JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, D18121, doi:10.1029/2009JD011960, 2009

Deterministic versus stochastic trends: Detection and challenges

S. Fatichi,[1] S. M. Barbosa,[2] E. Caporali,[1] and M. E. Silva[3]

Received 23 February 2009; revised 22 June 2009; accepted 29 June 2009; published 29 September 2009.

[1] The detection of a trend in a time series and the evaluation of its magnitude and statistical significance is an important task in geophysical research. This importance is amplified in climate change contexts, since trends are often used to characterize long-term climate variability and to quantify the magnitude and the statistical significance of changes in climate time series, both at global and local scales. Recent studies have demonstrated that the stochastic behavior of a time series can change the statistical significance of a trend, especially if the time series exhibits long-range dependence. The present study examines the trends in time series of daily average temperature recorded in 26 stations in the Tuscany region (Italy). In this study a new framework for trend detection is proposed. First two parametric statistical tests, the Phillips-Perron test and the Kwiatkowski-Phillips-Schmidt-Shin test, are applied in order to test for trend stationary and difference stationary behavior in the temperature time series. Then long-range dependence is assessed using different approaches, including wavelet analysis, heuristic methods and by fitting fractionally integrated autoregressive moving average models. The trend detection results are further compared with the results obtained using nonparametric trend detection methods: Mann-Kendall, Cox-Stuart and Spearman's ρ tests. This study confirms an increase in uncertainty when pronounced stochastic behaviors are present in the data. Nevertheless, for approximately one third of the analyzed records, the stochastic behavior itself cannot explain the long-term features of the time series, and a deterministic positive trend is the most likely explanation.

Citation: Fatichi, S., S. M. Barbosa, E. Caporali, and M. E. Silva (2009), Deterministic versus stochastic trends: Detection and challenges, *J. Geophys. Res.*, *114*, D18121, doi:10.1029/2009JD011960.

1. Introduction

[2] The evaluation of trends in hydro-climatic time series has always received a noticeable interest from the scientific community and even from professionals and companies involved in risk analysis and long-term design of infrastructures. The presence of deterministic trends in the analyzed time series may, in fact, provide information about the future evolution of the process or at least on the modifications occurred. In practical engineering the knowledge of the trend for a given variable of interest may help to forecast future realizations and to design future scenarios. Nowadays, with the growing importance of climate change assessment, trend detection and evaluation is a subject of intensive scientific research [*Brunetti et al.*, 2001; *Burn et al.*, 2002; *Groisman et al.*, 2004; *Cohn and Lins*, 2005; *Barbosa et al.*, 2008], as also testified in the recent fourth assessment report of the Intergovernmental Panel on Climate Change (IPCC) [*IPCC*, 2007]. One branch of climate change science is devoted to analyzing the past climate events and inside this branch trend detection and statistical significance testing assume an important role [*Trenberth et al.*, 2007].

[3] Notwithstanding this major attention to the topic, deriving a clear conclusion about the existence of a trend in a time series is far from a trivial task. The definition of trend is in someway a thorny problem as underlined by the tautological definition of *Cairncross* [1971]: "a trend is a trend is a trend." As already proposed by *Barbosa et al.* [2008] we prefer to consider the opposite and more robust concept of stationarity for which the statistical and mathematical properties are well defined [*Bras and Rodriguez-Iturbe*, 1994; *Beran*, 1994]. The trend detection problem is thus transformed into a nonstationarity identification problem. The identification of a trend in a time series, may not be a direct consequence of a deterministic process, since other statistical behaviors such as Long-Range Dependence (LRD) (also called long memory or long-term persistence), may induce trend-like behaviors, that influence trend detection tests, providing a false rejection of the null hypothesis of stationarity. Some authors argue that the concept of nonstationarity, even if well defined for the theory of stochastic processes, applies poorly to observed time series of climate realizations, because of their intrinsic limitations [*Koutsoyiannis*, 2006]. The effect of long-range dependence

[1]Department of Civil and Environmental Engineering, University of Firenze, Florence, Italy.
[2]Faculty of Science, University of Porto, Porto, Portugal.
[3]Faculty of Economics, University of Porto, Porto, Portugal.

0148-0227/09/2009JD011960$09.00

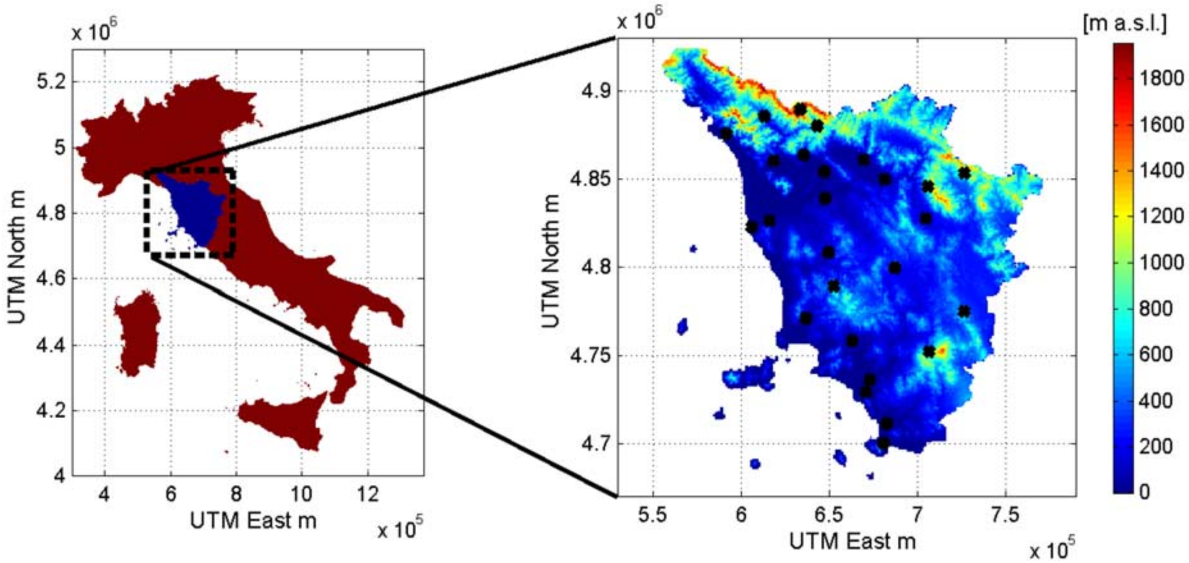

Figure 1. Digital elevation model and analyzed temperature gauges distribution in the Tuscany region (Italy).

in trend analysis has been the source of a recent challenging scientific discussion. The role of long-range dependence in the ''famous trend'' detected in the instrumental temperature time series of the Northern Hemisphere [*Mann et al.*, 1999; *McIntyre and McKitrick*, 2003; *Moberg et al.*, 2005] has been analyzed by *Cohn and Lins* [2005] and *Rybski et al.* [2006]. The two research groups have shown that the LRD hypothesis is compatible with both the instrumental records and the reconstructed time series of global temperature. To account for LRD in statistical tests they have suggested some modifications of the common statistical tests for the significance of a linear trend. Their conclusions about the statistical significance of the increasing trend in the time series of temperature are rather different, but both show a modification of the statistical significance due to LRD. *Cohn and Lins* [2005] show for example that orders of magnitude of significance can be lost in the statistical tests when LRD is present. Starting from this idea, a recent work of *Koutsoyiannis and Montanari* [2007] explains theoretically and numerically the implications of LRD in hydroclimatic research. They underline that the statistical uncertainty is dramatically increased in the presence of serial dependence, especially if the latter is LRD. Regarding this discussion the importance of reducing the uncertainty in trend detection appears once more, as well as the recognition of the important effect of the dependence structure, since the consequences may extend far beyond a statistical or scientific discussion with implications for policy makers and society.

[4] Here several issues related to trend identification, starting from the concept of trend itself, are discussed in the scope of the analysis of 26 temperature time series from the Tuscany region (Italy). The analyzed time series are tested for non stationarity through two parametric statistical tests: the Phillips-Perron (PP) test [*Phillips and Perron*, 1988] and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [*Kwiatkowski et al.*, 1992]. The Phillips-Perron (PP) test has been designed to test the null hypothesis of a unit root against a trend stationary alternative, while the KPSS test complements unit root tests by testing the null hypothesis of trend stationarity against the alternative hypothesis of difference stationarity.

