Time series Modelling – 2

Stationary and Non-stationary Time Series

Stochastic modelling methods

Introduction

- A time series is known to be stationary if the statistical properties of the time series remain constant over time.
- This property is known as stationarity. The order of the stationarity represents the highest central moment (moment around the mean), which remain constant over time.
- For instance, first-order stationarity indicates timeinvariant mean or mean does not change over time.
- Similarly, if both mean and variance (second-order central moment) remain constant over time, the time series is known to be second-order stationary or weakly stationary.

- If mean, variance, and all higher-order moments are constant over time, the time series is called strict sense stationary or simply stationary.
- In hydroclimatic applications, second-order stationary can be safely assume to be satisfactory. However, impacts of climate change may impart non-stationarity in many hydrologic time series.
- If the statistical properties of a time series change or vary with time, it is known as non-stationary time series.
- Apart from various other causes, presence of trend, jump, periodicity, and a combination thereof, cause non-stationarity in the time series. **These are generally deterministic components that should be removed to obtain the stochastic component of the time series.**
- However, their removal does not always guarantee stationarity

Examples of non-stationary features

- Trend Detection and Removal

-- MK Test

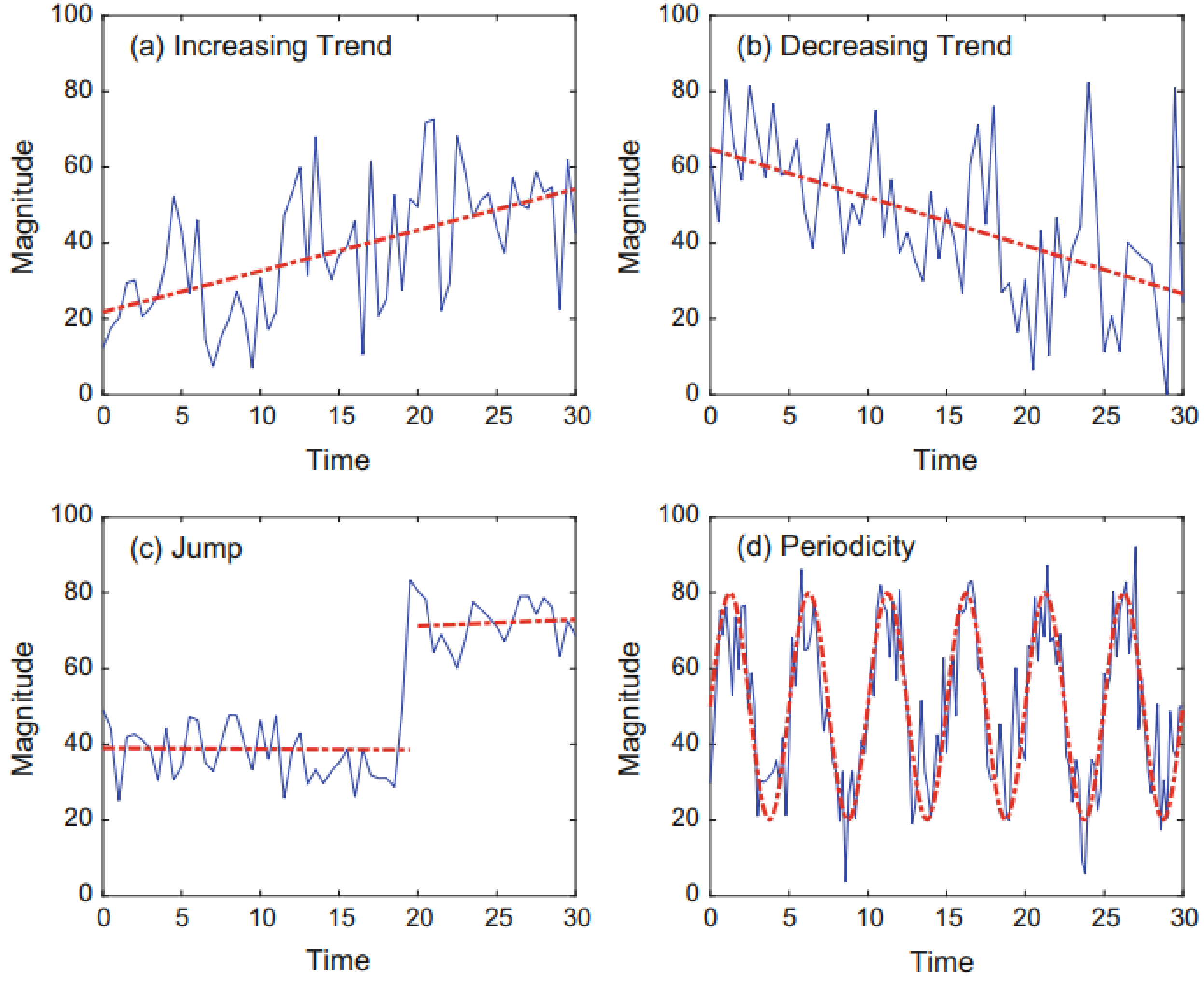
--Kendall tau test

--Sen’s Slope estimate

- Periodicity

- Fourier Analysis

- Wavelet Analysis



Time Series Modeling Climate data

- After removal of deterministic components (like trend, periodicity, or jump) of time series, different time series modeling approaches can be used for modeling stochastic component of the time series.
- Some of the popular linear models for time series prediction/forecast are following:

