

U.K. Rainfall Data: A Long-Term Persistence Approach

LUIS A. GIL-ALANA

University of Navarra, Faculty of Economics, Pamplona, Spain

(Manuscript received 14 September 2011, in final form 17 January 2012)

ABSTRACT

This paper looks at the analysis of U.K. monthly rainfall data from a long-term persistence viewpoint. Different modeling approaches are considered, taking into account the strong dependence and the seasonality in the data. The results indicate that the most appropriate model is the one that presents cyclical long-run dependence with the order of integration being positive though small, and the cycles having a periodicity of about a year.

1. Introduction

Time series modeling of hydrological and climatological time series data is a major issue in the planning, operating, and decision making concerning water resources and the investigation of climatic fluctuations. Also, it is important for data generation, forecasting, estimating missing data, and extending data records (Delleur et al. 1976; Salas 1993; Hipel and McLeod 1994).

Rainfall data are difficult to model given that they are not perfectly linear and are non Gaussian, present seasonally in some cases, and possess a degree of dependence across time. Throughout this paper we present a variety of model specifications that take into account all the above characteristics, focusing mainly on long-memory or long-term persistence processes. These processes were introduced in the hydrological community by Hurst (1951), who heuristically detected the presence of long-range dependence in the well-known series of annual minima of the Nile River. These processes are so called because they display a high degree of association between observations that are distant in time. Following Hurst's work, extensive research has been carried out to detect long-term persistence or long memory in hydrological data (Hipel and McLeod 1994; Montanari et al. 1996; Montanari and Rosso 1997; Pelletier and Turcotte 1997; Corduas and Piccolo 2006; Koutsoyiannis and Montanari 2007; Gil-Alana 2009).

In the context of rainfall data, long-memory models fit well in the drought hazard assessment. Bras and Rodriguez-Iturbe (1985) started with the discussion of long-range dependence in the context of droughts. They showed that standard autoregressive integrated moving average (ARIMA) models often underestimate the frequency of historical drought. The use of fractional models was also advocated by Booy and Lye (1989) for use in flood frequency analysis. Montanari et al. (1996) estimated an autoregressive fractionally integrated moving average (ARFIMA) model to the monthly and daily inflows of Lake Maggiore, Italy, finding that the ARFIMA models provide a much better fit than the traditional ARIMA models. Other papers dealing with long memory in rainfall data are Burlando et al. (1996), Brath et al. (2001), Brockwell and Chan (2006), Kantelhardt et al. (2006), and Gil-Alana (2004, 2009). On the other hand, several authors have found significant trends in rainfall data. Thus, for example, Camuffo (1984) used one of the longest precipitation records available in Italy to detect the possible presence of tendencies or cyclical patterns in its behavior and found evidence of wavy trends, not always in phase with the frequency trend. Montanari et al. (1996) analyzed monthly and annual rainfall data in several locations in Italy and found decreasing trends in the precipitation amounts observed in the last century. Using long-range dependence techniques these authors conclude, however, that some of the trends were statistically insignificant.

This paper innovates with respect to other articles in long-memory models for rainfall data in the use of a cyclical approach based on the Gegenbauer processes. We show that the U.K. monthly rainfall data can be well

Corresponding author address: Luis A. Gil-Alana, University of Navarra, Faculty of Economics, Edificio Biblioteca, Entrada Este, E-31080 Pamplona, Spain.
E-mail: alana@unav.es

described in terms of a cyclical precipitation(d) model with a positive though small differencing parameter d . From a physical perspective, development of rainfall systems depends on the local environmental conditions and if we consider synoptic scales can be traced back to previously occurred weather systems. Nevertheless, though the evidence of long memory has been verified in many environmental time series (e.g., Bunde et al. 2003a,b) its evidence in rainfall data is weak (Pelletier and Turcotte 1997).

In this paper we focus on the U.K. monthly rainfall data and estimate different long-range dependence models based on their degrees of seasonal and non-seasonal persistence. The possibility of linear trends is also investigated. The results indicate that U.K. rainfall data do not contain linear trends though present a persistent cyclical pattern of length of approximately one year, and with the effect of the shocks disappearing relatively fast in time. The outline of the article is as follows: Section 2 presents the statistical framework. Section 3 describes the data and the models employed. Section 4 displays the empirical results, and section 5 concludes the article.

2. The statistical framework

For the purpose of the present work, we define an I(0) process as a covariance stationary process with spectral density function $f(\lambda)$ that is positive and finite at any frequency. Alternatively, it can be defined in the time domain as a process such that the infinite sum of the autocovariances is finite. This includes a wide range of model specifications such as the white-noise case, the stationary autoregression (AR), moving average (MA), and stationary autoregressive moving average (ARMA) models. Usually, the I(0) condition is a prerequisite for statistical inference in time series analysis.

Within the I(0) context, the ARMA model is commonly employed in time series analysis, and in particular, the AR(1) model given by

$$u_t = \rho u_{t-1} + v_t, \quad t = 1, 2, \dots, \quad (1)$$

with $|\rho| < 1$, where ρ is the AR coefficient and v_t is white noise. This model has been widely employed in the hydrological and climatological community because of its relation with the stochastic first order differential equation. In case of seasonal data, (1) can be replaced by the seasonal AR(1) process:

$$u_t = \rho u_{t-s} + v_t, \quad t = 1, 2, \dots, \quad (2)$$

where s indicates the number of time periods per year (e.g., $s = 4$ with quarterly data, and $s = 12$ with monthly

data). A strong limitation of this approach is that the coefficient ρ in (2) is fixed across seasons (months) implying that the rainfall amount for a certain month depends solely on the rainfall amount of the same season of the previous year.

