11th EMS Annual Meeting
10th European Conference on Applications of Meteorology

Forecasting the weather Ensemble techniques in probabilistic weather prediction

Berlin | Germany | 12 – 16 September 2011

TREND OF INTER-ARRIVAL TIMES OF RAINFALL EVENTS FOR ITALIAN SUB-ALPINE AND MEDITERRANEAN AREAS

C. Agnese[1], G. Baiamonte[1], I. Bevilacqua[2], C. Cammalleri[3], D. Cat Berro[4], S. Ferraris[2], and L. Mercalli[4]

1) Dept. SAGA - Università degli Studi di Palermo (IT)
2) Dept. EIAFA - Università di Torino (IT)
3) Dept. DICA - Università degli Studi di Palermo (IT)
4) Società Meteorologica Italiana, Castello Borello, Bussoleno (IT)

Viale delle Scienze Ed. 8 – 90128 Palermo (Italy).
e-mail: cammalleri@idra.unipa.it; Tel: +39(0)91 23896574; Fax: +39(0)91 6657749.

MOTIVATION AND RESEARCH GOALS

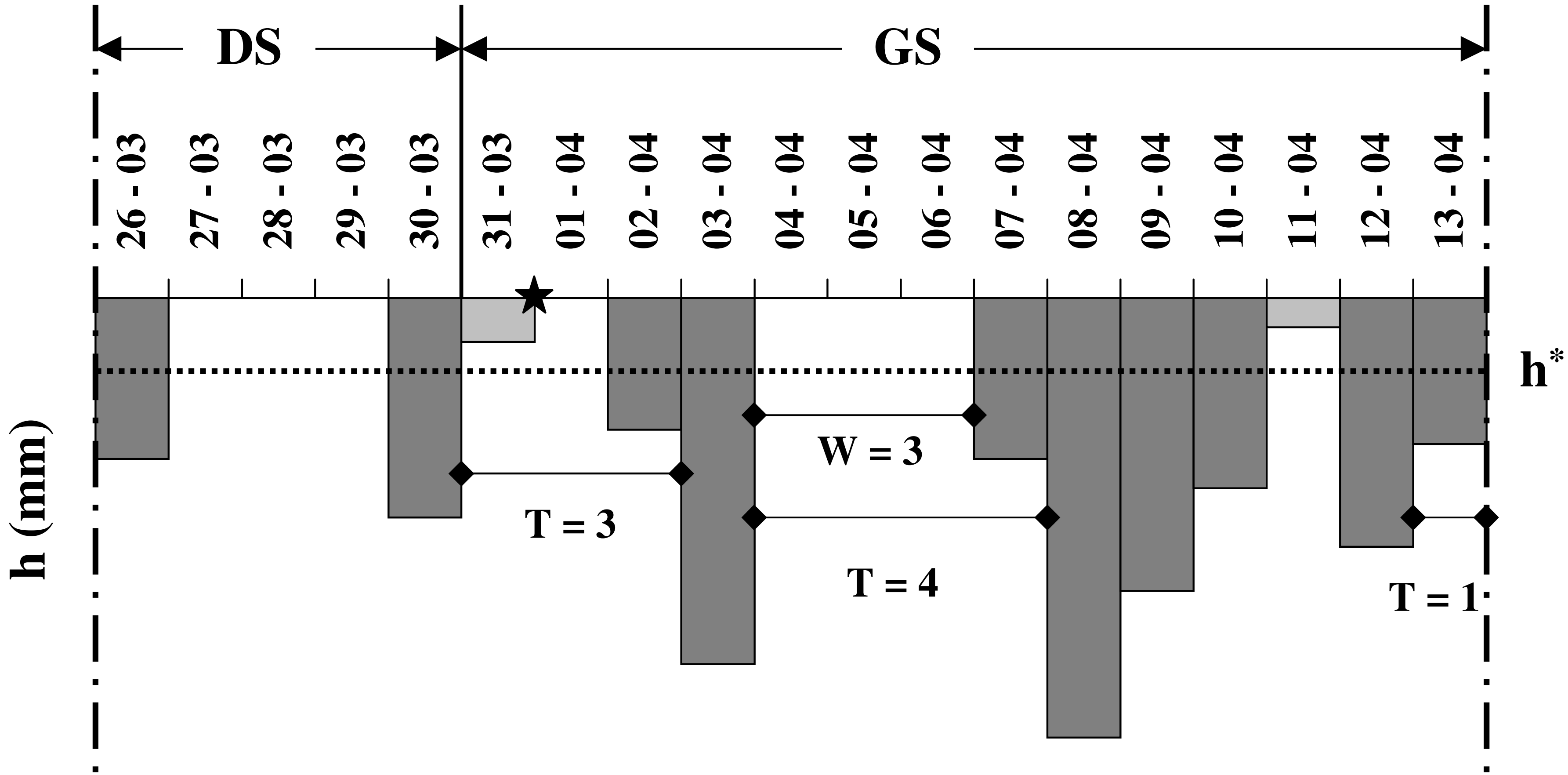
Modeling of rainfall statistical structure represents an important research area in hydrology, meteorology, atmospheric physics and climatology, because of the several theoretical and practical implications in both predictive analyses and simulation studies.

At the daily scale, rainfall process could be represented by a time-series in which each element (day) is marked by a rainfall depth, $h \geq 0$. A day is rainy if $h > h^*$ (threshold).

This work deals with the temporal structure of the rainfall, aiming to reproduce the most important observed features of rainfall time-series (intermittency, persistency, non-stationarity) by means of parsimonious probabilistic distribution, i.e., one- or two- parameter.

Theoretical representation of rainfall temporal structure allows more robust analysis of the time series to detect temporal trends.

INTER-ARRIVAL TIME: Definition



T = {4, 3, 1, 4, 1, 1, 1, 2, 1} **Inter-arrival times**

W = {3, 2, 3, 1} **Waiting times**

INTER-ARRIVAL TIME: Properties

Some fundamental recursive properties were observed on inter-arrival time series:

- **the inter-arrival time (in a defined time interval, i.e., days) is a discrete random variable;**
- **the observed frequencies of T decrease monotonically and they are considerable for T = 1 (≈ 50%);**
- **the observed frequencies of relatively long inter-arrival times are small but still significant (10^{-3} – 10^{-4}) in the studied climates;**
- **rainfall phenomenon shows the tendency to clustering.**

A discrete, monotonically decreasing probability function with a long (or heavy) tail and an intrinsic ability to reproduce clusters.

INTER-ARRIVAL TIME: Modelling (1/2)

On the hypothesis that *T* are independent and identically distributed (iid) random variables, the number of rainfall events that occur in a defined interval can be described by discrete counting process (Bernoulli trials).

Logarithmic distribution (m)

$$P\{T=k\}=-\frac{m^k}{k\ln(1-m)} \qquad (k\in\mathbb{N};\quad 0<m\le 1)$$

Yule distribution (ρ)

$$P\{T=k\}=\frac{\rho\Gamma(k)\Gamma(\rho+1)}{\Gamma(\rho+1+k)} \qquad (k\in\mathbb{N};\quad \rho>0)$$

$\Gamma(\cdot)$ = gamma function

Zeta distribution (s)

$$P\{T=k\}=\frac{1}{\zeta(s)k^s} \qquad (k\in\mathbb{N};\quad s>1) \quad \text{with} \quad \zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s} \qquad (n\in\mathbb{N};\quad s>1)$$

INTER-ARRIVAL TIME: Modelling (2/2)

E-Yule distribution **(ρ, λ)**

$$P\{T = k\} = \frac{(\rho+1)\lambda^{k-1}B(k,\rho+1)}{{}_2F_1(1,1;\rho+2,\lambda)} \qquad (k \in \mathbb{N}; \quad \rho > 0; \quad 0 < \lambda \leq 1)$$

where B(·) is the beta function and $_2F_1(\cdot)$ is the Gaussian hypergeometric function.

