



Statistical analysis of inter-arrival times of rainfall events for Italian Sub-Alpine and Mediterranean areas

C. Agnese¹, G. Baiamonte¹, C. Cammalleri², D. Cat Berro³, S. Ferraris⁴, and L. Mercalli³

¹Università degli Studi di Palermo, Dip. SAGA, Palermo, Italy

²Università degli Studi di Palermo, DICAM, Palermo, Italy

³Società Meteorologica Italiana, Castello Borello, Bussoleno (TO), Italy

⁴Università degli Studi di Torino, DIST, Torino, Italy

Correspondence to: C. Cammalleri (carmelo.cammalleri@unipa.it)

Received: 15 January 2012 – Revised: 28 August 2012 – Accepted: 30 August 2012 – Published: 17 September 2012

Abstract. In this work a set of time-series of inter-arrival times of rainfall events, at daily scale, was analysed, with the aim to verify the issue of increasing duration of dry periods. The set consists of 12 time-series recorded at rain gauges in 1926–2005, six of them belong to an Italian Sub-Alpine area (Piedmont) and six to a Mediterranean one (Sicily). In order to overcome the problem related to limited sample size for high values of inter-arrival times, the discrete probability polylog-series distribution was used to fit the empirical data from partial (20 yr) time-series. Moreover, a simple qualitative trend analysis was applied to some high quantiles of inter-arrival times as well as to the average extent of rain clusters. The preliminary analysis seems to confirm the issue of increasing duration of dry periods for both environments, which is limited to the “cold” season.

1 Introduction

Rainfall modelling represents an important issue in several research areas, such as hydrology, meteorology and climatology, because of both theoretical and practical implications. In particular, the assessment of the probabilistic structure of rainy days occurrence allows estimating the probability of both rainy and dry periods.

One interesting way to characterize the temporal structure of rainfall is the use of inter-arrival times series, which is constituted by the succession of the times, T , elapsed between a rainy day and the immediately preceding one (Waymire and Gupta, 1981). Because of some typical properties of rainfall, as rainy days clustering and dry periods persistence, the observed T -samples usually present some statistical peculiarities: very-high variability and skewness, relative high frequency associate to $T = 1$, “monotonically” decreasing frequencies with sporadic sampling for high values of T .

The main purpose of this work is to find a sign of an existing trend for some high T -quantiles (long dry periods). While a large sample is already available for some yearly-averaged features (i.e., the mean inter-arrival time), a con-

sistent sample for high T -quantiles can be obtained only by considering a rainfall record of several years of observation; a sampling window of 20-yr has been judged large enough to fulfil this purpose. However, a time-series of any high T -quantile obtained following such procedure contains few elements only, because of the limited length of available rainfall records (≈ 100 yr of observations). In order to overcome this problem, a partial time-overlap between two successive 20-yr sampling windows has been assumed. The introduction of this overlap allows obtaining more reliable time-series, but it makes tricky the detecting of possible trends using the procedures commonly adopted for yearly-averaged features.

Sample T 's-distributions have been fitted by (discrete) polylog-series probability distributions; that allowed defining un-biased high T -quantiles, removing the difficulty caused by “sparse sampling”. Besides, the use of theoretical distributions instead of empirical ones provides a more robust analysis tool. Furthermore, in the context of eco-hydrology, modelling of rainfall process at daily scale enables to infer statistics for some derived processes, as the mean soil moisture and/or the actual evapotranspiration, allowing us to analyse possible trends.

A set of 12 long time-series of inter-arrival times was analysed, six of them belong to an Italian Sub-Alpine area (Piedmont) and six belong to a Mediterranean one (Sicily); these two areas are characterized by different rainfall regimes and seasonality. Time-series were obtained by daily rainfall data recorded in the period 1926–2005, subdivided in 13 sub-sets of 20-yr duration progressively shifted of 5-yr. Time-series of high quantiles of associate probability distribution and theoretically-derived average cluster size were analysed for detecting possible temporal trend.

