Arnold Schwarzenegger
Governor

QUANTIFYING VOLATILITY AND THE PROBABILITY OF DAILY PRECIPITATION EXTREMES

PIER PROJECT REPORT

Prepared For:
California Energy Commission
Public Interest Energy Research Program

Prepared By:
Scripps Institution of Oceanography
University of Nevada, Reno

October 2007
CEC-500-2007-034

California Climate Change Center
Report Series Number 2007-012

Prepared By:
A. Gershunov[1], A. K. Panorska[2], and T. J. Kozubowski[2]
[1]Climate Research Division, Scripps Institution of Oceanography
University of California, San Diego
9500 Gilman Drive
La Jolla, California 92093-0224
[2]Department of Mathematics and Statistics, University of Nevada, Reno

Commission Contract 500-02-004
Commission Work Authorization MR-025

Prepared For:
Public Interest Energy Research (PIER) Program
California Energy Commission

Beth Chambers
Contract Manager

Kelly Birkinshaw
Program Area Lead
Energy-Related Environmental Research

Laurie ten Hope
Office Manager
Energy Systems Research

Martha Krebs
Deputy Director
ENERGY RESEARCH & DEVELOPMENT DIVISION

B. B. Blevins
Executive Director

DISCLAIMER

This report was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees or the State of California. The Energy Commission, the State of California, its employees, contractors and subcontractors make no warrant, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the California Energy Commission nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

Acknowledgments

This work was supported through National Science Foundation (NSF) grant ATM-0236898 "Modeling, variability and predictability of North American hydrologic extremes." Additional funding was provided by the California Energy Commission via the Public Interest Energy Research (PIER) Program through the Scripps California Climate Change Center. We have enjoyed and benefited from interesting discussions with D. Cayan, M. Dettinger, K. Redmond, and T. Cavazos, and we gratefully acknowledge their input. We thank D. Rominger for programming support. Results of this study contribute to the North Atlantic Treaty Organisation (NATO) Science for Peace and Prosperity project (SFP 981044) "Extreme precipitation events: their origins, predictability and impacts."

Please cite this report as follows:

Gershunov, A., A. K. Panorska, and T. J. Kozubowski. 2007. *Quantifying Volatility and the Probability of Daily Precipitation Extremes.* California Energy Commission, PIER Energy-Related Environmental Research Program. CEC-500-2007-034.

Preface

The Public Interest Energy Research (PIER) Program supports public interest energy research and development that will help improve the quality of life in California by bringing environmentally safe, affordable, and reliable energy services and products to the marketplace.

The PIER Program, managed by the California Energy Commission (Energy Commission), conducts public interest research, development, and demonstration (RD&D) projects to benefit California's electricity and natural gas ratepayers. The PIER Program strives to conduct the most promising public interest energy research by partnering with RD&D entities, including individuals, businesses, utilities, and public or private research institutions.

PIER funding efforts are focused on the following RD&D program areas:

- Buildings End-Use Energy Efficiency
- Energy-Related Environmental Research
- Energy Systems Integration
- Environmentally Preferred Advanced Generation
- Industrial/Agricultural/Water End-Use Energy Efficiency
- Renewable Energy Technologies
- Transportation

In 2003, the California Energy Commission's Public Interest Energy Research (PIER) Program established the **California Climate Change Center** to document climate change research relevant to the states. This Center is a virtual organization with core research activities at Scripps Institution of Oceanography and the University of California, Berkeley, complemented by efforts at other research institutions. Priority research areas defined in PIER's five-year Climate Change Research Plan are: monitoring, analysis, and modeling of climate; analysis of options to reduce greenhouse gas emissions; assessment of physical impacts and of adaptation strategies; and analysis of the economic consequences of both climate change impacts and the efforts designed to reduce emissions.

The California Climate Change Center Report Series details ongoing Center-sponsored research. As interim project results, the information contained in these reports may change; authors should be contacted for the most recent project results. By providing ready access to this timely research, the Center seeks to inform the public and expand dissemination of climate change information, thereby leveraging collaborative efforts and increasing the benefits of this research to California's citizens, environment, and economy.

Quantifying Volatility and the Probability of Daily Precipitation Extremes is the final report for the Climatic Data Collection, Analyses, and Modeling project (contract number 500-02-004, work authorization number MR-025) conducted by Scripps Institution of Oceanography and University of Nevada, Reno.

For more information on the PIER Program, please visit the Energy Commission's website www.energy.ca.gov/pier/ or contract the Energy Commission at (916) 654-5164.

Table of Contents

Preface iii
Abstract vii
Executive Summary 1
1.0 Introduction 3
2.0 Project Approach 5
3.0 Results 15
4.0 Conclusions and Recommendations 17
5.0 References 19

List of Figures

Figure 1. Log likelihood ratio (L) computed for daily excesses over local median at each of the 560 stations. Values close to zero (blue and green circles) represent approximately exponential tails, while yellow, red, and black circles represent heavier tails. 8

Figure 2. Log likelihood ratio (L), as in Figure 1. Plates (a) and (b) show L for precipitation in the winter half-year (November–April) and the summer half-year, respectively. 9

Figure 3. Probability plots for excesses over threshold (local median, see Table 1) at selected stations. Sorted observed excesses displayed in mm along the x-axis are plotted against the corresponding theoretical quantiles derived from the fitted exponential (blue e's) and Pareto (red p's) models. 11