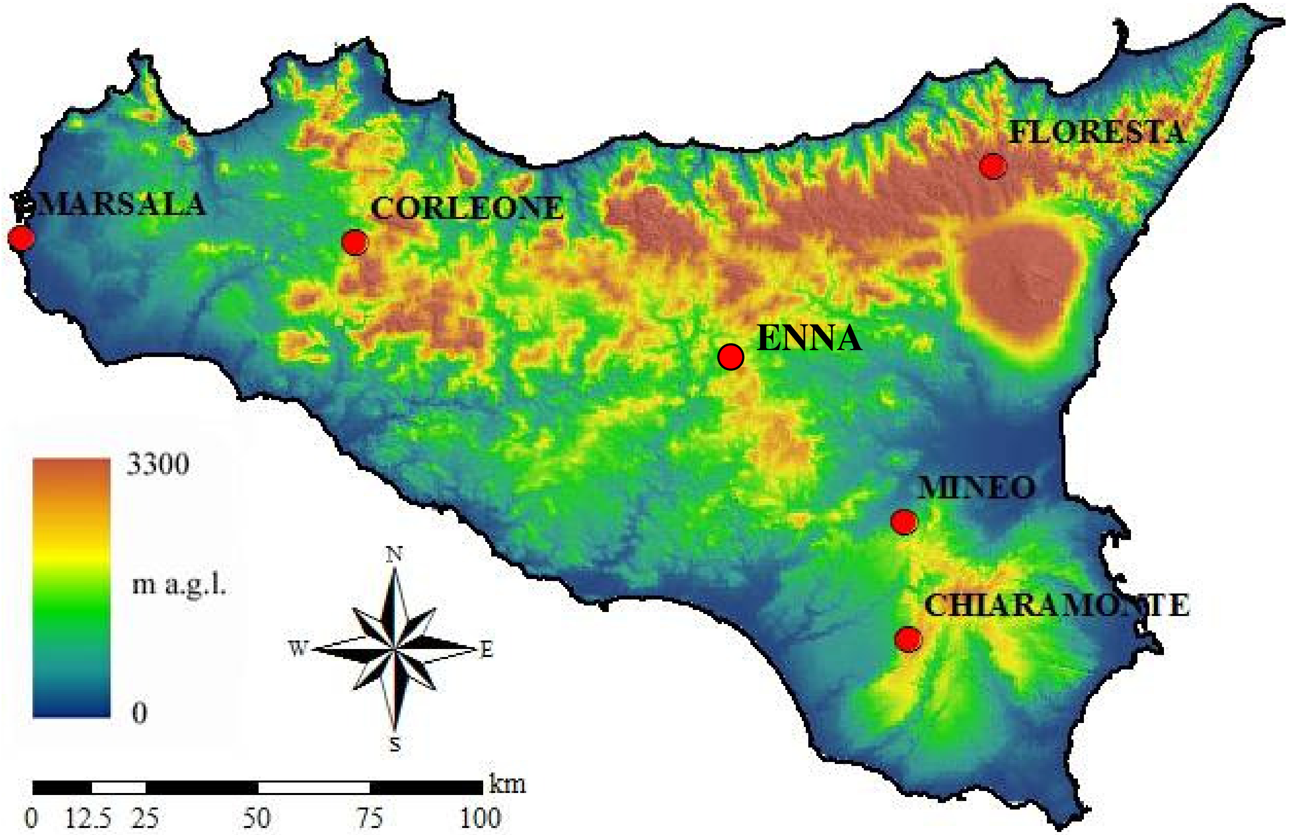
PolyLog distribution **(s, w)**

$$P\{T = k\} = \frac{w^k}{L_i(w,s)k^s} \qquad (k \in \mathbb{N}\,; 0 < w \leq 1)$$

$$L_i(w,s) = \sum_{n=1}^{\infty} \frac{w^n}{n^s} \qquad (n \in \mathbb{N}\,; 0 < w \leq 1)$$

***L_i* is the polylogarithmic function, also known as the Jonquière's function.**

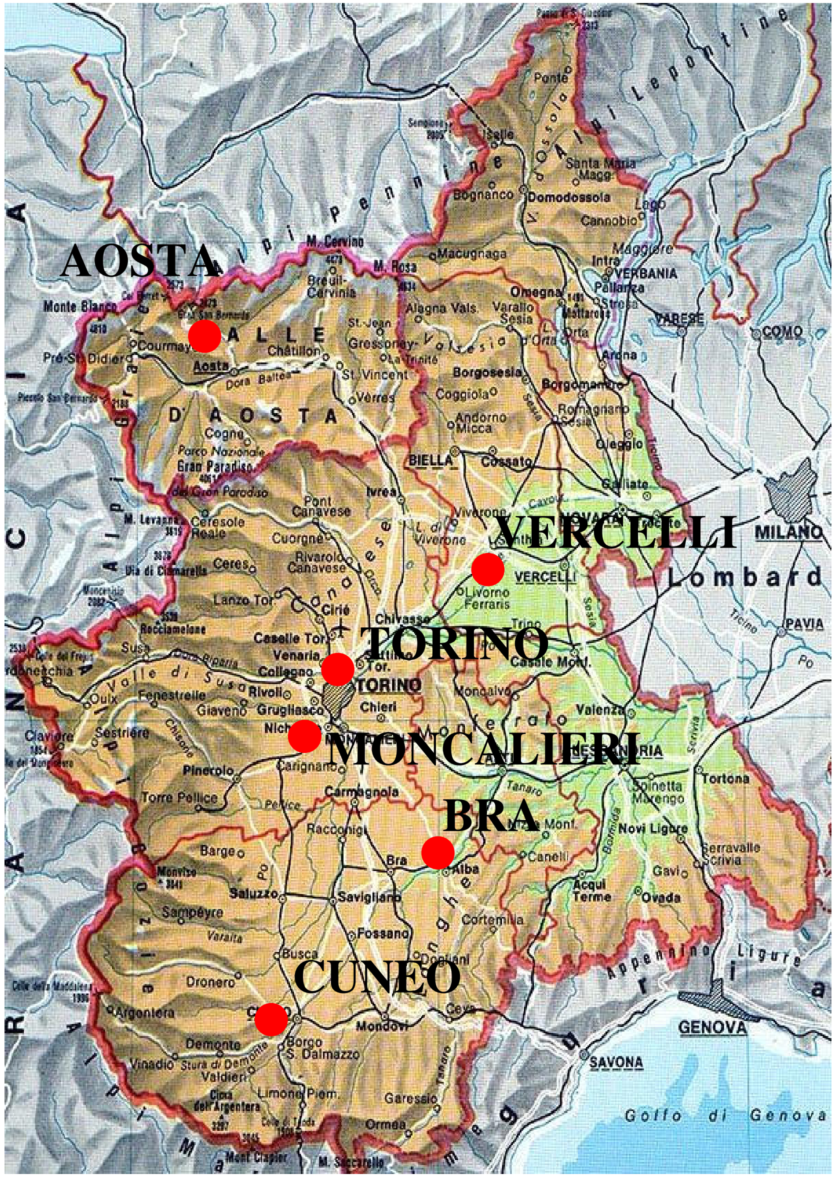
DATASET (1/3)



Station	Start	End
CHIARAMONTE	1916	2005
CORLEONE	1917	2005
ENNA	1919	2004
FLORESTA	1916	2005
MARSALA	1919	2005
MINEO	1916	2003

Mediterranean Environment (SIC)

DATASET (2/3)



Station	Start	End
AOSTA	1891	2010
BRA	1862	2010
CUNEO	1901	2010
MONCALIERI	1866	2010
TORINO	1802	2010
VERCELLI	1927	2007

Sub-Alpine Environment (PMT)

DATASET (3/3)

Statistical analyses were performed on reduced time-series (from 1926 to 2005, 80 years) by considering for each station sub-periods of 20 years overlapped for 15 years (13 sub-sets).

The analyses were repeated for:

1. Whole year (Y);
2. Growing season (GS), from April to September;
3. Dormant season (DS) from October to March.

468 fittings were performed on these datasets

(12 stations × 3 seasons × 13 sub-periods).

