Review

International Journal of Sports Science & Coaching

International Journal of Sports Science & Coaching
0(0) 1–12

DOI: 10.1177/1747954118808068
journals.sagepub.com/home/spo
SAGE

Swimming performance index based on extreme value theory

Daniel T Gomes[1] and Lígia Henriques-Rodrigues[2]

Abstract

The International Swimming Federation has developed a points system that allows comparisons of results between different events. Such system is important for several reasons, since it is used as a criterion to rank swimmers in awards and selection procedures of national teams. The points system is based entirely on the world record of the correspondent event. Since it is based on only one observation, this work aims to suggest a new system, based on the probability distribution of the best performances in each event. Using extreme value theory, such distribution, under certain conditions, converges to a generalized Pareto distribution. The new performance index, based on the peaks over threshold methodology, is obtained based on the exceedance probabilities correspondent to the swimmers' times that exceed a given threshold. We work with 17 officially recognized events in 50 m pool, for each women and men, and considered all-time rankings for all events until 31 December 2016. A study on the adequacy of the proposed generalized Pareto distribution index and a comparison between the performances of Usain Bolt and Michael Phelps are also conducted.

Keywords

Extreme quantiles, extreme value theory, generalized Pareto distribution, sports statistics, swimming

Introduction

Competitive swimming has a non-subjective ranking system, based on times. When one analyzes the results of a given event, such criterion is free from arbitrariness and is the universal way to choose the "best (fastest) swimmers." However, for some reasons, it might be reasonable to compare performances in different events. For example, the winning time of the 100 m freestyle in a given competition is better than the winning time of the 200 m butterfly?

Obviously, the concept of "better" needs to be defined. Intuitively, a great performance is the one that, in its respective event, corresponds to a discrepant result in comparison to the others.

Such comparison is made by the International Swimming Federation (FINA) through a points system. It is important, since it is used to determine the winners of the World Cup.[1] It is also used by several national federations as the criterion to determine national teams, such as Brazil.[2]

In 2015, the FINA points system was used for the first time to select the best performances of that year. In 2016, Katie Ledecky (United States) and Adam Peaty (Great Britain) were chosen the best performers of the year, by their performances in women's 800 m freestyle and men's 100 m breaststroke, respectively, in Rio de Janeiro Olympic Games.[3]

The FINA points system is based on what is called the *base times*.[4] The base times are defined every year, based on the latest world record that was approved by FINA. The base times are defined with the cut date of 31 December. So, in a given event, let $T_i(t)$ be the time obtained by the swimmer i, in seconds, at the year t, and $B(t-1)$ the base time, which is the world record, in

Reviewer: John Einmahl (Tilburg University, Netherlands)
Ian G. McHale (University of Liverpool, UK)
Jonathan Tawn (Lancaster University, UK)

[1]National School of Statistics, Brazilian Geography and Statistics Institute, Rio de Janeiro, Brazil
[2]Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

Corresponding author:
Daniel T Gomes, National School of Statistics, Brazilian Geography and Statistics Institute, Rio de Janeiro, Brazil.
Email: daniel.gomes@ibge.gov.br

seconds, at the end of the year t–1. The points of the swimmer i at the year t, $P_i(t)$, are given by

$$P_i(t) = \left\lfloor 1000\left(\frac{B(t-1)}{T_i(t)}\right)^3 \right\rfloor \qquad (1)$$

with $\lfloor x \rfloor$ denoting the integer part of x. The fragility of this methodology is the fact that the points system uses little information—in fact, only the world record—and does not have scientific support. World records of different events do not necessarily represent the same difficult level, and because of this they do not represent a fair ground for comparisons.[5]

The motivation to obtain an alternative criterion is to use a set of times as the base, and not only one time, as the FINA points system, in equation (1), and the evaluation of the data distribution.

Usually, studies about probability models in sports data refer to records behavior and forecasts. The majority of works is about athletics data and extreme value theory.

Smith[6] proposed a maximum likelihood (ML) method for fitting models to a series of records and applied his method to athletic records for the 1500 m and the marathon, comparing normal and extreme value distributions. Robinson and Tawn[7] considered the annual best times in the women's 3000 m event and applied several extreme values methods to evaluate how discrepant was the 1993 world record time established by the Chinese runner Wang Junxia. Barão and Tawn[8] studied the evolution of 1500 m and 3000 m annual best times in a bivariate set-up. The analysis conducted in these papers was based on the development of top performances over time. This is not the case in this paper. We are using only the top performances associated with a set of n athletes, as in Einmahl and Magnus[9] and Henriques-Rodrigues, Gomes and Pestana,[10] who evaluated how good was a world record in comparison to the others and what were the estimated ultimate world records. Adam and Tawn[11] performed an analysis of swimming times data over different events and modeling evolution over time using a bivariate extreme value analysis. Stephenson and Tawn[12] used a Bayesian framework and Markov chain Monte Carlo methods to simultaneously model performances over both time and event distance in athletics.

In this paper, some ideas explored by the papers cited above will be used to evaluate a theoretically justified criterion alternative to the FINA's one. Such methodology could be applied to athletics as well. Regarding the data, they will be used in the same approach as Einmahl and Magnus,[9] who analyzed athletics historical rankings. In this paper, the aim is to compare performances that do not necessarily represent world records. Our approach is based on Fisher-Tippet and PickandsBalkemade Haan theorems and the peaks over threshold (POT) methodology.[13–15]

The paper is organized as follows. Firstly, we describe the data related with the swimming events under study. Next, we present some preliminary results in extreme value theory related with the *peaks over threshold* (POT) methodology and present the new generalized Pareto distribution (GPD) *performance index* proposed for performance comparisons between different swimming events. Then, we present the application of the POT methodology to the data sets under consideration, the performance comparisons, the GPD's suitability and a comparison between Usain Bolt and Michael Phelps best performances. Finally, some conclusions are drawn regarding the FINA points system and the new GPD performance index.

Data

In this paper, the aim is to evaluate swimming performances from 2016 to point out the better results of that year.

There are 34 officially recognized individual swimming events in 50 m pool, 17 for women and 17 for men, in the following distances and styles: 50 m, 100 m, 200 m, 400 m, 800 m, and 1500 m freestyle; 50 m, 100 m, and 200 m butterfly; 50 m, 100 m, and 200 m backstroke; 50 m, 100 m, and 200 m breaststroke; and 200 m, and 400 m individual medley. The all-time rankings for all events are considered, until 31 December 2016.

The database is comprised of world rankings published by FINA (*http://www.fina.org*) and SwimNews (*http://www.swimnews.com*) websites from 1990 on and an all time world ranking published by Swimming World Magazine website (*http://www.swimmingworldmagazine.com*) in 2007.

