A Fully Probabilistic Approach to Extreme Rainfall Modeling

Stuart Coles
University of Bristol, U.K.

Luis Raúl Pericchi* and Scott Sisson
University of Puerto Rico Río Piedras, U.S.A.

Abstract

It is an embarrassingly frequent experience that statistical practice fails to foresee historical disasters. It is all too easy to blame global trends or some sort of external intervention, but in this article we argue that statistical methods that do not take comprehensive account of the uncertainties involved in both model and predictions, are bound to produce an over-optimistic appraisal of future extremes that is often contradicted by observed hydrological events. Based on the annual and daily rainfall data on the central coast of Venezuela, different modeling strategies and inference approaches show that the 1999 rainfall which caused the worst environmentally related tragedy in Venezuelan history was extreme, but not implausible given the historical evidence. We follow in turn a classical likelihood and Bayesian approach, arguing that the latter is the most natural approach for taking into account all uncertainties. In each case we emphasize the importance of making inference on predicted levels of the process rather than model parameters. Our most detailed model comprises of seasons with unknown starting points and durations for the extremes of daily rainfall whose behavior is described using a standard threshold model. Based on a Bayesian analysis of this model, so that both prediction uncertainty and process heterogeneity are properly modeled, we find that the 1999 event has a sizeable probability which implies that such an occurrence within a reasonably short time horizon could have been anticipated. Finally, since accumulation of extreme rainfall over several days is an additional difficulty – and indeed, the catastrophe of 1999 was exaggerated by heavy rainfall on successive days – we examine the effect of timescale on our broad conclusions, finding results to be broadly similar across different choices.
Keywords: Annual Maximum; Bayes; Declustering;Generalized Extreme Value Distribution; Generalized Pareto; Distribution; Seasonal Series.

1 Introduction

There is a growing dissatisfaction with the use of standard statistical tools for the prediction of extremes and rare events. Examples abound in several scientific disciplines of gross underestimation, based on historical data, of the probabilities of extreme events that subsequently

*Corresponding Author. Email: pericchi@goliath.cnnet.clu.edu

occur and cause catastrophic damage. This is not restricted to hydrological events: examples appear in ecology (estimating the probability of species extinction, Ludwig, 1996); fish management (estimating the exhaustion of fisheries, Hilborn and Mangel, 1997, Malakov, 1999); insurance (calculating the probabilities of enormous claims, Smith and Goodman, 2000); and geophysical sciences (Smith, 1989). Recently Pinter *et al.* (2001) have also emphasized the need for a reassessment of flood hazards, highlighting the potential impact of the dynamical structure of rivers which have been modified due to different land usage patterns.

Standard methodology for modeling extremes consists of adopting an asymptotic model to describe stochastic variation at extreme levels of a process, inference, and forecasting on the basis of the inferred model. Asymptotically motivated models remain the centerpiece of our modeling strategy, since without such an asymptotic basis, models have no rationale for extrapolation beyond the level of observed data. However, in this article we argue that there are two principal reasons why naive adoption of this paradigm leads to systematic underestimation of the probability of disastrous events. First, model and prediction uncertainty are often overlooked or ignored. Since such uncertainties can be substantial, designs made without allowance for these effects can be disastrously anti-conservative. Moreover, it is commonplace to reduce the dimensionality of extreme value models and to proceed on the basis of a simplified model. This procedure may lead to similar estimated models, but is likely to lead to overly confident measures of precision as a principal source of model uncertainty is artificially discarded.

The second aspect is that of model homogeneity. We argue that a false assumption of a stationary model may also lead to considerable underestimation of the probability of a disastrously extreme event. While recent techniques, such as the local likelihood method of Ramesh and Davison (2002), identify local temporal variation by parameter estimation at each individual time point via appropriate data weightings, we take a different approach. With no empirical evidence of temporal trends, we model non-stationary effects through the assumption of within-season homogeneity for model parameters within a number of seasons, whilst allowing for variations across seasons whose starting point and duration are treated as unknown.

Our main departure from standard methodology is a preference for Bayesian inference. Though we also include classical likelihood analyses for comparison, we argue that the Bayesian approach provides a more coherent framework for keeping track of, and incorporating, all of the uncertainties involved in the prediction process. Computations for such models are intractable using conventional techniques, but are now almost routine using stochastic algorithms such as Markov chain Monte Carlo (MCMC).

We take as a case study daily rainfall measurements recorded on the Venezuelan central coast at Maiquetia station (International Airport). This set of data is especially pertinent, since an extreme daily rainfall in December 1999 caused what has been termed the worst environmentally related tragedy of Venezuelan history and one of the largest historical rainfal-induced debris flows documented in the world. The event was about three times as large as the previous maximum daily rainfall event in a 50-year period. Massive flooding and landslides washed away an entire state, producing a number of deaths that has been estimated at between 15,000 and 50,000. Around 8000 residences and 700 apartment buildings were destroyed or badly affected with damages being estimated at around $2 Billion. (Corporación Andina de Fomento, 2000; CIA, 2000; Wieczorek et al (2001)).

Standard interpretations of models fitted to the pre-1999 data attach a probability of virtually zero to the actual 1999 event. It was therefore argued that the event itself was impossible to foresee. Implicitly this suggests a model failure, either because of a violation of the asymptotic assumptions on which the model was built, or because of a sudden change in the meteorological climate (there is no evidence for any gradual change). Whilst both these interpretations are possible, we argue that a more critical implementation of standard extreme value models can also lead to an analysis that attaches significant probability to an event of the order of the actual 1999 event, implying that an event of this magnitude should have been foreseeable. Furthermore, we show that, in contrast to other models, our model adapts well to the inclusion of the 1999 information.

Our article investigates the role of different components in the analysis of the Venezuelan data: data selection, model choice, mode of inference and method of interpretation. We begin with an exploratory data analysis in Section 2. In Section 3 we analyze the annual data from 1951 until 1999, using both the Gumbel and the generalized extreme value distributions. Each of these models has an asymptotic justification, and it is common to use the data to determine which of the two models is most appropriate. We argue, however, that even if data support a reduction to the Gumbel family, allowance for the uncertainty of this decision must be made. Furthermore, inferences based on annual maxima are generally imprecise as a consequence of the limited amount of data available. This is resolved in Section 4 by applying a threshold model to the daily rainfall observations which are available for the period 1961 to 1999. This comprises a Poisson rate of threshold exceedances whose magnitudes follow a generalized Pareto distribution. Again, this model is derived from asymptotic considerations. However, there is external and empirical evidence for a nonstationarity in the model due to seasonal effects. This is found to be most easily identified through a Bayesian analysis which, if required, can also be used to set the seasonal structure in a classical likelihood analysis. An additional difficulty with the modelling of extreme rainfall, compared with other environmental variables, is the cumulative effect of its impact. Therefore, in Section 5 we establish the robustness of our general conclusion to the choice of timescale. In Section 6 we suggest some further refinements of the model and outline proposals for future work.

