Research Article **AJSRE (2016), Vol. 1, Issue 01**

American Journal of Scientific Research and Essays (ISSN:2475-7527)

Time Series Analysis: A Hydrological Prospective

Sirisha Adamala

[1]Applied Engineering Department, Vignan's Foundation for Science, Technology and Research University (VFSTRU), Vadlamudi-522213, Guntur, Andhra Pradesh, India.

ABSTRACT

The analysis of time series is based on the assumption that successive values in the data file represent consecutive measurements taken at equally spaced time intervals. There are two main goals of time series analysis: (a) identifying the nature of the phenomenon represented by the sequence of observations, and (b) forecasting or predicting future values of the time series variable. Both of these goals require that the pattern of observed time series data is identified and more or less formally described. Once the pattern is established, one can interpret and integrate it with other data (i.e., Use it in the theory of the investigated phenomenon, e.g., Seasonal commodity prices). Regardless of the depth of one's understanding and the validity of our interpretation (theory) of the phenomenon, one can extrapolate the identified pattern to predict future events. This paper discusses about how to analyze time series data, what are its goals, types of time series data, and models available to analyze time series data.

Key words:

Time series, ARIMA, Box-Jenkins, rainfall, climate, hydrology.

***Correspondence to Author:**

Sirisha Adamala, Applied Engineering Department, Vignan's Foundation for Science, Technology and Research University (VFSTRU), Vadlamudi-522213, Guntur, Andhra Pradesh, India. E-mail: sirisha@agfe.iitkgp.ernet.in

How to cite this article:

Sirisha Adamala.Time Series Analysis: A Hydrological Prospective. American Journal of Scientific Research and Essays, 2016,1(1): 0031-0040.

eSciencePublisher

eSciPub LLC, Houston, TX USA. Website: http://escipub.com/

INTRODUCTION

How does someone measure the year's rainfall, snow melt, the amount of evaporation, water level or how much water is flowing down the river? Hydrologists and hydraulic engineers capture such data in time series. Time series data are used to describe many aspects of the hydrologic cycle. To a hydrologist or a hydraulic engineer, time series data are used to define the resources and conditions of a water basin. Hydrologists use time series data for displaying the amount of rainfall that has fallen within a catchment for the past day, year or 10 years. With this information, combined with additional time series data, hydrologists calculate the amount of runoff water and determine the total discharge of the catchment. Hydraulic engineers use measured time series of discharge and water level of a river for designing new river systems. They also use time series data for defining any input or withdrawal of water from the river.

Time series methods for analysis of hydrological data have a history of about half a century and continue to be an intense research topic. Time series have already played an important role early in the natural sciences. Babylonian astronomy used time series of the relative positions of stars and planets to predict astronomical events. Observations of the planets' movements formed the basis of the laws Johannes Kepler discovered. The analysis of time series helps to detect regularities in the observations of a variable and derive 'laws' from them, and/or exploit all information included in this variable to better predict future developments. The basic methodological idea behind these procedures, which were also valid for the Babylonians, is that it is possible to decompose time series into a finite number of independent but not directly observable components that develop regularly and can thus be calculated in advance. For this procedure, it is necessary that there are different independent factors which have an impact on the variable.

However, since the 1970's, a totally different approach has increasingly been applied to the statistical analysis of time series. The purely descriptive procedures of classical time series analysis were abandoned and, instead, results and methods of probability theory and mathematical statistics have been employed. This has led to a different assessment of the role of stochastic movements with respect to time series. Whereas the classical approach regards these movements as residuals without any significance for the structure of time series, the modern approach assumes that there are stochastic impacts on all components of a time series. Thus, the 'law of movement' of the whole time series is regarded as a stochastic process, and the time series to be analyzed is just one realization of the data generating process.

Evgenij Evgenievich Slutzky and the British statistician George Udny Yule at the beginning of the last century reviewed that time series with cyclical properties similar to economic (and other) time series can be generated by constructing weighted or unweighted sums or differences of pure random processes. The authors abandoned the idea of different components and assumed that there was a common stochastic model for the whole generation process of time series[2]. Firstly, this method identifies a specific model on the basis of certain statistical figures. Secondly, the parameters of this model are estimated. Thirdly, the specification of the model is checked by statistical tests. If specification errors become obvious, the specification has to be changed and the parameters have to be re-estimated. This procedure is re-iterated until it generates a model that satisfies the given criteria. This model can finally be used for forecasts.

