QUANTILE ESTIMATION IN A NON-STATIONARY CONTEXT: A BAYESIAN GEV-B-Spline APPROACH

Bouchra Nasri (1), Salaheddine El Adlouni (2), Taha .B.M.J Ouarda (1, 3)

(1) Canada Research Chair on the estimation of Hydrometeorological variables, INRS-ETE.

(2) Université de Moncton, Département de Mathématique et de Statistique.

(3) Masdar Institue of Science and Technology.

Introduction

Many fields of modern science and engineering have to deal with rare events with significant consequences. Extreme value theory (EVT) allows to provide the basis for the statistical modeling of such extremes. The EVT shows that the maxima, Independent and Identically Distributed (iid), is asymptotically Generalized Extreme Value (GEV) distributed (Jenkinson, 1955). In practice, the hypotheses of the EVT are generally not fulfilled, and a classical frequency analysis of independent, homogeneous and stationary samples, is considered with a large range of probability distributions to estimate the occurrence of extreme events

In the context of the above statement, the effect of a covariate can be considered in a polynomial form (e.g. Coles 2001, El Adlouni et al. 2007) for nonstationarity analysis of extreme events. These polynomial forms for estimating the GEV parameters were developed by the introduction of covariates in polynomial forms such as linear or quadratic function. However, the dependence between covariate and variable of interest can take different dependence structures.

The main objective of the present study is to develop the Generalized Extreme Value model with covariates where the dependence structure is represented by a B-spline in Bayesian framework. Prior distributions have been proposed and the posterior distribution is simulated through the Metropolis Hasting (MH) algorithm (Metropolis et al., 1953 and Hastings, 1970).

The proposed GEV-B-Spline model is considered to estimate the return period of the maximum annual precipitation (MAR) at Randsburg station, California, where the Southern Oscillation Index (SOI) and Pacific Decadal Oscillation (PDO) index are considered as the covariates.

Methods

The extreme value theory introduced by Fisher and Tippett [1928], shows that the limiting distribution of the maximum is one of the distributions: Gumbel, Frechet or Weibull. These three distributions can be grouped in a single Generalized Extreme Value (GEV) distribution:

$$F(y,\mu,\sigma,\xi)=\exp\left[-\left(1+\xi\left(\frac{y-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right] \qquad \xi\neq 0$$

$$F(y,\mu,\sigma,\xi)=\exp\left[-\exp\left(-\frac{y-\mu}{\sigma}\right)\right] \qquad \xi=0$$

So the quantile of the GEV distribution is:

$$y_T=F^{-1}\left(1-\frac{1}{E\ t}\right)=\mu-\frac{\sigma}{\xi}\left[1-\left(\log\left(1-\frac{1}{T}\right)\right)^{-\xi}\right]$$

In the non-stationary case, the parameters of the GEV are functions of time or other covariates. So the quantiles y_T depend on time or covariates. In the present study, the parameters σ and ξ are supposed constants. Let Y a random variable that follows the GEV(μ_x,σ, ξ) and X=(X_1, X_2,.., X_p) a vector of covariates. The location parameter of GEV model is a function of covariates :

$$\mu_{\underline{x}}=\sum_{i=1}^{p} f_i(X_i)=f_1(X_1)+f_2(X_2)+\cdots+f_p(X_p)$$

The fi can be decomposed in the form of basic spline functions.

$$f_i\ \ x_i\ \ =\beta_0+\sum_{j=1}^{m}\beta_j B_{j,d}\ \ x_i$$

$B_{j,d}\ x$ is a polynomial of degree *d* on each interval, *m* is the number of control points.

The GEV-B-Spline is considered in a fully Bayesian framework. For a given parameter prior distribution,π(θ) , the Bayes theorem allows to define the posterior distribution :

$$f\ \ \theta\,|\,y\ \ =\frac{f\ \ y\,|\,\theta\ \ *\pi\ \ \theta}{f\ \ y}$$

where

$$\theta=\ \mu_t,\sigma,\xi\ \ =\ \boldsymbol{\beta}_0,\boldsymbol{\beta},\sigma,\xi$$

The prior distribution are:

$$\xi\ \sim\ beta\left(6,9\right)$$

and

$$\left(\boldsymbol{\beta}_0,\boldsymbol{\beta}\right)\sim\ N\ \left(0,\sigma_\beta^2\right)$$

Results & Conclusion

For model development, the following function is first fitted:

GEV-B-Spline :

$$MAR\sim GEV\left(f_1(SOI)+f_2(PDO),\sigma,\xi\right)$$

f1 and f2 are independent spline functions, for which the degree and the number of nodes should be determined. In this application both the number of nodes and the degree of the B-spline functions are fixed to 3.

