CrossMark

ORIGINAL PAPER

A stochastic model for the analysis of maximum daily temperature

B. Sirangelo[1] · T. Caloiero[2] · R. Coscarelli[3] · E. Ferrari[4]

Abstract In this paper, a stochastic model for the analysis of the daily maximum temperature is proposed. First, a deseasonalization procedure based on the truncated Fourier expansion is adopted. Then, the Johnson transformation functions were applied for the data normalization. Finally, the fractionally autoregressive integrated moving average model was used to reproduce both short- and long-memory behavior of the temperature series. The model was applied to the data of the Cosenza gauge (Calabria region) and verified on other four gauges of southern Italy. Through a Monte Carlo simulation procedure based on the proposed model, 10^5 years of daily maximum temperature have been generated. Among the possible applications of the model, the occurrence probabilities of the annual maximum values have been evaluated. Moreover, the procedure was applied for the estimation of the return periods of long sequences of days with maximum temperature above prefixed thresholds.

✉ R. Coscarelli
coscarelli@irpi.cnr.it

[1] Department of Environmental and Chemical Engineering (DIATIC), University of Calabria, Rende (CS), Italy

[2] National Research Council of Italy, Institute for Agriculture and Forest Systems in the Mediterranean (CNR-ISAFOM), Rende (CS), Italy

[3] National Research Council of Italy, Research Institute for Geo-hydrological Protection (CNR-IRPI), Rende (CS), Italy

[4] Department of Computer Engineering, Modeling, Electronics, and Systems Science (DIMES), University of Calabria, Rende (CS), Italy

1 Introduction

Nowadays, investigation on air temperature has achieved relevant importance because of its influence on all natural systems and human activities, such as crop growth (Verdoodt et al. 2004; Bechini et al. 2006), agro-ecological zoning (Caldiz et al. 2001; Ye et al. 2008), and food security assessment (Ye and Van Ranst 2009, 2002; Ye et al. 2013). Moreover, high temperatures can cause an increase of death rates, especially when the data rise above critical values (Kunst et al. 1993; Curriero et al. 2002; Hajat et al. 2002; Keellings and Waylen 2012).

Stochastic modeling and simulation of daily meteorological data is a prominent subject in literature since several decades. A classical approach for the analysis of temperature data, based on the procedures presented by Yevjevich (1972), was proposed by Richardson (1981), who considered maximum and minimum temperatures as continuous multivariate stochastic processes. Generally, the stochastic models, used to reproduce climatic and hydrological series, are the autoregressive (AR), the moving average (MA), the autoregressive moving average (ARMA), and the autoregressive integrated moving average (ARIMA) models (Box and Jenkins 1976; Grimaldi et al. 2005). A limitation of these models is that they can only capture the short-range dependence, thus presenting a lack of flexibility in reproducing the combined effect of short- and long-memory (Box and Jenkins 1976). Since Hurst (1951) detected the presence of long-term persistence in data series studying the Nile River levels, the need of long memory in time series modeling has been pointed out. In fact, long-range dependence has been encountered in various hydrological data (Lye and Lin 1994; Pelletier and Turcotte 1997; Koscielny-Bunde et al. 2006; Ehsanzadeh and Adamowski 2010). Doukhan et al. (2003) evidenced that long-range dependent processes are

characterized by a hyperbolic decrease of the autocorrelation function and are closely related to self-similarity. Recently, Prass et al. (2012) found that long-range dependence may affect the performance of time series models with a short time step. Moreover, incorporating long-range dependence into time series modeling is also conceptually important, since the model should capture the behavior of the data as realistically as possible. To this aim, Granger and Joyeux (1980) and Hosking (1981) proposed the fractionally differenced ARIMA models (FARIMA or ARFIMA) as an extension of the ARIMA models. The differencing order of the FARIMA models can be fractional, thus providing flexibility in that they can capture both short- and long-memory behavior, by varying the autoregressive and moving average components and using few parameters (Lohre et al. 2003; Montanari et al. 1997; Prass et al. 2012). This is an advantage of the FARIMA modeling framework because other long-memory models, capable of reproducing the Hurst phenomenon, such as the fractional Gaussian noise, have no flexibility in the choice of the short-memory autocorrelation structure (Koutsoyiannis 2002). In fact, it is shown that the fractionally integrated time series models are much more accurate than the traditional autoregressive models employing a similar number of parameters (Caballero et al. 2002).

Many studies have been developed on FARIMA models in literature. Bisaglia and Grigoletto (2001) proposed a bootstrap-based method to construct prediction intervals for FARIMA processes. Rupasinghe and Samaranayake (2012) and Rupasinghe et al. (2014) introduced a simpler alternative method, based on the sieve bootstrap approach of Alonso et al. (2002) and Alonso et al. (2003). Several contributions have been proposed in hydrology (e.g., Hosking 1984), climatology (e.g., Baillie and Chung 2002), and temperature (e.g., Smith 1993). As regards in particular hydrological studies, Montanari et al. (1997) proposed a FARIMA model for the analysis of monthly and daily inflows of Lake Maggiore (Italy). Montanari et al. (2000) applied a special form of the generalized FARIMA process to the Nile River monthly flows at Aswan. They combined the generalized FARIMA approach with a multiplicative ARIMA approach, which allowed to model seasonal and non-seasonal long and short memory. Sheng and Chen (2011) developed a new model, based on FARIMA with stable innovations, to analyze the data and predict the future elevation levels of Great Salt Lake. Yang and Bowling (2014) used a FARIMA model to estimate the long memory in daily stream flow for basins in the Upper Great Lakes region.

In this paper, a stochastic model which adopts the FARIMA approach is proposed for the analysis of daily temperature. Specifically, the different steps of the proposed model concern the data deseasonalization, by means of a truncated Fourier series expansion; the data normalization, through the transformations introduced by Johnson (1949); and the

analysis of the correlation structure through a FARIMA model. The model was applied to a maximum temperature series and verified on other four southern Italian gauges. As an example of the possible model applications, some features of the maximum daily temperature have been evaluated by means of a Monte Carlo simulation procedure.

2 Stochastic modeling of daily temperature

Let us indicate with $i = 0, 1, 2, \ldots$ the days from a generic starting point, and let us define $T(i)$ as the temperature characterizing the generic i-day. The $T(i)$ values can be the maximum, $T_{\max}(i)$, or the minimum, $T_{\min}(i)$, or even the mean temperature, $T_{\text{mean}}(i)$, of the day. A sequence of $T(i)$ observations can be explained as a realization of a discrete parameter stochastic process which shows a cyclostationarity condition, with period D equal to a year ($D = 365.25$ days).

2.1 Deseasonalization

The temperature is not a stationary variable (Jewson and Caballero 2003; Campbell and Diebold 2005), where the non-stationarity can be caused by a deterministic seasonal component or by a monotonic trend. Assuming the trend component equal to zero, the $T(i)$ process can be reduced to a weakly stationary standardized process, $Y(i)$, through the transformation (Grimaldi 2004; Montanari et al. 1997, 2000; Prass et al. 2012)

$$Y(i) = \frac{T(i) - \mu_T(i)}{\sigma_T(i)} \tag{1}$$

where $\mu_T(i)$ and $\sigma_T(i)$ are the mean and the standard deviation functions of the $T(i)$ process, respectively. The mean and the variance functions can be described by means of the truncated Fourier series

$$\mu_T(i) = \frac{1}{2} a_{\mu,0} + \sum_{j=1}^{n_{h,\mu}} \left[a_{\mu,j} \cos\left(\frac{2\pi j}{D} i\right) + b_{\mu,j} \sin\left(\frac{2\pi j}{D} i\right) \right] \tag{2}$$

$$\sigma_T^2(i) = \frac{1}{2} a_{\sigma^2,0} + \sum_{j=1}^{n_{h,\sigma^2}} \left[a_{\sigma^2,j} \cos\left(\frac{2\pi j}{D} i\right) + b_{\sigma^2,j} \sin\left(\frac{2\pi j}{D} i\right) \right] \tag{3}$$

in which $n_{h,\mu}$ and n_{h,σ^2} are the number of harmonics, while $a_{\mu,0}, a_{\mu,j}, b_{\mu,j}$ and $a_{\sigma^2,0}, a_{\sigma^2,j}, b_{\sigma^2,j}$ are the coefficients of the Fourier expansion of the mean and the variance functions, respectively.

Given a sample of observed temperature, $t(i_k)$, with $k = 1, 2, \ldots K$ which corresponds to the days $i_1, i_2, \ldots, i_k$, the functions $\mu_T(i)$ and $\sigma_T^2(i)$ can be estimated through their analogous sample values, $m_T(i)$ and $s_T^2(i)$.

By using the least squares method, the Fourier coefficients of the mean function can be estimated by minimizing the following function:

$$S_\mu^2\left(a_{\mu,0}, a_{\mu,1}, b_{\mu,1}, \ldots\right)$$
$$= \sum_{k=1}^{K} \left\{ t(i_k) - \frac{1}{2} a_{\mu,0} - \sum_{j=1}^{n_{h,\mu}} f_\mu\left(i_k j; a_{\mu,j}, b_{\mu,j}\right) \right\}^2 \quad (4)$$

where $f_\mu(i_k j; a_{\mu,j}, b_{\mu,j}) = a_{\mu,j} \cos(2\pi i_k j/D) + b_{\mu,j} \sin(2\pi i_k j/D)$.

