

## Generation of 10-day flow of the Brahmaputra River using a time series model

M. Shahjahan Mondal and Jahir Uddin Chowdhury

### ABSTRACT

Time series models are used in hydrology for the generation of river flow data. The development of such a model, namely deseasonalized Autoregressive Moving Average (ARMA), for the generation of 10-day flows of the Brahmaputra River in Bangladesh is described. The model was fitted following systematic stages of identification, estimation and diagnostic checking of model building. A negative power transformation for the Brahmaputra flow was found to be necessary for model construction. The seasonality of the flow was removed by Fourier analysis using five harmonics for 10-day means and 13 harmonics for standard deviations. The fitted model was ARMA (1, 3) having one autoregressive parameter and three moving average parameters. The validation forecasts made with the model indicated that the deseasonalized ARMA model could capture the 10-day variability of the Brahmaputra flow reasonably well. To further validate and verify the model 200 synthetic flow sequences, each with a length of 50 years, were generated. The fitted ARMA model was found to be capable of preserving both short-term statistics (variance and autocorrelation) and long-term statistics (Hurst coefficient and rescaled adjusted range) of the historic Brahmaputra flow.

**Key words** | ARMA model, Brahmaputra River, deseasonalization, synthetic flow

**M. Shahjahan Mondal** (corresponding author)  
**Jahir Uddin Chowdhury**  
Institute of Water and Flood Management,  
Bangladesh University of Engineering and  
Technology,  
Dhaka, 1000,  
Bangladesh  
E-mail: [mshahjahanmondal@iwfm.buet.ac.bd](mailto:mshahjahanmondal@iwfm.buet.ac.bd)

### INTRODUCTION

Time series models are popular and widely used tools in stochastic river hydrology throughout the world, mainly for medium-range forecasting and the generation of synthetic flows (McKerchar & Delleur 1974; Delleur & Kavvas 1978; Govindasamy 1991; Hipel & McLeod 1994; Young *et al.* 1999; Papamichail & Georgiou 2001; Mondal & Wasimi 2005a, 2005b, 2006, 2007a; Artis *et al.* 2007). Synthetic data are now widely used in water resources planning and simulation studies throughout the developed nations. Typical uses include estimation of command area of a water development project, sizing of water retention structures, evaluation of risk-based performance indicators for water supply or setting a reservoir/barrage operation policy. Mondal & Wasimi (2007b) and Mondal *et al.* (2010) have recently used two such models in a risk-based evaluation of the Ganges and Brahmaputra water developments, respectively, in meeting future water demands within Bangladesh. Time series models of univariate and multivariate,

periodic and non-periodic, and seasonal and non-seasonal types are in use.

There are a number of commonly used univariate models for seasonal forecasting and data generation, such as exponential smoothing, Markov models, Holt-Winters method, Box-Jenkins multiplicative Seasonal Autoregressive Integrated Moving Average (SARIMA) class of models, deseasonalized models, periodic models and disaggregation models. The selection of an appropriate model for analysing a particular problem depends on many factors, such as: number of series to be modelled; modelling costs; required accuracy; ease of use of the models; and ease of interpretation of the results. In the published literature (Newbold & Granger 1974; Hipel & McLeod 1978; Pankratz 1983; Chatfield 1996), it is noted that, when the number of series to be modelled is relatively few and a large expenditure of time and effort can be justified (as in the case of the Brahmaputra River), the Box-Jenkins

doi: 10.2166/nh.2012.242

method (SARIMA) is generally preferred. This choice is due to its inclusion of a family of models which can be fitted to a wide variety of time series processes. An inherent advantage of the SARIMA family of models is that only few model parameters are required for describing time series which exhibit non-stationarity both within and across seasons. Some useful applications of these models in seasonal river flow forecasting are reported in McKerchar & Delleur (1974), Panu *et al.* (1978), Cline (1981), Govindasamy (1991), Irvine & Eberhardt (1992), Sidhu (1995), Papamichail & Georgiou (2001), Mondal (2005) and Mondal *et al.* (2007).

In the case of the SARIMA model, seasonal and/or non-seasonal differencing is applied to remove the intra- and/or inter-year non-stationarity, respectively. Kavvas & Delleur (1975) have shown from both analytical and empirical results that seasonal and/or non-seasonal differencing, although very effective in the removal of hydrologic periodicities, distorts the original spectrum, thus making it impractical or impossible to fit an Autoregressive Moving Average (ARMA) model for hydrologic simulation or synthetic generation. McKerchar & Delleur (1974) and Delleur *et al.* (1976) have also shown that forecasting capabilities of seasonally differenced models may be impaired by the fact that they may not take into account the variation in the seasonal standard deviations. In addition, non-seasonal differencing does not preserve the seasonal structure in forecasting.

When simulation, as well as forecasting, is an objective, another class of hydrologic models called structural models can be used (e.g. Tao & Delleur 1976; Salas *et al.* 1981; Vecchia 1985; Hipel & McLeod 1994). This class of models is suitable for seasonal hydrologic time series which exhibit an autocorrelation structure that depends not only on the time lag between observations but also on the season of the year and which, except for some random variation, possesses second-order stationarity within individual seasons across years (Hipel & McLeod 1994). There are two types of structural models: deseasonalized and periodic. In deseasonalized modelling, the seasonal component of the time series to be modelled is removed by first subtracting each seasonal mean from the corresponding seasonal observations, and then dividing by the respective seasonal standard deviation (if necessary). An appropriate ARMA model is then fitted

to the resulting deseasonalized time series (Lungu & Sefe 1991; Hipel & McLeod 1994). In periodic modelling, the model parameters, as well as model types and orders, are allowed to vary depending on the season of the year. The advantage of a periodic model is that it can account for variability in seasonal standard deviations and correlations that a SARIMA model cannot (Delleur & Kavvas 1978; Hipel & McLeod 1994). However, a potential drawback of using a periodic model in an application is that the model often requires the use of a substantial number of parameters.

In this paper, we develop a deseasonalized ARMA model for the generation of 10-day flow of the Brahmaputra River at Bahadurabad. The necessity for development of such a model emerges due to the fact that available records of the Brahmaputra flow within Bangladesh are of limited lengths, and therefore not suitable for detailed evaluation of development options for the Brahmaputra water in meeting future water demand in the Brahmaputra Dependent Area.

## DESEASONALIZED ARMA MODEL

### Formulation, identification, estimation and diagnostic checking

Let  $x_{r,s}$  represent a time series value in the  $r$ th year and  $s$ th season. For 10-day series of data,  $s = 1, 2, \dots, 36$ . Year and season indices follow modulo  $s$  arithmetic such that  $x_{r,s} = x_{r \pm m, s \pm 36m}$  for 10-day data, where  $m$  is any real integer. If the variable  $x_{r,s}$  is skewed, then an appropriate transformation may be undertaken to make the transformed series  $y_{r,s}$  approximately normal. After transformation, the seasonal mean is removed from the series  $y_{r,s}$  by subtracting the seasonal mean  $\mu^{(s)}$  from each observation and then dividing the result by the corresponding seasonal standard deviation  $\sigma^{(s)}$ . Seasonal means and standard deviations can be obtained by parametric or non-parametric analysis. The non-parametric method requires many parameters to remove seasonality, particularly when a time series spans a week or 10 days. The problem of requiring many parameters can be overcome by using

the parametric method, which is based on the Fourier series approach (Salas *et al.* 1988).

