Analysis of Nonstationary Extreme Events

Norbert A. Agana; Mohammad Gorji Sefidmazgi; Abdollah Homaifar

Department of Electrical Engineering, North Carolina A&T State University, Greensboro, NC, USA.
{naagana, mgorjise}@aggies.ncat.edu, homaifar@ncat.edu

Abstract

Extreme events by definition are rare events that occur infrequently but their impacts on both physical and socioeconomic resources are very enormous. Extreme climate events such as heavy precipitation, drought, tropical cyclones, hurricanes and heat waves are known to have tremendous impact on the society. Over the last few decades, our understanding of the mean behavior of the climate and its normal variability has improved to a large extend but the same cannot be said of climate extremes. Climate extremes represent nonlinear systems that are very hard to study and even harder to make predictions on them. The objective of this paper is to assess how these extreme events relate to modes of climatic variability such as El Nino–Southern Oscillation, the Pacific Decadal Oscillation and the North Atlantic Oscillation by utilizing the familiar distributions that arise out of the extreme value theory such as the generalized extreme value distribution and the generalized Pareto distribution. Nonstationarity is ensured by expressing the parameters of the distribution as functions of the covariates.

Keywords: Extreme Events**;** Covariates; Maximum likelihood; Bayesian Information Criterion.

Introduction

Extreme events are rare events that occur infrequently but have enormous impact on both physical and socioeconomic resources. Extreme climate events such as heavy precipitation, temperature, drought, tropical cyclones, hurricanes and heat waves have had tremendous impact on the society, costing lives and property. There is therefore no surprise that considerable attention has been given to climate studies in the past decades. However, in analyzing precipitation events, the majority of the existing methods are based on the assumption that precipitation time series are stationary, implying that the distribution of precipitation events is not significantly affected by climatic trends, long-term cycles or modes of climate variability (Gorji Sefidmazgi, Sayemuzzaman, and Homaifar 2014; Gorji Sefidmazgi, Sayemuzzaman, et al. 2014; Agana, Sefidmazgi, and Homaifar 2014; Vogel, Yaindl, and Walter 2011; AghaKouchak et al. 2013).

Although relaxing the assumption of stationarity can lead to accurate models, the results can be potentially misleading. Hence, there is the need to modify the assumption of a series of independently and identically distributed data with constant properties through time (stationarity) to reflect the effect of long-term climate change on the variable of interest. For instance, the maximum time series of climatic variables such as temperature and precipitation could show trends over time (Panagoulia, Economou, and Caroni 2014). Also, due to natural climate variability or anthropogenic climate change, there is evidence that the hydroclimatic extreme series are not stationary (Jain and Lall 2001; Milly et al. 2008). Large-scale modes of climate variability such as El Nino–Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), and the North Atlantic Oscillation (NAO) are known to have profound impacts on the precipitation regimes, especially during the winter season over North America. A number of researchers have studied the impact of modes of climate variability on climate extremes and have shown that these variables have great influence on extreme precipitation and temperature (Zhang et al. 2010; Griffis and Stedinger 2007).

ENSO events, in particular have influence on the occurrence of precipitation events. It has also been shown that there is a well-established connection between the two phases of ENSO and the North American precipitation (Cayan, Redmond, and Riddle 1999; Gershunov and Barnett 1998; Ropelewski and Halpert 1986; Shabbar, Bonsal, and Khandekar 1997). El Nino events influence the frequency of occurrence of different daily precipitation magnitudes in Western U.S. winters and tend to be associated with an increase in the frequency of high daily precipitation over the Southwest but a decrease in the Northwest (Cayan, Redmond, and Riddle 1999). In order to Model nonstationary extreme events within the framework of the GEV distribution, the GEV distribution requires extended models with covariate-dependent changes in at least one of the distribution’s parameters (Coles 2001).

The objective of this research is to assess how these extreme events relate to modes of climatic variability such as the ENSO, NAO and PDO. Similar work has been carried out on non-stationary extreme events where they considered only trend in their analysis (AghaKouchak et al. 2013; Katz, Parlange, and Naveau 2002; Feng, Nadarajah, and Hu 2007). Instead of analyzing only trend, we have also analyzed the effect of ENSO on extreme precipitation and sea level rising. We achieved these by utilizing the familiar generalized extreme value (GEV) distribution that arises out of the extreme value theory (EVT) (Coles 2001). Non-stationarity is ensured by expressing the parameters of the GEV distribution as functions of time and ENSO. The maximum likelihood estimation (MLE) method is employed to estimate the distribution parameters (Coles 2001; Katz, Parlange, and Naveau 2002; Vogel, Yaindl, and Walter 2011). We applied the model to precipitation data in Pasquotank, North Carolina and also to the sea level data at Pensacola, Florida. Furthermore, we have compared the different fitted models and selected the best model based on the Bayesian Information Criterion (BIC). This paper demonstrates that covariate-dependent models are necessary for analyzing extreme events, especially precipitation extremes. Also based on the results obtained from this work, it is observed that linear parameter-covariate dependence might not be able to relate the dependence of the parameters on the covariates well and therefore nonlinear dependent models might be appropriate.

