Engineering Hydrology

Week-11

CHAPTER-5 STOCHASTIC HYDROLOGY

Sisay Demeku Derib (PhD, PE)
Sisay.demeku@aastu.edu.et
Addis Ababa Science and Technology University (AASTU),
Department of Civil Engineering
May, 2025

Lecture contents of the last weeks (Week-10)

CHAPTER-5 STOCHASTIC HYDROLOGY

5.1 Introduction.

5.2 Time Series

5.3 Analysis of Hydrologic Time Series

5.4 Time Series Synthesis

5.5 Some Stochastic Models

Home assignment:

Based on the last week part of this chapter lecture, I hope you are sure you area able to exercise on:

- Discuss on analysis of hydrologic time series
- Describe time series synthesis
- List some stochastic models

Lecture Learning Outcomes

Course Learning Outcomes: After completion of this Lecture, you will be able to:

CLO-1: Apply measurement techniques of the components of the hydrologic cycle, water balance and filling of missed data;

CLO-2: Examine rainfall-runoff relationship and hydrograph;

- Apply flood routing

CLO-3: Examine the probability of occurrence;

- Analyze of hydrologic time series
- Compare some stochastic models

;

CLO-4: Analyze the water movement in to, over, and through the soil surface;

CLO-5: Design capacity of reservoir;

CLO-6: Design runoff volume and time of distribution of the runoff hydrograph from urbanization effect.

Lecture contents of the week (Week-11)

CHAPTER-5 STOCHASTIC HYDROLOGY

On this second part of the chapter, we will cover details of:

5.3 Analysis of Hydrologic Time Series

5.5 Stochastic Models comparison

Revision: Time series

- A time series is a sequence of data points recorded over a period of **time**, usually at regular intervals (hourly, daily, monthly, annually).

Characteristics and Concepts:

- **Order**: It is **ordered** by time, making it distinct from other types of data where the **order is not significant**.
- **Interval**: It is often collected at **regular** intervals.
- **Pattern/trend**: Time series can exhibit patterns like trends, seasonality, and cycles, which are important for analysis and forecasting.
- **Application**: Time series analysis is used in diverse fields like finance, weather forecasting, healthcare, and marketing.

Time Series

Examples of time series data:

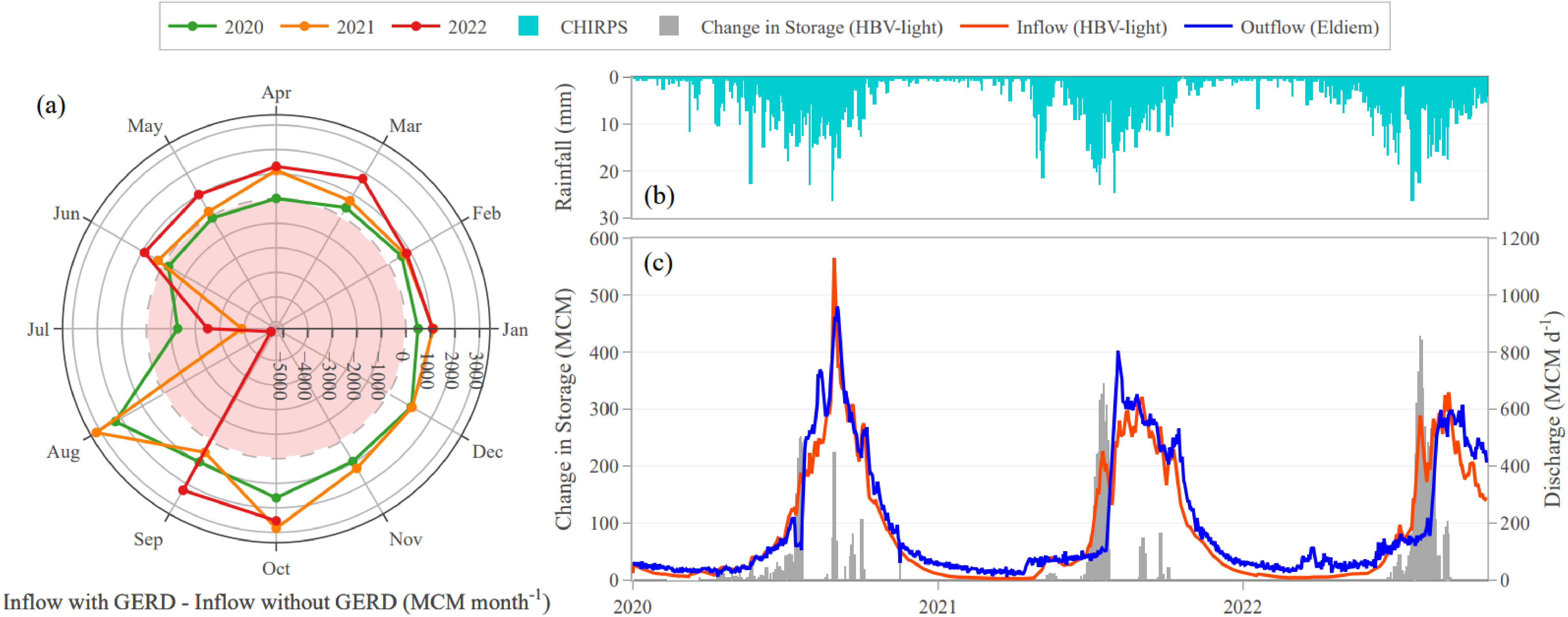
- Hydrograph, hyetograph,
- Hourly temperature readings from a weather station.
- 5 minutes rainfall data from AASTU weather kit
- Daily stage measurements of a river gage station

Others:

- Daily stock prices of a company.
- Monthly sales data of a product.
- Patient heart rate over time during a medical procedure.

Time Series

Examples of time series data: GERD filling on inflow to Sudan [1]



Analysis of hydrologic time series

- Hydrologic time series analysis involves examining patterns and trends in hydrological data over time, such as:
  - Precipitation, river flow, temperature, hydrographs over time.
- It helps to understand the variability and behavior of hydrological processes,
- It is crucial for **water resource** management, **flood** and **drought** prediction, and understanding the impact of **climate change**.
- It needs historical data collection, data cleaning like:
  - Address missing values, outliers, and inconsistencies in the data to ensure accuracy and reliability of the time series data before analysis.

Analysis of hydrologic time series: **Trends**

Detecting and analyzing **trends in hydrological** time series, e.g., :

- increasing or decreasing precipitation patterns for understanding long-term climate change impacts.
- to identify trends in river flow, potentially indicating changes in water availability or flood frequency.
- Examining changes in groundwater levels over time to detect trends related to groundwater depletion or recharge.
- changes in evaporation rates, which can be influenced by temperature and other factors, to assess their impact on water balance.
- Analyzing temperature trends, often alongside rainfall trends, to assess the influence of climate change on hydrological processes.
- Analyzing historical flood records to determine if the frequency or magnitude of floods has changed over time, potentially due to climate change or other factors.

Analysis of hydrologic time series: **Trends**

Tools and Techniques trends in hydrological time series, e.g., :

1. **Non-parametric** tests like the **Mann-Kendall test** are commonly used to detect monotonic trends in these time series data.
   - Parametric tests assume specific distributions (like normality) and use parameters like the mean with larger samples,
   - Non-parametric tests don't make those assumptions and can be used with various data types, including nominal and ordinal with smaller samples.