(i) Autoregressive model

(ii) Moving average model

(iii) Autoregressive moving average model

(iv) Autoregressive integrated moving average model

Linear Models for time series

- All of these models are linear regression model and try to relate the present value of time series with the previous values. Being linear, these models rely on mutual linear association between time series values.
- These linear associations are expressed in term of autocorrelation function and partial autocorrelation function in time series.

Measures of Linear Association in Time Series

- Climatic time series often have linear association between its successive values.
- These linear association can be utilized in developing the structure of the linear models for analysis/prediction of the time series.
- Two linear association measures for time series are autocorrelation and partial autocorrelation functions

Autocorrelation Function

- Autocorrelation is a measure of linear association between the values of same time series separated by some time lag/steps (say k)
- For a time series X(t), and the same time series with lag k (represented by X(t − k)), the linear association is measured by autocovariance.
- The term auto is used as the values are from same series but with some lags.
- The autocovariance function for lag k (represented by Ck) is given by

$$C_k = E(X(t), X(t-k))$$

- The autocorrelation function for lag k is defined as

$$\rho_k = \frac{C_k}{\sqrt{E\left[(X(t) - \overline{X}(t))^2\right] E\left[(X(t-k) - \overline{X}(t-k))^2\right]}} = \frac{C_k}{\sigma_t \sigma_{t-k}}$$

where σ_t and σ_{t-k} are standard deviation for time series X(t) and X(t − k), respectively

If the time series is second-order or higher-order stationary (standard deviation does not change over time), then for the rho_k can be expressed as

$$\rho_k = \frac{C_k}{\sigma^2}$$

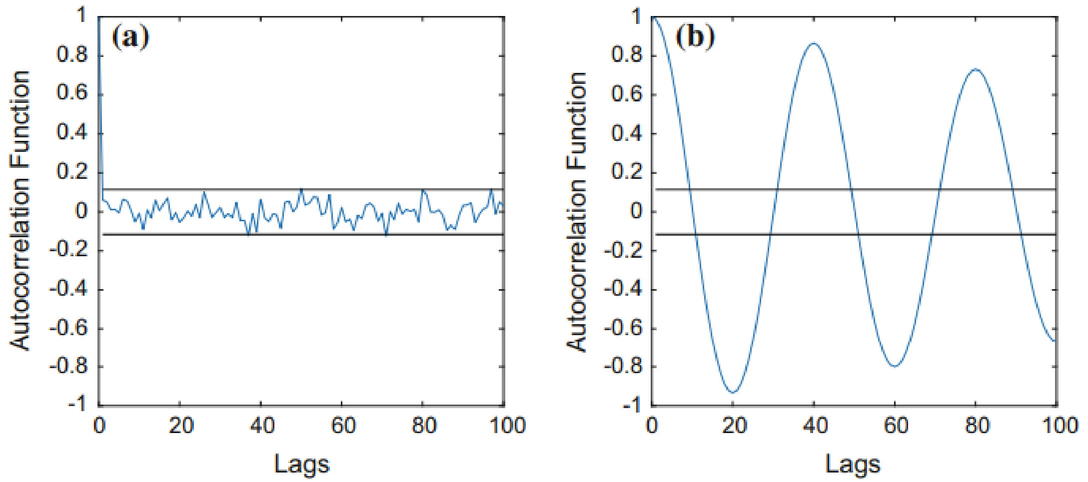
here σ is the standard deviation of time series X(t)

