STATISTICAL MODELING OF CLIMATE PARAMETERS

Sabita Madhvi Singh¹, P. R. Maiti¹, Sanwita Shaiwalini²

¹Department of Civil Engineering, IT- Banaras Hindu University, Varanasi, India
²Research Scholar, Department of Geography, Patna University, Patna, India

ABSTRACT

Climate variations more or less decide the environmental dynamics of an area. Extreme changes of climatic conditions cause problems in anywhere, especially to the agricultural sector. In this respect, knowledge of the likely climate and its impact could add value to Agro-environmental management. In this study, a forecasting approach that incorporates climatic variability is presented. Using meteorological data collected, the time series model: Auto Regressive Integrated Moving Average (ARIMA) is applied to estimate future variation in meteorological parameters. The parameter of each model is estimated with the time series module of software SPSS. The annual rainfall, temperature, relative humidity is forecasted with the observed data.

INTRODUCTION

The efficient management of water and other natural resources will likely be an increasingly significant issue in future years. Growing social pressure on the available resources requires the development of method for all phase of resources management; data collection, planning, development and management. Any decision regarding the planning or operation of water resources development and flood control projects requires the prediction of the characteristics and quantity of water available. While the historical hydrologic data would indicate the characteristics of the river flow. The river flow is treated as a random process. The appropriate word for this is “stochastic”. It justifies that river flow is a function of precipitation and other process, which at the present level of knowledge seem to evolve randomly in time and space. Even if the underlying phenomena and their interactions were thoroughly understood, it would not be able to describe mathematically the rate of discharge in a natural watercourse without involving unsystematic unknown effects. Non-availability of long-term sequence needs a mean by which sufficient data can be generated to overcome the problem caused by short term and which requires an adequate mathematical model. Since the river flow and other hydrological sequences are characterized by variability and oscillatory behavior this highlights the importance of studying time series, the properties of which are of great significance in planning, designing and operation of water resources systems.

Meteorological data are frequently needed while evaluating the long term effects of proposed man made hydrologic changes. Such evaluations are often undertaken by using mathematical models of hydrologic processes which use whether data as the input. The metrological variables required for most of the hydrologic models include precipitation, maximum and minimum temperature, solar radiation and relative variables. It is essential to have a sufficiently long-term historic data to achieve an operating mathematical rule.

Records, which consist of short-term data, are not suitable for proper planning and management processes and also these consist of only one time realization. Long-term climate data are often not available for a location of interest. In such circumstances, it becomes necessary to extrapolate the data of any other location. Most hydrologic systems have both deterministic as well as stochastic component, but stochastic time series models are widely used in regeneration of steam flow.

Regression techniques do not define the time-based structure of a set of data. Time series analysis is used to describe statistical properties, explain, and understand mechanism of variations and periodic future values of data arranged sequentially in time. Such an approach does not require assumptions as in case of regression analysis. Time series modeling of stream flow has mainly two uses in hydrology and water resources; viz, for generation of synthetic hydrologic time series and for forecasting future hydrologic series.

Forecast of hydrologic series are generally needed for short term planning of reservoir operation, for real time and short-term operations of river basins or systems for planning operations an ongoing drought and similar application. A time series is defined as a set of observations arranged chronologically i.e. a sequence of observations usually
ordered in time but may be ordered according to some other dimensions. The principal aim of a time series analysis is to describe the history of movements in time of some variable at a particular site. The objective is to generate data having properties of the observed historical record. To compute properties of a historical record, the historical record or time series is broken into separate components and analyzed individually to understand the casual mechanism of different components. Once properties of these components are understood, these can be generated with similar properties and combined together to give a generated future time series. Analysis of a continuously recorded rainfall data time series is performed by transforming the continuous series into a discrete time series of finite time interval. For such analysis daily, weekly, monthly and annual time series data are mostly used. Mathematical modeling of rainfall data as stochastic process is of interest for reservoir operation and irrigation planning. Several mathematical models based on the probability concept are available. These models help in knowing the probable weekly, monthly or annually rainfall but are not suitable for estimation of the total amount of expected rainfall. Over the past decade or so, a number of models have been developed to generate monthly rainfall and runoff for use in the design.

As Lovelock (1995) states: “there is no such thing as a perfect comprehensive model which is mathematically or spatially perfect; even if there was one it will be completely useless.” Further, Lenton (1998) observed that the modeling is a ‘mimicking’ exercise and should never be confused with reality. Models can be used as general guidelines to understand probable future conditions that may occur.

### 1. The need of Forecast

![Significance of meteorological forecasts for operational water management applications](image)

The broad term of time series analysis encompasses activities like
- Definition, classification, and description of time series
- Model building using collected time series data
- Forecasting or prediction of future values.

For forecasting the future values of a time series a wide spectrum of methods is available.

