

Occurrence analysis of daily rainfalls through non-homogeneous Poissonian processes

B. Sirangelo, E. Ferrari, and D. L. De Luca

Soil Conservation Department, University of Calabria, Rende (CS), Italy

Received: 17 June 2010 – Revised: 5 January 2011 – Accepted: 11 March 2011 – Published: 14 June 2011

Abstract. A stochastic model based on a non-homogeneous Poisson process, characterised by a time-dependent intensity of rainfall occurrence, is employed to explain seasonal effects of daily rainfalls exceeding prefixed threshold values. The data modelling has been performed with a partition of observed daily rainfall data into a calibration period for parameter estimation and a validation period for checking on occurrence process changes. The model has been applied to a set of rain gauges located in different geographical areas of Southern Italy. The results show a good fit for time-varying intensity of rainfall occurrence process by 2-harmonic Fourier law and no statistically significant evidence of changes in the validation period for different threshold values.

1 Introduction

Investigations on daily rainfall time series may contribute at various levels to the assessment of the existence of temporal and spatial variability in the rainfall field characteristics. The different models that can be used may be classified into four categories (Onof et al., 2000). The first category is formed by meteorological models, which involve complex sets of differential equations in order to represent the complete physical processes controlling precipitation and other weather variables. The second one is represented by stochastic models which describe the spatial evolution of the rainfall process in a scale-independent approach, like the multi-scaling models (Lovejoy and Schertzer, 1990; Gupta and Waymire, 1994). A third one is formed by statistical models (single or multi-site temporal), able to represent spatial non-stationarity and temporal trend (Stern and Coe, 1984). Finally, the fourth

category concerns temporal and spatial-temporal stochastic process models which try to summarize the main rainfall features by a few parameters, providing an understanding of the hierarchical structure of the rainfall process. In particular, these models follow a point process approach based upon the use of Poisson and Poisson-cluster processes for explaining the joint or separate arrival and magnitude processes of daily rainfalls (Waymire and Gupta, 1981a, b, c; Kavvas and Delleur, 1981; Yevjevich and Dyer, 1983; Rodriguez-Iturbe et al., 1984, 1987; Onof et al., 2000).

This latter approach has been used for a long time as a tool for the stochastic modelling of rainfall, combining in different ways the basic Poisson process of storm arrivals with the Neyman-Scott or the Bartlett-Lewis point processes. In more recent developments, rainfall variables, such as the duration and intensity of a rainy event, have been treated as correlated variables or, alternatively, correlation has been introduced using independent superposed point processes to represent the different kinds of storms in a region (Cowpertwait et al., 2007; Leonard et al., 2008). The statistical properties of the superposed process are normally obtained by aggregation of the properties of each independent process. In other words, spatial-temporal models can be flexibly structured with continuous distributions of storm types, thus providing a parameter configuration that can be efficiently fitted to a large range of historical rainfall records (Cowpertwait, 2010). Point process rainfall models are generally fitted to the observed sample properties using derived moments and model functions. A good fit is usually assessed by comparing observed properties of data not used to estimate the model parameters with equivalent properties simulated using the fitted model (Entekhabi et al., 1989; Cowpertwait et al., 2007). This procedure, if successfully extended to extreme values, can certify the model suitability, due to the sensitivity of the distributions of the extremes to any departures in the good fitting of the depth distributions.



Correspondence to: E. Ferrari
(ferrari@dds.unical.it)

Alternatively, in order to reproduce the temporal and/or spatial non-stationary features in both the analysis and the simulation of rainfall variability, the generalized linear models (GLMs) can be applied (Chandler and Wheater, 2002; Wheater et al., 2005). These models explain daily rainfall variability by using probability distributions in which the probability of rainfall occurrence is modelled separately from the amount of rain if non-zero. By developing appropriate methods of spatial-temporal disaggregation of daily rainfalls, GLMs can also be linked to the single-site models to provide multi-site sequences incorporating sub-daily temporal structure.

With reference only to the temporal variability of the rainfall process, the stochastic point process models, described above, are widely used for at-site analysis of storm occurrences, in order to explain the intermittent feature of rainfalls and simulate inter-storm periods. Nevertheless, the stochastic models often hypothesize temporal stationary processes. In this case, the apparent seasonal periodicity exhibited by the frequency of the rainfall phenomena can be modelled as a stationary stochastic process in short temporal intervals, in which the behaviour can be considered reliably homogeneous.

Another approach of functional type, aimed to cope with seasonal and trend variability of the storm occurrences, is based on temporally non-homogeneous stochastic point processes with time depending parameters (Sirangelo, 1994; Pujol et al., 2007). In this case, great attention has to be paid to the parsimony of the models, as regards the number of parameters and the bias introduced into the generation of synthetic series, and to the influence of threshold values in extracting a peak storm database from daily rainfalls data.

In this work, the procedure adopted for explaining the occurrence process of daily rainfalls greater than the prefixed threshold values is a non-homogeneous Poisson model, with temporal variation of intensity expressed by using Fourier series. The model has been applied to a set of rain gauges located in different geographical areas of Southern Italy. The proposed procedure constitutes the first step of a more comprehensive peak-over-threshold (POT) analysis, in which the occurrence and exceedance of rainy events globally depend on rainfall threshold values.

2 Non-homogeneous Poisson model for storm occurrence

The storm occurrence process, as successive clustered events, has been explained by several researchers as the result of specific atmospheric conditions able to generate climatic perturbations (Waymire and Gupta, 1981a, b, c; Smith and Karr, 1983; Chang et al., 1984; Onof et al., 2000). This process is generally non-homogeneous in time, as shown for a typical rain gauge of Southern Italy where a marked reduction of occurrence intensity related to daily rainfalls during

spring-summer period is detected (Fig. 1). Moreover, the statistical features of the storm occurrence process can be influenced by the threshold chosen for extracting exceedance values from the basic rainfall time series. Both the problems of choosing threshold value and selecting the criteria for retaining peaks represent some of the main difficulties classically associated with the POT analysis. This approach, a kind of compromise between annual maxima frequency and classical time series modelling, requires the analysis of both the magnitude and the arrival time of storm peaks, neglecting the autocorrelation structure of the basic data. The higher rainfall values of different sets of consecutive rainy days are generally assumed as rainfall peaks, thus, practically assuring the independence required for the Poisson process. For the choice of threshold value, which directly influences the distribution of peaks, some statistical tests and practical guidelines are available (Lang et al., 1999).

