Extreme value theory for precipitation: sensitivity analysis for climate change

R.W. Katz 1

Environmental and Societal Impacts Group, National Center for Atmospheric Research, Boulder, CO 80307, USA

Received 6 October 1997; received in revised form 6 September 1998; accepted 6 April 1999

Abstract

Extreme value theory for the maximum of a time series of daily precipitation amount is described. A chain-dependent process is assumed as a stochastic model for daily precipitation, with the intensity distribution being the gamma. To examine how the effective return period for extreme high precipitation amounts would change as the parameters of the chain-dependent process change (i.e., probability of a wet day, shape and scale parameters of the gamma distribution), a sensitivity analysis is performed. This sensitivity analysis is guided by some results from statistical downscaling that relate patterns in large-scale atmospheric circulation to local precipitation, providing a physically plausible range of changes in the parameters. For the particular location considered in the example, the effective return period is most sensitive to the scale parameter of the intensity distribution. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Chain-dependent process; Design value; Gamma distribution; Markov chain; Return period

1. Introduction

Physical considerations and experiments using numerical models suggest an intensification of the hydrologic cycle as part of the enhanced greenhouse effect [10]. Empirical evidence from observations indicates increasing trends in precipitation in many regions [18], as well as consistent patterns in other hydrologic variables including evaporation [3]. Other results are not as definitive, but more subtle effects such as an increased frequency of intense precipitation have been detected [9]. Clearly, our methods of analysis of the precipitation process need to be examined in detail, not simply considering monthly/seasonal totals, but also frequency of precipitation occurrence and parameters of the intensity distribution.

Stochastic models for precipitation are especially useful for investigating issues such as those that arise with climate change. They can account for the essential features of the precipitation process; namely, its intermittency, the tendency of wet or dry spells to persist, and the positively skewed distribution of intensity. A particularly simple form of stochastic model for precipitation is known as a “chain-dependent process” [11]. It involves a first-order Markov chain for the occurrence of precipitation [23], and the assumption that the amounts are conditionally independent and identically distributed given whether or not it has occurred. Of course, many more complex forms of stochastic model for precipitation exist [21], but this complexity makes their use in climate change studies somewhat problematic. In particular, such models involve parameters about which insufficient information is available, either from observations or from numerical models of the climate system.

The definition of a chain-dependent process and its fit to an example are described in Section 2. Then extreme value theory is reviewed in Section 3, with the key idea being to explicitly account for the fact that such extremes involve a random number of individual precipitation intensities (some theoretical details are relegated to the appendix). Making use of this theory, a sensitivity analysis is performed in Section 4 to examine how the effective return period for extreme high precipitation amounts (loosely speaking, “flood” events) would change as the various parameters of the chain-dependent process change (e.g., probability of a wet day). To make the sensitivity analysis more physically meaning-
ful, it is guided by some results from statistical downscaling, in which the parameters of the chain-dependent process are allowed to vary conditionally on an index of large-scale atmospheric circulation [14]. Finally, Section 5 consists of a discussion and conclusions.

2. Stochastic model

A chain-dependent process has the desirable feature of requiring only a relatively small number of parameters, while still accounting for the most important characteristics of precipitation time series [11,20]. Its simple structure enables the analytical determination of many of its properties, including extreme values [12]. Ignoring any annual cycles by restricting consideration to a single month or season, the stochastic model for the precipitation process is assumed stationary. Nevertheless, the sensitivity analysis to be performed involves varying the model parameters, in effect, reflecting a change to another stationary climate.

The tendency of wet spells (i.e., runs of consecutive days on which precipitation occurs) or of dry spells to persist is represented by a two-state, first-order Markov chain model for daily precipitation occurrence [23]. Let \( J_t = i, i = 1, 2, \ldots \) denote the sequence of daily precipitation occurrence (i.e., \( J_t = 1 \) indicates a “wet day,” \( J_t = 0 \) a “dry day”). The Markov property consists of the future state of the chain being conditionally independent of all the past states given the present state; that is,

\[
\Pr\{J_{t+1} = j | J_t = i, J_{t-1}, \ldots, J_1\} = \Pr\{J_{t+1} = j | J_t = i\} \equiv P_{ij}, \quad i, j = 0, 1.
\]

Here the \( P_{ij} \)'s are termed transition probabilities and completely characterize the model. For example, \( P_{01} \) represents the conditional probability that precipitation will occur tomorrow given that it did not occur today.

