Detecting non-stationarity in extreme rainfall data observed in Northern Italy

A. BRATH, A. CASTELLARIN & A. MONTANARI

Dipartimento di Ingegneria delle Strutture, dei Trasporti, delle Acque, del Rilevamento, del Territorio, University of Bologna, Via del Risorgimento 2, I-40136 Bologna, Italy - e-mail: armando.brath@mail.ing.unibo.it

ABSTRACT

Recent analyses have claimed the possible presence of non-stationarity, produced by the presence of either trend or long-term climatic fluctuations, in some historical hydrometeorological records observed in Europe as well as in other countries. Such non-stationarity might exert a remarkable effect on the estimation of the frequency distribution of the extreme events. However, it is well known that a reliable assessment of the presence of non-stationarity in hydrological records is not an easy task, because of the limited extension of the available data sets. This makes difficult distinguishing between non-stationarity, sample variability and long-term climatic fluctuations. This paper analyses two long rainfall records observed in the Emilia-Romagna region, in Northern Italy, in order to assess whether non-stationarity might be present. The results are compared with the outcomes of regional analyses aimed at assessing the presence of non-stationarity at regional scale. Moreover, some synthetic data are analysed, in order to assess how much the sample variability of a short stationary series might induce effects which could at first glance be ascribed to non-stationarity.

1 Introduction

In the recent past many analyses have been performed in order to detect the possible presence of non-stationarity in hydrometeorological records. These efforts have been mainly motivated by the results of some meteorological and hydrological studies which claimed the possible presence of irreversible climatic change, due to global climate forcing, such as increasing atmospheric CO_2. (Jones *et al*., 1986; Hansen and Lebedeff, 1987; Bloomfield, 1992). The awareness of the significant effects that such a global change might exert even on the optimal design of urban and land drainage networks and flood protection works has motivated a number of studies in order to detect evidences of climatic changes even at local scale. This is not an easy task, since the limited sample size of the climatic observations today available often does not allow a reliable identification of patterns in the data.

A significant number of rainfall series were recently analysed in Italy, where some long precipitation records are available such as, for instance, the daily rainfall series observed in Padova (Italy), which covers a very long observation period (since 1725) and is one of the longest daily rainfall record available in the world. This series was studied by Camuffo (1984), in order to detect the possible presence of tendencies or cyclical patterns and thus to gain an insight into the possible mechanism of climatic fluctuations. He found evidences of a secular wave in the yearly frequency of dry days, that apparently is not leading to irreversible changes. Even the precipitation amount was found to show a wavy trend of different period, not always in phase with the frequency trend. This resulted in a cyclical variation also of the precipitation intensity. The periodic pattern of the oscillations found by Camuffo (1984) highlights the particular care that should be taken when analysing short climatic records, since increasing or decreasing trends detected in a short series might be the effect of a longer cycle, thus not leading to irreversible changes.

Burlando (1989) analysed the series of the daily rainfall data observed in Florence (Italy), which collects data observed since 1813 and is another of the longest daily rainfall records available today in Italy. He found significant changes along time of the extreme storms structure and, in particular, a decrease of the number of storm events and a corresponding increase of their intensity in the latter decades. Analogous results were found by Montanari (1998) who analysed four long rainfall series observed in the cities of Sondrio, Milan, Florence and Genoa (Italy). These results might explain the apparent increase of the magnitude of the extreme storm events in the recent past, not necessarily leading to irreversible changes.

Montanari *et al.* (1996) analysed the monthly and annual rainfall records observed in Sondrio, Milan, Parma, Genoa, Florence and Rome and found decreasing trends in the precipitation amounts observed in the last century, in phase with the increase of the precipitation intensities previously detected. In order to verify whether or not the detected trends might be due to long-term climatic fluctuations, rather than non-stationarity, they performed a long-memory analysis on the available data. Using different recently developed techniques for long-memory detection and estimation, including maximum likelihood, they found significant evidences of long-memory effects in the series observed in Parma and Rome, which allowed to conclude that the detected trends in the precipitation amounts are never statistically significant. The results of this analysis highlighted that the estimation of trends and tendencies, when dealing with hydrometeorological variables, should always take into account the effects of the possible presence of long-term persistence.