Adopting a non-homogeneous Poisson process, the occurrence probability of n events in a time period $[t, t + \Delta t]$ is given by:

$$p_{N_{t,\Delta t}}(n) = \frac{\Lambda_{t,t+\Delta t}^n e^{-\Lambda_{t,t+\Delta t}}}{n!} \quad \Lambda_{t,t+\Delta t} > 0; \quad n = 0, 1, \dots \quad (1)$$

The mean number of events $\Lambda_{t,t+\Delta t}$ depends on both time t and interval Δt :

$$\Lambda_{t,t+\Delta t} = f(t, \Delta t) = \int_t^{t+\Delta t} \lambda(x) dx \quad (2)$$

where the law $\lambda(\cdot)$ represents the temporal variation of the intensity function of the occurrence process. For a homogeneous Poisson process the function $\lambda(\cdot)$ is constant, therefore, it holds $\Lambda_{t,t+\Delta t} = \lambda \cdot \Delta t$.

Actually, the temporal variation of intensity implies changes in some properties of the homogeneous Poisson processes. In particular, the times between consecutive events, which are independent and exponentially distributed in the case of homogeneous process, are distributed with the following conditional probability density:

$$p_{T_n|T_{n-1}\dots T_1}(\tau_n|\tau_{n-1}\dots\tau_1) = \lambda(t + \tau_1 + \dots + \tau_n) e^{-\int_{t+\tau_1+\dots+\tau_{n-1}}^{t+\tau_1+\dots+\tau_n} \lambda(x) dx} \quad (3)$$

With this position, the waiting times form a Markov chain with transition density (Snyder, 1975):

$$p_{W_i|W_{i-1}}(w_i|w_{i-1}) = \lambda(w_i) e^{-\int_{w_{i-1}}^{w_i} \lambda(x) dx} \quad (4)$$

characterised by the following joint probability density:

$$p_{W_1,\dots,W_n}(w_1,\dots,w_n) = \left[\prod_{i=1}^n \lambda(w_i) \right] e^{-\int_t^{w_n} \lambda(x) dx} \quad t < w_1 < \dots < w_n. \quad (5)$$

Table 1. Main statistics of daily rainfall occurrence process for the selected rain gauges.

Code	Rain gauge	Elevation (m)	Years	No. of data	Rainy days	Frequency of rainy days
Western rain gauges						
1010	Cosenza	250	80	28 791	9118	0.317
1030	S. Pietro in Guarano	640	71	25 230	7284	0.289
1040	Rende	482	71	25 474	7720	0.303
3060	Paola	97	69	24 715	7690	0.311
Eastern rain gauges						
1590	Belvedere Spinello	330	69	24 577	5098	0.207
1620	Verzino	550	67	24 345	6116	0.251
1650	Strongoli	342	69	24 917	5571	0.224
1680	Crotone	6	79	27 801	6647	0.239

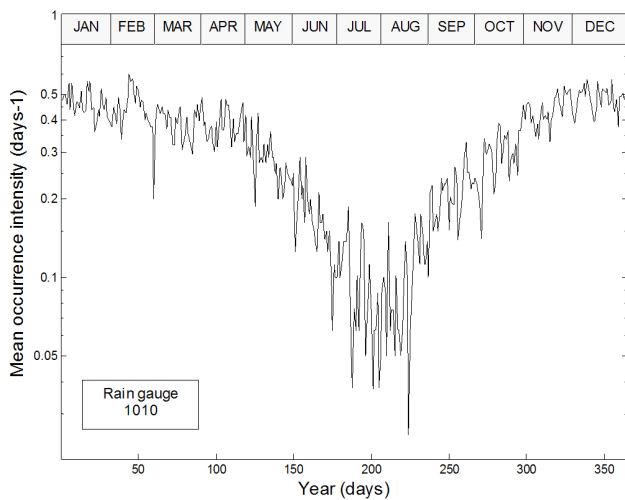


Fig. 1. Example of time variability of the intensity function of daily rainfall occurrences.

2.1 Temporal variation of rainfall occurrence intensity

Meteorological phenomena show statistically significant seasonal and daily features due to the revolution of the Earth around itself and the sun. Thus, the temporal variation of rainfall occurrence intensity $\lambda(t)$ can be expressed through the Fourier series as a function of period $P = 1$ year:

$$\lambda(t) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} [a_k c_k(t) + b_k s_k(t)] \tag{6}$$

where $c_k(t)$ and $s_k(t)$ indicate the functions:

$$s_k(t) = \sin\left(\frac{2\pi}{P} k t\right) \quad c_k(t) = \cos\left(\frac{2\pi}{P} k t\right) \tag{7}$$

To ensure parsimony of the model, Fourier series has to be limited to an optimal number of harmonics, n_h , matched with the goal of the proposed procedure:

$$\lambda(t) = \frac{1}{2}a_0 + \sum_{k=1}^{n_h} [a_k c_k(t) + b_k s_k(t)] \tag{8}$$

By integrating the Eq. (2) with the expressions (7) and (8) one obtains:

$$\Lambda_{t, t+\Delta t} = \frac{1}{2}a_0\Delta t + \frac{P}{\pi} \sum_{k=1}^{n_h} \frac{1}{k} s_k\left(\frac{\Delta t}{2}\right) \left[a_k c_k\left(t + \frac{\Delta t}{2}\right) + b_k s_k\left(t + \frac{\Delta t}{2}\right) \right] \tag{9}$$

For time intervals equal to 1 yr, independently on the time value t , the mean number of events $\Lambda_{t, t+\Delta t}$ becomes:

$$\Lambda_{t, t+P} = \frac{1}{2}a_0P \tag{10}$$

Moreover, it is well known that a non-homogeneous Poisson process for rainfall occurrences can be transformed into a homogeneous and unit one through a suitable transformation of the time variable (Cox and Isham, 1980). The transformation is based on the inverse function of Eq. (2), i.e.:

$$\Delta t^*(\Delta t) = \inf\{w : \Lambda_{t, t+w} \geq \Delta t\} \quad \Delta t > 0 \tag{11}$$

Operatively, starting from the representation of the function $\Lambda_{t, t+\Delta t}$ in which $t = 0$ and $\Delta t = w$ varying from 0 to P (Fig. 2), a subdivision of the interval $[0; \Lambda_{0, P}]$ into n_s equal subperiods is carried out on the vertical axis. Therefore, it is possible to determine the corresponding partition of the period P on the horizontal axis, thus, providing n_s time subperiods.

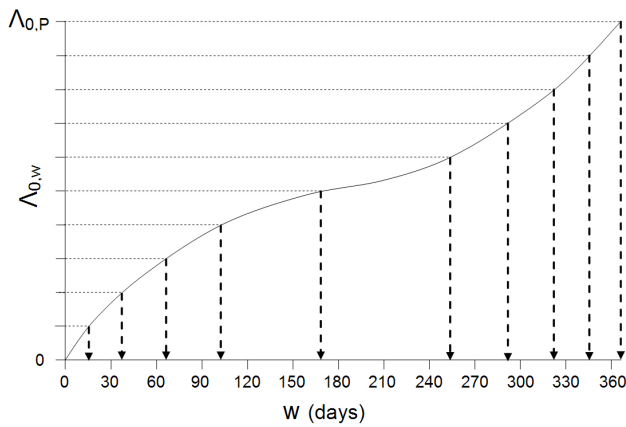


Fig. 2. Subdivision of the period P into n_s subperiods (in this example $n_s = 10$).

