

L-moments based assessment of a mixture model for frequency analysis of rainfall extremes

V. Tartaglia, E. Caporali, E. Cavigli, and A. Moro

Dipartimento di Ingegneria Civile, Università degli Studi di Firenze, Italy

Received: 2 December 2004 – Revised: 2 October 2005 – Accepted: 3 October 2005 – Published: 23 January 2006

Abstract. In the framework of the regional analysis of hydrological extreme events, a probabilistic mixture model, given by a convex combination of two Gumbel distributions, is proposed for rainfall extremes modelling. The papers concerns with statistical methodology for parameter estimation. With this aim, theoretical expressions of the L-moments of the mixture model are here defined. The expressions have been tested on the time series of the annual maximum height of daily rainfall recorded in Tuscany (central Italy).

1 Introduction

In hydrology the most frequent use of statistics has been that of frequency analysis (Haan, 1994). This applies in particular when the magnitude of an event with a given return period or the rarity of a hydrological event, has to be estimated. Often the common use of the asymptotic extreme value type I, or Gumbel, distribution does not represent an acceptable fitting of the model to the observations of rare frequency, i.e. it is not able to model the upper tail of the hydrological extreme events. This limit has been overcome, for example, by the Two Components Extreme Value (TCEV) distribution (Rossi et al., 1984). Observing the presence of outliers data, the TCEV shows, in fact, a better fitting to the data sample skewness (Matalas et al., 1975), but it needs of a larger sample in order to obtain a robust estimation of the parameters. For this reason such kind of distributions are often used on regional basis. The assumption that a region is a set of gauging sites whose hydrologic events frequency behaviour is homogeneous, in some quantifiable manner (Cunnane, 1988), allows to estimate hydrologic variables quintiles even in ungauged sites. In Italy a regionalized procedure have been developed in the VaPI (Valutazione delle Piene in Italia) project, based on the TCEV distribution (Rossi et al., 1992). In the framework of the regional analysis of rainfall

events in Tuscany (central Italy), a statistical methodology for parameter estimation of a probabilistic mixture model is here developed. The use of a probabilistic mixture model in the regional analysis of rainfall extreme events, comes from the hypothesis that rainfall phenomenon arises from two independent populations (Becchi et al., 2000; Caporali et al., 2002). The first population describes the basic components corresponding to the more frequent events of low magnitude, controlled by the local morpho-climatic conditions. The second population characterizes the outlier component of more intense and rare events, which is here considered due to the aggregation at the large scale of meteorological phenomena. As the rare component is due to the large scale state of energy and to the morpho-climatic conditions, it is assumed that any regionalization procedure can be reduced to the use of parameters of the basic or ordinary component.

The contribution of the basic and the outlier components is presented as a mixture model, i.e. a convex combination of two independent populations having Gumbel distribution. Linearity of “sum” – operator leaves separated the contributions of the two populations, and it allows to evaluate separately their statistical parameters (Becchi et al., 2000).

Being very suitable for frequency analysis of extreme values, the L-moments method is proposed for parameter estimation (Hosking and Wallis, 1993). The L-moments are in fact expectations of linear combinations of order statistics (Hosking, 1990) and are more robust to the data outliers and virtually unbiased for small samples. Moreover the L-moments have the very important advantage, over the conventional moments, of being less affected from the effects of sampling variability being linear functions of the data.

In the following the probabilistic model and the parameter estimation method are described. A procedure to find an estimation of the outlier percentage is also investigated.

2 Model development and parameter estimation

The probability model is defined by the following convex combination or mixture of two Gumbel distributions.

$$F_X(x) = (1 - \alpha) \cdot \exp\left[-\Lambda_1 \exp\left(-\frac{x}{\theta_1}\right)\right] + \alpha \cdot \exp\left[-\Lambda_2 \exp\left(-\frac{x}{\theta_2}\right)\right] \tag{1}$$

with : $\Lambda_1 = \exp\left(-\frac{\mu_1}{\theta_1}\right)$ and $\Lambda_2 = \exp\left(-\frac{\mu_2}{\theta_2}\right)$ (2)

where μ_1, μ_2 are the modal values and θ_1, θ_2 are proportional to the standard deviations of the two Gumbel distributions appearing in the mixture. The indexes 1 and 2 indicate respectively the basic and the outlier component. Using the α proportion, the model underlines the contribution of the two independent components to the rainfall extreme values.

The fitting of the mixture model to the time series of rainfall extreme values is evaluated comparing the L-moments of the at-site data (Hosking, 1990) with their theoretical expressions.