Autocorrelogram

- A plot of autocorrelation function with corresponding lag is called autocorrelogram. For a stationary time series the autocorrelation become insignificant with increasing lag. However, for a periodic time series the autocorrelation is also periodic and decreases slowly with damping peaks

The confidence limits of autocorrelation function for α significance level are given as follows,

$$\frac{-Z_{(\alpha/2)}}{\sqrt{N-k}} \leq \rho_k \leq \frac{Z_{(\alpha/2)}}{\sqrt{N-k}}$$



Typical autocorrelogram for a random/stationary time series b periodic time series

where Z(α/2) is standard normal variate at (1−α/2)×100% non-exceedance probability, i.e., P(Z > Z(α/2)) = α/ 2

Partial Autocorrelation Function (PACF)

- **Partial correlation** is the measure of linear association between two random variables when effect of other random variables is removed.

- For instance, let X, Y , and Z be three random variables. The partial correlation between X and Y , when the effect of Z is removed, represented as $\rho_{XY/Z}$, is the correlation between the residuals R_y and R_x resulting from linear regression of Y and X with Z, respectively. Hence, $\rho_{XY/Z}$ is expressed as

$$\rho_{XY/Z} = \frac{E(R_x R_y)}{\sqrt{Var(R_x)Var(R_y)}}$$

With an assumption that all involved variables are multivariate Gaussian distributed, if X is conditionally independent of Y given Z, then ρXY /Z is zero

- Partial autocorrelation function (PACF) of a time series X at lag k is defined as

$$\varphi_k = \rho_{X_0 X_k / \{X_1, X_2, \ldots, X_{k-1}\}}$$

Statistical Operators on Time Series

- Backward Shift Operator
- Backward shift operator or Backshift operator (represented as B(•)) returns the immediate previous value of time series. For a time series X(t), backshift operation is represented by:

$$BX(t) = X(t-1)$$
$$B^2X(t) = X(t-2)$$
$$B^nX(t) = X(t-n)$$

- Forward Shift Operator

Forward shift operator (represented as F(•)) returns the immediate next value of time series. It works opposite of backshift operator, and thus, also represented as B−1. For a time series X(t) it is represented by

$$FX(t) = B^{-1}X(t) = X(t+1)$$
$$F^2X(t) = B^{-2}X(t) = X(t+2)$$
$$F^nX(t) = B^{-n}X(t) = X(t+n)$$

- Difference Operator
- Difference operator returns the difference of the current and previous time step value in a time series. It is expressed as:

$$
\begin{aligned}
\nabla(X(t)) = (1 - B)X(t) \quad &= X(t) - X(t-1) \\
\nabla^2(X(t)) = (1 - B)^2 X(t) \quad &= (1 - 2B + B^2)X(t) \\
&= X(t) - 2X(t-1) + X(t-2) \\
\nabla^n X(t) = (1 - B)^n X(t) \quad &
\end{aligned}
$$

- Moving Average—Low Pass Filtering

Moving average (also known as rolling or running average) tries to reduce the shortterm fluctuations in time series by taking the average of the neighboring (say n) values of the time series.

Moving average works as a low pass filter and reduces the high-frequency oscillation in the time series. A ‘n’ term or n window Moving average is expressed as

$$Y(n) = \frac{1}{n}(X(1) + X(2) + \cdots + X(n))$$

$$Y(n+1) = \frac{1}{n}(X(2) + X(3) + \cdots + X(n+1))$$

In these equations, the moving average is assigned at the end of the window over which average is captured. Sometimes, the moving average values are assigned to the central value of the window selected.

