सत्यमेव जयते

Government of India

MANUAL ON RAINFALL DATA VALIDATION

MINISTRY OF WATER RESOURCES,
RIVER DEVELOPMENT & GANGA REJUVENATION

NHP

NATIONAL HYDROLOGY PROJECT

सत्यमेव जयते

Government of India

Ministry of Water Resources, River Development and Ganga Rejuvenation

Manual on
Rainfall Data Validation

NHP

National Hydrology Project

THE WORLD BANK
IBRD • IDA

May 2019

Ministry of Water Resources, River Development and Ganga Rejuvenation
National Hydrology Project
2nd Floor, Rear wing, MDSS (MTNL) Building, 9,CGO Complex,
Lodhi Rd, New Delhi 110003
E-mail: nhp-mowr@nic.in
Website: http://nhp.mowr.gov.in

Title:

Manual on Rainfall Data Validation

Summary:

This manual has been prepared by the Technical Assistance and Management Consultancy for the Ministry of Water Resources, River Development and Ganga Rejuvenation under the National Hydrology Project. It presents a compilation of standard methods used for the validation of rainfall data. The manual updates and further develops the manuals prepared under the previous HP-I and HP-II projects.

Disclaimer:

In the preparation of this document every effort has been made to correctly present information, methods and formulae from reliable sources. However, neither the compilers of the manual nor the Ministry assume any responsibility, explicit or implicit, to the analysis results produced by other parties and the consequences arising thereof.

Date:

May 2019

Sh. Akhil Kumar
Joint Secretary
Ministry of Water Resources,
River Development and Ganga Rejuvenation

Preface

Hydro-meteorological observations and the statistical analyses thereon are the basis for efficient and sustainable water management. They play a crucial role in supporting informed decision making in water resources planning, flood forecasting and flood management. Telemetry based rainfall monitoring integrated with real time Decision Support Systems (DSS) has a huge potential to provide rapid and efficient knowledge tools for flood forecasting and the operation of hydraulic infrastructure during flood events, considering a range of operating criteria and flood scenarios. Seasonal forecasts and strategic river basin modelling allow to make the best possible use of water resources and effectively manage water scarcity. The basis for all this are reliable hydro-meteorological observations.

Through the Hydrology Project Phase-I, the concept of Hydrological Information Systems was introduced imparting training to Implementing Agencies and preparing training manuals on the analysis and multi-stage validation of hydro-meteorological data. Being an extension of HP-I, Phase-II of the Hydrology Project focused on the analytical framework for state-of-art flood forecasting and water resources planning and management.

Under the National Hydrology Project (NHP) the monitoring networks and the analytical tools for water resources planning and management will be further enhanced, building upon the practices established in HP-I and HP-II, improving real-time monitoring and developing web-based Water Resources Information Systems comprising time series data, geographic databases & developing various applications and dissemination portals.

Hydro-met data can be subject to errors at various levels, including erroneous field measurements, data entry and transfer of information. Data analysis and validation ensure that the information which reaches water resource planners, designers and managers is reliable and free from errors. This Manual describes the techniques of data analysis and validation of rainfall data.

Water resources planning, real-time forecasting and systems operation require adequate information on the hydro-meteorological regime. Poor availability of comprehensive and good quality data often leads to unsound designs and operation. With this need in mind, the current Manual is an effort to provide a ready reference for a variety of users, including water resources planners, hydrologists, site and field engineers, designers, and water systems operators.

Comprising of eight chapters, the Manual describes in detail the concepts of primary and secondary data validation, correction, compilation, completion, analysis, and report generation of rainfall data.

I am confident that this document will be of great use for a wide range of water professionals at different levels, not only from Implementing Agencies of NHP but for a larger audience in the water sector. It will bring us one step closer to addressing the challenges of water resources management in India.

27/05/2019

Sh. Akhil Kumar
Joint Secretary
MoWR, RD & GR

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ARF	Areal Reduction Factor
ARG	Automatic rain gauge
cm	Centimetre
CWC	Central Water Commission
DAD	Depth Area Duration
H	Height
HP	Hydrology Project
IDW	Inverse distance weighted (interpolation)
IMD	India Meteorological Department
km	Kilometre
m	Metre
MM	Millimetres
N	Number
NHP	National Hydrology Project
P	Precipitation
Pest	Estimated precipitation
PMP	Probable Maximum Precipitation
Pobs	Observed precipitation
PWM	Probability Weighted Moments
Racc	Accumulated rainfall
SRG	Standard rain gauge
Stdev	Standard deviation
TBRG	Tipping Bucket Rain Gauge
WMO	The World Meteorological Organisation

1 INTRODUCTION

1.1 Background

The knowledge of the amount and distribution of rainfall in time and space is an essential element of water and energy balance studies, planning of agriculture, and research in meteorology and climatology. It is an important input into runoff computations, flood forecasting and various engineering design computations. proper collection and processing of rainfall data is a prerequisite to carrying out any hydrological analysis.

The effect of wind induced errors on measurement of rainfall has been widely reported over the decades. However, owing to the complexity and uncertainty in the choice of a universally acceptable wind correction factor, and paucity of the wind data, it still resides in the domain of research. This topic therefore has not been addressed in the current manual, prepared primarily to deal with the needs of the engineer at the field.

Rainfall is arguably the most frequently measured hydro-meteorological variable. It is also one that is most useful, particularly in the countries like India, where long term observed records of variables representing other components of the hydrological cycle are either non-existent or scant. Therefore, there often arises a need to estimate the amount representing these other variables like runoff, evaporation, transpiration, infiltration, based on the measured rainfall and available assessment procedures.

The India Meteorological Department (IMD) is the prime organisation for collection, storage and dissemination of all data related to meteorological variables in the country, which maintains their network of rain gauges and weather stations. As water resources projects are often carried out at remote locations, this sometimes leads to forcible use of data maintained by other state organisations like the Water Resources Department, the Agricultural Department, the Disaster Management Department, Universities and the like.

It is a common experience that the rainfall data in its raw form contain many gaps and inconsistent values. The procedures to check and validate the rainfall records are not very well studied in the academic institutions at the undergraduate engineering programmes. Nor are they compiled into a single organised document that is readily available.

This Manual on Procedures for Handling and Processing of Rainfall Data is designed to help practitioners deal with the issues related to data maintenance, validation and processing. It contains the overview of the most commonly available procedures to accomplish those tasks.

1.2 The need for the manual

The primary goal of the Ministry of Water Resources, River Development and Ganga Rejuvenation is to ensure optimal sustainable development, maintenance of quality and efficient use of water resources to match with the growing water demands of the country. The Ministry is responsible for laying down policy guidelines and programmes for the development and regulation of country's water resources. This includes providing technical guidance, scrutiny, clearance and monitoring of all aspects of water use.

Individual module reports were published under the earlier hydrology projects (HP-I and HP-II), which catered to the specific objectives of meeting the training needs on carrying out primary validation of rainfall data, carrying out secondary validation of rainfall data, correction and completion of rainfall data records, compilation of rainfall data, analysis of rainfall data, and the preparation of rainfall data reports. To make it compatible with the slow internet speed available in those days, there were restrictions on the file size to ensure its successful download. Subsequent developments of hardware and software, and the wide range availability of freeware have simplified many cumbersome tasks. The availability of data for hydrologic analyses has also significantly improved with the additional developments of the India WRIS website.

The National Hydrology Project has been approved by the Cabinet on 6.4.2016 as a central sector scheme, with a further objective to improve the extent, quality, and accessibility of water resources information, decision support systems for floods and basin level resource assessment and planning, and to strengthen the capacity of targeted water resources professionals and management institutions in India. It includes the development of a series of new and revised manuals and guidelines. These include guidance documents that could be applied nationwide, such as this one dealing with the Procedures for Handling and Processing Rainfall Data.

1.3 Purpose and scope of the manual

Recently, the Ministry has approved sharing of restricted data of the Ganga Basin to the concerned states, to users with due administrative privileges. This is expected to open up huge possibilities for studies related to conception, planning and optimisation of future and existing projects, ensuring betterment of future water resource management. The goal of this manual is to compile the available techniques of rainfall data processing, validation and analysis under a single volume that is available free of charge. Apart from professional practitioners, it is also expected to benefit the research community and the students in India and other countries.

Even though it is advisable to follow the procedures described here, most of which are commonly accepted among practitioners, neither the authors of this manual nor the Ministry accept explicitly or implicitly any responsibility resulting from errors or erroneous use of these methods.

1.4 Publication and contact information

This document is available on the website for the National Hydrology Project

http://www.nhp.mowr.gov.in/

For any further information contact:

Senior Joint Commissioner 1
National Hydrology Project
Ministry of Water Resources, River Development and Ganga Rejuvenation
2nd and 3rd Floor, Rear Wing
9, CGO Complex
MTNL Building, Lodhi Road
New Delhi – 110003
Email: sjc1nhp-mowr@gov.in

1.5 Acknowledgments

Preparation of this manual relied on the work of others in India, the Netherlands, and elsewhere. Grateful appreciation is extended to the following organizations whose publications and websites provided valuable information:

- Hydrology Project (HP)
- Bureau of Indian Standards
- The World Meteorological Organization (WMO)
- The original authors of the Hymos software system which served as a basis for establishing the background information for this manual in the previous phases of the HP-I and HP-II projects

2 DATA VALIDATION

2.1 Role of data validation

The statistics of hydro-meteorological data underpin the water management policies and practices of Water Resource initiatives of a nation. However, hydro-meteorological observations are subject to errors arising at various levels from field measurement, data entry, data computation, transfer or correction. Data Validation is a process to ensure that the value stored is reliable and the best possible representation of true value of variable at the measurement site at a given time or in a given interval of time. The processes under Data Validation are multi-level and parameter specific, broadly covered under a series of functionalities, depicted in Figure 2.1.

Data Validation is carried out mainly for three reasons:

1. To correct errors in the recorded data wherever possible,
2. To assess the reliability of a record where it is not possible to correct errors
3. To identify the source of errors and to ensure that such errors are not repeated in future.

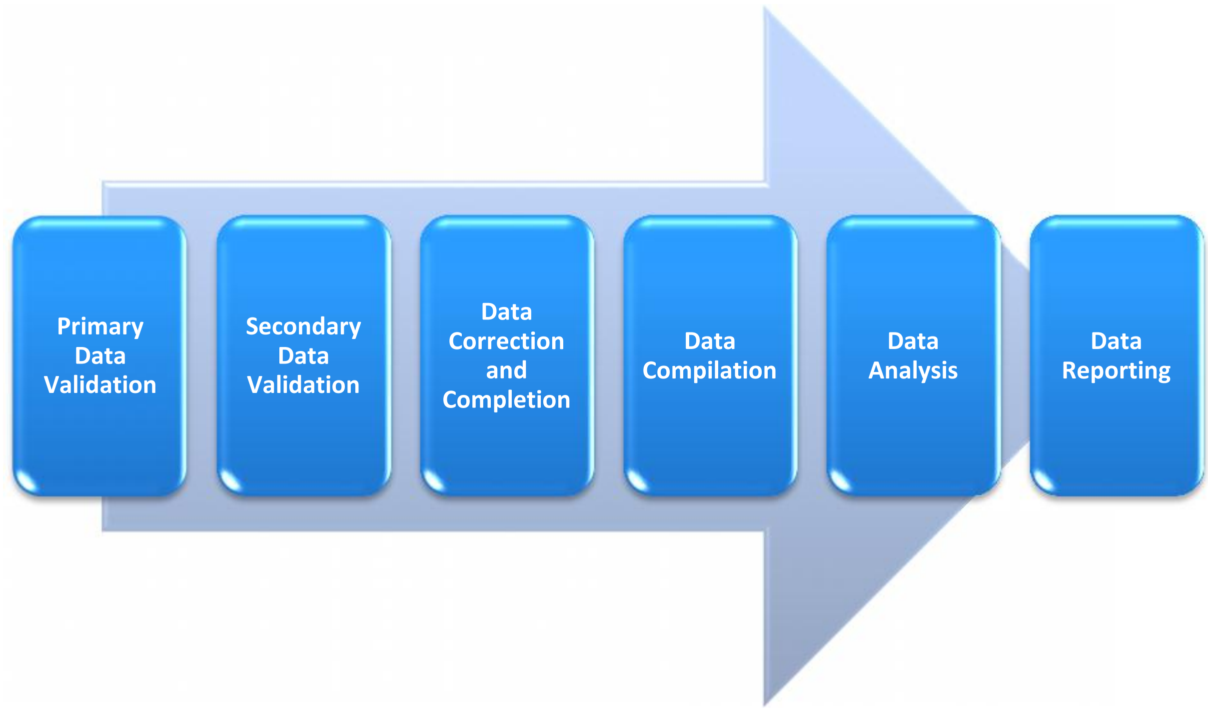


Figure 2.1: Multi level processes in data validation

By their nature, errors can be classified as random, systematic or spurious:

- **Random errors** are sometimes referred to as experimental errors and are equally distributed about the mean or 'true' value. The errors of individual readings may be large **o**r small, e.g. the error in a rainfall gauge reading where the water surface is subject to wind action, but they tend to compensate with time or by taking sufficient number of measurements.

- **Systematic errors** imply the existence of a systematic difference, either positive or negative, between the measured value and the true value, where the situation is not improved by increasing the number of observations. For example, Hydrometric field measurements are often subject to a combination of random and systematic errors. Systematic errors are generally more serious and are what the validation process is designed to detect and if possible, to correct.
- **Spurious errors** are sometimes distinguished from random and systematic errors as arising due to some abnormal external factor. Such errors may be readily recognized but cannot easily be statistically analysed and the measurements often are discarded.

2.1.1 Levels of validation

It is desirable to carry out data validation as soon as the data is observed. However, complete validation close to observation sites is impractical both in terms of computational support from equipment and staffing. The sequence of validation process has therefore been divided so that those which primarily require interaction with the observation station, are carried out in close proximity (i.e. at State Sub-divisional office) whereas the more complex validation procedure is carried out at higher levels. Essentially, data validation is a multi-stage process and sometimes a two-way approach.

Based on the sequence and level, data validation can be grouped into two major categories: Primary data validation and Secondary data validation.

2.1.1.1 Primary data validation

Primary data validation is presumed to be carried out immediately after the observations are made or data extracted from charts or downloaded from loggers. This ensures that any obvious errors coming from the observer or instrument are spotted at the earliest and resolved. Primary validation is primarily involved with comparison of variable observation records restricted within a single data series with pre-set limits, statistical range, or in conformance with the expected hydrological behaviour. However, data from stations in close proximity may also sometimes be available and this may be used in primary validation.

Primary data validation highlights those data which are not within the expected range or are not hydrologically consistent. These data are then revisited in the data sheets or analogue records to see if there were any errors while making computations in the field or during keying-in the data. If it is found that the entered value(s) are different than the recorded ones, then such entries are immediately corrected. Where such data values are found to have been correctly entered, they are then flagged as doubtful with a remark against the value in the computer file, indicating the reason for such doubt.

Apart from data entry errors, the suspected values are identified and flagged but not amended at the Sub-divisional level. However, the flag and remarks provide a basis for further consideration of action at the time of secondary data validation.

2.1.1.2 Secondary data validation

Secondary data validation consists of comparison of the variables at two or more stations. This is carried out to test the data against expected behaviour of the system on a spatial scale. The underlying assumption is that the variables under consideration have adequate spatial correlation within their distance. This correlation is derived on the basis of historical records and the statistics and utilized to validate the data. For certain hydrological variables like water level and discharge, which bear a very high degree of dependence or correlation between adjoining stations, the inter-relationship can be established with a comparatively higher level of confidence. However, for some variables which lack serial correlation and show great spatial variability (e.g. convectional rainfall), it is difficult to ascertain the behaviour with the desired level of confidence. In such circumstances, it becomes very difficult, if not impossible, to detect errors.

While validating data on the basis of group of surrounding stations, the strategy must always be to rely on certain key stations known to be of good quality. If all the observation stations are given the status of being equally reliable, data validation will become comparatively more difficult. Field experience shows that the quality of data received from some stations are better than that received from other stations. Also, the process of allocating greater weights to a limited number of stations makes the data validation procedure simpler and faster to carry out. This may be due to physical conditions at the station, quality of instruments, or reliability of staff. It must always be remembered that these key or reliable stations can also report incorrect data and they do not enjoy the status of being absolutely perfect.

Similar to primary data validation, the guiding factor for secondary data validation is that none of the test procedures should be considered objective on their own. They must always be taken as tools to screen out suspect data values. The validity of each of these suspect values is then confirmed on the basis of other tests and corroborative facts based on the information received from all stations. It is only when it is clear that a certain value is incorrect and an alternative value provides a more reliable indication of the true value of the variable that suitable correction should be applied and the value be flagged as corrected.

If it is not possible to confidently conclude that the suspect value is incorrect, then such values should be left as recorded with a proper flag indicating doubt. All data which have been identified as suspicious at the level of primary validation are to be validated again on the basis of additional information available from a larger surrounding area. All such data which are supported by additional spatial information must be accepted as correct and accordingly the flags indicating them as doubtful must be removed at this stage.

2.2 Data in-filling (completion) and correction

Raw observed data may have missing values or sequence of missing values due to factors like equipment malfunction, observer absence, etc. These gaps should, where possible, be filled to make the series complete. In addition, all values flagged as doubtful in validation must be reviewed to decide whether they should be replaced by corrected values or whether doubt remains but a more reliable correction is not possible and the original value remains with a flag.

In-filling or completion of a data series is done in a variety of ways depending on the length of gap, nature of the variable and availability of suitable records for estimation. The simplest case is where variables are observed with more than one instrument at the same site (e.g. daily rain gauge and recording gauge); the data from one gauge can be used to complete the data from the other gauge. For a single value or short gaps in a series with high serial correlation, simple linear interpolation between known values or values filled with reference to the graphical plot of the series may be acceptable. Gaps in series with high random component and little serial correlation such as rainfall cannot be filled in this way and must be completed with reference to neighbouring stations through spatial interpolation. Longer gaps can be filled through the appropriate regression analysis. However, it must be emphasized here that various methods used for in-filling or correction will affect the statistics of the variable unless care is also taken with respect to its probability distribution function.

Data correction is to be done using similar procedures as for completing the data series. In case of rainfall, there can be a shift in recorded values. The possible reasons can be due to identified systematic error or due to the relocation of on observation station. The data correction can involve techniques like Double Mass curve to adjust the portion of shift for the record to be consistent with the present and continuing data.

2.3 Data compilation

Compilation refers primarily to the transformation of data observed at a certain time interval to a different interval, e.g. hourly to daily, daily to weekly, weekly to monthly, etc. This is done by a process of aggregation. Occasionally, disaggregation, or a conversion from longer to shorter time steps, may also be required, but it is usually not recommended due to the loss of accuracy of the resulting disaggregated data.

Compilation also refers to the transformation of point rainfall to areal rainfall. Both areal averaging and aggregation may be required for validation, for example in rainfall runoff comparisons, but also to provide a convenient means of summarizing large data volumes.

Derived series can also be created from the raw data. The examples of this include the maximum, minimum and mean statistics for selected time intervals, or a listing of peaks over thresholds, to which a variety of hydrological analyses may be applied.

2.4 Data analysis

Procedures used in data validation and reporting have wide analytical use. The following are examples of the available techniques:

1) Basic statistics (e.g. mean, standard deviations, etc.)
2) Statistical tests
3) Fitting of frequency distributions
4) Flow duration series
5) Regression analysis
6) Rainfall Depth-Area-Duration
7) Rainfall Intensity-Frequency-Duration

2.5 Data reporting

Data reporting includes periodic publications of special reports showing long term statistics for selected stations, or special reports of unusual events. This can be prepared in digital form or provided on web enabled information media through a wide range of PDF and graphical formats available. Examples of these include the comparisons of the current year values with the long-term statistics, thematic maps of variables such as annual and seasonal rainfall, duration and frequency curves, etc. More detailed information such as stage discharge ratings can be provided to meet specific needs.

The annual report shows how observations at individual stations are integrated in the network in which rainfall and other data are transferred in stages from the field to local and regional offices for data entry, processing and validation.

3 PRIMARY VALIDATION OF RAINFALL DATA

3.1 General

Improvement in computing facilities now enables the validation to be carried out at primary level whereas in the past, considering the volume of data and the time required to carry out, comprehensive manual validation was prohibitive.

Primary validation of rainfall data can be carried out at the Sub-divisional level and is concerned with data comparisons at a single station:

- for a single data series, between individual observations and pre-set physical limits
- between two measurements of a variable at a single station, e.g. daily rainfall from the standard rain gauge and an accumulated total from an digital recorder

Before carrying out the Primary Validation, it is presumed that data entry checks have already been conducted to ensure that there has been no transcription error from field sheets to the database. Some doubtful values may already have been flagged by the field supervisor.

The high degree of spatial and temporal variability of rainfall compared to other climate variables make validation of rainfall more difficult. This is particularly the case on the Indian sub-continent, experiencing a monsoon type of climate involving convective precipitation.

3.2 Instruments and observational methods

The method of measurement or observation influences our view of why the data are suspect. To understand the source of errors we must understand the method of measurement or observation in the field and the typical errors of given instruments and techniques.

Data validation is not a purely statistical or mathematical exercise. Staff involved in it must understand the field practice. Three basic instruments are in use for measurement of daily and short duration rainfall:

i. Standard daily rain gauge
ii. Siphon gauge with chart recorder
iii. Tipping bucket gauge with digital recorder

These will be separately described with respect to the typical errors that occur with each gauge or observation method, and the means by which errors might be detected.

3.2.1 Standard rain gauge (SRG)

3.2.1.1 Instrument and procedure

Daily rainfall can be measured using the familiar standard rain gauge (SRG). This consists of a circular collector funnel with a brass or gun metal rim and a collection area of either 200 cm^2 (diameter 159.5 mm) or 100 cm^2 (diameter 112.8 mm), leading to a base unit partly embedded in the ground and containing a polythene collector bottle. The gauge is read once or twice daily and any rain held in the polythene collector is poured into a measuring glass to determine rainfall in millimetres. Typical measurement errors are:

- Observer reads measuring glass incorrectly
- Observer enters amount incorrectly in the field sheet
- Observer reads gauge at the wrong time (the correct amount may thus be allocated to the wrong day
- Observer enters amount to the wrong day
- Observer uses wrong measuring glass (i.e. 200 cm^2), glass for 100 cm^2 gauge, giving half the true rainfall or 100 cm^2 glass for 200 cm^2 gauge giving twice the true rainfall
- Observed total exceeds the capacity of the gauge
- Instrument fault - gauge rim damaged so that collection area is affected
- Instrument fault - blockage in rain gauge funnel so that water does not reach collection bottle and may overflow or be affected by evaporation
- Instrument fault - damaged or broken collector bottle and leakage from gauge

It may readily be perceived that errors from most of these sources will be very difficult to detect from the single record of the standard rain gauge, unless there has been a gross error in reading or transcribing the values. These kinds of errors are described in more detail in sub-section 3.4 and 3.5.

Errors at a station are more readily detected if there is a concurrent record from an autographic rain gauge or from a digital record obtained from a tipping bucket rain gauge (TBRG). As these too are subject to errors (of a different type), comparisons with the daily rain gauge will be followed for the descriptions of errors for these gauges.

The final check by comparison with rain gauges at neighbouring stations will show up further anomalies, especially for those stations which do not have an autographic or digital rain gauge at the site. This is carried out under Secondary Validation at the Divisional office where more gauges are available for comparison.

3.2.2 Autographic rain gauge (natural syphon)

3.2.2.1 Instrument and procedure

In the past short period rainfall has been measured almost universally using the natural syphon rain gauge. The natural syphon rain gauge consists of the following parts:

- a circular collector funnel with a gun metal rim, 324 cm^{2}in area and 200 mm in diameter and set at 750 mm above ground level, leading to float chamber in which is located a float which rises with rainfall entering the chamber
- a syphon chamber is attached to the float chamber and syphon action is initiated when the float rises to a given level. The float travel from syphon action to the next represents 10 mm rainfall.
- a float spindle projects from the top of the float to which is attached
- a pen which records on a chart placed on a clock drum with a mechanical clock

The chart is changed daily at the principal recording hour. During periods of dry weather, the rainfall traces a horizontal line on the chart; during rainfall it produces a sloping line, the steepness of which defines the intensity of rainfall. The chart is graduated in hours and the observer extracts the hourly totals from the chart and enters it in a register and computes the daily total.

3.2.2.2 Typical measurement errors

Potential measurement faults are now primarily instrumental rather than caused by the observer and include the following:

- Funnel is blocked or partly blocked so that water enters the float chamber at a different rate from the rate of rainfall
- Float is imperfectly adjusted so that it syphons at a rainfall volume different from 10 mm
- During heavy rainfall the float rises and syphons so frequently that individual pen traces cannot be distinguished
- Clock stops; the rainfall is not recorded or clock is either slow or fast and thus timings are incorrect
- Float sticks in float chamber, hence the rainfall is not recorded or it is recorded incorrectly
- Observer extracts information incorrectly from the pen trace

In addition, differences may arise from Ordinary Rain Gauge due to different exposure conditions arising from the effects of different level of the rim and larger diameter of collector, apart from other possible reasons. It is the usual practice to give priority to daily SRG whenever there is a discrepancy between the two.

3.2.3 Tipping bucket rain gauge

3.2.3.1 Instrument and procedures

Short period rainfall is more readily digitised using a tipping bucket rain gauge. It consists of the following components.

- A circular collector funnel with a brass or gunmetal rim of differing diameters, leading to a tipping bucket arrangement which sits on a knife edge. It fills on one side, and then tips filling the second side and so on.
- A reed switch actuated by a magnet registers the occurrence of each tip

- A data logger which records the occurrence of each tip and places a time stamp with each occurrence

The logger stores the rainfall record over an extended period and it may be downloaded as required. The logger may rearrange the record from a non-equidistant series of tip times to an equidistant series with amounts at selected intervals. The digital record thus does not require the intervention of the field observer. For field calibration, a known amount of rainfall is periodically poured into the collector funnel and checked against the number of tips registered by the instrument.

3.2.3.2 Typical measurement errors

- Funnel is blocked or partly blocked so that water enters the tipping buckets at a different rate from the rate of rainfall
- Buckets are damaged or out of balance so that they do not record their specified tip volume
- Reed switch fails to register tips
- Reed switch double registers rainfall tips as bucket bounces after tip (better equipment includes a de bounce filter to eliminate double registration.
- Failure of electronics due to lightning strike etc. (though lightning protection usually provided)
- Incorrect set up of measurement parameters by the observer or field supervisor

Differences may arise from the daily rain gauge (SRG) for reasons of different exposure conditions in the same way as the autographic rain gauge.

3.2.4 Real Time Data Acquisition System (RTDAS)

Under the National Hydrology Project, tipping bucket rain gauges with telemetering facilities are being installed. The system comprises of data collection platform, telemetry device and database management systems on servers installed at centralised locations. Through this, real time rainfall data will be shared through the internet and made available to all relevant agencies. Please refer to the guideline "*An Introduction to Real-time Hydrological Information System*" published under the NHP (MoWR RD & GR, 2018) for details.

3.3 Comparison of daily time series for manual and autographic or digital data

3.3.1 General description

If a standard rain gauge is available at stations where rainfall is measured at short durations using an autographic or a digital recorder, rainfall data at daily time interval is available from two independent sources. The rainfall data at hourly or smaller interval is aggregated at the daily level and then a comparison is made between the two. The differences which are less than 5% can be attributed to exposure, instrument accuracy and precision in tabulating the analogue records and are ignored. Any appreciable difference (more than 5%) between the two values must be probed further.

Given the fact that there is a higher degree of possibility of malfunctioning of autographic or digital recorders owing to their mechanical and electromechanical systems, the observation made using a standard rain gauge is considered more reliable. However, significant systematic or random errors are also possible in the daily Rain gauge as shown above.

If the error is in the autographic or digital records, then it must be possible to relate it either to instrumental or observational errors. Moreover, such errors tend to repeat under similar circumstances.

3.3.2 Data validation procedure and follow up actions

This type of validation can be carried out in tabular or graphical form. For both approaches, the values of hourly data are aggregated to daily values to correspond to those observed using a standard rain gauge. A comparison is made between the daily rainfall observed using standard and automatic gauges. Percent discrepancy can be shown by having a second axis on the plot. Tabular output for those days for which the discrepancy is more than 5% can be obtained. A visual inspection of such a tabulated output will ensure screening of all the suspect data with respect to this type of discrepancy.

The following provides a diagnosis of the likely sources of error with discrepancies of different sorts along with the corresponding actions:

1) Where the recording gauge gives a consistently higher or lower total than the daily gauge, then the recording gauge could be out of calibration and either tipping buckets (TBRG) or floats (ARG) need recalibration.

Action: Accept SRG and adjust ARG or TBRG

2) Where agreement is generally good, but the difference increases in high intensity rainfall suggest that for the ARG:
    - the syphon is working imperfectly in high rainfall, or
    - the chart trace is too close to distinguish each 10 mm trace (underestimate by multiples of 10 mm)

    For the TBRG:

    - gauge is affected by bounce sometimes giving double tips

Action: Accept SRG and adjust ARG or TBRG

3) Where a day of positive discrepancy is followed by a negative discrepancy and rainfall at the recording gauge was occurring at the observation hour, and then it is probable that the observer read the SRG at a different time from the ARG. The sum of SRG readings for successive days should equal the two-day total for the TBRG or ARG

Action: Accept TBRG or ARG and adjust SRG

4) Where the agreement is generally good but isolated days have significant differences, then the entered hourly data should be checked against the manuscript

values received from the field. Entries resulting from incorrect entry are corrected. Check that water added to the TBRG for calibration is not included in rainfall total. Otherwise there is probable error in the SRG observation.

Action: Accept ARG or TBRG and adjust SRG

In certain cases the values reported for daily rainfall by SRG and ARG match one to one on all days for considerable period notwithstanding the higher rainfall values etc. It is very easy to infer in those situations that there has been an attempt by the observer to match these values forcefully by manipulating one or both data series. It is not expected that both these data series should exactly match in magnitude, since such variation should exist due to variance in the catch and instrumental and observation variations.

Example 3-1

Consider the daily totals of hourly rainfall (observed by an autographic rain gauge) and the daily rainfall observed by the standard rain gauge (SRG) at station Askheda of Pargaon catchment. The graphical and tabular comparison of these two data series for the period from 1/9/1996 to 31/10/99 is given in Table 3.1 and Figure 3.1, respectively.

It is clear from these graphical and tabular outputs that there has been a marked difference between the reported daily rainfall as observed from a standard rain gauge and that obtained by compiling the hourly values, tabulated from autographic chart, to daily level.

Table 3.1: Comparison of daily rainfall obtained from SRG and ARG at the same station Aksheda

Year	Month	Day	SRG	ARG	% Diff.
1996	9	1	0.00	0.00	-
1996	9	2	0.00	0.00	-
1996	9	3	0.00	0.00	-
1996	9	4	0.00	0.00	-
1996	9	5	0.00	0.00	-
1996	9	6	18.70	18.50	-1.10
1996	9	7	0.00	0.00	-
1996	9	8	3.70	4.00	8.10
1996	9	9	0.00	0.20	-
1996	9	10	0.00	0.00	-
1996	9	11	0.00	0.0	-
1996	9	12	0.00	0.00	-
1996	9	13	5.00	0.00	-100.00
1996	9	14	0.00	4.80	-

Year	Month	Day	SRG	ARG	% Diff.
1996	9	15	3.90	0.00	-100.00
1996	9	16	3.80	4.80	26.30
1996	9	17	7.20	3.50	-51.40
1996	9	18	0.00	6.90	-
1996	9	19	2.00	0.00	-100.00
1996	9	20	0.00	2.00	-
1996	9	21	0.00	0.00	-
1996	9	22	0.00	0.00	-
1996	9	23	14.00	0.00	-100.00
1996	9	24	13.20	14.80	12.10
1996	9	25	3.80	13.50	255.30
1996	9	26	6.80	3.50	-48.50
1996	9	27	3.00	7.20	140.00
1996	9	28	0.00	3.00	-
1996	9	29	2.70	0.00	-100.00
1996	9	30	0.00	2.40	-
1996	10	1	0.00	0.00	-
1996	10	2	19.00	18.30	-3.70
1996	10	3	75.80	0.50	-99.30
1996	10	4	2.80	1.50	-46.40
1996	10	5	4.00	4.50	12.50
1996	10	6	0.00	0.00	-
1996	10	7	0.00	0.00	-
1996	10	8	0.00	0.00	-
1996	10	9	0.00	0.00	-
1996	10	10	0.00	0.00	-
1996	10	11	0.00	0.00	-
1996	10	12	0.00	0.00	-
1996	10	13	0.00	0.00	-
1996	10	14	0.00	0.00	-
1996	10	15	0.00	0.00	-
1996	10	16	0.00	0.00	-
1996	10	17	0.00	0.00	-
1996	10	18	0.00	0.00	-
1996	10	19	0.00	0.00	-
1996	10	20	2.00	1.80	-10.00
1996	10	21	50.30	50.10	-0.40
1996	10	22	0.70	1.50	114.30
1996	10	23	70.00	70.50	0.70
1996	10	24	9.00	8.90	-1.10
1996	10	25	0.00	0.00	-

Year	Month	Day	SRG	ARG	% Diff.
1996	10	26	0.00	0.00	-
1996	10	27	0.00	0.00	-
1996	10	28	0.00	0.00	-
1996	10	29	6.00	6.00	-
1996	10	30	0.00	0.00	-
1996	10	31	0.00	0.00	-

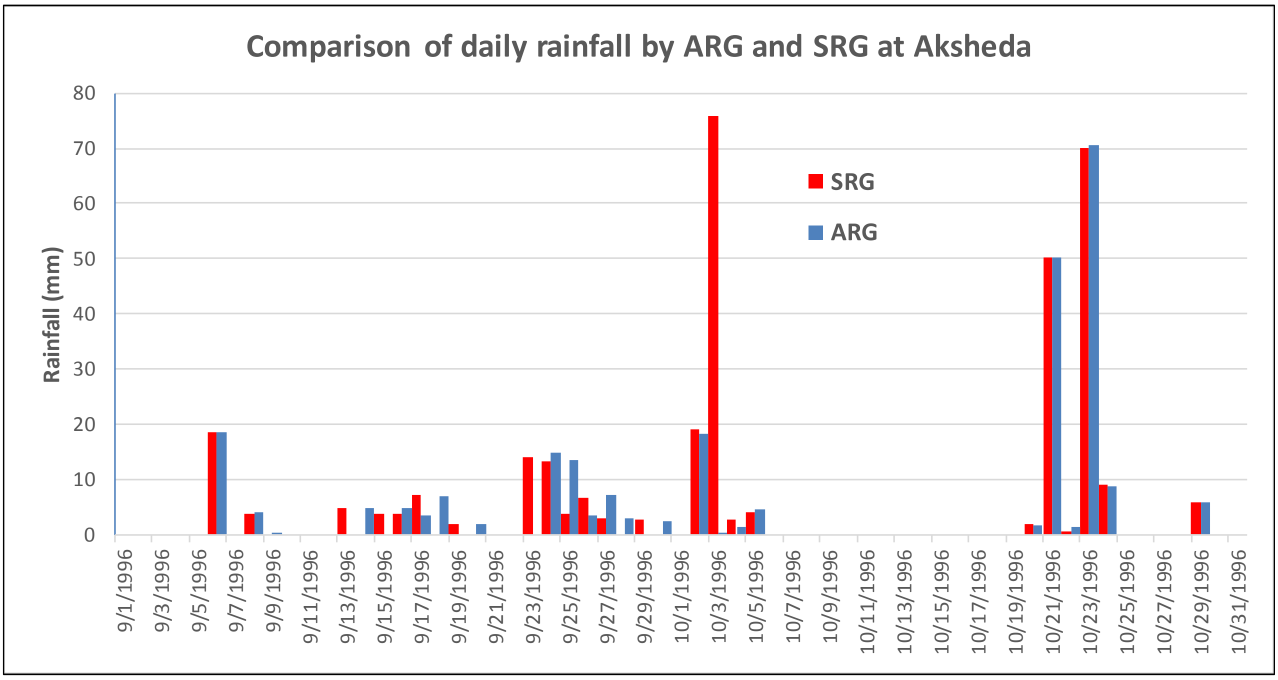


Figure 3.1: Graphical comparison of daily rainfall obtained from SRG and ARG at the same station

The following points can be noticed:

a. The difference in daily values from SRG and ARG varies considerably; from a very reasonable deviation like 1.1, 0.4, 0.7 % (on 6/9/96, 21/10/96 and 23/10/96 respectively) to unacceptably high values like 51.4, 255.3, 99.3 % (on 17/9/96, 25/9/96 and 3/10/96 respectively).

b. In this example, the resulting errors can be categorised into three major classes:

- There are many instances where a larger difference is caused by shifting of one of the data series by one day. From 13/9/96 to 31/9/96 a shift of one day in one of the series can be very clearly noticed. This shift is not present before and after this period. Such errors are not exactly due to the differences in the two observations but are the result of recording or entering one of the data series inappropriately on the wrong date. However, even if this time shift were not present, there would have been substantial differences in the corresponding values as can be easily inferred from the tabulated values.

- There are a few instances where the difference is due to mistake in recording or entering or failure of ARG. Such differences like the one on 3/10/96 where SRG record shows 75.8 mm whereas ARG data shows 0.5 mm are clear cases of mistakes. Such errors are very easy to be detected.
- There are number of instances where the percent difference is moderate to high which can be attributed to observational errors, instrumental errors and the variation in the catch in the two rain gauges. Most of these high percentage differences are for the very low rainfall values which also highlight the variation in the catch or the accuracy of equipment at such low rainfall events.

The following actions must be taken as a follow-up of data validation:

a. The cause of the shift in one of the data series can be very easily detected and removed after looking at the dates of the ARG charts and corresponding tabulated data.

b. Cause of mistake like that on 3/10/96 can be removed if ARG chart also shows comparable rainfall. If ARG data is found correct according to the chart and there does not seem to be any reason to believe instrumental failure, then the daily rainfall as reported by the SRG can be corrected to correspond to the ARG value. Else, if there is any ambiguity then the daily data has to be flagged and it has to be reviewed at the time of secondary validation on the basis of rainfall recorded at the adjoining stations.

c. Moderate to large differences (more than 5%) in the two data series are to be probed in detail by looking at the ARG chart and corresponding tabulations. Inspection of the differences in this case shows that there is no particular systematic error involved. Sometimes SRG value is more by a few units and sometimes ARG is more by similar magnitude. This might be due to observation SRG at non-standard times or incorrect tabulation of the ARG chart. At low rainfall these differences can also be due to variation in the catch or due to inaccuracy of the equipment. In both circumstances, it must be ensured whether standard equipment and exposure conditions are maintained at the station.

3.4 Checking against maximum and minimum data limits

3.4.1 General description

Rainfall data, whether daily or hourly must be validated against limits within which it is expected to physically occur. Such limits are required to be wide enough to avoid the possibility of rejecting true extreme values. For rainfall data, it is obvious that no data can be less than zero which perfectly serves as the limiting minimum value. However, it is quite difficult to assign an absolute maximum limit for the rainfall data in a given duration occurring at a particular station. Nevertheless, on the basis of past experience and physical laws governing the process of rainfall, it is possible to arrive at such maximum limits which in all probability will not be exceeded. The limit may be set as the maximum capacity of the rain gauge, but care should be taken in rejecting values on this basis where the gauge observer has read the gauge several times to ensure the gauge capacity was not exceeded.

Maximum limits also vary spatially over India with climatic region and orography. Also, this maximum limit has a strong non-linear relationship with rainfall duration. For example, for any place, the maximum limit for daily rainfall is not equal to 24 times the maximum limit for hourly rainfall. It is certainly much lower than this amount. For this reason, it is essential to set maximum limits for rainfall for durations other than 1-day. Limits for 1-day and 1-hour should be set and this will generally be sufficient to identify gross errors over the intervening range of duration.

For 1-day duration, the India Meteorological Department and Indian Institute of Tropical Meteorology have prepared atlas for 1-day Probable Maximum Precipitation (PMP) which gives the expected maximum amount that can physically occur in a given duration at a given location. Values extracted from this map should be applied or else could be derived as the historical maximum value from long term records now available for most regions. There might be some variation in the values obtained from both of these atlases, but such differences may be ignored for the purpose of prescribing the maximum limits. Alternatively, the derived information on observed maximum 1-Day point rainfall, which is available for scores of stations across the country from long term records, can be used as a reasonably good estimate of the maximum limit of rainfall.

Similarly, maps showing 50 year – 1-hour maximum rainfall developed by India Meteorological Department can be used for prescribing the maximum data limit for the case of hourly rainfall data. Such initial estimates can be adjusted using local judgement adjusted on the basis of experience or of local research studies based on either:

- Storm maximisation by considering precipitable moisture and inflow of moist air
- Statistical analysis of observed extreme values for shorter durations.

3.4.2 Data validation procedure and follow up actions

Setting minimum and maximum limits ensures filtering of values outside the specified limits. Such values are considered suspect. They are first checked against manuscript entries and corrected if necessary. If manuscript and entry agree and fall outside prescribed limits, the value is flagged as doubtful. Where there are some other corroborative facts about such incidents, available in manuscript or notes of the observer or supervisor, they must then be incorporated with the primary data validation report. This value has to be probed further at the time of secondary data validation when more data from adjoining stations become available.

When the data being entered exceed the prescribed limits while the high rainfall events have been experienced by the staff and recorded by other nearby stations, the maximum limit is reset to a suitable higher value. If there is no justifiable basis for setting the new maximum value, the new value is reported in the form of a remark which can be reviewed at the secondary validation stage.

Example 3-2

Consider the long-term daily rainfall at Megharaj station in Kheda catchment as shown in Figure 3.2

This is a long-term rainfall data of the station (37 years) and it can be seen that the maximum daily rainfall in any year has usually been about 100 mm on average. The daily rainfall has been more than 150 mm a few times in this period, and the value exceeded 200 mm three times in the 3-year period. Only once have the values exceeded 250 mm. Setting the value of about 320 – 325 mm as the maximum limit for daily rainfall at this station can then help in future data validation. The values of arbitrary locations can be derived from the isoclines of maps giving observed maximum 1-Day rainfall or 1-Day PMP values.

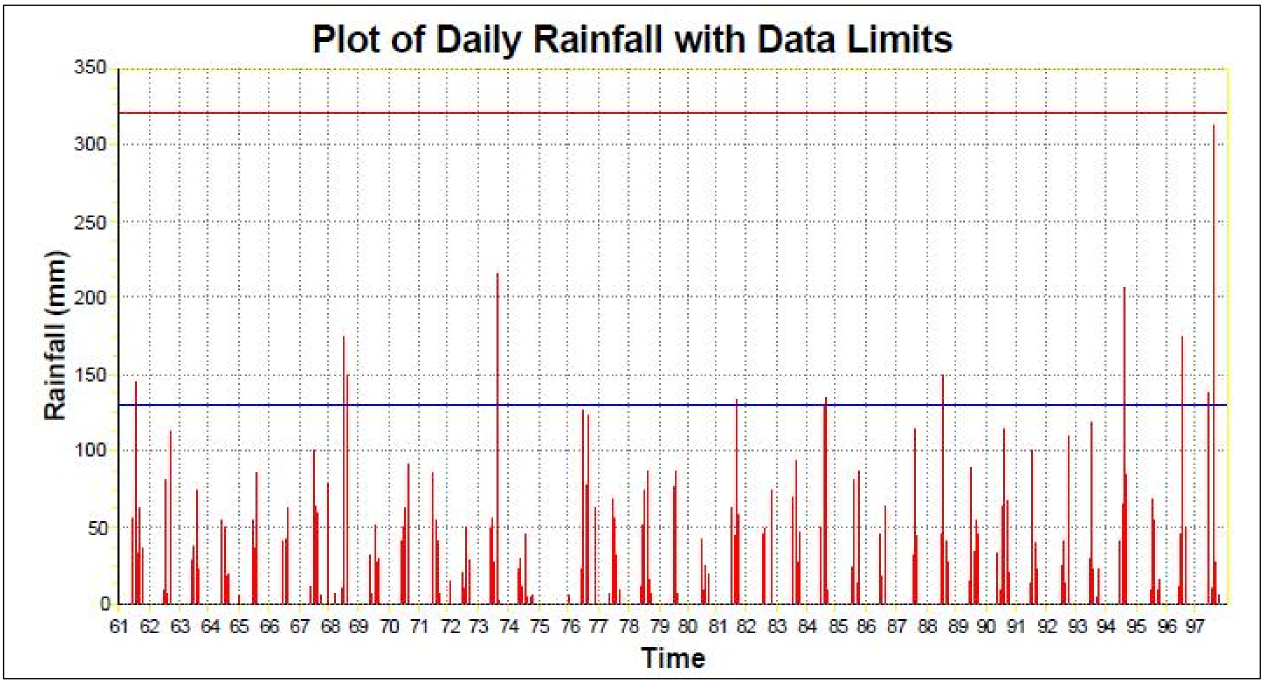


Figure 3.2: Physical significance of maximum limit and upper warning level

3.5 Checking against upper warning level

3.5.1 General description

Validation of rainfall data against an absolute maximum value does not detect possible frequent occurrence of erroneous data which are below the prescribed maximum limit. Because of this, it is advantageous to consider one more limit, called the upper warning level, which can be employed to see if any of the data value has violated it. This limit should take into account the seasonal character of daily rainfalls, by subdividing them in monthly sub-sets, and by calculating the daily data statistics for each month individually. The warning levels should then be set to the mean monthly values plus 1.96 times the standard deviation for a particular month, with the intention of flagging the high data values which are not expected to occur frequently. The underlying purpose of carrying out such a test is to consider a few high data values with suspicion and subsequent scrutiny. Other limits can be used instead that are based on the statistical analyses (e.g. 98 percentile) of daily data that belong to particular months.

Similar statistic can be employed for obtaining suitable value for upper warning level for hourly rainfall data. The central idea while setting these upper warning levels is that the higher rainfall data is screened adequately, that is the limits must be such that it results in not too many and not too few data values being flagged for validation.

3.5.2 Data validation procedure and follow up actions

Setting warning limits in the Primary module results in filtering values outside the specified range. Values are checked against manuscript entries and corrected if necessary. Remaining values are flagged as doubtful, and any associated field notes or corroborative facts are incorporated with the primary validation report and forwarded to the Divisional Data Processing Centre for secondary validation.

Example 3-3

Data from the Purna at Lakhpouri station (198-2016) was used in this example. Daily data were sub-divided into monthly bins, based on the month to which they belong, and the two statistics were calculated, as shown in Table 3.2 below.

Table 3.2: Example of calculation of the upper warning data limit

Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
98 Percentile	4.80	2.00	7.06	1.40	5.26	49.61	59.30	55.84	44.88	18.28	4.97	0.00
mean+1.96 StDev.	6.37	4.02	7.27	1.95	5.16	30.85	36.83	37.02	27.13	22.17	9.89	4.49
StDev.	3.06	1.95	3.46	0.93	2.45	13.46	15.29	15.70	11.55	10.37	4.78	2.20
Mean	0.37	0.20	0.49	0.14	0.36	4.47	6.85	6.25	4.50	1.83	0.52	0.18

4 SECONDARY VALIDATION OF RAINFALL DATA

4.1 General

It is presumed that the rainfall data received at divisional offices have already received primary validation on the basis of knowledge of instrumentation, conditions at the field station and information contained in Field Record Books. Secondary Validation consists of a set of functions aimed to identify suspect values by comparison with neighbouring stations, as shown in Figure 4.1. Some of the checks which can be made are oriented towards specific types of errors known to be made by observers, whilst others are general in nature and lead to identification of spatial inconsistencies in the data.

Secondary Validation is mainly carried out on a Division level. However, since comparison with neighbouring stations is limited by Divisional boundaries, the validation of some stations near the Divisional boundaries will have to await compilation of data at the State Data Processing Centre.

Rainfall poses special problems for spatial comparisons because of the limited or uneven correlation between stations. When rainfall is convectional in type, it may rain heavily at one location while another location may remain dry only a few miles away. Over a month or monsoon season such spatial unevenness tends to get smoothened out and aggregated totals are much more closely correlated.

Spatial correlation thus depends on:

- Duration (smaller at shorter duration)
- Distance (decreasing with distance)
- Type of precipitation
- Physiographic characteristic of a region

Correlation structure inherent in the data can be determined on the basis of historical rainfall data for different durations. A study for determining such correlation structures for yearly duration for the entire country has been made *(Upadhyay et al.,1990) Mausam 41, 4, 523-530*). In this study, the correlation field has been determined for 21 meteorological homogeneous regions which cover almost the entire country using 70 years of data (1900 - 1970) and about 2000 stations. However, for the purpose of data validation on an hourly or daily basis are not readily available. It will be possible to determine such structures on the basis of available rainfall data.

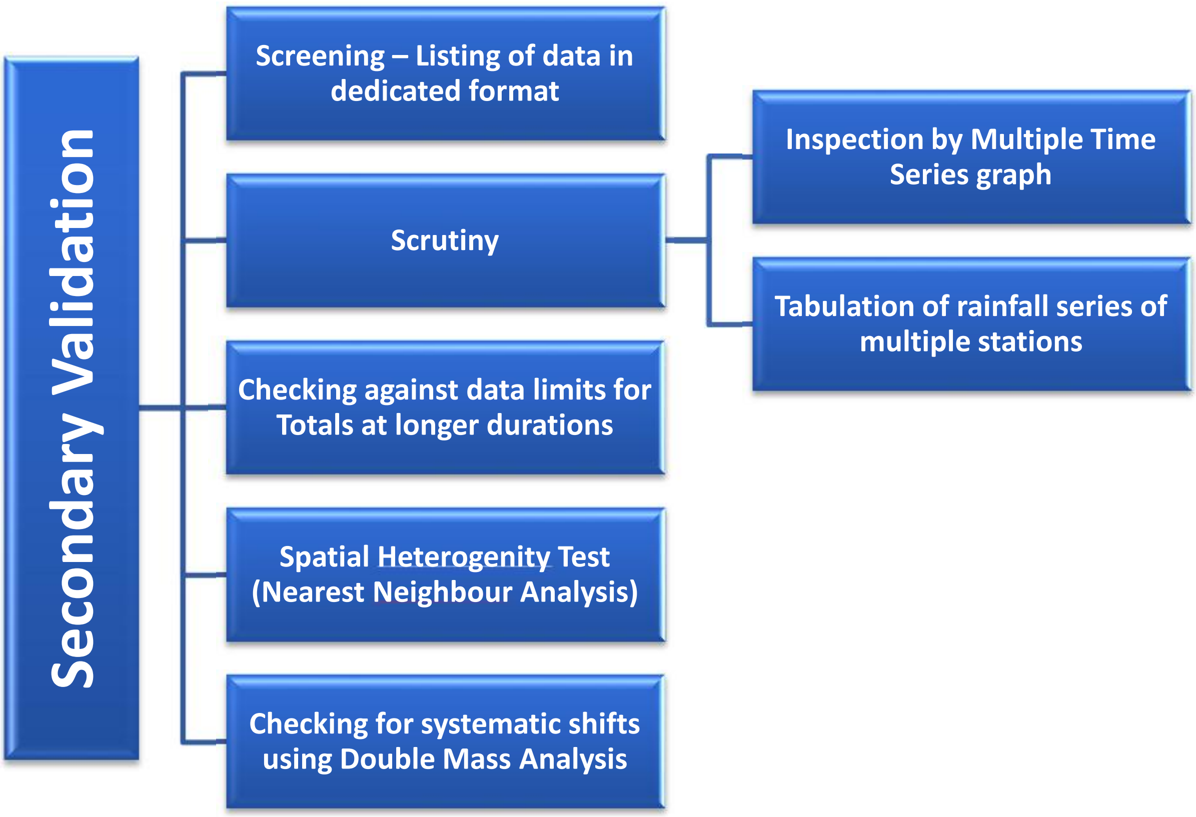


Figure 4.1: Processes involved in secondary validation

Example 4-1

The effect of aggregation of data to different time interval lengths and that of the inter-station distances on the correlation structure is illustrated below.

The scatter plot of correlation between various rainfall stations of the Kheda catchment for the daily, ten daily and monthly rainfall data is shown in Figure 4.2, Figure 4.3 and Figure 4.4, respectively.

From the corresponding correlation for same distances in these three figures, it can be noticed that aggregation of data from daily to ten daily and further to monthly level increases the level of correlation significantly. At the same time, it can also be seen that the general slope of the scatter points become flatter as the aggregation is done. This demonstrates that the correlation distance for monthly interval is much more than that for ten-day interval. Similarly, the correlation sharply reduces for the case of daily time interval.

