

A hidden Markov model for daily rainfall occurrence over southern Calabria Ionica (Italy) ^(*)

HMMs per le occorrenze di pioggia sulla Calabria Ionica meridionale

B. Betrò¹, A. Bodini¹, Q. A. Cossu², G. Gullà³, O. Terranova³

¹ CNR-IMATI, Milano

e-mail: antonella.bodini@mi.imati.cnr.it

² SAR, Sassari

³ CNR-IRPI, Cosenza

Riassunto: Si presenta un'analisi statistica della serie temporali giornaliere delle piogge, schematizzate come fenomeno binario, relativamente ad un'area della Calabria Ionica Meridionale. Fra le 45 stazioni disponibili per quest'area sono state selezionate quelle che possiedono un periodo di osservazione comune sufficientemente lungo e completo. In particolare si è considerato il trentennio 1961-1990. Il fenomeno è stato quindi modellato tramite un cosiddetto *hidden Markov process*.

Keywords: hidden Markov model, rainfall occurrence.

1. Introduction

The main goal of this work is to examine the probability distribution of local daily rainfall occurrences in part of southern Calabria Ionica to sketch a framework for further studies aiming to define possible scenarios of landslide-triggering due to rainfall. Despite a large literature, the high frequency of landslide events makes necessary deep researches to determine a complete background where the problem of soil slip could be analyzed in a systematic way.

2. Observed rainfall data

We use daily rainfall data collected at 5 stations from the southern Calabria Ionica, listed in Table 1. Among the 45 available, these stations are close to the area of interest and have complete records for the standard 30 years period 1961-1990 (SP), according to the *World Meteorological Organization* suggesting a length of about 30 years for the analysis of weather features, (Thom, 1966). In fact, the current standard period is 1971-2000; our choice is due to the outstanding lack of data for the period 1991-2000. We select the 182-days period beginning October 1 (October-March, OM) corresponding to about 80% of the annual precipitation and to 70% of the raining days. A wet day is defined as having a rainfall of over 0.1mm. In the case of leap years, the rainfall amount at the extra day (i.e., February 29) is added to March 1. The geographical coordinates of the stations together with some rainfall features are shown in Table 1. For comparison, Table 1 contains both

(*) The authors are very grateful to Ing. G. Iiritano (Arpacal, Cosenza, Italy) for useful explanations.

the means over the OM period and the corresponding standard normals⁽¹⁾.

Table 1: *The stations.*

	Gioiosa I.	Riace	Roccella I.	Simbario	Staletti
Egb (1)	2633831.31	2649046.09	2642346.50	2635522.64	2653859.28
Ngb (1)	4244136.10	4253219.10	4242617.83	4274467.10	4291867.00
altitude (meters)	125	304	5	760	390
N. of wet days (year)	113.4	109.2	85.5	114	87.1
N. of wet days (OM)	80.2	77.9	61.0	76.9	61.3
Persistency (OM) (2)	0.63	0.62	0.51	0.60	0.57
Daily rainfall (year) (3)	2.37	2.42	1.91	4.22	2.46
Daily rainfall (OM)	3.79	3.89	3.04	6.36	3.85

(1) Gauss–Boaga system. (2) Frequency of wet days following a wet day. (3) Expressed in mm.

From now on only the OM period will be considered. The analysis of rainfall occurrences shows that one-day periods of rain have the highest relative frequency (from 0.4 to 0.5), while the relative frequency of raining periods of more than 5 days is smaller than 0.1. The analysis of the mean daily rainfall does not show any trend along the SP, while a clear decreasing trend is highlighted in the mean fraction of wet days. This clearly reflects the well-recognized fact that extreme weather conditions due to climate changes are occurring in the Mediterranean area. Concentration of heavy rainfalls in few days emphasizes weathering processes of superficial rocks, which turn out to be more vulnerable in front of severe events (cfr Gullà *et al.*, 2005).

3. The hidden Markov model

In a hidden Markov model (HMM) an underlying and unobserved sequence of states follows a Markov chain with a finite state space, and the probability distribution of the observation at any time is determined only by the current state of the Markov chain. More formally, let $X_t = (X_{t1}, \dots, X_{tq})$ be a multivariate random vector of rainfall occurrences for a network of q rain stations. Let the observed value $x_{ti} = 1$ if rain is observed on day t at station i and $x_{ti} = 0$ otherwise. Let C_t be the hidden rainfall state for day t , taking on one of K values. By $X_{1:T}$ and $C_{1:T}$ we will denote the sequence of T daily precipitation occurrences and hidden rainfall states, respectively. A (homogeneous) HMM defines the joint distribution on $X_{1:T}$ and $C_{1:T}$ by assuming that $C_{1:T}$ is a first-order stationary Markov chain and $P(X_t = x | X_{1:t-1}, C_{1:t}) = P(X_t = x | C_t)$. Following Zucchini and Guttorp (1991), we further assume that given the state C_t the components of X_t (i.e., the occurrences at the different stations) are independent and that the (conditional) distribution of X_{tj} does not depend on time:

$$P(X_t = (x_1, \dots, x_q) | C_t = i) = \prod_{j=1}^q P(X_{tj} = x_j | C_t = i) = \prod_{j=1}^q p_{ij}^{x_j} (1 - p_{ij})^{(1-x_j)}$$

where $p_{ij} = P(X_{tj} = 1 | C_t = i)$. This hypothesis ascribes to the hidden states only the dependence among stations and has to be carefully considered when medium or small area

⁽¹⁾ A climatological standard normal is the average of climatological data computed for consecutive periods of 30 years.

are studied. In our case, it can be justified by the different geographical characteristics of selected stations. HMMs for climate analysis were further developed by Hughes and Guttorp (1994) and Hughes *et al.* (1999) among others. Their interest lies in the underlying correspondence between the hidden states and the concept of discrete weather states, as described for instance by Bardossy and Plate (1992); instead of explicitly define the weather states, HMMs allow to define them according to the observed data. To our knowledge, these models are seldom applied in hydrological investigations but we believe that their contribution might be substantial.