The parametric representation of  $\mu^{(s)}$ , denoted in general as  $\hat{\mu}^{(s)}$ , can be obtained by:

$$\hat{\mu}^{(s)} = \bar{\mu} + \sum_{i=1}^h [A_i \cos(2\pi is/S) + B_i \sin(2\pi is/S)] \quad (1)$$

where  $\bar{\mu}$  is the mean of  $\mu^{(s)}$ ;  $A_i$  and  $B_i$  are the Fourier series coefficients;  $i$  is the harmonic, and  $h$  is the total number of harmonics which is equal to  $S/2$  when  $S$  is even and  $(S-1)/2$  when  $S$  is odd. For instance, a 10-day series has  $S = 36$  and  $h = 18$ . The Fourier coefficients are determined by:

$$A_i = \frac{2}{S} \sum_{s=1}^S \mu^{(s)} \cos(2\pi is/S) \quad \text{and} \\ B_i = \frac{2}{S} \sum_{s=1}^S \mu^{(s)} \sin(2\pi is/S), \quad i = 1, 2, \dots, h \quad (2)$$

When  $S$  is even, the last coefficients  $A_h$  and  $B_h$  are given by:

$$A_h = \frac{1}{S} \sum_{s=1}^S \mu^{(s)} \cos(2\pi hs/S) \quad \text{and} \quad B_h = 0 \quad (3)$$

When  $\hat{\mu}^{(s)}$  of Equation (1) is determined using all the harmonics  $i = 1, 2, \dots, h$ , (i.e. all the coefficients  $A_i$  and  $B_i$ ),  $\hat{\mu}^{(s)}$  is exactly the same as  $\mu^{(s)}$  for all values of  $s = 1, 2, \dots, S$ .

To find out the required number of harmonics and corresponding Fourier coefficients necessary for a good fit in Equation (1) a cumulative periodogram test, which is a graphical test, is usually conducted. This test is the most accurate for selecting the number of significant harmonics (Salas *et al.* 1988). The test is carried out by computing first the mean squared deviation (MSD) of  $\mu^{(s)}$  around  $\bar{\mu}$ :

$$\text{MSD}(\mu) = \frac{1}{S} \sum_{s=1}^S (\mu^{(s)} - \bar{\mu})^2 \quad (4)$$

$\text{MSD}(\mu)$  is composed of the  $\text{MSD}(i)$  of each harmonic  $i$ , which is determined by:

$$\text{MSD}(i) = \frac{1}{2} (A_i^2 + B_i^2), \quad i = 1, 2, \dots, h \quad (5)$$

After computation of all the values of  $\text{MSD}(i)$ , they are arranged in descending order so that  $\text{MSD}(h_i)$  represents the ordered sequence,  $h_1$  being the harmonic with the highest MSD and  $h_h$  with the lowest MSD.  $P_i$ , which is the ratio of the sum of the first  $i$  MSDs to the  $\text{MSD}(\mu)$ , is then computed via:

$$P_i = \frac{\sum_{j=1}^i \text{MSD}(h_j)}{\text{MSD}(\mu)}, \quad i = 1, 2, \dots, h \quad (6)$$

The plot of  $P_i$  versus  $i$  is called the cumulative periodogram, which is composed of two distinct parts: a faster increasing periodic part and a slower increasing sampling part. The two parts are approximated by two smooth curves, the intersecting point of which provides the number of significant harmonics. The separation of the two parts becomes difficult when their intersecting point is in such a position that the two curves are almost one continuous curve. In such a situation, a criterion often used in practice is to select the number of harmonics which explain 90% or 95% of variance,  $\text{MSD}(i)$  (Salas *et al.* 1988).

Let the deseasonalized series be written as:

$$z_t = z_{r,s} = \frac{y_{r,s} - \mu^{(s)}}{\sigma^{(s)}} \quad (7)$$

where  $t = 1, 2, \dots, n$  and  $n$  is the number of observations. An ARMA model is then fitted to the deseasonalized time series  $z_t$ . The general equation of the ARMA model (Box *et al.* 1994) for a variable  $z_t$  is given by:

$$\varphi_p(B)z_t = \theta_q(B)a_t \quad (8)$$

where the polynomials  $\varphi_p(B)$  and  $\theta_q(B)$  are autoregressive (AR) and moving average (MA) operators of order  $p$  and  $q$ , respectively, i.e.

$$\varphi_p(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p) \quad \text{and} \\ \theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \quad (9)$$

$a_t$  is a stochastic random shock component with zero mean, constant variance and no serial correlation (i.e. white noise).

The model selection process consists of three iterative stages: (1) model identification; (2) model parameter

estimation; and (3) diagnostic checking of the model residuals. A detailed description of each of these stages is provided by [Box \*et al.\* \(1994\)](#). At the identification stage, the orders of AR and MA parameters are chosen from the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the variable  $z_t$ . At the estimation stage, maximum likelihood estimates of different model parameters are obtained. Stationarity and invertibility conditions of the AR and MA parameters, respectively, are checked at this stage. Mathematical definitions and physical interpretations of stationarity of AR and invertibility of MA parameters can be found in [Delleur & Kavvas \(1978\)](#), [Vandaele \(1997\)](#) and others. There are some additional statistical tools, such as Akaike Information Criterion (AIC; [Akaike 1974](#)) and Bayes Information Criterion (BIC; [Schwartz 1978](#)), which can be used to select the best model from several possible models.

At the diagnostic checking stage, a decision is made about whether the selected model from the estimation stage is statistically adequate. For this, the model residuals are checked to determine whether they satisfy the assumptions of independence, normality and homoscedasticity (constant variance). The most important of these assumptions is that the random shocks from the estimated model are independent. Residual ACF is the basic analytical tool to test the null hypothesis of white noise residuals through either  $t$ - or  $\chi^2$ -test ([Box & Pierce 1970](#); [Ljung & Box 1978](#)). [Box \*et al.\* \(1994\)](#) have further suggested the cumulative periodogram test for the detection of periodic patterns in a background of white noise.

The hypothesis that a given time series is normal can be tested through a normal probability plot, detrended normal plot, Shapiro–Wilk test ([Shapiro & Wilk 1965](#)), Lilliefors test ([Lilliefors 1967](#)),  $\chi^2$ -test or skewness test ([Salas \*et al.\* 1988](#)). To check for violations of equality-of-variance assumption, a scatter-plot of residuals against fitted values can be produced. If the spread of the residuals does not appear to be increasing or decreasing with the magnitude of the fitted values, the assumption of constant variance is met. Some informal techniques, such as a time series plot of residuals, over-fitting of the selected model and fitting models to subsets of data, often provide valuable information on adequacy of the model and reveal

important clues on how an inadequate model can be reformulated. When the model residuals pass all these checks, the model is considered statistically adequate. In a situation where a model fails to pass any of the checks, the model is reformulated by changing its form or order or by choosing an appropriate data transformation, such as Box–Cox transformation ([Box & Cox 1964](#)).