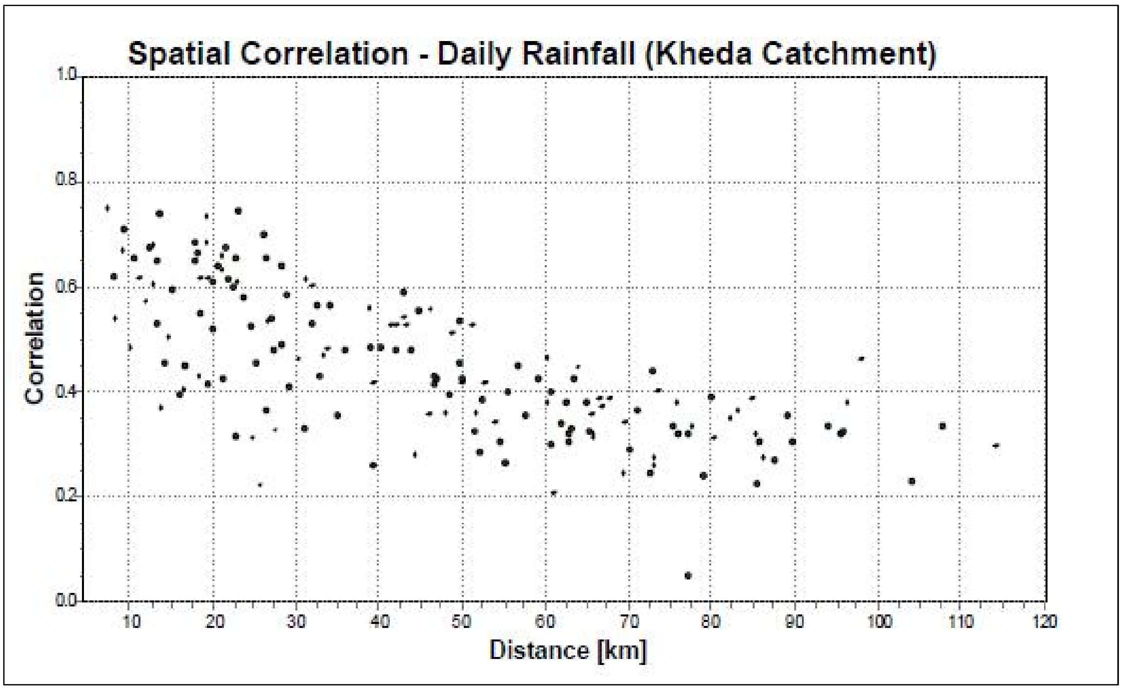


Figure 4.2: Plot of correlation with distance for daily rainfall data

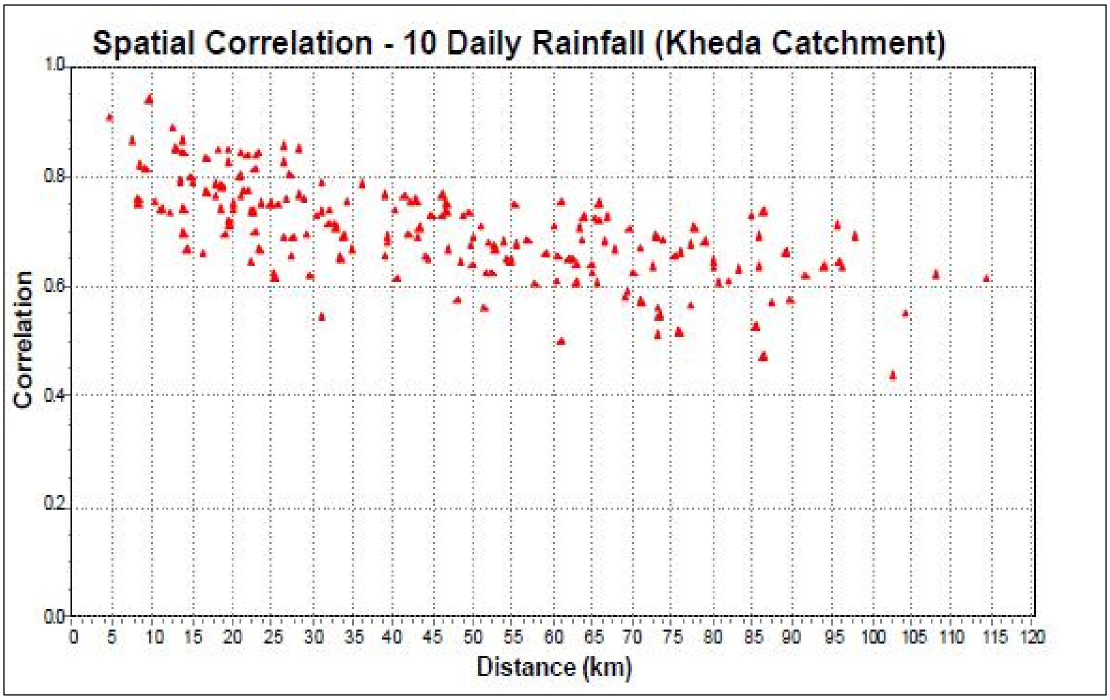


Figure 4.3: Plot of correlation with distance for ten-daily rainfall data

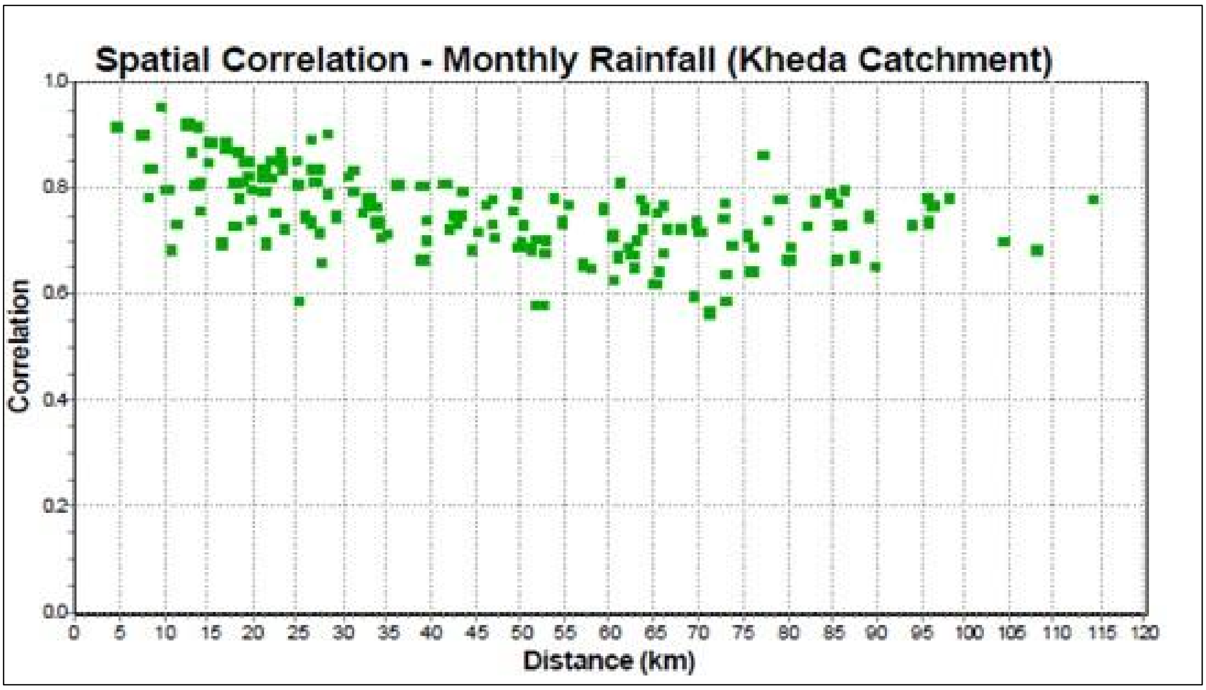


Figure 4.4: Plot of correlation with distance for monthly rainfall data

Example 4-2

The effects of physiographic characteristics over the correlation structure is illustrated by considering monthly rainfall for two groups of stations in the Pargaon catchment.

Figure 4.5 shows the scatter plot of the correlation among some 20 stations in small hilly region (elevations ranging from 700 m to 1250 m) in the lower left part of the catchment (see Figure 4.6). This small region can be considered as homogeneous, which is also substantiated by the scatter plot of the correlation. Monthly rainfall data has been considered for this case and it is clear from the plot that there is a very high level of correlation among stations. The general slope of the scatter diagram indicates a high value of the correlation for higher distances.

However, Figure4.7 shows the scatter plot of the correlation among monthly rainfall at some 34 stations in a region which includes the hilly region together with an extended portion in the plain region (the plains ranging from 700 m to 600 m with very low and scattered hills in between) of the catchment (see Figure 4.8).

It is apparent from Figure4.7 that when a few stations from the hilly region and another lot from the adjoining plain region are analysed together, then the resulting correlation plot shows a weaker correlation structure. The correlation decays very quickly against the distance and even for shorter distances it is becomes diffused. In fact, the level of variability for the group of stations in the hilly region is much lower than that of the remaining stations in the plain region. This is what is exhibited by Figure4.7 in which lot of scatter is shown even for smaller inter station distances.

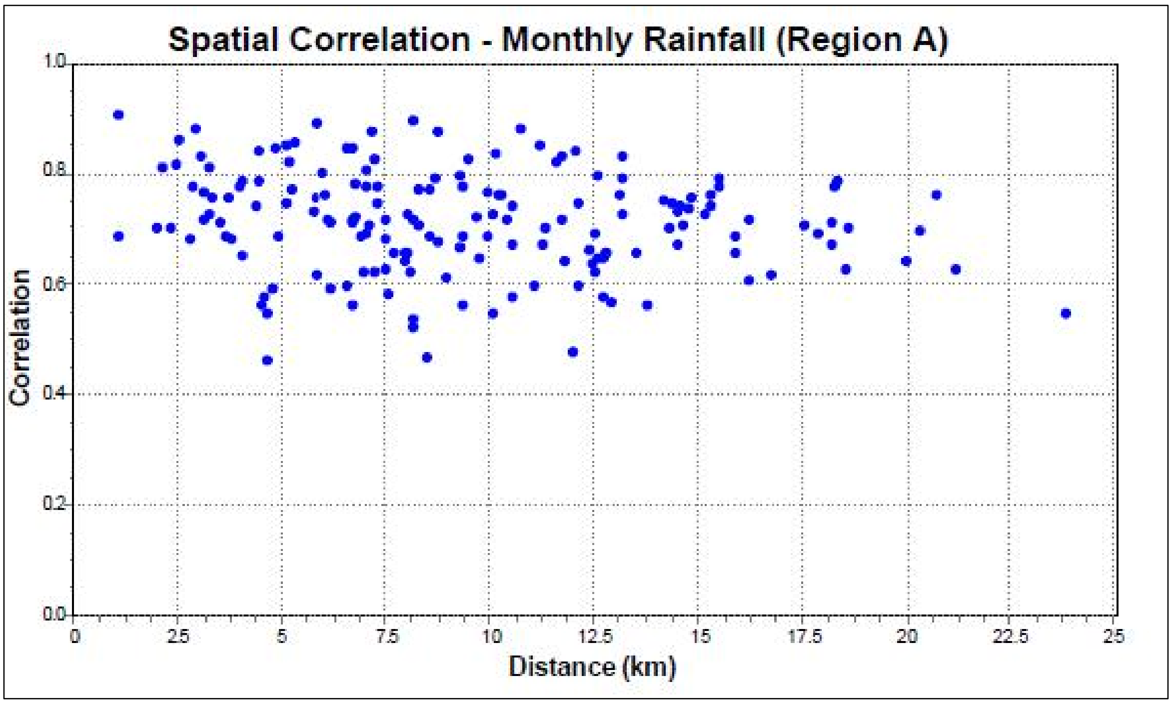


Figure 4.5: Scatter plot of correlation for monthly rainfall in the small hilly region

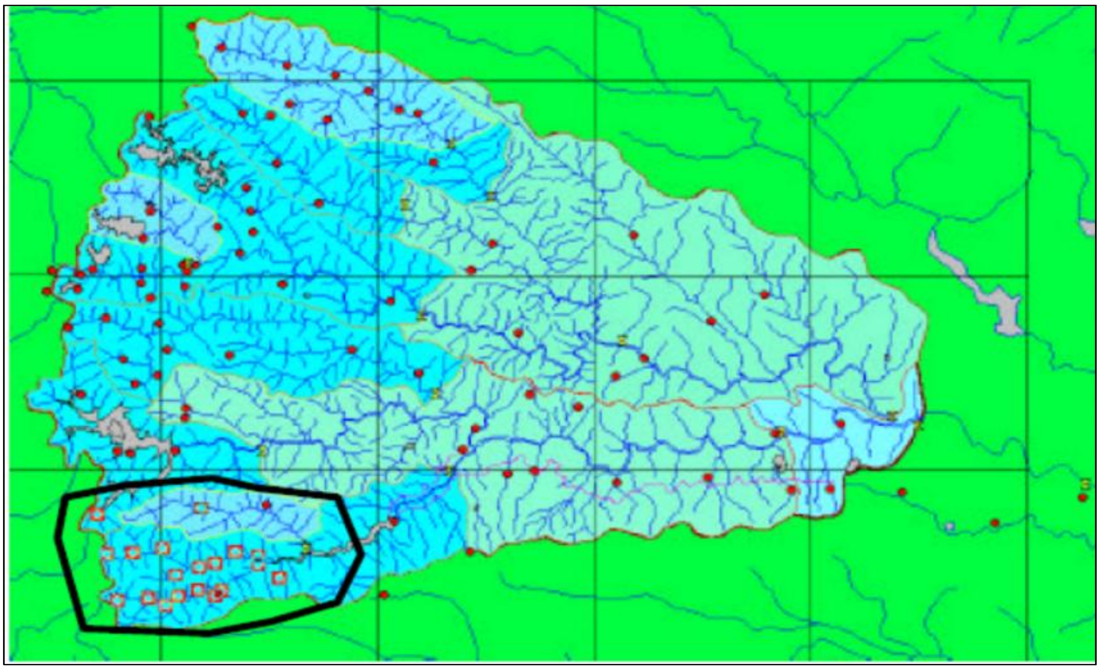


Figure 4.6: Selection of group of some 20 stations in the hilly region of the catchment

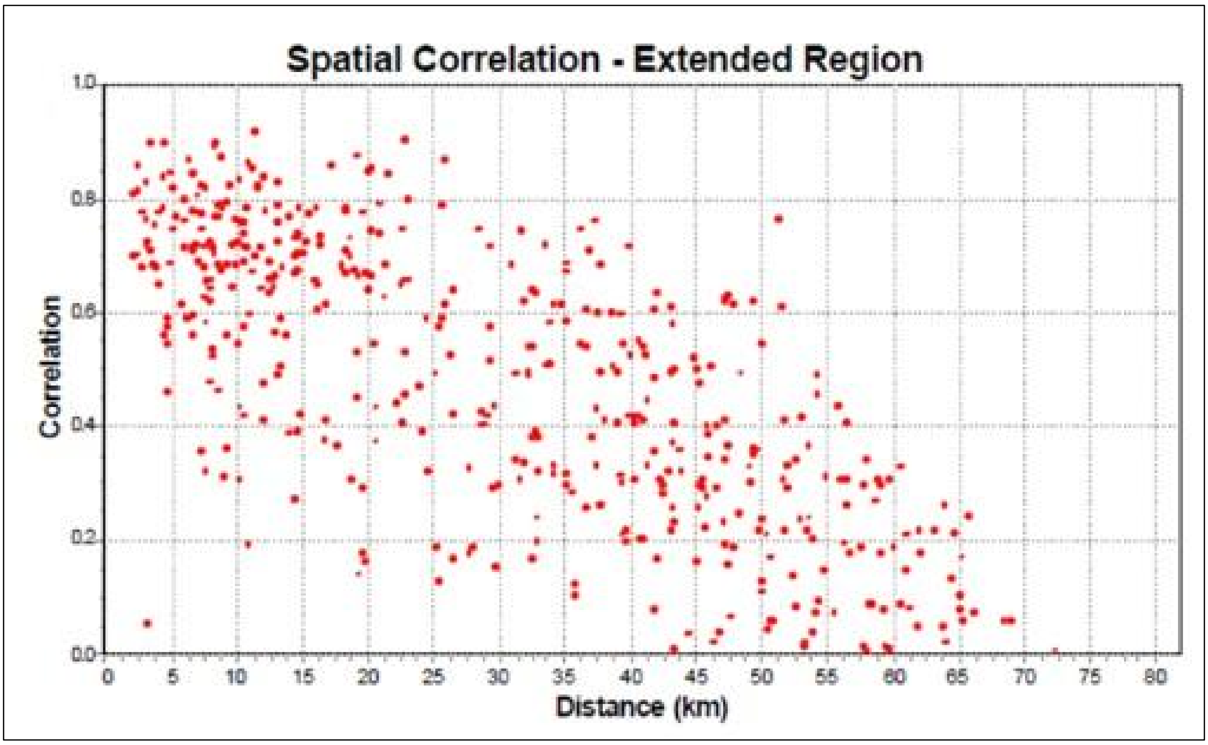


Figure4.7: Scatter plot of correlation for monthly rainfall in the extended region

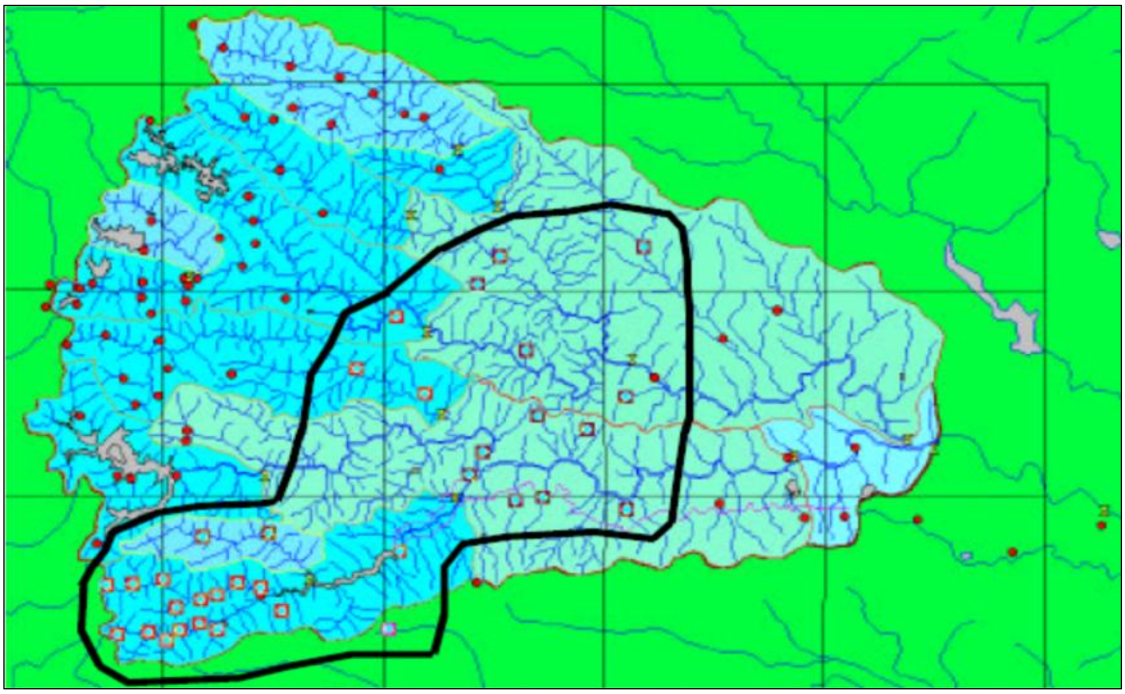


Figure 4.8: Selection of a group of some 34 stations in the extended region of the catchment

Note: Spatial correlation can be used as a basis for spatial interpolation and correction. However, there is a danger of rejecting good data which is anomalous as well as accepting bad data. A balance must be struck between the two. In considering this balance, it is good practice to give weights to the previous performance of the station and the observer.

One must particularly be wary of rejecting extreme values, as the true extreme values are the most interesting and useful for design purposes. True extreme values (like the false ones) will often be flagged as suspect by validation procedures. Before rejecting such values, it is advisable to refer both to field notes and to confer with Sub-divisional staff.

The data processor must continue to be aware of field practice and instrumentation and the associated errors which can arise in the data, as described in Chapter 03: Validation of rainfall data.

4.2 Screening of data series

After the data from various Sub-Divisional offices have been received at the respective Divisional office, they are organised and imported into the temporary databases of secondary data validation module of assigned data processing software.

The first step towards data validation is making the listing of data for various stations in the form of a dedicated format. Such listing involves, for daily rainfall data, flagging of all those values which are beyond the maximum data limits or the upper warning level. It also prepares the data in a well-organised matrix form in which various months of the year are given as separate columns and various days of the month are given as rows. The monthly and yearly basic statistics like total rainfall, maximum daily rainfall, number of rainy days etc. are listed at the bottom of the table. The number of instances where the data is missing or has violated the data limits is also given.

This listing of the screening process and basic statistics are very useful in seeing whether the data has come in the databases in a desired manner or not, and whether there is any marked inconsistency vis-à-vis the expected hydrological patterns.

Example 4-3

An example of the listing of a screening process for the Dhalegaon station of Godavari catchment for the year 2010 is given in Table 4.1 The flagging of a few days of high rainfall shows that these values have crossed the Upper Warning Level. Such flagged values can then be subsequently attended to when comparing with adjoining stations. This particular year shows a few days of very heavy rainfall, one in fact making the recorded maximum daily rainfall (i.e. 312 mm on 27 July). Monthly and yearly statistics are also viewed for appropriateness.

Table 4.1: Result of the screening process of daily rainfall data for one year

Day	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
1	0.0	0.0	192.5*	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
2	0.0	0.0	15.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
3	0.0	0.0	1.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
4	0.0	0.0	0.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
5	0.0	0.0	0.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
6	0.0	0.0	0.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
7	0.0	0.0	1.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
8	0.0	0.0	32.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
9	0.0	0.0	1.0	25.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
10	0.0	0.0	0.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
11	0.0	0.0	0.0	14.5	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
12	0.0	0.0	7.0	1.5	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
13	0.0	0.0	1.0	4.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
14	0.0	0.0	0.5	0.5	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
15	0.0	0.0	1.0	1.0	5.5	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
16	14.0	0.0	0.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
17	0.0	0.0	0.5	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
18	0.0	0.0	0.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
19	0.0	10.0	12.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
20	0.0	0.0	1.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
21	0.0	2.0	6.5	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
22	0.0	1.0	0.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
23	12.0	0.0	9.5	2.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
24	9.0	0.0	125.5	27.5	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
25	138.0*	1.0	11.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
26	132.0*	4.0	54.5	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
27	38.0	312.0*	1.0	0.0	0.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
28	54.0	32.5	0.0	0.0	0.0	-99.0*	-99.0*	-99.0*		-99.0*	-99.0*	-99.0*
29	0.0	4.5	0.5	0.0	0.0	-99.0*	-99.0*	-99.0*		-99.0*	-99.0*	-99.0*
30	0.0	12	0.5	0.0	0.0	-99.0*	-99.0*	-99.0*		-99.0*	-99.0*	-99.0*
31	22.0	0.0	0.0	0.0	0.0	-	-99.0*	-99.0*		-99.0*		-99.0*
Data	30.0	31.0	31.0	30.0	31.0	30.0	31.0	31.0	28.0	31.0	30.0	31.0
Eff.	30.0	31.0	31.0	30.0	31.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Miss	0.0	0.0	0.0	0.0	0.0	30.0	31.0	31.0	28.0	31.0	30.0	31.0
Sum	397.0	401.0	474.5	76.0	5.5	-	-	-	-	-	-	-
Mean	13.2	12.9	15.3	2.5	0.2	-	-	-	-	-	-	-
Min.	0.0	0.0	0.0	0.0	0.0	-	-	-	-	-	-	-
Max.	138.0	312	192.5	27.5	5.5	-	-	-	-	-	-	-
High	130	130	130	130	130	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Num	2.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Low	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Num	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Summary

Annual	values:				
Data	365.0	Sum	1354.0	Minimum	0.0
Effective	153.0	Mean	8.8	Maximum	312.0
Missing	212.0				

Exceedance of:

- Lower bound (0.00) marked with *
- Upper bound (130.00) marked with *
- Rate of rise (320.00) marked with +
- Rate of fall (320.00) marked with –
- Missing data marked with -99.0

4.3 Scrutiny by multiple time series graphs

Inspection of multiple time series graphs may be used as an alternative to inspection of tabular data. Some processors may find this a more accessible and comprehensible option. This type of validation can be carried out for hourly, daily, monthly and yearly rainfall data. The validation of compiled monthly and yearly rainfall totals helps in bringing out those inconsistencies which are either due to a few very large errors or due to small systematic errors which persist unnoticed for much longer durations. The procedure is as follows:

- Choose a set of stations within a small area with an expectation of spatial correlation
- Include, if possible, in the set one or more stations which historically have been more reliable
- Plot rainfall series as histograms stacked side by side and preferably in different colours for each station. Efficient comparison on the magnitudes of rainfall at different stations is possible if the individual histograms are plotted side by side. On the other hand, a time shift in one of the series is easier to detect if plots of individual stations are plotted one above the other.
- After inspection for anomalies and comparing with climate, all remaining suspect values are flagged, and comment inserted as to the reason for suspicion.

Example 4-4

Consider that a few of the higher values at Anior station of Kheda catchment during July and August 1996 are suspect. Comparison with adjoining available stations Bhempoda, Rellawad and Megharaj is made for this purpose. Figure 4.9 gives the plot of daily rainfall for these multiple stations during the period under consideration.

It may be noticed that rainfall of about 165 mm and 70 mm are observed at Anior and Bhempoda stations which are virtually no more than 5 km apart. Such variation is possible but rare, and they can be identified for events to cross check with other information. On checking with the hourly observations available at Anior station, it was noticed that the compiled daily rainfall is only 126 mm. This substantiates the earlier suspicion of it being comparatively larger.

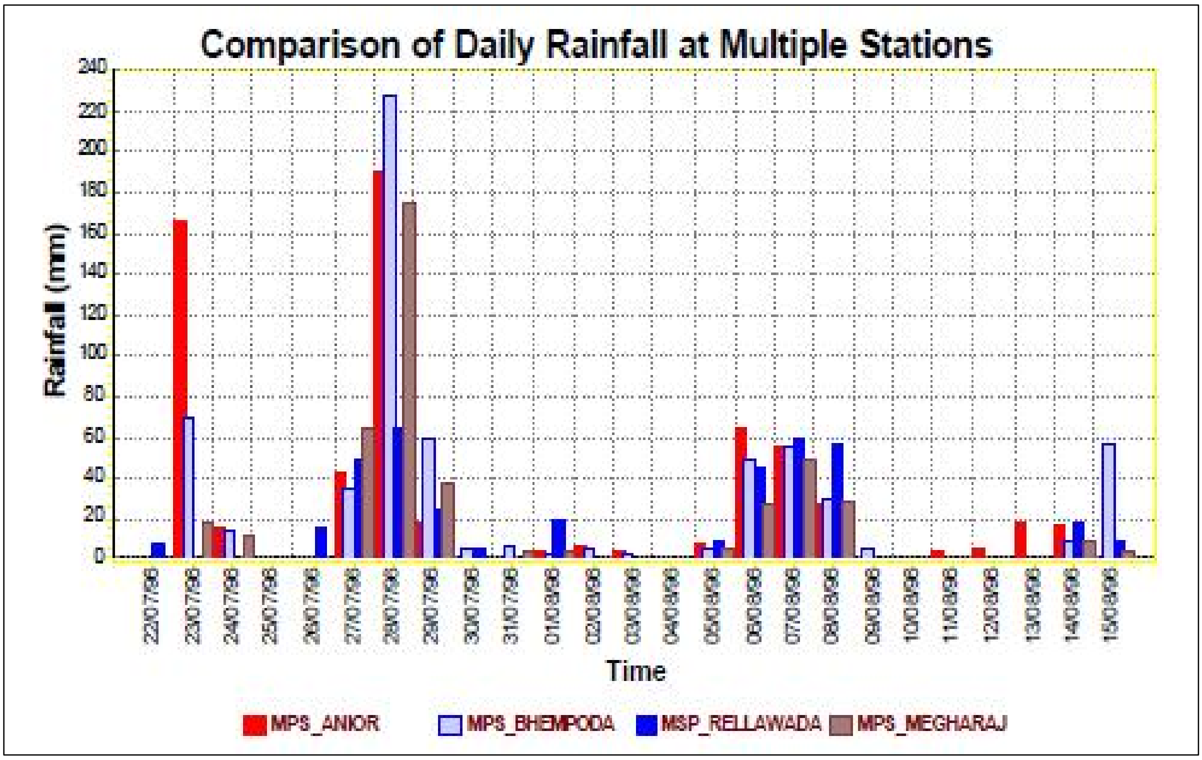


Figure 4.9: Comparison of multiple time series plot of daily rainfall data

Further it may be noticed from the plot that the daily rainfall for 12th and 13th August at Anior seems to be shifted ahead by a day. This shifting is also confirmed when the ARG record is compared with the SRG record. The time shifting error is clearly in the SRG record of Anior station. Thus, inspection of the record sheets, visit to site and interaction with the observer can be helpful in getting more insight into the probable reasons of such departures.

4.4 Scrutiny by tabulations of daily rainfall series of multiple stations

In the case of rainfall (unlike other variables), a tabular display of daily rainfall in a month, listing several stations side by side can reveal anomalies which are more difficult to see on multiple time series graphs (see Figure 4.9), plotted as histograms. Scanning such tabular series will often be the first step in secondary data validation. The following questions may be posed when looking for anomalies:

- Do the daily blocks of rainy days generally coincide in start day and finish day?
- Are there exceptions that are misplaced, starting one day early or late?
- Is there a consistent pattern of misfit for a station through the month?
- Are there days with no rainfall at a station when (heavy) rainfall has occurred at all neighbouring stations?

Field entry errors to the wrong day are particularly prevalent for rainfall data and especially for stations which report rainfall only. This is because rainfall occurs in dry and wet spells and observers may fail to record the zeros during the dry spells and hence lose track of the date when the next rain arrives. When ancillary climate data are available, this may be used to compare with rainfall data. For example, a day with unbroken sunshine in which rain has been reported suggests that rainfall has been reported for the wrong day. However, most comparisons are not so clear cut and the processor must be aware that there are a number of possibilities:

- Rainfall and climate data both reported on the wrong day - hence no anomaly between them but discrepancy with neighbouring stations
- Rainfall data only on the wrong day - anomalies between rainfall and climate and between rainfall and neighbouring rainfall
- Rainfall and climate both reported on the correct day - the anomaly was in the occurrence of rainfall. For example, no rainfall at one site but at neighbouring sites. In this case climatic variables are likely to have been shared between neighbouring stations even if rainfall did not occur.

Example 4-5

As a routine process of scrutinising daily data for a common error of time shift in one or more data series, consider Kapadwanj, Kathlal, Mahisa, Savlitank and Vadol stations of Kheda catchment. These stations are within a circle of 25 km diameter and thus are expected to experience similar average rainfall.

For easy scrutiny of the data series for possible time shift in one or more series the data series are tabulated side by side as shown in Table 4.2 for the period of the 1st August to the 21st August 1984. A very casual look at this tabulation reveals that there is very high possibility of a one-day time shift in the data of Savlitank station. Data series of Savlitank station appears to be having a lag of one day in consecutive rainfall events. Exactly the same shift is persisting for all 21 days and is confirmed by closely looking at the start and end times of five rainfall events (highlighted by underlining) one after another.

Note: identification of a possible time shift must be followed by first a closer look at the manuscript record to see if the shift has been made during entering or managing the data series. If it is found that the shift has been due to data handling during or after data entry, then it is corrected accordingly. If the manuscript record also shows the same series then the observer can be asked to tally it from the field note book. The feedback from the observer will help in settling this type of discrepancy and also will encourage observer to be careful subsequently.

Table 4.2: Scrutiny of possible error in the timing of daily rainfall data

Tabulation of series from 1/8/1984 to 21/8/1984

Date	Kapadwanj	Kathlal	Mahisa	Savlitank	Vadol
8/1/1984	0.00	0.00	0.00	0.00	0.00
8/2/1984	0.00	0.00	0.20	0.00	0.00
8/3/1984	152.40	99.30	157.40	0.00	39.30
8/4/1984	104.10	50.20	87.00	150.00	59.20
8/5/1984	7.70	12.00	18.00	76.00	13.10
8/6/1984	1.50	35.00	0.00	16.00	0.00
8/7/1984	0.00	0.00	0.00	3.00	0.00
8/8/1984	1.30	0.00	0.00	0.00	0.00
8/9/1984	0.00	13.00	0.00	0.00	0.00
8/10/1984	231.20	157.00	179.00	0.00	17.30
8/11/1984	43.20	18.30	64.00	201.00	63.20
8/12/1984	0.00	0.00	0.00	26.00	33.3
8/13/1984	0.00	0.00	0.00	0.00	13.10
8/14/1984	0.00	0.00	20.00	0.00	0.00
8/15/1984	0.00	0.00	0.00	30.00	0.00
8/16/1984	2.60	8.30	16.50	0.00	16.30
8/17/1984	0.00	0.00	0.00	20.0	20.20
8/18/1984	32.00	50.30	25.60	0.00	37.20
8/19/1984	16.50	8.20	15.00	27.00	19.30
8/20/1984	0.00	0.00	0.00	13.00	0.00
8/21/1984	0.00	0.00	0.00	0.00	0.00

It is clearly visible from the tabular data that entry of Savlitank rainfall data is shifted by a day.

4.5 Checking against data limits for totals at longer durations

4.5.1 General description

Many systematic errors are individually so small that they cannot easily be noticed. However, since such errors are present till suitable corrective measures are taken, they tend to accumulate with time and therefore tend to be visible more easily. Also, some times when the primary data series (e.g. daily rainfall series) contain many incorrect values frequently occurring for a considerable period (say a year of so), primarily due to negligence of the observer or at the stage of handling of data with the computer, then the resulting series compiled at larger time interval also show the possible incorrectness more visibly. Accordingly, if the observed data are accumulated for longer time intervals, the resulting time series can again be checked against the corresponding expected limits. This check applies primarily to daily rainfall at stations at which there are no recording gauges.

4.5.2 Data validation procedure and follow up actions

Daily data are aggregated to monthly and yearly time intervals for checking if the resulting data series are consistent with the prescribed data limits for such time interval.

Together with the upper warning level or maximum limit for monsoon months and yearly values, the use of lower warning level data limit can also be made to see if certain values are unexpectedly low and thus warrant a closer look. Aggregated values violating the prescribed limits for monthly or annual duration are flagged as suspect and inappropriate.

Remarks are made in the data validation report stating the reasons for such flagging. These flagged values must then be validated on the basis of data from adjoining stations.

Example 4-6

The daily data of Vadol station (in Kheda catchment) is considered and the yearly totals are derived. The period from 1970 to 1997 is taken for compilation wherein two years of data, i.e. 1975 & 1976, are missing.

The plot of these yearly values is shown in Figure 4.10. In this case of yearly rainfall data, the values can be validated against two data limits as the upper and lower warning levels. The values of such limits can be drawn from the statistical distribution of the yearly rainfall in the region. In this case, the mean of the 26 yearly values is about 660 mm with a standard deviation of 320 mm with a skewness of 0.35. With an objective of only flagging a few very unlikely values for the purpose of scrutiny, a very preliminary estimate of the upper and lower warning levels is arbitrarily obtained by taking them as:

Lower warning level = mean – 1.645 × (standard deviation) = 660 – 1.645 × 320 = 133.6 mm [assuming a normal distribution, even though rainfall values seldom follow it]

Upper warning level = mean + 1.96 × (standard deviation) = 660 + 1.96 × 320 = 1287.2 mm [assuming a normal distribution again]

The multipliers to the standard deviation for lower and upper warning levels have been taken differently, as it is in general more plausible to get higher than lower rainfall values. It is important to retain rainfalls that are capable to cause floods. Such limits can be worked out on a regional basis on the basis of the shape of distribution and basically with the aim to demarcate highly unlikely extremes. An alternative procedure to determine the upper and lower warning level for total annual precipitation can be accomplished by using the previously validated annual data to construct a probability plot based on the Weibul plotting position formula (probability = m/(n+1) where m is the rank and n is the number of data points, and determine the 5% and 95% thresholds as the minimum and maximum warning levels.

The Upper and Lower warning limits based on the formulas the use the mean and the standard deviation statistics have been shown in the plot of yearly values and it may be seen that there are a few instances where the annual rainfall values come very close or go beyond these limits. For example, in year 1997 a high value of more than 1329 mm is reported and similarly for year 1974 the reported rainfall is as low as 92.6 mm (years 1975 and 1976 have missing data). The use of secondary information (e.g. rainfall catch at nearby stations or records of the downstream floods) is also suggested to verify the occurrence of data outliers, which can suggest which historic years should be examined in more detail on a daily basis.

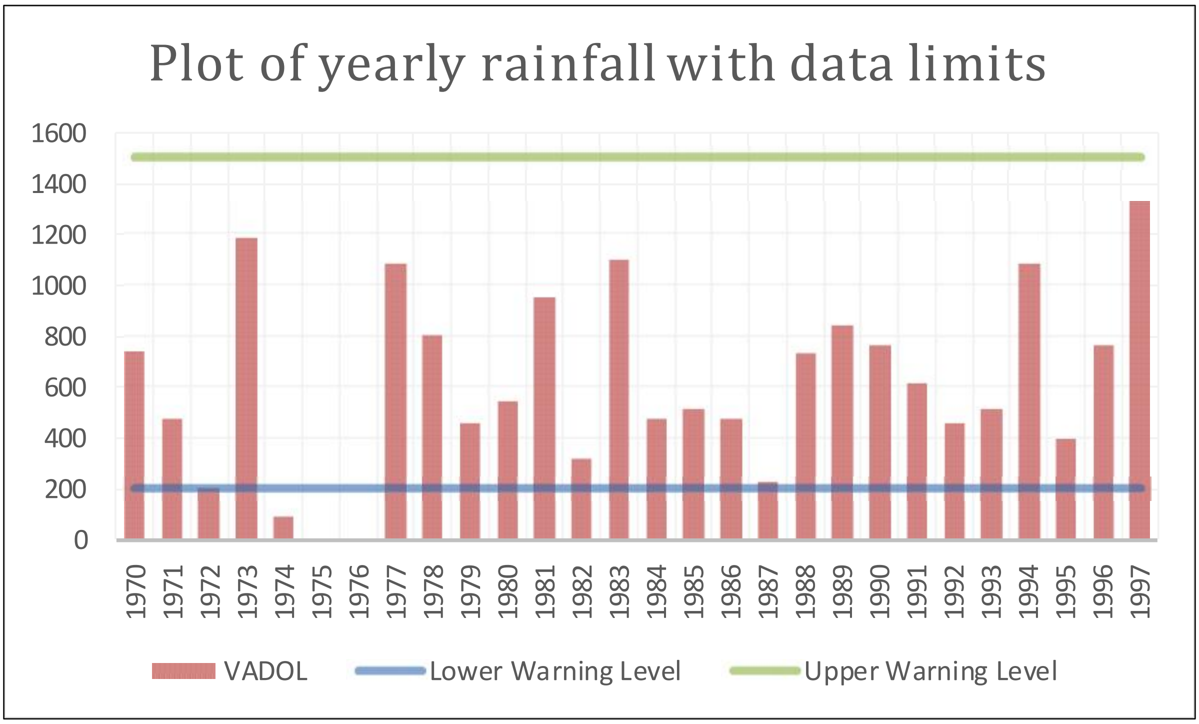


Figure 4.10: Plot of rainfall data compiled at yearly interval

After screening such instances of extreme values in the data series compiled at longer time intervals, it is essential that for such instances the values reported for the station under consideration is compared with that reported at the neighbouring stations. For this, the yearly data at five neighbouring stations including the station under consideration, i.e. Vadol, is tabulated together as in Table 4.3 for easy comparison.

Table 4.3: Tabulation of yearly rainfall at five neighbouring stations

Year	Balasinor	Kapadwanj	Savlitank	Vadol	Vagharoli
1970	802.80	927.20	-99.00	739.80	-99.00
1971	546.70	569.50	-99.00	475.00	-99.00
1972	338.20	291.00	-99.00	198.20	-99.00
1973	1061.20	1305.00	1226.00	1186.40	1297.40
1974	338.10	421.00	268.50	92.60	-99.00
1975	-99.00	-99.00	-99.00	-99.00	-99.00
1976	-99.00	-99.00	-99.00	-99.00	-99.00
1977	1267.20	1217.50	1168.90	1083.50	1575.80
1978	672.80	507.50	517.00	801.40	1347.00
1979	437.50	428.50	525.50	455.60	1197.00
1980	551.30	661.60	378.00	545.70	892.00
1981	917.70	1273.60	1004.00	950.70	722.00
1982	302.10	540.20	376.00	320.10	267.00
1983	1028.00	1088.50	1020.00	1099.10	1110.00
1984	523.10	882.90	888.00	475.10	649.60
1985	438.90	661.50	1101.00	510.80	1173.00
1986	526.90	474.90	256.00	470.70	505.00
1987	257.00	256.00	209.00	227.50	232.00
1988	-99.00	1133.00	826.00	734.50	849.40
1989	1088.00	1064.00	787.00	840.80	-99.00
1990	1028.10	971.00	1042.00	761.00	1174.00
1991	451.00	815.00	523.00	618.10	628.00
1992	421.10	1028.00	469.00	459.60	606.00
1993	531.00	410.50	781.00	512.80	781.00
1994	1085.00	1263.00	1039.00	1083.30	1332.00
1995	590.00	528.00	422.00	399.60	525.00
1996	1397.00	968.00	760.00	762.60	1050.00
1997	1272.00	1876.00	1336.20	1329.00	950.00

Missing values: -99.00

It may be seen from this table that for year 1997 the reported rainfall is very high at most of the neighbouring stations and is about 1876 mm for Kapadwanj station. At two other stations, it is in the range of 1200 to 1300 mm except that for Vagharoli, where it is only 950 mm for this year. Thus, as far as the suspect value of 1329 mm at Vadol station is concerned, the suspicion may be dropped in view of the high values reported nearby. Comparison for the year 1974 shows that although all the stations seem to have experienced comparatively lower amount of rainfall (about 340, 420 and 270 mm), the rainfall at Vadol station is extremely low (i.e. 92.6 mm). Such a situation warrants that the basic daily data for this test station must be looked more closely for its appropriateness.

The 1974 data given in Table 4.3 is shown in more detail for the selected period in May on a daily basis in Table 4.4.

Though there are comparatively more zeros reported for the Vadol station then other stations for many rainfall events during the season, based on the variability of data at the neighbouring stations, the Vadol station data might be accepted. However, there is one significant event in the month of May which is reported elsewhere and for which zero rainfall is reported at Vadol. This may seem to be an error due to non-observation or incorrect reporting. It is necessary to refer to the manuscript for this year and to see if data in the database corresponds with it. It may also be possible that the observations have not really been taken by the observer on this particular station for this period during which it is normally not expected to rain. On the basis of the variability experienced between various stations in the region it may then be decided to consider some of the reported zero values as doubtful at Vadol station.

Table 4.4: Tabulation of daily rainfall

Year	Month	Day	Balasinor	Kapadwanj	Savlitank	Vadol	Vagharoli
1974	5	23	0.00	0.00	0.00	0.00	-99.00
1974	5	24	0.00	0.00	0.00	0.00	-99.00
1974	5	25	0.00	0.00	0.00	0.00	-99.00
1974	5	26	4.20	75.00	73.00	0.00	-99.00
1974	5	27	23.00	30.00	19.00	0.00	-99.00
1974	5	28	0.00	0.00	0.00	0.00	-99.00
1974	5	29	12.00	0.00	0.00	0.00	-99.00
1974	5	30	0.00	0.00	0.00	0.00	-99.00
1974	5	31	0.00	0.00	0.00	0.00	-99.00

4.6 Spatial homogeneity testing of rainfall (nearest neighbour analysis)

4.6.1 General description

As mentioned above, rainfall exhibits some degree of spatial consistency. The degree of consistency is primarily based on the actual spatial correlation. The expected spatial consistency is the basis of investigating the observed rainfall values at the individual observation stations. An estimate of the interpolated rainfall value at a station is obtained on the basis of weighted average of rainfall observed at the surrounding stations. Whenever the difference between observed and estimated values exceed the expected limiting value, such values are considered as suspect values and they are then flagged for further investigation for possible causes of their discrepancies.

4.6.2 Data validation procedure and follow up actions

The estimation of the spatially interpolated rainfall value is made at the station under consideration. The station being considered is the suspect station and is called the test station. The interpolated value is estimated by computing the weighted average of the rainfall observed at neighbouring stations. Ideally, the stations selected as neighbours should be physically representative of the area in which the station under scrutiny is situated. The following criteria are used to select the neighbouring stations (see Figure 4.11):

a) The distance between the test and the neighbouring station must be less than a specified maximum correlation distance, say Rmax (kms)
b) A maximum of 8 neighbouring stations can be considered for interpolation
c) To reduce the spatial bias in selection, it is appropriate to consider a maximum of only two stations within each quadrant

The estimate of the interpolated value at the test station based on the observations at N neighbouring stations is given as

$$P_{est}(t) = \frac{\sum_{i=1}^{N} P_i(t)/D_i^b}{\sum_{i=1}^{N} 1/D_i^b} \qquad \text{Eqn. 4.1}$$

Where:

$P_{est}(t)$ = estimated rainfall at the test station at time t

$P_i(t)$ = observed rainfall at the neighbour station i at time t

D_i= distance between the test and the neighbouring station i

N = number of neighbouring stations considered.

b = power of distance (typically equal to 2 or slightly below 2)

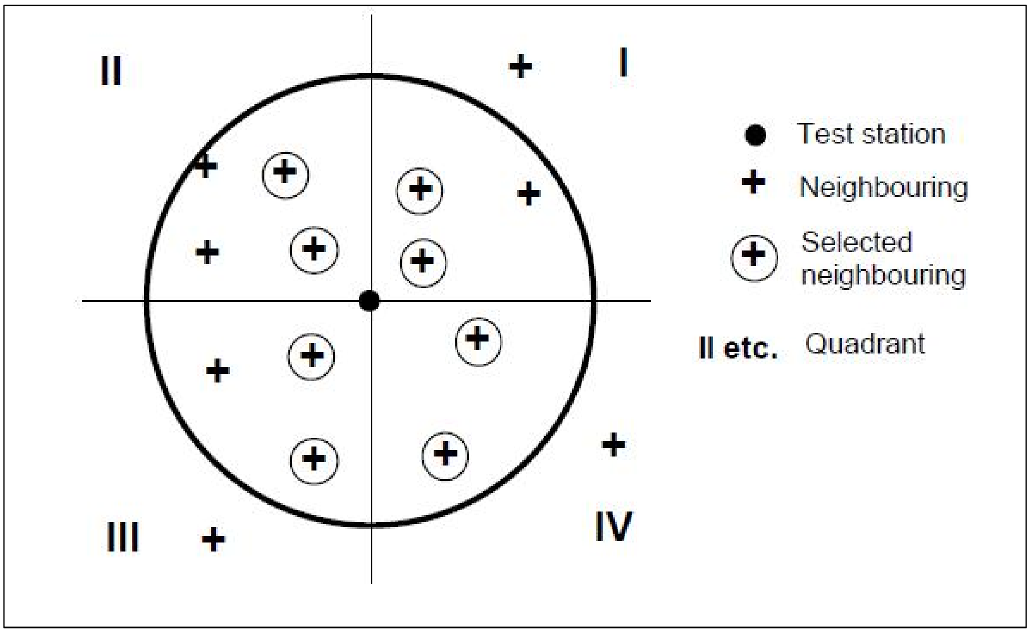


Figure 4.11: Definition sketch of test and base (neighbouring) stations

This estimated value is compared with the observed value at the test station and the difference is considered as insignificant if the following conditions are met:

$$|P_{obs}(t) - P_{est}(t)| = X_{abs} \qquad \text{Eqn. 4.2}$$

$$|P_{obs}(t) - P_{est}(t)| \leq X_{ref}.S_{P_{est}}(t) \qquad \text{Eqn. 4.3}$$

Where:

X_{abs} = admissible absolute difference

$S_{Pest(t)}$ = standard deviation of neighbouring values

X_{rel}= multiplier of standar[illegible]eviation

$$S_{P_{est}}(t) = \sqrt{\sum_{i=1}^{N}(P_i(t) - \text{[illegible]}))^2} \qquad \text{Eqn. 4.4}$$

Where departures are unacceptably high, the recorded value is flagged “+” or “-”, depending on whether the observed rainfall is greater or less than the estimated one. The limits X_{abs} and X_{rel} are chosen by the data processor and have to be based on the spatial variability of rainfall. They are normally determined on the basis of experience with the historical data with the objective of flagging a few suspect values (e.g. those beyond the 5 or 95 percentile).

It is customary to select a reasonably high value of X_{abs} to avoid having to deal with a large number of different values in the lower range. In the example illustrated below, X_{abs} = 50

It should be noted that where X_{rel} only is applied (i.e., X_{abs} is large), the test also picks up an excessive number of anomalies at low annual or seasonal rainfalls where $X_{rel} \cdot S_p$ has a small value. Such differences at low rainfall are both more likely to occur and have less effect on the overall rainfall total, so it is important to select a value of X_{rel} to flag a realistic number of suspect values. In the example shown $X_{rel} = 2$.

This check for spatial consistency can be carried out for various durations of rainfall accumulations. This is useful in case smaller systematic errors are not detectable at lower level of aggregation. The relative limit X_{rel} is less for daily data than for monthly data because of relatively higher S_{Pest}.

Typical rainfall measurement errors show up with specific patterns of "+" and "-" in the spatial homogeneity test and will be mentioned in the following sections to aid interpretation of the flagged values.

Example 4-7

A test is performed for reviewing the spatial homogeneity of the daily rainfall data at Savlitank station in Kheda catchment. An area within a radius of 25 km around Savlitank station is considered for selecting the base stations (see Figure 4.12). Absolute and Relative errors admissible for testing are kept as 50 mm and a multiplier of 2 with standard deviation respectively. Report on the result of the analysis of spatial homogeneity test is given in Table 4.5

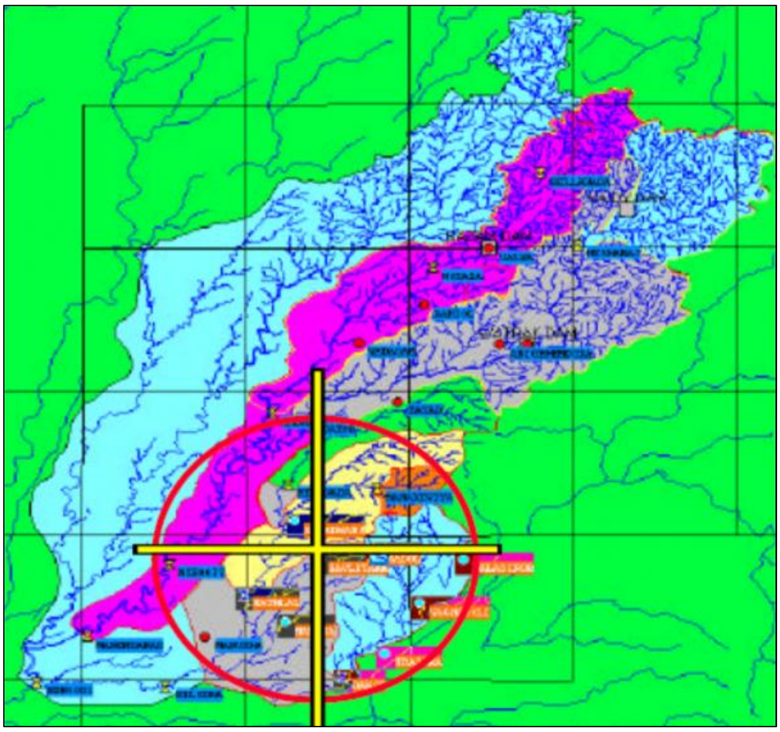


Figure 4.12: Selection of test station Savlitank and neighbouring base stations

Table 4.5: Results of the spatial homogeneity test

Test station	: Savlitank
Radius of circle of influence	: 25 Km
Station weights proportional to	: $1/D^2$
Admissible absolute Error	: 50
Multiplier to stdev. of neighbours	: 2

Selected neighbour stations:

Quadrant	Station	Distance (km)
1	Vadol	9.23
2	Kapadwanj	8.14
3	Mahisa	13.48
3	Kathlal	13.89
4	Vagharoli	17.87
4	Thasara	21.17

Year	Month	Day	Hr	Si	P_{obs}	Flag	P_{est}	StDev.	N
1984	6	14	0	1	9.00	+	0.00	0.00	6
1984	6	15	0	1	14.00	+	0.00	0.00	6
1984	6	16	0	1	23.00	+	0.00	0.00	6
1984	7	2	0	1	52.00	+	14.52	9.71	6
1984	7	6	0	1	47.00	+	2.13	4.51	6
1984	7	25	0	1	25.00	+	0.32	1.21	6
1984	***8***	***3***	***0***	***1***	***0.00***	**-**	***96.59***	***65.7***	***6***
1984	***8***	***4***	***0***	***1***	***150.00***	**+**	***78.44***	***38.47***	***6***
1984	8	5	0	1	76.00	+	20.64	36.20	6
1984	***8***	***10***	***0***	***1***	***0.00***	**-**	***128.36***	***93.57***	***6***
1984	***8***	***11***	***0***	***1***	***201.00***	**+**	***59.25***	***42.04***	***6***
1984	8	15	0	1	30.00	+	0.50	1.89	6
1984	8	19	0	1	27.00	+	16.81	4.91	6
1984	8	28	0	1	8.00	+	0.00	0.00	6
1985	6	13	0	1	9.00	+	0.00	0.00	6
1985	6	14	0	1	14.00	+	0.00	0.00	6
1985	6	16	0	1	8.00	+	0.00	0.00	6
1985	7	2	0	1	21.00	+	0.07	0.37	6
1985	7	6	0	1	47.00	+	0.73	3.73	6
1985	7	19	0	1	60.00	+	16.05	15.49	6
1985	7	21	0	1	29.00	+	10.41	7.93	6
1985	7	23	0	1	12.00	+	0.15	0.75	6

Year	Month	Day	Hr	Si	P_{obs}	Flag	P_{est}	StDev.	N
1985	7	25	0	1	25.00	+	3.15	3.78	6
1985	8	1	0	1	10.00	+	0.48	1.97	6
1985	***8***	***4***	***0***	***1***	***150.00***	***+***	***82.57***	***76.84***	***6***
1985	8	5	0	1	76.00	+	15.06	37.51	6
1985	***8***	***11***	***0***	***1***	***201.00***	***+***	***11.39***	***53.59***	***6***
1985	8	15	0	1	30.00	+	0.29	1.49	6
1985	8	17	0	1	20.00	+	1.09	5.59	6
1985	8	19	0	1	27.00	+	1.75	8.94	6
1985	8	28	0	1	8.00	+	0.00	0.00	6
1985	9	14	0	1	17.00	+	0.00	0.00	6
1985	9	15	0	1	3.00	+	0.00	0.00	6
1985	***10***	***8***	***0***	***1***	***145.00***	***+***	***70.17***	***67.38***	***6***
1985	***10***	***9***	***0***	***1***	***0.00***	***-***	***86.03***	***116.43***	***6***

Note: The bold numbers indicate suspect values based on the filters explained below.

