N.A. Bhatia , T.M.V.Suryanarayana/ International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 1, Issue 3, pp.1072-1079

Study of Time Series and Development of System Identification Model for Agarwada Raingauge Station

N.A. Bhatia[1] and T.M.V.Suryanarayana[2]

[1]Teaching Assistant,
[2]Assistant Professor,
Water Resources Engineering and Management Institute,
Faculty of Technology and Engineering,
The Maharaja Sayajirao University of Baroda, Samiala
Ta. & Dist.: Vadodara, Gujarat, India

ABSTRACT

Rainfall data from the year 1965 to 2005 have been analyzed and then plotted with respect to time to analyze time series plot and autocorrelation plot.. From the autocorrelation plots, it is observed that upto lag 4 to lag 5, which is around 10% of the length of the data, the results shows better correlations and for lags beyond 10% of the record length, there exist less correlation and more variations. Further, the rainfall data have been used to develop and analyze model using soft computing technique called System Identification Model (SIM) to predict and forecast future values. An attempt has been made to develop the best fit model for Agarwada Station of Panam catchment area by trying different models like linear parametric models, process models, correlation models and non-linear models. Model is developed for 70% data and validated for remaining 30% data. Best Fit values of each model has been checked and compared for analysis purpose. In present study, the best fit value comes out to be 100 and it is concluded that process models are the best preferred models for Agarwada Station According to Akaike's theory, the most accurate model has the smallest FPE and Loss Function. The results obtained for the same are 3.46e-024 and 2.19e-024 respectively, which are almost equal to zero.

Key words : Time Series Plot, Autocorrelation Plot, System Identification Model, Model Residual Plot, Poles And Zeros

1. INTRODUCTION

Arrangement of statistical data in chronological order i.e., in accordance with occurrence of time, is known as “Time Series”. In other words, a time series is a set of observed data recorded at specified times, usually spaced at equal intervals. Hydrological processes are intrinsically continuous in time. Measurements of hydrological parameters are inherently discontinuous, or discrete in time. Retaining the information content from the continuous to the discrete domains requires significant planning, and often tradeoffs in experimental design. Mathematically, a time series can be expressed by the values Z1, Z2, Z3, etc. of a variable Z at times t1, t2, t3, etc. The primary purpose of the analysis of time series is to discover and measure all types of variations which characterize a time series. The central objective is to decompose the various elements present in a time series and to use them in decision making.

System Identification model lets one estimate linear and nonlinear mathematical models of dynamic systems from measured data. One can use the resulting model to simulate the output of a system for a given input and analyze the system response, predict future system outputs based on previous inputs and outputs. Different models like process models, linear parametric models, non-linear models and spectral models can be tried out under SIM. After trying different models, one can make out the best fit model to predict future values for that particular station.

Forecasting runoff ratio variation is important since it can be used to calculate the runoff of the river catchment for a given rainfall on a particular day. De Silva (2006) developed a model SARIMA

(seasonal auto regressive integrated moving average) for this purpose which was based on autocorrelation function (ACF) and partial autocorrection function (PACF) and was fitted to the runoff ratio data of the Rantanpura catchment of the Kalu Ganga basin. The appropriateness of the same model was tested against two other catchments Dela and Ellagawa of the same river basin and found to be suitable.

Geophysical time series often contain missing data, which prevents analysis with many signal processing and multivariate tools. Schoellhamer (2001) developed a modification of singular spectrum analysis for time series with missing data and successfully tested with synthetic and actual incomplete time series of suspended-sediment concentration from San Francisco Bay. This method can also be used to low pass filter incomplete time series.

Time series analysis and forecasting has become a major tool in different applications in hydrology and environmental management fields. Among the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins, ARIMA (Autoregressive Integrated Moving Average). Momami (2009) used Box-Jenkins methodology to build ARIMA model for monthly rainfall data taken for Amman airport station for the period from 1922-1999 with a total of 936 readings. Results: In the research, ARIMA (1, 0, 0) (0, 1, 1)12 model was developed. He found out that this model is used to forecasting the monthly rainfall for the upcoming 10 years to help decision makers establish priorities in terms of water demand management. Conclusion/Recommendations: An intervention time series analysis could be used to forecast the peak values of rainfall data.