2 Fitting sample distribution of inter-arrival times through polylog-series probability distribution

Assuming that T-values are independent and identically distributed (i.i.d.), them can be studied as random variables that generate a renewal process (Feller, 1968), which can be seen as a generalisation of the Poisson process (implying an exponential inter-arrival times distribution). Several studies already suggested to interpret the rainfall occurrence process as a renewal process (see for example: Buishand, 1978; Foufoula-Georgiou and Lettenmaier, 1987); however, because seasonality character is frequently present in T-series, the renewal property was generally applied to each homogeneous period, distinctly. In order to describe statistical peculiarities of observed T-samples, some discrete probability distributions have been tested: the log-series distribution and the 2-parameter Extended-Yule (Martínez-Rodríguez et al., 2011) and polylog-series distributions (Gupta et al., 2008). The 2-parameter polylog-series distribution allowed to capture the main characters of T's-samples (clustering, very high variability and skewness, persistence) with better accuracy than the other distributions; accordingly, following analyses referred to this distribution only, which is defined as:

$$P\{T = k\} = \frac{w^k}{k^s Li_s(w)} \quad (1)$$

where w and s are the parameters of the polylog-series distribution, $Li_s(\cdot)$ is the polylogarithm (or Jonquière's) function:

$$Li_s(w) = \sum_{k=1}^{\infty} \frac{w^k}{k^s} \quad (2)$$

It can be observed that for $s = 1$, the Jonquière's function (2) reduces to $-\ln(1 - w)$, and the polylog-series distribution reduces to the logarithm-series distribution (Feller, 1968), often used in the past in order to regularise dry spell data (Chatfield, 1966; Kottogoda and Rosso, 1997).

The hypothesis that T s are i.i.d. also implies that the expected cluster size, $\langle cl \rangle$, can be derived by the following simple relationship:

$$\begin{aligned} \langle cl \rangle &= \sum_{k=1}^{\infty} k P_1^{k-1} (1 - P_1) = (1 - P_1) (1 + 2P_1 + \dots + kP_1^{k-1} + \dots) \\ &= \sum_{k=0}^{\infty} P_1^k = \frac{1}{1 - P_1} \end{aligned} \quad (3)$$

where $P_1 = P\{T = 1\}$. Thus, a good estimate of the observed average cluster size mainly depends on the capability of the distribution to well reproduce $P\{T = 1\}$.

The estimate of the two parameters was performed by means of the maximum likelihood method, which consists of the solution of a system of two equations in the unknown w and s (for the details see Gupta et al., 2008). Goodness of fit was tested according to Martínez-Rodríguez et al. (2011); these authors highlight that the classical chi-square test is not appropriate to evaluate the goodness-of-fit for distributions characterized by long tail (as polylog-series one), because expected frequencies less than 5 are common. Also, it has to be pointed out that grouping the classes with expected frequencies less than 5 could lead to some problems (Spierdijk and Voorneveld, 2007). So, an alternative procedure to test the goodness of fit was adopted by numerically reconstructing the null hypothesis of chi-square (according to polylog-series) on the base of a Monte Carlo simulation. A set of 2000 replicates of the observation was produced by randomly sampling from the theoretical discrete distribution (using Monte Carlo approach), and the p-value associated to the actual observation was computed from the numerically reconstructed null hypothesis (Martínez-Rodríguez et al., 2011).

