STOCHASTIC MODELLING OF DISSOLVED OXYGEN IN THE YAMUNA RIVER AT DELHI

The monthly observation for dissolved oxygen of the river Yamuna at upstream and downstream of Delhi were used for the water quality analysis. Fig. 1 and 2 shows the time series of mean monthly dissolved oxygen concentrations for the period 1977-1985. The data were observed and edited by the CWC.

TREND COMPONENT

The turning point test and Kandall's rank correlation test were carried out for the detection of trend. For U/S section the z value (-1.64) indicated no evidence of trend at the 5% level of significance (standard normal variate from published table at 95% level of significance is ± 1.96). This was confirmed by tests for the detection of linear trends (t value -1.64 which is less than t-critical ± 1.98). For D/S section the z value (-3.08) indicate the exitance of trend component. Further it also shows the exitance of linear trend (t value -3.45).

Kottegoda (1980) suggest that only annual data should be used for the analysis of trend by virtue of which the periodic component P_t is suppressed. The Kendall's rank co-rrelation tests were carried out using annual data which indicated that there were no exitance of trend component at both U/S and D/S sections of the river as the calculated z-values are less than their critical values.

For U/S section z = -0.147, and
For D/S section z = -1.67

PERIODIC COMPONENT

Periodicities in the series were identified through the construction of auto-correlogram and the spectrum analysis.

For U/S section auto-correlogram (Fig. 3) and the spectral density function (Fig. 4) show the periodic character of the Dissolved oxygen concentration. From the spectral density function it is clear that only first two harmonics are significant for U/S section. Table 1 summarizes the results of the harmonic analysis for U/S section.

Table 1 : Summary of Harmonic Analysis of U/S Data

Harmonic	A_i	B_i	P_i
1	-0.032	-0.464	0.017
2	-0.010	-0.016	0.017
3	0.301	0.330	0.032
4	-0.684	0.262	0.074
5	0.459	0.265	0.096
6	-0.437	0.0	0.111

Similarly, for D/S section auto-correlogram (Fig. 5) and spectral density function (Fig. 6) show the periodic characteristics of dissolved oxygen data. The Sp density function shows that all the six harmonics are significant and hence should be retained while substracting the periodic component from the original series. Table 2 summarises the results of harmonic analysis at D/S section.

Table 2 : Summary of Harmonic analysis of D/S Data

Harmonic	A_i	B_i	P_i
1	0.628	0.409	0.098
2	-0.024	-0.270	0.111
3	0.166	-0.075	0.117
4	-0.274	-0.019	0.130
5	0.047	0.530	0.180
6	0.008	0.0	0.180

DEPENDENT STOCHASTIC COMPONENT

Several ARMA & ARIMA models were used to describe the dependent structure of the stochastic component from the trend & periodicity free series.

For U/S section, auto-correlation function (acf) and partial auto-correlation function (pacf) were calculated for lags 1 through 27. The acf and pacf were plotted in Figs. 7 and 8. The estimated acf and pacf suggested a higher order model which is practically not justified as the dissolved oxygen depend on the past days rather than past months. Further, the estimated acf drops off slowly toward zero, which gives an indication of non-stationarity character of series mean. The differencing technique were applied to remove non-stationarity as suggested by Box-Jenkins (1976).

The first differences of DO data at U/S do not show any periodicity as shown by Sp. density function plotted in Fig. 9, which was removed automatically by differencing and hence no harmonic analysis was required for the differenced data. The estimated acf of first differences of DO data drops off rapidly to zero showing that the non-stationarity is no more present in the differenced series. The acf and pacf were plotted in Figs. 10 and 11 respectively. On the basis of estimated acf and pacf it appears reasonable to try the ARIMA (1, 1, 0), ARIMA (2, 1, 0) and ARIMA (1, 1, 1) models. The residual acf were used for diagonostic checking to test the hypothesis that the shocks of the applied model are statistically independent. The residual acf were calculated along with their t-values. The residual acf were plotted for ARIMA (1, 1, 0), ARIMA (2, 1, 0) and ARIMA (1, 1, 1) in Figs. 12, 13 and 14 respectively. Out of these three models ARIMA (2, 1, 0) is the best choice as the t-value for first 5 lags is insignificant in the residual acf of ARIMA (2, 1, 0), while ARIMA (1, 1, 0) residual has a significant t-value (more than 2) at lags 3, 6, 9 and 15 and ARIMA (1, 1, 1) residuals has t-value in the same pattern as ARIMA (2, 1, 0). Further, it is clear that among the ARIMA (2, 1, 0) and ARIMA (1, 1, 1) models, the ARIMA (2, 1, 0) is the best alternative as it has smaller t-values in the initial lags. The residuals of ARIMA (2, 1, 0) is tabulated in table 3.