Methodology

The foundation of Extreme Value Theory (EVT) is the Generalized Extreme Value (GEV) distribution (AghaKouchak et al. 2013; Coles 2001). The GEV distribution classically models block maxima (or minima) of data over a certain period of time such as daily, monthly or annual maxima. The block maxima refers to the number of years (for annual maxima) from which the maxima is taken. The justification of the GEV arises from an asymptotic argument that postulates that as the sample size increases, the distribution of the sample maxima, for example X, follow a Frechet, Weibull or Gumbel distribution. The EVT characterize rare events by describing the tail behavior of the underlying distribution. Let the time series denoted by $\{X_1, X_2 \ldots X_n\}$ be independent random variables having a distribution function G.

Let $M_n = max\{X_1, X_2, \ldots, X_n\}$ suppose there exist normalizing constants $a_n>0$ and $b_n>0$ such that

$pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x)$ as $n \rightarrow \infty$ then the cumulative distribution function for the GEV distribution is defined as shown in Equation (1) (Coles 2001; Katz, Parlange, and Naveau 2002). If n is the number of observations in a year, then M_n is the annual maximum.

$$G(x,\mu,\sigma,\xi) = \begin{cases} \exp\left[-\left(1+\frac{\xi(x-\mu)^{\frac{1}{\xi}}}{\sigma}\right)\right] \\ 1+\frac{\xi(x-\mu)}{\sigma} > 0, \xi \neq 0 \end{cases} \tag{1}$$

Where μ, σ>0, ξ are the location, scale and shape parameters respectively. The expression in Equation (1) can be made non-stationary by expressing the parameters of the distribution as linear functions of covariates which have influence on the occurrence of extreme events. In our case, we only expressed the location parameter as a function of time and the El Nino Southern Oscillation (Nino 3.4), which are shown in Equations (3) to (5).

Model 1: $\mu = \mu_o$ (2)

Model 2 : $\mu(t) = \mu_o + \mu_1 t$ (3)

Model 3: $\mu(y) = \mu_o + \mu_1 y$ (4)

Model 4 : $\mu(t, y) = \mu_o + \mu_1 t + \mu_2 y$ (5)

The combined effect of both time and the El Nino Southern Oscillation is shown in Equation (5) where *t* (time) is the year in which the maxima is taken and *y* the covariate representing the Nino 3.4. The above GEV distribution models are fitted to both the annual monthly maxima of precipitation data at Pasquotank, North Carolina and the mean sea level data at Pensacola, Florida.

Parameter Estimation

All the model parameters are obtained using the maximum likelihood estimation (MLE) procedure. Although other methods such as the *Method of Moments* (MOM), *Probability Weighted Moments* (PWM) can be used, we exclusively used the MLE because of its easy adaptability to non-stationary conditions (Katz, Parlange, and Naveau 2002; El Adlouni et al. 2007). Also, the advantage of maximum-likelihood estimators is that they can employ censored information without difficulty (Martins and Stedinger 2000).

If $G(x(t);\mu(t),\sigma(t),\xi(t))$ is the probability density function of a random variable x with $\mu(t),\sigma(t)$ and $\xi(t)$ as parameters, the log likelihood of the GEV distribution is simply given by (Coles 2001):

$$L(\mu,\sigma,\xi) = \prod_{t=1}^{n} g(x(t);\mu(t),\sigma(t),\xi(t)) \tag{6}$$

Both the stationary and nonstationary models of the GEV distribution can be fitted to the time series of the random variable x by maximizing the log-likelihood of the function (Coles 2001):

$$l(\mu,\sigma,\xi) = -n\log\sigma(t) - (1+1/\xi(t))\sum_{i=1}^{n}\log\left(1+\xi(t)\frac{x_i-\mu(t)}{\sigma(t)}\right) - \sum_{i=1}^{n}\left(1+\xi(t)\frac{x_i-\mu(t)}{\sigma(t)}\right)^{-1/\xi(t)} \quad (7)$$

which is obtained by taking the log of Equation (6).The maximum likelihood estimates are the values of the parameters μ, σ and ξ that maximize the log likelihood function in (7). Time is used as an explanatory variable (covariate). The parameters μ, σ and ξ can also be expressed as functions of other explanatory variables and similar procedure is followed to estimate them. Instead of maximizing the log likelihood, we can rather minimize the negative log likelihood of Equation (7). Numerical methods such as the Newton-Raphson iteration algorithm can be used to solve Equation (7) (Martins and Stedinger 2000; Hosking 1985).

