The modeling of extremum daily temperature series by spectral methods

V. Malvestuto[1], O. Testa[1], M.C. Beltrano[2]

Corresponding Author: v.malvestuto@isac.cnr.it

Abstract - We examine here daily minimum and maximum temperatures recorded at 7 climatic stations all located in Lazio region, Italy. These 14 time series were provided by the Italian "Agro-meteorological National Data Base" (BDAN) of the National Agricultural Information System (SIAN) and cover the second half of the XX century. The purposes of the signal processing were, first, to extract the linear trend and the two main seasonal cycles present in the series, second, after their subtraction from the signal, to assess the relative importance of the residual stochastic component and, finally, identifying a stochastic model for the latter, in order to arrive at an artificial simulation of the original series. After retrieving and filling the data gaps, we obtained uninterrupted series of daily data. Then, after detrending and filterinfg away the seasonal components (with 6-month and 12-month periods), it was possible to determine correlograms and power spectra of the residual zero-mean stochastic component. Also, a successful attempt was carried out to model this stochastic residual by means of an AR(1) process, thus yielding an efficient representation of the time variability of each of the 14 temperature series. In all cases the residual white noise obtained is definitely non gaussian. This model including the trend, the seasonal oscillation and the AR(1) process permitted a fairly good artificial reconstruction of the given temperature series via computer simulations specific for each given climatic station. This reconstruction, on capturing the essential features of each given series, represents a useful tool to describe and understand the recurrence of weather patterns and the possible occurrences of weather-linked phenomena interesting the local vegetation and the related biological processes. As a by-product, the analysis has permitted to evaluate the relative incidence of the two main seasonal components, and their importance with respect to the residual variability associated to purely stochastic fluctuations. From a comparison with the results of other similar studies, carried out in other countries of Europe and Oceania, it appears that the trends found by us for both minimum and maximum temperature daily series, when statistically significant, are generally lower than the corresponding values reported by the last IPCC (2007) for those areas that, at least from a geographical viewpoint, appear similar to ours.

Keywords - spectral analysis, stochastic methods, daily thermometric series.

1. INTRODUCTION

The spectral analysis of time series, adopted in the study of the most various physical phenomena, represents a valid method of investigation also in the field of meteorological and climatic phenomena [1, 2, 3]. The partitioning of the signal via Fourier transform in separate periodic components allows to obtain a quantitative assessment of the relative incidence of each simple harmonic oscillation into which the full signal can be resolved. In this way one can easily recognize the presence within the series of characteristic simple periodic phenomena like, e.g. in the meteorological series, of annual and seasonal cycles, and to weigh their importance with respect to the level of the always present erratic component, representing the contribution to the signal of the random fluctuations. This erratic component, on assuming that it has a stationary character, can be in turn separately modelled by the standard tools of stochastic analysis [4]. However, a full spectral analysis can be carried out in a reliable way only on complete, regularly sampled data series, with no gaps.

This methodology has been employed to study some thermometric series of Lazio, a region of Central Italy. For each series the deterministic components have been first detected and quantitatively assessed: these include a linear trend and two seasonal cycles, whose frequencies and energies have been determined; then, after subtracting these components from the original signal, a residual signal of stochastic nature was obtained, which was subsequently analyzed by means of a simple AR(1) model, thus separating the full erratic component in an auto-regressive part and a residual white noise, the intensity of which has been finally estimated.

The results of the overall analysis permitted us to artificially simulate the temperature signal at each station, thus opening the way toward the possibility of reconstructing the series both in the past and, cautiously, in the near future.

1. *CNR-ISAC, Italian National Research Council - Institute of Atmospheric Sciences and Climate; Via Fosso del Cavaliere, I-00133 Rome ; ph. +390645488266; fax +390620660291*
2. *CRA-CMA Agricultural Climatology and Meteorology Research Unit; Via del Caravita 7a, I-00186 Rome*

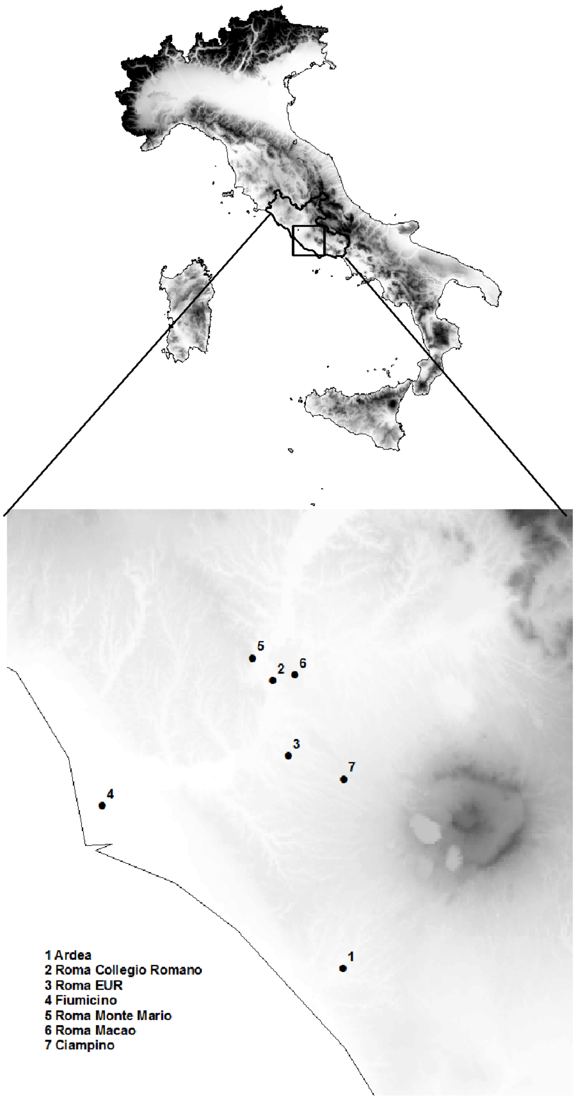

Fig. 1: Map of the seven weather stations analyzed in the Lazio area

In recent years signals of a climate change in progress have been documented in many locations throughout the world. In particular, positive rainfall trends have been reported in Argentina [5], Australia and New Zealand [6, 7]. Negative rainfall trends have been reported in the Russian Federation [8], Turkey [9], Africa [10, 11] and in China [12]. On the other hand, in nineteen weather stations of Northern and Central Europe Heino *et al.* [13] found no changes in precipitation extremes. In accordance with the latter study, Testa *et al.* [14, 15], after examining the records in 31 Central Italy weather stations, found no changes as well in the amounts of annual and seasonal rainfall, and in precipitation extremes. For what concerns temperature records, positive trends in annual minimum and maximum temperatures have been reported in many countries; in particular positive trends in maximum and mean temperatures have been evidenced in Northern and Central Europe, as well as over the Russian Federation, and over Canada [16], in Australia and New Zealand [7]. Prior to 1950, the available data were insufficient to develop global-scale maps of maximum and minimum temperatures. Then, most of the relevant data analysis begins from 1950. The report of the last IPCC [17] indicates that the annual trend in minimum and maximum air temperatures over regions with a discrete amount of data were +0.2 °C per decade and +0.14 °C per decade, respectively, with a trend in diurnal temperature range (DTR) of -0.07 °C per decade [18]. Also, Smith *et al.* [19] have reported that mid-latitude regions, such as the Mid-Western USA, Southern Europe and Asia, are becoming warmer and drier, while the lower latitudes are becoming warmer, too, but wetter.

