Engineering Hydrology

Week-10

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
April, 2025

Lecture contents of the last weeks (Week-8-9)

CHAPTER-4 FREQUENCY ANALYSIS

Week-8

4.2 Flow Frequency

4.3 Flood Probability

4.4 Methods of Peak flood determination

4.5 Regional Frequency Analysis

Week-9

4.6 Low Flow Analysis

4.6.1 Definitions and Basic Concepts

4.6.2 Low Flow Frequency Analysis

4.6.3 Drought Analysis

4.7 Precipitation Probability

4.8 Risk, Reliability and Safety Factor

Home assignment:

I hope you are sure you area able to exercise on:

- Define flood frequency, probability and recurrence interval
- List methods of flood frequency analysis
- Analyze design flood using Gumbel's and log-pearson methods
- Describe regional frequency and low flow analysis
- Explain risk, reliability and safety factor

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;

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

Introduction.

- **Be remember the word statistics from your previous course.**

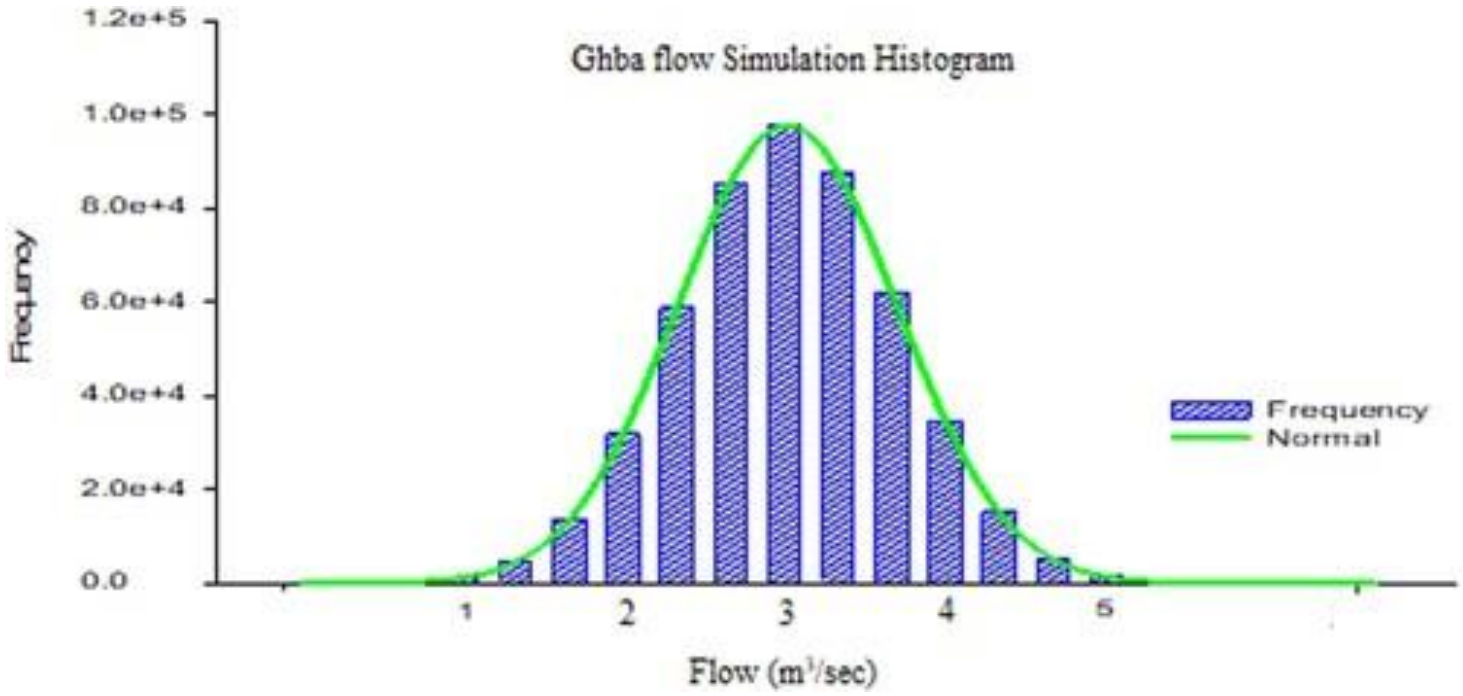
The word “Stochastic” :