It is convenient to reparameterize the Markov chain in terms of the probability of a wet day, \( \pi = \Pr\{J_t = 1\} \), and the first-order autocorrelation coefficient (or “persistence parameter”), \( d = \text{Corr}(J_t, J_{t+1}) \). These two parameters, \( \pi \) and \( d \), are related to the transition probabilities [Eq. (1)] by

\[
\pi = P_{01} / (P_{01} + P_{10}), \quad d = P_{11} - P_{00}.
\]

The number of occurrences of precipitation (i.e., the number of wet days) within some time period of length \( T \) days (e.g., a month or season) is denoted by \( N(T) \), with \( N(T) = J_1 + J_2 + \cdots + J_T \). It is important to recognize that \( N(T) \) is itself a random variable, fluctuating about the expected number of wet days, \( E[N(T)] = T\pi \). This counting process \( N(T) \) plays a pivotal role in extreme value theory of precipitation amounts.

Let \( Z_k > 0 \) denote the “intensity” corresponding to the \( k \)th occurrence of precipitation (i.e., the amount on the \( k \)th wet day), \( k = 1, 2, \ldots, N(T) \). These intensities are taken independent and identically distributed (i.i.d.), say with common cumulative distribution function (c.d.f.) \( F(z) = \Pr\{Z_k \leq z\} \). A positively skewed distribution, such as the exponential, lognormal, or based on a power transformation (e.g., square root or cube root) to normality, is assumed by \( F \). In the present paper, the gamma distribution is utilized, with probability density function [i.e., \( f(z) = F'(z) \)] given by

\[
f(z) = \left[ (z/\beta)^{x-1} e^{-z/\beta} \right] / \Gamma(x), \quad z, \alpha, \beta > 0.
\]

Here \( \alpha \) and \( \beta \) are the shape and scale parameters, respectively, and \( \Gamma \) denotes the gamma function [23].

The mean of the gamma distribution is related to the parameters \( \alpha \) and \( \beta \) by \( E(Z_k) = \alpha \beta \), and the variance by \( \text{Var}(Z_k) = \alpha \beta^2 \). The shape parameter \( \alpha \) is dimensionless, as it governs only the “shape” of the distribution (e.g., corresponding to the exponential distribution for \( \alpha = 1 \)), with the degree of skewness decreasing as \( \alpha \) increases. On the other hand, the scale parameter \( \beta \), in the units in which precipitation is measured, has no bearing on the shape of the distribution. When a gamma distribution is “rescaled” (i.e., multiplied by a constant), the resultant distribution is still the gamma with the same shape parameter, but a modified scale parameter. For instance, the rescaled random variable \( Z_k / \beta \) has a gamma distribution with shape parameter \( \alpha \) and unit scale parameter.

3. Extreme value theory

3.1. Basic theory

The maximum amount of daily precipitation over a time period of length \( T \) days involves taking the largest of a random number, \( N(T) \), of intensities \( Z_k \); that is,

\[
M_T = \max\{Z_1, Z_2, \ldots, Z_{N(T)}\}.
\]

For the maximum of a fixed, nonrandom number of i.i.d. random variables (whose c.d.f. is any of those mentioned in Section 2 as having been fitted to precipitation intensity), classical extreme value theory includes the following result [17]

\[
\Pr\{aT < x \Rightarrow G(x) \equiv \exp(-e^{-x}), \quad -\infty < x < \infty, \quad a \rightarrow \infty.
\]

