



# Structural break or long memory: an empirical survey on daily rainfall data sets across Malaysia

F. Yusof<sup>1</sup>, I. L. Kane<sup>1</sup>, and Z. Yusop<sup>2</sup>

<sup>1</sup>Department of Mathematical Sciences, Faculty of Science University of Technology Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

<sup>2</sup>Institute of Environmental and Water Resources Management, University of Technology Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

Correspondence to: I. L. Kane (ibrahimlk131@gmail.com)

Received: 8 October 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 31 October 2012

Revised: 25 January 2013 – Accepted: 17 March 2013 – Published: 8 April 2013

**Abstract.** A short memory process that encounters occasional structural breaks in mean can show a slower rate of decay in the autocorrelation function and other properties of fractional integrated  $I(d)$  processes. In this paper we employed a procedure for estimating the fractional differencing parameter in semiparametric contexts proposed by Geweke and Porter-Hudak (1983) to analyse nine daily rainfall data sets across Malaysia. The results indicate that all the data sets exhibit long memory. Furthermore, an empirical fluctuation process using the ordinary least square (OLS)-based cumulative sum (CUSUM) test for the break date was applied. Break dates were detected in all data sets. The data sets were partitioned according to their respective break date, and a further test for long memory was applied for all subseries. Results show that all subseries follows the same pattern as the original series. The estimate of the fractional parameters  $d_1$  and  $d_2$  on the subseries obtained by splitting the original series at the break date confirms that there is a long memory in the data generating process (DGP). Therefore this evidence shows a true long memory not due to structural break.

in distinguishing between non-stationarity, sample variability and long-term climatic fluctuations (Brath et al., 1999).

Studies by Cheung (1993) and Diebold and Inoue (2001) have shown that there is a bias in favour of finding long-memory processes when structural breaks are not accounted for in a time series. Observed long memory behaviour can be due to neglected structural breaks. The presence of breaks in a time series and/or of long memory behaviour in the break-free series could indicate whether the series shows real breaks or long memory. A further insight can be obtained by estimating the difference parameter on the subseries identified by splitting the original series according to the estimated break dates: in the case of erroneous long memory identification, the subseries are expected to show short memory.

If a time series has one or more structural breaks, then the series has one or more discontinuities in the data-generating process (DGP). In this case a structural break method will report a number of breaks which will divide the series into regimes which are of different subpopulations. The statistical properties of these subpopulations within the regimes will need to be estimated. The estimated differences will be the result of actual differences between the samples. The task of estimating the break dates can be accomplished within the framework of least squares regression (Rea et al., 2009).

In the last decade, a lot of interest has been paid to the issue of confusing long memory and occasional structural breaks in mean (see among others Diebold and Inoue, 2001; Granger and Hyung, 2004; and Smith, 2005). Indeed, there is evidence that a stationary short memory process that encounters occasional structural breaks in mean can show a slower

## 1 Introduction

Recent analyses have claimed the possible presence of non-stationarity is produced by the presence of either trend or long-term cyclic fluctuations. However, it is well known that a reliable assessment of the presence of non-stationarity in hydrological records is not an easy task because of the limited extension of the available data sets. This causes difficulty

rate of decay in the autocorrelation function and other properties of fractionally integrated ( $I(d)$ ) processes (Cappelli and Angela, 2006). Therefore, a time series with structural breaks can generate a strong persistence in the autocorrelation function, which is an observed behaviour of a long memory process. On the contrary, long memory processes may cause breaks to be detected spuriously. In this paper, we used the term true long memory to refer to fractionally integrated series.

The literature on the tests to distinguish between true long memory and various spurious long memory models has been steadily growing. For example, Berkes et al. (2006) and Shao (2011) proposed a testing procedure to discriminate a stationary long memory time series from a short-range dependent time series with change points in the mean. Their null hypothesis corresponds to changes in the mean, and their alternative is that the series is stationary with long memory. The test statistic is a modification of the cumulative sum (CUSUM-type) test, which is quite popular in the literature of change point detection (e.g. Galeano, 2007; Shao and Zhang, 2010).

