J. Indian Inst. Sci., Jan.–Apr. 2004, **84**, 1–20
© Indian Institute of Science.

A fractionally integrated model for some Indian monthly rainfall data

L. A. GIL-ALANA

University of Navarre, Campus Universitario, Faculty of Economics, Edificio Biblioteca, Entrada Este, E-31080 Pamplona, Spain.
email: alana@unav.es; Phone: +34-948 425 625; Fax: +34-948 425 626.

Abstract

We examine some Indian monthly meteorological data by means of fractional integration. The results show that long memory is present in the monthly structure of various rainfall data. Moreover, they are homogeneous across the regions, with the values of d ranging between 0.25 and 0.75. Attempting to summarize the conclusions for the individual months, the degree of dependence between the observations during May–September seems to be higher than for the remaining months.

Keywords: Fractional integration, long memory.

1. Introduction

Time-series analysis has been applied to many situations in recent years, including several applications in water-related areas such as stream flow modelling [1], event rainfall data in semi-arid climates [2], detection of climate changes [3], water quality analysis [4], rainfall storm flow assessment [5], etc. In this study, an attempt has been made to apply time-series analysis to some Indian monthly rainfall data for the time period 1871–1999. The importance of the Indian rainfall data when modelling and forecasting the monsoon rainfall at different spatial and temporal scales has been in vogue for nearly a century. The idea is to develop suitable mathematical models in order to get a better understanding of its behaviour. Broadly speaking, these models may be classified as empirical or dynamical. The present work deals with the empirical models in the sense that it will be based on the variability of past observations. A basic feature of rainfall data is its non-gaussianness across different temporal and spatial scales. However, most of the statistical techniques, usually employed, require gaussianity in order to make statistical valid inference. In this paper, we use a methodology that, though based on the likelihood function, does not require gaussianity; a moment condition only of order 2 is required.

Since excellent reviews of the empirical models used for prediction of Indian rainfall are available [6–9], we only mention a few important facets here. A large number of potential predictors have been used in the analysis of these data, including factors such as El Niño, southern oscillation, snow over the Himalayas and Eurasia, and some global and regional conditions on spatial scales. Additionally, in the last two decades, new techniques based on auto-regressive moving average (ARMA) models [10], power (nonlinear) regression models

[11, 12], dynamic stochastic transfer models [13], as well as neural network models [14, 15], have been used, and a model that utilizes 16 parameters to provide qualitative predictions on the basis of the fraction of favourable parameters can be found in Gowariker *et al.* [11, 12]. On the other hand, the neural network model [14] uses only information on past history of rainfall data. The present paper deals with the latter approach in the sense that we use a univariate model, based on past information, following the line of research based on '*let the data speak for themselves*'.

We focus on the long memory property of the data and, in particular, on the fractional differencing parameter in some monthly rainfall data corresponding to several regions in India. For this purpose, we use a parametric testing procedure, proposed by Robinson [16], that has several distinguishing features compared with other methods. Thus, Robinson's method permits us to test unit and/or fractional roots at zero and the seasonal frequencies. The tests have standard null and local limit distributions, and this standard behaviour holds independent of the way of modelling of $I(0)$ disturbances.

For the purpose of the present paper, we define an $I(0)$ process $\{u_t, t = 0, \pm 1, ...\}$ as a covariance stationary process with spectral density function that is positive and finite at the zero frequency. In this context, we say that a given raw time series $\{x_t, t = 0, \pm 1, ...\}$ is $I(d)$ if:

$$(1-L)^d \quad x_t = u_t, \qquad t = 1, 2, \dots, \tag{1}$$

$$x_t = 0, \quad t \le 0,$$

where u_t is $I(0)$ and L, the lag operator ($Lx_t = x_{t-1}$). Note that the polynomial above can be expressed in terms of its binomial expansion, such that for all real d,

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

The literature has usually stressed the cases of $d = 0$ and 1, however, d can be any real number. Clearly, if $d = 0$ in (1), $x_t = u_t$, and a 'weakly autocorrelated' x_t is allowed for. However, if $d > 0$, x_t is said to be a long memory process, also called 'strongly autocorrelated', and so-named because of the strong association between observations widely separated in time. As d increases beyond 0.5 and through 1, x_t can be viewed as becoming 'more nonstationary', in the sense, for example, that the variance of partial sums increases in magnitude. (Models with d ranging between –0.5 and 0 are short memory and have been addressed as anti-persistent [17], because the spectral density function is dominated by high-frequency components). Fractional processes were introduced by Granger [18, 19], Granger and Joyeux [20], and Hosking [21] (though earlier work [22, 23] shows an awareness of its representation). They were theoretically justified in terms of aggregation of ARMA processes with randomly varying coefficients in Robinson [24] and Granger [18]. Similarly, others [25–28] also use aggregation to motivate long-memory processes, while Parke [29] uses a closely related discrete time error duration model. Time series with this characteristic has been found to be present in hydrology [30, 31], economics [32, 33], high-speed networks [34, 35] and in other areas.

To determine the appropriate degree of integration in raw time series is important from a statistical point of view. If $d = 0$, the series is covariance stationary and possesses ‘short memory’, with the autocorrelations decaying fairly rapidly. If d belongs to the interval (0, 0.5), x_t is still covariance stationary; however, the autocorrelations take much longer time to disappear than in the previous case. If $d \in [0.5, 1)$, the series is no longer covariance stationary, but still mean reverting, with the effect of the shocks dying away in the long run. Finally, if $d \geq 1$, x_t is nonstationary and non-mean reverting. Thus, the fractional differencing parameter d plays a crucial role in describing the persistence in the time series behaviour: higher the d, higher will be the level of association between the observations.

There exist many approaches of estimating and testing the fractional differencing parameter d. Many of the estimators are graphical in nature (heuristic estimators), while some involve numerical minimisation of a likelihood-type function [36–40]. However, several papers in a hydrological context showed that the presence of periodicities might influence the reliability of the estimators [31, 41, 42]. Analysing the series of the monthly flows of the Nile River at Aswan, it was found that many heuristic estimators gave a positive value for d, indicating long memory where none was present. In another paper [43], an extensive Monte Carlo investigation was performed to find out how reliable the estimators of long memory were in the presence of periodicities. The conclusions were that the best results were those obtained using the likelihood-type methods.

In this article, we use a parametric testing procedure of Robinson [16] described in Section 2. In Section 3, the tests are applied to some Indian monthly rainfall data, while Section 4 contains some concluding comments.

2. The testing procedure

Most of commonly used unit-root tests existing in the literature [44–46] have been developed in autoregressive (AR) alternatives of form:

$$(1 - r L) x_t = u_t, \tag{2}$$

where the unit root null corresponds to

$$H_o\colon r = 1. \tag{3}$$

Conspicuous features of these methods for testing unit roots are the nonstandard nature of the null asymptotic distributions involved, and the absence of Pitman efficiency. However, these properties are not automatic, but rather depend on what might be called a degree of ‘smoothness’ in the model across the parameters of interest, in the sense that the limit distribution does not change in an abrupt way with small changes in the parameters. This is associated with the radically variable long-run properties of AR processes around the unit root. In (2), for $|r| > 1$, x_t is explosive, for $|r| < 1$, x_t is covariance stationary, and for $r = 1$ it is nonstationary but non-explosive. In view of these abrupt changes, the fractional processes have become a rival class of alternatives to the AR model in the case of testing unit roots. Robinson [16] proposes a Lagrange Multiplier (LM) test of the null hypothesis:

$$H_o\colon d = d_o, \tag{4}$$

for any real value d_o in a model given by (1), and where x_t can be the errors in a regression model:

$$y_t = b' z_t + x_t, \tag{5}$$

where $b' = (b_1, \ldots, b_k)$ is a $(k \times 1)$ vector of unknown parameters, and z_t is a $(k \times 1)$ vector of deterministic regressors that may include, for example, an intercept, (e.g., $z_t \equiv 1$), or an intercept and a linear time trend (in the case of $z_t = (1, t)'$). Clearly, the unit root corresponds then to the null hypothesis:

$$H_o\colon d = 1. \tag{6}$$

Fractional and AR departures from (3) and (6) have very different long-run implications. In (6), x_t is nonstationary but non-explosive for all $d \geq 0.5$. As d increases beyond 0.5 and through 1, x_t can be viewed as becoming 'more nonstationary', but it does so gradually, unlike in the case of (2) around (3). Specifically, the test statistic proposed by Robinson [16] is given by:

$$\hat{r} = \left(\frac{T}{\hat{A}}\right)^{1/2} \frac{\hat{a}}{\hat{s}^2}, \tag{7}$$

where T is the sample size and

$$\hat{a} = \frac{-2p}{T} \sum_{j=1}^{T-1} y(l_j) g(l_j; \hat{t})^{-1} I(l_j) \quad \hat{s}^2 = s^2(\hat{t}) = \frac{2p}{T} \sum_{j=1}^{T-1} g(l_j; \hat{t})^{-1} I(l_j)$$

$$\hat{A} = \frac{2}{T} \left(\sum_{j=1}^{T-1} y(l_j)^2 - \sum_{j=1}^{T-1} y(l_j) \hat{e}(l_j)' \times \left(\sum_{j=1}^{T-1} \hat{e}(l_j) \hat{e}(l_j)' \right)^{-1} \times \sum_{j=1}^{T-1} \hat{e}(l_j) y(l_j) \right)$$

$$y(l_j) = \log \left| 2 \sin \frac{l_j}{2} \right|; \hat{e}(l_j) = \frac{\partial}{\partial t} \log g(l_j; \hat{t}); \quad l_j = \frac{2p\, j}{T}.$$