3 Dataset

A set of 12 time-series of inter-arrival rainfall times was obtained by daily rainfall data recorded in the period 1926–2005 by 12 rain gauges. As it is shown in Fig. 1, 6 stations are located in the Sicily Island (hereafter referred as SIC), which is characterized by a typical Mediterranean climate, whereas the other 6 stations are placed in a sub-Alpine region including Piedmont and Aosta Valley Italian regions (from here on referred to as PMT). The 12 series were selected in order to satisfy the following criteria: (i) minimum gaps in the 80-year recorded data (no gap filling was performed); (ii) homogeneity of the time series. Relatively to the latter point, both control of data quality and specific homogeneity tests have been carried out; in particular, the procedure suggested by Wijngaard et al. (2003) has been followed. Each time-series has been subdivided in 13 sub-sets of 20-yr duration; the first subset refers to the period 1926–1945, the other subsets were obtained by progressively shifting of 5-yr the sampling, then obtaining an overlap of 15 yr between two successive series (1931–1950, 1936–1955, ..., 1986–2005). Moreover, the analysis was performed for three different datasets separately: (1) whole year (Y); (2) six-months relatively “warm” season (WS – from April to September) and (3) six-months relatively “cold” season (CS – from October to March). As it

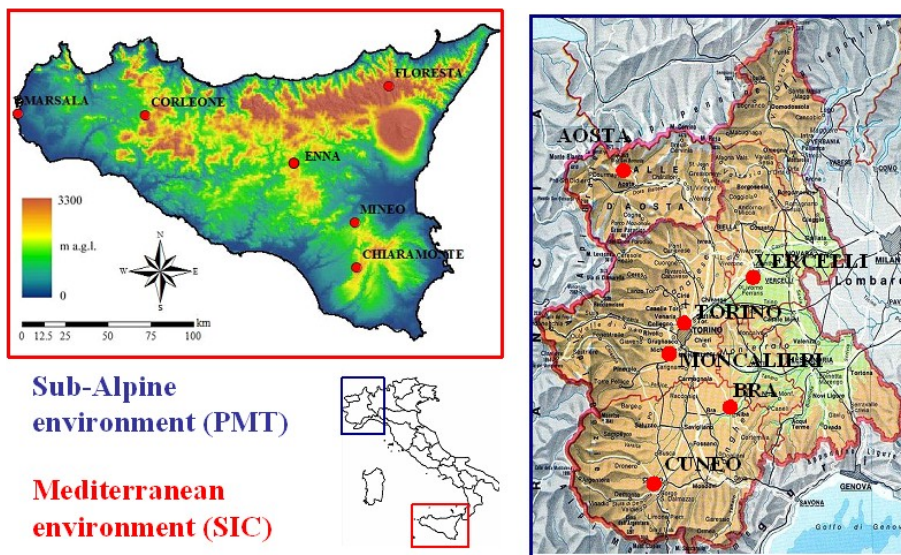


Figure 1. Spatial location of the rain gauges that provided rainfall time series used for the analysis. Left panel reports the 6 stations located in the Mediterranean area (SIC), whereas right panel reports the 6 stations located in the sub-Alpine (PMT) area.

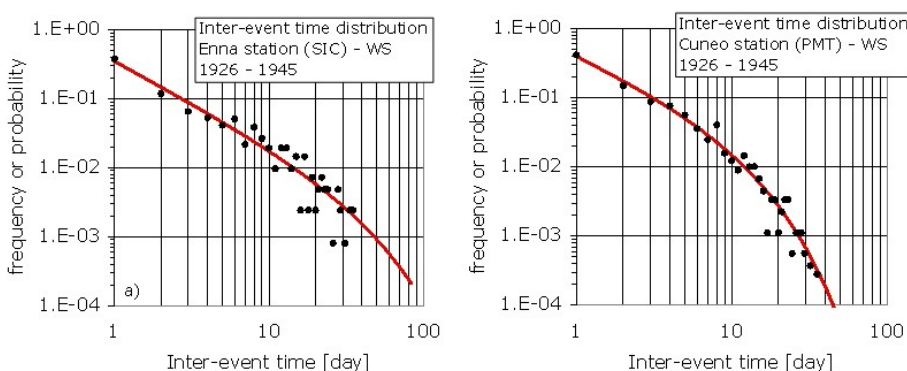


Figure 2. Fitting frequency distribution of inter-event times, T , with the polylog-series distribution, for the warm season and for Enna (SIC – left panel) and Cuneo stations (PMT – right panel). Sample data refer to the period 1926–1945.

is known, Sicily’s rainfall regime is characterised by a strong seasonality, with a “warm” (dry) and a “cold” (wet) season. During the WS, rainfall is usually lower than atmospheric demand, whereas the converse is true during the CS. Rainfall regime in Piedmont and Aosta Valley is almost virtually opposite to the one in Sicily; indeed, WS is generally wetter than CS. In this phase, no additional disaggregation has been considered, mainly to not further reduce the sample size for high T -values (or for not increase the length of sampling window); therefore, similar analyses for three-month seasons have not been performed.