Time series data contains several pieces of information that can be utilized by the user for various analytical reasons. The data are usually collected at regular intervals referred as the time step and stored as an integer in this data model with the time step property. This type of time series data is known as "measured time series data". The hydrologist might also vary the time step for recording more data during a flood or during the rainy season to obtain more accurate data for depicting the current conditions. On some occasions, and for various reasons, some time steps are either skipped or no data was collected for a time step or number of time steps. In this case, the values may sometimes be interpolated. Hydrologic models can also generate time series data. This type of time series is referred to as a "generated time series".

A time series is a chronological sequence of observations on a particular variable. Usually the observations are taken at regular intervals (days, months, and years), but the sampling could be irregular. It consists of two steps: (1) building a model that represents a time series, and (2) using the model to predict (forecast) future values.

Based on the theory of probability and stochastic processes and more recently complemented by advances on the study of chaotic nonlinear dynamical systems, time series analysis provides a repertoire of mathematical tools for the modeling of hydrological systems. Such tools have proved very effective and useful in numerous applications and case studies. The effectiveness of stochastic descriptions of hydrological processes may reflect the enormous complexity of the hydrological systems, which makes a purely deterministic description ineffective.

Stochastic approaches in analysis of observed time series have considered by many as ‘‘black box’’ approaches that do not help understanding of the system at hand. In this respect, they have been contrasted to deterministic approaches, which reveal the causative mechanisms of the natural processes. Such criticism of stochastic approaches may be valid in several cases, in which the focus was on the algorithmic details of the analyses. However, in their generality, stochastic approaches and tools are not ‘‘black boxes’’ and blind recipes.

The authors applied time series models and especially transfer function noise (TFN) models to analyze hydrological systems for many years[17]. In groundwater hydrology, the main applications of TFN modeling are decomposition of groundwater level fluctuations into natural and anthropogenic fluctuations and prediction of the effects of interventions. The authors applied modern time-series analysis, particularly complex demodulation to precipitation series in order to investigate temporal fluctuation in the decadal averages of the annual precipitation cycle in the British Isles and neighboring areas[14]. The authors applied time series analysis methodology and techniques in water resources management to explore the relationship between rainfall and stream flows[12].

The authors applied the Hinish tests[2] for Gaussianity and linearity to selected stationary segments of four kinds of such series, namely, stream flow, temperature, precipitation, and Palmer’s drought severity index to detect non-linearity in monthly hydrologic time series[4,5]. [7] studied the time series analysis of storm behavior in Pennsylvania water resources by quantifying the place of Karst aquifers in the groundwater to surface water continuum. Non-linear time variant analysis, such as wavelet transforms further defines rainfall-runoff relationships in Karst springs and may enable better prediction of input-output relations, where non-stationary behavior occurs. [6]developed an approach for river flood prediction using time series data mining (TSDM), which combines chaos theory and data mining to characterize and predict events in complex, non-periodic, and chaotic time series. They applied the TSDM methodology in the prediction of floods to the river discharge data at the St. Louis Gauging Station. [16] used time series analysis to develop an alternative method for reconstructing the natural flow regimes in the Murrumbidgee (Murray-Darling) river basin using pre-regulation climatic variables and river flow data. The method to reconstruct river flow time series illustrated in the study can be adopted to evaluate the effects of river management plans and policies.

Goals of Time Series Analysis

Time series analysis can be used to accomplish different goals as mentioned below:

1) Descriptive analysis determines what trends and patterns a time series has by plotting or using more complex techniques. The most basic approach is to graph the time series and look at the overall trends (increase/decrease), cyclic patterns (seasonal effects), outliers (points of data that may be erroneous), and turning points (different trends within a data series that changes over time and does not repeat or at least does not repeat with our range captured by our data).

2) Spectral analysis is carried out to describe how variation in a time series may be accounted for by cyclic components. This may also be referred to as “frequency domain”. With this an estimate of the spectrum over a range of frequencies can be obtained and periodic components in a noisy environment can be separated out.

3) Forecasting can do just that - if a time series has behaved a certain way in the past, the future behavior can be predicted within certain

confidence limits by building models.

4) Intervention analysis can explain if there is a certain event that occurs that changes a time series. This technique uses a lot of the time in planning experimental analysis. In other words, 'Is there a change in a time series before and after a certain event?'