MODELS FITTING: Maximum Likelihood (1/2)

For the selected distribution, the ML (Maximum Likelihood) parameters estimation was performed.

Logarithmic distribution (m)

$$\frac{\partial}{\partial m}\{\ln L\} = \left[\frac{m}{(m-1)\log(1-m)} - \frac{1}{n}\sum_{i=1}^{n}\tau_i\right]^2$$

Yule distribution (ρ)

$$\frac{\partial}{\partial \rho}\{\ln L\} = n\ln(\rho+1) + \sum_{i=1}^{n}\ln \mathrm{B}(\tau_i, \rho+1) - n\ln\left(\frac{\rho+1}{\rho}\right)$$

Zeta distribution (s)

$$\frac{\partial}{\partial s}\{\ln L\} = -\sum_{i=1}^{n}\ln \tau_i - n\frac{\zeta'(s)}{\zeta(s)}$$

MODELS FITTING: Maximum Likelihood (2/2)

E-Yule distribution (ρ, λ)

$$\begin{cases} \ln L = n\ln(\rho+1) + \left(\sum_{i=1}^{n}\tau_i - n\right)\ln\lambda + \\ \sum_{i=1}^{n} B(\tau_i, \rho+1) - n\ln {}_2F_1(1, 1, \rho+2\ \lambda) \end{cases}$$

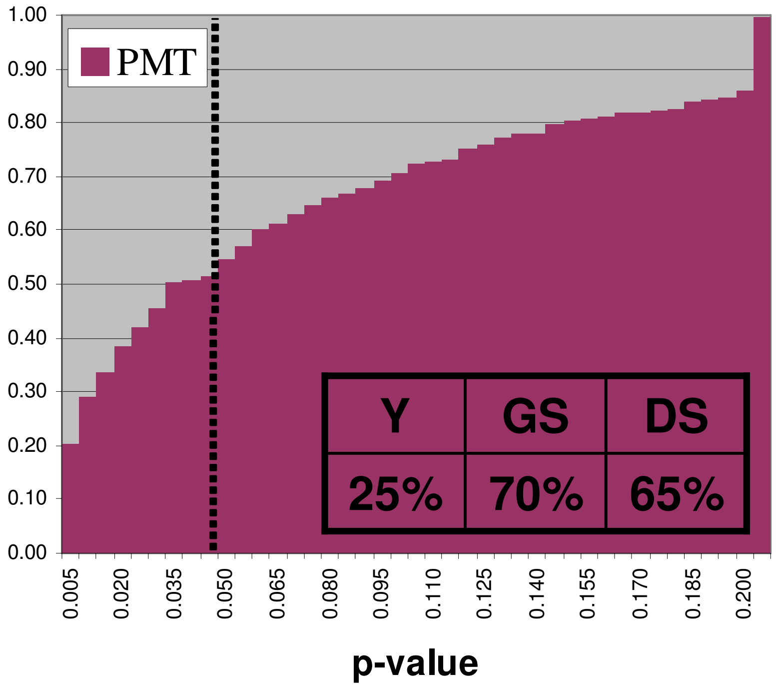
PolyLog distribution (s, w)

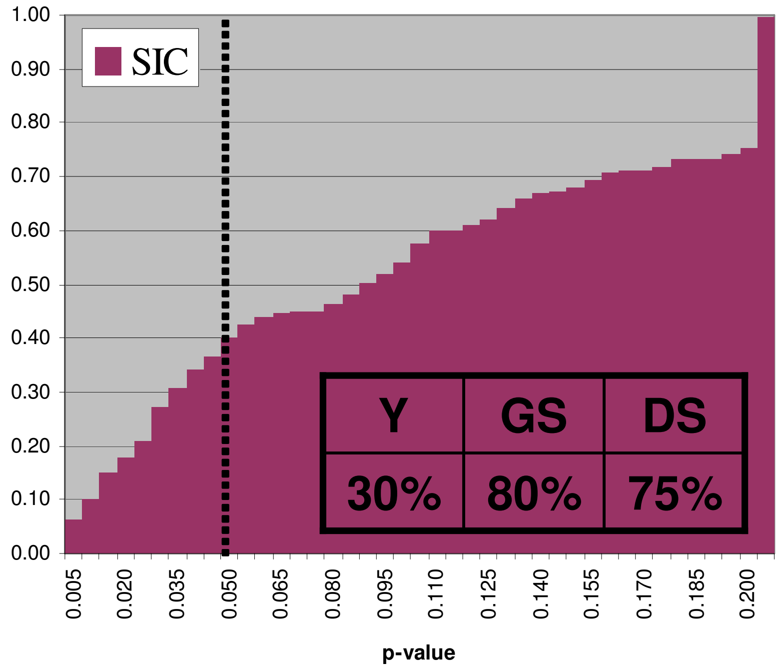
$$\begin{cases} \frac{\partial}{\partial w}\{\ln L\} = \frac{\sum_{i=1}^{n}\tau_i}{w} - n\frac{\frac{\partial}{\partial w}L_i(w,s)}{L_i(w,s)} \\ \frac{\partial}{\partial s}\{\ln L\} = -\sum_{i=1}^{n}\ln\tau_i - n\frac{\frac{\partial}{\partial s}L_i(w,s)}{L_i(w,s)} \end{cases}$$

TEST FITTING: Chi-square (1/3)

The **goodness of fit** was tested by calculating the **p-value** associated with that statistic value for a **Monte Carlo test** with **2000** replicates. This because is not appropriate the use of "standard" chi-square test when the distributions have long/heavy tails.

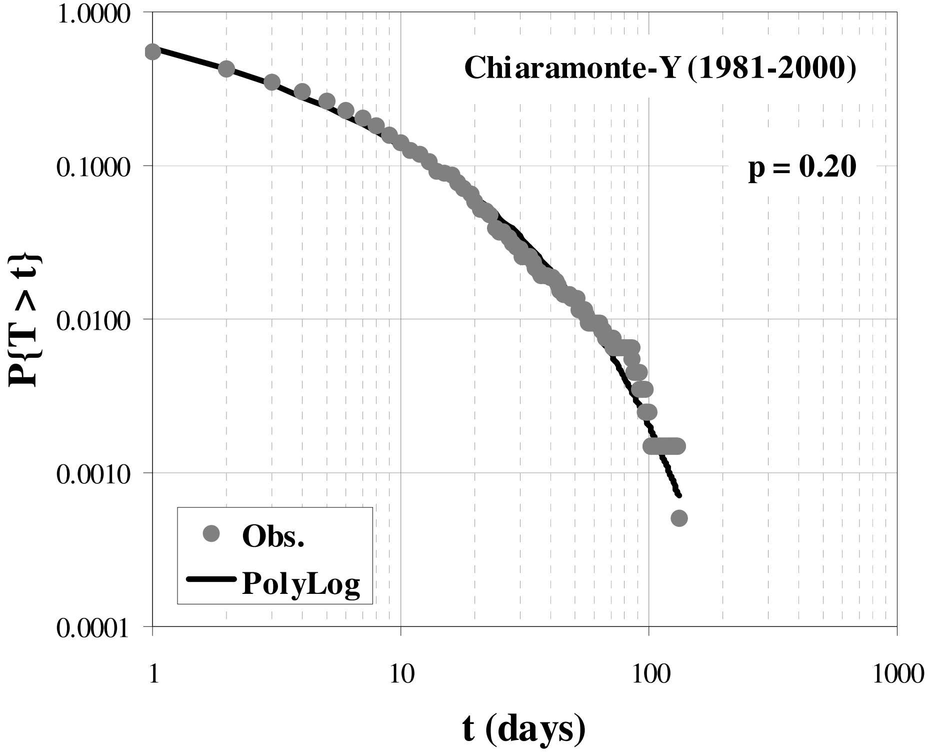
The **best** results were obtained for the **Polylog distribution**, so the results are reported just for this one.