List of Tables

Table 1. Precipitation statistics at selected stations for the common observational period 1950–2001 10

Abstract

Standard statistical approaches to modeling daily precipitation cannot account for observed volatility, and they typically underestimate the probability of extreme events. This can result in dangerously incorrect risk assessment for practical applications and hinder climatic studies of extreme meteorological events. This study used stochastic theory to quantify volatility of daily precipitation amounts at hundreds of weather stations across North America and to identify a mathematically and physically consistent model for the probability of daily precipitation extremes. The study's results indicate that, characteristically, precipitation volatility can be explained by heavy-tailed probability models but not by the customary distributions with exponential tails. Moreover, by examining geographical structures in extreme precipitation behavior, this study shows that the degree of volatility observed at specific locations is determined by the diversity in precipitation-producing mechanisms. These results represent a fundamental development in stochastic modeling of precipitation extremes in climate, and they have immediate and vital practical implications.

Keywords: Precipitation, extremes, probability, weather, climate, geography, families of distributions, exponential tails, heavy tails, Pareto distribution

Executive Summary

Introduction

Standard statistical approaches to modeling daily precipitation cannot account for observed volatility, and they typically underestimate the probability of extreme events. This can result in dangerously incorrect risk assessment for practical applications and hinder climatic studies of extreme meteorological events. To quantify probabilities of extreme precipitation in current and future climates, analysts need to use a mathematical model that is consistent with observations and theory.

Purpose

This study used stochastic theory to quantify volatility of daily precipitation amounts at hundreds of weather stations across North America and to identify a mathematically and physically consistent model for the probability of daily precipitation extremes.

Project Objectives

By investigating the nature of probability tails across many different climatic regimes, the authors aim to understand and explain the climatic and geographic factors that determine precipitation volatility—the quality that fundamentally determines the size and frequency of precipitation extremes.

Project Outcomes

The study's results indicate that, characteristically, precipitation volatility can be explained by heavy-tailed probability models but not by the customary distributions with exponential tails. Moreover, by examining geographical structures in extreme precipitation behavior, this study shows that the degree of volatility observed at specific locations is determined by the diversity in precipitation-producing mechanisms.

Conclusions

These results represent a fundamental development in stochastic modeling of precipitation climate extremes, and they have immediate and vital practical implications. Wherever stochastic modeling of regional precipitation extremes is of interest, a classification of local precipitation into exponential or heavy-tailed families of probability distribution functions is the logical first step toward choosing a specific probability distribution function to represent local precipitation. This work provides, for the first time, an efficient method to do just that.

Recommendations

The authors recommend that the approach presented here for determining the form of the correct stochastic model be applied for estimating probabilities of specific precipitation

extremes, n-year events, return periods, and others, and for the quantification of extreme precipitation probability for use in climate research.

Benefits to California

This research provides a more robust quantification of extreme event probabilities essential to safe engineering design, hazard assessment, and other applications—as well as for fostering further investigations of hydrologic weather extremes and climate in the state.

1.0 Introduction

A correct parametric approach to modeling the probability of high-frequency hydrologic extremes is required to appreciate weather in the context of climate variability and change, and for the accurate estimation of extreme event probabilities necessary for important practical applications. The urgency of this problem is amplified by the prospect that global hydrologic change may disproportionately manifest itself in increased frequency of extreme precipitation. This view is supported by theoretical reasoning (Allen and Ingram 2002; Trenberth et al. 2003; Karl et al. 2003) and climate projections (Zwiers and Kharin 1998; Semenov and Bengtsson 2002; Hegerl et al. 2004; Wehmer 2004; Groisman et al. 2005), as well as empirical evidence for increasing trends in the frequency of extreme daily precipitation worldwide (Groisman et al. 2005; Tsonis 1996; Easterling et al. 2000; Folland and Karl 2001; Groisman et al. 1999; Groisman et al. 2004).

It is difficult, however, to define and account for high frequency precipitation extremes without a reasonable probability distribution model. Results derived from probability theory were used to develop an efficient automated scheme to distinguish between heavy and exponential precipitation probability tails in hundreds of daily station records spanning five decades over the North American continent.[1] These results suggest that daily precipitation extremes at a vast majority of the stations are not exponential, but more closely follow a power law—which means that statistical distributions traditionally used to model daily rainfall (e.g., exponential, Weibull, Gamma, lognormal) generally underestimate the probabilities of extremes.

The degree of this distortion—that is, its volatility—depends on regional and seasonal climatic peculiarities. It is most severe in regions that experience precipitation from a variety of large-scale (also known as *synoptic*) atmospheric systems producing wildly different precipitation rates. However, even at stations that experience precipitation exclusively from consistent synoptic sources, the heavy-tailed models may be superior to the exponential models and can lead to more realistic and safer estimates of extreme event probabilities, return periods, n-year events, and design limits. The correct choice of the distribution function is therefore essential to safe engineering design, hazard assessment, and other applications. It will also facilitate further investigations of extreme weather events and climate.

A plethora of theoretical probability distribution functions (PDFs) is available for modeling any type of environmental data, including rainfall. The correct PDF choice is not trivial. At best, it can be determined from a combination of physical reasoning and a careful examination of the connection between data and mathematics. At worst, it is dictated by convention and involves no thorough investigation of the reason why a particular PDF should fit the data.

