IWRS Journal of Indian Water Resources Society
Vol. 2 No. 3, July 1996

# SPECTRAL ANALYSIS OF CAUVERY RIVER FLOWS

**D . Nagesh Kumar**
Dept. of Civil Engineering
Indian Institute of Technology
Kharagpur (WB)

**P. P. Majumdar**
Dept. of Civil Engineering
Indian Institute of Science
Bangalore

**ABSTRACT**

*A detailed study of the monthly Cauvery River flows at the Krishna Raja Sagara (KRS) reservoir is carried out by using the techniques of spectral analysis. The correlogram and power spectrum are plotted and used to identify the periodicities inherent in the monthly inflows. The statistical significance of these periodicities is tested. Apart from the periodicities at 12 months and 6 months, a significant periodicity of 4 months was also observed in the monthly inflows. The analysis prepares ground for developing an appropriate stochastic model for the time series of the monthly flows.*

## INTRODUCTION

Time series analysis plays an important role in hydrology. The data available for many hydrologic variables, such as streamflows, rainfall and evaporation cover a smaller time span compared to the long duration for which the designs are required. As a result, data generation becomes imperative in most of the cases. In the past, hydrologic systems have been analysed with the main assumption of stationarity of the data. The use of time series analysis for the hydrologic data provides a good insight into the structure of the time series. Use of generated data overcomes design difficulties to a great extent. Spectral analysis, which forms an important part of time series analysis, provides valuable information about the structure of the time series which is essential in construction of the series. In particular, spectral analysis helps identify the significant periodicities present in the time series and thus provides a valuable clue to the particular class of stochastic models to be considered for modelling the series.

An objective of the present study is to identify the significant periodicities in the monthly flows in the Cauvery River. Data on monthly flows in the years 1934 to 1974 are used for the purpose. The results of this study are used subsequently to develop appropriate stochastic models for simulation and forecasting of the flows (Mujumdar and Nagesh Kumar, 1990). The details of spectral analysis and testing the significance of periodicities are discussed in standard text books such as Jenkins and Watts (1968), Kashyap and Rao (1976), Chatifield (1980) and Pandit and Wu (1983) among others, and hence the theoretical discussion is kept brief.

### Preliminary Investigation

As a first step the monthly means and standard deviations of the inflows are computed as follows:

$$\overline{X}_i = \sum_{j=1}^{N_y} X_{ij} / N_y \qquad (1)$$

Where i denotes the month, j denotes the year,

$x_i$ is the sample mean of the month i,
$x_{ij}$ is the inflow in the month i of the year j,
and $N_y$ is the number of years for which the data is available.

and,

$$S_i = \sqrt{\frac{\sum_{j=1}^{N_y} (X_{ij} - X_i)^2}{N_y - 1}} \qquad (2)$$

where, $S_i$ is the standard deviation of the inflows during the month i.

Figures 1 and 2 show respectively the mean and standard deviation of the inflows for each month. IN these figures and subsequent discussion, a year is the period from June 1st to May 31. The inflow data are arranged as a time series starting with the first month of the first year (June, 1934) through to the last month of the last year (May, 1974). Fig. 3 shows the plot of this time series. For further analysis we will focus on this series alone. All transformations are done only with respect to this series. This series consists of 480 values.

### Covariance Function and the Correlogram

The covariance function reflects the relationship between any value in the series to a value that is at some specified time periods away from it. This distance in time is called the lag. The covariance function at lag k, $R_k$, is computed as follows:

$$R_k = \frac{\sum_{t=1}^{N-K} (X_t - \overline{X})(X_{t+k} - \overline{X})}{N} \quad (3)$$

