Regional–scale analysis of extreme rainfalls via HMM

B. Betrò[1], A. Bodini[1,*] and Q.A. Cossu[2]

[1] *Institute of Applied Mathematics and Information Technology (CNR–IMATI), via E. Bassini 15, 20133 Milano (Italy); bruno.betro@mi.imati.cnr.it, antonella.bodini@mi.imati.cnr.it*
[2] *Sardinian Environmental Protection Agency (ARPA Sardegna), via Rockefeller 58–60, 07100 Sassari (Italy); qa.cossu@arpas.sardegna.it*
[*] *Corresponding author*

Abstract. *In [1], a hidden Markov model (HMM) with emissions distribution being a mixture of a Dirac distribution and two Weibull distributions proved to be able in identifying the recurrent patterns in the occurrence of extreme events over a small area of Sardinia (Italy). However, to reduce computational difficulties due to the high number of parameters, estimation was carried out after data transformation, and by exploitation of the relationship between Weibull distribution and Exponential distribution. In this paper, parameter estimation has been improved and the HMM applied to a few stations over all the Island, those having the highest 50yr return level. The obtained result confirm that mixture of Weibull distributions are able to adequately capture patterns of extreme events. However, the higher flexibility also introduces sensitivity of the model to spurious features of the data, like low values frequently occurring due to the instrumental inaccuracy, reducing the model smoothness.*

Keywords. *Extreme events; hidden Markov models; Mixtures; Weibull distribution.*

1 Introduction

In Sardinia (Italy), the highest frequency of extreme rainfall events is recorded in the Central-East area (3-4 events per year). The presence of high and steep mountains near the sea on the central and south-eastern coast, causes an East-West precipitation gradient in autumn especially, due to hot and moist currents coming from Africa. The occurrence and severity of extreme events in this area has been extensively studied in [1] and [2]. However, a few other areas in Sardinia are affected by extreme events. In November 2008, even two severe events caused floods in a wider area, including the southern cost: to face heavy damages to private houses, household goods, public infrastructure and agriculture, the

Regional Government allocated over 70 million euro (source: Regional Government of Sardinia).

In [1], the capability of a HMM in identifying possible recurrent patterns in the occurrence of extreme events over a small area of Central-East Sardinia has been investigated. Mixtures of Weibull distributions have been considered to capture extreme daily rainfall amounts: to reduce computational difficulties, parameter estimation was carried out after data transformation, and by exploitation of the relationship between Weibull distribution and Exponential distribution. The good performance obtained suggested to improve the estimation method and then to extend the analysis to the whole island. In this paper, parameter estimation has been improved and the HMM applied to a few stations over all the island, those having the highest 50 year daily precipitation return level. Finally, a comparison between the estimated sequence of hidden states with two different weather type classifications has been considered.

2 Area under study and data

We analysed daily rainfall data referring to the standard 30 year climatological period from 1961-1990, in accordance with WMO. Data are collected at 7 Governmental Hydrographic Service pluviometric stations and available at the Sardinian Environmental Protection Agency (henceforth ARPAS, Agenzia Regionale per la Protezione dell'Ambiente della Sardegna, in Italian). These stations have the highest 50 year daily precipitation return level in Sardinia, varying from 215 mm to 414 mm according to a GEV distribution ([4]). We selected the 153-day period from September-January, the period of highest occurrence of extreme events. Figure 1 shows the location of the stations together with a map of the area of Sardinia affected by the latest floods (2008; source: Servizio difesa del suolo, Regione Autonoma della Sardegna).

