J2.6 HYDROLOGIC ASSESSMENT: APPLICATION OF EXTREME VALUE THEORY FOR CLIMATE EXTREMES SCENARIOS CONSTRUCTION

Jeanna Goldstein*, Monirul Mirza, David Etkin and Jennifer Milton
Environment Canada

1. INTRODUCTION

The Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC, 2002) indicates that changes in frequency and intensity of extreme climate events can be expected. Changes in extremes are particularly important for human adaptation issues since they, by definition, occur outside of societies coping range, and therefore have large impacts. In fact, the global economic costs of natural disasters have increased manyfold in the past several decades, highlighting the importance of this topic. According to the IPCC report, it is very likely that daytime maximum and minimum temperatures will increase, heat waves will become more frequent, and that the number of cold waves and frost days in some Canadian regions will decline. Increases in high intensity precipitation events are also likely to occur. Kharin and Zwiers (2000) found distinctly different future changes in extremes of daily maximum and daily minimum temperatures in Canada. They also found non-linear effects, with changes in extreme precipitation being larger than changes in total precipitation. Non-linear effects are also present in the occurrence of temperature extremes, with heat waves becoming increasingly frequent at higher thresholds (Colombo et al., 1999). There have been numerous examples of natural disasters resulting from extreme events in recent decades. Recent major floods in Canada and other regions of the world are a reminder that vulnerability to floods remains high (de Loe, 2000 ; Mirza, 2002). Dore (2002) concluded that floods are the most common Hydrological Disasters (HDs) and are increasing in Canada over time. The distribution of floods by province indicates that a quarter of HDs affect Ontario and Quebec, the most densely populated provinces of Canada. Most notably, the Saguenay flood that occurred in 1996 in Quebec caused extensive property damage and claimed ten lives. Temperature variability and extreme temperature events further demonstrate a lack in adaptive capacities due to infrastructure limits. The heat waves and drought of the summer of 2001 in Ontario and Quebec, and of 2002 in Canadian Prairies sent shock waves to the energy industry, farming and water resources infrastructures and the health care system in terms of system vulnerability and management. So far, most of the climate change impact studies in Canada have been carried out based on changes in means. However, changes in extremes will likely expose both infrastructure and social systems in Canada to greater hazards in the future. At the same time, many trends are increasing the vulnerability of these systems, including increased reliance upon external lifelines, population growth and concentration, and changes in the distribution of wealth (Etkin, 1999). Extreme climate scenarios will be of great importance for impact researchers, for the modellers community and for other stakeholders interested in climate change impacts, and measures that can be taken to adapt to them. Most environmental research studies assume a stationary climate. Extreme Value Theory (EVT) is frequently used in environmental studies (Smith, 2001; Katz et al., 2002) and financial studies (McNeil, 1998) to produce distributions to fit data consisting of maxima or minima in random samples for fixed intervals, as well as to model the distribution of excess over thresholds, and to estimate parameters of arbitrary distributions. The IPCC Workshop on Changes in Extreme Weather and Climate Events (2002) highlighted that there are still few weather and climate studies that use EVT. The purpose of this paper is to conduct a literature survey of existing methods of extreme value analysis used in disciplines such as hydrology, climatology, insurance, finance, engineering and environmental science, to demonstrate that EVT is one of the most advanced theories for this type of investigation, to make an attempt to fill gaps existing in climate research of extremes and to justify a choice of EVT for climate extreme scenarios construction for Canada. The second section presents statistical methods of probability estimations of extreme events. The EVT forms are discussed in the third section. The fourth section is devoted to the selection of EVT form for climate studies**.** Fitting methods and uncertainty evaluation are described in the fifth section. The sixth section discusses methodologies of climate extreme scenarios construction. Conclusions can be found in the seventh section.

2. STATISTICAL METHODS OF PROBABILITY ESTIMATIONS OF EXTREME EVENTS

There are several approaches to simulate the frequency of extreme events, and to reflect stochastic volatility and leptokurtosis of the return distributions, these being (1) parametric, (2) non-parametric, (3) stochastic methods