For what concerns the Mediterranean and, particularly, the Italian region, there are several interesting studies of long temperature series covering the second half of the last century and some longer ones starting even in the XIX century [20 , 21, 22, 23]. These studies show that the series of annual and seasonal T_{min} and T_{max} exhibit significant positive trends during the period 1865-1996, trends greater in Central-Southern Italy than in Northern Italy, that in both subregions mostly depend on the behavior of data in the last few decades of XX century. Also, these studies found a positive significant trend for the diurnal temperature range (DTR), higher for Northern Italy than for Central and Southern Italy.

This vast and increasing amount of results motivated our investigation, by which we intended, first, to ascertain if the existence of similar trends could be demonstrated in some areas of Central Italy in view of an effective detrending of our time series before investigating their stochastic nature, which is the main purpose of the present work, and, second, to ascertain if the trends in minimum and maximum temperatures are comparable with the results of the many other studies referring to other countries in Northern and Central Europe, in the Russian Federation [8], in Canada [16, 24], in Australia and New Zealand [7]. Indeed, temperature trends observed in Europe and other continents can hardly be directly compared with analogous trends documented in the Italian region, where local trends will be crucially affected by regional features, such as peculiar changes in the large scale circulation patterns and their modifications due to the particular orography and to the penisular character of the most part of Italy. For example, the global maps of temperature anomaly reported by IPCC present a large spatial variability, so trends found in a limited area might sensibly differ from the average global behaviour. But, it is interesting here to note that the reported mean temperature trend of 1°C per century is not only confirmed by local studies regarding the Italian Alpine area, but is even found to be double than that [22, 23, 25].

It can be anticipated that the results of the present work generally confirm the presence of positive trends in minimum temperatures in all the examined sites, but the estimated trend values are mostly below 1°C per century, that is, the trend values found by us are less than the half of those reported in the last IPCC (2007) for similar areas.

2. DAILY DATA: TREATMENT AND METHODS

2.1. Introducing the database

The data examined in this paper are daily minimum and maximum temperatures recorded at 7 climatic stations located in Lazio region. These 14 time series of daily data, expressed in °C, cover the second half of the XX century, from the beginning of 1951 to the end of 1999, but almost all series presented some gaps within them and/or at their ends. The stations studied in the present work were selected from a larger set as those for which it was possible to fill all the existing data gaps, thus obtaining uninterrupted series of daily data. This permitted us to carry out a complete stochastic analysis of these series, without resorting to questionable zero-padding techinques when evaluating the spectra. The seven climatic stations studied are located in different districts of the Lazio region, some on the coastline, others in the urban centre of Rome and others in its suburbia. For a complete list of these stations see Figure 1 and Table 1. For a description of the data, of their quality and of the gap-filling process they were subjected to, see Appendix A.

2.2. The necessity of filling the data gaps

The temperature is a climatic parameter that, unlike precipitation, shows a variability continuous in time and space, so that it is possible to use methods based either on multiple cross-correlations among neighbouring stations or on simple auto-regression in order to fill in the missing data. Generally, time series having gaps at the beginning and at the end have been previously trimmed. Afterwards, the remaining inner data gaps have been filled using an auto-regression method illustrated below. The filling of the gaps is required when an estimation of the power spectrum is desired, since such estimation is based on the assumption that a regularly sampled record having no gaps be available

2.3. The method used to fill the data gaps

Let us define as "gap" an uninterrupted run of one or more missing data in a time series. As mentioned above, the single missing data is indicated by a fictitious numeric code (e.g., -999), so that a gap consists of a more or less long sequence of these special no–data codes. Sometimes a time series is preceded or terminated by a gap, and sometimes both circumstances are true. Such gaps are denoted "terminal gaps" and, as a rule, these gaps have been simply ignored by shortening the duration of the time series.
The method used by us to fill in the inner gaps have been described by the authors elsewhere (Malvestuto, Beltrano, Testa, 2011: Eur.Phys.J.Plus (2011) 126:25 DOI: 10.1140/epjp/i2011-11025-9). Proceeding in this way, we succeeded in filling all the inner gaps for above mentioned 7 stations, so as to obtain 14 daily time series, 7 of minimum and 7 of maximum temperatures, without any data gaps.

3. DATA ANALYSIS OF THE 14 COMPLETE DAILY TEMPERATURE SERIES

3.1. Analysis of a series in the time domain and in the frequency domain

In general a meteorological data series may be regarded as the superposition of several simpler components:

- a trend, representing the deterministic tendential behavior on the large time scale. It indicates the background tendency of the phenomenon, and generally it is supposed to be a linear function of time;
- some cyclic oscillations, among which we can detect those, called *seasonal*, which have periods (one year, half year, etc.) that are directly linked to the astronomic cycles of the planet Earth, and other having either shorter periods or multi-year, pluri-decadal, or even quasi-secular, periods (for example, phenomena connected to the synodical cycle, or the large time-scale oscillations related to global phenomena like ENSO, NAO and MO, as well as the almost-periodical perturbations associated with the evolution of the main meteorological systems).
- the erratic component (also called "stochastic residual"), that represents the overall environmental noise; the latter includes the local meteorological variability at small scales and all the natural random fluctuations.

The analysis of the historical series is usually based on the assumption of stationarity of the erratic component, according to which the random factors, which have affected the behaviour of the time series in the past, continue to act unchanged in the present, and, probably, will even continue to exert their effects in the future. Thus, on following the standard approach of time series analysis, the erratic component of the time series is here assumed to be the realization of a stationary stochastic process, which one can attempt to describe by means of some parametric probability model.
After detecting and removing the trend and the main cyclic components from the original data series, the attention of our analysis will be focused onto the erratic component, that is, on the zero-mean signal obtained from the original signal after subtraction of

all its detected deterministic components. First, the correlogram of the erratic component is computed in order to get an estimate of the autocorrelation function of the underlying stochastic process. Second, a visual inspection of the shape of such correlogram may allow to hypothesize a parametric stochastic model, belonging to the so-called ARIMA class, that may hopefully be used to represent it in the best way (in the sense of a least-mean square optimization). The total variability associated with the erratic component will thus be partitioned in a variance due to the internal memory (or, to the degree of the auto-correlation) and a variance attributable only to a residual white noise, typically a normal white noise.

This model, on putting together the separate computer simulations of the deterministic and of the erratic components, would thus permit the artificial reconstruction in the past of the values of the given meteorological series and, possibly, its cautious prediction in the next future. Moreover, the choice of a well determined model of autocorrelation function permits the direct calculation of the power spectrum of the series by a simple application of the Wiener-Khinchin theorem (i.e., via the Blackman-Tuckey method of spectrum estimation).