2.2 Parameters estimation

The evaluation of the optimal number of harmonics n_h in Eq. (8) plays a crucial role in parameters estimation of $\lambda(t)$ function. In technical literature, cross-validation techniques are mostly adopted (Picard and Cook, 1984; Cressie, 1993; Balachandrudu et al., 2009), but a limit of these methods consists of the assumption that the residuals, between sample and theoretical values, are independent and Gaussian distributed. These hypotheses need further studying (Cressie, 1993). Consequently in this work, besides the application of cross-validation techniques and, in particular, the evaluation of Cross Validation Error (CVE) (Balachandrudu et al., 2009), we propose a more robust procedure, which is based on the property that a non-homogeneous Poisson process can be transformed into a homogeneous one (Sect. 2.1).

In detail, for each rain gauge and for a prefixed number of harmonics n_h , the cross-validation approach concerns the following steps:

- a1. the whole data sample is partitioned into N not overlapped subsamples;
- b1. a single subsample is withheld and the remaining subsamples are used to estimate the parameters a_0 and a_k, b_k , with $k = 1, \dots, n_h$, of Eq. (8) by using least squares method;
- c1. the mean-squared difference between the theoretical and sample values related to the withheld subsample is computed;
- d1. the point (b1) and (c1) are repeated for each subsample. The average of the squared differences over the number of withheld subsamples provides the cross validated square error; its square root represents the CVE.

The optimal number of harmonics n_h is characterised by the minimum value of the CVE.

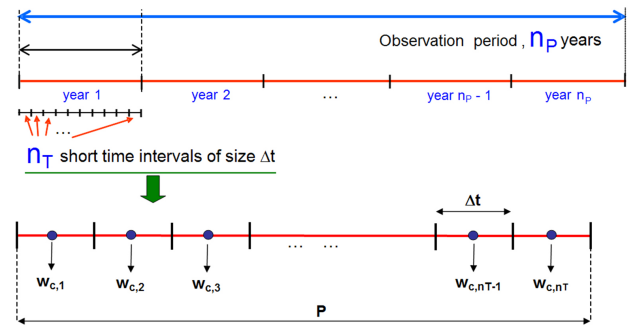


Fig. 3. Subdivision of the period P (1 yr) into n_T short time intervals of size Δt and localization of $w_{c,m}$ ($m = 1, \dots, n_T$).



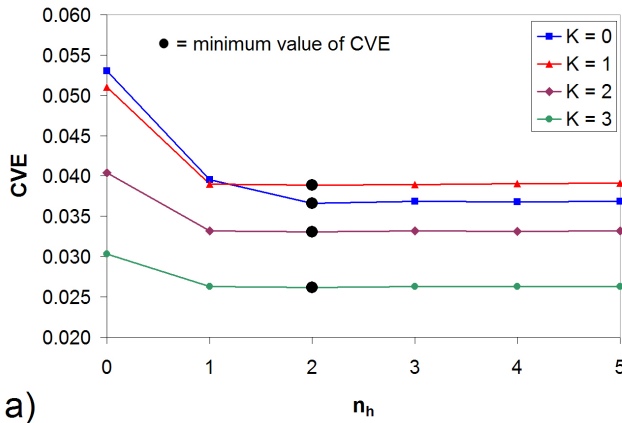
Fig. 4. Daily rainfall series of Calabria region used in the present study.

Instead, the procedure based on the transformation of a non-homogeneous process into a homogeneous one consists of the following steps, again for each rain gauge and for a prefixed number of harmonics n_h :

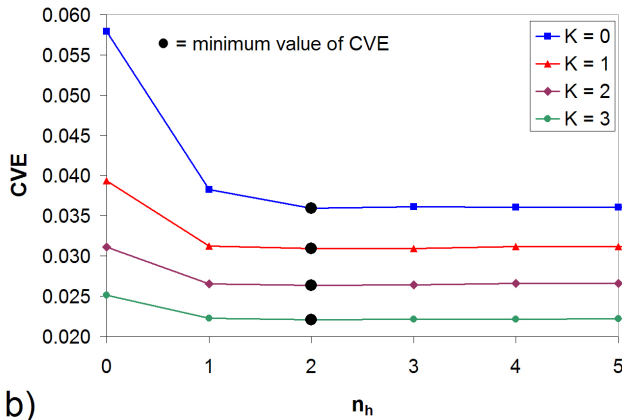
- a2. evaluation of the parameters a_0 and a_k, b_k , with $k = 1, \dots, n_h$, of Eq. (8) by using least squares method and considering the whole data sample;
- b2. subdivision of the period P into n_s subperiods through the partition of the segment $[0; \Lambda_{0,P}]$ (Fig. 2);
- c2. application of two Hald tests (1952), aimed at verifying that the number of occurrences, related to each subperiod, may be derived from the same homogeneous Poisson process (see Appendix A1).

In this case, the optimal number of harmonics n_h coincides with the minimum number which satisfies the statistical tests concerning the point (c2).

As regards the application of the least squares method to both the procedures (points a1 and a2), firstly a sample estimation of the occurrence intensity may be obtained. For this scope, considering $P = 1$ yr and consequently n_p years of rainfall observations, each year is subdivided into n_T very short time interval of size Δt (three or five days); in this way $n_{m,p}$ represents the number of sample occurrences related to



a)



b)

Fig. 5. Cross validated error (CVE) values estimated for each threshold value and the number of harmonics for (a) Cosenza and (b) Crotone rain gauges. The minimum CVE values occur for $n_h = 2$.

the m^{th} interval of the p^{th} year (Fig. 3). Therefore, the sample evaluation of occurrence intensity is obtained as:

$$\lambda_m = \frac{1}{n_p} \sum_{p=1}^{n_p} \frac{n_{m,p}}{\Delta t} \quad m = 1, 2, \dots, n_T \quad (12)$$

and applying the least squares method the parameters of Eq. (8) are evaluated as:

$$\begin{aligned} \frac{1}{2}a_0 &= \frac{1}{n_T} \sum_{m=1}^{n_T} \lambda_m & a_k &= \frac{2}{n_T} \sum_{m=1}^{n_T} \lambda_m c_k(w_{c,m}) \\ b_k &= \frac{2}{n_T} \sum_{m=1}^{n_T} \lambda_m s_k(w_{c,m}) & k &= 1, 2, \dots, n_h \end{aligned} \quad (13)$$

where $w_{c,m}$ is the temporal centre of the m^{th} time interval Δt .