---

* *Corresponding author address :* Dr. Jeanna Goldstein, Environment Canada Quebec Region, 100 Alexis-Nihon Blvd, 3rd Floor, Saint-Laurent, Quebec, Canada, H4M 2N8 ;
e-mail : Jeanna.Goldstein@ec.gc.ca
http://www.criacc.qc.ca/scenario/cs/climate_scenario.html

and (4) Extreme Value. The parametric method is based upon fitting some particular distribution to a set of observed or simulated returns. This method is well known in climatology as a percentile method or as a return period approach (Jones and Reid, 2001). The drawback of this approach is that usually return period data distributions derived using this approach is not representative for tail estimation. These distributions of extreme returns are far from being asymptotic. An historical or non-parametric approach addresses evaluation of appropriate return period histograms. A quantile approach (Karl and Knight, 1998; Jones and Reid, 2001; Rusticucci and Vargas, 2002) could be an example of historical method in the climatological studies. Non-parametric approach does not take into consideration events beyond sample range and also does not indicate the tail form. Following this method it is very difficult to estimate extreme quantiles. Stochastic methods (Monte Carlo) generate repeated situations that simulate returns based on random traction from some stochastic projections. These approaches assume normality and thus do not accommodate observed fat tails in return data. Monte Carlo techniques could be successfully carried out for data already simulated from EV distribution (Palutikof, 1999). Stochastic simulations of some extreme variables give an indication of climate conditions to be changed to non-stationary (Burlando and Rosso, 2002) thus demanding consideration of the conditional return distribution. The EVT approach is designed specifically for tail estimation, for recognition and modelling leptokurtic distributions, for dealing with non-stationary distribution and for the determination of current volatility. The EVT is able to estimate extreme quantiles for a short record of data. McNeil (1998) considers EVT to be the most honest approach to measure the uncertainty inherent to the problem. The IPCC Workshop on Changes in Extreme Weather and Climate Events (IPCC, 2002) pointed out gaps in extreme weather and climate events investigations. It recommended an EVT like a tail modelling approach, which has many potential advantages over other existing approaches. For example, descriptive indices of the extremes (such as percentiles, growing season length and wet/dry day duration) were addressed at this workshop as measures that do not fully summarise all the important attributes of extremes. The EVT operates with all attributes of extremes including frequency, intensity, volatility and clustering, thus this gap in investigation of the extremes can be fulfilled.

3. FORMS OF EXTREME VALUE THEORY

Extreme Value Theory exists in conventional, modern and intermediate forms. The conventional form was produced as a result of scientific investigations based on the ‘’Three types theorem’’ (Fisher and Tippett, 1928) and studies of Gumbel (1958). These scientists stated and justified that under certain conditions the distribution of the standardized maxima/minima converges to the three limiting distributions (Gumbel, Frechet and Weibull) as the size of the series increases (Gnedenko, 1943). A standard combination of these three basic families is called the Generalized Extreme Value (GEV) distribution. This technique is often referred to as ‘’the method based on limit theorems for block maxima’’, or as ‘’the annual maximum method of return time estimation’’, or as ‘’the annual maxima’’. The modern form of the EVT is known as ‘’threshold’’ form and is based on the Generalized Pareto Distribution (GPD) which is the analogue of the GEV distribution for annual maxima. GPD has proven to be more flexible than annual maximum methods (Smith, 2001) and can deal with asymmetries in the tails (McNail and Frey, 2000). The intermediate form is based on the r-largest order statistics method. The appropriate joint distribution is fitted to the r largest values in each year (r equals 1 is classical GEV method). The IPCC Workshop on Changes in Extreme Weather and Climate Events (2002) stated that application of peaks over threshold (POT) technique is more recommended than the annual maxima method. Katz et al. (2002) considered POT approaches to supply more accurate estimates of the parameters and quantiles of the extremes under the condition of obtaining additional information about the extreme tails. The POT method could be suggested for climate extreme scenarios construction attempting to model current and future meteorological extremes, to derive a natural model for:

- a Poisson process for the occurrence of an exceedance of a high threshold
- a generalized Pareto distribution for the excess over the threshold.

According to Smith (2001) Poisson-GPD probability distribution function of an annual maximum of the process is less than x is

$$\Pr\left\{ \max_{1\le i\le N} Y_i \le x \right\} = \exp\left\{ -\lambda \left(1+\xi\,\frac{x-u}{\sigma}\right)^{-1/\xi} \right\} \qquad (1)$$

where x > u, ξ - shape parameter which determines the nature of the tail of the distribution, λ - intensity, u - threshold, σ - location parameter which depends on u.

