# Ultimate Olympics Records in Athletics Using Extreme Value Theory 

Fumio Maruyama<br>Department of Sports and Health Science, Matsumoto University, Matsumoto, Japan<br>Email: fmaruya@nagoya-u.jp

How to cite this paper: Maruyama, F. (2022) Ultimate Olympics Records in Athletics Using Extreme Value Theory. Open Journal of Applied Sciences, 12, 541-554.
https://doi.org/10.4236/ojapps.2022.124038

Received: March 13, 2022
Accepted: April 26, 2022
Published: April 29, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/


#### Abstract

Extreme value theory provides methods to analyze the most extreme parts of data. We used the generalized extreme value (GEV) distribution to predict the ultimate $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}, 4 \times 100 \mathrm{~m}$ relay, and long jump records of male gold medalists at the Olympics. The diagnostic plots, which assessed the accuracy of the GEV model, were fitted to all event records, validating the model. The $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}, 4 \times 100 \mathrm{~m}$, and long jump records had negative shape parameters and calculated upper limits of $9.58 \mathrm{~s}, 19.18 \mathrm{~s}, 42.97 \mathrm{~s}$, 36.71 s , and 9.03 m , respectively. The calculated upper limit in the 100 m (9.58 s) was equal to the record of Usain Bolt (August 16, 2009). The 100 m and 200 m world records were close to the calculated upper limits, and achieving the calculated limit was difficult. The 400 m and $4 \times 100 \mathrm{~m}$ relay world records were almost equal to the calculated upper limits and the 500-year return level estimate, and slight improvement was possible in both. At the Tokyo Olympics in August 2021, in the $100 \mathrm{~m}, 200 \mathrm{~m}$, and $4 \times 100 \mathrm{~m}$, in one year the probability of occurrence for a record was about $1 / 30$. In the 400 m and long jump, it was about $1 / 20$. In the $100 \mathrm{~m}, 200 \mathrm{~m}$, and $4 \times 100 \mathrm{~m}$ relay, more difficult records show that a fierce battle has taken place.


## Keywords

Athletics, 100 m, 200 m, $400 \mathrm{~m}, 4 \times 100 \mathrm{~m}$ Relay, Long Jump, Extreme Value Theory, GEV Model

## 1. Introduction

Extreme value theory (EVT) has emerged as one of the most important statistical disciplines in applied science. Extreme value techniques are also widely used in other disciplines, such as financial market risk assessment and telecommunications traffic prediction [1]. EVT deals with statistical problems concerning the
far tail of the probability distribution and is unique as a statistical tool since it develops models and techniques to describe the unusual event rather than the usual. Using EVT, the theoretical distribution and its population parameter that the maximum value follows are estimated from long-term observation data. Additionally, the maximum or large value that occurs every period can be predicted based on the calculated results. For instance, EVT was developed in the 1920s and has been used to predict events such as droughts and flooding [2] or financial risk [3]. The application of extreme value modeling has been used in the fields of biomedical data processing [4], thermodynamics of earthquakes [5], and public health [6].

In this paper, we focus on popular events in athletics: the $100 \mathrm{~m}, 200 \mathrm{~m}, 400$ $\mathrm{m}, 4 \times 100 \mathrm{~m}$ relay, and long jump for men at the Olympics. The Olympic Games were reintroduced in 1896 by Pierre de Coubertin. World record collection shows the progression of human performance as elite athletes. In the 100 m , Usain Bolt (Jamaica) is the first of men: 9.58 s (August 16, 2009). In the 200 m , Usain Bolt (Jamaica) is the first of men: 19.19 s (August 20, 2009). In the 400 m , Wayde Van Niekerk (Republic of South Africa) is the first of men: 43.03 s (August 15, 2016). In the $4 \times 100 \mathrm{~m}$ relay, Jamaica is the first of men: 36.84 s (August 11, 2012). In the long jump, Mike Powell is the first of men: 8.95 m (August 30, 1991). The ultimate records in athletics were calculated using extreme value theory [7] [8] [9]. This study predicts the $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}, 4 \times 100 \mathrm{~m}$ relay, and long jump records for men at the Olympics using the extreme value theory.

## 2. Data and Method of Analysis

### 2.1. Data

We used the men's $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}, 4 \times 100 \mathrm{~m}$ relay, and long jump records of gold medalists for men at the Olympics [10].