Table 3 : Residual acf of ARIMA (2, 1, 0) for U/S section

lag	acf	t-value
1	0.012	0.072
2	0.058	0.600
3	0.137	1.417
4	-.036	-.365
5	-.182	-1.84
6	-.226	-2.23
7	-.115	-1.08
8	-.112	-1.04
9	-.246	-2.27
10	-.114	-1.00
11	0.062	0.546
12	0.076	0.665
13	0.101	0.873
14	0.084	0.728
15	0.336	2.881
16	-.002	-.019
17	-.082	-.658
18	0.101	0.804
19	-.140	-1.107
20	-.142	-1.112
21	-.038	-.292
22	-.146	-1.124

The coefficients of ARIMA (2, 1, 0) model were estimated using the maximum likelyhood criterion. The estimated coefficients were:

ϕ_1 = -0.588 and ϕ_2 = -0.15

This model satisfy the stationarity requirements:

$$| \phi_2 | < 1$$
$$| \phi_2 + \phi_1 | < 1$$
$$| \phi_2 - \phi_1 | < 1$$

For D/S section acf and pacf were calculated for lag 1 to 27 and plotted in figs. 15 and 16. The estimated acf drops off quickly to zero which shows that the mean of the series is stationary and hence there is no need of differencing. Further, both acf and pacf coefficients are significant up to lag 2 and then oscillates with in the confidence band suggesting AR models to explain the D/S DO data.

AR(1) and AR(2) models were tried and the residual acf are shown in figs. 17 and 18. The estimated AR(1) coefficient ϕ_1 (equal to 0.25) is significantly different from zero and satisfy the stationarity condition. However, the residual acf is not good at all: the absolute t-value at lag 2 is 3.39 which is far larger than the residual acf short lag warning level of 1.25. This indicate that AR(1) model fail to give independent residuals and hence it is not an adequate model to explain the D/S data.

Fig. 18 and table 4 show that AR(2) model is satisfactory. All estimated coefficients (ϕ_1=0.167 and ϕ_2=0.334) are statically significant and they meet the stationarity conditions. Diagonostic checking using the residual acf indicates that AR(2) is an adequate model: as none of the residual auto-correlations in table 5 and fig. 18 has an absolute t-value larger than our practical warning value (1.25). Furthermore, according to the chi-square test the residual auto-correlations are not significantly different from zero as a set. The estimated chi-squared statistics (equal to 22) is not significant. For 25 degree of freedom this statistic would have to exceed 34 to indicate statistical dependence in the random shocks at the 10% level.

Table 4 : Residual acf of ARIMA (2, 0, 0) for D/S section

lag	acf	t-value
1	0.014	0.084
2	-.023	-.232
3	-.050	-.519
4	0.061	0.631
5	0.022	0.222
6	0.014	0.148
7	0.078	0.793
8	0.034	0.348
9	0.001	0.008
10	0.116	1.180
11	0.080	0.797
12	-.066	-.660
13	-.029	0.292
14	-.005	-.053
15	-.093	-.917
16	-.003	-.034
17	0.167	1.640
18	0.195	1.870
19	0.057	0.5287
20	-.148	-1.36
21	0.064	0.528
22	-.074	-.672
23	0.032	0.291
24	0.012	0.111
25	0.168	1.516
26	-.022	-.195
27	-.164	-1.45

INDEPENDENT RESIDUAL COMPONENT

The dependent stochastic part represented by ARMA (p, q) model was substracted from the series. The remaining series containing independent stochastic part is called as the residual series which can only be described by some probability distribution function. In Box-Jenkins ARIMA modelling it is assumed that the independent residual component is normally distributed. Therefore, if the ARIMA modelling is correct the residual part must follow a normal distribution otherwise one should think to improve the ARIMA model so that the residual part may follow a normal distribution.