4 Results

4.1 Analysis of inter-arrival times series distribution

The polylog-series distribution was fitted to the 468 available datasets (12 stations \times 3 periods \times 13 sub-series). Test

of goodness of fit was performed separately for each data set, rejecting the null hypothesis when the p -value associated to the fitted distribution was < 0.05 .

Synthetically, the chi-square test results indicate that the null hypothesis is frequently rejected for the yearly datasets for both SIC and PMT stations (74.4 % for SIC and 78.2 % for PMT), while only in few cases it is rejected for WS and CS data sets (14.1 % and 26.9 % for WS-SIC and WS-PMT, respectively; 18.2 % and 27.1 % for CS-SIC and CS-PMT, respectively). These results generally show how polylog distribution better agrees with SIC seasonal datasets than with PMT ones.

Figure 2 shows two examples of data fitting with polylog-series distribution; for both cases the null hypothesis has not been rejected and it can be seen how the visual agreement results satisfactory.

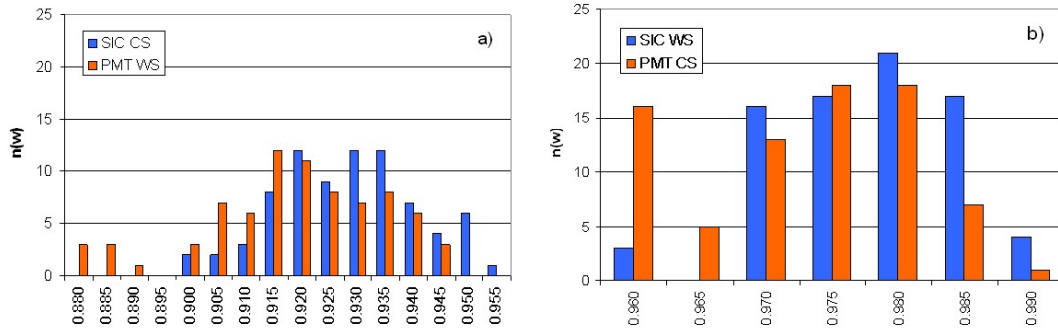


Figure 3. Frequency distribution of w parameter for the two studied areas (SIC and PMT), for warm season (WS) and cold season (CS).

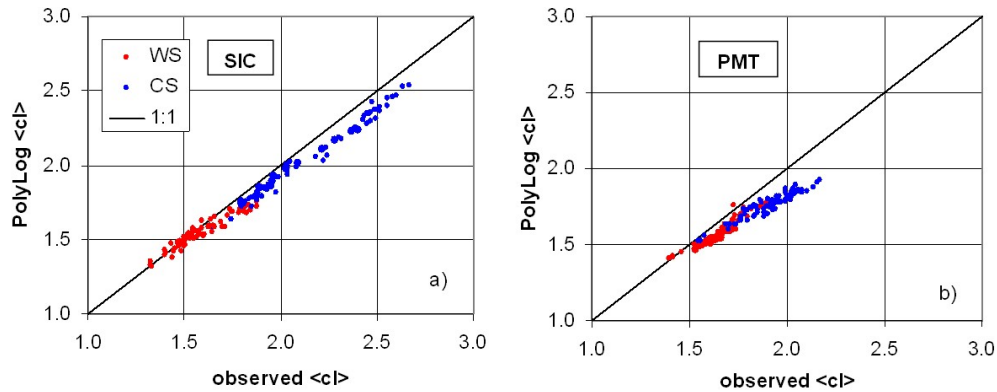


Figure 4. Comparison between observed and estimated average cluster size $\langle cl \rangle$ for the Mediterranean (SIC, panel a) and sub-Alpine (PMT, panel b) areas in both cold (CS) and warm (WS) season.