Here \( a_T > 0 \) and \( b_T \) are normalizing constants that depend on the exact form of “parent” c.d.f. \( F \) from which the maximum is being sampled. The limiting c.d.f. \( G \) that appears on the right-hand side of Eq. (5) is called the Type I extreme value distribution. It is expressed in standardized form, with zero location parameter and unit scale parameter. If the parent c.d.f. \( F \) is the gamma [Eq. (3)], then one choice of normalizing constants is (see Gumbel [6] for a special case of this result)

\[
a_T = 1/\beta, b_T / \beta = c_T + (x - 1) \ln(c_T),
\]

where \( c_T > 0 \) is the “location” parameter, which precipitation is measured, has no bearing on the shape of the distribution. When a gamma distribution is “rescaled” (i.e., multiplied by a constant), the resultant distribution is still the gamma with the same shape parameter, but a modified scale parameter. For instance, the rescaled random variable \( Z_k / \beta \) has a gamma distribution with shape parameter \( \alpha \) and unit scale parameter.
where
\[ c_T = \ln[T/\Gamma(z)]. \] (7)

In the design of engineered systems to allow for extreme events, it is convenient to introduce the concept of a design value. Suppose that it is desired to determine the value of the maximum of a sequence for which the probability of exceedance is \( p \). That is,
\[ p = \Pr\{M_T > x(p)\}, \] (8)
where \( x(p) \) is termed the design value corresponding to a return period of 1/\( p \) yrs (e.g., a probability of \( p = 0.01 \) corresponds to a return period of 100 yrs). Applying the extreme value approximation Eq. (5) to Eq. (8) and solving for \( x(p) \), the design value for the Type I extreme value distribution can be expressed as (e.g., Ref. [4])
\[ x(p) = b_T - \ln[-\ln(1 - p)]/a_T. \] (9)

3.2. Chain-dependent process

There are two approaches to extending the classical extreme value theory to the situation of precipitation extremes being generated by a chain-dependent process. The first approach (followed by Katz [12]) treats the unconditional c.d.f. of precipitation amount (i.e., both zero and nonzero amounts) as a mixture, involving the intensity c.d.f. as well as the probability of a wet day; that is, with c.d.f.
\[ F(z) = (1 - \pi) + \pi F_1(z), \quad z \geq 0. \] (10)
Here \( F \) now denotes the unconditional c.d.f. of precipitation amount, whereas \( F_1 \) denotes its conditional distribution given occurrence. Except for being weighted by the probability of a wet day \( \pi \), the right-hand tail of the unconditional c.d.f. in Eq. (10) is identical to that for the intensity c.d.f. (i.e., \( 1 - F(z) = \pi[1 - F_1(z)] \)). Hence the same limiting extreme value theory as applies to the c.d.f. \( F \) would be anticipated to apply to the c.d.f. \( F_1 \), except for a modification to the normalizing constants that appear in Eq. (5). This approach also requires use of the result that the classical extreme value theory still holds when the time series is dependent; in particular, under the dependence for precipitation amounts induced by the first-order Markov chain for occurrence. In fact, such a relatively “weak” form of dependence requires no change in the normalizing constants that appear in Eq. (5) [17].

The second approach is somewhat more theoretically appealing, in that it makes explicit use of the representation in Eq. (4) of the maximum of a random number of intensities [1,19]. In particular, this approach makes clear the straightforward extension of extreme value theory to point processes other than a first-order Markov chain (e.g., a higher than first-order chain) for generating the counting process \( N(T) \). Because \( N(T)/T \to \pi \) as \( T \to \infty \) (in probability) for a Markov chain (i.e., a version of the law of large numbers),
\[ \Pr\{a_T(M_T - b_T) \leq x\} \to [\exp(-e^{-x})]^\pi \]
\[ = \exp[-e^{-(x - \ln \pi)}] \quad \text{as} \quad T \to \infty. \] (11)
Here the normalizing constants, \( a_T \) and \( b_T \), are again given by Eq. (6) and Eq. (7), but now the c.d.f. on the right-hand side of Eq. (11) is the Type I extreme value with nonzero location parameter \( \ln \pi \) (i.e., depending on the probability of a wet day).

Either of these two approaches produces the following theoretical result that corresponds to an appealing heuristic as well (see appendix for more detailed derivations). The length of time \( T \) in the expressions for the normalizing constants [Eqs. (6) and (7)] is simply replaced by the expected number of wet days, say \( T^* = T \pi \). In other words, the only change is that now Eq. (5) holds with
\[ c_T = \ln[T\pi/\Gamma(z)], \] (12)
instead of Eq. (7). Because \( c_T \) depends on the probability of a wet day \( \pi \), the design value [Eq. (9)] is a function of three of the parameters of a chain-dependent process, \( \pi, \alpha \) and \( \beta \). It is independent of the remaining parameter \( d \) for the reason mentioned earlier in this subsection.