4. SELECTION OF EVT FORM

Extreme Value Analysis demands that the sequence of extremes be independent and identically distributed (IID). Extraction of the dependent extremes is determined by a strong serial correlation in daily variables. Kharin and Zwiers (2000), and Smith (2001) referring to Leadbetter (1983) indicate that extremes of the non-IID process in the case of series with large numbers (Gnedenko, 1943) could be properly evaluated by a GEV distribution. McNeil and Saladin (1998), Smith (2001), Katz et al. (2002) suggest that Poisson-GPD model could be used for the processes consisting of IID random values. The POT-method is suggested to be used for dependent processes (McNeil and Saladin, 1998; Smith, 2001). The POT method is also called the partial duration serious (PDS) method. Independent series are generated separating picks within the clusters, thus becoming applicable for a Poisson-GPD technique. The parent distribution is effectively censored by the POT methods implying a low limit on the selection of extremes. Applications of the EVT forms are restricted by the length of a data set, and serial daily correlation of a meteorological parameter i.e. the size of the tail and time dependency. Smith (2001) recommended treating non-IID data as follows: remove seasonal trend from the data before simulating Pareto distribution function; subdivide the year into homogeneous seasons and apply the Poisson-GPD model separately within each; expand the Poisson-GPD model to include covariates. The IPCC Workshop on Changes in Extreme Weather and Climate Events (2002) recommended the use of the POT method rather than annual maxima.

5. FITTING METHODS AND UNCERTAINTY EVALUATION

There are several methods used to evaluate parameters of the applied distribution in order to estimate how well a parametric model fits the data. The most often used techniques are: Maximum Likelihood (ML), Bayesian, L-moments and graphical. A choice of the parameter estimation technique depends on the EVT form applied for the investigation. For example, application of the Poisson-GPD model demands the use of the ML or Bayesian methods for meteorological and hydrological studies (El-Jabi, 1998; Smith, 2001) and supplies more information about the presence of a heavy tail than used with block maxima model (Katz et al., 2002). Smith (2001) advocates ML and Bayessian methods for the series of data generated by GCMs and RCMs. The ML method is recommended for application to provide estimations of conditional volatility (McNail, 2000) and could be used in the presence of covariates (Katz et al. , 2002). L-moment theory offers a parameter estimation tool used in recent environmental sciences and preferably applied when dealing with small samples sizes (Kharin and Zwiers, 2000). The disadvantage of this method is that it does not readily incorporate covariates. (Katz et al, 2002). L-moments technique shall be recommended for parameter estimation along with utilization of the block maxima method (Kysely, 2002). Graphical techniques include examination of the quantile-quantile (Q-Q) plot or probability plot correlation coefficient (PPCC). Booij (2002) referred to PPCC method as a simple and powerful method. The Q-Q method is widely used to explore data and to carry on fitness tests.

The Likelihood function that corresponds to the Poisson-GPD (Smith, 2000) for N exceedances $Y_1, Y_2, ..., Y_N$ over T-year period is

$$L_{N,Y}(\lambda,\sigma,\xi) = N\log\lambda - \lambda\, T - N\log\sigma - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{N}\log\left(1+\xi\frac{Y_i}{\sigma}\right) \qquad (2)$$

provided $1+\xi\, Y_i/\sigma \succ 0$ for all i. Smith (2000) noted some practical constraints for the maximization of $L_y(\lambda,\sigma,\xi)$ and gave detailed instructions regarding specific applications of numerical maximum likelihood estimation of the Poisson-GPD parameters. For example, for ξ < 0, as $\sigma \rightarrow -\xi\, Y_{\max}$ there is a singularity in the likelihood and $L_y(\lambda,\sigma,\xi) \rightarrow \infty$. The correct procedure is to ignore the singularity and to find the local maximum.

The choice of the statistical technique to estimate parameter distribution and quantiles of the extreme events is a source of uncertainty in a cascade of uncertainties related to extreme climate scenario construction. Uncertainty estimation could be defined by the standard error (Hoskings and Wallis, 1987). Extreme value analysis is characterised as successfully obtaining small standard errors. Another way to estimate precision is by introducing the Monte Carlo method or ''bootstrap'' technique (Palutikof, 1999; Kharin and Zwiers, 2000). Danielson and de Vries (1997a) suggest the use of a semiparametric approach based on the Hill-estimator for the tail index. They continue their research introducing a two step sub-sample bootstrap method to estimate first and second order return distribution parameters (Danielsson et al., 2001).