Since the work of Hurst (1951) that detected the presence of long-term persistence in minimum annual flow series of the Nile River, a lot of studies have been carried out for testing and modelling long memory in hydrological processes. For example, Montanari et al. (1997) studied the monthly and daily inflows of Lake Maggiore, Italy, using fractional an autoregressive integrated moving average (ARIMA) model. Rao and Bhattacharya (1999) studied the memory of four hydrologic time series in Midwestern United States by testing the null hypothesis that there is only short-term memory in the series using a modified version of the rescaled range method. The main conclusion from the study is that there is little evidence of long-term memory in monthly hydrologic series. However, for annual series, the evidence for lack of long-term memory is inconclusive, mainly because the number of observations is small and the power of the test based on modified rescaled range is low with small samples. Koutsoyiannis (2002) proposed a simple explanation of the Hurst phenomenon based on the fluctuation of a hydrological process at different temporal scales. The stochastic process that was devised to represent the Hurst phenomenon, i.e. the fractional Gaussian noise, was also studied. Based on its studied properties, three simple and fast methods to generate fractional Gaussian noise are proposed. Wang et al. (2005) analysed two daily stream flow series of the Yellow River, and found that both daily stream flow processes exhibit a strong long memory. In the subsequent study, Wang et al. (2007) applied four methods: the Lo's modified rescaled adjusted range test (R/S test), the GPH test (Geweke and Porter-Hudak, 1983) and two approximate maximum likelihood estimation (MLE) methods, i.e. Whittle's estimator (W-MLE) and another one implemented in S-Plus (S-MLE) to the daily average discharge series recorded at 31 gauge stations with different drainage areas in eight river basins

in Europe, Canada and USA to detect the existence of long memory. The results show that 29 out of the 31 daily series exhibit long memory as confirmed by at least three methods, whereas the other two series are indicated to have long memory with two methods. Gil-Alana (2012) analysed the UK monthly rainfall data from a long-term persistence viewpoint using different modelling approaches, 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. Other applications of long memory models to hydrological time series can be found in Hosking (1984), Koutsoyiannis (2003, 2005a, b, 2006), Koscielny-Bunde et al. (2006), Rybski et al. (2006), Langousis and Koutsoyiannis (2006), and Mudelsee (2007).

Several authors did a reasonably good job in testing for long memory in hydro-meteorological time series; however, is it really a long memory process or a short memory with structural break? This is the main concern of this paper. In order to have an insight on "which is which", we propose fitting long memory and structural break separately following Cappelli and Angela (2006). In case both provide a plausible explanation of the DGP of the data at hand, the long memory and structural break analysis are repeated on the break-free series and on the filtered series, respectively. However in Malaysia and other tropical regions, little or not much has been done in this area.

Heavy rainfall could bring disaster such as floods and landslides. Of course, the shortage of rainfall could also affect the water management system in such a way it could cause problems to the economic activities. Therefore, there is a need to investigate the characteristic of rainfall of a country intensively and comprehensively. Modelling of daily rainfall using various mathematical models has been done throughout the world to give a better understanding about the rainfall pattern and its characteristics (e.g. Suhaila and Abdul Aziz, 2008).

The purpose of this paper is to apply a simple strategy to study whether the daily rainfall data sets of nine weather stations across Malaysia exhibit true long memory behaviour, since the presence of long memory in hydrological time series is a well-known phenomenon and, on the other hand, the presence of structural breaks in these series represents a relevant environmental issue.

The long memory or long-term dependence property describes the high-order correlation structure of a series. Thus, autocorrelation provides initial information relevant to the internal organisation of each set of time series data (Mahdi and Petra, 2002). The prevalence of autocorrelation in a data series is also an indication of persistence in the series of observations. The autocorrelation coefficients provide an essential hint whether forecasting models can be developed based on the given data (Janssen and Laatz, 1997). If a series exhibits long memory, there is persistent temporal dependence even between distant observations; such series are characterized

by distinct but non-periodic cyclical patterns. Fractionally integrated processes can give rise to long memory (Alptekin, 2006). Fractional integration is part of the larger classification of time series, commonly referred to as long memory models. Long memory models address the degree of persistence in the data. In empirical modelling of long memory processes, the autoregressive fractionally integrated moving average (ARFIMA) model that was proposed by Granger and Joyeux (1980) and Hosking (1981) is used.

## 2 Materials and methods

### 2.1 Statistical model

A time series process  $\{X_t, t = 0, \pm 1, \dots\}$  is said to be (co-)variance stationary if the mean and the variance do not depend on time, and the covariance between any two observations depends on the temporal distance between them but not on their specific location in time. This is a minimal requirement in time series analysis to make statistical inference.

Given a zero-mean covariance-stationary process  $\{X_t, t = 0, \pm 1, \dots\}$ , with autocovariance function  $\gamma_\mu = E\{X_t, X_{t+\mu}\}$ , we say that  $X_t$  is integrated of order zero (denoted by  $X_t \approx I(0)$ ) if

$$\sum_{\mu=-\infty}^{\infty} |\gamma_\mu| < \infty. \tag{1}$$

If a time series is non-stationary, one possibility for transforming the series into a stationary one is to take first differences of the series, such that

$$(1 - B)X_t = \mu_t, \quad t = 1, 2, \dots \tag{2}$$

where  $B$  is the lag-operator ( $BX_t = X_{t-1}$ ) and  $\mu_t$  is  $I(0)$ . In such a case,  $X_t$  is said to be integrated of order 1, denoted  $I(1)$ . Likewise, if two differences are required, the series is integrated of order 2, denoted  $I(2)$ . If the number of differences required to get  $I(0)$  stationary is not an integer value but a fractional one, the process is said to be fractionally integrated or  $I(d)$ . Therefore,  $X_t$  is  $I(d)$  if