[5] The scaling properties of the time series are then analyzed in order to assess long-range dependence features. The analysis is carried out in different ways: in the wavelet domain, through heuristic methods, in the frequency domain with the Geweke and Porter-Hudak's Method and by fitting a fractionally integrated autoregressive moving average model ARFIMA to the time series. This framework together with the comparison of the results with nonparametric trend detection methods allows to draw conclusions about the possible presence of a deterministic trend. This is a more reliable alternative for trend detection than simply verifying the presence of a trend through linear regression or nonparametric statistical tests and provides further insights on nonstationarity analysis. The proposed approach may be a little more laborious but the large uncertainties present and the growing importance of trend significance justify the prudence in asserting a deterministic trend.

2. Data Description and Deseasonalization

[6] The time series considered here are the daily average temperature records from 26 stations in the region of Tuscany (Italy). The stations, as illustrated in Figure 1, are widespread throughout the whole of the region. The time series cover a period varying from 15 to 47 years in the second half of the 20th century. The data set is extracted from the record of the Italian National Hydrographic Service, including only the time series longer than 15 years and without relevant gaps. A complete overview of the data is provided in Table 1. Having a homogeneous data set with few gaps is important when the long-range dependency characteristic is considered [*Rust et al.*, 2008].

[7] The daily temperature time series are obviously affected by seasonality, thus the variation of the sample mean, variance as well as of other statistics may be periodical. The seasonality should be removed before analyzing the scaling

Table 1. Analyzed Temperature Gauge Records

ID	Station	Elevation[a] (m asl)	Period	Percent Missing
1	S. Miniato	124	1980–1995	0
2	Castel di Pietra	56	1980–1995	4.4
3	Boscolungo	1340	1980–1998	1.3
4	Montepulciano	607	1980–1998	0.4
5	Castelmartini	23	1980–1997	0
6	Alberese	17	1980–1998	0.4
7	S. Donato	19	1980–1998	0.4
8	Orbetello	1	1980–1998	0.8
9	Castelnuovo Garfagnana	276	1982–1998	0
10	S. Marcello Pistoiese	625	1980–1999	5.5
11	Mutigliano	62	1980–1998	0
12	S. Giovanni Valdarno	132	1980–1998	0
13	Volterra	476	1983–2000	0.4
14	Nugola	69	1980–1998	0
15	Castel del Piano	639	1980–1998	0
16	Camaldoli	1111	1980–1999	0
17	Pescia	62	1980–1998	0
18	Suvereto	112	1980–2000	1.0
19	Vallombrosa	955	1980–2000	0
20	Prato	70	1980–1997	4.9
21	Larderello	400	1980–2000	0
22	Siena	348	1985–2001	0
23	Grosseto	8	1980–1998	0
24	Massa	65	1980–1999	0.1
25	Firenze Ximeniano	51	1953–1998	0.9
26	Livorno	3	1951–1998	0.3

[a]Here asl, above sea level.

behavior of the time series [*Montanari et al.*, 1997]. In order to remove periodicity, transformations or specific procedures like the STL procedure based on loess smoothing [*Cleveland et al.*, 1990] are often necessary. In this study we prefer to avoid transforming the data [*Montanari et al.*, 2000] and we use a classical approach consisting of estimating and subtracting a periodic deterministic component from the series. Operatively, only the mean and the variance of the seasonal cycle are removed, because their components are usually more meaningful than other higher-order statistics. This technique consists in dividing an observed time series X_t in 365 subseries, created with the available data for every day of the year. For every subseries, mean and variance are estimated, yielding 365 values of mean and 365 values of variance function of the Julian day [*Grimaldi*, 2004]. The seasonality in each value of X_t is then removed by subtracting the mean and dividing by the standard deviation of the correspondent day of the year. Such a procedure does not allow, in general, a complete removal of the seasonality. Nevertheless, in many cases the deseasonalization of the two first order statistics is enough to eliminate any significant periodicity from the autocorrelation function of the data, in order to consider the time series deseasonalized for the proposed analysis [*Montanari et al.*, 1997].

3. Nonstationarity Assessment

[8] The hypothesis of nonstationarity is tested with two parametric statistical tests borrowed from econometrics and aimed at discriminating between stationarity, a deterministic trend and non stationarity in the form of a unit root (including random walk). Difference stationarity and trend stationarity are two specific forms of nonstationarity. A process X_t is difference stationary or integrated if its dth backward difference $(1 - B^d)X_t$ is stationary and invertible, where B is the backshift operator and d is a integer. In the particular and frequent case of $d = 1$, a process is described as a unit root or integrated of order 1. The distinction between a unit root and a stationary process is important to avoid on one hand over differencing (differencing a stationary process) leading to difficulties in estimation and forecasting due to noninvertibility, and on the other hand to avoid spurious time regressions and overoptimistic forecasts at long lead times by failing to difference a nonstationary process. A process X_t is trend stationary if it is a combination of a deterministic trend with a stationary and zero mean uncorrelated process. Thus, a trend stationary process is not difference stationary since its dth backward difference is not invertible. Statistical tests have been developed to assess these two specific forms of nonstationarity. The Phillips-Perron (PP) test [*Phillips and Perron*, 1988] has been designed to test the null hypothesis of a unit root against a trend stationary alternative. The Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) [*Kwiatkowski et al.*, 1992] test complements unit root tests by testing the null hypothesis of trend stationarity (in the form of a constant level or a deterministic trend plus a stationary stochastic noise) against the alternative hypothesis of difference stationarity.

[9] The Phillips-Perron (PP) test is based on the model:

$$X_t = \eta + \beta t + \pi X_{t-1} + \psi_t, \tag{1}$$

where η and β are the parameters of a first-order polynomial regression and the stationary process ψ_t is not assumed to be white noise, with serial correlation and heteroscedasticity in the ψ_t term being handled directly in the test statistic. The unit root null hypothesis is expressed by $H_0{:}\pi = 1$ versus $H_1{:}\pi < 1$.

[10] The Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) test is based on the model:

$$X_t = \beta t + r_t + \nu_t, \tag{2}$$

where r_t is a random walk, $r_t = r_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and ν_t is a stationary process. The hypothesis for the KPSS test are $H_0{:}\sigma_\varepsilon^2 = 0$ (and $\beta = 0$) versus $H_1{:}\sigma_\varepsilon^2 \neq 0$ (for the particular case of trend stationarity in the form of a constant level) and $H_0{:}\sigma_\varepsilon^2 = 0$ (and $\beta \neq 0$) versus $H_1{:}\sigma_\varepsilon^2 \neq 0$ in the general case of trend stationarity.

[11] These tests are complementary and should be jointly employed. If both null hypotheses are rejected alternative parameterizations like LRD should be considered. If both tests fail to reject the null hypothesis, then the time series (or the tests) are not sufficiently informative for discriminating the kind of stationary behavior. Rejection of the unit root hypothesis in the PP test and no rejection of the KPSS test null hypothesis does not preclude a deterministic trend, while no rejection of the unit root null hypothesis in the PP test and rejection of KPSS's null hypothesis indicates a unit root process.

[12] The results of the PP and KPSS tests for the time series analyzed are summarized in Table 2. From the PP test results one can conclude that the unit root (random walk) hypothesis is rejected for all the analyzed time series; this is not surprising since in contrast with economic time series,

Table 2. P Values From KPSS and PP Statistical Tests

ID	Station	PP Test[a]	KPSS Test[b]	Rejection of Deterministic Trend?
1	S. Miniato	<0.01	>0.1	no
2	Castel di Pietra	<0.01	0.024	yes
3	Boscolungo	<0.01	>0.1	no
4	Montepulciano	<0.01	>0.1	no
5	Castelmartini	<0.01	>0.1	no
6	Alberese	<0.01	0.074	no
7	S. Donato	<0.01	<0.01	yes
8	Orbetello	<0.01	<0.01	yes
9	Castelnuovo Garfagnana	<0.01	>0.1	no
10	S. Marcello Pistoiese	<0.01	<0.01	yes
11	Mutigliano	<0.01	<0.01	yes
12	S. Giovanni Valdarno	<0.01	>0.1	no
13	Volterra	<0.01	>0.1	no
14	Nugola	<0.01	0.015	yes
15	Castel del Piano	<0.01	>0.1	no
16	Camaldoli	<0.01	>0.1	no
17	Pescia	<0.01	>0.1	no
18	Suvereto	<0.01	<0.01	yes
19	Vallombrosa	<0.01	0.037	yes
20	Prato	<0.01	>0.1	no
21	Larderello	<0.01	<0.01	yes
22	Siena	<0.01	<0.01	yes
23	Grosseto	<0.01	0.020	yes
24	Massa	<0.01	>0.1	no
25	Firenze Ximeniano	<0.01	0.064	no
26	Livorno	<0.01	0.011	yes

[a]H_0: random walk.
[b]H_0: trend stationarity.

hydro-climatic time series rarely exhibit a random walk behavior [*Barbosa et al.*, 2008]. The KPSS test rejects the hypothesis of a deterministic trend in 12 of the 26 analyzed stations. Thus trend-like features in these time series should be considered a result of stochastic behavior, rather than resulting from a deterministic trend. This outcome of the stationarity tests limits the possibility of a deterministic trend to a subsample of only 14 stations. Given the small number of stations analyzed, 26, we expect that the results are not affected by the multiple site testing issue [*Ventura et al.*, 2004].