Legend:

n = number of neighbour stations

+ = P_obs - P_est > 0 - = P_obs - P_est < 0 * = P_est is missing

Six neighbouring stations are considered eligible for making the spatial estimate. Comparison of observed and estimated daily rainfall value is made and those instances where the difference between observed and estimated value is more than the test criteria (i.e. absolute or relative difference) a flag is put. Listing of these instances can be seen in the analysis report given above.

The following can be easily deduced from the above listing:

- There are quite a few very large differences in the observed and the estimated values e.g. those on 3rd, 4th, 10th, 11th of August 1984; the 4th, 11th of August 1985 and 8th, 9th of October 1985 (highlighted in the table). Such large differences warrant a closer look at the observed values in conjunction with the rainfall at the neighbouring stations.
- A few of these instances of large differences are preceded or followed by zeros, which indicates that either the rainfall is accumulated or there is a possibility of time shift in the data. However, presence of a large amount of standard deviation points to the fact that the variability of rainfall at these instances is quite high among the neighbouring stations and it may not be impossible to observe such large variations at the test station as well. Another possibility is that there has been some time shift in the data of one or more of the base stations as well. When all the stations considered are also likely to have similar errors, this option can be ruled out. The tabulation of data at these base stations in fact reveals the possibility of such shifting.

- Some of the instances when the rainfall has been very low and the standard deviation among the neighbouring stations is also very low are also listed (especially those with zero rainfall at all the neighbouring stations and thus zero standard deviation and a very low rainfall at the test station). Such differences would normally be picked up by the relative error test owing to very small standard deviations and can be overlooked if the value at test station is also meagre. It can be noticed that on the estimated rainfall is 0 in June implies that there has been zero rainfall reported at all the six neighbouring stations. Since the resulting standard deviation is also zero, it is very likely that at all these neighbouring stations observation of rainfall is started from 16th June of every year and thus the first observation is available only for 17th of June. Inadvertently, all these missing data on and before 16th June have been reported as 0 mm. Further, Savlitank station is on a reservoir site where there might be a requirement of having continuous observations throughout the year, so it may also happen that the reported rainfall values are correct.
- As explained above, for the listed inconsistencies possible scenarios are required to be probed further and only then a judicious corrective measure can be exercised. In case none of the corroborative facts substantiates further suspicion, either the value can be left as suspect or if the variability of the process is considered very high such suspect values can be subsequently accepted.

4.7 Identification of common errors

In the following sections, procedures for identification of common errors in rainfall data are discussed with reference to either Graphical and tabular (Section 4.3 and 4.4) or Spatial homogeneity tests (Section 4.6)

Typical errors are:

- Entries on the wrong day - shifted entries
- Entries made as accumulations
- Missed entries
- Rainfall measurement missed on days of low rainfall.

4.8 Checking for entries on wrong days - Shifted entries

4.8.1 General description

Since the record of rainfall data is interspersed with many entries having zero values, values may be entered against wrong days. This is due to the fact that while entering the data one or more zero entries may get omitted or repeated by mistake. For daily data, such mistakes are more likely when there are a few non-zero values in the middle and most of the entries at the beginning and end of the month as zero values. These result in shifting of one or more storms by a day or two, which normally tend to get corrected with the start of the new month. This is because for the next month the column or page starts afresh in the manuscript from which the data are entered.

4.8.2 Data validation procedure and follow up actions:

Shift errors in rainfall series can often be spotted in the tabulated or plotted multiple series, especially if they are repeated over several wet/dry spells. It is assumed that no more than one of the listed series will be shifted in the same direction in the same set. With respect to spatial homogeneity testing, application of the test will generate a + at the beginning of a wet spell and a - at the end (and possibly others in between) if the data are shifted forward, and the reverse if the data are shifted backward.

A shift to coincide with the timing of adjacent stations and rerun of the spatial homogeneity test will generally result in the disappearance of the + and - flags, if our interpretation of the shift was correct.

The re-shifted series is then adopted as the validated series for the station/period in question.

Example 4-8

Spatial homogeneity test for daily rainfall series of Vadagam station in Kheda catchment is carried out with neighbouring stations Modasa, Rahiol, Bayad and Anior as base stations. The result of this test is reported in Table 4.6 below.

It may be noticed from above listing that a negative flag together with 0 mm observed rainfall followed by a positive flag, both with very high value of absolute difference between the observed and estimated daily rainfall is shown on 5th and 7th August 1988. Such flagging indicates a possible shift in the data at this station. Other instances listed in the test report are primarily due to small standard deviation among base stations during low rainfall days and may be ignored.

This suspicion is confirmed after looking at the tabulation of this station data along with the other four base stations since month of July as given in Table 4.7

It may be seen that except for the event starting on the 5th of August, most of the other rain events at these five stations correspond qualitatively with respect to timings. Data for this event seems to have shifted forward (i.e. lagging in time) by one day. This shifting has been the reason for the negative flag and 0 observed rainfall followed later by the positive flag in the recession phase of the event.

Table 4.6: Result of the spatial homogeneity test at Vadagam station

Test station	: Vadagam
Start date	: 1988
End date	: 1988
Radius of circle of influence	: 25 km
Station weights proportional to	: $1/D^2$
Admissible absolute error	: 50
Multiplier to Standard Deviation of neighbours	: 2

Selected neighbour stations:

Quadrant	Station	Distance (km)
1	Rahiol	12.61
1	Modasa	18.69
4	Bayad	12.88
4	Anior	21.83

Year	Month	Day	Hr	Si	P_obs	Flag	Pest	Stdev	N
1988	8	1	0	1	0.50	-	8.32	3.83	4
1988	8	5	0	1	0.00	-	182.00	45.70	4
1988	8	7	0	1	161.00	+	14.23	8.32	4
1988	8	8	0	1	4.00	-	11.98	3.06	4
1988	8	9	0	1	18.00	+	7.12	1.72	4
1988	8	11	0	1	4.20	+	0.59	1.43	4

This shift was confirmed by looking at the manuscript and thus implies that this has occurred at the time or after the data has been entered into the computer. The shift was corrected by removing one day lag in this storm event and stored as a temporarily (Data type TMA). When the spatial homogeneity test was carried out again with this corrected series, the following results were obtained:(given in Table 4.8)

It may now be seen that there is no negative or positive flag with 0 observed rainfall and large difference in observed and estimated value. The rainfall on 6th August is still flagged because of larger difference in observed and estimated rainfall as against the permissible limit. Thus, in this way the time shifts may be detected and removed by making use of spatial homogeneity test.

Table 4.7: Tabulation of daily rainfall at neighbouring stations

Date			Observed Rainfall at Neighbouring Stations (mm)				Rainfall at Vadagam (mm)	
Year	Month	Day	Anior	Bayad	Modasa	Rahiol	Estimated P_{est}	Observed
			Weights					
			0.121	0.349	0.166	0.364		
1988	7	12	0.00	0.00	0.00	0.00	0.00	0.00
1988	7	13	0.00	0.00	11.62	0.00	11.60	0.00
1988	7	14	3.99	22.69	12.45	10.92	50.00	14.00
1988	7	15	0.97	6.21	2.08	1.82	11.10	3.00
1988	7	16	3.24	4.89	5.15	22.06	35.30	40.00
1988	7	17	0.65	0.42	1.66	0.73	3.50	1.00
1988	7	18	0.00	0.70	0.00	0.00	0.70	1.00
1988	7	19	4.84	20.17	0.42	18.49	43.90	35.00
1988	7	20	6.56	16.05	9.96	11.94	44.50	46.00
1988	7	21	0.85	5.93	0.66	1.46	8.90	19.00
1988	7	22	13.67	27.36	20.58	33.42	95.00	82.00
1988	7	23	0.00	3.91	2.49	2.48	8.90	16.30
1988	7	24	1.57	0.00	4.81	2.69	9.10	0.00
1988	7	25	0.97	4.89	7.22	13.03	26.10	23.10
1988	7	26	2.18	9.42	0.17	0.00	11.80	4.20
1988	7	27	3.75	0.35	0.00	1.24	5.30	1.20
1988	7	28	3.51	14.66	1.16	3.64	23.00	23.00
1988	7	29	0.00	4.89	2.49	1.46	8.80	10.00
1988	7	30	1.62	0.00	7.14	0.73	9.50	0.00
1988	7	31	0.51	5.93	1.00	0.00	7.40	0.00
1988	8	1	0.97	1.05	2.16	4.15	8.30	0.50
1988	8	2	0.48	0.00	0.33	0.00	0.80	0.00
1988	8	3	0.00	0.00	2.82	8.01	10.80	4.00
1988	8	4	0.00	0.35	0.17	0.00	0.50	0.00
1988	8	5	30.61	47.12	26.73	77.46	181.90	0.00
1988	8	6	16.82	32.81	18.59	40.26	108.50	140.00
1988	8	7	2.42	8.38	0.66	2.77	14.20	161.00
1988	8	8	1.36	2.79	1.83	6.01	12.00	4.00
1988	8	9	1.09	2.79	1.49	1.75	7.10	18.00
1988	8	10	0.31	1.05	1.33	0.36	3.10	1.20
1988	8	11	0.42	0.00	0.17	0.00	0.60	4.20
1988	8	12	0.00	0.00	0.50	0.00	0.50	3.00
1988	8	13	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.8: Results of the spatial homogeneity test on the corrected series

Spatial homogeneity check

Test station : Vadagam

Radius of circle of influence : 25 (km)

Station weights proportional to : 1/D^2.00

Admissible absolute error : 50

Multiplier to st.dev of neighbours : 2

Selected neighbour stations:

Quadrant	Station	Distance (km)
1	Rahiol	12.606
1	Modasa	18.689
4	Bayad	12.882
4	Anior	21.829

Year	Month	Day	Hr	Si	P_obs	Flag	P_est	St dev	N
1988	8	1	0	1	0.50	-	8.32	3.83	4
1988	8	6	0	1	161.00	+	108.49	16.13	4
1988	8	9	0	1	1.20	-	7.12	1.72	4
1988	8	25	0	1	32.00	+	1.97	4.34	4
1988	9	6	0	1	9.50	+	0.00	0.00	4
1988	9	29	0	1	12.00	+	1.09	1.30	4

4.9 Entries made as accumulations

4.9.1 General description

The rainfall observer is expected to take rainfall observations every day at the stipulated time, without discontinuity for holidays, weekends or sickness. Nevertheless, it is likely that on occasions the rain gauge reader will miss a reading for one of the above reasons. The observer may make one of three choices for the missed day or sequence of days.

- Enter the value of the accumulated rainfall on the day on which he/she returned from absence and indicate that the intervening values were accumulated (the correct approach).
- Enter the value of the accumulated rainfall on the day on which he/she returned and enter a zero (or no entry) in the intervening period.
- Attempt to guess the distribution of the accumulated rainfall over the accumulated period and enter a positive value for each of the days.

The third option is probably the more common as the observer may fear that he will be penalised for missing a period of record even for a legitimate reason. The second option is also common. Observers must be encouraged to follow the first option, as a more satisfactory interpolation can be made from adjacent stations than by the observer’s guess.

4.9.2 Data validation procedure and follow up actions

If accumulations are clearly marked by the observer then the accumulated value can readily be distributed over the period of absence, by comparison with the distribution over the same period at adjacent stations.

For not indicated accumulations with a zero in the missed values, the daily tabulation will indicate a gap in a rainy spell in comparison to neighbouring stations. Of course, an absence during a period of no rain will have no impact on the reported series. Spatial homogeneity testing will show a negative flag on days on which there was significant rain during the period of accumulation and a positive flag on the day of accumulation.

The data processor should inspect the record for patterns of this type and mark such occurrences as suspect. In the first instance, a reference is made to the field record sheet to confirm that the data were entered as recorded. Then, this being so, a search is made backward from the date of accumulated total to the first date on which a measurable rainfall has been entered and an apportionment made on the basis of neighbouring stations.

The apportioning is done over the period which immediately preceded the positive departure with negative departures and zero rainfall. The accumulated rainfall is apportioned in the ratio of the estimated values on respective days as:

$$P_{appor,i} = \frac{P_{est,i} * P_{tot}}{\sum_{i=1}^{N_{acc}} P_{est,i}} \qquad \text{Eqn. 4.5}$$

Where:

P_{tot} = accumulated rainfall as recorded

N_{acc} = number of days of accumulation

$P_{est,I}$ = estimated daily rainfalls during the period of accumulation on the basis of adjoining stations

$P_{appor,I}$ = apportioned value of rainfall for each day of accumulation period

Where it is not possible to adequately reason in favour or against such an accumulation, then the suspect value can be left labelled as doubtful. On the other hand, if the period of such accumulation is clearly marked by the observer, then the apportionment for the said period can be done directly without checking for the period of accumulation.

The field supervisor should be informed of such positively identified or suspicious accumulations and requested to instruct the field observer in the correct procedure.

Example 4-9

As a routine secondary validation, spatial homogeneity test for station Dakor (Kheda catchment) for the year 1995 is carried out considering a few neighbouring stations. The test results are as given below (Table 4.9):

On examining the above results, it can be apparent that there are a few negative flags having nil observed rainfall which is followed by a positive flag having a very high rainfall value. Such combination indicates a possible accumulation of rainfall for one or more days prior to 28 July 95 and warrants a closer look at this suspect scenario at Dakor station.

The listing of daily rainfall for neighbouring stations considered for the above spatial homogeneity test is as given in Table 4.10. Upon careful examination it can be seen that at Dakor station the rainfall recorded for a few consecutive days during 11 July 1995 to 27 July 1995 is nil, while most of other neighbouring stations have received significant rainfall on these days. On the next day (28th of July), there was a very large value recorded for Dakor station whereas the other nearby stations did not record high rainfall. Such situation does not rule out an un-indicated accumulation of rainfall at Dakor for one or more days prior to 28 July.

At this stage the manuscripts of daily rainfall at Dakor station must be revisited to confirm if the data in the databases are properly recorded. If the data are as per the records then based on the feedback from the observer about his absence/holidays etc. and upon overall reliability of the station in the past, it can be decided to flag such un-indicated accumulations for subsequent correction using spatial interpolation (see Chapter 04).

Table 4.9: Result of spatial homogeneity test at station Dakor

Test station : Dakor

Start date : 1995

End date : 1995

Radius of circle of influence : 25

Station weights proportional to : $1/D^2$

Admissible absolute error : 50

Multiplier to St Dev of neighbours : 2

Selected neighbouring stations:

Quadrant	Station	Distance (km)
1	Thasara	8.25
1	Vagharoli	18.98
2	Mahisa	13.95

Quadrant	Station	Distance (km)
2	Kathlal	22.22
2	Mahudha	22.69
2	Savlitank	23.40

Year	Month	Day	Hr	si	P_obs	Flag	P_est	Stdv	N
1995	7	15	0	1	0.00	-	56.64	20.50	6
1995	7	18	0	1	0.00	-	8.79	3.34	6
1995	7	19	0	1	0.00	-	21.24	8.73	6
1995	7	20	0	1	0.00	-	36.82	15.42	6
1995	7	28	0	1	97.50	+	18.12	13.28	6
1995	7	30	0	1	6.80	-	48.59	16.20	6

Legend:

n = number of neighbour stations

+ = P_obs - P_est > 0

- = P_obs - P_est < 0

* = P_est is missing

Table 4.10: Tabulation of daily rainfall for neighbouring stations

Year	Month	Day	Dakor	Kathlal	Mahisa	Mahudha	Savlitank	Thasara
1995	7	11	0.00	7.00	10.00	1.50	27.00	9.00
1995	7	12	0.00	0.00	3.00	2.00	3.00	17.00
1995	7	13	0.00	45.00	0.00	0.00	0.00	0.00
1995	7	14	0.00	10.00	20.00	7.50	0.00	7.00
1995	7	15	0.00	14.00	50.00	33.50	24.00	77.00
1995	7	16	0.00	0.00	8.00	9.50	25.00	8.00
1995	7	17	0.00	20.00	4.00	1.00	0.00	22.00
1995	7	18	0.00	10.00	8.00	1.00	6.00	11.00
1995	7	19	0.00	23.00	20.00	43.00	27.00	16.00
1995	7	20	0.00	0.00	35.00	32.50	14.00	48.00
1995	7	21	0.00	57.00	27.00	23.00	14.00	56.00
1995	7	22	0.00	0.00	6.00	7.00	4.00	0.00
1995	7	23	0.00	0.00	4.00	12.00	2.00	27.00
1995	7	24	0.00	10.00	0.00	0.00	0.00	0.00
1995	7	25	0.00	11.00	10.00	3.00	6.00	3.00
1995	7	26	0.00	25.00	0.00	10.00	5.00	8.00

Year	Month	Day	Dakor	Kathlal	Mahisa	Mahudha	Savlitank	Thasara
1995	7	27	0.00	18.00	3.00	4.00	25.00	9.00
1995	7	28	97.50	25.00	24.00	46.00	3.00	12.00
1995	7	29	16.70	40.00	4.00	6.00	0.00	0.00
1995	7	30	6.80	45.00	34.00	22.00	62.00	52.00
1995	7	31	0.00	10.00	3.00	13.00	39.00	9.00

4.9.3 Screening for accumulations on holidays and weekends

To screen for accumulated values on holidays and weekends on stations where such readings are missing, a comparison is made between observed and estimated values of daily rainfall of the station under consideration for the period of holidays and weekends and a day following it. While comparing the two sets, the data points having significant positive difference between observed and estimated values on the day following the holidays or weekends are picked up.

4.10 Missed entries

4.10.1 General description

Values may be missed from a record either by the observer failing to do the observation, failing to enter a value in the record sheet or as the result of a missed entry. A zero may have been inserted for the day (or days), however, zero means no rain, while missing data does not exclude the possibility that there was rainfall in the missing data period. Similarly, some longer periods may have missed readings without an accumulated value at the end, for example resulting from breakage of the measuring cylinder.

4.10.2 Data validation procedure and follow up actions

For rainy periods, missing data will be anomalous. In the multiple station tabulation and plots, this will be indicated by a series of negative departures in the spatial homogeneity test.

Where such missed entries are confidently identified, the missed values will be replaced by the estimates derived from neighbouring stations by the Spatial Homogeneity test. Where there is some doubt as to the interpretation, the value will be left unchanged but flagged as suspect.

Example 4-10

The spatial homogeneity test for Bhempoda station (Kheda catchment) for the year 1997 is carried out. The results of the test are given below in Table 4.11.

On examining the above tables, it can be noticed that there are many instances in succession which are flagged negative and also have nil (0 mm) observed rainfall. At the same time, on these days of negative flag and 0 mm observed rainfall a considerable rainfall at the neighbouring stations has been reported. Such an inference leads to suspicion that at this test station the rainfall has either not been observed and wrongly reported as 0 mm or has been observed but has been wrongly entered.

The above suspicion is very strongly corroborated after looking at the tabulation of the neighbouring stations given in Table 4.12

It is almost certain that the rainfall at Bhempoda station has been entered incorrectly from the second week of August 97 onwards for most of the rainy days reported at the neighbouring stations. These rainfall values must be checked with the records of the data at Bhempoda station and if the values available in the records are different from those available in the database, then the same must be corrected. Instead, if the manuscript also shows the same values, then these have to be flagged for necessary correction using spatial interpolation.

Table 4.11: Results of spatial homogeneity at Bhempoda station

Test station	: Bhempoda
Start date	: 1997
End date	: 1997
Radius of circle of influence	: 25 (km)
Station weights proportional to	: $1/D^2$
Admissible absolute error	: 40
Multiplier to StDev of neighbours	: 2

Selected neighbouring stations:

Quadrant	Station	Distance (km)
1	Megharaj	20.90
2	Rahiol	17.90
3	Anior	4.54
3	Bayad	23.26

Year	Month	Day	Hr	Si	P_obs	Flag	P_est	Stdev	N
1997	6	9	0	1	9	+	0.00	0.00	4
1997	6	14	0	1	3	+	0.00	0.00	4
1997	6	22	0	1	20	+	4.79	2.38	4
1997	6	23	0	1	17	+	2.11	4.20	4
1997	6	25	0	1	165	-	205.65	33.94	4
1997	6	27	0	1	173	+	71.55	37.77	4
1997	7	10	0	1	0	-	1.31	0.65	4
1997	7	20	0	1	3	+	1.34	0.65	4
1997	7	21	0	1	29	-	80.48	34.46	4
1997	7	26	0	1	1	-	12.73	4.42	4
1997	7	27	0	1	125	-	225.13	58.75	4
1997	7	28	0	1	280	-	376.98	153.43	4
1997	8	2	0	1	94	+	36.15	21.21	4
1997	8	8	0	1	0	-	20.98	5.32	4
1997	8	9	0	1	0	-	2.37	0.56	4
1997	8	11	0	1	0	-	0.44	0.22	4
1997	8	14	0	1	0	-	2.66	1.14	4
1997	8	19	0	1	0	-	48.96	18.63	4
1997	8	24	0	1	0	-	87.56	42.17	4
1997	9	11	0	1	0	-	18.50	6.03	4
1997	9	13	0	1	0	-	15.36	5.79	4

Table 4.12: Tabulation results for daily rainfall at neighbouring stations

Year	Month	Day	Hr	Si	P_obs	Flag	P_est	Stdev	N
1997	6	9	0	1	9.00	+	0.00	0.00	4
1997	6	14	0	1	3.00	+	0.00	0.00	4
1997	6	22	0	1	20.00	+	4.79	2.38	4
1997	6	23	0	1	17.00	+	2.11	4.2	4
1997	6	25	0	1	165.00	-	205.65	33.94	4
1997	6	27	0	1	173.00	+	71.55	37.77	4
1997	7	10	0	1	0.00	-	1.31	0.65	4
1997	7	20	0	1	3.00	+	1.34	0.65	4
1997	7	21	0	1	29.00	-	80.48	34.46	4
1997	7	26	0	1	1.00	-	12.73	4.42	4
1997	7	27	0	1	125.00	-	225.13	58.75	4
1997	7	28	0	1	280.00	-	376.98	153.43	4
1997	8	2	0	1	94.00	+	36.15	21.21	4
1997	8	8	0	1	0.00	-	20.98	5.32	4
1997	8	9	0	1	0.00	-	2.37	0.56	4

Year	Month	Day	Hr	Si	P_obs	Flag	P_est	Stdev	N
1997	8	11	0	1	0.00	-	0.44	0.22	4
1997	8	14	0	1	0.00	-	2.66	1.14	4
1997	8	19	0	1	0.00	-	48.96	18.63	4
1997	8	24	0	1	0.00	-	87.56	42.17	4
1997	9	11	0	1	0.00	-	18.50	6.03	4
1997	9	13	0	1	0.00	-	15.36	5.79	4

4.11 Rainfall observation missed on days with low rainfall - rainy days check

4.11.1 General description

While it is required that observers inspect the rain gauge for rain each day, the practice of some observers may be to visit the gauge only when they know that rainfall has occurred. This will result in zeros on a number of days on which a small amount of rain may have occurred. The totals will be generally correct at the end of the month, but the number of rainy days may be anomalously low. In addition, spatial homogeneity testing may not pick up such differences.

Owing to spatial homogeneity with respect to the daily rainfall, it is expected that the number of rainy days in a month or year at the neighbouring stations will not differ much. Presently, there are two definitions for number of rainy days: some agencies consider a minimum of 0.1 mm (minimum measurable) in a day to be eligible for the rainy day whereas some use 2.5 mm and above as the deciding criteria. The latter is used more often in the agriculture sector. For the hydrological purpose it is envisaged that the definition of minimum measurable rainfall (i.e. 0.1 mm) will be used for the data validation.

It is good to check if the observed data follow such characteristics. A graphical or tabular comparison of the differences in the number of rainy days for the neighbouring stations for the monthly or yearly period will be suitable in bringing out any gross inconsistency. The tolerance in the number of rainy days between the stations has to be based on the variability experienced in the region and can easily be established using the historical data. If the difference is more than the maximum expected, the data may be considered suspect. Any gross inconsistency noticed must then be probed further by looking at the manuscript and seeking a report on, or inspecting the functioning and behaviour of the observer.

4.11.2 Data validation procedure and follow up actions

First of all, with the help of historical daily rainfall data belonging to a homogenous region, the expected maximum variation in the number of rainy days for each month of the year and for year as a whole is found out. A group of stations being validated is then chosen and the number of rainy days at each station within the month(s) or year obtained. The number of rainy days at each station is then compared with every other station in the group. All those instances when the expected variation is exceeded by the actual difference in the number of rainy days are presented in tabular or graphical form. It is appropriate to present the output in a matrix form in which the stations are listed as rows and columns of the table or the graph. In case the presentation is on monthly basis then each tabular or graphical matrix can accommodate a period of one year.

Any visible departure in the number of rainy days at one or more stations can be apparent by inspecting the matrix. The station for which the number of rainy days is significantly different from the others will have the column and row with lower (or occasionally higher) values. The data pertaining to such months or years of the station(s) for which the difference in the number of rainy days is beyond the expected range is considered suspect and has to be probed further. The original observer's manuscript for the suspect period can be compared with the values available in the database. Any discrepancy found between the two can be corrected by substituting the manuscript values. Where the manuscript matches with the data available in the database, a comparison with other related data like temperature and humidity at the station, if available, can be made. Together with the analytical comparison, feedback from the observer or supervisor will be of a great value in checking this validation especially where it is done within one or two months of the observations. If the related data corroborate the occurrence of such rainy days then the same can be accepted.

Where there is strong evidence to support the view that the number of rainy days derived from the record is incorrect, then the total may be amended by reference to the neighbouring stations. Such action implies that there are unreported errors remaining in the time series, which have not been possible to identify and correct. A note to this effect should be included with the station record and provided with the data to users.

As a follow up measure, a report can be sought on the functioning and behaviour of the observed.

4.12 Checking for systematic shifts using double mass analyses

4.12.1 General description

Double mass analysis is a technique that is effective in detecting a systematic shift, like abrupt or gradual changes in the mean of a series persisting in the record for a considerable period of time. Rainfall records may contain such inconsistencies for a considerable period of time. Inconsistencies present in the rainfall data of a station can occur for various reasons:

- The rain gauge might have been installed at different sites in the past
- The exposure conditions of the gauge may have undergone a significant change due to the growth of trees or construction of buildings in its proximity
- There might have been a change in the instrument, say from 125 mm to 200 mm rain gauge
- The rain gauge may have been faulty for a considerable period of time

Such inhomogeneity in the data set must be removed before any statistical inference can be drawn. The double mass analysis tests the record for its inconsistency and accuracy and provides a correction factor to ensure that the data series is reasonably homogeneous throughout its length and is related to a known site. A note may be available in the station registers of the known changes of site and instruments and can corroborate the detection of inconsistency using this technique. The application of double mass analysis to rainfall data will not be possible until a significant amount of historical data has been entered into the database.

4.12.2 Description of the method

Double mass analysis is a technique for detection of possible homogeneities in time series data by investigating the ratio of accumulated values of two series, which are:

- the series to be tested, and
- the base series

The base series is generally a composite series, i.e. the average of reliable series of nearby stations (usually 3 as a minimum), which are assumed to be homogenous.

First of all, the accumulated test and base series are obtained as two vectors (e.g. Yi and Xi respectively, for i = 1, N). The double mass analysis then considers the following ratio:

$$rc_i = \frac{\sum_{j=1}^{i} Y_j}{\sum_{j=1}^{i} X_j} \qquad \text{Eqn. 4.6}$$

or expressed as a ratio of the percentages of the totals for N elements:

$$pc_i = \frac{\sum_{j=1}^{i} Y_j \sum_{j=1}^{N} X_j}{\sum_{j=1}^{N} Y_j \sum_{j=1}^{i} X_j} \qquad \text{Eqn. 4.7}$$

These ratios in absolute and percent form give the overall slope of the double mass plot from the origin to the selected duration of analysis.

A graph is plotted between the cumulative rainfall of the base series as abscissa and the cumulative rainfall of test station as the ordinate. The resulting plot is called the double mass curve. If the data of test station is homogeneous and consistent with the data of the base series, the double mass curve will show a straight line. An abrupt change in the test-series will create a break in the double mass curve, whereas a trend will create a curve. Graphical inspection of the double mass plot provides the simplest means of identifying such inconsistencies but significance tests may also be used to identify breaks and jumps. A change in slope is not usually considered significant unless it persists for at least 5 years and there is corroborating evidence of a change in location or exposure or some other change.

Limitations of this technique is evident when there is a regional consistency in precipitation pattern for long periods of time but this consistency becomes less pronounced for shorter periods. Therefore, the double mass technique is not recommended for adjustment of daily or storm rainfalls. It is also important to mention here that any change in regional meteorological or weather conditions would not have any influence on the slope of the double mass curve because the test station as well as the surrounding base stations would have been equally affected.

4.12.3 Data validation procedure and follow up actions

For analysing the rainfall data for any persistent systematic shift, the accumulated rainfall for longer duration at the station under consideration (called the test station) is compared with another accumulated rainfall series that is expected to be homogeneous. Homogeneous series for comparison is derived by averaging rainfall data from a number of neighbouring homogenous stations (called base stations).

Accumulation of rainfall can be made from daily data to monthly or yearly duration. The double mass plot between the accumulated values in percent form at the test and the base station is drawn and observed for any visible change in its slope. The tabular output giving the ratio between the accumulated values at test and base station in absolute and percent is also obtained. In case there are some missing data points within each duration of analysis, a decision can be made about the number of elements which must essentially be present for that duration to be considered for analysis. The analysis, if required, can also be carried for only a part of the years or months.

Where there is a visible change in the slope of the double mass plot after certain period, then such a break must be investigated further. Possible reasons for the inhomogeneity in the data series are explored and suitable explanation prepared. If the inhomogeneity is caused by changed exposure conditions or shift in the station location or systematic instrumental error then the data series must be considered suspect. The data series can then be made homogeneous by suitably transforming it before or after the period of shift as required.

Transformation for inconsistent data is carried out by multiplying it with a correction factor which is the ratio of the slope of the adjusted mass curve to the slope of the unadjusted mass curve (see Chapter 04 for details).

Example 4-11

Double mass analysis for Vadagam station (in Kheda catchment) is carried out considering two stations Megharaj and Bayad as the base stations for the period from 1968 to 1996. A period of only three months from July to September (92 days) has been taken into consideration while carrying out the analysis. The reliability of records and the homogeneity of these base stations have to be ascertained before considering them for the analysis. In this case it has been assumed that they are reliable stations. It can be seen from double mass plot of this analysis, as shown in Figure 4.13, that the data of Vadagam station are fairly consistent throughout the period of analysis (1968 to 1997) with respect to the other two base stations. Baring a few short-lived very small deviations from the ideal curve (of 45 degree), the plot shows a similar trend throughout the period.

Table 4.13. The yearly rainfall and the rainfall accumulated in time for the base and test station is given in columns 2, 3 and 5, 6 respectively. These cumulative rainfall values are then expressed in percent form in columns 4 and 7 respectively. The ratio of these cumulated values in absolute in percent form is given in the last two columns 8 and 9.

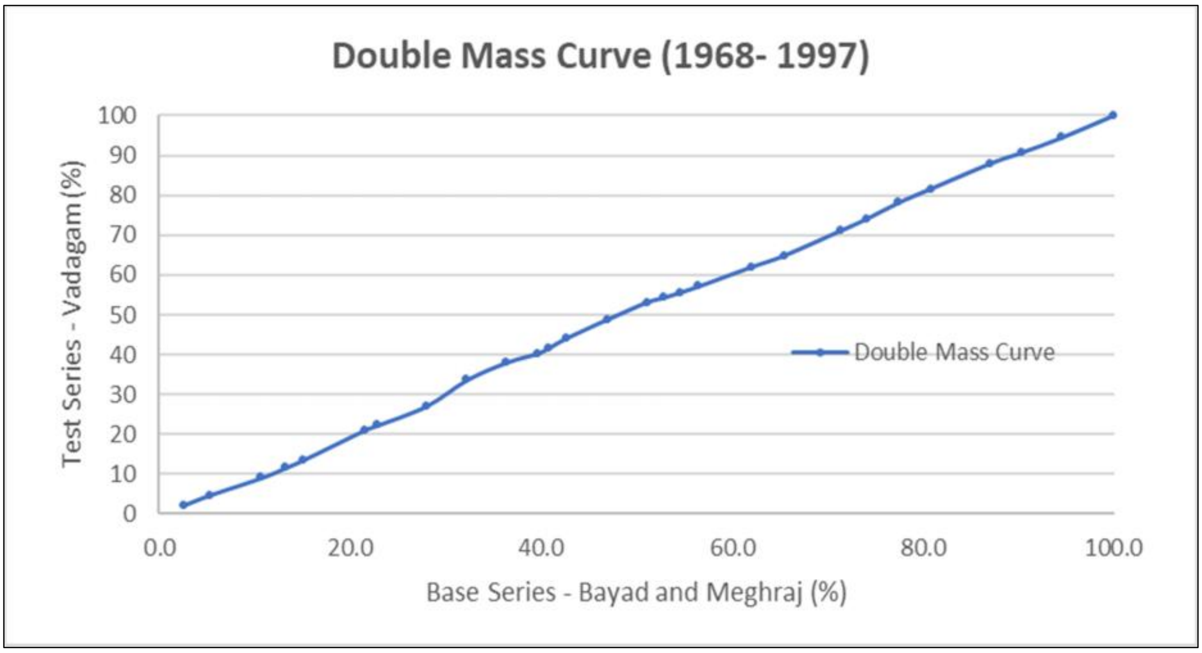


Figure 4.13: Double mass plot showing near consistent trend at test station

Table 4.13: Analysis result of the double mass analysis

Test series: Vadagam

Base series: Megharaj Weight 0.5

Bayad Weight 0.5

Period	BASE			TEST			Ratios	
	Amount	Cum	Perc	Amount	Cum	Perc	(8)	(9)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(6)/(3)	(7)/(4)
1968	451.50	451.50	2.50	382.40	382.40	2.20	0.85	0.88
1969	487.50	939.00	5.30	437.00	819.00	4.80	0.87	0.90
1970	957.40	1896.00	10.70	743.10	1563.00	9.10	0.82	0.85

Period	BASE			TEST			Ratios	
	Amount	Cum	Perc	Amount	Cum	Perc	(8)	(9)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(6)/(3)	(7)/(4)
1971	462.30	2359.00	13.30	443.40	2006.00	11.70	0.85	0.88
1972	332.10	2691.00	15.20	339.10	2345.00	13.70	0.87	0.90
1973	1124.80	3816.00	21.50	1266.30	3611.00	21.00	0.95	0.98
1974	247.80	4063.00	22.90	214.90	3826.00	22.30	0.94	0.97
1976	910.20	4974.00	28.00	831.60	4658.00	27.10	0.94	0.97
1977	751.00	5725.00	32.20	1124.10	5782.00	33.70	1.01	1.04
1978	735.00	6460.00	36.40	748.20	6530.00	38.00	1.01	1.05
1979	576.00	7036.00	39.60	389.10	6919.00	40.30	0.98	1.02
1980	205.30	7241.00	40.80	234.30	7154.00	41.70	0.99	1.02
1982	323.60	7565.00	42.60	417.70	7571.00	44.10	1.00	1.03
1983	766.30	8331.00	46.90	817.40	8389.00	48.90	1.01	1.04
1984	737.80	9069.00	51.10	737.00	9126.00	53.20	1.01	1.04
1985	312.40	9381.00	52.80	198.40	9324.00	54.30	0.99	1.03
1986	313.80	9695.00	54.60	229.60	9554.00	55.70	0.99	1.02
1987	337.30	10032.00	56.50	261.90	9816.00	57.20	0.98	1.01
1988	986.00	11018.00	62.10	837.70	10653.00	62.10	0.97	1.00
1989	605.80	11624.00	65.50	493.00	11146.00	64.90	0.96	0.99
1990	1047.80	12672.00	71.40	1065.50	12212.00	71.10	0.96	1.00
1991	481.00	13153.00	74.10	508.50	12720.00	74.10	0.97	1.00
1992	596.80	13750.00	77.50	697.00	13417.00	78.20	0.98	1.01
1993	598.00	14348.00	80.80	599.00	14016.00	81.70	0.98	1.01
1994	1101.00	15449.00	87.00	1079.50	15096.00	87.90	0.98	1.01
1995	592.50	16041.00	90.40	478.50	15574.00	90.70	0.97	1.00
1996	746.80	16788.00	94.60	647.60	16222.00	94.50	0.97	1.00
1997	963.00	17751.00	100.00	944.00	17166.00	100.00	0.97	1.00

Total number of periods of analysis: 28

Example 4-12

The long-term data series of rainfall for the period 1970 to 1996 is considered at Vadol station (in Kheda catchment) for double mass analysis taking three nearby stations Kapadwanj, Mahisa and Thasara. Unlike the previous example, which is a case of the test station being homogeneous in time, this example illustrates a case where the test station records show that there has been a significant change in the amount of rain over a period of time.

It can be easily seen from the double mass curve shown in Figure 4.14 that the behaviour of the test station suddenly changes after about half of the time period under consideration.

This turning point corresponds with the year 1984 and is also apparent from the values of the ratios of accumulated rainfall at test and base stations as given in Table 4.14 showing the results of the test.

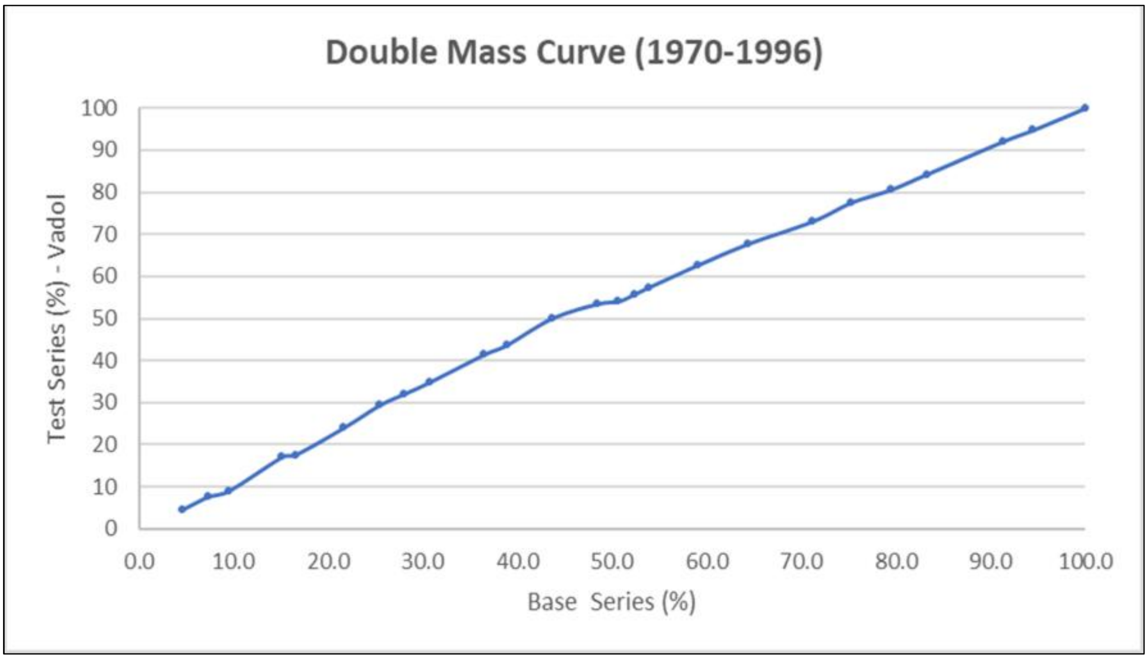


Figure 4.14: Double mass curve for station Vadol

Table 4.14: Results of the double mass analysis

Period	BASE			TEST			Ratios	
	Amount MM	Cum MM	Perc	Amount MM	Cum MM	Perc	(6)/(3)	(7)/(4)-
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1970	767.00	767.00	4.60	624.00	624.00	4.50	0.81	0.98
1971	454.00	1221.00	7.30	426.00	1050.00	7.60	0.86	1.04
1972	373.00	1594.00	9.50	198.00	1248.00	9.00	0.78	0.94
1973	935.00	2529.00	15.10	1114.00	2363.00	17.00	0.93	1.13
1974	240.00	2770.00	16.60	73.00	2435.00	17.60	0.88	1.06
1977	844.00	3613.00	21.60	883.00	3318.00	23.90	0.92	1.11
1978	646.00	4260.00	25.50	759.00	4077.00	29.40	0.96	1.15
1979	437.00	4696.00	28.10	370.00	4447.00	32.10	0.95	1.14
1980	450.00	5147.00	30.80	389.00	4836.00	34.90	0.94	1.13
1981	950.00	6097.00	36.50	898.00	5734.00	41.40	0.94	1.13
1982	404.00	6500.00	38.90	320.00	6054.00	43.70	0.93	1.12
1983	801.00	7302.00	43.70	882.00	6936.00	50.00	0.95	1.15
1984	806.00	8108.00	48.50	475.00	7411.00	53.50	0.91	1.10
1985	364.00	8472.00	50.70	83.00	7494.00	54.10	0.88	1.07
1986	282.00	8753.00	52.30	234.00	7728.00	55.70	0.88	1.06

Period	BASE			TEST			Ratios	
	Amount MM	Cum MM	Perc	Amount MM	Cum MM	Perc	(6)/(3)	(7)/(4)-
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1987	258.00	9011.00	53.90	228.00	7956.00	57.40	0.88	1.06
1988	866.00	9877.00	59.10	735.00	8690.00	62.70	0.88	1.06
1989	877.00	10754.00	64.30	693.00	9384.00	67.70	0.87	1.05
1990	1145.00	11899.00	71.20	746.00	10130.00	73.10	0.85	1.03
1991	683.00	12582.00	75.20	618.00	10748.00	77.50	0.85	1.03
1992	698.00	13280.00	79.40	422.00	11170.00	80.60	0.84	1.01
1993	640.00	13919.00	83.20	513.00	11683.00	84.30	0.84	1.01
1994	1350.00	15269.00	91.30	1083.00	12766.00	92.10	0.84	1.01
1995	525.00	15794.00	94.50	372.00	13138.00	94.80	0.83	1.00
1996	927.00	16721.00	100.00	725.00	13863.00	100.00	0.83	1.00

Total number of period of analysis: 25

It is clear that from the year 1985 onwards the test station Vadol started receiving rainfall which is less than what it used to receive before that time. And this change in behaviour is not short lived, but is continuous thereafter. The reasons for such variations need to be ascertained. Various factors which could result in such a change can be: (a) a systematic error in the observation of rainfall after the year 1983 or (b) a possible change in the meteorological factors around the test station (which is very unlikely since any meteorological change would generally be spread wide enough to cover more neighbouring stations). For both possibilities, the reasons have to be identified beyond any doubt before any corrective measure can be taken. A visit to the station, checking the exposure conditions, and taking the history from the observer will be very useful in trying to establish the reasons of this change in the behaviour.

5 CORRECTION AND COMPLETION OF RAINFALL DATA

5.1 General

Various primary and secondary validation tests create data outputs with can be used to flag the suspect values. Some records may also be missing due to non-observation or loss during recording or transmission. This identifies the need to fill data gaps and correction of errors. The process of filling the missing data by estimated values based on other observations is referred to as "Data Completion".

The methodology of data filling depends on the type of error, length of gap and the availability of suitable source records for estimation. After primary and secondary validation, a number of values will be flagged as incorrect or doubtful. Some records may be missing due to non-observation or loss on recording or transmission.

Incorrect and missing values will be replaced where possible by estimated values based on other observations at the same station or at neighbouring stations.

It must be recognised that values estimated from other gauges are inherently less reliable than the properly measured values. Doubtful original values will therefore be generally given the benefit of the doubt and will be retained in the record with a flag. Where no suitable neighbouring observations or stations are available, missing values will be left as 'missing' and incorrect values will be set to 'missing'.

Procedures for correction and completion shown in Figure 5.1 depend on the type of errors and the availability of suitable source records used as a basis for generating data estimates.

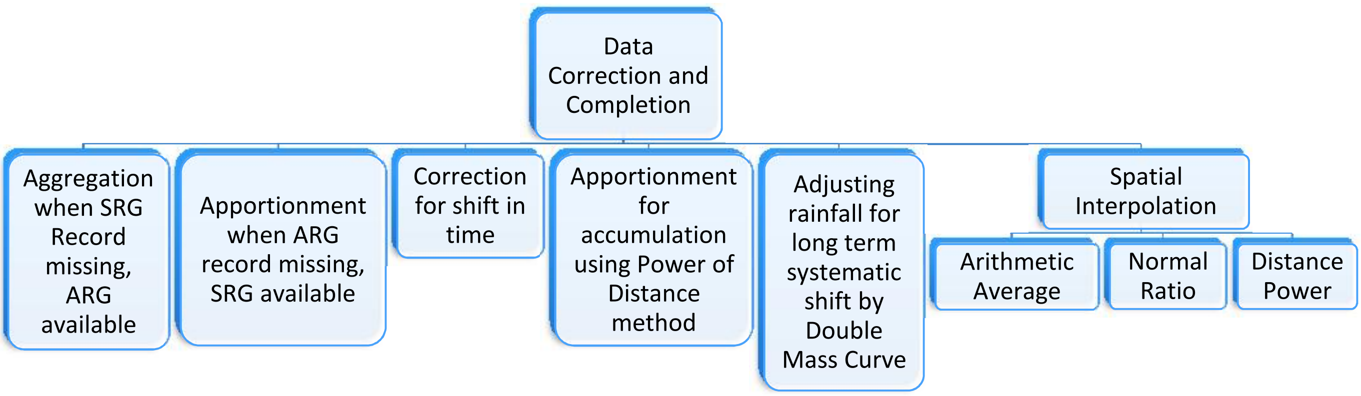


Figure 5.1: Methods for data correction and completion

5.2 Use of ARG and SRG data at one or more stations

5.2.1 General description

All observational stations equipped with an automatic rain gauge (ARG) should also have an ordinary or standard rain gauge (SRG) installed. One instrument can be used as a back-up and for correcting errors in the other in the event of failure of the instrument or the observer. The retention of an SRG at stations with an ARG is based on the view that the chances of malfunctioning of automatic type of equipment are higher.

Where an autographic record at a station is erroneous or missing and there are one or more adjoining stations at which autographic records are available, these may possibly be used to complete the missing values.

5.2.2 Data correction or completion procedure

Correction and completion of rainfall data using ARG and SRG data depends on which one of them has failed and the nature of the failure. The procedures to be followed in typical situations are explained below.

5.2.2.1 SRG record missing or faulty - ARG available

The record from the standard rain gauge may be missing or faulty due to a poor observation technique, a wrong or broken measuring glass or a leaking gauge. In these circumstances, it is reasonable to correct the erroneous standard rain gauge, or complete the data using the autographic records of the same station. The standard rain gauge data in such cases are made equal to that obtained from the autographic records. The standard rain gauges are normally observed one or two times in the day i.e. at 0830 hrs and 1730 hrs. The estimated values for such observations can be obtained by aggregating the hourly autographic records corresponding to these timings.

Example 5-1

Referring back to Example 4-4 it was found during scrutiny of rainfall data of neighbouring stations by multiple graphs that a few daily values at Anior station (Kheda catchment) are doubtful. One of these suspect values is 165 mm on 23/07/96 and there are a couple of instances (12th & 13th of Aug. 1996) where the values seem to have been shifted by a day.

Since autographic chart recorder (ARG) is also available at Anior station, it is possible to make a one-to-one comparison of daily rainfall totals obtained from both recorders. For this, the hourly data series obtained from ARG are used to compile the corresponding daily totals. The daily rainfall records thus obtained from SRG and ARG are tabulated together for an easy comparison, as shown in Table 5.1.

Table 5.1: Tabulation result for daily rainfall series obtained from SRG & ARG

Year	Month	Day	Anior MPS (ARG)	Anior MPA (SRG)
1996	7	16	11.00	11.00
1996	7	17	20.00	20.00
1996	7	18	8.00	8.00
1996	7	19	0.50	0.50
1996	7	20	12.00	12.00
1996	7	21	0.00	0.00
1996	7	22	0.00	0.00
1996	7	23	126.00	165.00
1996	7	24	15.50	15.50
1996	7	25	0.00	0.00
1996	7	26	0.00	0.00
1996	7	27	42.00	42.00
1996	7	28	190.00	190.00
1996	7	29	17.50	17.50
1996	7	30	0.00	0.00
1996	7	31	0.50	0.50
1996	8	1	3.50	3.50
1996	8	2	5.50	5.50
1996	8	3	3.50	3.50
1996	8	4	7.00	7.00
1996	8	5	0.00	0.00
1996	8	6	63.00	63.00
1996	8	7	55.00	55.00
1996	8	8	26.00	26.00
1996	8	9	0.00	0.00
1996	8	10	0.00	0.00
1996	8	11	2.50	2.5.00
1996	8	12	0.00	4.00
1996	8	13	4.00	18.00
1996	8	14	18.00	17.00
1996	8	15	17.00	0.00
1996	8	16	0.00	0.00
1996	8	17	0.00	0.00
1996	8	18	0.00	0.00
1996	8	19	0.00	0.00
1996	8	20	0.00	0.00
1996	8	21	0.00	0.00

Both of the above-mentioned suspicions are cleared after examining the tabulation results. Rainfall obtained from SRG (data type MPS) and ARG (data type MPA) on 23/07/96 is 165 and 126 mm respectively. At this stage the manuscript of SRG record and hourly tabulation of ARG record is referred to and confirmation made. Assuming that in this case the daily value of ARG record matches with the manuscript, and a look at the corresponding chart record confirms proper hourly tabulation, then the daily value is accordingly corrected from 165 mm to 126 mm making it equal to ARG daily total.

Secondly, the doubt regarding shift in SRG data around 12th, 13th August is also substantiated by the above tabulation results. The daily SRG data exhibits shift of one day from two independent comparisons and this does not warrant further confirmation from the manuscript. In such a straight forward situation, the SRG data of 12th, 13th& 14thAugust can be shifted forward by one day, i.e. to 13th, 14th& 15th August and the resulting void on 12th is to be filled by 0 mm rainfall (Refer Figure 5.2).

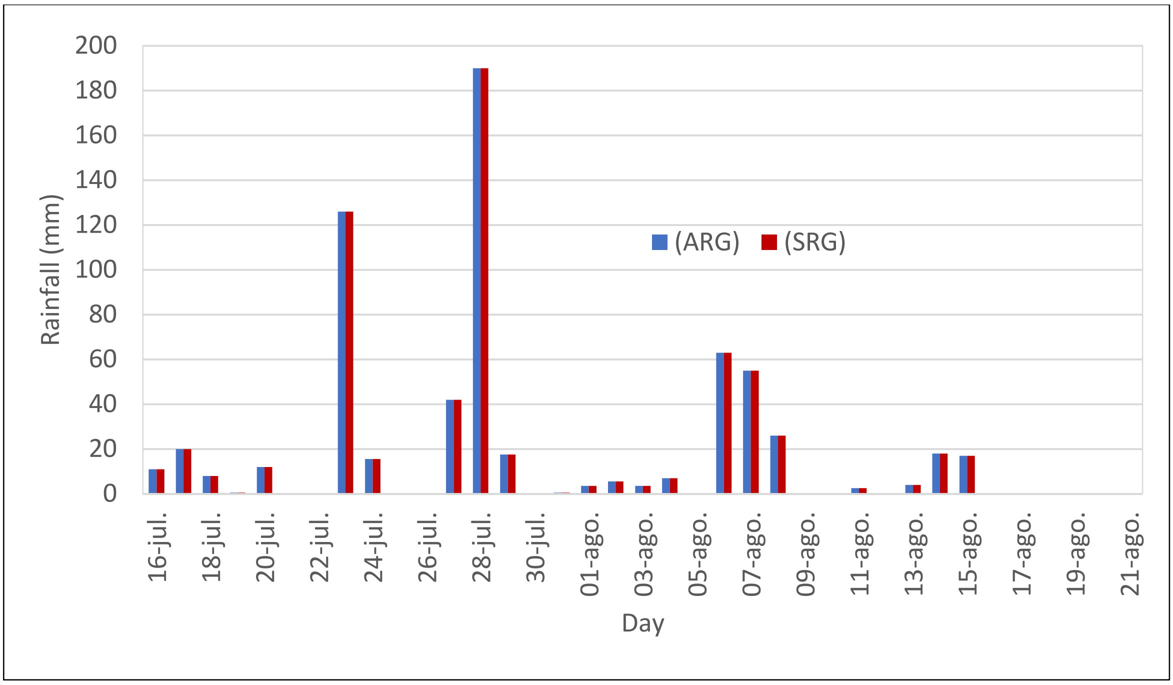


Figure 5.2: Corrected rainfall at Anior for July-August 1996

5.2.2.2 ARG record missing or faulty SRG available

The autographic record may be missing for various reasons such as for example the failure of the recording mechanism or blockage of the funnel. Under such situations, records from autographic gauges at neighbouring stations can be used in conjunction with the SRG at the station to complete the records. Essentially, this involves hourly distribution of the daily total from the SRG at the station by using reference to the hourly distribution at one or more neighbouring stations. Donor (or base) stations are selected by making comparison of cumulative plots of events in which autographic records are available at both stations and selecting the best available for data filling procedure.