$$(1 - B)^d X_t = \mu_t, \quad t = 1, 2, \dots \tag{3}$$

with  $\mu_t$  equal to  $I(0)$ . The expression in the left-hand side in Eq. (3) can be presented in terms of a binomial expansion, such that, for all real  $d$ ,

$$\begin{aligned} (1 - B)^d &= \sum_{r=0}^d \binom{d}{r} B^r (-1)^r \\ &= 1 - dB + \frac{d(d-1)}{2} B^2 - \frac{d(d-1)(d-2)}{3!} B^3 + \dots \end{aligned}$$

Therefore, Eq. (3) can be written in the following form:

$$x_t = dx_{t-1} - \frac{d(d-1)}{2} x_{t-2+\dots+u_t}. \tag{4}$$

If  $d$  is a positive integer value,  $X_t$  will be a function of a finite number of past observations, while if  $d$  is not an integer,  $X_t$  depends strongly upon values of the time series far in the past (e.g. Granger and Ding, 1996; Dueker and Asea, 1995). Moreover, the higher the value of  $d$ , the higher will be the level of association between the observations (Gil-Alana, 2009).

The parameter  $d$  plays an important role from a statistical viewpoint. Thus, if  $-0.5 < d < 0.5$ ,  $\mu_t$  is a stationary and ergodic process with a bounded and positively valued spectrum at all frequencies. One important class of processes occurs when  $\mu_t$  is  $I(0)$  and is covariance stationary. For  $0 < d < 0.5$ , the process exhibits long memory in the sense of Eq. (1); its autocorrelations are all positive and decay at a hyperbolic rate. For  $-0.5 < d < 0$ , the sum of absolute values of the process autocorrelations tends to constant, so that it has short memory according to Eq. (1). In this situation the ARFIMA  $(0, d, 0)$  process is said to be anti-persistent or to have intermediate memory, and all its autocorrelations excluding lag zero are negative and decay hyperbolically to zero. As  $d$  increases beyond 0.5 and through 1 (the unit root case),  $X_t$  can be viewed as becoming “more nonstationary” in the sense, for example, that the variance of the partial sums increases in magnitude. This is also true for  $d > 1$  (Gil-Alana, 2009).

### 2.2 Test and estimation of order of integration

There exist several procedures for estimating the fractional differencing parameter in semiparametric contexts. Of these, the log-periodogram regression estimate proposed by Geweke and Porter-Hudak (1983) has been the most widely used (Shimotsu, 2002).

Given a fractional integrated process  $\{Y_t\}$ , its spectral density is given by

$$f(\omega) = [2 \sin(\omega/2)]^{-2d} f_u(\omega), \tag{5}$$

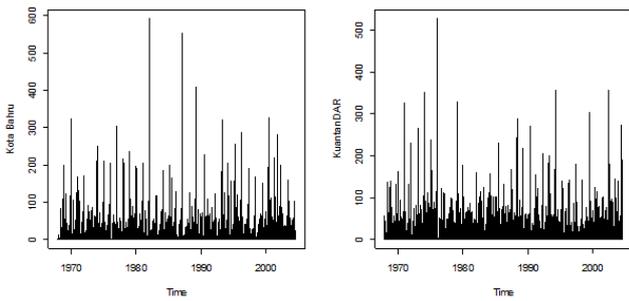
where  $\omega$  is the Fourier frequency,  $f_u(\omega)$  the spectral density corresponding to  $u_t$ , and  $u_t$  a stationary short memory noise with 0 mean. Consider the set of harmonic frequencies  $\omega_j = (2\pi j/n)$ ,  $j = 0, 1, \dots, n/2$ , where  $n$  is the sample size. Taking the logarithm of Eq. (4) we have

$$\ln f(\omega_j) = \ln f_u(0) - d \ln[4 \sin^2(\omega_j/2)]. \tag{6}$$

Equation (5) can be re-written in an alternative form following (Wang et al., 2007) as

$$\ln f(\omega_j) = \ln f_u(0) - d \ln[4 \sin^2(\omega_j/2)] + \ln \left[ \frac{f_u(\omega_j)}{f_u(0)} \right]. \tag{7}$$

The fractional differencing parameter  $d$  can be estimated by the regression equations constructed from Eq. (6). Using the periodogram estimate of  $f(\omega_j)$ , if the number of frequencies  $m$  used in Eq. (6) is a function  $g(n)$  (a positive integer) of the sample size  $n$ , where  $m = g(n) = n^\alpha$  with  $0 < \alpha < 1$ , it can be demonstrated that the least squares estimate  $\hat{d}$  using



**Fig. 1.** Time series plots of daily rainfall series of Kota Bharu and Kuantan.

the above regression is asymptotically normally distributed in large samples (Geweke and Porter-Hudak, 1983).

$$\hat{d} \sim N \left( d, \frac{\pi^2}{6 \sum_{j=1}^{g(n)} (U_j - \bar{U})^2} \right),$$

where  $U_j = \ln[4 \sin^2(\omega_j/2)]$  and  $\bar{U}$  is the sample of  $U_j$ ,  $j = 1 \dots g(n)$ .