For U/S section the independent residue component has the following parameters:

Mean	=	0.02
Stand. Dev.	=	1.82
Skewness	=	0.645
t-statistic	=	-1.98

For normality t-test was carried out, and the t-statistic is calculated by the following formula:

$$\text{t-statistic} = \frac{\textit{calculated value} - \textit{hypothised value}}{\textit{standard error of estimate}}$$

where, estimated standard error is given by:

$$Se(g) = \sqrt{\frac{6N(N-1)}{(N-2)\,(N+1)\,(N+3)}}$$

Using this equation, Se(g) for U/S = 0.2335 and t-value calculated using the equ. (53) is found to be -1.98 while critical t-value is 1.98 for 95% confidence limit with two tailed test. Since calculated t-value is not smaller than critical t-value, we can't accept the hypothesis at 5%, level of significance, and hence the independent residue component can not be assumed normally distributed. As, already stated that the ARIMA model at the U/S section needs further improvement. However, as the calculated t-value for the sample is just equal to the critical t-value, one may roughly asume that the residue component is normally distributed. The independent residue can be generated using the following normal distribution equation evaluated by least square fitting:

$$a_t\,(c) = 0.0189 + 1.809\, z_{at}$$

in which,

$$z_{a_t} = \text{reduced variate} = \frac{\hat{a}_t - \bar{a}_t}{\alpha_{a_t}}$$

$a_t(c)$ = independent residue

The coeff. of co-relation is 0.997. The fitted normal distribution and the independent residual is shown in Fig. 19.

For D/S section the independent residual part has the following statistics:

Mean	=	0.02
Stand. Dev.	=	1.58
Skewness	=	0.058
t-statistic	=	0.248

Because the calculated t-value is less than the critical t-value, the null hypothesis is accepted. The independent residue component can be generated using the following normal distribution:

$$a_t\,(c) = 0.0182 + 1.5895\, z_{at}$$

where,

z_{at} = reduced variate
$a_t\,(c)$ = independent residue

The coeff. of co-relation is 0.999. The fitted normal distribution and independent residue is shown in Fig. 20.

RESULTS AND DISCUSSION

In this study, an attempt has been made to analyze mean monthly dissolved oxygen concentrations observed at the U/S and D/S sections of the Yamuna river in Delhi. The results of the analysis are discussed hereunder.

There was no indication of exitance of trend in the annual data of dissolved oxygen concentrations at both the U/S and D/S sections. However, the monthly data at D/S section indicated the exitance of trend. This may be due to the highly periodic nature of D/S DO data which was supposed to be removed in the analysis of annual data.

The periodicity was detected in both the series at U/S and D/S sections with the help of auto-correlogram and spectral density function. At U/S section only first two harmonics were found significant while at D/S section all the six harmonics were found significant. This may be due to the reflection of seasonal character of both the withdrawal of river water and waste disposal in between the U/S and D/S sections.

The stochastic component is comprised of dependent stochastic component and independent residue component. The dependent stochastic component is modelled using ARIMA models whose results are given in the tables 5.

Table 5 : Chosen ARIMA models and their parameter

Station	ARIMA model	Parameters of model	Model
U/S	ARIMA (2,1,0)	ϕ_1 = -0.588 ϕ_2 = -0.147, C=0.0	$w_t = -.558w_{t-1} - .147w_{t-2} + a_t$ $w_t = \nabla^1 z_t = (1-B)^1 z_t$
D/S	ARIMA (2,0,0)	ϕ_1 = -0.167 ϕ_2 = -0.334, C=0.0	$z_t = .167z_{t-1} + .3342_{t-2} + a_t$

The ARIMA (2, 1, 0) model needs further improvements as it fail to give residual auto-correlations that were significantly different from zero as a set (chi-square test). However, higher order model does not seem to be appropriate in the case of dissolved oxygen concentration as it depends on previous days but not on the previous 3 or 4 months. Further, there may be some another causes which influence the DO concentration at U/S section. For D/S section ARIMA (2, 0, 0) is an adequate model as it give residual auto-correlations that are not significantly different from zero as a set.

Independent residue component was represented by normal distributions as given in Table 6.