$I(l_j)$ is the periodogram of $\hat{u}_t$, where

$$\hat{u}_t = (1-L)^{d_o} y_t - \hat{b} w_t, \quad w_t = (1-L)^{d_o} z_t; \hat{b} = \left(\sum_{t=1}^{T} w_t w_t' \right)^{-1} \sum_{t=1}^{T} w_t (1-L)^{d_o} z_t,$$

and g above is a known function coming from the spectral density of u_t:

$$f(l_j; t) = \frac{s^2}{2p} g(l_j; t).$$

Note that these tests are purely parametric and, therefore, require specific modelling assumptions to be made regarding the short memory specification of u_t. Thus, for example, if

u_t is white noise, $g \equiv 1$, and if u_t is AR(1) of form: $u_t = t\,u_{t-1} + e_t$, $g(l_j; t) = |1 - t\,e^{i l_j}|^{-2}$, with $s^2 = V(e_t)$, so that the AR coefficients are functions of t.

Robinson [16] showed that under certain regularity conditions (which are very mild, and concern technical assumptions to be satisfied by $y(l)$):

$$\hat{r} \rightarrow_d N(0,1) \quad \text{as} \quad T \rightarrow \infty. \tag{8}$$

Thus, an approximate one-sided $100a$%-level test of H_o (4) against the alternative: H_a: $d > d_o$ ($d < d_o$) will reject H_o if $\hat{r} > z_a$ ($\hat{r} < -z_a$), where the probability that a standard normal variate exceeds z_a is a. Furthermore, he shows that the above test is efficient in the Pitman sense, i.e. that against local alternatives of form: H_a: $d = d_o + dT^{-1/2}$, with $d \neq 0$, the limit distribution is normal with variance 1 and mean that cannot (when u_t is gaussian) be exceeded in absolute value by that of any rival regular statistic. Therefore, we are in a classical large sample testing situation by reasons described in Robinson [16]. Empirical applications based on this version of Robinson's tests can be found in [47, 48], and other versions of his tests, based on seasonal (quarterly and monthly) and cyclical data, are presented in [49–51].

There exist other procedures for estimating and testing the fractionally differenced parameter, some of them also based on the likelihood function. We believe that as in other standard large-sample testing situations, Wald and LR test statistics against fractional alternatives will have the same null and local limit theory as the LM tests of Robinson. Sowell [38] employed essentially such a Wald testing procedure but it requires an efficient estimate of d, and while such estimates can be obtained, no closed-form formulae are available and so the LM procedure of Robinson seems computationally more attractive. In the following section, the versions of the tests described above will be applied to some Indian meteorological data.

3. Data and empirical results

The time-series data analysed in this section correspond to the monthly observations of the homogeneous Indian rainfall datasets for the time period 1871–1999, for all India and six subdivisions (Core monsoon; North east; Central west; Central northeast; North west; and Peninsular. See Fig. 1), obtained from the Indian Institute of Tropical Meteorology (India Meteorological Department).

Any modelling effort on this dataset will have to be based on an understanding of the variability of past data. Thus, considerable literature is available on the analysis of the Indian rainfall data [10, 52–56]. Some of these papers, for example [10], assume that the series of interest is nonstationary, and first differences are adopted in order to examine the short-run behaviour throughout the ARMA structures. In other words, it has imposed an order of integration equal to 1 as opposed to the case of $d = 0$ if the series is stationary. In this paper, we permit the order of integration to be a real value and, in doing so, we allow for a much richer degree of flexibility in its dynamic behaivour. Other recent empirical papers based on forecasting Indian monsoon rainfall data are those of Gadgil *et al.* [57] and Iyengar and Raghu Kanth [58].

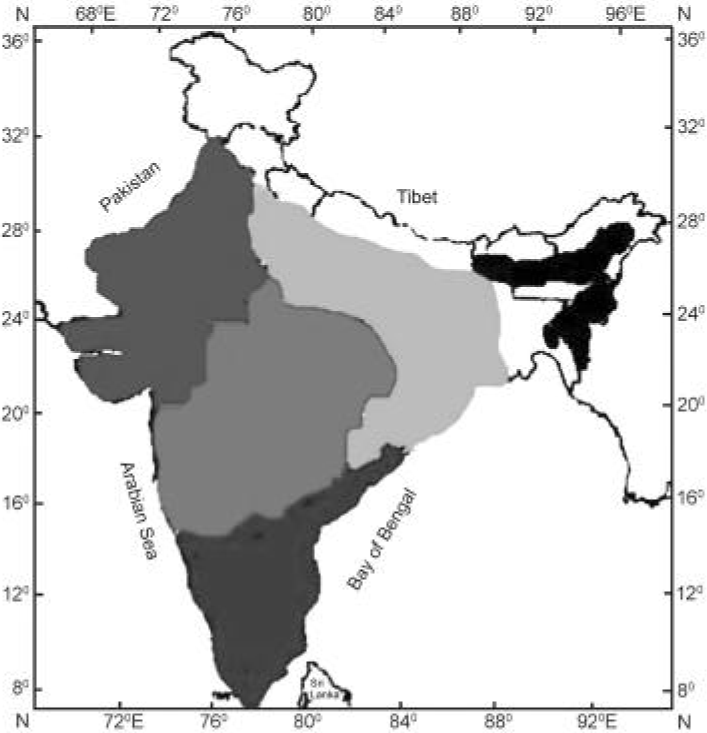

FIG. 1. Regions: ■ Northwest India; ■ West central India; ■ Central northeast India; ■ Northeast India; ■ Peninsular India; Homogeneous regions of India.

Tables I and II display some salient features of the data. In Table I, we report the mean and the standard deviation for all India and each of the homogeneous regions. We observe that for north east, central northeast and peninsular, the mean values are above the mean of the whole country. The same statistics were also computed for each month in Table II. Here, we observe large differences across months, the highest values for the mean obtained during the months from June to September.

Figure 2 displays plots of the original data for the whole country along with its first seasonal (monthly) differences. We see that the original series has a strong seasonal component, while the first differences may be stationary. Figure 3 displays the correlograms and the periodograms of both the series and we observe that the differenced series may be over-differenced in relation to its seasonal structure, with a large negative value in the correlogram at lag 12, and the periodogram with values close to zero at the seasonal frequencies.

Table I
Salient statistics for All India and homogeneous regions (mm/month)

		All India	Core monsoon	North east	West central	Central northeast	North west	Peninsular
Total	Mean	909.61	800.17	1725.23	898.93	1002.50	455.65	968.10
	Std dev.	954.81	1096.84	1528.17	1130.68	1201.91	718.57	787.65