Analysis of hydrologic time series: **Trends**

Mann-Kendall test statistic, ***S***, is calculated using given equation ,

$$s = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(x_j - x_i)$$

where,

- n is the number of data points.
- X_j and X_i are the data value in time series i and j (j >i), respectively, and,
- $\operatorname{sgn}(X_j - X_i)$ is the sign function as: .

$$\operatorname{sgn}(x_j - x_i) = \begin{cases} +1, & \text{If } x_j - x_i > 0 \\ 0, & \text{if } x_j - x_i = 0 \\ -1, & \end{cases}$$

Source: (Kamal, Neel and S. Pachauri, 2018)

2. Sen's Slope Estimator:

- Is used to estimate the magnitude of the trend in time series data.
- is a robust nonparametric method for estimating the slope of a linear trend in time series data
- the estimation of the slope of a trend in the sample of N pairs of data:
- Calculates slopes for each pair of X_j and X_k, where j > k, calculate the slope (Q_i) using the formula: $Q_i = (X_j - X_k) / (j - k)$.

2. Sen's Slope Estimator:

- Q_i are ranked from the least to the highest and the median slope or Sen's slope estimator was calculated as:

$$Q_{med} = \begin{cases} Q_{\left[\frac{N+1}{2}\right]}, & \text{if N is odd} \\ \dfrac{Q_{\left[\frac{n}{2}\right]} + Q_{\left[\frac{N+2}{2}\right]}}{2}, & \text{if N is even} \end{cases}$$

- A positive Sen's slope indicates an increasing trend, while a negative slope indicates a decreasing trend.

Source: Frimpong et al. (2022)

Analysis of hydrologic time series: **Trends**

3. Detrended Fluctuation Analysis:

- A method for analyzing the long-range dependence of time series data.
- It involves partitioning a time series into windows and quantifying the fluctuations within each window, taking into account local trends.

