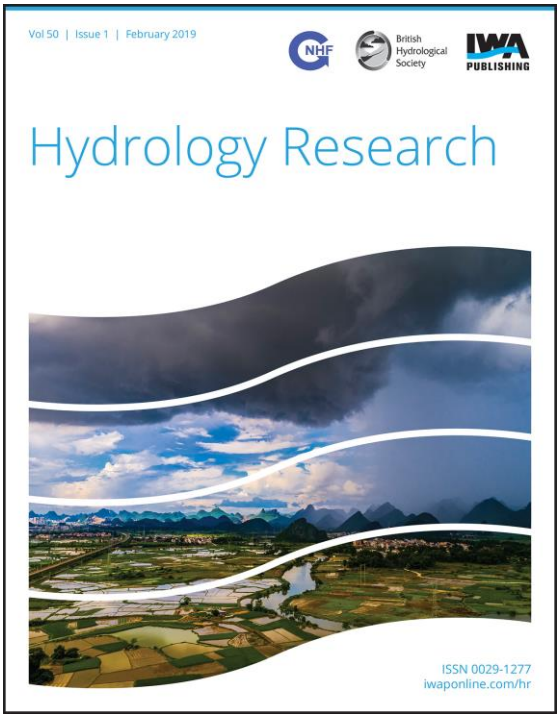

Modelling the frequency distribution of inter-arrival times from daily precipitation time-series in North-West Italy

Giorgio Baiamonte, Luca Mercalli, Daniele Cat Berro, Carmelo Agnese and Stefano Ferraris

ABSTRACT

The discrete three-parameter Lerch distribution is used to analyse the frequency distribution of inter-arrival times derived from 26 daily precipitation time-series, collected by stations located throughout a 28,000 km^2 area in North-West Italy (altitudes ranging from 113 m to 2,170 m a.s.l.). The precipitation regime of these Alpine regions is very different (latitude 44.5 to 46.5 N) from the typical Mediterranean precipitation regime of the island of Sicily (latitude 37 to 38 N), where the Lerch distribution has already been tested and whose results are compared. In order to verify the homogeneity of the precipitation time series, the Pettitt test was preliminarily performed. In this work, a good fitting of the Lerch distribution to NW Italy is shown, thus evidencing the wide applicability of this kind of distribution, also allowing to jointly model dry spells and wet spells. The three parameters of the Lerch distribution showed rather different values than the Sicily ones, likely due to the very different precipitation regimes. Finally, a relevant spatial variability of inter-arrival times in the study area was revealed from the regional scale application of the probability distribution here described. The outcomes of this study could be of interest in different hydrologic applications.

Key words | inter-arrival times, Lerch probability distribution, rainfall regime, water scarcity, wet spells and dry spells

Giorgio Baiamonte (corresponding author)
Carmelo Agnese
Department of Agricultural, Food and Forest Sciences,
Università di Palermo,
viale delle Scienze, 13, 90128 Palermo,
Italy
E-mail: *giorgio.baiamonte@gmail.com*

Luca Mercalli
Daniele Cat Berro
Società Meteorologica Italiana,
Via Real Collegio, 30, 10024 Moncalieri (TO),
Italy

Stefano Ferraris
Interuniversity Department of Regional and Urban Studies and Planning,
Politecnico e Università di Torino,
viale Mattioli, 39, 10125 Torino,
Italy

INTRODUCTION

A robust statistical inference of the alternation of wet spells and dry spells (*DS*) is still an open issue, despite its fundamental role in several common hydrological applications, including irrigation management (Rallo *et al.* 2014; Canone *et al.* 2015), the analysis of precipitation temporal trends (Agnese *et al.* 2012) and the eco-hydrological modelling of land–atmosphere feedbacks (Agnese *et al.* 2016).

In most studies, the statistical analysis of the sequences of rainy days, previously denoted wet spells (*WS*), and that of no-rainy days, *DS*, was carried out separately. In particular, the observed *WS* frequencies were commonly modelled by the geometric distribution, consequent to the statement that a rainy day is a 'success' outcome of classical Bernoulli trials, with constant rain probability.

As showed by Chatfield (1966), this assumption of constant rain probability is commonly unrealistic for modelling *DS* frequency distribution; in fact, as stated by the author, 'the probability that a dry day will be followed by a dry day does increase with the previous number of consecutive dry days'. By using daily rainfall data of London from 1958 to 1965, Chatfield successfully fitted the dry spell frequencies through the logarithmic-series distribution; this distribution allows accounting for the observed persistence of dry states, because of its peculiar pattern having a slowly decaying tail (Kottegoda & Rosso 1997). In a previous study, Agnese *et al.* (2014) demonstrated that the two-parameter polylogarithmic distribution significantly outperforms the logarithmic distribution, in order to statistically infer the observed *DS* frequencies.

doi: 10.2166/nh.2018.042

As mentioned above, here the joint modelling of both *WS* and *DS* is performed, instead of treating them separately. In fact, a single occurrence process regulates the alternating sequences of rainy and no-rainy days. A similar approach was carried out by Strupczewski *et al.* (2013), who derived a two-component model of flood dynamics that combines the duration of high waters and the exceeding of a discharge threshold, thus distinguishing two possible states, namely, flood and no-flood.

A joint modelling of both *WS* and *DS* can be obtained by considering the observed frequencies of the so-called 'inter-arrival times (IT)' of precipitation events, *ITs*, which represent the series of times elapsed between two subsequent rainy days, instead of the commonly used number of dry days. Indeed, the probabilities of rainy day sequences (i.e., wet spells, *WS*) and of dry sequences (i.e., dry spells, *DS*), can be easily derived from the *IT* distribution. *IT* series can be viewed as a realization of a renewal process (Feller 1968; Cox 1970; Buishand 1977), by assuming that *ITs* are independent and identically distributed (*i.i.d.*) random variables. Accordingly, the *WS* geometric distribution directly emerges from the renewal property, whereas the *DS* probabilities can be easily calculated by rearranging the *IT* probabilities.

Another important aspect of this work is to use a discrete probability distribution. In the literature, both discrete and continuous probability distributions were used to model the frequency distribution of inter-arrival time-series (Foufoula-Georgiou & Lettenmaier 1986; Rodriguez-Iturbe *et al.* 1987; de Groen & Savenije 2006; Manfreda & Rodrıguez-Iturbe 2006), although the inter-arrival time is properly a discrete random variable. In fact, Foufoula-Georgiou & Lettenmaier (1986) demonstrated the statistical advantages of treating the rainfall occurrence as a discrete process, rather than a continuous one. In order to describe the rainfall duration, Manfreda & Rodrıguez-Iturbe (2006) considered the exponential distribution, which is the extension to the continuum of the discrete geometric distribution. Eichner *et al.* (2011) discussed the influence of the persistence of the daily rainfall process on that of the IT by using long-term correlated data sets.

In a previous work, the *ITs* have already been modelled by Agnese *et al.* (2014) using the discrete three-parameter Lerch distribution. This distribution is able to account for some peculiar features usually observed in *IT* samples, which other distributions are not capable of reproducing as well. These features are the high standard deviation and the high positive skewness, the monotonically decreasing pattern of frequencies with a slowly decaying tail and a distinctive 'drop' marking the passage from the wet status ($IT = 1$) to the dry status ($IT = 2$). The ability of the Lerch distribution in reproducing all of these features allows the use of this distribution for a variety of hydrological applications. In a previous work regarding the period 1975–2005, Agnese *et al.* (2014) successfully applied the Lerch distribution to the *IT* frequencies derived from 55 daily rainfall series recorded in Sicily.

However, rainfall regime in Sicily is characterized by a strong seasonality, splitting of the year into two parts, a wet cold season (from October to March) and a dry warm one (from April to September). Consequently, data have been preliminarily partitioned into two groups falling in those two different periods of the year, and then they were fitted separately via the Lerch distribution. For data belonging to the 'wet' semester, all records satisfied the chi-square test of the goodness of fit; whereas for data of the 'dry' semester, 11 series did not satisfy the test, likely because of the small available sample size.

The primary objective of this work is to test the Lerch distribution in an area ranging from the highest peaks of the Alps to the Tyrrhenian Sea. In order to do that, the *IT* series were derived from long daily precipitation series (70/90 years duration), collected in 26 stations located throughout the Piedmont and Aosta Valley, in typical Alpine and sub-Alpine areas of North-West Italy. The climate of the studied area is strongly spatially variable (Figure 1) and was classified as sub-litoranean, sub-alpine, pre-alpine and sub-continental, and differs completely from that of Sicily which is typical Mediterranean. The different parameter values derived by the statistical inference were compared to those obtained in the study about the Sicily region and results were discussed. Therefore, the application of the three-parameter Lerch distribution also to the area of NW Italy represents an important test to establish the wide applicability of the proposed approach. An attempt to explore the validity of the hypothesis (usually accepted) that *ITs* are *i.i.d.s* was carried out.

Moreover, an objective of this paper is also to use the Lerch distribution in order to describe the sequences of rainy days, interrupted by 1-day DS, which could have a

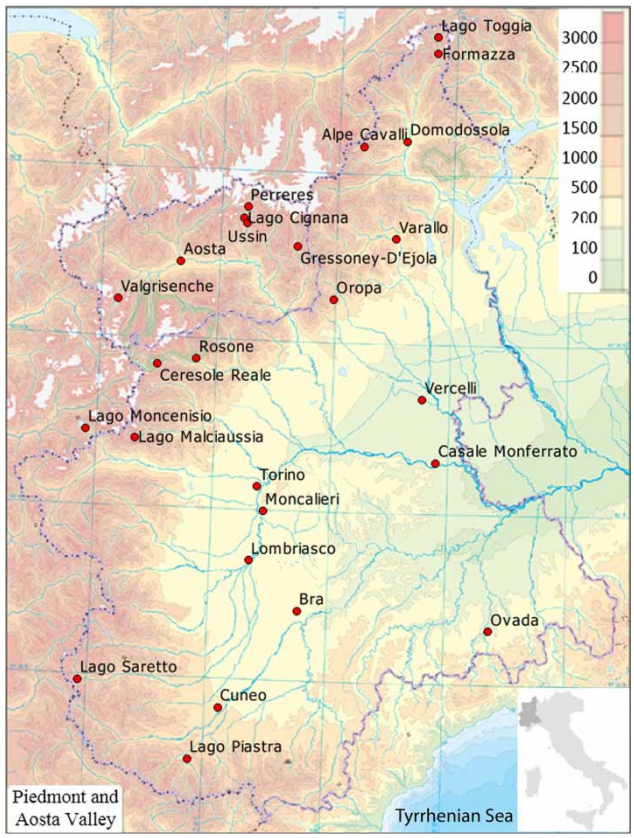

Figure 1 | Physical map of Piedmont and Aosta Valley with the station locations (dots) and the elevation legenda of the altitudes (in metres) at the right top of the figure.

certain interest from a hydrological point of view, representing in a broad sense, a unique perturbed period, not significantly altering the antecedent soil moisture status. The equivalence on using a single *IT* three-parameter Lerch distribution, in place of using two distributions, namely *DS* two-parameter polylogarithmic distribution and *WS* one-parameter geometric distribution, is also argued.