Once partitioned the period P into n_s subperiods (Fig. 2), on the basis of the function $\Lambda_{t,t+\Delta t}$ which depends on the number of harmonics n_h (Eq. 9), it has been indicated as $n_{i,j}^{(n_h)}$ the sample occurrences related to the j^{th} sub-period ($j = 1, \dots, n_s$) into the i^{th} year ($i = 1, \dots, n_p$).

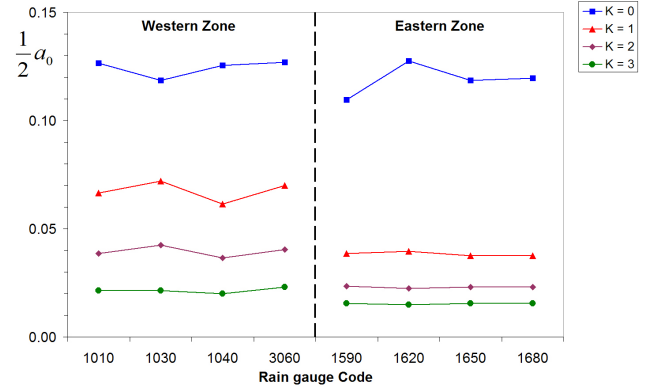


Fig. 6. Evaluation of $a_0/2$ for each analysed rain gauge and threshold values.

Considering the sum of the occurrences:

$$S_j^{(n_h)} = \sum_{i=1}^{n_p} n_{i,j}^{(n_h)} \quad j = 1, 2, \dots, n_s \quad (14)$$

for each of the joint set of sub-periods n_s , it can be hypothesized that variable $S_j^{(n_h)}$ is Poisson-distributed. The application of Hald tests (see Appendix A1) is aimed at verifying that all the $S_j^{(n_h)}$ variables derive from the same homogeneous Poisson process.

The procedures described in this section can be used for modelling the temporal variability of intensity referred to storm occurrences of daily rainfall heights exceeding any threshold value T , defined as:

$$T = \mu + K \cdot \sigma \quad (15)$$

where μ is the expected value of daily rainfall, σ is the standard deviation and $K = 1, 2, 3, \dots$. If $K = 0$, it is assumed $T = 0.1$ mm.

3 Dataset

The procedure has been applied to 8 daily rainfall series (Table 1) located in Calabria region (Southern Italy), out of which 4 belong to the western zone and 4 to the eastern zone of the region (Fig. 4). The choice of the database has been based on length and completeness of the time series. In particular, the time series have been selected on the basis of availability of at least 50 yr in the time period 1921–1986, chosen as the calibration period, and of all the years of observation in the subsequent validation period 1987–2001, where the daily rainfall occurrence process variability is under hypothesis. The mean number of observation years of the dataset during the period 1921–2001 is 71.9, with a maximum of 80 yr (Cosenza rain gauge), while the mean values of daily rainfall frequency are 0.305 and 0.230, respectively, for the western and eastern sets of rain gauges.

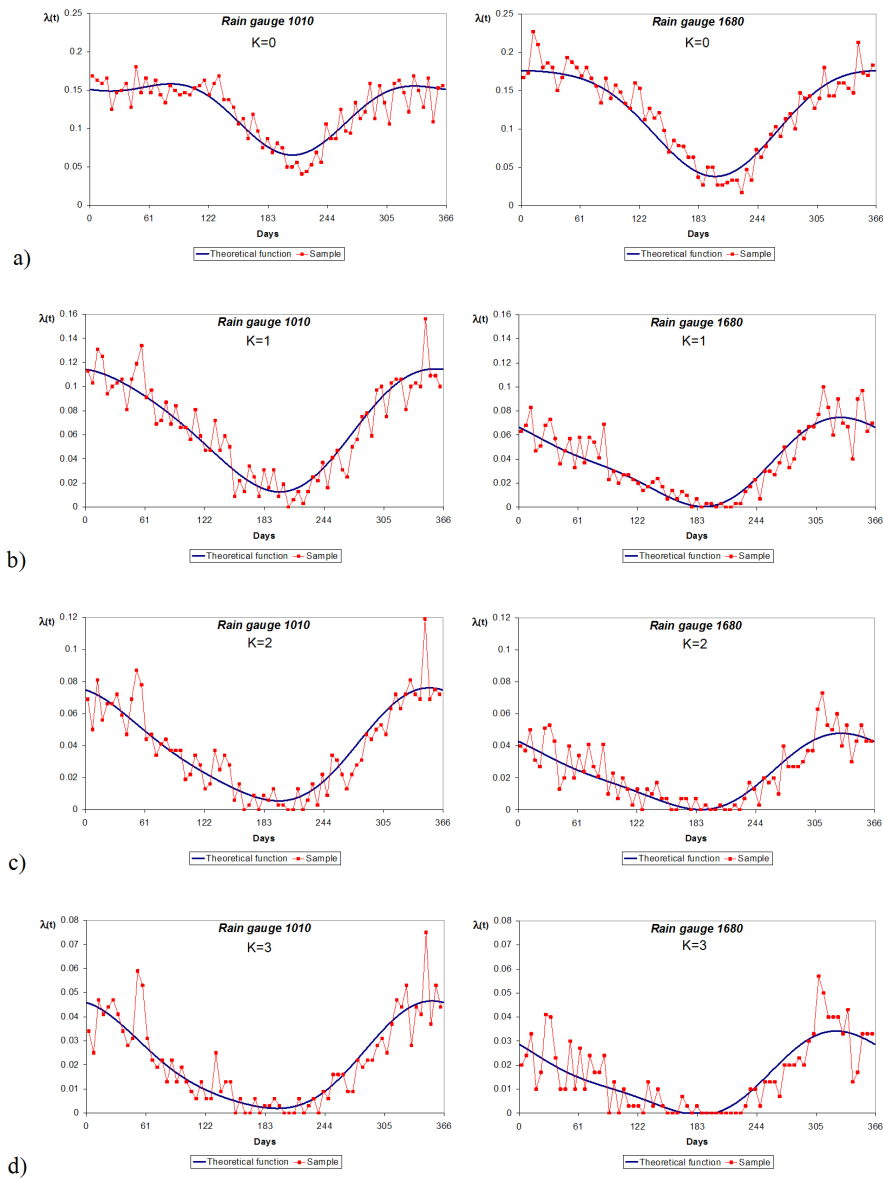


Fig. 7. Cosenza (1010) and Crotona (1680) rain gauge: sampling vs. theoretical distribution of intensity function $\lambda(t)$ by using Fourier series with 2 harmonics, for $K = 0$ (a), $K = 1$ (b), $K = 2$ (c) and $K = 3$ (d).