Paudel et al. (2005) developed a profit maximization model and an ARIMA model to forecast water demand for broiler production. Broiler production decisions were made in three successive stages -- primary broiler breeding flock, hatchery flock, and finishing broiler production. The forecasted numbers of broilers from structural and ARIMA models diverge significantly from a USGS physical model. Analysis indicated 15% slippage in water demand forecasting related to disregarding the role of economic variables. They found that an appropriate lag structure can fully capture the information used in structural models, assuming no structural change.

2. METHODOLOGY

Different plots have been developed for Agarwada Raingauge Station and are explained as follows:

2.1 Time Series Plot

Time series plot is used to develop and analyze time series. Time series plot shows the rainfall pattern with respect to regular interval of time. By studying rainfall pattern for past values one can analyze the trend and by making a suitable rainfall model one can predict the future values and can also find the missing data of that station and of nearby station also. By plotting data as a function of time, one can quickly gain insight into the following data features: Outliers, or values that do not appear to be consistent with the rest of the data, Discontinuities, Trends, Periodicities, Time intervals containing the data of interest. These features, when considered in the context of the data, enable one to plan their analysis strategy.

2.2 Autocorrelation Plot

A correlogram checks whether a data set or time series is random or not. Random data should not exhibit any identifiable structure in the lag plot. A lag is a fixed time displacement. Correlogram can be generated for lags from 1 time unit to a maximum lag of approximately 10% of the record length. So the correlogram should be checked for different time lags to know the behavior of the rainfall data for higher lags.

2.3 System Identification Model

Different models have been constructed for the rainfall data of Agarwada Raingauge Station by system identification model. After estimating for different models, the model having best fit with the original data is taken into consideration. The model having best fit value is taken into consideration for validation.

The Fit (%) is the mean square error between the measured data and the simulated output of the model. 100% corresponds to a perfect fit (no error) and 0% to a model that is not capable of explaining any of the variation of the output and only the mean level.

Different models available for estimation purpose are explained as follows.

The general model introduced by Box and Jenkins (1976) includes autoregressive as well as moving average parameters, and explicitly includes differencing in the formulation of the model. Specifically, the three types of parameters in the model are: the autoregressive parameters (p), the number of differencing passes (d), and moving average parameters (q). In the notation introduced by Box and Jenkins, models are summarized as ARIMA (p, d, q); so, for example, a model described as (0, 1, 2) means that it contains 0 (zero) autoregressive (p) parameters and 2 moving average (q) parameters which were computed for the series after it was differenced once.

For a single-input/single-output system (SISO), the ARX model structure

is:

$$y(t)+a_1 y(t-1)+\ldots+a_{na}\,y(t-n_a)= b_1 u(t-n_k)+\ldots+b_{nb}u(t-n_k-n_b+1)+e(t)$$

y (t) represents the output at time *t*, *u (t)* represents the input at time *t*, *na* is the number of poles, *nb* is the number of zeros plus 1, *nk* is the input delay—the number of samples before the input affects the system output (called delay or dead time of the model), and *e (t)* is the white-noise disturbance).

One should specify the model orders *na*, *nb*, and *nk* to estimate ARX models. One can change the values of na, nb, and nk and try different models to get the best fit model.

Example: (3 3 2) Here na, nb and nk are 3, 3 and 2 respectively. Similarly one can try other values to get another model.

2.3.1 Process Model

Process models are popular for describing system dynamics in many industries and apply to various production environments. The primary advantages of these models are that they provide delay estimation, and the model coefficients have a physical interpretation.

One can create different model structures by varying the number of poles, adding an integrator, or adding or removing a time delay or a zero. One can specify a first-, second-, or third-order model, and the poles can be real or complex (underdamped modes). Thus models like P1DZ, P1DIZ, P2DZ, P2DIZ can be developed.

2.3.2 Non-Linear Models

Nonlinear ARX models describe nonlinear structures using a parallel combination of nonlinear and linear blocks. The nonlinear and linear functions are expressed in terms of variables called regressors.