[1] The tail of an exponential probability distribution function decays exponentially, while that of a heavy-tailed PDF decays as a power law, allowing more probability of occurrence for extreme events. When seen in a graph, heavy probability tails are thicker than exponential tails, providing reasonable probability for the occurrence of observed extremes—events that tend to be treated as "outliers" or improbable events according to the exponential model.

Needless to say, the two approaches can lead to significantly different results. The latter paradigm seems to be the typical approach to modeling daily rainfall with exponentially tailed PDFs. This approach, however, often leads to discarding extremely large events as outliers. If the extreme events are of primary importance, a more flexible approach is needed to encompass them into the body of the statistical model.

An alternative is provided by heavy-tailed distributions, such as Pareto (Johnson et al. 1994) or the stable laws (Samorodnitsky and Taqqu 1994). The stable laws arise naturally as approximations to sums of random quantities such as total precipitation accumulated over a time period. Heavy-tailed behavior has been documented in precipitation and streamflow data measured at specific locations or regions for time series of extreme values, such as annual peak record streamflow (Smith 1989; Katz et al. 2002) and precipitation (Katz et al. 2002), as well as threshold exceedances for streamflow (Anderson and Meerschaert). Heavy tails have also been suggested as reasonable models for daily precipitation amount at a large number of U.S. stations (Smith 2001). However, in cutting edge climate and hydrology research exponentially tailed PDFs are still typically used to model daily rainfall, including extremes (Tsonis 1996; Groisman, Karl, and Easterling 1999; Zolina et al. 2004).

2.0 Project Approach

This study sought to investigate the heaviness of the PDF tail for all heavy and extreme daily precipitation events at hundreds of weather stations over North America and to examine the spatial structure of the result. To document heavy-tail behavior, the research team focused on an approach for differentiating between exponential and heavy-tailed precipitation distributions.

The team was specifically interested in extreme precipitation observations—events that are traditionally called "outliers" because they are unaccounted for by the tails of traditional PDFs, those with exponentially decaying probabilities of extreme events. According to power laws for the heavy-tailed distributions, these probabilities decay much slower. Thus, the heavy-tailed models allow greater variability and can explain observed extreme events larger than those predicted by the exponential models. This study applied these laws to daily precipitation extremes observed at many stations and compared the performance of heavy tails with that of the traditional exponential tails.

The research team was immediately confronted by the fact that many choices exist for the specific PDF form with exponential (e.g., Gamma; Weibul; double, stretched or extended exponential) and heavy tails (e.g., Cauchy, Pareto). To resolve the problem of distributional multiplicity (that is, to avoid having to choose specific PDFs from the exponential and heavy-tailed families), researchers applied the results of the peak-over-threshold (POT) theory (Reiss and Thomas 2001). Because the researchers were primarily interested in estimating probabilities of large (extreme) events, the POT method helps focus the search for the reasonable models and provides a rigorous mathematical foundation for the results. The POT method involves examination of the data falling above a threshold—that is, excesses over the threshold. For any data value X and threshold u, $X^{[u]}$ is the exceedance; that is, the (conditional) value of X given that X exceeds u. Excesses $X^{[u]}$-u are considered over the threshold u.

The Balkema-de Haan-Pickands theorem (Balkema and de Haan 1974; Pickands 1975) provides the limiting distribution of excesses. The theorem states that when the threshold (u) increases, the distribution of the excesses $X^{[u]}$-u converges to a Generalized Pareto (GP) distribution. Any GP distribution has to be one of three kinds: exponential, Pareto, or Beta. The importance of this result for practical applications is: no matter what the specific original distribution of X is, the excess $X^{[u]}$-u over any threshold u is (approximately) one of only three distributions. The three distributions of the excesses correspond to the tails of the original distribution of X. If X has exponential tail, then $X^{[u]}$-u will have an approximately exponential distribution. If X has heavy tail, then $X^{[u]}$-u will have an approximately Pareto distribution. Beta distribution has finite support and it is not considered here. The problem of finding the correct PDF for describing daily extreme event probabilities reduces, therefore, to classifying excesses at a particular station. It is necessary to seek a decision rule (ideally a formal statistical test) for classifying the over-threshold precipitation excesses into either exponential or Pareto models.

The research team approached this problem using ideas from the theory of likelihood ratio tests (Lehmann 1997). The approach was to consider the ratio of the maxima of the likelihoods of the

observed sample under the null (Pareto in the numerator) and alternative (exponential in the denominator) models. The logarithm of the likelihood ratio statistic is:

$$L = \log\left(\frac{\sup\limits_{\alpha>0, s>0} L_{Pareto}(\overset{\rho}{x}|\alpha, s)}{\sup\limits_{\sigma>0} L_{\exp}(\overset{\rho}{x}|\sigma)} \right),$$

where $\overset{\rho}{x}$ is the observed sample (of excesses) and $L_{Pareto}(\overset{\rho}{x}|\alpha,s)$ and $L_{\exp}(\overset{\rho}{x}|\sigma)$ are the likelihood functions of the sample under Pareto and exponential hypotheses, respectively. Researchers used a Pareto distribution with the survival function $S(x) = P(X>x) = (1/(1+x/s\alpha))^{\alpha}$ and an exponential distribution with the survival function $S(x) = P(X>x) = \exp(-x/\sigma)$.