3.2. Strategy for the analysis of the temperature series

All the 14 time series (7 of daily minimum, 7 of daily maximum temperature) were subjected to spectral analysis after correlogram evaluation. Every signal has been first processed in view of detecting the deterministic component (a trend plus some seasonal oscillations) and then extracting by subtraction a zero-mean residual, that represents the erratic component. Successively, on working on this component, that is always present with a conspicuous variance, a formal, standard representation has been tried via an autoregressive model of low order. In all cases a simple AR(1) model, with an optimal parameter value, depending on the station and on the type of temperature record, has proven to be sufficient. Moreover, even considering all the 14 time series, the spread of the values of such parameter proved to be rather narrow (see column 7 in Table 1), as if all the temperature records shared some universal feature characterizing their erratic components.

We recall that a stochastic process is said to be autoregressive of order p , briefly, an AR(p) process, if the value observed at time t_j is correlated to all the p previous values of the same process, that is, to the values observed at the instants t_{j-1} , t_{j-2} , ..., t_{j-p} . Such a process is characterized by the property of possessing an internal "memory" of indefinite length, but gradually vanishing with lag-time and appreciably different from zero only for time delays not much longer than p sampling intervals.

The procedure followed to process each time series always started with a preliminary computation just intended to recognize the frequencies of the relevant seasonal cycles present in the signal:

- first, a zero-mean signal was obtained by subtracting from the original series its mean value computed over the whole period;
- second, an explorative periodogram has been computed permitting a quick detection of the main deterministic frequencies involved in the series.

This preliminary treatment always resulted in pinning down the same pair of main deterministic frequencies, one corresponding to a 1-year period, the other one to a 6-month period.

As a consequence of this preliminary analysis, the following time series model has been adopted:

$$T(t) = a + bt + A_1 \cos(\omega_1 t - \varphi_1) + A_2 \cos(\omega_2 t - \varphi_2) + \tau(t) \quad (1)$$

where a and b are the linear-trend parameters, ω_1 and $\omega_2 = 2\ \omega_1$ are the pulsations corresponding exactly to a 1-year and 6-month periods, respectively, A_1 and A_2 are their corresponding centered amplitudes, φ_1 and φ_2 the respective angular phases, and $\tau(t)$ is the erratic component of the time series, namely, its stochastic residual, to be further modelled via a suitable auto-regressive process.

The six free parameters of the deterministic component, that is, the two linear-trend parameters (the slope b and the intercept a) and the four harmonic parameters (the two amplitudes A_1 and A_2 and the two phases φ_1 and φ_2) characterizing the two above mentioned seasonal oscillations, have been simultaneously determined via a multiple regression. The residuals with respect to this multiple regression are used to define the stochastic residual $\tau(t)$, that is assumed to have a purely stochastic nature.

Starting now to work exclusively on this residual signal, its correlogram has been computed in order to obtain an estimate of the underlying autocorrelation function. Then, an autoregressive AR(1) model has been fitted to the correlogram according to the formula

$$\tau(t) = \alpha\tau(t-1) + Z(t) \quad (2)$$

where α is a suitable constant and $Z(t)$ is a white noise with a variance to be estimated. The constant α , having an absolute value less than 1, is the only characteristic parameter of the AR(1) process and gives the shape of the corresponding theoretical autocorrelation function - see Eq.(3). The optimal value of α for the given time series is determined by the slope of the regression line of $\tau(t)$ vs. $\tau(t-1)$, while the variance of the white noise $Z(t)$, that must be equal to a percentage $1-\alpha^2$ of the variance of the full erratic component $\tau(t)$, is measured by the spread of the residuals around the regression line itself. In this way, the total variance of the stochastic component $\tau(t)$ is separated into a portion due to its auto-regressive behaviour and a portion connected to the residual white noise, which, by its very nature, cannot be given any further explanation.

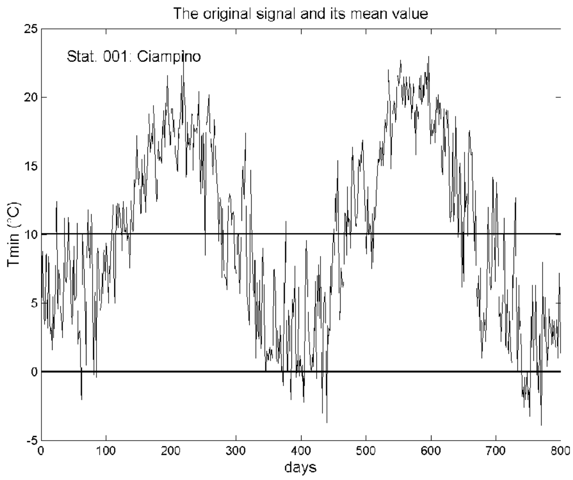

Fig. 2. Ciampino: A 2-year segment of the daily minimum temperatures recorded at Ciampino station from 1951 to 1999: about 800 days are shown. The thin horizontal line near 10°C evidences the average of the signal over the whole 50-year period.

The confirmation that the latter part of the variance is essentially due to a white noise is obtained on the ground of a practically flat appearance of the power spectral density of the residual signal *Z(t)*, the latter signal being computed at each time via Eq.(2), starting from $\tau(t)$ and after an optimal determination of the α parameter. As shown in Table 1, this modelling effort produces good results for both daily temperature series at each of the stations under examination.

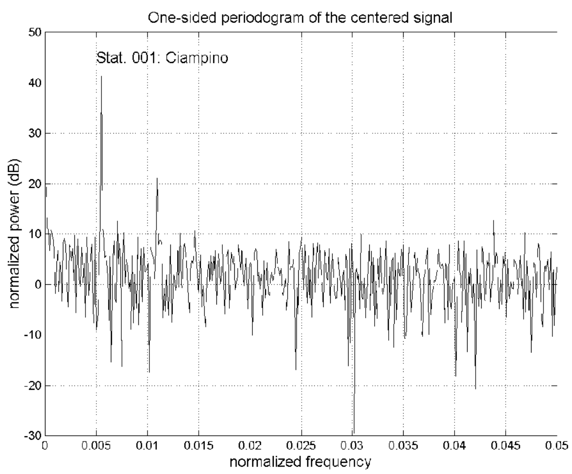

Fig. 3. Ciampino: daily minimum temperatures 1951-1999. The explorative periodogram with ordinates in log-scale, in a very short segment of frequencies near the zero frequency. The two frequencies of the main seasonal components are evident as sharp peaks emerging above the otherwise almost flat power level.

3.3 - Analysis of the temperature series

Each uninterrupted daily time series of temperature has been presented versus time, as in Figure 2a, where, for example, the series of daily minimum temperatures at Ciampino has been plotted on limiting the presentation to a short time interval of about 800 days, starting with January 1, 1951.