2 Exploring the Venezuelan Coastal Rain Data

Our data comprise records of daily rainfall from Maiquetia international airport, located on the North Central Coast of Venezuela, which is a tropical zone. The data were recorded by the Venezuelan Air Force *'FAV'* (González and Córdova, 2000a,b). Daily records are available for the period January 1961 to December 1999 (excluding 2 months of missing data), while a longer record of annual maxima dates back to 1951. The hydrometeorology of the exceptional December 1999 storm is that of the interaction of a cold front with moist southwesterley flow from the Pacific Ocean towards the Caribbean Sea produced an extremely long and wet period ("vaguada" in Spanish) – much more persistent than normal – over coastal northern Venezuela, a highly populated zone where there is a high mountain chain very close to the sea. For further details see MARN (2000) and Wieczorek et al (2001).

The annual maximum data are shown in Figure 1. No obvious trend is present, and this feature can be confirmed by regression-type modeling. The outstanding feature is that the 1999 maximum is almost three times the second largest maximum, suggesting that the data may be heavy-tailed and also that anticipating the event from the historical records is likely to

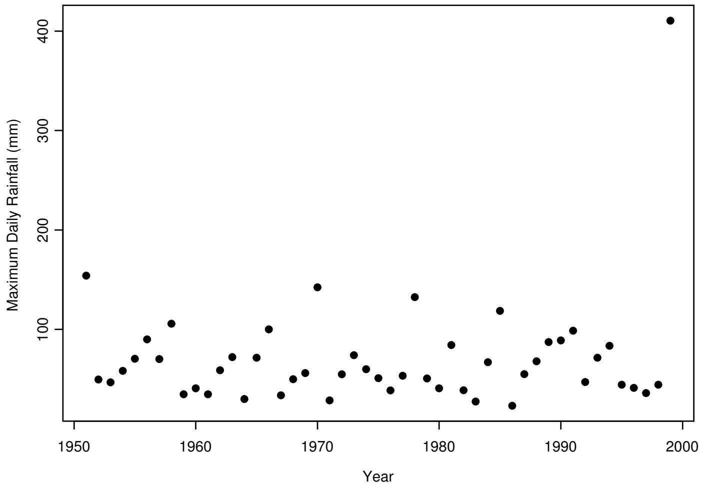

Figure 1: Annual maximum daily rainfall values recorded in Maiquetia, Venezuela

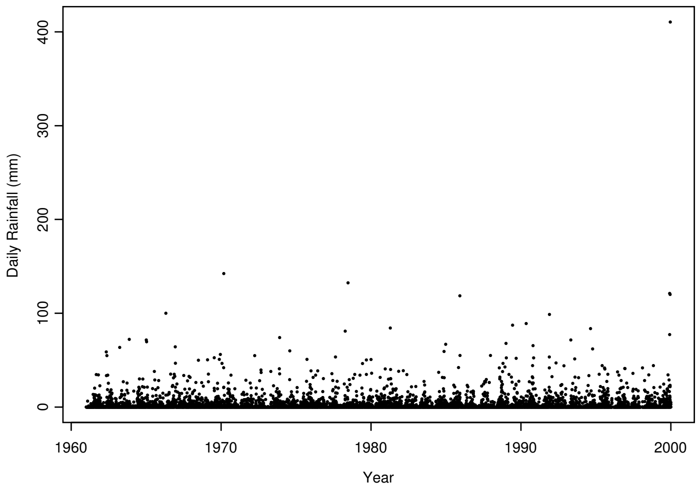

Figure 2: Daily rainfall values recorded in Venezuela.

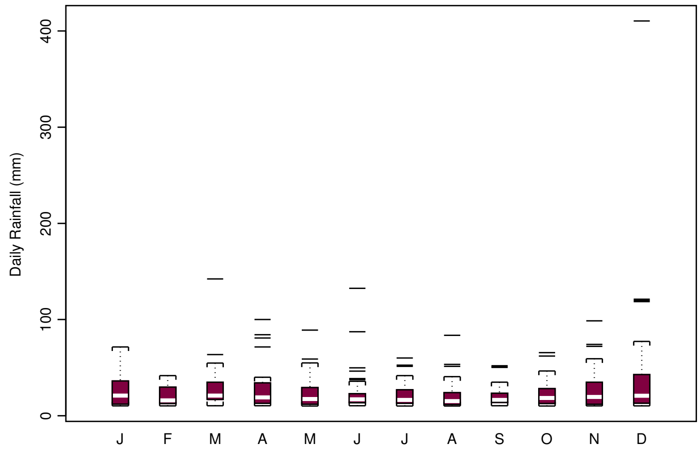

Figure 3: Monthly boxplots of exceedances of 10mm threshold.

be a severe statistical challenge. A time series of the daily data is shown in Figure 2. Again, there are no obvious trends and no visible signs of large-scale clustering, features which can again be verified through empirical checks or detailed modeling. Further exploratory analysis also reveals only a very slight correlation with various indices that serve as proxies for the el Niño phenomena, an observation that is consistent with several previous studies (M. C. Larsen of the USGS, Puerto Rico; personal communication). Therefore, because of the limited information gain provided by such indexes, relative to the other factors we consider here, we do not pursue further the effect on rainfall of the el Niño phenomena.

Anticipating the possibility of seasonal variation, Figure 3 shows the daily data in the form of box-plots stratified by month. To avoid the compression of the figures due to the large number of near-zero observations, we have produced plots only for excesses over a threshold of 10 mm, which we adopt as a modeling threshold in subsequent sections. No very obvious patterns emerge, although established meteorological wisdom suggests that there should be two seasons displaying different kinds of meteorological phenomena: a 'cold fronts' season, roughly from November until at least February, and an 'inter-tropical convergence zone' for the rest of the year (González and Córdoba, 2000a). That there is little graphical support for this interpretation adds weight to the argument for using carefully chosen statistical models that have the capacity to detect seasonal structure.