4. Estimated parameters and hidden states

The direct numerical maximization of log-likelihood is essentially based on an EM algorithm: computations have been performed by Fortran code referred to in Hughes *et al.* (1999). The Bayesian Information Criterion (BIC) allows to select the 5-state HMM, although the 4-state model has a BIC value only slightly higher than the minimum one (23193 versus 23189) so that it turns out to be preferable for the sake of parsimony. More important, as we will see soon after, the estimated 4 states have a clear meteorological interpretation, therefore we adopt the 4-state model for the analysis. The values on the main diagonal of the transition matrix (not shown here) are relatively large (i.e., > 0.5) in the case of states 1 and 4 only. The direct transition from state 2 to state 3 has the highest probability (~ 0.28). Table 2 shows the estimated probabilities of a wet day in each station for the different hidden states. State 1 turns out to be clearly associated to a frontal perturbation, while state 4 to a stationary high pressure system. States 2 and 3 seem to behave in a complementary way with respect to the five stations, thus capturing the local instability possibly due to different altitudes (see Table 1).

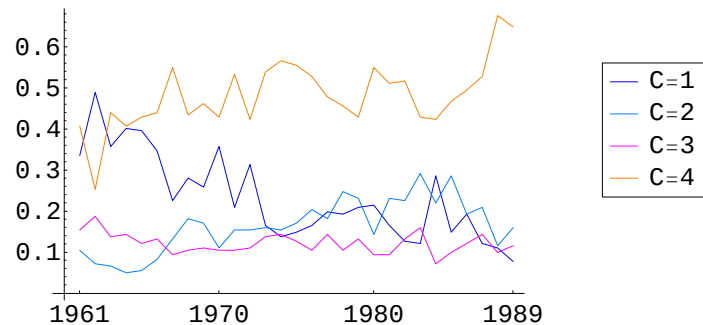
Table 2: *Rainfall probabilities*

	C=1	C=2	C=3	C=4
Gioiosa I.	0.9733	0.8651	0.2975	0.0391
Riace	0.9703	0.9121	0.2183	0.0307
Roccella I.	0.8895	0.7272	0.0475	0.0092
Simbario	0.9170	0.5709	0.4997	0.0407
Staletti	0.9570	0.2555	0.3518	$4.47 \cdot 10^{-8}$

The Viterbi algorithm allows to compute the most likely state transition path given the sequence of observations. In Figure 1 we show the variation of the seasonal frequency of the different states in the estimated state-sequence (ESS). Note that the mean frequency of state 1 decreases in time: this fact might correspond to the observed decrease of the number of wet days along the SP.

To test the ability of the model to reproduce the data, we compare observed and estimated: (a) rainfall probabilities at each station; (b) correlations between rainfall occurrences at different stations and (c) runlength distributions. We obtain an highest absolute difference $< 10^{-4}$ in case (a); of 0.01 in case (b) and of 0.049 in case (c). This latter value corresponds to the difference between the relative frequency and the estimated probability of one-day periods of rain at Simbario. These results suggest that the fit is

Figure 1: Variation of the seasonal frequencies in ESS.



sufficiently good.

5. Conclusions

In the light of the obtained results, two kind of generalizations have to be considered in future work. The decreased number of wet days we observed suggests that the hypothesis of stationarity could not be adequate here. Moreover, because of the relatively small extension of the considered area, alternative models to conditional independence could be considered to further improve the fitting and the predictive capabilities. Finally, non-homogeneous models will be considered in presence of further atmospheric data, to downscale General Circulation Models at the local level, as in Hughes *et al.* (1999).

References

- Bardossy A., Plate E. J. (1992) Space-time models for daily rainfall using atmospheric circulation patterns, *Water Resources Research*, 28, 1247–1259.
- Gullà G., Mandaglio M.C., Moraci N. (2005) Influence of degradation cycles on the mechanical characteristics of natural clays, in *Proceedings of the 16th International Conference on Soil Mechanics and Geotechnical Engineering*, Osaka, Japan.
- Hughes J. P., Guttorp P. (1994) A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, *Water Resources Research*, 30, 1535–1546.
- Hughes J. P., Guttorp P., Charles S. P. (1999) A nonhomogeneous hidden Markov model for precipitation occurrence, *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 48, 15–30.
- Thom, H. (1966) *Some Methods of Climatological Analysis*, WMO Technical Note Number 81, Secretariat of the World Meteorological Organization, Geneva, Switzerland.
- Zucchini W., Guttorp P. (1991) A hidden Markov model for space-time precipitation, *Water Resources Research*, 27, 1917–1923.