The data correspond to the best times over all times until 31 December 2016. Only the best time of each athlete is considered in each event. Obviously, one athlete may have several times registered, which are correlated. So, considering only the best time is a way to guarantee the independence assumption. We collected data on the personal best performances of as many of the top athletes in each event as we could, taking care that no "holes" occur in the list. Thus, if an athlete appears on our list, then all athletes with a better personal mark will also appear on our list. Table 1 gives an overview of the number of athletes (the depth) and the worst and best results for each event in the sample.

It is very frequent to obtain repeated values (clusters) for several athletes, due to the lack of precision of the measurements and discretization of data, the times being registered to the second decimal place. Because clusters can cause problems in the estimation and in the

Table 1. Summary data until 31 December 2016.

	Women			Men		
Event	Depth (n)	Best	Worst	Depth (n)	Best	Worst
50 m freestyle	595	23.73	25.75	486	20.91	22.66
100 m freestyle	546	52.06	55.79	458	46.91	49.68
200 m freestyle	557	1:52.98	2:00.39	491	1:42.00	1:49.29
400 m freestyle	488	3:56.46	4:12.83	438	3:40.07	3:52.33
800 m freestyle	440	8:04.79	8:40.38	485	7:32.12	8:06.85
1500 m freestyle	457	15:25.48	16:43.58	469	14:31.02	15:24.64
50 m butterfly	602	24.43	27.46	499	22.43	24.32
100 m butterfly	527	55.48	1:00.08	373	49.82	53.16
200 m butterfly	445	2:01.81	2:11.96	562	1:51.51	1:59.70
50 m backstroke	649	27.06	29.54	594	24.04	26.19
100 m backstroke	645	58.12	1:02.54	550	51.85	55.83
200 m backstroke	589	2:04.06	2:13.99	626	1:51.92	2:01.75
50 m breaststroke	509	29.28	32.29	537	26.42	28.48
100 m breaststroke	489	1:04.35	1:09.71	469	57.13	1:01.92
200 m breaststroke	556	2:19.11	2:30.34	525	2:07.01	2:14.68
200 m ind. medley	583	2:06.12	2:16.26	534	1:54.00	2:02.78
400 m ind. medley	620	4:26.36	4:48.79	667	4:03.84	4:23.87

evaluation of the goodness-of-fit tests, we smoothed the data, according to the procedure suggested by Einmahl and Magnus.[9] So, if r athletes complete the 100 m freestyle for women in the time of t_0 seconds (s), the r results are smoothed in the interval $]t_0 - 0.005, t_0 + 0.005[$ in the following way, where $ts_k(.)$ is the time after smoothing of the kth athlete, in s

$$ts_k(t_0) = (t_0 - 0.005) + 0.01\,\frac{2k-1}{2r}, \quad k = 1, \ldots, r.$$

This procedure was conducted for all the observed clusters, resulting in a new data set denoted by $T = \{t_j, j = 1, \ldots, n\}$.

Besides that, one more transformation is required for the subsequent analysis. Usually, the extreme value theory deals with inference based on sample maxima. Since swimming times implies a sample minima analysis, we have to transpose the smoothed swimming times to a maxima context. This is done converting the times in swimming speeds, so that a lower swimming time corresponds to a higher speed. This way, selecting the lowest time for an athlete is equivalent to selecting the highest speed.

Preliminary results in extreme value theory

Let us think on any swimming event, like for instance the women's 100 m freestyle. Let X_i, $1 \le i \le N$, denote the best personal mark of the ith athlete, N the number of athletes being considered. By "best personal mark," we mean the swimming time transformed to speed, so the higher the speed, the best. Note that N is different from n, which refers to the sample size available, whereas N is the number of athletes that has swum the event in the world.

We consider these N personal bests as independent, identically distributed (i.i.d.) observations from some distribution function (d.f.) F. Let $X_{1:N} \le X_{2:N} \le \ldots \le X_{N:N}$ be the associated order statistics, so that $X_{N:N}$ denotes the world record.

Suppose that there exists real sequences $a_N > 0$ and $b_N \in \mathbb{R}$ such that $(X_{N:N} - b_N)/a_N$ converges in distribution for some non-degenerate d.f. $G(x)$, i.e.

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{X_{N:N} - b_N}{a_N} \le x\right) = \lim_{N\to\infty} F^N(a_N x + b_N) = G(x). \tag{2}$$

If this condition holds, we say that F is in the max-domain of attraction of G and we write $F \in D(G)$. According to the Fisher-Tippet theorem,[15] if $F \in D(G)$, the d.f. G in equation (2) is in the family of G_ξ, where G_ξ is the d.f. of the generalized extreme value (GEV) distribution, given by

$$G_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\}, & \text{if } \xi \neq 0 \\ \exp\{\exp(-x)\}, & \text{if } \xi = 0 \end{cases} \tag{3}$$

where $1 + \xi x > 0$. The shape parameter ξ is directly related with the right-tail of the d.f. G_ξ and it determines the weight of the right-tail of G_ξ. The shape parameter is also known as the tail index and is one of the main parameters in extreme value theory. For $\xi = 0$, we say that d.f. $G_0(x)$, in equation (3), is in the max-Gumbel-type domain of attraction, which contains exponential right-tailed distributions such as normal, exponential, Gama and Gumbel itself. If $\xi > 0$, $G_\xi(x)$, in equation (3), is in the max-Fréchet-type domain of attraction, which contains heavy right-tailed distributions such as Pareto, Student-t and Fréchet itself. If $\xi < 0$, $G_\xi(x)$, in equation (3), is in the max-Weibull-type domain of attraction, which contains light right-tailed distributions such as Beta and Weibull itself.[15]

POT methodology

The POT methodology refers to the analysis of random variables that exceed a given high and fixed threshold u. The observations that exceed such threshold u are called exceedances over u.

Let x^F be the finite or infinite right endpoint of the support of F, i.e. $x^F = \sup\{x \in \mathbb{R} : F(x) < 1\}$. The conditional d.f. of the exceedances over u is given by

$$F_u(x) = \mathbb{P}(X - u \le x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}$$

for $0 \le x \le x^F - u$. Balkema and de Haan[16] and Pickands[17] proved that, for a sufficiently high threshold u and under the validity of equation (2), the conditional d.f. F_u may be well approximated by GPD, they proved that the GPD is the limiting distribution of suitably normalized exceedances when $u \to x^F$. Such result is known as the Pickands–Balkema–de Haan theorem,[14] and the GPD is expressed by

$$G_{\theta,\xi}(x) = \begin{cases} 1 - (1 + \xi x/\theta)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\{-x/\theta\}, & \text{if } \xi = 0, \end{cases} \tag{4}$$

where $\theta > 0$ is a scale parameter and the support is $x \ge 0$ if $\xi \ge 0$ and $0 \le x \le -\theta/\xi$ if $\xi < 0$.[13]

The theorem suggests that, for a sufficiently high threshold u, the d.f. of the exceedances, i.e. d.f of the athletes' performances that are higher than u, can be approximated by a GPD.