3.3. Example

A time series of daily precipitation amount in January for 78 yrs during the period 1907–1988 (with four years eliminated because of missing observations) at Chico, California, was analyzed by Katz and Parlange [14]. They fit a chain-dependent process to this data, the only difference being their use of a power transformation (fourth root) to normality, instead of the gamma distribution, for intensity. The transition probabilities of the Markov chain model for occurrence are estimated from the transition counts for precipitation occurrence, \( n_{ij} \), \( i,j = 0,1 \), denoting the number of times in the sample that state \( i \) is followed by state \( j \) (e.g., \( n_{00} \) denotes the number of times a dry day is followed by a wet day). Specifically, the maximum likelihood estimators of the transition probabilities are given by
\[ \hat{P}_{ij} = n_{ij}/(n_{00} + n_{i0}), \quad i,j = 0,1. \] (13)

The approximate maximum likelihood estimates of the shape and scale parameters of the gamma distribution for intensity are obtained by a method suggested by Greenwood and Durand [5]. This method (as well as the exact method) requires the calculation of the statistic
\[ D = \ln \tilde{z} - \sum_k \ln Z_k/N(T), \quad \tilde{z} = \sum_k Z_k/N(T). \] (14)
In other words, \( D \) is the difference between the logarithm of the mean and of the geometric mean for the sample of daily intensities. The estimate \( \tilde{z} \) of the shape parameter is a ratio of polynomials in \( D \) (see Wilks [23]), with the
estimate $\hat{\beta}$ of the scale parameter being obtained by $\hat{\beta} = \bar{z}/\hat{\alpha}$.

The parameter estimates of the fitted chain-dependent process for the January Chico data are as follows:

\[ \hat{\alpha} = 0.329, \quad \hat{\beta} = 0.360, \quad \hat{x} = 0.875, \quad \hat{\beta} = 15.26 \text{ mm}. \]  

Here relationship Eq. (2) has been used to convert the transition probability estimates to the corresponding estimates of the parameters $\pi$ and $d$. So precipitation occurs on roughly one-third of the days in January at Chico, with the conditional probability of tomorrow being a wet day being about 0.36 higher if today were wet than if today were dry (i.e., reflecting the tendency of wet or dry spells to persist). The estimated shape parameter of the gamma distribution is less than one, indicating more skewness in the intensity distribution than that for an exponential distribution, with the sample mean daily intensity being about 13.4 mm.

Next the estimation of design values is considered. When the time period over which the maximum is taken is $T = 31$ days as in the Chico data set, the normalizing constants for the Type I extreme value distribution (substituting the parameter estimates [Eq. (15)] into Eq. (6) and Eq. (12)) are $\hat{a}_{31} = (15.26 \text{ mm})^{-1}$ and $\hat{b}_{31} = 32.59 \text{ mm}$. Using Eq. (9), the estimated design values for January are listed in Table 1 for return periods of 10 and 20 yrs, respectively. A time period of $T = 90$ days will be treated in the sensitivity analysis in Section 4 to approximate the length of the peak wet period in Chico (i.e., roughly December through February). For simplicity, the same parameters of the chain-dependent process for January are used (instead of refitting the model to December and February data as well). In this case, the normalizing constant $\hat{a}_{31} = \hat{a}_{31}$, whereas $\hat{b}_{90} = 48.11 \text{ mm}$, and the estimated design values are necessarily somewhat larger than for January alone (see Table 1).

These estimates of the normalizing constants and design values for the maximum could be termed “theoretical,” because they are derived from a stochastic model for daily precipitation. As a check on this approach, “empirical” estimates are obtained by directly fitting the generalized extreme value distribution (of which the Type I is a special case with zero shape parameter) to the maximum daily precipitation amount in January at Chico (i.e., 78 observed values of $M_{31}$). Maximum likelihood estimates and approximate standard errors [7] (i.e., measures of the uncertainty in the estimates due to the limited sample size) are obtained using the software discussed in Farag and Katz [4]. The estimated shape parameter of the generalized extreme value distribution is quite small (in absolute value) relative to its standard error, indicating that assumption of an approximate Type I distribution for the maximum is reasonable.