Analogously, the coefficients of the Fourier expansion of the variance function can be estimated using the same sample, by minimizing the following function:

$$S_{\sigma^2}^2\left(a_{\sigma^2,0}, a_{\sigma^2,1}, b_{\sigma^2,1}, \cdots\right) \quad (5)$$
$$= \sum_{k=1}^{K} \left\{ [t(i_k) - \mu_T(i_k)]^2 - \frac{1}{2} a_{\sigma^2,0} - \sum_{j=1}^{n_{h,\sigma^2}} f_{\sigma^2}\left(i_k j; a_{\sigma^2,j}, b_{\sigma^2,j}\right) \right\}^2$$

where $f_{\sigma^2}\left(i_k j; a_{\sigma^2,j}, b_{\sigma^2,j}\right) = a_{\sigma^2,j} \cos(2\pi i_k j/D) + b_{\sigma^2,j} \sin(2\pi i_k j/D)$.

If the temporal span of the sample is a multiple of the period D, and the series does not have any missing days, the trigonometric interpolation theory provides the estimation of the coefficients in explicit form (Prass et al. 2012). The number of harmonics ($n_{h,\mu}$ and n_{h,σ^2}) must be estimated by assuring the absence of the periodicity of the process $Y(i)$, with respect to the criterion of the parameter parsimony. The $Y(i)$ data obtained through the deseasonalization of $t(i_k)$ are generally correlated, but the tests employed for the estimation of the number of harmonics require a random sample. Therefore, for each prefixed couple $n_{h,\mu}$ and n_{h,σ^2}, a transformed random subsample of $y(i_k)$, namely, $y(i_{k*}) = [t(i_{k*}) - \mu_T(i_{k*})] / \sigma_T(i_{k*})$ with $k_* = 1, 2, \ldots, K_*$, can be created. This subsample can be obtained by extracting values sampled every δ days (temporal span), with δ long enough to limit the stochastic dependence effect, however assuring a reliable sample length. The subsample is subdivided into M classes, assigning to each generic m^{th} class the N_m values of $y(i_{k*})$ for which it holds

$$(m-1)/M \le (i_k \bmod D)/D < m/M \quad m = 1, 2, \ldots M \quad (6)$$

Indicated as $\mu_{Y,m}$ and $\sigma_{Y,m}^2$ the mean and the variance values of each class, it is possible to test the hypotheses $H_{0,\mu}: \mu_{Y,1} = \mu_{Y,2} = \ldots = \mu_{Y,M} = \mu_Y$ and $H_{0,\sigma^2}: \sigma_{Y,1}^2 = \sigma_{Y,2}^2 = \ldots = \sigma_{Y,M}^2 = \sigma_Y^2$. The hypothesis $H_{0,\mu}$ is tested through the statistics S_V^2, approximately distributed according to a Fisher variance-ratio law $v^2(f_1, f_2)$. The hypothesis H_{0,σ^2} is tested through the Barlett's test (Snedecor and Cochran 1989) based on the statistics S_B^2 approximately distributed according to a $\chi^2(f_1)$ law. Both the procedures are detailed in Appendix 1.

2.2 Gaussianization procedure

The Gaussianization procedure (e.g., Chen and Gopinath 2000; Hólm et al. 2002; Servidio et al. 2011) is a necessary condition to respect the coherence of the linear stochastic model. In this case, the sample values $y(i_k)$ of the random variable Y have a null mean value, $m_Y = 0$, and a unit variance, $s_Y^2 = 1$, but generally show skewness ($g_{1,Y}$) and kurtosis ($g_{2,Y}$) coefficients, which significantly differ from the theoretical values expected for a Gaussian variable ($\gamma_{1,Y} = 0$ and $\gamma_{2,Y} = 3$, respectively). In this case, it is possible to transform the original variable, Y, into a standardized Gaussian variable, $Z = f(Y)$. For this purpose, the transformation functions introduced by Johnson (1949) are well suited,

$$Z = \eta + \theta \ln\left[f_Y^{(\cdot)}(y; \alpha; \beta) \right] \quad (7)$$

where $-\infty < \eta < +\infty$, $\theta > 0$, $-\infty < \alpha < +\infty$, and $\beta > 0$ are the parameters of the transformation and $f_Y^{(\cdot)}(y; \alpha, \beta)$ can take one of the following forms:

$$f_Y^{(U)}(y; \alpha, \beta) = (y-\alpha)/\beta + \sqrt{1 + (y-\alpha)^2/\beta^2} \quad -\infty < y < +\infty \quad (8)$$

$$f_Y^{(L)}(y; \alpha, \beta) = (y-\alpha)/\beta \quad y > \alpha \quad (9)$$

$$f_Y^{(B)}(y; \alpha, \beta) = (y-\alpha)/(\alpha + \beta - y) \quad \alpha < y < \alpha + \beta \quad (10)$$

Specifically, Eqs. (8) and (10) are known as unbounded and bounded Johnson transformations, respectively, while Eq. (9) implies that the random variable y is distributed according to a log-normal law with three parameters. The choice of the function $f_Y^{(\cdot)}(y; \alpha, \beta)$ to be adopted depends on the sample values of the skewness coefficient, $g_{1,Y}$, and the kurtosis coefficient, $g_{2,Y}$. In fact, given that

$$R_1 = \sqrt[3]{1 + g_{1,Y}^2/2 + \sqrt{g_{1,Y}^2\left(1 + g_{1,Y}^2/4\right)}} \quad (11)$$

$$R_2 = \sqrt[3]{1 + g_{1,Y}^2/2 - \sqrt{g_{1,Y}^2\left(1 + g_{1,Y}^2/4\right)}} \quad (12)$$

$$G_2 = 6\left(R_1^2 + R_2^2\right) + \left(g_{1,Y}^2 - 8\right)(R_1 + R_2) - 2g_{1,Y}^2 + 7 \quad (13)$$
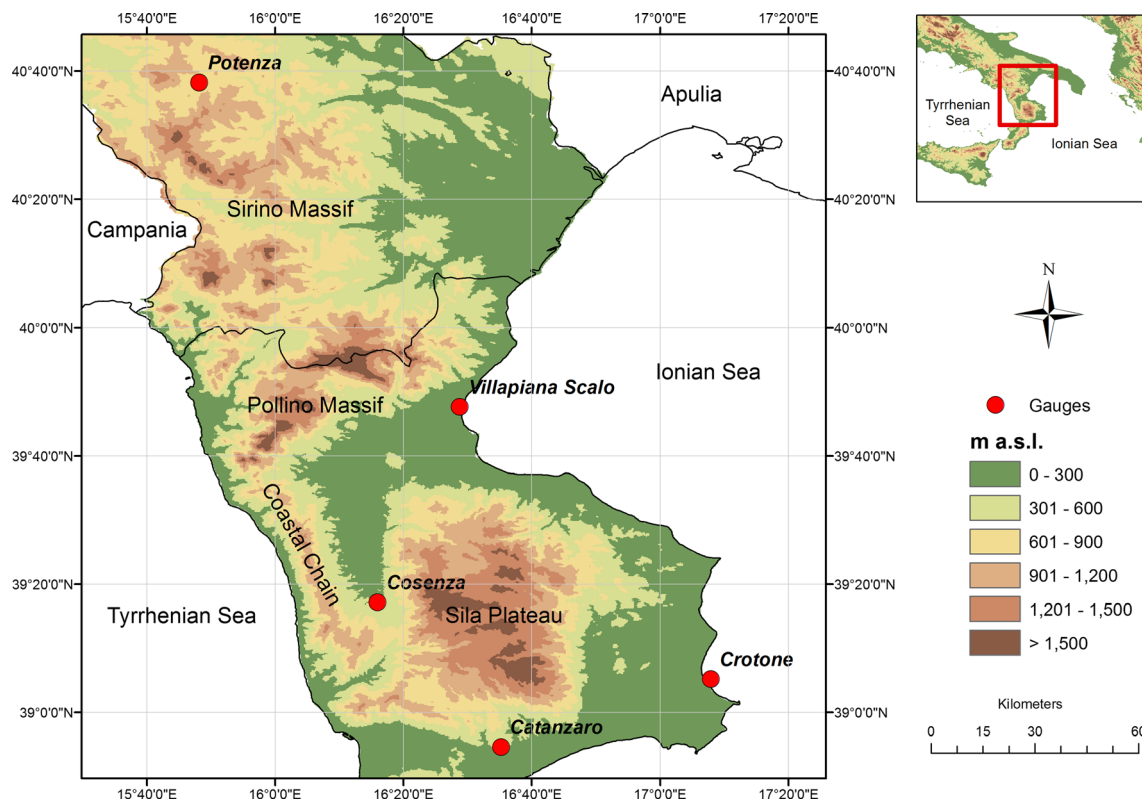
Fig. 1 Localization of the temperature gauges on a DEM

Equation (8) has to be used if $g_{2,Y} > G_2$, Eq. (10) if $g_{2,Y} < G_2$, while Eq. (9) concerns only the special case $g_{2,Y} = G_2$.

The different techniques used for the estimation of the transformation parameters are based on the method of moments. In the proposed model, only the parameters η and θ have to be effectively estimated, since α and β are linked to the former through analytical expressions, being $\mu_Y = 0$ and $\sigma_Y^2 = 1$.