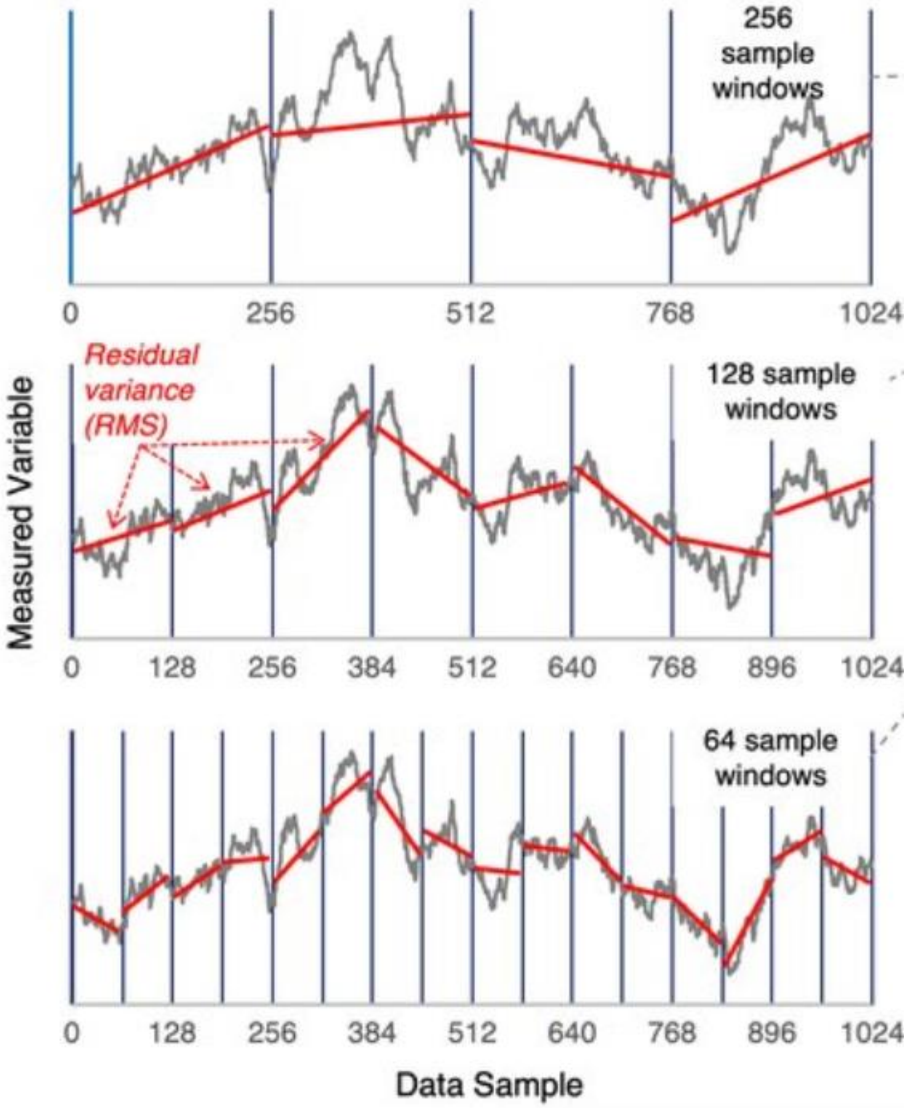


Illustration of Detrended Fluctuation Analysis (DFA) (Source: Rigoli et al (2020)

Analysis of hydrologic time series: **Trends**

Statistical Packages:

- like **R** or **Python**, with libraries like ***statsmodels*** or ***pandas***, are used for performing trend analysis and data visualization.
- R boasts a wide array of specialized packages for time series analysis, including forecast, tseries, and xts, which offer advanced statistical models and functionalities.
- R's ggplot2 package provides excellent data visualization

Analysis of hydrologic time series: **Trends**

Statistical Packages:

- Python has a large and active community with many libraries and frameworks,
    - including pandas, statsmodels, scikit-learn, and Prophet, which are widely used for time series analysis.
- Python has a **more readable** syntax and a smoother learning curve compared to R, making it easier for beginners to pick up.

Covariance and correlation

Covariance:

- Covariance measures how much **two variables vary together**.
  - A tendency to increase or decrease when the other variable increases or decreases
- A **positive covariance** means that as one variable increases, the other also tends to increase, and as one decreases, the other tends to decrease.
- A negative covariance means that as one variable increases, the other tends to decrease, and vice versa.

Covariance and correlation

Covariance:

- Is calculated using a formula that involves the expected value of the **product of** the **deviations** of each variable from their respective **means**.
  - $Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$/N, where X and Y are the random variables, μ_X and μ_Y are their means, and E represents the expected value.

Covariance and correlation

Example 1: Calculate the covariance of X and Y:

X	Y	$X - \mu_X$	$Y - \mu_Y$	$[X - \mu X) * (Y - \mu Y)$
1	2	-1.5	-4.5	6.75
2	5	-0.5	-1.5	0.75
3	8	0.5	1.5	0.75
4	11	1.5	4.5	6.75
				15

Step 1: Calculate the means of X and Y :

- Mean of X (μX) = (1+2+3+4)/4 = 2.5
- Mean of Y (μY) = (2+5+8+11)/4 = 6.5

Step 2: Calculate the deviations:

- (X - μX) = {1-2.5, 2-2.5, 3-2.5, 4-2.5} = {-1.5, -0.5, 0.5, 1.5}
- (Y - μY) = {2-6.5, 5-6.5, 8-6.5, 11-6.5} = {-4.5, -1.5, 1.5, 4.5}

Step 3: Multiply the deviations and sum them:

- ∑[(X - μX) * (Y - μY)] = (-1.5 * -4.5) + (-0.5 * -1.5) + (0.5 * 1.5) + (1.5 * 4.5) = 6.75 + 0.75 + 0.75 + 6.75 = 15

Step 4: Divide by the sample size minus 1 (for sample covariance) or the sample size (for population covariance):

- Sample covariance (Cov(X,Y)) = 15 / (4 - 1) = 5
- Population covariance (Cov(X,Y)) = 15 / 4 = 3.75

Covariance and correlation

Covariance: Example 1: Calculate the covariance of X and Y:

X	Y	X - µX	Y - µY	[X - µX) * (Y - µY)
1	2	-1.5	-4.5	6.75
2	5	-0.5	-1.5	0.75
3	8	0.5	1.5	0.75
4	11	1.5	4.5	6.75
				15

- Sample covariance (Cov(X,Y))
- = 15 / (4 - 1) = 5
- Population covariance (Cov(X,Y))
- = 15 / 4 = 3.75

Interpreting the covariance

Consider two variables: X: Rainfall (in mm) and Y: Humidity (in %).