Empirical estimates of the normalizing constants for the Type I extreme value distribution (with standard errors in brackets), again fit by the method of maximum likelihood [4], are

\[ \hat{a}_{31} = (15.55 \text{ mm})^{-1} [s(\hat{a}_{31}) = 1.37 \text{ mm}], \]
\[ \hat{b}_{31} = 29.08 \text{ mm} [s(\hat{b}_{31}) = 1.85 \text{ mm}]. \]  

The theoretical estimates are within two standard errors of these empirical estimates, suggesting that the observed differences could well be attributed solely to sampling fluctuations. Consequently, the corresponding empirical estimates of the design values ($\hat{x}(0.1) = 64.1$ mm and $\hat{x}(0.05) = 75.3$ mm) are reasonably close to the theoretical estimates given in Table 1.

### 4. Sensitivity analysis

#### 4.1. Procedure

A sensitivity analysis is performed to study how return periods for extreme precipitation events might be affected by climate change. Both because of the long life of many engineered systems and because of the uncertainty about changes in precipitation, it is unrealistic to necessarily expect a design value to be revised in anticipation of potential climate change. Rather, the impact of climate change is examined in terms of how the “effective” return period associated with a design value for the present climate would deviate from its desired length (or, equivalently, how the probability of an event would deviate from its desired likelihood).

Specifically, the values of the parameters of the chain-dependent process, $\pi$, $\alpha$, and $\beta$, on which a design value depends are varied. As opposed to simply blindly perturbing these parameters, some results from the statistical downscaling study of Katz and Parlange [14] are utilized. They fit two conditional chain-dependent processes to the time series of January daily precipitation amount at Chico, depending on whether January mean sea level pressure at a grid point off the coast of California is below or above normal. The range of these parameters between the two conditional models is used to guide the sensitivity analysis. Such a range is physically plausible, because it corresponds to a change in the frequency of occurrence of above (or below) normal
pressure. Notwithstanding the model parameters having different units, this approach provides justification for directly comparing the individual sensitivity curves.

4.2. Results

The example is treated in which design values are adopted whose return periods under the present climate are either 10 or 20 yrs (i.e., design values of 82.5 or 93.4 mm from Table 1) for a winter season (i.e., $T = 90$ days) at Chico. The parameters and corresponding design values of the chain-dependent process are assumed known (i.e., any sampling error in the estimates [Eq. (15)] is ignored). Figs. 1–3 show the sensitivity of the effective return period to the three individual parameters of the chain-dependent process, $\pi$, $\alpha$ and $\beta$. In these figures, the range of the parameters is selected to coincide with the range of values for the two conditional models [14].

These plots indicate that the effective return period is much more sensitive to $\beta$ see (Fig. 3), the scale parameter of the gamma distribution, than to the other two parameters, $\pi$ see (Fig. 1) and $\alpha$ (Fig. 2). It is less obvious that the effective return period is slightly more sensitive to $\pi$, the probability of precipitation, than to $\alpha$, the shape parameter of the gamma distribution. No matter which parameter, the sensitivity curves are necessarily steeper for the more extreme event of an effective

---

**Fig. 1.** Effective return period as function of $\pi$, probability of precipitation ($T = 90$ days and return periods of 10 and 20 yrs for present climate).

**Fig. 2.** Same as Fig. 1 except $\alpha$, shape parameter of gamma distribution, is varied.

**Fig. 3.** Same as Fig. 1 except $\beta$, scale parameter of gamma distribution, is varied.
None 0.329 0.875 15.26 10.0 20.0
π decreased 0.254 0.875 15.26 12.7 25.5
π increased 0.413 0.875 15.26 8.1 16.2
z decreased 0.329 0.805 15.26 11.4 23.0
z increased 0.329 0.964 15.26 8.5 17.0
β decreased 0.329 0.875 13.35 21.1 47.3
β increased 0.329 0.964 15.72 8.6 16.8
High pressure 0.254 0.805 13.35 30.7 69.2
Low pressure 0.413 0.964 15.72 6.0 11.5