- derived from the Greek word “Stochasticos” or “Stochazesthai”
- Means to **shoot** (an arrow), to guess, to imaginate, to think deeply, to contemplate, to cogitate and to mediate **at a target.** [1]
- Stochastic methods aim at predicting the value of some variable:
  - at non-observed **times** or at non-observed **locations**, while
  - also stating **how uncertain** we are when making these predictions
    - Bierkens, & van Geer (2007)

Introduction

Statistics of data distribution:

- Let's look in to the graph of Ghba flow simulation [3] (Ethiopia) (Gebreyohannes et al. 2024)
- The horizontal axis represents the flow rate, ranging from 0 to 5 m³/sec, while
- pattern (a **normal distribution**).



Source: Gebreyohannes et al. (2024)

- the vertical axis indicates the frequency of each flow rate occurrence, up to 1.2e+5.
- The peak of both the histogram and the normal distribution curve is around 3 m³/sec, the most frequently occurring flow rate.
- The distribution is symmetrical, indicating a typical bell-curve pattern (a **normal distribution**).

Introduction.

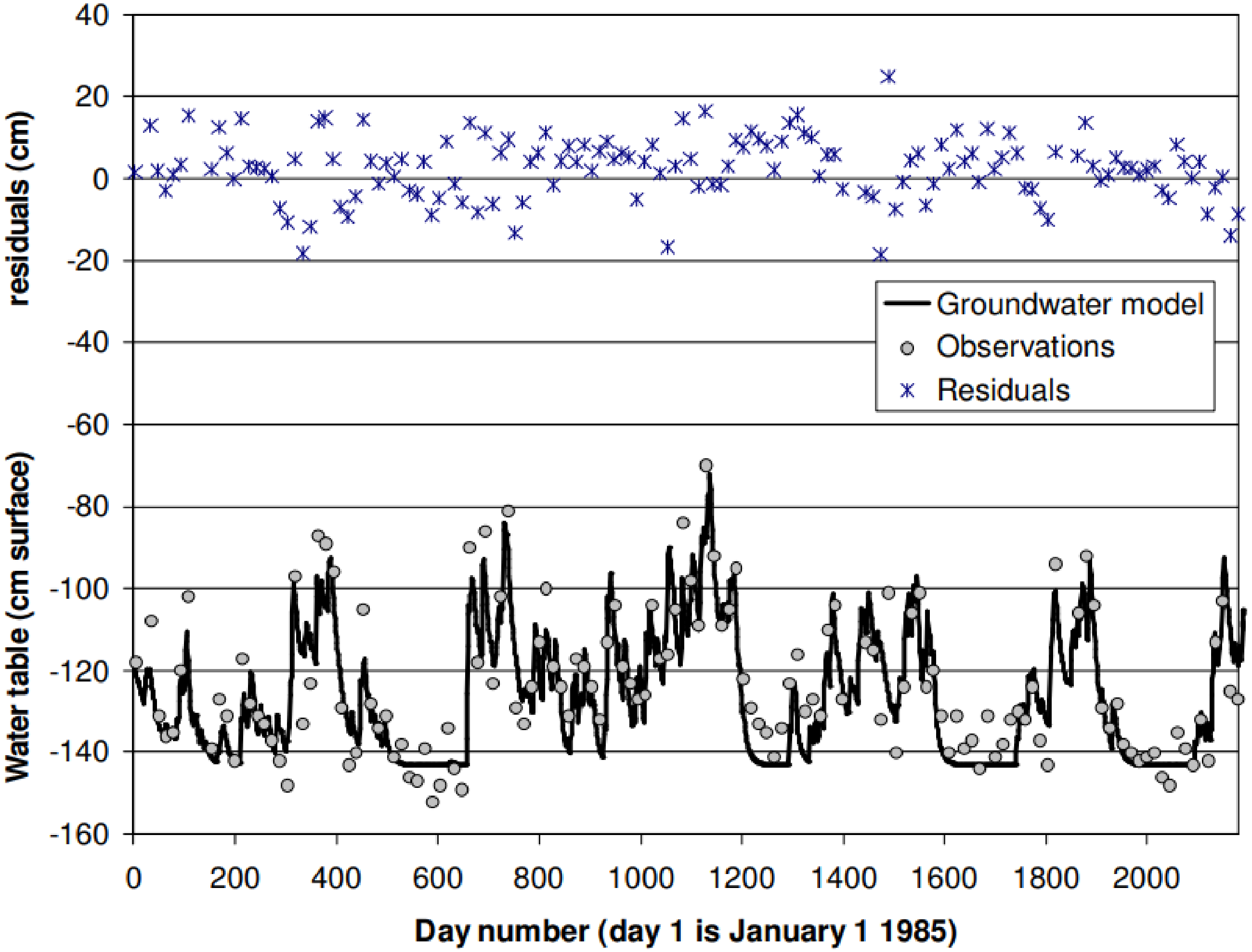
Uncertainty associated with predictions:

- Observation errors
- Errors in **boundary** and **initial** conditions and input
- Unknown **heterogeneity** and **parameters**
- Scale discrepancy
- Model or system errors

Introduction.

Figure: Observed water table depths and water table depths predicted with a groundwater model at the same location (Bottom).

- **Residuals:** the differences between model outcome and observations (Top).



(Bierkens & van Geer, 2007).

Introduction.

Expected values:

- In statistics, "E" often refers to the **expected value** or **mathematical expectation** of a random variable in statistics.
- It represents the long-run average value of that variable, weighted by its probability distribution.
- The expected value of a random variable X, is denoted as $E(X)$.

Discrete and continuous random variables

Discrete random variables:

- A random variable that takes a finite or countable values (eg. No. cars, peoples, …)
- Frequency is the number of occurrences of a specific event.
- Relative frequency is resulting from dividing frequency by the total number of events. e.g. if n = No. of years having exactly 50 rainy days;
  - Let n=10 years and N=100 years.
  - Then, the frequency of having exactly 50 rainy days is 10 and
  - The relative frequency of having exactly 50 rainy days in 100 years is $n/N = 0.1$.

Discrete and continuous random variables

Discrete random variables:

- E(X) is calculated as the sum of each possible value of the variable multiplied by its corresponding probability.
  - If X is a random variable with possible values x_1, x_2, … , x_n and corresponding probabilities $P(x_1), P(x_2), \dots, P(x_n)$, then $E(_X) = x_1 P(x_1) + x_2 P(x_2) + \dots + x_n * P(x_n)$.
- Probably Density Function (PDF):
  - the must satisfy $\sum_{xi} P(x) = 1$ and $1 \geq p(x) \geq 0$ for all values of x.
- Cumulative Distribution Function
  - $\Pr(X \leq x_o) = \sum p(x_i)$

Discrete and continuous random variables

Continuous random variables:

- A random variable that takes a infinite or any values within the given range (eg. Measurements for ***height, temperature, time*** …)
- Frequency needs to be defined for a **class interval**.
- A plot of frequency or relative frequency versus class intervals is called **histogram** or **probability polygon**.
- As the number of sample gets infinitely large and class interval length approaches to zero, the histogram will become a smooth curve, called **probability density function** (PDF).

Discrete and continuous random variables

Continuous random variables:

- The expected value, $E(X)$, is the integral of the variable multiplied by its probability density function over the entire range of possible values. .
  - If X is a continuous random variable with probability density function $f(x)$, then $E(X) = \int x * f(x)\, dx$, where the integral is taken over all possible values of x.
- Probably Density Function (PDF):
  - the must satisfy $\int_{-\infty}^{\infty} P(x)dx = 1$ and $f(x) \geq 0$ for all values of x.
- Cumulative Distribution Function
  - $\Pr(X \leq x_o) = \int_{-\infty}^{x_o} P(x)dx$

Descriptors and Moments

- Descriptors are used to show statistical properties of a sample of the population.
  - are tools used to summarize and describe the main features of a dataset.
- They indicates sample /population measures of:

(1) **Central tendency** – describe the center or typical value of a dataset (**mean, median, mode**).

(2) **Dispersion** – how spread out or dispersed the data from the central point (**range, variance, standard deviation**).

(3) **Asymmetry or shape** – describes how evenly distributed data is around its central point (**skewness, kurtosis**)

- These statistical descriptors are related to statistical moments at indicated on the previous presentation.