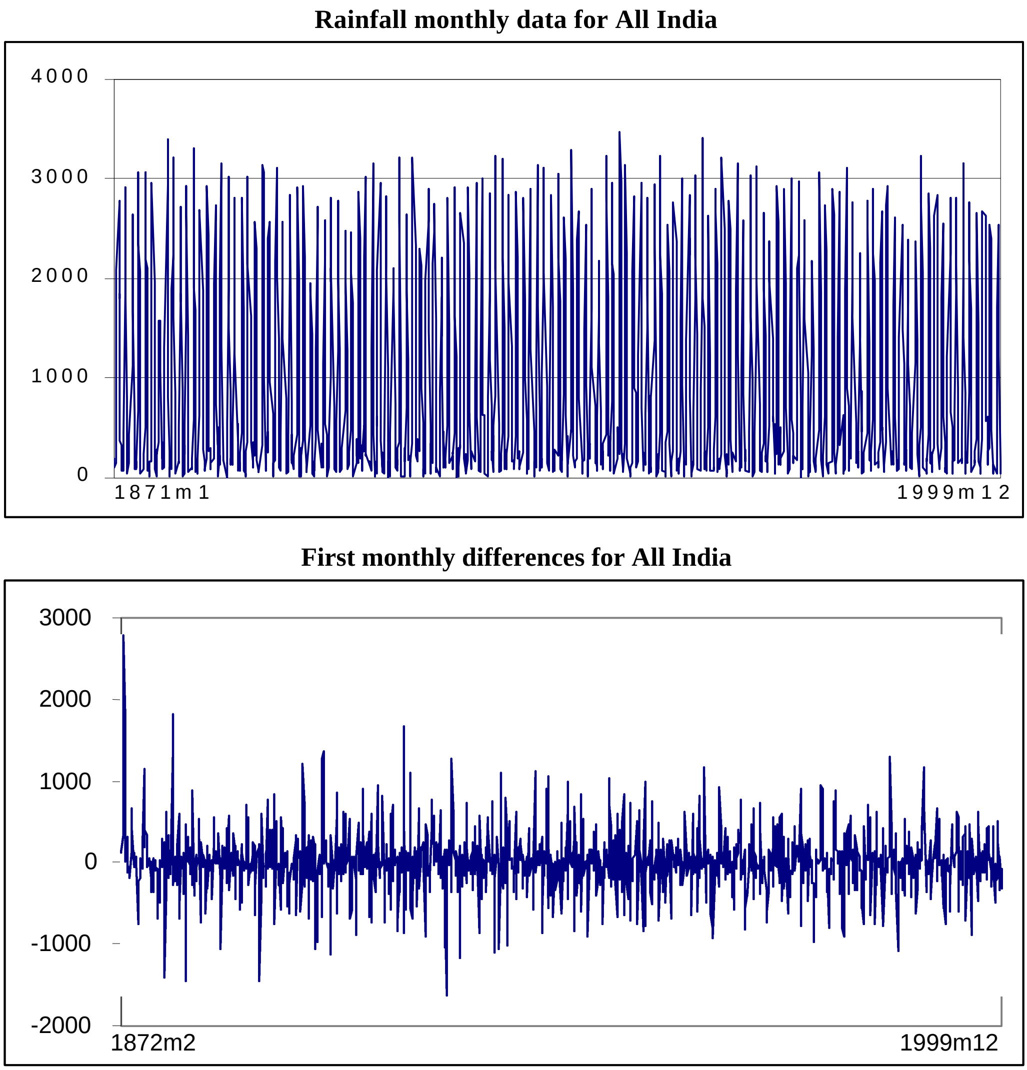

FIG. 2. Original series and first monthly differences.

Table II
Salient statistics for each month in All India (mm/month)

January	Mean	111.33	Std dev.	77.36	July	Mean	2737.76	Std dev.	362.22
February	Mean	127.36	Std dev.	88.08	August	Mean	2433.79	Std dev.	380.01
March	Mean	151.89	Std dev.	91.23	September	Mean	1712.29	Std dev.	373.83
April	Mean	262.72	Std dev.	89.66	October	Mean	779.30	Std dev.	285.46
May	Mean	527.89	Std dev.	161.27	November	Mean	315.54	Std dev.	183.95
June	Mean	1633.98	Std dev.	361.24	December	Mean	121.72	Std dev.	98.22

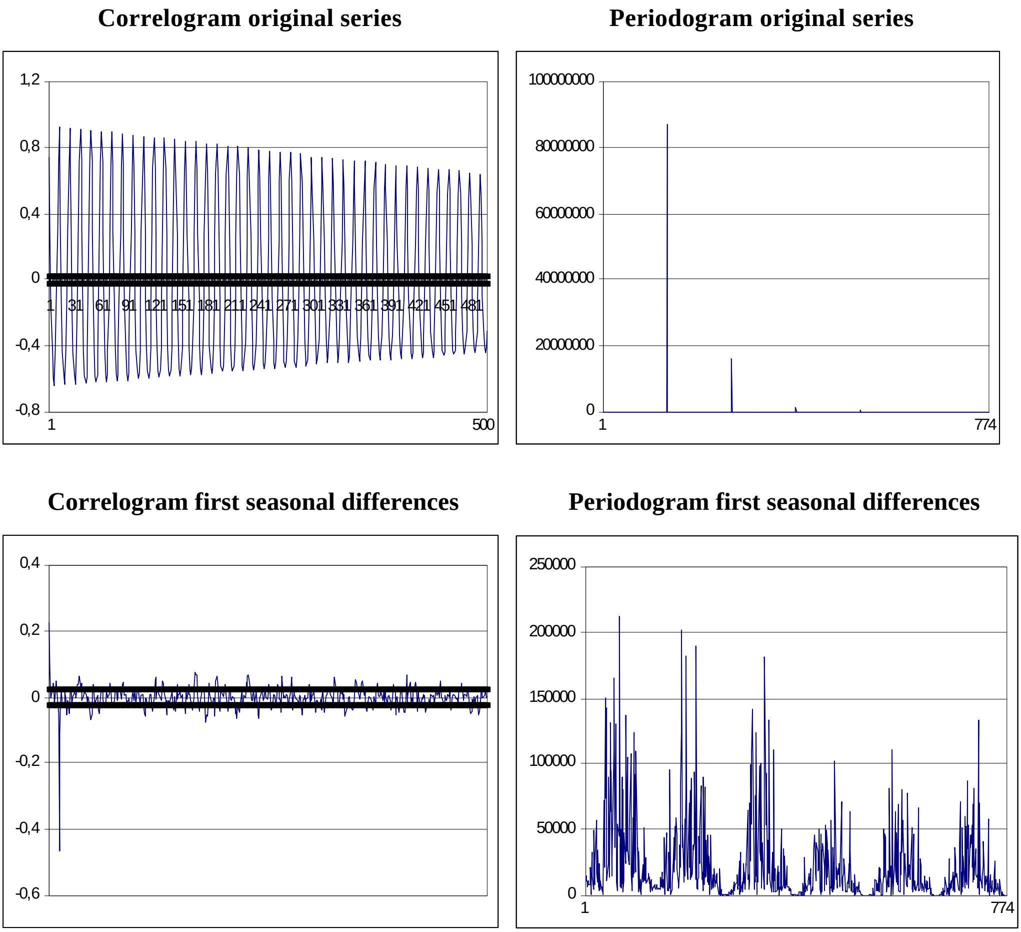

FIG. 3. Correlograms and periodograms of the original series and first monthly differences. The large sample standard error under the null hypothesis of no autocorrelation is $1/\sqrt{T}$ or roughly 0.025 for the series used in this application. The periodograms are drawn for the discrete frequencies $l_j = 2pj/T$, $j = 0, \ldots, T/2$.

Figures 4 and 5 display respectively the plots of the original series and the first seasonal differences for each of the six regions in India. Similar to the data for the whole country, the original series clearly appear nonstationary with a strong seasonal pattern. The seasonal differences, however, may be stationary.

Denoting each of the time series by y_t, we employ throughout the model given by (1) and (5), with $z_t = (1, t)'$, $t \geq 1$, $z_t = (0, 0)'$. Thus, under the null hypothesis (4),

$$y_t = b_0 + b_1 t + x_t, \qquad t = 1, 2, \ldots \tag{9}$$

$$(1 - L)^{d_o} x_t = u_t, \qquad t = 1, 2, \ldots. \tag{10}$$

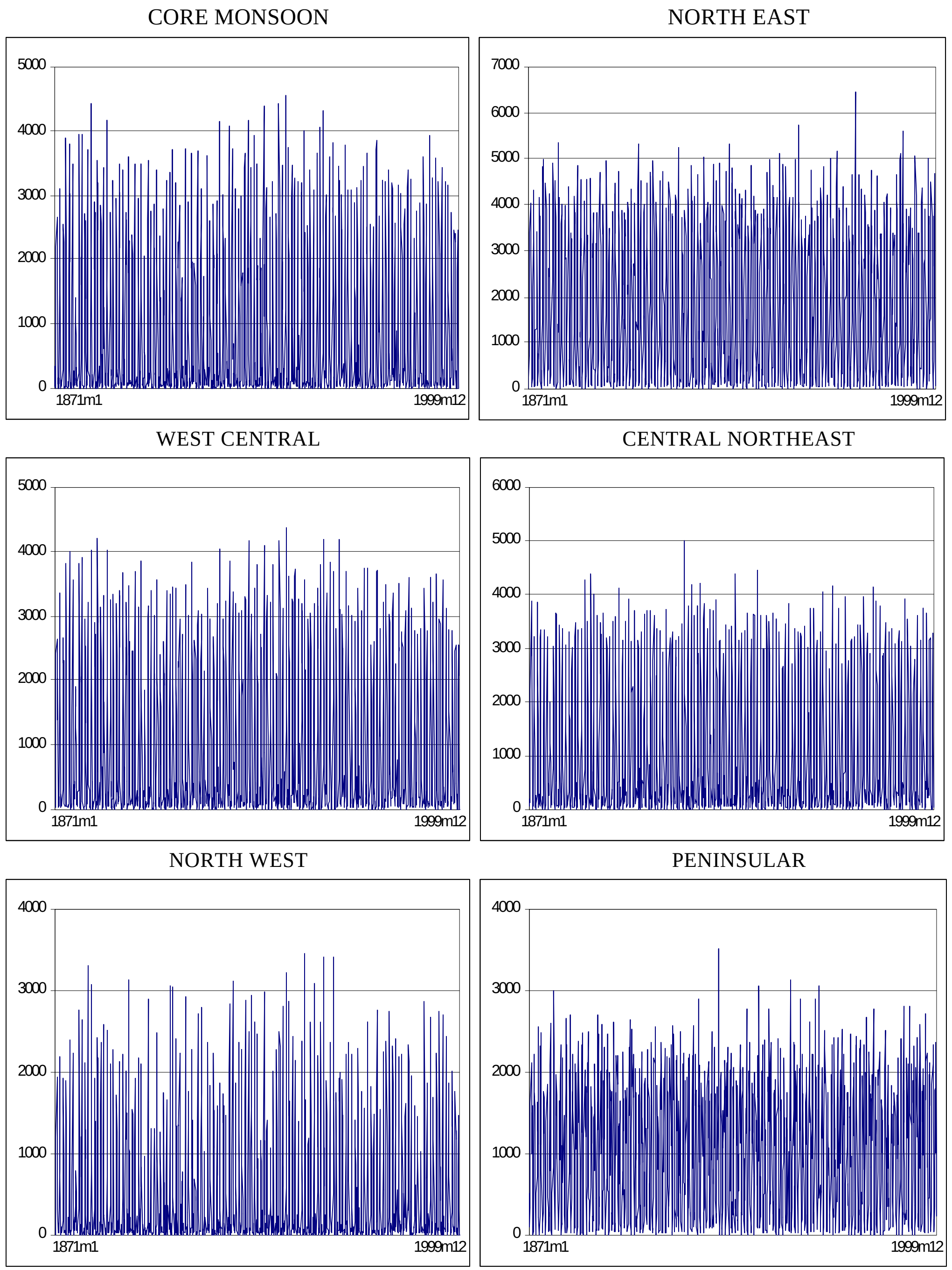

FIG. 4. Plots of the original series.