In the Pareto case, the α parameter determines the thickness of its tail and is of primary importance. The scale parameter s is of secondary importance. In the exponential case, σ is the scale parameter. Note that as the tail parameter α increases, the Pareto survival function converges to the exponential one. This fact adds particular difficulty to differentiating between the two distributions as α increases.

To compute the log likelihood ratio (L), both likelihoods are first maximized (via maximum likelihood estimates of the parameters), and then the natural logarithm of their ratio is taken as the likelihood ratio statistic. The intuition behind this approach is as follows. If the sample (i.e., excesses over threshold) is more likely to come from the Pareto distribution, then the maximum Pareto likelihood will be larger than the exponential one. Thus, the ratio will be greater than one ($L > 0$). If the sample is most likely to come from the exponential distribution, the ratio will be less than one ($L < 0$).

The research team programmed and tested the computation of L from the numerical point of view, ran several simulation studies, and applied the method to the precipitation data. Although researchers chose the thresholds in an intuitive manner, fairly robust and meaningful results were obtained for daily precipitation.[2] The problem of choosing the optimal threshold is recognized as a very difficult and still open one (Smith 1987; Davidson and Smith 1990; Smith 1994; Gross et al. 1994).

The research team has computed L for a subset of over ten thousand available station records of daily rainfall distributed across Canada (Vincent and Gullett 1999), the United States (Groisman et al. 2004), and Mexico (Miranda 2003)—all quality controlled and homogenized at the National Climatic Data Center. The five-hundred-sixty best-quality stations were selected to give a reasonable coverage (with at least one station within a 70 kilometer radius) over North America. The selection used the most stringent quality standards in well-sampled regions and

[2] Here, results are presented based on median threshold, although reasonable deviations in threshold choice do not substantially affect the results in most regions.

relaxed these standards to include all stations in regions with the poorest coverage: Alaska, Northern Canada, and parts of Mexico. In mountainous regions, the researchers always included the highest station along with the best-quality station. At least 80% of the data was required to be present at all stations for the common observational time period: January 1, 1950 to December 31, 2001. This study compared fits to excesses over local medians derived from Pareto probability models to those obtained from the widely used exponential models. Figure 1 shows the L at each station.

To help interpret the result presented in Figure 1, note again that the Pareto distribution can approach the exponential; therefore, values of L smaller than zero, indicating exponential tails, rarely occur (fewer than 5% of the stations, blue circles on Figure 1). Values close to zero, therefore, represent approximately exponential tails (blue and green circles on Figure 1), while the positive magnitude of L is, in theory, a measure of the "heaviness" of the empirical precipitation tail. Although a formal likelihood ratio test for the acceptance or rejection of the null hypothesis (exponential tails) is being developed (Kozubowski et al. 2007), this study's current, rather informal result suggests a spatially coherent dependence of tail type on geographical location with respect to climate.

First, the vast majority of local tails are, strictly speaking, non-exponential (i.e., $L > 0$). The degree of departure from the exponential model depends on the peculiarities of regional climatic regimes. Precipitation at stations with a large variety of meteorological influences (i.e., frontal, thunderstorm, tropical cyclone, intense and organized convection) tends to display highly non-exponential tails. This occurs especially where extreme weather interacts with noteworthy topography (such as the Mexican Gulf coast or the front range of the Rockies), the regions of hurricane landfalls (such as the Gulf and East coasts of the United States), and along the Tornado Alley in the middle of the country. In the Great Plains and the Midwest, especially the north, heavy precipitation events can appear more extreme relative to typically light winter snow accumulations (see below).

Close to exponential behavior, on the other hand, occurs where precipitation is reasonably frequent and preferably of one type (i.e., generated by similar-type systems). For example, the exclusively convective summer precipitation along the central Mexican high plateau, as well as isolated regions of almost exclusively frontal-system-generated precipitation along the windward side of gently sloping topography (such as the Sierra Nevada and Rocky Mountains), result in nearly exponential tails.

A simple way to classify precipitation into frontal and convective types is to consider winter and summer data separately. Figure 2 presents L computed for the cold and warm half-year separately.

Log likelihood ratio, all data

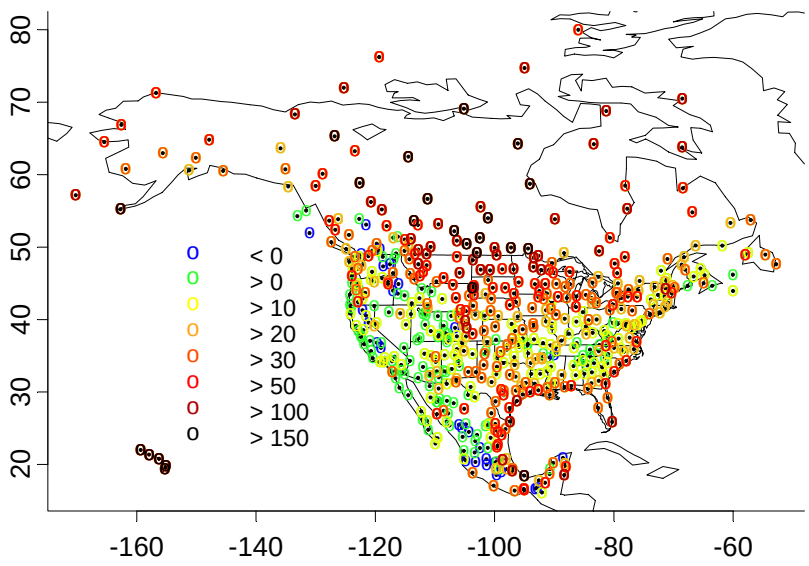


Figure 1. Log likelihood ratio (L) computed for daily excesses over local median at each of the 560 stations. Values close to zero (blue and green circles) represent approximately exponential tails, while yellow, red, and black circles represent heavier tails.