Table 6 : Characteristics of residual series and parameters of chosen distributions

Characteristics of residual series obtained after ARIMA model					Parameter of fitted normal distribution: $a_t(c) = \alpha + \beta z_{at}$			
Obs. St.	Mean	Skewness		Std. dev.	α	β	Corr coeff	Std. Error of reg.equ.
		Coeff	t-val					
U/S	0.02	-.4645	-1.98	1.82	.0189	1.809	0.997	0.30
D/S	0.02	0.0580	0.248	1.58	.0182	1.589	0.999	0.14

The residual series obtained after ARIMA model for U/S section are skewed towards right (skewness=-.4645) and calculated t-value is just equal to the critical t-value which indicate that the residuals have some what dependency which should be removed by further improving the ARIMA model. For D/S section, the residuals follow a normal distribution having more or less zero skewness (0.058) and insignificant t-value.

CONCLUSIONS

The D/S data series were adequately described by low-order ARMA models (p, q $\leq$ 2). But for the U/S data, the residual series showed some dependence even after fitting ARIMA (2, 1, 0) model. A higher order model, ARIMA (p, q $\geq$ 2), should therefore be used for this data series. The dependence of the residual series for low order ARIMA model for this series may imply that there are still some deterministic elements remaining in the series even after removing trend and periodicity by the methods described.

The two annual data series for U/S and D/S sections did not show any trend, which could be due to the reason of availability of small size of data (only 9 years), otherwise the increase in the effluent discharges from industrial, agricultural, and domestic sectors that have rapidly expanded in the recent past could have been reflected by showing the trend component. The seasonal variations in water quality is greatly influenced by the annual weather cycle and the cyclic pattern of the hydrological inputs of the river water environment. Uncertainties that are of a random nature, e.g., measurement errors, unexpected high levels of effluent discharges near the sampling site, and variation of the sampling point within the river cross-section, are reflected in the residue series.

The D/S series showed strong periodicity in comparison of U/S data series. This periodicity could be induced due to the periodic nature of raw water withdrawal and disposal of effluent discharges from various sources e.g. industrial, agricultural, and domestic sources etc.

The Box-Jenkins method for time series analysis was successful in modelling the monthly water quality data in the Yamuna river near Delhi. The models were parsimonious and physically reasonable. Dissolved oxygen data for U/S required a first difference-moving average model, while DO data for D/S section required moving average model without any difference.

The coefficients of selected ARIMA models were estimated by method of maximum likelyhood. However, for precise estimation of model coefficients one should use a better optimisation technique such as grid search technique or Marquardt's compromise scheme.

The Box-Jenkins technique was able to employ defective data, containing an oscillation believed to be machine-induced, to obtain a workable model. This method provides the water quality analyst with a new technique which may succeed where other methods would not. It can also serve as an alternative, and perhaps superior approach in situations where other methods can be employed.

** *** **

DATA SOURCE : ANNEX 4.7/2, STATUS REPORT ON WATER QUALITY FOR YAMUNA SYSTEM , CENTRAL WATER COMMISSION, NEW DELHI, APRIL ,1990

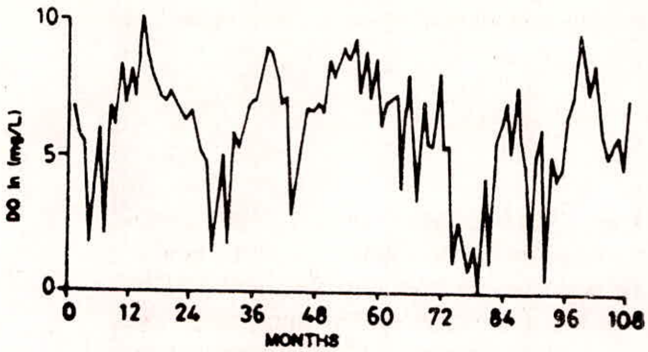

FIG. 1 TIME SERIES OF MONTHLY MEAN DO CONCENTRATIONS AT U/S SECTION IN YAMUNA AT DELHI FOR THE PERIOD NOV. 1981 - OCT. 1990.