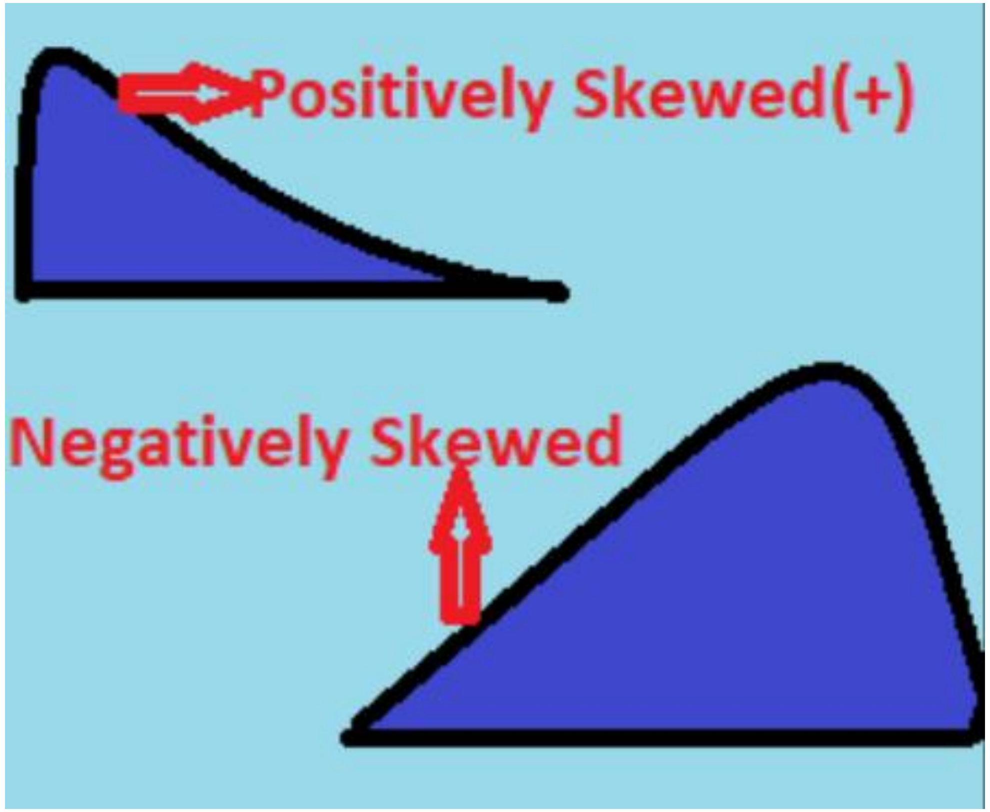
Descriptors and Moments

Descriptors …

- You are well aware of measure central tendency and spread.
- The two most common measure of data asymmetry or shape of spread:

Skewness – measures the asymmetry of the distribution.

- **Zero** skewness is perfectly symmetrical.
- **Positive** skewness indicates a longer right tail,
- **Negative** skewness indicates a longer left tail.