- Differencing

Differencing is a high pass filtering method that removes low-frequency oscillation from the time series. The nth-order differencing is expressed as:

$$
\begin{aligned}
Y_1(t) &= X(t) - X(t-1) && \text{for } t = 2, 3, \ldots \\
Y_2(t) &= Y_1(t) - Y_1(t-1) && \text{for } t = 3, 4, \ldots \\
Y_n(t) &= Y_{n-1}(t) - Y_{n-1}(t-1) && \text{for } t = n+1, n+2, \ldots
\end{aligned}
$$

Auto-Regressive (AR) Model

- Autoregressive model tries to estimate the current value of time series using linear combination of weighted sum of previous values of the same time series.
- AR models are extensively used in hydroclimatic time series as current values of the time series are expected to be affected by the previous values.
- This characteristic of hydroclimatic variables is also referred as memory component.
- The number of lagged values being considered (say p) is called order of AR model

- pth-order AR model (AR(p))is given by

$$X(t) = \sum_{i=1}^{p} \Phi_i X(t-i) + \varepsilon(t)$$

where i(for i ∈ {1, 2,..., p}) are called autoregressive coefficients and ε(t) is uncorrelated identically distributed error with mean zero, also known as white noise.

Time series X(t) is obtained after removing the deterministic components like trend and periodicity. Using the backshift operator ARMA(p) can also be written as,

$$X(t) - \Phi_1 B(X(t)) - \Phi_2 B^2(X(t)) - \cdots - \Phi_p B^p(X(t)) = \varepsilon(t)$$

$$\text{or, } \Phi(B)X(t) = \varepsilon(t)$$

where $\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p$ for AR(p) model.

As an initial guess, the order p is decided from partial autocorrelation function.

Number of lags for which partial autocorrelation is significant is considered as p. Hence, for a AR(p) model, all partial autocorrelation with lag more than p should be zero and autocorrelation decays exponentially to zero

Different AR models are fitted using the slight variation in initial guess of AR order, the best model out of all fitted models is chosen on the basis of their parsimony

Following assumptions are made while developing an AR model.

$$E(\varepsilon(t)) = 0$$
$$E(\varepsilon(t)\varepsilon(t-k)) = E(\varepsilon(t)X(t-k)) = 0 \qquad \text{for } k = 1, 2, \ldots, p$$

How to obtain the model parameters of AR (p) model

- The parameters of a pth-order AR model are obtained by Yule–Walker equations.
- Yule–Walker equations are derived by taking expectation of p different equations obtained by multiplying lagged values of time series, i.e., X(t − 1), X(t − 2), ..., X(t − p) with the general form of AR model
- Yule–Walker equations are derived by taking expectation of p different equations obtained by multiplying lagged values of time series, i.e., X(t − 1), X(t − 2), ..., X(t − p) with the general form of AR model

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \vdots \\ \Phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{bmatrix}$$

where ρi is autocorrelation coefficient at lag i. It should be noted that ρ0 = 1, hence, the above Yule–Walker equation can also be written as:

$$\begin{bmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & \rho_0 & \rho_2 & \cdots & \rho_{p-2} \\ \rho_2 & \rho_1 & \rho_0 & \cdots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & \rho_0 \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \vdots \\ \Phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{bmatrix}$$

$$\text{or, } \sum_{i=1}^{p} \rho_{k-i} \Phi_i = \rho_k$$

It should be noted that p is the partial autocorrelation at lag p (φp)

At a location, the rainfall data is found to follow gamma distribution. For 20 consecutive days the recorded rainfall (in mm/day) are 2.89, 7.39, 23.88, 10.59, 5.91, 1.53, 3.48, 56.54, 26.19, 6.35, 38.09, 0.01, 3.03, 41.57, 44.73, 21.39, 15.87, 1.22, 21.75, and 0.21, respectively.

Calculate the autocorrelation at lags 0, 1, and 2. Calculate the 95% confidence limits for autocorrelation at lags 1 and 2.

Derive the nature of autocorrelation function for AR(1) and AR(2) models.

For time series X(t), the first-order autoregressive model AR(1) is given by,

From the Yule–Walker equation we can write,

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \vdots \\ \Phi_n \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_n \end{bmatrix}$$

For AR(1) model 2 = 3 =⋯= n = 0. Hence, the autocorrelation function is given by:

$$\begin{aligned} \rho_1 &= \Phi_1, \\ \rho_2 &= \Phi_1\rho_1 = \rho_1^2 \\ &\vdots \\ \rho_n &= \Phi_1\rho_{p-1} = \rho_1^p \end{aligned}$$

Hence, the autocorrelation function of AR(1) model decays exponentially to zero for positive values of ρ1. For negative values of the autoregressive coefficient, the autocorrelation function is damped and oscillates around zero.