The frontal-system-generated precipitation over the southwestern and eastern United States in winter (Figure 2a) and the thunderstorm-generated rainfall over the Mexican high plateau in summer (Figure 2b) tend to result in lighter, more exponential tails.

On the other hand, seasonal precipitation PDFs with notably heavy tails are produced by all of the following:

- Landfalling tropical cyclones along the East and Gulf Coasts in summer.
- Occasional extreme snowstorms in winter and thunderstorms in summer mixed in with the usually mild precipitation over the northern Great Plains.
- Rare convective storms in otherwise dry California summer.
- Rare but intense winter precipitation associated with the nortes (on-shore winds) along the Mexican Gulf Coast from southern Texas south to the Isthmus of Tehuantepec.

a. Cold season data (November – April) b. Warm season data (May – October)

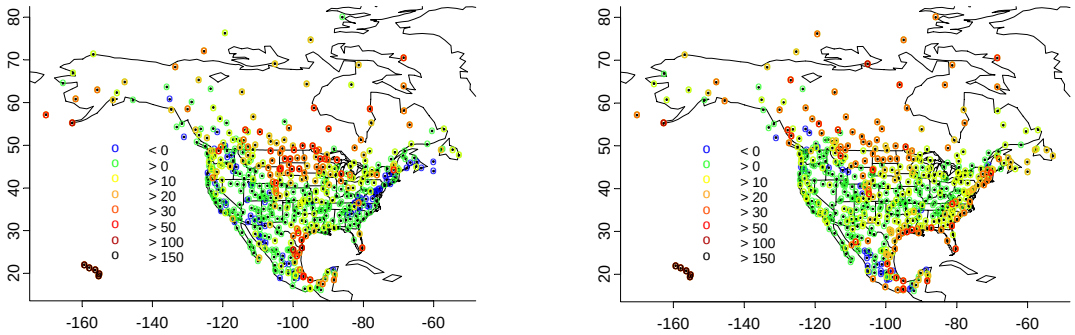


Figure 2. Log likelihood ratio (L), as in Figure 1. Plates (a) and (b) show L for precipitation in the winter half-year (November–April) and the summer half-year, respectively.

Generally, L computed for all precipitation data (Figure 1) is an aggregation of seasonal values, reflecting the volatility of precipitation in different seasons. It is clear, for example, that the volatility (large values of L) observed along the East Coast of the United States is primarily due to summertime precipitation, while the lighter-tailed behavior of precipitation in the southwestern United States, generally, is mostly due to consistent midlatitude cyclones in winter.[3]

The most general comparison of Figures 1 and 2, however, indicates that daily precipitation tails are heavier when all seasons' data is considered compared to individual seasons, clearly illustrating the idea that the diversity of precipitation-producing systems is essentially responsible for precipitation volatility. Moreover, notably heavy precipitation tails observed in all seasons along the precipitous Mexican part of the Gulf Coast (as well as along the steep front range of the Rockies) testify to the additional effect of sharp topography in enhancing precipitation volatility.

Figure 3 helps provide a visual, intuitive feel for what the measure L represents. with respect to the form of the probability tails and their fit to the data. It shows probability plots for stations selected to represent specific precipitation regimes and to span (roughly) the range of L. These probability plots show how heavy and extreme events are accounted for by exponential and

[3] Summer storms in this region of infrequent but intense winter accumulations are generally rare and localized in space and time, so as to make their effect on the whole insignificant. A short-lived convective downpour, even if it happens to rain out over a gauge, will not typically add up to the amount generated over a day by an intense frontal cyclone in winter. In the northern plains, however, where summertime daily accumulations, are significantly larger than those in winter, the effect of including all seasons' data strongly inflates the cumulative tail relative to the already heavy seasonal tails.

Pareto models. Figure 3 presents graphical pairs—scatter plots of the quantiles of the empirical distribution against the quantiles of a theoretical model: exponential or Pareto.

As in the calculation of L (Figures 1 and 2), the daily station data used are excesses over local median of the daily amounts recorded on wet days (i.e., moderate to extreme events). In this type of plot, the best model is the one that scatters along a straight line. Table 1 presents precipitation statistics at these selected stations for the common observational period 1950–2001: L, probability of precipitation (that is, percent of days with recorded precipitation), median daily total on days with precipitation, maximum recorded daily total, the estimated 100-year event assuming exponential and Pareto tails, and the Pareto probability of exceeding the 100-year event estimated assuming an exponential tail.