4. Long-Range Dependence Assessment

[13] When time series exhibit a slow decay of autocorrelation function for large lags their behavior is called Long-Range Dependence (LRD) or long memory, or long-term persistence [*Beran*, 1994; *Taqqu and Teverovsky*, 1997; *Koutsoyiannis*, 2003; *Doukhan et al.*, 2003; *Robinson*, 2003; *Sibbertsen*, 2003; *Diebolt and Guiraud*, 2005]. LRD behavior may be represented both in the time domain and in the frequency domain. In the time domain, the autocorrelation function of the long-memory process is characterized by a hyperbolical decay:

$$\lim_{k \to \infty} \frac{\rho(k)}{k^{2d-1}} = c, \tag{3}$$

where $\rho(k)$ is the autocorrelation function at lag k, c is a constant and d is the long-memory parameter (or fractional differencing parameter). It can be shown [*Granger and Joyeux*, 1980] that with a long-memory d parameter in the range $0 < d < 0.5$, the time series is stationary and exhibits long memory with intensity d.

[14] The pioneering studies on long-memory processes came from the famous English hydrologist *Hurst* [1951]. The famous Hurst parameter H is related to d: $H = d + 0.5$. After him it was found in several fields and applications that time series may exhibit the phenomenon of long-range dependence [*Haslett and Raftery*, 1989; *Montanari et al.*, 1997; *Stephenson et al.*, 2000; *Syroka and Toumi*, 2001; *Caballero et al.*, 2002; *Koutsoyiannis*, 2002]. The physical explanation for observed behaviors characterized by LRD and the quantification of LRD are open discussions in the scientific literature [*Koutsoyiannis*, 2005]. Nevertheless the aim of this study is not to deal with the physical explanation of this observed behavior that has animated the scientific community in the last decades (references may be found in the works of *Klemeš* [1974], *Beran* [1994], *Montanari et al.* [1997], *Koutsoyiannis* [2003, 2005], *Mudelsee* [2007], and *Rust et al.* [2008]), but to provide an evaluation of the long-range dependence (or long memory) d parameter used to characterize LRD. Among the existent methods proposed in the last three decades [*Mandelbrot and Ness*, 1968; *Bras and Rodriguez-Iturbe*, 1994; *Beran*, 1994; *Taqqu et al.*, 1995], multiple approaches are applied herein. First the time series are analyzed in the wavelet domain, then with three different heuristic methods in the time domain and with the Geweke and Porter-Hudak's method in the frequency domain, finally with a parametric approach by fitting the series with a fractionally integrated autoregressive moving average model (ARFIMA).

4.1. Wavelet Analysis

[15] The discrete wavelet transform is a powerful tool for the analysis of geophysical time series (see *Percival* [2008] for a review) and specifically for assessing the long-range dependent behavior of a geophysical time series [*Percival and Walden*, 2000; *Barbosa et al.*, 2006a, 2006b]. The maximal overlap discrete wavelet transform (MODWT) yields a decomposition of the "energy" of a time series as a scale-by-scale wavelet variance analysis [*Percival*, 1995]. The wavelet spectrum (the log-log plot of wavelet variance versus scale) provides an estimate of the spectral density function, and its slope yields an estimate of the scaling exponent for a power law process. The wavelet approach is quite robust to smooth, polynomial trends [e.g., *Jach and Kokoszka*, 2008] and it is able to act as a decorrelating transform, converting long-range dependence in the time domain into shorter-range dependence in the wavelet domain, rendering it particularly appealing for the analysis of long-range dependent behavior in geophysical time series. The wavelet spectra based on a MODWT with a Daubechies filter of length 4 and periodic boundary conditions are shown in Figure 2 for the temperature time series at two sample stations (Firenze [ID = 25] and Livorno [ID = 26]). The spectra show a linear decay approximately for scales larger than 10 days. The slope of the wavelet spectrum is therefore computed only for scales longer than 10 days by weighted least squares [*Abry et al.*, 2003]. The results of the wavelet analysis are shown in Figure 3a. The value of the long-memory parameter d does not suggest strong long-range dependence: only for six stations (*ID* = 6, 7, 8, 18, 22, 26) we found a value of d larger than 0.2. In order to assess uncertainties for the estimated wavelet slopes, a simple ordinary nonparametric bootstrap is applied

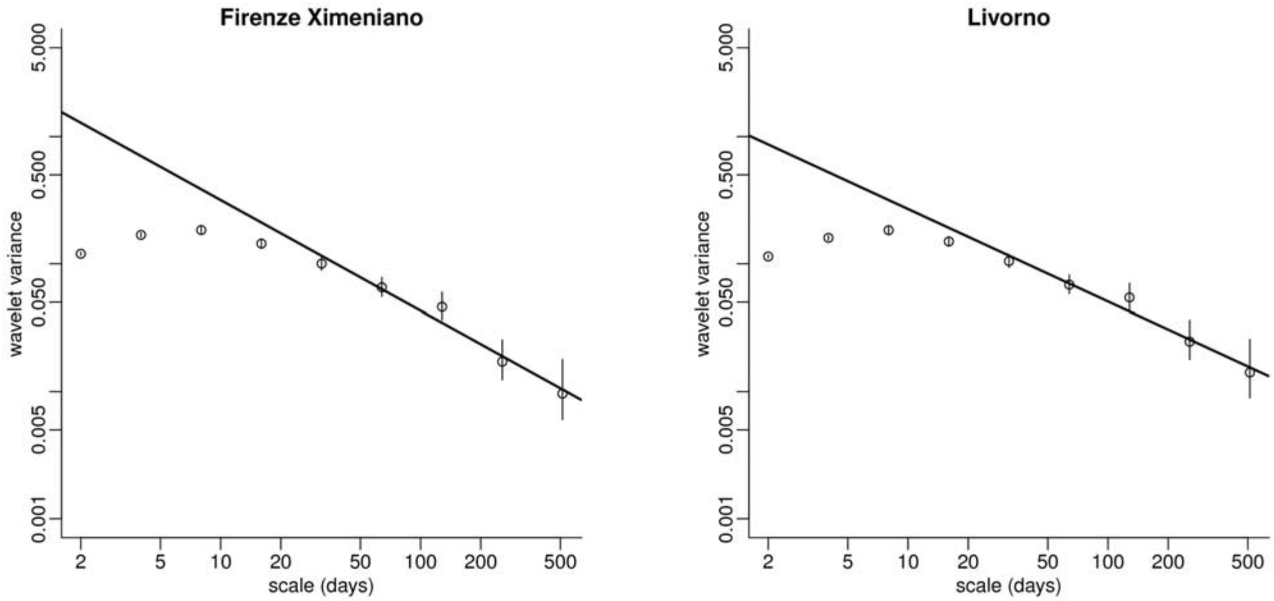

Figure 2. Wavelet spectrum for the gauge stations Firenze Ximeniano (ID = 25) and Livorno (ID = 26).