5) Explanative analysis (Cross-correlation) using one or more variable time series, a mechanism that results in a dependent time series can be estimated. A common question to be answered by this analysis would be "What relationship is there between two time series data sets?"

Types of Time Series Data

Continuous vs. Discrete

Continuous - observations made continuously in time. *Examples*: Seawater level as measured by an automated sensor and Carbon dioxide output from an engine. Discrete - observations made only at certain times. *Examples*: Measurements of Rainfall, evapotranspiration, infiltration, solar radiation etc.

Stationary vs. Non-stationary

Stationary - Data that fluctuate around a constant value. The properties of time series do not change with time. For a stationary series p $(x; t_1)$ = p $(x; t_2)$, where t_1 and t_2 represent any two different possible times. Non-stationary - A series having parameters of the cycle (i.e., Length, amplitude or phase) changes over time. For a non-stationary series p $(x; t_1)$ not equal to p $(x; t_2)$.

Deterministic vs. Stochastic

Deterministic time series - This data can be predicted exactly as embedded in a built model. Stochastic time series - Data are only partly determined by past values and future values have to be described with a probability distribution.

Trend Analysis

Trends in a hydrologic time series can result from gradual natural or man-induced changes in the hydrologic environment producing the time series. The trend is a long term movement in a time series. It tells whether a particular data series have increased or decreased over the period of time. Detecting changes in climate and hydrologic time series are called as trend analysis. For detecting trend in time series data, one can use two types of tests: Parametric test and Non-Parametric test along with different confident tests.

Linear Regression Test

This is a parametric test that assumes that the data are normally distributed. It tests whether there is a linear trend by examining the relationship between time (*x*) and the variable of interest (*y*). The regression gradient is estimated by:

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad (1)$$

and the intercept is estimated as:

$$a = \overline{y} - b\overline{x} \quad (2)$$

The test statistic *S* is:

$$s = b / \sigma \quad (3)$$

Where,
$$\sigma = \sqrt{\frac{12\sum_{i=1}^{n}(y_i - a - bx_i)}{n(n-2)(n^2-1)}} \quad (4)$$

The test statistic S follows a Student-t distribution with *n*-2 degrees of freedom under the null hypothesis (critical test statistic values for various significance levels can be obtained from Student's t statistic tables). The linear regression test assumes that the data are normally distributed and that the errors (deviations from the trend) are independent and follows the same normal distribution with zero mean.

Mann-Kendall Test

This method tests whether there is a trend in the time series data. It is a non-parametric test. Then time series values (x_1, x_2, x_3,.., x_n) are replaced by their relative ranks (R_1, R_2, R_3, ….., R_n) (starting at 1 for the lowest up to n).

The test statistic S is:

$$S = \sum_{i=1}^{n-1}\left[\sum_{j=i+1}^{n}\text{sgn}(R_j - R_i)\right] \quad (5)$$

Where, sgn(x) = 1 for x > 0, sgn(x) = 0 for x = 0, and sgn(x) = -1 for x < 0

If the null hypothesis H_o is true, then S is approximately normally distributed with: μ = 0 and variance, σ = n (n – 1) (2n + 5) / 18

The z-statistic is therefore (critical test statistic values for various significance levels obtained from normal probability tables) as:

$$z = |S| / \sigma^{0.5} \tag{6}$$

A positive value of S indicates that there is an increasing trend and vice versa.

Autocorrelation Function (ACF)

It is a parametric test for randomness. The lag-one autocorrelation coefficient is calculated as:

$$r_1 = \frac{\left[\sum_{i=1}^{n-1}(x_i - \bar{x})(x_{i+1} - \bar{x})\right]}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]} \tag{7}$$

If the time series data come from a random process, then the expected value and variance of r_1 are:

$$E(r_1) = -1/n \text{ and } Var(r_1) = \left(n^3 - 3n^2 + 4\right) / \left[n^2\left(n^2 - 1\right)\right] \tag{8}$$

The z-statistic is therefore (critical test statistic values for various significance levels it can be obtained from normal probability tables):

$$z = \left|r_1 - E(r_1)\right| / Var(r_1)^{0.5} \tag{9}$$

Partial Autocorrelation Function (PACF)

Measure degree of association between Y_t and Y_{t-k}. PACF can be calculated by regression equation as below:

$$Y_t = b_0 + b_1 Y_{t-1} + \ldots + b_k Y_{t-k} \tag{10}$$

Where, critical value of PACF is (± 2/√ N).