On the other hand, the series might be nonstationary, in the sense that the mean, the variance, or the autocovariance may change across time. In this context we can take ρ in (1) equal to 1, and u_t is said to be integrated of order 1 [and denoted by $u_t \sim I(1)$]. In such a case, the first differences $(1 - L)u_t = u_t - u_{t-1}$ are stationary I(0), and statistical inference must be relied on the first differenced process. In other words, u_t is said to be I(1) if

$$(1 - L)u_t = v_t, \quad t = 1, 2, \dots, \quad (3)$$

where L is the lag operator ($Lu_t = u_{t-1}$) and v_t is I(0) as defined above. If v_t is ARMA(p, q), then u_t is said to be an ARIMA($p, 1, q$) model.¹ Similarly, for the seasonal case, we can take seasonal first differences on u_t , such that

$$(1 - L^s)u_t = v_t, \quad t = 1, 2, \dots, \quad (4)$$

and v_t is I(0). There exist various test statistics for testing models like (3) (i.e., unit roots; e.g., Dickey and Fuller 1979; Phillips and Perron 1988; Elliot et al. 1996; Ng and Perron 2001) and (4) (seasonal unit roots; Dickey et al. 1984; Hylleberg et al. 1990; Beaulieu and Miron 1993) that have been widely employed in the time series literature. These methods, however, focus on the I(0) (stationarity) and I(1) (nonstationary) cases and do not take into account other more flexible alternatives such as the fractional ones that will be examined in this work.

The above models have been extended in recent years to the fractional case. The number of differences required to render a series stationary I(0) may not necessarily be an integer value (usually 1) but a fractional one. In this context, u_t is said to be I(d) if

$$(1 - L)^d u_t = v_t, \quad t = 1, 2, \dots, \quad (5)$$

where d can be any real value, $u_t = 0, t \leq 0$, and v_t is again I(0). Note that the polynomial on the left-hand side of (5) can be expanded, for all real d , as

$$(1 - L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^j = 1 - dL + \frac{d(d-1)}{2} L^2 - \dots$$

¹ See Salas et al. (1980) for a complete review of these models and their characteristics.

Thus, if d in (5) is an integer value, u_t will be a function of a finite number of past observations, while if d is a non-integer, u_t depends upon values of the time series in the distant past, and the higher the value of d , the higher the level of dependence between the observations.² Examples of applications of $I(d)$ models of form as in (5) in hydrological data are the papers of Montanari et al. (1996) and Montanari and Rosso (1997), and other applications of these models in meteorological and climatological data can be found in Bloomfield (1992), Smith (1993), Lewis and Ray (1997), Maraun et al. (2004), and Gil-Alana (2003, 2005).

If $d > 0$ in (5) u_t displays long-range dependence (LRD). We can provide two definitions of LRD, one in the time domain and the other in the frequency domain. The time domain definition of LRD states that given a covariance stationary process $(u_t, t = 0, \pm 1, \dots)$, with an autocovariance function $E[(u_t - Eu_t)(u_{t-j} - Eu_t)] = \gamma_j$, u_t displays LRD if

$$\lim_{T \rightarrow \infty} \sum_{j=-T}^T |\gamma_j|$$

is infinite. A frequency domain definition is as follows. Suppose that u_t has an absolutely continuous spectral distribution, so that it has a spectral density function, denoted by $f(\lambda)$, and defined as

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j \cos \lambda j, \quad -\pi < \lambda \leq \pi.$$

Then, u_t displays LRD if the spectral density function has a pole at some frequency λ in the interval $[0, \pi]$, that is,

$$f(\lambda) \rightarrow \infty \quad \text{as} \quad \lambda \rightarrow \lambda^*, \quad \lambda^* \in [0, \pi]. \quad (6)$$

Most of the empirical literature has concentrated on the case where the singularity or pole in the spectrum occurs at the zero frequency ($\lambda^* = 0$). In fact, the $I(d)$ model, defined as in (5), is characterized because the spectral density function is unbounded at the origin.³ However, there might be situations where the singularity or pole in the spectrum takes place at other frequencies. This is the case of the seasonal fractional process, which is basically an extension of model (5) to

² Though not explicitly mentioned, u_t in (5) admits an infinite AR representation, implying that u_t can be expressed in terms of all its past history.

³ These processes were introduced by Granger (1980, 1981) and Hosking (1981), and they were justified in terms of the aggregation of individual heterogeneous AR(1) processes by Robinson (1978) and Granger (1980).

$$(1 - L^s)^d u_t = v_t, \quad t = 1, 2, \dots, \quad (7)$$

where d may again be any real value, and its binomial expansion provides that, for all real d ,

$$(1 - L^s)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^{js} = 1 - dL^s + \frac{d(d-1)}{2} L^{2s} - \dots$$

Thus, if d in (7) is noninteger, u_t depends on past (multiple of s) values. In this case, the spectral density function is unbounded at the zero but also at the $s - 1$ seasonal frequencies. Models of this form have been examined by, among others, Porter-Hudak (1990), Ray (1993), Sutcliffe (1994), and Gil-Alana and Robinson (2001).

Another specification in the context of LRD models is the one where the spectrum contains a single pole at a frequency away from zero. In this case, the process still displays the property of LRD but the autocorrelations present a cyclical structure that decays very slowly. This is the case of the Gegenbauer processes defined as

$$(1 - 2 \cos w_r L + L^2)^d u_t = v_t, \quad t = 1, 2, \dots, \quad (7a)$$

where w_r and d are real values, and v_t is $I(0)$. For practical purposes we define $w_r = 2\pi r/T$, with $r = T/c$, and thus c will indicate the number of time periods per cycle, while r refers to the frequency that presents a pole or singularity in the spectrum of u_t (λ^*). Note that if $r = 0$ (or $c = 1$), the fractional polynomial in (7a) becomes $(1 - L)^{2d}$, which is the polynomial associated with the common case of fractional integration at the long-run or zero frequency.

Gray et al. (1989, 1994) showed that the polynomial in (7a) can be expressed in terms of the Gegenbauer polynomial, such that, denoting $\mu = \cos w_r$, for all $d \neq 0$,

$$(1 - 2\mu L + L^2)^{-d} = \sum_{j=0}^{\infty} C_{j,d}(\mu) L^j,$$

where $C_{j,d}(\mu)$ are orthogonal Gegenbauer polynomial coefficients recursively defined as

$$\begin{aligned} C_{0,d}(\mu) &= 1, \quad C_{1,d}(\mu) = 2\mu d, \\ C_{j,d}(\mu) &= 2\mu \left(\frac{d-1}{j} + 1 \right) C_{j-1,d}(\mu) \\ &\quad - \left(2 \frac{d-1}{j} + 1 \right) C_{j-2,d}(\mu), \quad j = 2, 3, \dots, \end{aligned}$$

[see, for instance, Magnus et al. (1966) and Rainville (1960) for further details on Gegenbauer polynomials]. Gray et al. (1989) showed that u_t in (7a) is (covariance) stationary if $d < 0.5$ for $|\mu = \cos w_r| < 1$ and if $d < 0.25$ for $|\mu| = 1$.