We perform the analysis of maxima of speeds, instead of minima of times. According to Wadsworth et al.,[18] the settings can result in potentially different results. However, we opted to analyze the maxima of speeds because POT methodology is based on the conditional d.f. of the exceedances over threshold. To perform the analysis of minima of times, one could negate the data and analyze that.

The choice of the threshold u is still an open problem. Some methods are suggested in the literature (see Vicente[19] for an overview on the subject).

Davison and Smith[20] suggest the use of the Mean Excess function, $e(u) = \mathbb{E}(X - u | X > u)$. Instead of working with the exceedances over u, we can work with the excesses, represented by the r.v. $Y := X - u$. If Y follows a GPD distribution, then $e(u) = (\theta + \xi u)/(1 - \xi)$, if $\zeta < 1$ So, if the GPD assumption is true, the plot of $e(u)$ *versus* u, called mean excess plot (ME-Plot), should follow a straight line with intercept $\theta/(1 - \xi)$ and slope $\xi/(1 - \xi)$. In practice, based on a sample size n, $x_1 \cdots x_n$, $e(u)$ is estimated by its empirical counterpart, the sample mean excess function

$$\hat{e}_n(u) = \frac{\sum_{i=1}^{n} x_i I(x_i > u)}{\sum_{i=1}^{n} I(x_i > u)} - u$$

where $I(x_i > u)$ is the indicator function of the event (set) $\{x_i : x_i > u\}$, given by 1, if $x_i > u$ and 0 otherwise.

We generally construct the sample ME-plot with pairs x_{n-kn} and $\hat{e}_n(x_{n-k:n})$, $k = 1, \ldots, n-1$ where $x_{k:n}$ is the kth order statistics. If the data over a high threshold is well fitted by a GPD, we would expect the sample ME-plot to become linear. But even for data that are genuinely GP-distributed, the sample ME-plot is seldom perfectly linear, particularly toward the right-hand side, where we are averaging a small number of large excesses. We often omit the final few points from consideration, as they can severely distort the plot.[13] Consequently, the threshold u is chosen at the point to the right of which a rough linear pattern appears in the plot.

Coles[21] suggests to fit the GPD at a range of thresholds, and to look for stability of parameter estimates.

If a GPD is a reasonable model for the excesses over a threshold u_0, then the excesses over a higher threshold u, $u > u_0$, should also follow a GPD. The shape parameters of the two distributions are identical, as well as the reparameterized scale parameter

$$\theta^* = \theta_u - \xi u.$$

Consequently, estimates of both θ^* and ξ should be constant above u_0, if u_0 is a valid threshold for excesses to follow the GPD. This argument suggests plotting both $\hat{\theta}^*$ and against u, together with confidence intervals for each of these quantities, and selecting u_0 as the lowest value of u for which the estimates remain near-constant. Sampling variability means that the estimates of these quantities will not be exactly constant, particularly towards the right-hand side, where we deal with a small number of large excesses.

After the determination of threshold, the estimation of θ and ξ should be conducted, based on the available data. There are several estimation methods available in the literature. For a deeper discussion about its practical aspects, please refer to Bermudez and Kotz.[22]

Given a threshold u, the exceedance probability p of a high level x_0, in a GPD context, can be estimated through its d.f. by $\hat{p} = 1 - G_{\hat{\theta},\hat{\xi}}(x_0 - u)$, with $G_{\theta,\xi}$ given in equation (4). Such value is the *GPD performance index* proposed in this article for performance comparisons between different swimming events or different sports events.

Let A and B be two distinct swimming events and $(u_A; n_{u_A}; \hat{\theta}_A, \hat{\xi}_A)$ and $(u_B; n_{u_B}; \hat{\theta}_B, \hat{\xi}_B)$ the thresholds, u, number of exceedances above the threshold u, n_u, and the scale and shape estimators, $(\hat{\theta}, \hat{\xi})$, associated to the events A and B, respectively. Let x_A be the mark obtained by an athlete in event A and x_B the mark obtained by other athlete in event B and let us assume that . With $G_{\theta,\xi}$ given in equation (4), the exceedance probabilities associated to x_A and x_B are, respectively,

$$\hat{p}_A = 1 - G_{\hat{\theta}_A,\hat{\xi}_A}(x_A - u_A) \quad \text{and}$$
$$\hat{p}_B = 1 - G_{\hat{\theta}_B,\hat{\xi}_B}(x_B - u_B).$$

We say that x_A is better than x_B, i.e. we say that the mark attained by an athlete in the event A, x_A, is better than the mark attained by another athlete in the event B, x_B, if and only if $\hat{p}_A < \hat{p}_B$ or, similarly, if and only if $1 - \hat{p}_A > 1 - \hat{p}_B$, the criterion used later on in the data sets under consideration.

Application

Preliminary data analysis and threshold selection

Consider one swimming event and let $X_1, X_2, \ldots, X_N$ denote the best personal marks, in m/s, of all N athletes in the world in this event. The definition and possible measurement of N are difficult. Fortunately, the value of N turns to be unimportant.[9]

We consider these N best personal marks as i.i.d. observations from some d.f. F. According to the results presented in the previews section, for a high enough threshold u, the distribution of X_i conditional to $X_i > u$ converges to a GPD, $1 \leq i \leq N$, under certain conditions. Such result can be used to model the tail of the distribution of the best personal marks and give a measure of how much the ith athlete is discrepant from the distribution of all swimmers in the world.

As mentioned before, we use ME-plots and parameter stability plots to determine the threshold associated to each distribution.

If the GPD assumption is valid, the ME-plot should follow a straight line from the threshold on.[14] Figures 1 and 2 display the ME-plots for the women's and men's 400 m freestyle, where we can see decaying patterns. However, apart from the high volatility characterizing the top sample, it is possible to note two linear patterns, shown by the lines indicated by 1 and 2. The two linear patterns are separated by a kink, visible at 1.59 and 1.73, respectively. As the objective of the POT methodology is to induce a cut-off in the sample above which the sample ME-plot follows a linear trend, these values are good candidates for the thresholds of the GPD.

Figures 3 and 4 display the stability plots for the same events, women's and men's 400 m freestyle. The parameters are estimated by ML. Note that both parameter estimates remain near-constant above the mentioned values of 1.59 and 1.73, respectively.