Model Selection

Model choice is usually necessary when you have more than one model to choose from. For instance, to compare two nested models (usually between a simpler model and a complex model), we can easily apply the *Likelihood Ratio Test* (LRT) to select the best model by computing the test statistic and determining whether it is significant or not. However, the use of the likelihood ratio test becomes cumbersome when there are more than two models to choose from. Model selection techniques such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) can be used to select the best model among a collection of nested models. The BIC selection criterion is applied here to select the best model among a collection of nested models (AghaKouchak et al. 2013; Gorji Sefidmazgi, Moradi Kordmahalleh, et al. 2014). The BIC selects the model that minimizes the quantity:

$$BIC(k) = -2l(k) + k\ln n \quad (8)$$

where l is the log-likelihood which is obtained from Equation (6). Also, k and n are the number of parameters and number of block maxima respectively (number of years in this case).

Data Sets

Monthly precipitation data in North Carolina for the time period 1950-2008 was obtained from the National Climatic Data Center (NCDC). We selected the station near the Atlantic Ocean. Figure 1 shows a time series plot of the annual maxima of monthly precipitation data at Pasquotank as well as a scatter plot showing how the precipitation vary with the El Nino Southern Oscillation Index (Nino3.4).

Also, we have analyzed the annual mean sea level at Pensacola, Florida during the same time period of 1950-2008. The sea level data was obtained from the University of Hawaii Sea Level Center. Figure 2 shows a time series plot of the mean sea level data as well as a scatter plot of the sea level versus the El Nino Southern Oscillation Index (Nino3.4). Most climate indices such as ENSO, NAO and PDO do not contain values beyond 1950. Hence, in order to analyze the effects of these indices on climate extremes, we chose a time period 1950-2008 so as to have the same data length.

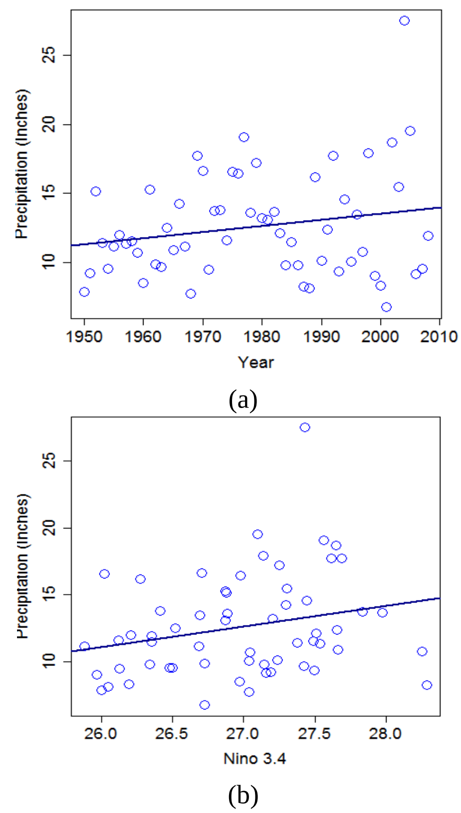
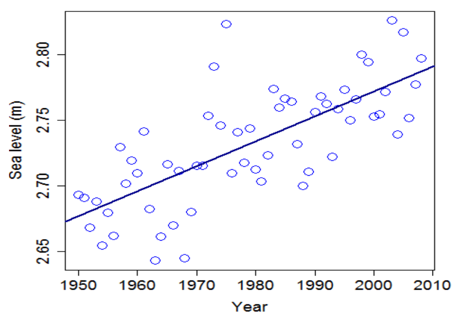


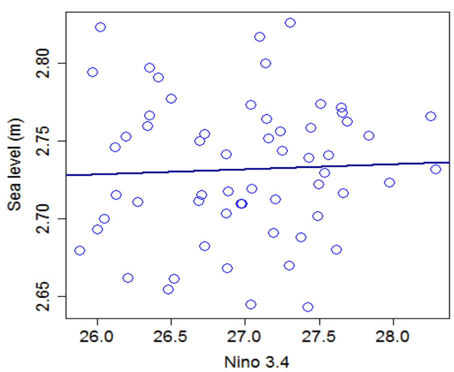
Figure. 1. (a) Time series plot of annual maxima of precipitation and (b) Scatter plot of precipitation and El Nino 3.4 at Pasquotank, North Carolina. Trend is indicated by the solid line.