Recently, De Michele *et al* (1998) analysed the same time series considered by Montanari (1998), but focusing on the extreme values. They highlighted the effect that the climatic fluctuations can induce on the estimation of the critical design storm. Their analysis was performed by essentially estimating the daily rainfall depth expected for a return period of 30 years (30-years critical design storm, *CDS*), in each year of the observation period of each time series, using four different extreme value distributions and fitting only the annual maximum daily rainfall depths observed in the latter 30 years before the year of estimation. The progress of their *CDS* along the observation period of the time series resulted impressive. For instance, the *CDS* referred to the Genoa series increased of more than 40% from 1880 to 1980. Rosso (1999, personal communication) coupled the output of this analysis with a rainfall-runoff model and proved that the above mentioned trend in the *CDS* is likely responsible of the increase of the peak flows of the Bisagno River which took place in the last decades and has caused a corresponding increase of the flood events which have interested the city of Genoa.

The outcomes of this latter analysis highlighted that the detection of climate change at local scale, and therefore of non-stationarity in hydrological records, has relevant implications in the design of the river engineering and drainage facilities, and consequently is not only a matter of ecological concern. Although it is still not clear whether or not the detected tendencies are clues of global climate change, they are worth analysing from an operational point of view.

The present study first of all performs an analysis similar to the one developed by De Michele *et al.* (1998), considering more recent data collected in a geographically different region, in order to verify if the previously obtained results apply also when referring to the very recent past and different contexts. Secondly, the study is aimed at better inspecting the possible causes of the detected increase of the *CDS* in the latter decades. Accordingly, the estimated *CDS* are compared with the analogous obtained by analysing synthetic stationary series whose sample probability distribution is identical to the one of the observed data. The aim of this comparison is to assess how much the sample variability of a stationary series can give raise to effects which at first glance could be attributed to non-stationarity. Moreover, a regional study is performed, in order to verify whether or not the detected non-stationarity is still present at regional scale, thus excluding the possible presence of local behaviour and measurements errors. The estimation of long-memory in the rainfall data finally helps in assessing the possible presence of long-term fluctuations in the analysed series.

The next section of the paper describes the rainfall series analysed here and the regional data available. The third section describes the procedure for the estimation of the *CDS* and its progress along the observation period. The fourth section is devoted to the description of the regional study while the fifth section describes the results of the analysis of the synthetic rainfall data. The sixth section reports the results of the estimation of long memory and the last section of the paper summarises the conclusions of the study.

2 THE RAINFALL DATA

Two long daily rainfall series observed in the Emilia-Romagna region, in Northern Italy, are analysed here. They were collected in the cities of Parma and Ferrara, which are located within the Po River valley. The location of the two raingauges is showed in Figure 2, along with the delimitation of the homogeneous pluviometric regions which will be described later. The altitude of Parma and Ferrara is respectively 52 *m* and 10 *m* above sea level. The observation period of the Parma series is extended from January 1st 1878, to December 31st 1997, so that the sample size counts 43830 daily observations. The Ferrara series was observed from January 1st, 1906 to December 31st, 1998, but three years are missing so that the sample size amounts to 32874 data.

Following the framework of the analysis proposed by De Michele *et al.*(1998), only the annual maxima daily rainfall were considered here, whose main statistical properties are reported in Table 1.

RAINGAUGE	SAMPLE SIZE	MEAN	STD. DEV.	ASYMMETRY
Parma	120	54.9 *mm*	19.71 *mm*	1.43
Ferrara	90	48.2 *mm*	19.55 *mm*	2.11

Table 1. Main statistical properties of the analysed series of annual maximum daily rainfall depths.

The plots presented in Figure 1 report the two annual maxima time series versus time.

An extensive analysis has been recently performed for the geographical area showed in Figure 2 (Emilia-Romagna ad Marche administrative regions) aimed at identifying homogenous regions with respect to the frequency of the annual maximum daily rainfall depths (see Brath *et al*., 1998).

