J. Comp. Int. Sci. (2016) 7(2):25-29
http://epacis.net/jcis/PDF_JCIS/JCIS-0107.pdf
jcis@epacis.net

Pan-american Associacion of Computational Interdisciplinary Sciences

Computational statistical analysis of estimated river discharge with Generalized Pareto Distribution

Henrique R. A. Freitas[a1],
[a]The Graduate Program in Applied Computing (CAP),
National Institute for Space Research (INPE),C.P. 515, S. J. dos Campos - SP, 12001-970, Brazil.

Received on July 01, 2016 / accepted on August 31, 2016

Abstract

Time series analysis is important for the detection of extreme events that occur in processes of nature. This work considers the analysis of estimated river discharge data generated from a hydrological model by comparing the properly rescaled data of the time series to a function of the Generalized Pareto Distribution (GPD) with particular parameters estimated from a GPD algorithm. Results indicate that the patterns of the rescaled data are well represented by the GPD function, and also that positive shape parameter values correspond to the presence of extreme events.

Keywords: *Time series analysis, river dynamics, generalized Pareto distribution, computational statistics.*

1. Introduction

Several processes of nature are analyzed from observations made by considering time as a discrete variable. These observations constitute a time series that includes measures of some information of the process that is being studied. In addition, the time series can be used for the detection of patterns in order to improve the understanding of how the information is distributed.

In hydrology, the representation of the water flow on rivers for the calculation of discharge estimates is fundamental to analyze extreme events such as floods and droughts [Fan et al. 2014]. In this work, time series of discharge data are generated from a hydrological model called MGB [Collischonn et al. 2007] that represents the hydrological processes in large-scale watersheds.

After a rescaling procedure, the time series of estimated river discharges are compared to a statistical distribution, in this case, the Generalized Pareto Distribution (GPD). Results indicate that this distribution provides a well-suited representation of the rescaled information of the time series.

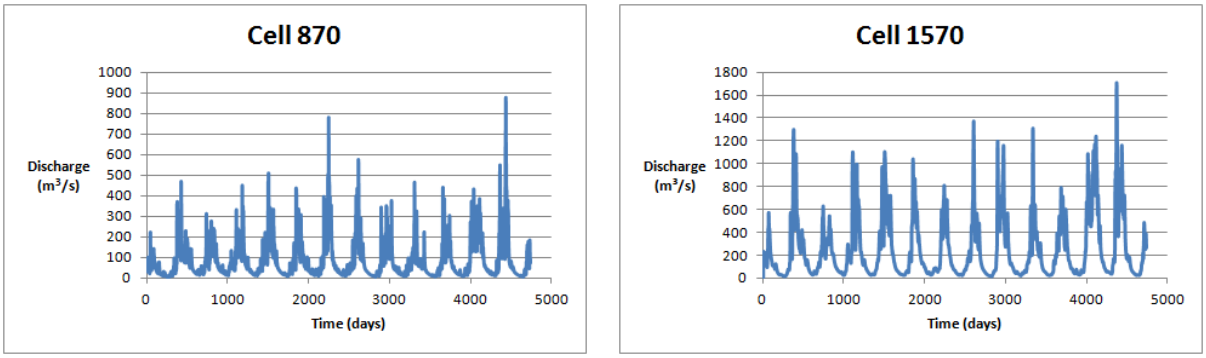
The paper is organized as follows. The next section describes the main equations of the MGB hydrological model used for the calculation of the discharge data. Section 3 briefly presents the rescaling procedure of the time series information, and the GPD formulation used. In section 4, the rescaled discharge data is compared to the GPD showing that the patterns of discharge are well represented by the GPD with particular parameters. The conclusions are given in section 5. A brief description of the GPD algorithm is given in the Appendix.

2. MGB Model

The MGB model (Modelo de Grandes Bacias, in Portuguese) is developed at the IPH-UFRGS [IPH 2015] for applications that include analysis of extreme events,forecast of discharge, effects of climate change,

[1]E-mail Corresponding Author: henrique.renno@inpe.br

quality of water, and many others. The model considers three different spatial units, namely, catchments (or cells), subbasins, and hydrological response units (HRU). The model simulates the conversion of rainfall to runoff, and also other processes of the hydrological cycle such as evapotranspiration estimates, interception by vegetation etc [Paiva et al. 2011]. The main equations of the model are comprised of a simplification of the Saint-Venant equations (inertial equations) [Fan et al. 2014] that are solved numerically for the propagation of the water flow on rivers. Figure 1 exhibits examples of time series of discharge data generated from the MGB model.



(a) (b)

Figure 1 - Time series of discharge data for (a) cell 870, and (b) cell 1570.

The inertial equations used in the MGB model are given below, where h is water depth, q is discharge, y is water level, g is the acceleration of gravity, and n is the Manning coefficient:

$$\frac{\partial h}{\partial t} + \frac{\partial q}{\partial x} = 0$$

$$\frac{\partial q}{\partial t} + gh\frac{\partial y}{\partial x} + gh\frac{|q|.q.n^2}{h^{7/3}} = 0$$

3. Rescaling procedure and Generalized Pareto Distribution

The first step for the analysis of each time series is to randomly shuffle the data in order to generate the surrogate time series. Then a rescaling procedure is applied, which consists of modifying the data according to the formulation $(x - \bar{x})/s$, where $\bar{x}$ and s are, respectively, the mean and the standard deviation of each time series [Ramos et al. 2006].