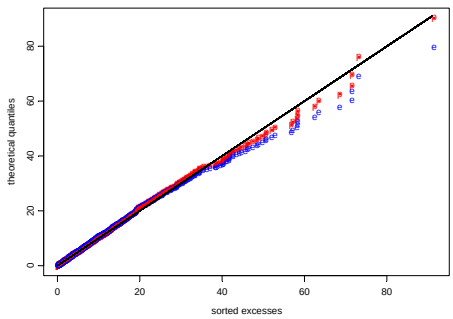
Station	Log likelihood ratio (L)	P[p > 0] (%)	median($p_{p>0}$) (mm)	$Max_{obs}(p)$ (mm)	100-yr event Exp and Pareto (mm)	Pareto $P[p > p_{exp}^{100}]$ (%)
Sacramento	1.38	16	4.3	95	79 and 90	2.3
Nashville	6.32	26	7.1	153	115 and 145	4.7
St. Louis	20.9	30	4.1	142	101 and 158	9.4
Houston	51.3	27	5.1	253	167 and 343	15.7
Miami	118.8	36	4.6	377	149 and 412	28.0
Williston	149.1	26	1.5	125	49 and 200	35.1

Table 1. Precipitation statistics at selected stations for the common observational period 1950–2001

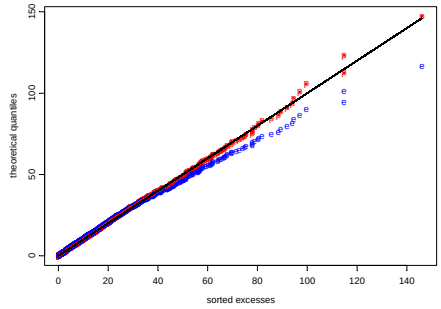
One example of an approximately exponential station, is the Sacramento airport (Sacramento FAA AP) (40.52°N, 123.82°W, 640m, Figure 3a), located in California's Central Valley, where rain is produced almost exclusively by frontal systems in the winter half year. Both exponential and Pareto models fit these data reasonably well, and L is close to zero. However, the shape of the Pareto tail still differs significantly from the exponential tail, especially beyond the observed range. In the 52 years of analyzed data at Sacramento, the median daily total precipitation was 4.3 millimeters (mm) and the maximum was 95 mm. Although both models seem to fit the observed data well, model choice can lead to significant differences in inferences regarding the probabilities of hitherto unobserved extremes. For example, based on the observed record and assuming the exponential (Pareto) model, the estimated 100-year event is 79 (90) mm/day. The Pareto model implies that the 100-year event estimated by the exponential model is 2.3 times as likely to occur—that is, a 43-year event.

PROBABILITY PLOTS FOR SELECTED STATIONS

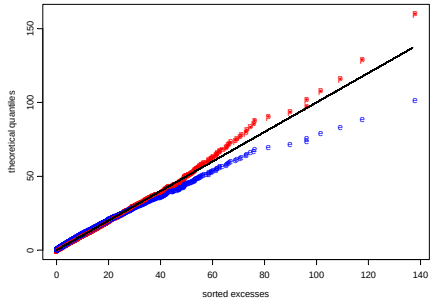
a. Sacramento, L = 1.38



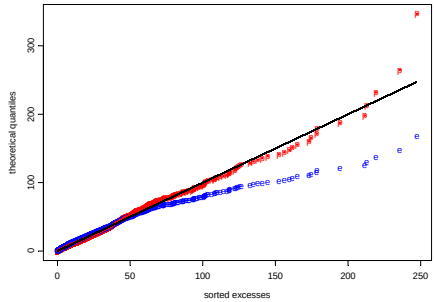
b. Nashville, L = 6.32



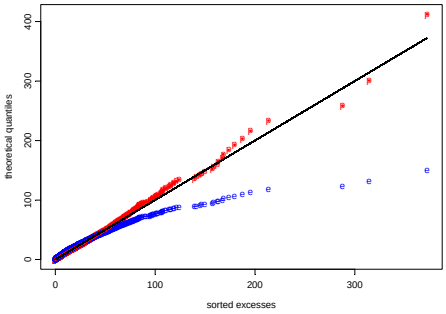
c. St. Louis, L = 20.9



d. Houston Hobby Airport, L = 51.3



e. Miami WSCMO Airport, L = 118.8



f. Williston ND, L = 149.1

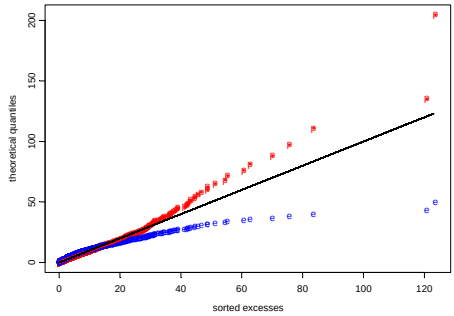


Figure 3. Probability plots for excesses over threshold (local median, see Table 1) at selected stations. Sorted observed excesses displayed in mm along the x-axis are plotted against the corresponding theoretical quantiles derived from the fitted exponential (blue e's) and Pareto (red p's) models.

Exponential tails become progressively less adequate as stations with progressively larger L are considered. This is readily visible on the probability plots (Figure 3b–f) and by comparing expected values of 100-year events with the observed 52-year station maxima (Table 1). Pareto tails perform reasonably well—certainly better by comparison with the exponential for all stations considered. This is not to say that they perform perfectly well, though. For large values of L, Pareto tails tend to overestimate the observed probabilities of extremes, as exemplified by

above-median precipitation recorded at Williston, North Dakota (Figure 3f). The true probabilities of extremes appear to lie somewhere between those given by exponential tails and those given by Pareto tails.