## Data generation

[McLeod & Hipel \(1978\)](#) developed an exact simulation procedure for the univariate ARMA model and its subsets. They suggested using a theoretically correct variance-covariance structure to initialize the generation process and to avoid systematic bias in the generated sequences. Suppose that it is required to generate  $N$  terms of an ARMA ( $p, q$ ) model with innovations that are normally and independently distributed (NID) ( $0, \sigma_a^2$ ). The data generation procedure ([McLeod & Hipel 1978](#)) is as follows.

1. The theoretical auto-covariance function  $\gamma_j$  for  $j = 0, 1, \dots, (p-1)$  is obtained first. For this, the ARMA model is first written in the difference form and then multiplied by  $z_{t-k}$  and  $a_{t-k}$ , and the resulting equations are solved for  $\gamma_j$ . The notation  $k$  indicates a time lag.
2. The random shock coefficients  $\psi_j$  for  $j = 1, 2, \dots, (q-1)$  is then determined by equating the coefficients of like powers of  $B$  from both sides of  $z_t = (\theta(B)/\varphi(B))a_t = \psi(B)a_t$ .
3. The variance-covariance matrix  $\Delta$  of  $z_p, z_{p-1}, \dots, z_1, a_p, a_{p-1}, \dots, a_{p-q+1}$  is formed as:

$$\Delta = \begin{bmatrix} \left( \begin{matrix} \gamma_{i-j} & \psi_{j-i} \\ \psi_{i-j} & \delta_{i,j} \end{matrix} \right)_{p \times q} & \left( \begin{matrix} \psi_{j-i} \\ \delta_{i,j} \end{matrix} \right)_{p \times q} \\ \left( \begin{matrix} \psi_{i-j} \\ \delta_{i,j} \end{matrix} \right)_{q \times p} & \left( \begin{matrix} \delta_{i,j} \end{matrix} \right)_{q \times q} \end{bmatrix}_{(p+q) \times (p+q)}$$

where the  $(i, j)$ th element and dimension of each partitioned matrix are indicated. The values of  $\delta_{i,j}$  are 1 or 0 according to whether  $i = j$  or  $i \neq j$ , respectively. When  $i - j < 0$ , then  $\gamma_{i-j} = \gamma_{j-i}$  and  $\psi_{i-j} = 0$ .

4. The lower triangular matrix  $M$  is determined by Cholesky decomposition or any other matrix method such that

$$\Delta = MM^T$$

where the notation  $T$  indicates a matrix transpose.

5. Two random sequences  $e_1, e_2, \dots, e_{p+q}$  and  $a_{p+1}, a_{p+2}, \dots, a_N$  are generated, where both  $e_t$  and  $a_t$  sequences are NID  $(0, \sigma_a^2)$ .

6. A  $w_t$  sequence of  $p$  terms  $w_1, w_2, \dots, w_p$  is calculated from

$$w_{p+1-t} = \sum_{j=1}^t m_{t,j} e_j, \quad t = 1, 2, \dots, p$$

where  $m_{t,j}$  is the  $(t, j)$ th entry in the matrix  $M$ .

7. A residual sequence of  $a_t$  with  $q$  terms  $a_{p-q+1}, a_{p-q+2}, \dots, a_p$  is determined from

$$a_{p+1-t} = \sum_{j=1}^{p+1} m_{t+p,j} e_j, \quad t = 1, 2, \dots, q$$

8. Finally, the remaining  $(N - p)$  terms,  $w_{p+1}, w_{p+2}, \dots, w_N$  of  $w_t$  sequence are obtained using

$$w_t = \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + \dots + \varphi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

where  $t = p + 1, p + 2, \dots, N$

Steps (5) through (8) are repeated for generation of every new series of length  $N$ .

## APPLICATION OF THE MODEL

The Brahmaputra is a major trans-boundary river and contributes about two-thirds of the total dry season flows in Bangladesh. It plays an important role in the overall socio-economic development of the country. An application of the above deseasonalized ARMA model was made to the 10-day flow of this river at Bahadurabad. Before proceeding to the model application, a general description of the Brahmaputra River and its flow characteristics is provided.

### The Brahmaputra River

The Brahmaputra River is one of the largest rivers in the world. It originates in the Jima Yangzong glacier near the Mount Kailas in the northern Himalayas. It has a long course for about 1,700 km through the dry and flat region of southern Tibet. Throughout this upper course, the river is generally known as the Tsang-Po (FAP 24 1996a). At its easternmost point, the river bends around the Namcha

Barwa peak. As the river enters Arunachal state of India, it is called Siang. The Brahmaputra appears in the Assam valley as the Dihang River. It flows for about 268 km through Arunachal state and 640 km through Assam. The Dihang is joined by the Dibang and the Lohit from the east near Sadiya in northeast Assam. From this point of confluence, the river is called the Brahmaputra. As Brahmaputra, the river flows through the entire stretch of Assam and sweeps round the Garo Hills and enters Bangladesh. In Bangladesh, the Brahmaputra flows southwards for nearly 240 km before joining the Ganges at Goalanda (FAP 24 1996b). The total length of the Brahmaputra River is therefore about 2,848 km, of which about 8.4% lies within Bangladesh.

The catchment area of the river is about 550,000 km<sup>2</sup> stretching over Tibet, India, Bangladesh and Bhutan, of which about 8% is within Bangladesh (FAP 24 1996b). However, this 8% is equivalent to about 32% of the area of Bangladesh. A gauge station of the river is located at Bahadurabad, which is at 10 km downstream of the off-take of the Old Brahmaputra. The distance between Bahadurabad and Aricha is about 130 km. The width of the river varies spatially and temporally, and the overall width ranges from 6 to 14 km (FAP 24 1996b).

### Flow characteristics of the Brahmaputra River

Discharge data of the Brahmaputra at Bahadurabad station have been available for this study for a period of 49 years (April 1956–February 2005). The data are missing for 18 months (October 1963–March 1964 and April 1971–March 1972). Inconsistencies have been detected in the Bangladesh Water Development Board (BWDB) discharge data for a period of 56 months (August 1988–March 1993) in the FAP 24 (1996a) report. The data for this period have been replaced with the data derived from the three rating equations suggested in FAP 24 (1996a).

An average, the Brahmaputra flow reaches its peak during the second 10-day period of July and trough during the last 10-day period of February. From the second 10-day period of June to the first 10-day period of October, flows are much higher compared to the rest of the year. There is a strong seasonal pattern in the Brahmaputra flow, so the flow is intra-year non-stationary.



To check whether or not the flow is inter-year stationary, a total of 36 time series (one for each 10-day period of each month of the year) was plotted. A linear regression line was superimposed on each of these plots. The slopes of the least-squares lines were found to be negative for the periods of May II, May III, Jun III, Aug I and Aug II (I, II and III indicating the first, second and third 10-day periods, respectively, in the relevant month), and positive for the remaining periods. The flow therefore has generally increasing trends except for some 10-day periods in the pre-monsoon and monsoon periods. However, the slopes of the increasing trend lines were generally low. The per year increase was found to vary between 0.10% for the second 10-day period June to 1.16% for the third 10-day period of November of the respective 10-day average flows. The values of the coefficient of determination ( $R^2$ ) of the trend lines were also low (0–16.6%). The small increasing trends found in some 10-day periods were therefore ignored in subsequent analyses.