Simulation Results

We investigated the use of the GEV distribution to model both extreme precipitation and sea level in North Carolina and Florida respectively. We modeled these events using both stationary and non-stationary models for the time period 1950-2008.



(a)



(b)

Figure. 2. (a) Time series plot of mean Sea level and (b) Scatter plot of Sea level and Nino 3.4 at Pensacola, Florida. Trend indicated by the solid line.

The effects of both time and El Nino Southern Oscillation index (Nino 3.4) were taken into account. The results are summarized in Tables 1 and 2. The values in parenthesis are the standard errors of the estimates for the parameters. The minimized negative log-likelihood as well as the BIC values is shown in the tables. From the results, it can be seen from Table 1 that when the ENSO was used as a covariate, the negative log-likelihood was minimum as observed in model 3. However, there was an increase in the BIC value. This increase in the BIC value implies that though its introduction has an effect, the change is not significant as compared to the computation complexity involved. In Table 2, model 2 for the Florida sea level has the most minimized negative log likelihood as well the least BIC value, and hence is selected by the BIC as the best model. This means that the model with the linear trend in time is the most appropriate model to be considered for the sea level data at Pensacola.

Conclusion

The effects of both time and ENSO have been analyzed in this research unlike previous work where only linear trend in time was analyzed. From the results, it is realized from the log likelihood values that the sea level at Pensacola is affected by both time and the ENSO as seen in Model 4. Model 4, which is a combination of both time and the ENSO, has the lowest negative log likelihood as compared to the stationary model in Model 1. Comparing the results of models 2 and 4, it is observed that the combined effect of both time and the ENSO is greater than that of time alone but due to computational complexity the BIC results favor the time dependent model in Model 2. Similar observations can be made of from the precipitation data at Pasquotank. However, for this station, the impact of ENSO seems to be greater than that of time as can be seen from both the negative log likelihood and BIC values. This is also evident from the time series plots shown in Fig.1. Again, due to computational complexity, the BIC results favors the stationary model. This suggests that the effect is not significant enough as compared to the computation complexity involved. Hence, the stationary model may be suitable for the precipitation data according to the BIC values. The work presented here only considered simple forms of nonstationarity, where we only relied on linear models of the covariates. The linear models used might have influenced the less significance of the effects of the covariates. As such, as future work, we will consider nonlinear forms of nonstationarity such as vector generalized additive models (VGAM) or generalized additive models for location, scale and shape (GAMLSS) parameters of the distribution.

Table 1 Parameter estimates and standard errors for the GEV distribution fitted to the annual maxima of precipitation (inches) at Pasquotank, North Carolina (values in brackets are standard errors)

	Model 1	Model 2	Model 3	Model 4
σ	2.7326 (0.310)	2.7200 (0.315)	2.7013 (0.3065)	2.7272 (0.318)
ξ	0.0743 (0.112)	0.0817 (0.1195)	0.0636 (0.1128)	0.0769 (0.121)
μ_0	10.8150 (0.409)	10.9124 (0.6479)	-9.7803 (14.2051)	9.7730 (14.092)
μ_1	-	-0.0036 (0.0190)	0.7654 (0.5281)	-0.0048 (0.019)
μ_2	-	-	-	0.5232
-log *l*	155.062	155.043	153.9863	153.956
BIC	322.3567	326.397	324.2828	328.300

Table 2 Parameter estimates and model selection for the GEV distribution fitted to the time series of mean sea level (mm) at Pensacola Florida (values in brackets are standard errors)

	Model 1	Model 2	Model 3	Model 4
σ	52.3875 (6.576)	50.8995 (5.631)	52.799 (6.552)	49.2521 (5.394)
ξ	0.0658 (0.152)	-0.1265 (0.116)	0.0474 (0.148)	-0.1251 (0.114)
μ_0	2840.55 (8.389)	2793.74 (15.217)	2840.57 (8.324)	2787.8965 (14.806)
μ_1	-	1.8946 (0.506)	4.3535 (7.181)	2.0330 (0.475)
μ_2	-	-	-	13.3157
-log *l*	329.1631	320.8547	328.9922	319.0587
BIC	670.5588	658.0195	674.6331	661.5552

Acknowledgments This work is partially supported by the Expeditions in Computing by the National Science Foundation under Award CCF-1029731.