Where N is the total observations in the series (480 in this case). At k =0,

$$R_0 = \frac{\sum_{t=1}^{N} (X_t - \overline{X})^2}{N} \quad (4)$$

which is the variance of the series. The ratio $R_k/R_o$ is defined as the serial correlation coefficient, $C_k$, at lag k. Thus,

$$C_k = R_k / R_o \quad (5)$$

For a time series, normally the correlation coefficients are computed upto a lag of 0.15N, although theoreitcally they may be computed upto a lag of N-1. To visualise the strength of relationship between values, in this case, the correlation coefficients are computed upto a lag of 100, and are plotted against the lag. This plot called the correlogram, is shown in Fig. 4. It is abvious that the correlogram oscillates almost like a sine wave signifying the presence of a strong periodicity. The correlation coefficient reaches peaks at lags 12, 24, 36, 48,... This means that there is a strong correlation between the inflow during a month to that during the same month of the previous years. In hydrology this is a common feature becauses of the yearly cyclicity. Another important aspect of the correlogram is that it is slowly decaying with the lags. This means that the correlation between the inflow during a month in a year to the inflow durint the same month during previous years decreases as more data are considered. Therefore, for example, last year's June inflow will be similar to the present year's June inflow. TheJune inflow of the year previous to the last year would have a smaller resemblance to this years June inflow and so on.

The first conclusion dranw from the corrleogram is, therefore, that there is a strong periodicity in the data and that there is a cyclicity of 12 months. But a closer inspection of the correlogram reveals that at lags 6, 18,30,42,...the correlations are quite high and so also at lags 4, 16, 28, 40,... Merely by looking at the correlogram, therefore, it would not be possible to draw a conclusion about exactly which of these lags influence the inflow during a period significantly. For this purpose furher analysis is necessary.

### Spectral Analysis

The spectral analysis is necessary to identify the periodicities in the data. In this study, the spectrum of KRS inflows is plotted using two different functions as ordinates as explained below: (a) The first estimate results from expressing the time series ($x_t$) as a fourier series, (Chatfield, 1980). ie.,

$$x_t = a_o + \sum_{p=1}^{N/2-1} [\alpha_p \text{Cos}(w_p t/N) + \beta_p \text{Sin}(w_p t/N)$$
$$+ a_{N/2} \text{Cos}(\pi t) \text{ for } t = 1, 2, \ldots\ldots, N \quad (6)$$

where,

$$w_p = \frac{2\pi p}{N}$$

$$a_o = \mu$$

$$a_{n/2} = \sum_{t=1}^{N} (-1)^t X_t / N$$

$$\alpha_p = 2\sum_{t=1}^{N} X_t \text{Cos}(w_p t/N) \quad (7)$$

$$\beta_p = 2\sum_{t=1}^{N} X_t \text{Sin}(W_p t/N) \quad (8)$$

The plot of $(\alpha^2+\beta^2)N/2$ versus $w_p$ is called a line spectrum. Fig. 5 shows the line spectrum of the monthly inflows to KRS reservoir. It musth be noted that $w_p$ can be experessed as a frequency (in cycles per month, cpm ) by dividing it by $2\pi$.

The line spectum shows 4 prominent peaks. These peaks represent the periodicities inherent in the data. It may be seen from the line spectrum (Fig. 5) that, as a rough approximation the first two periodicities can be taken as significant and the others may be neglected. However, for an accurate inference further analysis is necessary. The period (in months) corresponding to any value of $w_p$ may be computed by $2\pi / w_p$, so that the first peak which corresponds to $w_p = 0.5238095$ results in a periodicity of 12 months. The second peak at $w_p = 1.047619$ shows that the next level of periodicity is at 6 (12/2) months, the third at 4 (12/3) months, the fourth at 3 (12/4) months and so on.

The line spectrum thus transforms the information from tie domain to the Frequency domain. While the correlogram indicate merely the presence of periodicities in he data, the spectral analysis helps identify accurately the signficant periodicities them selves. A difficulty that exists in the line spectrum is that it is not a consistent estimator of the spectrum. The second estimate of the spectrum overcomes this difficulty.