To see whether or not the annual hydrograph of the Brahmaputra River exhibits a trend in the annual peak or trough, the highest and lowest flows of each year were found out from the 365 or 366 daily values (Mondal *et al.* 2007). The analysis of the two extreme value series did not indicate the presence of any linear trend in either series. To check if there has been any temporal change in the annual peak and low flows, the dates of occurrences of the highest and lowest water levels were determined for each year. It is found that the median date of occurrence of peak flow is 30 July with a standard deviation of 36 days and the median date of occurrence of the lowest flow is 27 February with a standard deviation of 13 days. Dividing the peak and low-flow time series into two halves (each half with a 24-year length), it is found that there is no significant difference in the time of occurrence of either the peak discharge or the low discharge between the two halves of the available periods. Furthermore, no trend is found in the two time series of the dates of occurrences of the highest and lowest flows.

Ten-day means and standard deviations of the Brahmaputra flow were found to be roughly proportional (see Mondal *et al.* 2007). The existence of such proportionality indicates that a power or logarithmic transformation should be applied to the raw data before model construction. This conclusion is also justified from the fact that the skewness of all months

except June and July decreases due to the natural logarithmic transformation, as reported in Mondal *et al.* (2007). A common way of investigating the relationship between the average value or expected level of a variable and the variability or spread associated with it is to plot the values of spread and level for each period. If there is no relationship, the points would align around a horizontal line. Otherwise, we can use the observed relationship between the two variables to choose an appropriate transformation. To determine an appropriate power for transforming the data we can plot, for each period, the logarithm of the median against the logarithm of the inter-quartile range. Figure 1 shows such a plot for the Brahmaputra flow data. We can see that there is a fairly strong linear relationship between spread and level with a  $R^2$  value of about 95%. The slope of the least-squares line is about 1.20, so the power for the transformation is  $-0.20$ . After applying this power transformation, a spread versus level plot was again obtained. No further relationship was evident from such a plot.

Normality checks of the negative power-transformed data were made with normal probability plots as well as with tests of normality. Both Kolmogorov–Smirnov’s test with Lilliefors significance correction (Lilliefors 1967) and Shapiro–Wilk’s test (Shapiro & Wilk 1965) indicated that the power transformation improved the normality of the data significantly. The box-and-whisker plot and the stem-and-leaf plot (Tukey 1977) also indicated that the number of outliers/extremes is reduced by to the transformation.

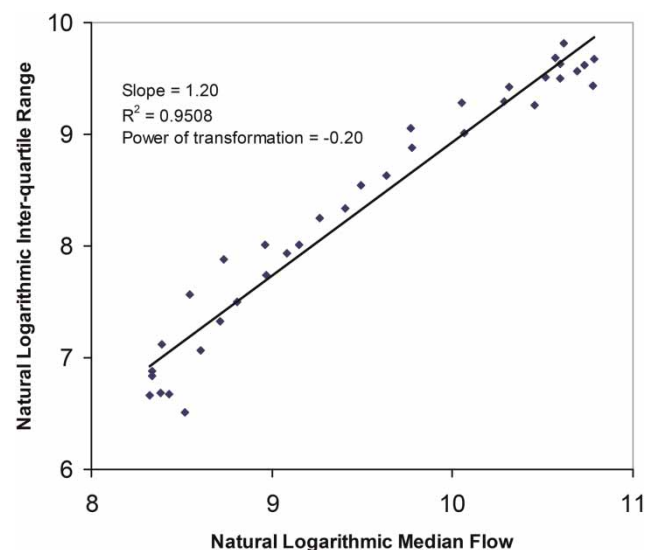


Figure 1 | Spread versus level plot of the Brahmaputra flow at Bahadurabad.

The negative power-transformed data were therefore used for model building in the following sections.

**Fitting deseasonalized ARMA model to the Brahmaputra flow**

To fit a deseasonalized ARMA model, the 10-day mean was subtracted from each 10-day observation. The result was then divided by the corresponding 10-day standard deviation to obtain the deseasonalized series. The 36 10-day means and standard deviations for the transformed flows were obtained by the parametric method. The first five and 13 harmonics were found to be significant for the 10-day means and standard deviations, respectively. To identify the significant harmonics, the graphical criterion of separating the harmonics into the periodic and sampling variation parts from the plot of the cumulative periodogram, as outlined earlier, was followed. Figure 2 shows the cumulative periodogram for the 10-day means.

After removal of the seasonal component, autocorrelations and partial autocorrelations at different lags of the deseasonalized series were estimated and are given in Figure 3. It is seen from the figure that the ACF has an exponentially decaying pattern and the PACF has significant values until lag 5, except for lag 2, also with a decaying pattern. These patterns indicate that the model can be a mixed model having both AR and MA parameters. After a few iterations, the model that appeared to be suitable was ARMA (1,3). The estimated parameters of the fitted model are given in Table 1. It is seen from the last column of the table that three parameters

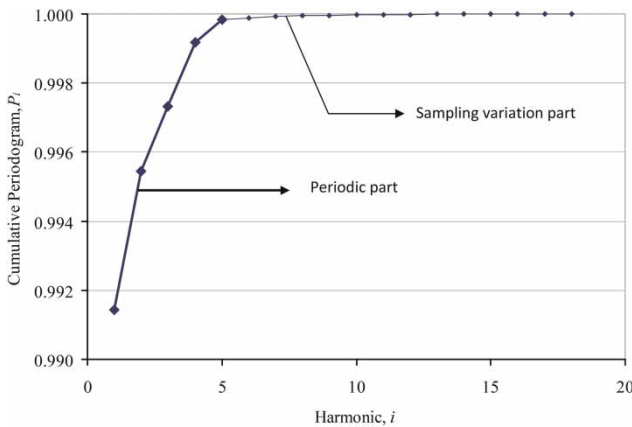


Figure 2 | Separation of the cumulative periodogram of 10-day means into two parts: periodic (significant) and sampling variation (insignificant).

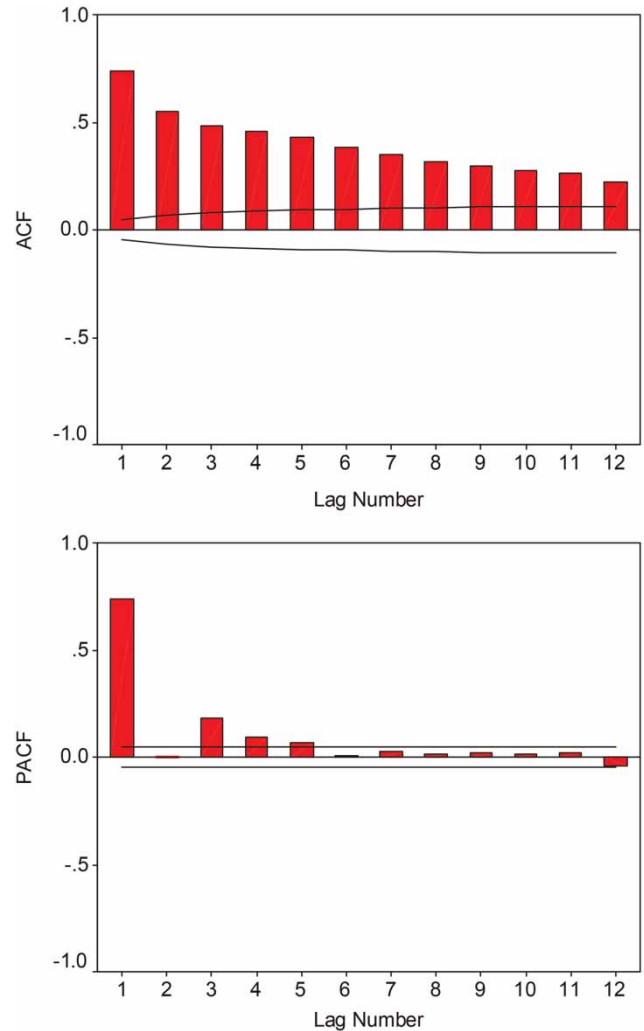


Figure 3 | ACF and PACF along with the 95% confidence limits of the deseasonalized, power-transformed 10-day flow of the Brahmaputra River at Bahadurabad.