The reliability of the obtained subdivision was tested by verifying the capability of the Generalised Extreme Values probability distribution (*GEV*) to fit the dimensionless annual maximum rainfall depths, for fixed duration *d*, observed in all the raingauges located within each homogeneous region. This analysis was performed for different storm durations, namely 1, 3, 6, 12 and 24 hours. The dimensionless rainfall depth is computed by dividing each observation by the mean value of the whole data collected at the same raingauge for the corresponding storm duration.

The identified homogenous regions are shown in Figure 2.

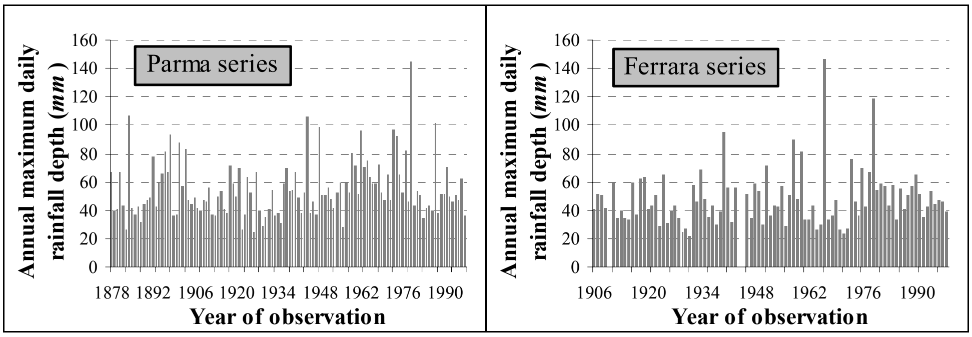

Figure 1 – Annual maximum daily rainfall depths observed in Parma (1878 – 1997) and Ferrara (1906 – 1998).

3 EVIDENCES OF NON-STATIONARITY IN EXTREME DAILY RAINFALL DEPTHS

The effects of the possible presence of non-stationarity in the rainfall data on the magnitude of the expected extreme values can be assessed by following the same approach proposed by De Michele *et al.* (1998), who focused their attention on the time series of the annual maximum daily rainfall depths. In detail, the procedure is based on the estimation of the daily rainfall depth corresponding to a return period of 20 years (*CDS*). The estimation is repeated for each year of the observation period, starting from the 30^{th}, and is performed by fitting only the historical annual maximum daily rainfall collected in the previous 30 years. These data are fitted using a *GEV* distribution which allows the computation of the expected *CDS* corresponding to the assigned year and return period. If the data were stationary, the resulting *CDS* should result constant along the observation period, apart from fluctuations due to sample variability. The progress of the obtained *CDS* along time allows one to assess the effects of non-stationarity on the estimation of the *CDS* itself. However, it should be noted that discriminating between non-stationarity and the above mentioned sample variability is not a simple task. Sample variability itself might in fact give raise to significant changes in the *CDS*, especially when dealing with phenomena characterised by long-tailed probability distributions, as it is often the case for rainfall.

The Generalised Extreme Value distribution (GEV) is written as,

$$F(x) = \exp\left\{-\left[1-\frac{k(x-\varepsilon)}{\alpha}\right]^{1/k}\right\} \quad \text{for } k \neq 0 \tag{1}$$

and

$$F(x) = \exp\left\{-\exp\left[-\frac{(x-\varepsilon)}{\alpha}\right]\right\} \quad \text{for } k = 0, \qquad (2)$$

where α, ε and k are the parameters of the distribution. As shown by (2), when $k = 0$ the GEV distribution reduces to the Gumbel distribution.

Parameter estimation was performed here by using the *L*-moments method (Hosking, 1990).