We see that, toward the right-hand side, the confidence intervals for the parameters seem unreliable. In fact, one can expect to face numerical problems with

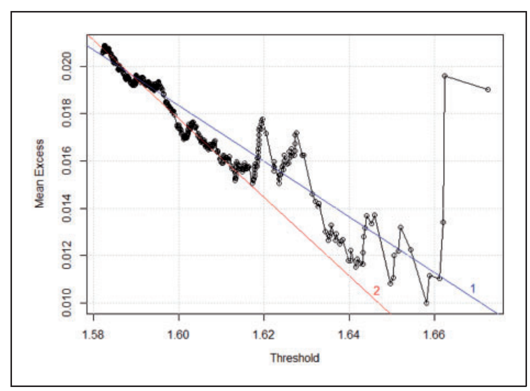

Figure 1. Women's 400 m freestyle ME-plot.

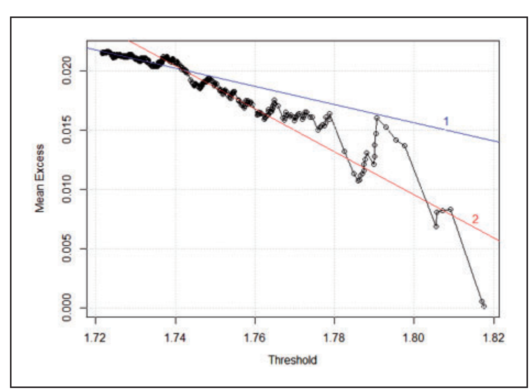

Figure 2. Men's 400 m freestyle ME-plot.

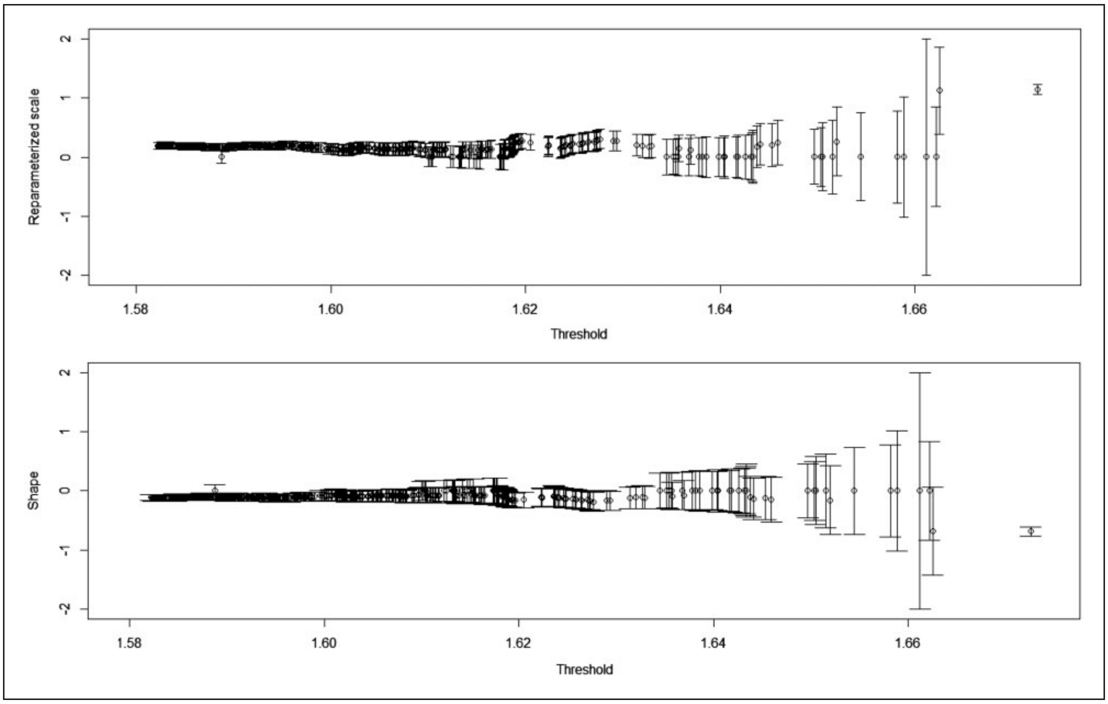

Figure 3. Reparameterized scale and shape parameters estimates versus threshold, women's 400 m freestyle.

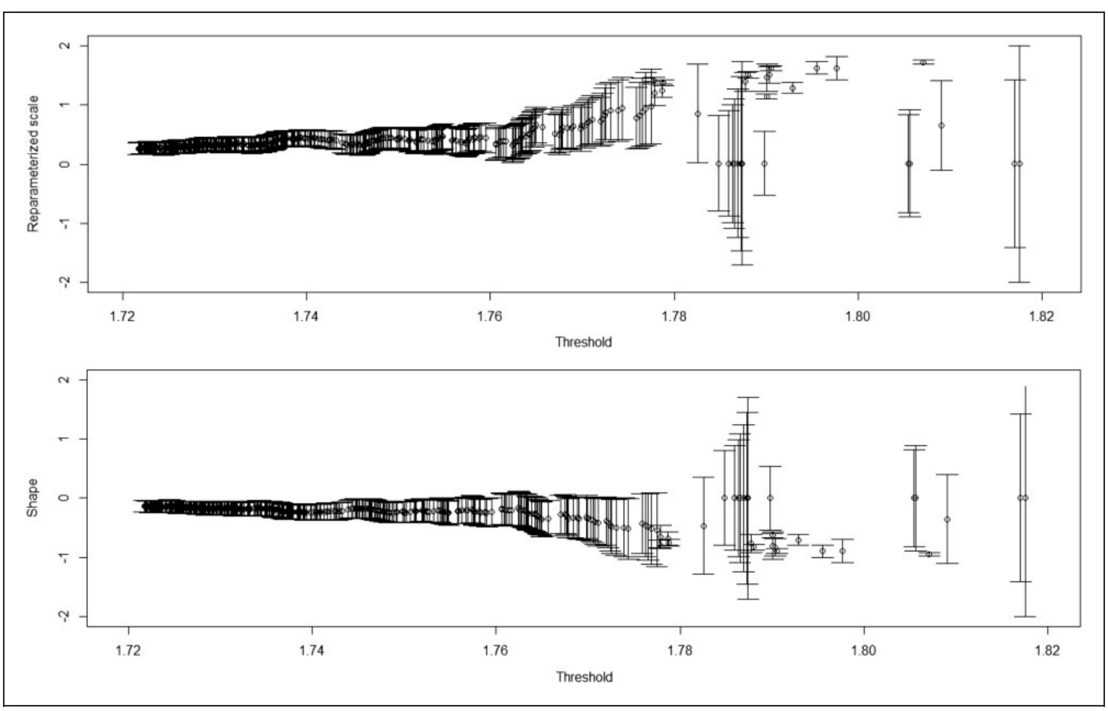

Figure 4. Reparameterized scale and shape parameters estimates versus threshold, men's 400 m freestyle.

ML estimates of a GPD when dealing with a small number of large excesses.[22] This is an important issue to be discussed in the next section in order to choose the value of the thresholds.