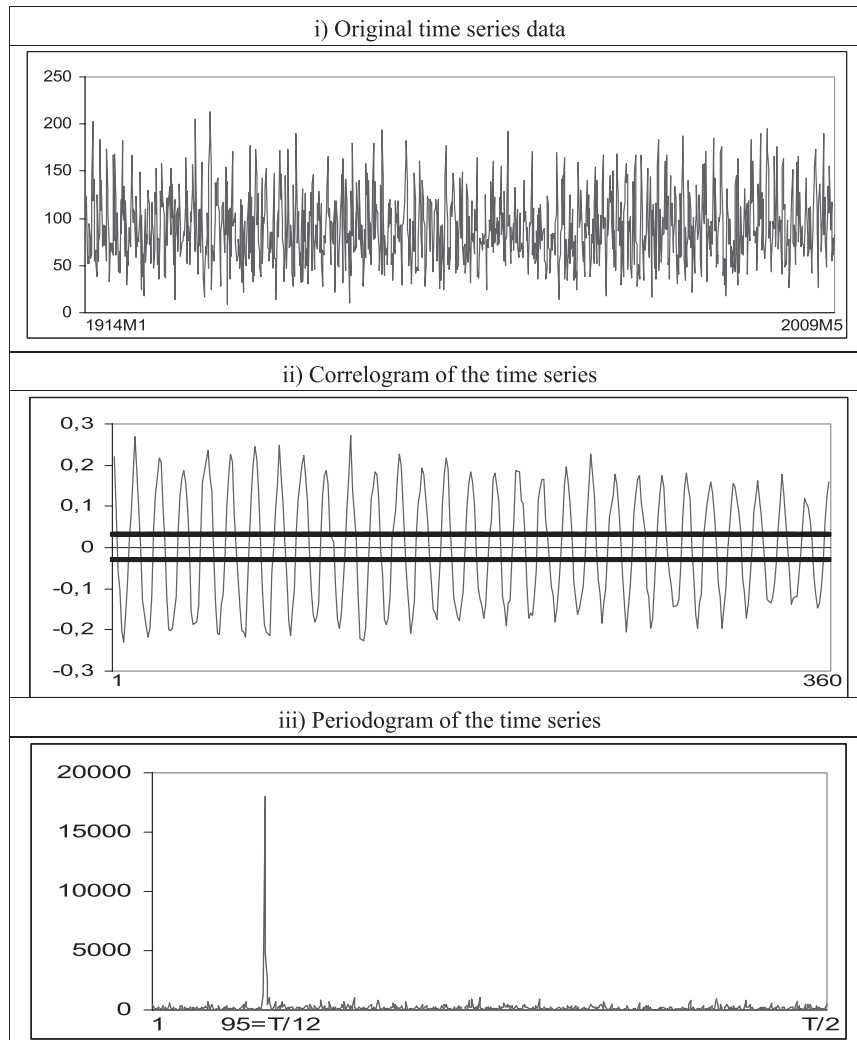


FIG. 1. (top to bottom) Original time series, the correlogram, and the periodogram.

This type of process described in (7a) was introduced by Anel (1986) and subsequently analyzed by Gray et al. (1989, 1994), Chung (1996a,b), Gil-Alana (2001), and Dalla and Hidalgo (2005) among many others.

All the above models display LRD, they are characterized because the spectral density is unbounded at one (or more) frequencies in the interval $[0, \pi]$, and they will all be employed in the following sections when analyzing the U.K. monthly rainfall data.

3. Data and models

The data employed in this article correspond to the monthly U.K. rainfall (mm) data, areal series, from January 1914 to December 2008. They were obtained from the Met Office website (<http://www.metoffice.gov.uk>).

Figure 1 displays the time series plot along with its corresponding sample correlogram and periodogram.

The data have an appearance of stationarity though with some type of seasonal (cyclical) pattern, which is also observed through the correlogram and the periodogram. In fact, the correlogram displays values with a cyclical structure that is decaying very slowly, which may be consistent, among many other specifications, with some of the LRD models described in section 2. Moreover, the periodogram presents a large peak at a frequency away from zero, giving further support in favor of the cyclical $I(d)$ model described by (7a). Although the periodogram is not a consistent estimator of the spectral density function, it is an asymptotically unbiased estimator of it and, evaluated at the discrete Fourier frequencies $\lambda_j = 2\pi j/T$, $j = 1, 2, \dots, T/2$, can give us an indication about the length of the cycles. In our case, we observe that the highest peak takes place at $\lambda_j^* = 2\pi 95/T$, which implies that the cycles have a periodicity of $T/95 = 1140/95 = 12$ periods (=months)/cycle.

In what follows, we assume that y_t is the time series we observe (U.K. monthly rainfall data), initially we suppose that it contains deterministic terms such as an intercept and/or a linear time trend,⁴

$$y_t = \beta_0 + \beta_1 t + x_t, \quad t = 1, 2, \dots, \quad (8)$$

and, based on the above evidence, we examine the following models for the detrended series:

- 1) a seasonal AR(1) process

$$y_t = \beta_0 + \beta_1 t + x_t; \quad x_t = \rho x_{t-12} + \varepsilon_t; \quad (8a)$$

- 2) an $I(d)$ process with seasonal AR(1) disturbances

$$y_t = \beta_0 + \beta_1 t + x_t; \quad (1-L)^d x_t = u_t; \\ u_t = \rho u_{t-12} + \varepsilon_t; \quad (8b)$$