Finally, the expected return periods of dry and wet spells at the regional scale are determined, in order to quantify the spatial variability of the precipitation regimes in the investigated area, and to provide suitable results for the sake of agricultural water management.

STUDY AREA AND DATA

In this work, the daily precipitation data long series (70/90 years) from 26 stations have been used. They belong to the available databases of the National and Regional Observatories (ARPA Piemonte and Centro Funzionale Valle d'Aosta), and of the Italian Meteorological Society (Mercalli & Cat Berro 2008). These stations are located throughout the Piedmont and Aosta Valley area (Italy) and cover an area of 28,000 km^2, characterized by different climatic regimes, ranging from sub-continental to sub-litoranean ones, in between 113 and 2,170 m altitude. It is an area characterized by very high mountains like Mont Blanc, Matterhorn and Gran Paradiso, quite close to the plains and to the sea. In addition to the mentioned orographic issue, the irrigated agricultural Po Plain plays a major role in the air humidity pattern because of the widespread surface irrigation, which saturates wide areas. This produces a notable variation in the average statistics of precipitation occurrence through the year and through the study area.

The stations were selected in order to have highly reliable and long records, with few gaps in the data series. Figure 1 shows the location of the stations (dots) on a physical map of the area. Table 1 reports the available years of observations, the elevation, the latitude and the longitude (UTM 32 T, m). For the purpose of this study, a common starting year (the year 1919, if available) was considered for all 26 data series. The *IT* series were obtained from the recorded precipitation observations using the precipitation threshold $h^* = 1$ mm (Figure 2), which is usually adopted in Italy by regional observatories in order to discriminate between rainy and no-rainy days.

A simple analysis of the monthly-averaged frequencies of wet days through the year (Table 2), revealed the presence of a 'wet' period in spring (peak in May) and a 'dry' period in winter (minimum in January), for all the stations. Furthermore, another 'wet' period in autumn is observed in the area located at low–medium altitudes, whereas a secondary peak in August is present at the high altitudes. These facts complicate the splitting of the year into two homogeneous periods valid for the whole area, as for the Sicily case.

In spite of these problems, with almost similar characters for all the sites, a simple partition of the whole year into two contiguous periods was chosen. Accordingly, a first period, denoted as the 'cold season', including the four months (from December to March) in which lowest precipitation day frequencies are observed 'on average'

Table 1 | Years of observations, elevation and coordinates of the selected 26 stations

ID	Station	From	To	Available records (years)	Elevation (m a.s.l.)	North latitude UTM 32 T (m)	East longitude
1	Alpe Cavalli	1929	2010	82	1,510	431,301	5,104,900
2	Aosta	1891	2010	120	583	369,928	5,066,980
3	Bra	1862	2010	149	290	408,452	4,949,620
4	Casale Monferrato	1913	2011	99	113	455,028	4,999,120
5	Ceresole Reale	1926	2011	86	1,579	361,956	5,032,690
6	Cuneo	1901	2010	110	536	382,020	4,917,460
7	Domodossola	1871	2013	143	302	445,620	5,106,821
8	Formazza	1913	2008	96	1,270	456,009	5,136,150
9	Gressoney-D'Ejola	1927	2011	85	1,850	408,907	5,071,847
10	Lago Cignana	1927	2011	85	2,170	390,934	5,081,400
11	Lago Malciaussia	1936	2011	76	1,810	354,297	5,008,050
12	Lago Moncenisio	1931	2010	80	2,000	337,919	5,011,030
13	Lago Piastra	1916	2010	95	956	371,812	4,900,240
14	Lago Saretto	1913	2010	98	1,633	335,203	4,927,000
15	Lago Toggia	1932	2010	79	2,160	456,050	5,141,710
16	Lombriasco	1928	2012	85	241	392,405	4,966,572
17	Moncalieri	1866	2010	145	225	397,101	4,983,130
18	Oropa	1961	2012	94	1,180	420,972	5,053,905
19	Ovada	1913	2010	98	187	472,477	4,942,550
20	Perreres	1928	2011	84	1,750	392,292	5,085,080
21	Rosone	1938	2011	74	714	375,024	5,034,280
22	Torino	1802	2010	209	238	395,177	4,991,310
23	Ussin	1929	2007	79	1,322	392,196	5,079,520
24	Valgrisenche	1913	2007	95	1,780	348,869	5,054,490
25	Varallo	1913	2011	99	453	441,947	5,074,045
26	Vercelli	1927	2010	84	135	450,518	5,020,260

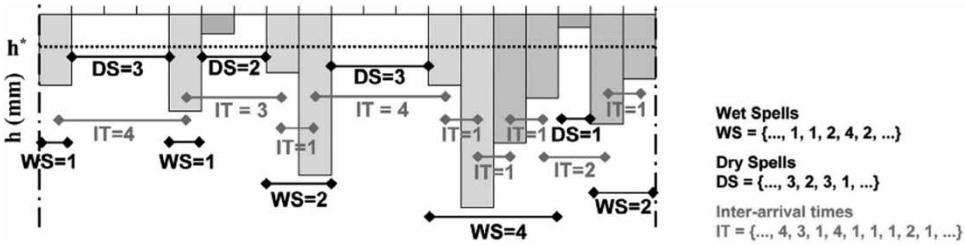

Figure 2 | Sketch of inter-arrival times (IT), dry spells (DS) and wet spells (WS). A day is a rainy day if the daily precipitation depth h is greater than a threshold h^*.

and a second period, namely, the 'warm season', covering the remaining months (from April to November), were selected. This choice is supported by Figure 3, which reports, for all the considered stations, the average number of events per month, <# events/month>, during the 'cold season' and the 'warm season'. As can be observed, all the

Table 2 | Average number of events per month, <# events/month>, of the 26 selected stations

	Station	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1	Alpe Cavalli	6.25	6.84	7.67	10.32	12.83	11.22	9.06	10.15	9.06	8.87	8.50	6.72
2	Aosta	5.08	5.01	5.31	5.80	6.63	6.77	5.59	5.96	5.71	6.26	6.45	5.76
3	Bra	3.90	4.49	5.59	7.39	7.99	6.08	3.77	4.73	5.25	6.28	5.88	5.00
4	Casale Monferrato	5.03	4.97	5.51	6.72	7.22	5.89	4.08	4.57	4.81	6.26	6.52	5.34
5	Ceresole Reale	5.74	6.22	6.81	9.26	11.11	9.83	7.64	8.74	7.98	7.84	7.51	6.35
6	Cuneo	4.46	5.00	7.00	8.67	9.14	7.63	5.02	5.28	6.29	7.35	6.47	5.37
7	Domodossola	5.07	5.23	6.61	9.05	11.12	9.26	7.84	8.47	7.58	7.44	7.79	5.46
8	Formazza	7.48	7.59	7.84	9.34	11.13	10.36	9.27	10.07	8.74	8.85	8.83	7.73
9	Gressoney-D'Ejola	6.74	7.21	7.62	9.84	12.28	11.62	9.32	10.22	7.98	7.87	7.99	7.42
10	Lago Cignana	7.82	8.01	8.15	9.19	11.48	10.76	8.50	9.78	8.11	7.93	8.22	7.78
11	Lago Malciaussia	6.53	6.69	7.47	9.62	11.69	10.53	6.90	8.37	7.70	8.19	7.43	6.83
12	Lago Moncenisio	6.31	6.18	6.43	7.49	9.96	9.06	6.60	7.17	7.34	7.19	6.91	6.48
13	Lago Piastra	5.10	5.16	6.63	9.04	9.50	7.71	5.35	5.95	6.46	7.42	6.62	5.55
14	Lago Saretto	5.48	5.96	6.87	8.74	9.61	9.47	6.35	6.88	7.11	7.84	7.02	6.52
15	Lago Toggia	10.15	10.12	9.59	10.68	12.26	12.23	10.48	11.94	10.09	9.84	11.22	10.24
16	Lombriasco	4.24	4.36	5.08	7.05	8.22	6.51	4.05	5.31	5.36	5.68	5.82	4.49
17	Moncalieri	3.91	4.58	5.41	7.61	9.08	7.28	5.18	5.45	5.59	6.26	5.63	4.80
18	Oropa	5.29	6.39	7.98	11.67	13.53	11.48	8.41	9.03	8.44	8.08	7.56	5.45
19	Ovada	5.17	5.29	5.60	6.41	5.97	4.16	2.80	3.35	4.66	6.50	7.66	5.83
20	Perreres	7.19	7.31	7.29	8.73	11.11	10.67	8.92	9.62	8.28	7.87	8.07	7.29
21	Rosone	5.12	5.72	6.36	9.58	11.60	10.15	7.28	8.54	7.26	7.60	6.55	5.53
22	Torino	4.10	4.73	5.73	8.26	9.85	8.29	5.54	5.88	6.09	6.57	5.95	4.97
23	Ussin	6.13	6.10	6.50	7.79	10.34	9.74	7.77	8.41	7.29	7.84	7.29	6.26
24	Valgrisenche	7.61	7.66	8.02	8.57	8.79	9.59	7.62	8.62	7.69	7.88	8.41	7.84
25	Varallo	3.88	4.39	6.11	9.26	11.04	9.87	7.62	7.64	7.15	7.39	6.96	4.42
26	Vercelli	4.69	4.61	5.51	7.29	8.12	6.73	4.54	5.48	5.43	6.62	6.98	5.22
	<# events/month>	5.71	5.99	6.72	8.59	10.06	8.96	6.75	7.52	7.06	7.45	7.32	6.18

months of the 'cold season' show lower average number of events per month than the months of the 'warm season'.

Then, for each season, a statistical inference was performed by fitting the *IT* data frequencies through the three-parameter Lerch (*Lch*) distribution. For the sake of comparison, the statistical inference was also applied to the whole year data set (Y).