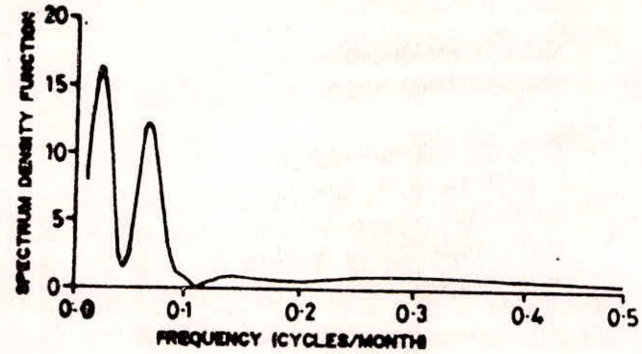

FIG. 4 POWER SPECTRUM OF U/S DO SERIES AT DELHI ON RIVER YAMUNA

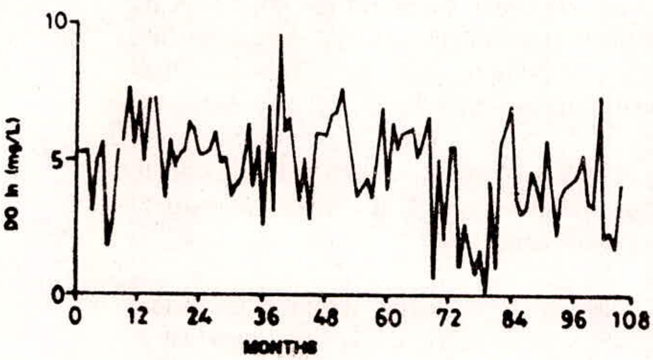

FIG. 2 TIME SERIES OF MONTHLY MEAN DO CONCENTRATIONS AT D/S SECTION IN YAMUNA AT DELHI FOR THE PERIOD NOV., 1981-OCT., 1990.

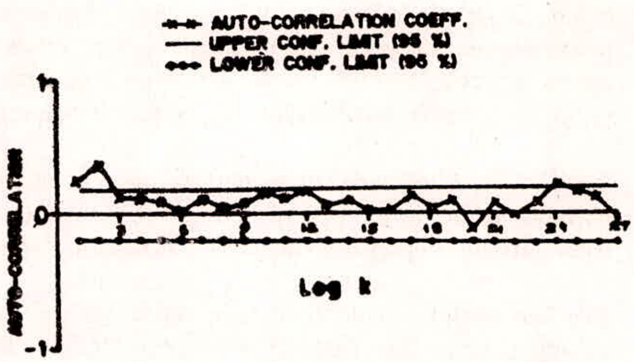

FIG. 5 AUTO-CORRELATION FUNCTION OF HISTORIAL DO SERIES AT D/S OF YAMUNA IN DELHI.

YAMUNA IN DELHI.

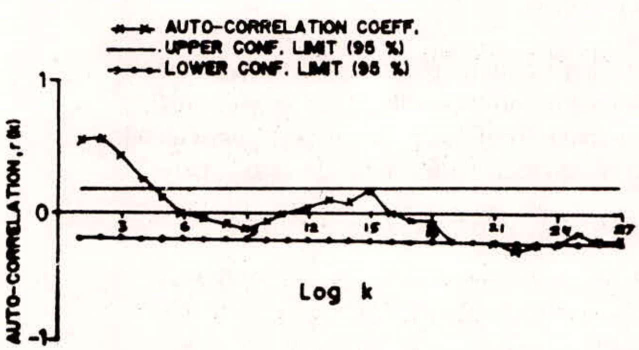

FIG. 3 AUTO-CORRELATION FUNCTION OF HISTORICAL DO SERIES AT U/S OF YAMUNA IN DELHI.

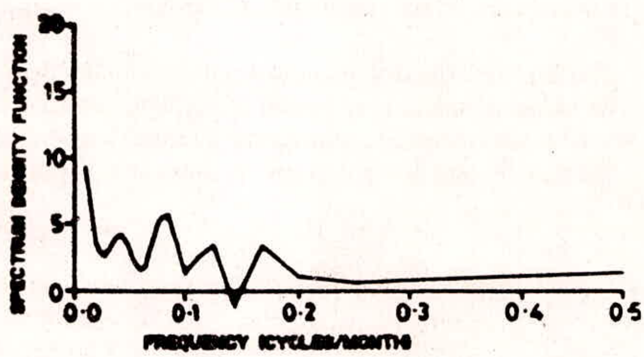

FIG. 6 POWER SPECTRUM OF D/S DO SERIES AT DELHI ON RIVER YAMUNA.