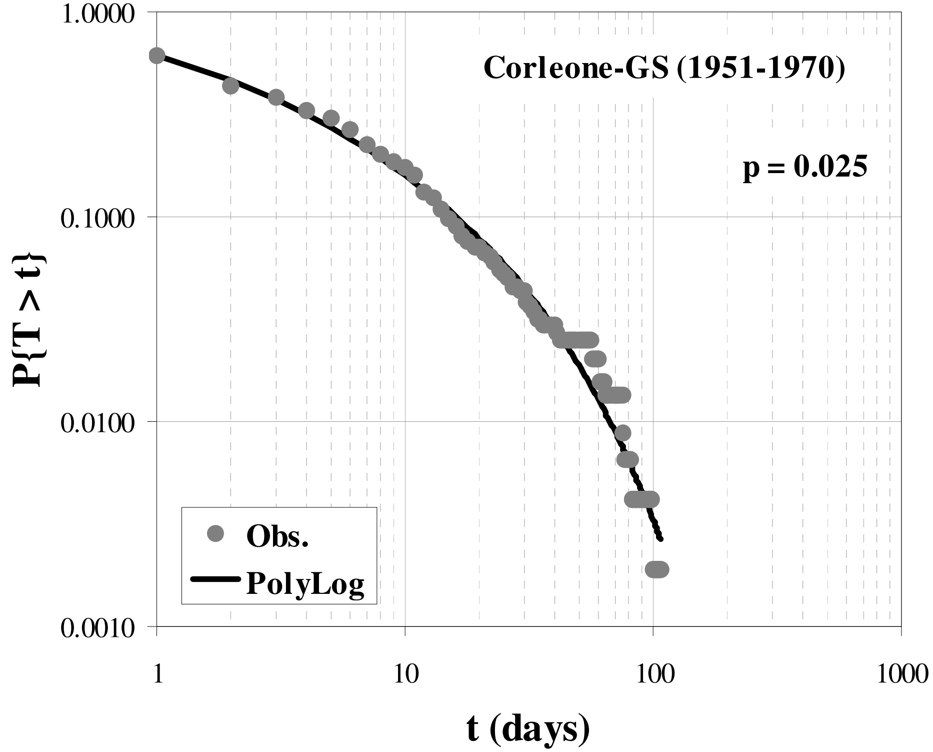


TEST FITTING: Chi-square (2/3)

s = 1.385; w = 0.977



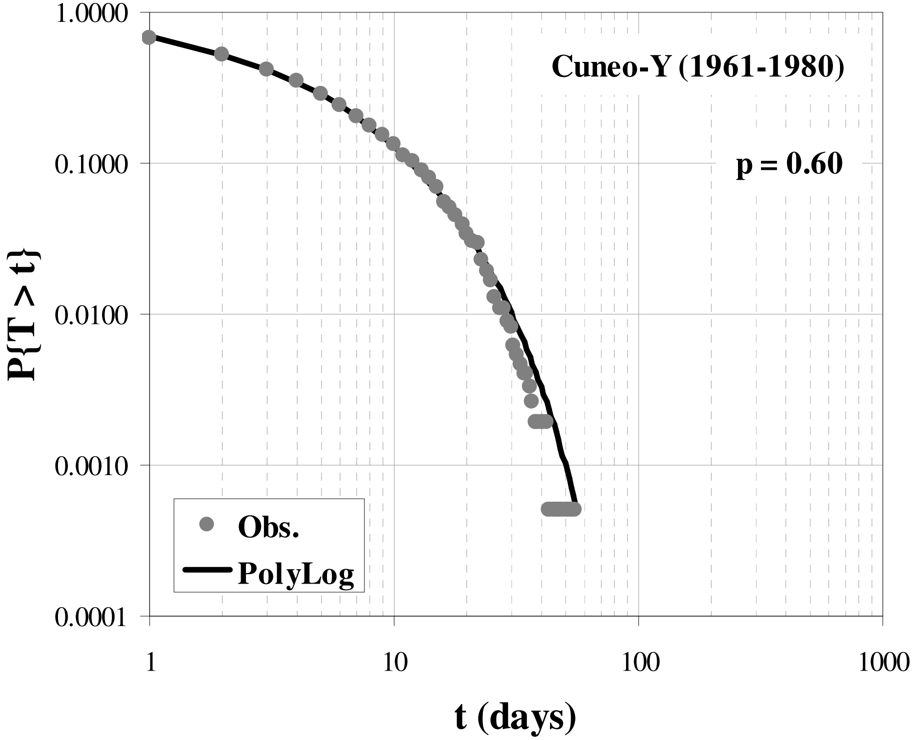
s = 1.325; w = 0.979



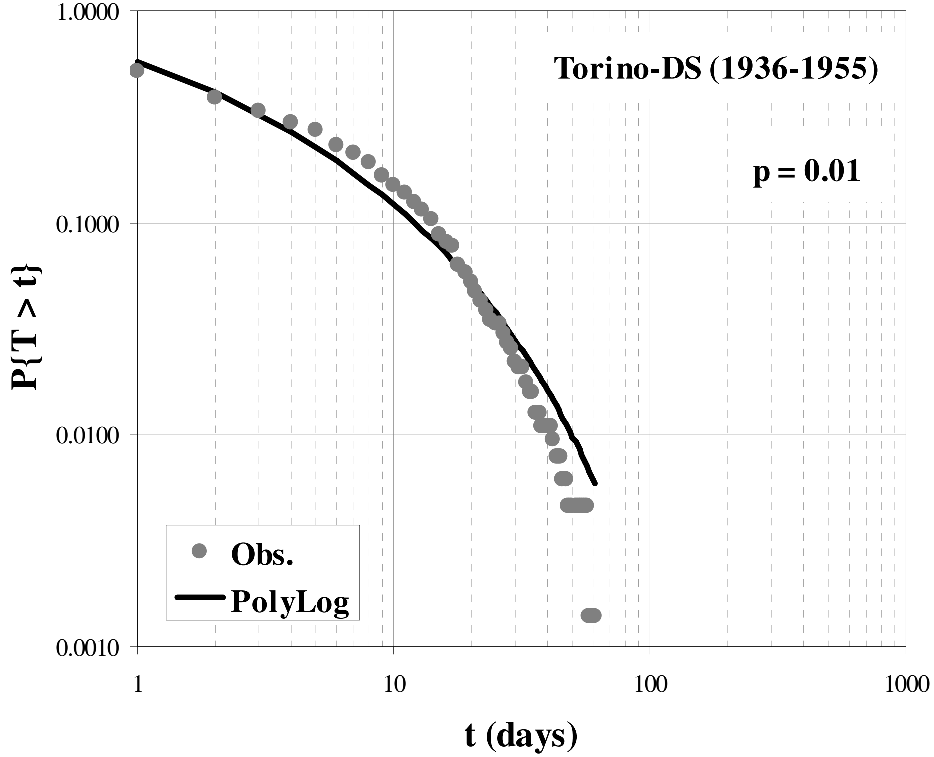
The empirical data seem well reproduced in the cases of p > 0.05 either for p ≈ 0.025, suggesting that a confidence level lower than 0.05 can be assumed as acceptable in some cases.

TEST FITTING: Chi-square (3/3)

s = 0.825; w = 0.911



s = 1.395; w = 0.973

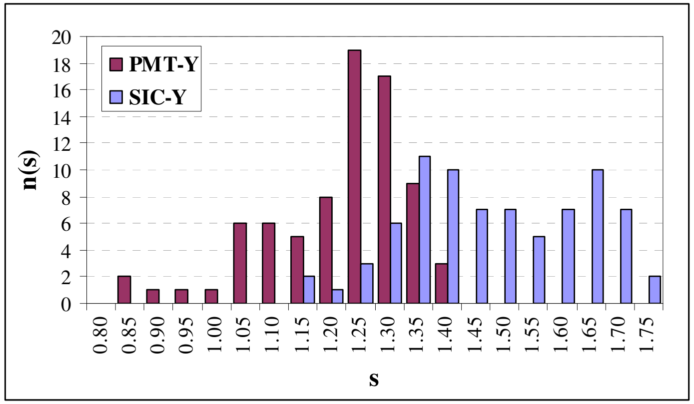


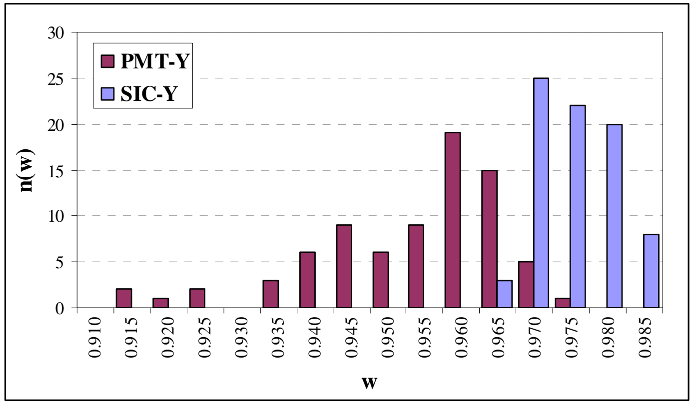
The data poorly reproduced are the ones characterized by $p < 0.02$ in the Montecarlo test.