### 2.2. Method

EVT concerns with phenomena of extreme data. We used the block maxima method. A method for modeling the extremes of a stationary time series is block maxima, in which consecutive observations are grouped into non-overlapping blocks of length $n$, generating a series of $m$ block maxima, $\mathrm{Mn}, 1, \ldots, \mathrm{Mn}, \mathrm{m}$, to which the Generalized Extreme Value (GEV) distribution can be fitted for some large value of $n$. The usual approach considers blocks of a given time length, thus yielding maxima at regular intervals [1]. Here a block was considered as a year, i.e., annual maxima values were used. Although the block maxima method is suitable for analyzing maximum value data, it has the disadvantage of being easily affected by one realization value and having a large variance in the estimator.

When data are taken to be the maxima (or minima) over certain blocks of time (such as annual maximum precipitation), it is appropriate to use the GEV distribution:

$$
G(z)=\left\{\begin{array}{l}
\exp \left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1 / \xi}\right\}, \xi \neq 0,  \tag{1}\\
\exp \left\{-\exp \left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\}, \xi=0
\end{array}\right.
$$

where $z$ are extreme values from blocks, $\mu$ a location parameter, $\sigma$ a scale parameter, and $\xi$ a shape parameter. $G(z)$ is defined for all z such that $(1+\xi(z-$ $\mu) / \sigma)>0$ for $\xi \neq 0$, and all $z$ for $\xi=0$. Three families of GEV distributions are defined depending on the value of $\xi$. We get the Fréchet distribution with a heavy tail for $\xi>0$, the Gumbel distribution with a lighter tail for $\xi=0$, and the Weibull distribution with a finite tail for $\xi<0$.

We want to know how small the value is as the fastest speed, hence, it is necessary to multiply the $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}$, and $4 \times 100 \mathrm{~m}$ relay data by -1 to put it in the framework of extreme value statistics that consider the maximum.

If a GEV distribution is fitted to observations, it becomes possible to estimate the probability of an event that has not yet been observed. Estimates of extreme quantiles of the annual maximum distribution are obtained by inverting Equation (1):

$$
z_{p}=\left\{\begin{array}{l}
\mu-\frac{\sigma}{\xi}\left[1-\{-\log (1-p)\}^{-\xi}\right], \xi \neq 0  \tag{2}\\
\mu-\sigma \log \{-\log (1-p)\}, \xi=0
\end{array}\right.
$$

where $G\left(z_{p}\right)=1-p . z_{p}$ is the return level associated with the return period $1 / p$, since $z_{p}$ is expected to be exceeded on average once every $1 / p$ year with a reasonable degree of accuracy. More accurately, $\mathrm{z}_{\mathrm{p}}$ is exceeded by the annual maximum in any particular year with probability $p$ [1].

Modeling was performed using the evd package in R for the GEV calculations. We also tried a non-stationary model in the GEV, but it did not work.

## 3. Results

The men's 100 m and 200 m records of gold medalists at the Olympics are shown in Figure 1. The change in the 100 m was small and that in the 200 m was large. The 200 m record has a larger decrease and a larger growth margin. The records of the men's 400 m and $4 \times 100 \mathrm{~m}$ relay of gold medalists at the Olympics are shown in Figure 2. The change in the $4 \times 100 \mathrm{~m}$ relay was small and that at 400 m was large. The 400 m record has a larger decrease and a larger growth margin. The records of the men's long jump of gold medalists at the Olympics are shown in Figure 3.

### 3.1. 100 m Run

The GEV parameter estimates, which were the results of the GEV modeling on the 100 m records of male gold medalists using the block maxima method, are shown in Table 1. The GEV parameters were estimated using the maximum


Figure 1. Plot of the 100 m and 200 m records of male gold medalists.


Figure 2. Plot of the 400 m and $4 \times 100 \mathrm{~m}$ relay records of male gold medalists.


Figure 3. Plot of the long jump records of male gold medalists.