The first preliminary step of the analysis, as stated above, consists of merely subtracting from the complete signal its time mean, so as to obtain a centered signal, that is, a zero-mean signal (Figure 2b). In order to detect the presence of any cyclic components in the series, a periodogram of the whole 50-year centered signal has been calculated (see Figure 3). On the *x* axis we have the "normalized frequency" of the single harmonic component, $x \equiv \omega\tau/\pi = 2\tau/T$, that is a dimensionless quantity, where T is its period, $\omega \equiv 2\pi/T$ its pulsation, and τ is the sampling time-step (1 day) of the series. Note that the range of *x*-values over which the spectrum can be determined (namely, from zero frequency to the Nyquist frequency) is $0 \le x \le 1$. Obviously, once this normalized frequency is known, the period of the single cyclic component is easily computed. For example, if the time unit is 1 day (so that $\tau = 1$) a cyclic component having $x = 0.04$ has a period equal to 2(1 / 0.04) = 50 days.

On the other hand, on the *y* axis we have, for each normalised frequency, the power density expressed in decibel (dB). The shape of the periodogram shows how the variability of the signal is distributed over frequencies, namely, as a function of *x*. From the segment of periodogram shown in Figure 3 it is clear that the frequencies carrying most of the signal energy are the two ones corresponding to periods of 1 year ($x_1 = 0.0054$) and 6 months ($x_2 = 0.011 \approx 2x_1$).

After having detected the two above mentioned main frequencies, the two parameters of the linear trend (intercept and slope) and the 4 harmonic parameters (the two amplitudes and the two phases) of these two main oscillations have been simultaneously estimated by one multiple regression. In this way it has been possible to evaluate all the six parameters fixing the time behaviour of the whole deterministic component (the trend plus two seasonal oscillations) according to the model equation (1). As evident from Figure 4, the deterministic component, when plotted versus the original signal, already appears to capture most of its time variability. By subtracting now the deterministic component from the full original signal, a new signal, called "stochastic residual" or "erratic component", is obtained (see Figure 5).

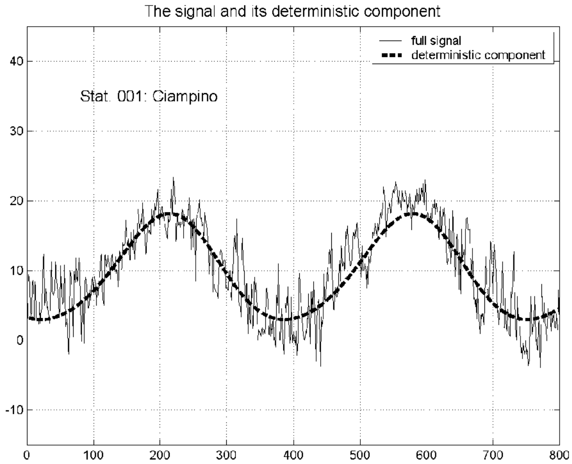

Fig. 4. The deterministic component of the signal (dashed line) plotted for comparison over the full original signal. The time span is again limited for both signals to the first 800 days.

We have no deterministic explanation for this residual signal. However, we want to be sure that this erratic component contains no more significant deterministic cycles. To this purpose, its periodogram has been computed and shown in Figure 6.
From a comparison with the previous periodogram shown in Figure 3, it is evident, first, that the two main peaks corresponding to the strong seasonal components are now absent (see lower panel) and, second, that the dB range of the ordinates has shrunk significantly (they vary now from -20 to +10 dB only), implying a rather more uniform distribution of the residual variability over frequency. On the other hand, this distribution is not so uniform as a pure white noise would imply.

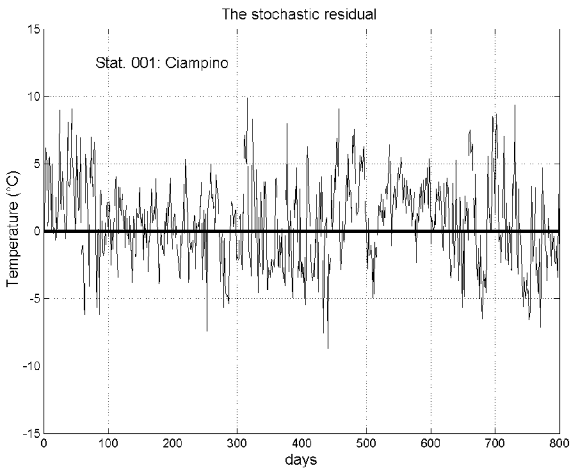

Fig. 5. Ciampino: daily minimum temperatures in the period 1951-1999. The stochastic residual of the signal, obtained after subtracting the deterministic component (namely, the trend and the two main seasonal cycles). Only the first 800 days are shown.

Therefore, the shape of this periodogram, while suggesting that the erratic component of our original signal is by now a genuine stochastic process containing no significant deterministic component, it also indicates that this residual stochastic process possesses some degree of internal memory, that is responsible for the evident monotonic decrease of the energy distribution over frequency by a factor of over 100 (see upper panel).
In the effort of appreciating the extent of this internal memory, the correlogram of the stochastic residual has been computed by means of the well known formula:

$$r_k = \frac{\sum_{j=1}^{N-k}\left(s_j-\mu\right)\left(s_{j+k}-\mu\right)}{\sum_{j=1}^{N}\left(s_j-\mu\right)^2} \qquad (3)$$

$$(k = 1, 2, ..., m)$$

where s_j represents the signal at time j and μ its overall mean value, whereas k indicates the time delay expressed in days, or in multiples of the basic sampling time step. In Figure 7 the behaviour of this correlogram vs. time-lag has been shown (dashed line) up to a maximum time delay of 30 days (k_{max} = 40). The inspection of the auto-correlogram of a stationary series should permit to quantify the level of serial dependence among the data and thus to quantitatively evaluate the link existing between one value of the series and the few previous ones as a function of their separation in time. The generic r_k, being substantially a correlation coefficient, assumes values between -1 and +1, and the nearer its absolute value is to one, the more correlated the original data series is with its lagged copy, namely with the series obtained by shifting it k time steps apart (negative values of r_k obviously indicating an anti-correlation).

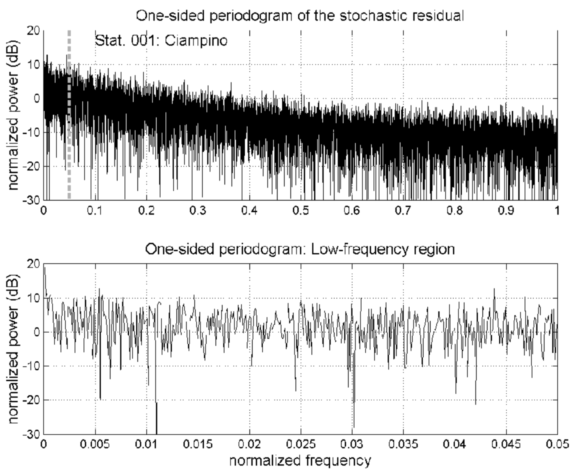

Fig. 6. Ciampino: daily minimum temperatures in the

period 1951-1999. Upper panel: the periodogram of the stochastic residual over the whole frequency range. The vertical dashed line marks the band of the low frequencies that is zoomed in the lower panel. Here, it can be seen that the two peaks corresponding to the main seasonal cycles, well evident in Figure 3, are no longer present.