6. CLIMATE EXTREME SCENARIOS CONSTRUCTION METHOD

Climate Extreme Scenarios (CES) construction is based on the fitting of observations and climate models output to the POT model with a Poisson point process approach, on extending POT to the trend models, and on considering all possible types of the trend models. Frequency scenarios are defined according to homogeneous and non-homogeneous Poisson processes with constant and time-dependent intensity for different trend models. Stress scenarios are developed by considering GPD and stressed GPD models with and without trend models. The full set of the scenarios represents a combination of frequency and stress scenarios (McNeil and Saladin, 2000). There is a tendency to consider a climate model grid box simulated variable value as representative of an areal quantity (Hoff, 2001; Booij, 2002). The use of reduction methodology (Sivapalan and Bloschl, 1998; Booij, 2002) should be taken into consideration when constructing CES from the climate model data in order to transform an areal quantity to local point value. Skelly and Henderson-Sellers (1996) suggested a grid-box approach for subgrid-scale interpolation of GCMs data. A homogeneous data set of returns is needed for a valid data analysis. Data homogeneity increases if a threshold value is not sufficiently high, but the data set itself becomes shorter, which implies large standard errors. This is a reason why thresholds should be chosen very thoroughly. The advanced form of EVT that describes the behaviour of extreme events for stochastic processes evolving dynamically in time and space should be applied to model multivariate extremes. Embrechts et al (2001) discussed joint distribution and the use of copulas for modelling multivariate processes that are consistent with prespecified marginal distributions and correlations. An attempt to model climate and weather extreme multivariate events can be found in Embrechts et al (2000).

7. CONCLUSION

A literature survey shows Extreme Value Theory (EVT) to be a reliable tool for climate extreme scenarios construction. A peaks over threshold (POT) method with Generalized Pareto (GP) - distributed peak values - is a preferred technique to investigate the behaviour of upper quantile values and to assess the uncertainty of these estimates. A maximum likelihood method is recommended for the evaluation of GP distribution parameters that depend upon covariates. Thus, construction of climate extreme scenarios for the future, which is consistent with observed trends and climate models, is based upon sensible limiting modelling and assesses uncertainty applying extreme value theory. Suggested approaches include the investigation of the most important attributes of extreme events: frequency, intensity, volatility and clustering.

8. REFERENCES

Beersma, J., M. Agnew, D. Viner, and M. Hulme (eds), 2000 : Climate scenarios for water-related and coastal impacts. *Eclat-2 Workshop Report N.3 KNMI, the Netherlands*, 10-12 May 2000, CRU, Norwich, UK, 144pp.

Bensalah,Y, 2000 : Steps in applying extreme value theory to finance : a review. *Bank of Canada Working Paper* 2000-20.

Booij, M.J., 2002. Extreme Daily Precipitation in Western Europe with Climate Change at Appropriate Spatial Scale, *International Journal of Climatology*, **22**, 69-85.

Burlando, P., and R. Rosso, 2002 : Effects of transient climate change on basin hydrology. 1. Precipitation scenarios for the Arno River, central Italy. *Hydrol. Process.*, **16**, 1152-1175.

Colombo, A.F., D. Etkin, and B.W. Karney, 1999 : Climate variability and the frequency of extreme temperature events for nine sites across Canada: Implications for power usage. *Journal of Climate,* **12**, 2490-2502.

Danielsson, J., and C. de Vries, 1997a : Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, **4**, 241-257.

Danielsson, J., L. de Haan, L. Peng, and C. de Vries, 2001 : Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, **76**, 226-248.

Dore, M.H.I., 2002. Forecasting the Conditional Probabilities of Natural disasters in Canada as a Guide for Disaster Preparedness, *Natural Hazards* (in press).

El-Jabi, N., F.Ashkar, and S. Hebabi, 1998 : Regionalization of floods in New Brunswick (Canada). *Stochastic Hydrology and Hydraulics*, **12(1)**, 65-82.

Embrechts, P, L. de Haan, and X. Huang, 2000 : Modelling Multivariate Extremes. In: Extremes and Integrated Risk Management, P. Embrechts (Editor), *Risk Books, Risk Waters Group*, London, 59-67.

Embrechts, P., A. McNeil, D. Straumann, D., 2001: Correlation and dependence in risk. Properties and Pitfalls. *To appear in Risk Management: Value at Risk and Beyond*, M. Dempster (Ed.), Cambridge University Press.

Etkin, D.A., 1999 : Risk transference and related trends: driving forces towards more mega-disasters. Environmental Hazards, **1**, 69-75.