The extreme values in PMT are generally lower than the SIC ones.

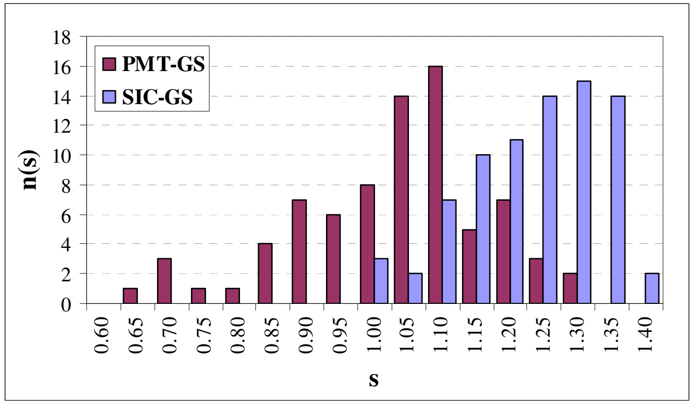
PARAMETERS COMPARISON

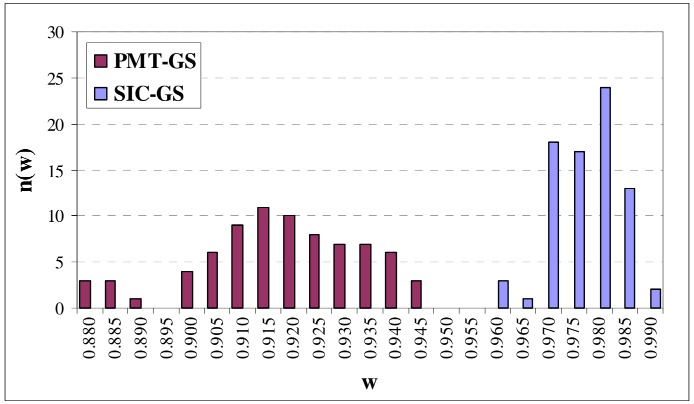
Y



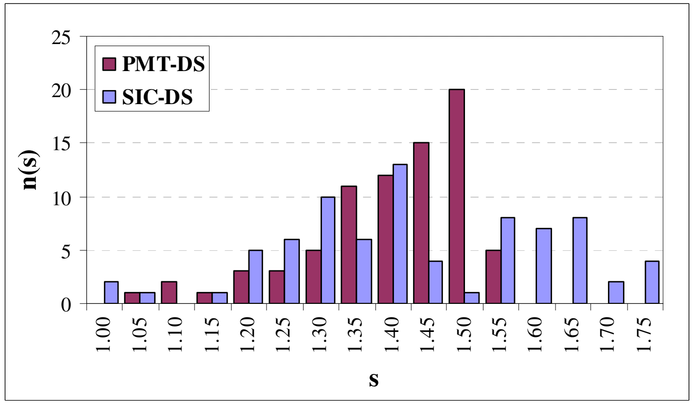


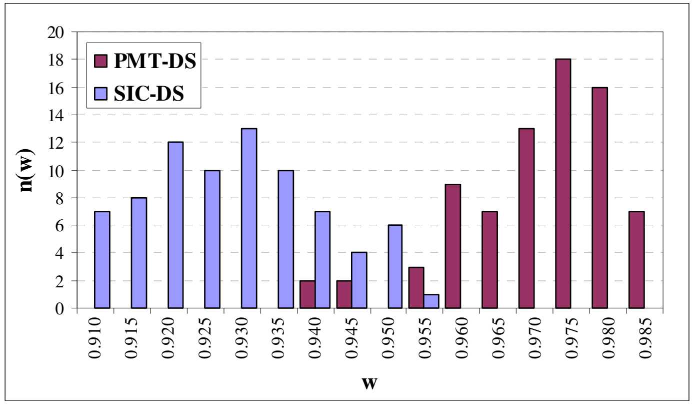
GS





DS



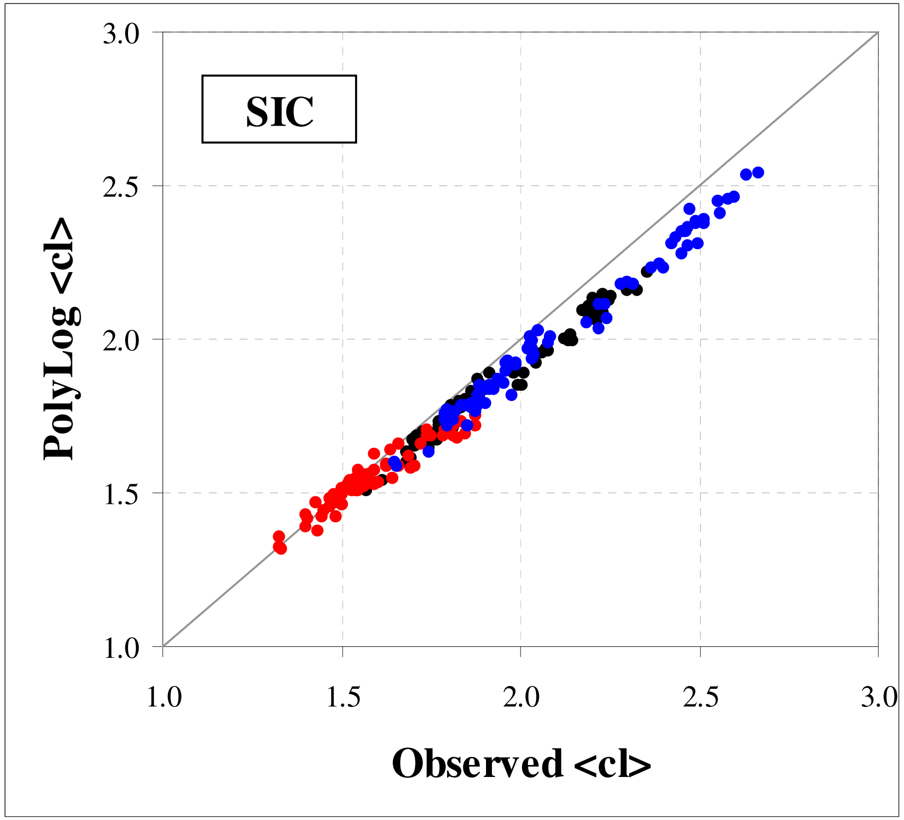


A stronger seasonality was observed in Mediterranean environment with respect to Sub-alpine one. The range of yearly s (w) parameter in SIC is very similar to the DS (GS) one.