Now, if the signal consisting of the stochastic residual under study were the realization of an autoregressive AR(1) process with basic parameter α , its theoretical autocorrelation function would have the following shape:

$$\rho_k = \alpha^{|k|} \qquad (4)$$

$$(-1 < \alpha < 1)$$

no matter which the sampling time step was, and its spectrum would exhibit a distinct peak in the origin of frequencies, like that one visible in Fig. 6. In the case $0 < \alpha < 1$ the behaviour (4) corresponds to an exponential monotonic decrease with increasing time delay *k*.

For a comparison with the actual correlogram, such ideal behaviour has been shown in Figure 7 (solid curve) after having determined the parameter α via an optimization. More precisely, an optimal estimate of α is given by the slope of the regression line through the scatterplot of points in the plane (S_j, S_{j+1}), S_j denoting the value of the stochastic residual at time j . Such scatterplot is presented in Figure 8a.

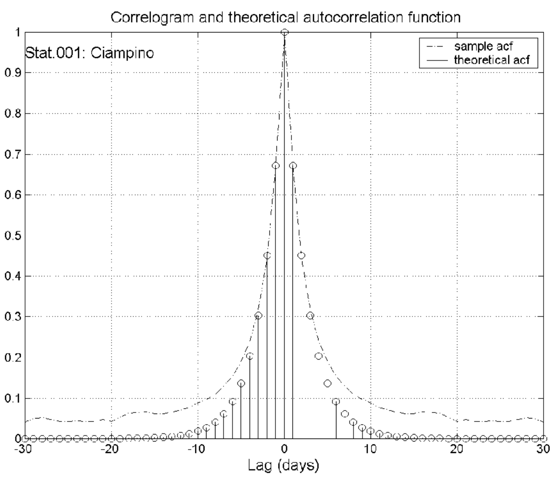

Fig. 7. Ciampino: daily minimum temperatures in the period 1951-1999. Correlogram of the stochastic residual (dashed curve) and the hypothetical auto-correlation function (stem-plot) of a related AR(1) process with its parameter optimally determined through a linear regression (see Figure 8a) .

Figure 7 shows that the sample correlogram (the dashed curve) follows approximately the shape of the hypothesized theoretical correlation function (4), with the optimal α determined via linear regression, at least far enough from the tails. This congruence remains true for all the stations examined, up to time delays less than about one week, while for longer time lags it is seen that the correlogram values, though slowly decreasing toward zero, remain positive and definitely higher than the corresponding theoretical tails. In other words, the agreement with the theoretical autocorrelation function of an AR(1) process is satisfactory only in a short range of small time lags. Nevertheless, we can attempt to formulate a very concise model of each observed correlogram by using the optimal AR(1) process consistent with it, as stated by the model equation (2), the parameter α and the variance of the white noise *Z(t)* having to be contextually determined during the optimization process.

In order to appreciate the extent of the auto-regressive character of the time series, and so to check the goodness of fit of the adopted model (2), we have shown in Figure 8a, for the examined station, the regression line superimposed to the scatterplot of data points (S_j, S_{j+1}). Hence one can see the strong dependence existing between the temperature value measured today and the one measured tomorrow. Far from forming a random distribution, the data points exhibit a neat tendency to align along a certain straight line having a positive slope, in this case near to 2/3, implying a correlation coefficient as high as 67%.

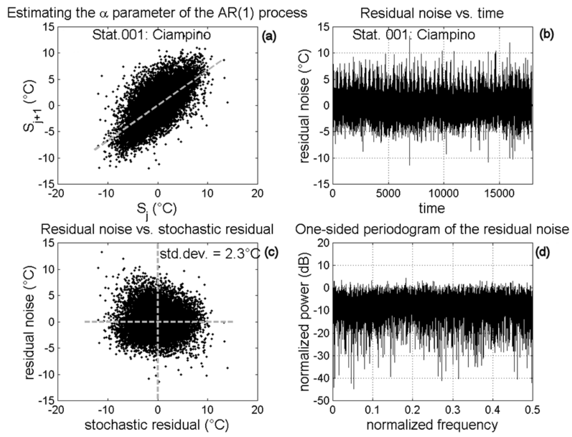

Fig. 8. Ciampino station. (a) The plot of the stochastic residual vs. itself one day before shows a definite slope permitting the AR(1) parameter to be estimated via regression; (b) The full plot vs. time of the residual noise, obtained by subtracting the autoregressive component from the stochasti residual (c) The plot of the residual noise vs. the full stochastic residual no longer shows a preferred direction of alignment for data-points, which suggests we are left with a white noise; (d) this is confirmed by the flat appearance of the periodogram over the whole frequency range (0.5 is the dimensionless Nyquist frequency).

The residual variability associated to the deviations of the data points with respect to their regression line is quantitatively estimated by the standard deviation of the residuals, which coincides with the square mean amplitude of the residual white noise *Z(t)* appearing in the model equation (2). The significantly high correlation coefficient associated to this linear regression confirms the existence of a close relationship between the temperature of one

day and the one of the next day and justifies the use of the AR(1) model to concisely represent the internal memory inherent to the stochastic residual extracted from the original signal.
A stochastic model is good if it is able to catch the main part of the variability associated to the given time series, that is, if the time series of the residual signal *Z(t)* is similar to the realization of a white noise. In our analysis the goodness of the proposed model is confirmed by the nature of the residual noise *Z(t)*, illustrated in Figure 8b-8c. In particular, a first assessment of the goodness of the adopted model (2) can be already obtained by looking at the scatter plot in Figure 8c, whence it appears how random the distribution of the residual noise is in comparison with the data points in Fig. 8a and considering how difficult it would be now to try predicting the value of this remnant signal by using the simultaneous value of the full erratic component. The vertical extent of the scatter plot in Figure 8c, when compared to its horizontal extent, also evidences how the standard deviation associated to the residual noise (about 5°C) is almost two times smaller than that of the full stochastic residual (close to 10 °C). It can be therefore concluded that the auto-regressive character of the process is able alone to explain about the half of the variability carried by the full erratic component extracted from the original temperature signal, while the other half has no apparent explanation. Indeed, the appearance of the residual noise, unlike the full stochastic component of the signal, shown in Figure 5, is quite similar to that of a white noise, since its power spectrum appears practically flat, as evident from Figure 8d. More precisely, since the variance of this noise is about half the variance of the full stochastic residual, we can conclude that, by means of the AR(1) model, a consistent part of its variability has been successfully explained on the base of its auto-regressive structure.
As for the sample probability distribution underlying this white noise, shown in figure 9, it is a bell-shaped curve, apparently not far from a Gaussian, but significantly different from it (even adopting a conservative 95 % confidence level in the normality test) because of the high χ^2 value.