The System identification model computes regressors by performing transformations of the measured input *u (t)* and output *y (t)* signals based on the model order that one specify. The predicted output of a nonlinear model at time *t* is given by the following general equation:

$$\hat{y}(t) = F(x(t))$$

where *x (t)* represents the regressors. *F* is a nonlinear regression function, which is approximated by a nonlinearity estimator.

After trying out different models to get best fit model, the best obtained model is then simulated for 30% of the data to check its validity. The best model obtained in such a way is called best fit model for that station which is further analysed by plot of poles and zeros and plot of model residual.

Residual analysis consists of the whiteness test. According to the whiteness test criteria, a good model has the residual autocorrelation function inside the confidence interval of the corresponding

estimates, indicating that the residuals are uncorrelated.

In the figure, the top axes show the autocorrelation of residuals for the output (whiteness test). The horizontal scale is the number of lags, which is the time difference (in samples) between the signals at which the correlation is estimated. Any fluctuations within the confidence interval are considered to be insignificant. A good model should have a residual autocorrelation function within the confidence interval, indicating that the residuals are uncorrelated.

Model information can be obtained from the model to check its FPE and Loss Function. Both the parameters should be nearer to zero. But the basis for selection of model in our study is the FIT value. Thus in this way by trying different models, one can develop model for different raingauge station to forecast and predict future values.

3. RESULTS

3.1 Time Series Plot

The daily rainfall data of various years have been collected from State Water Data Centre, Gandhinagar.

Time series plot of Agarwada station is plotted.

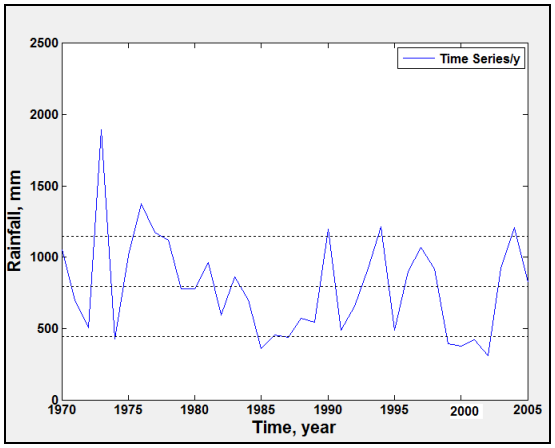
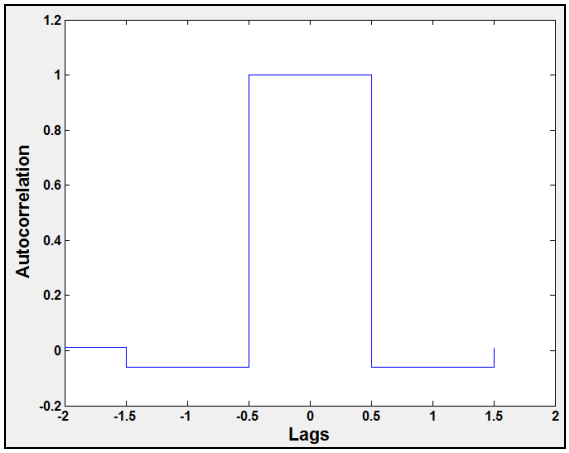
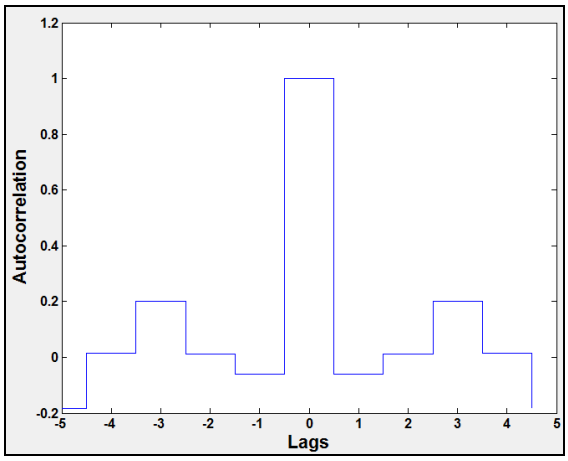


Fig 1: time series plot for agarwada raingauge station

3.2 Autocorrelation Plot

Correlogram summarizes characteristic features of the time series, viz. randomness, rising or declining trend, oscillation, etc. A correlation plot shows correlation coefficients on the vertical axis, and lag values on the horizontal axis. An example of Agarwada raingauge station is given below.