- 3) a seasonal $I(d)$ model with AR(1) disturbances

$$y_t = \beta_0 + \beta_1 t + x_t; \quad (1-L^{12})^d x_t = u_t; \\ u_t = \rho u_{t-1} + \varepsilon_t; \quad (8c)$$

and

- 4) a cyclical $I(d)$ model with white noise (8d) and seasonal AR(1) (8d') disturbances, that is,

$$y_t = \beta_0 + \beta_1 t + x_t; \quad (1 - 2 \cos \omega_r L + L^2)^d x_t = \varepsilon_t \quad (8d)$$

and

$$y_t = \beta_0 + \beta_1 t + x_t; \quad (1 - 2 \cos \omega_r L + L^2)^d x_t = u_t; \\ u_t = \rho u_{t-12} + \varepsilon_t, \quad (8d')$$

in all cases with white noise ε_t . Note that model (8a) imposes an $I(0)$ structure for the detrended time series x_t ; model (8b) is more general, allowing for fractional integration or $I(d)$, and including model (8a) as a particular case when $d = 0$. In this context, the parameter d describes the long-run persistence of the series while ρ measures the short-run seasonal behavior. Model (8c) also uses fractional integration, though now d describes the long-run seasonal evolution of the series, while ρ measures the short-run dynamics.⁵ Finally, in models

(8d) and (8d') the parameter d indicates the cyclical long-range persistence and the short-run dynamics are described in terms of a seasonal AR process in (8d') to incorporate potential periodicities. Nonseasonal AR disturbances were also considered in this model and they were discarded since we do not find statistically significant cases.

The inclusion of intercepts and linear trends may appear unrealistic in the context of the present research. However, these terms disappear when interacting with the LRD polynomials. Note that the model $y_t = \beta_0 + \beta_1 t + x_t$, $\rho(L; d)x_t = u_t$, where $\rho(L; d)$ is the LRD polynomial, can be expressed as $y_t^* = \beta_0 l_t^* + \beta_1 t_t^* + u_t$, where $y_t^* = \rho(L; d)y_t$; $l_t^* = \rho(L; d)l_t$; and $t_t^* = \rho(L; d)t_t$ and the time trend asymptotically disappears for $d > 0$. In any case, the presence of deterministic trends in rainfall data has been observed in some regions (e.g., Camuffo 1984; Montanari et al. 1996), which could be due to increasing moisture content in the atmosphere associated with anthropogenic global warming and/or increasing rainfall from severe weather systems. The significance of these deterministic terms will be examined by means of the t values noting that the residuals in the differenced processes are supposed to be well behaved $I(0)$.⁶

The LRD models (8b)–(8d') were estimated using the Whittle function in the frequency domain (see Dahlhaus 1989) and also using a very general testing procedure derived by Robinson (1994) that permits us to consider each of the above specifications as particular cases of interest. This method is based on the Lagrange multiplier (LM) principle and tests the null hypothesis $H_0: d = d_0$ in (8b)–(8d') for any real value d_0 , thus encompassing stationary ($d_0 < 0.5$) and nonstationary ($d_0 \geq 0.5$) hypotheses. Moreover, this method is the most efficient one in the context of fractional integration under the assumption of Gaussianity of the error term.⁷ The functional form of this method can be found in any of the numerous empirical applications based on this approach (Gil-Alana and Robinson 1997, 2001; Gil-Alana 2001).

4. Empirical results

We start by estimating model (8a), the simple seasonal AR(1) process. We display in Table 1 the estimates for the three cases of no deterministic terms [i.e., $\beta_0 = \beta_1 = 0$]

⁴ Seasonal dummy variables were not considered given the lack of systematic patterns in the monthly observations (see Fig. 2).

⁵ These two models [(8b) and (8c)] may look inappropriate in view of the correlogram and periodogram of the data. However, in these two plots displayed in Fig. 1, seasonality may be obscuring other important features of the data.

⁶ Some authors eliminate trends in the data by using techniques such as wavelets (Arneodo et al. 1996) or detrended fluctuation analysis (DFA; Bunde et al. 2003a). See also Yue et al. (2010) for applications of these techniques in precipitation.

⁷ Note, however, that Gaussianity is not a requirement in this procedure with a moment condition of only order 2 required.

