Return period of extreme rainfall at George, South Africa

Jean-Luc Mélice[a*] **and Chris J.C. Reason**[b]

The torrential rains of August 2006 in the southern Cape of South Africa were the most intense observed in the region. Here we use the longest-available daily rainfall series at George (from 1941 to 2006), in the vicinity of which the most destructive floods were observed, together with an extreme value model to estimate the return period of such an extreme event. According to this model, the greatest annual maximum daily rainfall of 230 mm, observed at the town on 1 August 2006, has a return period of 1222 years, whereas the second-largest observed annual maximum daily rainfall (132 mm in September 1964) has a return period of 23 years. This shows that the August 2006 extreme rainfall at George can be considered as a particularly rare event.

Introduction

The floods of August 2006 in the southern Cape region of South Africa were the most destructive observed in the region. The flooding resulted in the loss of almost a dozen lives and caused much damage to housing, roads and other infrastructure along the so-called 'Garden Route'. The section of the N2 highway near George was severely damaged by the floods and was closed to traffic. This damage isolated one of the country's major transport arteries and had serious economic implications for Garden Route towns. The extreme rainfall at George and elsewhere on the southern Cape coast on 1 August, 2006 seems to have resulted from an intense cold front with a secondary depression developing behind it (with a cut-off low at mid- and upper levels), and a strong anticyclone extending from the subtropics to well south of 55°S farther to the west. As a result, the air advected northwards towards the southern Cape mountains behind the front was very cold (there were extensive snowfalls over the interior on 1–2 August) but was strongly de-stabilized on crossing the warm Agulhas Current. Strong uplift of this low-level flow on approaching the coastal mountains led to heavy falls at George and vicinity as has been experienced at other times in the southern Cape when a low-level jet interacts with the topography.[1,2]

In this paper, we use the 1941–2006 daily rainfall series at George (33°58'S, 22°29'E) to estimate the return period of these extreme floods. This rainfall series is the longest available for the region and it was in the vicinity of the town that the most destructive floods were observed.

Extreme value theory provides efficient techniques for estimating probabilities of future extreme levels of a process given historical data. There are many variations on this theme, but the simplest is when the historical data consist of a sequence of annual maximum observations. Standard extreme value arguments imply that a family of distributions with which to model such data is the extreme value family. These distributions are among the most common probabilistic models used for hydrological and meteorological extremes and have been widely employed for quantifying risk associated with extreme rainfall.[3] From the statistical theory of extremes, we show that the first extreme value, or Gumbel, distribution fits the annual maximum daily rainfall at George very well and can therefore be useful for estimating the return period of rainfall extremes. Before showing these results, we first summarize the theory of extremes.

Statistical distribution of extremes

A series of size N (here N is the number of days from 1941 to 2006) is given as a sample of a random variable x (x being the daily rainfall at George). This sample is divided into n subsamples each of size m (n years of $m \cong 365$ days), so that $N = nm$. From each subsample, the extreme (the largest value) is selected, so that the n subsamples of size m provide a new sample of size n (the n extremes of the n years of record). With very few restrictive considerations concerning the law of the x variable, the statistical distribution of the series of the n extremes approaches asymptotically a simple probability law: the Fisher-Tippett 3 asymptotes, also called extreme values distributions EV1, EV2, and EV3.[4]

The EV1 distribution holds for an initial distribution of the exponential type and the EV2 for the 'Cauchy' type. In both cases, the initial variables are unlimited to the right (i.e. for the largest value), or in both directions. Thus, if the x variable initial distribution belongs to the exponential type, the statistical distribution of the n extremes described above will tend asymptotically to the first EV1 distribution. Most types of parent distribution functions that are used in hydrology and meteorology, such as exponential, gamma, Weibull, normal, and lognormal belong to the exponential type and to the domain of attraction of the EV1 distribution, which is also commonly named the Gumbel distribution. If the x variable initial distribution belongs to the Cauchy type, the statistical distribution of the extremes will tend asymptotically to the EV2 distribution. In contrast, the domain of attraction of the EV2 distribution includes less commonly met parent distributions like Pareto, Cauchy, and log-gamma. The EV3 distribution, which holds for initial distributions bounded towards the right, has no physical significance in most practical applications of extreme values, as the extremes would be limited in this case.

The EV1 distribution takes the form:

$$F(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right],$$

where $F(x)$ is the cumulative distribution function, μ is the location parameter, and σ is the scale parameter.

The EV2 distribution takes the form:

$$F(x) = \exp\left[-\left(\frac{\alpha-\beta}{x-\beta}\right)^{\gamma}\right],$$

where $F(x)$ is the cumulative distribution function, $x \geq \beta, 0 < \beta \leq \alpha$, and $\gamma > 0$. A special case is obtained when the left bound β becomes zero. This can be justified if we consider that the lowest extreme rainfall value can be equal to zero. In such a case it can be seen that the logarithm of the EV2 x variable follows an EV1 distribution.