If we find a positive covariance between X and Y,

- it suggests that high rainfall amount tend to be the larger the humidity of the air, and vise versa.
- A covariance of zero would suggest no linear relationship between rainfall and humidity.

Covariance and correlation

Correlation:

- quantifies both the direction and the strength of the linear relationship between two variables.
  - A correlation of +1 indicates a perfect positive linear relationship.
  - A correlation of -1 indicates a perfect negative linear relationship.
  - A correlation of 0 indicates no linear relationship.
- Covariance alone doesn't tell us how strong the relationship is, while correlation does, giving us a standardized measure of strength and direction.

Covariance and correlation

Example 2: Calculate the correlation of X and Y:

X	Y	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$	r
1	2	2.25	20.25	6.75	
2	5	0.25	2.25	0.75	
3	8	0.25	2.25	0.75	
4	11	2.25	20.25	6.75	
2.5	6.5	**5**	**45**	**15**	**0.07**

Step 1: Calculate the means of X and Y :

- Mean of X (μX) = (1+2+3+4)/4 = 2.5
- Mean of Y (μY) = (2+5+8+11)/4 = 6.5

Step 2. Calculate the sum of squared deviations:

- $\Sigma(x - \bar{x})^2$ and $\Sigma(y - \bar{y})^2$

Step 3: Calculate the sum of cross-products:

- $\Sigma(x - \bar{x})(y - \bar{y})$

Step4: Calculate the correlation coefficient:

- $r = \Sigma(x - \bar{x})(y - \bar{y}) / \sqrt{[\Sigma(x - \bar{x})^2 * \Sigma(y - \bar{y})^2]}$
- r= 15/(5*45)= 0.07

Interpretations:

- A correlation coefficient of near 1, suggests a very strong positive relationship, while -1 indicates a very strong negative relationship.
- A value of 0, like on this example, indicates no correlation.

Covariance and correlation

Covariance:

- Is calculated using a formula that involves the expected value of the **product of** the **deviations** of each variable from their respective **means**.
  - $Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$, where X and Y are the random variables, μ_X and μ_Y are their means, and E represents the expected value.

Autocovariance and autocorrelation

- are both statistical measures that quantify the relationship between a **time series** and its **lagged versions**.
- Autocovariance measures the covariance between a **time series** at **different points in time**, while
- Autocorrelation is a standardized measure of autocovariance, making it easier to compare across **different series** and interpret.

Autocovariance and autocorrelation

Autocovariance:

- Measures the covariance of a **time series** with itself at different **lags (time intervals)**.
- It quantifies how much the values of a time series at **one point** in time are correlated with its values at **past or future time** points.

Autocovariance and autocorrelation

Autocorrelation:

- Autocorrelation is a **normalized version** of autocovariance, typically expressed as a coefficient between -1 and 1.
- It is denoted as $\rho(k)$, $\rho(k) = \gamma(k) / \gamma(0)$, where $\gamma(0)$ is the variance of the time series.
- Autocorrelation is widely used in time series analysis for **forecasting**, **modeling**, and identifying **patterns** in time series data.

Autocovariance and autocorrelation

Interpretation:

- A high **positive autocorrelation** (close to +1) means that **past** values tend to **predict future** values.
- A high **negative autocorrelation** (close to -1) means that past values tend to be **negatively correlated with future** values.
- A **low autocorrelation** (close to 0) means that past values are **not strongly related** to future values.