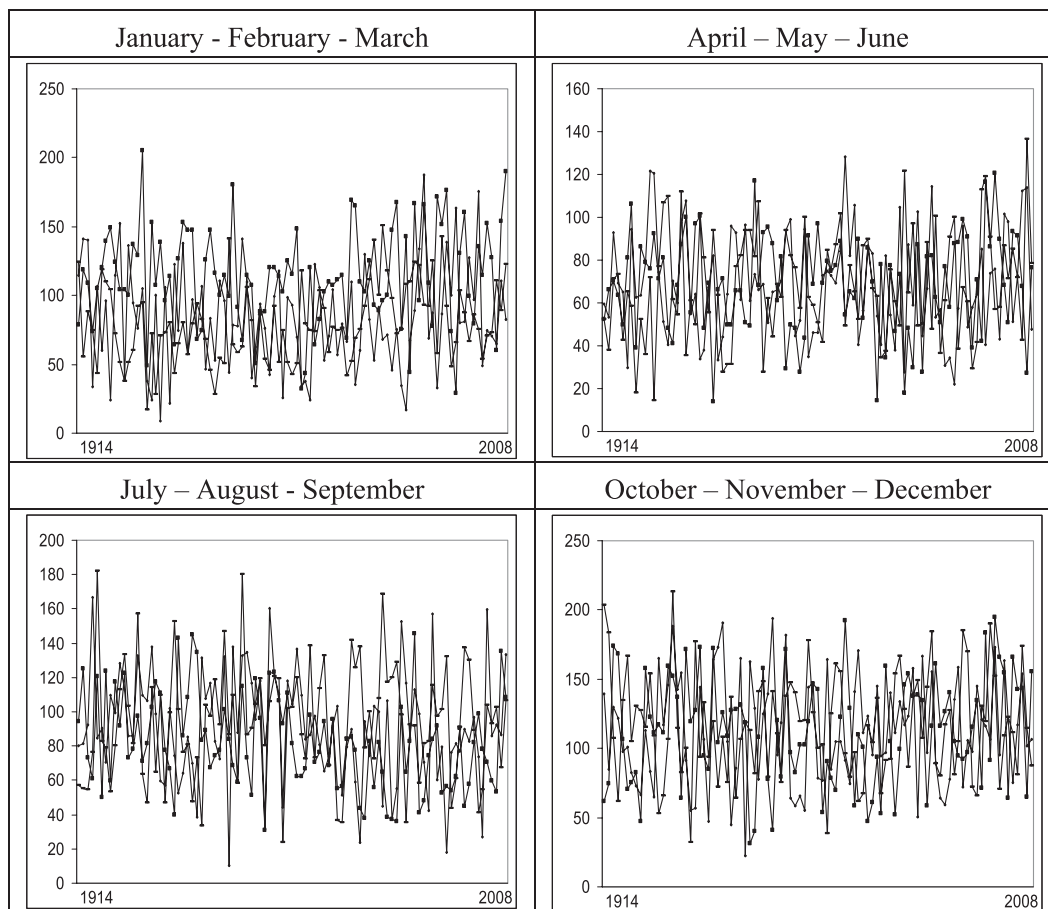


FIG. 2. Original time series for three-month groupings of January–March, April–June, July–September, and October–December for the years from 1914 to 2008.

a priori in (8)], an intercept (β_0 unknown and $\beta_1 = 0$), and an intercept with a linear time trend (β_0 and β_1 unknown). In the latter case, the trend coefficient appears statistically insignificantly different from zero, while the intercept is significant. In the three cases, the AR coefficient is found to be around 0.27. As earlier mentioned, this model seems unrealistic since it imposes the same value ρ across months.⁸

Next, we allow for the possibility of fractional integration at the zero frequency and consider model (8b) as a potential specification. The results—again for the three cases of no regressors, an intercept, and an intercept with a linear time trend—are displayed in Table 2. We see here that the estimated value of d is significantly positive in the three cases, therefore ruling out model (8a), which would

correspond to the case of $d = 0$. If no regressors are considered, the estimated value of d is equal to 0.138, and including an intercept and/or a linear time trend, the value is slightly smaller (0.076), though still statistically significant at the 5% level. The t values of the deterministic terms indicate that a time trend is not required, while the intercept appears statistically significant. The seasonal AR coefficient is about 0.25 for the three cases. This specification is fairly realistic, with the current value being a function of all its past history, and higher weights are given to the values in the same month as the current one in previous years.⁹

Next, we consider model (8c), that is, the long-run behavior is now described through the seasonal fractional differencing parameter, d , while the short-run dynamic is described by means of a nonseasonal AR(1) process. The results for this case are reported in Table 3. We observe

⁸ Performing AR(1) models for each month, the estimated parameters were $-0.03, -0.05, 0.27, 0.03, 0.08, 0.06, 0.12, -0.07, 0.05, 0.08, -0.00,$ and 0.10 , clearly showing a distinct pattern across months.

⁹ As earlier mentioned, though long-range persistence exists in temperature, its evidence is weak in rainfall data.

TABLE 1. Coefficient estimates in a monthly seasonal AR(1) process. The model is $y_t = \beta_0 + \beta_1 t + u_t$; $u_t = \rho u_{t-12} + \varepsilon_t$. The selected model is in boldface font.

	No regressors	An intercept	A linear time trend
Intercept (<i>t</i> value)	—	90.7942 (82.634)	89.0708 (40.506)
Time trend (<i>t</i> value)	—	—	0.003 02 (0.904)
Monthly seasonal AR	0.270	0.270	0.270

that the seasonal differencing parameter is slightly negative in the three cases (about -0.004) and the $I(0)$ hypothesis cannot be rejected at conventional statistical levels. Here, the estimated AR(1) coefficient is about 0.22. As in the previous case, the time trend seems not to be required while the intercept is significantly different from zero. Note that in this model the current value of the series is also a function of all its past history though the autocorrelations are now decaying hyperbolically with respect to the seasonal monthly structure, and exponentially faster for the remaining values.

In Tables 4 and 5, we consider an alternative specification given by (8d) and (8d'). Here we suppose that the disturbances are white noise (in Table 4) and seasonal AR(1) (in Table 5). According to these two models, the U.K. monthly rainfall data follow a cyclical $I(d)$ process with a periodicity of 12 months for the two cases of uncorrelated and AR(1) errors.

The results are similar in the two cases. The cyclical order of integration d is positive though small, the time trends are insignificant, while the intercepts are statistically different from zero. If seasonal autoregressions are allowed, the coefficient is found to be about 0.24. Noting that model (8d) is nested in (8d') we performed a likelihood ratio (LR) test on these two specifications, testing the null of $\rho = 0$ in (8d'). The result of the test statistic (1.391) indicates that the model with white noise disturbances [i.e., (8d)] is preferred.

According to the above specifications, and noting that model (8a) is rejected in favor of (8b), and (8d) is preferred to (8d'), we consider the following three potential models to describe the persistent behavior of the series:

$$M1: y_t = 90.9472 + x_t; \quad (1 - L)^{0.076} x_t = u_t; \\ u_t = 0.252u_{t-12} + \varepsilon_t,$$

$$M2: y_t = 90.7618 + x_t; \quad (1 - L^{12})^{-0.004} x_t = u_t; \\ u_t = 0.223u_{t-1} + \varepsilon_t,$$

TABLE 2. Coefficient estimates in an $I(d)$ model with seasonal AR(1) disturbances. The model is $y_t = \beta_0 + \beta_1 t + u_t; (1 - L)^d x_t = u_t; u_t = \rho u_{t-12} + \varepsilon_t$. The selected model is in boldface font.

	No regressors	An intercept	A linear time trend
Intercept (<i>t</i> value)	—	90.9472 (50.542)	89.0555 (26.760)
Time trend (<i>t</i> value)	—	—	0.003 33 (0.675)
d (95% interval)	0.138 (0.027, 0.218)	0.076 (0.020, 0.141)	0.076 (0.019, 0.141)
Monthly seasonal AR	0.252	0.252	0.252

and

$$M3: y_t = 90.7825 + x_t; \quad (1 - 2 \cos w_{12} L + L^2)^{0.014} x_t = \varepsilon_t.$$

Remember that model M1 refers to the case of fractional integration or $I(d)$ behavior (at the long-run or zero frequency) with seasonal AR disturbances, and model M2 refers to seasonal fractional integration with non-seasonal AR disturbances, while model M3 is the one related to cyclical fractional integration of the form based on the Gegenbauer processes. We examined the residuals of the three selected models based on diagnostic tests of serial correlation and functional form, and the results strongly support the cyclical $I(d)$ model M3.