3.1 Calibration of the non-homogeneous Poisson model

For each prefixed threshold value ($K = 0, 1, 2, 3$), the parameters estimation of the non-homogeneous Poisson model for different numbers of harmonics has been obtained with reference to the observed data during the period 1921–1986 (calibration period) under the hypothesis of unchanged daily rainfall process. The values of the parameters a_0 , $a_k^{(n_h)}$ and $b_k^{(n_h)}$ have been obtained by means of least squares method through Eq. (13) by considering Δt equal to 5 days (Fig. 3 and Table 2). For each threshold, Hald tests showed that $n_h = 2$ satisfies the hypothesis of homogeneity for the transformed occurrence process by using $n_s = 10$ (see Appendix A2).

The result $n_h = 2$ is also confirmed by the application of the cross-validation approach. As examples, Fig. 5 illustrates the CVE evaluation related to Cosenza and Crotona rain gauges, for which the data sample was partitioned into $N = 10$ subsamples.

For each analysed rain gauge and threshold value, Fig. 6 shows the evaluation of $a_0/2$, which represents the expected value of intensity $\lambda(t)$ on the whole period P (see Eqs. 8 and 13). For $K = 0$ (i.e. considering the occurrences of daily rainfall heights greater than 0.1 mm) there are no significant differences between the two zones. Focusing attention on each threshold $K \geq 1$, it can be noticed that the western zone is characterised by higher values with respect to the eastern

Table 2. Parameters estimation for non-homogeneous Poisson model through least square method applied to daily rainfall occurrence process (data belong to the calibration period 1921–1986; Fourier series with $n_h = 2$ harmonics).

Code	Rain gauge	K	a_0	a_1	b_1	a_2	b_2
1010	Cosenza	0	0.253	0.0374	0.0186	-0.0127	-0.0143
		1	0.133	0.0496	0.0057	-0.0018	-0.0058
		2	0.077	0.0341	-0.0014	0.0022	-0.0046
		3	0.043	0.0218	-0.0019	0.0026	-0.0024
1030	S. Pietro in Guarano	0	0.237	0.0394	0.0173	-0.0124	-0.0146
		1	0.144	0.0502	0.0078	-0.0069	-0.0052
		2	0.085	0.0378	-0.0007	0.0012	-0.0043
		3	0.043	0.0207	-0.0022	0.0011	-0.0041
1040	Rende	0	0.251	0.0388	0.0178	-0.0161	-0.0147
		1	0.123	0.0475	0.0056	-0.0036	-0.008
		2	0.073	0.0345	0.0003	0.0019	-0.0032
		3	0.040	0.0223	-0.0018	0.0045	-0.0027
3060	Paola	0	0.254	0.0391	0.0178	-0.0164	-0.0156
		1	0.140	0.0465	0.0015	-0.0053	-0.0079
		2	0.081	0.0304	-0.0018	-0.0017	-0.0077
		3	0.046	0.0183	-0.003	0.0002	-0.008
1590	Belvedere Spinello	0	0.219	0.0463	0.0108	-0.0110	-0.0102
		1	0.077	0.0255	-0.0048	-0.0028	-0.005
		2	0.047	0.0199	-0.0051	-0.0008	-0.0035
		3	0.031	0.0137	-0.0052	-0.0003	-0.0036
1620	Verzino	0	0.255	0.0502	0.0169	-0.0110	-0.0083
		1	0.079	0.0289	-0.0035	-0.0003	-0.0063
		2	0.045	0.0193	-0.0047	-0.0021	-0.0042
		3	0.030	0.0143	-0.0034	-0.0011	-0.0022
1650	Strongoli	0	0.237	0.0611	0.0168	-0.0134	-0.0078
		1	0.075	0.0287	-0.0038	-0.0025	-0.0051
		2	0.046	0.0206	-0.004	-0.0011	-0.004
		3	0.031	0.015	-0.0027	-0.0009	-0.0037
1680	Crotone	0	0.239	0.0663	0.0181	-0.0099	-0.0083
		1	0.075	0.0329	-0.0083	-0.0037	-0.0072
		2	0.046	0.0213	-0.0064	-0.0014	-0.0045
		3	0.031	0.0147	-0.0063	-0.0015	-0.0037

zone (about twice), that is the frequency of heavy daily rainfalls is greater in the western part of the region.

As an example, the fit of $\lambda(t)$ theoretical distribution through the Fourier series with 2 harmonics on sampling occurrence process referred to each threshold value is shown in Fig. 7 for Cosenza and Crotone rain gauges, as the representative stations of western and eastern zones, respectively. Analysing the diagrams, it can be noted that, except for $K = 0$, in the winter period the occurrence intensity $\lambda(t)$ of Cosenza is about twice as high as the Crotone one, confirming what has been shown in Fig. 6.

3.2 Check on occurrence process changes

In order to test the hypothesis that the daily rainfall occurrence process preserves the same behaviour in more recent time periods, the theoretical distribution of intensity function $\lambda(t)$, evaluated for each threshold value in the calibration period 1921–1986 (Sect. 3.1), has been also adopted for the 15-yr observation data collected in the validation period 1987–2001.

For each rain gauge, by using the Monte Carlo approach, 1000 synthetic generations of a 15-yr period of daily rainfall occurrences have been carried out and, for every simulation, sample $\lambda(t)$ has been estimated. This procedure has been adopted because of the complexity of determining analytical

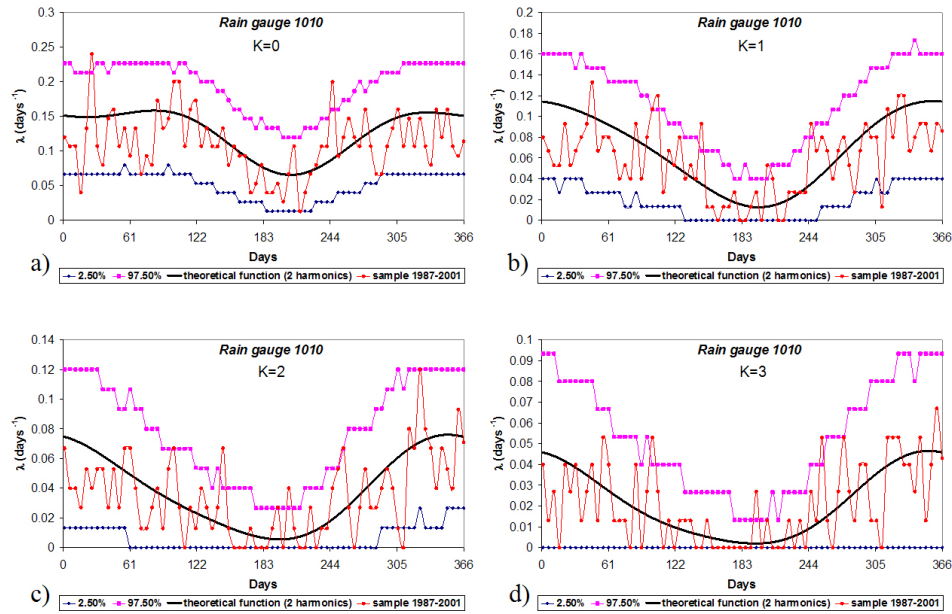


Fig. 8. Cosenza rain gauge: confidence interval of sample intensity function $\lambda(t)$ referred to the validation period for $K = 0$ (a), $K = 1$ (b), $K = 2$ (c) and $K = 3$ (d).