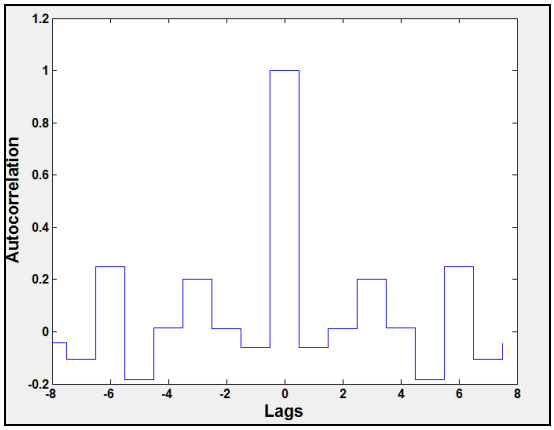


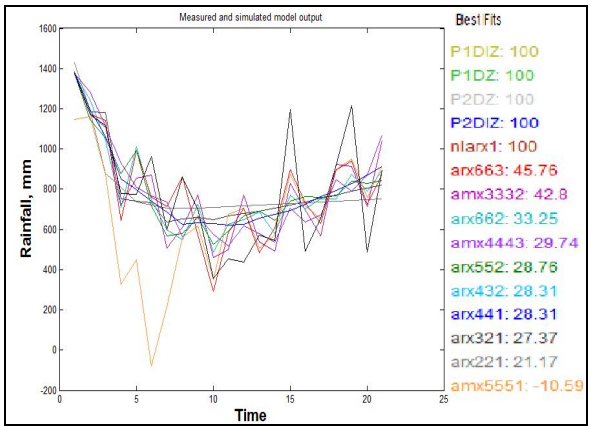
Fig 2: autocorrelation plot of agarwada station for lag 2, lag 5 and lag 8 respectively

.

3.3 System Identification Model

Fig.3. shows all the models estimated to get the best fit model for Agarwada Raingauge station. The models estimated on Agarwada Raingauge station are linear parametric models, process models, non-linear models and correlation models. Out of all these models the best fit model is applied to predict future values for Agarwada Raingauge station.

Fig.4. shows the best fit models obtained for Agarwada Raingauge station. But these models are further validated and checked for remaining 30% data of Agarwada Raingauge station. And final best model is obtained.



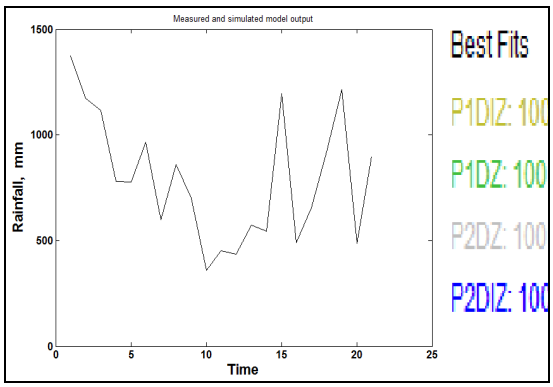


Fig. 3 & 4: different models estimated and best model obtained respectively for agarwada Station

Fig.5 shows the simulated data window. Here when the data was validated for the model P1DIZ i.e. one pole delay integration zero model and non-linear model, the FIT value was 100. This shows that this method is best applicable on the data set. From the graph it can be seen that process model and non-linear models are best suited for Agarwada station. For other models the data did not gave good results.