(b) An estimate of the power spectrum is given by,

$$f(w) = \frac{1}{\pi} [\lambda_0 R_0 + 2 \sum_{k=1}^{M} \lambda_k R_k \mathrm{Cos}\,(wk)] \qquad (9)$$

where [$\lambda_k$ ] are a set of weights called Lag Windows and $R_k$ is covariance function at lag k. For estimating [$\lambda_k$], many functions, called the windows are proposed in the literature (e.g., Tukey Window and Parzen Window, Chatfield (1980). In this study, the Tukey Window is used. It is given by ,

$$\lambda_k = \frac{1}{2} [ 1+\mathrm{Cos}\,(\frac{\pi k}{M})] \quad k = 0,1,2, \ldots\ldots M \qquad (10)$$

The value of M, called the truncation point, which represents the maximum number of lags, was chosen as 48 (0.1N). Fig. 6 shows the plot of $f(w_p)$ versus $w_p$ for p = 0, 1,2,..240. This is the smoothened spectrum of the inflows. The peaks occur at the same values of $W_p$ as in the previous case (Fig. 5), underlining the presence of periodicities of 12, 6, 4 and 3 months. The estimate defined by equation (9) is statistically consistent, where as the estimate used in Fig. 5, $(a^2 + \beta^2)N/2$, is statistically inconsistent (Chatfield, 1980).

An obvious question that remains to be answered is the one relating to the significance of the periodicities. To test the significance the first two periodicities which are assumed to be significant are removed from the original series to get a new series $\{Z_t\}$, where

$$Z_t = X_t - Y_t \qquad (11)$$

$$Y_t = \mu + \hat{\alpha}_1 \mathrm{Cos}\,(w_1 t) + \hat{\beta}_1 \mathrm{Sin}\,(w_1 t) + \hat{\alpha}_2 \mathrm{Cos}\,(w_2 t) + \hat{\beta}_2 \mathrm{Sin}\,(w_2 t) \qquad (12)$$

where $w_1$ is the $w_p$ value in Fig. 6 corresponding to the first peaks and $w_2$ to the second peaks.

The corresponding coefficients $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\alpha}_2$, $\hat{\beta}_2$ are estimated from equations (7) and (8). The new series $\{Z_t\}$ is free of the first two periodicities. This can be seen from Fig. 7, which shows the line spectrum of the series $\{Z_t\}$. But there appear three peaks which are prominent in the spectrum shown in Fig. 7. A hurried and wrong conclusion from Fig. 7 may be that these periodicities are indeed significant. However, they need to be analysed for their statistical significance.

### Significance of the Periodicities

The periodicities are tested for significance following the procedure given by Kashyap and Rao (1976). A statistic $\cap$ is defined as,

$$\cap = \frac{\gamma^2(N-2)}{4\hat{\rho}_1} \qquad (13)$$

where,

$\gamma^2 = \alpha^2 + \beta^2$ and

$$\hat{\rho}_1 = \frac{1}{N} \left[ \sum_{t=1}^{N} \left\{ X_t - \hat{\alpha}\ \mathrm{Cos}\,(w_p t) - \hat{\beta}\ \mathrm{Sin}\,(w_p t) \right\} \right] \qquad (14)$$

The periodicity corresponding to $w_p$ is significant at level $\alpha$ only if

$\cap \geqslant F_\alpha (2, N-2)$

where F denotes the F-distribution.

For the KRS inflows, this test was done for periodicities of 4,3 and 2 months at 95% significance level. It must be noted that this test examines the significance of one periodicity at a time, and should be carried out on a series from which all the periodicities, previously found significant are removed. That is, to test the significance of the first periodicity, the original series is considered. If this is found significant, for testing the significance of the second periodicity, the test should be done on a series from which the first periodicity is removed and so on. Table 1 gives the results of these tests for different periodicities of the KRS inflows. It can be seen that all the periodicities given in the table 1 are significant for the original series.

A necessary condition in stochastic models of the ARMA (auto regressive moving averge ) family is that the series being modelled must be free from any significant periodicities. One way of removing the periodicities from the time series is to simply transform the series into a standradized one. One method of standardising a series $\{X_t\}$ is by expressing $\{X_t\}$ as a new series $\{Z_t\}$ where,

$$Z_t = (X_t - \bar{X}_i)/S_i \qquad (15)$$

where $X_i$ is the estimate of the mean of the inflows of the month i to which period t belongs and $S_i$ is the estimate of the standard deviation of the inflows of the month i. The series $\{Z_t\}$ has zero mean and unit variance.