Consider that the daily rainfall (from 08:30 hrs on previous day to 08:30 hrs on the day under consideration) at the station under consideration is D_{test} and the hourly rainfall for the same period at the selected adjoining station are $H_{base,i}$ (i = 1, 24). Then the hourly rainfall at the station under consideration, $H_{test,i}$ is obtained as:

$$H_{test,i} = D_{test} \cdot \frac{H_{base,i}}{\sum_{i=1}^{24} H_{base,i}} \qquad \text{Eqn. 5.1}$$

The procedure may be repeated for more than one base station and the average or resulting hourly totals calculated.

Example 5-2

Hourly rainfall data at Rahiol station (Kheda catchment) is considered for the period of July-August 1996. Though there is no missing data during the in this period under consideration, it is assumed that the rainfall values during 27–29 July 1996 are not available and are thus tried to be estimated on the basis of hourly distribution of rainfall at neighbouring stations.

Four neighbouring stations (Anior, Megharaj, Vadagam & Bayad) are available around this Rahiol station at which two days of hourly rainfall is required to be estimated. For this, first of all the hourly rainfall pattern of Rahiol station is tried to be correlated with one or more of the neighbouring stations. Data of a rainfall event in the area during 5-7 August 1996 is considered for identifying suitable neighbouring stations for estimates of hourly distribution. For this, hourly rainfall pattern of Rahiol station can be correlated with one or more of the neighbouring stations

Figure 5.3 shows the comparison of cumulative hourly rainfall between these five neighbouring stations. Vadagam and Anior stations show quite a high level of similarity with the Rahiol station. Distribution at Bayad station is also not very different from that at Rahiol. Megharaj station though shows a distinct behaviour then the rest four stations. Thus, for this case both Vadagam and Anior stations can be considered as the basis for estimating hourly distribution at Rahiol station.

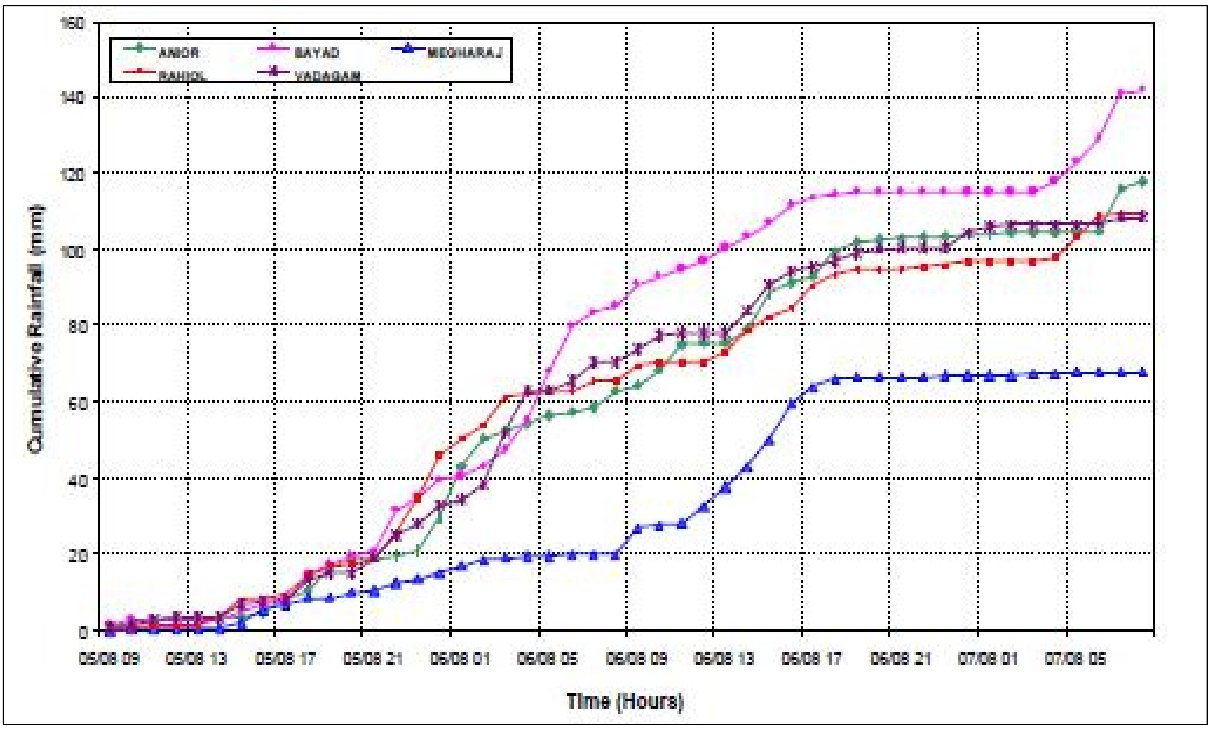


Figure 5.3: Plot of hourly rainfall distribution at Rahiol and surrounding stations

Hourly rainfall data at these three stations during the period 27-29 July 1996 for which it is assumed that the data is missing at Rahiol station is given in Table 5.2. The daily rainfall totals at Anior and Vadagam are found from hourly data for 28th and 29th July and are 190.0 & 17.5 mm and 168.0 & 24.0 mm respectively. Observed daily rainfall (SRG record) at Rahiol station for these dates are 152.0 mm and 28.0 mm respectively. It may be noted that totals as compiled from the hourly data (which is assumed to be missing in this example and would be so if such method is to be applied for the purpose of filling-in) is 144.0 mm and 28.0 mm respectively and is slightly different from the SRG value. The hourly values estimated for Rahiol ($P_{Rahiol,\ est,\ i}$) for 28th and 29th on the basis of that observation ed at Anior station ($P_{Anior,\ obs,\ i}$) are worked out as:

$$P_{Rahlol,est,i} = P_{Anior,obs,i} \times \frac{152.0}{190.0} \quad for\ each\ i^{th}\ hour\ on\ 28^{th} \qquad \text{Eqn. 5.2}$$

and

$$P_{Rahlol,est,i} = P_{Annor,obs,i} \times \frac{28.0}{17.5} \quad for\ each\ i^{th}\ hour\ on\ 29^{th} \qquad \text{Eqn. 5.3}$$

Similar estimates can be made on the basis of hourly rainfall observed at Vadagam. Both these estimates are averaged to get an overall estimate of the hourly rainfall distribution at Rahiol. These computations are self-explanatory from the Table 5.2

Table 5.2: Hourly distribution of observed daily rainfall by SRG on the basis of nearby hourly rainfall by ARG

Date/ Time	Observed Hourly rainfall (mm)			Estimated Rainfall at Rahiol (mm)		Avg.	Cum. Rahiol Observed	Cum. Rahiol Estimated
	Anior	Rahiol	Vadagam	As per rain distribution at				
		Assumed missing		Anior	Vadagam			
27/07/96 09:30	4.00	7.00	5.10	3.20	4.60	3.90	7.00	3.90
27/07/96 10:30	6.50	5.50	5.10	5.20	4.60	4.90	12.50	8.80
27/07/96 11:30	3.50	12.50	4.10	2.80	3.70	3.30	25.00	12.10
27/07/96 12:30	4.50	5.50	5.50	3.60	5.00	4.30	30.50	16.40
27/07/96 13:30	10.0	3.50	6.50	8.00	5.90	6.90	34.00	23.30
27/07/96 14:30	6.00	2.50	6.50	4.80	5.90	5.30	36.50	28.60
27/07/96 15:30	2.00	3.50	6.50	1.60	5.90	3.70	40.00	32.40
27/07/96 16:30	9.50	6.00	0.55	7.60	0.50	4.00	46.00	36.40
27/07/96 17:30	6.50	0.50	1.00	5.20	0.90	3.10	46.50	39.50
27/07/96 18:30	2.50	1.00	4.60	2.00	4.16	3.10	47.50	42.60
27/07/96 19:30	0.50	2.5.00	9.60	0.40	8.69	4.50	50.00	47.10
27/07/96 20:30	1.00	0.00	7.50	0.80	6.79	3.80	50.00	50.90
27/07/96 21:30	5.50	3.00	7.50	4.40	6.79	5.60	53.00	56.50
27/07/96 22:30	7.00	4.50	10.50	5.60	9.50	7.60	57.50	64.00
27/07/96 23:30	2.00	2.50	11.10	1.60	10.04	5.80	60.00	69.90
28/07/96 00:30	6.00	8.00	13.20	4.80	11.94	8.40	68.00	78.20
28/07/96 01:30	8.50	17.00	12.55	6.80	11.35	9.10	85.00	87.30
28/07/96 02:30	24.50	28.00	7.50	19.60	6.79	13.20	113.00	100.50
28/07/96 03:30	16.50	7.50	7.10	13.20	6.42	9.80	120.50	110.30
28/07/96 04:30	9.00	6.50	8.05	7.20	7.28	7.20	127.00	117.60
28/07/96 05:30	15.00	4.00	5.00	12.00	4.52	8.30	131.00	125.80
28/07/96 06:30	7.50	2.00	6.50	6.00	5.88	5.90	133.00	131.80
28/07/96 07:30	12.00	11.00	16.10	9.60	14.57	12.10	144.00	143.80
28/07/96 08:30	20.00	0.00	0.00	16.00	0.00	8.00	144.00	151.80
28/07/96 09:30	3.00	1.00	0.00	4.80	0.00	2.40	145.00	154.20
28/07/96 10:30	1.50	1.50	7.50	2.40	8.75	5.60	146.50	159.80
28/07/96 11:30	3.00	3.50	9.00	4.80	10.5	7.70	150.00	167.50
28/07/96 12:30	1.00	4.00	5.50	1.60	6.42	4.00	154.00	171.50
28/07/96 13:30	3.00	5.50	1.50	4.80	1.75	3.30	159.50	174.80
28/07/96 14:30	4.00	3.00	0.50	6.40	0.58	3.50	162.50	178.20
28/07/96 15:30	1.00	2.00	0.00	1.60	0.00	0.80	164.50	179.00
28/07/96 16:30	0.50	0.50	0.00	0.80	0.00	0.40	165.00	179.40
28/07/96 17:30	0.00	0.00	0.00	0.00	0.00	0.00	165.00	179.40
28/07/96 18:30	0.00	0.00	0.00	0.00	0.00	0.00	165.00	179.40
28/07/96 19:30	0.00	0.00	0.00	0.00	0.00	0.00	165.00	179.40
28/07/96 20:30	0.00	0.00	0.00	0.00	0.00	0.00	165.00	179.40
28/07/96 21:30	0.00	0.00	0.00	0.00	0.00	0.00	165.00	179.40
28/07/96 22:30	0.00	0.50	0.00	0.00	0.00	0.00	165.50	179.40
28/07/96 23:30	0.50	3.50	0.00	0.80	0.00	0.40	169.00	179.80
29/07/96 00:30	0.00	0.00	0.00	0.00	0.00	0.00	169.00	179.80
29/07/96 01:30	0.00	0.00	0.00	0.00	0.00	0.00	169.00	179.80
29/07/96 02:30	0.00	3.00	0.00	0.00	0.00	0.00	172.00	179.80

Date/ Time	Observed Hourly rainfall (mm)			Estimated Rainfall at Rahiol (mm)		Avg.	Cum. Rahiol Observed	Cum. Rahiol Estimated
	Anior	Rahiol	Vadagam	As per rain distribution at				
		Assumed missing		Anior	Vadagam			
29/07/96 03:30	0.00	0.00	0.00	0.00	0.00	0.00	172.00	179.80
29/07/96 04:30	0.00	0.00	0.00	0.00	0.00	0.00	172.00	179.80
29/07/96 05:30	0.00	0.00	0.00	0.00	0.00	0.00	172.00	179.80
29/07/96 06:30	0.00	0.00	0.00	0.00	0.00	0.00	172.00	179.80
29/07/96 07:30	0.00	0.00	0.00	0.00	0.00	0.00	172.00	179.80
29/07/96 08:30	0.00	0.00	0.00	0.00	0.00	0.00	172.00	179.80

For judging the efficacy of the procedure, a comparison is made between the observed (which was not missing actually) and estimated hourly rainfall values at Rahiol and is shown in Figure 5.4. It may be observed that there is a fairly good level of matching between the observed and the estimated hourly rainfall values. However, on many occasions the matching may not be so good and even then, it may be acceptable in view of no other way of estimation.

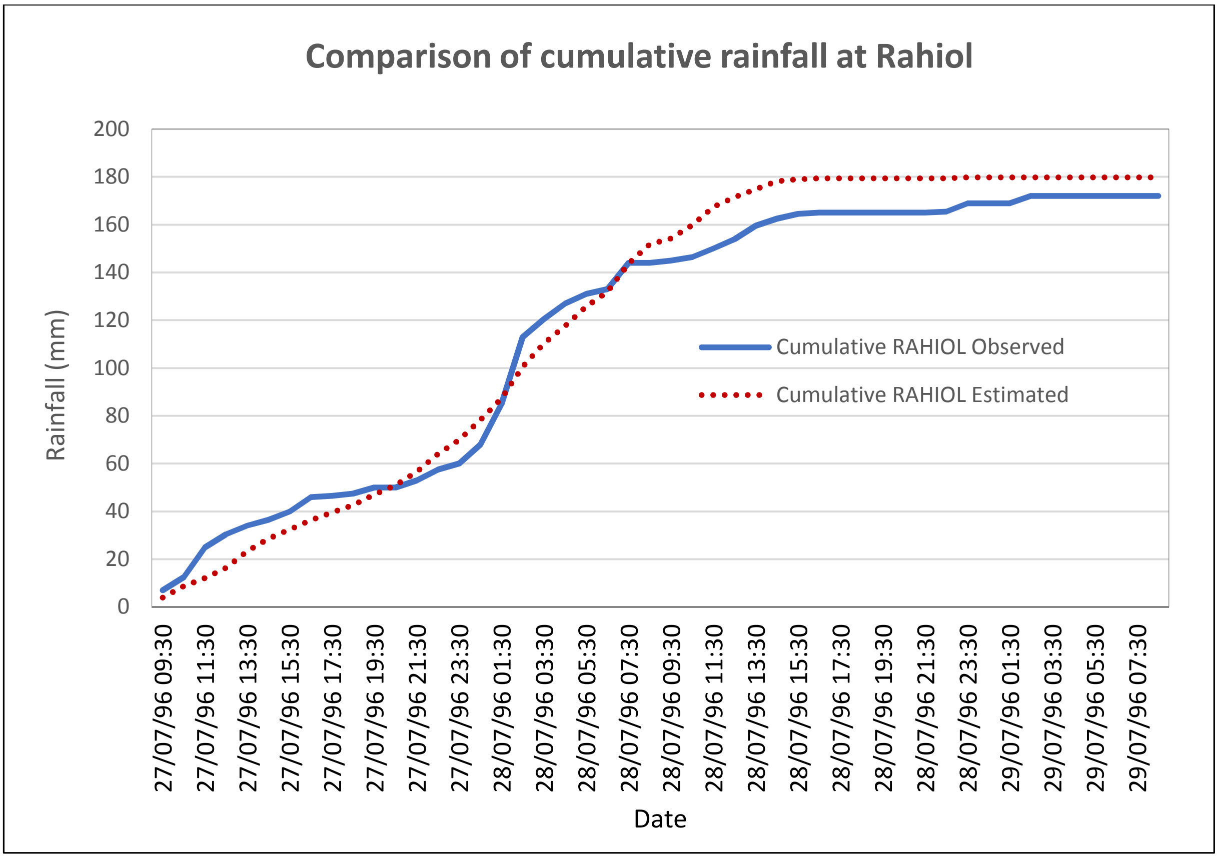


Figure 5.4: Plot comparison of observed and estimated hourly rainfall at Rahiol

5.3 Correcting for entries to wrong days

5.3.1 General description

Daily rainfall data are often entered to the wrong day especially following a period when no rainfall was observed. Identification of such mistakes is explained under secondary validation which identifies the occurrence of the time shift and quantifies its amount.

Correction for removing the shift in the data is done by either inserting the missing data or deleting the extra data points causing the shift (usually zero entries). While inserting or deleting data points care must be taken that only those data values are shifted which are affected by the shift. Though this type of correction is required frequently for daily data a similar procedure may be employed for other time intervals if a shift is identified.

5.3.2 Data correction procedure:

There are two important things to be considered while correcting the data for the identified shift in the data series.

1. the amount of shift and
2. the extent of data affected by the shift

The amount of shift is the number of days by which a group of daily data is shifted. For example, if the daily data in a certain month is shifted forward 2 days, then the amount of shift is 2 days. The extent of the shift may be longer or shorter than a month, and it is expressed in the number of days within each consecutive month. The data must be corrected by first deleting the unwanted data points from the desired location in the database. This deletion must be followed by shifting the affected data to fill up the deleted locations. Obviously, this will result in making a gap before the period where rainfall values were entered to the correct day. These must be filled with suitable entries (possibly by using other nearby stations). For example, if a 2-day shift extends to the end of the month, then the last 2 data points in the previous month must similarly be filled up with suitable entries. If a 2-day shift continues into the following month, the first two values of the next month are adjusted as well.

Example 5-3

Referring back to Example 4-5 wherein during validation by tabulation a time shift of one day was found to be present at Savlitank station. The tabulation of the data series of the nearby stations for the month of August 1984 is given in Table 5.3

It is clear from the tabulation that there is a one-day time shift in the data of Savlitank station. The data series of Savlitank station appears to be having a lag of one day in consequent rainfall events. The same shift is persisting for all 20 days as it can be confirmed by closely looking at the start and end times of five rainfall events (highlighted) one after another. If the manuscript records do not show any shift, it means that there has been an error while entering or handling the data that must therefore be corrected accordingly. Even if the records also show the same shift at Savlitank station, in clear-cut cases such as this one, it can be confidently attributed to the incorrect recording by the observer.

The corrected data series for Savlitank station is shown in the last column of Table 5.3 It may be seen that the data from the 3rd to the 20th of August is advanced by one day using simple copying and pasting option while editing the data series.

Table 5.3: Correction for shift in time in daily rainfall at station Savlitank

Date	Observed					Corrected
	Kapadwanj	Kathlal	Mahisa	Savlitank	Vadol	Savlitank
1/8/1984	0.00	0.00	0.00	0.00	0.00	0.00
2/8/1984	0.00	0.00	0.20	0.00	0.00	0.00
3/8/1984	152.40	99.30	157.40	0.00	39.30	150.00
4/8/1984	104.10	50.20	87.00	150.00	59.20	76.00
5/8/1984	7.70	12.00	18.00	76.00	13.10	16.00
6/8/1984	1.50	35.00	0.00	16.00	0.00	3.00
7/8/1984	0.00	0.00	0.00	3.00	0.00	0.00
8/8/1984	1.30	0.00	0.00	0.00	0.00	0.00
9/8/1984	0.00	13.00	0.00	0.00	0.00	0.00
10/8/1984	231.20	157.00	179.00	0.00	17.30	201.00
11/8/1984	43.20	18.30	64.00	201.00	63.20	26.00
12/8/1984	0.00	0.00	0.00	26.00	33.30	0.00
13/08/84	0.00	0.00	0.00	0.00	13.10	0.00
14/08/84	0.00	0.00	20.00	0.00	0.00	30.00
15/08/84	0.00	0.00	0.00	30.00	0.00	0.00
16/08/84	2.60	8.30	16.50	0.00	16.30	20.00
17/08/84	0.00	0.00	0.00	20.00	20.20	0.00
18/08/84	32.00	50.30	25.60	0.00	37.20	27.00
19/08/84	16.51	8.20	15.00	27.00	19.30	13.00
20/08/84	0.00	0.00	0.00	13.00	0.00	0.00

5.4 Apportionment for indicated and unindicted accumulations

5.4.1 General description

Where the daily rain gauge has not been read for a period of days and the total record represents an accumulation over a period of days identified in validation, the accumulated total is distributed over the period of accumulation by reference to rainfall at neighbouring stations over the same period.

5.4.2 Data correction procedure

The accumulated value of the rainfall and the affected period due to accumulation is known before initiating the correction procedure. Consider that:

number of days of accumulation = N_{acc}

accumulated rainfall as recorded = R_{acc}

Estimates of daily rainfall, for each day of the period of accumulation, at the station under consideration is made using spatial interpolation from the adjoining stations (in the first instance without reference to the accumulation total) using:

$$P_{est,j} = \frac{\sum_{i=1}^{N_{base}} \frac{P_{ij}}{D_i^{\,b}}}{\sum_{i=1}^{N_{base}} \frac{1}{D_i^{\,b}}} = \sum_{i=1}^{N_{base}} \left(P_{ij} \frac{\frac{1}{D_i^{\,b}}}{\sum_{i=1}^{N_{base}} \frac{1}{D_i^{\,b}}} \right) \qquad \text{Eqn. 5.4}$$

Where:

$P_{est,\,j}$ = estimated rainfall at the test station for j^{th} day

P_{ij} = observed rainfall at i^{th} neighbour station on j^{th} day

D_i = distance between the test and i^{th} neighbouring station

N_{base} = number of neighbouring stations considered for spatial interpolation

b=power of distance used for weighting individual rainfall value. Usually taken as 2.

The accumulated rainfall is then apportioned in the ratio of the estimated values on the respective days as:

$$P_{appor,j} = \frac{P_{est,j} * P_{tot}}{\sum_{j=1}^{N_{acc}} P_{est,j}} \qquad \forall j = 1 \text{ to } N_{acc} \qquad \text{Eqn. 5.5}$$

Where:

P_{tot}= accumulated rainfall as recorded

N_{acc}= number of days of accumulation

$P_{appor,j}$ = apportioned rainfall for j^{th} day during the period of accumulation

Example 5-4

Referring back to Example 4-9 wherein during validation of data at Dakor station, there was suspicion that there has been an accumulation of rainfall during the month of July 1995 which has not been recorded by the observer. The tabulation of data of Dakor and other neighbouring stations is given in Table 5.4

After verifying from the field observer, it may be possible to know the exact number of days for which accumulated value on 28th July has been reported. Assuming that it has been indicated by the observer that the value of 97.5 mm on 28th July is an accumulation of observations from the start of the 21st day onwards, this accumulated value should be distributed over 8 days. This accumulated value is distributed in proportion of the corresponding estimated values at Dakor station.

Table 5.4: Tabulation of daily rainfall for neighbouring stations

Test Station: Dakor station

On JULY 1995 Accumulation: 21st July to 28th July period of acc.= 8 days

Year	Month	Day	Dakor	Kathlal	Mahisa	Mahudha	Savlitank	Thasara
1995	7	11	0.00	7.00	10.00	1.50	27.00	9.00
1995	7	12	0.00	0.00	3.00	2.00	3.00	17.00
1995	7	13	0.00	45.00	0.00	0.00	0.00	0.00
1995	7	14	0.00	10.00	20.00	7.50	0.00	7.00
1995	7	15	0.00	14.00	50.00	33.50	24.00	77.00
1995	7	16	0.00	0.00	8.00	9.50	25.00	8.00
1995	7	17	0.00	20.00	4.00	1.00	0.00	22.00
1995	7	18	0.00	10.00	8.00	1.00	6.00	11.00
1995	7	19	0.00	23.00	20.00	43.00	27.00	16.00
1995	7	20	0.00	0.00	35.00	32.50	14.00	48.00
1995	7	21	0.00	57.00	27.00	23.00	14.00	56.00
1995	7	22	0.00	0.00	6.00	7.00	4.00	0.00
1995	7	23	0.00	0.00	4.00	12.00	2.00	27.00
1995	7	24	0.00	10.00	0.00	0.00	0.00	0.00
1995	7	25	0.00	11.00	10.00	3.00	6.00	3.00
1995	7	26	0.00	25.00	0.00	10.00	5.00	8.00
1995	7	27	0.00	18.00	3.00	4.00	25.00	9.00
1995	7	28	97.50`	25.00	24.00	46.00	3.00	12.00
1995	7	29	16.70	40.00	4.00	6.00	0.00	0.00
1995	7	30	6.8.00	45.00	34.00	22.00	62.00	52.00
1995	7	31	0.00	10.00	3.00	13.00	39.00	9.00

The estimation procedure is outlined in the description above assuming the value of exponent is 2.0. The distances and computation of weights of the neighbouring stations computed is given in Table 5.5.

The estimated daily rainfall based on the weighted average of the neighbouring station is computed and given in Table 5.6. The sum of this estimated daily rainfall for the 8 days of accumulation from 21st to 28th is found to be equal to 104.1 mm. The spatially averaged rainfall estimate is proportionally reduced so that the total of this apportioned rainfall equals the accumulated total of 97.5 mm. This is done by multiplying the spatial estimate by a factor of (97.5/104.1) as shown in Table 5.6.

Table 5.5: Computation of normalised weights for neighbouring stations on the basis of the Distance Power method

Sl. No.	Name of neighbouring stations	Distance from Dakor	Factor	Station weight
		Di	$(1/Di)^2$	$\{(1/D_i)^2\}/\sum\{(1/D_i)^2\}$
1	Thasara	8.250	0.015	0.573
2	Mahisa	13.950	0.005	0.200
3	Kathlal	22.120	0.002	0.080
4	Mahudha	22.700	0.002	0.076
5	Savlitank	23.400	0.002	0.071
	SUM		0.026	1.000

Table 5.6: Computation of spatial estimate during period of accumulation and its distribution

Date	Observed	Weighted Rainfall (for Dakor) at					Weighted Average	Corrected
	Dakor	Kathlal	Mahisa	Mahudha	Savlitank	Thasara		Dakor
	Station weight	0.080	0.200	0.076	0.071	0.573	Rest j	Rest j * 97.5 / 104.544
11/7/95	0.00	0.56	2.00	0.11	1.92	5.16	9.75	0.00
12/7/95	0.00	0.00	0.60	0.15	0.21	9.74	10.71	0.00
13/07/95	0.00	3.60	0.00	0.00	0.00	0.00	3.60	0.00
14/07/95	0.00	0.80	4.00	0.57	0.00	4.01	9.38	0.00
15/07/95	0.00	1.12	10.00	2.55	1.70	44.12	59.49	0.00
16/07/95	0.00	0.00	1.60	0.72	1.78	4.58	8.68	0.00
17/07/95	0.00	1.60	0.80	0.08	0.00	12.61	15.08	0.00
18/07/95	0.00	0.80	1.60	0.08	0.43	6.30	9.21	0.00
19/07/95	0.00	1.84	4.00	3.27	1.92	9.17	20.19	0.00
20/07/95	0.00	0.00	7.00	2.47	0.99	27.50	37.97	0.00
21/07/95	0.00	4.56	5.40	1.75	0.99	32.09	44.79	41.77
22/07/95	0.00	0.00	1.20	0.53	0.28	0.00	2.02	1.88

Date	Observed	Weighted Rainfall (for Dakor) at					Weighted Average	Corrected
	Dakor	Kathlal	Mahisa	Mahudha	Savlitank	Thasara		Dakor
	Station weight	0.080	0.200	0.076	0.071	0.573	Rest j	Rest j * 97.5 / 104.544
23/07/95	0.00	0.00	0.80	0.91	0.14	15.47	17.33	16.16
24/07/95	0.00	0.80	0.00	0.00	0.00	0.00	0.80	0.75
25/07/95	0.00	0.88	2.00	0.23	0.43	1.72	5.25	4.90
26/07/95	0.00	2.00	0.00	0.76	0.36	4.58	7.70	7.18
27/07/95	0.00	1.44	0.60	0.30	1.78	5.16	9.28	8.65
28/07/95	97.50	2.00	4.80	3.50	0.21	6.88	17.39	16.21
29/07/95	16.70	3.20	0.80	0.46	0.00	0.00	4.46	16.70
30/07/95	6.80	3.60	6.80	1.67	4.40	29.80	46.27	6.80
31/07/95	0.00	0.80	0.60	0.99	2.77	5.16	10.31	0.00

5.5 Adjusting rainfall data for long term systematic shifts

5.5.1 General description

Double mass analysis is a technique to ensure that the data series is reasonably homogenous before any statistical inference can be drawn. The possible non-homogeneities in series such as jumps, trends or long-term systematic shifts in rainfall series are detected by investigating the ratio of accumulated values of two series. Double Mass Analysis is normally used with the aggregated series. The double mass analysis technique is used in data validation to detect significant long-term systematic shift in rainfall data. The same technique can be used to adjust the suspect data. Inconsistency in data is demonstrated by a distinct change in the slope of the double mass curve and may be due to a change in instrument location or exposure or measurement technique. It does not imply that either period is incorrect - only that it is inconsistent. The double mass curve shows a straight line if the test-series is homogeneous. A jump in the test-series creates a break in the double mass curve, whereas a trend creates a curved line. When there is a visible change in slope of the double mass plot after certain period, the break needs to be investigated. The data can be made consistent by adjusting so that there is no break in the resulting double mass curve. The existence of a discontinuity in the double mass plot does not in itself indicate which part of the curve should be adjusted (before or after the break). It is usual practice to adjust the earlier part of the record so that the entire record is consistent with the present and continuing record. There may be circumstances however, when the adjustment is made to the latter part, where an erroneous source of the inconsistency is known or where the record has been discontinued. The correction procedure is described below.

5.5.2 Data correction procedure

Consider a double mass plot shown in Figure 5.5. There is a distinct break at point A in the double mass plot and records before this point are inconsistent with present measurements and require adjustment. The adjustment consists of either adjusting the slope of the double mass curve before the break point to confirm to the slope after it or adjusting the slope in the later part to confirm with that of the previous portion. The decision to be considered for the period of adjustment depends on the application of data and on the reasons for the exhibited in-homogeneity. For example, if the change in behaviour after a certain point in time is due to an identified systematic error then obviously the portion after the break point will be adjusted. On the other hand, if shift is due to the relocation of an observation station in the past then for making the whole data set consistent with the current location the portion before the break needs to be corrected.

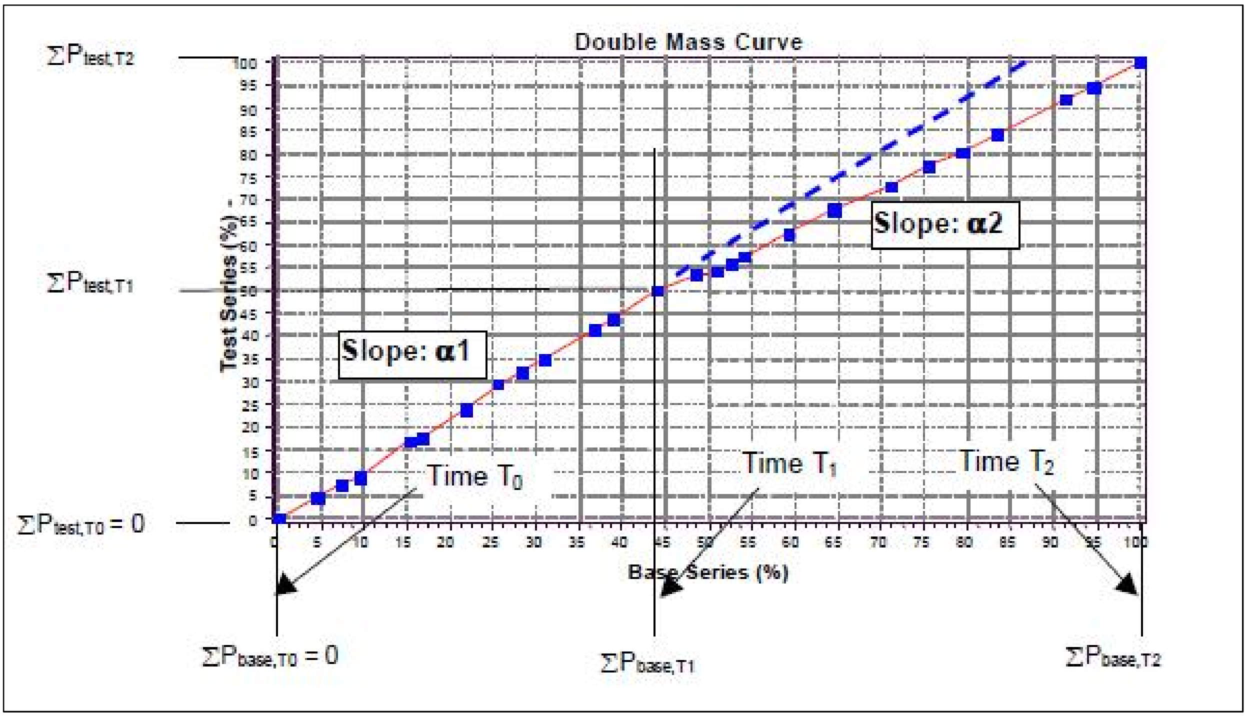


Figure 5.5: Definition sketch for double mass analysis

Considering the double mass plot shown in Figure 5.5, the break points occurs at time T_1 and if the start and end times of the period under consideration are T_0 and T_2 respectively, then the slopes of the curve before and after the break point can be expressed as:

$$\alpha_1 = \frac{\sum_{i=1}^{T_1} P_{test,i}}{\sum_{i=1}^{T_1} P_{base,i}} \qquad \text{Eqn. 5.6}$$

and

$$\alpha_2 = \frac{\sum_{i=T_0}^{T_2} P_{test,i} - \sum_{i=T_0}^{T_1} P_{test,i}}{\sum_{i=T_0}^{T_2} P_{base,i} - \sum_{i=T_0}^{T_1} P_{base,i}} \qquad \text{Eqn. 5.7}$$

In case the earlier portion between T_0 and T_1 is correction factor and the corrected observations respectively as:

$$P_{corr,i} = P_{test,i} \times \frac{\alpha_2}{\alpha_1} \qquad \text{Eqn. 5.8}$$

After making such correction the double mass curve can again be plotted to see that there is no significant change in the slope of the curve.

The double mass curve technique is usually applied to aggregated data and carried out annually. Aggregated daily data should be used to determine precisely when the change in the data trend begun, it However, there are circumstances where the technique might be applied to daily data to date the beginning of an instrument fault such as a leaking gauge. Once an inconsistency has been identified, the adjustment should be applied to all subsequent data intervals as long as the modified slope persists.

Example 5-5

The long-term data series of rainfall for the period 1970 to 1996 was considered at Vadol station (in Kheda catchment) for double mass analysis taking three nearby stations Kapadwanj, Mahisa and Thasara. It was observed that the test station (Vadol) records shows that there has been a significant change in the amount of rain received after the year 1983. This can be easily seen from break point marked in the double mass curve shown in Figure 5.6 that the behaviour of the test station suddenly changes after about half of the time period under consideration.

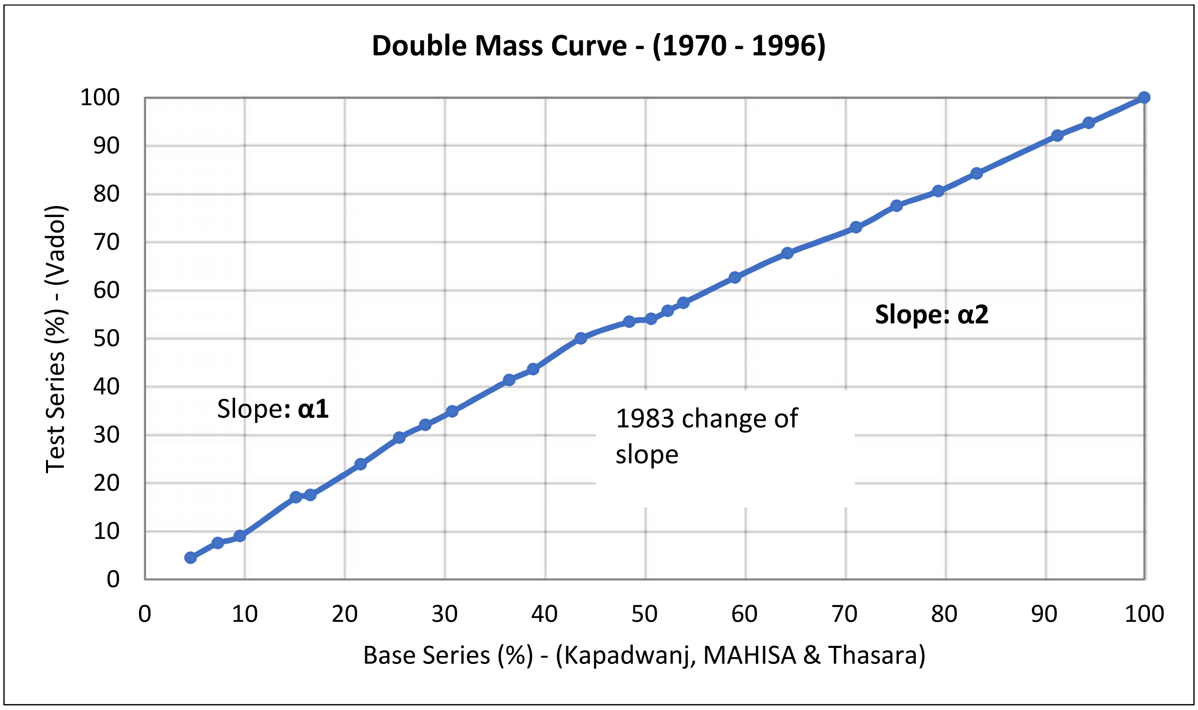


Figure 5.6: Double mass curve for station Vadol showing a change in slope of the curve after about half of the period under consideration.

Assuming that, on the basis of a visit to the station and feedback from the observer, it has been found that the exposure conditions at the rain gauge site have not been up to the desired standards. If the lower rainfall catches after 1983 can be confidently attributed to such improper exposure conditions then the second half of the data series after year 1983 can be adjusted so as to correspond to the actual rainfall occurring at the station had the normal exposure conditions existed. This is done carrying out following computations:

As is apparent from Figure 5.7 and the results of the Double Mass analysis given in Table 5.7, that from the year 1984 onwards, the rainfall received at Vadol station is comparatively less than in the previous13 year period in relation to the base stations Kapadwanj, Mahisa & Thasara around it.

Table 5.7: Results of the Double Mass Analysis

Test series: Vadol

Base series: Kapadwanj Weight: 0.33

Mahisa Weight: 0.33

Thasara Weight: 0.33

Period	BASE Station			TEST Station			Ratios	
	Rainfall mm	Cum mm	Perc	Amount mm	Cum mm	Perc	(6)/(3)	(7)/(4)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1970	767.40	767.00	4.60	624.40	624.40	4.50	0.81	0.98
1971	454.00	1221.00	7.30	426.00	1050.40	7.60	0.86	1.04
1972	372.50	1594.00	9.50	197.90	1248.30	9.00	0.78	0.94
1973	935.30	2529.00	15.10	1114.20	2362.50	17.00	0.93	1.13
1974	240.30	2770.00	16.60	72.80	2435.30	17.60	0.88	1.06
1977	843.80	3613.00	21.60	882.80	3318.10	23.90	0.92	1.11
1978	646.40	4260.00	25.50	758.80	4076.90	29.40	0.96	1.15
1979	436.70	4696.00	28.10	370.20	4447.10	32.10	0.95	1.14
1980	450.20	5147.00	30.80	388.90	4836.00	34.90	0.94	1.13
1981	950.00	6097.00	36.50	898.10	5734.10	41.40	0.94	1.13
1982	403.60	6500.00	38.90	320.10	6054.20	43.70	0.93	1.12
1983	801.40	7302.00	43.70	882.10	6936.30	50.00	0.95	1.15
1984	806.00	8108.00	48.50	475.10	7411.40	53.50	0.91	1.10
1985	364.20	8472.00	50.70	82.80	7494.20	54.10	0.88	1.07
1986	281.50	8753.00	52.30	234.00	7728.20	55.70	0.88	1.06
1987	257.70	9011.00	53.90	227.50	7955.70	57.40	0.88	1.06
1988	866.10	9877.00	59.10	734.50	8690.20	62.70	0.88	1.06
1989	877.00	10754.00	64.30	693.30	9383.50	67.70	0.87	1.05
1990	1145.00	11899.00	71.20	746.00	10129.50	73.10	0.85	1.03
1991	682.70	12582.00	75.20	618.10	10747.60	77.50	0.85	1.03
1992	697.70	13280.00	79.40	422.20	11169.80	80.60	0.84	1.01
1993	639.80	13919.00	83.20	512.80	11682.60	84.30	0.84	1.01
1994	1350.00	15269.00	91.30	1083.30	12765.90	92.10	0.84	1.01
1995	525.00	15794.00	94.50	371.60	13137.50	94.80	0.83	1.00
1996	926.70	16721.00	100.00	725.00	13862.50	100.00	0.83	1.00

Total number of periods analysis: 25

The average slopes of the double mass curve before and after this break can be worked out from the computations shown in Table 5.7

as:

$$\alpha_1 = \frac{\sum_{i=1}^{T_1} P_{test,j}}{\sum_{i=1}^{T_1} P_{base,i}} = \frac{6936}{7302} = 0.9498 \qquad \text{Eqn. 5.9}$$

and

$$\alpha_2 = \frac{\sum_{i=T_0}^{T_2} P_{test,i} - \sum_{i=T_0}^{T_1} P_{test,i}}{\sum_{i=T_0}^{T_2} P_{base,i} - \sum_{i=T_0}^{T_1} P_{base,i}} = \frac{13862-6936}{16721-7302} = 0.7353 \qquad \text{Eqn. 5.10}$$

Thus, the correction factor, if the latter portion is to be corrected to exhibit an average slope of α_1 is:

Correction Factor $= \frac{\alpha_2}{\alpha_1} = \frac{0.9498}{0.7353} = 1.2916$

Thus, all the rainfall values after the year 1983 have to be increased by a factor of 1.2916 to correct the rainfall data at Vadol for improper exposure condition and thus to make it consistent in time. This is done by carrying out data series transformation using linear algebraic option.

Such a correction when employed would make the double mass curve correspond to the dashed line shown after the break point in Figure 5.6. The double mass curve after adjusting the data series is given in Figure 5.7 and the corresponding tabular analysis results in Table 5.8. It may be noted that the double mass curve after the data series is corrected beyond 1983 shows a consistent trend throughout.

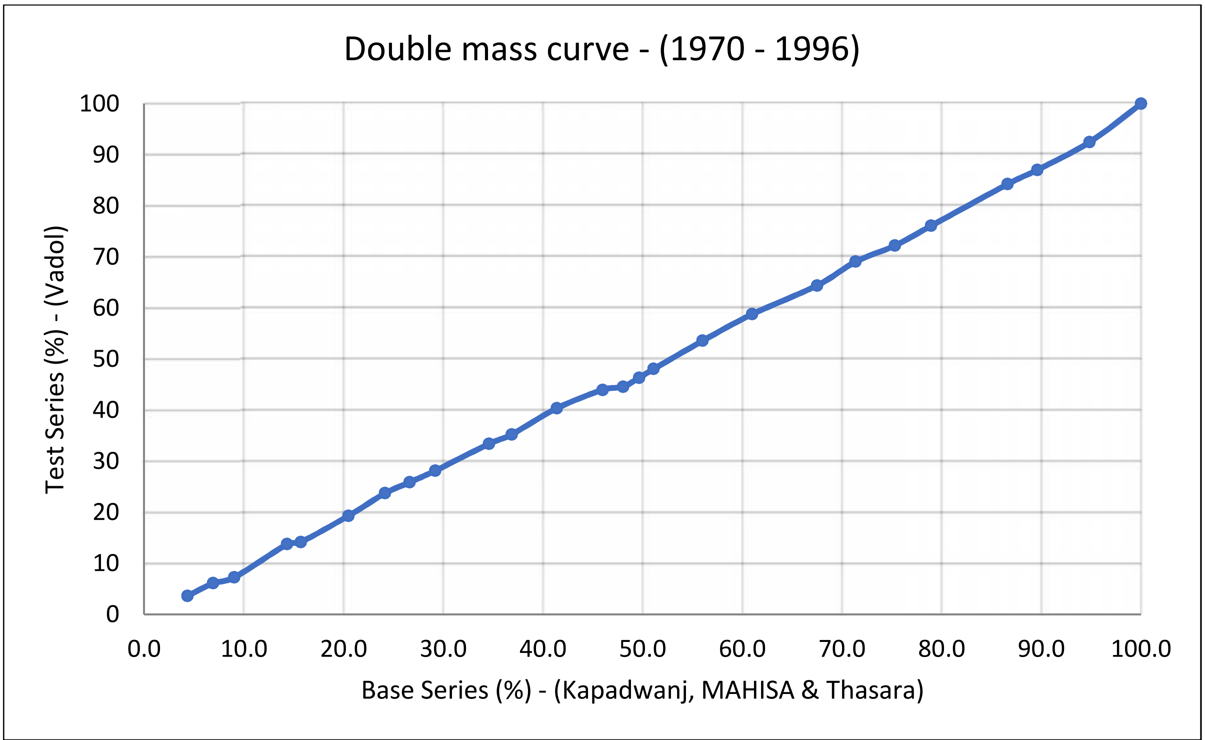


Figure 5.7: Double mass plot after adjusting data for the period of inconsistency

Table 5.8: Results of double mass analyses after adjusting data for the period of inconsistency

Double mass analysis

Test series:	Vadol	
Base series:	Kapadwanj	Weight 0.33
	Mahisa	Weight 0.33
	Thasara	Weight 0.33

Period	BASE			TEST			Ratios	
	Rainfall mm	Cum mm	Perc	Amount mm	Cum mm	Perc	(6)/(3)	(7)/(4)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1970	767.40	767.00	4.40	624.40	624.40	3.60	0.81	0.84
1971	454.00	1221.00	6.90	426.00	1050.40	6.10	0.86	0.88
1972	372.50	1594.00	9.00	197.90	1248.30	7.30	0.78	0.80
1973	935.30	2529.00	14.30	1114.20	2362.50	13.80	0.93	0.96
1974	240.30	2770.00	15.70	72.8.00	2435.30	14.20	0.88	0.90
1977	843.80	3613.00	20.50	882.80	3318.10	19.30	0.92	0.94
1978	646.40	4260.00	24.20	758.80	4076.90	23.70	0.96	0.98
1979	436.70	4696.00	26.60	370.20	4447.10	25.90	0.95	0.97
1980	450.20	5147.00	29.20	388.90	4836.00	28.20	0.94	0.96
1981	950.00	6097.00	34.60	898.10	5734.10	33.40	0.94	0.97
1982	403.60	6500.00	36.90	320.10	6054.20	35.30	0.93	0.96
1983	801.40	7302.00	41.40	882.10	6936.30	40.40	0.95	0.98
1984	806.00	8108.00	46.00	613.80	7550.10	44.00	0.93	0.96
1985	364.20	8472.00	48.10	107.00	7657.10	44.60	0.90	0.93
1986	281.50	8753.00	49.70	302.30	7959.40	46.40	0.91	0.93
1987	257.70	9011.00	51.10	293.90	8253.30	48.10	0.92	0.94
1988	866.10	9877.00	56.00	948.90	9202.20	53.60	0.93	0.96
1989	877.00	10754.00	61.00	895.70	10097.90	58.80	0.94	0.96
1990	1145.00	11899.00	67.50	963.80	11061.70	64.40	0.93	0.95
1991	682.70	12582.00	71.40	798.50	11860.20	69.10	0.94	0.97
1992	697.70	13280.00	75.30	545.50	12405.70	72.30	0.93	0.96
1993	639.80	13919.00	79.00	662.50	13068.20	76.10	0.94	0.96
1994	1350.00	15269.00	86.60	1399.50	14467.70	84.30	0.95	0.97
1995	525.00	15794.00	89.60	480.10	14947.80	87.10	0.95	0.97
1996	926.70	16721.00	94.90	936.60	15884.40	92.50	0.95	0.98
1997	907.00	17628.00	100.00	1283.90	17168.30	100.00	0.97	1.00

Total number of periods of analysis: 26

5.6 Using spatial interpolation to interpolate erroneous and missing values

5.6.1 General description

Spatial Interpolation using neighbouring stations is widely applied to fill-in missing data or correct rainfall values identified as erroneous. The adjoining stations are selected on the basis of the proximity criterion (i.e. that they must lie within the specified radius from the test station where data are filled).

Missing data and data identified as erroneous by validation can be substituted by interpolation from neighbouring stations. These procedures are widely applied to daily rainfall. Estimated values of the rainfall using such interpolation methods are obtained for as many data point as required. However, in practice only a limited number of data values will be estimated at a stretch. Three analytical procedures for estimating rainfall using such spatial interpolation methods are described below:

5.6.2 Arithmetic average method

This method is applied if the average annual rainfall of the station under consideration is within 10% of the average annual rainfall at the adjoining stations. The erroneous or missing rainfall at the station under consideration is estimated as the simple average of neighbouring stations. Thus, if the estimate for the erroneous or missing rainfall at the station under consideration is P_{test} and the rainfall at M adjoining stations is $P_{base,\,i}$ (i = 1 to M), then:

$$P_{test} = \frac{1}{M}(P_{base,1} + P_{base,2} + P_{base,3} + \ldots\ldots + P_{base,M}) \qquad \text{Eqn. 5.11}$$

Usually, averaging of three or more adjoining stations is considered to give a satisfactory estimate.

Example 5-6

Consider the station Balasinor (in Kheda catchment) at which the daily rainfall record is not available for the year 1988. There are a few stations like Mahisa, Savlitank and Vadol around this station at which daily observation are available. It is desired to see the appropriateness of the arithmetic average method of spatial interpolation at station Balasinor for the missing period on the basis of these neighbouring stations.

First the long-term average of these stations is considered to get an idea of variability. The annual rainfalls at these stations are:

For Balasinor = N_{test} = 715 mm

For Mahisa = $N_{base,2}$ = 675 mm

For Savlitank = $N_{base,5}$ = 705 mm

For Vadol = $N_{base,4}$ = 660 mm

It may be seen that the difference in the normal annual rainfall at the three base stations is about 5.5, 1.3 and 7.8 %, and thus the simple arithmetic average method for obtaining the estimates of daily rainfall at Balasinor station can be employed.

The arithmetic averaging can be carried out by employing the process of algebraic series transformation on the three base series taken together and multiplying them with an equal weight of 0.333. Table 5.9 shows the computation of the daily rainfall estimates at Balasinor station on the basis of above three adjoining (base) stations.

Table 5.9: Estimation of daily rainfall at station Balasinor by arithmetic average method

Date	Observed Rainfall (mm)			Estimated Rainfall (mm) Balasinor
	Mahisa	Savlitank	Vadol	
	Station Weights			
	0.333	0.333	0.333	
12/07/88	0.00	0.00	0.00	0.00
13/07/88	13.00	0.00	2.00	5.00
14/07/88	25.00	50.00	37.20	37.40
15/07/88	46.00	30.00	42.00	39.33
16/07/88	97.00	50.00	17.00	54.67
17/07/88	4.00	3.00	5.00	4.00
18/07/88	8.00	3.00	14.00	8.33
19/07/88	7.00	15.00	16.00	12.67
20/07/88	21.00	28.00	18.50	22.50
21/07/88	6.00	6.00	3.00	5.00
22/07/88	62.00	45.00	28.00	45.00
23/07/88	15.00	18.00	38.00	23.67
24/07/88	5.00	8.00	4.00	5.67
25/07/88	18.00	10.00	4.80	10.93
26/07/88	6.00	15.00	20.00	13.67
27/07/88	43.00	0.00	12.00	18.33
28/07/88	40.00	125.00	47.40	70.80
29/07/88	11.00	21.00	17.60	16.53
30/07/88	0.00	5.00	6.60	3.87
31/07/88	11.00	11.00	5.20	9.10

5.6.3 Normal ratio method

This method is preferred if the average (or normal) annual rainfall of the station under consideration differs from the average annual rainfall at the adjoining stations by more than 10%. The erroneous or missing rainfall at the station under consideration is estimated as the weighted average of adjoining stations. The rainfall at each of the adjoining stations is weighted by the ratio of the average annual rainfall at the station under consideration and average annual rainfall of the adjoining station. The rainfall for the erroneous or missing value at the station under consideration is estimated as:

$$P_{test} = \frac{1}{M}\left(\frac{N_{test}}{N_{base,1}}P_{base,1} + \frac{N_{test}}{N_{base,2}}P_{base,2} + \frac{N_{test}}{N_{base,3}}P_{base,3} + \ldots + \frac{N_{test}}{N_{base,M}}P_{base,M}\right) \qquad \text{Eqn. 5.12}$$

Where:

N_{test}= annual rainfall at the station under consideration

$N_{base,\,i}$= annual rainfall at the adjoining stations (for i = 1 to M)

A minimum of three adjoining stations must be generally used for obtaining good estimates using the Normal Ratio method.