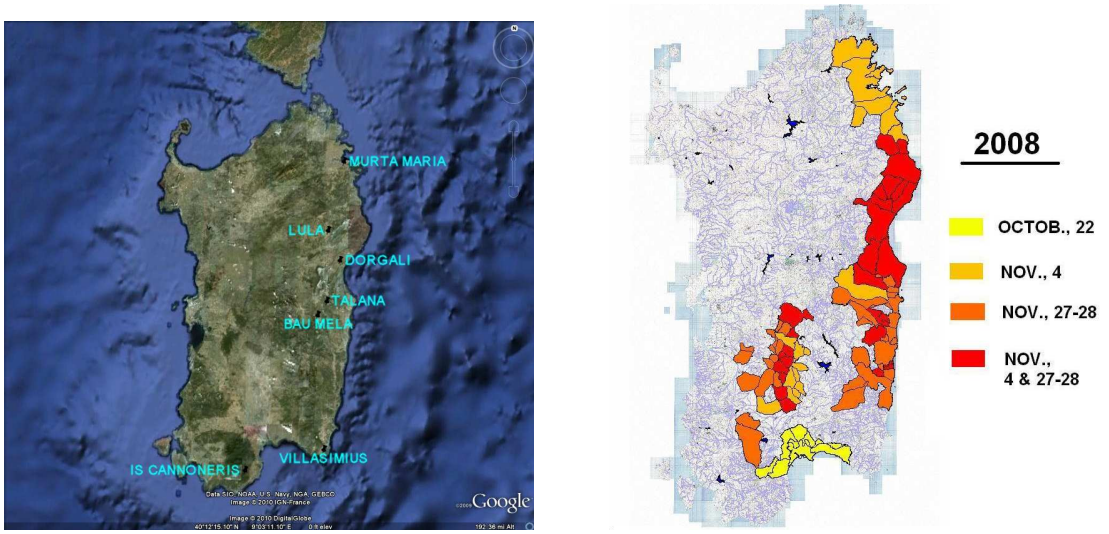

Figure 1: Left: location of the considered stations. Right: area of Sardinia affected by the latest floods (2008. Source: Servizio difesa del suolo, Regione Autonoma della Sardegna).

3 The hidden Markov model

Let $X_t = (X_{t1}, \ldots, X_{tq})$, $t = 1, 2, \ldots$, be a random vector representing the rainfall amounts measured at day t for a network of q rain stations. In a homogeneous HMM, process $\{X_t\}_t$ is assumed to be driven by a non observable (first order) Markov chain $\{C_t\}_t$ with finite state space $\{1, \ldots, K\}$, and the probability distribution of the observation X_t at any time t is determined only by the current state C_t. Following [5], we further assume that given the state C_t, the rainfall amounts at the different stations are independent

$$P(X_{t1} \leq r_1, \ldots, X_{tq} \leq r_q | C_t = i) = \prod_{j=1}^{q} P(X_{tj} \leq r_j | C_t = i), \quad \forall\, (r_1, \ldots, r_q) \in \mathbb{R}^q \tag{1}$$

and that the conditional distribution $P(X_{tj}|C_t)$ does not depend on time. In [1], $P(X_{tj}|C_t)$ was assumed to be a mixture of a Dirac measure on 0 and N Weibull distributions: $\alpha\beta x^{\alpha-1} \exp[-\beta x^{\alpha}]$, $x > 0$; $\alpha > 0, \beta > 0$. To reduce the computational difficulties due to the high number of estimating parameters, it has been proposed to firstly fitting a Weibull distribution to the seasonal maxima at each station and then a HMM to the transformed data $Y_{tj} = (X_{tj}/10)^{\beta_j}$, where β_j are the estimated shape parameters. In this paper, the shape parameters have been estimated along with other parameters, by the EM algorithm. Values X_{tj} are scaled for computational convenience.

4 Results

For purposes of comparison, the HMM has been firstly applied to the data presented in [1] and the parameters of the model have been reestimated: the BIC criterion selects the model with two Weibull components and discards the two-Gamma and the two-Exponential models. Goodness–of–fit diagnostics generally improve, thanks to the increased degree of flexibility. Six hidden states are obtained as in [1], which are well separated.
Then, the model has been estimated from the new data set, presented in Section 2. The BIC criterion selects $K = 4$ and the estimated model reproduces the main observed frequencies very closely. By examining both the estimated rainfall probabilities and the conditional distributions of positive rainfall amounts, a dry state and 3 rainy states are recognized, the heaviest tails being obtained under state 1. The two Weibull components under state 1 are shown in Figure 2.

