore, Ellis (2005). "Making Sense of Sports." *Google Books*. Taylor & Francis. Web. 31 July 2015.

Grubb, H.J. (1998). "Models for comparing athletic performances." *The Statistician 47,* 509-521.

iaaf.org,. 'IAAF: Results | Iaaf.Org'. N.p., 2015. Web. 15 Nov. 2015.

Kuper, G and E. Sterken (2008). "Modelling the development of world records in running." *Statistical Thinking In Sports.* (pp 7-31). Boca Raton, FL: Taylor & Francis Group, LLC

Nevill, A.M and G. Whyte (2005). "Are there limits to running world records?" *Medicine and Science in Sports and Exercise 37,* 1785-1788.

Olympic.org,. 'Athletics | Olympic Marathon, Sprint, Decathlon Videos, Medals'. N.p., 2015. Web. 15 Nov. 2015 "World Records-Men." *Men's Outdoor World Records*. Web. 31 July 2015.