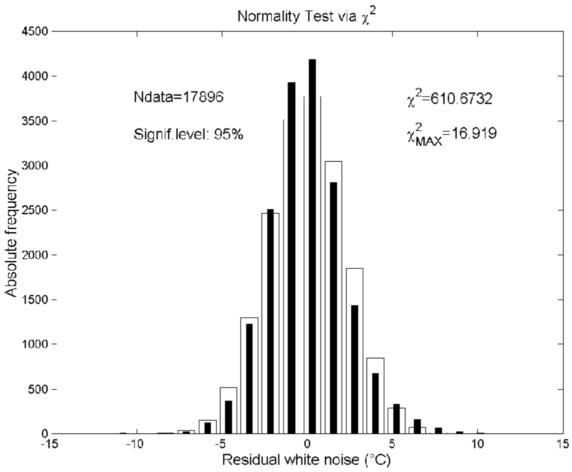

Fig.9: A comparison between the observed values (narrow bars) of the residual noise and the corresponding predicted ones (wide bars) by a Gaussian distribution with the same mean and variance. The very high χ^2 value indicates that the normality hypothesis must be rejected. In fact the observed histogram is more sharply peaked and also fairly skewed with a definite bias toward negative values.

This efficient and succinct modelling of the erratic component of the original temperature series has proved to fit well for all the seven stations examined for both minimum and maximum temperatures.
Finally, it can be observed that, if one is able to generate a noise with the right standard deviation (the latter being a very modest part of the whole variability of the original temperature series), it is possible to build, for any real temperature series, an artificial signal as the sum of a trend, two seasonal oscillations and an AR(1) stochastic process, including the given noise. By means of this artificial signal the real temperature signal can be adequately simulated, in the sense that the artificial and the real signals can hardly be distinguished from each other. This is evident from the comparison between the two panels (a) and (d) in Figure 10, the first of which shows the real signal as measured at Ciampino station, whereas the second is the artificial signal built as described above (a gaussian white noise was used).

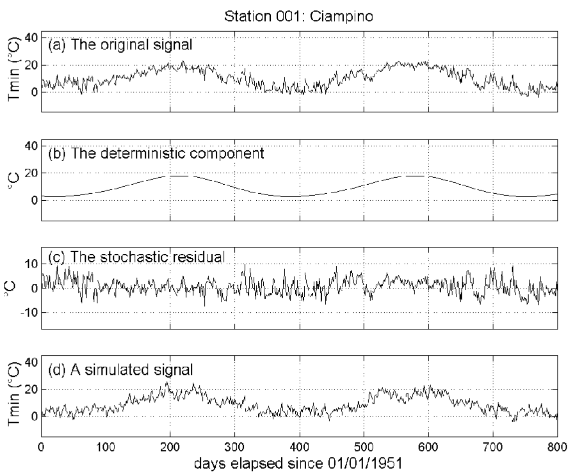

Fig. 10. Ciampino: (a) The original signal (minimum temperatures); (b) The deterministic component of the signal. (c) The stochastic residual of the signal. (d) An artificial signal generated by computer simulation on using an AR(1) process with the right α parameter and with the noise variance typical of the given climatic station.

4. Results and discussion

The results obtained via spectral analysis for the daily minimum and maximum temperature series are quite similar for all the stations, thus confirming that the stochastic model of type AR(1) is able to adequately represent the erratic component of all temperature records. The detailed quantitative outcomes for all the 7 climatic stations are summarized in Table 1.

Table I - Results of spectral analysis: Parameters characterizing the deterministic part (first 6 columns) and the stochastic part (last 2 columns) of the two temperature series observed at each station. Slope of the trend (and the associated value of F-statistic) are in column 1; the trend intercept is in column 2; the amplitudes of the two seasonal components, A_1 e A_2 are in columns 3 and 5, while the corresponding phases, φ_1 e φ_2 , expressed in degrees, are in columns 4 and 6. The basic parameter α of the modelling AR(1) process is in column 7, and the standard deviation of the residual noise Z_t is in the last column.

	Trend Slope	Intercept al 1975	A_1	φ_1	A_2	φ_2	AR(1) α	Noise Z_t std. dev.
Station name	°C/century (F_{crit} =3.84)	°C	°C	degrees	°C	degrees	---	°C
Column	1	2	3	4	5	6	7	8
Daily Maximum Temperatures								
Ciampino	-1.70 (13.95)	20.7	9.4	204	1.1	90	0.71	2.0
Roma Collegio Romano	0.39 (2.66)	20.4	9.5	202	1.0	86	0.76	1.7
Fiumicino	2.03 (30.76)	20.4	8.1	208	0.9	92	0.67	1.8
Ardea	0.89 (10.83)	19.9	8.3	207	0.8	91	0.72	2.1
Roma Eur	0.43 (2.94)	21.5	9.5	203	1.0	92	0.76	2.1
Roma Macao	0.12 (0.87)	20.7	9.3	204	1.0	95	0.73	2.4
Roma Monte Mario	-0.12 (0.13)	20.4	9.7	204	1.2	94	0.75	2.4
Daily Minimum Temperatures								
Ciampino	0.68 (8.09)	10.0	7.6	208	0.7	80	0.67	2.3
Roma Collegio Romano	2.40 (64.91)	12.1	7.8	206	0.8	82	0.68	2.0
Fiumicino	0.49 (4.11)	10.6	7.4	210	0.7	84	0.62	2.5
Ardea	-3.36 (88.62)	9.8	7.2	211	0.6	83	0.70	2.6
Roma Eur	0.84 (9.91)	10.8	7.6	208	0.7	64	0.70	2.6
Roma Macao	0.63 (6.63)	11.7	7.6	206	0.7	82	0.70	2.3
Roma Monte Mario	0.97 (13.75)	11.2	7.5	209	0.7	83	0.71	2.4

Each series was decomposed into a deterministic component, made of a linear trend (see the first 2 columns in the table) and of two seasonal harmonic oscillations (columns 3 to 6), the larger having a 1-year period, the smaller a 6-month period, and a stochastic residual, which has been successfully modelled by means of an AR(1) process, whose computed parameters are listed in the last two columns of Table 1.

The average daily maximum temperatures over the period 1951-1999 range from 19.9 to 21.5°C among stations, while the average daily minimum temperatures range from 10 to 11.7°C (see column 2 of the table), with a difference between minimum and maximum temperatures of about 10°C for all stations. As for the trends of minimum temperatures (column 1), with the exception of Ardea station, where a rather important negative trend appears (a -

3.36°C per century that deserves further attention), all the stations present positive trends, that are all statistically significant at the 95% level, since the corresponding F-values exceed the critical threshold F_{crit} =3.84 appropriate to the given number of degrees of freedom. In particular, Roma-Collegio Romano presents a positive trend of about 2.5°C per century. It is likely that these high values are due to the *heat-island effect* enhancing locally the general tendency to a slow increase of temperature with time. However, this tendency is not in line with the results we obtained for the daily maximum temperatures: in fact the F-values allow to regard as significant only three of the generally positive, but modest, trends found at the 7 stations, and one of them is even negative. The three stations are Fiumicino with an extant +2.03°C/century, Ardea with a more modest 0.9°C/century and Ciampino that, against the general tendency, features a distinct -1.7°C /century. However a best assessment of the trends can be done on considering yearly averaged data, or at most monthly data, rather than daily data as those we have processed. As a matter of fact, in the present work the estimation of the trends is of secondary importance with respect to the main objective, that consists of investigating the stochastic properties of the temperature signals.