3 Fitting the Annual Data

Denoting daily observations by $X_1, X_2, \ldots$, the classical model for extremes is obtained by studying the behavior of $M_n = \max\{X_1, \ldots, X_n\}$ for large values of n. With $n = 365$, M_n corresponds naturally to the annual maximum. Asymptotic considerations suggest that the distribution of M_n should be approximately that of a member of the generalized extreme value (GEV) family (NERC, 1975; Leadbetter et al., 1983, for example), having distribution function

$$F(z|\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

with parameter space $\{(\mu,\sigma,\xi) : \mu \in \mathbb{R}, \sigma > 0, \xi \in \mathbb{R}\}$. The special case of the GEV distribution obtained by letting $\xi \to 0$ is the Gumbel distribution, with distribution function

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

The Gumbel distribution is often itself used – perhaps even more often than the full GEV family – as a model for annual maxima. In some cases there is empirical evidence to support this; in others it is sometimes argued that, since there are many distributions for the X_i which would lead to a limiting Gumbel distribution for M_n – the normal, lognormal and gamma distributions, for example – this is a more appropriate family than the GEV. We will argue below that this is a very risky strategy, even when formal hypothesis tests and model diagnostics support the model reduction.

Various techniques are available for extreme value model parameter estimation. In our view, not least for the capacity of model-building and extension, inference methods based on the likelihood function are preferable. Assuming independent annual maxima $Z_1, \ldots, Z_m$, this takes the form

$$L(\mu,\sigma,\xi; Z_1,\ldots,Z_m) = \prod_{i=1}^{m} f(z_i \mid \mu,\sigma,\xi),$$

where f is the Gumbel or GEV density function as required. There are two contrasting ways to utilize the likelihood function to obtain parameter estimates. The first is maximum likelihood in which L (or, for numerical convenience, $\log L$) is maximized with respect to the parameters μ, σ and ξ. This is numerically straightforward and also has the convenience that various standard large sample theory results are available to enable the numerical calculation of standard errors and confidence intervals. An alternative strategy is adoption of the Bayesian paradigm. This requires a prior distribution on the parameters (μ,σ,ξ) that is intended to represent beliefs about parameter values, prior to the availability of data. In some contexts it may be appropriate to use this device to apply genuine knowledge about the process under study. More commonly, there is no such knowledge, so essentially arbitrary distributions with a large variance are adopted to reflect this prior ignorance. Denoting the prior distribution by $\pi(\cdot)$, Bayes' theorem states that

$$f(\mu,\sigma,\xi \mid Z_1,\ldots,Z_m) \propto L(\mu,\sigma,\xi; Z_1,\ldots,Z_m) \times \pi(\mu,\sigma,\xi),$$

leading to a so-called posterior distribution of the parameters that is a modification of the prior distribution, due to the information contained in the data expressed through the likelihood function. The outcome of a Bayesian analysis is an entire distribution on the parameter set, which represents a considerable advantage over classical methods: rather than just a point estimate, we obtain a complete probabilistic distribution of the parameter values. If a point estimate is required it is usual to give a simple summary statistic of the posterior – the posterior mean or mode, say.

Direct implementation of Bayes' theorem is complicated in general, and for our model in particular, by the fact that the proportionality factor implies the necessity of an integration over the parameter space. This is not possible analytically, and may be difficult using standard numerical methods. Fortunately, standard Markov chain Monte Carlo methods routinely allow approximation of such integrals (see Hastings, 1970, for an early reference, Gamerman, 1997, for a more complete, though introductory, overview and also Coles (2001) for a specific explanation in the context of extreme value modeling.).

Returning to the specifics of an extreme value analysis, it is natural to consider models in terms of their implications for future extreme values of the process under study, rather than in terms of their parameter values. In the simplest case, denoting a future annual maximum variable by Y with distribution function F_Y, and setting

$$H_Y(y|\mu, \sigma, \xi) = 1 - F_Y(y|\mu, \sigma, \xi),$$

the value

$$N(y) = \frac{1}{H_Y(y|\mu, \sigma, \xi)}$$

is the 'return period' of y, since y approximates to the level that will be exceeded on average once every $N(y)$ years. In practice, it is usual to invert this expression, and determine the 'return level' y corresponding to a fixed return period of, say, 100 or 500 years. Return levels or periods can be estimated easily by 'plugging-in' maximum likelihood estimates. Standard likelihood theory can also be invoked to obtain standard errors or confidence intervals for these quantities. Again however, there are advantages to adopting a Bayesian alternative. In this case

$$F(y|\boldsymbol{z}) = \int F(y|\mu, \sigma, \xi)\pi(\mu, \sigma, \xi|Z_1, \ldots, Z_m)d\mu d\sigma d\xi,$$

is the 'predictive distribution' of Y, incorporating uncertainty both in the future value of Y and in the parameter values themselves.

Use of the Bayesian predictive distribution is not new in hydrological applications, and in fact has a long history in flood frequency analysis. Beard (1960) formulated an expected probability method, predating a Bayesian approach formalized in the U.S. Manual Bulletin 17B (1982). In greater formality, use of the full Bayesian predictive distribution, termed the design flood distribution, was suggested by Stedinger (1983), with further modifications proposed by Kuczera (1999).

	MLE		Bayes	
	Gumbel	GEV	Gumbel	GEV
μ	50.9 (3.3)	49.0 (3.4)	50.8 (3.4)	49.0 (3.6)
σ	21.5 (2.5)	19.9 (2.7)	22.5 (2.8)	21.1 (3.0)
ξ	—	0.17 (0.14)	—	0.182 (0.15)
nllh	224.9	224.1	—	—

Table 1: Maximum likelihood estimates and posterior means in Gumbel and GEV analyses of annual maximum rainfall data. Figures in parentheses are, in the case of the MLEs, standard errors, and in the case of the Bayesian analysis, posterior standard deviations. For the MLE analyses the negative log-likelihoods (nllh) are also reported.