Under the null hypothesis, of no long memory ( $d = 0$ ),

the  $t$  statistic  $t_{d=0} = \hat{d} \left( \frac{\pi^2}{6 \sum_{j=1}^{g(n)} (U_j - \bar{U})^2} \right)$  has limiting standard

normal distribution.

The value of the power factor  $\alpha$  is the main determinant of the ordinates included in the regression. Traditionally the number of periodogram ordinates  $m$  is chosen from the interval  $[T^{0.45}, T^{0.55}]$ . However, Hurvich and Deo (1998) showed that the optimal  $m$  is of order  $O(T^{0.8})$ .

**2.3 OLS-based CUSUM test**

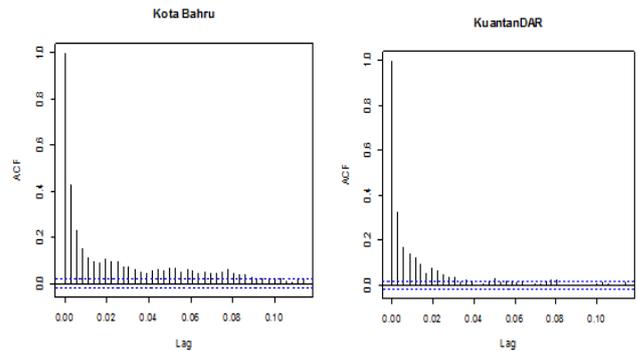
The CUSUM tests are concerned with testing against the alternative that an unknown coefficient vector varies over time (Zeileis, 2000). Ploberger and Kramer (1992) proposed a test based on ordinary least square (OLS) residuals. The OLS-based CUSUM test uses the OLS residuals  $\hat{u}_t = y_t - x_t^T \hat{\beta}$ . The OLS-based CUSUM-type empirical fluctuation process is defined as

$$W_n^0(t) = \frac{1}{\hat{\sigma} \sqrt{n}} \sum_{i=1}^{[nt]} \hat{u}_i, (0 \leq t \leq 1),$$

where  $\hat{\sigma} = \sqrt{\frac{1}{n-k} \sum_{t=1}^n \hat{u}_t^2}$ . This path will always start with 0 at  $t = 0$  and it also returns to 0 for  $t = 1$ , but if there is structural change at  $t_0$  it should have a peak close to the break point  $t_0$ . The null hypothesis  $H_0$  is rejected if the path crosses either

**Table 1.** Selected weather stations in Malaysia and their general geographic information.

Station	Latitude (North)	Longitude (East)	Ht. a.m.s.l. (m)	Duration (Daily)
Alor Setar	6°12'	100°24'	3.9	01/01/1968–31/12/2003
Bayan Lepas	5°18'	100°16'	2.8	01/01/1968–31/12/2003
Kota Bharu	6°10'	102°17'	4.6	01/01/1968–31/12/2003
Ipoh	4°34'	101°06'	40.1	01/01/1968–31/12/2003
Kuantan	3°47'	103°13'	15.3	01/01/1968–31/12/2003
Malacca	2°16'	102°15'	8.5	01/01/1968–31/12/2003
Mersing	2°27'	103°50'	43.6	01/01/1968–31/12/2003
Sitiawan	4°13'	100°42'	7.0	01/01/1968–31/12/2003
Subang	3°07'	101°33'	16.5	01/01/1968–31/12/2003



**Fig. 2.** Autocorrelation function of the data sets of Kota Bharu and Kuantan.

of the boundaries  $(-\lambda, \lambda)$ , which is equivalent to rejecting when the test statistic

$$S^0 = \sup_{0 \leq t \leq 1} |W_n^0(t)|$$

is larger than  $\lambda$ , which determines the significance level of the test. As  $n \rightarrow \infty$ ,

$$W_n^0(t) \xrightarrow{d} B^0(t),$$

where  $B^0(t)$  is the standard Brownian bridge. Details on this test can be found in Ploberger and Kramer (1992).