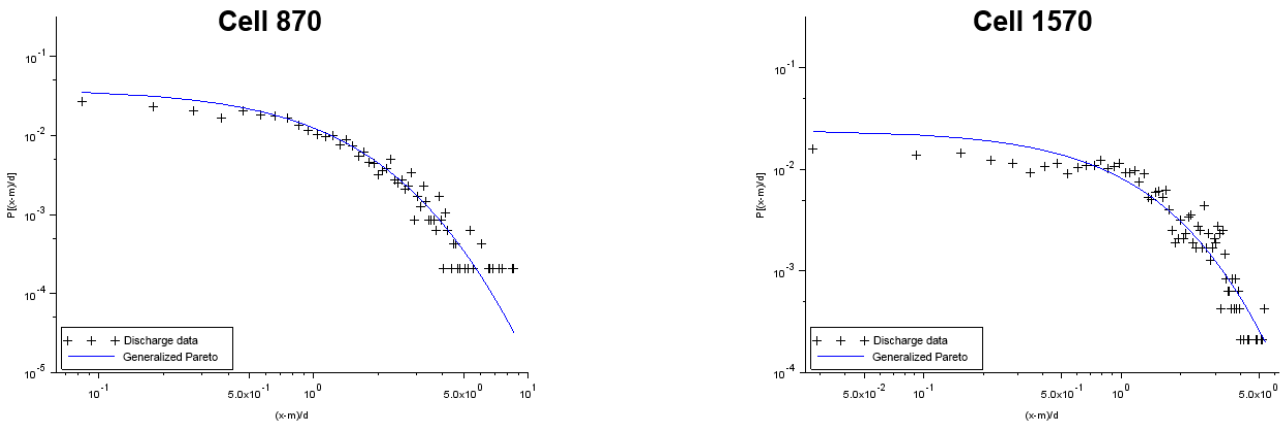
Many formulas are related to the Generalized Pareto Distribution and this work considers the use of the *complementary cumulative distribution function* (ccdf) given by

$$ccdf(z) = (1 + \xi z)^{-1/\xi} \tag{1}$$

where $z = (x - \mu)/\sigma$, for parameters of location (μ), scale (σ), and shape (ξ).

4. Results

The time series of discharge data generated from the MGB model, after being properly rescaled, are compared to the ccdf of the GPD in a log-log plot. Figure 2 shows the empirical histogram log-log plots obtained by considering 100 sample intervals for all the rescaled measurements of each time series illustrated in Figure 1. The ccdf of the GPD for particular parameter values is also present in the plot, showing that it is a suitable representation of the pattern described by the rescaled time series information.



(a) (b)

Figure 2 - Log-log plots of rescaled time series and GPD with parameters values (a) $\mu = -2.3$, $\sigma = 0.6$, $\xi = 0.1$, and (b) $\mu = -2.7$, $\sigma = 0.6$, $\xi = 0.1$.

It is worth mentioning that the parameters of the GPD can be chosen from an optimization algorithm that calculates the curve that better fits to the set of data of the time series. Another important observation is that positive shape parameters ($\epsilon = 1$) corresponds to a heavy-tailed distribution with the presence of extreme events [Ramos et al. 2006].

5. Concluding remarks

This work considered the time series analysis of estimated river discharge data generated from a hydrological model by comparing the data to a function of the Generalized Pareto Distribution with particular parameters. The patterns of the rescaled time series mostly agreed to the curve defined by the GPD function, indicating that this distribution is well-suited to represent the information of the discharge data that constitute the time series. Moreover, positive shape parameter values correspond to the presence of extreme events. As future work, one intends to use a particular algorithm for the optimization of the GPD parameters to find the curve that fits best to the data of the time series.

Appendix

The GPD distribution functions are:
dgpd: Density of the GPD Distribution,
pgpd: Probability function of the GPD Distribution,
qgpd: Quantile function of the GPD Distribution,
rgpd: random variates from the GEV distribution,
gpdMoments: computes true mean and variance,
gpdSlider: displays density or rvs from a GPD.

Usage
$dgpd(x, xi = 1, mu = 0, beta = 1, log = FALSE)$
$pgpd(q, xi = 1, mu = 0, beta = 1, lower.tail = TRUE)$
$qgpd(p, xi = 1, mu = 0, beta = 1, lower.tail = TRUE)$
$rgpd(n, xi = 1, mu = 0, beta = 1)$

Acknowledgments: *This work was supported by CNPq-Brazil. The authors thank R.R.Rosa for discussion on statistical analysis technique introduced in this paper.*

References

[1] Collischonn, W., Allasia, D., Silva, B. C., Tucci, C. E. M. (2007). The MGB-IPH model for large-scale rainfall-runoff modelling. Hydrological Sciences, v. 52, n. 5, p. 878 895.

[2] Fan, F. M., Pontes, P. R. M., Paiva, R. C. D. (2014). Avaliao de um mtodo de propagao de cheias em rios com aproximao inercial das equaes de SaintVenant. Revista Brasileira de Recursos Hdricos, v. 19, n. 4, p. 137147.

[3] IPH. (2015). Modelo de grandes bacias, Fortran 90 source code. Instituto de Pesquisas Hidrulicas. Available from ¡https://www.ufrgs.br/hge/mgb-iph¿.

[4] Paiva, R. C. D., Collischonn, W., Tucci, C. E. M. (2011). Large scale hydrologic and hydrodynamic modeling using limited data and a GIS based approach. Journal of Hydrology, v. 406, p. 170181.

[5] Ramos, F. M., Lima, I. B. T., Rosa, R. R., Mazzi, E. A., Carvalho, J. C., Rasera, M. F. F. L., Ometto, J. P. H. B., Assireu, A. T., Stech, J. L. (2006). Extreme event dynamics in methane ebullition fluxes from tropical reservoirs. Geophysical Research Letters, v. 33, p. 14.