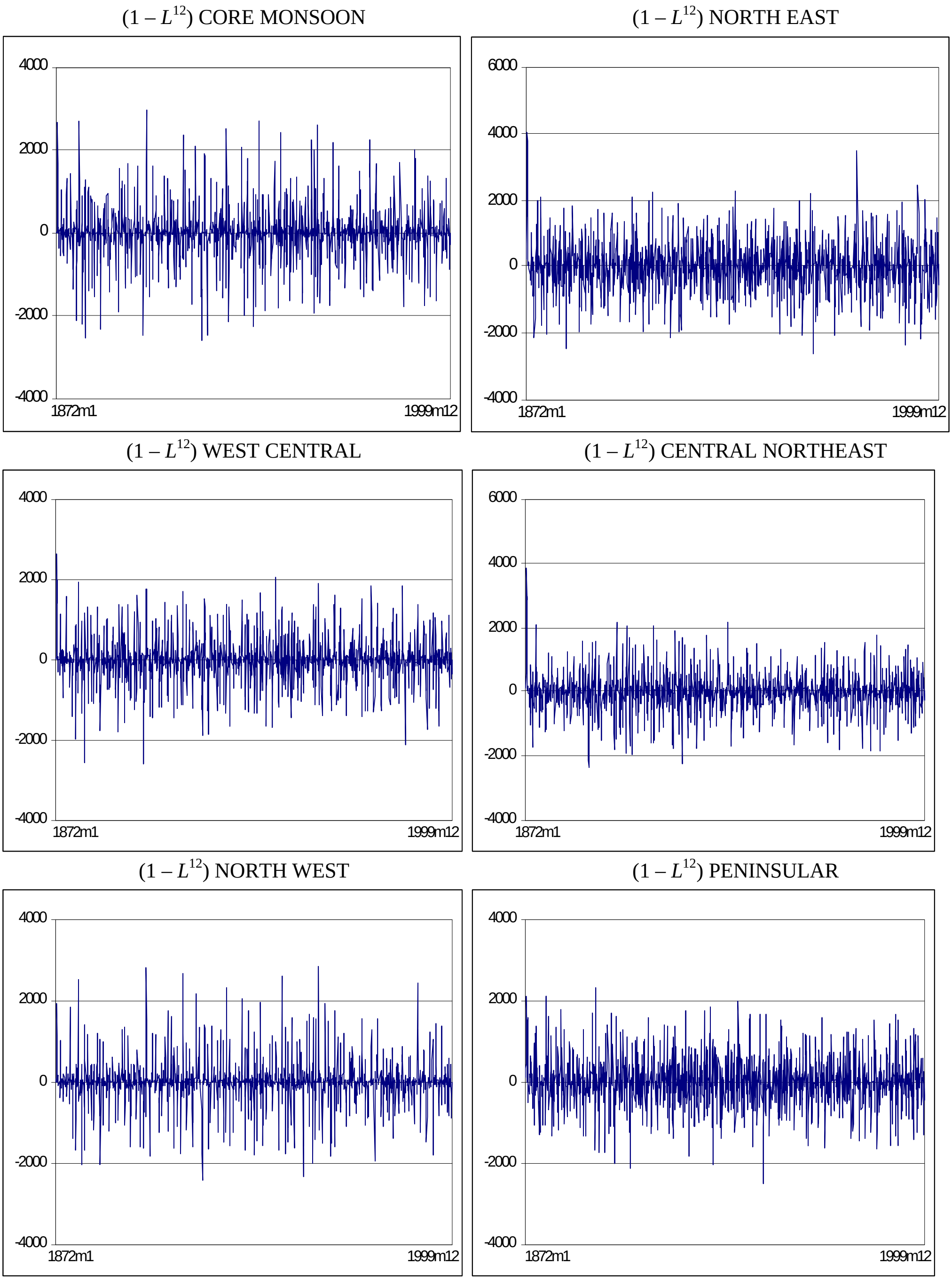

FIG. 5. Plots of the first monthly differences.

and treat separately the cases $b_0 = b_1 = 0$ a priori; b_0 unknown and $b_1 = 0$ a priori; and b_0 and b_1 unknown, i.e. we consider respectively the cases of no regressors in the undifferenced regression (9), an intercept, and an intercept and a linear time trend. However, given the similarities obtained in the results across the three cases, we report in the tables the values based on the case of no regressors. (The coefficients corresponding to the intercept and the linear trend were insignificant in all the cases where H_o cannot be rejected. They are based on the null model, which is short memory, and thus standard *t*-tests apply). We report the test statistics not merely for the null $d_o = 1$, (i.e. a unit root), but also for $d_o = 0$, (0.25), 2, thus including a test for stationarity ($d_o = 0.5$), for $I(2)$ processes ($d_o = 2$), as well as other fractionally integrated possibilities.

The test statistic reported in Table III is the one-sided one corresponding to $\hat{r}$ in (7), so that significantly positive values of this are consistent with orders of integration higher than d_o, whereas significantly negative ones are consistent with alternatives of form: $d < d_o$. A notable feature observed in Table III (i), in which u_t is taken to be white noise, is that the value of the test statistic monotonically decreases with d_o. This is something to be expected in view of the fact that it is a one-sided statistic. Thus, for example, if H_o (4) is rejected with $d_o = 1$ against the alternative $d > 1$, an even more significant result in this direction should be expected when $d_o = 0.75$ or 0.50 are tested. We see that the results change substantially depending on the series under study. Starting with the data corresponding to the whole country, we observe that the unit root null hypothesis (i.e. $d = 1$) is rejected in favour of higher orders of integration. In fact, the only value of d where H_o cannot be rejected corresponds to $d = 1.25$. The unit root null is also rejected in favour of higher values of d for north east: it is nonrejectable for west central and central northeast, while for the other three regions (core monsoon, north west and peninsular), it is rejected in favour of smaller values of d. The last column of the table reports the 95%-confidence intervals of those values of d_o where H_o cannot be rejected. We see that for all series, except All India and north east, the intervals include the unit root, the values ranging from (0.44–0.62) for north west to (0.97–1.18) for central northeast. For the whole India, the interval is (1.16–1.36).

The significance of the above results, however, may be in large part due to the unaccounted for $I(0)$ autocorrelation in u_t. Thus, we also performed the tests, imposing autocorrelated disturbances. We use AR(1) (in Table III (ii)) and Bloomfield [59] disturbances (in Table III (iii)). The latter is a nonparametric approach of modelling the $I(0)$ disturbances in which u_t is exclusively specified in terms of its spectral density function, which is given by:

$$f(l;t) = \frac{s^2}{2p} \exp\left(2 \sum_{r=1}^{m} t_r \cos(l\, r) \right). \qquad (11)$$

The intuition behind this model is the following. Suppose that u_t follows an ARMA process,

$$u_t = \sum_{r=1}^{p} f_r u_{t-r} + e_t - \sum_{r=1}^{q} q_r e_{t-r},$$

where e_t is a white noise process and all zeros of $f(L) = (1 - f_1 L - \ldots - f_p L^p)$ lie outside the unit circle and all zeros of $q(L) = (1 - q_1 L - \ldots - q_q L^q)$ lie outside or on the unit circle. Clearly, the spectral density function of this process is then,

Table III

Values of Robinson's test statistic ($\hat{r}$) testing H_o: $d = d_o$ in the model $(1 - L)^d x_t = u_t$