Fisher, R.A., and L.H.C. Tippett, 1928 : Limiting form of the frequency distributions of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.*, **24,** 180-190.

Frey, R and A.J. McNeil, 2001 : VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights *(preprint).*

Gambel, E.J., 1958 : Statistics of extremes. Columbia University Press, New York, 375pp.

Gnedenko, B.V., 1943 : Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.

Hoff,C., 2001 : Spatial and temporal persistence of mean monthly temperature on two GCM grid cells. *Int. J. Climatol.*, **21**, 731-744.

Hoskings, J.R.M. and J.R. Wallis, 1987 : Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339-349.

IPCC, 2001 : Climate Change 2001 : The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., Y. Ding, D.J.

Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.*)]. Cambridge University Press, Cambridge*, United Kingdom and New York, NY, USA, 881 pp.

IPCC, 2001 : Climate Change 2001 : Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change [McCarthy, J.J., O.F. Canziani, N.A. Leary, D.J. Dokken and K.S.White (eds.)]. *Cambridge University Press, Cambridge*, United Kingdom and New York, NY, USA, 1032.

IPCC Workshop on Changes in Extreme Weather and Climate Events, workshop report, Beijing, China, 11-13 June, 2002, pp. 107.

Jones, P.D., and P.A. Reid, 2001 : Assessing future changes in extreme precipitation over Britain using regional climate model integrations. *Int. J. Climatol*., **21**, 1337-1356.

Karl, T.R., and R.W. Knight, 1998 : Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc*., **79**, 231-241

Katz, R.W., M.B. Parlange, and P. Naveau**,** 2002: Statistics of extremes in hydrology. Advances in Water Resources *(in press).*

Katz, R.W., 2002: "Techniques for estimating uncertainty in climate change scenarios and impact studies." *Climate Research*, **20**, 167-185.

Kharin V.V. and Zwiers F.W., 2000 : Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *Journal of Climate*, **13**, 3760-3788.

Kysely,J., 2002: Probability estimates of extreme temperature events: stochastic modelling approach vs Extreme value distributions. *Stud. Geophys. Geod*., **46**, 93-112.

Leadbetter, M.R., G. Lindgren, and H. Rootzen, 1983 : Extremes And related prppeties of random sequences and processes. Springer Verlag, 336 pp.

de Loe, R. , 2000. Floodplain Management in Canada: Overview and Prospects, *Canadian Geographer*, **44**, 355-368.

McNeil, A.J., 1998 : Calculating quantile risk measures for financial return series using extreme value theory. Mimeo. *ETH Zentrum*, Zurich.

McNeil AJ and Frey R, 2000: Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach . *Journal of Empirical Finance*, **7,** 271-300.

McNeil AJ and Saladin T, 2000: Developing scenarios for future extreme losses using the POT method. In Extremes and Integrated Risk Management , edited by Embrechts PME, *published by RISK books*, London.

Mirza, M. M. Q., 2002 : Global warming and changes in the probability of occurrence of floods in Bangladesh and implications. Global Environmental Change, **12**, 127-138.

Osborn, T., 1997 : Areal and point precipitation intensity changes : Implications for the application of climate models. *Geophysical research letters*, **24**, No 22, 2829-2832.

Palutikof, J.P., B.B. Brabson, D.H. Lister, and S.T. Adcock, 1999 : A review of methods to calculate extreme wind speeds. *Meteor. Appl*., **6**, 119-132.

Rusticucci, M, and W. Vargas, 2002 : Cold and warm events over argentina and their relationship with the ENSO phases : risk evaluation analysis. *Int. J. Climatol*, **22**, 467-483.

Skelly, W., and A. Henderson-Sellers, 1996 : Grid box or grid point : what type of data do GCMs deliver to climate impact researchers? *Int. J. Climatol*,., **16**, 1079-1086.

Sivapalan M., and G. Bloschl, 1998 : Transformation of point rainfall to areal rainfall : intensity-duration-frequency curves. *Journal of Hydrology*, **204**, 150-167.

Smith, R.L., 2001 : Extreme value statistics in meteorology and environment. Environmental Statistics, Chapter 8, pp.300-357. [available at : *http ://www.stat.unc.edu/postscript/rs/envstat/env.html*].

Voss, R., W. May, and E. Roeckner, 2002 : Enhanced resolution modelling study on anthropogenic climate change : changes in extremes of the hydrological cycle. *Int. J. Climatol*,., **22**, 755-777.