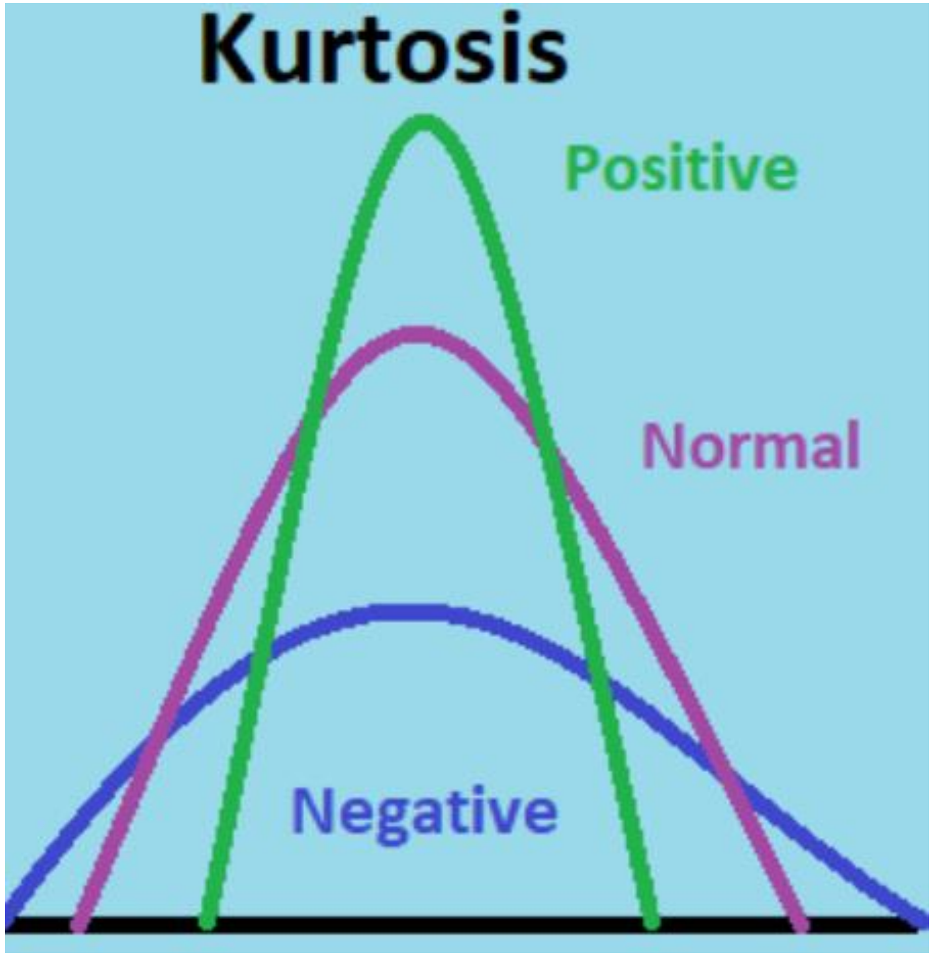
www.statisticalaid.com

Descriptors and Moments

Descriptors …

Kurtosis – measures the "tailedness" or the heaviness of the tails of the distribution.

- **Mesokurtic:** A normal distribution with a kurtosis value of 3.
- **Leptokurtic**: A distribution with high kurtosis (greater than 3), indicating more **extreme values**.
- **Platykurtic**: A distribution with low kurtosis (less than 3), indicating **fewer extreme values**.



www.statisticalaid.com

Descriptors and Moments

Statistical moments : remember **moment in physics** due to the location of application of force.

- Moments in statistics are used to describe the characteristic of a data distribution.
- They tell us much about our data like **mean, variance, skewness**, and **kurtosis**.
- Two commonly used types of statistical moments in hydrosystem engineering applications are:

  (1) Product-moments and

  (2) L-moments or linear - moments

Descriptors and Moments

(1) Product-moments:

- are traditional measures of central tendency and dispersion,
- are standard statistical measures like:
  - mean,
  - variance,
  - skewness, and
  - kurtosis,
- They are derived from the product of values in a dataset.

Descriptors and Moments

(1) Product-moments …

- are standard statistical measures like mean, variance, skewness, and kurtosis, derived from the product of values in a dataset.
- describe the shape, spread, and symmetry of a distribution, allowing for comparison between different datasets.
- **Limitations**: sensitive to **outliers** and may **not be well-defined** for all distributions, like those with heavy tails or extreme values.

Descriptors and Moments

(2) L-moments or Linear moments:

- are an alternative set of statistics calculated using linear combinations of order statistics,
- provide a more robust way to characterize probability distributions, especially when dealing with outliers or small sample sizes.
- are defined as linear combinations of order statistics (L-statistics), which are values from a sorted dataset.

Descriptors and Moments

Key Differences:

Feature	Product Moments	L-Moments
Definition	Standard measures derived from the product of values in a dataset	Linear combinations of order statistics
Sensitivity to Outliers	Highly sensitive to outliers	Less sensitive to outliers
Sample Size	May not be reliable for small samples	Can be reliable even with small samples
Existence	May not exist for all distributions	Exist if the mean is finite

Descriptors and Moments

Moments, expectation, covariance and correlation:

- A single time series is considered to be a stochastic process that can be characterized by its (central) statistical moments.
- The **first** and **second** order moments are the **mean** value, the **variance** and the **autocorrelation** function
- The expected value of a random variable, X, is its mean, often denoted as E(X) or μ: $E(X) = \mu$

Expected Value: $\mu_X = E(X) = \sum x f(x)$

Variance: $\sigma_X^2 = V(X) = \sum (x - \mu_X)^2 f(x) = \sum x^2 f(x) - (\mu_X)^2$

Standard Deviation: $\sigma_X = \sqrt{\sigma_X^2}$

Bivariate statistics

- We have delt about single variable called **univariate statistics**.
- statistics of two variables are considered, called **bivariate statistics,** deals with two variables measured simultaneously at a **single location** or at **a single time**.