Series/d_o	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	Conf. intervals
(i) *With white noise disturbances*										
All India	16.12	13.03	11.12	7.85	4.11	**0.17**	–3.65	–7.08	–9.95	[1.16–1.36]
Core monsoon	12.72	8.94	5.07	**1.00**	–2.96	–6.55	–9.58	–12.01	–13.89	[0.72–0.91]
North east	17.81	15.39	12.21	7.94	2.75	–2.70	–7.64	–11.54	–14.32	[1.06–1.20]
West central	13.95	10.99	7.29	3.40	**–0.57**	–4.36	–7.71	–10.49	–12.69	[0.87–1.06]
Central northeast	14.14	11.70	8.35	4.80	**1.11**	–2.52	–5.88	–8.80	–11.22	[0.97–1.18]
North west	9.60	4.90	**0.44**	–3.55	–6.99	–9.84	–12.10	–13.85	–15.20	[0.44–0.62]
Peninsular	13.51	9.30	3.66	–1.76	–6.55	–10.36	–13.14	–15.06	–16.36	[0.60–0.74]
(ii) *With AR(1) disturbances*										
All India	20.62	10.31	**–0.33**	–6.87	–9.07	–9.26	–10.04	–10.04	–10.37	[0.45–0.53}
Core monsoon	1.97	–5.76	–7.81	–8.17	–8.88	–8.76	–9.02	–9.59	–10.35	[0.01–0.09]
North east	16.53	2.81	–6.93	–7.77	–7.86	–8.55	–8.89	–9.90	–9.97	[0.28–0.33]
West central	8.34	–1.75	–7.53	–8.20	–8.89	–8.93	–8.94	–9.30	–9.93	[0.15–0.23]
Central northeast	12.90	2.63	–5.15	–8.65	–9.57	–9.58	–9.69	–9.75	–9.83	[0.28–0.36]
North west	–5.85	–8.49	–9.03	–9.05	–9.22	–9.69	–10.37	–11.15	–11.96	—
Peninsular	–5.51	–6.53	–6.56	–7.18	–7.97	–8.03	–8.83	–10.80	–12.54	—
(iii) *With Bloomfield (1) disturbances*										
All India	–4.39	–4.87	–5.59	–6.50	–7.26	–7.92	–8.78	–9.22	–10.09	—
Core monsoon	–4.86	–6.00	–6.66	–7.45	–8.39	–8.88	–9.54	–10.33	–10.96	—
North east	–2.09	–2.14	–2.82	–3.25	–4.06	–4.40	–5.17	–5.90	–6.21	—
West central	–4.70	–5.73	–6.38	–7.19	–7.83	–8.68	–9.38	–9.95	–10.64	—
Central northeast	–4.97	–5.76	–6.70	–7.41	–8.29	–8.70	–9.60	–10.06	–10.64	—
North west	–5.51	–6.99	–7.75	–8.84	–9.93	–10.06	–10.84	–11.23	–11.72	—
Peninsular	–3.35	–3.87	–4.94	–5.68	–6.26	–7.16	–7.66	–8.47	–9.22	—
(iv) *With monthly AR(1) disturbances*										
All India	4.37	–3.21	–8.38	–11.50	–13.52	–15.17	–16.83	–18.63	–20.54	[0.08–0.19]
Core monsoon	1.90	–5.66	–10.36	–13.39	–15.75	–17.94	–20.10	–22.19	–24.12	[0.02–0.11]
North east	**1.17**	–5.70	–9.94	–12.23	–13.92	–16.07	–19.03	–22.34	–25.36	[–0.01–0.08]
West central	2.39	–5.15	–9.81	–12.73	–14.88	–16.82	–18.78	–20.77	–22.68	[0.02–0.11]
Central northeast	**0.63**	–5.75	–9.94	–12.52	–14.38	–16.10	–17.95	–19.99	–22.10	[–0.02–0.07]
North west	2.66	–5.36	–10.55	–14.17	–17.10	–19.71	–22.04	–24.07	–25.79	[0.03–0.12]
Peninsular	1.66	–4.89	–9.46	–13.16	–16.74	–20.22	–23.26	–25.64	–27.41	[0.01–0.11]
(v) *With monthly AR(2) disturbances*										
All India	3.99	–3.56	–8.69	–11.55	–13.16	–14.18	–14.92	–15.53	–16.83	[0.07–0.17]
Core monsoon	**0.74**	–6.11	–10.42	–12.44	–13.30	–13.78	–14.34	–15.12	–16.04	[–0.02–0.06]
North east	**0.77**	–6.11	–10.42	–12.44	–13.30	–13.78	–14.34	–15.12	–16.04	[–0.01–0.07]
West central	**1.38**	–5.98	–10.18	–12.51	–13.89	–14.83	–15.56	–16.19	–16.74	[0.00–0.08]
Central northeast	**0.90**	–5.71	–9.97	–12.30	–13.62	–14.45	–15.05	–15.54	–15.99	[–0.01–0.08]
North west	**1.46**	–5.89	–9.96	–12.27	–13.73	–14.78	–15.61	–16.30	–16.88	[0.00–0.08]
Peninsular	**0.79**	–5.74	–9.49	–11.62	–13.07	–14.30	–15.38	–16.28	–17.00	[–0.01–0.07]

Figures in bold represent nonrejection values at 5% significance level.

$$f(l;t) = \frac{s^2}{2p}\left|\frac{1-\sum_{r=1}^{q} q_r e^{irl}}{1-\sum_{r=1}^{p} f_r e^{irl}}\right|^2, \qquad (12)$$

where t corresponds to all the AR and MA coefficients and s^2 is the variance of e_t. Bloomfield showed that the logarithm of an estimated spectral density function is often found to be a fairly well-behaved function and can thus be approximated by a truncated Fourier series. He showed that (11) approximates to (12) well, where p and q are small values. Like the stationary AR(p) model, the Bloomfield [59] model has exponentially decaying autocorrelations and thus we can use a model like this for u_t in (10). Formulae for Newton-type iteration for estimating the t_l are very simple (involving no matrix inversion), updating formulae when m is increased is also simple, and we can replace $\hat{A}$ in (7) by the population quantity,

$$\sum_{l=m+1}^{\infty} l^{-2} = \frac{p^2}{6} - \sum_{l=1}^{m} l^{-2},$$

which indeed is constant with respect to t_r (unlike what happens in the AR case). The Bloomfield model, involving fractional integration has not been used very much in previous econometric models. Though it is a well-known model in other disciplines [60], one by-product of the present work is the emergence of that model as a credible alternative to the fractional ARIMAs, which have become conventional in parametric modelling of long memory. Amongst the few empirical applications found in the literature are those of Gil-Alana and Robinson [47], Velasco and Robinson [61], and more recently Gil-Alana [62]. Reverting to the results in Table III, we observe that using AR(1) u_t, the confidence intervals are higher than 0, but smaller than 1, in all series except north west and peninsular and, imposing Bloomfield (with m = 1) disturbances, they are smaller than 0 in all cases. However, these results should be considered with care since the previous specifications did not consider the seasonal patterns observed in Figs 2–5. So, we also performed the tests imposing seasonal autoregressions of form:

$$u_t = \sum_{i=1}^{m} f_i u_{t-12i} + e_t,$$

with m = 1 and 2. The results are shown in Tables III (iv) and (v). Here, we observe that the degrees of integration are very small, fluctuating around 0 in practically all cases. This can be explained by the fact that the seasonal AR coefficients are competing with d in describing the nonstationary component of the series. Note that the estimates are Yule–Walker and thus, though they are smaller than 1 in absolute value, they can be arbitrarily close to 1, this being perhaps the reason for the nonrejection of the null when d = 0.

As mentioned in Section 1, several papers by Montanari and others showed that the presence of periodicities in the data may be affecting the degree of integration in the long run or zero frequency, implying long memory when it is not present. In Table IV, we specifically

Table IV
Values of Robinson's test statistic ($\hat{r}$) testing H_o: $d = d_o$ in the model $(1 - L^{12})^d x_t = u_t$