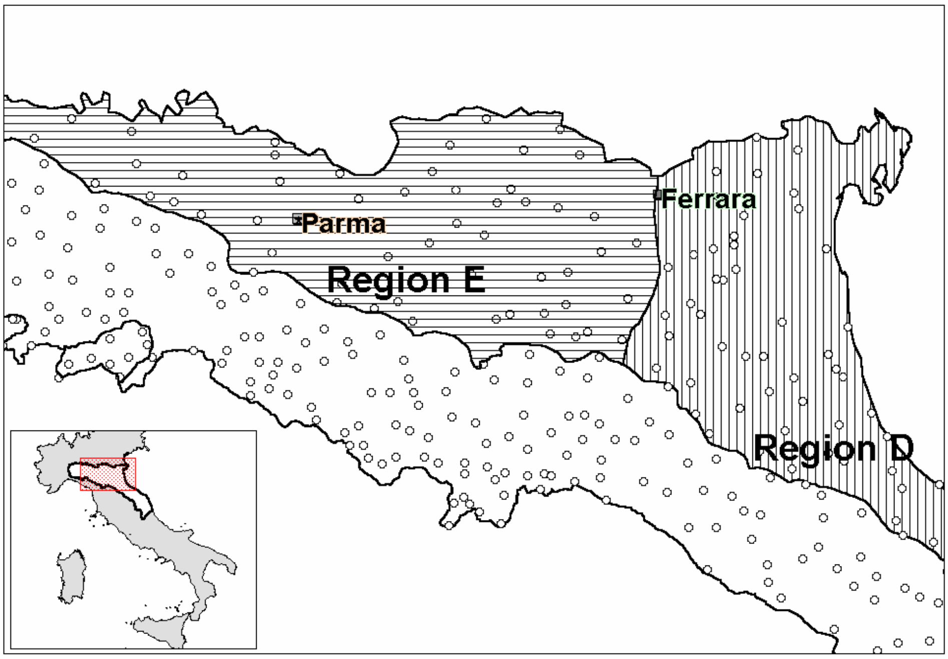

Figure 2 – Location of Parma and Ferrara; limits of the homogeneous regions D and E and location of the raingauges which have been considered in the regional analysis (circles).

Figure 3 shows the obtained progress of the *CDS* for both series. It can be seen that sizeable increasing trends, beginning from the 30's, are present. This tendency is evident in both series and is consistent with the results obtained by De Michele *et al.*(1998). However, the data sets considered here, which are more extended in the recent past, show that such increasing tendency seems not to persist after 1990.

The availability of many annual maximum daily rainfall series for the Emila-Romagna and Marche regions, observed in the period from 1937 to 1998, suggests to verify whether or not the increase of the *CDS* is present also at regional scale.

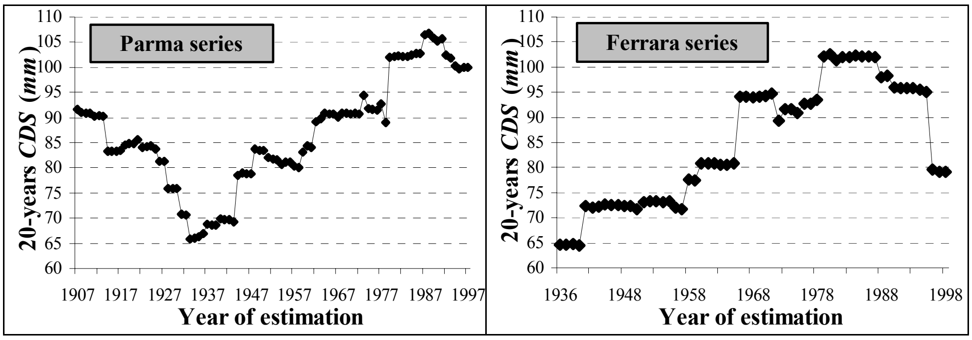

Figure 3 – Progress of the *CDS* along the observation period for the series of Parma (1878 – 1997) and Ferrara (1906 – 1998).

4 REGIONAL ANALYSES

Short to medium-range records of annual maximum daily rainfall depths are available for the Emilia-Romagna and Marche administrative regions, referred to quite uniformly distributed raingauges. These data were processed here by performing a regional analysis which is structured as follows. For each one of the two previously considered raingauges, the annual maxima daily rainfall observed in the corresponding pluviometric homogeneous region, within a distance of 80 *km*, were collected. Only the raingauges having at least 15 years of rainfall observations were considered for the analysis. The dimensionless annual maxima daily rainfall depths were computed and divided in two groups, in order to pool together all the data observed in two different and non-overlapping periods, extended respectively from 1937 to 1968 and from 1969 to 1998. The two regional growth curves referred to these periods were computed and then compared, in order to assess the presence of differences due to the climatic fluctuation. As can be observed from Figure 2, the Parma raingauge is located within the homogenous region E, while the Ferrara raingauge is instead located near the border line between the regions *D* and *E*. Therefore it seemed reasonable, in this latter case, to pool the pluviometric information from the neighbouring raingauges belonging to both the homogenous regions. Figure 1 shows the position of all the available raingauges and the extension of the pluviometric homogeneous regions.