Fig. 8 shows the the correlogram of the standardised series. It also shows the standard error ($+ 2/\sqrt{N}$) of the correlation coeficients. It is seen that correlations of only a few lags are significant. Also, the periodicities that were so prominently seen in the correlogram of the original series are absent.This absence of periodicities is also seen in the spectrum of the standardised series (Fig. 9).This spectrum is estimated by using the Tukey window for $\{\lambda_k\}$ in equation (9). It is seen from Table 1 that the periodicities given in the table are insignificant for the standardised data. Thus, by standardising the data, we are able to remove all the periodicities. This is a significant property of standardisation and comes as a handy tool in the synthetic generation and forecasting of the inflows.

## CONCLUSIONS

A detailed spectral analysis of the Cauvery river flows at the Krishna Raja Sagar (KRS) reservoir site is presented in this paper. Procedures commonly employed for identifying the significant periodicities in the data are explained in detail.

Forty years of data on monthly flows of the Cauvery river have been used to identify the significant periodicities in the flows. The correlogram- which gives a first, rough estimate fo the periodicities - showed the presence of a periodicity of 12 months. For most sites in India, the monthly streamflows may exhibit a strong periodicity of 12 months because of a fairly regular pattern of the monsoon rainfall, which is the major source for streamflows in India. However, for an accurate modelling of the streamflows, periodicities of less than one year (12 months) need to be identified. This is not possible with the correlogram alone, and a frequency domain analysis becomes necessary. The spectral analysis presented in this paper is one such procedure of identifying all the significant periodicities present in the flow data. A line spectrum, that transforms the data from time domain to the frequency domain, showed that the Cauvery river flows exhibit periodicities of 12, 6, 4 and 3 months. The significance of each of these periodicities is examined and it has been found that all theses periodicities are statistically significant. In the time domain modelling of the flows, one preliminary condition that needs to be satisfied often is that the series should be free of any significant periodicities. It has been verified in the present study that a simple standardisation of the series removes all the periodicities in the data. This transformed series may then be used for constructing the stochastic models of the flows (Mujumdar and Nagesh Kumar, 1990).

## ACKNOWLEDGEMENT

This work originated from a series of lectures on Time Series Analysis delivered at the Indian Institute of Science, Bangalore by Dr. A. Ramchandra Rao, Professor, Purdue University. The authors thank him for introducing them to the important field of Time Series Analysis.

## REFERENCFS

1. Chatfield C. 'Analysis of Time Series : An Introduction', Chapman & Hall, London, 1980.
2. Jenkins G.M., and Watts D.C., 'Spectral Analysis ofTime Series', Holden - Day, Sanfransisco, 1968.

3. Kashyap R.L. and Ramchandra Rao A. 'Dynamic Stochastic Models From Empirical Data', Academic Press, New York, 1976.
4. Mujumdar P.P. and Nageshkumar D., 'Stochastic Models of Steamflow-some case studies', Hydrological Sciences Journal, UK, 35(4), 1990, pp. 395-410.
5. Pandit Sudhankar M and Wu Shien-Ming. 'Time Series and System Analysis with Applications', Wiley, New York, 1983.

Table 1. Test for Significance of Periodicities

| Value for Periodicity | Actual data | Standardised Data | $F_{0.95}(2, N-2)$ |
|---|---|---|---|
| 12 Months | 5.987E + 19 | 0.241E-02 | 3.00 |
| 6 Months | 1.762E + 19 | 0.804E-02 | 3.00 |
| 4 Months | 0.791E + 19 | 1.031E-02 | 3.00 |