Finally, we compare the above models in terms of their forecasting performance by means of an in-sample forecasting experiment. Standard measures of forecast accuracy are the following: the mean absolute percentage error (MAPE), the mean-square error (MSE), the root-mean-square error (RMSE), the root-mean-percentage-square error (RMPSE), and mean absolute deviation (MAD). Some of these measures were implemented and, though not reported because of limited space, model M3 produced the best results in the majority of the cases. On the other hand, it may be argued that the above measures are purely descriptive devices. There exist now statistical tests for comparing different

TABLE 3. Coefficient estimates in a seasonal $I(d)$ model with AR(1) disturbances. The model is $y_t = \beta_0 + \beta_1 t + u_t; (1 - L^{12})^d x_t = u_t; u_t = \rho u_{t-1} + \varepsilon_t$. The selected model is in boldface font.

	No regressors	An intercept	A linear time trend
Intercept (<i>t</i> value)	—	90.7618 (83.385)	88.8979 (40.115)
Time trend (<i>t</i> value)	—	—	0.003 23 (0.965)
Seasonal d_s (95% interval)	-0.004 (-0.029, 0.033)	-0.004 (-0.031, 0.025)	-0.004 (-0.031, 0.025)
AR(1) coef	0.223	0.223	0.223

TABLE 4. Coefficient estimates in a cyclical $I(d)$ model with white-noise disturbances. The model is $y_t = \beta_0 + \beta_1 t + u_t; (1 - 2 \cos \omega_{12} L + L^2)^d x_t = \varepsilon_t$. The selected model is in boldface font.

	No regressors	An intercept	A linear time trend
Intercept	—	90.782 50	89,027 94
(<i>t</i> value)		(82.579)	(40.371)
Time trend	—	—	0.003 07
(<i>t</i> value)			(0.917)
Cyclical <i>d</i> (95% interval)	0.017 (−0.015, 0.048)	0.014 (0.007, 0.025)	0.014 (0.008, 0.034)
<i>R</i>	12	12	12

forecasting models. One of these tests, widely employed in the time series literature, is the asymptotic test for a zero expected loss differential of Diebold and Mariano (1995).¹⁰ However, for a given covariance stationary sample realization, $(d_t)_{t=T+h, \dots, T+n}$, Harvey et al. (1997) noted that the Diebold–Mariano test statistic could be seriously oversized as the lead time h increases. These authors provided a modified Diebold–Mariano test statistic given by

$$M - DM = DM \sqrt{\frac{n + 1 - 2h + h(h - 1)/n}{n}}$$

where DM is the original Diebold–Mariano statistic.

Based on the $M - DM$ test, we make pairwise comparisons between the three models using the RMSE and the MAPE as the criteria employed, and for $h = 12, 24, 36,$ and 48 in a 48-period (in sample) horizon. We report in Table 6 the values of the $M - DM$ statistic for the case of the RMSE and $h = 48$ clearly observing the superiority of the model M3 over the other two. Similar results were obtained with other criteria and (lead) time horizons.

5. Concluding comments

In this paper we have examined long-range persistence in the U.K. monthly rainfall data by using a variety of model specifications to describe the long-run dependence in the data. Specifically, we used the following three models: 1) a fractionally integrated or $I(d)$ model with seasonal AR disturbances, 2) a seasonally fractionally integrated model with nonseasonal AR disturbances, and 3) a cyclically fractionally integrated model. This latter model is based on the Gegenbauer processes and assumes a cyclical structure in the autocorrelations decaying slowly

¹⁰ An alternative approach is the bootstrap-based test of Ashley (1998), though this method is computationally more intensive.

TABLE 5. Coefficient estimates in a cyclical $I(d)$ model with seasonal AR(1) disturbances. The model is $y_t = \beta_0 + \beta_1 t + u_t; (1 - 2 \cos \omega_{12} L + L^2)^d x_t = u_t; u_t = \rho u_{t-12} + \varepsilon_t$. The selected model is in boldface font.

	No regressors	An intercept	A linear time trend
Intercept	—	90.7826	89.024 13
(<i>t</i> value)		(82.657)	(40.407)
Time trend	—	—	0.003 07
(<i>t</i> value)			(0.920)
Cyclical <i>d</i> (95% interval)	0.015 (−0.019, 0.047)	0.010 (0.004, 0.022)	0.009 (0.003, 0.030)
Seasonal AR coef	0.231	0.247	0.249
<i>R</i>	12	12	12

to zero. This process is also characterized because the spectral density function is unbounded at a nonzero frequency. Note that all these models imply long-run persistence, which is a commonly observed feature in hydrological and meteorological data. Also, the specifications employed in this work nest standard models widely employed in modeling rainfall data, including stationarity and nonstationary models, with and without deterministic trends, with short-memory and long-memory dependencies, and, in the latter case allowing the pole in the spectrum to occur at zero and nonzero frequencies. Our results clearly reject the hypothesis of short memory or $I(0)$ stationarity in favor of long-memory models. Also, diagnostic tests carried out on the residuals along with an in-sample forecasting experiment support the view that the cyclical $I(d)$ model is the most correct specification for this series. This result is consistent with the peak observed in the periodogram of the data at the $T/12$ frequency, suggesting cycles of length of approximately one year. The estimated order of integration is found to be positive though small, implying that the autocorrelations cyclically decay relatively fast (though hyperbolically) to zero.

With respect to the deterministic components (in particular, the intercepts that are found to be statistically significant) it may be argued that they do not have a physical meaning in the context of rainfall data. However, as we argue earlier in this work, they disappear in the long run as the fractional differencing parameter is found to be strictly higher than 0. Moreover,

TABLE 6. Pairwise comparisons between the models using $M - DM$ statistic (RMSE, $h = 48$).