Table 2 resumes the number of raingauges with sufficiently long records along with the total number of dimensionless annual maxima referred to each one of the two series and for the two time periods.

The growth curves referred to the two raingauges and two periods are estimated by fitting a *GEV* distribution to the obtained dimensionless data, using the *L*-moments method.

SITE OF INTEREST	PERIOD	NO. OF RAINGAUGES WITH MORE THAN 15 OBSERVATIONS	NO. OF OBSERVED ANNUAL MAXIMA
Ferrara	1939 – 68	58 (Region D and E)	1407
Ferrara	1969 – 98	44 (Region D and E)	2378
Parma	1938 – 67	14 (Region E)	345
Parma	1968 – 97	26 (Region E)	472

Table 2. Number of raingauges with more than 15 observation and total number of annual maxima available for the regional analyses

Figure 4 and 5 report the results for both the regional analyses. For Parma and Ferrara the probability plot shows the regional *GEV* estimated by fitting all the available dimensionless data pooled from each homogeneous region, the regional *GEV* referred to the two non-overlapping periods, the at-site *GEV* estimated using the whole at-site data-set and the two at-site *GEV* referred to the two periods. The plots allow one to assess the differences between the at-site and regional *GEV* in the examined periods and, as such, whether or not the climatic fluctuations can be detected also at regional scale.

It can be seen that the at-site growth curves that refers to the two different periods are clearly different each other in both the raingauges. This was to be expected and confirms the outcomes of the previous at-site analysis. In Parma the 1937-1968 at site growth curve underestimates the adimensional rainfall depth with respect to the 1969-1998 one. The opposite result is found in Ferrara, where the decreasing tendency in the present decade, mentioned in the previous paragraph, is more pronounced (see Figure 3).

The comparison of the regional growth curves shows that the 1937-1968 ones lead to underestimate the rainfall depths in both regions. Therefore this result is similar to the outcome of the at-site analysis of Parma. The differences in the regional growth curves referred to the two periods indicate that there are indeed some evidences of non-stationarity, although with minor intensity with respect to the at-site cases. Thus the increasing tendency which was detected in the period 1940 – 1990 is not confined to the two long rainfall records considered here but it seems to be ascribed to a climatic fluctuation which is noticeable even at regional scale. This result is not surprising, since contemporaneous rainfall data are expected to be spatially correlated. Nevertheless it is significant, since it allows us to neglect the possibility to ascribe the detected tendencies to measurement errors or local behaviours.

5 ASSESSING THE EFFECTS OF SAMPLE VARIABILITY

It was mentioned before that it is not a simple task to assess whether the detected tendency in the *CDS* is indeed due to non-stationarity or sample