Table 1. GEV parameter estimates in the 100 m records of male gold medalists.

|  | $\mu$ | $\sigma$ | $\xi$ |
| :---: | :---: | :---: | :---: |
| Parameter estimate | 10.2 | 0.380 | -0.571 |
| Standard error | 0.0870 | 0.0709 | 0.162 |
| $95 \% \mathrm{CI}$ | $(10.1,10.4)$ | $(0.242,0.519)$ | $(-0.887,-0.254)$ |

likelihood estimation (MLE). The model has three parameters: location parameter, $\mu$; scale parameter, $\sigma$, and shape parameter, $\xi$. Because $\xi$ was negative, the 100 m records of male gold medalists had a finite upper limit.

Estimates of the maximum return levels for the return periods of $10,20,50$, 100 , and 500 years along with their respective $95 \%$ confidence intervals, CI, are shown in Table 2. We estimated the 10-year return level to be 10.03 s , with a $95 \%$ CI $(9.88,10.18)$. We estimated the 100 -year return level to be 9.69 s , with a $95 \%$ CI $(9.60,9.78)$. We explain it differently, so it means that there is approximately a $1 \%$ chance $(1 / 100)$ each year that the 100 m record will not exceed 9.69 s. There is approximately a $10 \%$ chance $(1 / 10)$ each year that the 100 m records will not exceed 10.03 s .

The diagnostic plots for assessing the accuracy of the GEV model fitted to the 100 m records for men are shown in Figure 4. Straight lines and curves in solid lines represent estimated functions. Each point plot and short-dashed line is a realization value. The points on both sides represent the $95 \%$ CI. Probability and quantile plots show the validity of the proposed model: each set of points follows a near-linear behavior. The corresponding density estimate is consistent with the data. The estimated curve is nonlinear in the return level curve due to the negative $\xi$. Consequently, the diagnostic plots supported the fitted GEV model.


Figure 4. Diagnostic plots for GEV fit to the 100 m records of male gold medalists.

Table 2. GEV return level estimates in the 100 m records of male gold medalists.

| Return period <br> (years) | 10 | 20 | 50 | 100 | 500 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Return level (s) | 10.03 | 9.86 | 9.74 | 9.69 | 9.62 |
| Standard error | 0.0772 | 0.0638 | 0.0491 | 0.0457 | 0.0568 |
| $95 \%$ CI | $(9.88,10.18)$ | $(9.74,9.99)$ | $(9.64,9.84)$ | $(9.60,9.78)$ | $(9.51,9.73)$ |

### 3.2. 200 m Run

Estimates of the GEV parameters, which were the results of the GEV modeling on the 200 m records using the block maxima method, are shown in Table 3. Because $\xi$ was negative, the 200 m records of male gold medalists had a finite upper limit. The maximum return levels are shown in Table 4. The diagnostic plots for assessing the accuracy of the GEV model fitted to the 200 m records are shown in Figure 5. The estimated curve is nonlinear in the return level curve due to the negative $\xi$. The diagnostic plots supported the fitted GEV model.

## 3.3. $\mathbf{4 0 0} \mathrm{m}$ Run and $\mathbf{4 \times 1 0 0} \mathrm{m}$ Relay

Estimates of the GEV parameters, which were the results of the GEV modeling on the 400 m records of male gold medalists using the block maxima method, are shown in Table 5. Since $\xi$ was negative, the 400 m records of male gold medalists had a finite upper limit. The maximum return levels are shown in Table 6. The diagnostic plots for assessing the accuracy of the GEV model fitted to the 400 m records for male gold medalists are shown in Figure 6. The estimated curve is nonlinear in the return level curve due to the negative $\xi$. The diagnostic plots gave the validity of the GEV model.

Estimates of the GEV parameters, which were the results of the GEV modeling on the $4 \times 100 \mathrm{~m}$ relay records of male gold medalists using the block maxima method, are shown in Table 7. Since $\xi$ was negative, the $4 \times 100 \mathrm{~m}$ relay records of male gold medalists had a finite upper limit. The maximum return levels are shown in Table 8. The diagnostic plots for assessing the accuracy of the GEV model fitted to the $4 \times 100 \mathrm{~m}$ records for male gold medalists are shown in Figure 7. The estimated curve is nonlinear in the return level curve due to the negative $\xi$. The diagnostic plots supported the fitted GEV model.

