Asian Journal of Water, Environment and Pollution, Vol. 3, No. 1, pp. 111-114.

Autoregressive Model for Flood Forecasting

Sutapa Chaudhuri and Surajit Chattopadhyay

Department of Atmospheric Sciences
University of Calcutta, Kolkata – 700 019
✉ chaudhuri_sutapa@yahoo.com

Received January 12, 2005; revised and accepted November 15, 2005

Abstract: The purpose of the present paper is to develop an autoregressive model for forecasting the frequency of occurrence of flood that India may countenance in a given year. The proposed model being autoregressive, prediction requires no other data than the temporal frequency of occurrence of flood. A comparative study is made among several orders of autoregressive models. The comparison is furnished on the basis of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) statistics for different orders of autoregressive equations. The result of the study reveals the suitability of third order autoregression as a predictive model. The result is qualitatively supported by learning the sample autocorrelations and autocorrelations theoretically implied by third order autoregressive model.

Key words: Flood, autoregressive model, AIC, BIC, autocorrelation.

Introduction

Many of the Asian and Pacific developing countries are situated in the world's hazard belts and are subjected to natural hazards like floods, droughts, cyclones, earthquakes, windstorms, tidal waves, land slides etc. The major natural disasters that occur periodically in this region are largely due to climatic and seismic factors. The region has suffered 50 per cent of the world's major natural disasters. Vulnerability to disasters has increased due to the increased aggregation of people in urban centres, environmental degradation, and a lack of planning and preparedness. Within the state of science, the role of the meteorologist is to keep natural hazards away from natural disasters. From an operational perspective, this process has three steps: (a) Observations: Initial conditions are determined from these observations. (b) A series of products is then developed. (c) The products are finally delivered to the users.

The present paper attempts to develop autoregressive model for prediction of frequency of flood in India. The basic reason of occurrence of flood is excessive rainfall. Rainfall exceeding the rate of infiltration through soil produces surface runoff. The runoff reaches the main stream of the catchment area and some other water channels intercept parts of it. When the runoff is large enough, the channels are filled completely and the water logging starts. India is a large country and prone to a number of natural hazards. Among all the natural disasters that country countenances, river floods are the most frequent and often devastating.

In spite of colossal advancement in science and technology in recent times, the occurrence of river floods is not absolutely avoidable. It is only possible to reduce the harmful socio-economic impact on human life. This is only possible if appropriate indigenous predictive model could be developed through proper exploration of the previous record of occurrences. The exploration has various dimensions. A proper map needs to be constructed so that the most affected areas can be properly identified. A statistical tabulation of the frequency of occurrences, affected areas and people affected need to be prepared. In recent times, geographical information system (GIS) is being applied by various scientists in the study of flood.

Present paper is an attempt to develop a predictive autoregression model to predict the future yearly frequencies of occurrence of flood in India. Reason behind the autoregressive model is that there exists inherent chaotic pattern in the time series of frequency of occurrence of floods. The most important constituent of flood is rainfall. Several scientists (e.g. Sivakumar, 2001; Sivakumar et al., 1998; Sharifi et al., 1999; Handerson and Wells, 1988) could establish the existence of chaos in the occurrence of rainfall. The second important constituent is the runoff process. Sthelik (1999) has shown the chaotic behaviour of the process of surface runoff. Sivakumar (2001) established that analysis of autocorrelation function is highly important to enlighten chaotic natural phenomena. The reason behind choosing the autoregressive model is thus two-fold. Firstly, the constituents of flood are chaotic and secondly, the chaotic occurrences are implied by auto-correlated occurrences.