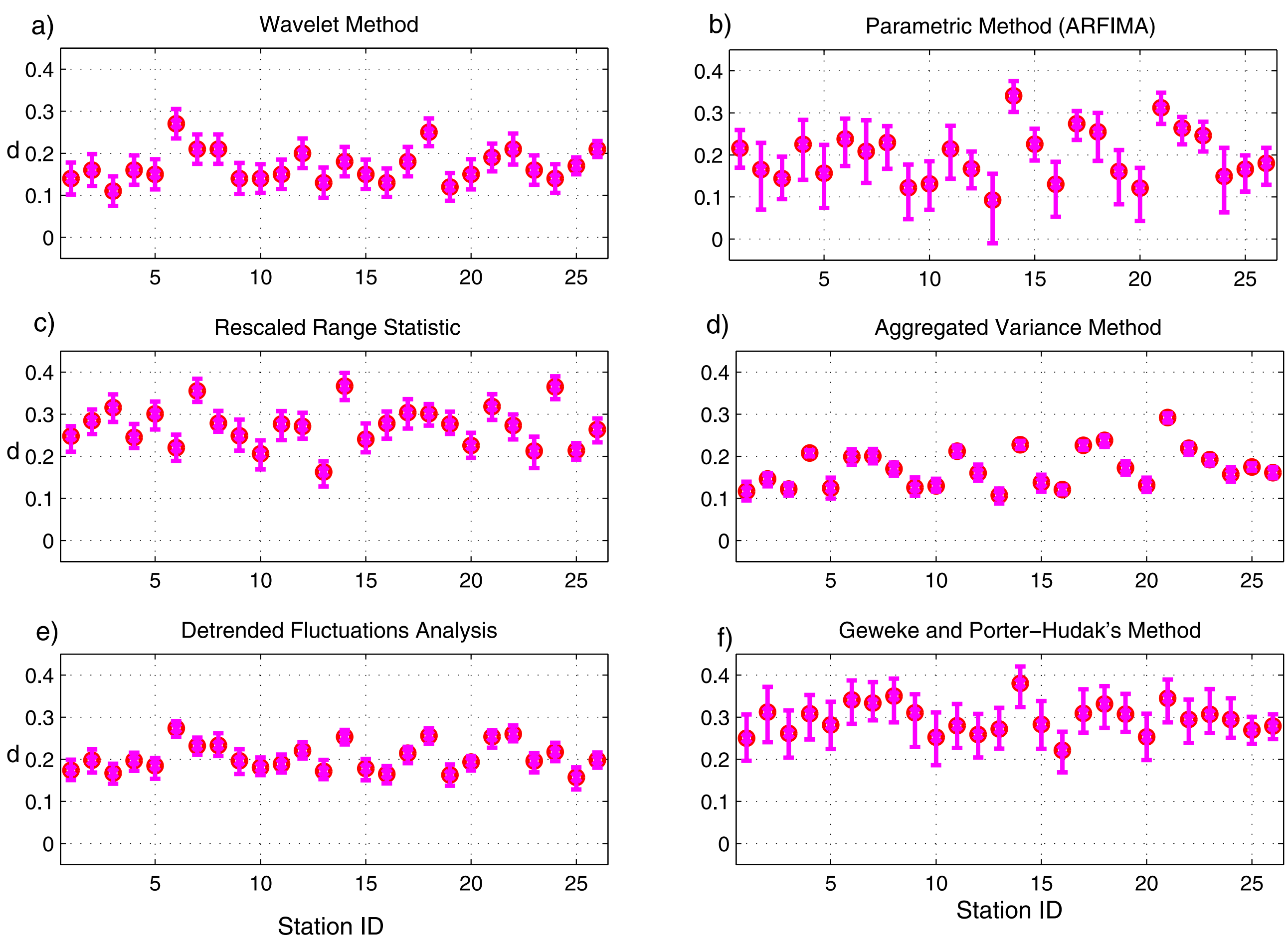

Figure 3. Results of the long-range dependence assessment, i.e., the *d* parameter. (a) Wavelet method. (b) Parametric method ARFIMA model. (c) Rescaled range statistic. (d) Aggregated variance method. (e) Detrended fluctuations analysis. (f) Geweke and Porter-Hudak's Method. The whiskers give 95% confidence intervals, except for wavelet method where they indicate standard error of the slope estimates.

[e.g., *Davison and Hinkley*, 1997]. Bootstrapping on the regression coefficient has been performed by considering 1000 bootstrap samples, each obtained by drawing a sample, with replacement, from the wavelet variance values. A linear regression model was then estimated from each bootstrap sample, yielding an estimate of the corresponding slope coefficient. Uncertainty bounds were obtained from the standard error of the slope estimates from the different replicates (Figure 3a).

4.2. Heuristic Methods and GPH

[16] Heuristic or semiparametric approaches do not require a complete parametrization of the process. Instead they use only asymptotic relationships to evaluate d. Many methods have been discussed in the literature [*Beran*, 1994; *Taqqu et al.*, 1995], for instance *Koutsoyiannis* [2002] recommends the use of the aggregated variance method. In climatic research many authors [*Kiràly and Janosi*, 2005; *Koscielny-Bunde et al.*, 2006; *Rybski et al.*, 2006] used the detrended fluctuation analysis (despite some caveats; see *Maraun et al.* [2004]), while other authors applied the Geweke and Porter-Hudak's method [*Wang et al.*, 2007; *Vyushin and Kushner*, 2009] or the periodogram method based in the frequency analysis of the time series [*Caballero et al.*, 2002]. In this study, the Geweke and Porter-Hudak's Method (GPH) and three heuristic methods are adopted, specifically: the Rescaled Range Statistic (RRS), the Aggregated Variance Method (AVM), and the Detrended Fluctuations Analysis (DFA) with linear polynomial. For further explanations about the various methods, refer to *Geweke and Porter-Hudak* [1983], *Beran* [1994], *Taqqu et al.* [1995], *Montanari et al.* [1997], *Kiràly and Janosi* [2005], and *Koscielny-Bunde et al.* [2006].

[17] The results obtained for the long-memory parameter d are displayed in Figures 3c–3f for the different methods including confidence intervals at the fixed nominal level of 95%. The confidence intervals for the parameter d from heuristic methods must be obtained with bootstrap techniques. Confidence intervals for GPH method can be obtained also directly [*Vyushin and Kushner*, 2009]. Bootstrap methods for time-dependent data are not as well understood as methods for independent random samples [*Buhlmann*, 2002; *Härdle et al.*, 2003] and there is a considerable undergoing research on the subject [*Härdle et al.*, 2003; *Franco and Reisen*, 2004; *Silva et al.*, 2006; *Franco and Reisen*, 2007]. A bootstrap in the residuals of the regression equation is applied as proposed by *Franco and Reisen* [2004, 2007] for RRS, AVM, DFA and GPH methods. This technique is based on resampling the residuals of the regression equations used to estimate d.

[18] The DFA method provides similar results to the wavelet analysis and indicates globally weak long memory with the larger values on the stations ID = 6, 7, 8, 14, 18, 21, 22, mostly the same stations indicated by the wavelet method. The GPH method confirms the results of DFA and wavelet methods in terms of the stations that exhibit the stronger long-range dependency but all the values of d are shifted toward higher values. More contrasting are the results obtained with the RRS and AVM methods, where there are larger differences between the d values for different stations. Nevertheless it is not surprising having quite different results from different long-memory assessment methods, given the different nature of the methods and the uncertainty present. In fact, when a short-memory process such as an ARMA model is superimposed to long-memory processes, the capability of these methods in detecting the right value of d decreases [*Franco and Reisen*, 2004]. Moreover, the heuristic methods, in contrast with the wavelet analysis, are less robust to the presence of deterministic trends that may distort the results, though if some discordant opinions on DFA exist [*Hu et al.*, 2001].

4.3. ARFIMA Models

[19] The fractionally integrated autoregressive moving average model [*Granger and Joyeux*, 1980; *Hosking*, 1981; *Beran*, 1994; *Reisen*, 1994], denoted by ARFIMA(p, d, q) is defined by the follow expression:

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t, \tag{4}$$

where X_t with t = 1, 2, ... n is the time series; B is the backshift operator, that means $BX_t = X_{t-1}$; $\Phi(B) = 1 - \Phi_1 B - \ldots - \Phi_p B^p$, and $\Theta(B) = 1 + \Theta_1 B + \ldots + \Theta_q B^q$ are the ordinary autoregressive and moving average polynomials respectively, ε_t is a white noise process. The expression $(1 - B)^d$ can be formally expanded as a power series:

$$(1-B)^d = \sum_{s=0}^{\infty} \frac{\Gamma(s-d)}{\Gamma(s+1)\Gamma(-d)} B^s. \tag{5}$$