- The second-order autoregressive model AR(2) has the form,

$$X(t) = \Phi_1 X(t-1) + \Phi_2 X(t-2) + \varepsilon(t)$$

From the Yule–Walker equations,

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \vdots \\ \Phi_n \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_n \end{bmatrix}$$

For AR(2) model $\Phi_3 = \Phi_4 = \cdots = \Phi_n = 0$. Hence, the autocorrelation function is given by:

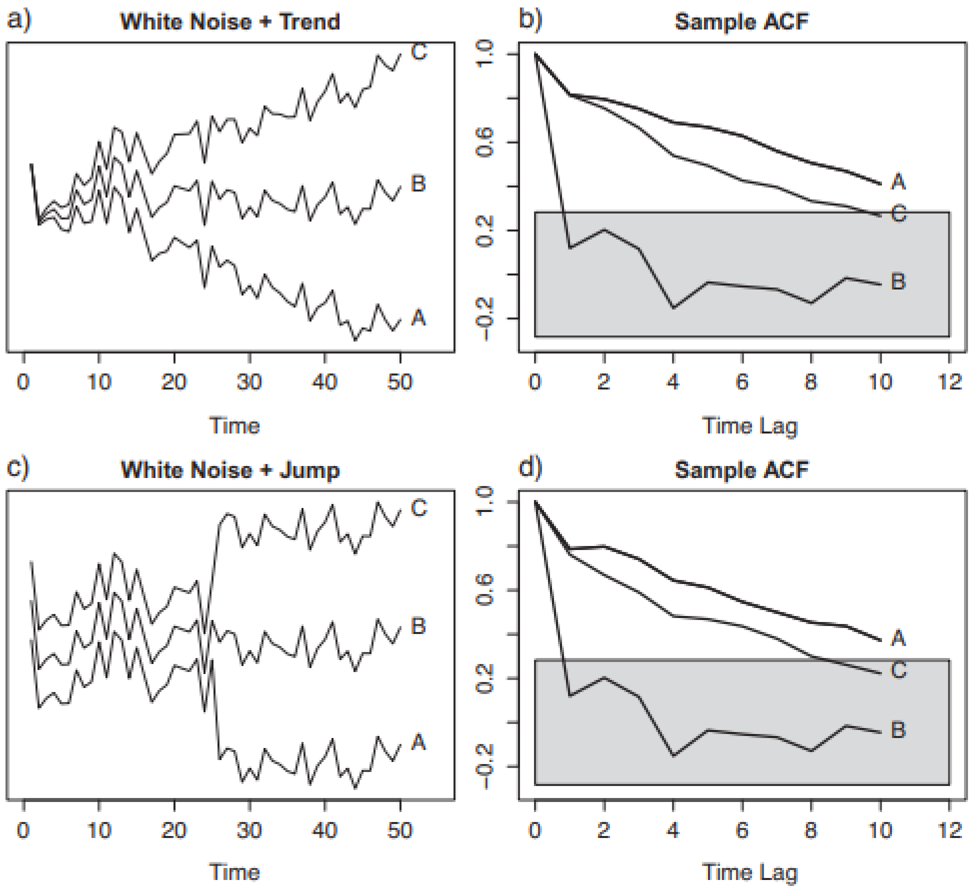
$$\begin{aligned}
\rho_1 &= \Phi_1 + \Phi_2\rho_1 \\
\rho_2 &= \Phi_1\rho_1 + \Phi_2 \\
\rho_3 &= \Phi_1\rho_2 + \Phi_2\rho_1 \\
&\vdots \\
\rho_n &= \Phi_1\rho_{n-1} + \Phi_2\rho_{n-2}
\end{aligned}$$

By solving the first two equations simultaneously, we get the following results,

$$\begin{aligned}
\Phi_1 &= \frac{\rho_1(1-\rho_2)}{(1-\rho_1^2)} \\
\Phi_2 &= \frac{(\rho_2-\rho_1^2)}{(1-\rho_1^2)}
\end{aligned}$$

For $k > 2$ the nature of autocorrelation function depends upon the values of Φ_1 and Φ_2. For instance, if $\Phi_1^2 + 4\Phi_2 \geq 0$ and $\Phi_1 > 0$ then the autocorrelation function decays exponentially to zero. However, if $\Phi_1^2 + 4\Phi_2 \geq 0$ and $\Phi_1 < 0$ then the autocorrelation function oscillates around zero. On the other hand, if $\Phi_1^2 + 4\Phi_2 < 0$ then the autocorrelation function is damped.