Such analysis was conducted similarly for all events. In Table 2, the chosen thresholds can be found, as well as the respective number of exceedances, n_u.

Table 2. Selected thresholds and number of exceedances.

	Women		Men	
Event	u	n_u	u	n_u
50 m freestyle	1.9735	218	2.2336	235
100 m freestyle	1.8194	214	2.0401	184
200 m freestyle	1.6900	173	1.8424	287
400 m freestyle	1.5955	261	1.7379	201
800 m freestyle	1.5614	174	1.6619	234
1500 m freestyle	1.5181	181	1.6407	209
50 m butterfly	1.8701	159	2.0968	156
100 m butterfly	1.6850	265	1.9014	188
200 m butterfly	1.5308	267	1.7020	177
50 m backstroke	1.7393	183	1.9596	152
100 m backstroke	1.6444	157	1.8216	223
200 m backstroke	1.5249	203	1.6783	175
50 m breaststroke	1.5871	150	1.7974	152
100 m breaststroke	1.4669	171	1.6467	151
200 m breaststroke	1.3534	235	1.5067	240
200 m ind. medley	1.4938	217	1.6612	153
400 m ind. medley	1.4064	253	1.5419	250

Fitting the GPD

Since the aim of this work is to compare performances in different events by the computation of exceedance probabilities, it is convenient that the distributions of different events be based on the same number of observations.

The chosen value for the threshold must be high enough to guarantee the GPD assumption, but not so high, so there can be a significant amount of observations for the estimation process.

If one particular GPD has threshold u, the observations above the threshold u_1 with $u_1 > u$ also follow a GPD, with the same ξ shape parameter.[14] All 34 events in study have more than 150 exceedances for the chosen thresholds. If we choose the thresholds that results in 150 exceedances in each event (these thresholds must be higher than the previously chosen thresholds), we can expect that these exceedances follow GPDs.

One of the key issues here is considering the data identically distributed. If the time trend effect was big, then a single GPD fitted to all the data would start to fail to be a good fit as the signal in the trend would outweigh the noise in the GPD. But, in fact, 86% of the data comes from the period 2007–2016, and only 4% were registered before 1996 (the earliest year from which a swimmer's time is included in the data analysis is 1976). We consider that it is reasonable to model the data as being identically distributed as the trends are small over the time range of the data relative to the natural variation in the data in any given year. Figure 5 shows the histogram of the years in which the data occur.

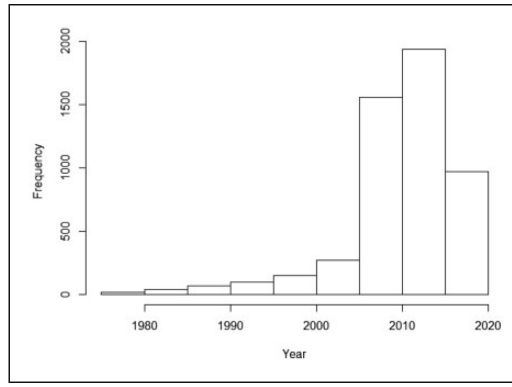

Figure 5. Histogram of the years in which the data occur.

Table 3 shows the thresholds for each event considering 150 exceedances, the estimated parameters computed by ML and its standard errors. We also present the p-values for the goodness-of-fit statistics Anderson-Darling (AD) and Cramér-von Mises for the adjustment of each event to the GPD. The p-values are obtained based on the tables given in Choulakian and Stephens.[23] Although not all p-values are above 0.05, they are, under the null hypothesis, uniformly distributed on (0, 1), as expected under the assumption of GPD data.

In order to check the stability of the method, we conducted the same analysis considering 160, 170, 180, 190, and 200 exceedances. The results are similar in terms of parameter estimates and goodness-of-fit tests.

Comparison of performances

For the comparison of performances in different events, the proposed criterion suggests the comparison of the exceedance probabilities of each event, based on estimated GPDs, according to the procedure described.

Table 4 shows the best performances of 2016, according to the proposed methodology, as well as the FINA points system and the correspondent rank. It is considered only the best performance of each athlete in each event. For women and men, the best performers, by the GPD, are the same as FINA ones: Katie Ledecky from United States and Adam Peaty from Great Britain, respectively. However, for FINA, Ledecky has the best performance for the 400 m freestyle, and by the GPD her best performance was obtained in the 800 m freestyle, which is in accord to what specialists say.[24] It is worth noting that in 2016, the 800 m freestyle performance by Katie Ledecky has a very small exceedance probability, which indicates that

Table 3. GPD fitting (150 observations in each event), ML estimates, standard errors, and goodness-of-fit tests.

Event	u	$\hat{\xi}$	SE($\hat{\xi}$)	$\hat{\theta}$	SE($\hat{\theta}$)	p AD	p CVM
Women							
50 m freestyle	1.986	−0.143	0.095	0.035	0.004	[0.25,0.5)	≥0.5
100 m freestyle	1.828	−0.261	0.067	0.031	0.003	[0.1,0.25)	[0.1,0.25)
200 m freestyle	1.692	−0.123	0.078	0.019	0.002	≥0.5	≥0.5
400 m freestyle	1.608	−0.118	0.075	0.019	0.002	[0.25,0.5)	[0.25,0.5)
800 m freestyle	1.565	−0.183	0.049	0.02	0.002	[0.005,0.01)	
1500 m freestyle	1.521	−0.174	0.063	0.025	0.003	≥0.5	≥0.5
50 m butterfly	1.874	−0.156	0.06	0.04	0.004	[0.1,0.25)	
100 m butterfly	1.704	−0.145	0.091	0.026	0.003	[0.025,0.05)	[0.025,0.05)
200 m butterfly	1.544	−0.194	0.065	0.026	0.003	≥0.5	≥0.5
50 m backstroke	1.746	−0.277	0.077	0.037	0.004	[0.25,0.5)	[0.25,0.5)
100 m backstroke	1.646	−0.183	0.093	0.024	0.003	[0.25,0.5)	[0.25,0.5)
200 m backstroke	1.532	−0.182	0.078	0.023	0.003	[0.25,0.5)	[0.25,0.5)
50 m breaststroke	1.587	−0.215	0.072	0.033	0.004	[0.25,0.5)	[0.25,0.5)
100 m breaststroke	1.471	−0.203	0.077	0.025	0.003	≥0.5	≥0.5
200 m breaststroke	1.365	−0.259	0.081	0.026	0.003	[0.05,0.1)	[0.25,0.5)
200 m ind. medley	1.502	−0.136	0.08	0.023	0.003	[0:25;0:5)	[0.1,0.25)
400 m ind. medley	1.422	−0.178	0.071	0.022	0.002	[0.005,0.01)	[0.01,0.025)
Men							
50 m freestyle	2.25	−0.031	0.106	0.033	0.004	[0.05,0.1)	[0.025,0.05)
100 m freestyle	2.047	−0.278	0.079	0.031	0.003	[0.25,0.5)	[0.25,0.5)
200 m freestyle	1.86	−0.06	0.073	0.02	0.002	[0.1,0.25)	[0.05,0.1)
400 m freestyle	1.747	−0.186	0.084	0.022	0.003	≥0.5	≥0.5
800 m freestyle	1.673	−0.143	0.074	0.024	0.003	≥0.5	≥0.5
1500 m freestyle	1.649	−0.281	0.075	0.026	0.003	≥0.5	≥0.5
50 m butterfly	2.099	−0.227	0.071	0.039	0.004	[0.025,0.05)	[0.025,0.05)
100 m butterfly	1.909	−0.19	0.07	0.028	0.003	≥0.5	≥0.5
200 m butterfly	1.706	−0.172	0.074	0.023	0.002	≥0.5	≥0.5
50 m backstroke	1.961	−0.311	0.064	0.044	0.004	≥0.5	≥0.5
100 m backstroke	1.837	−0.162	0.095	0.029	0.004	[0.25,0.5)	[0.25,0.5)
200 m backstroke	1.684	−0.174	0.075	0.028	0.003	≥0.5	≥0.5
50 m breaststroke	1.798	−0.244	0.063	0.029	0.003	[0.25,0.5)	[0.1,0.25)
100 m breaststroke	1.647	−0.104	0.073	0.021	0.002	[0.1,0.25)	[0.25,0.5)
200 m breaststroke	1.52	−0.522	0.054	0.031	0.002	[0.001,0.005)	[0.001,0.005)
200 m ind. medley	1.661	−0.103	0.076	0.022	0.002	≥0.5	≥0.5
400 m ind. medley	1.555	−0.14	0.08	0.022	0.003	[0.25,0.5)	[0.25,0.5)