In practical applications, the initial sampling distribution and its parameters are usually unknown, but we need not worry about its analytical expression because EV1's and EV2's parameters are estimated from the set of observed extreme values. The choice between EV1 and EV2 depends only upon the best fit of both distributions to the observed values.

[a]Institut de Recherche pour le Développement and Department of Oceanography, University of Cape Town, Rondebosch 7701, South Africa.
[b]Department of Oceanography, University of Cape Town.
*Author for correspondence. E-mail: jean-lucmelice@uct.ac.za

Fitting the annual maximum daily rainfall series at George with the EV1 distribution

The 66-year-long daily rainfall series at George from January 1941 to December 2006, is presented in Fig. 1. The annual maximum daily rainfall is extracted for each year (dots in the figure). The distribution of the annual maxima series was first studied using the EV1 distribution with the Gumbel probability plot. This plot is conceptually similar to the well-known normal probability plot. When the relation between the variable x and the reduced variable y is: $y = (x - \mu)/\sigma$, the graphical representation of this relation in a Cartesian linear x and y coordinated system is a straight line. Moreover, if the corresponding $F(y)$ values are reproduced along the linear y-axis, one obtains a graphical representation where the x scale is linear, the $F(y)$ scale is functional, and $x = x(F)$ is still represented by a straight line.

To construct the Gumbel probability plot (Fig. 2a), each value of the series of the $n = 66$ maxima is ranked i, and then plotted against its reduced value y_i. This reduced value is the double negative $\log_e$ expression of the datum rank – which is the distribution function for the EV1 distribution. It is given as follows: $y_i = -\log_e[-\log_e(p_i)]$, where $\Phi(y_i) = i/(n-1)$ is the observed cumulative function for the ascending series. The best fit for the EV1 distribution is obtained when all the points $[x_i, \Phi(y_i)]$ themselves lie on the a straight line which, theoretically, is represented by $x = \mu + \sigma y$. The parameters μ and σ are estimated using the maximum likelihood method.[5]

As seen in Fig. 2a, the annual maximum daily rainfall series is well fitted by the EV1 distribution. Statistical verification of the fit is given by the Kolmogorov-Smirnov test[6] and by a more precise test based on the extreme values and the median[5] (see the Appendix). The first test points out the largest difference between the estimated theoretical and observed cumulative frequency functions: $\Delta_i = |F(x_i) - \Phi(y_i)|$. The largest difference, $\Delta_1 = 0.094$, is smaller than $\Delta_0 = 1.07/\sqrt{n} = 0.13$, which is the largest theoretical difference acceptable by the test at the $\alpha_0 = 0.01$ level. The second test shows that the probabilities for the lowest value, the median, and the largest value of the sample are $\alpha'_1 = 0.39$, $\alpha'_2 = 0.15$, and, $\alpha'_3 = 0.11$, respectively. The probability for the Fisher test is $\alpha_1 = 0.12$. This means that the adjusted distribution is not rejected at the $\alpha = 0.05$ level.

One might suspect that 2006's largest rainfall extreme of 230 mm does not belong to the adjusted distribution. To verify that this value is admissible, we see that for $x = 230$, we have $F(y) = 0.9992$. In the series of 66 observations, the probability of the largest value being the greatest of the observed sample is $0.9992^{66} = 0.9485$. This value lies in the confidence interval [0.025, 0.975] at the $\alpha_0 = 0.05$ level. There is therefore no reason to suspect the largest 230 mm value. We then tested the stability of EV1 fit when excluding the 2006 extreme rainfall. The parameters for the entire period are: $\mu = 56.0$ and
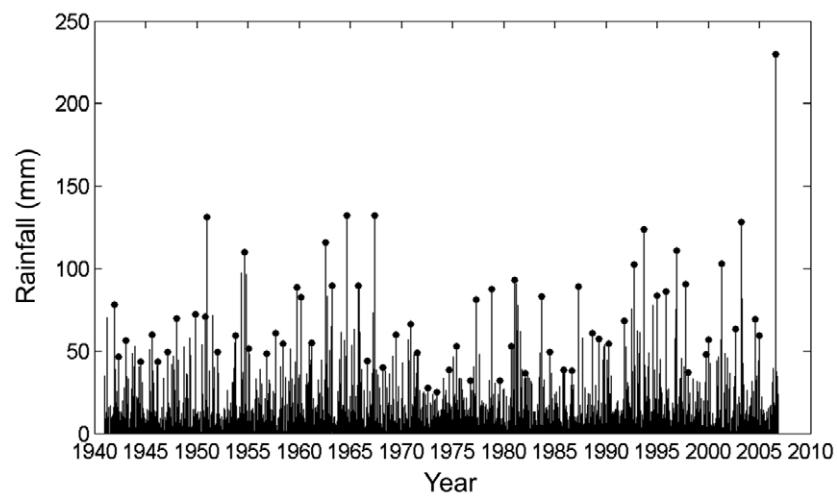