The presence of autoregressive and moving average components allows these models to also capture the short-memory high-frequency behavior, in contrast with the Fractional Gaussian Noise FGN [*Beran*, 1994]. The identification of the ARFIMA models is carried out using the procedure proposed by *Montanari et al.* [1997]. First an initial value of the long-memory d parameter is proposed, in this case the average d from the other methods. Then the time series of temperature X_t is filtered with the chosen value of d to remove the long-memory component $Y_t = (1 - B)^d X_t$. A traditional short-memory estimation technique is applied to the new series Y_t. In this study the Akaike Final Prediction Error is adopted to estimate the orders p and q of the short-memory ARMA model and an iterative search algorithm [*Ljung*, 1999] is used to find the ARMA model parameters Φ and Θ. The order of the ARMA model together with a initial set of the parameters $\theta = (\Phi, \Theta, d)$ is, thus, available. This initial set of parameters is used to identify the ARFIMA (p, d, q) model for each time series. Starting from this first trial parameter vector θ, the estimation of the ARFIMA model may be achieved with different techniques of maximization of likelihood [*Beran*, 1994; *Diebolt and Guiraud*, 2005]. We evaluated the ARFIMA parameters by minimizing the Whittle Estimator $W(\theta)$ [*Whittle*, 1953; *Montanari and Toth*, 2007; *Rust et al.*, 2008].

$$W(\theta) = \sum_{j=1}^{N/2} \left[\frac{J(\lambda_j)}{J_m(\lambda_j, \theta)} \right], \tag{6}$$

where $\lambda_j = 2\pi\ j/N$ are the Fourier frequencies; J is the periodogram (which is an estimate for the spectral density) of the time series of temperature and J_m is the spectral density of the ARFIMA model that depends on the parameter vector θ. The Whittle Estimator $W(\theta)$, as stated from *Montanari et al.* [1997] and *Rust et al.* [2008], has good statistical properties and shares the desired properties of consistency and asymptotic normality with the maximum likelihood estimator, while being faster.

[20] The results of the ARFIMA model fits (Figure 3b) are more spread than those obtained with DFA and wavelet analysis, and they show a stronger long-range dependency. The confidence intervals at the fixed nominal level of 95% are obtained with the nonparametric bootstrap in the residuals as described by *Franco and Reisen* [2004]. The uncertainty in the estimation of d with this technique is larger than for the other methods. This is probably due to the contemporary estimation of the complete ARFIMA vector parameters $\theta = (\Phi, \Theta, d)$. The time series with the highest d are the stations ID = 6, 8, 14, 17, 18, 21, 22, 23, thus mainly the same stations also highlighted by the other methods.

5. Nonparametric Trend Detection

[21] Usually common nonparametric tests are applied to investigate time series of hydro-climatic variables, the most popular is the Mann-Kendall test [*Mann*, 1945; *Kendall*, 1975; *Lettenmaier et al.*, 1994; *Burn et al.*, 2002; *Yue et al.*, 2002a]. We used this test together with other two tests, the Cox-Stuart test [*Cox and Stuart*, 1955] and the Spearman's ρ test [*Daniels*, 1950; *Yue et al.*, 2002a; *Khaliq et al.*, 2009] to compare the results obtained in the proposed framework and to provide further explanation on the reasons of trend detection. As stated previously for KPSS and PP tests, since the stations analyzed are only 26 we think the multiple site testing issue [*Benjamini and Yekutieli*, 2001; *Ventura et al.*, 2004] is not relevant in this study.

[22] The nonparametric Mann-Kendall statistical test allows searching for trends in the data without assuming any particular distribution. The null hypothesis for the Mann-Kendall test is that the data are independent and randomly ordered. The Cox-Stuart test allows to verify if a variable has a monotonically tendency (reject of null hypothesis of trend absence), and the test is very close to the sign test for two independent samples. More details may be found in *Conover* [1999]. Spearman's ρ test verifies if a variable changes monotonically in time against the null hypothesis of tendency absence. Spearman's ρ test has similar power to the Mann-Kendall test even if it is much less popular [*Yue et al.*, 2002a]. In all the tests the occurrence of a trend is suggested when the null hypothesis of trend absence is rejected, i.e., when the p value is below a given level of significance.

[23] Most trend detection studies using the Mann-Kendall test implicitly assume that the sample data are serially independent, although hydro-climatic variables often display statistically significant serial correlation. It has been shown that the impact of positive/negative serial correlation on the Mann-Kendall test is to increase/decrease the rejection rate of the null hypothesis [*Hamed and Rao*, 1998; *Yue and Wang*, 2002; *Yue et al.*, 2002b]. Thus the presence of positive correlation increases the possibility to reject by chance the null hypothesis of trend absence even if it is true, first type error in the hypothesis testing, while the presence of negative correlation increases the possibility to accept the null hypothesis, when it is false (second type error), reducing the power of the test.

[24] Several techniques have been proposed to remove the influence of serial correlation in the data such as removing the lag-1 autoregressive [AR(1)] process from the time series prior to applying the test. This operation is called "prewhitening," but it has been demonstrated by *Yue and Wang* [2002] that "prewhitening" may seriously distort the possibility of the test to detect a trend. Another approach uses the effective or Equivalent Sample Size (ESS), that allows to modify the variance of the Mann-Kendall statistic in order to reduce the influence of autocorrelation [*Yue and Wang*, 2004]. Among the ESS techniques we use the modification to the variance of the Mann-Kendall test proposed by *Hamed and Rao* [1998], even if some concern regarding this method has arisen [*Yue and Wang*, 2004]. In the case of the analyzed temperature time series, a certain number of first type errors in the hypothesis testing is expected, due to the positive autocorrelation structure of the time series. Recently *Hamed* [2008] proposed a modified version of the Mann-Kendall test to account directly for the effect of long memory in the data. He highlighted once more a decrease in the significance of trends when long memory is present. *Koutsoyiannis and Montanari* [2007] also show that the equivalent sample size n' of the independent identically distributed IID classical statistic of a LRD process is:

$$n' = n^{2(1-H)} \tag{7}$$

where n is the real length of the sample and H is the Hurst coefficient. Significance is thus reduced as H tends to 1.

[25] Figure 4 shows the p value obtained for all the stations using the different tests. If a common α level of significance equal to 0.05 is adopted, the original Mann-Kendall test detects a positive trend in 23 stations and only in the stations ID = 13, 18, 22 the trends are not significant. The Spearman's ρ test confirms this result since it excludes a trend only for two stations ID = 18, 22. The Cox-Stuart test instead identifies a significant trend, i.e., reject the null hypothesis with α = 0.05 only for 12 stations. Considering the autocorrelation in the data, using the modified test of Mann-Kendall provides interesting results (Figure 4d) generally reducing the significance of the trends, since with α = 0.05 trends are identified only in 16 stations. From this analysis emerges that the identification of 7 trends by the original Mann-Kendall test seems induced by autocorrelation in the time series, thus altering the performance of the test itself, specifically this takes place for the stations ID = 7, 8, 9, 10, 12, 16, 24, as shown in Figure 4.

6. Deterministic Trends Discussion

[26] The results of the trend detection analysis for all the temperature time series are summarized in Figure 5, along with the corresponding linear slope coefficients estimated by ordinary least squares.

[27] Starting from 26 time series, the first discrimination is based on the common Mann-Kendall nonparametric test where a trend identification is rejected for three stations