For each of the Gumbel and GEV models, parameter estimates obtained by both a maximum likelihood and a Bayesian analysis are shown in Table 1. The differences between

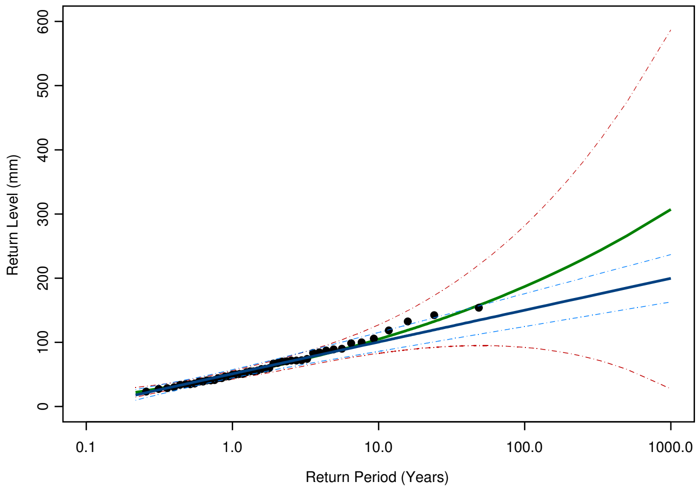

Figure 4: Return level plots for Gumbel and GEV models of Maiquetia annual maximum daily rainfall. Dashed curves correspond to 95% confidence intervals; points correspond to empirical estimates excluding the 1999 datum.

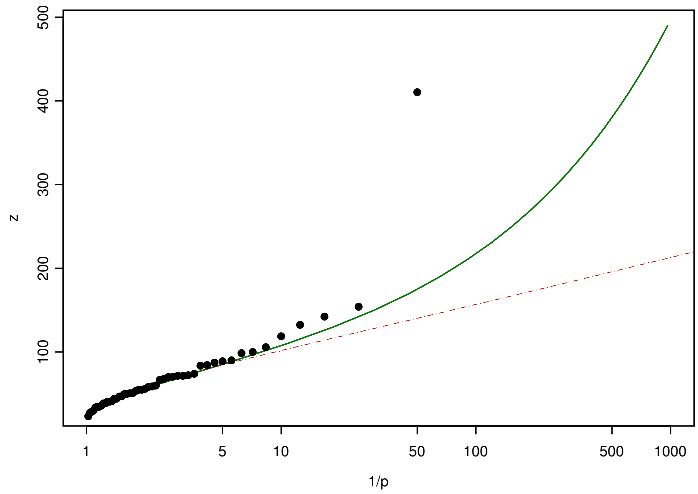

Figure 5: Predictive return level plots for Gumbel and GEV models of Venezuelan annual maximum daily rainfall. Solid curve is GEV model; dashed curve is Gumbel model. Points correspond to empirical estimates including the 1999 datum.

Mode of Inference	Model	Return period of 410.4 mm	
		1999 datum excluded	1999 datum included
MLE	Gumbel	17,600,000	737,000
	GEV	4280	302
Bayes	Gumbel	2,170,000	233,000
	GEV	660	177

Table 2: Return level estimates of 410.4 mm, the 1999 annual maximum, using different models and modes of inference. For the MLE analysis the values correspond to the maximum likelihood estimates of the return period. For the Bayesian analysis the values are the predictive return periods.

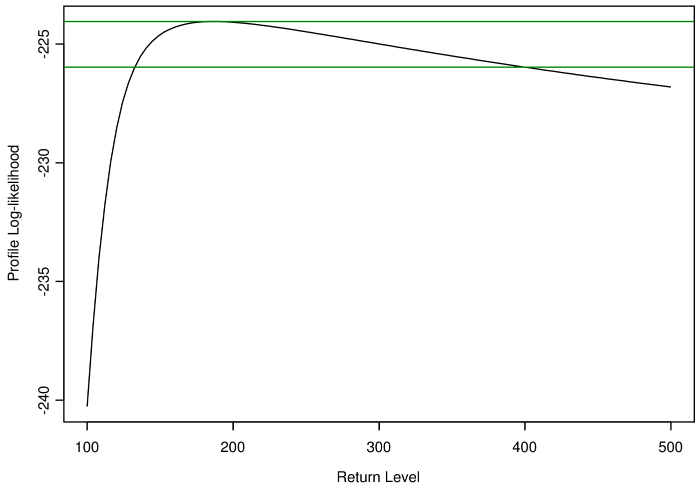

Figure 6: Profile likelihood of 100-year return level in GEV analysis of annual maximum rainfall data excluding 1999 datum. Horizontal line corresponds to $\ell_{\max} - q_{.95}$, where $\ell_{\max}$ is the maximum value of the log-likelihood and $q_{.95}$ is the 95% quantile of the χ^2_1 distribution.

the Bayesian and classical likelihood inferences are slight for each model. This is reassuring: we used diffuse priors and the likelihood surface is reasonably symmetric, so posterior means should lie close to the mode of the likelihood. As a consequence, there can be little objection to the adoption of a Bayesian mode of inference, although the advantages in doing so are yet to be determined.

Despite the similarity between the Gumbel and GEV estimated models there is a substantial difference between the models once measures of uncertainty are accounted for. This is most clearly seen in the return level plot of Figure 4. The maximum likelihood return level curves are almost identical for the two models, at least within the range of the observed data. However, 95% confidence intervals based on the GEV model are considerably wider than the corresponding intervals for the Gumbel model. Consequently, adoption of the Gumbel model, while not affecting point estimators much, leads to a dramatic increase in supposed precision, as the uncertainty in shape parameter estimation is eliminated. Obviously the reduction in variation achieved with the Gumbel model is useful in planning if the Gumbel model is correct. If it is not correct however, the mis-modeling may have catastrophic consequences. This was unfortunately the case with the Venezuelan data since, as discussed above, the 1999 event had a magnitude of 410.4 mm, a value that would have been regarded as virtually impossible under the Gumbel model. The issue is best viewed through a plot of the predictive return levels obtained from the Bayesian analysis. This is shown in Figure 5, where we have also plotted the empirical estimates including the 1999 data. For both models the predictive return levels are greater than the corresponding maximum likelihood estimates in Figure 4 due to the allowance for estimation uncertainty. Although the GEV model attaches low predictive probability to the actual 1999 event, it is orders of magnitude greater than the value obtained under the Gumbel model. This is confirmed in Table 2, where we have calculated maximum likelihood estimates and predictive values of the return period associated with the 1999 annual maximum of 410.4 mm. For future reference we also include the corresponding value obtained after inclusion of the 1999 data .

The contrast between the MLE and predictive Bayesian values is due principally to the implicit allowance for parameter uncertainty in the Bayesian estimates. A more careful likelihood analysis, based on refined asymptotic results, would lead to similar conclusions to the Bayesian analysis. For example, Figure 6 shows the profile log-likelihood of the 100-year return level for the annual maximum excluding the 1999 datum. Based on asymptotic theory, an approximate 95% confidence interval for the true value of the 100-year return level is given by the set of values whose profile log-likelihood is within $q_{0.95}$ of the maximized log-likelihood, where $q_{0.95}$ is the 95% quantile of the χ_1^2 distribution. From Figure 6 this interval is [134,401]. By this criterion the observed 1999 value of 410 mm is again not so very surprising.