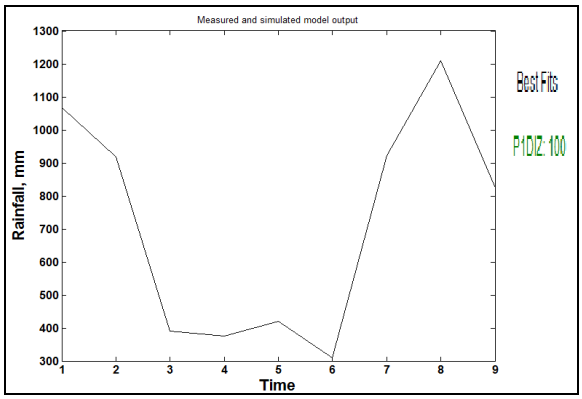
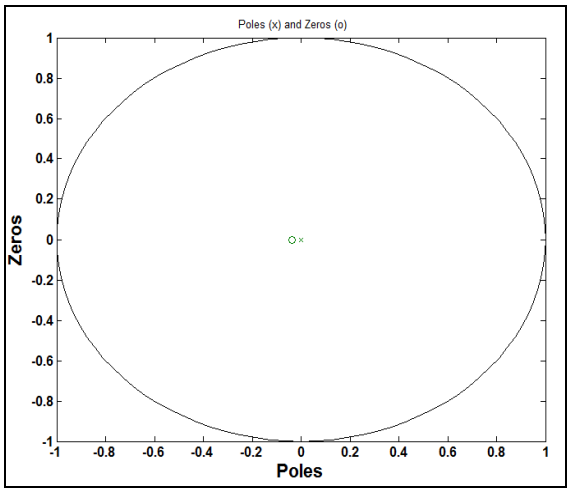


Fig.5: simulated model obtained for agarwada station

Further the result can be checked by examining poles and zeros and model residual method. Fig.6 shows the poles and zeros plot of the annual rainfall data of Agarwada Raingauge station. One can validate data by poles and zero plot. Poles are associated with the output side of the difference equation and zeros are associated with the input side of the equation.

Fig.7 shows the Residual analysis window which consists of two tests. The whiteness test and the independence test.



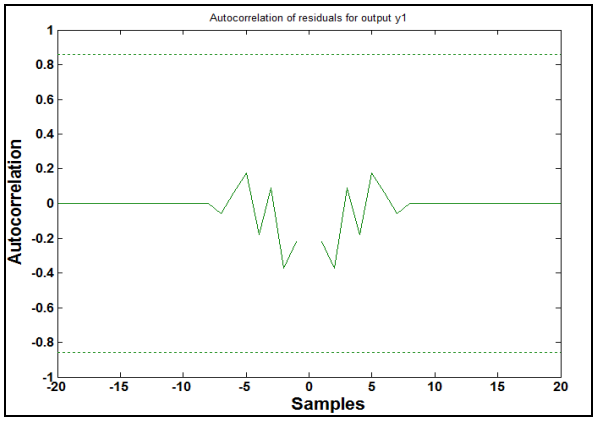


Fig.6 & 7: Poles and Zeros Plot and Model Residual Plot respectively for Agarwada Station

4. ANALYSIS

4.1 Time Series Plot

In the above time series plot, the centre line indicates the mean and the line below and above the mean indicates the standard deviation. Any data which is much deviated from the standard deviation line is known as erroneous data or outliers.

4.2 Autocorrelation Plot

Autocorrelation plot can be generated for lags from 1 time unit to a maximum lag of approximately 10% of the record length. Thus according to this criterion the annual data of 45 years is checked. From the above criteria one can make autocorrelation plot upto lag around 4.5. But here it is checked for lag 2, lag 5 and lag 8 .

From the above autocorrelation plot, we can analyze that the correlation of the time series data decreases as the time lag increases. Moreover, as the lag increases from lag 5, the data shows some cyclic variations in both directions. Thus, the above criterion holds good.

4.3 System Identification Model

In the above poles and zero plot the x indicate the pole and o indicates the zero. In the plot they are very close to each other which mean that the model applied is the best suited model for Agarwada station.

In the above model residual plot the values are within the confidence limit which means P1DIZ is the best suited model for Agarwada station. In the above plot there is no correlation between residual and input. Thus, residual analysis indicates that this model is best suited.

All the models like Process Model like P1DZ, P1DIZ, P2DZ and P2DIZ, Non-Linear Models, Auto-Regressive Models and Auto-Regressive Moving Average with different poles and delay have been tried under SIM for Agarwada Raingauge Station are shown below.