**3 Results and discussion**

Data of daily rainfall record for nine stations across Malaysia for the period 1 January 1968–31 December 2003 obtained from the Malaysian Meteorological Department were analysed in this study. Table 1 illustrates the generalized geographic information of the selected weather stations used for this study. We start the analysis by discussing the descriptive statistic of the considered daily rainfall series. Table 2a–c depict a summary of the statistical properties for the full rainfall series, subseries before break and the subseries after the break. It could be observed from the tables that both series follow the same statistical patterns with standard deviation greater than the mean and high values for kurtosis

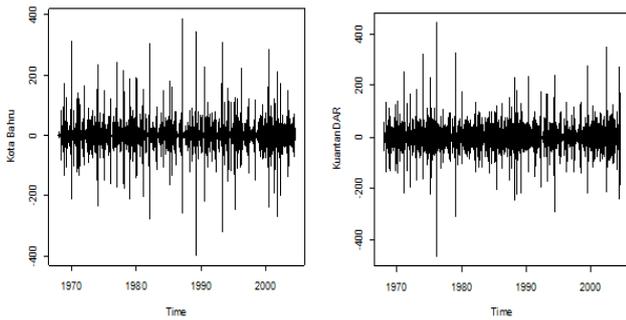


Fig. 3. Time plot of the first difference series of Kota Bharu and Kuantan.

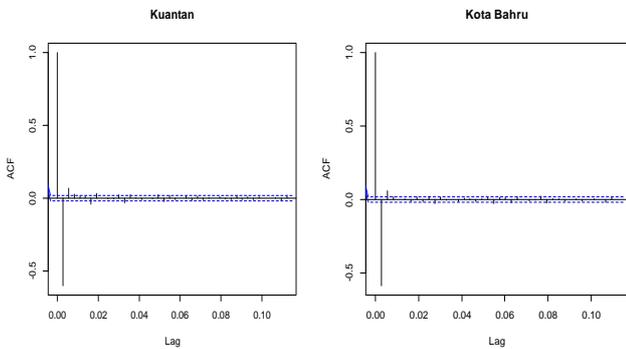


Fig. 4. Autocorrelation functions of the first difference series

and skewness. The skewness indicates that the mass of the distribution is concentrated on the right, or we may say that the distribution is left-skewed. The positive kurtosis shows a leptokurtic condition, and has a fatter tail than normal. The standard deviations given are greater than the corresponding mean values; this indicates that the daily rainfall fluctuates significantly through time.

Figure 1 depicts time series plots for Kota Bharu and Kuantan stations. From the figure, the time series plots show very persistent behaviour. The autocorrelation function (Fig. 2) provides a measure of temporal correlation between rainfall data points with different time lags. For a purely random event, all autocorrelation coefficients are zero, apart from  $r(0)$ , which is equal to 1. The autocorrelation function plots in Fig. 2 decay with hyperbolic rates and indicate that the time series are strongly correlated, i.e. decay up to long lags. These are the main characteristics of long memory appearance.

The series may now be stationary by taking the first difference displayed in Fig. 3, though the correlogram (Fig. 4) still shows significant values even at some lags relatively away from zero and clearly shows that the autocorrelation at lag one exceeds  $-0.5$ , which suggests overdifferencing (Van Beusekom, 2003). Therefore, taking the first difference is not appropriate for the series(s) under consideration. This may be an indication that fractional differencing smaller than or greater than 1 may be more appropriate than first differences.

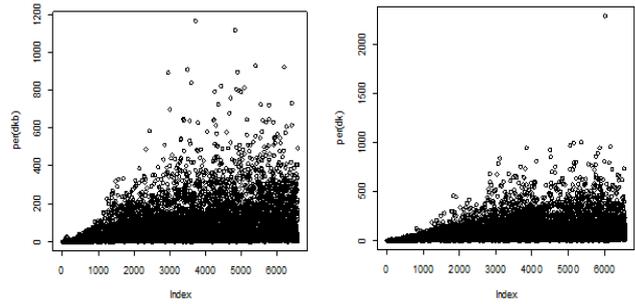


Fig. 5. Periodogram of the first difference series of Kota Bharu and Kuantan.

Table 2a. Summary statistics for the rainfall data sets.

Station	Mean	Standard deviation	Skewness	Kurtosis
Alor Setar	5.5405	12.760	4.0517	24.062
Bayan Lepas	6.6429	15.572	4.2215	27.578
Ipoh	6.7220	13.818	3.1798	12.550
Kota Bharu	7.1073	22.368	8.4517	120.06
Kuantan	7.9802	21.812	7.0057	80.590
Malacca	5.4457	12.966	4.3672	32.687
Mersing	7.3053	20.087	7.2462	83.754
Sitiawan	4.8253	11.760	4.1344	23.486
Subang	6.6938	13.8437	3.4519	16.405

In addition, the periodograms (Fig. 5) show values close to 0 at the zero frequency, which might suggest that these series are overdifferenced (Caporale and Gil-Alana, 2004).