Correlograms

The autocorrelation coefficient 'r_k' can then be plotted against the lag (k) to develop a correlogram. This will give us a visual look at a range of correlation coefficients at relevant time lags so that significant values may be seen.

Box-Jenkins Models

Box and Jenkins developed the Auto Regressive Integrative Moving Average (ARIMA) model which combined the Auto Regressive (AR) and Moving Average (MA) models developed earlier with a differencing factor that removes in trend in the data. This time series data can be expressed as: $Y_1, Y_2, Y_3, \ldots, Y_{t-1}, Y_t$ with random shocks (a) at each corresponding time: $a_1, a_2, a_3, \ldots, a_{t-1}, a_t$. In order to model a time series, we must state some assumptions about these 'shocks'. They have a mean of zero, constant variance, no covariance between shocks, and a normal distribution (although there are procedures for dealing with this).

Model Estimation and Testing

An ARIMA (p, d, q) model is composed of three elements: p: Auto regression, d: Integration or Differencing, and q: Moving Average. A simple ARIMA (0, 0, 0) model without any of the three processes above is written as: $Y_t = a_t$. Auto regression process [ARIMA (p,0,0)] refers to how important previous values are to the current one over time. A data value at t_1 may affect the data value of the series at t_2 and t_3. But the data value at t_1 will decrease on an exponential basis as time passes so that the effect will decrease to near zero. It should be pointed out that f_1 is constrained between -1 and 1 and as it becomes larger, the effects at all subsequent lags increase.

$$Y_t = f_1 Y_{t-1} + a_t \tag{11}$$

The integration process [ARIMA (0,d,0)] is differenced to remove the trend and drift of the data (i.e. makes non-stationary data stationary). The first observation is subtracted from the second and the second from the third and …. So the final form without AR or MA processes is the ARIMA (0,1,0) model:

$$Y_t = Y_{t-1} + a_t \tag{12}$$

The moving average process [ARIMA (0,0,q)] is used for serial correlated data. The process is composed of the current random shock and portions of the q previous shocks. An ARIMA (0,0,1) model is described as:

$$Y_t = a_t - f_1 a_{t-1} \quad (13)$$

Time Series Intervention Analysis

The basic question is "Has an event had an impact on a time series?" The null hypothesis is that the level of the series before the intervention (b_{pre}) is the same as the level of the series after the intervention:

$$(b_{post}) \text{ or } H_o: b_{pre} - b_{post} = 0 \quad (14)$$

After building the ARIMA model, an intervention term (I_t) can be added and the ARIMA equation is now a noise component:

$$(N_t): Y_t = f(I_t) + N_t \quad (15)$$

Models used for Time- Series Analysis

These models can be broadly divided into three groups: regression based methods, time series models and Auto Integrated (AI) based methods. Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the AR models, the *integrated* (I) models, and the MA models. These three classes depend linearly on previous data points. Combinations of these ideas produce Auto Regressive Moving Average (ARMA) and autoregressive integrated moving average (ARIMA) models. The Auto Regressive Fractionally Integrated Moving Average (ARFIMA) model generalizes the former three.

Among other types of non-linear time series models, there are models to represent the changes of variance along time (heteroskedasticity). These models are called Auto Rregressive Conditional Heteroskedasticity (ARCH). Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally-varying variability, where the variability might be modelled as being driven by a separate time-varying process, as in a doubly stochastic model.

Autoregressive Process

Most time series consist of elements that are serially dependent in the sense that one can estimate a coefficient or a set of coefficients that describe consecutive elements of the series from specific, time-lagged (previous) elements. This can be summarized in the equation:

$$x_t = \xi + \phi_1 * x_{(t-1)} + \phi_2 * x_{(t-2)} + \phi_3 * x_{(t-3)} + ... + \varepsilon \quad (16)$$

Where, ξ is a constant (intercept), and ϕ_1, ϕ_2, ϕ_3 are the autoregressive model parameters. Put in words, each observation is made up of a random error component (random shock, ε) and a linear combination of prior observations.