Models	$M - DM$ statistic	Selected model
M1 vs M2	−3.014	Model M1
M1 vs M3	2.905	Model M3
M2 vs M3	4.011	Model M3

we also conducted the analysis based on the mean-subtracted series and, though not reported, the estimated coefficients of the LRD coefficients, were practically the same as those obtained with the original data. In particular, with respect to model M3 the estimated value of d was found to be 0.017, slightly higher than the one reported in this work (0.014). The lack of linear trends in the data, at least for the U.K. case examined in this work, is another distinguishing feature in this paper, especially noting that linear trends are common in precipitation in many regions (see, e.g., Castañeda and Barros 1994, for Argentina; Trömel and Schönwiese 2008, for Germany; Yazdani et al. 2011, for Iran). Our results strongly support the view that there are no linear trends in the U.K. rainfall data. Nonlinear trends of the form employed by authors such as Karl et al. (2000) for global temperatures and Minetti et al. (2003) for Argentina and Chile will be examined in future papers.

The forecasting experiment conducted in this work also indicates that the cyclically fractionally integrated model outperforms noncyclical long-memory models, while these latter models are also found to be superior to others based on integer degrees of differentiation. How our cyclically $I(d)$ model predicts in out-of-sample contexts is a matter of research that will be investigated in future papers.

As a final remark, it is of interest to show whether the same result holds in other monthly rainfall data, noting that if that were the case, those specifications based on fractional (or integer) degrees of integration at the long-run or zero frequency would produce invalid results based on biased estimates of the fractional differencing parameter. Work in this direction is now in progress.

Acknowledgments. The author gratefully acknowledges financial support from the Ministerio de Ciencia y Tecnología (ECO2008-03035 ECON Y FINANZAS, Spain) and from a PIUNA Project at the University of Navarra. Comments from two anonymous referees and the editor of the journal are gratefully acknowledged.

REFERENCES

- Andel, J., 1986: Long memory time series models. *Kybernetika*, **22**, 105–123.
- Arneodo, A., Y. d'Auberton-Carafa, E. Bacry, P. V. Graves, J. F. Muzy, C. Thermes, 1996: Wavelet based fractal analysis of DNA sequences. *Physica D*, **96**, 291–320.
- Ashley, R., 1998: A new technique for postsample model selection and validation. *J. Econ. Dyn. Control*, **22**, 647–665.
- Beaulieu, J. J., and A. Miron, 1993: Seasonal unit roots in aggregate U.S. data. *J. Econom.*, **55**, 305–328.
- Bloomfield, P., 1992: Trends in global temperatures. *Climatic Change*, **21**, 275–287.
- Booy, C., and L. M. Lye, 1989: A new look at flood risk determination. *Water Resour. Bull.*, **25**, 933–943.
- Bras, R. L., and I. Rodriguez-Iturbe, 1985: *Random Functions in Hydrology*. Addison-Wesley, 559 pp.
- Brath, A., A. Castellarin, and A. Montanari, 2001: At site and regional assessment of the possible presence of non-stationarity in extreme rainfall in Northern Italy. *Phys. Chem. Earth*, **26**, 705–710.
- Brockwell, A. E., and N. H. Chan, 2006: Long memory dynamic Tobit model. *J. Forecasting*, **25**, 351–367.
- Bunde, A., J. Eichner, R. Govindan, S. Havlin, E. Koscielny-Bunde, D. Rybski, and D. Vjushin, 2003a: Power-law persistence in the atmosphere: Analysis and application. *Nonextensive Entropy—Interdisciplinary Applications*, M. Gell-Mann and C. Tsallis, Eds., Oxford University Press, 207–220.
- , —, S. Havlin, and J. W. Kantelhardt, 2003b: The effect of long term correlations on the statistics of rare events. *Physica A*, **330**, 1–7.
- Burlando, P., A. Montanari, and R. Rosso, 1996: Modelling hydrological data with and without long memory. *Mechanica*, **31**, 87–101.
- Camuffo, D., 1984: Analysis of the series of precipitation at Padova, Italy. *Climatic Change*, **5**, 55–77.
- Castañeda, E., and V. Barros, 1994: Las tendencias de la precipitación en el cono sur de América al este de los Andes (The precipitation tendencies in the southern cone of America to the east of the Andes). *Meteorológica*, **19**, 23–32.
- Chung, C.-F., 1996a: A generalized fractionally integrated autoregressive moving-average process. *J. Time Ser. Anal.*, **17**, 111–140.
- , 1996b: Estimating a generalized long memory process. *J. Econom.*, **73**, 237–259.
- Corduas, M., and D. Piccolo, 2006: Short and long memory unobserved components in hydrological time series. *Phys. Chem. Earth*, **31**, 1099–1106.
- Dahlhaus, R., 1989: Efficient parameter estimation for self-similar process. *Ann. Stat.*, **17**, 1749–1766.
- Dalla, V., and J. Hidalgo, 2005: A parametric bootstrap test for cycles. *J. Econom.*, **129**, 219–261.
- Delleur, J. W., P. C. Tao, and M. L. Kavass, 1976: An evaluation of practicality and complexity of some rainfall and runoff time series models. *Water Resour. Res.*, **12**, 953–970.
- Dickey, D. A., and W. A. Fuller, 1979: Distributions of the estimators for autoregressive time series with a unit root. *J. Amer. Stat. Assoc.*, **74**, 427–431.
- , D. P. Hasza, and W. A. Fuller, 1984: Testing for unit roots in seasonal time series. *J. Amer. Stat. Assoc.*, **79**, 355–367.
- Diebold, F. X., and R. S. Mariano, 1995: Comparing predictive accuracy. *J. Bus. Econ. Stat.*, **13**, 253–263.
- Elliot, G., T. Rothenberg, and J. H. Stock, 1996: Efficient tests for an autoregressive unit root. *Econometrica*, **64**, 813–836.
- Gil-Alana, L. A., 2001: Testing stochastic cycles in macroeconomic time series. *J. Time Ser. Anal.*, **22**, 411–430.
- , 2003: An application of fractional integration to a long temperature time series. *Int. J. Climatol.*, **23**, 1699–1710.
- , 2004: A fractionally integrated model for some Indian monthly rainfall data. *J. Indian Inst. Sci.*, **84**, 1–20.
- , 2005: Statistical model of the temperatures in the Northern Hemisphere using fractional integration techniques. *J. Climate*, **18**, 5357–5369.
- , 2009: Modelling Australian annual mean rainfall data: A new approach based on fractional integration. *Aust. Meteor. Ocean. J.*, **58**, 119–128.