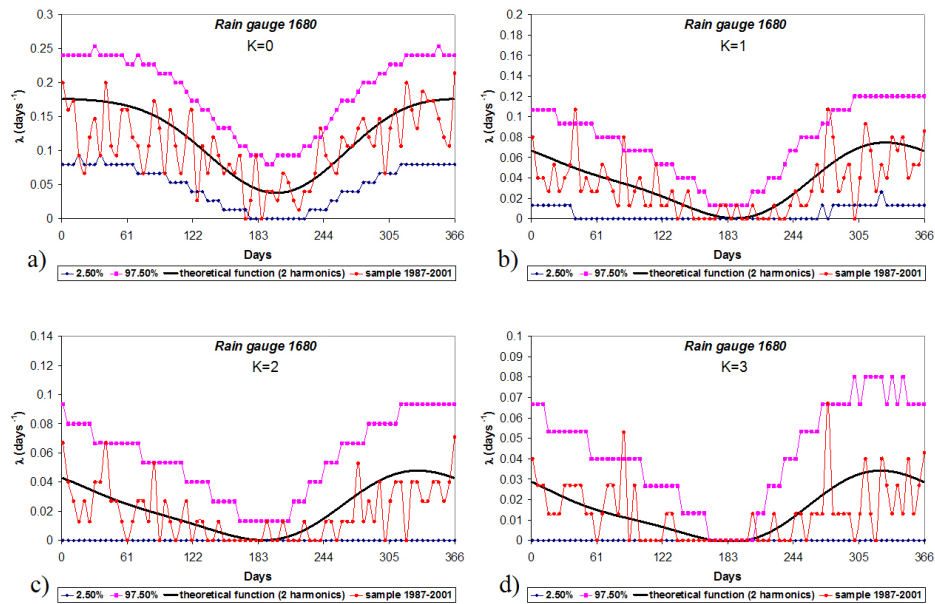


Fig. 9. Crotona rain gauge: confidence interval of sample intensity function $\lambda(t)$ referred to the validation period for $K = 0$ (a), $K = 1$ (b), $K = 2$ (c) and $K = 3$ (d).

statistical confidence intervals referred to the sample intensity $\lambda(t)$.

Sample intensity, theoretical function of the calibration period and 95% confidence interval, evaluated by Monte Carlo approach, are reported, as an example, for one rain gauge of each zone (Cosenza and Crotona) in Figs. 8–9. The differences between the sample and theoretical $\lambda(t)$ appear not

important, as more than 95% of the sample occurrence intensities are inside the statistical bounds evaluated by synthetic generation.

A further test has been performed for each rain gauge: considering the different threshold values, the root-mean-square error (RMSE), evaluated between the theoretical $\lambda(t)$ and the sample of recorded data, and the corresponding 95%

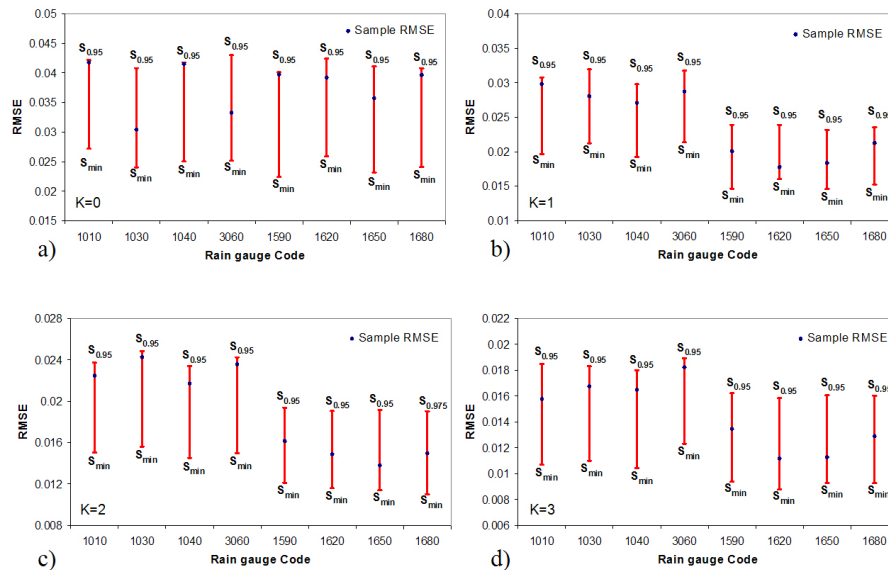


Fig. 10. Sample root-mean-square error (RMSE) and 95% one-tailed confidence intervals, for $K = 0$ (a), $K = 1$ (b), $K = 2$ (c) and $K = 3$ (d).

one-tailed confidence interval, evaluated from the RMSE values between the sample $\lambda(t)$ of each synthetic series and the theoretical one, were obtained and reported in Fig. 10. Analysing these results, it may be observed that the western zone is characterised by differences, between sample $\lambda(t)$ of the 1987–2001 period and theoretical $\lambda(t)$, which are higher than those concerning the eastern zone. This different behaviour between western and eastern zones emerges by considering threshold values $K \geq 1$. However, for each zone and threshold, these deviations are statistically acceptable, as all the sample RMSEs are inside their 95% confidence interval.

As a conclusion, the hypothesis that the occurrence process of the observed rainfall, referred to the validation period, derives from the same $\lambda(t)$ theoretical distribution of the calibration period cannot be rejected at 5% significance level.

4 Conclusions

In this paper, a non-homogeneous Poisson model has been adopted for the stochastic interpretation of the seasonal variability concerning the daily rainfall occurrence process in 8 rain gauges selected among the longest data series in two different zones of the Calabria region (Southern Italy).