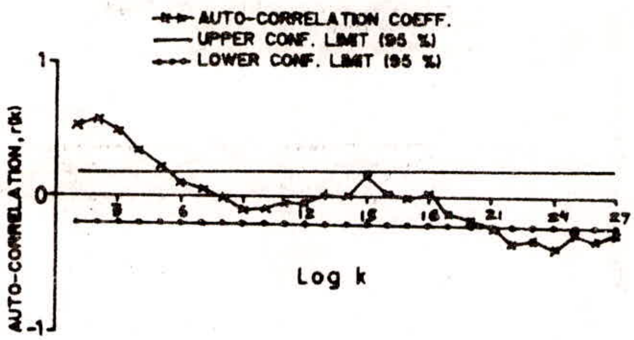

FIG. 7 AUTO-CORRELATION FUNCTION OF DO SERIES AT U/S OF DELHI ON RIVER YAMUNA (AFTER REMOVING PERIODIC COMPONENT)

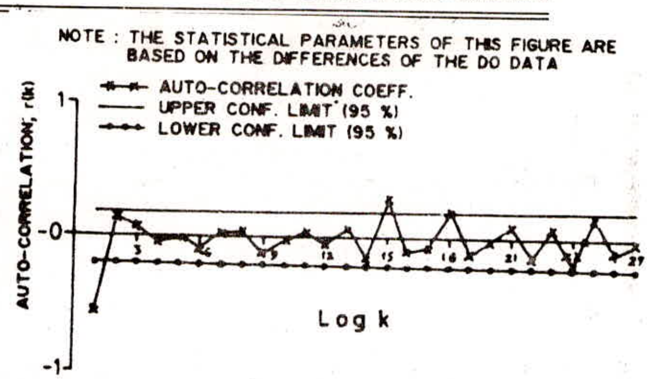

FIG. 10 AUTO-CORRELATION FUNCTION OF DO SERIES AT U/S OF DERLHI ON RIVER YAMUNA (AFTER REMOVING PERIODIC COMPONENT)

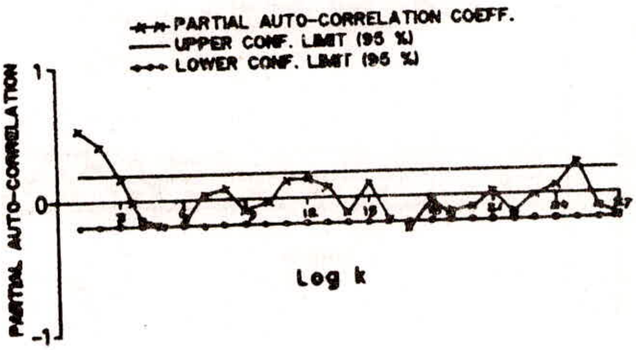

FIG. 8 PARTIAL AUTO-CORRELATION FUNCTION OF DO SERIES AT U/S OF DELHI ON RIVER YAMUNA (AFTER REMOVING PERIODIC COMPONENT)

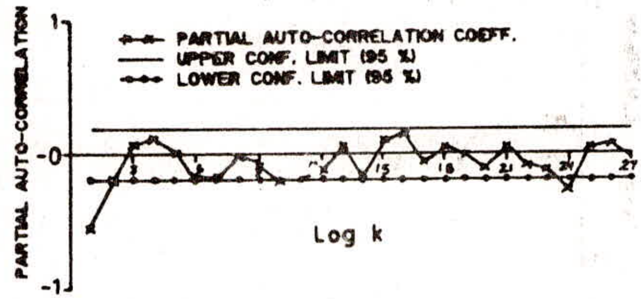

FIG. 11 PARTIAL AUTO-CORRELATION FUNCTION OF THE FIRST DIFFERENCES OF DO DATA AT U/S OF DELHI ON RIVER YAMUNA (AFTER REMOVING PERIODIC COMPONENT)

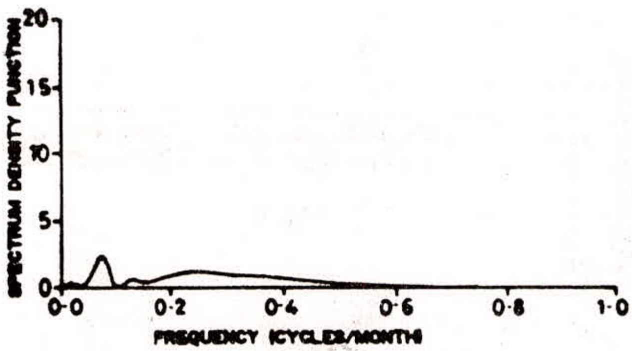

FIG. 9 POWER SPECTRUM OF FIRST DIFFERENCES OF DO DATA AT U/S OF YAMUNA IN DELHI.