If the unbounded Johnson transformation must be adopted, in order to estimate the parameters η and θ, the following equation in the variable $\omega = \exp(0/\theta^2)$ has to be numerically resolved (Tuenter 2001):

$$[\omega - \omega_*][2\omega + \omega_* + 3]^2 - 4g_{1,Y}^2 = 0 \qquad (14)$$

where

$$\omega_* = 2\left[\sqrt{1 + \frac{1}{2}\left(\omega^2 - \frac{g_{2,Y} + 3}{\omega^2 + 2\omega + 3}\right)} - 1\right] \qquad (15)$$

Otherwise, if the bounded Johnson transformation must be applied, the parameter estimation is less simple. In fact, for the estimation of the parameters η and θ, the following non-linear system should be numerically resolved:

$$\begin{cases} \gamma_{1,Y}(\eta, \theta) = g_{1,Y} \\ \gamma_{2,Y}(\eta, \theta) = g_{2,Y} \end{cases} \qquad (16)$$

in which also the values of the functions $\gamma_{1,Y}(\eta, \theta)$ and $\gamma_{2,Y}(\eta, \theta)$ should be numerically evaluated.

Table 1 Gauges used for the application

Gauge	Latitude	Longitude	Altitude (masl)	Number of data	Missing values (%)
Cosenza	39° 17′ 11″	16° 15′ 55″	242	20,694	5.6
Catanzaro	38° 54′ 39″	16° 35′ 09″	334	21,062	3.9
Crotone	39° 05′ 13″	17° 07′ 52″	5	20,433	6.8
Potenza	40° 38′ 12″	15° 48′ 05″	811	19,937	9.0
Villapiana Scalo	39° 47′ 42″	17° 29′ 55″	5	20,344	7.2

Table 2 Identification of the number of harmonics for the Cosenza gauge

$n_{h,\mu}$	S_V^2	$v_{0.95}^2$	$H_{0,\mu}$	n_{h,σ^2}	S_B^2	$\chi_{0.95}^2$	H_{0,σ^2}
1	5.72	1.79	Rejected	1	20.6	19.7	Rejected
1	5.86	1.79	Rejected	2	8.1	19.7	Not rejected
2	0.73	1.79	Not rejected	1	21.3	19.7	Rejected
2	0.72	1.79	Not rejected	2	8.8	19.7	Not rejected

2.3 Correlation structure

The sample data series $z(i_k)$ of the random variable Z, obtained from the Johnson transformations applied to the sample $y(i_k)$, usually shows a correlation structure characterized by a marked persistence, with values of autocorrelation coefficients slowly decreasing for growing lags. Assuming that the zero mean process $z(i)$ can be described by a FARIMA (p,d,q) model, the following relationship holds:

$$\Phi_p(B)(1-B)^d z(i) = \Psi_q(B)\varepsilon(i) \tag{17}$$

where B is the backward operator, $\Phi_p(B)$ is the p-order polynomial of the autoregressive component, $\Psi_q(B)$ is the q-order polynomial of the mean average component, $\varepsilon(i)$ is a sequence of i.i.d. random variables with mean zero, and d is the fractional order of differentiation. In other terms, the FARIMA (p,d,q) model can be considered as a composition between a fractional filter d and an ARMA (p,q) process

$$(1-B)^d z(i) = u(i) \tag{18}$$

$$\Phi_p(B)u(i) = \Psi_q(B)\varepsilon(i) \tag{19}$$

Equation (19) shows that the intermediate process $u(i)$ is an ARMA (p,q) process, which for $\Psi_0 = 1$ is

$$u(i) = \sum_{k_p=1}^{p} \varphi_{k_p} u(i-k_p) + \sum_{k_q=0}^{q} \psi_{k_q} \varepsilon(i-k_q) \tag{20}$$

By employing the series expansion of $(1-B)^d$, Eq. (18) can be described as

$$u(i) = \Gamma(d+1) \sum_{s=0}^{\infty} \frac{(-1)^s}{\Gamma(d-s+1)s!} z(i-s) \tag{21}$$

It is well known that the maximum likelihood estimation of the variance-covariance matrix of a multivariate normal distribution is the sample variance-covariance matrix (Anderson and Olkin 1985). By using a FARIMA model, the correlation structure depends on the model parameters. Thus, through a weighted least squares method, we find the best fitting of the maximum likelihood estimation of the variance-covariance matrix.

The estimation of the parameters d, φ_{k_p}, and ψ_{k_q} can be obtained following a trial-and-error procedure. Through Eq. (21), for each assigned value of the parameter d, it is possible to transform the sample $z(i_k)$ in a sample $u(i_k)$, which can be considered as a realization of the ARMA (p,q) process. Thus, once the sample autocorrelation values $r_{U,\lambda}(d)$ of $u(i_k)$ for $\lambda = 1, \ldots, p+q$ are evaluated, the estimated values $\hat{\varphi}_1(d), ..., \hat{\varphi}_p(d)$, $\hat{\psi}_1(d), ..., \hat{\psi}_q(d)$ can be obtained, by considering

$$\rho_{U,\lambda}\left[\hat{\varphi}_1(d), ..., \hat{\varphi}_p(d), \hat{\psi}_1(d), ..., \hat{\psi}_q(d)\right] = r_{U,\lambda}(d) \quad \lambda = 1, ..., p+q \tag{22}$$

where $\rho_{U,\lambda}$ are the theoretical autocorrelation values of the ARMA (p,q) process.

Table 3 Coefficients of the truncated Fourier expansions for the mean and the variance functions

Gauge	$\hat{n}_{h,\mu}$	$\hat{a}_{\mu,0}/2$	$\hat{a}_{\mu,1}$	$\hat{b}_{\mu,1}$	$\hat{a}_{\mu,2}$	$\hat{b}_{\mu,2}$	$\hat{n}_{h,\sigma^2}$	$\hat{a}_{\sigma^2,0}/2$	$\hat{a}_{\sigma^2,1}$	$\hat{b}_{\sigma^2,1}$	$\hat{a}_{\sigma^2,2}$	$\hat{b}_{\sigma^2,2}$
Cosenza	2	21.61	−9.34	−4.29	0.11	1.07	2	16.10	−0.82	2.81	−1.79	−1.17
Catanzaro	2	19.67	−7.70	−4.37	0.18	0.81	1	8.97	−1.62	1.67	–	–
Crotone	2	21.56	−8.20	−4.07	0.38	0.83	1	10.02	−0.78	−0.23	–	–
Potenza	2	16.92	−9.58	−4.23	0.16	0.84	2	16.53	0.09	3.14	−2.37	−0.27
Villapiana S.	2	21.82	−8.12	−4.13	0.41	0.98	1	8.83	−0.22	0.67	–	–

Fig. 2 Cosenza gauge: **a** comparison between the observed $m_T(i)$ and the estimated $\mu_T(i)$ functions of the mean, averaged for each day of the year, and **b** comparison between the sampling $s_T(i)$ and the estimated $\sigma_T(i)$ functions of the standard deviation, averaged for each day of the year
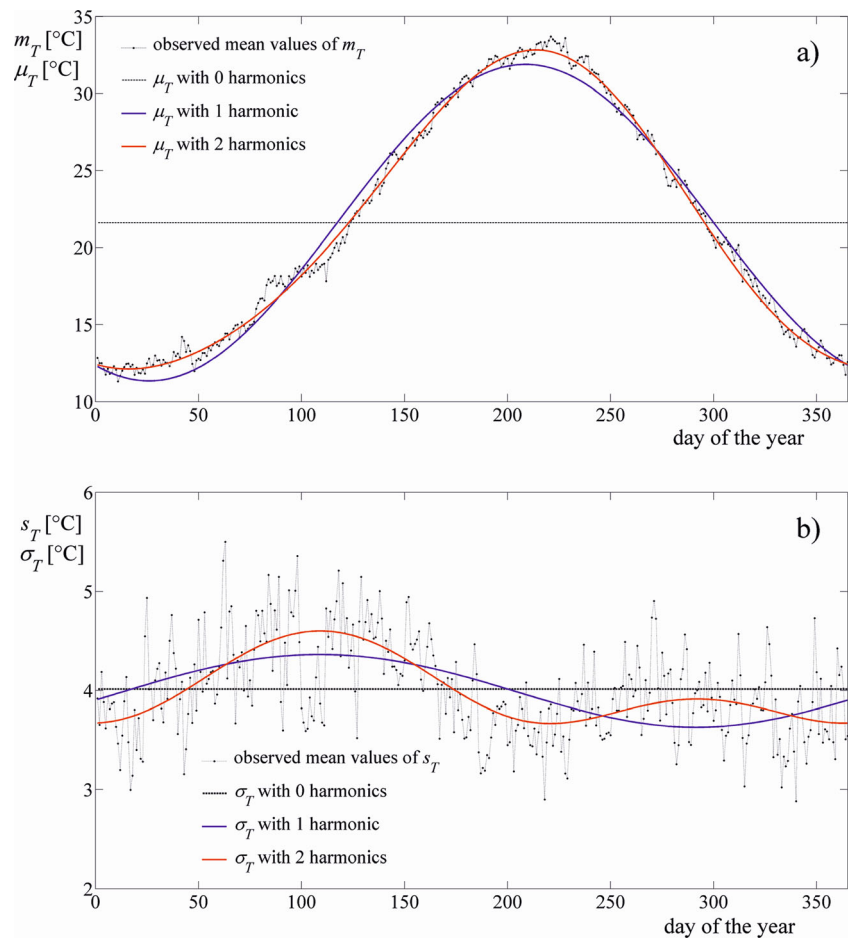
The actual value $\hat{d}$ of d can be estimated by minimizing the weighted mean square deviation

$$S_\rho^2 = \sum_{l=1}^{N_r} \omega_l \left\{ r_{Z,l} - \rho_{Z,l} \left[\hat{\varphi}_1(d), ..., \hat{\varphi}_p(d), \hat{\psi}_1(d), ..., \hat{\psi}_q(d), d \right] \right\}^2 \quad (23)$$

where $r_{Z,l}$ are the sample correlogram values derived by the sample $z(i_k)$; $\rho_{Z,l}$ are the theoretical autocorrelation values of the FARIMA (p,d,q) model; N_r is the maximum lag to which the calculation for the deviation can be extended; and ω_l, with $l = 1, \ldots, N_r$ and $\sum_{l=1}^{N_r} \omega_l = 1$, are the weighting coefficients which allow to distribute the quality of the data fit at varying of the lag l.