Series/d_o	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	Conf. intervals
(i) *With white noise disturbances*										
All India	4.32	**–0.07**	**–0.14**	**–1.60**	–4.04	–4.61	–4.75	–4.83	–4.88	[0.05–0.75]
Core monsoon	3.11	**–0.01**	**–0.33**	–2.12	–3.40	–3.73	–3.98	–4.02	–4.11	[0.04–0.69]
North east	3.02	**–0.06**	**–0.13**	**–1.42**	–3.49	–3.99	–4.15	–4.25	–4.32	[0.07–0.77]
West central	3.44	**–0.01**	**–0.24**	–1.92	–3.48	–3.87	–4.04	–4.16	–4.24	[0.03–0.72]
Central northeast	2.91	**–0.01**	**–0.28**	–2.17	–3.87	–4.22	–4.33	–4.39	–4.43	[0.08–0.71]
North west	2.34	**–0.06**	**–0.13**	**–1.42**	–3.49	–3.99	–4.15	–4.25	–4.32	[0.05–0.77]
Peninsular	2.58	**–0.08**	**–0.17**	**–1.57**	–3.03	–3.38	–3.54	–3.66	–3.75	[0.06–0.75]
(ii) *With AR(1) disturbances*										
All India	8.86	5.24	**–0.70**	**–1.50**	–2.60	–3.07	–3.13	–3.37	–3.46	[0.39–0.80]
Core monsoon	1.67	**–1.54**	**–0.28**	–2.00	–3.17	–3.52	–3.70	–3.81	–3.90	[0.04–0.52]
North east	2.60	1.71	**–0.39**	**–1.19**	–2.61	–3.05	–3.18	–3.34	–3.42	[0.42–0.79]
West central	2.14	**1.22**	**–0.16**	**–1.60**	–2.81	–3.19	–3.20	–3.49	–3.57	[0.21–0.57]
Central northeast	2.30	1.88	**–0.19**	**–1.50**	–2.79	–3.11	–3.15	–3.35	–3.42	[0.33–0.77]
North west	2.39	**1.05**	**–0.64**	–2.00	–3.25	–3.50	–3.57	–3.76	–3.87	[0.19–0.58]
Peninsular	4.47	**1.29**	**–0.14**	**–1.60**	–2.77	–3.13	–3.30	–3.42	–3.52	[0.22–0.61]
(iii) *With Bloomfield (1) disturbances*										
All India	7.74	4.02	**–0.55**	**–1.32**	–2.59	–3.17	–3.78	–4.09	–4.55	[0.36–0.79]
Core monsoon	1.66	**1.31**	**–0.11**	–1.78	–2.34	–3.00	–3.44	–3.76	–4.32	[0.09–0.73]
North east	2.33	**1.54**	**–0.35**	**–1.44**	–2.12	–2.97	–3.18	–3.43	–3.52	[0.23–0.77]
West central	2.24	**1.23**	**–0.23**	**–1.55**	–2.34	–3.09	–3.55	–3.79	–4.12	[0.18–0.78]
Central northeast	2.10	**1.54**	**–0.29**	**–1.56**	–2.11	–2.71	–2.99	–3.34	–3.78	[0.23–0.76]
North west	2.18	**1.22**	**–0.55**	–2.65	–3.35	–3.67	–3.77	–4.06	–4.22	[0.20–0.68]
Peninsular	3.31	**1.12**	**–0.34**	**–1.54**	–2.12	–2.68	–3.24	–3.57	–3.68	[0.19–0.78]
(iv) *With monthly AR(1) disturbances*										
All India	2.54	**1.17**	**–0.11**	**–1.57**	–2.90	–3.95	–4.07	–4.41	–4.54	[0.18–0.77]
Core monsoon	2.44	**1.23**	**–0.14**	**–1.18**	–2.10	–2.57	–2.99	–3.01	–3.17	[0.17–0.82]
North east	2.07	**1.51**	**–0.08**	**–1.10**	–2.42	–3.21	–3.50	–3.64	–3.78	[0.23–0.84]
West central	2.12	**1.60**	**–0.11**	**–1.20**	–2.11	–2.69	–2.70	–3.18	–3.35	[0.25–0.80]
Central northeast	2.35	**1.61**	**–0.13**	**–1.18**	–2.98	–3.73	–4.00	–4.17	–4.30	[0.26–0.79]
North west	2.25	**1.23**	**–0.30**	**–1.40**	–2.57	–2.92	–3.00	–3.33	–3.48	[0.21–0.79]
Peninsular	7.83	5.69	**1.44**	**–0.67**	–1.81	–2.33	–2.50	–2.79	–2.97	[0.36–0.83]
(v) *With monthly AR(2) disturbances*										
All India	2.35	**1.22**	**–0.15**	**–1.43**	–2.33	–2.89	–3.67	–3.99	–4.11	[0.17–0.77]
Core monsoon	2.14	**1.19**	**–0.54**	**–1.48**	–2.56	–2.98	–2.99	–3.14	–3.34	[0.18–0.76]
North east	2.65	**1.45**	**–0.33**	**–1.35**	–2.67	–3.32	–3.69	–4.09	–4.35	[0.20–0.80]
West central	3.11	**1.61**	**–0.22**	**–1.29**	–2.23	–2.78	–3.00	–3.19	–3.44	[0.23–0.82]
Central northeast	2.88	**1.64**	**–0.23**	**–1.44**	–2.18	–3.84	–4.21	–4.76	–4.90	[0.25–0.80]
North west	2.36	**1.33**	**–0.26**	**–1.37**	–2.33	–2.42	–2.69	–3.13	–3.99	[0.20–0.79]
Peninsular	6.57	5.69	**1.35**	**–0.89**	–1.77	–2.21	–2.33	–2.88	–3.11	[0.34–0.85]

Figures in bold represent nonrejection values at 5% significance level.

take into account the seasonal structure of the series and make use of another version of Robinson's tests that permits us to test unit and fractional roots not only at zero but also at the seasonal frequencies. Thus, instead of (1), we consider processes of form:

Table V
Values of Robinson's test statistic ($\hat{r}$) testing H_o: $d = d_o$ in the model $(1 - L)^d x_t = u_t$

Series/d_o	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	Conf. intervals
(i) *With white noise disturbances*										
January	**0.97**	**–1.52**	–3.18	–4.23	–4.86	–5.27	–5.55	–5.76	–5.92	[–0.04–0.26}
February	**0.47**	–2.24	–3.67	–4.50	–4.99	–5.33	–5.58	–5.77	–5.92	[–0.04–0.17]
March	–2.09	–2.91	–3.96	–4.68	–5.10	–5.37	–5.57	–5.73	–5.85	[–0.13–0.05]
April	**–1.30**	–2.83	–4.10	–4.97	–5.38	–5.61	–5.77	–5.90	–6.01	[–0.05–0.01]
May	**–0.62**	**0.03**	–1.83	–3.32	–4.17	–4.70	–5.07	–5.35	–5.56	[–0.08–0.47]
June	**0.08**	**1.63**	**–1.57**	–3.69	–4.62	–5.07	–5.33	–5.51	–5.65	[–0.08–0.51]
July	**0.23**	4.26	**0.49**	–2.82	–4.39	–5.10	–5.49	–5.73	–5.90	[–0.02–0.56]
August	**–1.48**	**1.13**	**–1.31**	–3.69	–4.80	–5.30	–5.58	–5.77	–5.91	[–0.03–0.00]
September	**–1.29**	**0.94**	**–1.55**	–3.62	–4.68	–5.24	–5.58	–5.80	–5.95	[–0.07–0.50]
October	**1.07**	**–0.04**	–1.94	–3.52	–4.46	–5.01	–5.36	–5.60	–5.77	[–0.02–0.46]
November	**–0.20**	–2.70	–4.09	–4.92	–5.38	–5.65	–5.83	–5.96	–6.07	[–0.06–0.09]
December	**–0.35**	–2.08	–3.42	–4.26	–4.78	–5.16	–5.45	–5.69	–5.87	[–0.11–0.18]
(ii) *With AR(1) disturbances*										
January	**–0.75**	**–1.30**	–2.09	–2.89	–3.47	–3.90	–4.23	–4.51	–4.74	[–0.24–0.37]
February	**0.41**	**–1.36**	–2.59	–3.38	–3.85	–4.17	–4.43	–4.65	–4.84	[–0.07–0.30]
March	–2.34	–2.17	–3.03	–3.81	–4.26	–4.53	–4.72	–4.86	–4.97	[–0.16––0.06]
April	**–0.05**	**–1.08**	–2.52	–3.87	–4.49	–4.75	–4.92	–5.08	–5.24	[–0.04–0.37]
May	**–1.08**	**1.26**	**–0.02**	–1.66	–2.76	–3.47	–3.99	–4.41	–4.75	[0.25–0.74]
June	**0.30**	2.82	**0.46**	–2.19	–3.72	–4.47	–4.89	–5.16	–5.35	[0.40–0.69]
July	**0.98**	3.04	2.37	**0.06**	–2.00	–3.21	–3.93	–4.42	–4.78	[0.59–0.94]
August	**–0.92**	2.24	**0.80**	–1.71	–3.35	–4.14	–4.57	–4.86	–5.09	[0.41–0.74]
September	**–1.50**	2.10	**0.85**	–1.22	–2.71	–3.63	–4.23	–4.65	–4.97	[0.40–0.81]
October	**–1.23**	**0.13**	**–0.59**	–2.02	–3.15	–3.92	–4.46	–4.85	–5.13	[–0.11–0.67]
November	**0.48**	**–0.75**	–2.13	–3.34	–4.08	–4.55	–4.87	–5.12	–5.33	[–0.07–0.41]
December	**–1.44**	–1.96	–2.87	–3.45	–3.72	–3.87	–4.01	–4.19	–4.37	[–0.23–0.13]
(iii) *With Bloomfield (1) disturbances*										
January	**–0.26**	**–1.21**	–2.00	–2.87	–3.32	–3.66	–3.90	–4.02	–4.15	[–0.11–0.36]
February	**0.46**	**–1.36**	–2.55	–3.38	–3.89	–4.19	–4.36	–4.52	–4.63	[–0.07–0.33]
March	–2.41	–1.90	–2.62	–3.48	–3.93	–4.11	–4.05	–3.91	–3.74	[–0.16–0.07]
April	**0.14**	**0.02**	**–1.23**	–2.99	–3.92	–4.17	–4.28	–5.05	–5.75	[–0.05–0.56]
May	**–1.26**	**0.42**	**–0.52**	–1.71	–2.54	–3.01	–3.25	–3.37	–3.53	[0.24–0.72]
June	**0.34**	2.98	**0.96**	**–1.14**	–2.56	–3.38	–3.87	–4.19	–4.39	[0.42–0.82]
July	**1.03**	4.40	2.53	**0.18**	–1.62	–2.67	–3.34	–3.67	–3.93	[0.60–1.00]
August	**–1.15**	2.73	**1.40**	**–0.95**	–2.60	–3.57	–3.99	–4.11	–4.24	[0.47–0.84]
September	**–1.58**	2.14	**0.81**	**–0.87**	–2.21	–3.07	–3.58	–3.87	–4.08	[0.39–0.86]
October	**–0.54**	**0.10**	**–0.47**	–1.65	–2.51	–3.14	–3.53	–3.94	–4.12	[–0.08–0.74]
November	**0.41**	**–0.47**	**–1.50**	–2.52	–3.08	–3.41	–3.46	–3.48	–3.52	[–0.06–0.55]
December	**–1.37**	–1.93	–2.91	–3.65	–4.05	–4.20	–4.26	–4.46	–4.61	[–0.21–0.10]