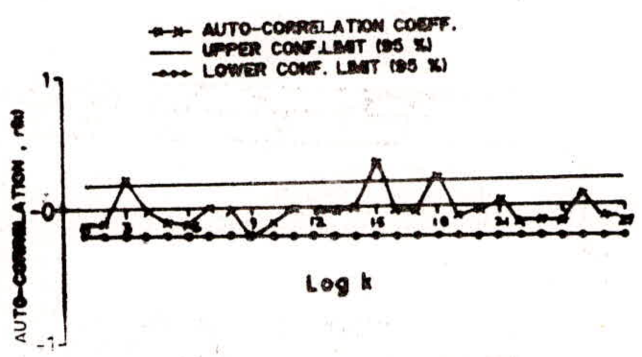

FIG. 12 RESIDUAL AUTO-CORRELATION FUNCTION FOR THE RESIDUALS FROM THE ARIMA (I,I,O) MODEL FOR THE U/S DO SERIES, YAMUNA, DELHI.

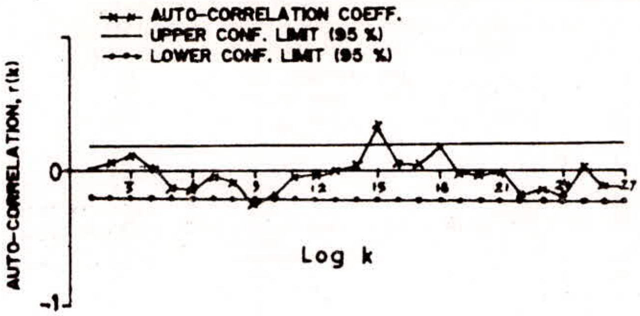

FIG. 13 RESIDUAL AUTO-CORRELATION FUNCTION FOR THE RESIDUALS FROM THE ARIMA (2,1,0) MODEL FOR THE U/S DO SERIES, YAMUNA, DELHI.

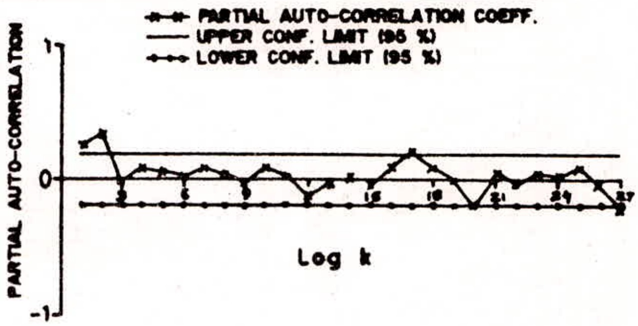

FIG. 16 PARTIAL AUTO CORRELATIN FUNCTION OF DO SERIES AT D/S OF DELHI ON RIVER YAMUNA (AFTER REMOVING PERIODIC COMPONENTS).

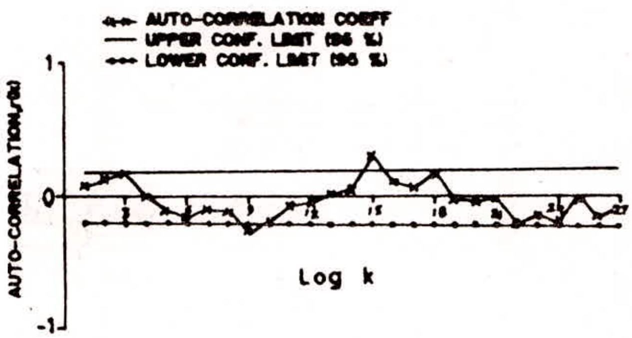

FIG. 14 RESIDUAL AUTO-CORRELATION FUNCTION FOR THE RESIDUALS FROM THE ARIMA (1,1,1) MODEL FOR THE U/S DO SERIES, YAMUNA, DELHI