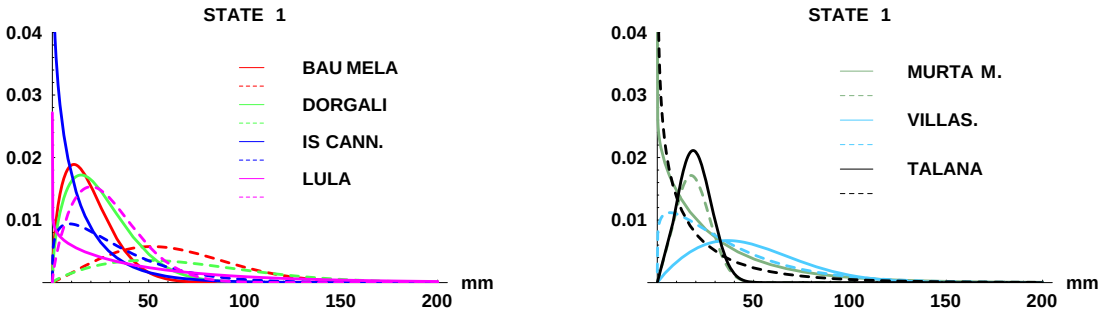

Figure 2: The Weibull components of state 1.

One of the main interests of modelling daily rainfall data by a HMM lies in the underlying correspondence between the hidden states and the concept of discrete weather states. Therefore, we compared

the most likely state transition path given the sequence of observations, or estimated state sequence (ESS), obtained using the Viterbi algorithm, with a) temporal clusters based on spatial similarity, in the mesoscale surface fields of precipitation as determined in ([3]); b) a weather types at 850 hPa classification (Dr. Mariani, personal communication). As far as the the former are concerned, a strong association between the ESS and the three clusters defined in [3] is obtained ($p \ll 0.01$). As in [3] a relationship between the clusters and different synoptic circulation types is indicated, the strong association obtained with the ESS suggests that this can provide a simple and alternative way to downscale atmospheric data. However, the fact that no association has been found with the more general weather types classification indicates that, as expected, the ESS mainly reflects certain identified spatial rainfall distribution patterns.

5 Discussion

In this work it has been established that mixture of Weibull distributions are able to adequately capture patterns of extreme events. The more general estimation procedure improves previous results, obtained in [1]. Therefore, mixture of Weibull distributions appear to be adequate at both the small and the large scale. Extreme events occurring at the sites considered in this paper seem to be triggered by the same atmospheric circulation, and this can be one of the reasons of the good results. This issue has to be further investigated. Finally, it has to be noted that the high degree of flexibility introduced by Weibull distributions can also allow for an insufficient model smoothness, as high values of α introduce sharp distributions, corresponding to very frequent data (like small daily rainfall values). This suggest that mixtures of Weibull distributions and of smoother distributions, like the Gamma distribution, can further improve the results. Such a mixtures could also be suitable for including in the analysis other sites, interested by extreme events at a lesser extent.

Acknowledgments. The authors are grateful to Dr. Luigi Mariani and to Dr. Piero Chessa, who kindly provided the weather types classifications. This work has been partially supported by the European PROTERINA–C project (2009–2010) – Programma operativo di Cooperazione Transfrontaliera Italia–Francia “Martittimo”, grant agreement ARPAS – IMATI CNR, nr. 1219.

References

[1] Betrò, B., Bodini, A., Cossu Q. A. (2008). Using a hidden Markov model to analyse extreme rainfall events in Central-East Sardinia. *Environmetrics* **19**, 702–713.

[2] Bodini, A., Cossu Q. A. (2010). Vulnerability assessment of Central-East Sardinia (Italy) to extreme rainfall events *Natural Hazard Earth System Science* **10**, 61–72.

[3] Chessa, P. A., Cesari, D., Delitala, A. M. S. (1999). Mesoscale regimes of precipitation and temperature over Sardinia (Italy) and the related synoptic-scale circulations. *Theoretical and Applied Climatology* **63**, 195–221.

[4] Cossu Q. A., Bodini, A., Canu, S. (2009). Vulnerabilità della regione Sardegna agli eventi estremi di precipitazione. In *Atti del 12mo Convegno nazionale di Agrometeorologia*, Sassari, 15-17 Giugno 2009. Rivista Italiana di Agrometeorologia, vol 14, n.2 (Italian).

[5] Zucchini, W., Guttorp, P. (1991). A hidden Markov model for space-time precipitation. *Water Resources Research* **27**, 1917–1923.