Before performing the statistical inference, the occurrence of a change-point in some average annual *IT* data series, <*IT*>, was detected. For this purpose, a number of different tests can be applied (Baiamonte & D'Asaro 2016; Pohlert 2018), such as either the Buishand test or the standard normal homogeneity test. However, for its simplicity and for its wide application in hydrology (Zhang *et al.* 2009), the Pettitt test (Pettitt 1979) was considered in this study. The Pettitt test localizes the year when most likely a breakpoint (discontinuity or inhomogeneity) occurs. Detecting the change-point allows dividing the entire data series into two intervals, before and after the change-point, forming two homogeneous groups.

The null hypothesis of the test is that the two sequences of random variables, $(x_1, x_2, \ldots x_t)$ and $(x_{t+1}, x_{t+2}, \ldots, x_T)$, have a common distribution function, against the alternative hypothesis that a change-point exists. The non-parametric

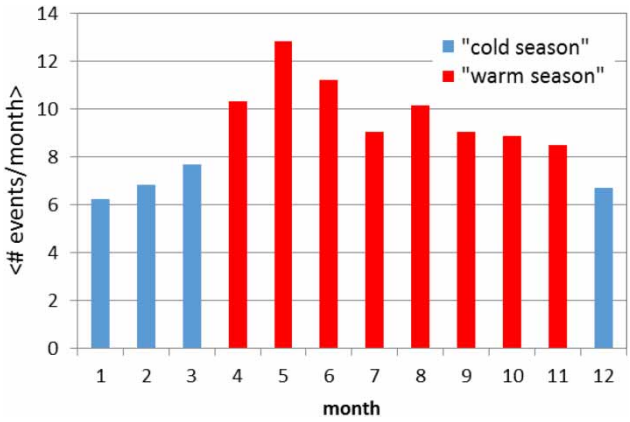

Figure 3 | Average number of events per month, <# events/month>, for the 'cold season' and for the 'warm season', of the selected 26 stations.

statistic $U_{t,T}$ is given by:

$$K_T = \max\{|U_{t,T}|\} \tag{1a}$$

where

$$U_{t,T} = \sum_{i=1}^{t}\sum_{j=t+1}^{T} Sign(x_t - x_j) \tag{1b}$$

For $p \le 0.05$, the significance probability of K_T is approximated by:

$$\rho = 2\exp\left(\frac{-6\,K_T^2}{T^3 + T^2}\right) \tag{1c}$$

If ρ is smaller than the specific significance level, p, the null hypothesis is rejected, then a change point exists, and the time series is divided into two parts at the location of the change-point.

For changing in one direction, the statistics $K_T^+ = \max\{U_t,T\}$ and $K_T^- = -\min\{U_{t,T}\}$ were calculated, with $K_T = \max\{K_T^+, K_T^-\}$, indicating that large K_T^+ corresponds to a shift down in level, whereas large K_T^- corresponds to a shift up in level from the beginning of the series, respectively.

Table 3 reports the statistics K_T^+, K_T^-, K_T, the p-value significance level, the year of changing and the shift upward or downward for the considered 26 stations.

The results showed that the null hypothesis was rejected (i.e., no change-point occurs) only for eight stations out of the selected 26 ones. Thus, 18 data series were considered as a whole, whereas in the remaining eight series, the period before the changing year (detected by the Pettitt test) was eliminated. In those cases, only the more recent homogeneous period was considered for the application of the Lerch distribution.

The cases where the null hypothesis was rejected are depicted in Figure 4, showing the <*IT*> time-series, the location of the change-point, and the disregarded period highlighted by a shaded area. As can be observed, the upward or downward shift agrees with the results of the Pettitt test reported in Table 3. However, it should be emphasized that, as recently observed by Serinaldi *et al.* (2018), such change-points should be accepted as physically meaningful if they can be related with a predictable process based on theoretical models and/or well identified physical dynamics justifying causality. In our case, neither of the occurrences was possible to find.

METHODS

Let $H = \{h_1, h_2, \ldots h_k.., h_N\}$ be a time-series of precipitation data spaced at a uniform time-scale, τ, e.g., day; a day k is a 'rainy' day if $h_k \ge h*$, otherwise k is a 'no-rainy' (or dry) day; $h* \ge 0$ is any precipitation threshold.

From the H series (of size N), the inter-arrival time series, $IT = \{IT_1, IT_2, \ldots IT_M\}$, representing the succession of the times elapsed between two subsequent rainy days, could be easily derived; IT size, M, equals the number of rainy days. Accordingly, an isolated rainy day is identified by the conditions $IT_j > 1$ and $IT_{j+1} > 1$, whereas a wet spell, WS, of duration $n+1$ is identified by any sequence, in IT, of n consecutive 1 values. Moreover, by neglecting any $IT_k = 1$, the DS sequence, $DS = \{DS_1, DS_2, \ldots, DS_L\}$, can be easily obtained, by $DS_j = IT_j - 1$; then, if r indicates the frequency of ITs equal to 1, the size L of DS (or, equivalently, of WS) results as: $L = M(1-r)$. Figure 2 reports an example of the terms and sequences defined above.

Sample means of IT, DS and WS (in τ units) can be easily expressed by the following relationships:

$$\langle IT\rangle = \frac{N}{M};\ \langle DS\rangle = \frac{N-M}{M(1-r)} = \frac{\langle IT\rangle - 1}{1-r};\ \langle WS\rangle = \frac{M}{M(1-r)} = (1-r)^{-1} \tag{2}$$

Table 3 | P value, statistics, K_T^+, K_T^-, K_T, and year of changing, of the selected 26 stations

#	Station	p value	K_T^+	K_T^-	K_T	Year of changing	Shift
1	Alpe Cavalli	1.000	165	198	198	1941	–
2	Aosta	0.519	414	129	414	1925	–
3	Bra	0.065	200	659	659	1980	–
4	Casale Monferrato	0.108	609	508	609	1950	–
5	Ceresole Reale	0.278	444	275	444	1949	–
6	Cuneo	0.000*	50	1,321	1,321	1964	Upward
7	Domodossola	0.582	94	396	396	1937	–
8	Formazza	0.068	645	274	645	1945	–
9	Gressoney-D'Ejola	0.020*	73	654	654	1958	Upward
10	Lago Cignana	0.200	82	471	471	1940	–
11	Lago Malciaussia	0.164	29	414	414	1996	–
12	Lago Moncenisio	0.000*	1,094	41	1,094	1949	Downward
13	Lago Piastra	0.000*	1,404	120	1,404	1970	Downward
14	Lago Saretto	0.005*	829	7	829	1967	Downward
15	Lago Toggia	0.480	266	338	338	2002	–
16	Lombriasco	0.031*	611	171	611	1952	Downward
17	Moncalieri	0.090	106	627	627	1978	–
18	Oropa	0.071	190	609	609	1972	–
19	Ovada	0.155	78	524	524	1973	–
20	Perreres	0.133	502	248	502	1976	–
21	Rosone	1.000	171	183	183	1944	–
22	Torino	0.143	165	579	579	1979	–
23	Ussin	0.010*	24	666	666	1947	Upward
24	Valgrisenche	0.001*	114	986	986	1940	Upward
25	Varallo	0.217	344	500	500	1957	–
26	Vercelli	0.091	70	546	546	1937	–

The shift for the stations for which the null hypothesis is rejected is also reported. The symbol *indicates the p value is less than 0.05.

With the hypothesis that ITs are independent random variables with identical distribution (*i.i.d.*), each sample series can be interpreted as a realization of a renewal process, the main feature of which is that it probabilistically re-starts at each arrival epoch, independently of the previous history (the so-called 'renewal property', Cox 1970); consequently, WS must be geometrically distributed:

$$P\{WS = k\} = p_1^{k-1}(1 - p_1) \tag{3}$$

with parameter $(1 - p_1)$, where p_1 equals the probability that $IT = 1$.

Furthermore, the occurrence probability of DS of size k, could be easily derived from IT distribution as follows:

$$P\{DS = k\} = \frac{P\{IT = k + 1\}}{1 - p_1} \tag{4}$$

where $P\{DS = k\}$ denotes the probability that DS equals k.

Conversely, if WS is geometrically distributed with parameter $(1 - p)$, the probability distribution of IT can be easily deduced as follows:

$$P\{IT = k\} = (1 - p)\, P\{DS = k - 1\} \text{ if } k > 1 \tag{5a}$$

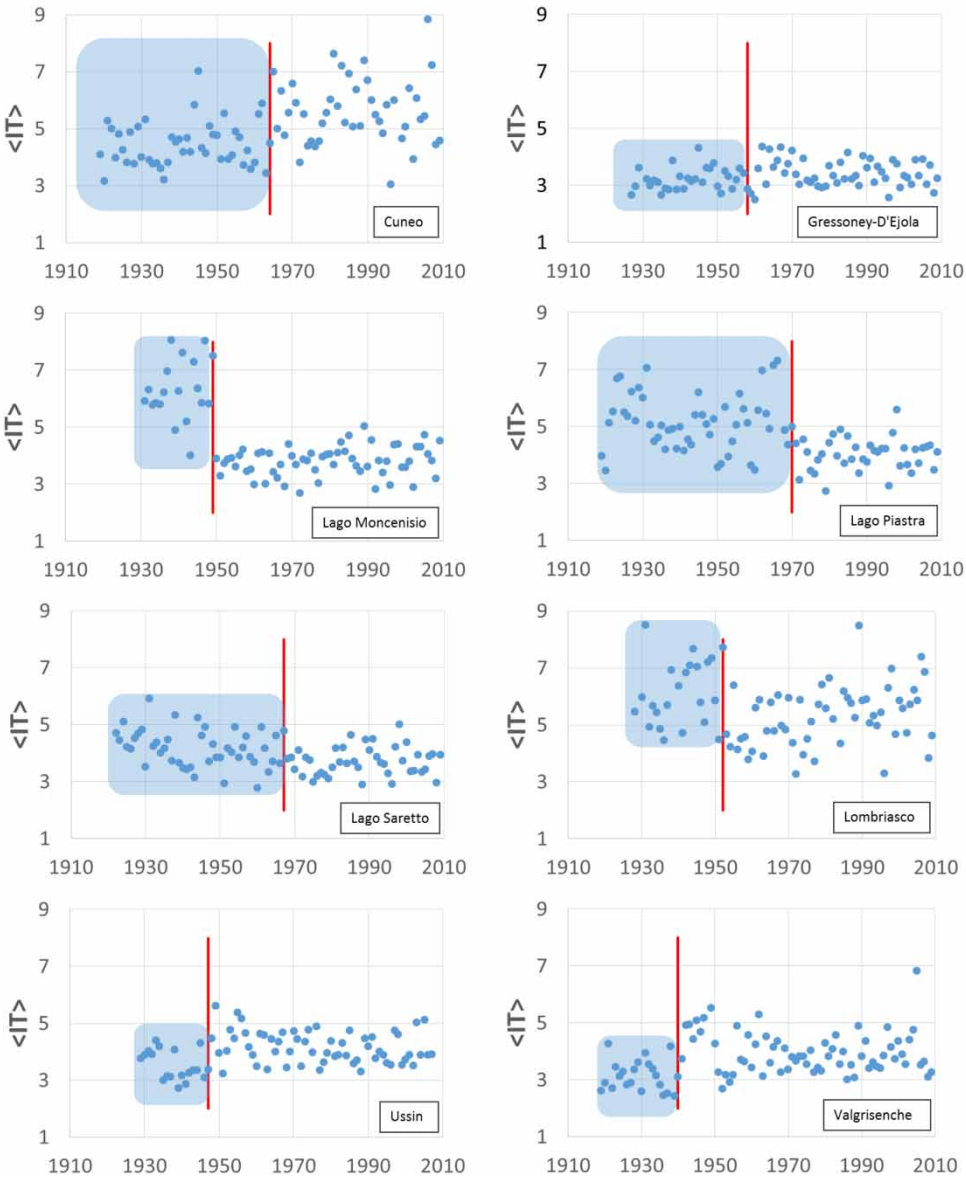

Figure 4 | Average annual *IT* data, <*IT*>, location of the change-point (vertical straight line), and the disregarded period highlighted by a shaded area.

and

$$P\{IT = k\} = p \text{ if } k = 1 \tag{5b}$$

The empirical frequency distributions of IT (or, equivalently of DS) at the daily scale, usually exhibited high standard deviation and skewness, and a monotonic decreasing pattern of frequencies with a slowly decaying tail. Moreover, *IT* distributions showed a peculiar 'frequency drop' from the modal value at $IT = 1$ to the value at $IT = 2$; this drop is generally absent in the *DS* distributions, since it properly accounts for the natural tendency of precipitation to produce clusters.

In order to reproduce the main features of the above-mentioned empirical *IT* distributions, Agnese *et al.* (2014) suggested applying the three-parameter Lerch distribution.

The three-parameter Lerch probability distribution represents the most general form of a family of discrete

probability distributions extensively used in the scientific literature (Aksenov & Savageau 2005; Kemp 2006; Gupta *et al.* 2008). The probability mass function (*pmf*) of the 'family' of Lerch distributions is:

$$P\{X=k\}=\frac{\theta^k}{(k+a)^s\Lambda(\theta,s,a)} \tag{6}$$

where θ, s and a are the distribution parameters, with the constraints indicated in Table 4, $\Lambda(\theta,s,a)=\sum_{k=1}^{\infty}(\theta^k/(k+a)^s)=\theta\,\Phi(\theta,s,a)$ and $\Phi(.,.,.)$ is the Lerch transcendent function (Erdélyi *et al.* 1981).

In the most general form, the distribution reported in Equation (5), hereafter denoted as *Lch*, depends on the three parameters θ, s and a (Table 4). Positive s values ensure that the probability mass function (*pmf*) is a monotonically decreasing function with mode $X=1$.

A special case of the *Lch* distribution is the polylogarithmic (hereafter referred to as *Poly*) distribution (Kemp 2006), which can be obtained by setting $a=0$ into Equation (5):

$$P\{X=k\}=\frac{\theta^k}{k^s Li_s(\theta)} \tag{7}$$

where $Li_s(\theta)=\sum_{k=1}^{\infty}(\theta^k/k^s)$ is the polylogarithm function (Jonquière 1889). Good (1953) and Kemp (2006) suggested using the *Poly* distribution in order to fit the observed population frequencies of species per genus. For $s=1$, Equation (6) reduces to the logarithm-series distribution (*Log*), widely used in the past for describing DS frequencies (Chatfield 1966; Buishand 1977, 1978; Kottegoda & Rosso 1997):

$$P\{X=k\}=\frac{\theta^k}{k\,Li_1(\theta)} \tag{8}$$

where $Li_1(\theta)=\sum_{k=1}^{\infty}(\theta^k/k)=-ln(1-\theta)$

For $s=0$, Equation (6) reduces to the well-known geometric series (hereafter referred to as *Geo*):

$$P\{X=k\}=(1-\theta)\theta^{k-1} \tag{9}$$

Therefore, the distributions expressed by Equations (8) and (9) are different forms (particular cases) of *Lch* distribution.

In Table 4, the Lerch family of the probability distributions and the corresponding parameters domain are resumed (ID = 1–7) (Zörnig & Altmann 1995).

In order to perform the statistical inference of the previously introduced *pmf*s, the well-known maximum likelihood estimation (MLE) method was used. The analytical solution of the *MLE* method for the general three-parameter *Lch* implies the equality of the arithmetic, geometric and harmonic expectations to the corresponding sample means (Gupta *et al.* 2008; Agnese *et al.* 2014). Obviously, two-parameter and one-parameter distributions of the Lerch family do not fulfil all equality conditions; in Table 4, conditions that are satisfied or not satisfied by each distribution are also indicated.

Table 4 | Lerch family of probability distributions with the corresponding parameters domains, and the theoretical means (arithmetic, AM, geometric, GM, and/or harmonic, HM) that match (x) or not (–) with the empirical ones, according to the MLE method

ID	Probability distribution	Code	θ	s	a	AM	GM	HM
1	3-par Lerch	*Lch*	$0<\theta<1$	>0*	>-1	x	x	x
2	2-par polylogarithmic	*Poly*	$0<\theta<1$	>0*	0	x	x	–
3	1-par logarithmic	*Log*	$0<\theta<1$	1	0	x	–	–
4	1-par geometric	*Geo*	$0<\theta<1$	0	0	x	–	–
5	2-par extended log	*Ext Log*	$0<\theta<1$	1	>-1	x	–	x
6	2-par Hurwitz	*Hur*	1	>1	>-1	–	x	x
7	1-par zeta (Zift)	*Zeta*	1	>1	0	–	x	–

*The condition $s>0$ allows obtaining monotonically decreasing *pmf*, with mode equal to 1.

The classical χ^2 procedure is not appropriate to test the goodness-of-fit of long-tailed distributions with several small classes (less than five elements) and strong asymmetry, as in our case, as pointed out by Martínez-Rodríguez *et al.* (2011). Therefore, the evaluation of the goodness-of-fit was performed by operating with artificial *IT* synthetic samples generated via Lerch distribution and numerically reconstructing the null hypothesis of χ^2 via Monte Carlo simulation. A set of 2,000 replicates of the *IT* sample was generated by randomly sampling from the theoretical distribution obtained by inference, and the associated *p*-value was defined as the fraction of the 2,000 replicates for which reconstructed χ^2_j values (with $j = 1, 2, \ldots, 2{,}000$) are greater than χ^2_{ref}, obtained by the classical procedure (Hope 1968). Accordingly, the null hypothesis was accepted if the *p*-value was greater than 0.05, the value that is usually assumed as the significance level.

Moreover, in order to establish if the adoption of the general three-parameter *Lch* is justified with respect to the other distributions with one or two parameters (Table 4), the likelihood-ratio (LLR) test was used, which is suggested when using distributions that are particular cases (logarithmic distribution, polylogarithmic distribution, etc.) of the general one (Lerch distribution) (Wilks 1938). In fact, it is obvious that the more general model will always fit at least as well as the special cases; the LLR test allows detecting if the improvement in terms of log-likelihood endorses the introduction of the extra parameters. Each of the two competing models, the null model (ID = 2–7) and the alternative model (*Lch*, ID = 1), is separately fitted to the data and the log-likelihood was calculated. The test statistic *D* is defined as twice the log of the likelihoods ratio:

$$D = -2 \ln\left(\frac{LLR\ null\ model}{LLR\ alternative\ model}\right) \tag{10a}$$

where the null model represents a special case of the alternative model and the likelihood function is defined by the following:

$$\ln L(\theta,\ s,\ a) = \ln\theta \sum_{i=1}^{M} IT_i - s\sum_{i=1}^{M}\ln(a + IT_i) - M\ln[(\theta,\ s,\ a)] \tag{10b}$$

where *M* is the sample size. The probability distribution of the test statistic is approximately a chi-squared distribution with degrees of freedom equal to $df_{alt} - df_{null}$, i.e., the number of free parameters of the alternative and the null model, respectively.

For the whole year data set and for all year records, Table 5 reports the comparison of the statistic, *D*, calculated for the alternative model (*Lch*) and that calculated for the null models (ID = 2–7). The critical chi-square statistic, at the significant level, $p = 0.05$, is also reported. The statistic *D* appears greater than the corresponding critical chi squared, in all the series and in all probability distributions, justifying the use of the three-parameter *Lch*.

Table 5 also shows that the extended log (hereafter referred to as *Ext log*) shows the best performance (lowest *D* values) among the other distributions, whereas the one-parameter *Geo* shows the worst performance (greatest *D* values).

For the purposes of this work, the three-parameter *Lch* distribution was fitted to the *IT* data series, which, according to Equations (2) and (3), allows the *DS* and *WS* distributions to be derived. Both of the latter could also be derived by directly fitting *Poly* and *Geo* distributions to the corresponding data series. Indeed, if θ and s are the parameters of the *Poly* distribution for *DS* series and p is the parameter of the *Geo* distribution for *WS* series, then the *IT* distribution can also be derived according to Equation (4):

$$P\{IT = k\} = \frac{\theta^{k-1}}{(k-1)^s Li_s(\theta)}(1-p) \text{ if } k > 1 \tag{11a}$$

$$P\{IT = k\} = p \text{ if } k = 1 \tag{11b}$$

The joint use of Equations (11a) and (11b) is actually equivalent to the three-parameter *Lch pmf* (Equation (5)) and requires the same number of parameters (θ, s, p), with p instead of a. Interestingly, θ and s values associated with the *Poly* distribution (for *DS* series, Ferraris *et al.* 2014) are very close to those associated with the *Lch* distribution (for *IT* series).