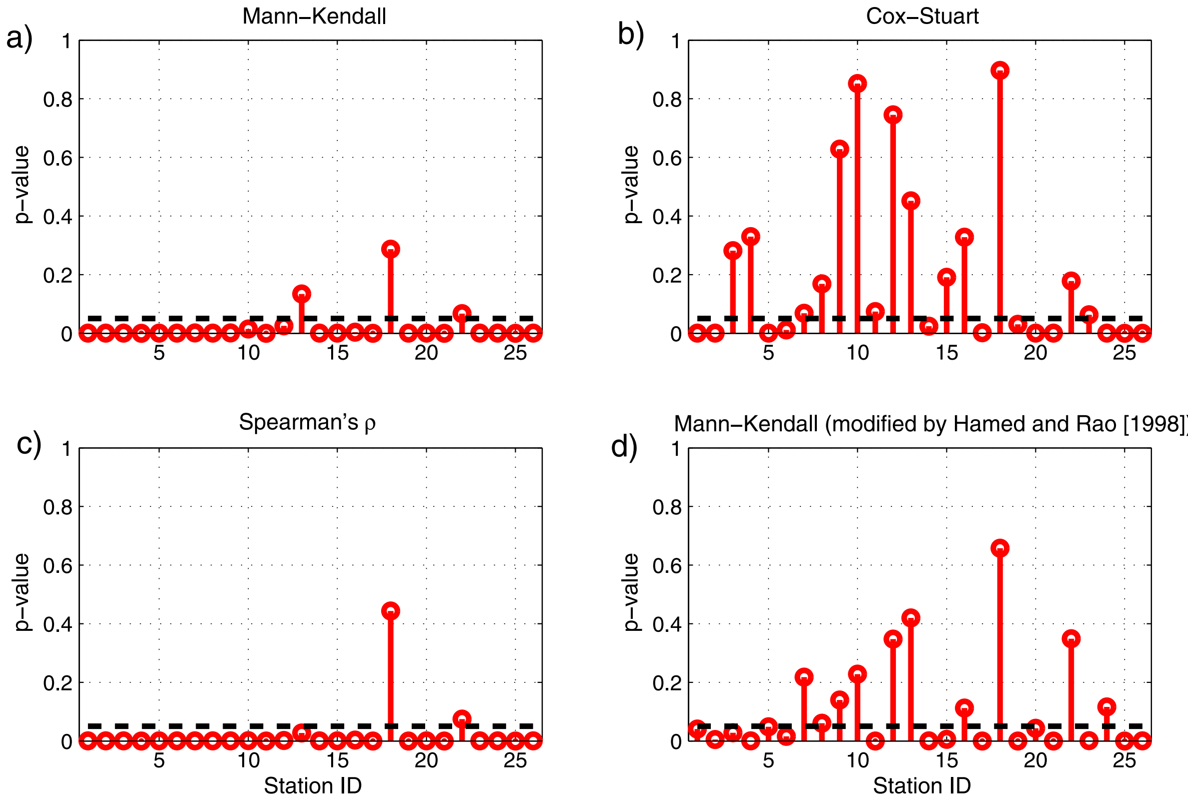

Figure 4. Results of the trend detection p value. (a) Mann-Kendall test. (b) Cox-Stuart test. (c) Spearman's ρ test. (d) Modified Mann-Kendall test [*Hamed and Rao*, 1998]. The dashed line indicates the 0.05 threshold.

ID = 13, 18, 22. Then on the basis of the Mann-Kendall test modified to take into account the autocorrelation of the process, the short memory appears to affect the test inducing a trend identification in 7 stations *ID* = 7, 8, 9, 10, 12, 16, 24. Regarding the remaining 16 stations a deterministic trend is not rejected (KPSS test) only for 9 stations *ID* = 1, 3, 4, 5, 6, 15, 17, 20, 25, for those stations thus a deterministic trend appears the most probable explanation of the tendency identified in the time series. Concerning the further 7 stations the trend identification may be induced by a long-range dependence behavior. This appears more probable for the stations *ID* = 14, 21 where the long-memory coefficient d is high, for the other stations this possibility cannot be excluded, but the evidence is weaker. As a matter of fact

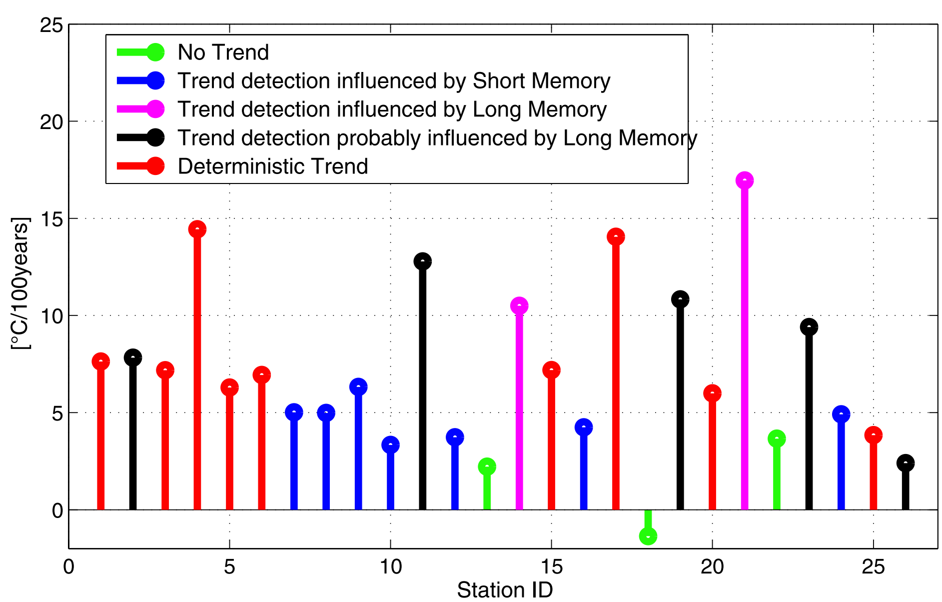

Figure 5. Magnitude of trend for the 26 stations, linear regression coefficient in [°C (100 years)$^{-1}$]. Interpretation of the possible reason for trend detection.

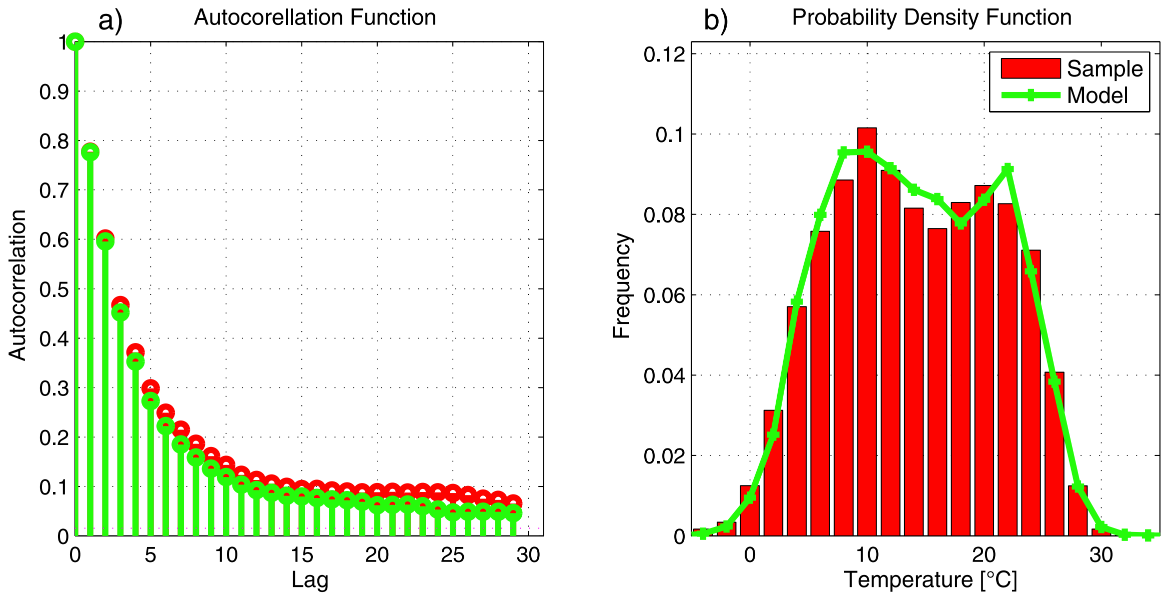

Figure 6. (a) Autocorrelation function and (b) probability density function for Firenze Ximeniano (ID = 25). Model refers to an ARFIMA(2,d,1) with $d = 0.1661$, $\Phi_1 = -0.2083$, $\Phi_2 = 0.5292$, $\Theta_1 = 0.8250$.

given the variability on the value of d obtained with the different methods (Figure 3) it does not seem appropriate to set a threshold to discriminate between time series that exhibit or not a long-memory behavior. Thus a definitive conclusion may not be drawn for the time series *ID* = 2, 11, 19, 23, 26. Some doubt also emerges for the time series *ID* = 4, 6, 17 where a deterministic trend is not rejected but the high values of d may be the reason why the nonparametric test results provide rejections of the null hypothesis.

[28] Finally a very simple experiment is carried out to assess the results obtained. Using the 26 ARFIMA(p, d, q) models estimated in section 4.3, we generate 100 times 26 time series of the same length of the analyzed records. The ARFIMA models are able to reproduce very well the autocorrelation structure and the probability density function of the analyzed time series, as shown in Figures 6a and 6b for Firenze Ximeniano, but it applies equally well to all the time series.