In summary, the Bayesian analysis does not give a radically different interpretation of the data, but does provide a more convenient and direct way of managing and expressing uncertainties. There are, however, fundamental differences between the Gumbel and GEV estimates: the ratio of estimates under the two models is greater than 3000, using either mode of inference. In particular, the Gumbel model attaches a probability of only 4.6×10^{-6} of an annual event as great as the 1999 observed value, even after allowance for parameter uncertainty. The corresponding GEV value of 1.5×10^{-3} seems much more realistic. So, despite the apparent support within the data for a reduction to a Gumbel model, adoption of this strategy would have led to a model giving virtually zero probability to a subsequently occurring event. Retainment of the GEV model would, in contrast, have attached a sizeable probability to the same event, so that the occurrence of such an event within a reasonable time-span could have been anticipated. In subsequent sections we seek to exploit additional information in order to improve further the accuracy estimated probabilities of extreme events.

4 Modeling Daily Data

The classical GEV or Gumbel models for extremes are based on asymptotic approximations to the sampling behaviour of block maxima. With block sizes of one year, the theory furnishes approximations for the annual maximum distribution. In most cases however, complete daily records of rainfall may be available, and there is a potential wastage in not using more of

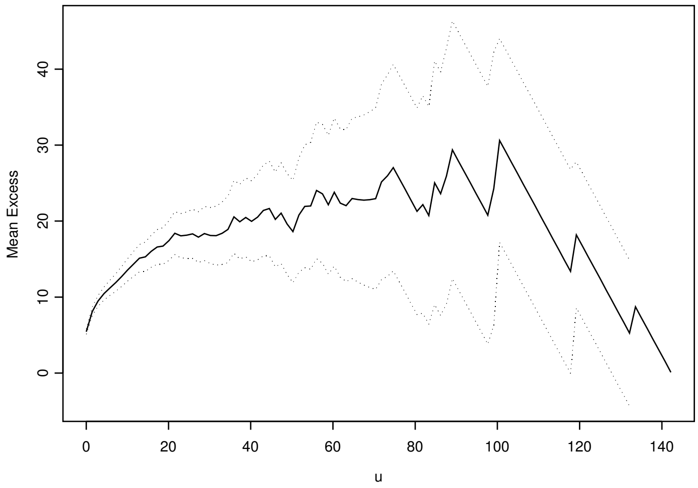

Figure 7: Mean residual life plot (with 95% confidence intervals) of daily rainfall data.

the available information on extremes that these provide. See Coles (2001) for a general discussion of these issues and an overview of the general modeling approach.

For the Venezuelan rainfall series, daily values are available in addition to the annual values for the period 1961–1999. Again based on asymptotic theory, taking the daily observations to be independent and identically distributed, it is common to model the extremes by means of a threshold model. This consists of two parts: a Poisson process of high threshold exceedances and a generalized Pareto distribution for the excess-over-threshold magnitudes. The form of the generalized Pareto distribution function is

$$F(v|\sigma,\xi) = 1 - \left(1 + \xi\frac{v}{\sigma}\right)^{-1/\xi}, \quad v > 0.$$

In the context of threshold exceedance modelling, the asymptotics suggest this model as an approximation to the distribution of the excesses, $V_i = X_i - u$, for a large enough threshold, u. The parameter ξ coincides with the shape parameter in the corresponding GEV distribution of the annual maximum, but the scale parameters of the two models are different. Furthermore, the value of σ in the generalized Pareto model is dependent on the chosen threshold.

Threshold choice requires some care, comprising a balance between bias and variance. Too low a value is likely to compromise the asymptotic justification of the model, leading to bias; too large a value will lead to few exceedances, and therefore a large variance of estimators. One commonly-used tool for threshold selection is a mean residual life plot (Davison and Smith, 1990), comprising a plot of the points

$$(u_i, e(u_i)) \tag{1}$$

for a range of possible thresholds u_i, where $e(u_i)$ is the empirical mean of the set $\{x_i - u : x_i > u\}$. Because of the identity

$$E(X - u \mid X > u) = \frac{\sigma - \xi u}{1 - \xi},$$

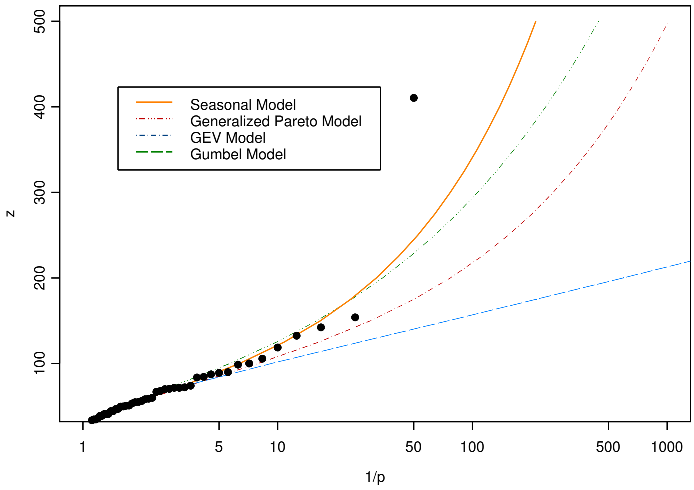

Figure 8: Predictive distributions for various models fitted to the Venezuelan rainfall data. Points correspond to empirical estimates.

valid for a generalized Pareto model of threshold excesses in the usual case of $\xi < 1$, it follows that the points in (1) should be approximately linear above a level u for which the model is valid. A mean residual life plot for the Venezuelan data, together with approximate 95% confidence intervals, is shown in Figure 7. Allowing for the confidence intervals, the plot displays curvature until around $u = 10$ mm, after which there is reasonable linearity. Subsequent model checks also find return level estimates to be robust to this choice.

	MLE	Bayes
σ	10.2 (0.8)	10.4 (0.8)
ξ	0.26 (0.08)	0.30 (0.06)

Table 3: Maximum likelihood estimates and posterior means in generalized Pareto analysis of daily rainfall data. Figures in parentheses are, in the case of the MLEs, standard errors, and in the case of the Bayesian analysis, posterior standard deviations.