RESULTS AND DISCUSSION

As can be noticed from the comparison of the elevation data reported in Table 1 and from Figure 1, the selected stations

Table 5 | Statistic D calculated for the alternative model (*Lch*) compared to the null models represented by the distributions, ID = 2–7, reported in Table 4, of the selected 26 stations. The critical chi-square statistic for $p = 0.05$, corresponding to $df_{alt} - df_{null}$ is also reported

		ID					
#	Station	2 $D_{Lch/Poly}$	3 $D_{Lch/Log}$	4 $D_{Lch/Geo}$	5 $D_{Lch/Ext\ Log}$	6 $D_{Lch/Hur}$	7 $D_{Lch/Zeta}$
1	Alpe Cavalli	363.01	905.63	10,174.71	116.64	725.89	849.32
2	Aosta	187.07	193.80	4,784.12	152.20	693.19	1,389.94
3	Bra	261.44	383.02	5,342.38	135.69	735.11	1,015.48
4	Casale Monferrato	270.65	365.52	5,322.77	152.97	765.52	1,084.49
5	Ceresole Reale	295.82	567.64	8,090.65	131.13	705.19	966.14
6	Cuneo	128.31	167.28	5,404.24	80.19	566.42	1,192.70
7	Domodossola	393.96	823.13	8,964.69	151.60	830.80	996.84
8	Formazza	324.69	832.04	11,006.51	102.45	730.87	928.62
9	Gressoney-D'Ejola	160.11	371.18	8,892.12	62.86	520.08	945.13
10	Lago Cignana	211.45	404.17	8,632.62	98.72	610.04	1,025.60
11	Lago Malciaussia	229.22	517.31	7,543.48	82.49	578.79	792.92
12	Lago Moncenisio	197.24	253.22	5,883.40	129.67	611.40	1,128.65
13	Lago Piastra	189.32	282.61	6,304.57	104.36	649.05	1,147.07
14	Lago Saretto	193.68	334.02	7,377.66	95.83	618.40	1,074.80
15	Lago Toggia	280.53	849.05	12,580.02	77.58	602.03	765.72
16	Lombriasco	194.24	300.33	4,725.47	94.64	599.61	868.17
17	Moncalieri	232.03	367.53	5,680.08	111.35	696.27	1,003.16
18	Oropa	383.96	1,035.83	11,369.72	107.58	773.40	894.23
19	Ovada	306.76	457.15	5,375.26	151.93	767.58	969.32
20	Perreres	188.86	360.69	8,166.55	88.13	578.96	1,016.16
21	Rosone	249.61	572.28	6,968.64	82.63	589.00	739.10
22	Torino	263.82	492.53	6,567.87	104.61	716.48	950.44
23	Ussin	180.54	316.97	6,760.57	88.85	569.42	982.46
24	Valgrisenche	195.13	321.18	8,302.92	104.86	634.70	1,201.97
25	Varallo	266.98	794.34	8,796.46	63.09	591.95	706.55
26	Vercelli	253.49	373.62	5,206.90	131.35	690.74	951.08
χ^2		3.841	5.991	5.991	3.841	3.841	5.991
$df_{null} - df_{alt}$		1	2	2	1	1	2

include a wide range of altitudes (from 113 m to 2,170 m a.s.l.), which is representative of the whole region. The comparison between the average IT, $<IT>$, reported in Table 6, highlights substantial differences occurring between the two seasons. However, a variety of conditions are present among the stations, within the same season, as is illustrated by the minimum and maximum (shown in bold in Table 6) of the average IT values, $<IT>$. Table 6 also reports the starting and the ending year of the observations, which accounts for the results obtained by the application of the Pettitt test, the average cluster size, $<WS>$, calculated according to Equation (2), which also characterizes each station.

In particular, Table 6 shows a minimum $<IT>$ value of 2.7 days in the 'warm season' for Lago Toggia station, one of the more northern stations, and a maximum one of 6.57 days in the 'cold season' for Cuneo station, one of the more southern ones. As can be observed, the p-values associated with the χ^2 test are greater than 0.05, for the 'cold

Table 6 | Sample size, average inter-arrival time, IT, average cluster size, <WS>, and *p* value, for the whole year (from October to September), for the 'cold season' (from December to March) and for the 'warm season' (from April to November) of the selected 26 stations

				From October to September (year)				From December to March ('cold season')				From April to November ('warm season')			
ID	**Station**	**Starting year**	**Ending year**	**Sample size**	**<IT> (days)**	**<WS> (days)**	***p*-value**	**Sample size**	**<IT> (days)**	**<WS> (days)**	***p*-value**	**Sample size**	**<IT> (days)**	**<WS> (days)**	***p*-value**
1	Alpe Cavalli	1929	2009	8,693	3.32	2.27	0.158	2,235	4.34	2.10	0.748	6,573	3.00	2.32	0.202
2	Aosta	1919	2009	6,427	5.04	1.67	0.112	1,912	5.67	1.72	0.802	4,593	4.88	1.64	**0.006**
3	Bra	1919	2009	5,942	**5.39**	1.79	0.169	1,615	6.32	1.86	0.530	4,408	5.20	1.75	0.200
4	Casale Monferrato	1919	2009	6,071	5.19	1.78	0.104	1,842	5.60	1.87	0.248	4,310	5.10	1.74	0.345
5	Ceresole Reale	1929	2009	8,002	3.71	2.05	**0.023**	2,085	4.75	1.93	0.225	5,974	3.38	2.09	0.121
6	**Cuneo**	1964	2009	3,783	4.34	1.83	0.497	767	**6.57**	**1.62**	0.523	2,311	5.07	**1.56**	0.070
7	Domodossola	1919	2009	8,282	3.91	2.10	0.073	2,003	5.34	2.05	0.910	6,385	3.52	2.10	0.059
8	Formazza	1919	2008	9,689	3.29	2.22	**0.022**	2,781	3.82	2.19	0.142	7,029	3.12	2.21	0.375
9	**Gressoney-D'Ejola**	1958	2009	5,553	3.26	2.07	0.123	1,516	4.13	1.88	0.302	4,085	3.08	2.04	0.627
10	Lago Cignana	1929	2009	8,650	3.38	2.03	0.188	2,599	3.79	2.06	0.888	6,106	3.26	2.02	0.140
11	Lago Malciaussia	1937	2009	7,167	3.62	2.08	**0.041**	1,973	4.34	1.93	0.174	5,260	3.41	2.13	0.718
12	**Lago Moncenisio**	1949	2009	5,110	4.22	1.80	0.108	1,663	4.31	1.86	0.630	4,245	3.54	1.90	0.163
13	**Lago Piastra**	1970	2009	2,848	5.00	**1.60**	0.221	925	4.96	2.12	0.421	2,732	3.65	2.02	0.425
14	**Lago Saretto**	1967	2009	3,506	4.09	1.81	0.283	1,114	4.28	2.01	0.757	3,070	3.47	2.02	0.086
15	Lago Toggia	1932	2009	10,023	**2.75**	**2.42**	0.093	3,135	**2.96**	**2.50**	0.457	6,991	**2.70**	**2.37**	**0.039**
16	**Lombriasco**	1952	2009	3,822	5.33	1.74	**0.002**	1,952	5.96	1.90	0.439	2,973	4.79	1.77	0.550
17	Moncalieri	1919	2009	6,305	5.12	1.80	0.142	1,553	6.44	1.86	0.836	4,826	4.80	1.77	0.082
18	Oropa	1919	2009	9,299	3.35	2.27	**0.021**	2,150	4.71	2.02	0.228	7,205	2.99	2.34	0.110
19	Ovada	1919	2009	5,727	5.25	1.84	0.523	1,918	5.12	1.96	0.402	3,876	**5.41**	1.77	0.948
20	Perreres	1929	2009	8,364	3.48	2.00	0.056	2,386	4.09	1.94	0.310	6,110	3.29	2.01	0.080
21	Rosone	1938	2009	6,549	3.93	2.06	0.085	1,615	5.28	1.91	0.267	5,053	3.52	2.11	0.192
22	Torino	1919	2009	6,793	4.75	1.89	0.469	1,646	6.21	1.91	0.697	5,234	4.39	1.86	0.685
23	**Ussin**	1947	2007	5,621	3.89	1.91	0.226	1,427	5.13	1.74	0.471	3,974	3.73	1.89	0.397
24	**Valgrisenche**	1940	2009	6,951	3.55	1.96	**0.012**	1,889	4.21	1.97	0.433	4,557	3.69	1.80	**0.036**
25	Varallo	1919	2009	7,696	3.65	2.18	**0.003**	1,621	5.48	1.98	0.155	6,172	3.22	2.22	**0.037**
26	Vercelli	1929	2009	5,885	4.99	1.83	**0.045**	1,613	5.81	1.91	0.241	4,360	4.75	1.79	0.405

The bold and underlined *p*-values are the ones lower than the assumed significance 0.05 level. Maximum and minimum averages are in bold. The bold and underlined stations are the ones with the null hypothesis of the Pettitt test verified ($p = 0.05$).

season', for all the *IT* data sets, whereas, for the 'warm season' and for the all-years record, only 22 and 18 data sets satisfy the χ^2 test, respectively.

For the 26 considered stations, Figure 5 illustrates the comparison between the observed frequencies and the theoretical probabilities calculated according to the *Lch* distribution. As can be observed, the dots are close to the line of perfect agreement, especially for high probabilities, showing evidence of the suitability of *Lch* in describing empirical *ITs*. In particular, the goodness-of-fit for $IT = 1$ also emphasized the *Lch* capability in describing the geometrical distributed *WS* (Equations (3), (5b) and (11b)).

Figure 6 reports two examples of *IT* data fitting with the *Lch* distribution for Lago Toggia and Torino stations that have different rainfall regimes, as can be observed in Table 6 from the corresponding <*IT*> and <*WS*> values. The comparison is performed in terms of cumulative distribution function, $P\{IT \leq t\}$, derived by the *Lch pmf*, according to:

$$P\{IT \leq k\} = \sum_{n=1}^{k} P\{IT = n\} \tag{12}$$

The cumulative *Lch* distributions shown in Figure 6 refer to the whole year (from October to September), and are associated with *p*-values equal to 0.093 and 0.469, for Lago Toggia and Torino, respectively. It is interesting to observe that a good agreement between empirical and theoretical distribution occurs in a wide range of probability, ranging three orders of magnitude. A good fitting also occurs for the low-level probability of exceedance (very high *IT* values), which is interesting from a practical point of view.

Figure 6 also illustrates the fitting of the more common polylogarithmic (*Poly*) and of the logarithmic distribution (*Log*), showing that with respect to the *Lch* distribution, they are not able to reproduce the observed *IT* frequencies.