Figures in bold represent nonrejection values at 5% significance level.

$$(1-L^{12})^d x_t = u_t, \quad t = 1, \ 2, \ \ldots, \tag{13}$$

where u_t is again $I(0)$, and test H_o for the same d_o values as in Table III. The test statistic then adopts a similar functional form as $\hat{r}$ in (7), the only difference being in $y(l_j)$ that takes the form:

$$y(\lambda_j) = \log\left|2\sin\frac{\lambda_j}{2}\right| + \log\left(2\cos\frac{\lambda_j}{2}\right) + \log\left|2\cos\lambda_j\right| + \log\left|2\left(\cos\lambda_j - \cos\frac{\pi}{3}\right)\right| +$$

$$\log\left|2\left(\cos\lambda - \cos\frac{2\pi}{3}\right)\right| + \log\left|2\left(\cos\lambda - \cos\ \frac{\pi}{6}\right)\right| + \log\left|2\left(\cos\lambda - \cos\frac{5\pi}{6}\right)\right|$$

and $\hat{u}_t$, which is now given by $(1-L^{12})^{d_o} x_t$. The test statistic still has normal null and local limit distributions (Note that the polynomial $(1-L^{12})$ can be decomposed into $(1-L)$ $(1 + L + \ldots + L^{11})$ and thus it includes the root at zero as part of the seasonal polynomial). We report the results based on white noise in Table IV, AR(1), Bloomfield (1) and seasonal AR(1) and AR(2) disturbances, and we see that the results are similar in all these cases. Thus, the unit root is rejected in all cases in favour of smaller degrees of integration, and H_o cannot be rejected when d is constrained between 0.25 and 0.75, implying long memory and mean reversion.

Finally, in Table V, we just concentrate on the data for the whole country. We decompose the time series into its monthly observations, testing the order of integration for each month in a similar way as in Table III, for the cases of white noise, AR(1) and Bloomfield $(m = 1)u_t$. We see that the results are similar for the three types of disturbances. The highest orders of integration are obtained during the months from May to September, i.e. including the Indian monsoon seasonal data, with d ranging between 0.25 and 0.75. On the other extreme, March appears as the most stationary series, with d smaller than 0 for the three types of disturbances.

4. Conclusions

We have examined the stochastic behaviour of several Indian rainfall datasets by means of fractional integration techniques. We have used a parametric testing procedure of Robinson [16] that has several distinguishing features compared with other methods. In particular, the tests have standard null and local limit distributions, which hold independently, of the inclusion or non-inclusion, of deterministic components and of the different types of $I(0)$ disturbances. In addition, they permit us to test unit (and fractional) roots, not only at zero but also at the seasonal frequencies. Moreover, they do not impose gaussianity in order to obtain a standard limit distribution, a condition that is rarely satisfied in this type of datasets [63].

The tests were applied to the monthly observations of the homogeneous rainfall data in India and six subdivisions. First, we performed a version of the tests with the root exclusively located at the long run or zero frequency. The results showed evidence of long memory, especially if the disturbances are white noise. However, this evidence of long memory might be due to the presence of periodicities in the data. So, we also performed another version of the tests, with the roots located at zero but also at the seasonal (monthly) frequencies. The results here suggest that the order of integration of the series ranges between 0.25 and 0.75, showing evidence of long memory and mean reverting behaviour. Finally, we also performed tests segregating the data across months, the results showing that from May to September, the degree of persistence is higher than for the remaining months.

We conclude this paper by saying that there is clear evidence of long memory in the Indian rainfall data. Thus, the standard approaches of assuming either stationarity (with $d = 0$, as in the ARMA case) or unit roots (i.e. $d = 1$, in ARIMA models) may be too restrictive, and so more detailed work into the fractional (seasonal)-type of models should be resorted to. Moreover, all the results reported here suggest that d is smaller than 1 (especially if the seasonal frequencies are considered) and thus the series is mean reverting, implying that any shock affecting them will disappear in the long run. This is in contrast to the $I(1)$ specifications, which imply that shocks persist for ever and require strong policy measures to bring the series back to their original long-term projections. Also, another fact noticeable is that the results are very similar for different regions, suggesting that there is not a different pattern for each of the areas in India. Finally, attempting to summarize the conclusions for the individual months, we are left with the impression that from May to September monsoon seasonal data), the degree of dependence between the observations is higher than for the remaining months with the implication that this might be due to the terms of modelling, policy and/or forecasting.

The problem of generating predictions of meteorological events (such as heavy rainfall over a region) is more complex than that of generating predictions of other time series. Gadgil *et al.* state [57]: "*This is because the atmosphere is unstable and the systems responsible for the events that we are trying to predict, such as clouds or a monsoon depression are the culmination of the instabilities and involve non-linear interaction between different spatial scales*". For long-range predictions of Indian summer monsoon rainfall, the empirical models seem to outperform the physical ones [64], the reason being that most of the atmospheric models have not been able to simulate accurately the inter-annual variability of the Indian summer monsoon rainfall. In this respect, the long-memory models employed in this paper can be considered as alternative approaches when modelling and forecasting the Indian rainfall data.

A potential drawback of the present work might be its univariate nature, with the limitation that it imposes in terms of theorising, policy-making or forecasting. Theoretical models and policy-making involve relationships between many variables, and forecast performance can be improved through the use of many variables (e.g. factor-based forecasts based on data involving hundreds of time series beat univariate forecasts [65]. Thus, it would also be possible to use climate-model-generated responses to forcing factors as covariates in place of using t as a covariate. This is the approach used, for example, in Smith *et al.* [66], where it is claimed that, after all forcing factors (including El Niño) are accounted for, the residuals in an annual time series may be modelled by a simple AR(1) structure. In that sense, the exponential decayment associated to the AR process might be replaced by the hyperbolic structure of the $I(d)$ models. However, the univariate approach adapted in this paper is useful in enabling us to determine the degree of dependence between the observations. Moreover, theoretical econometric models for fractional structures in a multivariate framework are not yet available. In this respect, the present paper can be considered as a preliminary step in the analysis of the Indian rainfall data from a different time-series perspective. Finally, climatological time series may, sometimes, present some properties (e.g. hidden trends, breaks, etc.) that may not be typical for time series in other areas. Robinson's procedure, described in this paper, allows us to include deterministic components (like an intercept, time trends or dummy variables for the breaks), with no effect on its standard null and local limit distributions. How the inclusion of these components may alter the results of the Indian rainfall data will be examined in future papers.

Acknowledgement

The author gratefully acknowledges the financial support from the Ministerio de Ciencia y Tecnologia (SEC2002-01839, Spain). Comments of two anonymous referees are also acknowledged.