Figures 1 and 2 show the estimated 2, 20 and 50-year return period maximum rainfall quantiles as function of the covariates (SOI and PDO).

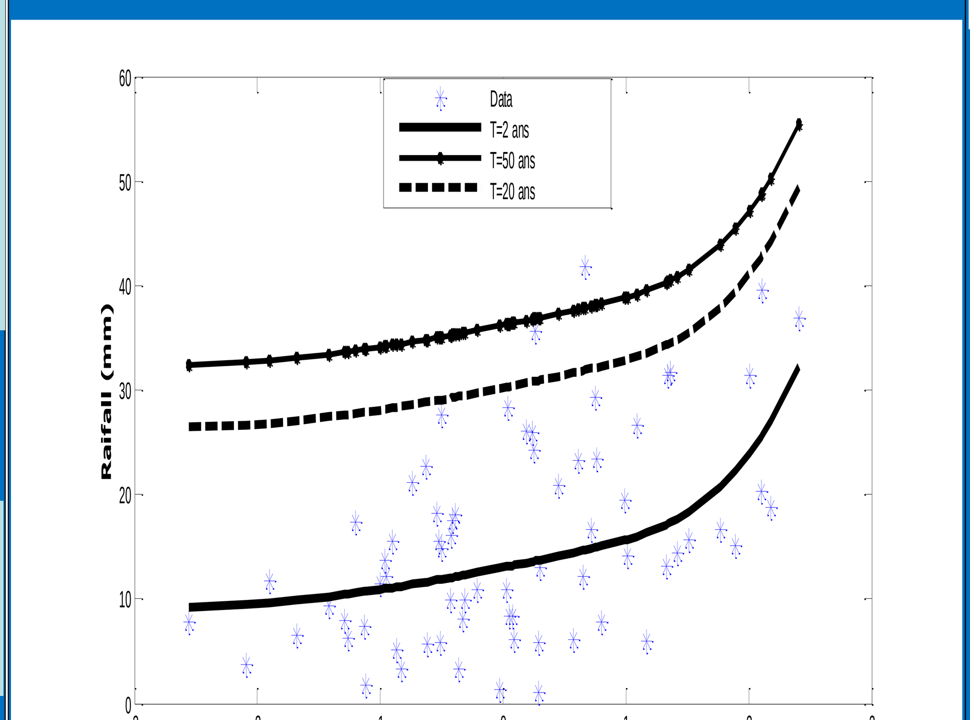

Figure 1

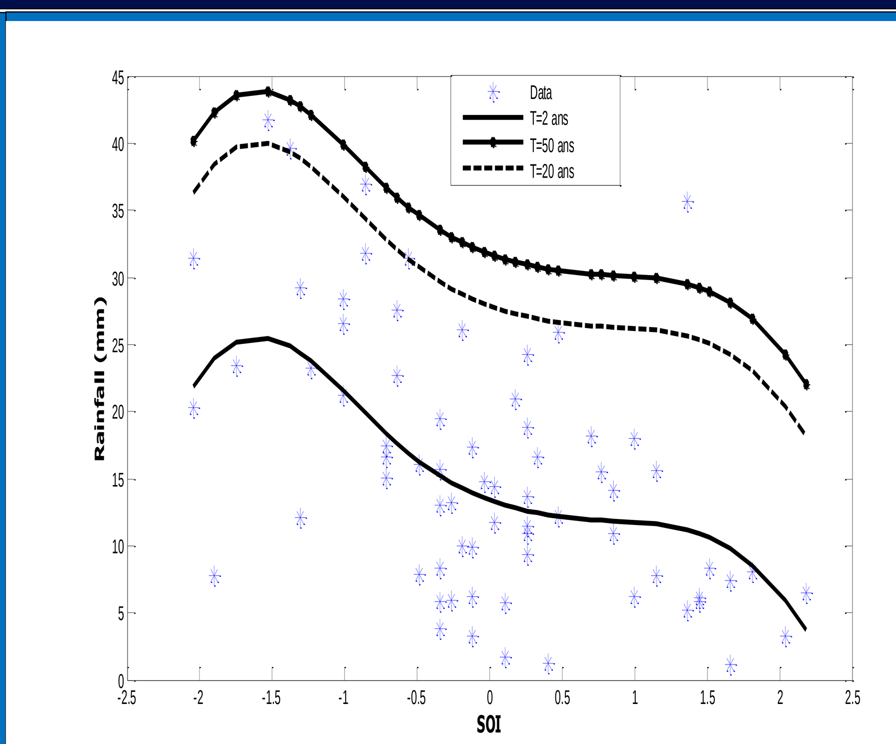

Figure 2

It can be seen that, generally, the SOI has a negative correlation with precipitation, while the PDO index is positively correlated with precipitation. The negative values of SOI (e.g. El Nino phase) and positive values of PDO (Warm Phase of PDO) coincide with the relatively high MAR observations. In the case of PDO, the change in MAR quantiles is very slightly increasing by increasing in PDO and then increases exponentially for PDO values greater than 1. On the other hand, different inflexion points, in the co-movement of SOI and MAR are observed (for example at SOI= -1.5, SOI=0 and SOI= 1.5), indicating more complex relationship between SOI and MAR that between PDO and MAR.

We propose a comparison between the Bayesian parameter estimation for GEV-B-Spline model (BAYES) and other estimation methods such as the conventional method of moments (MM) and the method of maximum likelihood (ML). The comparison is carried out using the bias and the root mean square error (RMSE) of quantile estimations at non-exceedance probabilities, p = 0.5, 0.8, 0.9, 0.99 corresponding to return periods of 2, 5, 10, 100. The results are given in Table 1.

Table 1

Probability	BIAS			RMSE		
	BAYES	MM	ML	BAYES	MM	ML
0.5	0,020	-0,075	0,052	0,403	0,715	0,435
0.8	-0,060	-0,094	0,090	0,418	0,901	0,514
0.9	-0,148	0,249	-0,177	0,450	1,847	1,525
0.99	-0,182	-0,655	-0,448	0,826	3,128	2,879

Results show that the Bayesian estimation for the GEV-B-Spline model in all cases represents the best results. It is interesting to note that for the case of low return periods, i.e. T = 2, T = 5, T = 10 year, the maximum likelihood gives almost comparable results with the Bayesian estimation. However, the error of ML method is rapidly increasing by the increase in the return periods and the method becomes increasingly less efficient.