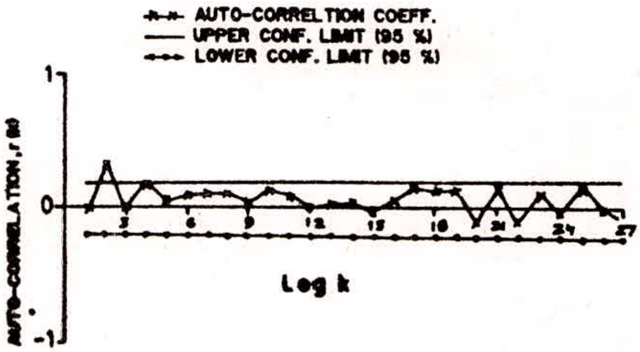

FIG. 17 RESIDUAL AUTO-CORRELATION FUNCTION FOR THE RESIDUALS FROM THE AR (1) MODEL FOR THE D/S DO SERIES, YAMUNA, DELHI.

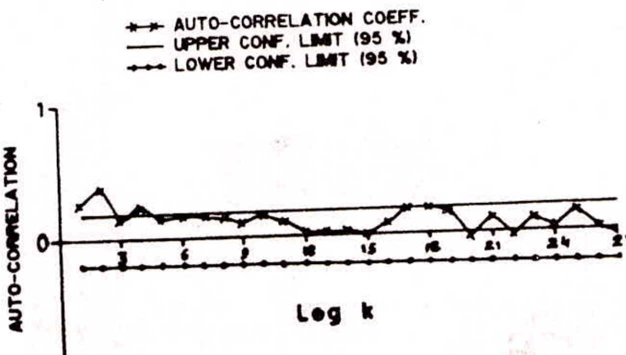

FIG. 15 AUTO-CORRELATION FUNCTION OF DO SERIES AT D/S OF DELHI ON RIVER YAMUNA (AFTER REMOVING PERIODIC COMPONENT).

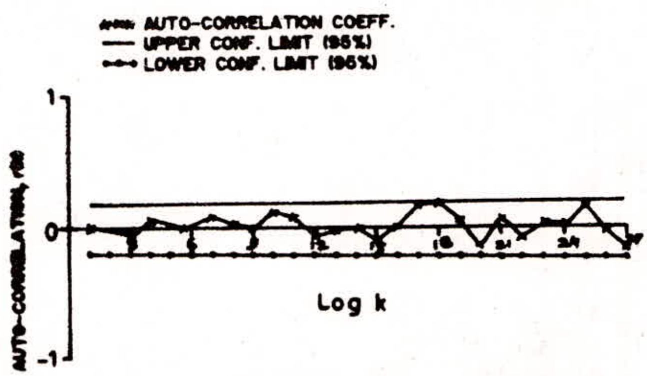

FIG. 18 RESIDUAL AUTO-CORRELATION FUNCTION FOR THE RESIDUALS FROM THE AR(2) MODEL FOR THE D/S DO SERIES, YAMUNA, DELHI

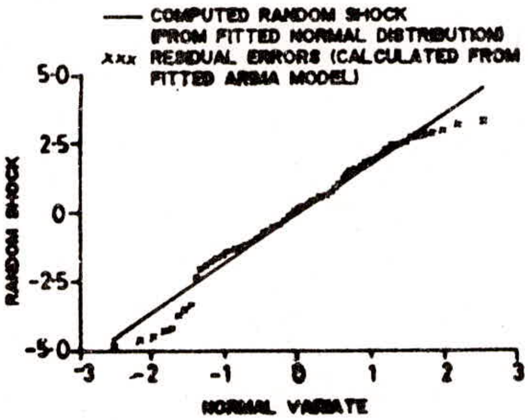

FIG. 19 NORMAL DISTRBUTION FITTED FOR THE DO AT U/S SECTION OF YAMUNA AT DELHI

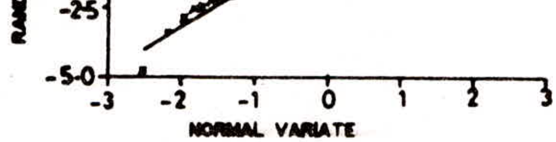

FIG. 20 NORMAL DISTRIBUTION FITTED FOR THE DO AT D/S SECTION OF YAMUNA AT DELHI.