### 3.4. Long Jump

Estimates of the GEV parameter estimates, which were the results of the GEV modeling on the long jump records of male gold medalists using the block maxima method, are shown in Table 9. Since $\xi$ was negative, the long jump records of male gold medalists had a finite upper limit. The predicted maximum return levels are shown in Table 10. The diagnostic plots for assessing the accuracy of the GEV model fitted to the long jump records of male gold medalists are shown in Figure 8. The estimated curve is nonlinear in the return level curve due to the negative $\xi$. The diagnostic plots supported the fitted GEV model.


Figure 5. Diagnostic plots for GEV fit to 200 m records of male gold medalists.


Figure 6. Diagnostic plots for GEV fit to 400 m records for male gold medalists.


Figure 7. Diagnostic plots for GEV fit to the $4 \times 100 \mathrm{~m}$ relay records of male gold medalists.


Figure 8. Diagnostic plots for GEV fit to the long jump records of male gold medalists.

Table 3. GEV parameter estimates in the 200 m records of male gold medalists.

|  | $\mu$ | $\sigma$ | $\xi$ |
| :---: | :---: | :---: | :---: |
| Parameter estimate | 20.74 | 0.981 | -0.629 |
| Standard error | 0.204 | 0.176 | 0.177 |
| $95 \% \mathrm{CI}$ | $(20.3,21.1)$ | $(0.637,1.33)$ | $(-0.975,-0.282)$ |

Table 4. GEV return level estimates in the 200 m records of male gold medalists.

| Return period <br> (years) | 10 | 20 | 50 | 100 | 500 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Return level (s) | 20.20 | 19.79 | 19.51 | 19.39 | 19.25 |
| Standard error | 0.180 | 0.140 | 0.0976 | 0.0903 | 0.123 |
| $95 \%$ CI | $(19.85,20.55)$ | $(19.51,20.06)$ | $(19.31,19.69)$ | $(19.21,19.56)$ | $(19.01,19.49)$ |

Table 5. GEV parameter estimates in the 400 m records of male gold medalists.

|  | $\mu$ | $\sigma$ | $\xi$ |
| :---: | :---: | :---: | :---: |
| Parameter estimate | 46.1 | 2.36 | -0.759 |
| Standard error | 0.470 | 0.427 | 0.144 |
| $95 \%$ CI | $(45.2,47.0)$ | $(1.52,3.19)$ | $(-1.52,-0.478)$ |

Table 6. GEV return level estimates in the 400 m records of male gold medalists.

| Return period <br> (years) | 10 | 20 | 50 | 100 | 500 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Return level (s) | 44.83 | 43.96 | 43.44 | 43.24 | 43.03 |
| Standard error | 0.368 | 0.270 | 0.169 | 0.122 | 0.0965 |
| $95 \% \mathrm{CI}$ | $(44.11,45.55)$ | $(43.43,44.49)$ | $(43.11,43.77)$ | $(43.00,43.48)$ | $(43.86,43.24)$ |

Table 7. GEV parameter estimates in the $4 \times 100 \mathrm{~m}$ relay records of male gold medalists.

|  | $\mu$ | $\sigma$ | $\xi$ |
| :---: | :---: | :---: | :---: |
| Parameter estimate | 39.4 | 1.64 | -0.621 |
| Standard error | 0.348 | 0.287 | 0.138 |
| $95 \% \mathrm{CI}$ | $(38.7,40.0)$ | $(1.08,2.21)$ | $(-0.891,-0.351)$ |

Table 8. GEV return level estimates in the $4 \times 100 \mathrm{~m}$ relay records of male gold medalists.

| Return period <br> (years) | 10 | 20 | 50 | 100 | 500 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Return level (s) | 38.45 | 37.75 | 37.28 | 37.08 | 36.85 |
| Standard error | 0.296 | 0.2379 | 0.174 | 0.148 | 0.155 |
| $95 \% \mathrm{CI}$ | $(37.86,39.03)$ | $(37.28,38.21)$ | $(36.93,37.62)$ | $(36.78,37.36)$ | $(36.54,37.14)$ |

Table 9. GEV parameter estimates in the long jump records of male gold medalists.

|  | $\mu$ | $\sigma$ | $\xi$ |
| :---: | :---: | :---: | :---: |
| Parameter estimate | 7.91 | 0.524 | -0.467 |
| Standard error | 0.113 | 0.0881 | 0.143 |
| $95 \%$ CI | $(7.69,8.13)$ | $(0.351,0.697)$ | $(-0.728,-0.198)$ |