The estimated values $\hat{p}$ and $\hat{q}$ must be fixed by respecting the principle of parametric parsimony (Box and Jenkins 1976). Akaike (1974) suggests a mathematical formulation of the parsimony criterion of model building, known as Akaike information criterion (AIC), for the purpose of selecting an optimal model fits to a given data. In this work, the Akaike information criterion in the correct form (AICc) has been used (Sugiura 1978; Burnham and Anderson 2002). The model showing the minimum value of the AICc can be considered as the one which better matches both the fitting to

the observed data and the principle of parametric parsimony. Before applying the AIC criterion, it has been verified that the residuals were white noise through the application of the Anderson-Darling test (Anderson and Darling 1952).

3 Application

3.1 Database

The proposed stochastic model has been applied to the maximum daily temperatures, $T(i) = T_{\max}(i)$, of the Cosenza gauge and verified on other four southern Italian gauges, managed by the Centro Funzionale Multirischi of the Calabria region (Fig. 1). The gauges are located in an area characterized by high climatic variability due to its geographic location and mountainous nature (Coscarelli and Caloiero 2012). Summers are typically dry, denoting a subtropical Mediterranean climate. The coastal areas present mild winters and hot summers with little precipitation. The Ionian side, influenced by air masses coming from Africa, records high temperatures and intense precipitation; on the Tyrrhenian side,

Fig. 3 Catanzaro gauge: **a** comparison between the observed $m_T(i)$ and the estimated $\mu_T(i)$ functions of the mean, averaged for each day of the year, and **b** comparison between the sampling $s_T(i)$ and the estimated $\sigma_T(i)$ functions of the standard deviation, averaged for each day of the year
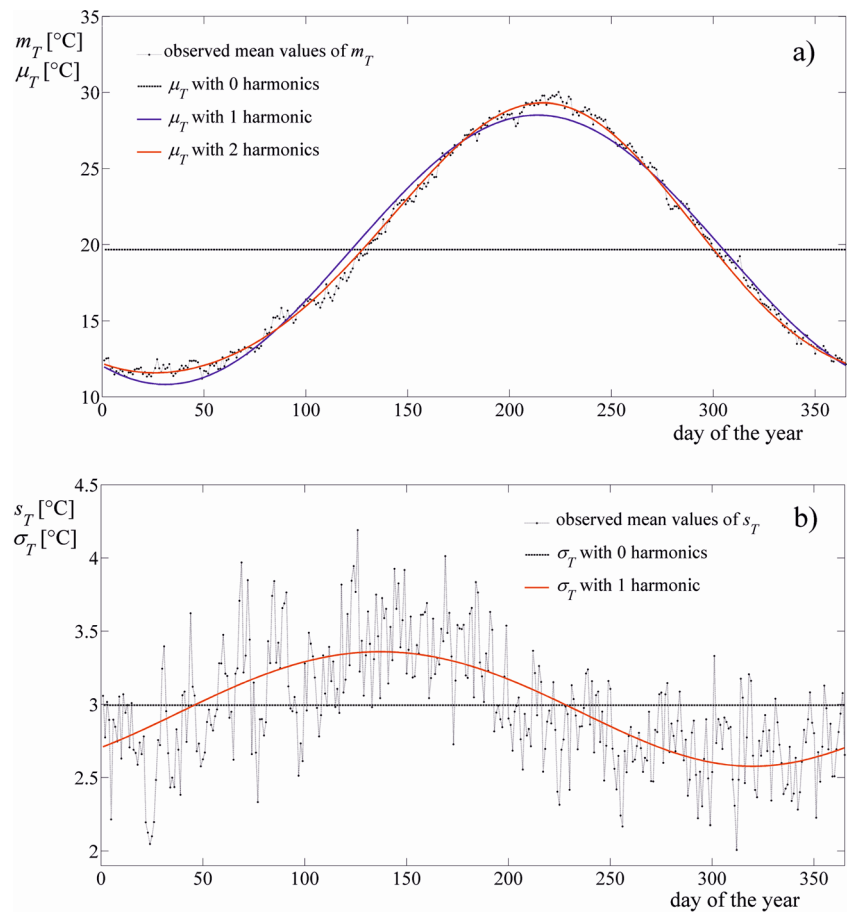
influenced by western air currents, milder temperatures and considerable orographic precipitation are observed. The internal areas are characterized by colder winters, sometimes snowy, and fresher summers with some precipitation (Ferrari et al. 2013; Buttafuoco et al. 2015; Caloiero et al. 2015a).

The main features of the selected gauges are presented in Table 1. The gauges are located at different altitudes, namely, from a few meters to about 800 m above sea level (a.s.l.).

Moreover, considering an observation period spanning from 1951 to 2010, the temperature gauges have less than 9 % missing days. Previous regional studies on the temperature data series of Calabria region (Caloiero et al. 2015b, c) did not show significant trends in maximum temperature data for the gauges used in this study. As a further analysis, in this study, the Mann-Kendall test (Mann 1945; Kendall 1962) has been applied to investigate the existence of monotonic

Fig. 4 Comparison among the autocorrelation values of the observed and the deseasonalized data (one and two harmonics) for a lag range centered to 365 days
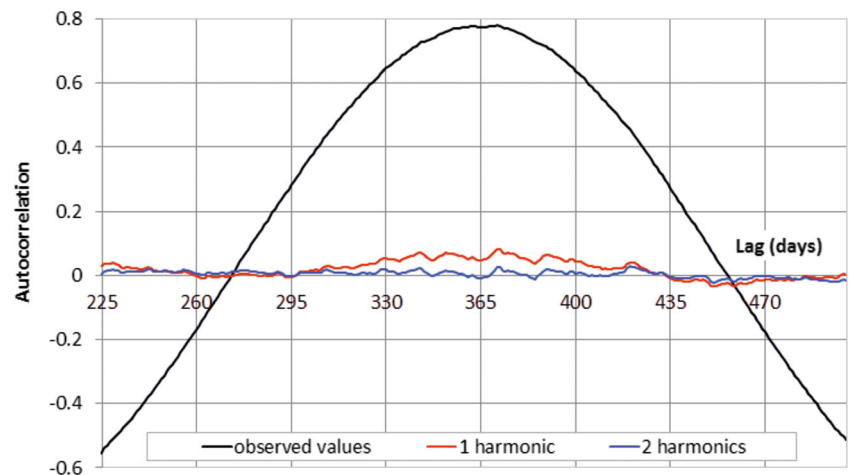
Table 4 Parameter estimation of the Johnson transformation

| Gauge	$g_{1,Y}$	$g_{2,Y}$	Distribution type	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\eta}$	$\hat{\theta}$
Cosenza	0.086	3.177	Unbounded	−0.701	4.895	−0.706	5.042
Catanzaro	0.176	3.596	Unbounded	−0.451	2.738	−0.455	2.941
Crotone	−0.067	3.293	Unbounded	0.324	3.783	0.325	3.923
Potenza	−0.128	2.864	Bounded	−7.806	13.57	−1.006	3.244
Villapiana Scalo	−0.060	3.470	Unbounded	0.182	3.007	0.183	3.168

trend in maximum daily temperature, as suggested also by Prass et al. (2012). Results confirmed that no statistical significant trends exist for the selected gauges.

3.2 Parameter estimation

The absence of trends in the data series considered in this study allows the use of Eqs. (2) and (3). Since the series of the maximum daily temperature, $t_{max}(i_k)$, show some missing data, the estimation of the coefficients of the truncated Fourier expansion referred to the mean, $\mu_T(i)$, and to the variance functions, $\sigma_T^2(i)$, has been performed by minimizing Eqs. (4) and (5).