This type of behavior is typical of stations with the largest observed values of L, which tend to cluster in the northern Great Plains and scatter into Canada. These high-latitude northern plains stations, where the most volatile precipitation behavior tends to be associated with snowfall (i.e., it occurs in winter, see Figure 2) and where light precipitation events (i.e., light and dry snow) are typically accompanied by strong winds, small precipitation totals tend to be severely underestimated (Golubev and Bogdanova 1996; Groisman, Peck, and Quayle 1999; Bogdanova 2002).

Heavy wintertime precipitation events are associated with wet snow that is efficiently caught by the gauge, and, therefore, totals on such days can appear more extreme in relation to the underestimated modest precipitation. Moreover, the effect of "false" precipitation due to blowing snow events increases the frequency of light precipitation, decreases the median, and further inflates the relative magnitude of heavy and extreme precipitation. Therefore, in high-latitude plains regions, the research team suspects that true precipitation tails may appear somewhat more in line with Pareto tails than observations suggest, although they are probably less heavy.

Considering summer and winter data separately results in a smaller L (30–50 for both seasons) and tails in line with the Pareto model. Increasing the threshold to the 70th–80th percentiles has a similar effect on all data at such stations (result not shown). Summer and winter daily (as well as total precipitation) amounts typically differ by a factor of three in these northern plains regions (result not shown) and mixing all seasons' data probably requires a higher threshold choice. The "blowing snow" effect may aggravate this necessity. In either case, the example of northern plains stations results points to the need of an objective system for choosing thresholds. Nevertheless, the choice of median as threshold adapted here works well for daily precipitation at most other North American locations considered.

The above-median data at a vast majority of the stations is more in line with the Pareto model. Even at high-latitude plains stations, where the truth appears to lie between Pareto and exponential models, Pareto is clearly the safer choice. The comparison of 100-year events estimated using the exponential and Pareto models with the observed 52-year station maxima (Table 1) reinforces the choice of heavy over exponential models to describe probabilities of extreme precipitation events. Assuming, as this study's results suggest, that the Pareto model is correct in most cases, the choice of exponential model may lead to a severe underestimation of extreme precipitation risk. In Miami, for example, the 100-year event estimated by the exponential model corresponds to a 3.6-year event estimated by Pareto; that is, the Pareto model predicts that the 100-year event is 28 times more likely to occur than the exponential model does.

The elegant part of the statistical methodology adapted here is that it allows us to simply evaluate whether the tails of an empirical distribution are exponential or heavy, regardless of

the specific form of the PDF from the exponential or heavy-tailed families of distributions. So, it is not necessary to test whether the data is or is not exponential, Gamma, Weibull, Double (or sometimes “stretched” or “extended”) exponential, or any other exponentially tailed PDF commonly used to model daily rainfall. These are accepted or rejected under the broad umbrella of the exponentially tailed family of distributions. Although theoretical questions related to the choice of threshold and rigorous hypotheses testing remain to be solved, the Pareto/exponential likelihood ratio for peaks over threshold has already proven a useful and efficient tool in diagnosing tail behavior and distinguishing between exponential and heavy tails.

3.0 Results

Results reported here confirm that most of the precipitation exceedance probabilities follow the Pareto law and therefore, most precipitation PDF tails follow a power law (that is, they are heavy, not exponential). Threshold exceedances at stations that may have approximately exponential tails can also be thought of as limiting, special cases of the Pareto model (i.e., Pareto with a very large tail parameter). That is because as α increases, the Pareto distribution gets close to the exponential and the L statistics may not be sensitive enough for high values of α. But such cases appear to be extremely rare in North America and, in practice, the heavy tail is clearly superior. The divergence from exponential is a matter of degrees. The most closely exponential tails are found at stations where similar systems and consistent processes produce virtually all of the precipitation. How far from exponential the tails are depends on location and season. Places that receive precipitation from a variety of systems tend to exhibit severely non-exponential precipitation tails; diversity is the mother of volatility. Severe diversity in topography also appears capable to fatten precipitation tails.

4.0 Conclusions and Recommendations

This report is meant as a first step to relate hydrologic weather extremes to climate. It is based on a physically and mathematically consistent attempt to choose a correct model for describing the probability of extreme precipitation events. More rigorous statistical results are yet to be worked out. However, this study's empirical results, although admittedly involving comparisons and value judgments, strongly suggest the superiority of heavy tails to the traditional exponential. An intriguing consequence of the heavier tail is that shifts in central tendency are not necessarily amplified in the tail, as they are in the exponential model. Therefore, the effect of climate change on precipitation extremes may be least severe for stations with the heaviest tails. The authors believe that the spatial structure and coherence of these results and their solid grounding in climate and weather patterns over the North American continent shows the great scientific potential and practical benefits of using stochastic methods for the extremes in climate research.

5.0 References

Allen, M. R., and W. J. Ingram. 2002. "Constraints on future changes in climate and the hydrological cycle." *Nature* 419: 224–232.

Anderson, P. L., and M. M. Meerschaert. 1998. "Modeling river flows with heavy tails." *Water Resour. Res.* 34: 2271–2280.

Balkema, A. A., and L. de Haan. 1974. "Residual life time at great age." *Ann. Probab.* 2: 792–804.

Bogdanova, E. G., B. M. Ilyin, and I. V. Dragomilova. 2002. "Application of a Comprehensive Bias-Correction Model to Precipitation Measured at Russian North Pole Drifting Stations." *J. Hydrometeor.* 3: 700–713.