Moving Average Process

Independent from the autoregressive process, each element in the series can also be affected by the past error (or random shock) that cannot be accounted for by the autoregressive component, that is:

$$x_t = \mu + \varepsilon_t - \theta_1 * \varepsilon_{(t-1)} - \theta_2 * \varepsilon_{(t-2)} - \theta_3 * \varepsilon_{(t-3)} - ... \quad (17)$$

Where, μ is a constant, and $\theta_1, \theta_2, \theta_3$ are the moving average model parameters.

Put in words, each observation is made up of a random error component (random shock, ε) and a linear combination of prior random shocks.

Auto Regressive Moving Average (ARMA) Model

Auto Regressive Moving Average (ARMA) models were fitted to consider the stochastic nature of the stream flows. Traditional time series analysis uses Box-Jenkins ARMA models. An ARMA model predicts the value of the target variable as a linear function of lag values (this is the auto-regressive part) plus an effect from recent random shock values (this is the moving average part). While ARMA models are widely used, they are limited by the linear basis function.

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-n}) + e_t \quad (18)$$

For ARMA models proposed[2], it is assumed that the times series is stationary and follows the normal distribution. Since [3]presented significant developments in the form of ARMA models of the hydrologic times series, ARMA has been one of the most popular hydrologic times series models for reservoir design and optimization. ARMA is also applied to monthly hydrologic time series[9]. Extensive application and reviews of the several classes of such models proposed for the modeling of water resources time series were reported[5, 11, 13, 15, 1, 10].

Auto Regressive Integrated Moving Average (ARIMA)

The general model introduced[2] 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). The ARIMA method is appropriate only for a time series that is stationary (i.e., its mean, variance, and autocorrelation should be approximately constant through time) and it is recommended that there are at least 50 observations in the input data. It is also assumed that the values of the estimated parameters are constant throughout the series.

ArcGIS Hydro Data Model

Including time series data in the ArcGIS Hydro data model not only builds a complete hydrologic data model for use within the ArcInfo environment, but also it makes less distinct boundary between what is in a GIS database and what is used by a hydraulic model. Hydraulic models have been utilizing time series data for many years. The time series objects in the ArcGIS Hydro data model are designed for two purposes. One is the storage and delivery of a complete hydro data model to include time series data. Therefore, the storage and delivery of time series data within the Geo database is essential. The second intention is to provide access to this data structure for the hydraulic models so that the division between the GIS and the hydraulic model becomes less distinct.

River Analysis Package (RAP)

RAP (River Analysis Package) is a collection of three tools: Hydraulic Analysis (HA) - examines the hydraulic characteristics of river channels to determine the optimal discharge for a river reach based on specified rules, Time Series Analysis (TSA) - calculates summary statistics of time series data, including hydrological metrics, and Time Series Manager - manipulates and manages time series data. The TSA module has been designed to calculate summary metrics of daily discharge data, however it can handle other forms of time series data such as time series hydraulic data output from the HA module. The range of statistics calculated by the TSA module has been informed by a review of the literature, focusing on hydrological statistics used in ecological studies. The TSA module can present summary statistics based on the entire period of record, annually, seasonally, or monthly depending on the specific issue being investigated. The TSA module includes spell analysis, rates of hydrograph rise and fall, prediction of flood return interval (partial and annual series), baseflow, and seasonality. In addition to the numeric output, the TSA module has some neat visualization tools for plotting flow duration curves, flood frequency curves, and baseflow *vs* flood-flow. In addition to above, there is many more package tools are used for hydrologic studies. The hard of the decision is which one to use and it depends of the dataset i.e. Digital Terrain Model (DTM), Digital Elevation Model (DEM), Rugossity, grid spatial resolution, computer equipment, etc.

Analysis of Time-Series in Hydrology

Linear models and distributed variables are usually used in time series analysis mainly due to the convenience in studying relevant statistical properties. For models such as linear AR and ARMA, procedures for model identification and parameter estimation have been well formalized based on Gaussianity and linearity. However, non-linear mechanisms are often encountered in physical sciences. Non-Gaussian stationary time series may be generated as a result from a specific non-linear operation on a Gaussian input process. Therefore, non-linear modeling approaches have gained increasing attention from time series analysts. Although the rainfall-runoff process is widely perceived as being non-linear, the signatures of non-linearity are not all recognizable in hydrologic time series. By using Hinich test[8], non-linearity is detected in daily hydrologic time series, but not in annual series. Monthly hydrologic time series is seasonal with a cycle of 12 months. A standardization procedure is often applied to monthly hydrologic time series.