Example 5-7

Consider the station Balasinor (in Kheda catchment) again at which the daily rainfall record is not available for the year 1988. Assuming that the record for the neighbouring stations like Mahisa & Savlitank and Vadol around this station is also not available. However, records for two stations Kapadwanj and Thasara which are at comparatively farther distance from Balasinor station are available. It is desired to see the appropriateness of the arithmetic average and normal ratio method of spatial interpolation at station Balasinor for a test period during the year 1984.

First, the long-term average of these stations is considered to get an idea of variability. The annual rainfall at these stations is obtained from 20-25 years of data between 1970 to 1997 as:

Annual rainfall

Balasinor	=	N_{test}	=	715 mm
Kapadwanj	=	$N_{base\ 1}$	=	830 mm
Thasara	=	$N_{base\ 3}$	=	795 mm

It may be seen that difference in the normal annual rainfall at the two base stations is about 16.0 and 11.2 % respectively which exceeds the 10% criterion. Therefore, the Normal Ratio method for obtaining the estimates of daily rainfall at Balasinor station is applicable.

First, the normalised weights for the two stations are derived by obtaining the ratio of test station normal and base station normal. These are obtained as:

Normalised weight for Kapadwanj $= \frac{1}{M}\frac{N_{test}}{N_{Base,1}} = \frac{1\,715}{2\,830} = 0.431$ and

Normalised weight for Thasara= $\frac{1}{M}\frac{N_{test}}{N_{Base,2}} = \frac{1\,715}{2\,795} = 0.450$

The normalised averaging can be carried out by employing the process of algebraic series transformation on the two base series taken together and multiplying them with weights of 0.431 and 0.450 respectively. For a qualitative comparison, estimates by arithmetic averaging are worked out. Since the data for 1984 Balasinor are not actually missing, the observed data is also tabulated along with the two estimated records using the two methods in Table 5.10.

Table 5.10: Estimation of daily rainfall at station Balasinor by arithmetic average and normal ratio method

Date	Observed Rainfall (mm)		Rainfall at Balasinor (mm)		
			Estimated		Observed
	Kapadwanj	Thasara	Arithmetic weights	Normal Ratio	
			0.5 & 0.5	0.431 & 0.450	
25/08/73	0.00	0.00	0.00	0.00	8.00
26/08/73	0.00	4.40	2.20	2.00	2.00
27/08/73	0.00	4.00	2.00	1.80	2.00
28/08/73	0.00	0.00	0.00	0.00	2.00
29/08/73	35.00	8.60	21.80	19.00	24.00
30/08/73	86.00	33.00	59.50	51.90	54.00
31/08/73	119.00	170.80	144.90	128.10	130.00
01/09/73	36.00	107.00	71.50	63.70	71.80
02/09/73	25.00	6.00	15.50	13.50	20.00
03/09/73	35.00	21.00	28.00	24.50	20.00
04/09/73	12.00	34.00	23.00	20.50	30.00
05/09/73	17.00	21.00	19.00	16.80	15.00
06/09/73	8.00	3.00	5.50	4.80	5.6.00
07/09/73	71.00	54.00	62.50	54.90	58.00
08/09/73	113.00	43.80	78.40	68.40	66.00
09/09/73	4.00	0.00	2.00	1.70	0.00
10/09/73	0.00	0.00	0.00	0.00	2.00

It may be seen from the above results that on an average the observed and estimated rainfall matches fairly well. Since the above is a very small sample for judging the performance of the two averaging methods, but the suitability of the Normal Ratio method is implied since it would maintain the long-term relationship between the three stations with respect to the station normal rainfalls.

5.6.4 Distance power method

In precipitation, stations in closer proximity have better correlation with the test station. Therefore, in this method, missing data at a test station are estimated by weighted averages of observations at the neighbouring stations. The weights are inversely proportional with some power of distance between the Test station and the neighbouring stations. An exponent of 2 is most commonly used with the distances to obtain the weighted average. This method weights neighbouring stations on the basis of their distance from the station under consideration, on the assumption that the closer stations are better correlated than those further away, and that beyond a certain distance they are insufficiently correlated to be of use. Spatial interpolation is made by weighing the adjoining station rainfall as inversely proportional to some power of the distances from the station under consideration.

In this method four quadrants are delineated by north-south and east-west lines passing through the rain gauge station under consideration, as shown in Figure 4.11. A circle is drawn of radius equal to the distance within which significant correlation is assumed to exist between the rainfall data, for the time interval under consideration. The adjoining stations are now selected on the basis of the following:

- The neighbouring stations must lie within the specified radius having significant spatial correlation with one another.
- A maximum number of 8 neighbouring stations are sufficient for estimation of spatial average.
- An equal number of stations from each of the four quadrants is preferred for minimising any directional bias. However, due to the prevailing wind conditions or orographic effects, spatial heterogeneity may be present. In such cases normalised values rather than actual values should be used in interpolation.

The spatially interpolated estimate of the rainfall at the station under consideration is obtained as:

$$P_{est,j} = \frac{\sum_{i=1}^{M_{base}} P_{i,j}/{D_i}^b}{\sum_{i=1}^{M_{base}} 1/{D_i}^b} \qquad \text{Eqn. 5.13}$$

Where:

$P_{est,j}$ = estimated rainfall at the test station at time j

$P_{i,j}$ = observed rainfall at the neighbour station i at time j

D_i = distance between the test and the neighbouring station i

M_{base} =number of neighbouring stations taken into account.

B = power of distance D used for weighting rainfall values at individual station

To correct for the sources of heterogeneity, e.g. orographic effects, normalized values must be used in place of actual rainfall values at the neighbouring stations. This implies that the observed rainfall values at the neighbouring stations used above are multiplied by the ratio of the normal annual rainfall at the station under consideration (Test station) and the normal annual rainfall at the adjoining stations (Base stations), i.e.:

$P_{corr, i, j} = (N_{test}/N_{base, i})\ P_{i, j}$

Where:

$P_{corr, i, j}$ = rainfall corrected for heterogeneity by the neighbour station i at time j

N_{test} = Annual Normal rainfall at the station under consideration

$N_{base, I}$ = Annual Normal rainfall at the adjoining stations (for i = 1 to M_{base})

Station Normal are calculated from the historical records. Otherwise, they may be computed from established relationships, as a function of altitude (Basin specific), if sufficient data is not available at all stations for estimating the station normal. The relationship for station normal as a function of the station altitude (*H*) is of the form:

$N_i = a_1 + b_1 \cdot H_s \forall \quad H_s \leq H_1$ $\quad$ a_1 = Rainfall datum station

$N_i = a_2 + b_2 \cdot H_s \forall \quad H_s \geq H_1$ $\quad$ H_1 = elevation datum station

b_1 =rate of change in rainfall

H_s = altitude of station under consideration

Example 5-8

Daily rainfall data series at Savlitank station is taken for illustrating the procedure of estimating the missing data at a station by making use of data available at neighbouring stations and employing distance power method of spatial interpolation.

For this, the search for neighbouring stations (base stations) is made within a radius of 25 km by using the option of "Spatial Interpolation", and six such stations are identified. Selection of the test and base stations is also shown in Figure 4.12. The nearest six stations are chosen which fall within the circle of 25 km radius. These stations are listed in Table 5.11 along with the quadrant, distances and corresponding normalised weights.

Table 5.11: Distances and normalised weights of stations adjoining Savlitank

Quadrant	Station	Distance (km)	Station weights ($\alpha 1/D^2$)	
			($1/D^2$)	Normalised weights
I	Vadol	9.225	0.011	0.274
II	Kapadwanj	8.139	0.015	0.353
III	Mahisa	13.48	0.005	0.128
III	Kathlal	13.895	0.005	0.120
IV	Vagharoli	17.872	0.003	0.073
IV	Thasara	21.168	0.002	0.052
	Sum		0.041	1.000

Results of the spatial interpolation are presented in Table 5.12 for August-September 1994 wherein the observed rainfall at all six base stations is listed followed by the estimated rainfall at Savlitank station. Since the daily rainfall at Savlitank station is actually not missing, a dummy data series at this station is first created and the spatially estimated rainfall values are stored in it. This is given as the estimated series at Savlitank station in the table. The available observed daily rainfall at Savlitank station is also given in the last column of the table for better appreciation of the usability of such an estimation procedure. A quick qualitative comparison (see Figure 5.8) of these estimated and observed daily rainfall values indicate that the two matches quite well. There will always be a few small and big deviations expected here and there for the simple reason that the averaging procedure is never expected to yield exactly what would have been the actual rainfall. It may also be noted however, that by employing such spatial interpolation, it is very likely that the number of rainy days at the station for which the estimation has been done increases to a significant extent. This is due to the fact that if there is some rainfall even at one station out of six the number of base stations then there is going to be some amount of rainfall estimated at the test station. If the data of all the base stations has been checked and corrected before making such interpolation then at least such increase in number of rainy days can be avoided on account of shifting of rainfall values at one or more stations. In any case, the statistic on number of rainy days must take into account long periods of estimated data using spatial interpolation.

Table 5.12: Observed daily rainfall at base stations

Date	Observed Rainfall at Neighbouring Stations (mm)						Rainfall at Savlitank (mm)	
	Vadol	Kapadwanj	Mahisa	Kathlal	Vagharoli	Thasara	Estimated/ Observed	
	0.274	0.352	0.128	0.121	0.073	0.052		
15/08/94	0.00	13.00	0.00	0.00	9.00	20.00	6.30	0.00
16/08/94	0.00	3.00	0.00	0.00	3.00	0.00	1.30	2.00
17/08/94	8.00	0.00	0.00	6.00	15.00	8.00	4.40	2.00
18/08/94	0.00	2.00	0.00	0.00	2.00	22.00	2.00	0.00
19/08/94	18.00	4.00	0.00	10.00	6.00	0.00	8.00	0.00
20/08/94	68.00	50.00	0.00	15.00	120.00	132.00	53.70	60.00
21/08/94	0.00	14.00	5.00	3.00	0.00	5.00	6.20	7.00
22/08/94	14.00	0.00	0.00	0.00	5.00	0.00	4.20	2.00
23/08/94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
24/08/94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25/08/94	0.00	0.00	2.00	0.00	0.00	0.00	0.30	0.00
26/08/94	0.00	0.00	0.00	5.00	0.00	0.00	0.60	0.00
27/08/94	9.00	4.00	6.00	5.00	5.00	7.00	6.00	0.00
28/08/94	40.00	43.00	0.00	0.00	43.00	43.00	31.50	39.00
29/08/94	0.00	14.00	0.00	0.00	0.00	0.00	4.90	0.00
30/08/94	0.00	0.00	0.00	7.00	0.00	0.00	0.80	0.00
31/08/94	0.00	0.00	0.00	0.00	0.00	40.00	2.10	0.00
01/09/94	50.00	74.00	30.00	10.00	30.00	15.00	47.80	24.00

Date	Observed Rainfall at Neighbouring Stations (mm)						Rainfall at Savlitank (mm)	
	Vadol	Kapadwanj	Mahisa	Kathlal	Vagharoli	Thasara	Estimated/ Observed	
	0.274	0.352	0.128	0.121	0.073	0.052		
02/09/94	27.00	60.00	25.00	8.00	25.00	45.00	36.90	18.00
03/09/94	0.00	48.00	0.00	5.00	18.00	41.00	20.90	21.00
04/09/94	0.00	0.00	6.00	0.00	0.00	0.00	0.80	4.00
05/09/94	0.00	4.00	3.00	0.00	10.00	0.00	2.50	2.00
06/09/94	0.00	0.00	0.00	7.00	0.00	0.00	0.80	0.00
07/09/94	220.00	336.00	315.00	100.00	305.00	312.00	269.50	278.00
08/09/94	61.00	60.00	65.00	50.00	45.00	42.00	57.70	122.00
09/09/94	0.00	19.00	8.00	0.00	12.00	0.00	8.60	8.00
10/09/94	15.00	15.00	5.00	10.00	0.00	7.00	11.60	6.00
11/09/94	0.00	0.00	0.00	0.00	0.00	4.00	0.20	0.00
12/09/94	8.00	0.00	0.00	0.00	0.00	0.00	2.20	0.00
10/12/94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11/12/94	15.00	0.00	0.00	0.00	0.00	115.00	10.10	0.00
12/12/94	0.00	0.00	80.00	18.00	0.00	40.00	14.50	5.00
13/12/94	40.00	44.00	16.00	33.00	45.00	112.00	41.60	40.00
14/12/94	0.00	13.00	0.00	10.00	12.00	0.00	6.70	32.00
15/12/94	0.00	0.00	0.00	12.00	0.00	0.00	1.50	0.00
16/12/94	0.00	0.00	0.00	15.00	0.00	0.00	1.80	0.00
17/12/94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

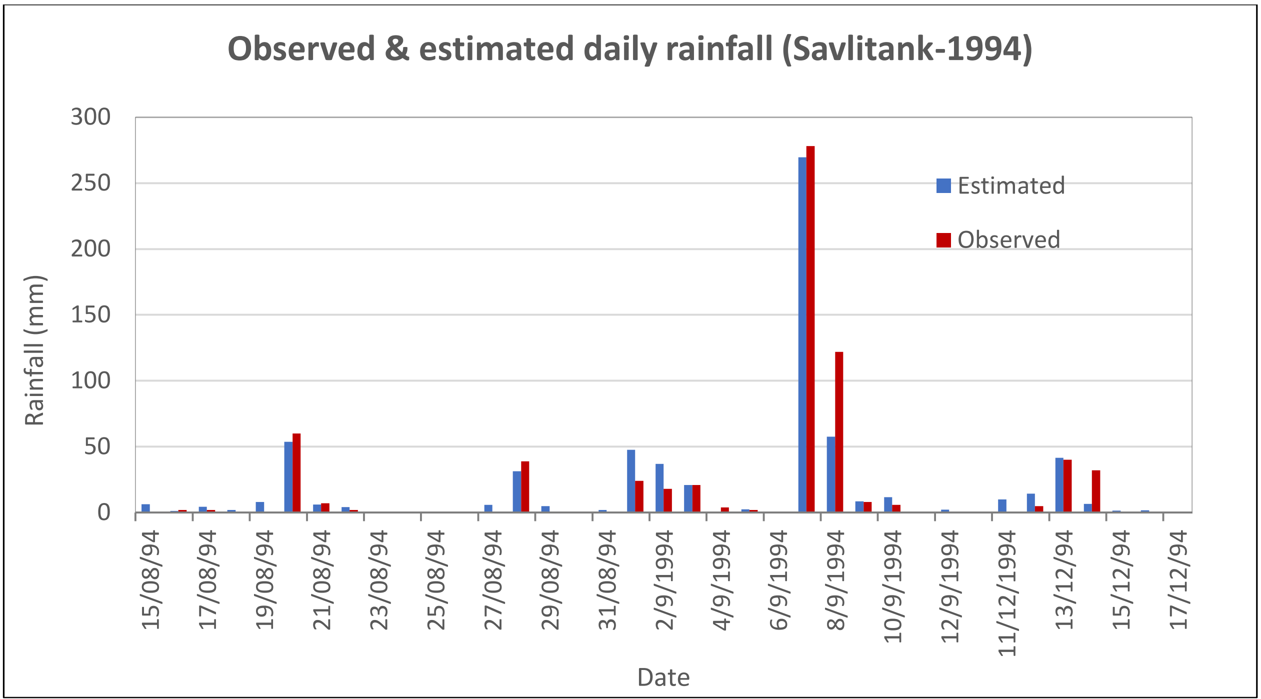


Figure 5.8: Comparison of observed and estimated rainfall at station Savlitank

After scrutiny and checking rainfall series the incorrect and missing values will be replaced where possible by estimated values based on other observations at the same station or at neighbouring stations. The process of filling in missing values is generally referred to as 'completion'.

Where no suitable neighbouring observations or stations are available, missing values will be left as 'missing' and incorrect values will be set to 'missing'. Procedures for correction and completion depend on the type of error and the availability of suitable source records with which to estimate, what should have been studied using the tools described in the previous section. The judgment of the hydrologist is critical at this stage. The newly calculated value will then be marked and it will not be missing any more. A label is attached to the new data value, implying that this data value has been completed.

5.7 Other methods of gap filing and data correction

There are few other methods of gap filling and correction of time series rainfall data such as

1. Relation curves
2. Drift correction

5.7.1 Relation curves

By Time Series Analysis/General Inspection of Series the relationship between stations are studied through regression curves. The regression curves analysed are stored if the user found them to be adequate. The best regression curve for each interval should be used for filling data gaps. The list of correlation equations that may be selected are:

- Linear. Only fill series with a few gaps: For rainfall, this method is recommended only for hourly data
- Polynomial
- Power
- Exponential

5.7.2 Drift correction

The pen of the autographic recorder may gradually drift from its true position. In this case, analogue observations may show deviation from the staff gauge observations. This deviation can be static or may increase gradually with time.

Where a digital record is produced from an analogue record using a pen-following digitizer, the annotated clock and recorder time and level can be fed into the digitizing program and an accumulative adjustment spread over the level recorded from the time the error is thought to have commenced till the error was detected or the chart removed. However, such procedure is not recommended as the actual reasons for the shift may still be unknown at the time of digitizing the charts. It is always appropriate to tabulate/digitize the chart record as it is in the first instance and then apply corrections thereafter.

This option for correcting the gradual spread of error in digital records is extracted from a chart recorder, with a growing adjustment from the commencement of the error until the end of the error detection. For example, let the error be ΔX observed at time t = i+k, where i is the time when the drift started, which the user has to determine using judgement. In that case, the correction that can be implemented to remove the drift in the data between times i and k could be based on the following formula:

$$X_{corr,j} = X_{means,j} - \frac{(j-i)}{k}\Delta X \quad \text{for } j=i, i+1, \ldots, k \qquad \text{Eqn. 5.14}$$

The above approach should be used with caution, since it may not be easy to determine the exact starting point where the adjustments for the data drift should begin.

6 COMPILATION OF RAINFALL DATA

6.1 General

Compilation is a process by which data at its observational/recorded time interval and units are transformed to another time interval or unit to facilitate analysis, validation or reporting. Under rainfall compilation, the observed rainfall is transformed:

- from one-time interval to another
- from one unit of measurement to another
- from point to areal values
- from non-equidistant to equidistant series

Compilation is carried out at the State Data Processing Centre. It may be carried out prior to data 'validation' if so required, but the final compilation is normally carried out after data 'correction and completion'.

6.2 Aggregation of data to longer durations

Rainfall from different sources is observed at different time intervals, but these are generally one day or shorter. For the standard rain gauge, rainfall is measured once or twice daily. For autographic records, a continuous trace is produced from which hourly rainfall is extracted. For digital rainfall recorders, rainfall is recorded at variable intervals with each tip of the tipping bucket. Hourly data are typically aggregated to daily; daily data are typically aggregated to weekly, ten daily, 15 daily, monthly, seasonal or annual.

Aggregation to longer time intervals is required for validation and analysis, as well as input into modelling. For validation, small persistent errors may not be detected at small time intervals, but may be detected at longer time interval.

6.2.1 Aggregation of daily to weekly data

Aggregation of daily to weekly time interval is usually done by considering the first 51 weeks of equal length (i.e. 7 days) and the last (52nd) week of either 8 or 9 days according to whether the year is a non-leap year or a leap year, respectively. The rainfall for such weekly time periods is obtained by simple summation of consecutive sets of seven-day rainfalls. The last week's rainfall is obtained by summing up the last 8 or 9 days of rainfall.

For some applications, it may be required to get the weekly compilation done for the exact calendar weeks (from Monday to Sunday). In such cases, the first week in any year will start from the first Monday in that year and thus there will be 51 or 52 full weeks in the year and one or more days left in the beginning and/or end of the year. The days left out at the end of a year or beginning of the next year could be considered for the 52nd week of the year under consideration. There will also be cases of a 53rd week when the

1st day of the year is also the first day of the week (for non-leap years) and 1st or 2nd day of the year is also first day of the week (for leap years).

6.2.2 Aggregation of daily to ten-day periods

Aggregation of daily to ten daily time intervals is usually done by considering each month of three ten daily periods. Hence, every month will have first two ten daily periods of ten days each and last ten daily period of either 8, 9, 10 or 11 days according to the month and the year. Rainfall data for such ten daily periods is obtained by summing the corresponding daily rainfall data. Rainfall data for 15 daily periods is also obtained in a similar manner for each of the two parts of every month.

6.2.3 Aggregation from daily to monthly

Monthly data are obtained from daily data by summing the daily rainfall data for the calendar months. Thus, the number of daily data to be summed up will be 28, 29, 30 or 31 according to the month and year under consideration. Similarly, yearly rainfall data are obtained by either summing the corresponding daily data or monthly data, if available.

6.2.4 Hourly to other intervals

It may sometimes be desired to obtain rainfall data for every 2 hours, 3 hours, 6 hours, 12 hours etc. for some specific requirement. Such compilations are carried out by simply adding up the rainfall data corresponding to the available shorter time intervals.

Example 6-1

Daily rainfall at Anior station (Kheda catchment) is observed with a standard rain gauge (SRG). An autographic rain gauge is also available at the same station for recording rainfall continuously, and hourly rainfall data is obtained by tabulating information from the chart records.

It is required that the hourly data be compiled to daily interval corresponding to the cut off time for the start/end of the day at 08:30 hrs. This compilation is done using the aggregation option and by converting from hourly to daily interval. The observed hourly data and compiled daily data are shown in Figure 6.1 and Figure 6.2 respectively.

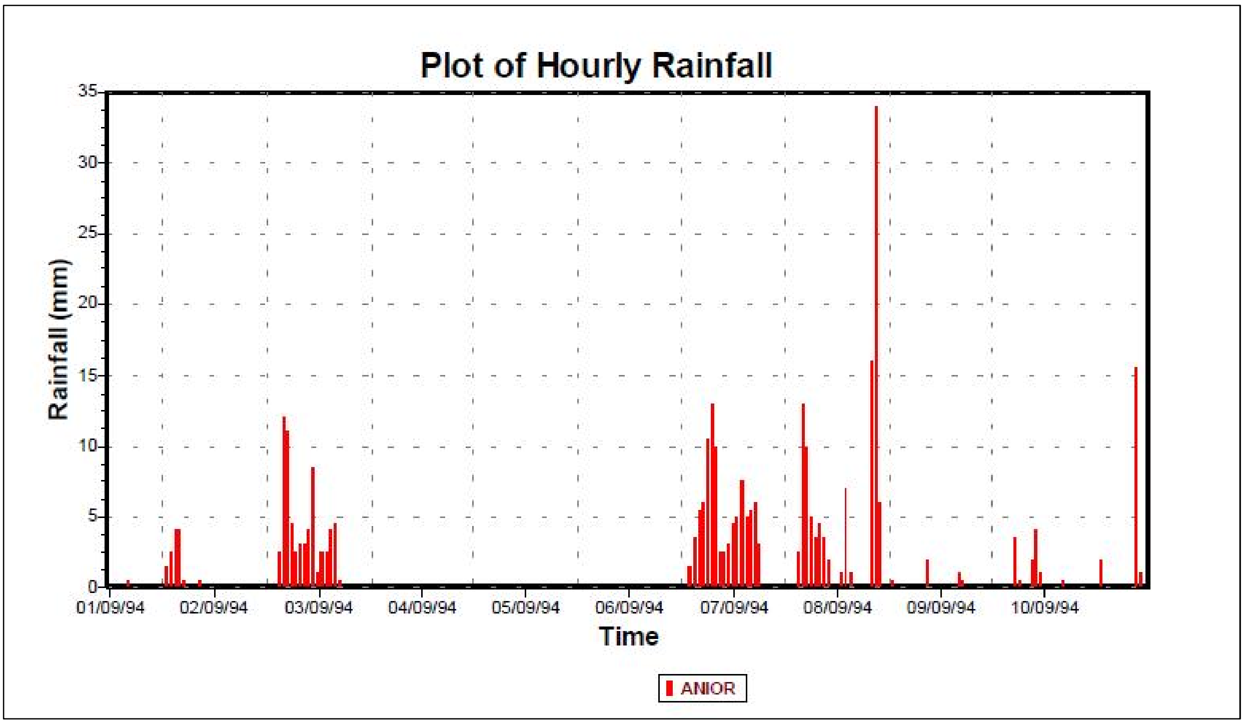


Figure 6.1: Plot of observed hourly rainfall data

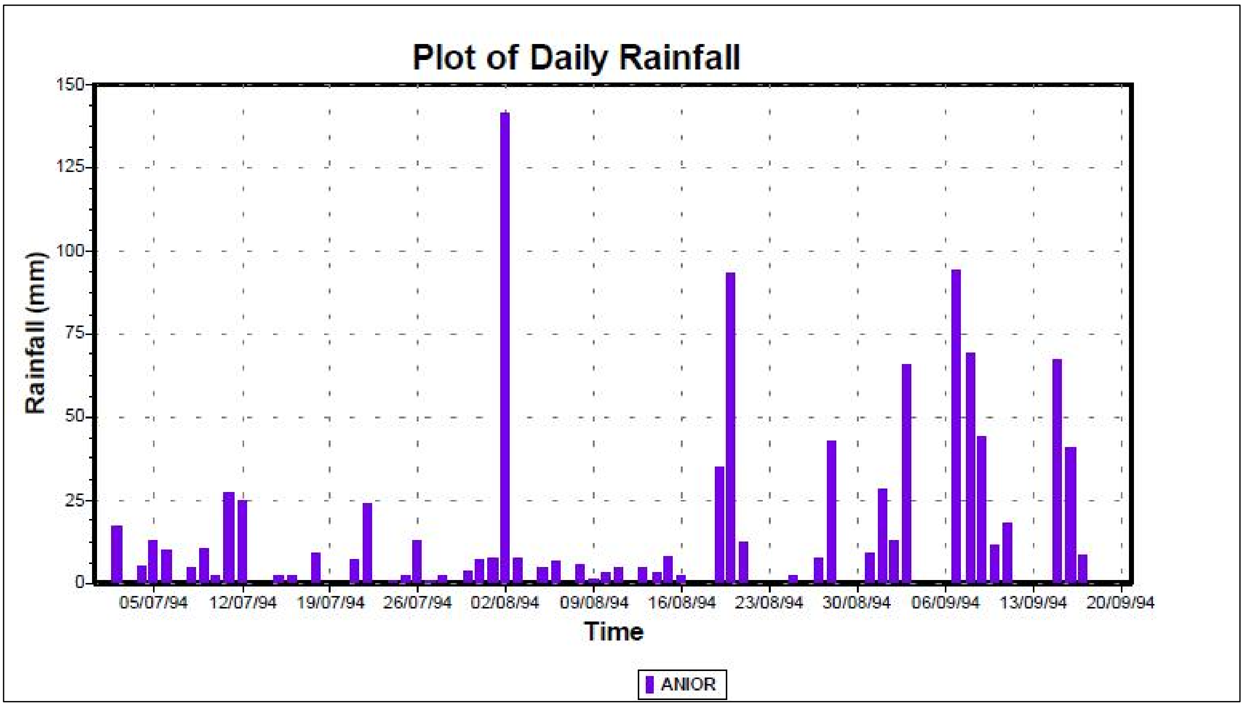


Figure 6.2: Compiled daily rainfall from hourly data tabulated from ARG charts

Similarly, daily data observed using SRG may be required at weekly, ten-daily, monthly and/or yearly intervals for various applications and for the purpose of data validation.

For this compilation, the daily data obtained using SRG is taken as the basic data and compilation is done to weekly, ten-daily, monthly and yearly intervals. These are illustrated in Figure 6.3 to Figure 6.6 respectively.

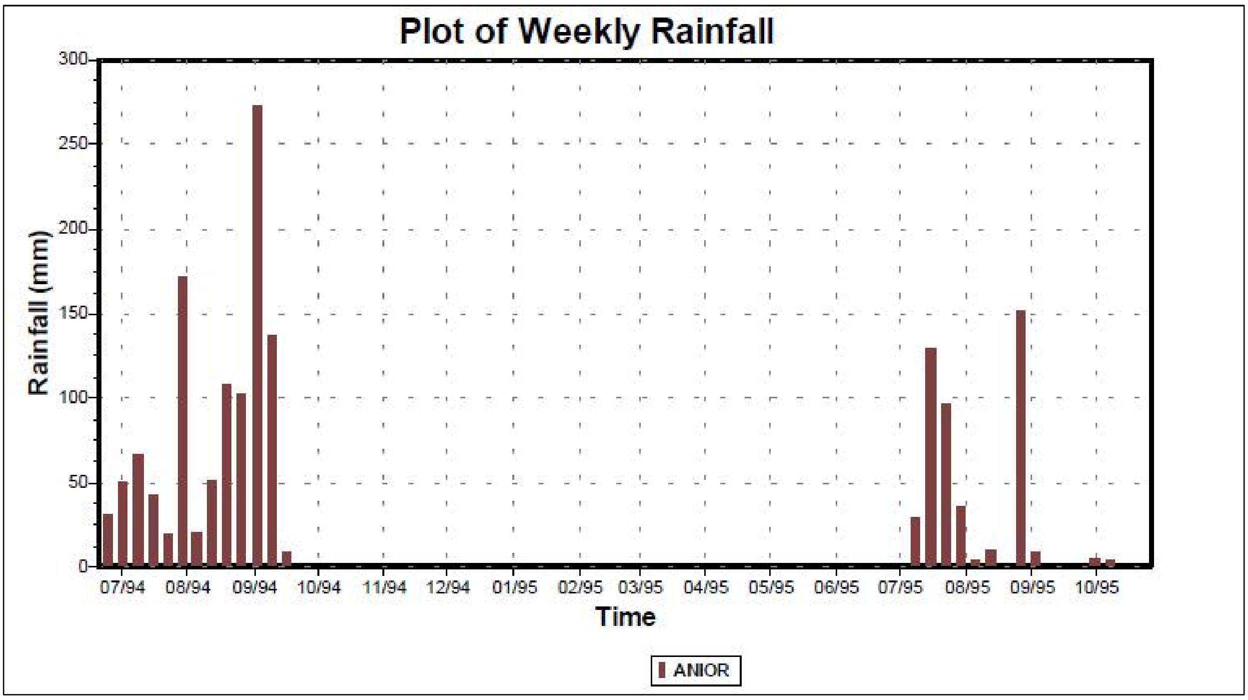


Figure 6.3: Compiled weekly rainfall from hourly data tabulated from ARG charts

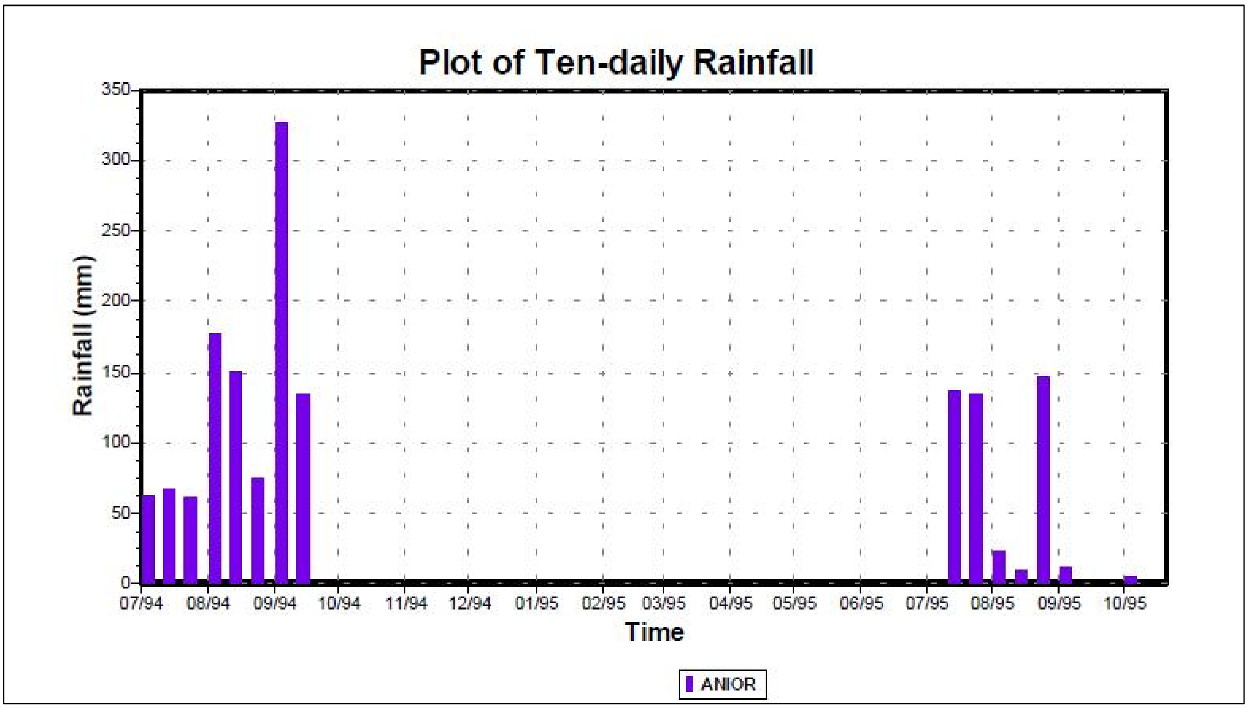


Figure 6.4: Compiled ten-daily data from daily data obtained from SRG records

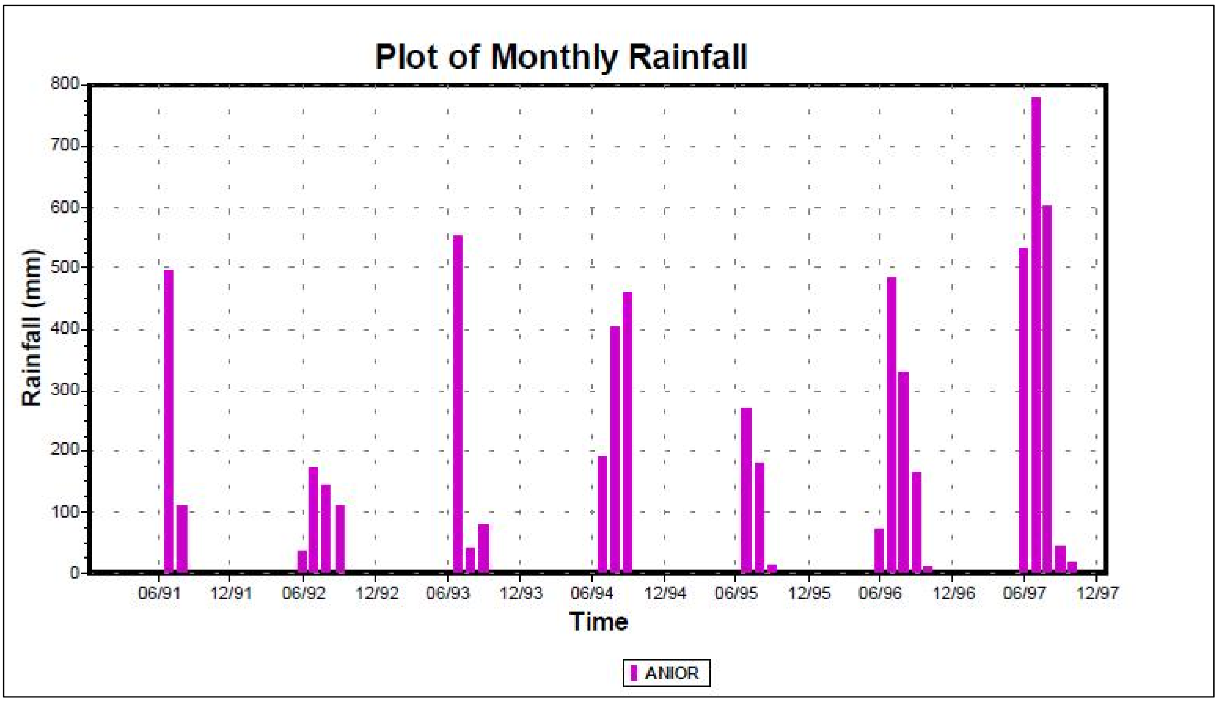


Figure 6.5: Compiled monthly data from daily data obtained from SRG records

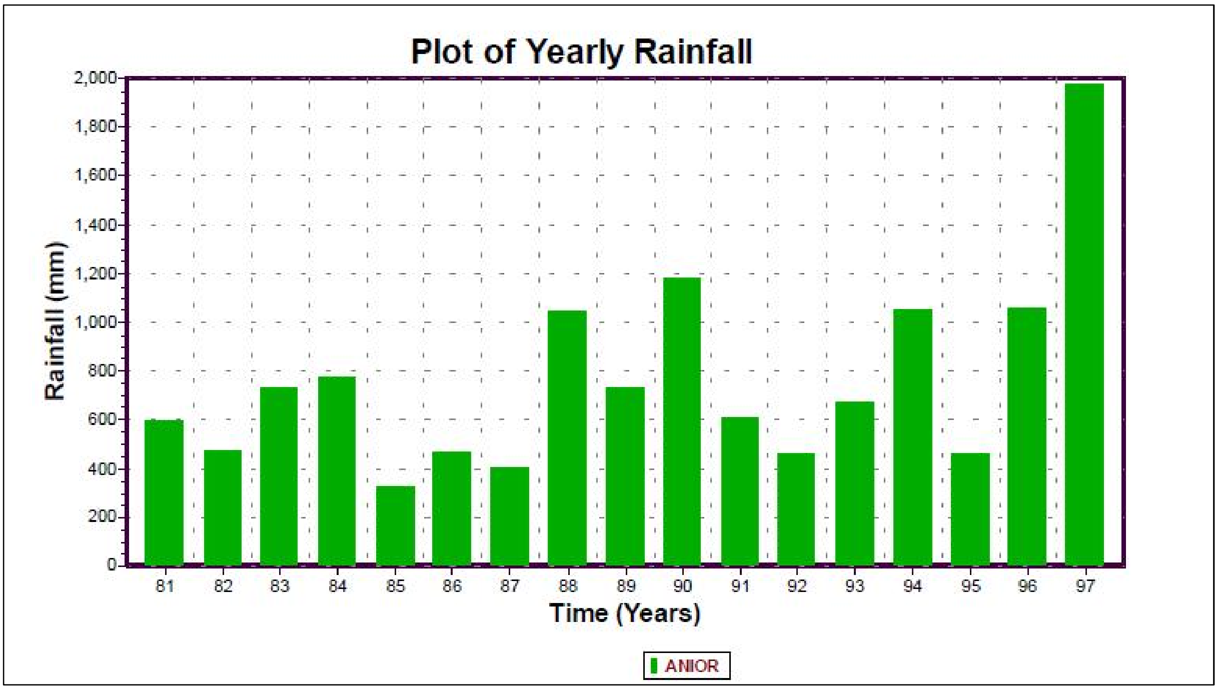


Figure 6.6: Compiled yearly data from daily data obtained from SRG records

6.3 Estimation of areal rainfall

6.3.1 General description

Rain gauges generally measure rainfall at individual points. A single point precipitation measurement is not a reliable representative of the volume of precipitation falling over a given catchment area. However, many hydrological applications require the average depth of rainfall occurring over an area which can then be compared directly with runoff from that area. The area under consideration can be a principal river basin or a component sub-basin. Occasionally, average areal rainfall is required for the entire basin, state or other administrative unit, and the areal average is obtained within the appropriate political or administrative boundary.

Since rainfall is spatially variable and the spatial distribution varies between events, point rainfall does not provide a precise estimate or representation of the areal rainfall. The areal rainfall will always be an estimate and not the true rainfall depth irrespective of the method.

There are number of methods which can be employed for estimation of the areal rainfall including:

- Arithmetic Average method
- Weighted Average method
- Thiessen Polygon method
- Spline method
- Kriging method

All these methods for estimation of areal average rainfall compute the weighted average of the point rainfall values; the difference between various methods is only in assigning the weights to these individual point rainfall values, the weights being primarily based on the proportional area represented by a point gauge. These methods are outlined below:

6.3.2 Arithmetic average

This is the simplest of all the methods and as the name suggests the areal average rainfall depth is estimated by simple averaging of all selected point rainfall values for the area under consideration.

If the rain gauges are uniformly distributed over the area and the rainfall varies in a regular manner, the results obtained by this method will be quite satisfactory and will not differ much than those obtained by other methods. This method gives equal weight to every station regardless of its location and can be used for the storm rainfall, monthly or annual rainfall average computations. This is given by:

$$P_{at} = \frac{1}{N}(P_{1t} + P_{2t} + P_{3t} + \ldots + P_{Nt}) = \frac{1}{N}\sum_{i=1}^{N} P_{it} \qquad \text{Eqn. 6.1}$$

Where:

P_{at} = estimated average areal rainfall depth at time t

P_{it} = individual point rainfall values considered for an area, at station i for i = 1,N) and time t,

N = total number of point rainfall stations considered

In this case, all point rainfall stations are allocated weights equal to the reciprocal of the total number of stations considered. Generally, stations located within the area under consideration are taken into account. However, it is good practice to also include such stations which are outside but close to the areal boundary and thus represent some part of the areal rainfall within the boundary. This method is also sometimes called as unweighted average method since all the stations are given the same weights irrespective of their locations.

This method gives satisfactory estimates and is recommended where the area under consideration is flat, the spatial distribution of rainfall is fairly uniform, and the variation of individual gauge records from the mean is not significant.

6.3.3 Weighted average using user defined weights

In the arithmetic averaging method, all rainfall stations are assigned equal weights. To account for orographic effects and especially where rain gauges are predominantly located in the lower rainfall valleys, it is sometimes required to weight the stations differently. In this case, instead of equal weights, user defined weights can be assigned to the stations under consideration. The estimation of areal average rainfall depth can be made as follows:

$$P_{wt} = \frac{1}{N}(C_1 P_{it} + C_2 P_{2t} + C_3 P_{3t} + \ldots + C_N P_{Nt}) = \frac{1}{N}\sum_{i=1}^{N} C_i P_{it} \qquad \text{Eqn. 6.2}$$

Where:

C_i=weight assigned to individual rain gauge station i (i=1, N)

To account for under-representation by gauges located in valleys, the weights do not necessarily need to add up to 1, although their sum should be close to 1.

6.3.4 Thiessen polygon method

This widely-used method was proposed by A.M. Thiessen in 1911. The Thiessen polygon method accounts for the variability in spatial distribution of gauges and the consequent variable area which each gauge represents. The areas representing each gauge are defined by drawing lines between adjacent stations on a map. The perpendicular bisectors of these lines form a pattern of polygons (the Thiessen polygons) with one station in each polygon (see Figure 6.7). Stations outside the basin boundary should be included in the analysis as they may have polygons which extend into the basin area. The ratio of the basin area of a polygon associated with an individual station to the total basin area represents the Thiessen weight for that station. Areal rainfall is thus estimated by first multiplying individual station totals by their Thiessen weights and then summing the weighted totals as follows:

$$P_{at} = \frac{A_1}{A} P_{1t} + \frac{A_2}{A} P_{2t} + \frac{A_3}{A} P_{3t} + \ldots + \frac{A_N}{A} P_{Nt} = \sum_{i=1}^{N} \left(\frac{A_i}{A}\right) P_{it} \qquad \text{Eqn. 6.3}$$

Where:

A_i= the area of Thiessen polygon for station i

A = total area under consideration

The Thiessen method is objective and readily computerized but is not ideal for mountainous areas where orographic effects are significant or where rain gauges are predominantly located at lower elevations of the basin. Altitude weighted polygons (including altitude as well as areal effects) exist, but they are not widely used.

Example 6-2

Estimate areal average rainfall for a catchment for the rainfall event of August 30,1982 on the basis of daily rainfall data observed at a number of rain gauges in and around the basin. Areal average is worked out using two methods: (a) Arithmetic average and (b) Thiessen method.

Data

Station		30 August 1982 storm Pi (mm)
1	Paikmal	338.0
2	Padampur	177.0
3	Bijepur	521.0
4	Sohela	262.0
5	Binka	158.0
6	Bolangir	401.6

a) Arithmetic Average

For the arithmetic average method rainfall stations located inside and very nearby to the catchment boundary are considered and equal weights are assigned to all of them. Since there are 6 stations considered the individual station weights work out as 0.167 and is given in Table 6.1 below. On the basis of these equal station weights daily areal average is computed.

Table 6.1: List of stations and corresponding weights as per arithmetic average method

Sl. No.	Station	30 August 1982 storm		Pi. Wi
		Pi(mm)	Weight Wi	
1	Paikmal	338.0	0.167	56.446
2	Padampur	177.0	0.167	29.559
3	Bijepur	521.0	0.167	87.007
4	Sohela	262.0	0.167	43.754
5	Binka	158.0	0.167	26.386
6	Bolangir	401.6	0.167	67.067
			SUM(Σ)=	310.219

Average precipitation for the Aug 30th 1982 storm is found as 310.22 mm.

b) Thiessen polygon method

Computation of areal average using Thiessen method is accomplished by first getting the Thiessen polygon layer (defining the boundary of Thiessen polygon for each contributing point rainfall station). The station weights are automatically worked out on the basis of the ratios of the areas of these polygons with respect to the total area of the catchment. The layout of the Thiessen polygons as worked out by the system is graphically shown in Figure 6.8 and the corresponding station weights are as given in Table 6.2. On the basis of these Thiessen polygon weights, the areal average of the basin is computed and this is shown in Table 6.2 for the year 1982. In this case it may be noticed that there is no significant change in the values of the areal rainfall (310 mm versus 30.94 cm) obtained by the two methods primarily on account of small variation in rainfall from station to station.

Table 6.2: Average precipitation by Thiessen-polygon method

No	Station	30 August 1982 storm	Area of influence		Weightage of each station	Pi.Wi
		(mm)	Pi (cm)	km^2	Wi	
1	Paikmal	338.00	33.80	572.12	0.10	3.51
2	Padampur	177.00	17.70	1374.04	0.25	4.42
3	Bijepur	521.00	52.10	1148.24	0.21	10.87
4	Sohela	262.00	26.20	517.63	0.09	2.46
5	Binka	158.00	15.80	934.56	0.17	2.68
6	Bolangir	401.60	40.16	958.40	0.17	6.99
	Sum			5504.99	1.000	30.935

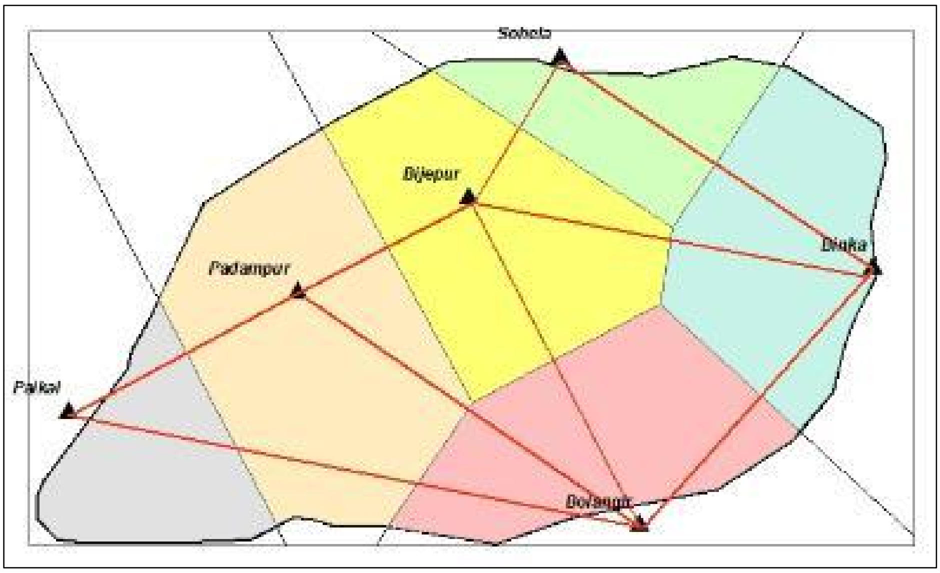


Figure 6.7: Example of Thiessen polygon prepared using Arc GIS software

6.3.4.1 Isohyetal method

The main difficulty with the Thiessen method is its inability to deal with orographic effects on rainfall. A method which can incorporate such effects is the isohyetal method, where lines of equal rainfall (isohyets) are being drawn by interpolation between point rainfall stations taking into account orographic effects.

In flat areas where no orographic effects are present, the method simply interpolates linearly between the point rainfall stations. Manually, the procedure is as follows: On a basin map, the locations of the rainfall stations within the basin and outside near the basin boundary are plotted. Next, the stations are connected with their neighbouring stations by straight lines. The positions of the isohyet(s) on these connecting lines are indicated depending on the rain depths for which isohyets are shown by linear interpolation between two neighbouring stations. After having completed this for all connected stations, smooth curves are drawn through the points marked on the straight lines between the stations connecting the concurrent rainfall values for which isohyets are to be shown, as shown in Figure 6.8. Drawing the isohyets relies on the personal experience with local conditions and information on storm orientation. Subsequently, the area between two adjacent isohyets and the catchment boundary is estimated using GIS. The average rainfall obtained from the two adjacent isohyets is assumed to have occurred over the entire inter-isohyet area. Hence, if the isohyets are indicated by P_1, P_2, …,, P_n with inter-isohyet areas a_1, a_2, …, a_{n-1} the mean precipitation over the catchment is computed from:

$$\overline{P} = \frac{a_1(\frac{P_1+P_2}{2})+\ldots+a_{n-1}(\frac{P_{n-1}+P_n}{2})}{A} \qquad \text{Eqn. 6.4}$$

It is noted that if the maximum and /or minimum point rainfall value(s) are within the catchment boundaries then P_1 and/or P_n is to be replaced by the highest and/or lowest point rainfall values. A slightly biased result will be obtained if the lowest (highest) isohyet is located outside the catchment area as the averaging over two successive isohyets will underestimate (overestimate) the average rainfall in the area bounded by the catchment boundary and the first inside isohyet.

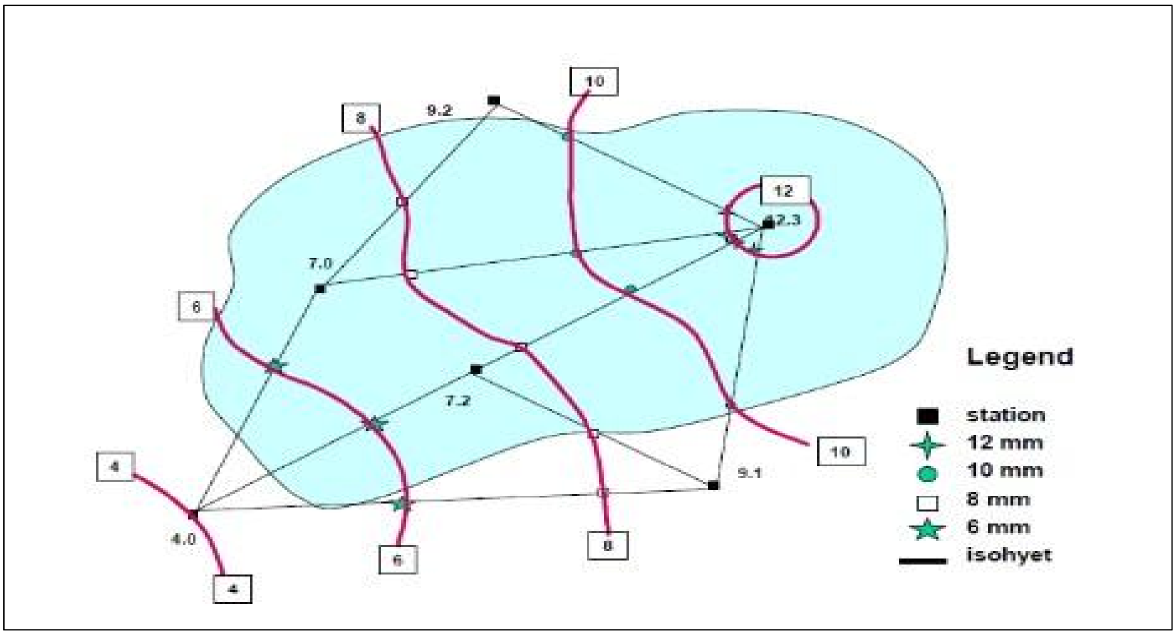


Figure 6.8: Example of isohyets prepared using linear interpolation

For flat areas the Isohyetal method is superior to the Thiessen method if individual storms are considered, as it allows for incorporation of storm features like the storm orientation. This feature is not available for monthly, seasonal or annual values. Its added value is of particular benefit when special meteorological features like orographic effects are present in the catchment rainfall. In such cases the above procedure is executed with a catchment map overlaying a topographical map to be able to draw the isohyets parallel to the contour lines. Also, the extent of rain shadow area at the leeward side of mountain chains can easily be identified from topographical maps. The computations are again carried out using equation 6.1. In such situations, the Isohyetal method can be is likely to be superior to the Thiessen method.

6.3.4.2 Iso-percental method

This method is recommended if long term seasonal topographical patterns are to be incorporated in the estimates of areal precipitation, which is achieved by drawing isohyets for individual storms or seasons. The assumption is that the long term seasonal topographical effect as displayed in the seasonal (or annual) isohyets are also applicable for individual storms and seasons. The procedure involves the following steps, and it is also demonstrated in Example 6-3.