- , and P. M. Robinson, 1997: Testing of unit roots and other nonstationary hypotheses in macroeconomic time series. *J. Econom.*, **80**, 241–268.
- , and —, 2001: Testing of seasonal fractional integration in the UK and Japanese consumption and income. *J. Appl. Econ.*, **16**, 95–114.
- Granger, C. W. J., 1980: Long memory relationships and the aggregation of dynamic models. *J. Econom.*, **14**, 227–238.
- , 1981: Some properties of time series data and their use in econometric model specification. *J. Econom.*, **16**, 121–130.
- Gray, H. L., N. Zhang, and W. A. Woodward, 1989: On generalized fractional processes. *J. Time Ser. Anal.*, **10**, 233–257.
- , —, and —, 1994: On generalized fractional processes—A correction. *J. Time Ser. Anal.*, **15**, 561–562.
- Harvey, D. I., S. J. Leybourne, and P. Newbold, 1997: Testing the equality of prediction mean squared errors. *Int. J. Forecasting*, **13**, 281–291.
- Hipel, K. W., and A. E. McLeod, 1994: *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, 1013 pp.
- Hosking, J. R. M., 1981: Fractional differencing. *Biometrika*, **68**, 168–176.
- Hurst, H. E., 1951: Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil*, **116**, 770–779.
- Hylleberg, S., R. F. Engle, C. W. J. Granger, and B. S. Yoo, 1990: Seasonal integration and cointegration. *J. Econom.*, **44**, 215–238.
- Kantelhardt, J. W., E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Haulin, 2006: Long-term persistence and multifractality of precipitation and river runoff records. *J. Geophys. Res.*, **111**, D01106, doi:10.1029/2005JD005881.
- Karl, T. R., R. W. Knight, and B. Baker, 2000: The record breaking global temperatures of 1997 and 1998. Evidence for an increase in the rate of global warming? *Geophys. Res. Lett.*, **27**, 719–722.
- Koutsoyiannis, D., and A. Montanari, 2007: Statistical analysis of hydroclimatic time series: Uncertainty and insights. *Water Resour. Res.*, **43**, W05429, doi:10.1029/2006WR005592.
- Lewis, P. A. W., and B. K. Ray, 1997: Modelling long-range dependence, nonlinearity and periodic phenomena in sea surface temperatures using TSMARS. *J. Amer. Stat. Assoc.*, **92**, 881–893.
- Magnus, W., F. Oberhettinger, and R. P. Soni, 1966: *Formulas and Theorems for the Special Functions of Mathematical Physics*. Springer, 508 pp.
- Maraun, D., H. W. Rust, and H. Timmer, 2004: Tempting long memory on the interpretation of DFA results. *Nonlinear Processes Geophys.*, **11**, 495–503.
- Minetti, J. L., W. M. Vargas, A. G. Poblete, L. R. Acuña, and G. Casagrande, 2003: Non-linear trends and low frequency oscillations in annual precipitation over Argentina and Chile, 1931–1999. *Atmósfera*, **16**, 119–135.
- Montanari, A., and R. Rosso, 1997: Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation. *Water Resour. Res.*, **33**, 1035–1044.
- , —, and M. Taqqu, 1996: Some long-run properties of rainfall records in Italy. *J. Geophys. Res.*, **101**, 431–438.
- Ng, S., and P. Perron, 2001: Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, **69**, 1519–1554.
- Pelletier, J. D., and D. L. Turcotte, 1997: Long range persistence in climatological and hydrological time series: Analysis, modelling and applications to drought hazard assessment. *J. Hydrol.*, **203**, 198–208.
- Phillips, P. C. B., and P. Perron, 1988: Testing for a unit root in a time series regression. *Biometrika*, **75**, 335–346.
- Porter-Hudak, S., 1990: An application of the seasonal fractionally differenced model to the monetary aggregate. *J. Amer. Stat. Assoc.*, **85**, 338–344.
- Rainville, E. D., 1960: *Special Functions*. MacMillan, 365 pp.
- Ray, B. K., 1993: Long range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model. *Int. J. Forecasting*, **9**, 255–269.
- Robinson, P. M., 1978: Statistical inference for a random coefficient autoregressive model. *Scand. J. Stat.*, **5**, 163–168.
- , 1994: Efficient tests of nonstationary hypotheses. *J. Amer. Stat. Assoc.*, **89**, 1420–1437.
- Salas, J. D., 1993: Analysis and modelling of hydrological time series. *Handbook of Hydrology*, D. R. Maidment, Ed., McGraw Hill, 19.1–19.72.
- , J. W. Delleur, V. M. Yevjevich, and M. L. Lane, 1980: *Applied Modeling of Hydrologic Time Series*. Water Resources Publications, 484 pp.
- Smith, R. L., 1993: Long-range dependence and global warming. *Statistics for the Environment*, V. Barnett, and K. F. Turkman, Eds., J. Wiley, 141–161.
- Sutcliffe, A., 1994: Time series forecasting using fractional differencing. *J. Forecasting*, **13**, 383–393.
- Trömel, S., and C. D. Schönwiese, 2008: Robust trend estimation of observed German precipitation. *Theor. Appl. Climatol.*, **93** (1–2), 107–115.
- Yazdani, M. R., K. D. Javad, M. Mohammad, and S. Ashish, 2011: Trend detection of the rainfall and air temperature data in the Zayandehrud Basin. *J. Appl. Sci.*, **11**, 2125–2134.
- Yue, J., Y. Zhao, and P. Shang, 2010: Effects of trends in detrended fluctuation analysis of precipitation series. *Math. Probl. Eng.*, **2010**, 749894, doi:10.1155/2010/749894.