The cumulative *Lch* distributions for the whole year data set, are compared with those corresponding to the 'cold' and to the 'warm' seasons, in order to demonstrate the seasonality effect, in Figure 7, for Lago Toggia and Torino stations. For both stations and with respect to the whole year *Lch* distributions, the figure shows greater differences in probability for the 'cold season' than for the 'warm season'.

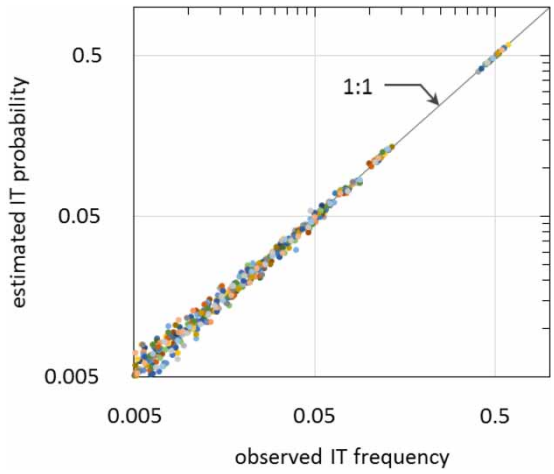

Figure 5 | Comparison between observed frequencies and theoretical probabilities calculated according to the *Lch* distribution, for the 26 considered stations.

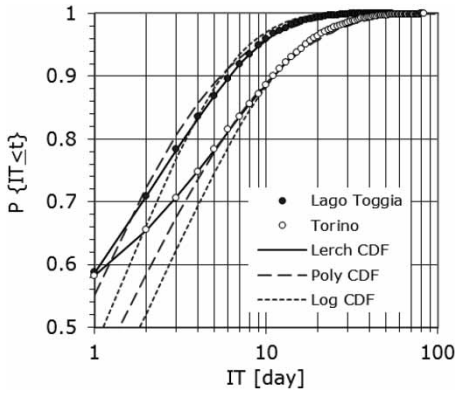

Figure 6 | Log–log plot displaying empirical cumulative distributions of the inter-arrival times (IT) and Lerch cumulative distributions (Lago Toggia and Torino stations, whole year data set).

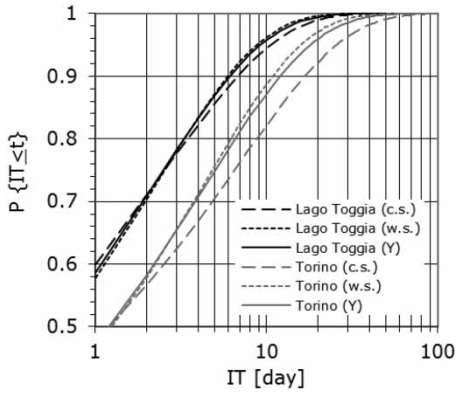

Figure 7 | Log–log plots displaying the Lerch cumulative distributions of the inter-event times (IT) (Lago Toggia and Torino stations' three data sets: 'cold season', 'warm season', whole year, Y).

With reference to the whole year data set, in order to synthetize results of the statistical inference via the Lerch distribution, for the all considered stations, Figure 8(a) compares the observed *WS* frequency distribution with the geometric distribution derived according to Equation (2), with parameter $1-p_1 \equiv \mathrm{P}\{IT>1\}$. Regarding the mentioned series of Lago Toggia and Torino stations, Figure 8(a) shows that the geometric distribution fits the empirical one well. Analogously, Figure 8(b) illustrates a good agreement of the *DS* empirical distribution with the theoretical polylogarithmic one derived from *IT* theoretical distribution, according to Equation (3). Thus, Figure 8(a) and 8(b) together confirm that both *WS* and *DS* distributions can be derived from *IT* distribution fitted with the three-parameter Lerch. The good agreement between the geometric *WS* distribution and the empirical one, illustrated in Figure 8(a), suggests that the assumption that *ITs* are independent random variables with identical distribution (*i.i.d.*) is well founded.

The *i.i.d.* hypothesis is also supported by the agreement between the frequencies of the sum of two observed *IT* values, randomly sampled, with those derived by the two-fold convolution of the *Lch* distribution with itself, as follows:

$$P\{IT_1+IT_2=n\}=\sum_{k=1}^{n-1} p_k\, p_{n-k} \tag{13}$$

As an example, for Bra station and for the whole year data set, this agreement can be observed in Figure 9.

Furthermore, the distribution of sequences of rainy days, interrupted by 1-day DS, was derived from the *Lch* distribution. Define l a number of days, in which a number n of wet days and a number h of 1-day DS occur. Obviously $l=n+h$, with the constraints $n\le l\le 2\ n-1$ and $0\le h\le n-1$. For $h=0$, the latter refers to an uninterrupted sequence of wet days, of probability $(1-p_1-p_2)$, which differs from the *WS* probability that is equal to $(1-p_1)$. The marginal probability $\mathrm{P}\{n=n_0\}$ can be written as:

$$P\{n=n_0\}=\sum_{h=0}^{n_0-1}\binom{n_0-1}{h} p_1^{n_0-1-h} p_2^{h}\,(1-p_1-p_2) \tag{14}$$

where $\binom{n_0-1}{h}$ is the combination of n_0-1 elements taken h at a time, $p_1=\mathrm{P}\{IT=1\}$, $p_2=\mathrm{P}\{IT=2\}$ and $(1-p_1-p_2)\equiv\mathrm{P}\{IT>2\}$. As can be observed in Figure 10, Equation (13) is able to reproduce wet days distribution with 1-day DS, which substantially differ from the *WS* distributions, for both the Lago Toggia and Torino stations.

The above described 'wet day sequences' can represent in a broad sense a unique perturbed period, even though just 1-day DS occur in between the same sequences. Equation (13) could have a certain interest from a hydrological point of view, if considering that 1 no-rainy day in between 2 rainy days does not significantly alter the root zone soil moisture status, mainly for clayey soils.

Regarding the parameters of the *Lch* distribution, for the whole year data set, the θ values varies in the interval $(0.84 \div 0.92)$ and, as expected, higher values are associated with the 'drier' stations and lower values with 'wetter'

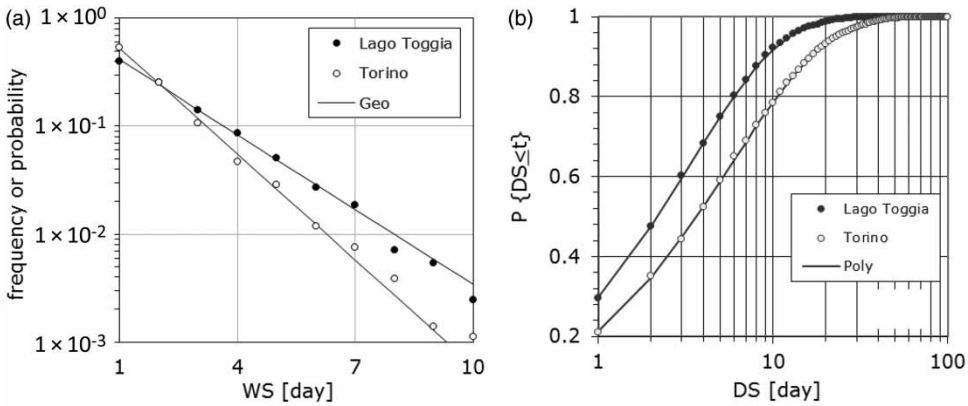

Figure 8 | (a) Empirical and theoretical wet spell, WS, (geometric distribution) and (b) empirical and theoretical dry spell, DS, (polylogarithmic distribution) (Lago Toggia and Torino stations).

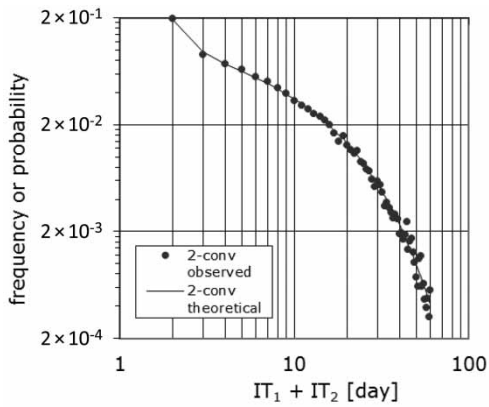

Figure 9 | Frequency of the sum of two observed IT values, IT1 + IT2, randomly sampled, and the fitting theoretical distribution (convolution of the Lerch distribution with itself) obtained by Equation (13).

ones. Moreover, the range of θ values associated with the 'warm season' ($0.82 \div 0.91$) is close to that of the whole year, whereas for the 'cold season' the range of θ values ($0.90 \div 0.95$) is narrower, if compared to that of the whole year data set. In Figure 11(a), the θ values are plotted vs. $<IT>$, for the all-year data set, for the 'warm season' and for the 'cold season'. As can be observed, 'θ behaves somewhat like a scale parameter' (Kemp 2006) and can be also related to a physical parameter, the elevation, which is used as the independent variable in Figure 11(b). In particular, θ values are negatively correlated with elevation, with coefficient of determination equal to 0.86, 0.86 and 0.78

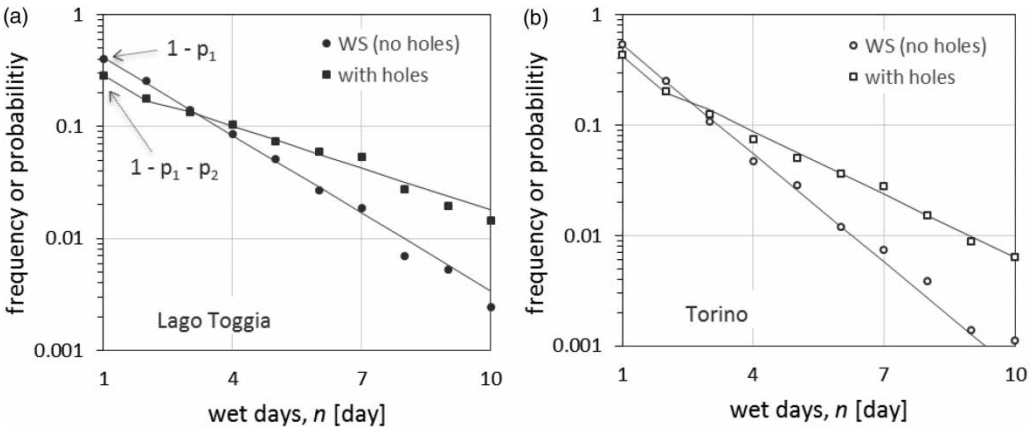

Figure 10 | Rainy days' sequences with 1-day dry spells (with gaps) for (a) Lago Toggia and (b) Torino stations. WS geometric distribution (no gaps), shown in Figure 7, is also reported for comparison.