AD: Anderson-Darling; CVM: Cramér-von Mises.

it corresponds to one of the greatest performances of all time. In other ranks, we can observe notable differences between the two criteria and eventually in other occasions the two criteria may point out different swimmers for the first position. The 95% C.I. for the exceedance probabilities presented in the table is computed by bootstrap methods (10,000 resamples).

Conducting the same analysis considering 160, 170, 180, 190, and 200 exceedances, the rankings are similar. For all number of exceedances, the rankings are composed by the same 40 performances shown in Table 4. Few changes in order are observed. For instance, considering 200 exceedances, only one switch of positions is observed in the women's ranking, and four are observed in the men's ranking.

GPD's adequacy

A question that arises is to know if the GPD performance index is more reliable than the FINA one. Also, it

Table 4. Performance index comparisons: GPD and FINA.

Athlete	Event	Time	Competition	$1-\hat{p}$	95% C.I. for $1-\hat{p}$	GPD rank	FINA points	FINA rank
Women								
K. Ledecky (USA)	800 m freestyle	8:04.79	Olympic Games	0.9996	(0.9962,1.0000)	1	1016	3
K. Ledecky (USA)	400 m freestyle	3:56.46	Olympic Games	0.9983	(0.9913,1.0000)	2	1024	1
K. Hosszu (HUN)	400 m ind. medley	4:26.36	Olympic Games	0.9976	(0.9889,1.0000)	3	1023	2
C. Campbell (AUS)	100 m freestyle	52.06	AUS Grand Prix	0.9968	(0.9896,1.0000)	4	1001	5
S. Sjoström (SWE)	100 m butterfly	55.48	Olympic Games	0.9959	(0.9860,1.0000)	5	1009	4
K. Hosszú (HUN)	200 m ind. medley	2:06.58	Olympic Games	0.9909	(0.9790,0.9996)	6	989	6
K. Ledecky (USA)	200 m freestyle	1:53.73	Olympic Games	0.9894	(0.9753,0.9992)	7	980	11
S. Sjoström (SWE)	50 m butterfly	24.99	European Champs.	0.9880	(0.9739,1.0000)	8	934	80
S. O'Connor (GBR)	200 m ind. medley	2:06.88	Olympic Games	0.9876	(0.9742,0.9987)	9	982	10
R. Kaneto (JPN)	200 m breaststroke	2:19.65	Japan Nationals	0.9868	(0.9749,0.9975)	10	988	7
C. Campbell (AUS)	50 m freestyle	23.84	AUS Olympic Trials	0.9847	(0.9728,0.9955)	11	986	8
S. Sjoström (SWE)	200 m freestyle	1:54.08	Olympic Games	0.9828	(0.9659,0.9974)	12	971	15
L. King (USA)	100 m breaststroke	1:04.93	Olympic Games	0.9814	(0.9665,0.9964)	13	973	12
K. Hosszú (HUN)	100 m backstroke	58.45	Olympic Games	0.9768	(0.9614,0.9910)	14	983	9
B. Campbell (AUS)	100 m freestyle	52.58	AUS Olympic Trials	0.9751	(0.9575,0.9958)	15	971	15
L. Smith (USA)	400 m freestyle	4:00.65	USA Olympic Trials	0.9717	(0.9495,0.9947)	16	972	14
F. Pellegrini (ITA)	200 m freestyle	1:54.55	Trofeo Sette Colli	0.9687	(0.9462,0.9899)	17	959	27
R. Meilutyte (LTU)	50 m breaststroke	29.98	Mare Nostrum	0.9686	(0.9486,0.9950)	18	950	48
S. Manuel (USA)	100 m freestyle	52.70	Olympic Games	0.9648	(0.9439,0.9897)	19	964	23
P. Oleksiak (CAN)	100 m freestyle	52.70	Olympic Games	0.9648	(0.9435,0.9901)	19	964	23
Men								
A. Peaty (GBR)	100 m breaststroke	57.13	Olympic Games	0.9989	(0.9926,1.0000)	1	1042	1
M. Phelps (USA)	200 m ind. medley	1:54.66	Olympic Games	0.9913	(0.9787,0.9997)	2	983	11
G. Paltrinieri (ITA)	1500 m freestyle	14:34.04	Olympic Games	0.9911	(0.9799,0.9999)	3	990	6
A. Peaty (GBR)	50 m breaststroke	26.61	Olympic Games	0.9909	(0.9796,1.0000)	4	979	16
R. Murphy (USA)	100 m backstroke	51.85	Olympic Games	0.9884	(0.9778,0.9970)	5	1005	2
J. Prenot (USA)	200 m breaststroke	2:07.17	USA Olympic Trials	0.9877	(0.9776,0.9991)	6	996	3
C. McEvoy (AUS)	100 m freestyle	47.04	AUS Olympic Trials	0.9873	(0.9761,0.9979)	7	992	5
K. Hagino (JPN)	200 m ind. medley	1:55.07	Japan Nationals	0.9864	(0.9711,0.9984)	8	972	21
K. Hagino (JPN)	400 m ind. medley	4:06.05	Olympic Games	0.9858	(0.9711,0.9978)	9	973	20
I. Watanabe (JPN)	200 m breaststroke	2:07.22	Olympic Games	0.9852	(0.9750,0.9980)	10	995	4
A. Govorov (UKR)	50 m butterfly	22.69	Open de France	0.9829	(0.9683,0.9975)	11	966	27
L. Cseh (HUN)	200 m butterfly	1:52.91	European Champs.	0.9807	(0.9639,0.9977)	12	963	30
C. Kalisz (USA)	400 m ind. medley	4:06.75	Olympic Games	0.9794	(0.9620,0.9950)	13	965	29
J. Schooling (SIN)	100 m butterfly	50.39	Olympic Games	0.9794	(0.9627,0.9976)	14	966	27
D. Plummer (USA)	100 m backstroke	52.12	USA Olympic Trials	0.9772	(0.9635,0.9904)	15	990	6
M. Horton (AUS)	400 m freestyle	3:41.55	Olympic Games	0.9735	(0.9556,0.9904)	16	980	15
R. Murphy (USA)	100 m backstroke	1:53.62	Olympic Games	0.9724	(0.9537,0.9918)	17	955	35
S. Yang (CHN)	400 m freestyle	3:41.68	Olympic Games	0.9709	(0.9520,0.9889)	18	978	18
D. Balandin (KAZ)	200 m breaststroke	2:07.46	Olympic Games	0.9696	(0.9562,0.9892)	19	989	8
C. Jaeger (USA)	1500 m freestyle	14:39.48	Olympic Games	0.9679	(0.9487,0.9897)	20	971	23