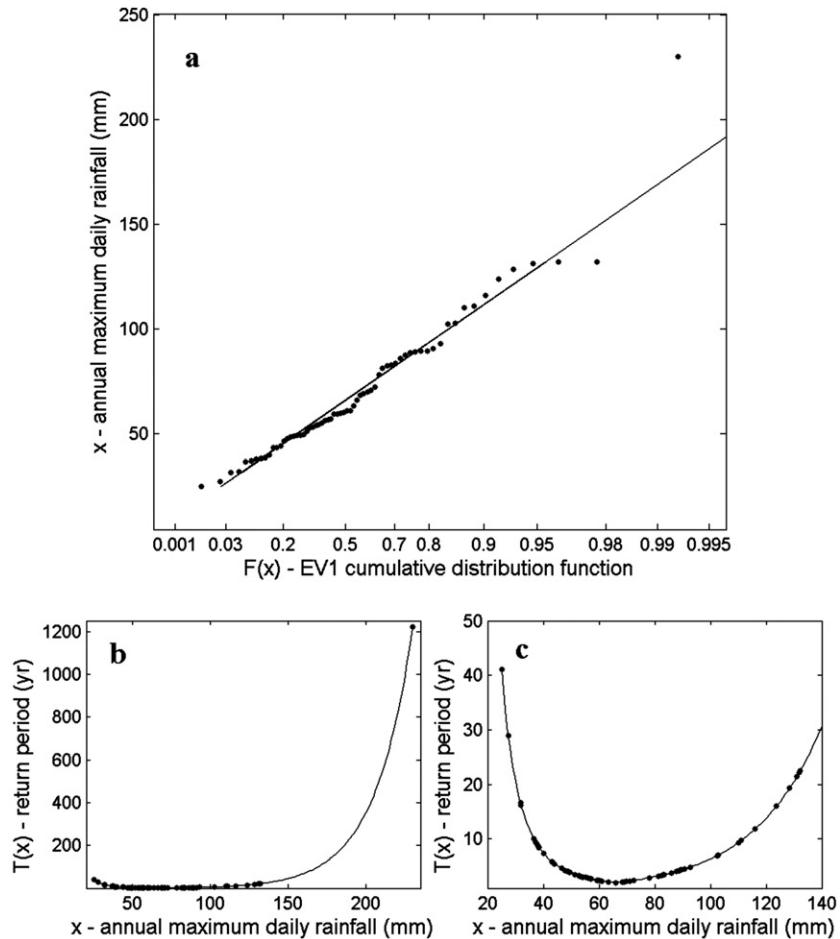**Fig. 1**. Daily rainfall and annual daily rainfall maxima (dots) at George, from January 1941 to December 2006.

Fig. 2. (**a**) Annual daily rainfall maxima at George from 1941 to 2006, adjusted to the first extreme value (Gumbel) distribution. (**b**) Return period (in years) of the maxima; (**c**) same as (**b**) but with 2006 maximum excluded.

$\sigma = 24.4$. When excluding 2006, they are $\mu = 55.6$ and $\sigma = 22.2$. The parameters change only slightly as a consequence of the inclusion of the 2006 data.

The annual maxima were then investigated with the simplified form of the EV2 distribution with $\beta = 0$. This simplified form is often used to test which of two distributions has to be chosen. By contrast to the EV1, the EV2 does not fit the empirical distribution well and is rejected at the $\alpha_0 = 0.05$ level by the test based on the extreme values and the median.

Return period of the annual maximum rainfall

The return period is defined here as the average time interval (expressed in years) between occurrences of a rainfall event of a given or greater magnitude. The return period denotes a recurrence interval. It is a statistical measure of how often a rainfall event of a certain magnitude is likely to happen. It is important in relating extreme rainfall to normal rainfall. Rainfall with a 10-year return period is expected to happen only every 10 years. A 100-year return period corresponds to such an extreme event that we expect it to occur only every 100 years. The return period is expressed as: $T = 1/F$ for $F \leq 0.5$ and $T = 1/(1 - F)$ for $F \geq 0.5$, where F is the cumulative distribution function. In our case, the annual rainfall maxima series is fitted to an EV1 distribution. The return periods of the annual maxima are displayed in Fig. 2b for the entire series and, for reason of readability, in Fig. 2c, where the 2006 maximum is omitted.