Therefore, the Bayesian method is a superior method for estimating the extreme rainfall quantiles for all return periods. In addition, the Bayesian method offers a general approach to combine observed and subjective information and the possibility to estimate the entire predictive distribution of the parameters and the quantiles.

Results show that the Bayesian estimation for the GEV-B-Spline model in all cases represents the best results. It is interesting to note that for the case of low return periods, i.e. T = 2, T = 5, T = 10 year, the maximum likelihood gives almost comparable results with the Bayesian estimation.

However, the error of ML method is rapidly increasing by the increase in the return periods and the method becomes increasingly less efficient. Therefore, the Bayesian method is a superior method for estimating the extreme rainfall quantiles for all return periods. In addition, the Bayesian method offers a general approach to combine observed and subjective information and the possibility to estimate the entire predictive distribution of the parameters and the quantiles.

The results indicate a better performance of the proposed Bayesian method for rainfall quantile estimation according to BIAS and RMSE criteria than the classical method of moments and maximum likelihood. It is also observed that the proposed Bayesian and maximum likelihood methods do not show a sharp difference of performance for a low-return period maximum rainfall while the BIAS and RMSE of maximum likelihood are relatively higher than those for GEV-B-Spline method for high-return period rainfall.

References

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

El Adlouni, S., Ouarda ,T.B.M.J., Zhang, X., Roy, R. and Bobee, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E.(1953) Equation of state calculations by fast computing machines.