Pitfalls in Interpreting ACFs



Sample time series (left) and corresponding sample autocorrelation functions (right). The top row shows results for three time series comprising white noise plus a straight line (known as a “trend”), and the bottom row shows results for three time series comprising white noise plus a “jump” (a function that is constant for the first half and a different constant for the second half of the period).

The shaded region in the ACF plots shows the region for insignificance at the 5% level.

Moving Average (MA) Model

- In MA model, the current time series values are modeled using linear association with the lagged residual values. The MA model of order q considers q lagged residual for developing the model.

- In general, the qth-order moving average model is expressed as:

$$X(t) = \varepsilon(t) - \sum_{i=1}^{q} \theta_i \varepsilon(t-i)$$

where θi and ε(t − i) are the MA parameter and residual at lag i respectively. Time series X(t) is obtained after removing the deterministic components like trend and periodicity

$X(t) = \varepsilon(t) - \theta_1 B(\varepsilon(t)) - \theta_2 B_2\ (\varepsilon(t)) -\cdots- \theta_q\ B_q\ (\varepsilon(t))$

The order of the MA model is estimated on the basis of autocorrelation function.

Prove that MA(1) model is equivalent to AR(∞) model

A MA(1) model for a time series $X(t)$ can be expressed as:

$$
\begin{aligned}
\varepsilon(t) &= X(t) + \theta_1 \varepsilon(t-1) \\
&= X(t) + \theta_1 (X(t-1) + \theta_1 \varepsilon(t-2)) \\
&= X(t) + \theta_1 (X(t-1) + \theta_1 (X(t-3) + \theta_1 \varepsilon(t-3))) \\
&\quad \cdots \\
&= \sum_{i=0}^{\infty} \theta_1^i X(t-i)
\end{aligned}
$$

- Prove that AR(1) model is equivalent to MA(∞) model

$$
\begin{aligned}
X(t) &= \Phi_1 B(X(t)) + \varepsilon(t) \\
&= \Phi_1(\Phi_1 B^2(X(t)) + B(\varepsilon(t))) + \varepsilon(t) \\
&= \Phi_1(\Phi_1(\Phi_1 B^3(X(t)) + B^2(\varepsilon(t))) + B(\varepsilon(t))) + \varepsilon(t) \\
&\quad \cdots \\
&= \sum_{i=0}^{\infty} \Phi_1^i B^i(\varepsilon(t))
\end{aligned}
$$

Auto-Regressive Moving Average (ARMA)

- Auto-Regressive Moving Average (ARMA) Model Auto-Regressive Moving Average (ARMA) Model is a linear regression model in which current value of time series is estimated using lagged values of time series and the lagged values of residuals.
- ARMA model is a combination of autoregressive (AR) and moving average (MA) models. In general, ARMA model with pth-order AR model and qth-order MA model (also represented as ARMA(p, q)) is expressed as

$$X(t) = \sum_{i=1}^{p} \Phi_i X(t-i) + \varepsilon(t) - \sum_{i=1}^{q} \theta_i \varepsilon(t-i)$$

Selection of Order of ARMA Model

- Order selection based on ACF and PACF: The order of the autoregressive component (p) is decided (initial guess) by using PACF.
- For an AR model, if first p partial autocorrelation coefficients are significant at given level of significance and the autocorrelation function is exponentially decaying, then order is taken as p.
- Similarly, order of moving average component (q) depends upon the number of significant ACF of the time series. If the first q partial autocorrelation functions are significant and autocorrelation function is exponentially decreasing for a time series, then the order of MA model is taken as q.

It should be noted that this method can be used for initial guess for the order of ARMA model. One needs to generate different ARMA models considering the some variation in the guessed order. The final selection of the most appropriate model order is done based on parsimony of the developed model

A parsimonious model aims to utilize a minimum number of parameters and adequately reproduce the statistics with the least variance

Parameter Estimation of ARMA(p, q) Mode

- Parameters of ARMA(p, q) model (as expressed in Eq. 9.68) can be estimated either by principle of least square or maximum likelihood. These methods are discussed below.