Covariance and correlation:

- Both statistical measures are used to assess the relationship between **two variables**, are ranging from -1 to 1.
- Covariance measures the direction and strength of the linear relationship, while correlation is a standardized version of covariance.
  - A **positive covariance** indicates that when one variable increases, the other tends to increase as well, while
  - A **negative covariance** suggests that they move in opposite directions.
- Correlation with -1 representing a perfect negative linear relationship, 1 representing a perfect positive linear relationship, and 0 indicating no linear relationship.

Descriptors and Moments

Covariance and correlation:

- The covariance measures linear co-variation of two datasets of variables z and y. It is calculated from the data as:

$$C_{zy} = \frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)(y_i - m_y) = \frac{1}{n}\sum_{i=1}^{n} z_i y_i - m_z m_y$$

- The correlation coefficient provides a measure of linear co-variation that is normalized with respect to the magnitudes of the variables z and y:

$$r_{zy} = \frac{C_{zy}}{s_z s_y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)(y_i - m_y)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - m_y)^2}} = \frac{n\sum_{i=1}^{n} z_i y_i - \sum_{i=1}^{n} z_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} z_i^2 - \left(\sum_{i=1}^{n} z_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

Covariance and correlation:

- The covariance measures linear co-variation of two datasets of variables z and y. It is calculated from the data as:

$$C_{zy} = \frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)(y_i - m_y) = \frac{1}{n}\sum_{i=1}^{n} z_i y_i - m_z m_y$$

- The correlation coefficient provides a measure of linear co-variation that is normalized with respect to the magnitudes of the variables z and y:

$$r_{zy} = \frac{C_{zy}}{s_z s_y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)(y_i - m_y)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - m_y)^2}} = \frac{n\sum_{i=1}^{n} z_i y_i - \sum_{i=1}^{n} z_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} z_i^2 - \left(\sum_{i=1}^{n} z_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

Break:
for rehearsal
on:

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:

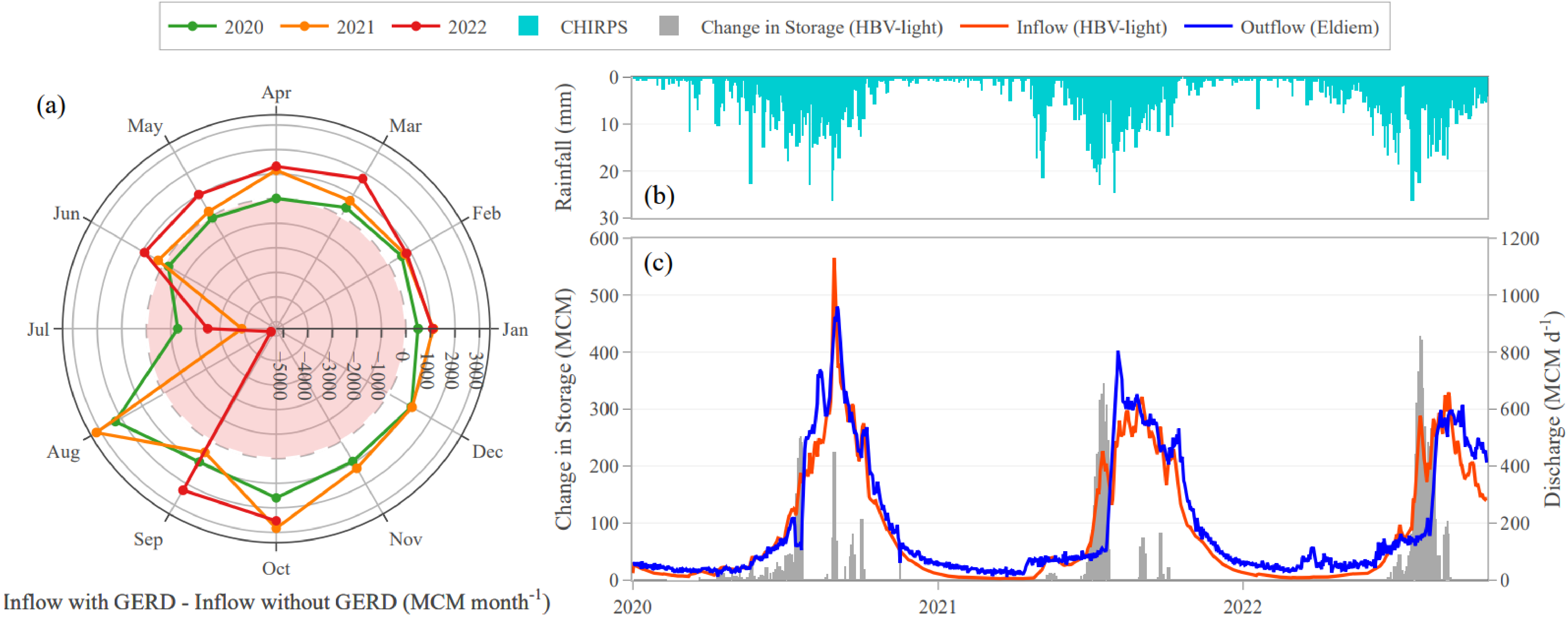
- Hydrograph, hyetograph,
- Hourly temperature readings from a weather station.
- 5 minutes rainfall data from AASTU weather kit
- Daily stage measurements of a river gage station