Table 1: All the Models Estimated on Agarwada Raingauge is shown below in the following table.

Sr. No.	Station	Model	Fit value
1.	Agarwada	P1DZ	100
		P1DIZ	100
		P2DZ	100
		P2DIZ	100
		nlarx	100
		arx663	45.76
		armax3332	42.8
		arx662	33.25
		armax4443	29.74
		arx552	28.76
		arx432	28.31
		arx441	28.31
		arx321	27.37
		arx221	21.17
		armax5551	-10.59

Table 2: Details of best fit model for Agarwada raingauge station.

STATION	MODEL	MODEL DETAILS
Agarwada	P1DIZ	Process model with transfer function 1+Tz*s G(s) = Kp * ---------- * exp(-Td*s) s(1+Tp1*s) with Kp = -14.825 Tp1 = 7.0739 Td = 30 Tz = 9596.2 An additive ARMA disturbance model ha y = G u + (C/D)e with C(s) = s + 6.713 D(s) = s + 3.635 Estimated using PEM using SearchMetho Loss function 2.19951e-024 and FPE 3.4

Final Prediction Error (FPE) criterion provides a measure of model quality by simulating the situation where the model is tested on a different data set. After computing several different models, one can compare them using this criterion. According to Akaike's theory, the most accurate model has the smallest FPE.

The Fit (%) is the mean square error between the measured data and the simulated output of the model. 100% corresponds to a perfect fit (no error) and 0% to a model that is not capable of explaining any of the variation of the output and only the mean level.

Loss function equals the determinant of the estimated covariance matrix of the input noise and its value should be equal to zero.

From the above models the best fitted model for Agarwada raingauge station is the model P1DIZ i.e. with one pole, derivative, integration, and zero as the FIT value comes out to be 100 for both 70% data and 30% data and value of FPE (Final Prediction Error) and Loss Function is 3.46e-024 and 2.19e-024 which is almost equal to zero. Thus the model obtained is best suited model for garwada rain gauge station.

5. CONCLUSIONS

It can be concluded from the time series plots that most of the data falls within the limits only i.e. only some of the rainfall data is deviated from the limits. Moreover, by studying rainfall pattern for past values we can analyse the trend and by making a suitable rainfall model we can predict the future values and can also find the missing data of that station and of nearby station also.

It can be concluded from autocorrelation plot that the criterion of choosing lag as 10% of the record length i.e. for 45 years it comes out to be 4.5 years is true. After increasing lag after lag 5, the correlation decreases. As we increase the lag, the correlation decreases and some cyclic variations can be seen in both the directions.

By looking at the results, the process models are the most accurate model having fit 100%. It can be concluded that process models with one pole, derivative, integration, and zero i.e. P1DIZ is the best model with FPE (Final Prediction Error) and Loss Function as 3.46e-024 and 2.19e-024 which is almost equal to zero. Thus they are best suited model for Agarwada Raingauge Station.

REFERENCE

[1] Bhatia, N.A. (2010) – "*Time Series Analysis and Development of System Identification Model for Panam Catchment Area*", A dissertation submitted to The Maharaja Sayajirao University of Baroda in partial fulfilment of degree of masters of Engineering (Civil) in Water Resources Engineering.

[2] De Silva, M.A.P. (2006) – "A time series model to predict the runoff ratio of catchments of the Kalu Ganga basin", *J.Natn.Sci.Foundation, Sri Lanka, 2006, 34 (2), Pp 103-105.*

[3] Schoellhamer, D.H. (2001) – "Singular Spectrum Analysis for Time Series with Missing Data", *Geophysical Research Letters, Vol. 1, No. 1, Pp 1-4*, 2001.

[4] Momani, P.E. (2009) – "Time Series Analysis Model for Rainfall Data in Jordan: Case Study for Using Time Series Analysis", *American Journal of Environmental Sciences 5 (5), Pp 599-604*, 2009.

[5] Houston, J.E., Adhikari, M., Paudel, L (2005) – "Broiler Water Demand: Forecasting with Structural and Time Series Models", *Proceedings of the Georgia Water Resources Conference*, April 2005.