Principle of least square: In this method, the Sum of squared residuals is minimized to get an estimate of ARMA model parameters. In terms of residual ARMA(p, q) is expressed as

$$\varepsilon(t) = X(t) - \sum_{i=1}^{p} \Phi_i X(t-i) + \sum_{i=1}^{q} \theta_i \varepsilon(t-i)$$

Autoregressive Moving Average Model with Exogenous Inputs (ARMAX)

- The models discussed above, i.e., AR, MA, ARMA, and ARIMA are developed using the information from the same time series, and these models do not consider any other variables/time series.

- However, in many cases in climatology, the time series under study (say precipitation) associated with other influencing time series (like air temperature, pressure), etc.
- Hence, for modeling these kind of interrelationships, the model should be able to use the information from the causal variable/time series known as exogenous input. A

- Autoregressive Moving Average Model with Exogenous Inputs (ARMAX) consists of an ARMA model and weighted sum of lagged values of exogenous time series.
- For an ARMAX model, if the r lagged value of exogenous time series is used and the ARMA part is of order (p, q), then the ARMAX model is said to be of the order of (p, q,r).
- In general, ARMAX model with order (p, q,r) is expressed as:

$$X(t) = \sum^{p} \Phi_i X(t-i) + \varepsilon(t) - \sum^{q} \theta_j \varepsilon(t-j) + \sum^{r} \psi_k I(t-k)$$

where $X(t)$ is stationary time series. ψ_k $(k = 1, 2, \ldots, r)$ is the weighting coefficients associated with lagged values of exogenous stationary time series $I(t)$. Φ_i $(i = 1, 2, \ldots, p)$ and θ_j $(j = 1, 2, \ldots, q)$ are autoregressive and moving average parameters, respectively.

Estimation of ARMAX Parameter

- The parameters of the ARMAX model are estimated by minimizing the sum of square of prediction errors.

$$
\begin{aligned}
\sum (\varepsilon(t))^2 &= \sum (X(t) - \hat{X}(t))^2 \\
&= \sum \left(X(t) - \sum_{i=1}^{p} \Phi_i X(t-i) + \sum_{j=1}^{q} \theta_j \varepsilon(t-j) - \sum_{k=1}^{r} \psi_k I(t-k) \right)^2
\end{aligned}
$$

Forecasting with ARMA/ARMAX

- Forecasting is the process of estimating future values of a time series, often using the past (or lagged) values of the same or other causal time series.
- Forecasting of hydroclimatic variables is important for making future plans/policies or preparedness for future extremes, if any.
- For instance, flood prediction system can be used for as early warning system and hence helps in evacuation.
- The procedure of forecasting can be used to estimate the past values of time series, this process is called hindcasting.

Forecasting with ARMA/ARMAX

- The forecast depends on the time step till which the information is being used (also known as origin of forecast).
- The difference in time step for which a forecast is made and the origin of forecast is called lead period.
- With the increase in lead period, the utility of forecast increases. However, the uncertainty in forecast also increases with increase in lead period.
- Hence, a suitable lead period can be used as compromise between two contrasting requirements.
- Further, a forecasting model can be static or dynamic. For static forecasting model the parameters once estimated do not change with time. However, for dynamic forecasting model the parameters change with time.

Analysis of Forecast Errors

- Mean Square Error

$$\mathrm{MSE} = \frac{1}{N}\sum_{l=1}^{N}\left(X(t+l) - \hat{X}_t(t+l)\right)^2$$

Mean Absolute Error

$$\mathrm{MAE} = \frac{1}{N}\sum_{l=1}^{N}\left|X(t+l) - \hat{X}_t(t+l)\right|$$

- The monthly average atmospheric pressure (in mb) measured at surface level for 24 consecutive months are 963.65, 965.03, 961.18, 959.43, 957.68, 953.42, 950.11, 952.44, 952.25, 956.88, 963.66, 963.36, 965.56, 964.5, 963.66, 960.91, 956.9, 952.18, 950.71, 952.54, 951.43, 955.06, 959.01, and 962.60.
- Find the autocorrelation and partial autocorrelation functions at lags 0, 1, 2, and 3..

- Autocorrelation function at lag 0, 1, 2, and 3 are 1, 0.782, 0.414, and 0.008 respectively. Partial autocorrelation function at lag 0, 1, 2, and 3 are 1, 0.807, −0.617, and −0.481 respectively