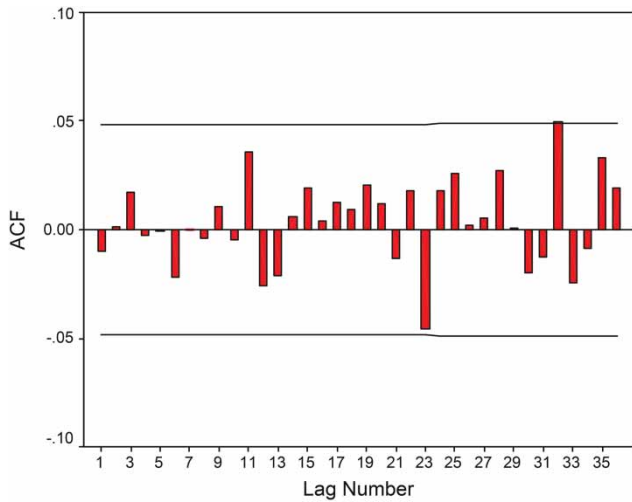
( $\hat{\varphi}_1$ ,  $\hat{\theta}_1$  and  $\hat{\theta}_2$ ) are significant at the 1% level and the remaining parameter ( $\hat{\theta}_3$ ) is significant at the 10% level.

The residual ACF of the above ARMA model is shown in Figure 4, which does not give any indication of non-whiteness of the model residuals. The cumulative periodogram of the residuals is shown in Figure 5. This figure does not show any periodic pattern in the model residuals. The fitted deseasonalized ARMA model can be expressed:

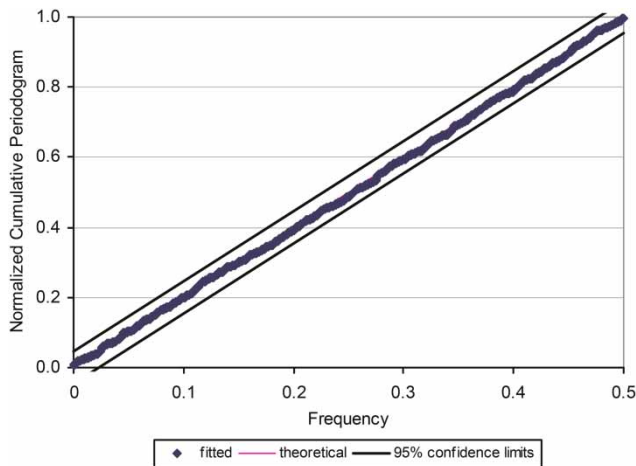
$$(1 - 0.92880B)z_t = (1 - 0.20725B - 0.28031B^2 - 0.05052B^3)\hat{a}_t \tag{10}$$

**Table 1** | Estimated values, standard errors, *t*-statistics, etc. of different parameters of the deseasonalized ARMA (1, 3) model. Residual variance, standard error, log-likelihood, AIC and BIC values were 0.4193, 0.6475, -1,680, 3,368 and 3,389, respectively, in the deseasonalized scale

Parameter	Estimated value	Standard error of estimate	<i>t</i> -ratio	Approximate probability
$\hat{\varphi}_1$	0.92880	0.01557	59.64	0.000
$\hat{\theta}_1$	0.20725	0.02908	7.13	0.000
$\hat{\theta}_2$	0.28031	0.02773	10.11	0.000
$\hat{\theta}_3$	0.05052	0.02651	1.91	0.057



**Figure 4** | The residual ACF along with the 95% confidence limits of the fitted deseasonalized ARMA model.



**Figure 5** | Cumulative periodogram with the 95% large-sample confidence limits for the residuals of the deseasonalized ARMA model.

where  $z_t$  is the parametrically deseasonalized power transformed 10-day flow of the Brahmaputra River. The total number of parameters in the deseasonalized ARMA model is 23 (five harmonics for 10-day means and 13 for standard deviations, one AR coefficient, three MA coefficients and one residual variance).

**Model validation**

To evaluate the performance of the model, validation forecasts were generated from the model. The procedures of forecast generation from the model are described by Mondal *et al.* (2007). The parameters of the model were estimated with the data up to February 1997 and the model was validated with the data from March 1997 to February 2005 using one-step-ahead validation forecasts. All the forecasted flows were found to be within the 95% large sample confidence limits (see Mondal *et al.* (2007) for details). Root mean square error (RMSE) and mean absolute error (MAE) of the forecasted flows (Table 2) were calculated to evaluate the performance of the model. The overall MAE

**Table 2** | RMSE and MAE ( $m^3 s^{-1}$ ) of one-step-ahead validation forecasts of the deseasonalized ARMA model fitted to the Brahmaputra flow

10-day period	RMSE	MAE	10-day period	RMSE	MAE
Jan I	729	568	Jul I	6,243	4,937
Jan II	388	292	Jul II	6,830	4,934
Jan III	404	269	Jul III	8,309	6,576
Feb I	406	300	Aug I	8,837	7,562
Feb II	363	279	Aug II	9,612	7,826
Feb III	692	412	Aug III	9,559	7,602
Mar I	367	267	Sep I	8,840	5,929
Mar II	299	238	Sep II	7,940	5,248
Mar III	962	659	Sep III	5,876	4,537
Apr I	1,898	1,175	Oct I	4,923	4,276
Apr II	1,436	1,192	Oct II	4,339	4,052
Apr III	2,306	1,815	Oct III	4,423	3,423
May I	4,443	3,269	Nov I	6,232	3,250
May II	3,010	2,320	Nov II	4,718	2,407
May III	2,516	1,874	Nov III	5,396	2,689
Jun I	6,699	5,649	Dec I	1,179	792
Jun II	8,038	6,014	Dec II	1,379	798
Jun III	7,203	5,766	Dec III	831	595



and RMSE were found to be about 3,050 and 4,100  $\text{m}^3 \text{s}^{-1}$ , respectively. Given the average flow of the Brahmaputra River of about 20,250  $\text{m}^3 \text{s}^{-1}$ , the performance of the model can be considered to be satisfactory. One-step-ahead validation forecasts from the deseasonalized ARMA (1,3) model along with the observed flows are given in Figure 6. It is seen from the figure that the fitted model captures the observed 10-day pattern of the Brahmaputra flow reasonably well. The model performs very well during the dry season, for which synthetic flow would basically be required.