GPD: generalized Pareto distribution; FINA: International Swimming Federation.

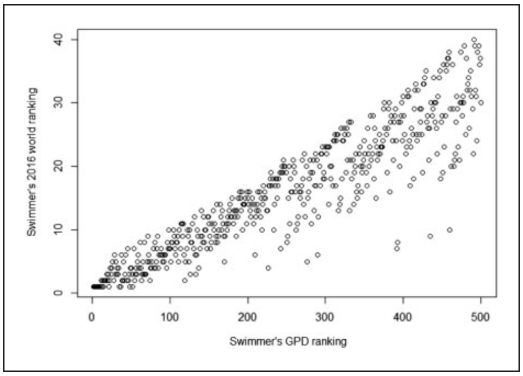

Figure 6. GPD adequacy: GPD × 2016 world rankings (women's events).

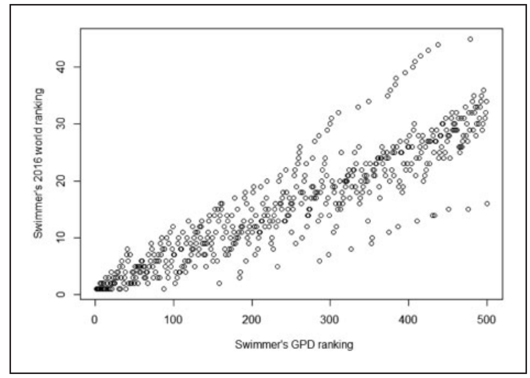

Figure 8. GPD adequacy: GPD × 2016 world rankings (men's events).

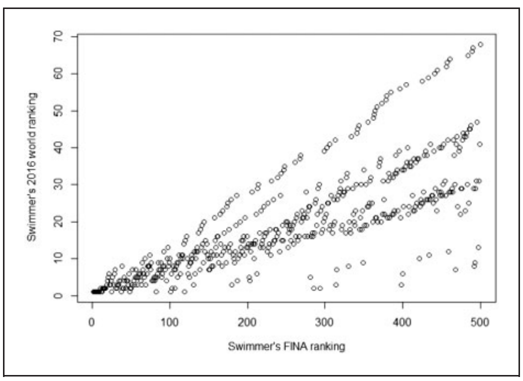

Figure 7. FINA adequacy: FINA × 2016 world rankings (women's events).

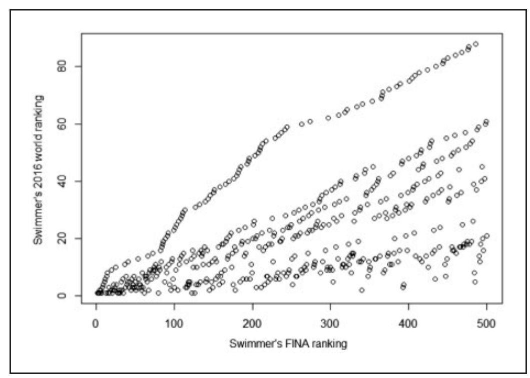

Figure 9. FINA adequacy: FINA × 2016 world rankings (men's events).

is important to know which of the "global rankings" offered by the two methods is in line with the annual world rankings of each event. For instance, in women's events, if the global ranking reflected exactly the 2016 world rankings, we would have the 17 swimmers best ranked in each event in the first 17 positions. In the subsequent 17 positions (18 to 34), we would have the 17 swimmers that ended 2016 ranked in second place in each event, and so on.

Consequently, in this ideal situation, the correlation between the 2016 world rankings and the global rankings would be close to 1.

Obviously we will never have such ideal situation, but we expect that the higher the correlation between the variables, the more adequate the global ranking (given by FINA and GPD methods) is.

Figures 6 to 9 show the scatter plots between the global rankings (GPD/FINA) and the 2016 world rankings, for women and men's events, for the 500 best ranked swimmers in their respective global ranking.

The correlation between the GPD ranking and the 2016 world rankings in women's events is 0.914. The bigger discrepancy is, according to GPD, a swimmer placed in 38th (100 m backstroke) has a better performance than a swimmer placed in 10th (1500 m freestyle) in the 2016 world rankings.

The correlation between the FINA points system and the 2016 world ranking in women's events is 0.781. The bigger discrepancy is, according to FINA points system, a swimmer placed in 67th (100 m backstroke) has a better performance than a swimmer placed in 8th (50 m backstroke) in the 2016 world rankings.

The correlation between the GPD ranking and the 2016 world ranking in men's events is 0.894. The bigger discrepancy is, according to GPD, a swimmer placed in 43th (100 m breaststroke) has a better performance than a swimmer placed in 14th (50 m butterfly) in the 2016 world rankings.

The correlation between the FINA points system and the 2016 world ranking in women's events is 0.566. The bigger discrepancy is, according to FINA points system, a swimmer placed in 87th (200 m breaststroke) has a better performance than a swimmer placed in 5th (50 m butterfly) in the 2016 world rankings.

It can be seen that the correlations are higher and the discrepancies are smaller when the GPD rankings are considered, in comparison to the FINA rankings.