If a time series has a regular pattern, then a value of the series should be a function of previous values. If Y is the target value that we are trying to model and predict, and Y_t is the value of Y at time *t*, then the goal is to create a model of the form:

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-n}) + e_t \quad (19)$$

Where, Y_{t-1} is the value of Y for the previous observation, Y_{t-2} is the value two observations ago, etc., and e_t represents noise that does not follow a predictable pattern (this is called a 'random shock'). Values of variables occurring prior to the current observation are called 'lag values'. If a time series follows a repeating pattern, then the value of Y_t is usually highly correlated with $Y_{\text{t-cycle}}$, where cycle is the number of observations in the regular cycle. For example, monthly observations with an annual cycle often can be modeled by:

$$Y_t = f(Y_{t-12}) \quad (20)$$

The goal of building a time series model is the same as the goal for other types of predictive models, which is to create a model such that the error between the predicted value of the target variable and the actual value is as small as possible. The primary difference between time series models and other types of models is that lag values of the target variable are used as predictor variables, whereas traditional models use other variables as predictors, and the concept of a lag value doesn't apply because the observations don't represent a chronological sequence.

REFERENCES

1. Arena VC, Mazumdar S, Zborowski JV, Talbott EO, He S, Chuang YH, Schwerha JJ (2006). A retrospective investigation of PM10 in ambient air and cardiopulmonary hospital admissions in Allegheny County, Pennsylvania: 1995–2000. Journal of Occupational and Environmental Medicine, 48:38-47.
2. Box GEP, Jenkins, GM (1976). Time series analysis: Forecasting and control. Prentice Hall, Upper Saddle River, NJ. pp. 575.
3. Carlson RF, McCormick JA, Watts DG (1970). Application of linear random models to four annual streamflow series. Water Resources Research, 6(4):1070-1078.
4. Chen HL, Rao AR (2002). Testing hydrologic time series for stationarity. Journal of Hydrologic Engineering, 7(2):129-136.
5. Chen HL, Rao AR (2003). Linearity analysis on stationary segments of hydrologic time series. Journal of Hydrology, 277(1-2):89-99.
6. Damle C, Yalcin A. (2007). Flood prediction using time series data mining. Journal of Hydrology, 333:305-316.
7. Dryden I, Markus L, Taylor C, Kovacs J (2005). Non-stationary spatio-temporal analysis of karst water levels. Journal of the Royal Statistical Society Series C, Applied Statistics, 54:673-690.
8. Hinich MJ (1982). Testing for gaussianity and linearity of a stationary time series. Journal of Time Series Analysis, 3(3):169-176.
9. Hipel KW, McLeod AI (1994). Time series modelling of water resources and environmental systems, Developments in water science, 45. Elsevier, New York.
10. Komornik J, Komornikova M, Mesiar R, Szokeova D, Szolgay J (2006). Comparison of forecasting performance of nonlinear models of hydrological time series. Physics and Chemistry of the Earth 31(18):1127-1145.
11. Salas JD (1993). Analysis and modeling of hydrologic time series. In: Maidment, D.R., (Ed.), McGraw Hill Handbook of Hydrology, pp. 19.5-19.9.
12. Sankarasubramanian A, Vogel RM (2002). Annual hydroclimatology of the United States. Water Resources Research, 38(6):1083-1095.
13. Srikanthan R, McMahon TA (2001). Stochastic generation of annual, monthly and daily climate data: A review. Hydrology and Earth System Sciences, 5(4):653-670.
14. Thompson R (1999). A time-series analysis of the changing seasonality of precipitation in the British Isles and neighbouring areas. Journal of Hydrology, 224:169-183.
15. Toth E, Brath A, Montanari A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology, 239:132-

147.

16. Wang WC, Chau KW, Cheng CT, Qiu L (2009). A Comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Journal of Hydrology, 374:294-306.

17. Young PC, Jakeman AJ, Post DA (1997). Recent advances in the data-based modelling and analysis of hydrological systems. Water Science Technology, 36(5):99-116.