With reference to the Cosenza gauge, Table 2 shows the results of the tests used for the identification of the minimum number of harmonics of the Fourier expansion, for which both the hypotheses H_{0,σ^2} and $H_{0,\mu}$ cannot be rejected (significant level = 0.05). The tests employed data sampled every 8 days ($\delta = 8$) and indicated that two harmonics are required for both the mean and the variance functions in order to remove the periodicity in the first- and second-order statistics of the observed series. The results of the same tests applied also to the other gauges confirm that two harmonics are needed for all the stations to remove the periodicity in the mean function (Table 3). Moreover, differently from the Cosenza gauge, for three series, the periodicity of the variance function can be

Fig. 5 Comparison between observed values and theoretical cumulative distribution function of the Z variable for two gauges (Cosenza and Villapiana Scalo)
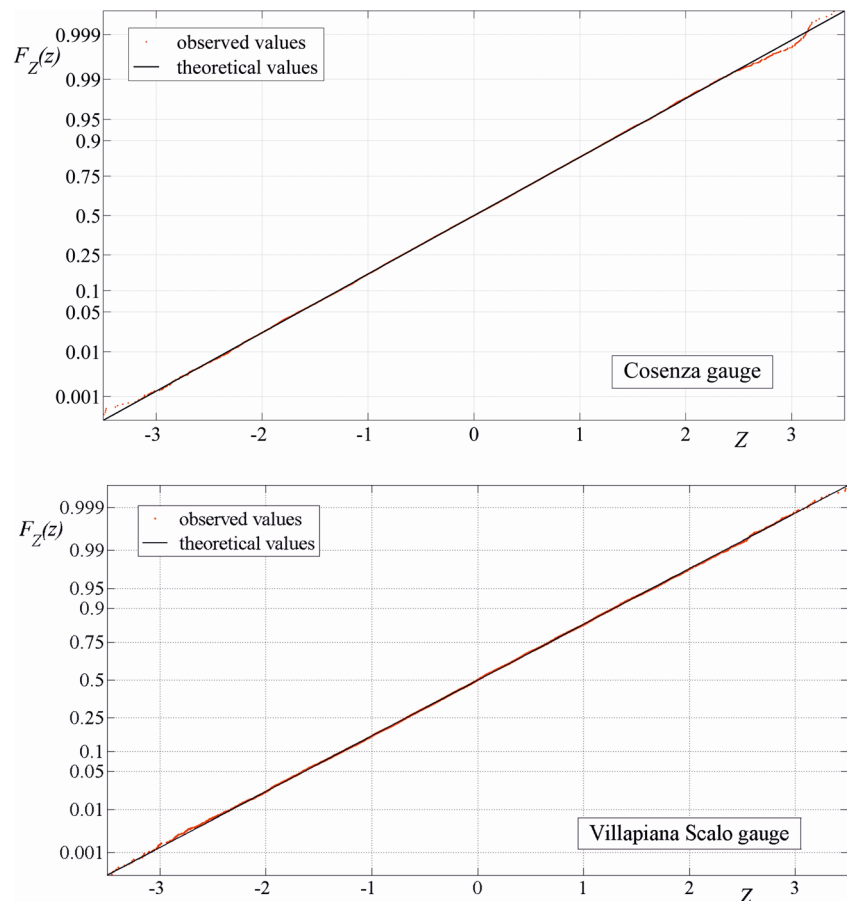
Table 5 Identification of the FARIMA model for the Cosenza gauge

FARIMA (p,d,q)	ν	$\hat{d}$	$\hat{\varphi}_1$	$\hat{\psi}_1$	$\hat{\varphi}_2$	$\hat{\psi}_2$	S_ρ^2	AICc
$(0,d,0)$	1	0.356	–	–	–	–	0.00798	−239.4
$(0,d,1)$	2	0.295	–	0.488	–	–	0.00141	−323.9
$(1,d,0)$	2	0.244	0.446	–	–	–	0.000344	−394.5
$(0,d,2)$	3	0.269	–	0.427	–	0.206	0.000465	−377.2
$(2,d,0)$	3	0.245	0.448	–	−0.0076	–	0.000341	−392.6
$(1,d,1)$	3	0.244	0.432	0.042	–	–	0.000338	−393.2

removed with only one harmonic. In Table 3, for each station, the estimated values of the coefficients of the truncated Fourier expansion for the mean and the variance functions are also reported. Figure 2 shows, for the Cosenza gauge, the comparisons between sample and estimated values of the mean and the standard deviation functions, respectively. In particular, the estimated values have been obtained for zero, one, and two harmonics for each day of the year. The same comparisons are shown for the Catanzaro gauge in Fig. 3. In Fig. 4, the comparison among the ACFs of the observed and the deseasonalized values (for a lag range centered to 365 days) is shown.

The Gaussianization procedure applied to the deseasonalized data series of Cosenza, $y(i_k) = [t_{\max}(i_k) - \hat{\mu}_T(i_k)] / \hat{\sigma}_T(i_k)$, was performed through the unbounded version of the Johnson transformations. Operatively, the parameters η and θ of the Johnson transformation have been estimated by numerically solving Eq. (14), thus allowing also the estimation of the derived parameters α and β (Table 4). The application to the other gauges evidenced that only the data series of the Potenza gauge has been transformed into a Gaussian process by means of the bounded function, thus requiring the numerical solution of the non-linear system (Eq. (16)). For all the gauges, Table 4 presents the sample values of the skewness and the kurtosis coefficients of the $y(i_k)$ series ($g_{1,Y}, g_{2,Y}$) and the set of the estimated values of the Johnson transformation parameters $(\hat{\alpha}, \hat{\beta}, \hat{\eta}, \hat{\theta})$.

The achievement of the Gaussian feature of the series $z(i_k)$ is evidenced by the comparison of the sampling values with the theoretical values of the standard normal variable on probabilistic plot, as shown for the Cosenza and the Villapiana gauges in Fig. 5.

The identification of the FARIMA (p,d,q) process aimed at describing the correlation structure of the Gaussian series. Specifically, the procedure, based on the AICc index, identified the orders $\hat{p}$ and $\hat{q}$, which better fitted the observed data and preserved the parametrical parsimony criterion. The results of this procedure for the Cosenza gauge are presented in Table 5. Namely, for increasing values of the orders p and q, the differential fractional order, d, was identified by minimizing the weighted mean square function, S_ρ^2 (Eq. (23)). In this equation, a maximum lag $N_r = 50$ has been fixed, and the weighting coefficients, $\omega_j = \omega_j^*/\sum \omega_j^*$, $j = 1, 2, \ldots, N_r$ with $\omega_j^* = 1 - [(j-1)/N_r]^c$ and $c = 1/2$, have been adopted. The final results for all the gauges are summarized in Table 6. Globally, for three stations, a FARIMA $(1, \hat{d}, 0)$ model was identified, while for the other two stations, a FARIMA $(1, \hat{d}, 1)$ model was detected. The comparisons between sampling and theoretical correlograms evidenced the ability of the proposed FARIMA model to reproduce the long-term memory (Fig. 6).
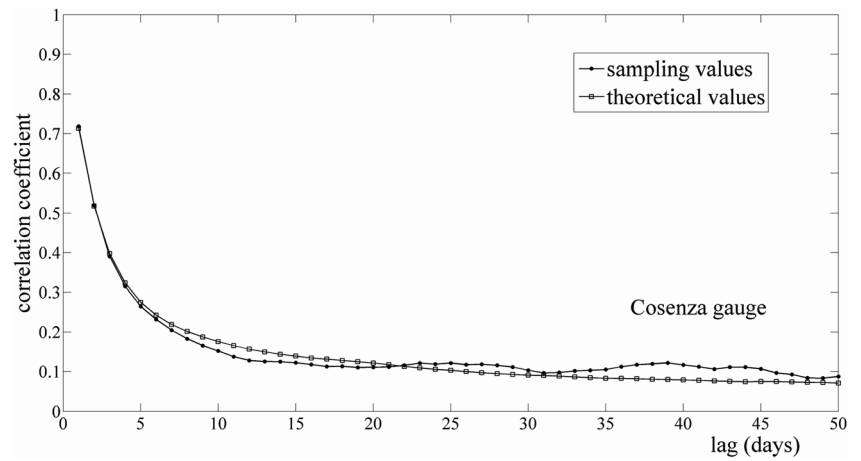
3.3 Analysis of maximum daily temperatures

The proposed model, through the application of a Monte Carlo simulation procedure (Appendix 2), can be useful for assessing various features of the temperature database. Specifically, the synthetic world simulation concerned 10^5 years, corresponding to a total generation of about $L_s = 4 \times 10^7$ values.