### 3. Objective of the presentation

- Potential application of time series to forecast the climatic parameters.
- To explore the trend and pattern of seasonal variation.
- Demonstration of performance of the model.

### 2. Methodology

Time series analysis deals with the problems of identification of basic characteristic features of time series, as well as with discovering - from the observation data on which the time series data are generated. In process control, the predicted time series data values help in deciding about the subsequent control actions to be taken.

### 2.1 Time Series Analysis

The main objectives of time series analysis are

- **Building of input-output models** that represent the equivalent transfer functions of processes behind the time series
- **Forecasting the future time series values** from the past values using the models developed
- **Control systems design**, based on the result of analysis.

Depending on the origin of the observation data, forecasting of future values of time series can also provide support in efficient process and production monitoring and failure diagnosis, in product quality inspection, etc., using the time-domain or frequency-domain approach. Once the time series model has been developed and tested it can be used for forecasting the future time series values at various time distances $d$. Of course, the forecasting does not deliver the exact future values of data that the given time series will really have, but rather their estimates. For example, using the autoregressive model

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \epsilon_t$$

(1)

based on a one-step movement along the time series
2.2.5 Estimation and Elimination of Trend and Demands

The two-steps ahead prediction, based on a two-steps movement along the time series, can be also formally written

\[ x_{t+2} = \Phi_1 x_{t+1} + \Phi_2 x_t + \varepsilon_{t+1} \]  
and the predicted value to be

\[ x_{t+2} = \Phi_1 x_{t+1} + \Phi_2 x_t \]  

2.2 Characteristic Features of time series

The major characteristic features of time series are the stationarity, linearity, trend, and seasonality. Although a time series can exhibit one or more of these features, for presentation, analysis, and prediction of time series values each feature is rather treated separately.

2.2.1 Stationarity

This property of a random process is related to the mean value and variance of observation data, both of which should be constant over time, and the covariance between the observations should only depend on the distance between the two observations and does not change over time.

2.2.2 Linearity

Linearity of a time series indicates that the shape of the time series depends on its state, so that the current state determines the local time series pattern. If a time series is linear, then it can be represented by a linear function of the present value and the past values. Examples of linear representations are the AR, MA, ARMA and ARIMA models, based on auto-regression and/or on a moving average technique. Nonlinear time series can be represented by the corresponding nonlinear or bilinear models.

2.2.3 Trend

The trend component of a time series is its long-term feature that is manifested through the local or global increase or decrease of data values as a consequence of superposition of true time series values and a disturbance with upward or downward trend. The presence of a disturbing component is detectable by pursuing the changes in the mean values in certain successive time intervals across the time series pattern. Trend analysis is important in time series forecasting.

2.2.4 Seasonality

The seasonality component of a time series is demonstrated through its periodically fluctuating pattern. This feature is more common in economic time series and in time series in which the observations are taken from real life, where the pattern may repeat hourly, daily, weekly, monthly, yearly, etc. Thus, the main objective of seasonal time series analysis is focused on the detection of the character of its periodical fluctuations and on their interpretation. In engineering, seasonal time series are found in the problems of power, gas, water, and other distribution systems, where the prediction of consumer demands represents the basic problem.

2.2.5 Estimation and Elimination of Trend and Seasonality

When two or more time series with different features are superimposed, or when a time series is superimposed by trend and/or seasonality component, decomposition analysis is needed to discriminate and separate individual components involved. More frequently, decomposition analysis is used for de-trending and deseasonalizing the time series data.

Anyhow, to make a proper forecast when a multi-component time series is given, it must first be identified to what extent the individual components are present in the time series data. This needs the decomposition of time series data to identify and extract the partial data superimposed to the main time series data. The time series decomposition process can be presented as shown in Figure.

[Figure 2.1. Time series decomposition process]

2.3 Forecasting Using the Box-Jenkins Method

Box and Jenkins have developed a general forecasting methodology for time series generated by a stationary autoregressive moving-average process. In the following, the methodology is explained on regressive models described previously.

2.3.1 Forecasting Using an ARIMA (autoregressive integrated moving average) Model

The forecasting approaches presented so far refer only to stationary models. In practice, however, many important time series are not stationary, so that they have to be transformed to stationary time series. For instance, the generalization of an ARMA model can be modified to provide a model for a time series that is nonstationary in the mean. The modified version of an ARMA is known as ARIMA (i.e. the autoregressive integrated moving average). The term integrated indicates the fact that the model is produced by repeated integrating or summing of the ARMA process. For example, by multiple summing the ARMA process we get the ARIMA model

\[ y_n = \sum_{i=1}^{p} a_i y_{n-i} + \mu + \sum_{j=1}^{q} \beta_i z_{n-j-1} \]  

for \( n \geq 0 \), where

\[ x_n = \sum_{i=1}^{s} y_i \]  