The procedure applied to a calibration time interval (1921–1986) shows that a Fourier series with 2 harmonics represents a good fit for explaining the variability of the occurrence intensity function $\lambda(t)$ for all the rain gauges, considering different threshold values (POT analysis). This result is obtained by carrying out both a classical approach concerning cross-validation and a more robust technique related to

the property that a non-homogeneous Poisson process can be transformed into a homogeneous one.

The theoretical distribution so obtained has been adopted to verify possible changes of $\lambda(t)$ function for the validation period (1987–2001), by using the Monte Carlo approach to generate synthetic series of rainfall occurrences. The results showed that the differences between the sample and the theoretical $\lambda(t)$ behaviour do not appear significant. Moreover, a statistical test based on RMSE has shown that the hypothesis of the same $\lambda(t)$ theoretical distribution for calibration and validation periods cannot be rejected at 5% significance level. Therefore, there is no statistically significant evidence of rainfall occurrence process changes for more recent periods in the analysed area.

Further applications of the non-homogeneous Poisson model will concern the joint analyses of the storm occurrence process with the precipitation height marks, interpreted by using a temporally homogeneous model in suitable sub-year intervals.

Appendix A

A1 Theory of Hald tests

Once partitioned the period P into n_s subperiods, on the basis of the function $\Lambda_{t,t+\Delta t}$ (Fig. 2) which depends on the number of harmonics n_h , let $n_{i,j}^{(n_h)}$ be the occurrences related to the j^{th} sub-period ($j = 1, \dots, n_s$) into the i^{th} year ($i = 1, \dots, n_p$), $n_{i,j}^{(n_h)}$ may be assumed as identically Poisson-distributed with parameter:

Table A1. Results from the application of Hald tests.

Code	Rain gauge	K	$n_h = 0$		$n_h = 1$		$n_h = 2$	
			Hald1	Hald2	Hald1	Hald2	Hald1	Hald2
1010	Cosenza	0	37.8%	204.4	57.8%	34.3	100.0 %	4.1
		1	20.0%	456.0	97.8%	8.9	100.0 %	4.3
		2	22.2%	378.7	80.0%	18.8	100.0 %	5.3
		3	24.4%	280.3	86.7%	15.6	100.0 %	11.0
1030	S. Pietro in Guarano	0	37.8%	194.8	62.2%	34.7	100.0 %	3.3
		1	17.8%	380.8	86.7%	14.5	100.0 %	2.0
		2	17.8%	350.9	93.3%	10.4	100.0 %	6.7
		3	22.2%	222.4	97.8%	9.9	100.0 %	8.5
1040	Rende	0	40.0%	192.3	53.3%	41.8	100.0 %	5.6
		1	20.0%	395.0	86.7%	18.0	100.0 %	4.9
		2	20.0%	339.5	86.7%	14.5	100.0 %	7.1
		3	26.7%	267.8	97.8%	10.0	100.0 %	4.6
3060	Paola	0	37.8%	186.0	55.6%	37.8	100.0 %	3.5
		1	20.0%	321.3	88.9%	16.6	100.0 %	4.8
		2	22.2%	242.7	80.0%	17.5	100.0 %	2.2
		3	26.7%	175.7	91.1%	14.0	100.0 %	1.7
1590	Belvedere Spinello	0	28.9%	229.7	62.2%	37.3	100.0 %	10.0
		1	22.2%	181.2	93.3%	10.8	100.0 %	3.6
		2	24.4%	183.3	97.8%	6.4	100.0 %	4.7
		3	35.6%	149.8	68.9%	23.1	100.0 %	11.0
1620	Verzino	0	31.1%	225.6	84.4%	15.4	100.0 %	5.5
		1	22.2%	213.6	88.9%	13.9	100.0 %	6.2
		2	28.9%	181.4	88.9%	13.0	100.0 %	7.9
		3	31.1%	155.7	93.3%	11.5	100.0 %	7.6
1650	Strongoli	0	24.4%	356.1	66.7%	30.7	100.0 %	2.3
		1	20.0%	231.7	93.3%	13.5	100.0 %	9.0
		2	24.4%	199.9	93.3%	12.8	100.0 %	9.1
		3	26.7%	159.1	97.8%	7.8	100.0 %	4.2
1680	Crotone	0	20.0%	475.4	68.9%	30.6	100.0 %	9.7
		1	13.3%	376.2	88.9%	14.3	100.0 %	6.7
		2	22.2%	262.3	97.8%	8.6	100.0 %	6.0
		3	20.0%	213.3	84.4%	20.9	100.0 %	11.5

$$\Lambda_j^{(n_h)} = \int_{w_{I,ij}}^{w_{F,ij}} \lambda^{(n_h)}(x) dx \tag{A1}$$

where $w_{I,ij}$ and $w_{F,ij}$ are, respectively, initial and final values of sub-period j in the original time scale. If we consider the sum of the occurrences:

$$S_j^{(n_h)} = \sum_{i=1}^{n_P} n_{i,j}^{(n_h)} \quad j = 1, 2, \dots, n_s \tag{A2}$$

for each of the joint set of sub-periods n_s , it can be hypothesized that variable $S_j^{(n_h)}$ is yet Poisson-distributed with parameters:

$$\theta_j^{(n_h)} = n_P \Lambda_j^{(n_h)} \tag{A3}$$

This hypothesis can be verified by means of two homogeneity tests both proposed by Hald (1952). The first test compares the values of the means $\theta_{j_1}^{(n_h)}$ and $\theta_{j_2}^{(n_h)}$ for any chain of period j_1 and j_2 . The test is verified if all comparisons between couples $j_1 = 1, 2, \dots, n_{s-1}$ and $j_2 = j_1 + 1, \dots, n_s$ are fulfilled. Statistically, the single hypothesis $\theta_{j_1}^{(n_h)} = \theta_{j_2}^{(n_h)}$ has to be rejected if:

$$F_{\text{sample}} = \frac{S_{\text{max}}^{(n_h)}}{1 + S_{\text{min}}^{(n_h)}} \geq F_{1-\alpha_0/2}(f_1, f_2) \tag{A4}$$

where $S_{\text{max}}^{(n_h)}$ and $S_{\text{min}}^{(n_h)}$ are maximum and minimum values of couple $[S_{j_1}^{(n_h)}, S_{j_2}^{(n_h)}]$, and $F_{1-\alpha_0/2}(f_1, f_2)$ is the $1-\alpha_0/2$

value of Fisher's distribution with freedom degrees $f_1 = 2[1 + S_{\min}^{(n_h)}]$ and $f_2 = 2S_{\max}^{(n_h)}$.

The second test concerns the verification of contemporary equal values of the means $\theta_j^{(n_h)}$ for all the sub-periods $j = 1, 2, \dots, n_s$. The test can be applied if $n_p > 5$ and $S_j^{(n_h)} > 5$.