TEST FITTING: Clusters (1/2)

The average cluster size can be generically computed as a simple function of the frequency of *T* = 1.

$$\langle cl \rangle = \frac{1}{1 - \mathrm{P}\{T = 1\}}$$



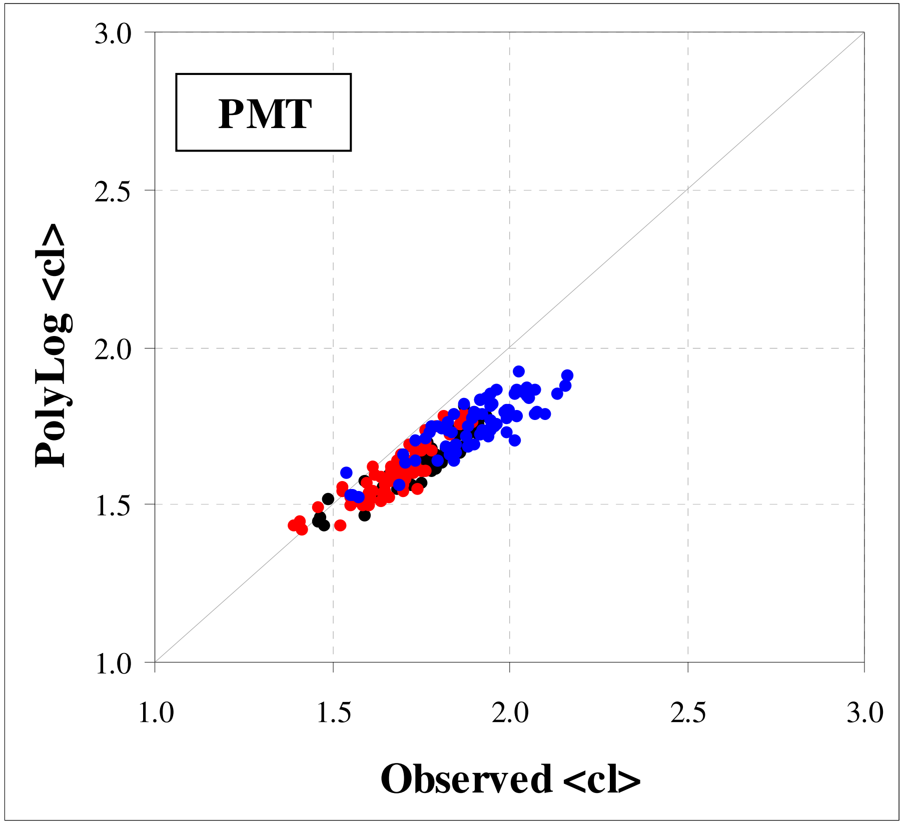
		Y	GS	DS
PMT	RE	7.3	5.0	8.3
	RMSD	0.13	0.08	0.16
SIC	RE (%)	4.6	3.1	4.6
	RMSD (d)	0.09	0.05	0.10

In SIC is possible to observe a significant seasonality in <cl>, also the global range of variability is quite wide (1.2 to 2.7).

TEST FITTING: Clusters (2/2)

The average cluster size can be generically computed as a simple function of the frequency of *T* = 1.

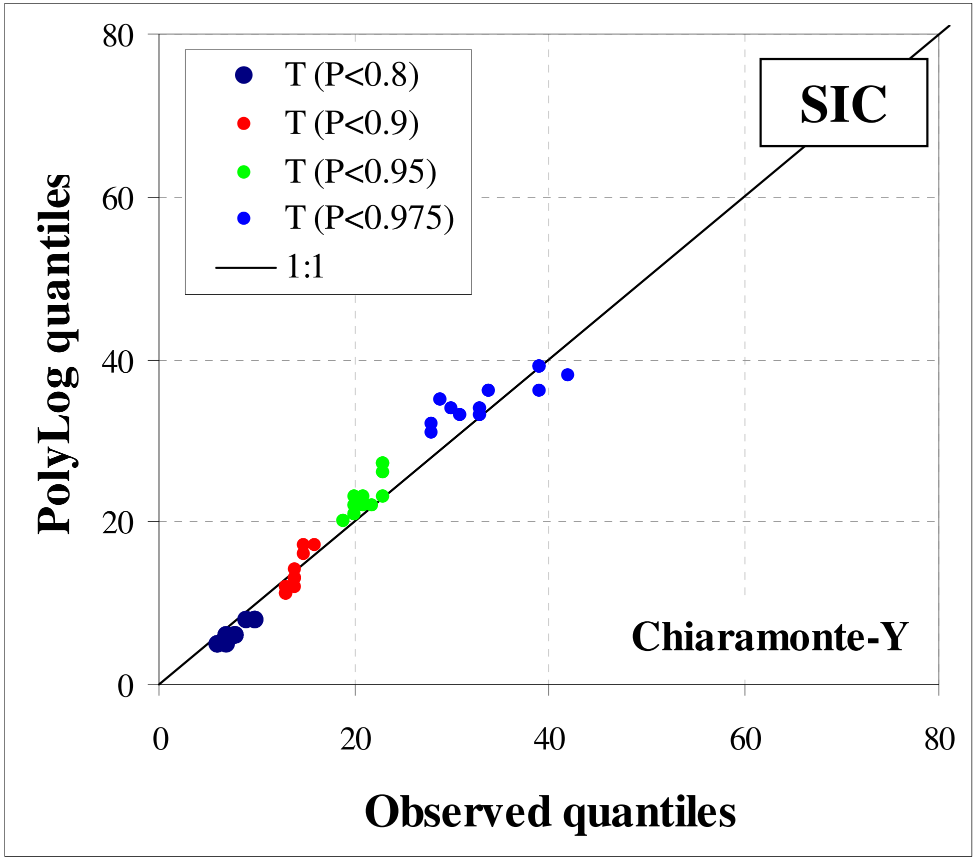
$$\langle cl \rangle = \frac{1}{1 - \mathrm{P}\{T = 1\}}$$

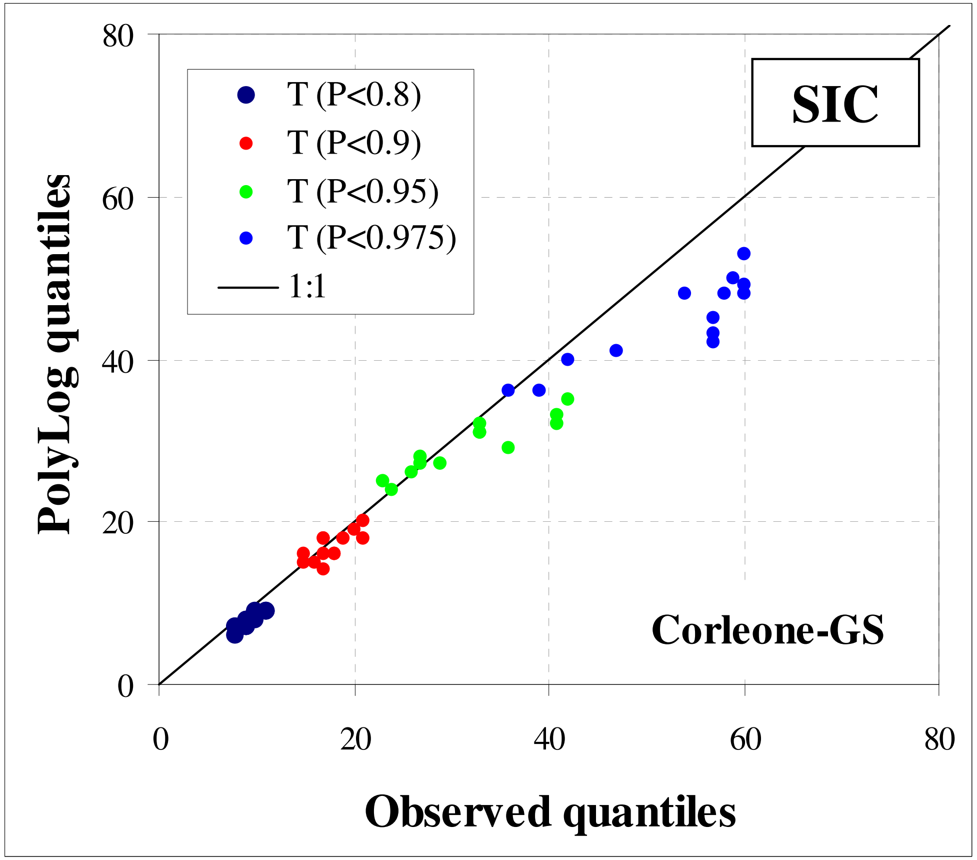


		Y	GS	DS
PMT	RE	7.3	5.0	8.3
	RMSD	0.13	0.08	0.16
SIC	RE (%)	4.6	3.1	4.6
	RMSD (d)	0.09	0.05	0.10

In PMT the range of variability of <cl> in the two seasons is quite similar, with values that ranging between 1.5 and 2.25.