A preliminary analysis of range of variability of distribution parameters w and s has been carried out. In particular, Fig. 3 reports the frequency distributions of w parameter, which is the one that mainly influences tail probabilities. Interestingly, w distributions of Sicily CS and Piedmont WS (Fig. 3a), as well as of Sicily WS and Piedmont CS (Fig. 3b), are rather similar in both range of variability and frequency distribution; that is likely due to the typical features of the two studies areas, which have somewhat opposite rainfall regimes.

4.2 Average cluster size

As previously introduced, the use of inter-arrival times allows characterizing the temporal structure of both rainy days and dry periods. A further test of the polylog fitting was performed in terms of average cluster size, computed with Eq. (3). In Fig. 4 observed vs. estimated $\langle cl \rangle$ are represented for the stations located in the two areas (SIC, Fig. 4a and PMT, Fig. 4b) and for the two periods (WS and CS). These plots highlight in general a good performance of the model, with a slightly systematic underestimation ($< 4\%$ on average), that becomes significant only for some recorded high values (occurred during CS for both SIC and PMT). Moreover, a significant seasonality for $\langle cl \rangle$, as well as a quite wide

range of variability (from 1.2 to 2.7 days), was observed for SIC region. On the contrary, in PMT sites the observed range of variability of $\langle cl \rangle$ is quite similar for the two periods, with a wide overlap range from 1.5 to 2.0 days, and with values exceeding 2 only for CS period. This further confirm how the seasonality is more evident in SIC area than in PMT one in term of clustering of rainy days.

4.3 Modelling of high quantiles

Table 1 reports root mean square difference (RMSD) statistics computed for some high T-quantiles (0.80, 0.85, 0.90 and 0.95) by comparing data observed and polylog-series estimates; for both investigated area, available sample size is equal to 78 (13 periods \times 6 stations). As can be noticed, RMSD values are generally very small, less than three days, with just few exceptions (SIC WS for $P \leq 0.95$; PMT CS for $P \leq 0.90$). The greatest discrepancies observed for SIC-WS could be likely explained by the reduction of the sample size with the increase of the quantiles considered. Similar considerations arise from the analysis of Fig. 5, where observed and estimated quantiles are compared for two stations (Corleone – WS and Torino – CS).

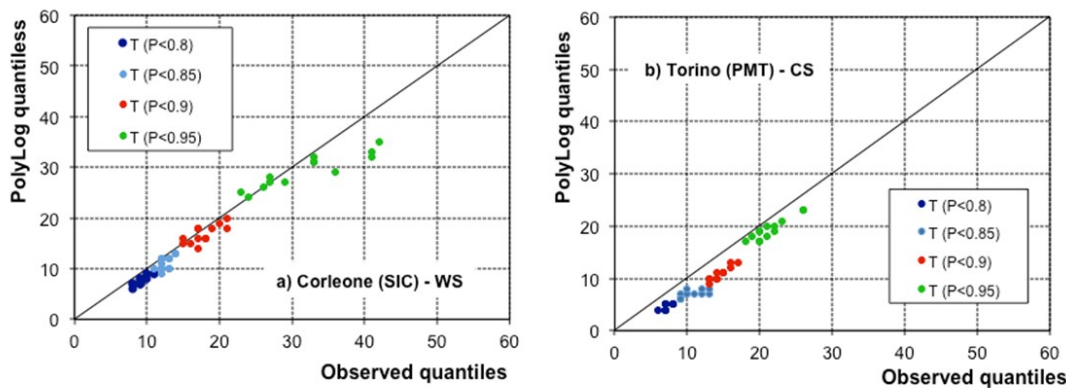


Figure 5. Comparison between some observed and estimated high T-quantiles for two sites: (a) Corleone station in SIC during WS; (b) Torino station in PMT during CS.