The theoretical expressions of the L-moments are defined as linear combinations of the Probability Weighted Moments (PWM) (Greenwood et al., 1979). The first four moments are:

$$\begin{aligned} \lambda_1 &= \beta_0; \\ \lambda_2 &= 2\beta_1 - \beta_0, \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0; \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 + \beta_0 \end{aligned} \tag{3}$$

where:

$$\beta_i = \int_{-\infty}^{+\infty} x \cdot F_X(x)^i dF(x) = \int_{-\infty}^{+\infty} x \cdot F_X(x)^i \cdot f(x) dx \tag{4}$$

is the i -th PWM given by:

$$\beta_0 = (1 - \alpha) \cdot \theta_1 \cdot \ln \lambda_1 + \alpha \cdot \theta_2 \cdot \ln \lambda_2 + \gamma [(1 - \alpha) \cdot \theta_1 + \alpha \cdot \theta_2] \tag{5}$$

$$\begin{aligned} \beta_1 &= (1 - \alpha)^2 \cdot I(1, 2, 1, 0) + (1 - \alpha)\alpha \cdot I(1, 2, 0, 1) \\ &+ (1 - \alpha)\alpha \cdot I(2, 1, 0, 1) + \alpha^2 \cdot I(2, 1, 1, 0) \end{aligned} \tag{6}$$

$$\begin{aligned} \beta_2 &= (1 - \alpha)^3 \cdot I(1, 2, 2, 0) + 2(1 - \alpha)^2\alpha \cdot I(1, 2, 1, 1) \\ &+ (1 - \alpha)\alpha^2 \cdot I(1, 2, 0, 2) + (1 - \alpha)^2\alpha \cdot I(2, 1, 0, 2) \\ &+ 2(1 - \alpha)\alpha^2 \cdot I(2, 1, 1, 1) + \alpha^3 \cdot I(2, 1, 2, 0) \end{aligned} \tag{7}$$

$$\begin{aligned} \beta_3 &= (1 - \alpha)^4 \cdot I(1, 2, 3, 0) + 3(1 - \alpha)^3\alpha \cdot I(1, 2, 2, 1) \\ &+ 3(1 - \alpha)^2\alpha^2 \cdot I(1, 2, 1, 2) + (1 - \alpha)\alpha^3 \cdot I(1, 2, 0, 3) \\ &+ (1 - \alpha)^3\alpha \cdot I(2, 1, 0, 3) + 3(1 - \alpha)^2\alpha^2 \cdot I(2, 1, 1, 2) \\ &+ 3(1 - \alpha)\alpha^3 \cdot I(2, 1, 2, 1) + \alpha^4 \cdot I(2, 1, 3, 0) \end{aligned} \tag{8}$$

where:

$$I(i, j, r, s) = \int_{-\infty}^{+\infty} x \cdot F_i^r(x) \cdot F_j^s(x) \cdot f_i dx \tag{9}$$

if $s=0$ is expressed by:

$$I(i, j, r, s) = \frac{\theta_i \cdot [\ln \lambda_i + \gamma + \ln(r + 1)]}{r + 1} \tag{10}$$

if $s>0$ and $\frac{s\lambda_j}{[(r+1)\lambda_i]^{\theta_i/\theta_j}} < 1$ is expressed by:

$$\begin{aligned} I(i, j, r, s) &= \frac{\theta_i}{r + 1} \ln((r + 1) \cdot \lambda_i) \cdot \left[\sum_{j=0}^{\infty} \frac{(-1)^j \lambda^j}{j!} \cdot \Gamma\left(\frac{j}{\theta} + 1\right) \right] + \\ &+ - \frac{\theta_i}{r + 1} \left[\sum_{j=0}^{\infty} \frac{(-1)^j \lambda^j}{j!} \cdot \left(-\gamma - \frac{\Gamma\left(\frac{j}{\theta} + 1\right)}{\frac{j}{\theta} + 1} \right) \right] \end{aligned} \tag{11}$$

here $\theta = \frac{\theta_j}{\theta_i}, \lambda = \frac{s\lambda_j}{[(r+1)\lambda_i]^{1/\theta}}$ and γ is the Euler's constant;

if $s>0$ and $\frac{s\lambda_j}{[(r+1)\lambda_i]^{\theta_i/\theta_j}} \geq 1$ is expressed by:

$$\begin{aligned} I(i, j, r, s) &= \frac{\lambda}{r + 1} \cdot \left[\ln(s \cdot \lambda_i) \cdot \theta_j \cdot \frac{1}{\lambda} \left[1 - \sum_{j=0}^{\infty} \frac{(-1)^j \lambda^j}{j!} \cdot \Gamma\left(\frac{j}{\theta} + 1\right) \right] \right] + \\ &- \frac{\lambda}{r + 1} \cdot \left[\frac{\theta_j^2}{\theta_i} \cdot \sum_{j=0}^{\infty} \frac{(-1)^j \lambda^j}{j!} \cdot \left[-\gamma - \frac{\theta}{j+1} \Gamma\left(\frac{j+1}{\theta}\right) \right] \right] \end{aligned} \tag{12}$$

here $\theta = \frac{\theta_j}{\theta_i}$ and $\lambda = \frac{(r+1)\lambda_i}{(s\lambda_j)^{1/\theta}}$ and γ is the Euler's constant.

3 Application and preliminary conclusion

The analysis of the capability of the mixture model to reproduce the L-moments of the sample values has been carried out using L-kurtosis ($\tau_4 = \lambda_4/\lambda_2$) and L-skewness ($\tau_3 = \lambda_3/\lambda_2$), evaluated from Eqs. (3) and (5)–(8).

The study has been developed using time series, more than 30 years long, of annual maximum of daily rainfall height, over a territory of about 22 000 km² in Tuscany. In particular 201 time series, from 1916 to 1996, having a low spatial correlation, have been extracted from a set of 584 rain gauges. Recently the database has been updated to the year 2000, and 307 time series have been extracted from a set of 880 rain gauges.

The analysis has been worked out on both the data sets and the L-moments, i.e. the L-kurtosis and the L-skewness, of the sample have been compared with the corresponding theoretical ones (Fig. 1). The comparison shows that the mixture model is able to reproduce the L-moments of the sample series on a wide range of values.

4 Future developments

4.1 Characterization of the outliers percentage

With the aim to develop a regionalization procedure for the parameters, based on the identification of homogeneous areas with respect to the basic component of the rainfall phenomena (Gabriele and Arnell, 1991), the spatial distribution of the outliers, in the studied region, has been investigated.

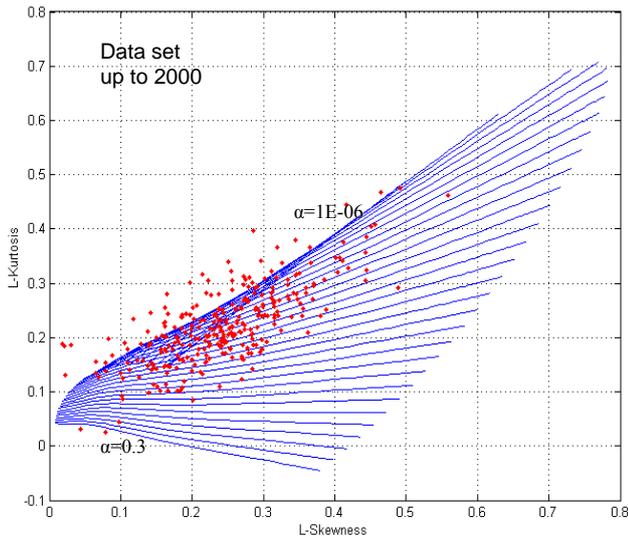


Fig. 1. Comparison of the sample L-skewness versus L-kurtosis (·) with the corresponding theoretical L-moments of the mixture model given the α proportion of rare events.

Particularly, the α proportion has been estimated and its spatial distribution has been analyzed. As a first approximation, the α proportion has been evaluated as the ratio of the number of outliers on the total number of events recorded in a rain gauge station. In future its relationships with territory characteristics (i.e. local elevation, slopes, exposure etc.) will be also investigated.

A box-and-whisker plot (Tukey, 1977) has been used to find the outliers of each site. The whisker extends to the farthest points that are considered as frequent events. Outliers' values are defined as values that are larger than the upper quartile plus 1.5 times the interquartile range. Each site is classified with the number of outliers per total number of data. The ratio of the outliers with the total number of data is the α proportion of the specific measurement site.

4.2 Outliers interpolation grid

To represent the particular trend of the outliers in the studied region, an inverse distance weighted interpolation has been used to predict the value for any unmeasured site. A 2×2 km grid has been generated (Fig. 2) assuming that each measured point has a local influence that decreases with the distance and data sets that are close are more similar than those that are farther apart.