It is well known fact that, when analysing climatic time series, distinguishing between long-term fluctuations and non-stationarity is not a simple task. Nevertheless, such a distinction would be extremely interesting. In fact, the presence of long-term climatic fluctuations, rather than non-stationarity, would imply that the patterns found in the data set could likely be attributed to cyclical behaviour rather than to irreversible tendencies. A useful tool for the above-mentioned distinction is the detection of the possible presence of long memory in the data. As mentioned before, one of the effects of long memory is the behaviour of the time series to be subjected to long-term cycles.

A time series which is generated by a true long memory process has a uniform data DGP throughout the entire series. Thus if a structural break location method is mistakenly applied to the series, it may report a number of breaks where no breaks exist. These spurious breaks will yield a number of partitions of differing lengths, but this partition will only be subsamples of a single population. Thus the subsamples will have the same statistical properties as the full series, because subsamples have been drawn from a single population. Any estimated differences will be the result of randomness and long range serial correlation but not differences between the samples. If a time series has one or more structural breaks, then the series has one or more discontinuities in the DGP. In

**Table 2b.** Statistics for the daily rainfall series before break.

Station	Mean	Standard deviation	Skewness	Kurtosis
Alor Setar	6.25	13.92	3.91	25.15
Bayan Lepas	6.46	15.12	4.00	24.99
Ipoh	6.57	13.62	3.20	15.63
Kota Bharu	6.74	22.24	9.49	153.17
Kuantan	7.72	21.26	7.15	89.79
Malacca	5.21	12.84	5.297	58.96
Mersing	7.55	20.89	7.23	85.32
Sitiawan	4.61	11.56	4.42	31.14
Subang	6.13	12.89	3.34	19.08

**Table 2c.** Statistics for the daily rainfall series after break.

Station	Mean	Standard deviation	Skewness	Kurtosis
Alor Setar	5.31	12.35	4.09	27.55
Bayan Lepas	7.61	17.78	4.81	43.95
Ipoh	7.58	14.89	3.05	14.82
Kota Bharu	8.03	22.65	6.01	53.50
Kuantan	9.39	24.66	6.31	59.78
Malacca	5.59	13.04	3.85	22.98
Mersing	6.44	16.94	6.84	79.77
Sitiawan	5.08	11.99	3.83	21.70
Subang	7.38	14.89	3.40	18.86

this case a structural break method will report a number of breaks which will divide the series into regimes, which are different subpopulations. The statistical properties of these subpopulations within the regimes will need to be estimated. The estimated differences will be the result of actual differences between the samples.

The results for the estimated values of the fractional differencing parameters along with ARFIMA ( $p, d, q$ ) model for all data sets are displayed in Table 3. From Table 3, we can observe that all the rainfall series can be described in terms of fractional integration, which is part of the larger classification of time series, commonly referred to as long memory models. In order to show how the above-described strategy can help in dealing with the discrimination between long memory and structural breaks, we have analysed the series(s) on which we have identified an ARFIMA (0,  $d$ , 0) model. The maximum likelihood estimate of the fractional parameter  $d$  was obtained by means of GPH. The presence of long memory in all data sets is displayed in Table 3. At the same time, we have detected the regime break using the OLS-based CUSUM test for break date. The break-date results found in these data sets differ from station to station. The series were partitioned according to their break dates, and all subseries were tested for long memory. Table 4 gives the break date along with the estimates of the differencing parameters  $d_1$ , and  $d_2$  for series before and after break respectively.

**Table 3.** Differencing parameter estimates with ARFIMA (0,  $d$ , 0) model.

Station	$d$	ARFIMA (0, $d$ , 0)
Alor Setar	0.03750534	$(1 - L)^{0.03750534} X_t = \mu_t$
Bayan Lepas	0.02780336	$(1 - L)^{0.02780336} X_t = \mu_t$
Ipoh	0.05657603	$(1 - L)^{0.05657603} X_t = \mu_t$
Kota Bharu	0.09306547	$(1 - L)^{0.09306547} X_t = \mu_t$
Kuantan	0.1329785	$(1 - L)^{0.1329785} X_t = \mu_t$
Malacca	0.06142958	$(1 - L)^{0.06142958} X_t = \mu_t$
Mersing	0.1072815	$(1 - L)^{0.1072815} X_t = \mu_t$
Sitiawan	0.04108584	$(1 - L)^{0.04108584} X_t = \mu_t$
Subang	0.05938356	$(1 - L)^{0.05938356} X_t = \mu_t$

**Table 4.** CUSUM test with corresponding break date and estimates of the differencing parameters  $d_1$  (before break) and  $d_2$  (after break).