Based on the generalized Pareto model for excesses of a 10 mm threshold, parameter estimates and measures of precision are included in Table 3. As before, the differences between the two modes of inference are slight. Clearer assessment of the models is made by looking at return levels, and again we favor the predictive return level plot because of its accommodation of estimation uncertainty. The resulting curve is compared in Figure 8 against the corresponding curves for the GEV and Gumbel models. Clearly, the extra information provided by the daily data leads to an inference that attaches yet higher probability to the subsequent 1999 event.

We also carried out one further analysis which we do not discuss in great detail here. As explained in the introduction, knowledge of local meteorology suggests a two-seasonal structure to the Venezuelan rainfall process. We therefore include in Figure 8 the predictive return level curve for a model in which a different generalized Pareto distribution is permitted

Mode of Inference	Model	Return period of 410.4 mm	
		1999 data excluded	1999 data included
MLE	Homogeneous GPD	752	251
	Seasonal GPD	282	84.4
Bayes	Homogeneous GPD	260	116
	Seasonal GPD	131	57.0

Table 4: Return level estimates of 410.4 mm, the 1999 annual maximum, using different one- and two-season threshold GPD models and two modes of inference. For the MLE analysis the values correspond to the maximum likelihood estimates of the return period. For the Bayesian analysis the values are the predictive return periods.

in each of two seasons (Coles and Pericchi, 2001). One novel aspect of this model is that the timing of the seasons are treated as unknown model parameters and are therefore included as part of the inference, a feature which is only feasible within the Bayesian framework. What is seen from Figure 8 is that this two-seasonal model improves again on the previous models in terms of assigning non-negligible probability to the subsequent 1999 event. This is largely a consequence of substantially different shape parameter estimates in each of the two seasons, $\hat{\xi} = 0.42$ and 0.22 respectively.

We summarize these final results in Table 4, which is an analogue of Table 2. We have also included a maximum likelihood analysis of the two-seasonal model, though we stress that such an analysis was only feasible after the Bayesian analysis had determined an appropriate timing for the seasonal changes. The results emphasize again the importance of taking estimation uncertainty into account through the Bayesian predictive analysis. Finally, although the threshold model is seen to improve considerably on the annual maximum models in terms of assigning probability to the 1999 event, we see that the two-seasonal model does better still. Indeed, the probability of having an annual maximum value greater than 410.4 mm is 1/130.6 according to this model. Consequently, the probability of having an event of this magnitude or greater at some point in a 49-year period is

$$1 - (1 - 1/130.6)^{49} = 0.31.$$

So, even by an analysis that excludes the 1999 event, there is a probability of around 30% that an event as large as 410 mm will occur in a 49-year period. Compare this interpretation with the naive Gumbel analysis which reported the 1999 event as the 'one-in-17 million year event', and we begin to understand the importance of taking seriously the process of statistical modeling of extreme value data.

5 Modeling Aggregate Data

The asymptotic arguments which lead to the generalized Pareto model as a distribution of threshold excesses by an independent sequence can be generalized to consider sequences of dependent, though stationary, series. Under a weak condition that limits the effect of any long-range dependence, it can be shown that the generalized Pareto family has the same status as a limiting family for threshold excesses in this more general setting (see Leadbetter *et al.*, 1987, for example). This provides some justification for use of the generalized Pareto distribution as a model for extreme daily rainfalls, even if there is some temporal dependence in the series. Strictly, some account should still be taken of the dependence in the series, since the inferential content of the data will be less than it would have been had the data been independent. But, as we have argued above, dependence in the series is not strong, and so our results so far are little affected by this misspecification.

It is a different story if the effects of aggregation are to be taken into account. The extreme rainfall event of 1999, measuring 410 mm, occurred on December 15. This followed heavy rainfall (120 mm) on the previous day. There is some uncertainty, however, concerning the rainfall that occurred on December 16. Our analyses so far were based on the original *FAV* source data, which report this rainfall as zero. However, there have been subsequent revisions of the data and it has been suggested that the event of December 16 was also extreme, the recording device having collapsed after the heavy rain of December 15. Because of uncertainty in the value of the December 16 rainfall, we have not modified the original value in our analyses so far. However, González and Córdova (2000b) estimate the December 16 rainfall as 290 mm. Including this additional datum in the previous models produces little substantive change, but the value, if accurate, increases substantially the three-day aggregate over the relavant period, and this is likely to have a considerable bearing on any analysis of time-aggregated rainfall. We therefore include the estimated value in an analysis of 3-day aggregate events.

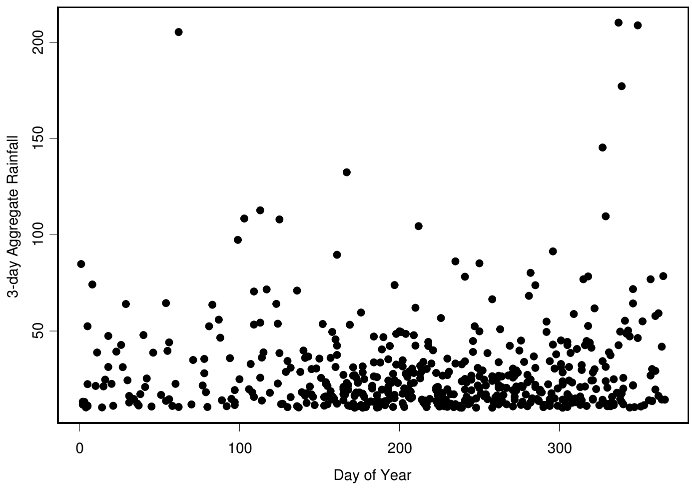

Figure 9: Declustered 3-day aggregate series of rainfall recorded in Venezuela excluding the December 1999 cluster.

We first define the series

$$X_i^{(3)} = \sum_{j=i}^{i+2} X_j,$$

corresponding to successive 3-day aggregates. These are then filtered to eliminate the obvious time-dependence induced by the aggregation procedure, leaving just the largest of the successive aggregates. Diagnostic checks, as described in Section 4, support the maintenance of

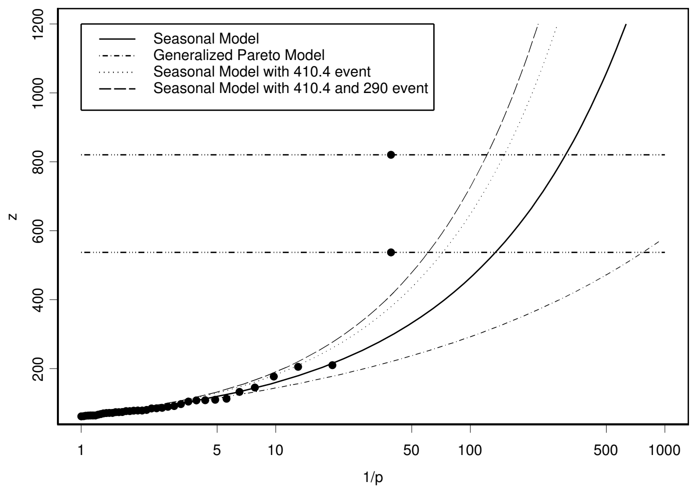

Figure 10: Bayesian predictive distributions for homogeneous and seasonal models fitted to the declustered 3-day aggregate Venezuelan data. Points correspond to empirical estimates.