Table 6 Identification of the FARIMA models for the all gauges used in the application

| Gauge	FARIMA (p,d,q)	ν	$\hat{d}$	$\hat{\varphi}_1$	$\hat{\psi}_1$	AICc
Cosenza	$(1,d,0)$	2	0.244	0.446	–	−394.5
Catanzaro	$(1,d,0)$	2	0.223	0.430	–	−407.7
Crotone	$(1,d,1)$	3	0.342	0.265	0.042	−332.4
Potenza	$(1,d,1)$	3	0.113	0.586	0.124	−490.7
Villapiana Scalo	$(1,d,0)$	2	0.236	0.417	–	−424.8

**Fig. 6** Comparison between sampling and theoretical autocorrelations values for Cosenza gauge

In this study, first, the proposed model has permitted the estimation of the probability $F_{W(\tau)}[w(\tau)] = P[W(\tau) \leq w(\tau)]$ of the annual maximum of the sequences of days, $W(\tau)$, with maximum daily temperature over a threshold, τ. Figure 7 shows, on a probabilistic Gumbel graph, the return periods $T_{W(\tau)}$ of the random variable $W(\tau)$, defined as $T_{W(\tau)} = 1/\{1 - F_{W(\tau)}[w(\tau)]\}$, corresponding to different threshold values

for the Cosenza station (altitude 242 m a.s.l.). In particular, for increasing threshold values, high return periods can be reached also for short-day sequences. As an example, considering the threshold $\tau = 40°$, the values of $W(\tau)$ shift from 2 to 5 days for return periods ranging from 10 to 100 years. Differently, for higher-elevation gauges, such as Potenza (811 m a.s.l.), the same range of values is

Fig. 7 Return periods $T_{W(\tau)}$ of the variable $W(\tau)$ for different values of the threshold τ for the Cosenza and Potenza gauges (*red* linear tendency lines)
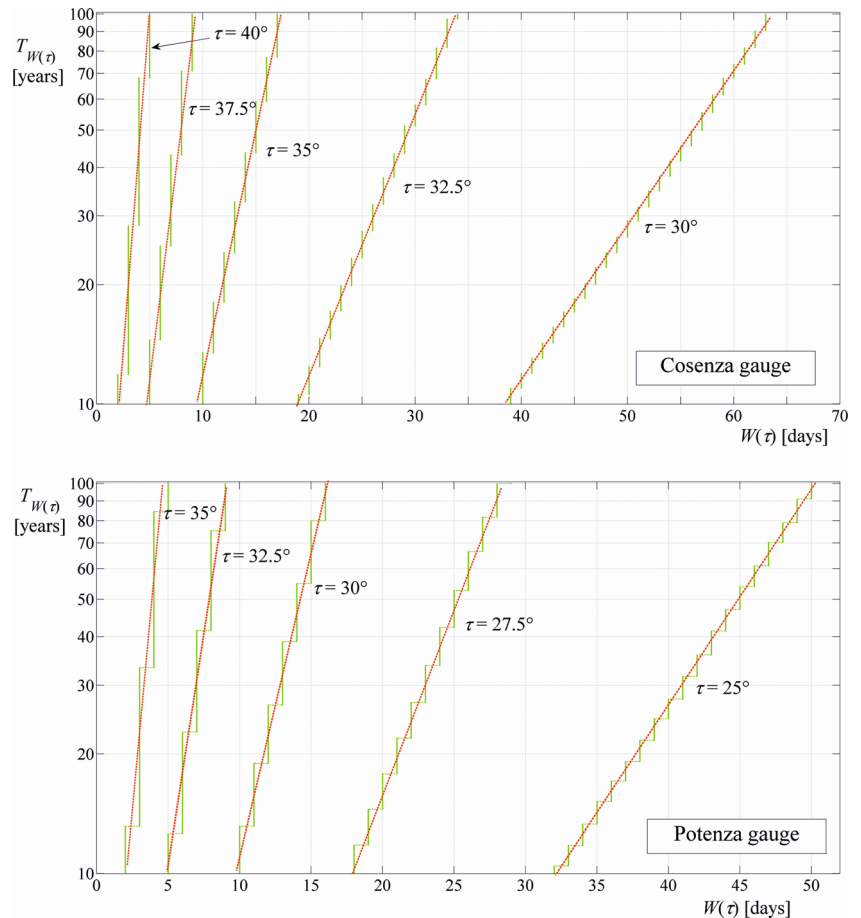
Fig. 8 Probability values $p_{IW(\tau)}$ of the temporal occurrence in the year $I_{W(\tau)}$ for different threshold values of temperature for the Cosenza gauge
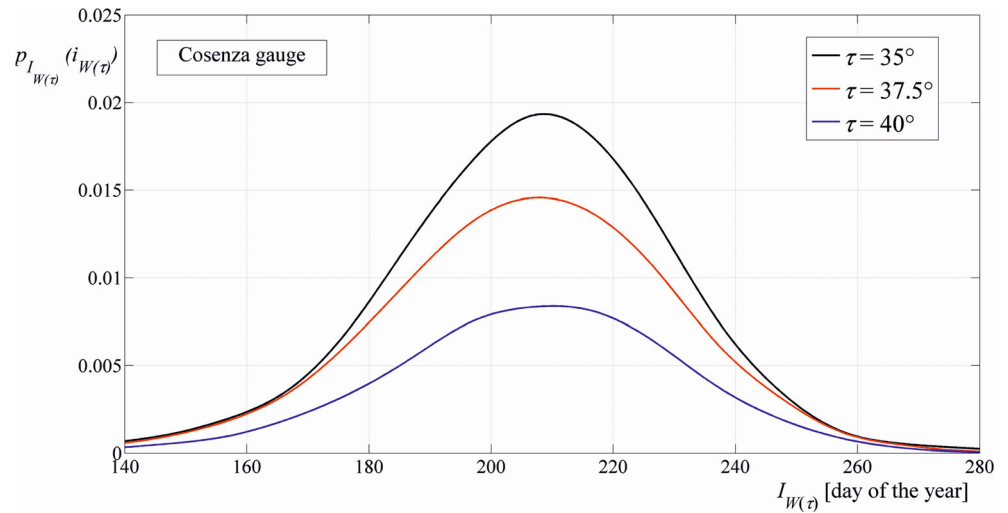

observed for a lower threshold value of temperature (about 35°). This behavior can be clearly connected to the influence of the gauge altitude on the variable $W(\tau)$.

As a further application, the probability values $p_{I_{W(\tau)}}\left(i_{W(\tau)}\right) = P\left[I_{W(\tau)} = i_{W(\tau)}\right]$ that such exceedances start in a specified

day, i_w, were also evaluated. The probability values $p_{I_{W(\tau)}}\left(i_{W(\tau)}\right)$ of the yearly temporal occurrence $I_{W(\tau)}$ of $W(\tau)$ are shown in Fig. 8 for various threshold values always for the Cosenza gauge. The maximum probability values have been detected in summer around the 210th day of the year.

Fig. 9 Return periods $T_{W(\kappa)}$ of the variable $W(\kappa)$ for different values of κ for the Cosenza and Crotone gauges (red linear tendency lines)
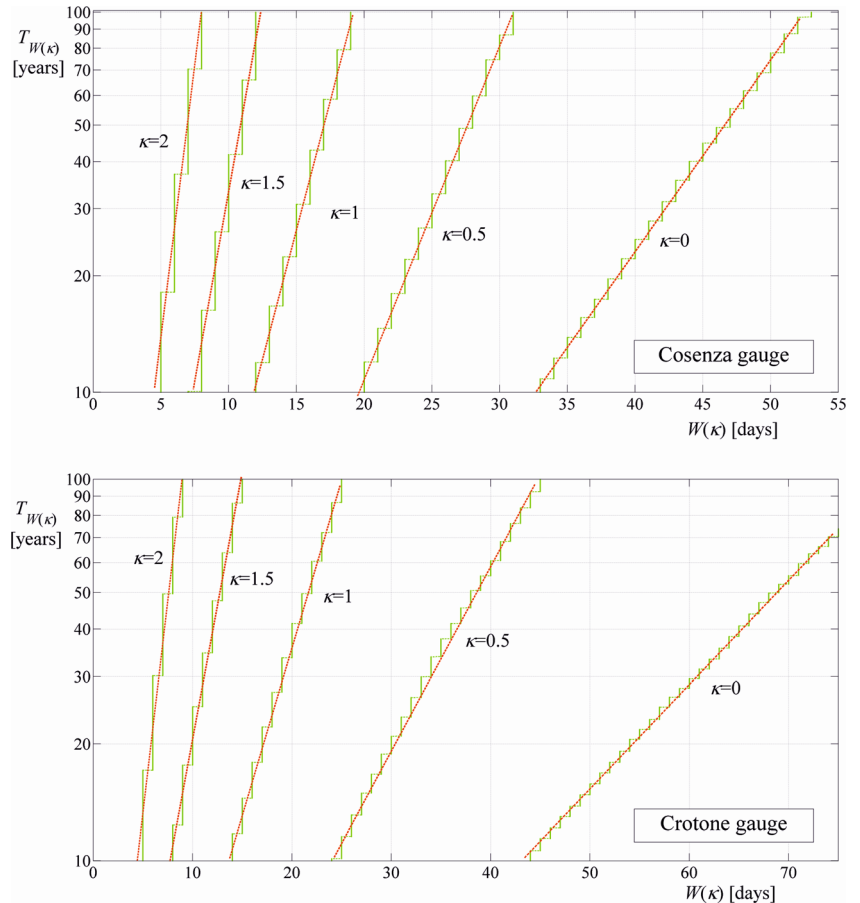
Finally, the probability values $F_{W(\kappa)}[w(\kappa)] = P[W(\kappa) \leq w(\kappa)]$, associated to the annual maximum of sequences of days $W(\kappa)$, with maximum daily temperature greater their expected value $\mu_T(i)$ plus κ-times the standard deviation, have been estimated. The return periods $T_{W(\kappa)}$ of the variable $W(\kappa)$, for different values of parameter κ, are presented in Fig. 9 for the Cosenza and the Crotone gauges. As a result, for fixed values of the return period and the parameter κ, the annual maxima values $W(\kappa)$ were greater for the Crotone gauge than for the Cosenza gauge. In particular, this behavior was more evident for the lower threshold values. For example, the $W(\kappa)$ values obtained for $\kappa = 0$, which corresponds to a return period $T_{W(\kappa)} = 100$ years, were 53 days for Cosenza and 80 days for Crotone, respectively.