Prediction of Time Series through Autoregressive Model

Autoregressive models (Box and Jenkins, 1976; Wilks, 1995) are highly suitable for prediction of time series. Such models can represent the correlation structure of a time series very successfully. The most widely used time-series model for prediction through autoregressive analysis is Box-Jenkins model (Box and Jenkins, 1976). The simplest Box-Jenkins model is the first order autoregressive [AR (1)] model, which is mathematically presented as

$$x_{t+1} - \mu = \phi(x_t - \mu) + \in_{t+1} \tag{1}$$

where μ is the mean of the time series; ϕ is the autoregressive parameter; and $\in_{t+1}$ is a random quantity corresponding to the residual in the ordinary regression.

The time series of x is assumed stationary, so that the mean is the same for each interval of time. The autoregressive model, as presented in equation (1), can represent the serial correlation of a time series. Parameter estimation in the first order autoregressive model is straightforward. The mean of the time series, μ, is simply the sample average of the dataset, under the assumption that the time series is stationary. The estimated autoregressive parameter is equal to the sample lag 1 autocorrelation coefficient. Mathematically, this estimation can be presented as

$$\hat{\phi} = r_1 \tag{2}$$

Here, r_1 is the sample autocorrelation coefficient for lag 1. For the resulting probability model to be stationary, it is required that $-1 < \phi < 1$. The full autocorrelation function for a time series governed by a first-order autoregressive process is given by (Wilks, 1995)

$$r_k = \phi^k \tag{3}$$

The first-order autoregressive model, as described in equation (1), predicts x_{t+1} using x_t as predictor. The first-order autoregressive process can readily be extended up to any order k to obtain autoregressive predictive models of higher order, that is, to predict x_{t+1} using predictors further back in time. The general autoregressive model of order K, or AR (K) model is

$$x_{t+1} - \mu = \sum_{k=1}^{K} \phi_k (x_{t-k+1} - \mu) + \in_{t+1} \tag{4}$$

Obviously, the AR (K) model requires K autoregressive parameters ϕ_K to be estimated. This estimation is most easily done by applying Yule-Walker equations (Katz, 1982)

$$\begin{aligned} r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 + \hat{\phi}_3 r_3 + \ldots\ldots\ldots\ldots + \hat{\phi}_K r_{K-1} \\ r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2 + \hat{\phi}_3 r_1 + \ldots\ldots\ldots\ldots + \hat{\phi}_K r_{K-2} \\ &\vdots \\ r_K &= \hat{\phi}_1 r_{K-1} + \hat{\phi}_2 r_{K-2} + \hat{\phi}_3 r_{K-3} + \ldots + \hat{\phi}_K \end{aligned} \tag{5}$$

Here, $\hat{\phi}_K = 0$ for $k > K$.

The theoretical autocorrelation function corresponding to a particular set of the ϕ_K values can be determined by solving (5) and then applying

$$\rho_m = \sum_{k=1}^{K} \phi_K \rho_{m-k} \tag{6}$$

Equation (6) holds for lags m greater than or equal to k, with $\rho_0 \equiv 1$.

Autoregressive model AR (2) and AR (3) are two special cases of equation (4). For AR (2), Yule-Walker equations will be

$$\begin{aligned} r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 \\ r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2 \end{aligned} \tag{7a}$$

For AR (3), Yule-Walker equations will be

$$\begin{aligned} r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 + \hat{\phi}_3 r_2 \\ r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2 + \hat{\phi}_3 r_1 \\ r_3 &= \hat{\phi}_1 r_2 + \hat{\phi}_2 r_1 + \hat{\phi}_3 \end{aligned} \tag{7b}$$

The sample residual variance $s_{\in}^2(m)$, also known as white-noise variance, is computed as (Katz, 1982)

$$s_{\in}^2(m) = \left[1 - \hat{\phi}_m^2(m)\right] s_{\in}^2(m-1) \tag{8}$$

Here, $s_{\in}^2(m)$ is the estimated white-noise variance of the mth auto regression and $s_{\in}^2(m-1)$ is estimated white-noise variance for the previously fitted (one order smaller) model. For autoregressive model of lag 0, the white-noise variance $s_{\in}^2(0)$ is simply the variance for the time series itself.