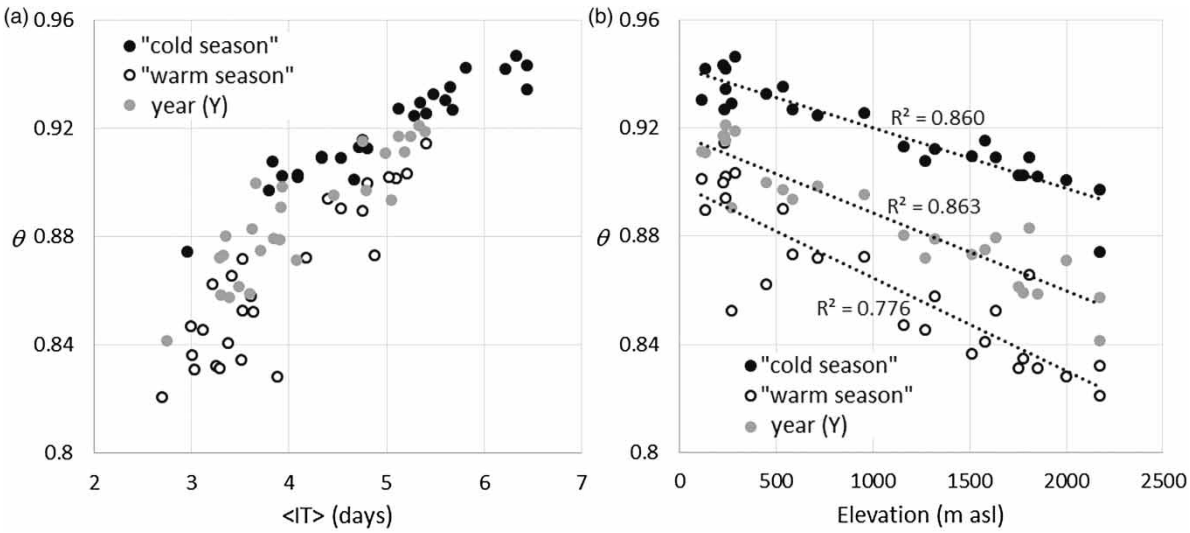

Figure 11 | θ parameter values of the Lerch *pmf* (a) vs. average IT and (b) vs. elevation, for the 'cold season', for the 'warm season' and for the whole year (Y) (all 26 stations).

for the 'cold season', the 'warm season' and for the whole year data sets, respectively.

Figure 12 compare parameters θ, s and a, of the 'warm season' with those of the 'cold season'. It is evident that θ values of the 'warm season' show a relationship with the corresponding θ values of the 'cold season', whereas s and a parameters do not show a clear behaviour, with a similar and narrow range of variability. Furthermore, no relationship of s and a parameters with other physical/climatic variables have been found. In the same figure, the corresponding values of *Lch* parameters obtained for the Sicily region in a previous work (Agnese *et al.* 2014) are reported. Similar considerations are valid for the parameters associated with the Sicily region, although the values are different, and their range of variability is wider, compared to that of the Piedmont and Aosta Valley area, likely indicating the different precipitation regime of the two areas. It is interesting that the warm season Sicily's θ, s and a values are in a completely different range. The θ values cold season also show a dfferent range.

To conclude the analysis of the results, Figure 13 shows the average cluster size $<WS>$, which according to the geometric distribution (see Equation (2)) is equal to $1/(1-p_1)$, as a function of the *IT* harmonic mean, for the 'cold season', for the 'warm season' and for the whole year data set. A clear relationship is detected, according to the fact that the harmonic mean enhances low *IT* values. Moreover, the considered seasons slightly influence this relationship.

An application of the described *Lch pdf* at the regional scale was carried out, on the basis of the good agreement between empirical and theoretical distribution. In particular, this application utilizes the good agreement in the wide range of probability values, including extremes with low-level probability of exceedance. Thus, the application was performed for a fixed long dry spell ($DS = 45$ days) and for a fixed wet spell ($WS = 5$ days), the values of which, and their associated spatial distribution, can play an important role in farm and environmental management. In fact, high duration of *DS* influences crops' growth, whereas the probability of occurrence of long precipitation spells is associated with environmental risks. Results are represented in terms of return period, T (years) and are illustrated in Figure 14(a) and 14(b). The expected return period was evaluated by using the known average *IT* values calculated for the whole year record (Y), reported in Table 4, according to $T = (1/(P\{DS \geq k\}\, N_{ev})) = (\langle IT \rangle / (P\{DS \geq k\}365))$. The probability $P\{DS \geq k\}$ was calculated for each station by Equation (3). The analysis was carried out for the whole year data set, thus separated 'cold season' and 'warm season' series (Table 4) were not considered. It is interesting to observe that the high elevation areas (compare Figure 1), are generally characterized by lower return period of $WS = 5$ days and higher return period of $DS = 45$ days, if compared with the low elevation areas. It is important to highlight that the expected return period for the above-mentioned was determined under the assumption that both *WS* and *DS* are time independent and uncorrelated random variables. In this particular case, the extended return period formulation introduced by Fernández & Salas (1999) and then by Volpi *et al.* (2015),

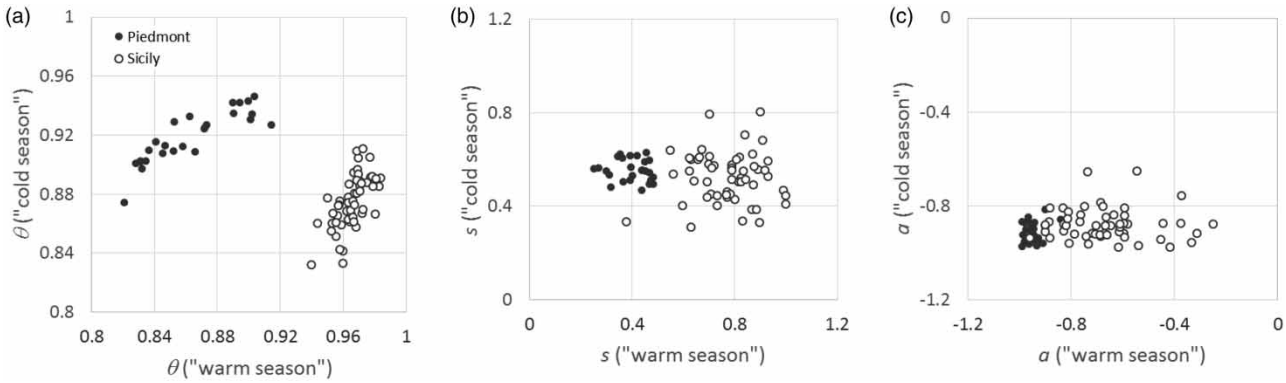

Figure 12 | Relation between 'cold season' and 'warm season' Lerch *pmf* parameters, θ, s and a, estimated for all 26 stations considered in this study, in Piedmont and Aosta Valley regions, and for the 55 Sicilian stations considered in Agnese *et al.* (2014).

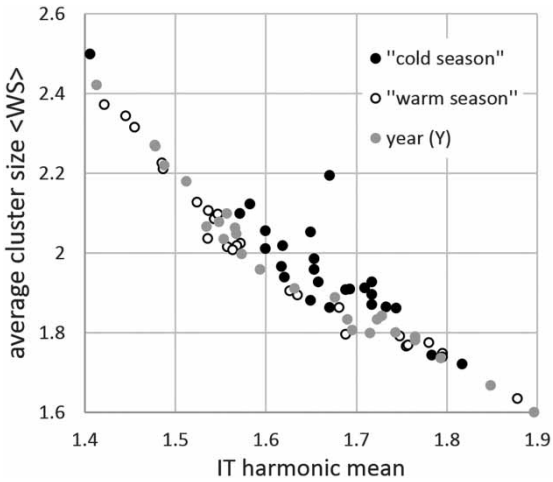

Figure 13 | Average cluster size <WS> vs. the IT harmonic mean, for the 'cold season', for the 'warm season' and for the whole year (Y), for the entire period (1919–2009) and for the considered 26 stations.

which differentiated the return period as an average occurrence interval or as an average recurrence interval, according to a not time-dependent or a time-dependent condition, respectively, leads to the same results.

Interestingly, for the case of time-dependent random variables, Volpi *et al.* (2015) also introduced an equivalent return period (ERP), which resembles the classical definition of return period, but in the case of independence. The ERP, which for this study matches the classical return period formulation considered above, is able to control the probability of the random variable under the time-dependent condition, preserving the virtue of the classical return period formulation that is insensitive to the time-dependent condition.

CONCLUSION

The statistical analysis performed in this work demonstrated the applicability of the discrete three-parameter Lerch probability distribution to the frequency distribution of the inter-arrival time-series in the NW of Italy, characterized by precipitation regimes ranging between that of hilly areas influenced by the Tyrrhenian Sea, to that of the Po Plain and to that of the Alps. In the past, the same distribution has already been successfully tested in the Mediterranean area (Sicily island) (Agnese *et al.* 2014), which has very different precipitation regimes from those investigated in this study.