References

Agana, Norbert, Mohammad Gorji Sefidmazgi, and Abdollah Homaifar. 2014. "Analysis of Extreme Precipitation Events." *Fourth International Workshop on Climate Informatics*.

AghaKouchak, Amir, David Easterling, Kuolin Hsu, Siegfried Schubert, and Soroosh Sorooshian. 2013. 'Extremes in a Changing Climate Detection, Analysis and Uncertainty', Springer.

Cayan, Daniel R., Kelly T. Redmond, and Laurence G. Riddle. 1999. 'ENSO and Hydrologic Extremes in the Western United States*', *Journal of Climate*, 12: 2881-93.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values* (Springer).

El Adlouni, S., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée. 2007. 'Generalized maximum likelihood estimators for the nonstationary generalized extreme value model', *Water Resources Research*, 43: W03410.

Feng, Song, Saralees Nadarajah, and Qi Hu. 2007. 'Modeling Annual Extreme Precipitation in China Using the Generalized Extreme Value Distribution', *Journal of the Meteorological Society of Japan. Ser. II*, 85: 599-613.

Gershunov, Alexander, and Tim P. Barnett. 1998. 'ENSO Influence on Intraseasonal Extreme Rainfall and Temperature Frequencies in the Contiguous United States: Observations and Model Results', *Journal of Climate*, 11: 1575-86.

Gorji Sefidmazgi, Mohammad, Mina Moradi Kordmahalleh, Abdollah Homaifar, and Stefan Liess. 2014. "Change Detection in Linear Trend of Temperature over US 1900-2012." In *Fourth International Workshop on Climate Informatics*.

Gorji Sefidmazgi, Mohammad, Mohammad Sayemuzzaman, and Abdollah Homaifar. 2014. 'Non-stationary Time Series Clustering with Application to Climate Systems.' in Mo Jamshidi, Vladik Kreinovich and Janusz Kacprzyk (eds.), *Advance Trends in Soft Computing* (Springer International Publishing).

Gorji Sefidmazgi, Mohammad, Mohammad Sayemuzzaman, Abdollah Homaifar, MK Jha, and Stefan Liess. 2014. 'Trend analysis using non-stationary time series clustering based on the finite element method', *Nonlinear Processes in Geophysics*, 21: 605-15.

Griffis, Veronica W., and Jery R. Stedinger. 2007

'Incorporating Climate Change and Variability into Bulletin 17B LP3 Model.' in, *World Environmental and Water Resources Congress 2007*.

Hosking, J. R. M. 1985. 'Algorithm AS 215: Maximum-Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34: 301-10.

Jain, Shaleen, and Upmanu Lall. 2001. 'Floods in a changing climate: Does the past represent the future?', *Water Resources Research*, 37: 3193-205.

Katz, Richard W., Marc B. Parlange, and Philippe Naveau. 2002. 'Statistics of extremes in hydrology', *Advances in Water Resources*, 25: 1287-304.

Martins, Eduardo S., and Jery R. Stedinger. 2000. 'Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data', *Water Resources Research*, 36: 737-44.

Milly, P. C. D., Julio Betancourt, Malin Falkenmark, Robert M. Hirsch, Zbigniew W. Kundzewicz, Dennis P. Lettenmaier, and Ronald J. Stouffer. 2008. 'Stationarity Is Dead: Whither Water Management?', *Science*, 319: 573-74.

Panagoulia, D., P. Economou, and C. Caroni. 2014. 'Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change', *Environmetrics*, 25: 29-43.

Ropelewski, C. F., and M. S. Halpert. 1986. 'North American Precipitation and Temperature Patterns Associated with the El Niño/Southern Oscillation (ENSO)', *Monthly Weather Review*, 114: 2352-62.

Shabbar, Amir, Barrie Bonsal, and Madhav Khandekar. 1997. 'Canadian Precipitation Patterns Associated with the Southern Oscillation', *Journal of Climate*, 10: 3016-27.

Vogel, Richard M., Chad Yaindl, and Meghan Walter. 2011. 'Nonstationarity: Flood Magnification and Recurrence Reduction Factors in the United States1', *JAWRA Journal of the American Water Resources Association*, 47: 464-74.

Zhang, Xuebin, Jiafeng Wang, Francis W Zwiers, and Pavel Ya Groisman. 2010. 'The influence of large-scale climate variability on winter maximum daily precipitation over North America', *Journal of Climate*, 23: 2902-15.