The hypothesis $\theta_{j_1}^{(n_h)} = \theta_{j_2}^{(n_h)} = \dots = \theta_{n_s}^{(n_h)}$ has to be rejected if:

$$\chi_{\text{sample}}^2 = n_p \frac{m'_{2S} - m'_{1S}{}^2}{m'_{1S}} \geq \chi_{1-\alpha_0}^2(f) \quad (\text{A5})$$

where $m'_{m_s} = \frac{1}{n_p} \sum_{j=1}^{n_p} [S_j^{(n_h)}]^m$ and $\chi_{1-\alpha_0}^2(f)$ is the $1-\alpha_0$ value of χ^2 distribution with $f = n_s - 1$ degrees of freedom.

A2 Results of Hald tests

Results obtained from the application of Hald tests are reported in Table A1, in which α_0 is set equal to 0.05. The percentage of couples satisfying the property $F_{\text{sample}} < F_{1-\alpha_0/2}(f_1, f_2)$ (see Eq. A4) is indicated with "Hald1" ($n_s = 10$ subperiods are considered in this context), while "Hald2" is related to value of χ_{sample}^2 (see Eq. A5) which must be less than $\chi_{1-\alpha_0}^2(f)$, equal to 16.9. By analysing the results, it is clear that the adoption of $n_h = 2$ harmonics satisfies both the statistical tests for each considered threshold K and, consequently, the transformed occurrence process may be considered as homogeneous.

Acknowledgements. The authors would like to thank the Editor and the reviewers for their useful suggestions that led to an improvement of the paper. This work has been developed as a research within the UNESCO MED-FRIEND project.

Edited by: A. Mugnai

Reviewed by: two anonymous referees

References

- Balachandrudu, N., DelSole, T., and Tippett, M. K.: Optimal Estimation of the Climatological Mean, *J. Climate*, 22, 4845–4859, 2009.
- Chandler, R. E. and Wheeler, H. S.: Climate change detection using generalized linear models for rainfall—a case study from the West of Ireland, *Water Resour. Res.*, 38(10), 1192, doi:10.1029/2001WR000906, 2002.
- Chang, T. J., Kavvas, M. L., and Delleur, J. W.: Daily precipitation modeling by discrete autoregressive moving average processes, *Water Resour. Res.*, 20(5), 565–580, 1984.
- Cowpertwait, P. S. P.: A spatial-temporal point process model with a continuous distribution of storm types, *Water Resour. Res.*, 46, W12507, doi:10.1029/2007WR006110, 2010.
- Cowpertwait, P. S. P., Isham, V., and Onof C.: Point process models of rainfall: developments for fine-scale structure, *Proc. R. Soc. London, Ser. A*, 463, 2569–2587, 2007.
- Cox, D. and Isham, V.: *Point processes*, Chapman & Hall, London, UK, 1980.
- Cressie, N. A. C.: *Statistics for Spatial Data*, Wiley, New York, NY, 1993.
- Entekhabi, D., Rodriguez-Iturbe, I., and Eagleson P.: Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation, *Water Resour. Res.*, 25, 295–302, 1989.
- Gupta, V. K. and Waymire, E. C.: A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteor.*, 32, 251–267, 1994.
- Hald, A.: *Statistical theory with engineering applications*, Wiley, New York, NY, 1952.
- Kavvas, M. L. and Delleur, J. W.: A stochastic cluster model of daily rainfall sequences, *Water Resour. Res.*, 17(4), 1151–1160, 1981.
- Lang, M., Ouarda, T. B. M. J., and Bobée, B.: Towards operational guidelines for over-threshold modelling, *J. Hydrol.*, 225, 103–117, 1999.
- Leonard, M., Lambert, M. F., Metcalfe, A. V., and Cowpertwait, P. S. P.: A space-time Neyman-Scott rainfall model with defined storm extent, *Water Resour. Res.*, 44, W09402, doi:10.1029/2010WR009728, 2008.
- Lovejoy, S. and Schertzer, D.: Multifractals, universality classes, and satellite and radar measurements of cloud and rain fields, *J. Geophys. Res.*, 95, 2021–2031, 1990.
- Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheeler, H. S., and Isham, V.: Rainfall modelling using Poisson-cluster processes: a review of developments, *Stoch. Environ. Res. Risk Assess.*, 14, 384–411, 2000.
- Picard, R. and Cook, D.: Cross-validation of regression models, *J. American Statistical Association*, 79(387), 575–583, 1984.
- Pujol, N., Neppel, L., and Sabatier, R.: Regional tests for trend detection in maximum precipitation series in the French Mediterranean region, *Hydrol. Sci. J.*, 52(5), 956–973, 2007.
- Rodriguez-Iturbe, I., Gupta, V., and Waymire, E.: Scale considerations in the modeling of temporal rainfall, *Water Resour. Res.*, 20(11), 1611–1619, 1984.
- Rodriguez-Iturbe, I., Febres De Power, B., and Valdes, J. B.: Rectangular pulses process models for rainfall: analysis of empirical data, *J. Geophys. Res.*, 92(D8), 9645–9656, 1987.
- Sirangelo, B.: Un modello poissoniano non omogeneo per l'analisi delle piogge cumulate su lunghi intervalli temporali, *Idrotecnica*, 3, 125–158, 1994 (in Italian).
- Smith, J. A. and Karr, A. F.: A point process model of summer season rainfall occurrences, *Water Resour. Res.*, 19(1), 95–103, 1983.
- Snyder, D. L.: *Random point process*, Wiley, New York, NY, 1975.
- Stern, R. D. and Coe, R.: A model fitting analysis of daily rainfall data, *J. Roy. Statist. Soc.*, 147(1), 1–34, 1984.
- Waymire, E. and Gupta, V. K.: The mathematical structure of rainfall representations, 1: a review of the stochastic rainfall models, *Water Resour. Res.*, 17(5), 1261–1272, 1981a.
- Waymire, E. and Gupta, V. K.: The mathematical structure of rainfall representations, 2: a review of the theory of point processes, *Water Resour. Res.*, 17(5), 1273–1285, 1981b.
- Waymire, E. and Gupta, V. K.: The mathematical structure of rainfall representations, 3: some applications of the point process theory to rainfall processes, *Water Resour. Res.*, 17(5), 1287–1294, 1981c.

- Wheater, H., Chandler, R., Onof, C., Isham, V., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., and Segond, M.-L.: Spatial-temporal rainfall modelling for flood risk estimation, *Stoch. Environ. Res. Risk Assess.*, 19, 403–416, 2005.
- Yevjevich, V. and Dyer, T. G. J.: Basic structure of daily precipitation series, *J. Hydrol.*, 64, 49–67, 1983.