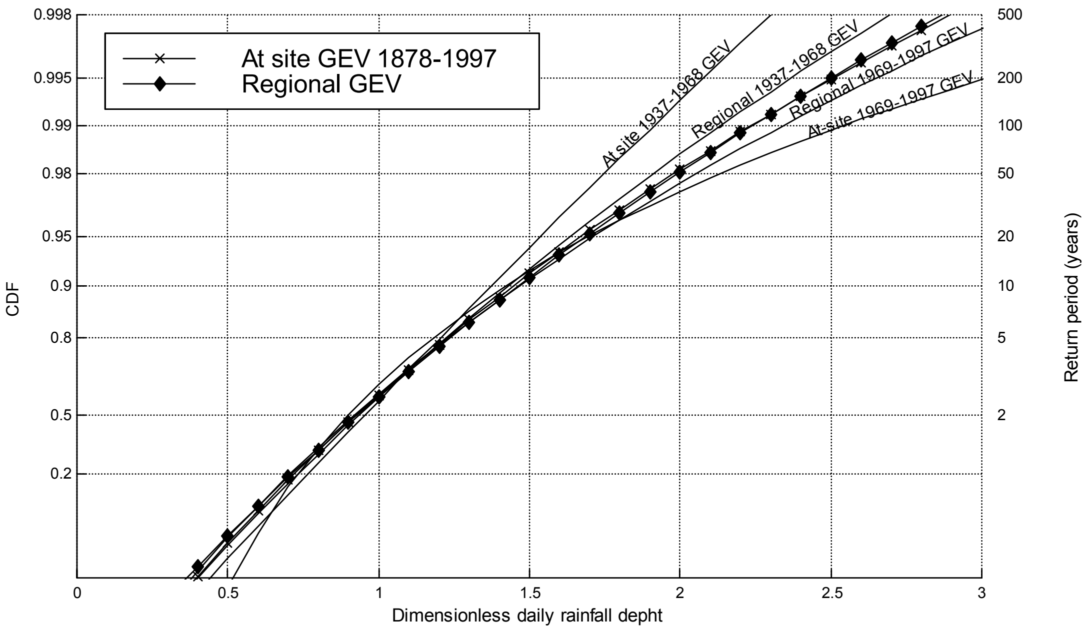

Figure 4 – Parma raingauge: comparison among at-site and regional *GEV* distributions for the whole period and the two subperiods (1937 – 1968 and 1969 – 1997)

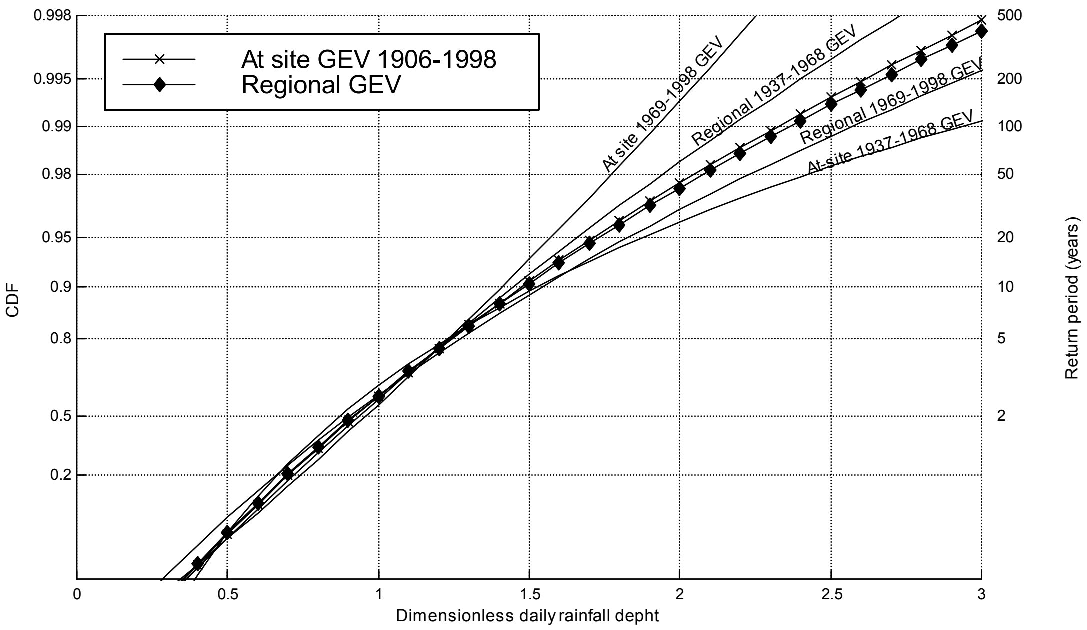

Figure 5 – Ferrara raingauge: comparison among at-site and regional *GEV* distributions for the whole period and the two subperiods (1937 – 1968 and 1969 – 1998).

variability. The continuous estimation of the *CDS* along the observation period, performed here, is in fact obtained through a moving window extended over the past 30 annual maxima; it follows that the series of the *CDS* are highly correlated and as such, for reasons of sample variability, they can show the presence of long-term cycles which could be falsely ascribed to non-stationarity, even if the original annual maxima series is stationary.