$u = 10$ mm as a threshold for the analysis of extremes of this series, even though such a level is obviously stochastically lower for 3-day aggregates than it is for daily events. With this threshold the 39 years of available daily rainfall data generate 520 3-day aggregate threshold exceedances. These values are illustrated in Figure 9, indexed by the day of occurrence of the maximum daily event contributing to the aggregate. Evidence of a two-season structure is now more apparent than we found in Figure 3.

For ease of comparison with the analysis of the daily data, similar analyses are performed using both maximum likelihood and Bayesian inference. Moreover, we consider both the inclusion and exclusion of the 1999 extreme aggregate event, with an additional sensitivity analysis obtained by exclusion of the reconstructed data point for the 16th December event. Maximum likelihood estimates and predictive return periods are summarized in Table 5, where the seasonal changepoints for the two-seasonal MLE analyses were fixed at the distributional modes of the relevant Bayesian analyses. As with the daily analysis, accounting for estimation uncertainty via the Bayesian predictive analysis improves considerably on the maximum likelihood approach, with the two-season model seeming to have the best performance. The resulting predictive curves for the Bayesian analyses are compared in Figure 10.

Based on the Bayesian seasonal analysis the probability of realizing a 3-day rainfall aggregate in excess of 537.5 mm or 820.4 mm at some point in a 49-year period is either

$$1 - (1 - 1/134)^{49} = 0.31 \qquad \text{or} \qquad 1 - (1 - 1/308)^{49} = 0.15$$

respectively, depending on whether the 1999 data are included or excluded. It seems, therefore, that similar benefits in terms of obtaining plausible probabilities of subsequently occurring events, as were obtained in the analysis of daily data, are also obtained in the analysis of 3-day aggregate data.

Was the Venezuelan Central Coast December 1999 an unique event? Wieczorek et al (2001) revealed that the 1999 event exposed an even greater event, that occurred within

Mode of Inference	Model	Return period	
		537.5 mm	820.4 mm
MLE	Homogeneous GPD	1482	8057
	Seasonal GPD	308	1034
	Seasonal GPD with 410 mm	142	386
	Seasonal GPD with 410/290 mm	112	289
Bayes	Homogeneous GPD	769	3112
	Seasonal GPD	134	308
	Seasonal GPD with 410 mm	72	149
	Seasonal GPD with 410/290 mm	60	122

Table 5: Return level estimates of 537.5 mm and 820.4 mm, the 1999 3-day aggregate cluster maximum excluding and including the reconstructed event of 290 mm respectively. Different one- and two-season threshold GPD models and two modes of inference are used, with increasing portions of the 1999 cluster included in the analysis. For the MLE analysis the values correspond to the maximum likelihood estimates of the return period. For the Bayesian analysis the values are the predictive return periods.

the span of about 500 years. In view of this, the return period of 308, given in Table 5 by the Bayesian analysis of the seasonal GPD for the 3-day cluster 1999 event, seems quite realistic and confirmed by an independent source of evidence.

6 Discussion

We hope in this article to have exposed a number of myths about extreme value analysis. A commonly-held perception of extreme value modeling is that, with such scarcity of data, there is little that can be achieved through serious statistical modeling. Our view, on the contrary, is that the lack of data makes statistical modeling, and especially the accounting of uncertainty, that much more important. Perhaps the biggest myth we hope to have exposed is the formulaic adoption of the Gumbel model when data seem to support such a reduction from the full GEV model. Our analysis here demonstrates the risks involved in adopting the Gumbel model without continuing to take account of the uncertainties such a choice involved. Our best advice, based on this and other analyses, would be always to work with the GEV model in place of the Gumbel, unless there is additional information supporting the Gumbel choice.

The preference for a Bayesian analysis over a classical likelihood analysis is also something we feel to be strongly supported by the Venezuelan data analysis. By working with diffuse priors the Bayesian framework is essentially formal, but it leads to an inference in which parameter uncertainty is properly formalized and for which inferences on predictions are naturally handled. If genuine prior information were also available, this could also be incorporated into the analysis (Coles and Tawn, 1996). The Bayesian framework also enabled us to fit a two-seasonal model when an appropriate timing for the seasons was unknown. Overall, it required the combination of model choice, maximal use of available data, the expression of uncertainty through the Bayesian model, and the modeling of seasonal heterogeneity to obtain an analysis that ascribed appreciable probability to the actual 1999 event.

We comment now that Tables 2 and 4 also include inferences that include the 1999 event. The Gumbel model, in particular, does not adapt well to the new information, leading to return level estimates that are totally unbelievable given the occurrence of the 1999 event. On the other hand, the seasonal model adapts especially well, leading to modified return level estimates that are entirely compatible with the occurrence of an event above 400 mm once in the 49-year observation period. We also comment that while there are more involved methods of attempting to model the clustering of extreme processes, the aggregate approach adopted here is intuitively appealing, and provides a natural physical interpretation despite the apparent simplicity involved.

There are a number of directions in which we intend to pursue this work further. The assumption of a Poisson rate of extreme events is not perfect for many environmental process because of a tendency for extreme events to cluster; for example, the Venezuelan 3-day December 1999 cluster. We have argued that such violations of the Poisson assumption are unlikely to have much impact on inference of return levels for daily values and have also presented a simple analysis of extreme 3-day aggregate rainfall. However, we have not tried explicitly to model the weak day-to-day dependence of extreme rainfall, and there may be some gain in precision in doing so. It is an open question whether such potential gain would outweigh the simplicity and robustness offered by our simple aggregation approach.

We also intend to examine the precipitation mechanism structure in greater detail, considering the possibility of a breakdown into a greater number of seasons. Our preliminary investigations suggest that the two-seasonal model captures most of the temporal structure in the process and that further modification is likely to have limited impact on the final results.