Table 10. GEV return level estimates in the long jump records of male gold medalists.

| Return period <br> (years) | 10 | 20 | 50 | 100 | 500 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Return level (m) | 8.21 | 8.47 | 8.68 | 8.78 | 8.91 |
| Standard error | 0.107 | 0.0958 | 0.0838 | 0.0849 | 0.113 |
| $95 \%$ CI | $(8.00,8.50)$ | $(6.66,10.28)$ | $(8.53,8.83)$ | $(8.61,8.94)$ | $(8.66,9.18)$ |

## 4. Discussion

The return level plot in the 100 m and 200 m records of male gold medalists in a $\log -\log$ plot is shown in Figure 9. Approximately straight lines are also shown. The 200 m record has a larger decrease and a larger growth margin. The calculated upper limit was 9.58 s in the 100 m , which is equal to the record of Usain Bolt (2009). Einmahl (2011) estimated the ultimate world record and found 9.51 s for men [8]. The calculated upper limit was 19.18 s in the 200 m , which was almost equal to the first record of Usain Bolt, 19.19 s (2009).

The return level plot in the 400 m and $4 \times 100 \mathrm{~m}$ relay in a log-log plot is shown in Figure 10. Approximately straight lines are also shown. The 400 m record has a larger decrease and a larger growth margin. The slope of the approximate straight line in Figure 9 and Figure 10 (" $a$ " is the slope in Table 11) is the largest at $200 \mathrm{~m}(-0.01130)$ and the smallest at $400 \mathrm{~m}(-0.009324)$. The larger the slope, the greater the improvement in future records. The calculated upper limit was 43.03 s and 36.71 s in the 400 m and $4 \times 100 \mathrm{~m}$ relay, respectively. In the 400 m , for the 500-year return period, the return level for men was obtained as 43.03 s , with a $95 \%$ CI $(42.86,43.24)$. Hence, the probability of occurrence in one year of the record of Wayde Van Niekerk, 43.03 s (2016), which was the first record in the world, was $1 / 500$. In the $4 \times 100 \mathrm{~m}$ relay, the 500 -year return level for men was obtained as 36.85 s , with a $95 \%$ CI (36.54, 37.14). Hence, the probability of occurrence in one year of the record in Jamaica, 36.84 s (2012), which was the first record in the world, was $1 / 500$.

The return level plot in the long jump records in a log-log plot is shown in Figure 11. An approximately straight line is also shown. The calculated upper limit in the long jump at the Olympics was 9.03 m . The long jump world record ( 8.95 m ) was close to the calculated upper limit. Growth was small after the 100-year return period.

Table 11 shows the results when the approximate straight line is $y=b x^{a}$ in Figures 9-11. All cases were well approximated to the straight lines, following a


Figure 9. Return level plot in the 100 m and 200 m records of male gold medalists in a $\log -\log$ plot.


Figure 10. Return level plot in the 400 m and $4 \times 100 \mathrm{~m}$ relay records of male gold medalists in a log-log plot.


Figure 11. Return level plot in the long jump records of male gold medalists in a $\log -\log$ plot.

Table 11. The results when the approximate straight line is $y=b x^{a}$ in $100 \mathrm{~m}, 200 \mathrm{~m}, 400$ $\mathrm{m}, 4 \times 100 \mathrm{~m}$, and Long jump. The correlation coefficient is indicated by r .

|  | 100 m | 200 m | 400 m | $4 \times 100 \mathrm{~m}$ | Long jump |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $a$ | -0.009958 | -0.01130 | -0.009324 | -0.01007 | 0.01973 |
| $b$ | 10.19 | 20.53 | 45.37 | 39.02 | 7.949 |
| $r$ | 0.9527 | 0.9409 | 0.9214 | 0.9427 | 0.9586 |

power law. A power law, known as a scaling law, is a relation of the type $y=b x^{a}$, where y and x are variables of interest, $a$ is called the power exponent, and b is a constant. This indicates a correlation between the return level and the return period. The difference in inclination is small and the $200 \mathrm{~m}(-0.01130)$ is the largest and $400 \mathrm{~m}(-0.009324)$ is the smallest among $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}$, and $4 \times$ 100 m .