Some stochastic models

- Models may be broadly classified, based on three different criteria used to generate synthetic sequences:
  1. Based on No. of variables: **univariate** models or **multivariate** models
  2. Based on the time interval: [**annual** model or **seasonal** model].
  3. Based on the choice of model: [autoregressive (**AR**), moving average (**MA**), autoregressive moving average (**ARMA**) models, autoregressive integrated moving average models (**ARIMA**), fractional Gaussian noise (**fGn**) models, etc]
- Are models used in time series analysis to forecast future values based on past observations

Some stochastic models

AR (Autoregressive) Model:

- predicts future values based on the linear combination of **past values** of the same time series.
- It assumes that the current value of a time series is **linearly dependent** on its previous values.
- Example:
  - An AR(1) model might predict tomorrow’s climate based on today’s climate.
  - An AR(2) model would use today's and yesterday’s climate to predict tomorrow’s climate.
- The 'p' in AR(p) represents the **order of the model**, indicating the number of past time steps to be considered in the prediction.

MA (Moving Average) Model:

- predicts future values based on the **linear combination of past errors** (differences between the observed values and their predictions).
- It's characterized by its reliance on **lagged error terms** (**residuals**) rather than past values of the series itself, unlike Autoregressive (AR) models.
- The 'q' in MA(q) represents the **order** of the model, indicating the **number of past errors** to be considered.

Some stochastic models

ARMA (Autoregressive Moving Average) Model:

- combines both **AR and MA** components.
- It predicts future values based on both **past values** of the time series and **past errors**.
- An ARMA model is denoted as ARMA(p, q), where 'p' represents the order of the AR component and 'q' represents the order of the MA component.

Some stochastic models

ARIMA (Autoregressive Integrated Moving Average) Model:

- extends the ARMA model by including an **integration** (differencing) component denoted by 'd'.
- It combines autoregressive (**AR**), integrated (**I**), and moving average (**MA**) components to model the relationship between a variable's current value and its past values, errors, and lagged forecast errors.
- **Differencing** involves subtracting consecutive observations to remove trends and seasonality, making the series stationary (meaning its statistical properties like mean and variance are constant over time).
- An ARIMA model is denoted as $\boldsymbol{ARIMA(p,d,q)}$, where 'p' represents the order of the AR component, 'd' represents the order of differencing, and 'q' represents the order of the MA component.

Home assignment:

Make sure that you can :

- Analysis of hydrologic time series
- Compare some stochastic models

References

Ali, A.M. Lieke A. M., and Adriaan J. T. 2023. Inferring reservoir filling strategies under limited-data-availability conditions using hydrological modeling and Earth observations: the case of the Grand Ethiopian Renaissance Dam (GERD), Hydrology and Earth System Sciences (HESS), 27(21): 4057–4086

Kamal, Neel and S. Pachauri. 2018. Mann-Kendall Test - A Novel Approach for Statistical Trend Analysis, International Journal of Computer Trends and Technology (IJCTT), 63(1): 18-21. Mann-Kendall Test - A Novel Approach for Statistical Trend Analysis

Frimpong, B.F., A. Koranteng, and F. Molkenthin. 2022. Analysis of temperature variability utilizing Mann–Kendall and Sen's slope estimator tests in the Accra and Kumasi Metropolises in Ghana, Environmental Systems Research, 11(24):1-13. https://doi.org/10.1186/s40068-022-00269-1

Rigoli, L. M., T. Lorenz, C. Coey, R. Kallen, S. Jordan & M. J. Richardson. 2020. Co-actors Exhibit Similarity in Their Structure of Behavioural Variation That Remains Stable Across Range of Naturalistic Activities, Scientific Reports, 10(6308) | https://doi.org/10.1038/s41598-020-63056-x

Thank you very much for your active attendance!!

Prepared By: Sisay Demeku Derib (PhD, Practicing Hydraulics Engineer, Assistant Prof.)
Affiliated to: Addis Ababa Science and Technology University –AASTU, Department of Civil Engineering
----
Scopus WOS ORCID ResearchGate Google Scholar LinkedIn YouTube