1. Compute point rainfall as a percentage of seasonal normal rainfall for all stations
2. Draw isopercentals (lines of equal actual point rainfall to station normal rainfall) on a transparent overlay
3. Superimpose the overlay on the seasonal isohyetal map
4. Mark each crossing of seasonal isohyets with isopercentals
5. Multiply for each crossing the isohyet with the isopercental value and add the value to the crossing on the map with the observed rainfall values; hence, the data set is extended with the rainfall estimated derived in step 4
6. Draw isohyets using linear interpolation while making use of all data points, i.e. observed and estimated data (see step 5)

6.3.4.3 Hypsometric method

Special attention is to be paid to situations where at the higher elevations rain gauge stations do not exist. Then the orographic effects have to be extrapolated from the lower reaches of the mountains by estimating a relationship between rainfall and elevation for the available range of values and extrapolating the same for higher elevations. Using this rainfall-elevation curve a number of points in the ungauged upper reaches are added to the point rainfall data to guide the interpolation process.

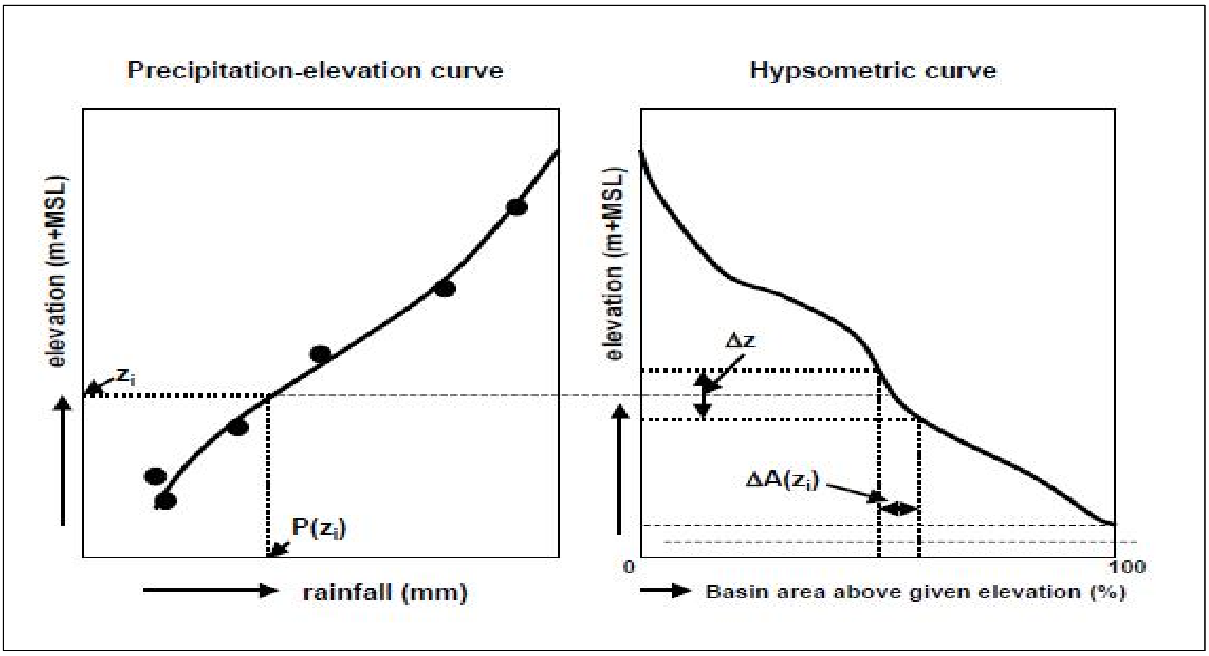


Figure 6.9: Principle of hypsometric method

A simple technique to deal with such situations is the **hypsometric method**, where a precipitation-elevation curve is combined with an area-elevation curve (called hypsometric curve) to determine the areal rainfall. The latter method avoids recurrent assessment of inter-isohyet areas, whereas the results will be similar to the isohyetal method. The precipitation-elevation curve has to be prepared for each storm, month, season or year, but its development will be guided by the rainfall-elevation curve, which is also represented using the orographic equation, often approximated by a simple linear relation of the form:

$$P(z) = a + bz \qquad \text{Eqn. 6.5}$$

This relation may vary systematically in a region (e.g. the windward side of a mountain range may have a more rapid increase in precipitation with elevation than the leeward side). In such cases separate hypsometric curves and orographic equations are established for the distinguished sub-regions. The areal rainfall is estimated by:

$$\overline{P} = \sum_{i=1}^{n} P(z_i)\, \Delta A(z_i)/A \qquad \text{Eqn. 6.6}$$

Where:

$\overline{P}$=areal rainfall

$P(z_i)$= rainfall read from precipitation-elevation curve at elevation z_i

$\Delta A(z_i)$ = percentage of basin area contained within elevation $z_i \pm 1/2\Delta z_i$

n= number of elevation interval in the hypsometric curve has been divided.

Example 6-3

The application of the iso-percental method is demonstrated in this example (NIH, 1988). The areal rainfall for the storm on the 30th of August 1982 has to be determined for the catchment shown in Figure 6.10. The total catchment area amounts 5,600 km^2. The observed and average annual rainfall amounts for the point rainfall stations in the area are given in Table 6.3.

Table 6.3: Storm rainfall and annual normal

Station	30 August 1982 storm	Normal annual rainfall	Storm rainfall as percentage of annual normal
	(mm)	(mm)	(%)
Primal	338.00	1728.00	19.60
Padampur	177.00	1302.00	13.60
Bijepur	521.00	1237.00	42.10
Sohela	262.00	1247.00	21.00
Binka	158.00	1493.00	10.60
Bolangir	401.60	1440.00	27.90

For each station the point rainfall as percentage of seasonal normal is displayed in the last column of Table 6.3. Based on this information isopercentals are drawn on a transparent overlay, which is subsequently superimposed on the annual normal isohyetal map. The intersections of the isopercentals and isohyets are identified and for each intersection the isopercental are multiplied with the isohyets to get an estimate of the storm rainfall for those points. These estimates are then added to the point rainfall observations to draw the isohyets, as seen in Figure 6.11. The inter-isohyet area is then determined and the areal rainfall is subsequently computed with the aid of equation 6.3 as shown in Table 6.4.

Table 6.4: Computation of areal rainfall by isohyetal/isopercental method

Isohyetal range	Mean rainfall	Area	Volume
(mm)	(mm)	(km²)	(km² × mm)
158-200	179.00	206.48	36959.22
200-300	250.00	1309.66	327414.79
300-400	350.00	1726.25	604188.27
400-500	450.00	2102.73	946229.40
500-521	510.50	159.87	81615.47
Total		5504.99	1996407.15

Volume/Area=362.65 mm = 36.2cm

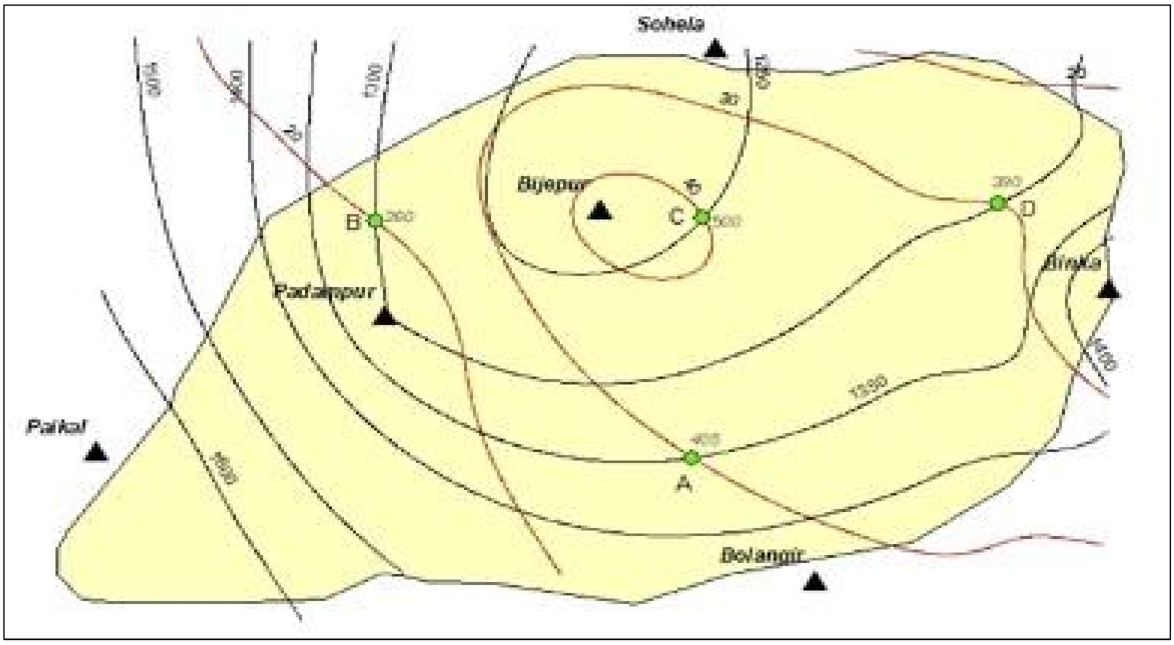


Figure 6.10: Isopercental Map

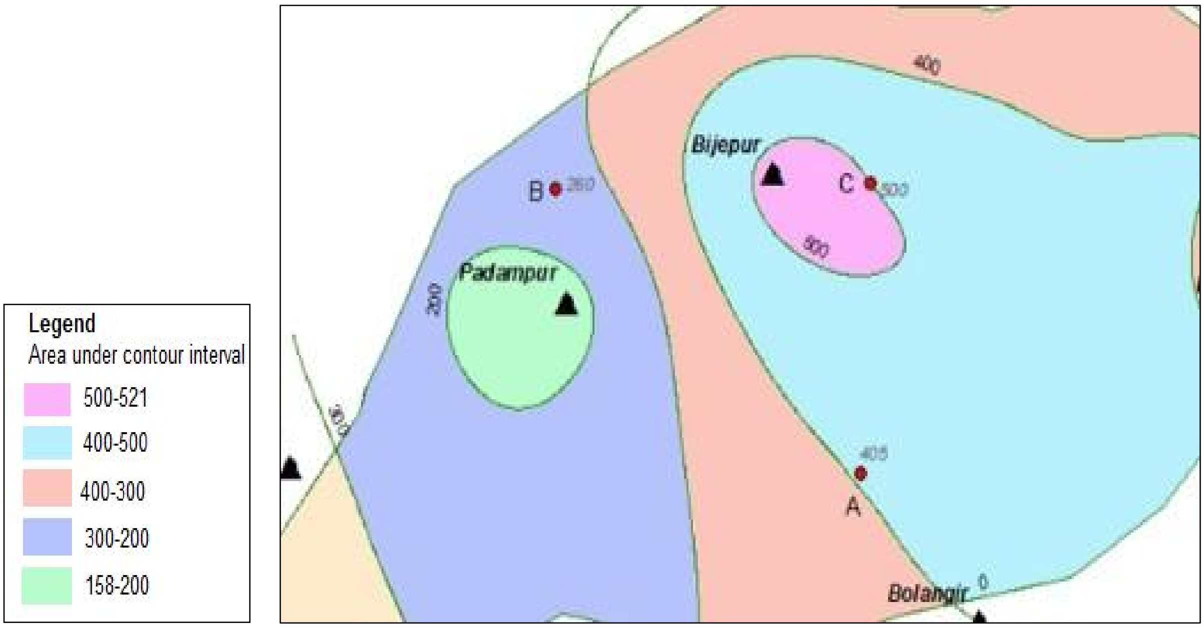


Figure 6.11: Isohyetal map drawn by isopercental method

6.3.5 Inverse distance weighted (IDW) method

Inverse distance weighted method is the simples and most commonly used method for interpolation of point rainfall data. It is a local method that assumes that the unknown value of a point is influenced more by nearby control points than those farther away. The degree of influence, or the weight, is expressed by the inverse of the distance between points raised to a power. In this method, the interpolated value is determined by:

$$P_0 = \frac{\sum_{j=1}^{N} \frac{P_j}{d_j{}^k}}{\sum_{j=1}^{N} \frac{1}{d_j{}^k}} \qquad \text{Eqn. 6.7}$$

It is observed that the weights are proportional to the distance between "0" and station j to some power p. The power factor k = 2 is commonly applied in rainfall estimation.

Compared with other methods, most notably the Kriging method, the Inverse Distance Weighted method is simpler and does not require pre-modelling or subjective assumptions in selecting the appropriate semi variogram model. The method runs faster, being of value in an emergency situation that requires rapid yet justifiable results. One major drawback of the inverse distance interpolation approach is that when two or more sampling points are close to each other (in the absence of measurement errors), the redundant information from these two stations is not ignored. Inverse distance interpolations commonly have a "duck-egg" pattern around solitary data points with values that differ greatly from the values at their surroundings.

The surface generated through inverse distance weighted interpolation is sensitive to the outliers, as it is an exact interpolator. The Inverse Distance Weighted Method does not take account of station clusters, which is convincingly shown in Table 5.5, last row; the estimate for "0" is seen to be almost entirely determined by the cluster (1, 2) which is nearest to "0". Hence, this method is to be applied only when the stations are more or less evenly distributed and clusters do not exist.

6.3.6 Spline

A spline is approximately a piecewise cubic polynomial that is continuous and has continuous first and second derivatives. Thin Plate Smoothing Splines are commonly applied for smooth multivariate interpolation of irregularly scattered noisy data. The geostatistical analysis tool in ArcGIS software uses a set of n basis functions, one for each data location. The predictor is a linear combination of the basis functions,

$$\hat{Z}(S_0) = \sum_{i=1}^{n} \omega_i \varphi([s_i - s_0]) + \omega_{n+1} \qquad \text{Eqn. 6.8}$$

where φ(r) is a radial basis function, $r = |s_i - s_0|$ is Euclidean distance between the prediction location s_0 and each data location s_i, and $\{\omega_i : i = 1,2, \dots , n + 1\}$ are weights to be estimated.

The radial basis functions commonly available with the software are: (a) Completely regularised spline function, (b) Spline with tension function, (c) Multiquadric function, (d) Inverse Multiquadric function, and (e) the Thin-plate spline function. The optimal smoothing parameter can be found by minimising the root mean square prediction error using cross validation. Because Splines are piecewise functions using a few points at a time, the interpolating values can be quickly calculated. In contrast to trend surfaces and weighted averages, Splines are able to retain small scale features. Unless the user has a strong background understanding of the rainfall process for the area and has sufficient time to develop good models for interpolation using Kriging, it is suggested that spline may be used, as a method more advanced than inverse distance weighted interpolation. This is a commonly available function in many GIS application packages.

6.3.7 Kriging method

6.3.7.1 General

The Kriging Method is an advanced interpolation method that takes care of the variation of rainfall with distance. It is a Geostatistical method for spatial interpolation named after the South African mining engineer D.G. Krige. It assumes that the spatial variation of an attribute is neither totally random, nor deterministic. The value of random variable Z at x, Z(x) can be expressed as the sum of three major components expressed as:

$$Z(x) = m(x) + \epsilon'(x) + \epsilon'' \qquad \text{Eqn. 6.9}$$

Where:

m(x) is a structural component, having a constant mean or trend $m(x)$

A random, but spatially correlated component, known as the variation of the regionalized variable $\epsilon'(x)$

A spatially independent Gaussian random noise or error term ϵ'' having zero mean and variance γ^2

This method provides rainfall estimates (or estimates of any other variable) at points (point-Kriging) or blocks (block-Kriging) based on a weighted average of observations made at surrounding stations. A dense grid is superimposed over the catchment as part of the application of the Kriging method for areal rainfall estimation. After having estimated the rainfall for the grid points, the areal rainfall is simply determined as the average rainfall of all grid points within the catchment.

It can be shown that the use of Kriging method of interpolation may lead to results with the smallest errors, particularly when the data are scant. However, it has also been widely reported in the literature that use of Kriging for interpolation without proper expertise and without devoting large amount of time to understand the spatial pattern of rainfall may lead to results that are grossly in error, even worse than those obtained with simpler methods like inverse distance weighted method. The sample data is often inadequate to realistically describe the spatial behaviour, and therefore the choice of variogram should be made with deep understanding of the spatial pattern based on physical observations and reasoning. Even though many GIS software platforms have incorporated tools to apply Kriging for interpolation, it is strongly advised to avoid its use unless a thorough understanding of the technique and the rainfall pattern is developed. It remains outside the scope of the present manual to deal with Kriging in such depth and details. Readers are referred to the standard texts on geo statistics for further reference (*Isaaks and Srivastava, 1989*). At each grid point the rainfall is estimated from:

$$\mathrm{Pe}_0 = \sum_{k=1}^{N} w_{0,k} . P_k \qquad \text{Eqn. 6.10}$$

Where:

Pe_0= rainfall estimate at some grid point "0"
$W_{0,k}$= weight of station k in the estimate of the rainfall at point "0"
P_k = rainfall observed at station k
N = number of stations considered in the estimation of Pe_0

The weights are different for each grid point and observation station. The weight given to a particular observation station k in estimating the rainfall at grid point "0" depends on the grid point-station distance and the spatial correlation structure of the rainfall field. The Kriging method uses weights which have the following properties:

- the weights are linear, i.e. the estimates are weighted linear combinations of the available observations
- the weights lead to unbiased estimates of the rainfall at the grid points, i.e. the expected estimation error at all grid points is zero
- The weights minimise the error variance at all grid points.

The procedure for estimation of weights can be found in any standard texts on Geostatistics, e.g. Clarke (1979), Isaaks and Srivastava (1989), Lloyd (2010) etc. The error in variance minimization distinguishes the Kriging method from other methods like the inverse distance weighting. The advantage of the Kriging method above other methods is that it also provides the best linear estimate of rainfall for a point on the grid in addition to the uncertainty in the estimate. The latter property makes the method useful if the locations of additional stations have to be selected when the network is to be upgraded, because then the new locations can be chosen such that overall error variance is minimized. These days Kriging method is available with many GIS application.

6.3.7.2 Bias elimination and error variance minimization

The claims of unbiasedness and minimum error variance require further explanation. Let the true rainfall at location 0 be indicated by P_0 then the estimation error at "0" becomes:

$e_0 = Pe_0 - P_0$ Eqn. 6.11

with Pe_0 estimated by (6.5). It is clear from (6.6) that any statement about the mean and variance of the estimation error requires knowledge about the true behaviour of the rainfall at unmeasured locations, which is not known. This problem is solved by hypothesising:

- that the rainfall in the catchment is statistically homogeneous so that the rainfall at all observation stations is governed by the same probability distribution
- Consequently, under the above assumption, the rainfall at ungauged locations in the catchment follows the same probability distribution as applicable to the observation sites.

Hence, any pair of locations within the catchment (measured or unmeasured) has a joint probability distribution that depends only on the distance between the locations and not on their actual locations. Hence:

- at all locations E[P] is the same, hence $E[P(x_1)] - E[P(x_1\text{-}d)] = 0$, where d refers to the distance between various locations
- the covariance between any pair of locations is only a function of the distance d between the locations and not dependent of the location itself: C(d).

The unbiasedness implies:

$E(e_0)=0$ So: $E\left[\sum_{k=1}^{N} W_{0,k}.P_k\right]-E[P]$ Eqn. 6.12

Or $E[P_0]\left(\sum_{k=1}^{N} w_{0,k}\right)=0$ Eqn. 6.13

Hence for each and every grid point the sum of the weights should be 1 to ensure unbiasedness:

$\sum_{k=1}^{N} w_{0,k} = 1$ Eqn. 6.14

The error variance can be shown to be (Isaaks and Srivastava, 1989):

$\sigma_e{}^2 = E[(Pe_0 - P_0)^2] = \sigma_p{}^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} w_{0,i} w_{0,j} c_{i,j} - 2 \sum_{i=1}^{N} w_{0,i} c_{0,i}$ Eqn. 6.15

where 0 refers to the site with unknown rainfall and i,j to the observation station locations. Minimising the error variance implies equating the N first partial derivatives of $\sigma_e{}^2$ to zero to solve for the $w_{0,i}$. In doing so, the weights $w_{0,i}$ will not necessarily sum up to 1 as it should to ensure unbiasedness. Therefore, in the computational process one more equation is added to the set of equations to solve $w_{0,\ i}$, which includes a Lagrangian multiplier μ. The set of equations to solve the stations weights, also called **ordinary Kriging system**, then reads:

$$C\,W = D$$ Eqn. 6.16

Where:

$$C = \begin{bmatrix} C_{11} \ldots\ldots\ldots C_{1N}\,1 \\ .. \\ ... \\ C_{N1} \ldots\ldots\ldots C_{NN}\,1 \\ 1 \ldots\ldots\ldots\ldots 10 \end{bmatrix}$$

$$W = \begin{bmatrix} W_{0,1} \\ . \\ . \\ W_{0,N} \\ \mu \end{bmatrix}$$

$$D = \begin{bmatrix} C_{0,1} \\ . \\ . \\ C_{0,N} \\ 1 \end{bmatrix}$$

Note that the last column and row in **C** are added because of the introduction of the Lagrangian multiplier μ in the set of N+ 1 equation. By inverting the covariance matrix, the station weights to estimate the rainfall at location 0 are obtained by solving (Eqn. 6.13):

$w = C^{-1}.D$ Eqn. 6.17

The error variance is then determined from:

$\sigma_e{}^2 = \sigma_p{}^2 - w^T.D$ Eqn. 6.18

From the above equations it is observed that $\mathbf{C^{-1}}$ is to be determined only once as it is solely determined by the covariances between the observation stations being a function of the distance between the stations only. Matrix **D** differs for every grid point as the distances between location “0” and the gauging stations vary from grid point to grid point.

6.3.7.3 Covariance and variogram models

To actually solve the above equations, a function is required which describes the covariance of the rainfall field as a function of the distance between stations. For this we recall the correlation structure between the rainfall stations discussed in Chapter 4. The spatial correlation structure is usually well described by an exponential relation of the following type:

$r(d)=r_0\exp(-d/d_0)$ Eqn. 6.19

Where: $r(d)$ = correlation coefficient as a function of distance

r_0 = correlation coefficient at small distance, with $r_0 \le 1$

d_0 = characteristic correlation distance.

Two features of this function are of importance:

- $r_0 \le 1$, where values < 1 are usually found in practice due to measurement errors or micro-climatic variations
- the characteristic correlation distance d_0, i.e. the distance at which r(d) reduces to $0.37r_0$. It is a measure for the spatial extent of the correlation, e.g. the daily rainfall d_0 is much smaller than the monthly rainfall d_0. Note that for $d = 3d_0$ the correlation has effectively vanished (only 5% of the correlation at d = 0 is left).

The exponential correlation function is shown in Figure 6.12.

The **covariance function** of the exponential model is generally expressed as:

$C(d)=C_0 + C_1$ for d=0 Eqn. 6.20

$C(d)=C_1\exp(-\frac{3d}{a})$ for d> 0 Eqn. 6.21

Since according to the definition $C(d) = r(d)\sigma_P^2$, the coefficients C_0 and C_1 in (6.12) can be related to those of the exponential correlation model in (6.11) as follows:

$C_0=\sigma_p^2(1-r_0)$; $C_1=\sigma_p^2 r_0$ and $a=3d_0$ Eqn. 6.22

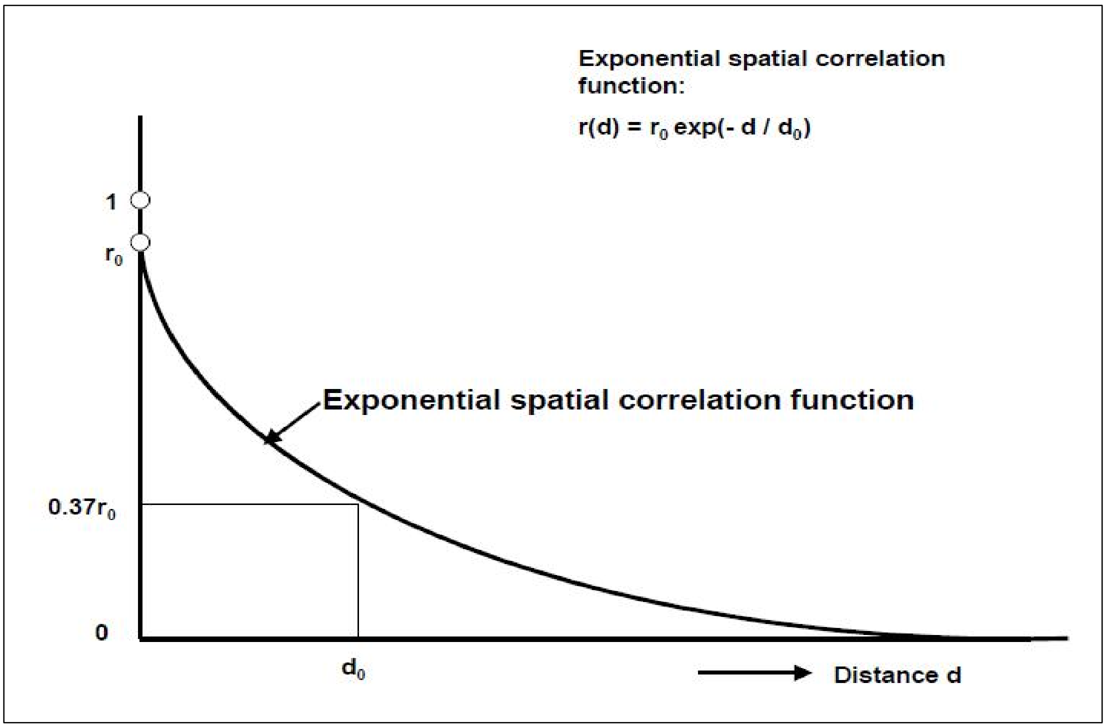


Figure 6.12: Spatial correlation structure of rainfall field

In Kriging literature, instead of using the covariance function C(d), the semi-variogram γ(d) is often used, which is half of the expected squared difference between the rainfall at locations distanced d apart; γ(d) is easily shown to be related to C(d) as:

$$\gamma(d) = \frac{1}{2}E[\{P(x_1)\text{-}P(x_1\text{-d})\}^2]=\sigma_p{}^2\text{-}C(d) \qquad \text{Eqn. 6.23}$$

Hence the **semi-variogram** of the exponential model reads:

$$\gamma(d)=0\,,\quad \text{for d=0}$$

$$\gamma(d)=C_0+C_1\left(1\text{-exp}(-\frac{3d}{a})\right)\text{for:}d > 0 \qquad \text{Eqn. 6.24}$$

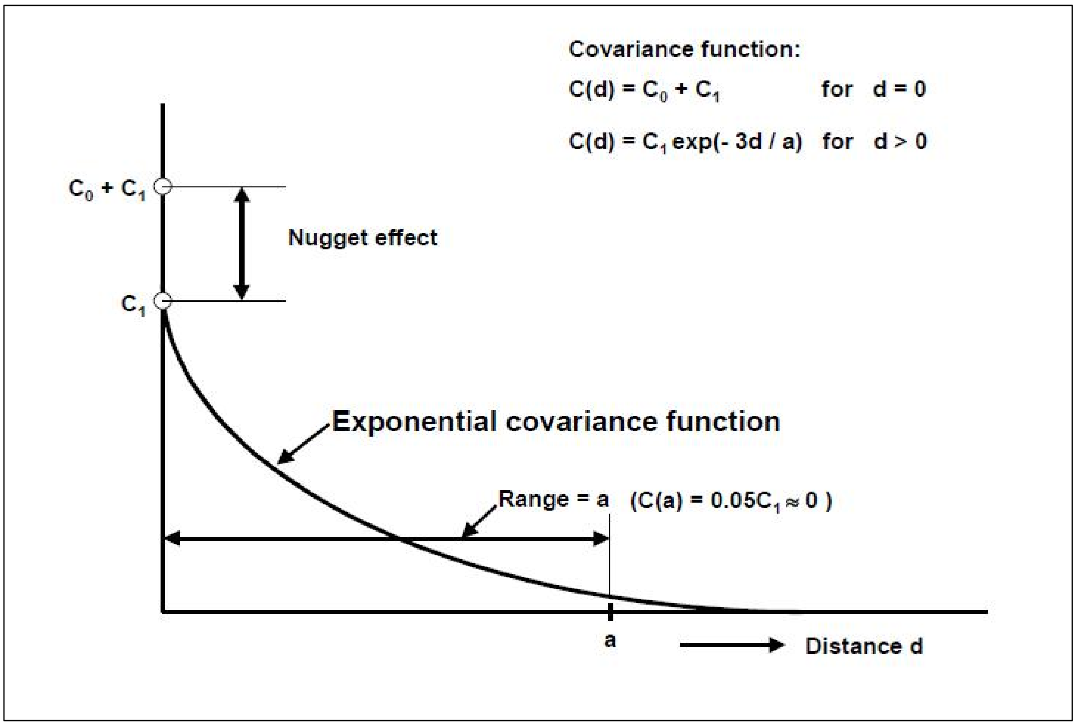


Figure 6.13: Exponential covariance model

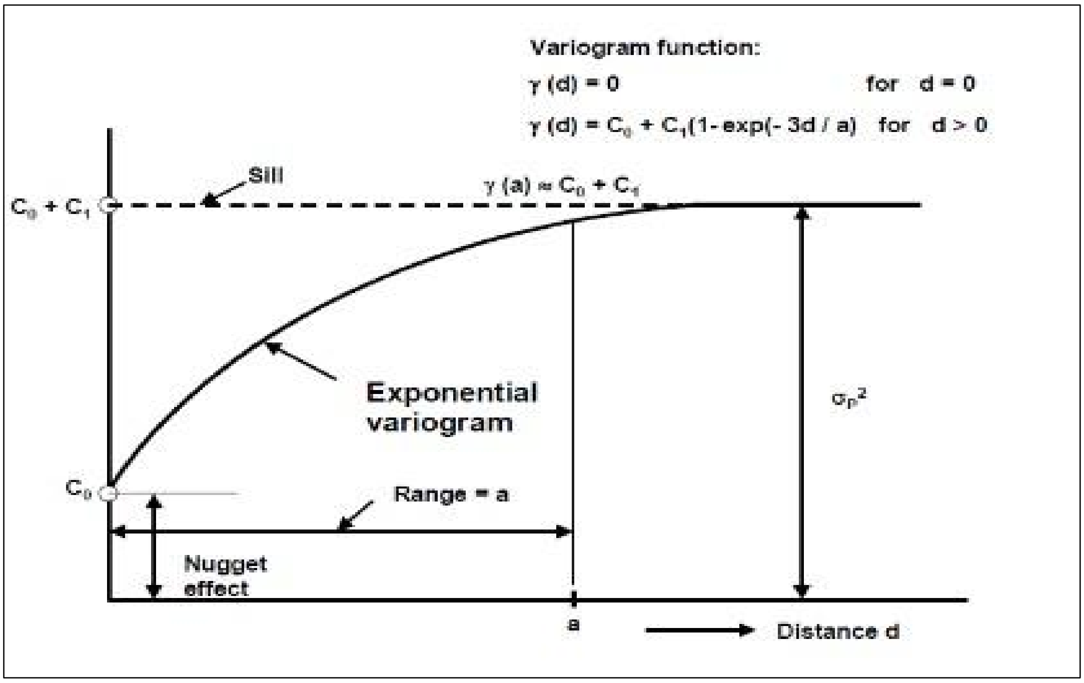


Figure 6.14:Exponential variogram model

The features of the exponential model are as follows:

- C_0, the **nugget effect** provides a discontinuity at the origin; according to (6.19): $C_0 = \sigma_P^2(1-r_0)$, hence in most applications of this model to rainfall data a small nugget effect is always present
- The distance 'a' in the covariance function and variogram is called the **range** and it refers to the distance above which the functions are essentially constant. For the exponential model the range of $a = 3d_0$ can be used.
- $C_0 + C_1$ is called the **sill** of the variogram and provides the limiting value for large distance and becomes equal to σ_P^2; it also gives the covariance for d = 0.

6.3.7.4 Other covariance and semi-variogram models

Beside the exponential model, other models are in use for ordinary Kriging, viz:

- Spherical model, and
- Gaussian model

These models have the following forms:

Spherical:

$$\text{if } d \leq a\,, \quad \gamma(d) = C_0 + C_1\left(\left(\frac{3}{2}\frac{d}{a}\right) - \frac{1}{2}\left(\frac{d}{a}\right)^3\right) \qquad \text{Eqn. 6.25}$$

Otherwise:

$$\gamma(d) = 1$$

Gaussian:

$$\gamma(d) = C_0 + C_1\left(1 - exp - \frac{3d^2}{a^2}\right) \qquad \text{Eqn. 6.26}$$

The Spherical and Gaussian models are shown with the Exponential Model in Figure 6.15.

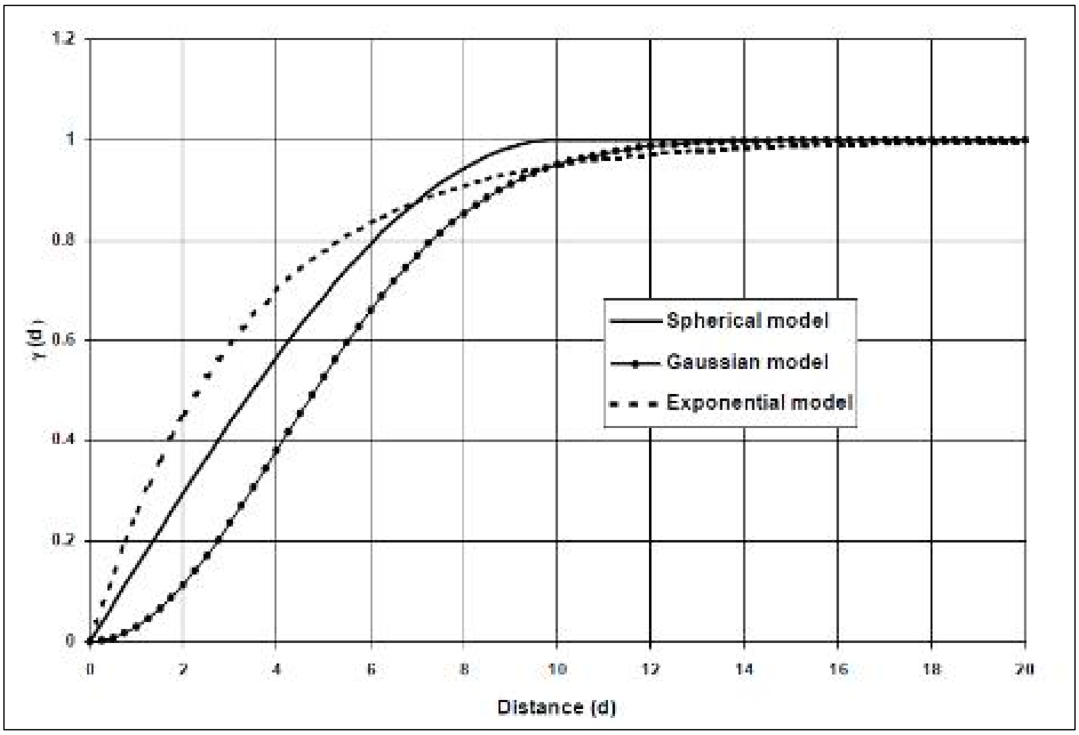


Figure 6.15: Example of spherical gaussian and exponential type of variogram models

6.3.7.5 Sensitivity analysis of variogram model parameters

To show the effect of variation in the covariance or variogram models on the weights attributed to the observation stations to estimate the value at a grid point, an example is presented by Isaaks and Srivastava (1989). Observations made at the stations as shown in Figure 6.16 are used. Some 7 stations are available to estimate the value at point '0' (65,137).

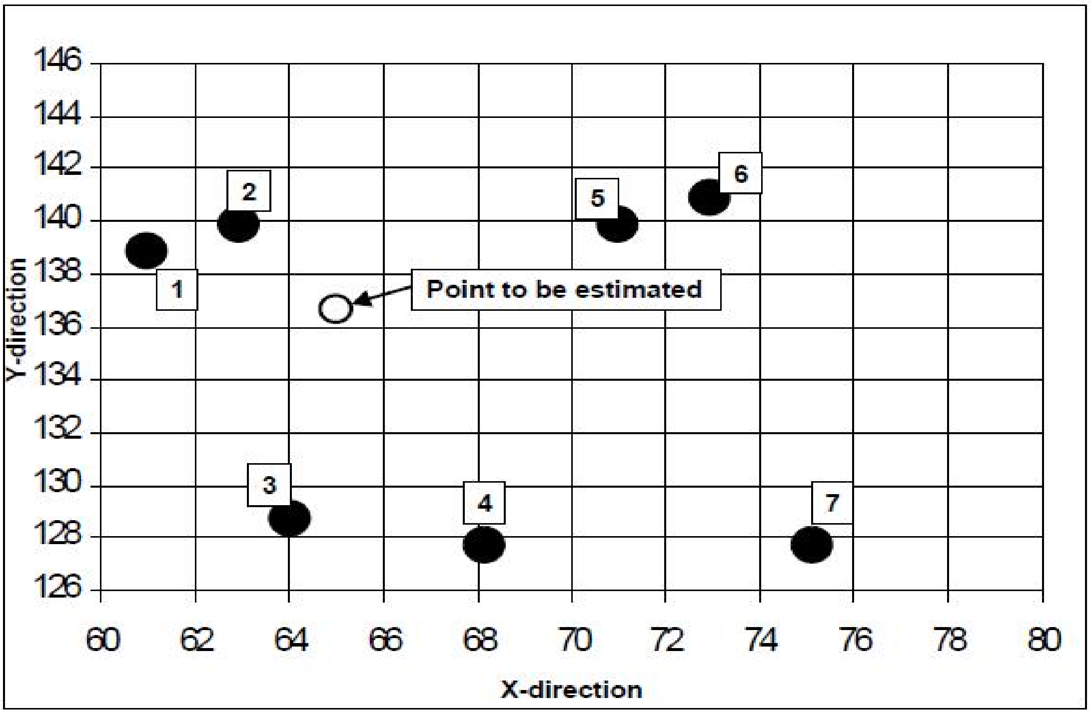


Figure 6.16: Layout of network with location of stations 1 to 7

Observations:

Station 1: 477
Station 2: 696
Station 3: 227
Station 4: 646
Station 5: 606
Station 6: 791
Station 7: 783

The following models (cases) have been applied to estimate the value for "0":

Case1:

$$\gamma_1(d) = 10\left(1 - \exp\left(-\frac{3d}{10}\right)\right), C_0 = 0\,,\ C_1 = 10\,,\ a = 10 \qquad \text{Eqn. 6.27}$$

Case 2:

$$\gamma_2(d) = 20\left(1 - exp(-\frac{3d}{10})\right) = 2\gamma_1(d),\ \ C_0 = 0\ , C_1 = 20\,,\ a = 10 \qquad \text{Eqn. 6.28}$$

Case 3:

$$\gamma_3(d) = 10\left(1 - exp(-3\left(\frac{d}{10}\right)^2)\right),\ \ C_0{=}0\ ,\ C_1{=}10\ ,\ \ \text{a=10(Gaussian)} \qquad \text{Eqn. 6.29}$$

Case 4:

$\gamma_4(d) = 0$ for d = 0

$$\gamma_4(d) = 5 + 5\left(1 - exp(-\frac{3d}{10})\right), \text{ for } d>0,\ C_0 = 5,\ C_1 = 10,\ a = 10 \qquad \text{Eqn. 6.30}$$

Case 5:

$$\gamma_5(d) = 10\left(1\text{-}\exp(-\frac{3d}{20})\right),\ C_0{=}0,\ C_1{=}10,\ a = 20 \qquad \text{Eqn. 6.31}$$

The covariance and variograms for the cases are shown in Figure 6.17 and Figure 6.18.

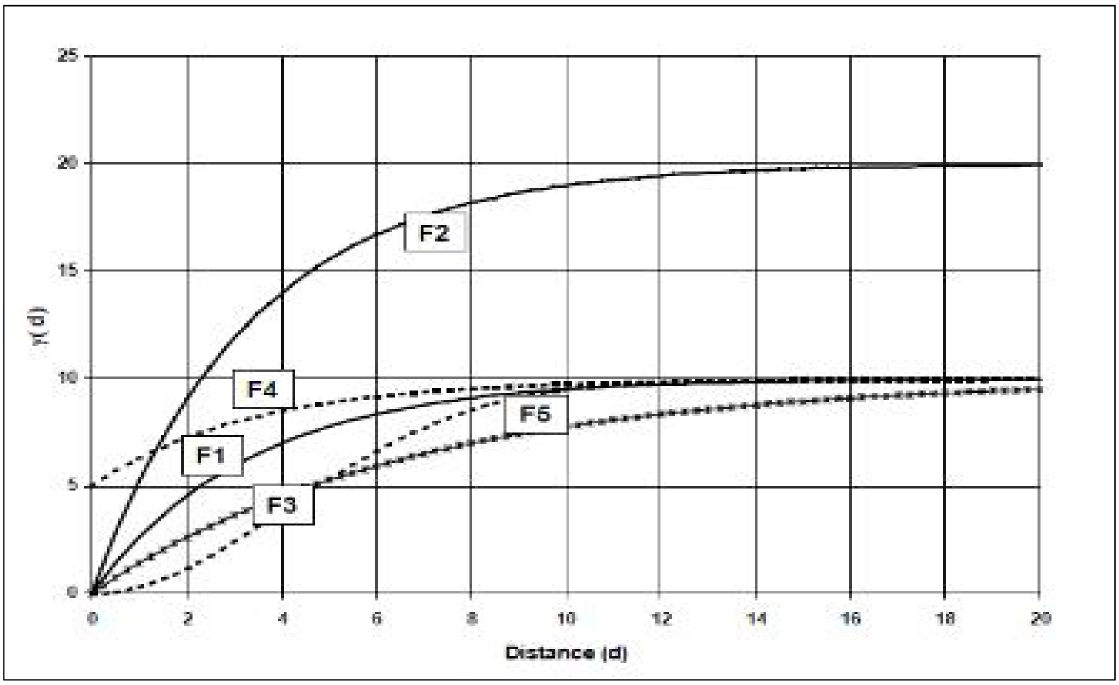


Figure 6.17: Covariance models for the various cases

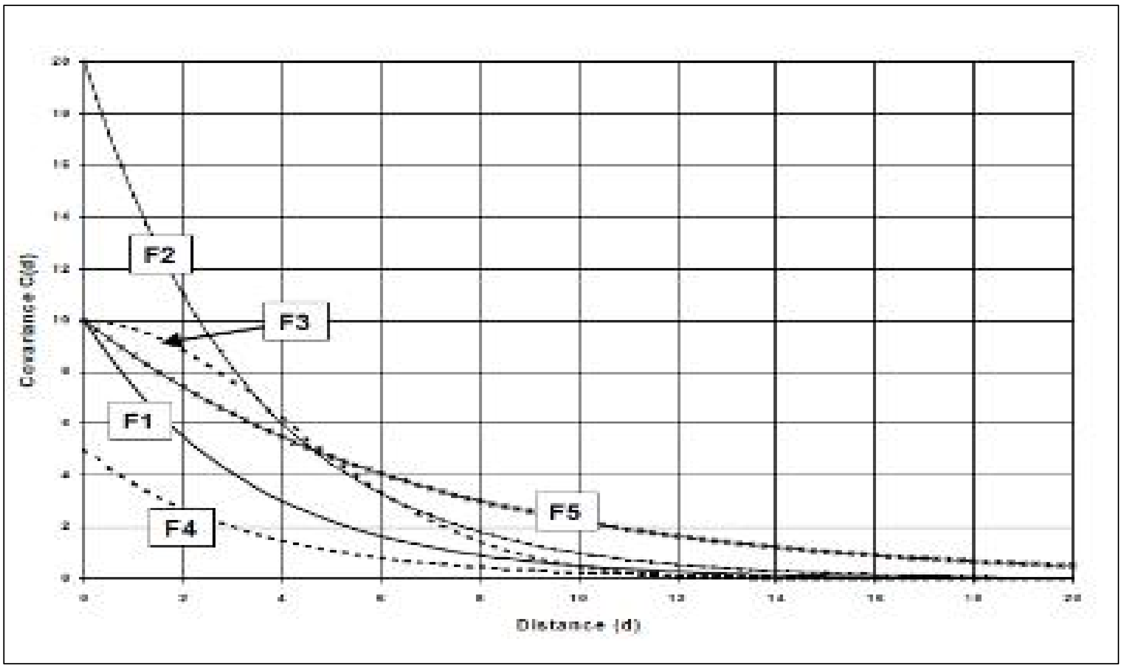


Figure 6.18: Semi-variograms for the various cases

The results of the estimate and variance at point “0” as well as the weights of the stations computed with the models in estimating point “0” are presented in Table 6.5.

Table 6.5: Results of computations for Cases 1 to 5 and IDW (Inverse Distance Weighted Method) with p = 2

Case	Estimate at “0” (mm)	Error variance	Station Number (Distance from 0)						
			1	2	3	4	5	6	7
			(4.47)	(3.61)	(8.06)	(9.49)	(6.71)	(8.94)	(13.45)
			Weights						
1	593	8.86	0.17	0.32	0.13	0.09	0.15	0.06	0.09
2	593	17.91	0.17	0.32	0.13	0.09	0.15	0.06	0.09
3	559	4.78	-0.02	0.68	0.17	-0.01	0.44	-0.29	0.04
4	603	11.23	0.15	0.18	0.14	0.14	0.13	0.13	0.14
5	572	5.76	0.18	0.38	0.14	0.07	0.2	0.0	0.03
ID	590	-	0.44	0.49	0.02	0.01	0.02	0.01	0.01

From the results the following can be concluded:

- **Effect of scale**: compare Case 1 with Case 2

In Case 2 the process variance, i.e. the sill is twice as large as in Case 1. The only effect this has on the result is a doubled error variance at “0”. The weights and therefore also the estimates remain unchanged. The result is easily confirmed from equations (3.9) and (3.10) as both **C, D** and σ_P^2 are multiplied with a factor 2 in the second case.

- **Effect of shape: compare Case 1 with Case 3**

In Case 3 the spatial continuity near the origin is much larger than in Case 1, but the sill is the same in both cases. It is observed that in Case 3 the estimate for “0” is almost entirely determined by the three nearest stations. Note that Kriging does cope with clustered stations; even negative weights are generated by stations in the clusters of stations (5, 6) and (1, 2) to reduce the effect of a particular cluster. Note also that the estimate has changed and that the error variance has reduced as more weight is given to the stations at sorter distance from station 0. It shows that due attention is to be given to the correlation structure at small distances as it affects the outcome significantly.

- **The nugget effect: compare Case 1 with Case 4**

In Case 4, which shows a strong nugget effect, the spatial correlation has substantially been reduced near the origin compared to Case 1. As a result, the model discriminates less among the stations. This is reflected in the weights given to the stations. It is observed that almost equal weight is given to the stations in Case 4. In case correlation would have been zero the weights would have been exactly equal.

- **Effect of range: compare Case 1 with Case 5**

The range in Case 5 is twice as large as in Case 1. It means that the spatial correlation is more pronounced than in Case 1. Hence, one would expect higher weights to the nearest stations and a reduced error in variance, which is indeed the case as can be observed from Table 6.5. Cases 1 and 5 basically are representative for rainfall at a low and high aggregation level, respectively (e.g. daily data and monthly data).

There are more effects to be concerned about, like the effects of anisotropy (spatial covariance being direction dependent) and spatial inhomogeneity (e.g. trends due to orographic effects). The latter can be dealt with by normalising or detrending the data prior to the application of Kriging and denormalise or re-invoke the trend after the computations. In case of anisotropy the contour map of the covariance surface will be elliptic rather than circular. Anisotropy will require variograms to be developed for the two main axis of the ellipse separately.

6.3.7.6 Estimation of spatial covariance function or variogram

Generally, the spatial correlation (and hence the spatial covariance) as a function of distance will show a huge scatter as shown in Figure 4.2 to Figure 4.5. To reduce the scatter, the variogram is being estimated from the average values per distance interval. The distance intervals are equal and should be selected such that sufficient data points are present in an interval but also that the correct nature of the spatial correlation is reflected in the estimated variogram.

Like other interpolation algorithms, the Kriging method tends to smooth out local details of the spatial variability of the attribute, leading to overestimation of small values and underestimation of large ones. The quality of estimates produced by ordinary Kriging depends on the time taken to choose an appropriate model of the spatial continuity. Ordinary Kriging with a poor model may produce worse estimates than the other simpler methods.

6.4 Transformation of non-equidistant to equidistant series

Data obtained from digital rain-gauges based on the tipping bucket may sometime be recording information at the time of each tip of the tipping bucket, i.e. a non-equidistant series.

Such non-equidistant series need to transfer to equidistant series by accumulating each unit tip measurement to the corresponding time interval. All those time intervals for which no tip has been recorded are filled with zero values.

6.5 Compilation of minimum, maximum and mean series

Daily maximum rainfalls (or instantaneous, if available) found within each year or season should be compiled into a maximum rainfall series as it is frequently used for flood analysis, while the minimum rainfall statistics on a seasonal or monthly basis may be required for drought analysis. The extraction of minimum, maximum, mean, median and any 25% and 90% percentile values (at a time) for any defined period within the year or for the complete year are given in the following example.

Example 6-4

A ten-daily data series is compiled from the daily rainfall records available for Megharaj station (Kheda catchment). For this ten-daily data series for the period 1961 to 1997, a few statistics like the minimum, maximum, mean, median and 25 & 90 percentile values are compiled specifically for the period between 1st July and 30th Sept. every year.

These statistics are shown graphically in

Figure 6.19 and are listed in tabular form in Table 6.6. Data of one of the years (1975) is not available and is thus missing. Many inferences may be derived from plots of such statistics. Different patterns of variation between 25 and 90 percentile values for similar ranges of values in a year may be noticed. Median value is always lower than the mean value suggesting higher positive skew in the ten-daily data (which is obvious owing to many zero or low values). A few extreme values have been highlighted in the table for general observation.

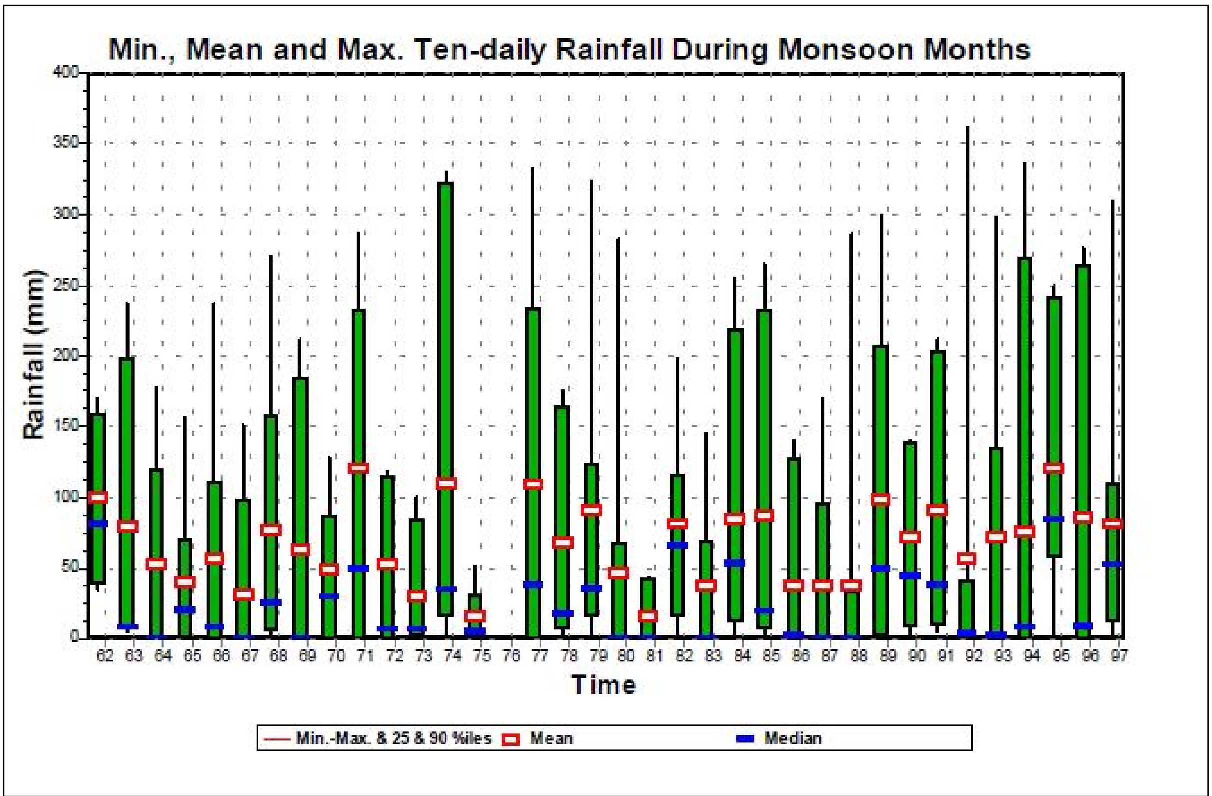


Figure 6.19: Plot of statistics of ten-daily rainfall series at station Megharaj

Table 6.6: Ten-daily statistics for station Megharaj between 1st July and 30st Sept

Year	Min.	Max.	Mean	Median	25 %ile	90 %ile
1961	34.54	170.39	99.6	81.03	39.36	158.47
1962	5.60	237.60	78.90	8.60	8.4	197.50
1963	0.00	177.44	53.00	0.00	0.00	119.10
1964	0.00	157.20	39.70	20.70	1.70	69.60
1965	0.00	237.00	56.30	8.00	0.00	110.60
1966	0.00	151.00	31.40	0.00	0.00	98.00
1967	0.00	270.00	75.90	26.00	6.00	158.00
1968	0.00	211.00	63.00	0.00	0.00	185.00
1969	0.00	128.00	49.20	30.00	0.00	87.00
1970	0.00	287.00	120.70	50.00	0.00	232.00
1971	0.00	118.50	53.10	7.00	0.00	114.00
1972	0.00	99.60	29.90	7.00	2.60	83.30
1973	0.00	330.40	110.80	34.80	17.00	322.60
1974	0.00	51.00	16.50	5.00	1.50	31.20
1976	0.00	333.40	108.80	38.20	0.00	234.20
1977	0.00	175.40	67.60	18.00	7.00	164.00
1978	0.00	324.00	90.30	36.00	16.00	123.00
1979	0.00	282.00	46.00	0.00	0.00	67.00
1980	0.00	43.00	15.30	0.00	0.00	42.00
1981	0.00	198.00	81.00	65.50	16.00	115.50
1982	0.00	144.00	38.50	0.00	0.00	69.00
1983	0.00	256.00	84.70	54.00	12.00	219.00
1984	0.00	265.00	87.00	19.50	7.50	231.50
1985	0.00	140.50	36.90	3.00	0.00	127.00
1986	0.00	170.00	38.40	0.00	0.00	94.50
1987	0.00	287.00	38.50	0.00	0.00	33.00
1988	0.00	300.00	99.00	50.00	3.00	207.00
1989	0.00	140.00	72.30	44.50	9.00	138.50
1990	5.00	211.50	91.10	38.50	10.00	203.50
1991	0.00	361.50	56.70	4.00	0.00	41.50
1992	0.00	298.00	72.20	3.00	0.00	134.00
1993	0.00	336.50	75.70	8.00	0.00	269.00
1994	0.00	249.00	121.10	85.00	58.50	241.50
1995	0.00	276.50	85.90	9.50	0.00	264.00
1996	0.00	309.00	81.90	52.50	13.50	109.00
1997	0.00	391.00	105.70	23.00	10.00	242.5
Full Period	0.00	391.00	68.70			

7 RAINFALL DATA ANALYSIS

7.1 General

Various kinds of analysis are required for data validation of rainfall time series, principally aimed to detect and describe quantitatively all generating processes underlying a given sequence of observations. Some analysis may further be required for data presentation and reporting.