To check how a disturbance in the current time period affects the current and future flows, the deseasonalized ARMA model was written in random shock form and the shock coefficients were estimated. A plot of the coefficients against lead time is shown in Figure 7. It is evident from the figure that any disturbance during a 10-day period has the most influence on flow for that time. The influence of the disturbance on future flows reduces with the increase in lead time. This is also understandable from a physical point of view. For example, if there is some rain in a time period, this rain will have the most influence on the current time period river flow. The effect of rain on river flow will decrease gradually with time. This physical explanation of the behaviour of the deseasonalized ARMA model gives it a strong basis for use in river hydrology.

## GENERATION OF SYNTHETIC FLOWS

The fitted deseasonalized ARMA model was employed to generate 10-day flows of the Brahmaputra River. The general algorithm described earlier for exact simulation with an ARMA model was used for the data generation. Portable independent normal variables  $e$  required in such simulations were generated using the SPSS (1995) package, with different random number seeds for different sequences. Two hundred synthetic traces, each trace with a length of 50 years, were generated with the developed model. For each generated sequence, the variance and the lag-1 to lag-7 autocorrelations were computed. The mean values and confidence limits of each of the six parameters were then obtained from each of the 200 values, and are given in column 4 of Table 3.

Stedinger & Taylor (1982) suggested two diagnostics – model verification and validation – in addition to the conventional diagnostic checks to evaluate the adequacy of a stochastic model. According to these authors, model verification is the demonstration that the developed model produces flows with the characteristics predicted by its theoretical prototype. For this test, the theoretical variance  $\gamma_0$  was obtained by multiplying  $z_t$  in Equation (7) with  $z_t$  and then taking expected values. The theoretical covariance  $\gamma_k$  at lag  $k$  was obtained by multiplying  $z_t$  with  $z_{t-k}$  and then taking expected values. The solution process involved a total

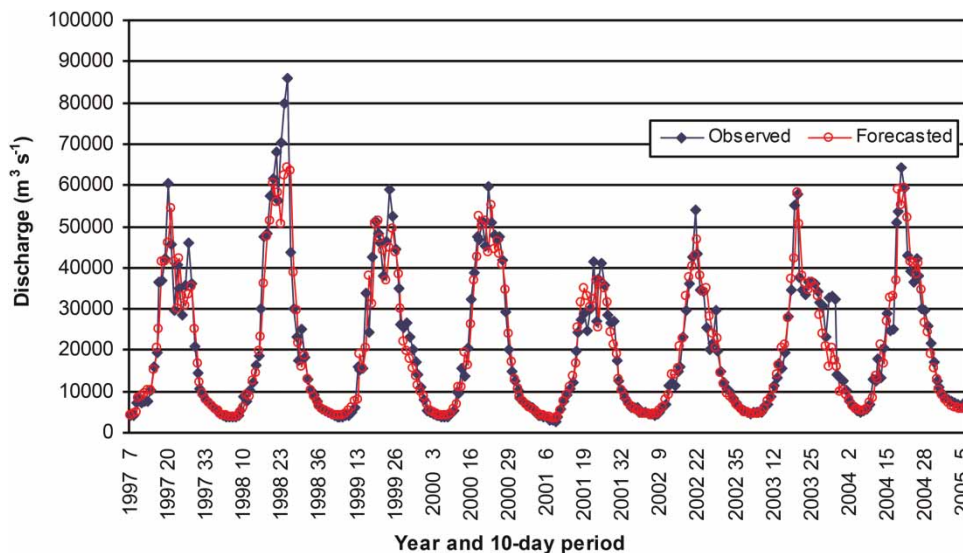


Figure 6 | One-step-ahead forecasted flows along with the observed flows from the first 10-day period of March 1997 to the last 10-day period of February 2005.

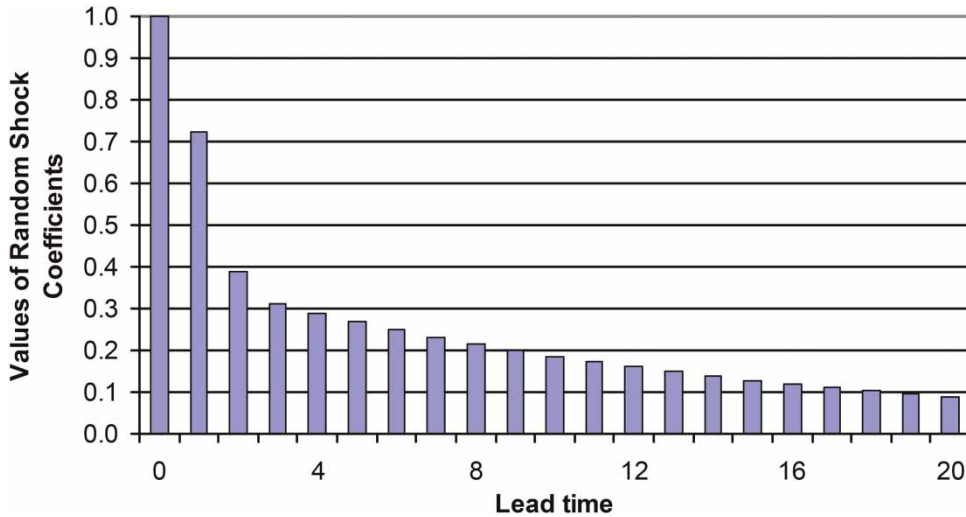


Figure 7 | Influence of the current period disturbance on current and future flows.

Table 3 | Observed, theoretical and generated variances, and lag-1 to lag-7 autocorrelations, for the deseasonalized 10-day period flows. Values in parentheses in the last column are 95% confidence limits

Parameter	Observed	Theoretical	Generated
Variance	0.982	0.998	0.977 (0.803, 1.151)
Lag-1 autocorrelation	0.741	0.748	0.741 (0.694, 0.789)
Lag-2 autocorrelation	0.548	0.562	0.551 (0.472, 0.630)
Lag-3 autocorrelation	0.486	0.501	0.490 (0.402, 0.577)
Lag-4 autocorrelation	0.459	0.465	0.453 (0.364, 0.542)
Lag-5 autocorrelation	0.431	0.432	0.419 (0.327, 0.512)
Lag-6 autocorrelation	0.385	0.401	0.388 (0.293, 0.482)
Lag-7 autocorrelation	0.347	0.373	0.360 (0.262, 0.457)

of ten equations with ten unknowns. Both the theoretical and observed variances and the lag-1 to lag-7 autocorrelations are given in Table 3. It is evident from the table that the theoretical, as well as the observed, values of all the parameters are well inside the 95% confidence limits of generated values. It can therefore be concluded that the generated sequences exhibit short-term characteristics, which are statistically indistinguishable not only from the theoretical prototype but also from the historical observations.

Stedinger & Taylor (1982) described model validation as the demonstration that the generated sequences preserve the long-term statistics. For this, the Hurst coefficient and Rescaled Adjusted Range (RAR) (Salas et al. 1979) were estimated for each of the 200 sequences, as well as for the

historical sequence of the Brahmaputra flows. These are reported in Table 4. It is seen from the table that the historical sequence has a Hurst coefficient of 0.669, whereas the generated sequences have an expected value of 0.667 with the 95% large sample confidence limits of 0.587–0.747. The observed sequence has a RAR of 90.85, and the generated sequences have a mean value of 95.98 with the 95% large sample confidence limits of 42.79–149.17. The probabilities of exceedence of the historical Hurst coefficient and RAR were found to be 48.0 and 50.5%, respectively. A probability of exceedence indicates the level of significance at which the null hypothesis (i.e. there is no difference between the observed and generated Hurst coefficient and RAR) would be rejected.