For what concerns the variability among the various stations of the two main seasonal cycles detected (see columns 3 to 6 of Table 1) , it can be noted that for both of them there is, as plausible, a certain uniformity over the whole geographic region. In particular, the yearly cyclic component of the daily maximum temperatures, having a yearly excursion comprised typically between 16 and 20°C (or a centered amplitude between 8.1 and 9.7°C), shows, as it is reasonable to expect, its peak in the second decade of July for all the stations (the yearly phases φ_1 being near 205 degrees) and its trough 6 month earlier, at a phase near 25 degrees (25 = 205 + 180 – 360), that is, within the third decade of January. In turn, the series of daily minimum temperatures present yearly excursions comprised typically between 14 and 15.5°C (or centered amplitudes between 7.2 e 7.8°C), the phases being almost identical to those of the daily maximum temperatures, what means: peaks in the second decade of July, though with a bias to culminate few days later, or toward the end of July, and troughs occurring in the last decade of January, again with a few-day delay with respect to the corresponding dates of daily maximum temperatures. For what concerns the 6-month seasonal cycle, we note that in all stations its centered amplitude is about 10 times weaker than that of the yearly cycle, and they are all near 0.7°C for the daily minimum and 1°C for the daily maximum (column 5 of the table); the corresponding angular phases (column 6) imply, for both temperature series, two peaks occurring in half February and half August and the corresponding two troughs in half May and half November (the values of angular phase being mostly in the interval 80 – 95 degrees). Daily minimum temperatures show however a systematic tendency to culminate few days earlier than the maximum do.

Columns 7 e 8 of Table 1 report, respectively, the optimal values of the α parameters (or the correlation coefficients) and of the standard deviation of the residual noise *Z(t)*. We again note a discrete uniformity of values among the various stations. In particular, once the deterministic part of the signal has been eliminated, the remaining erratic component can be modelled via an auto-regressive process of the AR(1) type, with an internal correlation coefficient α that is close to 0.7 for all stations (see column 7 of the Table 1). This level of internal correlation of 70%, given the high number of data points used, can be regarded as highly significant. The mean level of the residual white noise *Z(t)* (see column 8), which is obtained after subtraction from the erratic component of its auto-regressive part, is approximately 2°C for the maximum temperatures (1.7-2.4°C), and somewhat higher for the minimum (2-2.6°C). This quantifies the small part of the variability of the original temperature signal that is not predictable at all. The reason why the noise level is systematically higher for the minimum temperatures is not immediately clear and deserves further investigation. This noise background in the temperature signal may occasionally amplify the natural variability due to the deterministic part of the signal and, though of minor entity, can produce not negligible consequences on vegetation growth and, particularly, on crops. In fact, a greater amount of "energy" available for the processes of growth and for the development of crops results from a temperature increase. High temperatures in winter and spring tend to establish an early start of the growing season for crops, making them vulnerable to sudden drops in temperature, that can produce sensitive effects at the time of flowering. Second, a consequent lengthening of the growing season in situations of prolonged water scarcity, combined with exceptionally high temperatures, can lead to stress for crops, with serious consequences on productivity. Increasing temperatures also promote the proliferation of insect pests due to the lengthening of the growing season and to the increase in their survival probability across the winter.

5. ACKNOWLEDGMENTS

This study has been carried out thanks to financing by the Presidency of the Republic, within the research project "Impact of climate and of local atmospheric circulation on Mediterranean coastal ecosystems: Castelporziano Presidential Tenuta as a case's study."

6. REFERENCES

[1] VON STORCH H. AND ZWIERS F.W., *Statistical Analysis in Climate Research*, (Cambridge Univ. Press, 1999) pp. 484.

[2] EASTERLING D.R., EVANS J.L., GROISMAN P.YA., KARL T.R., KUNKEL K.E., AMBENJE P., *Bull. Am. Meteorol. Soc.*, **3** (2000) 417.

[3] DONNELLY, A. JONES, M.B. & SWEENEY, J, *International J. of Biometeorology*, **49** (2004) 1-12.

[4] BOX G.E.P., JENKINS G.M., *Time Series Analysis – Forecasting and Control.* (Holden Day, S.Francisco, 1976) pp. 576.

[5] VIGLIZZO E.F., ROBERTO Z.E., FILIPPIN M.C., PORDOMINGO A.J., *Agriculture Ecosystems and Environment*, **55** (1995) 7-16.

[6] SUPPIAH R. AND HENNESSY K. J., *International J. of Climatology,* **10** (1998) 1141–1164.

[7] PLUMMER N., SALINGER M.J., NICHOLLS N., SUPPIAH R., HENNESSY K.J., LEIGHTON R.M., TREWIN B., PAGE C.M. AND LOUGH , J.M., *Climatic Change* **42** (1999) 183–202.

[8] GRUZA G., RANKOVA E., RAZUVAEV V, BULYGINA O., *Climate Change*, **42** (1999) 219.

[9] TÜRKEŞ M., *International J. of Climatology*, **18** (1998) 649–680.

[10] HESS T.M., STEPHENS W., MARYAH U.M., *Agric. For Meteorol.,* **74** (1995) 87–97.

[11] S.J. MASON , *Climatic Change,* **32** (1996) 35–54.

[12] ZHAI P., SUN A., REN F., LIU X., GAO B. , ZHANG Q., *Climatic Change,* **42** (1999) 203–218.

[13] HEINO R., BRÁZDIL R., FØRLAND E., TUOMENVIRTA H., ALEXANDERSSON H., BENISTON M., PFISTER C., REBETEZ M., ROSENHAGEN G., RÖSNER S., WIBIG J., *Climatic Change*, **42** (1999) 151–181.

[14] TESTA O., LAVAGNINI A., MALVESTUTO V., DIODATO L., D'ABRAMO F., *Il Nuovo Cimento*, **29C** (2006) 323-340.

[15] TESTA O., COLACINO M., LAVAGNINI A., MALVESTUTO V., *Il Nuovo Cimento*, **29C** (2006) 191-213.

[16] BOOTSMA A., *Climatic Change*, **26** (1994) 65–88.

[17] TRENBERTH, K.E., P.D. JONES, P. AMBENJE, R. BOJARIU, D. EASTERLING, A. KLEIN TANK, D. PARKER, F. RAHIMZADEH, J.A. RENWICK, M. RUSTICUCCI, B. SODEN AND P. ZHAI, 2007: Observations: Surface and Atmospheric Climate Change. In: *Climate Change 2007: The Physical Science Basis.* Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[18] VOSE, R.S., EASTERLING D.R., AND GLEASON B., Maximum and minimum temperature trends for the globe: an update through 2004. *Geophys. Res. Letters*, **32**, L23822 (2005a) 5 pp., doi: 10.1029/2005GL024379.