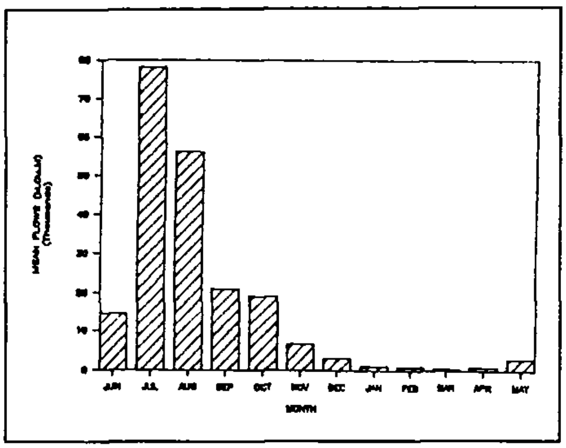

Fig. 1 Average Inflows into KRS Reservoir

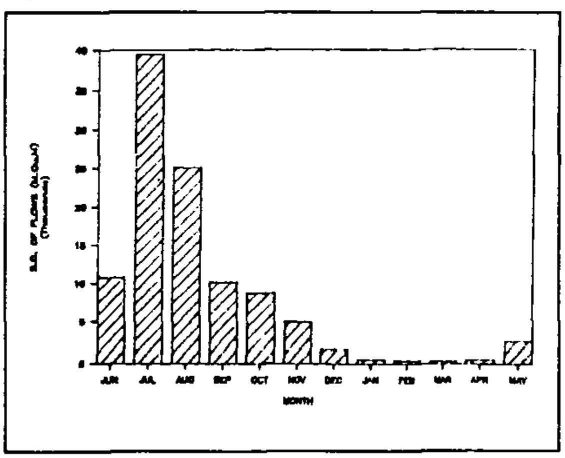

Fig. 2 Standard Deviations of Inflows into KRS Reservoir

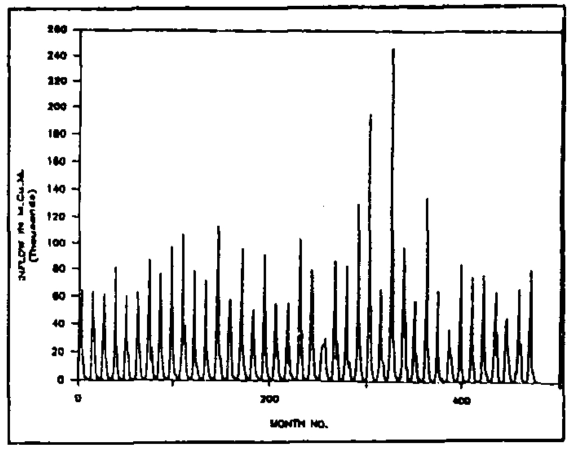

Fig. 3 Inflows into KRS Reservoir

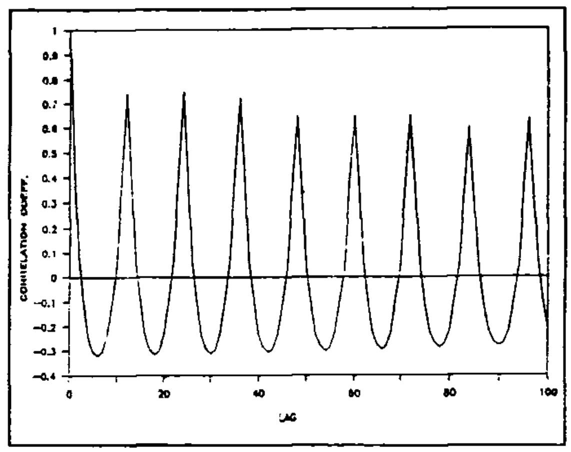

Fig. 4 Correlogram of Inflows

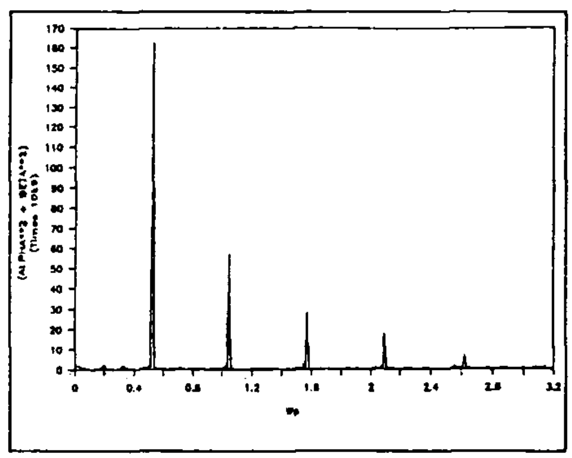