Since the exceedence probabilities are greater than the usual significance level of 5 or 10% employed in such testing using a  $\chi^2$  test (see Hipel & McLeod 1994), we cannot reject the null hypothesis. The developed deseasonalized ARMA model can therefore be considered able to preserve the

Table 4 | Hurst coefficient and rescaled adjusted range (RAR) for both observed and generated flows of the Brahmaputra River. Values in parentheses are 95% confidence limits

Parameter	Observed	Generated	Exceedence Probability
Hurst coefficient	0.669	0.667 (0.587, 0.747)	0.480
RAR	90.85	95.98 (42.79, 149.17)	0.505

important long-term statistics of the Brahmaputra flow. In fitting ARMA models to a number of different geophysical time series, Hipel & McLeod (1978, 1994), McLeod & Hipel (1978), Salas *et al.* (1988), Mondal (2005) and others have also demonstrated that these models can preserve the long-term statistics.

From the discussions above in this section and also from the results in Tables 3 and 4, it can be concluded that the

generated flows with the deseasonalized ARMA model have both short- and long-term statistical characteristics similar to the observed flows. Such flows have already been used in a risk-based evaluation of Brahmaputra water development in meeting future water demand (Mondal *et al.* 2010). Figure 8 shows one sequence (out of 200 sequences) of the discharge data generated with the ARMA model. Figure 9 shows the plot of a small portion

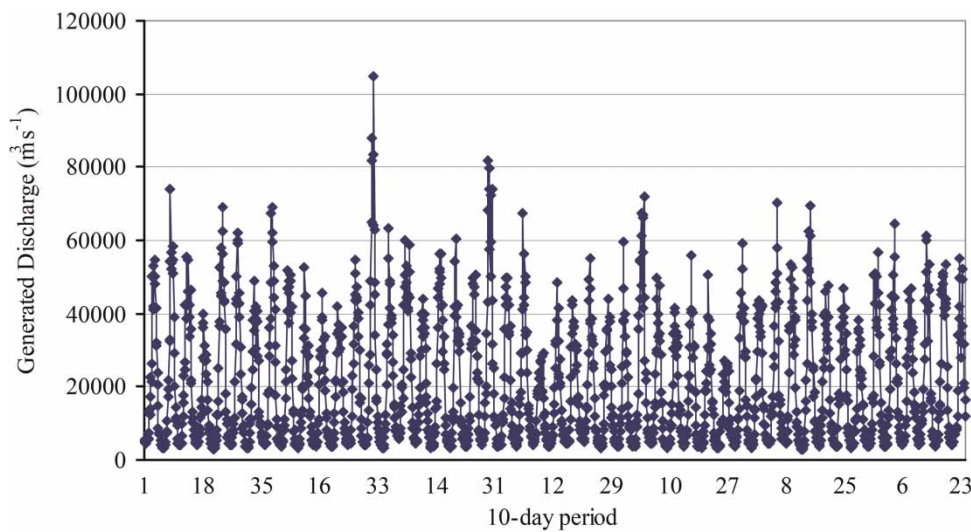


Figure 8 | One long time series (of the 200 available) of generated discharge data.

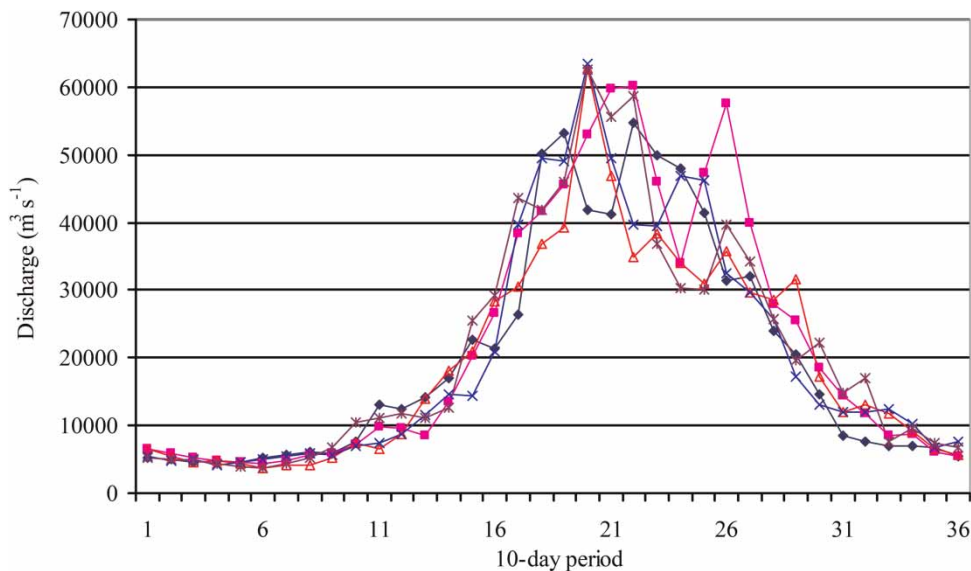


Figure 9 | Five time series (of the 200 available) of generated discharge data for 1 year.

(1 year) of the generated data for five sequences, so that we can get a view of the sequences together in a plot and have a visual impression of the underlying sampling variability.

## CONCLUSIONS

A deseasonalized ARMA model is fitted to the 10-day flow of the Brahmaputra River at Bahadurabad. The basic use of this model is to generate synthetic flows. Comparing the one-step-ahead validation forecasts with the observed flows using graphical plot as well as RMSE and MAE criteria, the deseasonalized ARMA model was found to be suitable for generation of the Brahmaputra flow. Further validation and verification of the model using synthetic flows showed that the ARMA model could preserve the important short- and long-term statistics of the Brahmaputra flow. The fitted model was used to generate 200 synthetic sequences, each of 50-year length, of the 10-day Brahmaputra flow. These sequences were used in risk-based evaluation of performance of the proposed Brahmaputra barrage at Bahadurabad in meeting future water demand of the Brahmaputra barrage command area.

## ACKNOWLEDGEMENTS

This research was carried out in collaboration with the Water Resources Planning Organization, the Center for Environmental and Geographic Information Services and the Bangladesh Agricultural Research Council, Dhaka. The comments provided by the two anonymous reviewers on the manuscript are gratefully acknowledged.