The relationship between sample data and associated value with each cell is:

$$p(i, j) = \frac{\sum_{k=1}^n p_k \cdot \frac{1}{D_{k-(i,j)}^\nu}}{\sum_{k=1}^n \frac{1}{D_{k-(i,j)}^\nu}} \quad (13)$$

where $p(i, j)$ is the value of the quantity associated with the ij cell; p_k is the value of the quantity in the site k ; $D_{k-(i,j)}$ is

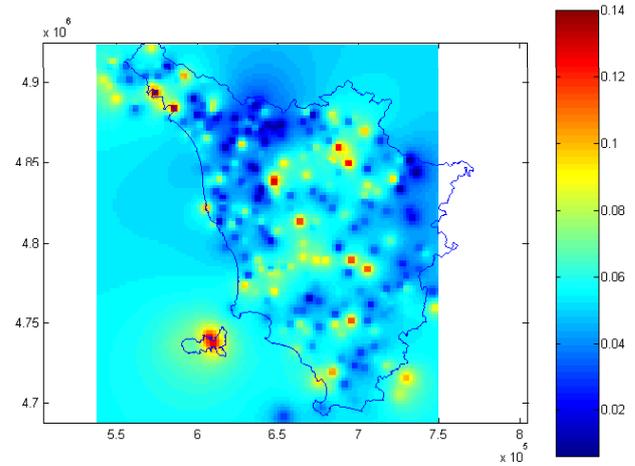


Fig. 2. Spatial distribution of outliers' percentage α in the region (UTM coordinates [m]).

the distance between the site k and the ij cell center; ν is the exponent.

To allocate values on the grid, distance between the ij cell and the measurement sites has been weighted. The decreasing influence of the rain gauge site is controlled by the exponent ν . After several tests, an interested value for the exponent seems to be $\nu = 2$. However, it still has to be proved that it is the optimal value, according to the minimization of the root mean square prediction error.

Acknowledgements. Authors would like to thank I. Becchi and A. Petrucci for suggestions and helpful discussions. This research has received support from the “Group for Prevention from Hydro-geological Disasters” (GNDCI) of the Italian National Research Council.

Edited by: L. Ferraris

Reviewed by: anonymous referees

References

Becchi, I., Caporali, E., Moro, A., and Sorbi, A.: Misure di distribuzione di valori estremi di precipitazione, XXVII Convegno di Idraulica e Costruzioni Idrauliche, II, 147–150, 2000.

Beran, M., Hosking, J. R. M., and Arnel, N.: Comment on “Two-Component Extreme Value Distribution for Flood Frequency Analysis” by Fabio Rossi, Mauro Fiorentino, and Pasquale Versace, *Water Resour. Res.*, 22, 263–266, 1986..

Caporali, E., Petrucci, A., and Tartaglia, V.: A mixture model of rainfall extreme values, 3rd Plinius Conference on Mediterranean Storms, edited by: Deidda, R., Mugnai, A., and Siccardi, F., GNDCI publ. n. 2560, 245–248, 2002.

Cunnane, C.: Methods and merits of regional flood frequency analysis, *J. Hydrol.*, 100(1–4), 269–290, 1988.

Gabriele, S. and Arnel, N. W.: A hierarchical approach to regional flood frequency analysis, *Water Resour. Res.*, 27, 1281–1289, 1991.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Probability weighted moments: definition and relation to

- parameters of several distributions expressible in inverse form, *Water Resour. Res.*, 15, 1049–1054, 1979.
- Haan, C. T.: *Statistical Methods in hydrology*, The Iowa State University Press/Ames., 1994.
- Hosking, J. R. M.: L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, *Journal of Royal Statistical Society, Series B*, 52, 105–124, 1990.
- Hosking, J. R. M. and Wallis, J. R.: Some statistics useful in regional frequency analysis, *Water Resour. Res.*, 29, 271–281, 1993.
- Matalas, N. C., Slack, R., and Wallis, J. R.: Regional skew in search of a parent, *Water Resour. Res.*, 11(6), 815–826, 1975.
- Rossi, F., Fiorentino, M., and Versace, P.: Two-Component Extreme Value Distribution for Flood Frequency Analysis, *Water Resour. Res.*, 20(7), 847–856, 1984.
- Rossi, F. and Villani, P.: A project for regional analysis of flood in Italy, *Proceedings of the NATO ADVANCED Study institute on Coping with Floods*, Erice, Italy, 193–217, 1992.
- Tukey, J. W.: *Exploratory data analysis*, Addison-Wesley, Reading, Mass., 1977.