Using the last equation we can build,

\[ x_n - x_{n-1} = y_n \]  

which, after applying the \( z \)-transformation, results in,

\[ y(Z) = (1 - z^{-1}) x(Z) \]  

so that the \( z \)-transformed ARIMA model is

\[ a(z) - z^{-d} x(z) = \beta(z) z d + \mu \]  

Again, after \( d \)-successive integrations, the last equation is converted to

\[ a(z) (1 - z^{-d}) x(z) = \beta(z) z d + \mu \]  

This is the ARIMA(\( p, d, q \)) model with \( p \) and \( q \) as the degrees of polynomials \( a(z) \) and \( \beta(z) \) respectively. We now consider the ARIMA value

\[ \hat{y}(t) = \sum_{i=1}^{s} \beta_i z(t-i) \]  

and the prediction

\[ \hat{y}(t + k) = \sum_{i=1}^{s} \beta_i z(t+k-i) \]  

and build the prediction error
3. Statistical Characteristics of spatial data

3.1 Mean
The mean \( m \) is the average value of the data and is a measure of locality, i.e. the centre of mass of the histogram. With \( n \) the number data and \( z_i \) the value of the \( i \)th observation we have:

\[
m = \frac{1}{n} \sum_{i=1}^{n} z_i
\]

3.2 Variance
The variance \( s_z^2 \) is a measure of the spread of the data and is calculated as:

\[
s_z^2 = \frac{1}{n} \sum_{i=1}^{n} (z_i - m)^2 = \frac{1}{n} \sum_{i=1}^{n} z_i^2 - m^2
\]

The larger the variance the wider is the frequency distribution. For instance in Figure 4.3 two histograms are shown with the same mean value but with a different variance.

3.3 Standard deviation
The standard deviation is also a measure of spread and has the advantage that it has the same units as the original variable. It is calculated as the square-root of the variance:

\[
s_z = \sqrt{s_z^2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (z_i - m)^2
\]

3.4 Coefficient of variation
To obtain a measure of spread that is relative to the magnitude of the variable considered the coefficient of variation is often used:

\[
CV_z = \frac{s_z}{m_z}
\]

Note that this measure only makes sense for variables with strictly positive values (e.g. hydraulic conductivity, soil moisture content, discharge).

3.5 Skewness
The skewness of the frequency distribution tells us whether it is symmetrical around its central value or whether it is asymmetrical with a longer tail to the left (<0) or to the right (>0)

\[
CS_z = \frac{1}{n} \sum_{i=1}^{n} (z_i - m_z)^3
\]

3.6 Kurtosis
The kurtosis measures the “peakedness” of the frequency distribution (Figure 4.5) and is calculated from the data as:

\[
CC_z = \frac{1}{n} \sum_{i=1}^{n} (z_i - m_z)^4
\]

3.7 Mean Absolute Error:
The performance of the model, were evaluated by mean absolute error was computed by following equation. (Raghuvanshi et al., 2000).

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |X_i(t) - X_0(t)|
\]

3.8 Mean Relative Error:
The mean relative error was computed by following equation:

\[
MRE = \frac{1}{n} \sum_{i=1}^{n} \frac{|X_i(t) - X_0(t)|}{X_0(t)}
\]

3.9 Mean Square Error:
The mean square error was computed by following equation:

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} [X_i(t) - X_0(t)]^2
\]

3.10 Root Mean Square Error:
The root mean square error was computed by following equation:

\[
RMSE = \left( \frac{1}{n} \sum_{i=1}^{n} [X_i(t) - X_0(t)]^2 \right)^{1/2}
\]

3.11 Integral Square Error:
The integral square error (ISE) is a measure of goodness of a fit of time series model. The integral square error is the relation of root mean square error to the mean observed...
data. It was estimated by following equation Singh et al., (1991):

\[ MAF = \left( \sum_{i=1}^{n} \left[ X_i(t) - C_i(t) \right]^2 \right) / \sum_{i=1}^{n} X_i(t) \]

### 3.12 Application of Time series Analysis

Time series model is developed for each parameter i.e. monthly rainfall, annual rainfall, annual maximum temperature, minimum temperature and relative humidity etc. Then forecasted values are compared with the actual data. The forecasted data and the actual data (on an annual scale) are compared in Figure.

**Figure 5:** Observed and Forecasted annual rainfall (with past data) of Mirzapur district through ARIMA analysis.

**Figure 6:** Observed and Forecasted annual RH (with past data) of Mirzapur district through ARIMA analysis.

**Figure 7:** Autocorrelations (ACF) and Partial Autocorrelations (PACF) plot

### CONCLUDING REMARKS

The results show that the rainfall becomes more elastic to the inter-annual climatic perturbations over time. Furthermore, it was also shown that forecasted data of rainfall, temperature, relative humidity, wind speed, sunshine hours etc. Patterns of several seasonal rainfall and temperature patterns are identified. This issue was promptly addressed which details the further analysis that was performed using an ARIMA forecasting simulation model.

### REFERENCES