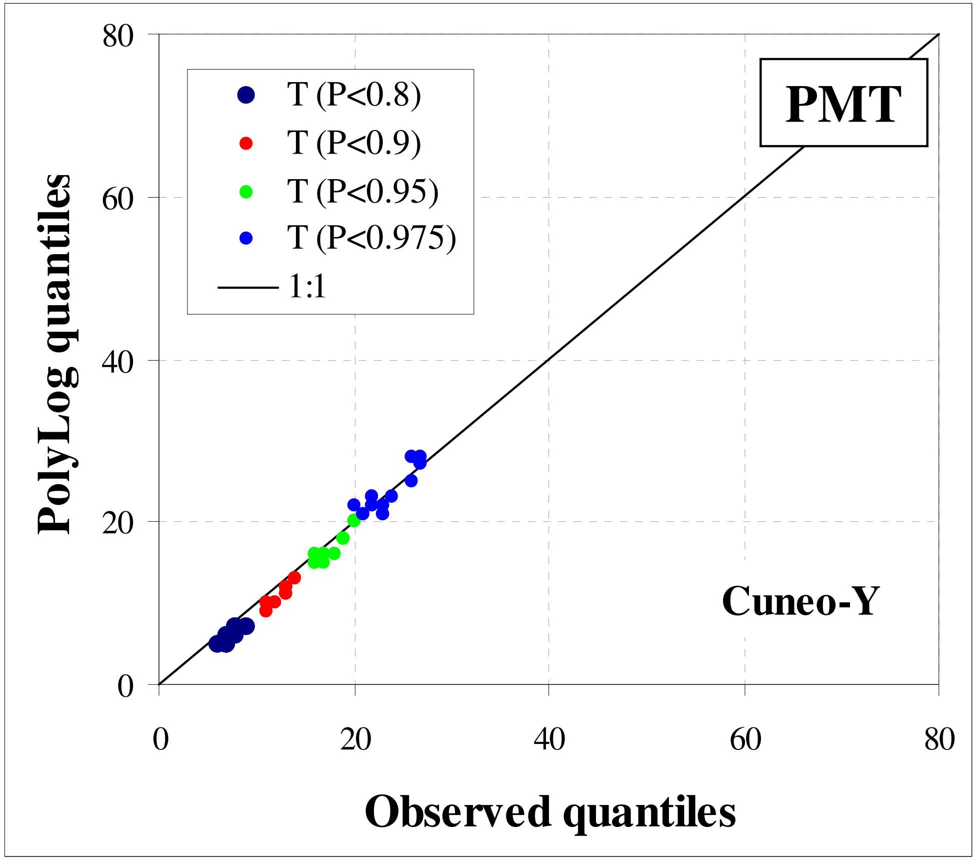
TEST FITTING: Quantiles (1/3)

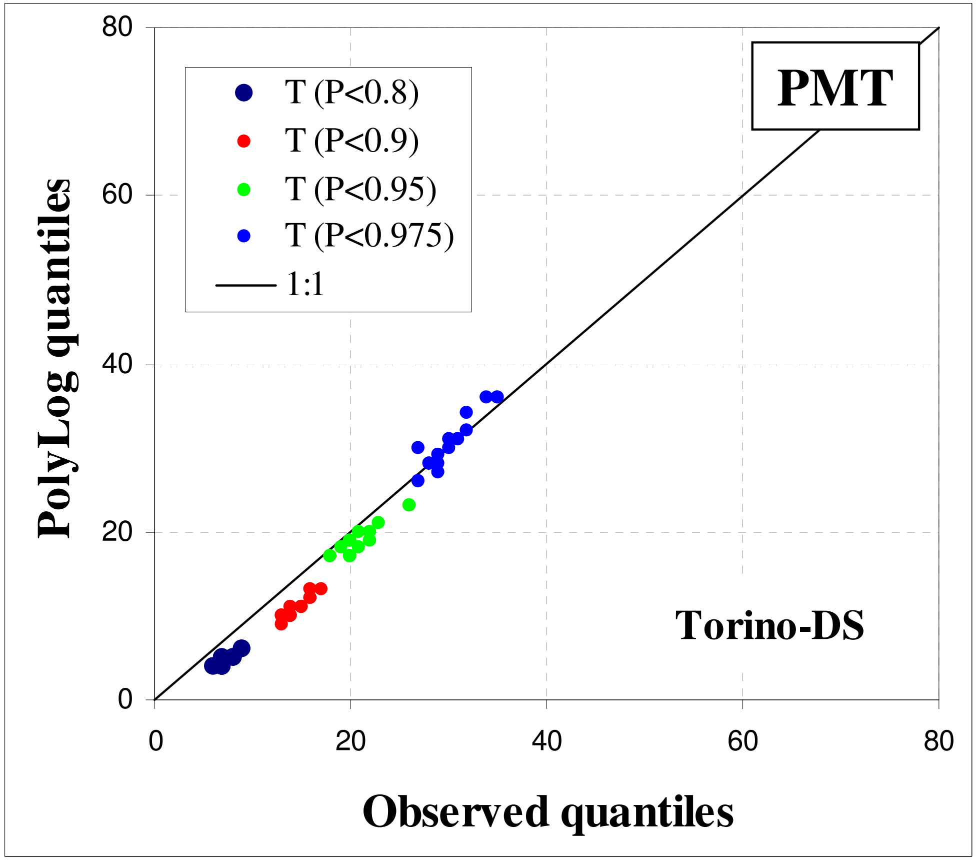




In SIC the observed quantiles (from 0.8 to 0.975) are well reproduced for all the stations, with few exceptions for the T(P<0.975) during GS (small dataset).

TEST FITTING: Quantiles (2/3)





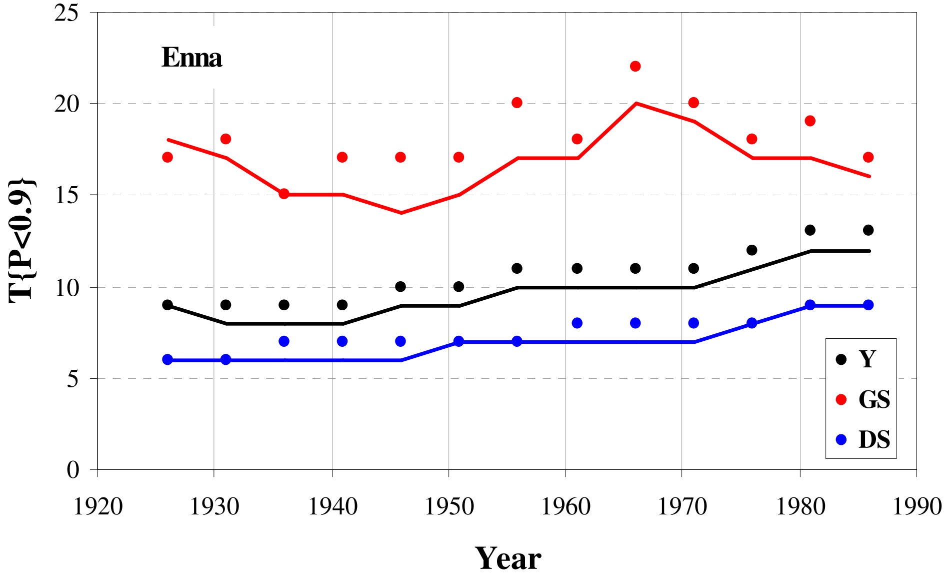
The results for the PMT stations show generally lower values compared to the SIC ones. Larger discrepancies were observed during the DS period especially for T(P<0.9).