## REFERENCES

- Akaike, H. 1974 [A new look at the statistical model identification](#). *IEEE Transactions on Automatic Control* **AC-19** (6), 716–723.
- Artis, M., Clavel, J. G., Hoffmann, M. & Nachane, D. 2007 Harmonic Regression Models: A Comparative Review with Applications, Working Paper Series ISSN 1424-0459, Institute for Empirical Research in Economics, University of Zurich.
- Box, G. E. P. & Cox, D. R. 1964 [An analysis of transformations](#). *Journal of the Royal Statistical Society, Series B, London* **26**, 211–252.
- Box, G. E. P. & Pierce, D. A. 1970 [Distribution of the residual autocorrelations in autoregressive integrated moving average models](#). *Journal of the American Statistical Association* **65**, 1509–1526.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. 1994 *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, New Jersey.
- Chatfield, C. 1996 *The Analysis of Time Series: An Introduction*. CRC Press, USA.
- Cline, T. B. 1981 [Selecting seasonal streamflow models](#). *Water Resources Research* **17** (4), 975–984.
- Delleur, J. W. & Kavvas, M. L. 1978 [Stochastic models for monthly rainfall forecasting and synthetic generation](#). *Journal of Applied Meteorology* **17**, 1528–1536.
- Delleur, J. W., Tao, P. C. & Kavvas, M. L. 1976 [An evaluation of the practicality and complexity of some rainfall and runoff time series models](#). *Water Resources Research* **12** (5), 953–970.
- FAP 24 1996a River survey project, Final Report, Annex 3: Hydrology. Flood Action Plan 24, Water Resources Planning Organization, Dhaka.
- FAP 24 1996b River survey project, Special Report No. 24: Morphological processes in the Jamuna River. Flood Action Plan 24, Water Resources Planning Organization, Dhaka.
- Govindasamy, R. 1991 [Univariate Box–Jenkins forecasts of water discharge in the Missouri River](#). *International Journal of Water Resources Development* **7** (3), 168–177.
- Hipel, K. W. & McLeod, A. I. 1978 [Preservation of the rescaled adjusted range, 2: Simulation studies using Box–Jenkins models](#). *Water Resources Research* **14** (3), 509–516.
- Hipel, K. W. & McLeod, A. I. 1994 *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam.
- Irvine, K. N. & Eberhardt, A. J. 1992 [Multiplicative, seasonal ARIMA models for Lake Erie and Lake Ontario water levels](#). *Water Resources Bulletin* **28** (2), 385–396.
- Kavvas, M. L. & Delleur, J. W. 1975 [Removal of periodicities by differencing and monthly mean subtraction](#). *Journal of Hydrology* **26**, 335–353.
- Lilliefors, H. W. 1967 [On the Kolmogorov–Smirnov test for normality with mean and variance unknown](#). *Journal of the American Statistical Association* **62**, 399–402.
- Ljung, G. M. & Box, G. E. P. 1978 [On a measure of lack of fit in time series models](#). *Biometrika* **65**, 297–303.
- Lungu, E. M. & Sefe, F. T. K. 1991 [Stochastic analysis of monthly streamflows](#). *Journal of Hydrology* **126**, 171–182.
- McKerchar, A. I. & Delleur, J. W. 1974 [Application of seasonal parametric linear stochastic models to monthly flow data](#). *Water Resources Research* **10** (2), 246–255.
- McLeod, A. I. & Hipel, K. W. 1978 [Simulation procedures for Box–Jenkins models](#). *Water Resources Research* **14** (5), 969–975.
- Mondal, M. S. 2005 Risk-based Evaluation of Performance in Water Management for the Ganges Delta within Bangladesh. PhD Thesis, Central Queensland University, Australia.



- Mondal, M. S., Chowdhury, J. U. & Ferdous, M. R. 2010 Risk-based evaluation for meeting future water demand of the Brahmaputra floodplain within Bangladesh. *Water Resources Management* **24** (5), 835–852.
- Mondal, M. S., Chowdhury, J. U. & Khatun, F. 2007 Risk-based evaluation of Brahmaputra water development in meeting future water demand. Technical Report 01: Generation of synthetic flows of the Brahmaputra River using time series models. Institute of Water and Flood Management, BUET, Dhaka.
- Mondal, M. S. & Wasimi, S. A. 2005a Periodic transfer function-noise model for forecasting. *Journal of Hydrologic Engineering* **10** (5), 353–362.
- Mondal, M. S. & Wasimi, S. A. 2005b Monthly flow forecasting of the Ganges river with PAR model. *Bangladesh Journal of Water Resource Research* **20**, 21–41.
- Mondal, M. S. & Wasimi, S. A. 2006 Generating and forecasting monthly flows of the Ganges river with PAR model. *Journal of Hydrology* **323** (2), 41–56.
- Mondal, M. S. & Wasimi, S. A. 2007a Choice of model type in stochastic river hydrology. Proceedings of the International Conference on Water and Flood Management, Institute of Water and Flood Management, BUET, 12–14 March, Dhaka, pp. 633–640.
- Mondal, M. S. & Wasimi, S. A. 2007b Evaluation of risk-related performance in water management for the Ganges Delta of Bangladesh. *Journal of Water Resources Planning and Management* **133** (2), 179–187.
- Newbold, P. & Granger, C. W. J. 1974 Experience with forecasting univariate time series and combination of forecasts. *Journal of Royal Statistical Society, Series A, London* **137**, 131–165.
- Pankratz, A. 1985 *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*. John Wiley & Sons, New York.
- Panu, U. S., Unny, T. E. & Ragade, R. K. 1978 A feature prediction model in synthetic hydrology based on concepts of pattern recognition. *Water Resources Research* **14** (2), 335–344.
- Papamichail, D. M. & Georgiou, P. E. 2001 Seasonal ARIMA inflow models for reservoir sizing. *Journal of the American Water Resources Association* **37** (4), 877–885.
- Salas, J. D., Boes, D. C., Yevjevich, V. & Pegram, G. G. S. 1979 Hurst phenomenon as a pre-asymptotic behaviour. *Journal of Hydrology* **44**, 1–15.
- Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. 1988 *Applied Modelling of Hydrologic Time Series*. Water Resources Publications, Fort Collins, Colorado, USA.
- Salas, J. D., Obeysekera, J. T. B. & Smith, R. A. 1981 Identification of streamflow stochastic models. *Journal of the Hydraulics Division* **107** (HY7), 853–866.
- Schwartz, G. 1978 Estimating the dimensions of a model. *Annals of Statistics* **6**, 461–464.
- Shapiro, S. S. & Wilk, M. B. 1965 An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611.
- Sidhu, A. 1995 Seasonal Flow Forecasting of Newfoundland Rivers. Master of Engineering Thesis, Memorial University of Newfoundland, Canada.
- SPSS 1995 *Base System User's Guide. Statistical Package for the Social Sciences*. SPSS Inc., Chicago, USA.
- Stedinger, J. R. & Taylor, M. R. 1982 Synthetic streamflow generation, 1: Model verification and validation. *Water Resources Research* **18** (4), 909–918.
- Tao, P. C. & Delleur, J. W. 1976 Seasonal and nonseasonal ARMA models in hydrology. *Journal of the Hydraulics Division* **102** (HY10), 1541–1559.
- Tukey, J. W. 1977 *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.
- Vandaele, W. 1997 *Applied Time Series and Box-Jenkins Models*. Academic Press, New York.
- Vecchia, A. V. 1985 Periodic autoregressive-moving average (PARMA) modelling with applications to water resources. *Water Resources Bulletin* **21** (5), 721–730.
- Young, P. C., Pedregal, D. J. & Tych, W. 1999 Dynamic harmonic regression. *Journal of Forecasting* **18**, 369–394.

First received 20 October 2011; accepted in revised form 17 July 2012. Available online 5 December 2012