The types of processing considered for rainfall data in the current Chapter are:

- checking data homogeneity
- computation of basic statistics
- annual exceedance rainfall series
- fitting of frequency distributions
- frequency and duration curves

Most of the hydrological analysis for purpose of validation will be carried out at the Divisional and State Data Processing Centres and for the final presentation and reporting at the State Data Processing Centres.

7.2 Checking data homogeneity

Ideally, rainfall data from a single series should ideally be homogeneous. This property implies that different portions of the data series do not vary significantly in statistical terms. Similarly, rainfall data for multiple series at neighbouring stations should ideally possess spatial homogeneity.

Tests of homogeneity are required for validation purposes, and there is a shared need for such tests with other climatic variables as well. Tests related to data validation, spatial homogeneity and data consistency using double mass curves are explained in other sections of this manual.

Single series tests of homogeneity include trend analysis, mass curves, residual mass curves, Student's t and Wilcoxon W-test on the difference of means and Wilcoxon-Mann-Whitney U test to investigate if the sample are from same population.

Multiple station validation includes comparison plots, residual series, regression analysis and double mass curves.

7.3 Computation of basic statistics

Basic statistics are widely required for validation and reporting. The following are commonly used:

- Arithmetic mean
- Median - the median value of a ranked series Xi
- Mode - the value of X which occurs with greatest frequency or the middle value of the class with greatest frequency
- Standard Deviation - the root mean squared deviation Sx

The Standard Deviation calculates the deviations of each data point from the mean, and squares the resulting sum of differences divided by the number of points. Standard deviation is equal to the square root of the variance:

$$S_x = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \overline{X})}{N-1}} \qquad \text{Eqn. 7.1}$$

- Skewness and Kurtosis

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if its statistical density distribution function looks the same to the left and right of the centre point. A distribution is skewed if one of its tails is longer than the other. The first distribution shown has a positive skew. This means that it has a long tail in the positive direction. The distribution below it has a negative skew since it has a long tail in the negative direction.

Skewness formula for a statistical sample is given by:

$$S = \frac{n\sqrt{n-1}}{n-2} \frac{\sum_{i=1}^{N}(X_i - X_{avg})^3}{(\sum_{i=1}^{N}(X_i - X_{avg})^2)^{3/2}} \qquad \text{Eqn. 7.2}$$

Kurtosis is a measure of whether the data are heavily-tailed or light-tailed relative to a normal distribution. Formula for coefficient of Kurtosis for a sample is given by

$$K = \frac{n(n+1)(n-1)}{(n-2)(n-3)} \frac{\sum_{i=1}^{N}(X_i - X_{avg})^4}{(\sum_{i=1}^{N}(X_i - X_{avg})^2)^2} \qquad \text{Eqn. 7.3}$$

In addition, empirical frequency distributions can be presented as a graphical representation of the number of data per class and as a cumulative frequency functions from which the exceedance probability values can be extracted, e.g. the daily rainfall which has been exceeded 1%, 5% or 10% of the time.

- Decile:

In statistics, a decile is any of the nine values that divide the sorted data into ten equal size bins, so that each bin (part) represents $1/10^{th}$ of the sampled population. A decile is one possible form of a quantile. The data series is sorted N data points (numbers) and the $n/10^{th}$ data point is the 1^{st} decile, $2n/10^{th}$ item is the 2^{nd} decile and so on. If indexes n/10, 2n/10, ..., 9n/10 are not integers, then we use interpolation between the nearest data points.

For example, for n=100 items, the first decile is the 10^{th} data point of ordered data set, 6^{th} decile is the 60^{th} data point, etc.

Example 7-1

The basic statistics for monthly rainfall data of Ahmednagar station are derived for the period 2004 to 2010. The analyses are carried out by taking the actual values and all the months in the year. The results are given in Table 7.1. The frequency distribution and the cumulative frequency is worked out for 7 classes between 0 and 350 rainfall data points and is given in tabular form and as the graph in Figure 7.1. Various decile values are also listed in the result of the analysis.

Since the actual monthly rainfall values are not normally distributed, the data will exhibit some skewness (1.26) and kurtosis (1.01). The value of mean is larger than the median value and the frequency distribution shows a positive skew. From the table of decile values, it can be seen that 50 % of the months receive less than 2.7 mm of rainfall. From the cumulative frequency table, it may be seen that 40 percent of the months receive zero rainfall (which can be expected in this catchment) and that there are very few instances when the monthly rainfall total is above 200 mm. A smaller size of the data bin (i.e. fewer than 50 data points) would increase the accuracy of the frequency curve. The frequency distribution function can be obtained by sorting all data points and applying the Weibull plotting position formula P=m/(n+1) where m is the rank of the sorted data point and n is the total number of data points. This calculation can also be performed using the percentile.exc() function available in excel.

Table 7.1: Computational results of the basic statistics for monthly rainfall at Ahmednagar

State / District	Maharashtra / Ahmednagar						
Year	2004	2005	2006	2007	2008	2009	2010
January	0.00	0.80	0.00	0.00	0.00	0.00	0.00
February	0.00	0.00	0.00	0.00	0.00	0.00	0.00
March	0.00	0.10	0.00	0.00	51.9	0.00	0.00
April	0.00	1.80	0.00	0.00	4.50	0.00	0.00
May	17.10	0.70	8.10	0.00	0.00	0.00	2.50
June	123.60	82.70	166.10	205.20	37.70	75.60	173.70
July	103.50	146.90	124.30	101.80	65.20	122.00	165.00
August	99.70	79.60	176.20	126.40	108.40	127.30	196.40
September	262.00	214.00	225.90	157.80	327.30	123.00	171.20
October	60.20	114.90	71.20	0.00	53.00	73.40	60.60
November	12.20	0.00	0.00	0.00	0.90	127.60	74.70
December	0.00	0.00	0.00	0.00	0.00	2.90	0.00
Annual Total	678.30	641.50	771.80	591.20	648.90	651.80	844.10

Statistics

Range	Frequency	Cumulative frequency
0-50	48	48
50-100	12	60
100-150	12	72
150-200	7	79
200-250	3	82
250-300	1	83
300-350	1	84

Calculation

Mean	57.4714
Standard deviation	76.5811
Skewness	1.26290
Kurtosis	1.01169

Decile	Value (mm)
1	0.00
2	0.00
3	0.00
4	0.00
5	2.70
6	60.20
7	91.20
8	124.30
9	172.45

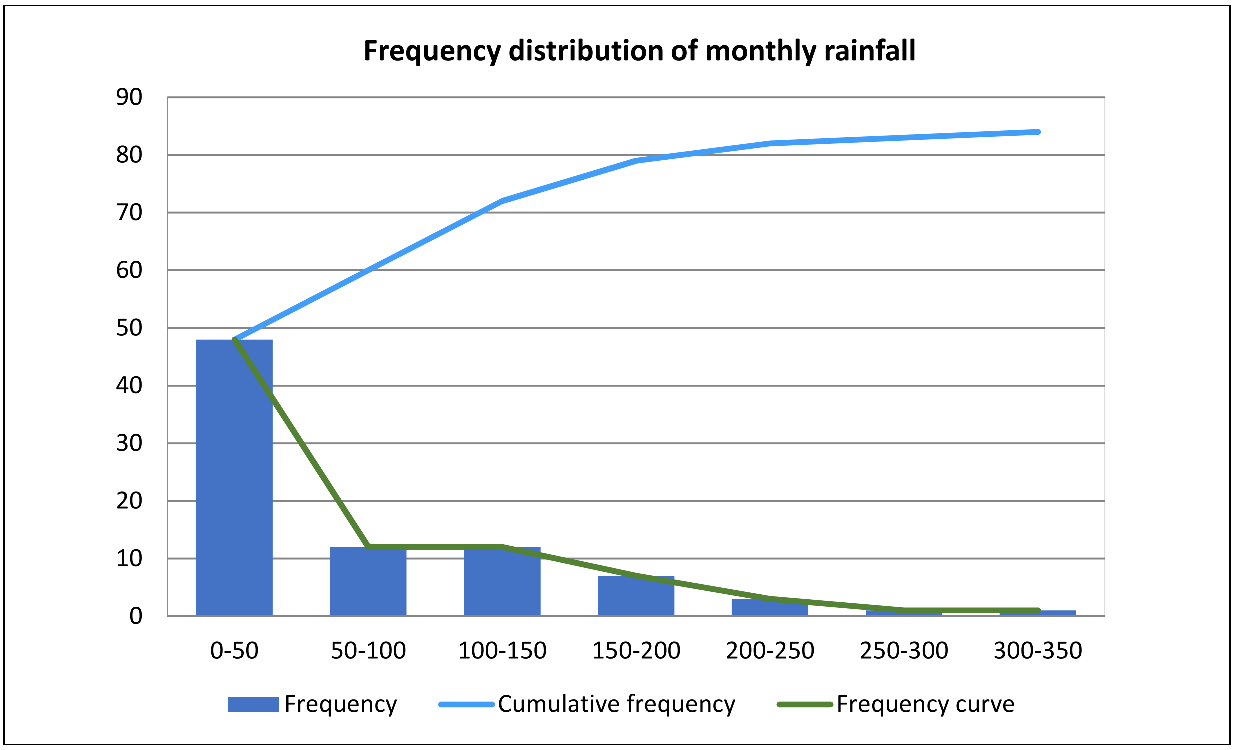


Figure 7.1: Frequency and cumulative frequency plot of monthly rainfall at Ahmednagar

7.4 Annual exceedance rainfall series

The following are widely used for reporting or for subsequent use in frequency analysis of extremes:

a) Maximum of a series - the maximum rainfall value of an annual series or of a month or season may be selected. Also, all values (peaks) over a specified threshold may be selected. For rainfall, daily maxima per year are commonly used, but hourly or N-hourly maxima may also be selected for specific analyses.
b) Minimum of a series - As the minimum daily value with respect to rainfall is frequently zero, so this statistic does not have the same importance as the maximums.

7.5 Fitting of frequency distributions

A common use of rainfall data is in the assessment of probabilities or return periods of given rainfall at a given location. Such data can then be used in assessing peak flood discharge for desired return periods through modelling, or by using a previously developed empirical formulas. Calibrated rainfall-runoff models can also be applied in flood forecasting and for the design of various hydraulic structures including bridges and culverts.

Frequency analysis usually involves fitting of theoretical frequency distribution using a selected fitting method, although empirical graphical methods can also be applied. The fitting of a particular distribution implies that the rainfall sample of annual maxima were drawn from a population of that distribution. For the purpose of application in design, it is assumed that future probabilities of exceedance will not change compared to those derived from the historic data. However, there is nothing inherent in the series to indicate whether one distribution is more likely to be appropriate than another and a wide variety of distributions and fitting procedures have been recommended for application in different countries and by different agencies.

Different distributions can give widely different estimates, especially when extrapolated or when an outlier (an exceptional value, well in excess of the second largest value) occurs in the data set. A degree of subjectivity is introduced in the selection of which distribution to apply.

Caution is advised when interpreting and reporting the results of the frequency analyses methods. Graphical as well as numerical output should always be inspected. The higher the degree of aggregation of input data, the closer the data fit to the normal distribution. The following frequency distributions are considered available in current practice

- Normal and log-normal distributions
- Pearson Type III or Gamma distribution
- Log-Pearson Type III
- Extreme Value type I (Gumbel), II, or III
- Generalized Extreme Value
- Goodrich/Weibull distribution
- Exponential distribution
- Pareto distribution

The following fitting methods available for fitting the above distributions are:

- Modified maximum Likelihood
- Method of Moments
- PWM methods (Probability Weighted Moments)
- L-moments

The following outputs are derived for each distribution:

- Estimation of parameters of the distribution
- A table of rainfalls of specified Exceedance Probabilities or Return Periods with confidence limits
- Results of Goodness of Fit Tests
- Graphical plot of the data fitted to the distribution

The above methods are very complex and details on these can be referred from Standard Textbooks on Frequency Analysis covered in Stochastic Hydrology.

7.6 Frequency and duration curves

A convenient way to show the variation of hydrological quantities through the year by means of frequency curves, where each frequency curve indicates the magnitude of quantity for a specific probability of non-exceedance. The duration curves are ranked representation of these frequency curves. The average duration curve gives the average number of occasions a given value was not exceeded in the years considered. The computation of frequency and duration curves is as given below:

7.6.1 Frequency curves

Consider "n" elements of rainfall values in a selected series that is statistically analysed. If the selected dataset is arranged in ascending order of magnitude, the probability that the i^{th} element X_i of this ranked sequence of elements is not exceeded is:

$$Fi = \frac{i}{n+1} \qquad \text{Eqn. 7.4}$$

The frequency curve connects all values of the quantity for j=1, n with the common property of equal probability of non-exceedance. Generally, a group of curves is considered which represents specific points of the cumulative frequency distribution for each j. Considering that curves are derived for various frequencies Fk {k=1, nf}, then values for rainfall Rk, j is obtained by linear interpolation between the probability values immediately greater (Fi) and lesser (Fi-1) to nk for each j as:

$$R_{k,j0} = R_{i-1} + (R_i - R_{i-1})\frac{F_k - F_{i-1}}{F_i - F_{i-1}} \qquad \text{Eqn. 7.5}$$

7.6.2 Duration curves

When the data Rk,j, k=1, nf and j=1, n is ranked in descending order for each k, the ranked matrix represents the duration curves for given probabilities of non-exceedance.

When all the data is considered without discriminating for different elements and ranked in the descending order of magnitude, then the resulting sequence shows the duration curve. This indicates how often a given quantity will be exceeded in a year (or month or day).

Example 7-2

A long-term monthly rainfall data series of Ahmednagar station (Kheda catchment) is considered for deriving frequency curves and duration curves. Analysis is done on yearly basis and the various frequency levels set are 10, 25, 50, 75 and 90%.

Figure 7.2 shows the frequency curves for various values (10, 25, 50, 75 and 90%) for each month in the year. Monthly rainfall distribution in the year 1982 is also shown superimposed on this plot for comparison. Minimum and maximum values for each month of the year in the plot give the range of variation of rainfall in each month. The results of this frequency curve analysis are shown in the tables below.

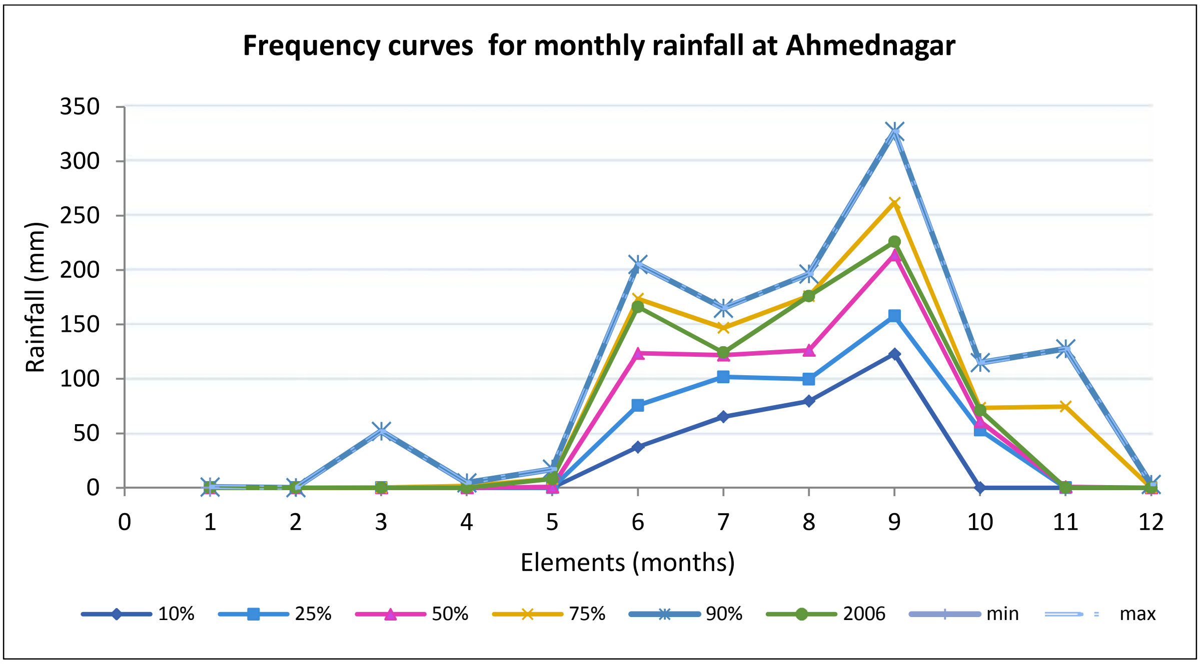


Figure 7.2: Monthly frequency curves for rainfall at Ahmednagar station

The plot in Figure 7.2 gives values of monthly rainfall which will not be exceeded for certain number of months in a year with the specified level of probability. The results of analysis for these duration curves are given in Table 7.2, Table 7.3 and Table 7.4. Monthly duration curve showing monthly values of rainfall which will likely not be exceeded in a year for the selected months is given as Figure 7.3.

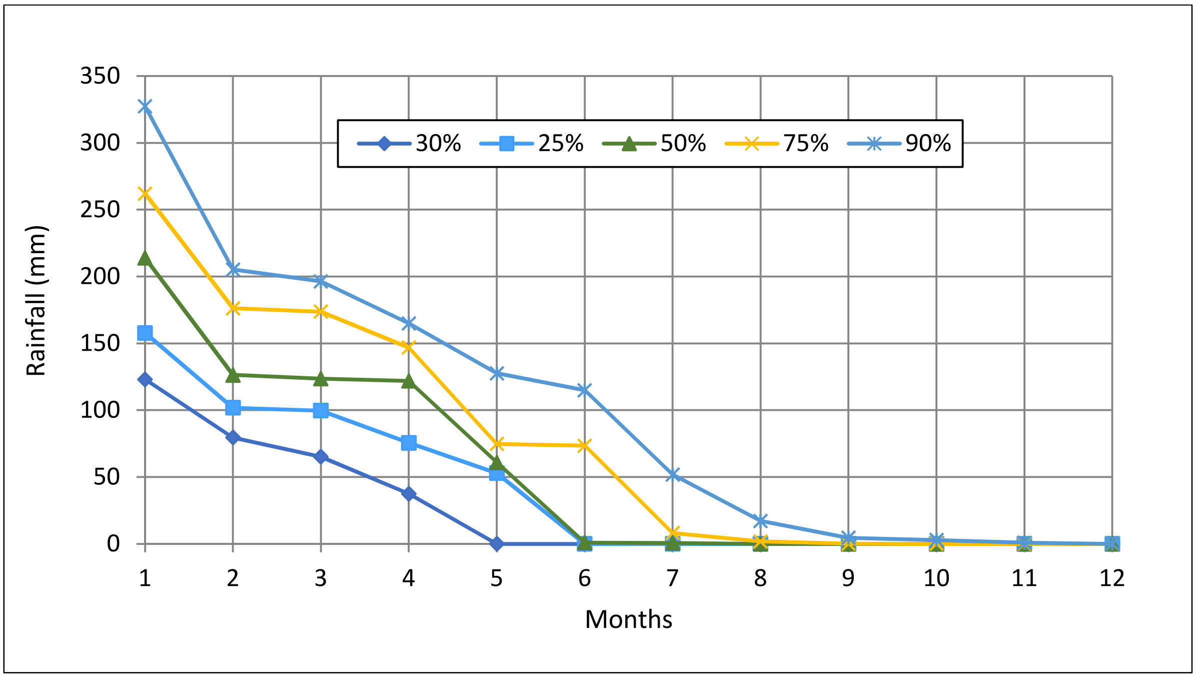


Figure 7.3: Duration for monthly rainfall series

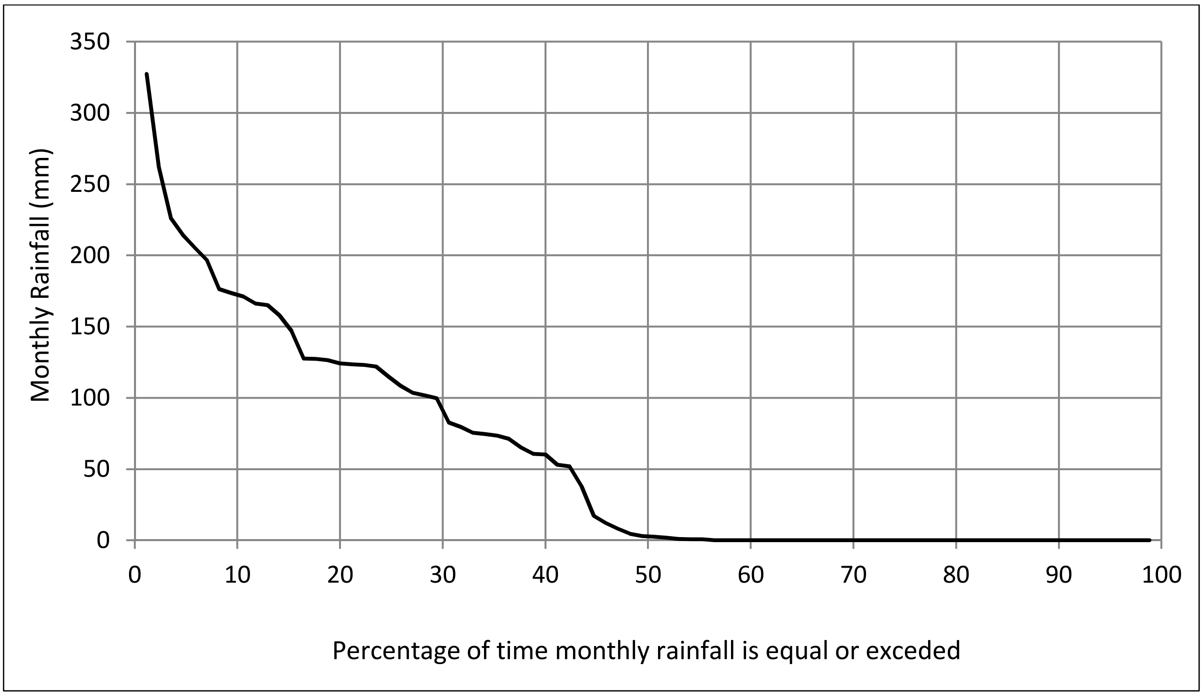


Figure 7.4: Data for Results of Analysis for monthly duration curves for Ahmednagar Station

Table 7.2: Results of analysis for frequency curves for monthly data for Ahmednagar station (rainfall values in mm)

Element	No of data	Frequency					Year 2006	Min	Max
		0.1	0.25	0.5	0.75	0.9			
1	7	0.00	0.00	0.00	0.00	0.80	0.00	0.00	0.80
2	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	7	0.00	0.00	0.00	0.10	51.90	0.00	0.00	51.90
4	7	0.00	0.00	0.00	1.80	4.50	0.00	0.00	4.50
5	7	0.00	0.00	0.70	8.10	17.10	8.10	0.00	17.10
6	7	37.50	75.60	123.60	173.70	205.20	166.10	37.70	205.20
7	7	65.20	101.80	122.00	146.90	165.00	124.30	65.20	165.00
8	7	79.60	99.70	126.40	176.20	196.40	176.20	79.60	196.40
9	7	123.00	157.80	214.00	262.00	327.30	225.90	1230	327.30
10	7	0.00	53.00	61.00	73.40	114.90	71.20	0.00	114.90
11	7	0.00	0.00	0.90	74.70	127.60	0.00	0.00	127.60
12	7	0.00	0.00	0.00	0.00	2.90	0.00	0.00	2.90

Table 7.3: Results of analysis for duration curves for monthly data for Ahmednagar station (rainfall values in mm)

Element	Frequency					Year 2006	Min	Max
	0.1	0.25	0.5	0.75	0.9			
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.80	0.00	0.00	0.80
3	0.00	0.00	0.00	0.00	2.90	0.00	0.00	2.90
4	0.00	0.00	0.00	0.10	4.50	0.00	0.00	4.50
5	0.00	0.00	0.00	1.80	17.10	0.00	0.00	17.10
6	0.00	0.00	0.70	8.10	51.90	0.00	0.00	51.90
7	0.00	0.00	0.90	73.40	114.90	8.10	0.00	114.90
8	0.00	53.00	61.00	74.70	127.60	71.20	0.00	127.60
9	37.50	75.60	122.00	146.90	165.00	124.30	37.70	165.00
10	65.20	99.70	123.60	173.70	196.40	166.10	65.20	196.40
11	79.60	101.80	126.40	176.20	205.20	176.20	79.60	205.20
12	123.00	157.80	214.00	262.00	327.30	225.90	123.00	327.30

Table 7.4: Results of analysis for average duration curves for monthly data for Ahmednagar station (rainfall values in mm)

Rank	Rain-fall value	Percent of time exceed	Rank	Rain-fall value	Percent of time exceed	Rank	Rainfall value	Percent of time exceed	Rank	Percent of time exceed
1	327.30	1.18	22	108.40	25.88	43	2.50	50.59	64	75.29
2	262.00	2.35	23	103.50	27.06	44	1.80	51.76	65	76.47
3	225.90	3.53	24	101.80	28.24	45	0.90	52.94	66	77.65
4	214.00	4.71	25	99.70	29.41	46	0.80	54.12	67	78.82
5	205.20	5.88	26	82.70	30.59	47	0.70	55.29	68	80.00
6	196.40	7.06	27	79.60	31.76	48	0.10	56.47	69	81.18
7	176.20	8.24	28	75.60	32.94	49	0.00	57.65	70	82.35
8	173.70	9.41	29	74.70	34.12	50	0.00	58.82	71	83.53
9	171.20	10.59	30	73.40	35.29	51	0.00	60.00	72	84.71
10	166.10	11.76	31	71.20	36.47	52	0.00	61.18	73	85.88
11	165.00	12.94	32	65.20	37.65	53	0.00	62.35	74	87.06
12	157.80	14.12	33	60.60	38.82	54	0.00	63.53	75	88.24
13	146.90	15.29	34	60.20	40.00	55	0.00	64.71	76	89.41
14	127.60	16.47	35	53.00	41.18	56	0.00	65.88	77	90.59
15	127.30	17.65	36	51.90	42.35	57	0.00	67.06	78	91.76
16	126.40	18.82	37	37.70	43.53	58	0.00	68.24	79	92.94
17	124.30	20.00	38	17.10	44.71	59	0.00	69.41	80	94.12
18	123.60	21.18	39	12.20	45.88	60	0.00	70.59	81	95.29
19	123.00	22.35	40	8.10	47.06	61	0.00	71.76	82	96.47
20	122.00	23.53	41	4.50	48.24	62	0.00	72.94	83	97.65
21	114.90	24.71	42	2.90	49.41	63	0.00	74.12	84	98.82

7.7 Intensity-duration-frequency analysis

7.7.1 General

If rainfall data from a recording rain gauge is available for long periods such as 25 years or more, the frequency of occurrence of a given intensity can also be determined, allowing the constriction of the intensity-frequency-duration curves. Such curves can be established for different parts of the year, e.g. a month, a season or the full year. This section describes a procedure to obtain such relationships for the entire year. The method for parts of the year is similar.

The entire rainfall record in a year is analysed to find the maximum intensities for various durations. Thus, each storm gives one value of maximum intensity for a given duration. The largest of all such values is taken to be the maximum intensity in that year for that duration. Likewise, the annual maximum intensity is obtained for different duration. Similar analyses yield the annual maximum intensities for various durations in different years. It will then be observed that the annual maximum intensity for any given duration is not the same every year but it varies from year to year. In other words, it behaves as a random variable. So, if 25 years of record is available then there will be 25 values of the maximum intensity of any given duration, which constitute a random variable sample. These 25 values of any one duration can be subjected to frequency analysis. The observed frequency distribution often fits the Gumbel distribution. A fit to a theoretical distribution function like the Gumbel is required if maximum intensities at return periods larger than the observed are required. Similar frequency analyses are carried out for other durations. The graphs of maximum rainfall intensity against the return period for various durations such as those shown in Figure 7.5 can be developed from the results of these analyses.

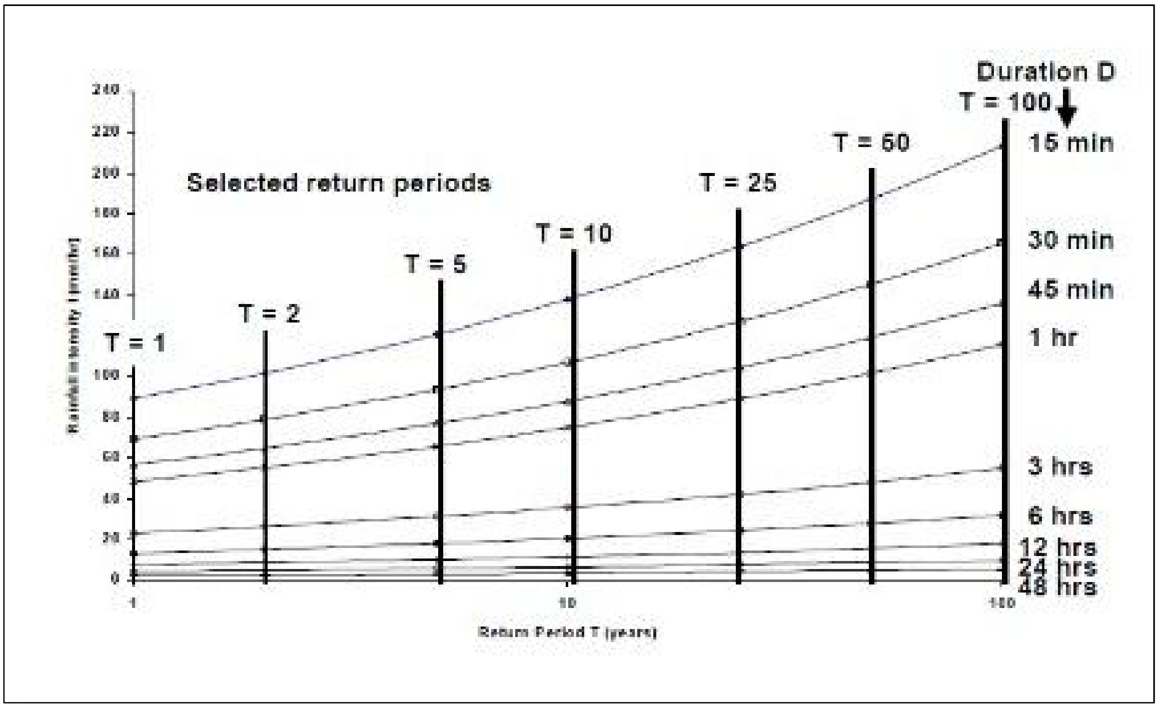


Figure 7.5: Intensity-duration -frequency curves

The intensity-duration curves can be created by reading each duration at distinct return periods. For this the rainfall intensities for various durations at concurrent return periods are connected as shown in Figure 7.6.

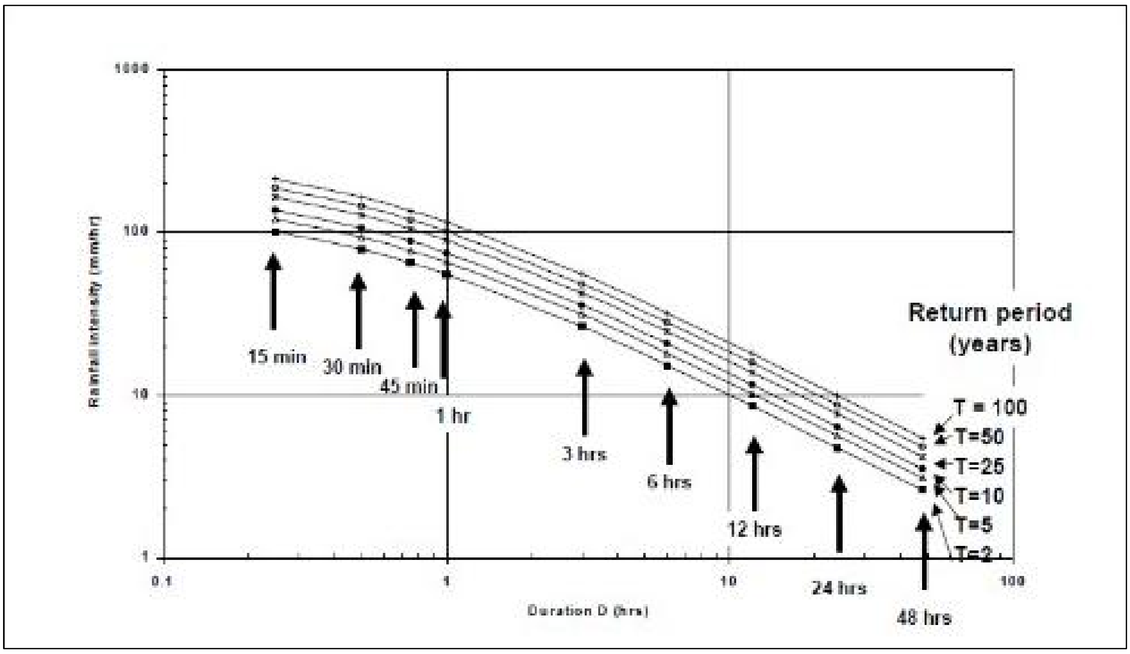


Figure 7.6: Intensity-duration -frequency curves for various return periods

The maximum intensity of rainfall for any duration and for any return period can be read from the curves in Figure 7.6.

7.7.2 Various empirical equations related to intensity duration

After analysing rainfall characteristics for 42 stations in India, Rambabu et al. (1979) presented IDF equation and nomographs in his publication. The general form of the formula to estimate rainfall intensity is known as Rambabu at Vasad & Kota. The general form of the equation is given as:

$$I = \frac{AT^x}{(t_r+B)^n} \qquad \text{Eqn. 7.6}$$

Where:

I = Intensity of rain in cm/hr

T = Return period in years

t_r = Duration of Rain in hrs

A, B, n, x = constants (refer to Table 7.5)

Table 7.5: Value of A, B, n, x at various zones

Zones	Location	A	x	B	n
Northern Zone	Agra	4.9100	0.1667	0.2500	0.6293
	Allahabad	8.5700	0.1692	0.5000	1.0190
	Amritsar	14.4100	0.1304	1.4000	1.2963
	Dehradun	6.0000	0.2200	0.5000	0.8000
	Jaipur	6.2190	0.1026	0.5000	0.8000
	Jodhpur	4.0980	0.1677	0.5000	1.1120
	Lucknow	6.0740	0.1813	0.5000	1.0031
	New Delhi	5.2080	0.1574	0.5000	1.1072
	Sri Nagar	1.5303	0.2730	0.2500	1.0636
	Northern Zone	5.9140	0.1623	0.5000	1.0127
Central Zone	Bagra-Tawa	8.5704	0.2214	1.2500	0.9331
	Bhopal	6.9296	0.1892	0.5000	0.8767
	Indore	6.9280	0.1394	0.5000	1.0651
	Jabalpur	11.379	0.1746	1.2500	1.1206
	Jagdalpur	4.7065	0.1084	0.2500	0.9902
	Nagpur	11.4500	0.1560	1.2500	1.0324
	Punase	4.7011	0.2608	0.5000	0.8656
	Raipur	4.6830	0.1389	0.1500	0.9284
	Thikrl	6.0880	0.1747	1.0000	0.8547
	Central Zone	7.4645	0.1712	0.7500	0.9599
Western Zone	Aurangabad	6.0810	0.1459	0.5000	1.0923
	Bhuj	3.8230	0.1919	0.2500	0.9902
	Mahabaleshwar	3.4830	0.1267	0.0000	0.4853
	Nandurbar	4.2540	0.2070	0.2500	0.7704
	Vengurla	6.8630	0.1670	0.7500	0.8683
	Veraval	7.7870	0.2087	0.5000	0.8908
	Western Zone	3.9740	0.1647	0.1500	0.7327
Eastern Zone	Agarthala	8.0970	0.1177	0.5000	0.8191
	Dumdum	5.9400	0.1150	0.1500	0.9241
	Gauhati	7.2060	0.1557	0.7500	0.9401
	Gaya	7.1760	0.1483	0.5000	0.9459
	Imphal	4.9390	0.1340	0.5000	0.9719
	Jamshedpur	6.9300	0.1307	0.5000	0.8737
	Jharsuguda	8.5980	0.1392	0.7500	0.8740
	North Lakhimpur	14.0700	0.1256	1.2500	1.0730
	Sagarisland	16.5240	0.1402	1.5000	0.9635
	Shillong	6.7280	0.1502	0.7500	0.9575
	Eastern Zone	6.9330	0.1353	0.5000	0.8801
Southern Zone	Banglorw	6.2750	0.1262	0.5000	1.1280
	Hyderabad	5.250	0.1354	0.5000	0.0295
	Kodaikanal	5.9140	0.1711	0.5000	1.0088
	Madras	6.1260	0.1664	0.5000	0.8027

Zones	Location	A	x	B	n
	Mangalore	6.7440	0.1395	0.5000	0.9374
	Tiruchirapalli	7.1350	0.1638	0.5000	0.9624
	Trivandrum	6.7620	0.1536	0.5000	0.8158
	Visakhapatnam	6.6460	0.1692	0.5000	0.9963
	Southern Zone	6.3110	0.1523	0.5000	0.9495

Rambabu et al. (1979) also gives monograph explaining how to convert one hour rainfall intensity into rainfall intensities of other durations.

For the locations Vasad and Kota, coefficients obtained by Central Water Conservation Research and Training Institute are as follows:

Location	A	x	B	n
Vasad	7.5060	0.1393	0.5000	0.3857
Kota	5.7900	0.2300	0.5000	0.8500

Another alternative is the Raudkivi (1979) equation:

$$I_t^T = \frac{C}{10}\frac{T^{0.20}}{t^{0.71}}\left(R_{24}{}^2\right)^{0.33} \qquad \text{Eqn. 7.7}$$

Where:

I_t^T = Rainfall intensity in cm/hr for T year returns period and T hour duration

$R_{24}{}^2$=24-hour 2-year return period rainfall in mm

T = Return period in years

t = Rainfall duration in hours

C = coefficient whose values for different regions in India are:

Geographical Region	Value of C
Northern India	8.00
Eastern India	9.10
Central India	7.70
Western India	8.30
Southern India	7.10

7.7.3 Developing intensity-duration-frequency curve from data set

For developing intensity-duration-frequency curves the following steps are required:

- Rainfalls for various time intervals like 5 min,10 min…600 min are used as input.
- Intensity duration frequency curve is generated using the Extreme Value-1 (EV-1) Distribution.
- Parameter estimation for EV–1 (Method of Moments)

  α=0.7797 · Standard deviation

  μ=Mean-0.45005 · Standard deviation

- Rainfall amount at T year return period or different frequency is given by

  Rainfall at T year= $\mu + \alpha(-\ln(-\ln(1-1/T)))$ Eqn. 7.8

- Rainfall intensity for T year return period is estimated as rainfall at T year return period divided by duration of rain (T_c).

Rainfall intensity estimation using Flood Estimation Report for various subzones in India is applicable only for 25, 50 and 100 year Return Period of 24-hour rainfalls. Rainfall values of return period (25, 50 and 100 year) are obtained from Isopluvial maps which are prepared by India Meteorological Department (IMD).

Rainfall total for t hours for a particular sub-catchment (where t may equal to the time of concentration Tc) may be obtained by multiplying 24-hour point rainfall with conversion factor corresponding to t hours for the sub-catchment (reference should be taken from "Flood Estimation Reports "Jointly published by Central Water Commission, Indian Meteorological Department and Ministry of Surface Transport. There are all together 26 reports for different Sub basin of India) The factors for the different sub-catchments are available in the PMP Atlas published by the CWC and freely downloadable from their website.

7.7.4 Annual maximum and annual exceedance series

The annual maximum series of rainfall intensities were considered in the procedure presented above. Distinction is to be made between the annual maximum and annual exceedance series in the application of frequency analyses. The latter is derived from partial duration series, which is defined as series of data above a threshold. The maximum values between each upstream crossing and the next downstream crossing (see Figure 7.7) are considered in the partial duration series. The threshold should be taken high enough to make successive maximums serially independent or a time horizon is to be considered around the local maximum to eliminate lower maximums exceeding the threshold but which are within the time horizon. If the threshold is taken such that the number of values in the partial duration series becomes equal to the number of years selected then the partial duration series is called annual exceedance series.

Since the annual maximum series consider only the maximum value each year, it may happen that the annual maximum in a year is less than the second or even third largest independent maximum in another year. Hence, the values at the lower end of the annual exceedance series will be higher than those of the annual maximum series. Consequently, the return period derived for a particular I(D) based on annual maximum series will be larger than one would have obtained from annual exceedances. The following relation exists between the return period based on annual maximum and annual exceedance series (Chow, 1964):

$$T_E = \frac{1}{ln(\frac{T}{T-1})} \qquad \text{Eqn. 7.9}$$

Where:

T_E=Return period for annual exceedance series

T=Return period for annual maximum series

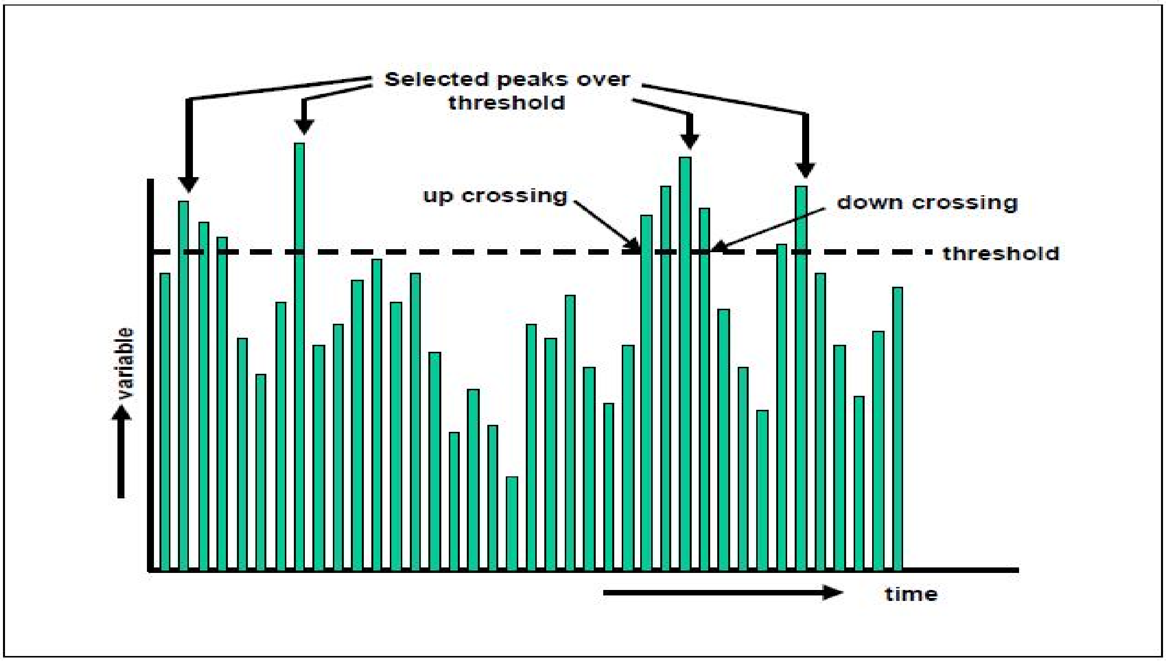


Figure 7.7: Definition of partial duration series

It is observed that the ratio (T_E/T) approaches 1 for large T. Generally, when T < 20 years, T has to be adjusted to TE for design purposes. Particularly for urban drainage design, where low return periods are used, this correction is of importance.

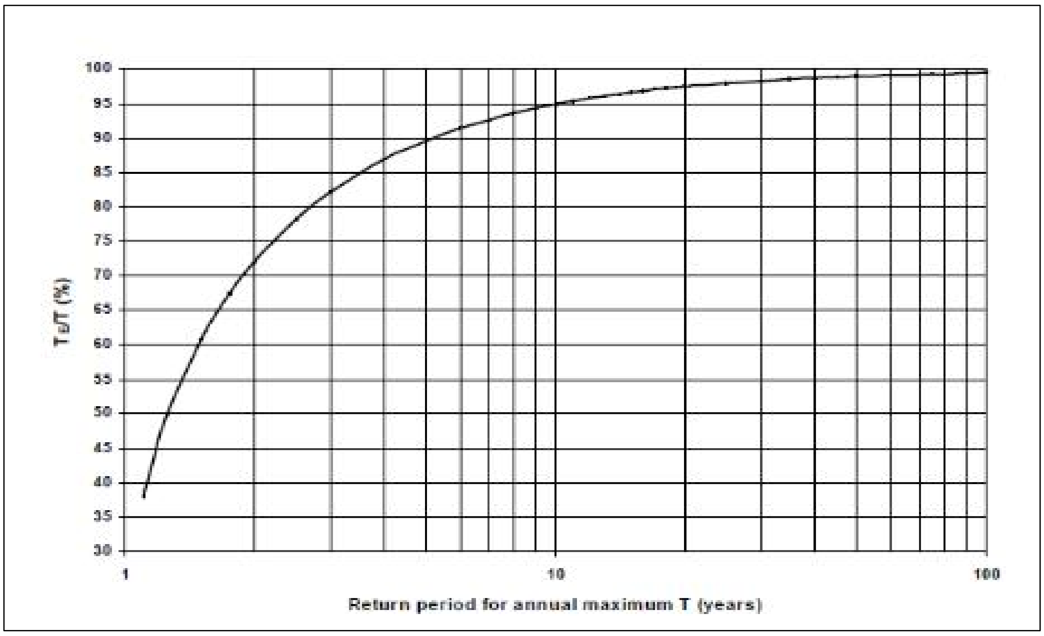


Figure 7.8: Relation between return periods annual maximum (T) and annual exceedance series (T_E)

Generally annual maximum series are used in the development of intensity-duration-frequency curves, which are fitted by a Gumbel distribution. Equation 6.3 is used to transform T into T_E for T < 20 years. Results can either be presented for distinct values of T or of T

Example 7-3

Rainfall data for the 1908-1911 monsoon season period has been summarized for various durations of rainfall in Table 7.6 below:

Table 7.6: Rainfall data

Time in min	5	10	15	30	60	90	120
Year	Rainfall in cm						
1908	0.85	1.20	1.40	1.74	2.15	2.46	2.97
1921	0.76	1.04	1.18	1.55	1.92	2.38	2.63
1915	0.73	0.93	1.11	1.36	1.70	2.14	2.34
1934	0.72	0.88	1.03	1.22	1.45	1.81	2.12
1929	0.66	0.84	0.97	1.18	1.40	1.65	1.83
1926	0.62	0.80	0.92	1.10	1.33	1.50	1.64
1931	0.51	0.78	0.90	1.05	1.25	1.40	1.55
1904	0.45	0.68	0.82	1.01	1.20	1.36	1.51
1917	0.36	0.52	0.67	0.95	1.14	1.34	1.46
1914	0.28	0.51	0.62	0.83	1.11	1.27	1.41
1911	0.21	0.39	0.5	0.79	1.09	1.23	1.34

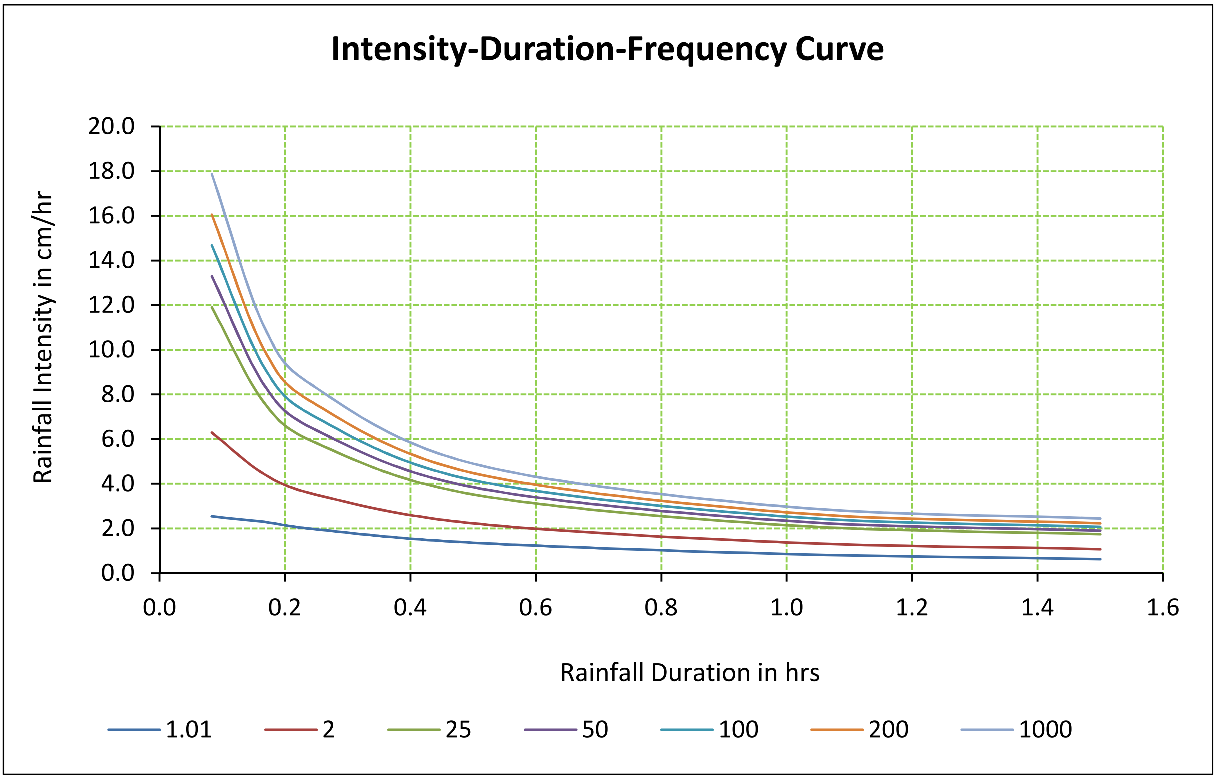


Figure 7.9: Fitting of Gumbel distribution to observed frequency distribution of hourly annual maximum series for monsoon season

Table 7.7: Analysis of data with Gumbel method

Time in min	5	10	15	30	60	90	120
Mean	0.56	0.78	0.92	1.16	1.43	1.69	1.89
Standard Deviation	0.21	0.24	0.26	0.29	0.35	0.45	0.55
Parameters (Gumbel Distribution)							
ά	0.2	0.2	0.2	0.2	0.3	0.4	0.4
ǔ	0.5	0.7	0.8	1.0	1.3	1.5	1.6

Table 7.8: Rainfall intensity-duration-frequency

Return Period T	Time in hrs					
	0.08	0.17	0.25	0.50	1.00	1.50
	Rainfall Intensity in cm/h					
1.01	2.54	2.29	1.96	1.36	0.85	0.63
1.05	3.36	2.76	2.30	1.55	0.97	0.73
1.11	3.91	3.08	2.52	1.68	1.04	0.79
1.25	4.62	3.48	2.82	1.84	1.14	0.88
1.5	5.38	3.92	3.13	2.02	1.25	0.97
2	6.29	4.44	3.51	2.23	1.37	1.07
5	8.54	5.72	4.43	2.75	1.68	1.34
10	10.02	6.56	5.05	3.09	1.89	1.52
20	11.45	7.38	5.64	3.42	2.09	1.68
25	11.90	7.64	5.82	3.52	2.15	1.74
50	13.29	8.43	6.40	3.84	2.34	1.90
100	14.67	9.22	6.97	4.16	2.53	2.07
125	15.12	9.47	7.15	4.27	2.59	2.12
200	16.05	10.00	7.54	4.48	2.72	2.23
500	17.87	11.04	8.29	4.90	2.97	2.44
1000	19.24	11.83	8.86	5.22	3.16	2.60

7.8 Depth-area-duration analysis

7.8.1 General

A storm of a given duration over a certain area rarely produces uniform rainfall depth over the entire area. The storm usually has a centre, where the rainfall P_o is maximum which is always larger than the average depth of rainfall P for the area as a whole. Generally, the difference between these two values, that is (P_o – P), increases with increase in area and decreases with increase in the duration. Also, the difference is more for convective and orographic precipitation than for cyclonic. To develop quantitative relationship between P_o and P, a number of storms with data obtained from recording rain gauges have to be analysed. The analysis of a typical storm is described below.