[29] From the generated time series, stationary and without trends, we estimated the number of stations with trend detected by means of the original and modified Mann-Kendall tests (α level of significance equal to 0.05). This experiment gives an estimate of how the correlation structure (long and short memory) influences a trend identification. The average number of stations with detected trend is 9.47 with the original Mann-Kendall test, which does not account for short memory, and 2.45 when short memory is taken into account. This experiment supports our results since we found almost the same number of trends identified because of short and long memory (7 and 2), corroborating the fact that deterministic trends may be a very likely explanation for the behavior of several temperature time series in the region of Tuscany.

7. Conclusions

[30] Trend assessment is a challenging task, since the stochastic behavior of a time series can induce trend-like features in the data, causing pure stochastic behavior to appear as a deterministic trend. This is particularly critical in the analysis of climate time series, since climate change is often identified by the existence of trends in the time series of climate observations. In this study we address the issues of trend identification, with a major focus on deterministic vs stochastic trends. Considering the impact of the stochastic behavior of a time series in trend detection, it is necessary to carefully assess the temporal dependence structure of the data, and specifically its long-range dependence characteristics. The comprehensive approach applied here shows how the application of complementary methods of trend detection allows to obtain further insights on the long-term variability of the time series.

[31] From the analysis of the 26 temperature time series in the Tuscany region, only for a subset of 9 stations a deterministic linear trend can be regarded as the most probable explanation for the rejection of the null hypothesis of the nonparametric tests. For the remaining records mixed conclusions are drawn, from no trend whatsoever to the influence of stochastic behavior in trends detection, especially long-range dependency. These findings are supported by the numerical experiment performed. This simple experiment cannot account separately for the influence of long memory, autoregressive model, moving average model and sample length on trend detection, for which more extensive Monte Carlo simulations would be necessary, but provides a robust validation of our findings concerning the analyzed time series. The results obtained may be affected by the length of the available records, that for most stations is limited, but the implications of the procedure discussed continue to hold and can be applied to different hydroclimatic data.

[32] It is confirmed in this work that the stochastic behavior of the time series, as already remarked by several authors [*Cohn and Lins*, 2005; *Kallache et al.*, 2005; *Koutsoyiannis and Montanari*, 2007; *Hamed*, 2008], may strongly influence trend detection and induce statistically

significant trends even in stationary time series. In particular the application of the original Mann-Kendall test or Spearman's ρ test may produce misleading results if the time series exhibit a strong memory dependence. As illustrated in this study, different methods can give different answers concerning the existence of long-range dependence and the specific value of the parameter d characterizing its strength. This issue requires particular attention and further investigation.

[33] The present study emphasizes the relevance of applying trend detection procedures before computing a deterministic linear trend and characterizing long-term climate features by the corresponding slope coefficient, particularly in climate change contexts.

[34] **Acknowledgments.** S. M. Barbosa wishes to thank FCT (grant SFRH/BPD/23992/2005), and all the authors are grateful to the anonymous reviewers for their extremely helpful and insightful comments on an earlier version of this manuscript.

References

Abry, P., P. Flandrin, M. S. Taqqu, and D. Veitch (2003), Self-similarity and long range dependence through the wavelet lens, in *Theory and Applications of Long Range Dependence*, edited by O. G. Doukhan and M. S. Taqqu, pp. 527–556, Birkhuser, Basel.

Barbosa, S. M., M. J. Fernandes, and M. E. Silva (2006a), Long range dependence in North Atlantic sea level, *Physica A*, *371*, 725–731.

Barbosa, S. M., M. E. Silva, and M. J. Fernandes (2006b), Wavelet analysis of the Lisbon and Gibraltar North Atlantic Oscillation winter indices, *Int. J. Climatol.*, *26*, 581–593.

Barbosa, S. M., M. E. Silva, and M. J. Fernandes (2008), Time series analysis of sea-level records: Characterising long-term variability, in *Nonlinear Time Series Analysis in the Geosciences—Applications in Climatology, Geodynamics, and Solar-Terrestrial Physics*, edited by R. V. Donner and S. M. Barbosa, pp. 157–173, Springer, Berlin.

Benjamini, Y., and D. Yekutieli (2001), The control of the false discovery rate in multiple testing under dependency, *Ann. Stat.*, *29*(4), 1165–1188.

Beran, J. (1994), *Statistic for Long-Memory Processes*, Chapman and Hall, New York.

Bras, R. L., and I. Rodriguez-Iturbe (1994), *Random Function and Hydrology*, vol. 21, Dover, New York.

Brunetti, M., M. Colacino, M. Maugeri, and T. Nanni (2001), Trends in the daily intensity of precipitation in Italy from 1951 to 1996, *Int. J. Climatol.*, *21*, 299–316.

Buhlmann, P. (2002), Bootstrap for time series, *Stat. Sci.*, *17*(1), 52–72.

Burn, D., A. Mohamed, and H. Elnur (2002), Detection of hydrologic trends and variability, *J. Hydrol.*, *255*, 107–122.

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

Cairncross, A. (1971), *Essays in Economic Management*, Allen and Unwin, London.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990), STL: A seasonal-trend decomposition procedure based on loess, *J. Off. Stat.*, *6*, 3–73.

Cohn, T. A., and H. F. Lins (2005), Nature's style: Naturally trendy, *Geophys. Res. Lett.*, *32*, L23402, doi:10.1029/2005GL024476.

Conover, W. J. (1999), *Practical Nonparametric Statistics*, 3rd ed., John Wiley, New York.

Cox, D. R., and A. Stuart (1955), Some quick tests for trend in location and dispersion, *Biometrika*, *42*, 80–95.

Daniels, H. E. (1950), Rank correlation and population models, *J. R. Stat. Soc. Ser. B*, *12*, 171–181.

Davison, A. C., and D. V. Hinkley (1997), *Bootstrap Methods and Their Application*, Cambridge Univ. Press, New York.

Diebolt, C., and V. Guiraud (2005), A note on long memory time series, *Qual. Quant.*, *39*, 827–836, doi:10.1007/s11135-004-0436-z.

Doukhan, P., G. Oppenheim, and M. S. Taqqu (2003), *Theory and Applications of Long-Range Dependence*, Birkhuser, Basel.

Franco, G. C., and V. A. Reisen (2004), Bootstrap techniques in semiparametric estimation methods for ARFIMA models: A comparison study, *Comput. Stat.*, *19*, 243–259.

Franco, G. C., and V. A. Reisen (2007), Bootstrap approaches and confidence intervals for stationary and non-stationary long-range dependence processes, *Physica A*, *375*, 546–562, doi:10.1016/j.physa.2006.08.027.

Geweke, J., and S. Porter-Hudak (1983), The estimation and application of long memory time series models, *J. Time Ser. Anal.*, *4*, 221–238.

Granger, C. W. J., and R. Joyeux (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 to daily hydrological series, *J. Hydrol. Eng.*, *9*(5), 383–391.

Groisman, P. Y., R. W. Knight, T. R. Karl, D. R. Easterling, B. Sun, and J. H. Lawrimore (2004), Contemporary changes of the hydrological cycle over the contiguous United States: Trends derived from in situ observations, *J. Hydrometeorol.*, *5*(1), 64–85.

Hamed, K. H. (2008), Trend detection in hydrologic data: The Mann-Kendall trend test under scaling hypothesis, *J. Hydrol.*, *349*, 350–363, doi:10.1016/j.jhydrol.2007.11.009.

Hamed, K. H., and A. R. Rao (1998), A modified Mann-Kendall trend test for autocorrelated data, *J. Hydrol.*, *204*, 182–196.

Härdle, W., J. Horowit, and J.-P. Kreiss (2003), Bootstrap methods for time series, *Int. Stat. Rev.*, *71*(2), 435–459.

Haslett, J., and A. E. Raftery (1989), Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource (with discussion), *Appl. Stat.*, *38*(1), 1–50.

Hosking, J. R. M. (1981), Fractional differencing, *Biometrika*, *68*, 165–176.

Hu, K., I. P. Ch., Z. Chen, P. Carpena, and H. E. Stanley (2001), Effects of trends on detrended fluctuation analysis, *Phys. Rev. E*, *64*, 011114, doi:10.1103/PhysRevE.64.011114.

Hurst, H. E. (1951), Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.*, *116*, 770–779.

Intergovernmental Panel on Climate Change (IPCC) (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, New York.

Jach, A., and P. Kokoszka (2008), Wavelet-domain test for long-range dependence in the presence of a trend, *Statistics*, *42*, 101–113.

Kallache, M., H. W. Rust, and J. Kropp (2005), Trend assessment: Applications for hydrology and climate research, *Nonlinear Process. Geophys.*, *12*, 201–210.