Our seasonal model would not be the only way of introducing physical structure into the statistical model. It could be argued, ignoring the seasonal timing of such events, that extreme events in tropical regions arise either as extremes of 'typical' storm events, or from infrequent cyclonic hurricane events, which have a different physical, and therefore statistical, behavior. The distribution of an extreme event might therefore be represented as a mixture distribution,

$$F(z|\mu, \sigma, \xi) = wF_1(z|\mu_1, \sigma_1, \xi_1) + (1 - w)F_2(z|\mu_2, \sigma_2, \xi_2), \tag{2}$$

where F_1 is a relatively non-extreme distribution, occurring with high probability w, whereas F_2 is a much more extreme distribution, but occurring with low probability $1-w$. This model can also be fitted by maximum likelihood or by MCMC methods (see Walshaw, 2001, for a similar model). In actual fact, this model is very similar to our 2-season model. Indeed, if all of the less extreme events occurred in one season, and the more extreme events in the other, the two models would be identical. The only substantive difference in the two models is that in the seasonal model the more extreme events are confined to a contiguous part of the annual cycle. This implies a wider flexibility in model (2) at the cost of not incorporating the metoerological beliefs about the tendency for the most extreme events to occur in a specific season. Nonetheless, it may be valuable to make a comparison between the two approaches.

A final issue is the notion of a maximum probable rainfall (MPR). On physical grounds it is often argued that an absolute upper-bound can be placed on the daily rainfall distribution. For the Venezuelan data we have been advised that 1700 mm is a reasonable value for the MPR (M. Gonzalez, personal communication). The Bayesian setting we have proposed provides an ideal framework within which to explore the impact of imposed MPR values, and this will be reported elsewhere. Nevertheless, the fact that the MPR of 1700 mm is so large relative to the observed data suggests that inferences will be little changed.

Acknowledgments

We are indebted to J.R. Córdova, M. González and M.C. Larsen, for providing the data and for several helpful discussions, and to I. Rodríguez-Iturbe, R. Smith and J. Tawn for being sources of inspiration over the years. EPSRC (United Kingdom) and UPR-RP support to one of the authors (LRP) is gratefully acknowledged.

References

Beard (1960) Probability estimates based on small normal distribution samples.*Journal of Geophysical Research,* 65(7): p. 2143-2148.

CIA (2000) *'Fact Book'*, www.cia.gov/cia/publications/factbook/geos/ve.html

Coles, S. (2001) *An introduction to statistical modeling of extreme values*, Springer: London.

Coles, S. and Pericchi, L.R. (2001) Anticipating catastrophes through extreme value modeling. Submitted.

Coles, S.G. and Tawn, J.A. (1996) A Bayesian analysis of extreme rainfall data, *Applied Statistics*, 45, 463–478.

Corporación Andina de Fomento (2000) Efectos de las lluvias caídas en Venezuela en Diciembre de 1999. CAF-PNUD.

Davison, A.C. and Smith, R.L. (1990) Model for exceedances over high thresholds. *Journal the Royal Statistical Society, Series B*, 52, No. 3, 393-442.

Gamerman, D. (1997) Markov chain Monte Carlo: Stochastic simulation for Bayesian inference (Texts in Statistical Science). Chapman and Hall, London.

González M. and Córdova J.R. (2000a) Consideraciones sobre la probabilidad de ocurrencia de lluvias máximas en la zona litoral del norte de Venezuela. Memorias del Seminario Internacional *Los Aludes Torrenciales de Diciembre 1999 en Venezuela*, Instituto de Mecánica de los Fluidos, Universidad Central de Venezuela, Diciembre 2000.

González M. and Córdova J.R. (2000b) Estudio de Crecidas de las Cuencas del Litoral Central D.F. Informe Final. Ministerio de Ciencia y Tecnología. Autoridad Unica del Estado Vargas.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.

Hilborn, R. and Mangel, M. (1997) The ecological detective: Confronting models with data. *Monographs on Population Biology. Princeton University Press.*

Kuczera (1999) Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. *Water Resources Research,* Vol. 35, No. 5, p. 1551-1557.

Leadbetter, M.R., Lindgren, G. and Rootzen, H. (1983) Extremes and related properties of random sequences and processes. Springer-Verlag, New York.

Ludwig, D. (1996) Uncertainty and the assessment of extinction probabilities. *Ecological Applications*, 6 (4), pp. 1067-1076.

Malakov, D. (1999) Bayes offers a *"New"* Way to make sense of numbers. *Science*, Vol. 286, pp. 1460-1464.

MARN (2000) Informe preliminar sobre los aspectos ambientales vinculadas al desastre natural ocurrido en Venezuela durante el mes de Diciembre de 1999. Ministerio del Ambiente y de Recursos Naturales, Venezuela. (Unpublished Report).
Pinter N., Thomas R. and Wlosinski H. (2001) Assessing flood hazards on dynamic rivers. *EOS, Transactions, American Geophysical Union*, Vol. 82, Number 31, 333 and 339-340.
Ramesh, N.I. and Davison, A.C. (2002) Local models for exploratory analysis of hydrological extremes, *Journal of Hydrology*, *256*, 106-119.
Stedinger (1983) Estimating a regional flood frequency distribution. *Water Resources Research,* Vol. 19, No. 2, p. 503-510.
Smith, R.L. (1989) Extreme value analysis of environmental times series: an example based on ozone data (with discussion), *Statistical Science*, 4, 367-393.
Smith, R.L. and D.J. Goodman (2000) Bayesian risk analysis. Technical Report, Dept. Statistics, UNC-Chapel Hill.
U.S. Manual Bulletin 17B (1982) Guidelines for determining flood flow frequency, *Bulletin 17B, of the Hydrology Subcommittee, March 1982, Interagency Advisory Committee on water data*, United States Department of the Interior Geological Survey. U.S. Water Resources Council.
Walshaw, D. (2000) Modelling extreme wind speeds in regiones prone to hurricanes, *Applied Statistics*, **49**, 51-62.
Wieczorek G.F, Larsen M.C., Eaton L.S., Morgan B.A. and Blair J.L. (2001) "Geologic Hazards Team: Debris-flow and flooding hazards associated with the December 1999 storm in coastal Venezuela and strategies for mitigation". *US Geological Survey.*(Open File Report 01-0144) *http://geology.cr.usgs.gov/pub/open-file-reports/ofr-01-0144/*