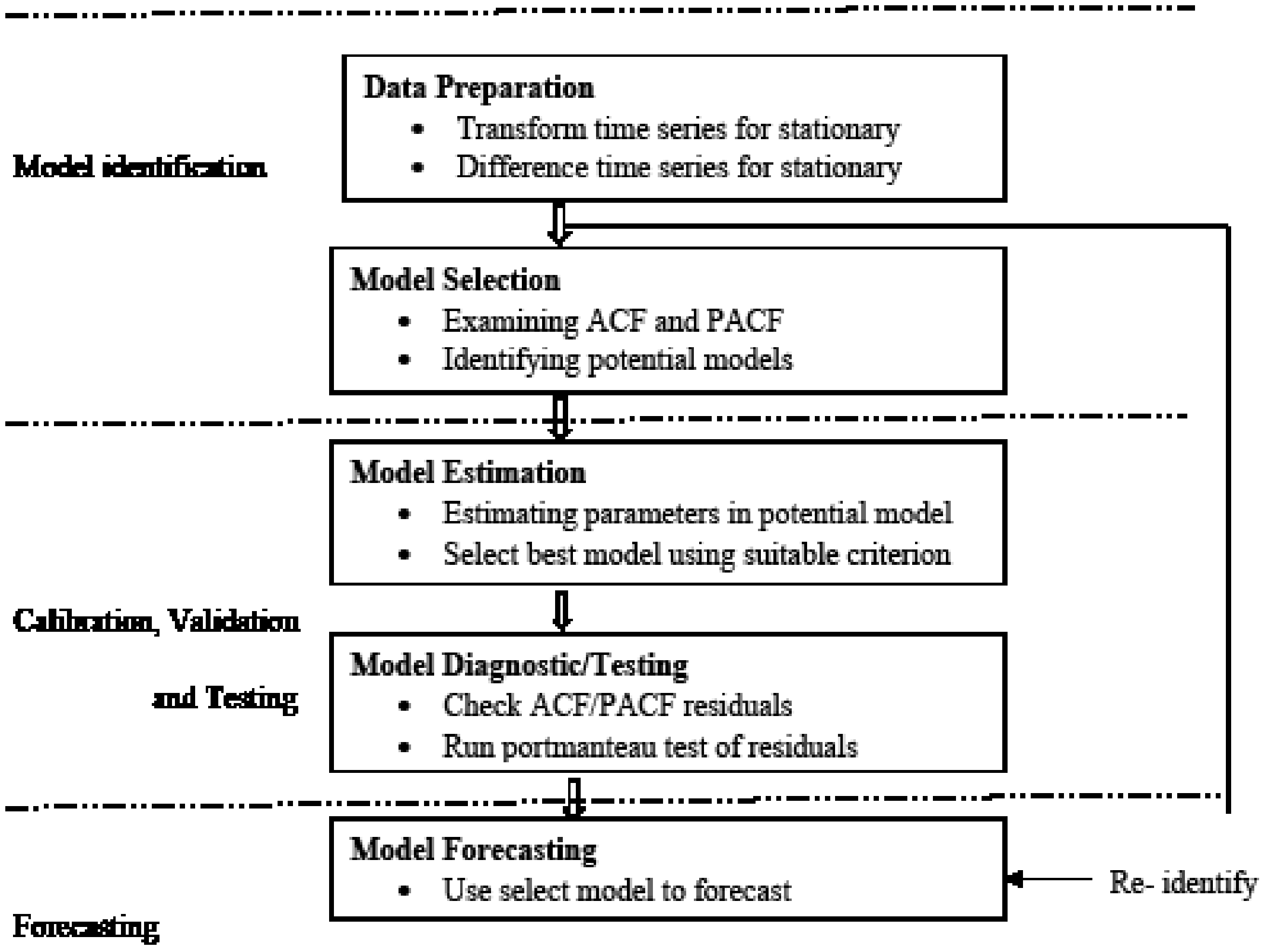

Figure 1. Flow chart of Box-Jenkins methodology.

ABBREVIATIONS

TFN	Transfer Function Noise
TSDM	Time Series Data Mining
PACF	Partial Autocorrelation Function
ACF	Autocorrelation Function
ARIMA	Auto Regressive Integrative Moving Average
AR	Auto Regressive
MA	Moving Average
AI	Auto Integrated
ARMA	Auto Regressive Moving Average
ARFIMA	Auto Regressive Fractionally Integrated Moving Average
ARCH	Auto Rregressive Conditional Heteroskedasticity
TSA	Time Series Analysis
HA	Hydraulic Analysis
DTM	Digital Terrain Model
DEM	Digital Elevation Model