In order to better clarify this aspect two synthetic 10000-years stationary series of annual maximum daily rainfall depths were generated, from a probability distribution equal to that observed in Parma and Ferrara. The estimation of the 20-years *CDS* was performed on these series by applying the same procedure used for the observed data. As it was expected, the results show that increasing and decreasing tendencies are present also in the *CDS* series derived by analysing the synthetic data.

In order to compare them with the ones found by analysing the observed series, the relative increment of the design storm detectable in a 50-years time span was chosen as testing measure. Such length of 50 years was selected in order to match the length of the increasing tendencies detected in Parma and Ferrara, extended from 1940 to 1990.

The 50-years relative increment of the *CDS*, ΔH, is defined as

$$\Delta H = \frac{H_b - H_a}{\overline{H}} , \tag{3}$$

where H_b and H_a are the values of the design storm respectively at the beginning and at the end of the 50-years time span, obtained from the linear regression of the observed estimates, i.e. assuming the existence of a linear trend, and $\overline{H}$ is the mean observed value of the 20-years design storm in the same period. The estimation of ΔH was performed for all the consecutive 50-years intervals of the synthetic series, so that the sample size of the resulting ΔH series amounts to 9922.

The 50-years relative increments found when analysing the synthetic series exceeded in some cases the maximum ones estimated on the two observed data series. The percentage of such exceedance was computed by dividing the number of greater relative increment cases by the total number of relative increment estimations performed on the synthetic series.

Table 2 reports the values of ΔH computed on the observed series along with the corresponding percentages of exceedance, estimated on the synthetic rainfall series.

RAINGAUGE	OBSERVED ΔH (1940-1990)	PERCENTAGE OF EXCEEDANCE
Parma	0.39	3.2%
Ferrara	0.49	2.0%

Table 3. Relative increments of the *CDS* detected in Parma and Ferrara in the period 1940-1990 along with their percentage of exceedance estimated by analysing the simulated 10000 size *CDS* series.

It can be seen that the maximum linear trends on a 50 years time span found in the *CDS* for Parma and Ferrara (period 1940 – 1990) were rarely exceeded when analysing the synthetic stationary series. The corresponding percentages of exceedance amount in fact to 2.0% for Ferrara and 3.2% for Parma.

The results indicates that the tendencies that were found in the observed data are not likely to be ascribed to sample variability, but should rather be attributed to either long-term climatic fluctuations or to non-stationarity.

6 LONG-MEMORY ANALYSIS OF THE ANNUAL MAXIMA

It is well-known that, when analysing climatic series, to distinguish between long-term fluctuations and non-stationarity (trend) is not a simple task, mainly because of the limited extension of the available data sets. Nevertheless, such a distinction would be extremely interesting. In fact, the presence of long-term climatic fluctuations, rather than non-stationarity, would imply that the patterns found in the *CDS* could likely be attributed to cyclical behaviour rather than to irreversible tendencies.

A useful tool for the above mentioned distinction is the detection of the possible presence of long memory in the data. In fact, it was mentioned before that one of the effects of long-memory is the attitude of the time series to be subjected to long-term cycles.

A stationary process possesses long memory if there exists a real number H ($0.5 < H < 1$), called the "Hurst exponent," and a constant c_H such that

$$\rho(k) \approx c_H k^{2H-2} \tag{4}$$

as k →:, where ρ (k) is the autocorrelation coefficient of the process at lag k.

Thus long memory implies that the autocorrelation function of the process decreases slowly, like a power function. The exponent H, first introduced by Hurst (1951) in a hydrological context, is also called the "intensity of long memory".

The detection of long memory in a time series can be done heuristically by estimating the value of H (Beran, 1994). An H value equal to 0.5 means absence of long memory. The higher the H, the higher the intensity of long-memory effects.

The H values were estimated here by applying two well-tested procedures, namely the *R*/*S* method and the Aggregated Variance method. These procedures will not be described here. A detailed description of them and other methods for long-memory detection and estimation can be found in Beran (1994).

The results of the estimation of the Hurst exponent performed on the two series of the annual maximum daily rainfall observed in Parma and Ferrara are reported in Table 4. It can be seen that both methods for long-memory estimation indicate the presence of long-memory in Parma and its absence in Ferrara. However, the Ferrara record resulted the one in which the patterns in the *CDS* are more pronounced. This seems to preclude the assessment of a direct link between the presence of long-memory and the detected patterns in the *CDS* estimated on the observed data. Thus, the presence of non-stationarity, and hence of irreversible tendencies, cannot be excluded.