Table 1. Mean (M) and root mean square difference (RMSD) for some quantiles, for both seasons and both investigated areas.

M (RMSD) in days		$T(P < 0.80)$	$T(P < 0.85)$	$T(P < 0.90)$	$T(P < 0.95)$
PMT	WS	6.9 (1.65)	8.6 (1.58)	11.1 (1.66)	15.1 (1.02)
	CS	7.5 (2.34)	10.3 (2.84)	14.2 (3.12)	21.4 (2.20)
SIC	WS	10.7 (1.83)	14.3 (1.79)	19.6 (1.72)	33.7 (4.41)
	CS	4.5 (1.47)	5.9 (1.57)	7.9 (1.66)	11.5 (1.47)

Table 2. Summary of the qualitative trend analysis carried out for a T-quantile ($P < 0.975$) and for the average cluster size. The symbol (+) indicates positive trend, (-) indicates negative trend and (n) no-trend.

Area	Station	$P < 0.975$		$\langle cl \rangle$	
		WS	CS	WS	CS
SIC	Chiaramonte	-	n	+	+
	Corleone	-	+	n	-
	Enna	n	+	-	-
	Floresta	n	n	+	n
	Marsala	n	+	n	-
PMT	Mineo	n	+	-	n
	Aosta	n	+	n	n
	Bra	n	+	n	n
	Cuneo	+	+	-	-
	Moncalieri	n	+	n	n
	Torino	n	+	n	n
	Vercelli	+	n	n	n

4.4 Preliminary analysis of the trends

The results reported in the previous sub-sections synthetically showed: (i) the polylog distribution usually provides a good fitting of sample T distribution for seasonal (WS and CS) data; (ii) the polylog distribution is able to well estimate the average cluster sizes as well; (iii) RMSD for some high T-quantiles (0.80, 0.85, 0.90 and 0.95) are generally low (with the above-mentioned exceptions). Accordingly, both estimated high T-quantiles and average cluster sizes were considered a more robust mean to analyse the trend of such quantities than empirical ones. Moreover, the analysis can be also extended to very high quantiles, for which sample size is small.

A simple qualitative trend analysis of the quantiles was performed in order to investigate the issue of increasing duration of dry periods, of which previous studies have been dealt with yet (Alpert et al., 2002; Dai et al., 1998); due to the overlapping window used in the reconstruction of samples, the analysis of trend cannot be performed with the usual methods, which are based on the hypothesis of independence among the single values of the series. For this reason, the analysis of trend is only qualitative at this point.

As synthetically shown in Table 2, a slightly increasing trend of very high T-quantile ($P < 0.975$) has been observed for almost all the PMT and SIC stations during CS periods. In general, no trend was observed for the WS period. It is noteworthy that no significant differences between the two