Station	break date (day)	@ observation	$d_1$	$d_2$
Alor Setar	1977 (15)	3255	0.062	0.054
Bayan Lepas	1998 (287)	11 087	0.051	0.171
Ipoh	1998 (358)	11 158	0.068	0.160
Kota Bharu	1994 (52)	9412	0.112	0.138
Kuantan	1999 (2)	11 162	0.135	0.238
Malacca	1981 (157)	4837	0.020	0.060
Mersing	1996 (190)	10 270	0.130	0.129
Sitiawan	1987 (304)	7144	0.081	0.034
Subang	1987 (357)	7197	0.079	0.060

#### 4 Conclusions

It is now well established that long memory and structural change are easily confused. However, most researchers choose to ignore the problem of structural break in testing for long memory. It is a known fact that short memory with structural break may exhibit the properties of long memory. The main contribution of the paper was to detect if the DGP of daily rainfall series of some locations across Malaysia is generated by a true long memory process. An approach based on fractional integration FI( $d$ ) process that can characterize a series was applied. The findings indicate that all the data sets exhibit long memory, but is it a true long memory? To answer this we employed a method which allowed detecting a structural break in each series. The series were partitioned according to the break date identified, and a similar test was applied to the subseries. We found that all subseries displayed the same properties as the original series. Therefore, all the rainfall data sets considered were generated by a true long memory process.

*Acknowledgements.* The authors thank the Malaysian Meteorological Department for providing the rainfall data. We also thank the Research Management Centre of UTM for facilitating this research and providing research grant.