[19] SMITH, T.M., AND REYNOLDS R.W., A global merged land and sea surface temperature reconstruction based on historical observations (1880–1997). *J. Clim.*, **18** (2005) 2021–2036.

[20] BRUNETTI M., BUFFONI L., MAUGERI M., NANNI T., *Theor. Appl. Climatol.*, **66** (2000a) 49-60.

[21] BRUNETTI M., MAUGERI M., NANNI T., *Theor. Appl. Climatol.* , **65** (2000b) 165-174.

[22] BÖHM R., AUER I., BRUNETTI M., MAUGERI M., NANNI T., Regional temperature variability in the Euopean Alps: 1760-1998 from homogenized instrumental time series, *International J. of Climatology*, **21** (2001) 1779-1801.

[23] CICCARELLI N., HARDENBERG J. VON, PROVENZALE A., RONCHI C., VARGIU A., PELOSINI R., *Global and Planetary Change,* **63** (2008) 185-195.

[24] BÉLANGER G., ROCHETTE P., CASTONGUAY Y., BOOTSMA A., MONGRAIN D., RYAN D.A.J., Climate Change and Winter Survival of Perennial Forage Crops in Eastern Canada. *Agronomy J.*, **94** (2002) 1120-1130.

[25] BENISTON M., *Pure Appl. Geophys.*, **165** (2005) 1587-1606.

Appendix A: The features of the database used in the present study

The data came to us stored in text files with a sampling time step equal to one day, the file format being the following:

yyyy (year) *mm* (month) *dd* (day) *ttt.d* (temperature in °C).

The first three columns contain invariably all the serial dates from 1/1/1951 to 31/12/1999, even in the presence of missing data, since in the latter case a conventional numeric code -999, instead of a normal temperature, was used to signal the absence of the data.

The quality of the data was assessed at IBIMET (a CNR Institute of Biometeorology in Florence), where a staff of dedicated personnel performed the quality control and the necessary homogenization of the time series under the cover of a specific project, financed by the National Academy of Sciences (alias "Academy of the XL"), named "*Impatto del clima e della circolazione atmosferica locale sugli ecosistemi costieri mediterranei: la Tenuta Presidenziale di Castelporziano come lo studio di un caso*". The final report, titled "Analisi climatica della zona costiera" was presented to the Academy in 2009 with authors: Marina Baldi, Paolo Coccimiglio, Alfonso Crisci, Francesco Primo Vaccari, and Gianpiero Maracchi, (26 pp). For the methods employed and the results reached in this preliminary validation of the various temperature series, we refer to this very detailed report.

For what concerns the missing data, the following two tables illustrate the status of the database for both T_{min} and T_{max} data in three successive stages: i) number of original gaps, as the data arrived to the above mentioned IBIMET validation group; ii) number of gaps after the preliminary treatment done at IBIMET, where a partial gap filling was also performed; iii) number of gaps remaining in the data after our additional gap filling process, illustrated in this paper (see section 2.3) .

Tmax period: 1951-1999 (17897 days)			
Station name / number	Gaps number		
	Original	After 1st treatment	After 2nd treatment
1) Roma Ciampino	208 (1.1%)	12 (0.1%)	0
2) Roma Collegio Romano	0.0 (0.0%)	12 (0.1%)	0
3) Roma Fiumicino (1959-1999)	3345 (18.3%)	2994 (16.7%)	0
4) Ardea	6297 (34.5%)	47 (0.3%)	0
5) Roma Eur	14001 (76.7%)	12 (0.1%)	0
6) Roma Macao	8998 (49.3%)	13 (0.1%)	0
7) Roma Monte Mario	12776 (70.0%)	13 (0.1%)	0

Tmin period: 1951-1999 (17897 days)			
Station name / number	Gaps number		
	Original	After 1st treatment	After 2nd treatment
1) Roma Ciampino	206 (1.1%)	12 (0.1%)	0
2) Roma Collegio Romano	0.0 (0.0%)	12 (0.1%)	0
3) Roma Fiumicino (1959-1999)	3332 (18.2%)	2984 (16.7%)	0
4) Ardea	6297 (34.5%)	46 (0.3%)	0
5) Roma Eur	14001 (76.7%)	12 (0.1%)	0
6) Roma Macao	8998 (49.3%)	13 (0.1%)	0
7) Roma Monte Mario	12776 (70.0%)	13 (0.1%)	0

Note that out of the original 15 stations for which we had data covering the period 1951-19999, we decided to proceed with our analysis using only the seven stations for which we succedeed in entirely filling the residual gaps by using our above mentioned method. The data of the other eight stations, where some gaps remained even after applying our method, were not used in this work. For the seven stations processed the temperature series extend all from 01-01-1951 to 31-12-1999, except the two Fiumicino series which extended from 01-01-1959 to 31-12-1999 (only 41 years of data, instead of 49).

Before proceeding to analyze the single stations a brief correlation analysis (at zero time-lag) was carried out revealing the following correlation structure among stations:

Correlation matrix for Tmin-series (lower left) and for Tmax-series (upper right):

Stat.	1	2	3	4	5	6	7
1	1	0.9861	0.9749	0.9525	0.9848	0.9726	0.9782
2	0.9715	1	0.9760	0.9551	0.9851	0.9711	0.9769
3	0.9665	0.9584	1	0.9506	0.9731	0.9591	0.9650
4	0.9364	0.9264	0.9294	1	0.9592	0.9494	0.9530
5	0.9695	0.9645	0.9597	0.9472	1	0.9852	0.9909
6	0.9638	0.9673	0.9495	0.9325	0.9715	1	0.9835
7	0.9718	0.9709	0.9618	0.9457	0.9849	0.9768	1

After subtraction of the trend, of the seasonal components and of the autoregressive stochastic part from the full signal, at each station a residual white-noise signal was obtained. These residual noises were again analyzed to ascertain the possibly remnant cross correlation, yielding the following situation.

Post-Correlation matrix for Tmin-series (lower left) and for Tmax-series (upper right):

Stat.	1	2	3	4	5	6	7
1	1	0.8143	0.6664	0.5189	0.8305	0.7182	0.7995
2	0.7625	1	0.6690	0.5038	0.8023	0.6866	0.7826
3	0.7383	0.6675	1	0.4497	0.6503	0.5333	0.6202
4	0.6824	0.6567	0.6276	1	0.5853	0.5200	0.5609
5	0.8282	0.7960	0.7208	0.7668	1	0.8489	0.9230
6	0.7748	0.7818	0.6457	0.7012	0.8636	1	0.8459
7	0.8252	0.8072	0.7094	0.7431	0.9211	0.8846	1

From the comparison of these two correlation matrices it appears that for almost all the pairs of stations a relevant part of the initial correlation remain even after the removal of the deterministic components of the signal and of the autoregressive part. This indicates that the white noises superimposed to the temperature series are likely to be referable to some synoptic scale fluctuations rather than to random disturbances of mere local nature.