The calculated upper limits and world records in the $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}, 4 \times$ 100 m relay, and long jump are shown in Table 12. The differences are $0,0.01$, 0.06 , and 0.13 s in the $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}$, and $4 \times 100 \mathrm{~m}$ relay, respectively. Hence, achieving the calculated limit was difficult in the 100 m and 200 m . However, in the 400 m , and $4 \times 100 \mathrm{~m}$ relay, a slight improvement was possible.

Denny (2008) followed an extreme value theory-related approach [11]. The predicted records for the $100 \mathrm{~m}, 200 \mathrm{~m}$, and 400 m are $9.48,18.63$, and 42.73 s , respectively. The results agree with our $9.58,19.18$, and 42.97 s for the 100 m , 200 m , and 400 m , respectively. In our previous paper [12], the calculated upper limit in the 100 m for men using the world records for 1970-2010 is 9.46 s , which agrees much with Denny's result.

The slopes of the regression lines of records are shown in Table 13. The slopes of regression lines were roughly proportional to the running distance for 100 m , 200 m , and $4 \times 100 \mathrm{~m}$ relay. In the 100 m , the slope of the regression line of records was the smallest and it was difficult to update records. From the density plot (Figure 4), in the 100 m there was almost one peak and the difference in records was small. The tendency was similar between the 100 m and 200 m , but the record update and the difference of records in the 100 m were smaller. In the 400 m , the slope of regression lines of records was the largest, and the record growth was large. The density plot (Figure 6) showed two peaks, and the records were distributed widely. The $4 \times 100 \mathrm{~m}$ had a smaller record update and a smaller record difference. In the 100 m , it is thought that an individual's ability from the beginning has a significant influence.

At the Tokyo Olympics in August 2021, the records of male gold medalists were $9.80 \mathrm{~s}, 19.62 \mathrm{~s}, 43.85 \mathrm{~s}, 37.50 \mathrm{~s}$, and 8.41 m in the $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}, 4 \times$ 100 m relay, and long jump, respectively. The record in the $100 \mathrm{~m}, 9.80 \mathrm{~s}$, was the 30-year return level estimate. In one year, the probability of occurrence for a record, 9.80 s was $1 / 30$. The record in the $200 \mathrm{~m}, 19.62 \mathrm{~s}$, was the 32 -year return level estimate. The record in the $400 \mathrm{~m}, 43.85 \mathrm{~s}$, was the return level estimate for the 23-year return period. The record in the $4 \times 100 \mathrm{~m}, 37.50 \mathrm{~s}$, was the 30 -year

Table 12. Calculated upper limits with standard errors and world records.

| Event | Calculated upper limit | Standard error | World record |
| :---: | :---: | :---: | :---: |
| 100 m | 9.58 s | 0.0861 s | 9.58 s |
| 200 m | 19.18 s | 0.981 s | 19.19 s |
| 400 m | 42.97 s | 2.36 s | 43.03 s |
| $4 \times 100 \mathrm{~m}$ relay | 36.71 s | 1.64 s | 36.84 s |
| Long jump | 9.03 m | 0.197 m | 8.95 m |

Table 13. The slopes of regression lines for records in the 100 m and $200 \mathrm{~m}, 400 \mathrm{~m}$ and 4 $\times 100 \mathrm{~m}$ relay, and long jump shown in Figures 1-3. The values of $\mathrm{R}^{2}$ are also shown.

| Event | Record $\times 10^{-2}$ | $\mathrm{R}^{2} \times 10^{-1}$ |
| :---: | :---: | :---: |
| 100 m | -1.03 | 8.66 |
| 200 m | -2.47 | 8.80 |
| 400 m | -6.44 | 8.19 |
| $4 \times 100 \mathrm{~m}$ relay | -4.87 | 8.99 |
| Long jump | 1.35 | 7.40 |

return level estimate. The record in the long jump, 8.41 m , was the 16 -year return level estimate. In the $100 \mathrm{~m}, 200 \mathrm{~m}$, and $4 \times 100 \mathrm{~m}$, in one year the probability of occurrence for a record was about $1 / 30$. In the 400 m and long jump, it was about $1 / 20$. In the $100 \mathrm{~m}, 200 \mathrm{~m}$, and $4 \times 100 \mathrm{~m}$ relay compared with the 400 $m$ and long jump, more difficult records show that a fierce battle has taken place.

