Chapter 10: Missing Rainfall Data – Estimation

Introduction

Rainfall data is a fundamental requirement in hydrology for the design of water resources projects like dams, canals, drainage systems, and flood control structures. However, due to equipment failure, human error, or natural calamities, rainfall records from certain rain gauge stations may be incomplete or missing. Accurate estimation of this missing data is crucial to maintain the consistency and reliability of hydrological analysis.

This chapter covers the various methods of estimating missing rainfall data, including consistency checks and the techniques recommended by the Indian Meteorological Department (IMD). It also elaborates on criteria for selecting a suitable method, along with step-by-step procedures and numerical examples where necessary.

10.1 Causes of Missing Rainfall Data

- **Instrumental Malfunction:** Damage or malfunctioning of rain gauges.
- **Human Error:** Delayed readings, recording mistakes, or failure to log data.
- **Natural Calamities:** Floods, storms, or earthquakes affecting stations.
- **Communication Issues:** Data not being transmitted to central repositories.
- **Operational Constraints:** Remote locations and lack of staff.

Understanding the causes is important to choose the correct method for estimation and whether the data gap is recoverable.

10.2 Importance of Estimating Missing Rainfall Data

- Ensures continuity in long-term records.
- Maintains the accuracy of hydrological analysis.
- Supports water resources project design and modeling.
- Facilitates computation of averages, intensity-duration-frequency (IDF) curves, and runoff estimation.

10.3 Criteria for Estimation Method Selection

The choice of estimation method depends on:

- Length and duration of missing record.

- Number of neighboring stations with complete data.
- Homogeneity and proximity of stations.
- Topographic and climatic similarity.
- Availability of long-term records.

10.4 Estimation Techniques

10.4.1 Arithmetic Mean Method

Applicability: When rainfall at surrounding stations is fairly uniform (variation <10%).

Formula:

$$P_x = \frac{1}{n} \sum_{i=1}^{n} P_i$$

Where: P_x = missing rainfall value at station X P_i = rainfall at i^{th} neighboring station n = number of neighboring stations

Advantages:

- Simple and quick.
- Good for flat terrains.

Limitations:

- Inaccurate in non-uniform orographic regions.

10.4.2 Normal Ratio Method

Applicability: When normal rainfall at surrounding stations differs by more than 10% from the station with missing data.

Formula:

$$P_x = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{N_x}{N_i} \cdot P_i \right)$$

Where: P_x = estimated rainfall at station X N_x = normal rainfall at station X N_i = normal rainfall at station i P_i = observed rainfall at station i

Advantages:

- Adjusts for climatic variability.

Limitations:

- Requires long-term normals.

10.4.3 Inverse Distance Weighting Method (IDW)

Applicability: When distances between the station with missing data and nearby stations are available.

Formula:

$$P_x = \frac{\sum_{i=1}^{n} \left(\frac{P_i}{d_i^2}\right)}{\sum_{i=1}^{n} \left(\frac{1}{d_i^2}\right)}$$

Where: P_x = estimated rainfall at station X P_i = rainfall at station i d_i = distance between station i and station X

Advantages:

- Considers geographical proximity.

Limitations:

- Requires accurate distance data.
- May not work well in regions with varying topography.

10.4.4 Multiple Regression Method

Applicability: When relationships between rainfall at different stations are linear and statistically significant.

General Form:

$$P_x = a_1 P_1 + a_2 P_2 + \ldots + a_n P_n + C$$

Where:

- $P_1, P_2, ..., P_n$ are rainfall values at neighboring stations
- $a_1, a_2, ..., a_n, C$ are regression coefficients determined statistically.

Procedure:

1. Collect data from nearby stations.
2. Conduct regression analysis.
3. Apply regression equation for estimation.

Advantages:

- Highly accurate if data is consistent.
- Captures correlation among stations.

Limitations:

- Requires computational tools.
- Sensitive to outliers.

10.5 Checking the Consistency of Rainfall Data

Before estimating missing data, it is crucial to verify the consistency of the available records. This is commonly done using a **Double Mass Curve**.

Double Mass Curve Method

Procedure:

1. Plot cumulative rainfall of the station under study against cumulative mean of nearby stations.
2. If the plot is linear → data is consistent.
3. If there's a break in slope → indicates inconsistency or change in gauge location/instrument.

Correction: Adjust the inconsistent portion using correction factors derived from the slope changes.

10.6 Homogeneity and Stationarity of Rainfall Data

- **Homogeneity**: Indicates whether the rainfall data comes from the same climatic regime.
- **Stationarity**: Assumes statistical properties (mean, variance) remain constant over time.

If the data is not homogeneous or stationary, the missing value estimation can be highly unreliable and may require additional statistical treatment or modeling.

10.7 Role of IMD Normals

The Indian Meteorological Department (IMD) provides:

- 30-year normals for various stations.
- Guidelines for estimation and comparison.

- Seasonal and annual normal data tables.

These IMD-provided normals are essential for the Normal Ratio Method and for consistency checks.

10.8 Use of GIS and Software Tools

With advances in technology:

- **GIS tools** can map stations and interpolate rainfall using IDW or Kriging.
- **Software packages** like MATLAB, Python (with pandas/geopandas), and commercial hydrology tools (e.g., HEC-HMS, SWAT) can automate missing data estimation.

These tools ensure spatially accurate and statistically sound estimation across large datasets.

10.9 Practical Guidelines

- Use at least 3–6 neighboring stations.
- Always cross-check estimated data against known meteorological patterns.
- Avoid using arithmetic mean in regions with high rainfall variability.
- Maintain metadata on how estimations were made for future reference.