5. Discussion and conclusions

Extreme value theory for a chain-dependent process has been described, a theory that explicitly accounts for the fact that such extremes involve a random number of individual precipitation intensities. For a physically plausible range of changes in the parameters based on statistical downscaling, the effective return period for extreme high precipitation amounts turns out to be most sensitive to the scale parameter of the intensity distribution in the specific example treated. The theory is quite general, in that it applies to other forms of stochastic model for precipitation (e.g., second-order, instead of first-order, Markov chain model for occurrence or power transform, instead of gamma, distribution for intensity). In order to extend the approach beyond sensitivity analysis, an interesting issue concerns how to generalize the concepts of “design value” and “return period” from a stationary climate to a gradually changing one [22].

This paper has dealt with the maximum of daily precipitation amount, whereas events involving extreme hourly amounts would be more relevant for certain hydrologic applications. In this regard, Katz and Parlange [15] showed how a chain-dependent process can be generalized to incorporate important characteristics of hourly precipitation, especially diurnal cycles and autocorrelated intensities. It would be reasonably straightforward to extend extreme value theory to cover such a stochastic model.

The sensitivity analysis has been guided by statistical downscaling involving two conditional chain-dependent processes. In principle, extreme value theory could be applied directly to these two stochastic processes, representing the distribution of maximum daily precipitation amount as a mixture of two extreme value distributions. In the same vein, Katz and Parlange [16] described some statistical properties of this combination of chain-dependent processes, termed an “induced” model for time series of daily precipitation amount.

Lastly, one way to partially verify the theory presented here would be to conduct a “regional analysis” of precipitation extremes, an approach that is popular in hydrology for estimating flood probabilities [8]. In other words, it would be determined to what extent spatial patterns in the observed relative frequency of occurrence of extreme high precipitation amounts could be explained by differences in the probability of occurrence and in the parameters of the intensity distribution.

Brown and Katz [2] performed a similar regional analysis for temperature extremes, suggesting that the...
approach could serve as a spatial analogue for climate change.

Acknowledgements

The comments of two anonymous reviewers are gratefully acknowledged. Research was partially supported by NSF Grant DMS-9312686 to the NCAR Geophysical Statistics Project.

Appendix A. Derivation of normalizing constants

First approach: The normalizing constants, \( a_T \) and \( b_T \), for the maximum of \( T \) gamma variables [Eq. (6)] are approximate solutions to (e.g., Ref. [19])

\[
1 - F(b_T) = 1/T, \quad a_T = T f(b_T) \quad (A1)
\]

with \( F \) and \( f \) denoting the gamma c.d.f. and density function [Eq. (3)], respectively. In the first approach, the normalizing constants, say \( a_T^* \) and \( b_T^* \), for the maximum of a random number of random variables [Eq. (4)] are determined directly from Eq. (A1), but now using the mixed cdf [Eq. (10)]. It follows from Eq. (10) that \( 1 - F(b_T^*) = 1/T \pi \). Because the mixed density function corresponding to the c.d.f. in Eq. (10) satisfies \( f(z) = \pi f_1(z), z > 0 \), it also follows that \( a_T^* = T \pi f(b_T^*) \).

So the adjustment to the normalizing constants does correspond to that for the maximum of a fixed number of random variables \( T' = T \pi \), that is,

\[
a_T^* = a_T \pi, \quad b_T^* = b_T \pi. \quad (A2)
\]

Second approach: By Eq. (11), the normalizing constants in the second approach are given by

\[
a_T^* = a_T, \quad b_T^* = b_T + \ln \pi. \quad (A3)
\]

Because \( a_T \) in Eq. (6) for a gamma variable does not depend on \( T \), it follows that \( a_T^* = a_T \). Although \( b_T^* \neq b_T \), it is straightforward to show that

\[
a_T^* (b_T^* - b_T^*) \to 0 \quad \text{as} \quad T \to \infty. \quad (A4)
\]

Consequently, the normalizing constants produced by the two approaches are asymptotically equivalent, in the sense that they produce the identical limiting distribution for the maximum, the Type I extreme value with zero location and unit scale parameters as appears on the right-hand side of Eq. (5) (e.g., Ref. [17]).

References