The analysis was successfully applied to 70–90 years' long inter-arrival time series derived from 26 daily precipitation stations records and the *i.i.d.* hypothesis of the *IT* random variable was then checked. Because of the climatic spatial variability of these regions, and of Sicily, the application of the three-parameter Lerch distribution has demonstrated its wide applicability. The spatial variability

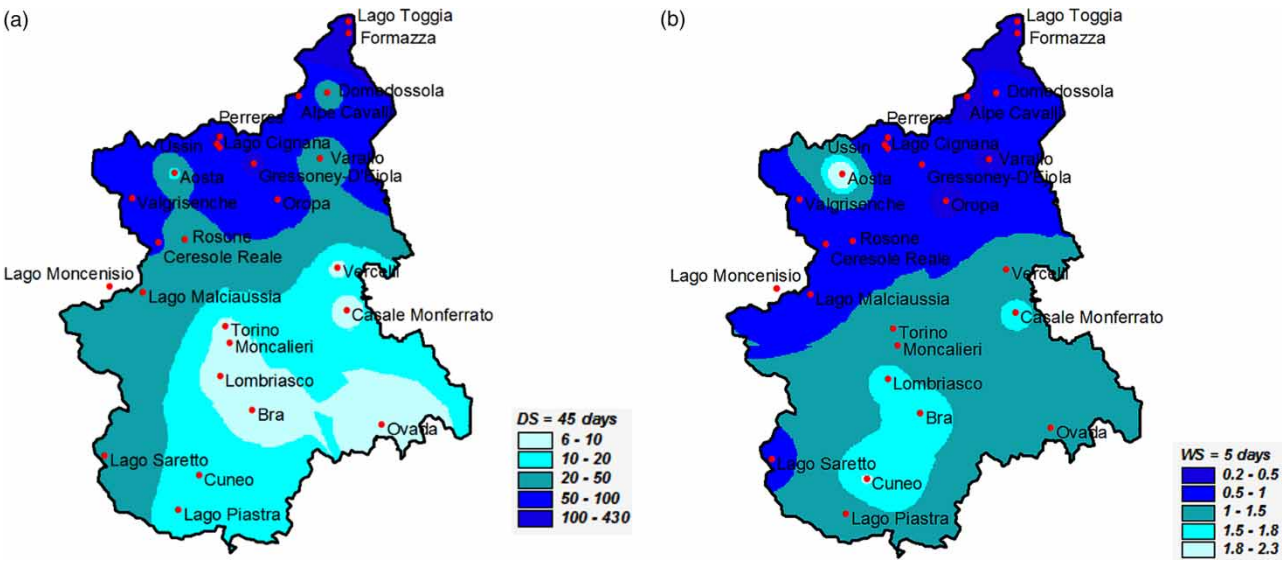

Figure 14 | Return period, T (years), of (a) event dry spell DS $\geq$ 45 days and (b) event wet spell WS $\geq$ 5 days.

of all the three parameters of the distribution, for both 'cold' and 'warm' seasons, is smaller than in Sicily. In particular, the differences found in the θ parameter, confirm its intrinsic character of scale parameter, whereas both s and a parameters appear to be purely fitting parameters, but with markedly different ranges during the cold season, with respect to Sicily. In this paper, it is shown that the Lerch distribution successfully describes the sequences of rainy days interrupted by 1-day DS, with useful practical applications from a hydrological point of view. Moreover, the equivalence on using a single *IT* three-parameter Lerch distribution, in place of two distributions, namely, the *DS* two-parameter polylogarithmic and *WS* one-parameter geometric ones, is also tested. Finally, an application at the regional scale describing the probability of exceedance of a long wet spell and a long dry spell, has confirmed the value of this distribution in quantifying the wide spatial variability of the precipitation statistics in the considered area.

ACKNOWLEDGEMENT

Research was supported by PRIN 2015 project granted by MIUR (Italian Ministry for University and Research) No. 2015AKR4HX, by the H2020 project 'ECOPOTENTIAL: Improving Future Ecosystem Benefits Through Earth Observations', coordinated by CNR-IGG (http://www.ecopotential-project.eu) (it has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 641762), and by the CNR flag project NEXTDATA.

REFERENCES

Agnese, C., Baiamonte, G., Cammalleri, C., Cat Berro, D., Ferraris, S. & Mercalli, L. 2012 Statistical analysis of inter-arrival times of rainfall events for Italian Sub-Alpine and Mediterranean areas. *Advances in Science and Research* **1**, 1–7. doi: 10.5194/asr-1-1-2012.

Agnese, C., Baiamonte, G. & Cammalleri, C. 2014 Modelling the occurrence of rainy days under a typical Mediterranean climate. *Advances in Water Research* **64**, 62–76. http://dx.doi.org/10.1016/j.advwatres.2013.12.005.

Agnese, C., Baiamonte, G., D'Asaro, F. & Grillone, G. 2016 Probability distribution of peak discharge at the hillslope scale generated by Hortonian runoff. *Journal of Irrigation and Drainage E-ASCE* **142**, 2. doi: 10.1061/(ASCE)IR.1943-4774.0000973., 04015052.

Aksenov, S. V. & Savageau, M. A. 2005 *Some Properties of the Lerch Family of Discrete Distributions*. http://arXiv.org/abs/math/0504485v1.

Baiamonte, G. & D'Asaro, F. 2016 Discussion of 'Analysis of extreme rainfall trends in Sicily for the evaluation of depth-duration-frequency curves in climate change scenarios' by Lorena Liuzzo and Gabriele Freni. *Journal of Irrigation and Drainage E-ASCE* **21**, 6. doi: 10.1061/(ASCE)HE.1943-5584.0001381, 07016005.

Buishand, T. A. 1977 *Stochastic Modelling of Daily Rainfall Sequences*. H. Veenman & Zonen BV, Wageningen, pp. 212.

Buishand, T. A. 1978 Some remarks on the use of daily rainfall models. *Journal of Hydrology* **36**, 295–308.

Canone, D., Previati, M., Bevilacqua, I., Salvai, L. & Ferraris, S. 2015 Field measurements based model for surface irrigation efficiency assessment. *Agricultural Water Management* **156**, 30–42.

Chatfield, C. 1966 Wet and dry spells. *Weather* **21**, 308–310.

Cox, D. 1970 *Renewal Theory*. Methuen & Co, London, pp. 142.

de Groen, M. M. & Savenije, H. H. 2006 A monthly interception equation based on the statistical characteristics of daily rainfall. *Water Resources Research* **42**, W12417. doi:10.1029/2006WR005013.

Eichner, J. F., Kantelhardt, J. W., Bunde, A. & Havlin, S. 2011 The statistics of return intervals, maxima, and centennial events under the influence of long-term correlations. In: *In Extremis: Disruptive Events and Trends in Climate and Hydrology* (J. Kropp & H.-J. Schellhuber, eds). Springer, Berlin, Heidelberg, pp. 2–43.

Erdélyi, A., Magnus, W., Oberhettinger, F. & Tricomi, F. G. 1981 *Higher Transcendental Functions*, Vol. 1. Krieger, New York, pp. 30–31.

Feller, W. 1968 *An Introduction to Probability Theory and Its Applications*, Vol. 1. John Wiley & Sons, New York, USA.

Fernández, B. & Salas, J. D. 1999 Return period and risk of hydrologic events. I: mathematical formulation. *Journal of Hydrologic Engineering* **4** (4), 297–307.

Ferraris, S., Baiamonte, G., Cammalleri, C., Agnese, C., Cat Berro, D. & Mercalli, L. 2014 Spatial variability of dry spells duration statistical distributions. In: *Trending Now: Water, 7th International Scientific Conference on the Global Water and Energy Cycle*, The Hague, The Netherlands, July 14–17.

Foufoula-Georgiou, E. & Lettenmaier, D. P. 1986 Continuous-time versus discrete-time point process models for rainfall occurrence series. *Water Resources Research* **22** (4), 531–542. doi:10.1029/WR022i004p00531.

Good, I. J. 1953 The population frequencies of species and the estimation of population parameters. *Biometrika* **40** (3–4), 237–264.

Gupta, P. L., Gupta, R. C., Ong, S. H. & Srivastava, H. M. 2008 A class of Hurwitz-Lerch Zeta distributions and their

applications in reliability. *Applied Mathematics and Computation* **196**, 521–531. doi:10.1016/j.amc.2007.06.012.

Hope, A. C. A. 1968 A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society B* **30**, 582–598.

Jonquière, A. 1889 Note sur la série $\sum_{n=1}^{n=8} (x^n/n^s)$. *Bulletin de la Société Mathématique de France*. **17**, 142–152.

Kemp, A. W. 2006 Polylogarithmic distributions. In: *Encyclopedia of Statistical Science* (S. Kotz, D. L. Banks & C. B. Read, eds). John Wiley & Sons, New York.

Kottegoda, N. T. & Rosso, R. 1997 *Statistics, Probability, and Reliability for Civil and Environmental Engineers*. McGraw-Hill, New York.

Manfreda, S. & Rodrıguez-Iturbe, I. 2006 On the spatial and temporal sampling of soil moisture fields. *Water Resources Research* **42**, W05409. doi:10.1029/2005WR004548.

Martínez-Rodríguez, A. M., Sáez-Castillo, A. J. & Conde-Sánchez, A. 2011 Modelling using an extended Yule distribution. *Computational Statistics and Data Analysis* **55** (1), 863–873. doi:10.1016/j.csda.2010.07.014.

Mercalli, L. & Cat Berro, D. 2008 *Atlante Climatico Della Valle D'Aosta*. Società Meteorologica Subalpina, Moncalieri, Torino, Italy, pp. 405.

Pettitt, A. N. 1979 A non-parametric approach to the change-point detection. *Applied Statistics* **28**, 126–135.

Pohlert, T. 2018 *Non-Parametric Trend Tests and Change-Point Detection*. Package 'trend', Version: 1.1.0, License: GPL-3. http://CRAN.R-project.org/package = trend.

Rallo, G., Baiamonte, G., Juárez, J. & Provenzano, G. 2014 Improvement of FAO-56 model to estimate transpiration fluxes of drought tolerant crops under soil water deficit: application for olive groves. *Journal of Irrigation and Drainage Engineering* **140** (9), 1–8. 10.1061/(ASCE)IR.1943-4774.0000693, A4014001.

Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. 1987 Some models for rainfall based on stochastic point process. *Proceedings of the Royal Society A* **410**, 269–288.

Serinaldi, F., Kilsby, C. G. & Lombardo, F. 2018 Untenable nonstationarity: an assessment of the fitness for purpose of trend tests in hydrology. *Advances in Water Resources* **111**, 132–155.

Strupczewski, W. G., Kochanek, K., Bogdanowicz, E. & Markiewicz, I. 2013 Inundation risk for embanked rivers. *Hydrology and Earth System Sciences* **17**, 3111–3125. doi:10.5194/hess-17-3111-2013.

Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F. & Koutsoyiannis, D. 2015 One hundred years of return period: strengths and limitations. *Water Resources Research* **51** (10), 8570–8585.

Wilks, S. S. 1938 The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* **9** (1), 60–62. http://dx.doi.org/10.1214/aoms/1177732360.

Zhang, W., Yan, Y., Zheng, J., Li, L., Dong, X. & Cai, H. 2009 Temporal and spatial variability of annual extreme water level in the Pearl River Delta region, China. *Global and Planetary Change* **69**, 35–47.

Zörnig, P. & Altmann, G. 1995 Unified representation of Zipf distributions. *Computational Statistics and Data Analysis* **19**, 461–473. http://dx.doi.org/10.1016/0167-9473(94)00009-8.

First received 15 March 2018; accepted in revised form 10 June 2018. Available online 16 July 2018