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 [4]



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

Main reasons for analyzing hydrological time series:

- Characterization:
  - analysis of properties like average values and probability of exceeding threshold values
  - analysis of characteristics such as seasonal behavior and trend
- Prediction and forecasting:
  - estimate the value of the time series at non-observed points in time
- Identify and quantify input-response relations:
  - quantify the effect of an individual influence and to evaluate water management measures

Analysis of hydrologic time series

Main reasons for analyzing hydrological time series:

- Characterization:
  - analysis of properties like average values and probability of exceeding threshold values
  - analysis of characteristics such as seasonal behavior and trend
- Prediction and forecasting:
  - estimate the value of the time series at non-observed points in time
- Identify and quantify input-response relations:
  - quantify the effect of an individual influence and to evaluate water management measures

Time series data synthesis

- It is a process that involves creating **artificial sequences** of data that resemble real-world time series data.
- It can be used for various purposes, including data **augmentation**, **privacy preservation**, and **model testing**.
  - **Data augmentation**: used to increase the size of the training dataset for machine learning models in case of data scarcity.
  - **Privacy preservation**: can be used to **protect sensitive** information in real time series data, allowing for safer sharing and analysis.
  - **Model testing**: can be used to test and refine **machine learning** models in controlled environments, ensuring their adaptability to **different scenarios** or scenario analysis ("What-If" scenarios)
- However, capturing temporal dependencies, correlations, and complication in modeling are challenges of time series synthesis.

Autocovariance and autocorrelation

- Both are measures of the correlation of a **time series** with **itself** at different **time lags**.
- Autocovariance is the **covariance** between a time series and its lagged version.
  - It quantifies the strength and direction of the linear relationship between the time series at **different time lags**.
  - It is used for identifying **patterns, seasonality,** and d**ependencies** in time series data.
  - It is denoted as $\gamma(k)$, measures the covariance between a time series {x(t)} at **time t** and its value at **time t-k**, where **k is the lag**.
  - $\gamma(k) = E[(x(t) - \mu)(x(t-k) - \mu)]$, where μ is the mean of the time series.

Autocovariance and 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.

Some stochastic models

CHAPTER-5 STOCHASTIC HYDROLOGY

- 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]

Home assignment:

Make sure that you can :

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

References

[1] Bierkens, M.F., & van Geer, F. C. (2007). Stochastic hydrology, Department of Physical Geography Utrecht University http://www.earthsurfacehydrology.nl/wp- content/uploads/2012/01/Syllabus_Stochastic-Hydrology.pdf

[2] Quinn, Julianne. 2019. Lecture 1: Stochastic Hydrology, Utah State University DigitalCommons@USU. All ECSTATIC Materials. Paper 81. https://digitalcommons.usu.edu/ecstatic_all/81

[3] Gebreyohannes, M.H., Admasu G.A. & Abraha A.A. 2024. Estimation of rainfall and streamflow missing data under uncertainty for Nile basin headwaters: the case of Ghba catchments, Journal of Applied Water Engineering and Research, 12(2):119-133

[4] 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

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