Comparing different sports

Several comparisons can be conducted and it is possible to compare different sports.

A natural comparison that arises is whether the runner Usain Bolt of Jamaica is better than the swimmer Michael Phelps of United States. The question is subjective and will never have a definitive answer. However, using the results of this paper, it is possible to check who has ever registered the best result in comparison to their rivals.

The following comparison will consider two of the most spectacular results registered by the athletes: the 100 m world record established by Bolt in the 2009 World Championships (9.58) and the 200 m individual medley world record established by Phelps in the 2003 USA Nationals (1:55.94).

According to the procedure already described, GPDs were adjusted to the two cases, considering 150 exceedances:

- Usain Bolt (2009, 100 m), we have got $u = 9.9300$, $(\mathrm{SE}(\hat{\theta}) = 0.0100)$ and $\hat{\xi} = -0.0613$ $(\mathrm{SE}(\hat{\xi}) = 0.0714)$.
- Michael Phelps (2003, 200 m individual medley), we have got $u = 1.6183$, $\hat{\theta} = 0.0213$ $(\mathrm{SE}(\hat{\theta}) = 0.0021)$ and $(\mathrm{SE}(\hat{\theta}) = 0.0616)$.

The AD and Kolmogorov-Smirnov p-values are higher than 0.5 in both cases, indicating that the adjusted distributions are reasonable for the data.

When we compare the exceedance probabilities, $1 - \hat{p}$, we got for Michael Phelps' time in 200 m individual medley the value 0.9993 (95% bootstrap I.C.: (0.9918,1.000)), slightly higher than the one of Usain Bolts' 100 m, 0.9988 (95% bootstrap I.C.: (0.9928,1.000)). Even though the difference is not statistically significant, the estimates of the exceedance probabilities allow us to rank the performances, similarly to the FINA points system. So, according to the GPD criterion, Michael Phelps' performance is better than the one of Usain Bolt's. Using FINA criterion Phelps' performance also would be better (1058×1034).

Conclusions

The FINA points system used for comparing performances from different swimming events is based solely in world records. In quest of an alternative criterion, it is proposed a new system, based on the probability distribution of the best performances of the 100 fastest swimmers of history in each event. Using extreme value theory, it is possible to get the approximate distribution of such data. Under certain conditions, the distribution of data that exceeds a threshold is GPD, assuming that the threshold is high enough.

The performance index based on the GPD was computed, considering swimming performances until the end of 2016. Each swimming time was used with the correspondent event's GPD, and the performance index was computed by exceedance probability. We argue that the proposed index is more reasonable than the FINA point system, which has no theoretical foundations and considers that world records of different events have the same difficulty level. Also, it is possible to use the proposed methodology in other sports, and even to compare performances between different sports.

Acknowledgements

The authors are very grateful to the referees for their careful reading of the paper and their constructive remarks and suggestions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The second author was partially supported by FCT – Fundação para a Ciência e a Tecnologia, project UID/MAT/0006/2013 (CEA/UL).

ORCID iD

Daniel T Gomes http://orcid.org/0000-0002-3106-861X

References

1. FINA. FINA/airwave Swimming World Cup 2016 – rules & regulations, www.fina.org/sites/default/files/fina-swc-2016_rulesregulations__final_v3.pdf (2016, accessed 28 June 2018).
2. CBDA. Tabela pontos – FINA – cálculo índice técnico, www.cbda.org.br/cbda/natacao/boletins/8891/tabela-pontos-fina-calculo-indice-tecnico (2016, accessed 28 June 2018).
3. Gyarfas T. What a really wonderful world. *FINA Aquatics World Magaz* 2016; 6: 10–13.
4. FINA. FINA points table, www.fina.org/sites/default/files/fina_points_table_1.pdf (2011, accessed 28 June 2018).
5. Bernhardt R. FINA points: illogical and impractical. *SwimSwam*, https://swimswam.com/fina-points-illogical-impractical (2014, accessed 28 June 2018).

6. Smith RL. Forecasting records by maximum likelihood. *J Am Stat Assoc* 1988; 83: 331–338.
7. Robinson ME and Tawn JA. Statistics for exceptional athletics records. *J R Stat Soc. Series C (Appl Stat)* 1995; 44: 499–511.
8. Barão MI and Tawn JA. Extremal analysis of short series with outliers: sea-levels and athletics records. *Appl Stat* 1999; 48: 469–487.
9. Einmahl JH and Magnus JR. Records in athletics through extreme-value theory. *J Am Stat Assoc* 2008; 103: 1382–1391.
10. Henriques-Rodrigues L, Gomes MI and Pestana D. Statistics of extremes in athletics. *Revstat* 2011; 9: 127–153.
11. Adam MB and Tawn JA. Bivariate extreme analysis of Olympic swimming data. *J Stat Theory Pract* 2012; 6: 510–523.
12. Stephenson AG and Tawn JA. Determining the best track performances of all time using a conceptual population model for athletics records. *J Quant Anal Sports* 2013; 9: 67–76.
13. Reiss RD and Thomas M. *Statistical analysis of extreme values*, 2nd ed. Basel: Birkhauser, 2001.
14. Beirlant J, Goegebeur Y, Segers J, et al. *Statistics of extremes: theory and applications*. London: Wiley, 2004.
15. De Haan L and Ferreira A. *Extreme value theory: an introduction*. New York: Springer Series in Operations Research and Financial Engineering, 2006.
16. Balkema AA and de Haan L. Residual life time at great age. *Ann Probab* 1974; 2: 792–804.
17. Pickands J. Statistical inference using extreme order statistics. *Ann Stat* 1975; 3: 119–131.
18. Wadsworth JL, Tawn A and Jonathan P. Accounting for choice of measurement scale in extreme value modeling. *Ann Appl Stat* 2010; 4: 1558–1578.
19. Vicente SLG. *Extreme value theory: an application to sports*. Master's dissertation, Lisbon University, Portugal, 2012.
20. Davison AC and Smith RL. Models for exceedances over high thresholds. *J R Stat Soc Series B* 1990; 52: 393–442.
21. Coles S. *An introduction to statistical modeling of extreme values*. London: Springer, 2001.
22. Bermudez PZ and Kotz S. Parameter estimation of the generalized Pareto distribution – part I. *J Stat Plan Inference* 2010; 140: 1353–1373.
23. Choulakian V and Stephens MA. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* 2001; 43: 478–484.
24. Sheinin D. How good is Katie Ledecky in 800 meters? The swimming world can't imagine. *The Washington Post*, www.washingtonpost.com/sports/olympics/how-good-is-katie-ledecky-in-800-meters-the-swimming-world-can-hardly-imagine/2016/08/11/35d00720-5fb8-11e6-8e45-477372e89d78_story.html?utm_term=.5182a55134d3 (2016, accessed 28 June 2018)).