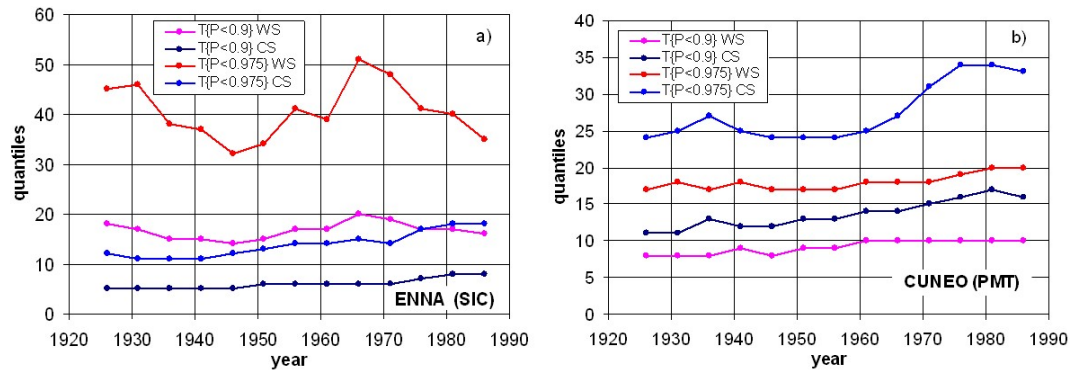


Figure 6. Time-series of two high quantiles $T\{P < 0.9\}$ and $T\{P < 0.975\}$, estimated by polylog-series distribution, for both season and for two sites: (a) Enna station (SIC) and (b) Cuneo station (PMT).

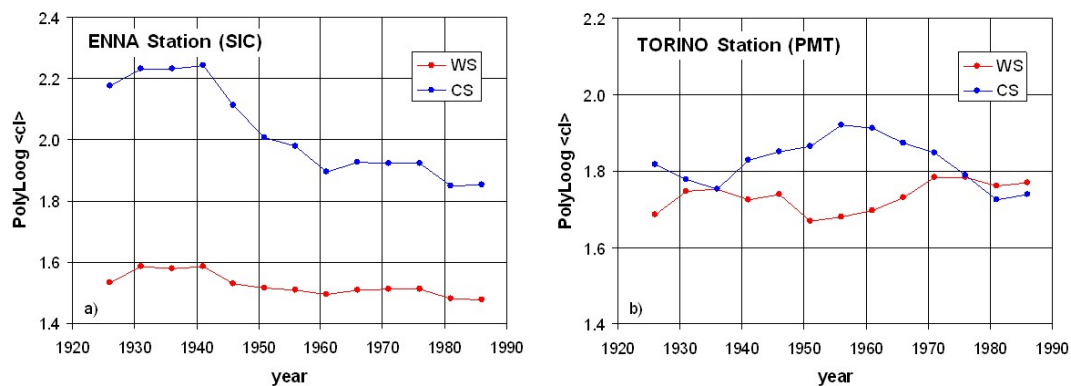


Figure 7. Time-series of expected cluster size estimated by polylog-series distribution, for both season and for two sites: (a) for Enna station (SIC); (b) for Torino station (PMT).

study areas were appreciated. As an example, in Fig. 6, time series of two estimated T-quantiles ($P \leq 0.9$, $P < 0.975$), for two stations, Enna (SIC) and Cuneo (PMT), are reported.

A similar qualitative trend analysis has also been performed for the average cluster size (Table 2). Generally, no trend was observed for Piedmont, while contrasting results were observed for Sicily. As an example, Fig. 7 shows the time-series of the expected cluster size for two sites, Enna (SIC) and Torino (PMT); a noticeable decreasing trend was observed for Enna, especially for CS.

5 Conclusions

The capability to reproduce the inter-arrival times statistical structure through a 2-parameter discrete distribution (polylog-series) was tested in two Italian regions: a Mediterranean area and a Sub-Alpine one.

A chi-square test based on the Montecarlo procedure suggests that polylog-series distribution seems to well reproduce the observed data in the whole range of variability of times and associated frequencies when the dataset is split in two sub-period: a relatively “warm” season (April–September)

and a relatively “cold” one (October–March). Poor performances were instead obtained when the whole year is analysed, mainly due to the strong seasonality characterising the two study areas.

The theoretical distribution allows also reproducing well some characteristics of rainfall as: average cluster sizes and high T-quantiles. In particular, the average cluster sizes results only slightly underestimated ($\cong 5\%$), whereas high quantiles ($P \leq 0.8$ to 0.95) are generally well reproduced with the only exception for $P \leq 0.95$ during WS (in SIC) and CS (in PMT). This latter result can be justified by the small sample size of empirical data for high inter-arrival times.

A preliminary qualitative analysis of the trends on high quantiles and average cluster size seems to confirm the issue of increasing duration of dry periods for both environments, which is limited to the “cold” season.