4 Conclusion

In this paper, a stochastic model developed to simulate daily maximum temperature series, coherently with observed time series, is proposed. The model was based on three different steps. The first step was data deseasonalization, obtained by means of a truncated Fourier series expansion. Subsequently, a normalization technique was performed through the Johnson transformation. Finally, a FARIMA model was applied for the analysis of the correlation structure of the normalized data, characterized by a marked persistence. The procedure has been first applied to the Cosenza gauge and then tested to a set of maximum temperature series registered in four gauges located in southern Italy. The model satisfactorily reproduced the long-term memory of the temperature series, also allowing the parametric parsimony criterion.

Moreover, through the application of a Monte Carlo simulation procedure, the proposed model allowed the evaluation of various features of the temperature database. First, the empirical probability distribution of the annual maximum of the sequences of days, with maximum daily temperature over fixed thresholds, has been obtained. Results showed high return periods also for short-day sequences at increasing threshold values. Successively, the probability values that such sequences can start in a specified day have been also evaluated, showing that the highest occurrence probabilities fall in summer periods. Finally, the return period values associated to annual maximum of sequences of days, characterized by maximum daily temperature greater than their expected value plus κ-times the standard deviation, have been obtained.

An important criterion for stochastic modeling is the reproducibility of the statistical characteristics of observed data (Lee 2015). Effectively, the proposed model allowed

the prediction of the statistical properties of temperatures, with few data required as input. Moreover, the stochastic model, not depending on station altitude and climatic zone, had the advantage of being applicable in a certain area, aiming to estimate occurrence probabilities and return periods associated with high-temperature events at any day of the year and/or at any gauge. This ability to extrapolate findings is particularly important when seeking to determine the risks of extremely rare events. For these reasons, the model is an attractive tool for management decision-making, and its basic structure can easily be applied to larger areas with spatially differentiated data.

Appendix 1: Estimation of the number of harmonics

The hypothesis, $H_{0,\mu}$: $\mu_{Y,1} = \mu_{Y,2} = \ldots = \mu_{Y,M} = \mu_Y$, can be verified by using the statistics

$$S_V^2 = s''^2_Y / s'^2_Y \tag{A1}$$

with

$$s'^2_Y = \frac{1}{K_* - M} \sum_{m=1}^{M} (N_m - 1) s_{Y,m}^2 \tag{A2}$$

$$s''^2_Y = \frac{1}{M-1} \sum_{m=1}^{M} N_m (m_{Y,m} - m_Y)^2 \tag{A3}$$

where $s_{Y,m}^2$ is the sample variance of the data referred to the mth class, $m_{Y,m}$ is the sample mean of the data referred to the mth class, and m_Y is the mean of the whole sample $y(i_{k_*})$. The statistics S_V^2 is approximately distributed according to a Fisher variance-ratio law $\nu^2(f_1, f_2)$ with $f_1 = M - 1$ and $f_2 = K_* - M$ degrees of freedom. For a significance level α_{SL}, the null hypothesis cannot be refused if $S_V^2 < \nu^2_{1-\alpha_{\text{SL}}}(f_1, f_2)$, where $\nu^2_{1-\alpha_{\text{SL}}}(f_1, f_2)$ is the $1 - \alpha_{\text{SL}}$ percentile of the ν^2 distribution.

The hypothesis H_{0,σ^2} : $\sigma_{Y,1}^2 = \sigma_{Y,2}^2 = \ldots = \sigma_{Y,M}^2 = \sigma_Y^2$ can be verified through the Bartlett's test (Snedecor and Cochran 1989), based on the statistics

$$S_B^2 = \frac{1}{c_B} \left[(K_* - M) \ln s'^2_Y - \sum_{m=1}^{M} (N_m - 1) \ln s_{Y,m}^2 \right] \tag{A4}$$

where

$$c_B = 1 - \frac{1}{3(M-1)} \left(\frac{1}{K_* - M} - \sum_{m=1}^{M} \frac{1}{N_m - 1} \right) \quad \text{(A5)}$$

The statistics S_B^2 is approximately distributed according to a $\chi^2(f_1)$ law, with $f_1 = M - 1$ degree of freedom. With a significance level equal to α_{SL}, the hypothesis cannot be rejected if $S_B^2 < \chi^2_{1-\alpha_{SL}}(f_1)$, where $\chi^2_{1-\alpha_{SL}}(f_1)$ is the $1 - \alpha_{SL}$ percentile of the χ^2 distribution.

The smallest values of $n_{h,\mu}$ and n_{h,σ^2}, for which both the hypotheses $H_{0,\mu}$ and H_{0,σ^2} cannot be rejected, detect the number of harmonics $\hat{n}_{h,\mu}$ and $\hat{n}_{h,\sigma^2}$ to be used in the truncated Fourier expansion for the functions $\mu_T(i)$ and $\sigma_T^2(i)$.

Appendix 2: Monte Carlo procedure

The Monte Carlo simulation procedure, used in this work to generate the daily maximum temperature $t_{\max}(i)$ series, can be schematized as follows:

1. By using L'Ecuyer random generator, a sequence $\upsilon(i)$, with $i = 1, 2, \ldots, L_s$, of random number uniformly distributed on the interval (0,1) is created.
2. The sequence $\upsilon(i)$ is transformed into a sequence $\varepsilon(i)$ of random numbers, distributed according to a normal law, with zero mean and variance σ_ε^2, through the Box and Müller technique.
3. According to an ARMA $(\hat{p}, \hat{q})$ model, initialized with $u(i) = 0$ and $\varepsilon(i) = 0$ for $i \leq 0$, a sequence of numbers is generated,

$$u(i) = \sum_{k_p=1}^{\hat{p}} \hat{\varphi}_{k_p} u(i - k_p) + \sum_{k_q=0}^{\hat{q}} \hat{\psi}_{k_q} \varepsilon(i - k_q)$$
$$i = 1, 2, \ldots, L_s$$
$$\text{(B1)}$$

4. By using the series development of the $(1-B)^{-\hat{d}}$ operator, the sequence $u(i)$ is transformed into a number sequence $z(i)$ corresponding to the FARIMA $(\hat{p}, \hat{d}, \hat{q})$ model with zero mean and unit variance

$$z(i) = \frac{1}{\Gamma(\hat{d})} \sum_{s=0}^{s_{\max}} \frac{\Gamma(\hat{d}+s)}{s!} u(i-s) \quad i = s_{\max} + 1, \ldots, L_s \quad \text{(B2)}$$

where

$s_{\max}$ is fixed so that $\Gamma(\hat{d} + s_{\max} + 1)/(s_{\max} + 1)!$ $< \xi \sum_{s=0}^{s_{\max}} \Gamma(d+s)/s!$, with $\xi = 10^{-4}$.

The value of the variance σ_ε^2 is fixed in order to obtain $\sigma_Z^2 = 1$. If a FARIMA $(1, \hat{d}, 0)$ model is employed, the value for σ_ε^2 is

$$\sigma_\varepsilon^2 = \frac{\Gamma^2(1-\hat{d})}{\Gamma(1-2\hat{d})} \cdot \frac{1 + \hat{\varphi}_1}{{}_2F_1(1, 1+\hat{d}, 1-\hat{d}; \hat{\varphi}_1)} \quad \text{(B3)}$$

where $\Gamma(.)$ and ${}_2F_1(.)$ indicate the complete gamma function and the hypergeometric function, respectively.

5. The sequence $z(i) = s_{\max} + 1, \ldots, L_s$ is transformed into the sequence $y(i) = s_{\max} + 1, \ldots, L_s$, by using the inverse function of the unbounded Johnson transformation

$$y(i) = \hat{\alpha} + \hat{\beta} \sinh\left[\frac{z(i) - \hat{\eta}}{\hat{\theta}}\right] \quad i = s_{\max} + 1, \ldots, L_s \quad \text{(B4)}$$

or the inverse function of the bounded Johnson transformation

$$y(i) = \frac{\hat{\alpha} + (\hat{\alpha} + \hat{\beta}) \exp\left[\frac{z(i) - \hat{\eta}}{\hat{\theta}}\right]}{1 + \exp\left[\frac{z(i) - \hat{\eta}}{\hat{\theta}}\right]} \quad i = s_{\max} + 1, \ldots, L_s \quad \text{(B5)}$$

6. The sequence of daily maximum temperature $t_{\max}(i) = s_{\max} + 1, \ldots, L_s$ is obtained as

$$t_{\max}(i) = \mu_T(i) + \sigma_T(i) y(i) \quad i = s_{\max} + 1, \ldots, L_s \quad \text{(B6)}$$

where

$$\mu_T(i) = \frac{1}{2} \hat{a}_{\mu,0} + \sum_{j=1}^{n_{h,\mu}} \left[\hat{a}_{\mu,j} \cos\left(\frac{2\pi j}{D} i\right) + \hat{b}_{\mu,j} \sin\left(\frac{2\pi j}{D} i\right) \right] \quad \text{(B7)}$$

$$\sigma_T(i) = \left\{ \frac{1}{2} \hat{a}_{\sigma^2,0} + \sum_{j=1}^{n_{h,\sigma^2}} \left[\hat{a}_{\sigma^2,j} \cos\left(\frac{2\pi j}{D} i\right) + \hat{b}_{\sigma^2,j} \sin\left(\frac{2\pi j}{D} i\right) \right] \right\}^{1/2} \quad \text{(B8)}$$