References

1. A. Avinash and D. Ghanshyam, Time series models of stream flow for a catchment of Ramganga River, *J. Instn Engrs India, Civil Engng Div.*, **88**, 228–230 (1988).
2. B. D. Janos, D. Lucien and H. R. Omar, Practical generation of synthetic rainfall event time series in a semi-arid climatic zone, *J. Hydrol.*, **103**, 357–373 (1988).
3. K. Geoff, Use of time series analysis to detect climatic change, *J. Hydrol.*, **111**, 259–279 (1989).
4. A. W. Jayarwardena and F. Lai, Time series analysis of water quality data in Pearl River, *China J. Env. Engng*, ASCE, **155**, 590–607 (1989).
5. R. J. Henderson, Rainfall time series for storm overflow assessment, *Wat. Sci. Technol.*, **21**, 1789–1791 (1989).
6. V. Thapliyal, Prediction of Indian monsoon variability evaluation and prospects including development of a new model. In *Climate of China and global climate* (D. Ye, C. Fu, J. Chano and M. Yoshino (eds)), China Ocean Press, 397–416 (1987).
7. K. Krishna Kumar, M. K. Soman and K. Rupa Kumar, Seasonal forecasting of Indian summer monsoon rainfall, A review, *Weather*, 449–467 (1995).
8. J. G. Charney and J. Shukla, *Predictability of monsoons. Monsoon dynamics* (Sir James Lighthill and R. P. Pearce (eds)), Cambridge University Press (1981).
9. M. Rajeevan, Prediction of Indian summer monsoon. Status, problem and prospects, *Curr. Sci.*, **81**, 1451–1457 (2001).
10. V. Thapliyal, Large-scale prediction of summer monsoon rainfall over India: Evolution and development of new models, *Mausam*, **41**, 339–346 (1995).
11. V. Gowariker, V. Thapliyal, R. Sarker, G. S. Mandal and D. R. Sikka, Parametric and power regression models. New approach to long range forecasting of monsoon rainfall in India, *Mausam*, **40**, 115–122 (1989).
12. V. Gowariker, V. Thapliyal, S. M. Kulshrestha, G. S. Mandal, N. Sen Roy and D. R. Sikka, A power regression model for long range forecast of southwest monsoon rainfall over India, *Mausam*, **42**, 125–130 (1991).
13. V. Thapliyal, Long range forecast of summer monsoon rainfall over India: Evolution and development of a new power transfer model, *Proc. Indian Natn. Sci. Acad. A*, **67**, 343–359 (2001).
14. H. D. Navone and H. A. Ceccatto, Predicting Indian monsoon rainfall. A neural network approach, *Clim. Dynamics*, **10**, 305–312 (1994).
15. P. Goswami and Srividya, A novel neural network design for long range prediction of rainfall pattern, *Curr. Sci.*, **70**, 447–457 (1996).
16. P. M. Robinson, Efficient tests of nonstationary hypotheses, *J. Am. Stat. Assoc.*, **89**, 1420–1437 (1994).
17. B. Mandelbrot, *Fractals: Form, chance and dimension*, Freeman (1977).
18. C. W. J. Granger, Long memory relationships and the aggregation of dynamic models, *J. Econ.*, **14**, 227–238 (1980).
19. C. W. J. Granger, Some properties of time series data and their use in econometric model specification, *J. Econ.*, **16**, 121–130 (1981).

20. C. W. J. Granger and R. Joyeux, An introduction to long memory time series and fractionally differencing, *J. Time Ser. Anal.*, **1**, 15–29 (1980).

21. J. R. M. Hosking, Modelling persistence in hydrological time series using fractional differencing, *Wat. Resources Res.*, **20**, 1898–1908 (1981).

22. R. K. Adenstedt, On large sample estimation for the mean of a stationary random sequence, *Ann. Stat.*, **2**, 259–272 (1974).

23. M. S. Taqqu, Weak convergence to fractional motion and to the Rosenblatt process, *Z. Wahrscheinlichkeitstheorie verw.* Geb., **31**, 287–302 (1975).

24. P. M. Robinson, Statistical inference for a random coefficient autoregressive model, *Scand. J. Stat.*, **5**, 163–168 (1978).

25. R. Cioczek-Georges and B. B. Mandelbrot, A class of micropulses and anti-persistent fractional Brownian motion, *Stoch. Proc. Appl.*, **60**, 1–18 (1995).

26. M. Chambers, Long memory and aggregation in macroeconomic time series, *Int. Econ. Rev.*, **39**, 1053–1072 (1998).

27. M. S. Taqqu, W. Willinger and R. Sherman, Proof of a fundamental result in self-similar traffic modelling, *Comp. Commun. Rev.*, **27**, 5–23 (1997).

28. M. Lippi and P. Zaffaroni, *Contemporaneous aggregation of linear dynamic models in large economies*, Manuscript, Research Department, Bank of Italy (1999).

29. W. R. Parke, What is fractional integration?, *Rev. Econ. Stat.*, **81**, 632–638 (1999).

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

31. A. Montanari, R. Rosso and M. S. Taqqu, Some long-run properties of rainfall records in Italy, *J. Geophys. Res. Atm.*, **101**, 431–438 (1996).

32. F. X. Diebold and G. D. Rudebusch, Long memory and persistence in the aggregate output, *J. Monetary Econ.*, **24**, 189–209 (1989).

33. R. T. Baillie, Long memory processes and fractional integration in econometrics, *J. Econ.*, **73**, 5–59 (1996).

34. J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, Long-range dependence in variable-bit-rate video traffic, *IEEE Trans. Commun.*, **43**, 1566–1579 (1995).

35. W. Willinger, M. S. Taqqu, W. E. Leland and D. V. Wilson, Self similarity in high-speed packet traffic. Analysis and modelling of ethernet traffic measurements, *Stat. Sci.*, **10**, 67–85 (1995).

36. R. Fox and M. S. Taqqu, Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Ann. Stat.*, **14**, 517–532 (1986).

37. R. Dahlhaus, Efficient parameter estimation for self-similar process, *Ann. Stat.*, **17**, 1749–1766 (1989).

38. F. Sowell, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *J. Econ.*, **53**, 165–188 (1992).

39. J. Smith, N. Taylor and S. Yadav, Comparing the bias and misspecification in ARFIMA models, *J. Time Ser. Anal.*, **18**, 507–527 (1997).

40. M. A. Hauser, Maximum likelihood estimators for ARFIMA models: A Monte Carlo study, *J. Stat. Plan. Inf.*, **80**, 229–255 (1999).

41. A. Montanari, R. Rosso and M. S. Taqqu, *A seasonal fractional differenced ARIMA model: An application to the River Nile monthly flows at Aswan*, Preprint (1995).

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

43. A. Montanari, M. S. Taqqu and V. Teverovsky, Estimating long range dependence in the presence of periodicity. An empirical study, *Math. Comp. Mod.*, **29**, 217–238 (1999).

44. D. A. Dickey, and W. A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, *J. Am. Stat. Assoc.*, **74**, 427–431 (1979).

45. P. C. B. Phillips and P. Perron, Testing for a unit root in a time series regression, *Biometrika*, **75**, 335–346 (1988).
46. D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root, *J. Econ*., **54**, 159–178 (1992).
47. L. A. Gil-Alana and P. M. Robinson, Testing of unit roots and other nonstationary hypotheses in macroeconomic time series, *J. Econ.*, **80**, 241–268 (1997).
48. L. A. Gil-Alana, Mean reversion in the real exchange rates, *Econ. Lett.*, **16**, 285–288 (2000).
49. L. A. Gil-Alana and P. M. Robinson, Testing seasonal fractional integration in the UK and Japanese consumption and income, *J. Appl. Econ.*, **16**, 95–114 (2001).
50. L. A. Gil-Alana, Testing fractional integration with monthly data, *Econ. Modelling*, **16**, 613–629 (1999).
51. L. A. Gil-Alana, Testing stochastic cycles in macroeconomic time series, *J. Time Ser. Anal.*, **22**, 411–430 (2001).
52. D. A. Mooley and R. Parthasarathy, Fluctuations in all India summer monsoon rainfall during 1871–1978, *Clim. Change*, **6**, 287–301 (1984).
53. S. Gregory, Macro-regional definition and characteristics of Indian summer monsoon rainfall, 1871–1985, *Int. J. Clim.*, **9**, 465–483 (1989).
54. K. Rupa Kumar, G. B. Pant, B. Parthasarathy and N. A. Sontakke, Spatial and subseasonal patterns of the long term trends of Indian summer monsoon rainfall, *Int. J. Clim.*, **12**, 257–268 (1992).
55. S. Hastenrath and L. Greischar, Changing predictability of Indian monsoon rainfall anomalies, *Proc. Indian Acad. Sci.*, **102**, 35–47 (1993).
56. R. N. Iyengar and P. Basak, Regionalization of Indian monsoon rainfall and long term variability signals, *Int. J. Clim.*, **14**, 1095–1114 (1994).
57. S. Gadgil, J. Srinivasan, Ravi S. Nanjundiah, K. Krishna Kumar, A. Munot and K. Rupa Kumar, On forecasting the Indian summer monsoon: The intriguing season of 2002, *Curr. Sci.*, **83**, 394–403 (2002).
58. R. N. Iyengar and S. T. G. Raghu Kanth, Empirical modelling of Indian monsoon rainfall, *Curr. Sci.*, **85**, 1189–1201 (2003).
59. P. Bloomfield, An exponential model in the spectrum of a scalar time series, *Biometrika*, **60**, 217–226 (1973).
60. J. Beran, Fitting long memory models by generalized linear regressions, *Biometrika*, **80**, 817–822 (1993).
61. C. Velasco and P. M. Robinson, Whittle pseudo-maximum likelihood estimation for nonstationary time series, *J. Am. Stat. Assoc.*, **95**, 1229–1243 (2000).
62. L. A. Gil-Alana, An exponential spectral model for the UK unemployment, *J. Forecasting*, **20**, 329–340 (2001).
63. B. Kedem and L. S. Chiu, On the log-normality of rain rate, *Proc. Natn. Acad. Sci. USA*, **84**, 901–905 (1987).
64. L. Goddard, S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher and M. A. Caner, *Int. J. Clim.*, **21**, 1111–1152 (2001).
65. J. H. Stock and M. W. Watson, Macroeconomic forecasting using diffusion indexes, *J. Bus. Econ. Stat.*, **20**, 147–162. (2002).
66. R. L. Smith, T. M. L. Wigley and B. D. Santer, A bivariate time series approach to anthropogenic trend detection in hemispheric time series, *J. Clim.*, **16**, 1228–1240 (2003).