The EV1-estimated return period of the 2006 daily rainfall maximum of 230 mm is 1222 years (Fig. 2b). The previously recorded maximum of 132 mm occurred in September 1964. Its magnitude is about half of the 2006 event and its return period is 23 years (Fig. 2c). We note that there are 5 annual maxima of about this magnitude in the record (Fig. 1). The rainfall magnitudes corresponding to the return periods of 10, 100, and 1000 years are 112, 170, and 225 mm, respectively.

Conclusion

The extreme daily precipitation event of 230 mm at George in August 2006 was almost twice the magnitude of the previously recorded maximum observed at the town. This extreme rainfall was probably the main cause of the destructive floods which occurred that day in the region. Extreme value theory was used to assess the likely return period of such extreme rainfall. We show that an extreme event of this magnitude, or greater, can be expected to occur, on average, once every 1222 years. We note that ~1000 years is not an unusual return period for the design of major flood protection work.[7] The second-largest observed annual maximum daily rainfall (132 mm in September 1964) has a return period of 23 years. This indicates that the August 2006 extreme rainfall at George can be considered as a particularly rare event.

1. Singleton A.T. and Reason C.J.C. (2006). Numerical simulations of a severe rainfall event over the Eastern Cape coast of South Africa: sensitivity to sea surface temperature and topography. *Tellus A* **58**, 355–367.
2. Singleton A.T. and Reason C.J.C. (2007). A numerical model study of an intense cut-off low pressure system over South Africa. *Mon. Wea. Rev.* **135**, 1128–1150.
3. Coles S. and Pericchi L. (2003). Anticipating catastrophes through extreme value modelling. *Appl. Statist.* **52**, 405–416.
4. Coles S. (2001). *An Introduction to Statistical Modeling of Extremes Values.* Springer, New York.
5. Sneyers R. (1990). *On the statistical analysis of series of observations.* WMO, Technical Note No. 143, p. 192. Geneva.
6. Chakravarti I.M., Laha R.G., and Roy J. (1967). *Handbook of Methods of Applied Statistics,* vol. 1, pp. 392–394. John Wiley, New York.
7. Koutsoyiannis D. (2003). On the appropriateness of the Gumbel distribution in modelling extreme rainfall. In *Hydrological Risk: recent advances in peak river flow modelling, prediction and real-time forecasting. Assessments of the impacts of land-use and climate changes,* pp. 303–319. *Proc. ESF LESC Exploratory Workshop,* Bologna, 24–25 October.

Appendix

Goodness of the fit test based on the extreme values and the median

Statistical verification of the goodness of the fit can be driven by testing the two extremes and the median, in the following way.[5] If x_1, x_m, x_n are respectively the lowest value, the median and the largest value of the following sample arranged in order of magnitude: $x_1 \leq \dots \leq x_m \leq \dots \leq x_n$, and if $F(x)$, $F(x_m)$, $F(x_n)$ are the cumulative distribution functions of these three values estimated by a certain distribution, then we can draw the following conclusions:

(i) The probability of the first observed value x_1 being the lowest of the observed sample is: $\alpha_1 V = [1 - F(x)]^n$. This value of x_1 is suspect and the adjusted distribution will be rejected at the α_0 significance level if $\alpha_1 \leq \alpha_0/2$ or $\alpha_1 \geq 1 - \alpha_0/2$.

(ii) As $F(x_m) = \beta$ is approximately normally distributed, $N(\mu, \sigma)$, with $\mu = 0.5$ and $\sigma = [(0.5)(0.5)/(n + 2)]^{1/2}$, it is possible to compute the probability $\alpha_2 = P[F(x_m) < \beta]$. The median x_m is then suspect and the adjusted distribution will be rejected at the α_0 significance level if $\alpha_2 \leq \alpha_0/2$ or $\alpha_2 \geq 1 - \alpha_0/2$.

(iii) The probability of the largest observed value x_n being the largest of the observed sample is: $\alpha_3 = [F(x_n)]^n$. This value of x_n is suspect and the adjusted distribution will be rejected at the α_0 significance level if $\alpha_3 \leq \alpha_0/2$ or $\alpha_3 \geq 1 - \alpha_0/2$.

Moreover, let us define α'_1 when $\alpha_i \leq 0.5$ or $\alpha'_i = 2[1 - \alpha_i]$ when $\alpha_i > 0.5$, with $i = 1, 2, 3$. The distributions of the 2 extreme values and the median being asymptotically independent, it is possible to combine the three probabilities α'_1, α'_2, and α'_3 by a Fisher global test; this test states that $X = -2 \sum_{i=1}^{k} \ln \alpha'_i$ follows a chi-squared distribution with $2k$ degrees of freedom (in this case, $k = 3$.). The adjusted distribution is finally rejected at the α_0 significance level when $P(\chi_{2k}^2 > X) \leq \alpha_0$.