References

Akaike H (1974) Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60:255–265

Alonso AM, Peña D, Romo J (2002) Forecasting time series with sieve bootstrap. J Stat Plan Infer 100:1–11

Alonso AM, Peña D, Romo J (2003) On sieve bootstrap prediction intervals. Statist Probab Lett 65:13–20

Anderson TW, Darling DA (1952) Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Ann Math Stat 23:193–212

Anderson TW, Olkin I (1985) Maximum-likelihood estimation of the parameters of a multivariate normal distribution. Linear Algebra Appl 70:147–171

Baillie RT, Chung S (2002) Modeling and forecasting from trend-stationary long memory models with applications to climatology. Int J Forecasting 18:215–226

Bechini L, Bocchi S, Maggiore T, Confalonieri R (2006) Parameterization of a crop growth and development simulation model at sub-model components level. An example for winter wheat (*Triticum aestivum* L.). Environ Model Softw 21:1042–1054

Bisaglia L, Grigoletto M (2001) Prediction intervals for FARIMA processes by bootstrap methods. J Stat Comput Simul 68:185–201

Box GEP, Jenkins GM (1976) Time series analysis forecasting and control. Holden-Day

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York

Buttafuoco G, Caloiero T, Coscarelli R (2015) Analyses of drought events in Calabria (southern Italy) using standardized precipitation index. Water Resour Manag 29:557–573

Caballero R, Jewson S, Brix A (2002) Long memory in surface air temperature: detection, modeling, and application to weather derivative valuation. Clim Res 21:127–140

Caldiz DO, Gaspari FJ, Haverkort AJ, Struik PC (2001) Agro-ecological zoning and potential yield of single or double cropping of potato in Argentina. Agric For Meteorol 109:311–320

Caloiero T, Coscarelli R, Ferrari E, Sirangelo B (2015a) Analysis of dry spells in southern Italy (Calabria). Water 7:3009–3023

Caloiero T, Buttafuoco G, Coscarelli R, Ferrari E (2015b) Spatial and temporal characterization of climate at regional scale using homogeneous monthly precipitation and air temperature data: an application in Calabria (southern Italy). Hydrol Res 46:629–646

Caloiero T, Callegari G, Cantasano N, Coletta V, Pellicone G, Veltri A (2015c) Bioclimatic analysis in a region of southern Italy (Calabria). Plant Biosystems, in press, doi:10.1080/11263504.2015.1037814

Campbell SD, Diebold FX (2005) Weather forecasting for weather derivatives. J Am Stat Assoc 100:6–16

Chen SS, Gopinath RA (2000) Gaussianization. Adv Neural Comput Syst 13:423–429

Coscarelli R, Caloiero T (2012) Analysis of daily and monthly rainfall concentration in southern Italy (Calabria region). J Hydrol 416–417:145–156

Curriero FC, Heiner KS, Samet JM, Zeger SL, Strug L, Patz JA (2002) Temperature and mortality in 11 cities of the eastern United States. Am J Epidemiol 155:80–87

Doukhan P, Oppenheim G, Taqqu MS (2003) Theory and application of long-range dependence. Birkhäuser, Boston

Ehsanzadeh E, Adamowski K (2010) Trends in timing of low stream flows in Canada: impact of autocorrelation and long-term persistence. Hydrol Process 24:970–980

Ferrari E, Caloiero T, Coscarelli R (2013) Influence of the North Atlantic oscillation on winter rainfall in Calabria (southern Italy). Theor Appl Climatol 114:479–494

Granger CWJ, Joyeux R (1980) An introduction to long-range time series models and fractional differencing. J Time Ser Anal 1:15–30

Grimaldi S (2004) Linear parametric models applied on daily hydrological series. J Hydrolog Eng 9:383–391

Grimaldi S, Serinaldi F, Tallerini C (2005) Multivariate linear parametric models applied to daily rainfall time series. Adv Geosc 2:87–92

Hajat S, Kovats RS, Atkinson RW, Haines A (2002) Impact of hot temperatures on death in London: a time series approach. J Epidemiol Community Health 56:367–372

Hólm E, Andersson E, Beljaars A, Lopez P, Mahfouf JF, Simmons AJ, Thépaut JN (2002) Assimilation and modelling of the hydrological cycle: ECMWF's status and plans. ECMWF Tech Memo 383, Reading

Hosking JRM (1981) Fractional differencing. Biometrika 68:165–176

Hosking JRM (1984) Modeling persistence in hydrological time series using fractional differencing. Water Resour Res 20:1898–1908

Hurst HE (1951) Long-term storage capacity of reservoirs. Trans Am Soc Civil Eng 116:770–799

Jewson S, Caballero R (2003) Seasonality in the statistics of surface air temperature and the pricing of weather derivatives. Meteorol Appl 10:367–376

Johnson NL (1949) Systems of frequency curves generated by methods of translation. Biometrika 36:149–176

Keellings D, Waylen P (2012) The stochastic properties of high daily maximum temperatures applying crossing theory to modeling high-temperature event variables. Theor Appl Climatol 108:579–590

Kendall MG (1962) Rank correlation methods. Hafner Publishing Company, New York

Koscielny-Bunde E, Kantelhardt JW, Braun P, Bunde A, Havlin S (2006) Long-term persistence and multifractality of river runoff records: detrended fluctuation studies. J Hydrol 322:120–137

Koutsoyiannis D (2002) The Hurst phenomenon and fractional Gaussian noise made easy. Hydrolog Sci J 47:573–595

Kunst AE, Looman CWN, Mackenbach JP (1993) Outdoor air temperature and mortality in the Netherlands: a time-series analysis. Am J Epidemiol 137:331–341

Lee T (2015) Stochastic simulation of precipitation data for preserving key statistics in their original domain and application to climate change analysis. Theor Appl Climatol. doi:10.1007/s00704-015-1395-0

Lohre M, Sibbertsen P, Könning T (2003) Modeling water flow of the Rhine River using seasonal long memory. Water Resour Res 39:1132

Lye LM, Lin Y (1994) Long-term dependence in annual peak flows of Canadian rivers. J Hydrol 160:89–103

Mann HB (1945) Nonparametric tests against trend. Econometrica 13:245–259

Montanari A, Rosso R, Taqqu MS (1997) Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation, and simulation. Water Resour Res 33:1035–1044

Montanari A, Rosso R, Taqqu MS (2000) A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. Water Resour Res 36:1249–1259

Pelletier JD, Turcotte DL (1997) Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment. J Hydrol 203:198–208

Prass TS, Bravo JM, Clarke RT, Collischonn W, Lopes SRC (2012) Comparison of forecasts of mean monthly water level in the Paraguay River, Brazil, from two fractionally differenced models. Water Resour Res 48:W05502

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17:182–190

Rupasinghe M, Mukhopadhyayb P, Samaranayakec VA (2014) Obtaining prediction intervals for FARIMA processes using the sieve bootstrap. J Stat Comput Sim 84:2044–2058

Rupasinghe M, Samaranayake VA (2012) Asymptotic properties of sieve bootstrap prediction intervals for FARIMA processes. Statist Probab Lett 82:2108–2114

Servidio S, Greco A, Matthaeus WH, Osman KT, Dmitruk P (2011) Statistical association of discontinuities and reconnection in magnetohydrodynamic turbulence. J Geophys Res 116:A09102

Sheng H, Chen YQ (2011) FARIMA with stable innovations model of Great Salt Lake elevation time series. Signal Process 91:553–561

Smith RL (1993) Long-range dependence and global warming. In: Barnett V, Turkerman KF (eds) Statistics for the environment. Wiley, New York, pp. 141–146

Snedecor GW, Cochran WG (1989) Statistical methods, 8th edn. Iowa State University Press, Iowa City

Sugiura N (1978) Further analysis of the data by Akaike's information criterion and the finite corrections. Commun Stat A-Theor 7:13–26

Tuenter HJH (2001) An algorithm to determine the parameters of the S_U-curves in the Johnson system of probability distributions by moment matching. J Stat Comput Sim 70:325–347

Verdoodt A, Van Ranst E, Ye L (2004) Daily simulation of potential dry matter production of annual field crops in tropical environments. Agron J 96:1739–1753

Yang G, Bowling LC (2014) Detection of changes in hydrologic system memory associated with urbanization in the Great Lakes region. Water Resour Res 50:3750–3763

Ye L, Tang H, Zhu J, Verdoodt A, Van Ranst E (2008) Spatial patterns and effects of soil organic carbon on grain productivity assessment in China. Soil Use Manage 24:80–91

Ye L, Van Ranst E (2002) Population carrying capacity and sustainable agricultural use of land resources in Caoxian County (North China). J Sustain Agr 19:75–94

Ye L, Van Ranst E (2009) Production scenarios and the effect of soil degradation on long-term food security in China. Global Environ Chang 19:464–481

Ye L, Xiong W, Li Z, Yang P, Wu W, Yang G, Fu Y, Zou J, Chen Z, Van Ranst E, Tang H (2013) Climate change impact on China food security in 2050. Agron Sustain Dev 33:363–374

Yevjevich V (1972) Structural analysis of hydrologic time series. Hydrol Pap 56, Colorado State University, Fort Collins (CO)