Yule-Walker equations can be potentially used to autoregressive models of any order. But, in order to have a predictive model for a time series, it is essential to decide which model is the best fit for the time series. The Bayesian Information Criteria popularly known as BIC statistic (Schwarz, 1978) and Akaike Information Criteria popularly known as AIC statistic (Akaike, 1974) are used to decide among orders of autoregressive models. For any order m, the BIC and AIC statistics are mathematically presented as

$$\text{BIC}(m) = n \ln\left[\frac{n}{n-m-1} s_{\in}^2(m)\right] + (m+1) \ln n \tag{9}$$

$$\text{AIC}(m) = n \ln\left[\frac{n}{n-m-1} s_{\in}^2(m)\right] + 2(m+1) \tag{10}$$

While experimenting with autoregressive models of different orders, the order having minimum AIC and minimum BIC is identified to give the best autoregressive predictive model for the time series.

Implementation of Model and the Results

The dataset explored in the present study consists of the yearly frequencies of occurrence of flood in India. Different flooded areas are considered and the total yearly occurrences of this weather event between 1975 and 2001 are considered as the input for the study. Since the years are considered without any gap (i.e. all the 27 years are considered), the frequencies can generate temporally equispaced data point to create a time series. The sample standard deviation is computed to be 2.13.

Thus, in the present study the lag 0 white-noise variance $s_{\in}^2(0)$ is 4.54. In the present study, lag 1, lag 2 and lag 3 autoregression equations are computed. The sample lag 1 autocorrelation is found to be 0.37. Thus, the estimated value of the autoregressive parameter ϕ_1 for lag 1 autoregression is 0.37. The other sample autocorrelations are also computed by increasing the powers of ϕ_1 with increase in lags, that is, by applying equation (3). The autoregressive parameters ϕ_1 and ϕ_2 for second-order autoregression are found to be 0.3 and 0.20 as computed from equation (7a). Applying equations (7b) the autoregressive parameters ϕ_1, ϕ_2 and ϕ_3 are found to be –0.7, 0.76 and 0.53 for the lag 3 autoregression equation. The sample white noise variances are computed from equation (8). The white-noise variances for autoregressions of different orders are shown in Figure 1. The figure shows that the white-noise variances have a decreasing trend with increase in the order of the autoregressive equation. The smaller white-noise variances for the higher orders imply that the higher-order models are exhibiting less residual uncertainty.

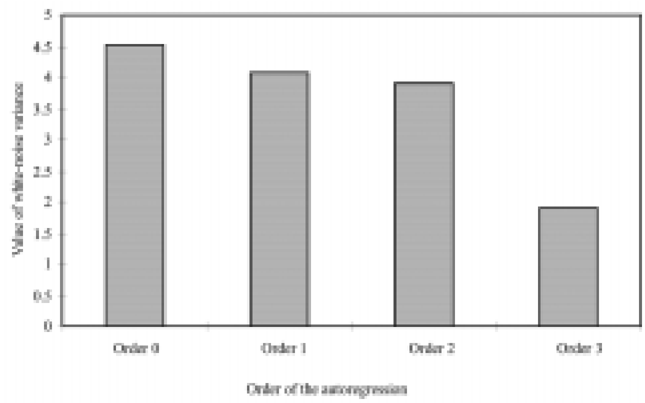

Figure 1: Schematic showing the white-noise variances corresponding to autoregressions of different orders.

The BIC and AIC statistics are computed for each autoregressive equations (9) and (10). The computed values are displayed in Table 1 and in Figure 2. Since the autoregressive model of third order produces the minimum for both the statistics, the third-order autoregressive model is identified as the best predictive model. As a qualitative support to the best fit of the third-order autoregressive model, the theoretical autocorrelation function implied by the third-order autoregressive model is computed from equation (6) and are compared with the sample autocorrelation function and this is made pictorially apparent in Figure 3.