No trend was instead observed for the average cluster size in Sub-Alpine areas, while discordant trends were observed in Mediterranean stations.

The polylog-series seems to account for peculiarities of inter-arrival times of the two areas by simple dividing the year in two six-month periods (a relatively “warm” season from April to September and a relatively “cold” season

from October to March). This result is promising for the use of polylog-series in ecohydrological models in those areas where classical exponential relationship fails to capture the characteristics of rainfall. However, in order to better catch the differences between the characters of seasonality in different climatic regions, the same procedure followed in this work could be applied to shorter periods (three months). Additionally, the requirement of several years to sufficiently sample high inter-arrival times (and the necessary overlapping between consecutive sampling windows) requires ad-hoc methodologies for the quantitative analysis of the trends that will be investigated in the next future. Finally, the possibility to address the observed limits of the adopted distribution by using a three-parameter Lerch distribution (which generalises the polylog-series distribution) have to be tested in future researches.

Acknowledgements. The authors thank Osservatorio delle Acque of the Assessorato Energia e Servizi di Pubblica Utilità della Regione Siciliana and ARPA (Agenzia Regionale per la Protezione dell’Ambiente) Piemonte for providing the meteorological dataset.

Edited by: I. Auer

Reviewed by: T. Nanni and three anonymous referees



The publication of this article is sponsored by the European Meteorological Society.

References

- Alpert, P., Ben-Gai, T., Baharad, A., Benjamin, J., Yekutieli, D., Colacino, M., Diodato, L., Ramis, C., Homar, V., Romero, R., Michaelides, S., and Manes, A.: The paradoxical increase of Mediterranean extreme daily rainfall in spite of decrease in total values, *Geophys. Res. Lett.*, 29, 1536, doi:10.1029/2001GL013554, 2002.
- Buishand, T. A.: Some remarks on the use of daily rainfall models, *J. Hydrol.*, 36, 295–308, doi:10.1016/0022-1694(78)90150-6, 1978.
- Chatfield, C.: Wet and dry spells, *Weather*, 21, 308–310, 1966.
- Dai, A., Trenberth, K. E., and Karl, T. R.: Global variations in droughts and wet spells: 1900–1995, *Geophys. Res. Lett.*, 25, 3367–3370, doi:10.1029/98GL52511, 1998.
- Feller, W.: An Introduction to probability theory and its applications, Vol. 1, J. W. A. Sons, New York, USA, 1968.
- Foufoula-Georgiou, E. and Lettenmaier, D. P.: A Markov renewal model for rainfall occurrences, *Water Resour. Res.*, 23, 875–884, doi:10.1029/WR023i005p00875, 1987.
- Gupta, P. L., Gupta, R. C., S.-huat Ong, S.-H., and Srivastava, H. M.: A class of Hurwitz-Lerch Zeta distributions and their applications in reliability, *App. Math. Comp.*, 196, 521–531, doi:10.1016/j.amc.2007.06.012, 2008.
- Kottogoda, N. T. and Rosso, R.: Statistics, probability, and reliability for civil and environmental engineers, McGraw-Hill, 1997.
- Martínez-Rodríguez, A. M., Sáez-Castillo, A. J., and Conde-Sánchez, A.: Modelling using an extended Yule distribution, *Comput. Stat. Data An.*, 55, 863–873, doi:10.1016/j.csda.2010.07.014, 2011.
- Spierdijk, L. and Vorneveld, M.: Superstars without talent? The Yule distribution controversy, *SSE/EFI Working Paper Series in Economics and Finance*, No. 658, <http://hdl.handle.net/10419/56199>, 2007.
- Waymire, E. and Gupta, V. K.: The mathematical structure of rainfall representations 1. A review of the stochastic rainfall models, *Water Resour. Res.*, 17, 1261–1272, 1981.
- Wijngaard, J. B., Klein Tank, A. M. G., and Konnen, G. P.: Homogeneity of 20th century European daily temperature and precipitation series, *Int. J. Climatol.*, 23, 679–692, 2003.