Kendall, M. G. (1975), *Rank Correlation Methods*, Griffin, London.

Khaliq, M. N., T. B. M. J. Ouarda, P. Gachon, L. Sushama, and A. St-Hilaire (2009), Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers, *J. Hydrol.*, *368*, 117–130, doi:10.1016/j.jhydrol.2009.01.035.

Kiràly, A., and I. M. Janosi (2005), Detrended fluctuation analysis of daily temperature records: Geographic dependence over Australia, *Meteorol. Atmos. Phys.*, *88*, 119–128.

Klemeš, V. (1974), The Hurst phenomenon—A puzzle?, *Water Resour. Res.*, *10*, 675–688.

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

Koutsoyiannis, D. (2002), The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrol. Sci. J.*, *47*(4), 573–595.

Koutsoyiannis, D. (2003), Climate change, the Hurst phenomenon and hydrological statistics, *Hydrol. Sci. J.*, *48*(1), 3–24.

Koutsoyiannis, D. (2005), Uncertainty, entropy, scaling and hydrological sthochastics, 2. Time dependence of hydrological processes, *Hydrol. Sci. J.*, *50*(43), 405–426.

Koutsoyiannis, D. (2006), Nonstationarity versus scaling in hydrology, *J. Hydrol.*, *324*, 239–254, doi:10.1016/j.jhydrol.2005.09.022.

Koutsoyiannis, D., and A. Montanari (2007), Statistical analysis of hydroclimatic time series: Uncertainty and insights, *Water Resour. Res.*, *43*, W05429, doi:10.1029/2006WR005592.

Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin (1992), Testing the null hypothesis of stationarity against the alternative of a unit root, *J. Econometrics*, 159–178.

Lettenmaier, D. P., E. F. Wood, and J. R. Wallis (1994), Hydro-climatological trends in the continental United States, 1948–88, *J. Clim.*, *7*, 586–607.

Ljung, L. (1999), *System Identification: Theory for the User*, Prentice-Hall, Upper Saddle River, N. J.

Mandelbrot, B. B., and J. W. V. Ness (1968), Fractional Brownian motions, fractional noises and application, *Soc. Ind. Appl. Math. Rev.*, *10*, 422–437.

Mann, H. B. (1945), Nonparametric tests against trend, *Econometrica*, *13*, 245–259.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1999), Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties and limitations, *Geophys. Res. Lett.*, *26*(6), 759–762.

Maraun, D., H. Rust, and J. Timmer (2004), Tempting long memory—On the interpretation of DFA results, *Nonlinear Process. Geophys.*, *11*, 495–503.

McIntyre, S., and R. McKitrick (2003), Corrections to the Mann et al. (1998) proxy data base and Northern Hemispheric average temperature series, *Energy Environ.*, *14*(6), 751–771.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlen (2005), Highly variable Northern Hemisphere temperatures reconstructed from low and high resolution proxy data, *Nature*, *433*(7026), 613–617.

Montanari, A., and E. Toth (2007), Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, *Water Resour. Res.*, *43*, W05434, doi:10.1029/2006WR005184.

Montanari, A., R. Rosso, and M. S. Taqqu (1997), Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation, *Water Resour. Res.*, *33*(5), 1035–1044.

Montanari, A., R. Rosso, and M. S. Taqqu (2000), A seasonal fractional ARIMA model applied to the Nile river monthly flows at Aswan, *Water Resour. Res.*, *36*(5), 1249–1265.

Mudelsee, M. (2007), Long memory of rivers from spatial aggragation, *Water Resour. Res.*, *43*, W01202, doi:10.1029/2006WR005721.

Percival, D., and A. T. Walden (2000), *Wavelet Methods for Time Series Analysis*, Cambridge Univ. Press, New York.

Percival, D. B. (1995), On the estimation of the wavelet variance, *Biometrika*, *82*, 619–631.

Percival, D. B. (2008), Analysis of geophysical time series using discrete wavelet transforms: An overview, in *Nonlinear Time Series Analysis in the Geosciences—Applications in Climatology, Geodynamics, and Solar-Terrestrial Physics*, edited by R. V. Donner and S. M. Barbosa, pp. 61–79, Berlin, Springer.

Phillips, P. C. B., and P. Perron (1988), Testing for a unit root in time series regression, *Biometrika*, *75*, 335–346.

Reisen, V. A. (1994), Estimation of the fractional differential parameter in the ARIMA (p, d, q) model using the smoothed periodogram, *J. Time Ser. Anal.*, *15*(3), 335–350.

Robinson, P. M. (2003), *Time Series With Long Memory, Advanced Texts in Econometrics*, Oxford Univ. Press, New York.

Rust, H. W., O. Mestre, and V. K. C. Venema (2008), Fewer jumps, less memory: Homogenized temperature records and long memory, *J. Geophys. Res.*, *113*, D19110, doi:10.1029/2008JD009919.

Rybski, D., A. Bunde, S. Havlin, and H. von Storch (2006), Long-term persistence in climate and the detection problem, *Geophys. Res. Lett.*, *33*, L06718, doi:10.1029/2005GL025591.

Sibbertsen, P. (2003), Log-periodogram estimation of the memory parameter of a long-memory process under trend, *Stat. Prob. Lett.*, *61*(3), 261–268.

Silva, E. M., G. C. Franco, V. A. Reisen, and F. R. B. Cruz (2006), Local bootstrap approaches for fractional differential parameter estimation in ARFIMA models, *Computat. Stat. Data Anal.*, *51*, 1002–1011, doi:10.1016/j.csda.2005.10.007.

Stephenson, D. B., V. Pavan, and R. Bojariu (2000), Is the North Atlantic Oscillation a random walk?, *Int. J. Climatol.*, *20*, 1–18.

Syroka, J., and R. Toumi (2001), Scaling and persistence in observed and modeled surface temperature, *Geophys. Res. Lett.*, *28*, 3255–3258.

Taqqu, M. S., and V. Teverovsky (1997), Estimating long-range dependence in finite and infinite variance series, in *A Practical Guide to Heavy Tails: Statistical Techniques for Analyzing Heavy Tailed Distributions*, edited by R. F. R. Adler and M. S. Taqqu, Birkhuser, Boston, Mass.

Taqqu, M. S., V. Teverovsky, and W. Willinger (1995), Estimators for long-range dependence: An empirical study, *Fractals*, *3*(4), 785–798.

Trenberth, K. E., et al. (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*, chap. 3, Cambridge Univ. Press, New York.

Ventura, V., C. J. Paciorek, and J. S. Risbey (2004), Controlling the proportion of falsely rejected hypothesis when conducting multiple tests with climatological data, *J. Clim.*, *17*, 4343–4356.

Vyushin, D. I., and P. J. Kushner (2009), Power-law and long-memory characteristics of the atmospheric general circulation, *J. Clim.*, doi:10.1175/2008JCLI2528.1, in press.

Wang, W., P. H. A. J. M. V. Gelder, J. K. Vrijling, and X. Chen (2007), Detecting long-memory: Monte Carlo simulations and application to daily streamflow processes, *Hydrol. Earth Syst. Sci.*, *11*, 851–862.

Whittle, P. (1953), Estimation and information in stationarity time series, *Ark. Mat.*, *2*, 423–434.

Yue, S., and C. Y. Wang (2002), The applicability of pre-whitening to eliminate the influence of serial correlation on the Mann-Kendall test, *Water Resour. Res.*, *38*(6), 1068, doi:10.1029/2001WR000861.

Yue, S., and C. Y. Wang (2004), The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series, *Water Resour. Manage.*, *18*, 201–218.

Yue, S., P. Pilon, and G. Cavadias (2002a), Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series, *J. Hydrol.*, *259*, 254–271.

Yue, S., P. Pilon, B. Phinney, and G. Cavadias (2002b), The influence of autocorrelation on the ability to detect trend in hydrological series, *Hydrol. Processes*, *16*, 1807–1829.

S. M. Barbosa, Faculty of Science, University of Porto, Rua do Campo Alegre 687, P-4169-007 Porto, Portugal. (susana.barbosa@fc.up.pt)

E. Caporali and S. Fatichi, Department of Civil and Environmental Engineering, University of Firenze, Via Saint Marta 3, I-50139 Firenze, Italy. (enrica.caporali@unifi.it; simone.fatichi@dicea.unifi.it)

M. E. Silva, Faculty of Economics, University of Porto, Rua Dr. Roberto Frias, P-4200-464 Porto, Portugal. (mesilva@fep.up.pt)