RAINGAUGE	*R*/*S* METHOD	AGGREGATED VARIANCE METHOD
Parma	0.72	0.70
Ferrara	0.48	0.43

Table 4. Results of the estimation of H performed on the two series of the annual maximum daily rainfall observed in Parma and Ferrara, using the *R*/*S* and the Aggregated Variance method

7 CONCLUSIONS

Two long series of annual maximum daily rainfall depth, observed in Northern Italy, were analysed here in order to detect whether or not non-stationarity might be present in the data. Such detection was performed by estimating the daily rainfall depth expected for a return period of 20 years (*CDS*) in each year of the observation period of each time series, fitting the annual maximum daily rainfall depths observed in the 30 years which precede the year of estimation.

The analysis has highlighted the presence of significant changes of the estimated *CDS*, which appear in particular to be affected by an increasing tendency in the period 1940 - 1990. This result is coherent with the ones that were obtained by De Michele *et al.* (1998) by analysing other Italian rainfall series. Therefore, a refined analysis has been performed here in order to better inspect the causes of the tendency mentioned above, which might be originated by non-stationarity, climatic fluctuations or sample variability. In particular, a regional analysis has shown that the increasing tendency is clearly detectable also at regional scale, and thus is not likely due to local behaviours or measurement errors. Moreover, the computation of the *CDS* on two synthetic series has shown that sample variability as well is not likely responsible for the variations in the *CDS* detected when analysing the observed data.

The small sample size of the analysed records does not allow to assess the possible presence of non-stationarity. However, the inversion of the increasing tendency of the *CDS* in the very recent past (1991 - 1998) might instead suggest the presence of cyclical fluctuations.

ACKNOWLEDGEMENTS

The work presented here has been partially supported by the European Community through the grant ENV4-CT97-0529 (FRAMEWORK project), by the National Research Council of Italy - Group for the Prevention of Hydrogeological Disasters, contract n° 98.00587.PF42 and by the Ministry of University and Research in Science and Technology (MURST) of Italy through its 40 % national grant to the program on "Climatic and anthropogenic effects on hydrological processes", contract n° 9908032871_001.

REFERENCES

Beran, J. Statistics for long-memory processes, Chapman & Hall, New York, United States, 1994.

Bloomfield, P. Trends in global temperature, *Climatic Change*, 1992, **21**, 1-16.

Brath, A., Franchini, M. & Galeati, G. Assessment of extreme rainfall frequency in Northern Italy (in Italian), *L'Acqua*, 1998, **4**, 15-27.

Burlando, P. Stochastic models for the prediction and simulation of precipitation in time (in Italian), Ph.D. Dissertation, Politecnico di Milano, Milano, 1989.

Camuffo, D. Analysis of the series of precipitation at Padova, Italy, *Climatic Change*, 1984, **6**, 55-77.

Dalrymple, T. Flood frequency analyses, Water Supply Paper 1543-A, U.S. Geological Survey, Reston, USA, 1960.

De Michele, C., Montanari, A. & Rosso, R. The effects of non-stationarity on the evaluation of the critical design storm, *Water Sci. and Tech.*, 1998, **37**, 187-193

Hansen, J. & Lebedeff, S. Global trends in measured surface air temperature, *J. of Geophys. Res.*, 1987, **D11**, 13345-13372.

Hosking, J. R. M. L-moments: analysis and estimation of distribution using linear combination of order statistics, *J. R. Statist. Soc.* B, 1990, **52**(1), 105-124.

Jones, P.D, Wigley, T.M.L. & Wright, P.B. Global temperature variations between 1861 and 1984, *Nature*, 1986, **322**, 430-434.

Montanari, A., Storm structure variability in historical rainfall data observed in Italy, *Ann. Geophysicae*, 1998, **16(2)**, 456.

Montanari, A., Rosso, R. & Taqqu, M.S. Some long-run properties of daily rainfall records in Italy, *J. of Geophys. Res. - Atmosphere*, 1996, **101**, 29431-29438