## 5. Conclusions

Extreme value theory can provide methods to analyze the most extreme parts of data. We used the generalized extreme value (GEV) distribution to predict the ultimate $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}, 4 \times 100 \mathrm{~m}$ relay, and long jump records of male gold medalists at the Olympics. The results are summarized as follows:

1) The diagnostic plots, which assessed the accuracy of the GEV model, were fitted to all event records, validating the model.
2) The $100 \mathrm{~m}, 200 \mathrm{~m}, 400 \mathrm{~m}, 4 \times 100 \mathrm{~m}$, and long jump records had shape parameters of $-0.571,-0.629,-0.759,-0.621$, and -0.467 and calculated upper limits of $9.58 \mathrm{~s}, 19.18 \mathrm{~s}, 42.97 \mathrm{~s}, 36.71 \mathrm{~s}$, and 9.03 m , respectively. The calculated upper limit in the 100 m ( 9.58 s ) was equal to the record of Usain Bolt (August 16,2009 ).
3) The 100 m and 200 m world records were close to the calculated upper limits, and achieving the calculated limit was difficult. The 400 m and $4 \times 100 \mathrm{~m}$ relay world records were almost equal to the calculated upper limits and the 500 -year return level estimate, and slight improvement was possible in both.
4) At the Tokyo Olympics in August 2021, in the $100 \mathrm{~m}, 200 \mathrm{~m}$, and $4 \times 100$ m , in one year the probability of occurrence for a record was about $1 / 30$. In the

400 m and long jump, that was about $1 / 20$. In the $100 \mathrm{~m}, 200 \mathrm{~m}$, and $4 \times 100 \mathrm{~m}$ relay, more difficult records show that a fierce battle has taken place.
5) The relationship between the return level of the records and the return period follows a power law.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

[1] Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values. Sprin-ger-Verlag, Berlin. https://doi.org/10.1007/978-1-4471-3675-0
[2] Katz, R.W., Parlange, M.B. and Naveau, P. (2002) Statistics of Extremes in Hydrology. Advances in Water Resources, 25, 1287-1304. https://doi.org/10.1016/S0309-1708(02)00056-8
[3] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) Modeling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin. https://doi.org/10.1007/978-3-642-33483-2
[4] Roberts, S.J. (2000) Extreme Value Statistics for Novelty Detection in Biomedical Data Processing. IEE Proceedings-Science Measurement and Technology, 147, 363-367. https://doi.org/10.1049/ip-smt:20000841
[5] Lavenda, B.H. and Cipollone, E. (2000) Extreme Value Statistics and Thermodynamics of Earthquakes: Aftershock Sequences. Annali di geofisica, 43, 967-982.
[6] Thomas, M., Lemaitre, M., Wilson, M. L., Vibound, C., Yordanov, Y., Wackernagel, H. and Carrat, F. (2016) Applications of Extreme Value Theory in Public Health. PLoS ONE, 11, e0159312. https://doi.org/10.1371/journal.pone. 0159312
[7] Ito, H. and Okano, S. (2005) Analysis of Changes in the 100 m Records in Japan and the World. Bulletin of Athletics Research, 1, 61-66.
[8] Einmahl, J.H.J. and Magnus, J.R. (2008) Records in Athletics through Extreme-Value Theory. Journal of the American Statistical association, 103, 1382-1391. https://doi.org/10.1198/016214508000000698
[9] Einmahl, J.H.J. and Smeets, S.G.W.R. (2011) Ultimate 100-m World Records through Extreme-Value Theory. Statistical Neerlandica, 65, 32-42. https://doi.org/10.1111/j.1467-9574.2010.00470.x
[10] https://www.olympic.org/athletics
[11] Denny, M. (2008) Limits to Running Speed in Dogs, Horses and Humans. Journal of Experimental Biology, 211, 3836-3849. https://doi.org/10.1242/jeb. 024968
[12] Maruyama, F. (2021) Analysis of Japan and World Records in the 100 m Dash Using Extreme Value Theory. Journal of Applied Mathematics and Physics, 9, 1442-1451. https://doi.org/10.4236/jamp. 2021.97097