Edited by: A. Langousis

## References

- Alptekin, N.: Long memory analysis of USD/trial exchange rate, *Int. J. Soc. Sci.*, 1, 111–112, 2006.
- Berkes, I., Horvath, L., Kokoszka, P., and Shao, Q.: On discriminating between long-range dependence and changes in mean, *Ann. Stat.*, 34, 1140–1165, 2006.
- Brath, A., Castellarin, A., and Montanari, A.: Detecting non stationarity in extreme rainfall data observed in Northern Italy, in: *Proceedings of EGS – Plinius Conference on Mediterranean Storms*, Maratea, 219–231, 1999.
- Caporale, G. M. and Gil-Alana, L. A.: Long-run and Cyclical Dynamics in the US Stock Market, *Economics Series*, 155, 1–29, ISSN: 1605-7996, 2004.
- Cappelli, C. and Angela, D.: Long memory and structural break analysis of environmental time series, available at: <http://www.sis-statistica.it/files/pdf/atti/Spontanee%202006\203-206.pdf> (last access: 30 October 2012), 2006.
- Cheung, Y. W.: Tests for fractional integration: a Monte Carlo investigation, *J. Time Ser. Anal.*, 14, 331–345, 1993.
- Diebold, F. X. and Inoue, A.: Long memory and regime switching, *J. Econ.*, 105, 131–159, 2001.
- Dueker, M. J. and Asea, P. K.: Non-Monotonic Long Memory Dynamics in Black-Market Exchange Rates, The Federal Reserve Bank of St. Louis, Working Paper 1995-003A, available at: <http://research.stlouisfed.org/wp/1995/95-003.pdf> (last access: 28 September 2012), 1–24, 1995.
- Galeano, P.: The use of cumulative sums for detection of change points in the rate parameter of a Poisson Process, *Journal Computational Statistics & Data Analysis archive*, 51, 6151–6165, 2007.
- Geweke, J. and Porter-Hudak, S.: The estimation and application of long memory time series models, *J. Time Ser. Anal.*, 4, 221–237, 1983.
- Gil-Alana, L. A.: Persistence and Time Trends in the Temperatures in Spain, *Adv. Meteorol.*, 2009, 415290, doi:10.1155/2009/415290, 2009.
- Gil-Alana, L. A.: U.K. Rainfall Data: A Long-Term Persistence Approach, *J. Appl. Meteorol. Clim.*, 51, 1904–1913, 2012.
- Granger, C. W. J. and Ding, Z.: Varieties of long memory models, *J. Econometrics*, 73, 61–77, 1996.
- Granger, C. W. J. and Hyung, N.: Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *J. Empir. Financ.*, 11, 213–228, 2004.
- Granger, C. W. J. and Joyeux, R.: An introduction to long-range time series models and fractional differencing, *J. Time Ser. Anal.*, 1, 15–30, 1980.
- Hosking, J. R. M.: Fractional differencing, *Biometrika*, 68, 165–176, 1981.
- Hosking, J. R. M.: Modeling persistence in hydrological time series using fractional differencing, *Water Resour. Res.*, 20, 1898–1908, doi:10.1029/WR020i012p01898, 1984.
- Hurst, H. E.: Long term storage capacities of reservoirs, *Trans. Am. Soc. Civ. Eng.*, 116, 776–808, 1951.
- Hurvich, C. M. and Deo, R. S.: Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long memory time series, *J. Time Ser. Anal.*, 3, 1–3, 1998.
- Janssen, J. and Laatz, W.: *Statistische Datenanalyse mit SPSS für Windows*, 2nd Edn., Springer-Verlag, Heidelberg, 636 pp., 1997.
- Koscielny-Bunde, E., Kantelhardt, J. W., Braun, P., Bunde, A., and Havlin, S.: Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies, *J. Hydrol.*, 322, 120–137, 2006.
- Koutsoyiannis, D.: The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrolog. Sci. J.*, 47, 573–595, 2002.
- Koutsoyiannis, D.: Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrolog. Sci. J.*, 48, 3–24, 2003.
- Koutsoyiannis, D.: Uncertainty, entropy, scaling and hydrological stochasticity, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrolog. Sci. J.*, 50, 381–404, 2005a.
- Koutsoyiannis, D.: Uncertainty, entropy, scaling and hydrological stochasticity, 2, Time dependence of hydrological processes and time scaling, *Hydrolog. Sci. J.*, 50, 405–426, 2005b.
- Koutsoyiannis, D.: Nonstationarity versus scaling in hydrology, *J. Hydrol.*, 324, 239–254, 2006.
- Langousis, A. and Koutsoyiannis, D.: A stochastic methodology for generation of seasonal time series reproducing overyear scaling behaviour, *J. Hydrol.*, 322, 138–154, 2006.
- Mahdi, O. and Petra, S.: A Preliminary Assessment of Characteristics and Long-term Rainfall Variability of Rainfall in Ethiopia-Basis for Sustainable Land Use and Resources Management, Conference on International Agricultural Research for Development, Witzenhausen, 9–11 October, available at: <http://www.tropentag.de/2002/abstracts/full/11.pdf> (last access: 30 October 2012), 2002.
- Montanari, A., Rosso, R., and Taqqu, M. S.: Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation, *Water Resour. Res.*, 33, 1035–1044, doi:10.1029/97WR00043, 1997.
- Mudelsee, M.: Long memory of rivers from spatial aggregation, *Water Resour. Res.*, 43, W01202, doi:10.1029/2006WR005721, 2007.
- Ploberger, W. and Kramer, W.: The CUSUM test with OLS residuals, *Econometrica*, 60, 271–285, 1992.
- Rao, A. R. and Bhattacharya, D.: Hypothesis testing for long-term memory in hydrological series, *J. Hydrol.*, 216, 183–196, 1999.
- Rea, W., Oxley, L., and Reale, M.: A New Procedure for Discriminating Between Long Memory and Shifting Means Alternatives, 18th World IMACS/MODSIM Congress, Cairns, Australia 13–17 July 2009, available at: <http://mssanz.org.au/modsim09> (last access: 22 January 2013), 2009.
- Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in climate and the detection problem, *Geophys. Res. Lett.*, 33, L06718, doi:10.1029/2005GL025591, 2006.
- Shao, X.: A simple test of changes in mean in the possible presence of long-range dependence, *J. Time Ser. Anal.*, 32, 598–606, 2011.
- Shao, X. and Zhang, X.: Testing for Change Points in Time Series, *J. Am. Stat. Assoc.*, 105, 1228–1240, doi:10.1198/jasa.2010.tm10103, 2010.
- Shimotsu, K.: Exact local whittle estimation of fractional integration with unknown mean and time trend, Department of Economics, University of Essex, 1–2, 2002.
- Smith, A. D.: Level shifts and the illusion of long memory in economic time series, *J. Bus. Econ. Stat.*, 23, 321–335, 2005.
- Suhaila, J. and Abdul Aziz, J.: Fitting the statistical distribution for daily rainfall in peninsular Malaysia based on AIC criterion, *J. Appl. Sci. Res.*, 4, 1846–1857, 2008.

- Van Beusekom, D. P.: Forecasting & Detrending of Time Series Models, BWI-paper, Free University of Amsterdam, available at: [www.few.vu.nl/nl/Images/werkstuk-beusekom\\_tcm38-91323.doc](http://www.few.vu.nl/nl/Images/werkstuk-beusekom_tcm38-91323.doc) (last access: 1 December 2012), 2003.
- Wang, W., van Gelder, P. H. A. J. M., and Vrijling, J. K.: Long memory properties of streamflow processes of the Yellow River. Proceedings of the International Conference on Water Economics, Statistics and Finance, 1, Rethymno-Crete, Greece, 481–490, 8–10 July 2005.
- Wang, W., Van Gelder, P. H. A. J. M., Vrijling, J. K., and Chen, X.: Detecting long-memory: Monte Carlo simulations and application to daily streamflow processes, *Hydrol. Earth Syst. Sci.*, 11, 851–862, doi:10.5194/hess-11-851-2007, 2007.
- Zeileis, A.: *p* Values and Alternative Boundaries for CUSUM Tests, available at: <http://www.wu-wien.ac.at/am>, 1–16, 2000.