Fig. 5 Line Spectrum of Inflows

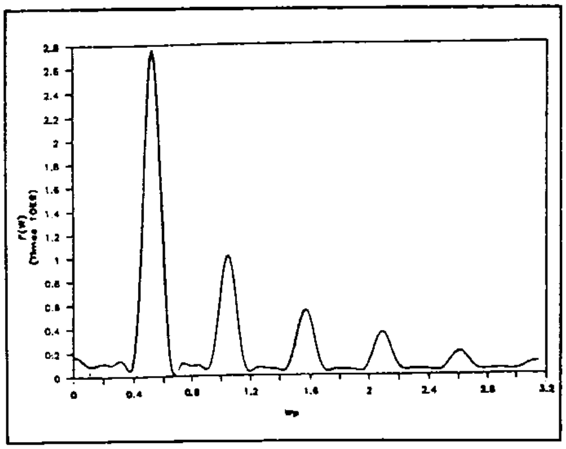


Fig. 6 Smoothed Spectrum of Inflows

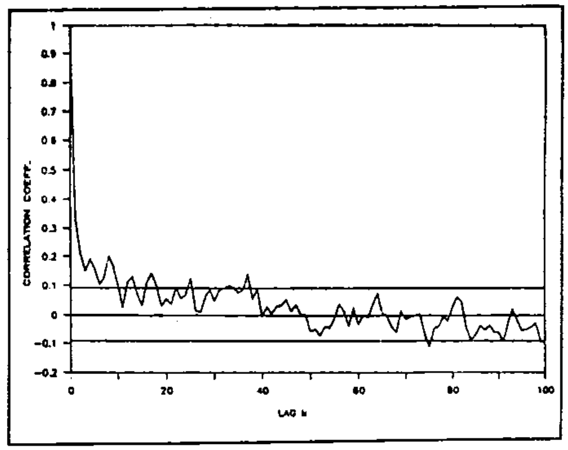


Fig. 8 Correlogram of Standradised Inflows

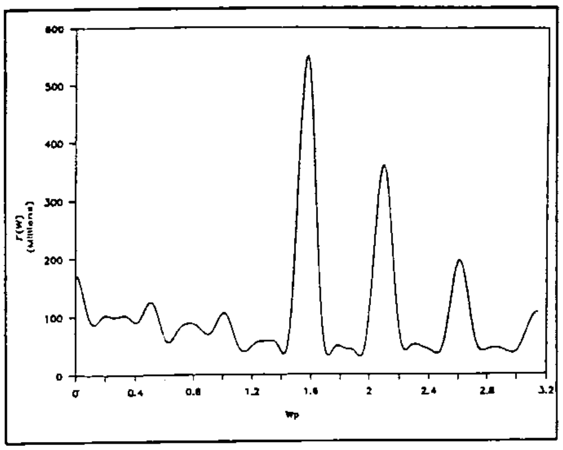


Fig. 7 Smoothed Spectrum of Residuals

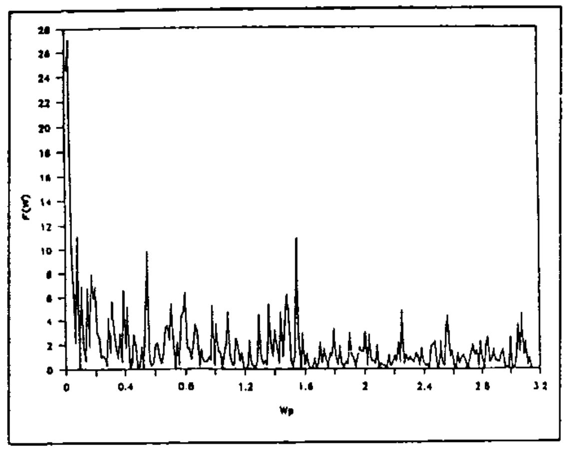


Fig. 9 Line Spectrum of Standardised Inflows