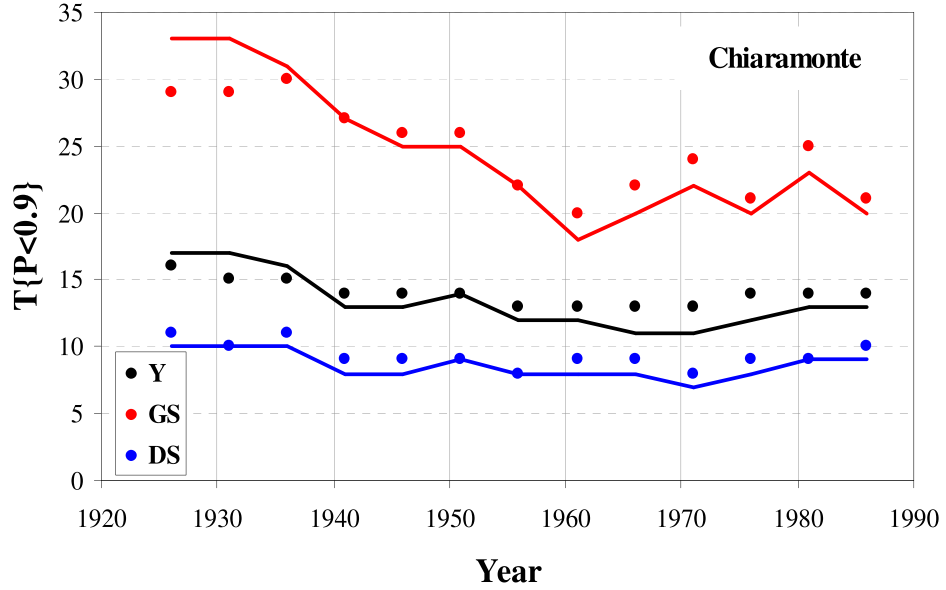
TEST FITTING: Quantiles (3/3)

RMSD (d)		T(P<0.80)	T(P<0.90)	T(P<0.95)	T(P<0.975)
PMT	Y	1.94	1.83	1.03	1.26
	GS	1.65	1.66	1.02	1.06
	DS	2.34	3.12	2.20	1.45
SIC	Y	1.50	1.19	1.41	2.40
	GS	1.83	1.72	4.41	8.91
	DS	1.47	1.66	1.47	1.13

RE (%)		T(P<0.80)	T(P<0.90)	T(P<0.95)	T(P<0.975)
PMT	Y	25.17	13.98	5.38	4.91
	GS	24.19	14.93	6.74	5.41
	DS	31.25	21.86	10.29	4.91
SIC	Y	24.78	11.22	7.62	8.52
	GS	17.64	9.26	12.59	15.51
	DS	33.16	22.02	13.43	7.54

PRELIMINARY-ANALYSIS OF TRENDS (1/2)





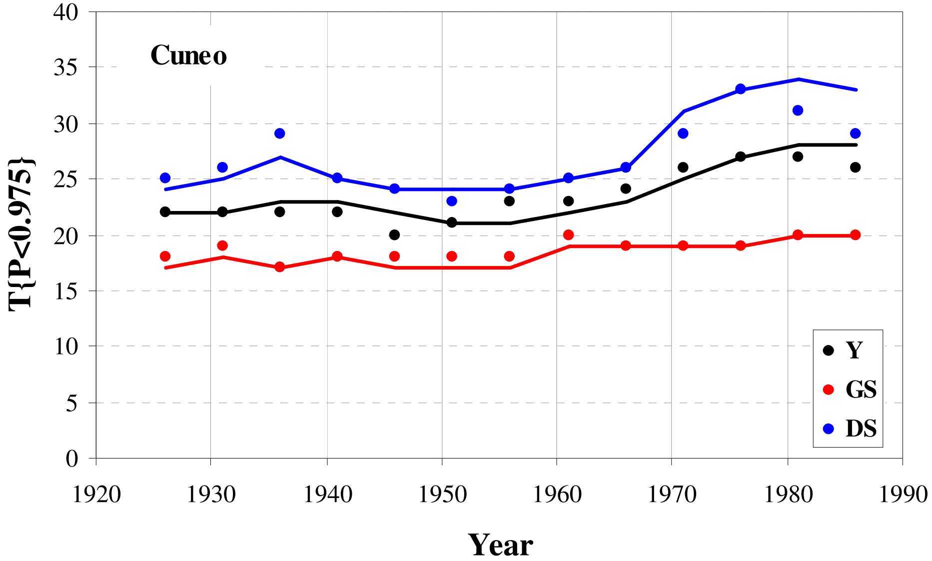
SIC

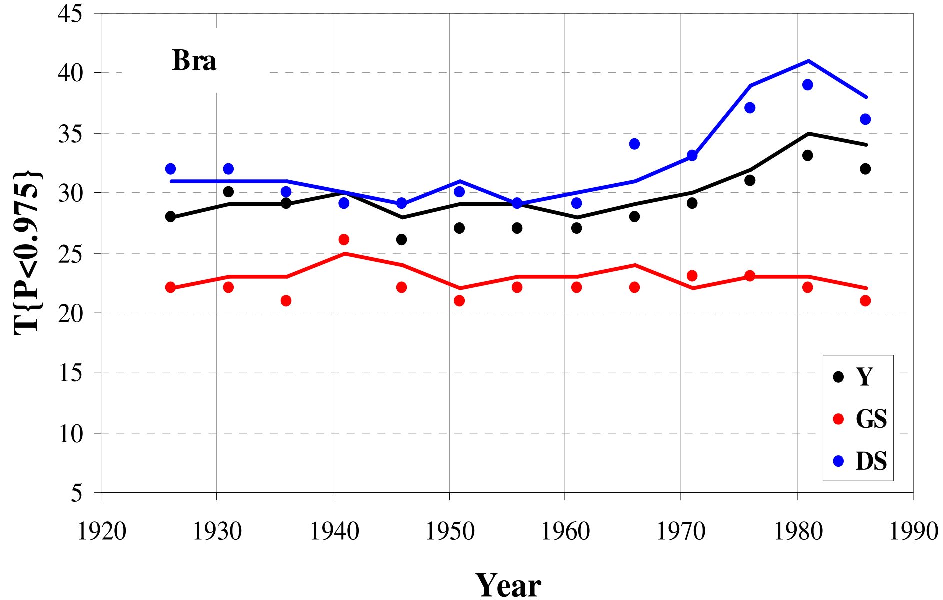
Station	Y	GS	DS
CHIARAMONTE	-	-	n
CORLEONE	+	-	+
ENNA	+	n	+
FLORESTA	n	n	n
MARSALA	+	n	+
MINEO	+	n	+

A qualitative analysis supports the existence of an increasing trend (excepted for Chiaramonte) during Y and DS (rainy season).

The results seem similar for all the analyzed quantiles.

PRELIMINARY-ANALYSIS OF TRENDS (2/2)





PMT

Station	Y	GS	DS
AOSTA	+	n	+
BRA	+	n	+
CUNEO	+	+	+
MONCALIERI	+	n	+
TORINO	+	n	+
VERCELLI	+	+	n

A slightly increasing trend was observed for all the PMT stations during both Y and DS periods.

No trend was observed for the GS season, similarly to SIC case.

SUMMARY AND CONCLUSIONS

- Several probability distributions were tested to model inter-arrival times in two Italian regions: a Mediterranean and a Sub-Alpine environment;
- The PolyLog distribution seems to well reproduce the observed data in the whole range of variability of times (~ 10^0 to 10^2 days) and frequencies (~ 10^0 to 10^4);
- The chi-square test based on the Montecarlo procedure suggests better performance for the seasons (GS and DS) compared to the whole year;
- The average cluster sizes are well reproduced by the model with only a slight underestimation (≈ 5%), catching the different dynamics in the two areas;
- The analysis of extreme quantiles (0.8 to 0.975) shows a good ability to reproduce them, with significant difference only during GS (in SIC) and DS (in PMT);
- A preliminary analysis of reproduced extreme quantiles seems to suggest the presence of an increasing trend in both areas, to be deepen.