Davidson, A. C., and R. L. Smith. 1990. "Models for exceedances over high thresholds." *J. R. Statistic. Soc. B.* 52: 393–442.

Easterling D. R., et al. 2000. "Observed variability and trends in extreme climate events: A brief review." *Bull. Am. Meteorol. Soc.* 81: 417–425.

Folland, C. K., and T. R. Karl. 2001. *Observed Climate Variability and Change. Climate Change 2001: The Scientific Basis.* Contribution of Working Group 1 to the Third IPCC Scientific Assessment. J.T. Houghton, et al. Eds. (Cambridge Univ. Press, Cambridge, U.K.). 99–181.

Golubev, V. S., and E. G. Bogdanova. 1996. *Accounting of blowing snow events in precipitation measurement in Russia.* Proc. of the ACSYS Solid Precipitation Climatology Project Workshop, Reston, Virginia, USA, 12–15 September 1995, World Meteorol. Organ., WMO/TD No. 739.

Groisman, P. Ya., T. R. Karl, and D. R. Easterling. 1999. Changes in the probability of heavy precipitation: Important indicators of climatic change. *Clim. Change* 42: 243–283.

Groisman, P. Ya., E. L. Peck, and R. G. Quayle. 1999. "Intercomparison of recording and standard non-recording U.S. gages." *J. Atmos. Ocean. Technol.* 16: 602–609.

Groisman P. Ya., et al. 2004. "Contemporary Changes of the Hydrological Cycle over the Contiguous United States: Trends Derived from In-Situ Observations." *J. Hydrometeor.* 5: 64–85.

Groisman, P. Ya., R. W. Knight, D. R. Easterling, T. R. Karl, and G. C. Hegerl. 2005. "Trends in intense precipitation in the climate record." *J. Clim.* 18: 1326–1350.

Gross, J., A. Heckert, J. Lechner, and E. Simiu. 1994. *Novel extreme value estimation procedures: Application to extreme wind data. Extreme Value Theory and Applications.* J. Galambos et al. Eds. Dordrecht, The Netherlands: Kluwer Academic Publishers. 139–158.

Hegerl, G. C., F. W. Zwiers, P. A. Stott, and V. V. Kharin. 2004. "Detectability of anthropogenic changes in temperature and precipitation extremes." *J. Clim.* 17: 3683–3700.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions, 2nd ed.* New York: Wiley.

Karl, T. R., and K. E. Trenberth. 2003. "Modern global climate change." *Science* 302: 1719–1723.

Katz, R.W., M. B. Parlange, and P. Naveau. 2002. "Statistics of extremes in hydrology." *Adv. Water Res.* 25: 1287–1304.

Kozubowski, T. J., A. K. Panorska, F. Qeadan, and A. Gershunov. 2007. Testing exponentiality versus Pareto distribution via likelihood ratio. Submitted to the *Journal of the American Statistical Association*.

Lehmann, E. L. 1997. *Testing Statistical Hypotheses, 2nd ed.* New York: Springer.

Miranda, S. 2003. *Actualización de la base de datos ERIC II.* (Final report of the project TH-0226, IMTA internal reports).

Pickands, J. 1975. "Statistical inference using extreme value order statistics." *Ann. Stat.* 3: 119–131.

Reiss, R.-D., and M. Thomas. 2001. *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields, 2nd ed.* Basel: Birkhäuser.

Semenov, V. A., and L. Bengtsson. 2002. "Secular trends in daily precipitation characteristics: Greenhouse gas simulation with a coupled AOGCM." *Clim. Dyn.* 19: 123–140.

Samorodnitsky, G., and M. Taqqu. 1994. *Stable non-Gaussian Random Processes.* New York: Chapman & Hall.

Smith, R. L. 1987. "Estimating tails of probability distributions." *Ann. Stat.* 15: 1174–1207.

Smith, J. A. 1989. "Regional flood frequency analysis using extreme order statistics of the annual peak record." *Water Resour. Res.* 25: 311–317.

Smith, R. L. 1994. *Multivariate threshold methods. Extreme Value Theory and Applications.* J. Galambos et al. eds. Dordrecht, The Netherlands: Kluwer Academic Publishers. 225–284.

Smith, R. L. 2001. "Extreme value statistics in meteorology and the environment." *Environmental Statistics.* Chapter 8, 300–357.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons. 2003. "The changing character of precipitation." *Bull. Amer. Meteorol. Soc.* 84, 1205–1217.

Tsonis. A. A. 1996. "Widespread increases in low-frequency variability of precipitation over the past century." *Nature* 382: 700–702.

Vincent, L. A., and D. W. Gullett. 1999. "Canadian historical and homogeneous temperature datasets for climate change analyses." *Int. J. Climatol.* 19: 1375–1388.

Wehmer, M. F. 2004. "Predicted Twenty-first-century Changes in Seasonal Extreme Precipitation Events in the Parallel Climate Model." *J. Clim.* 17: 4281–4290.

Zolina, O., A. Kapala, C. Simmer, and S. K. Gulev. 2004. "Analysis of extreme precipitation over Europe from different reanalyses: A comparative assessment." *Global Planet. Change* 44: 129–161.

Zwiers, F. W., and V. V. Kharin. 1998. "Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling." *J. Clim.* 11: 2200–2222.