Table 1: Order Selection for Autoregressive Models to Represent the Frequency of Occurrence of Flood in India between 1975 and 2001

Lag m	$s_{\in}^2$ *(m)*	*BIC (m)*	*AIC (m)*
0	4.54	45.16	44.87
1	4.07	46.56	43.98
2	3.91	49.88	45.99
3	1.91	34.98	29.80

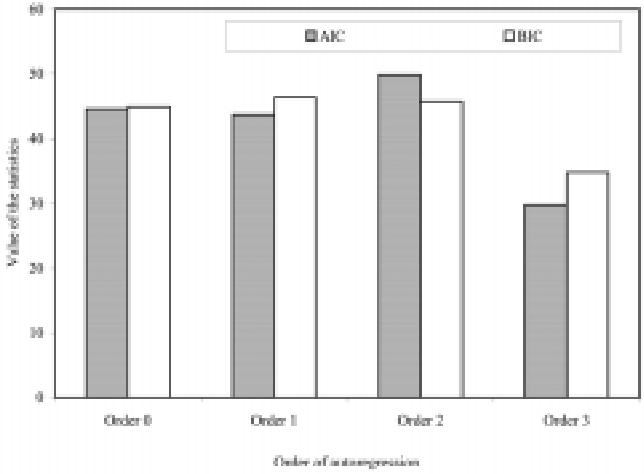

Figure 2: Schematic showing the AIC and BIC statistics corresponding to autoregressive equations of different orders.

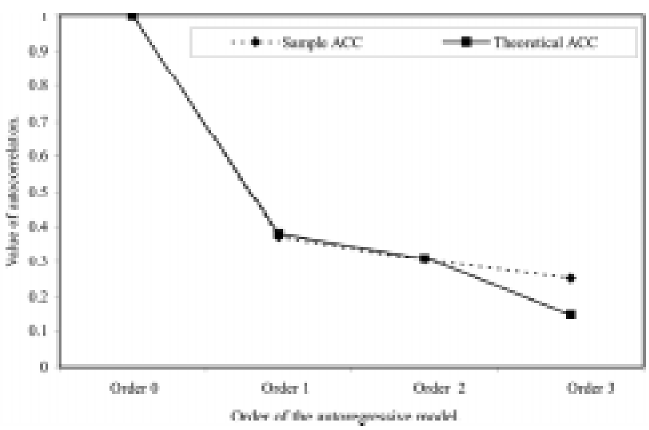

Figure 3: Schematic showing a close relationship between sample autocorrelation and theoretical autocorrelation implied by third order auto regression.

Conclusion

The above study leads to conclude that third order autoregressive model is a good predictive model for forecasting the frequency of occurrence flood in India.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control*, **AC-19:** 716-723.

Box, G.E.P. and G.M. Jenkins (1976). Time Series Analysis: Forecasting and Control. Holden Day: San Francisco.

Handerson, H.W. and R. Wells (1988). Obtaining attractor dimensions from meteorological time series. *Advances in Geophysics*, **30:** 205-237.

Katz, R.W. (1982). Statistical evaluation of climate experiments with general circulation models: A parametric time series modeling approach. *J. Atmos. Sci*., **39:** 1446-1455.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat*., **6:** 461-464.

Sivakumar, B. (2001). Is a chaotic multifractal approach to rainfall possible? *Hydrological Processes*, **15:** 943-955.

Sivakumar, B., Ling, S.Y. and C.Y. Liaw (1998). Evidence of chaotic behaviour in Singapore rainfall. *Journal of American Water Resources Association*, **34:** 301-310.

Stehlik, J. (1999). Deterministic chaos in runoff series. *Journal of Hydrology and Hydrodynamics*, **47:** 271-287.

Wilks, D.S. (1995). Statistical methods in Atmospheric Sciences. Academic Press, New York.