Rainfall data is plotted on the basin map and the isohyets are drawn by using a GIS tool. These isohyets divide the area into various zones. On the same map the Thiessen polygons are also constructed for all the rain gauge stations. The polygon of a rain gauge station may lie in different zones. Thus, each zone will be influenced by a certain number of gauges, whose polygonal areas lie either fully or partially in that zone. The gauges, which influence each zone along with their influencing areas, are noted. Next for each zone the cumulative average depth of rainfall (areal average) is computed at various time using the data of rainfall mass curve at the gauges influencing the zone and the Thiessen weighted mean method. In other words, the cumulative depths of rainfall at different times recorded at different parts are converted into cumulative depths of rainfall for the zonal area at the corresponding times. The mass curves of average depth of rainfall for accumulated areas are then computed starting from the zone nearest to the storm centre and by adding one more adjacent to it each time, using the results obtained in the previous step and applying Thiessen weight in proportion to the area of the zones. These mass curves are now examined to find the maximum average depth of rainfall for different duration and for progressively increasing accumulated area. The results are then plotted on semi-logarithmic paper. That is, for each duration the maximum average depth of rainfall on an ordinary scale is plotted against the area on logarithmic scale. If a storm contains more than one storm centre, the above analysis is carried out for each storm centre. An enveloping curve is drawn for each duration. Alternatively, for each duration a depth area relation of the form as proposed by Horton may be established:

$$\boldsymbol{P = P_0 e^{-kA^n}} \qquad \text{Eqn. 7.10}$$

Where

P_0= Highest amount of rainfall at the centre of the storm (A=25 km^2) for any given duration

P=Maximum average depth of rainfall over an area A (>25 Km^2) for the same duration

A=area considered for P

K, n=regression coefficients, which vary with storm duration and region

Example 7-4

The following numerical example illustrates the method described above. There are 7 Rain gauges in a catchment area of 2790 km^2 Rain Gauge as shown in Figure 7.10. The record of a severe storm measured in the catchment as observed at the 7 rain gauge stations is presented in Table 7.9 below.

Table 7.9: Cumulative rainfall for a severe storm at 7 rain gauges (A to G)

Time in hours	Cumulative rainfall in mm measured at rain gauge stations						
	A	B	C	D	E	F	G
4	0	0	0	0	0	0	0
6	12	0	0	0	0	0	0
8	18	15	0	0	0	6	0
10	27	24	0	0	9	15	6
12	36	36	18	6	24	24	9
14	42	45	36	18	36	33	15
16	51	51	51	36	45	36	18

Time in hours	Cumulative rainfall in mm measured at rain gauge stations						
18	51	63	66	51	60	39	18
20	51	72	87	66	66	42	18
22	51	72	96	81	66	42	18
24	51	72	96	81	66	42	18

The total rainfall of 51, 72, 96, 81, 66, 42 and 18 mm are indicated at the respective rain gauge stations A, B, C, D, E, F and G on the map. The isohyets for the values 30, 45, 60 and 75 mm are constructed. Those isohyets divide the basin area into five zones with areas as given in Table 7.10. The Thiessen polygons are then constructed for the given rain gauge network [A to G] on the same map. The areas enclosed by each polygon and the zonal boundaries for each rain gauge is also shown in Table 7.10.

Table 7.10: Zonal areas and influencing area by rain gauges

Zone	Area in km^2	Rain gauge Station area of influence in each zone (km^2)						
		A	B	C	D	E	F	G
I	415	0	105	57	253	0	0	0
II	640	37	283	0	20	300	0	0
III	1015	640	20	0	0	185	170	0
IV	525	202	0	0	0	0	275	48
V	195	0	0	0	0	0	37	158

As can be seen from Figure 7.10 Zone I (affected by the rainfall stations with the highest point rainfall amounts) is the nearest to storm centre while Zone V is the farthest.

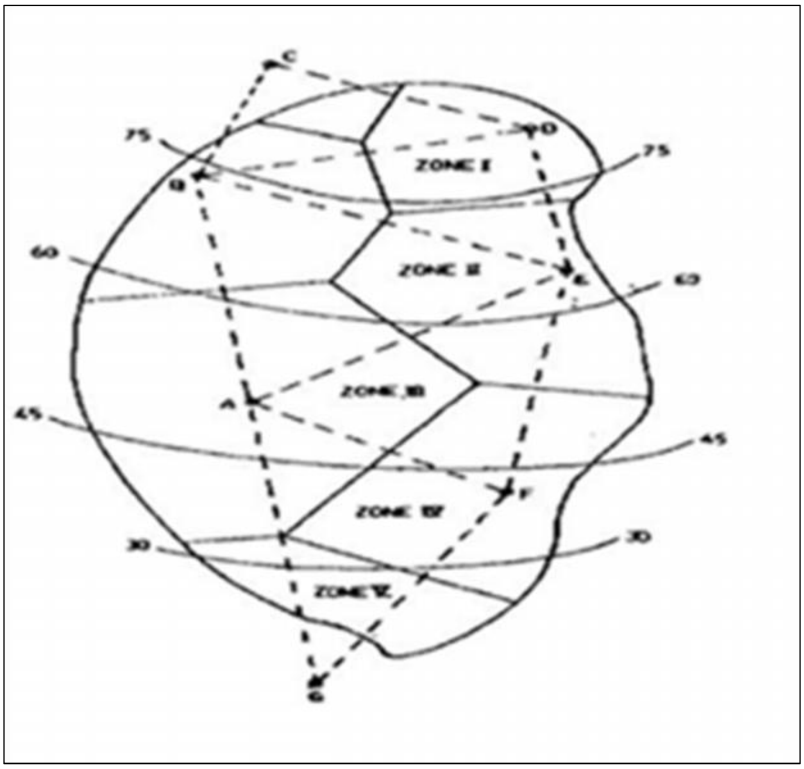


Figure 7.10: Depth-area-duration analysis

The cumulative average depth of rainfall for each zone is then computed using the data at Rain gauge stations A, B, C, D, E, F and G and the corresponding Thiessen weights. For example, the average depth of rainfall in Zone I at any time, P_I is computed from the following equation.

$$P_I = \frac{105\times P_B\times 57\times P_C+253\times P_D}{105+57+253} \qquad \text{Eqn. 7.11}$$

where P_B, P_C and P_D are the cumulative rainfalls at stations B, C and D at any given time. That is

$$P_I = 0.253\ P_B + 0.137\ P_C + 0.610\ P_D$$

Similarly, for Zone II, we have:

$$P_{II} = \frac{37\times P_A\times 283\times P_B+20\times P_D+300\times P_E}{37+283+20+300} \qquad \text{Eqn. 7.12}$$

Or:

$$P_{II}=0.058\ P_A+0.442\ P_B+0.031\ P_D+0.469\ P_E$$

and so on. These results are shown in Table 7.11. The calculation steps are shown in Table 7.12 and in Figure 7.11.

Table 7.11: Cumulative average depths of rainfall in various zones in mm

Time (Hrs)/ Area	Zone I	Zone 1+ II	Zone I+II+ III	Zone I+II+III+ IV	Zone I+II+III+IV+V
4	28.79	23.92	18.42	18.17	17.64
8	53.48	43.13	36.35	35.08	34.04
12	74.71	60.84	50.97	48.16	46.33
16	80.78	72.74	59.18	56.59	54.21
20	80.78	73.17	63.12	59.11	56.55

Table 7.12: Cumulative average rainfalls for accumulated areas in mm

Time (Hrs)/ Area (km^2)	Zone I	Zone I+ II	Zone I+II+ III	Zone I+II+III+ IV	Zone I+II+III+IV+V
	415	1055	2070	2595	2790
4	0.00	0.00	0.00	0.00	0.00
6	0.00	0.43	3.92	4.06	3.78
8	3.80	6.17	9.34	9.48	8.90
10	6.07	12.40	16.90	17.28	16.62
12	15.24	23.95	27.75	27.65	26.54
14	27.30	34.95	37.12	36.65	35.38
16	41.86	45.45	46.38	45.12	43.46
18	56.10	58.81	54.90	52.23	50.12
20	70.40	68.80	60.88	57.32	54.89
22	80.78	73.10	63.11	59.10	56.54
24	80.78	73.10	63.11	59.10	56.54

Calculation steps

Time (Hrs)/ Area	Zone I	Zone 1+ II	Zone I+II+ III	Zone I+II+III+ IV	Zone I+II+III+IV+V
4	28.79	23.92	18.42	18.17	17.64
8	53.48	43.13	36.35	35.08	34.04
12	74.71	60.84	50.97	48.16	46.33
16	80.78	72.74	59.18	56.59	54.21
20	80.78	73.17	63.12	59.11	56.55

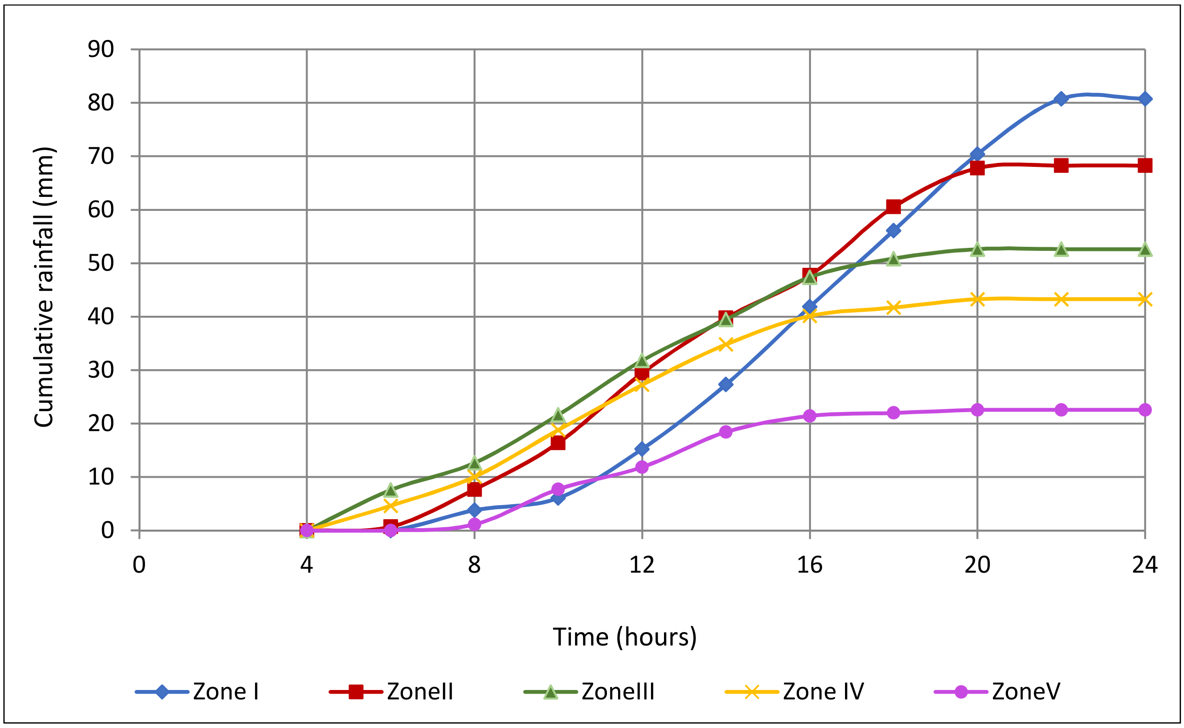


Figure 7.11: Cumulative average depths of rainfall in zones I to V

Calculation steps

Time (Hrs)/ Area	Zone I	Zone 1+ II	Zone I+II+ III	Zone I+II+III+ IV	Zone I+II+III+IV+V
	415.00	1055.00	2070.00	2595.00	2790.00
4	28.79	23.92	18.42	18.17	17.64
8	53.48	43.13	36.35	35.08	34.04
12	74.71	60.84	50.97	48.16	46.33
16	80.78	72.74	59.18	56.59	54.21
20	80.78	73.17	63.12	59.11	56.55

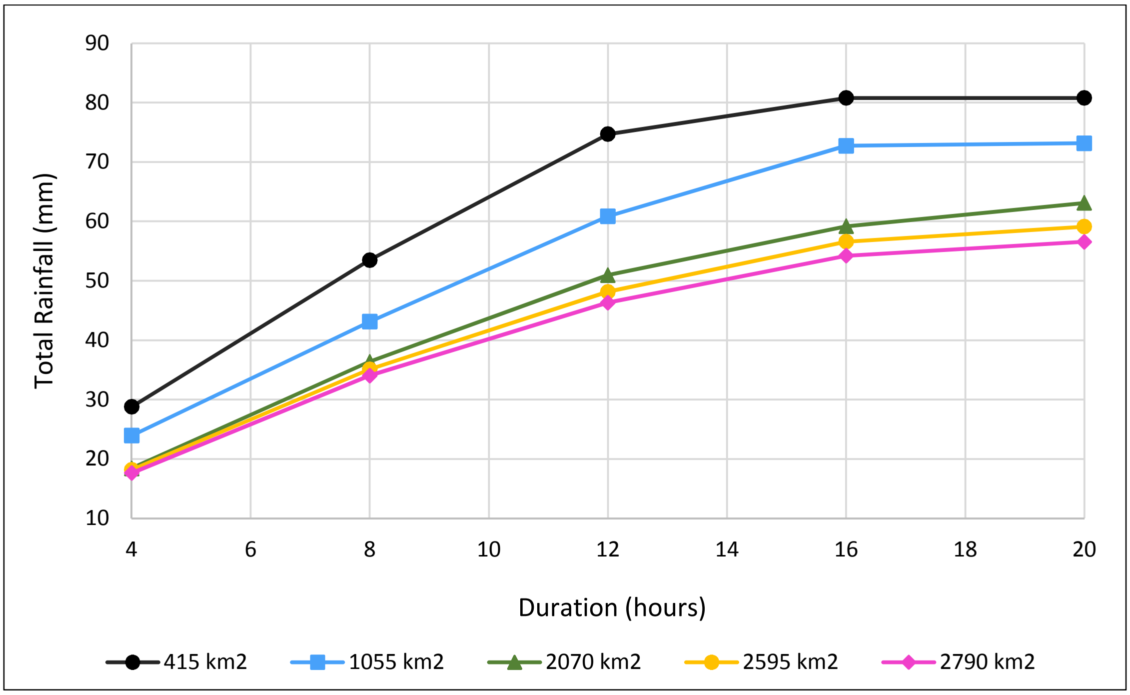


Figure 7.12: Cumulative average depths of rainfall in cumulated areas

For any zone the maximum average depth of rainfall for various durations of 4, 8, 12, 16 and 20 h can be obtained from Table 7.13 finding out maximum rainfall occurred at any 4/8/12/16/20 duration of that zone. The maximum value contained in the window of a particular width is presented in Figure 7.12 and Table 7.13.

Table 7.13: Maximum average depths of rainfall for accumulated areas

Zone	Zone 1	Zone1+ II	Zone I+II+ III	Zone I+II+ III+ IV	Zone I+II+ III+IV+V
4	28.79	23.92	18.42	18.17	17.64
8	53.48	43.13	36.35	35.08	34.04
12	74.71	60.84	50.97	48.16	46.33
16	80.78	72.74	59.18	56.59	54.21
20	80.78	73.17	63.12	59.11	56.55

For each duration, the maximum depths of rainfall are plotted against the area on logarithmic scale as shown in Figure 7.13.

By repeating this procedure for other severe storms and retrieving the maximum rainfall depths per duration for distinct areas from graphs like Figure 7.12, a series of storm rainfall depths per duration and per area is obtained. The maximum value for each series is retained to construct curves similar to those shown in Figure 7.13 (for larger range of areas the X-axis in Figure 7.13 is typically shown using the log scale). Consequently, the maximum rainfall depth for a particular duration as a function of area may be made using different storms to produce the overall maximum observed rainfall depth for a particular duration as a function of area to constitute the depth-area-duration (DAD) curve.

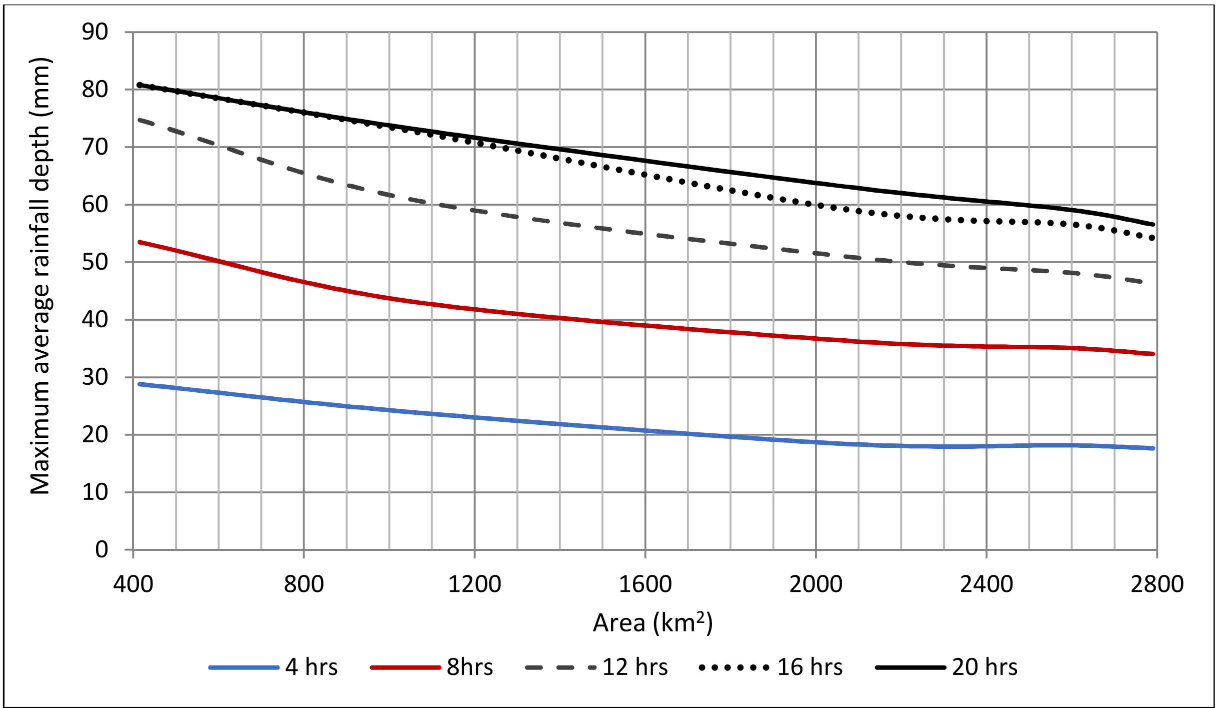


Figure 7.13: Depth -area -duration curves for a particular storm

For the catchment considered in the example these DAD curves will partly or entirely exceed the curves in Figure 7.13 unless the presented storm was the most extreme storm ever recorded in terms of the depth-area relationship.

7.8.2 Development of depth-area-duration (DAD) curves using ArcGIS

Another procedure for the development of DAD curves is by using the rainfall station locations and the historical station data.

The depths of precipitation are plotted on separate suitable base maps of the region showing rain gauge stations, height contours, etc., and isohyet lines are drawn. The area enclosed within the isohyet lines shall be measured by the help of the computer model (ArcGIS) and multiplied by mean isohyet values to find out the rainfall volume. Cumulative rainfall volume is then divided by the cumulative area to compute precipitation depth. Finally, the computed depth values and the corresponding areas are plotted to form the depth duration area curve for each rain storm of various durations.

A typical example of DAD curve generation with ArcGIS software for a sub basin of Godavari from a historical storm data described below.

Example 7-5

Processing gridded rainfall data for Godavari Basin in GIS

The methods generally applied for interpolation of point rainfall data to generate gridded rainfall output include the IDW, Spline, Kriging and the like. The topo-to-raster tool was created for the specific purpose of creation of DEM combining information from point heights, elevation contours and drainage network. The use of topo-to-raster for the purpose of creation of gridded rainfall data is described below, as for this specific data set it was found to yield results that appear to be more acceptable. This is not to recommend the use of this method as a preferable one over the others.

Step 1: Open ARC map and add basin boundary /sub basins and Gridded data with Rainfall value (Figure 7.14).

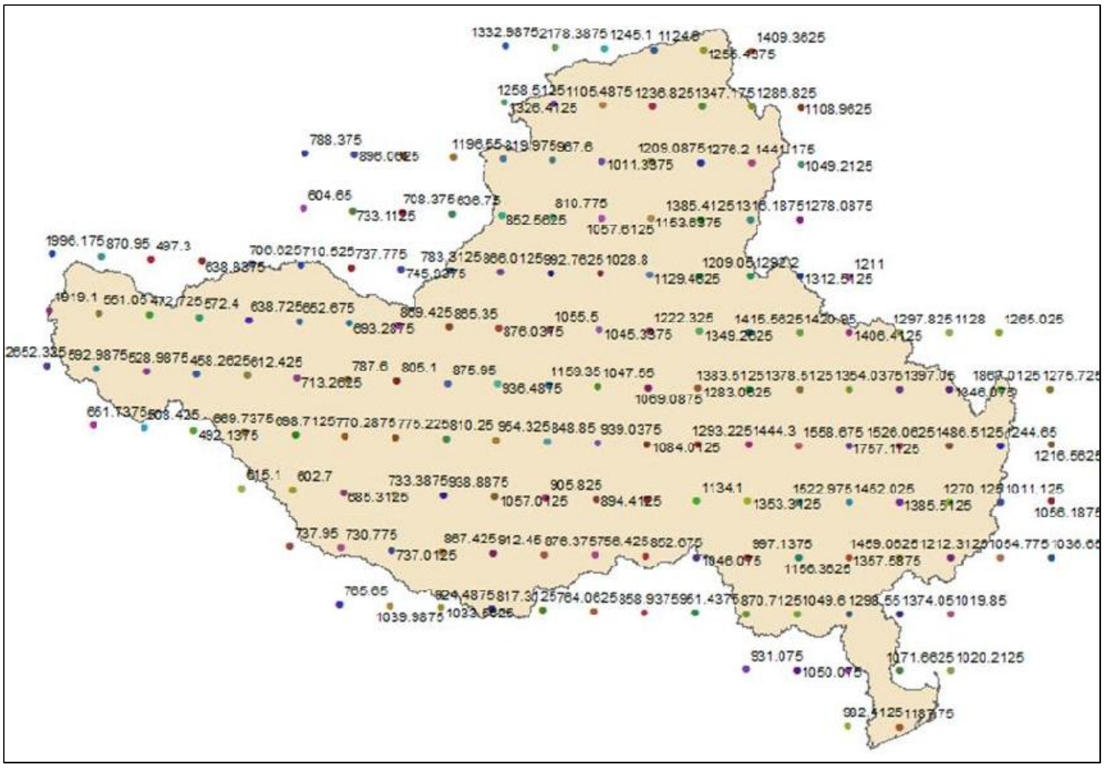


Figure 7.14: Basin with gridded rainfall data

Step 2: Create surface using topo-to-raster tool using required field from station layer (Figure 7.15 and Figure 7.16).

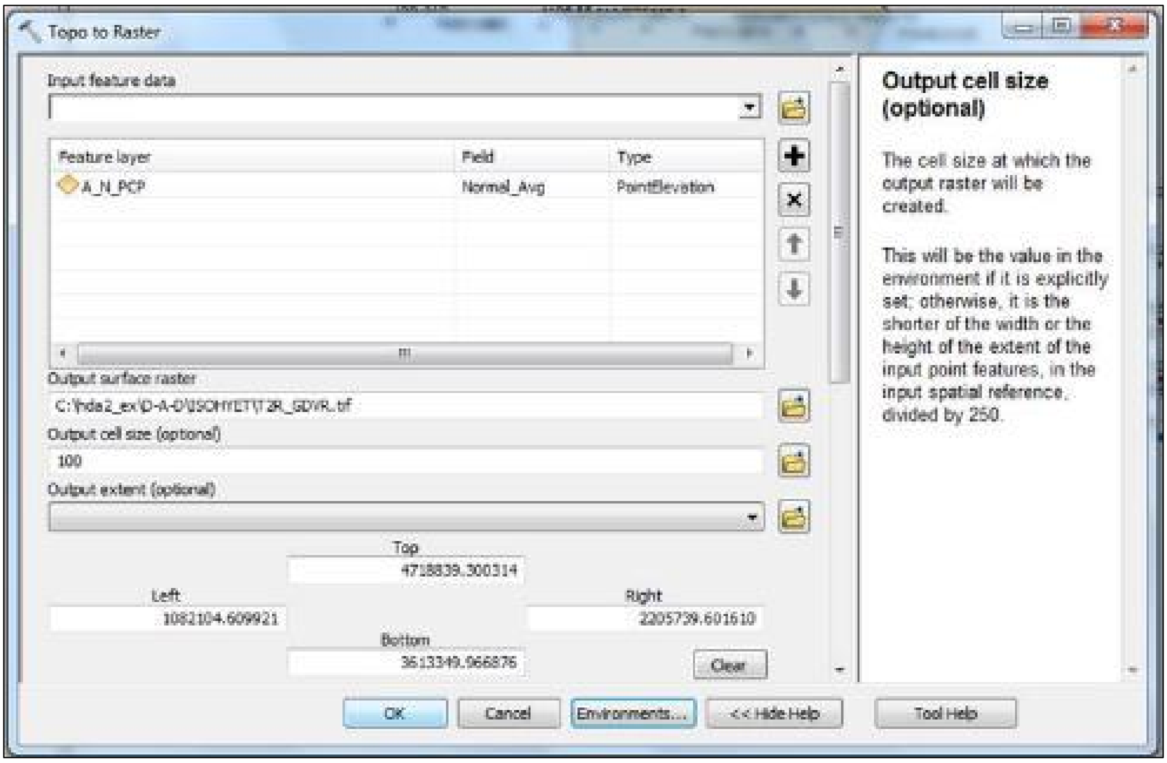


Figure 7.15: Topo to raster tool

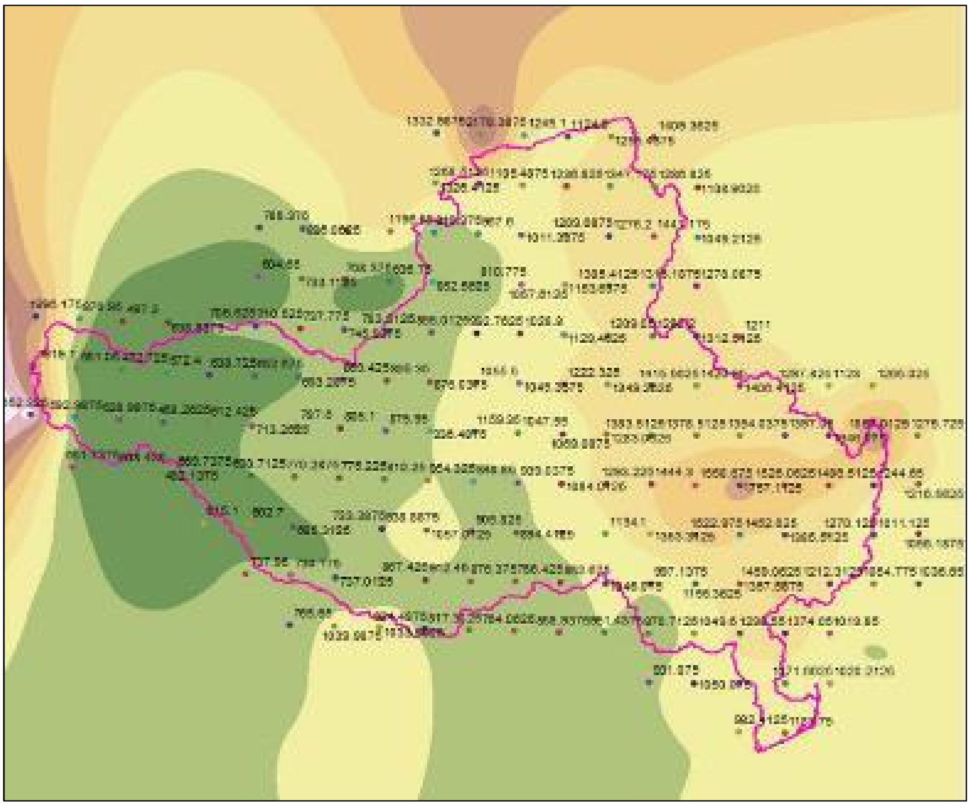
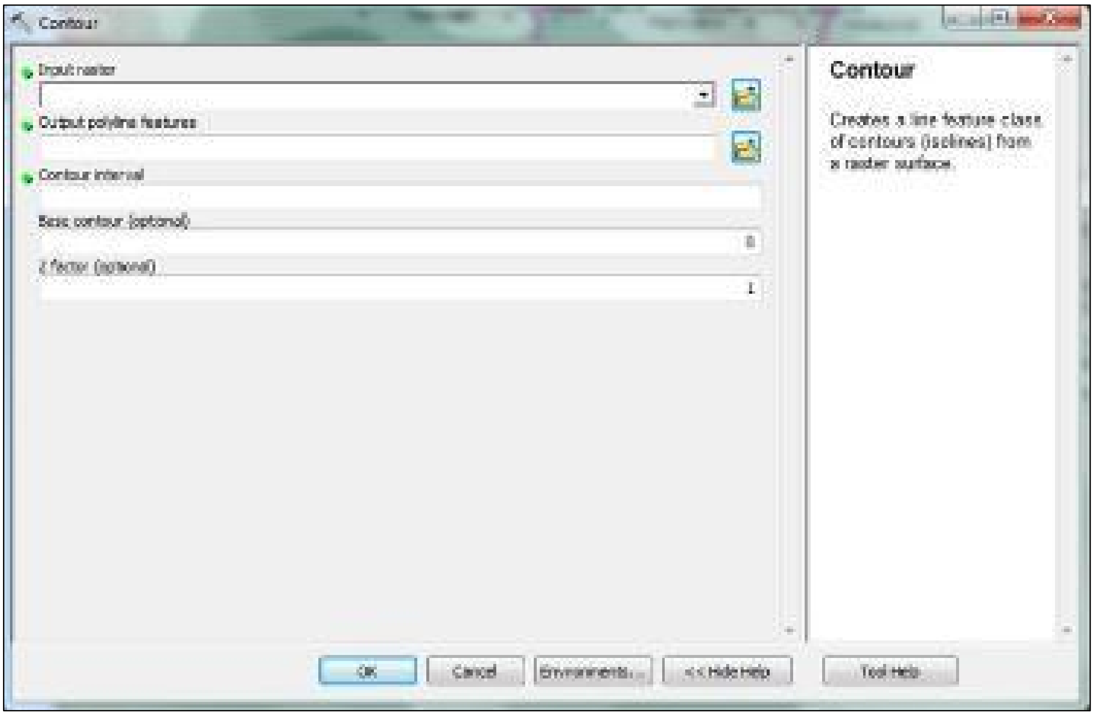


Figure 7.16: Topo to raster tool output

Step 3: Create isohyets using the contour tool with a specified contour interval (Figure 7.17, Figure 7.18, Figure 7.19).



,

Figure 7.17: Contour tool

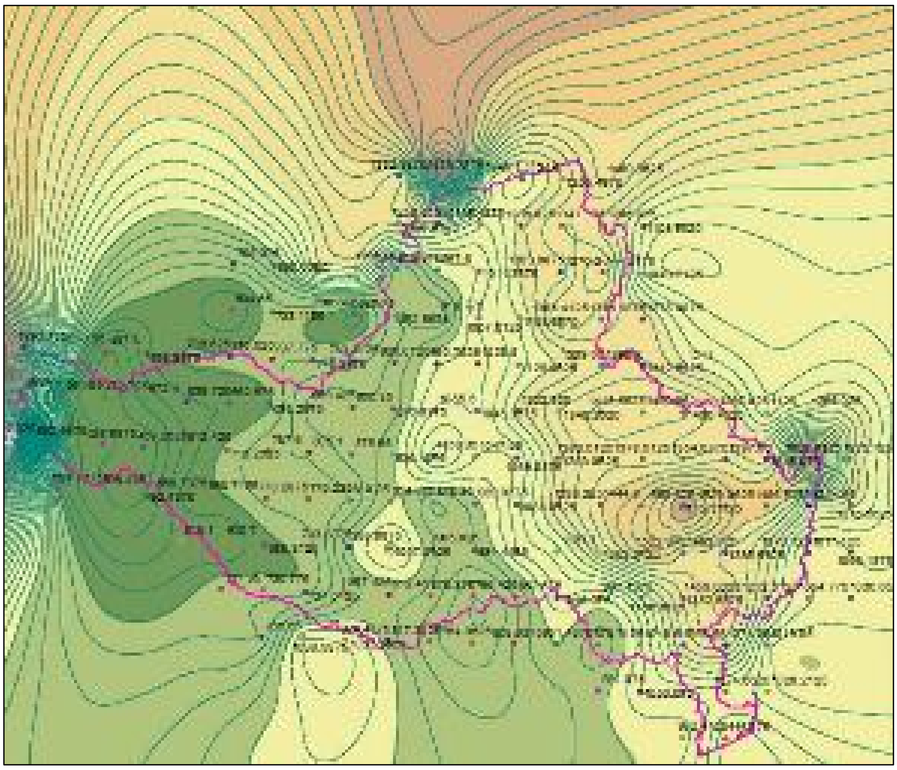


Figure 7.18: Output of contour tool

Step 4: Display levels of the Contour

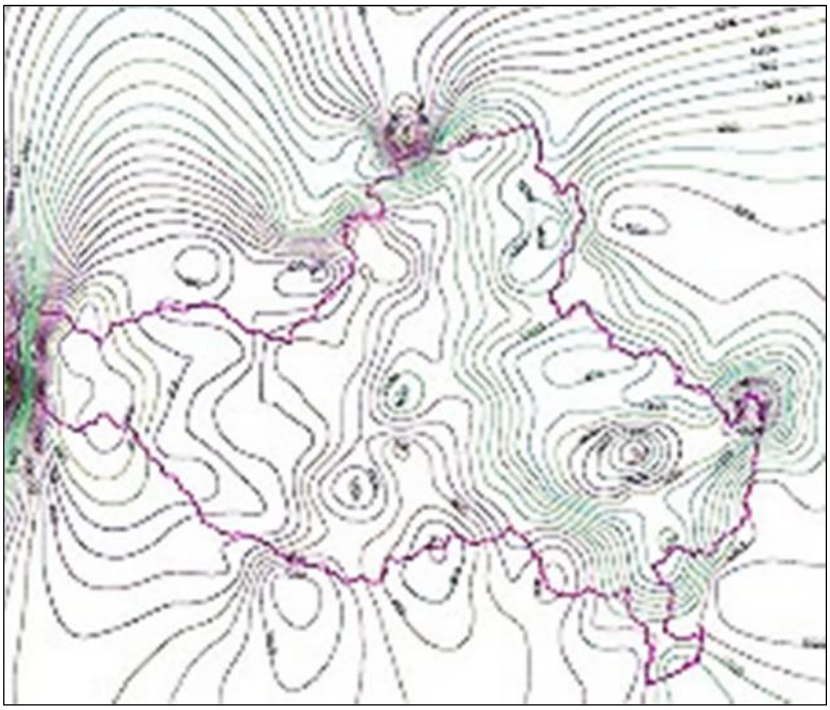


Figure 7.19: Contour with values

Step 5: Clip the isohyets within the sub basin boundary (Figure 7.20, Figure 7.21)

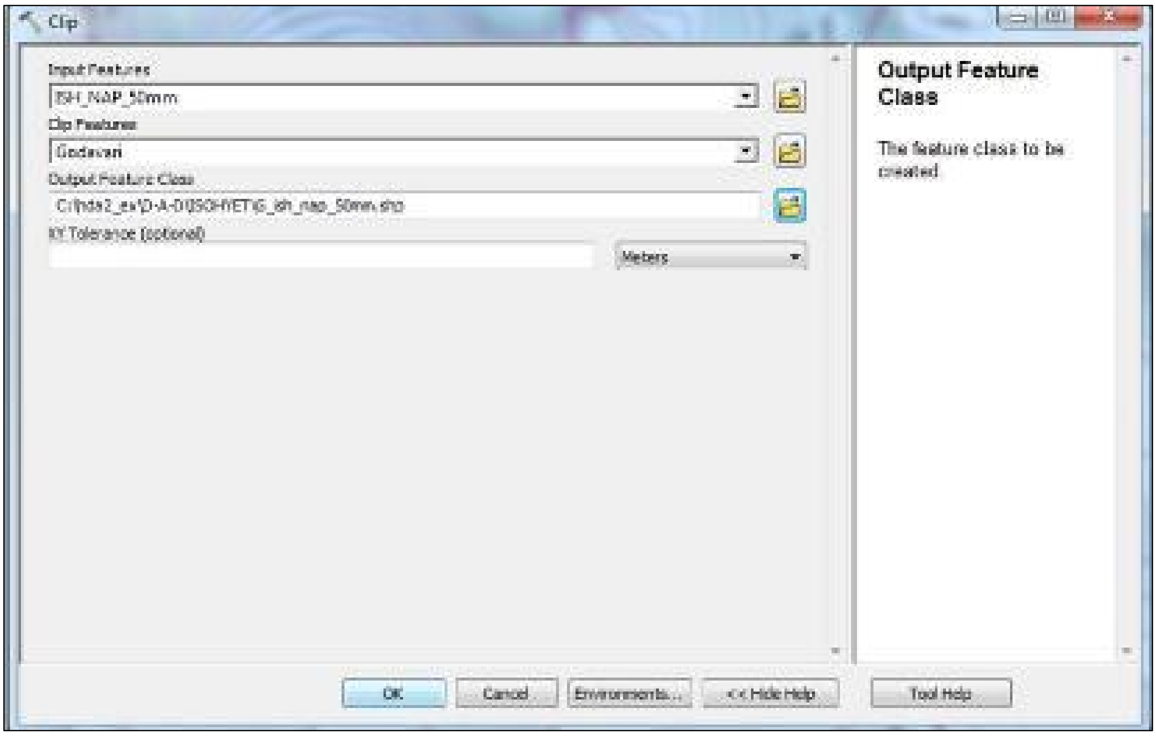


Figure 7.20: Clip tool

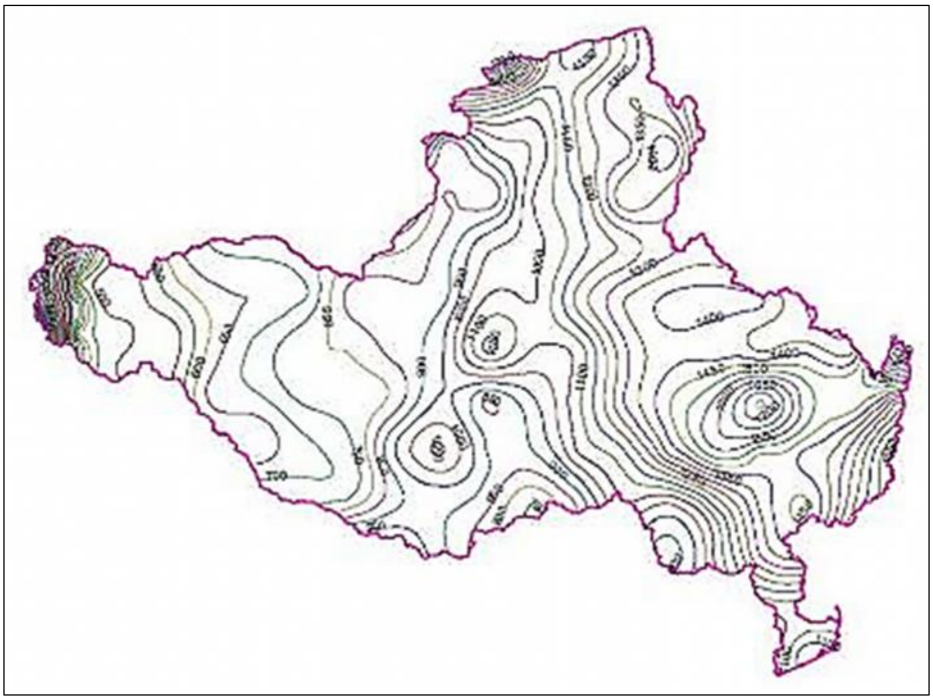


Figure 7.21: Output of clip tool

Repeat the steps with 24 hrs, 48hrs and 72 hrs storms value (if available).

Now further steps of processing of isohyetal area between contours are done and shown for a smaller basin of Godavari catchment (Wardha) with a 24 hrs storm Data (steps are same up to step 5, for 24 hrs data as well). DAD curve is prepared accordingly for a 24 hrs duration storms

Step 6: Convert the catchment boundary to polyline then copy and paste to contour layer. (Figure 7.22)

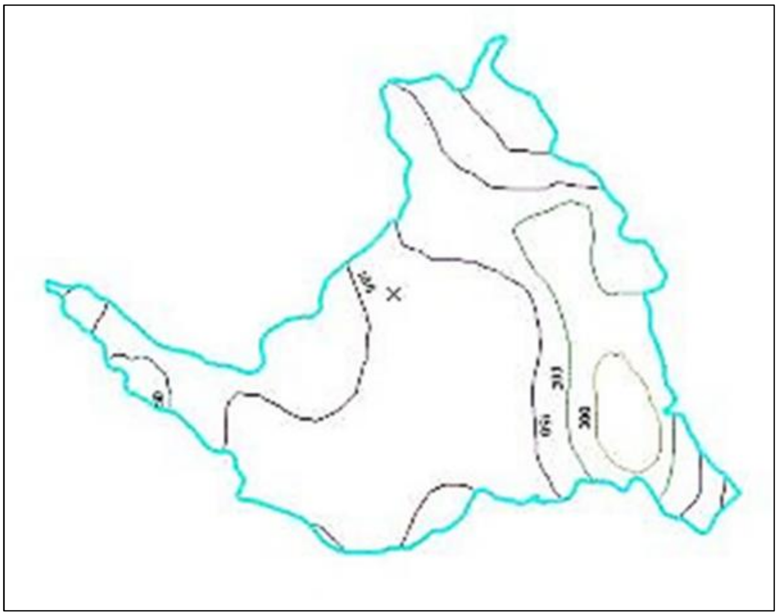


Figure 7.22: Catchment border polyline pasted onto contour

Step 7: Convert Contour lines to polygon (Figure 7.23)

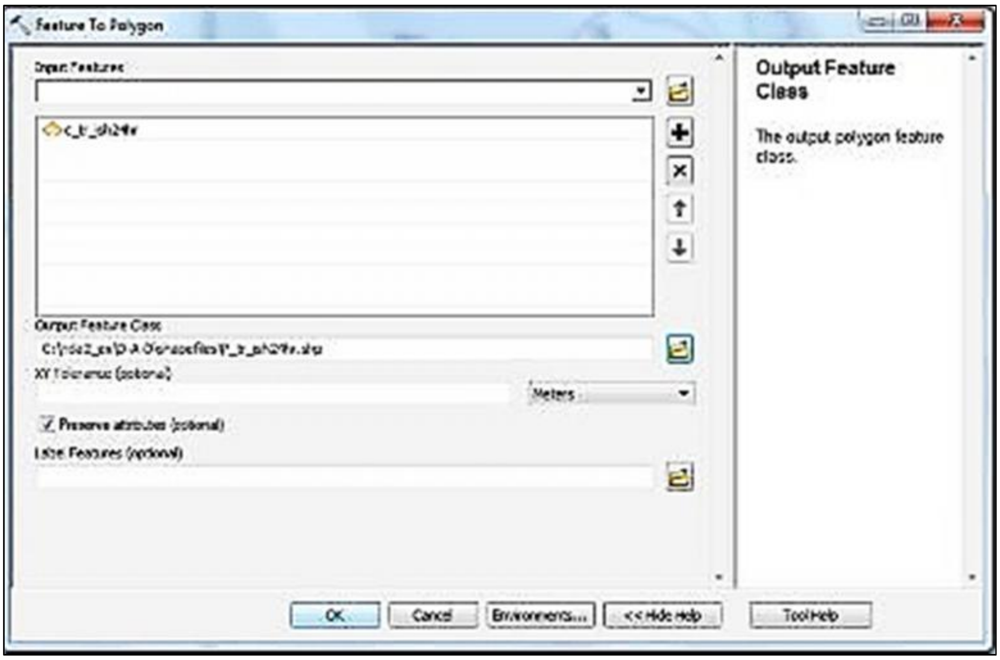


Figure 7.23: Feature to polygon tool

Step 8: Add field named area and calculate geometry in SQ Km (unit) (Figure 7.24)

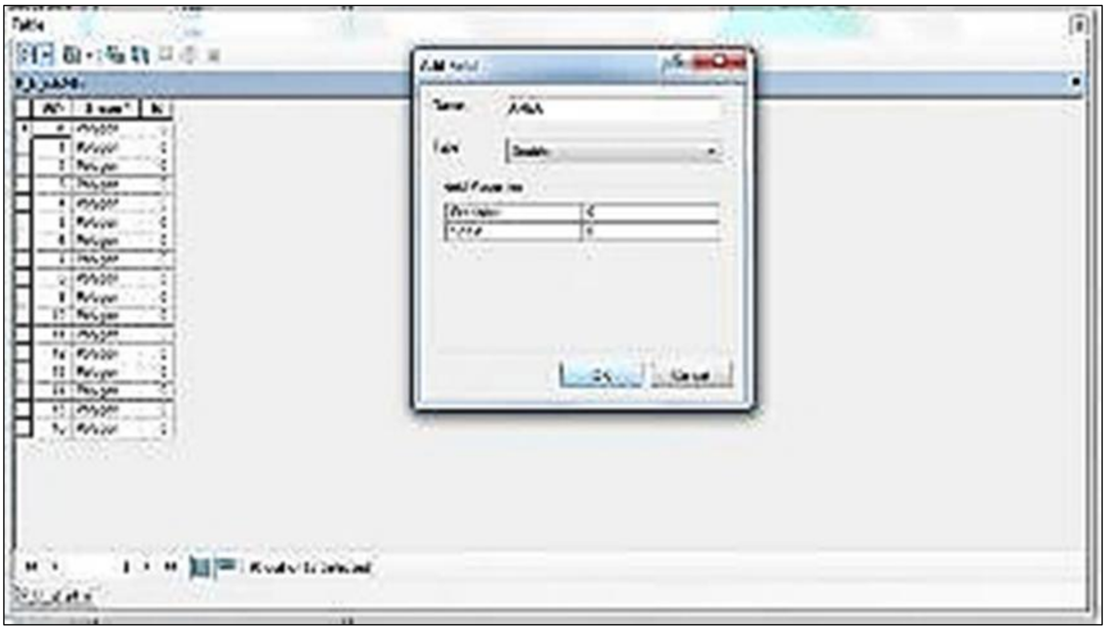


Figure 7.24: Add field tool

Step 9: Find out area between contour intervals (Figure 7.25, Figure 7.26)

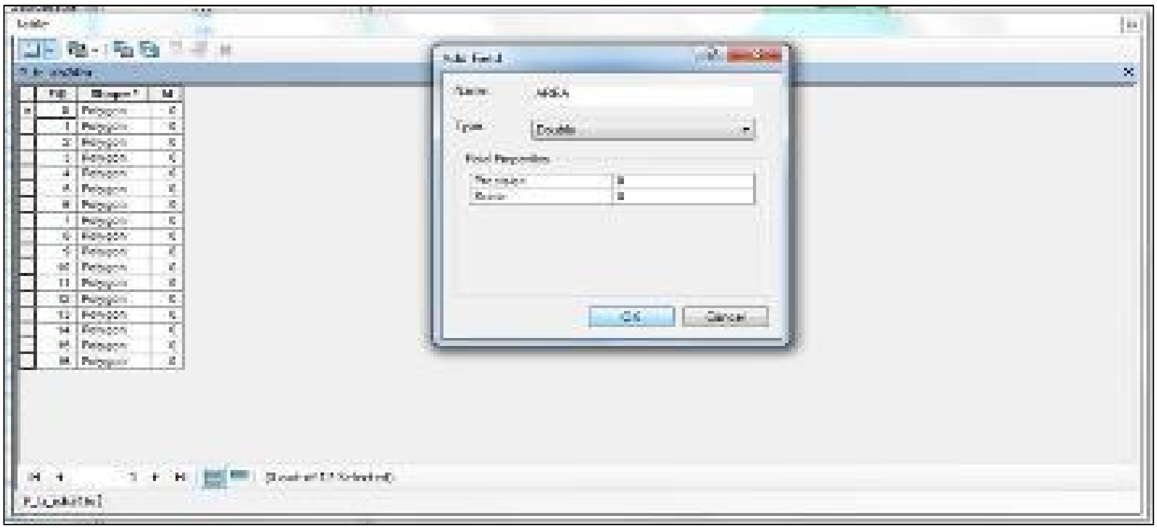


Figure 7.25: Calculate geometry tool

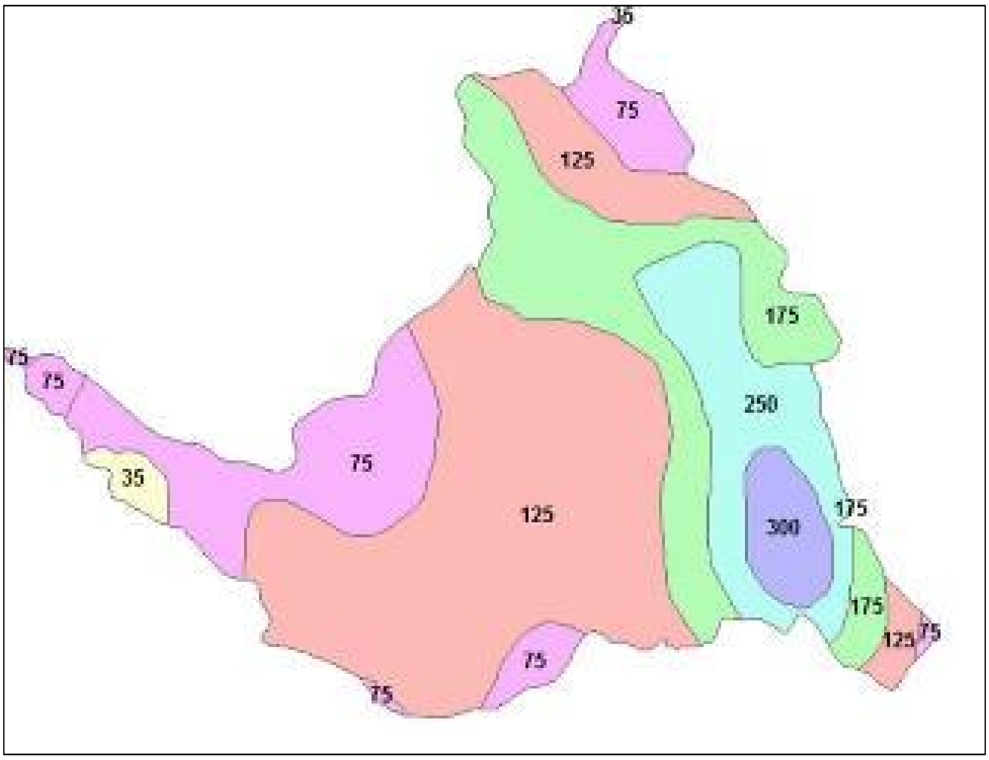


Figure 7.26: Areas under contour

Table 7.14 has been developed following the above procedure.

Table 7.14: Rainfall depth in mm vs area in square km

Average Rainfall depth between isohyets (mm)	Area (km²)	Accumulated Area (km²)	Rainfall volume in (mm-km²)	Accumulated Rainfall volume (mm-km²)	Rainfall Depth (mm)
300	1838.96	1838.96	551688.53	551688.53	300.00
250	5587.19	7426.15	1396796.52	1948485.05	262.38
175	9215.77	16641.92	1612759.41	3561244.46	213.99
125	21659.87	38301.79	2707483.77	6268728.22	163.67
75	8748.62	47050.40	656146.22	6924874.45	147.18
35	596.63	47647.03	20882.05	6945756.50	145.78

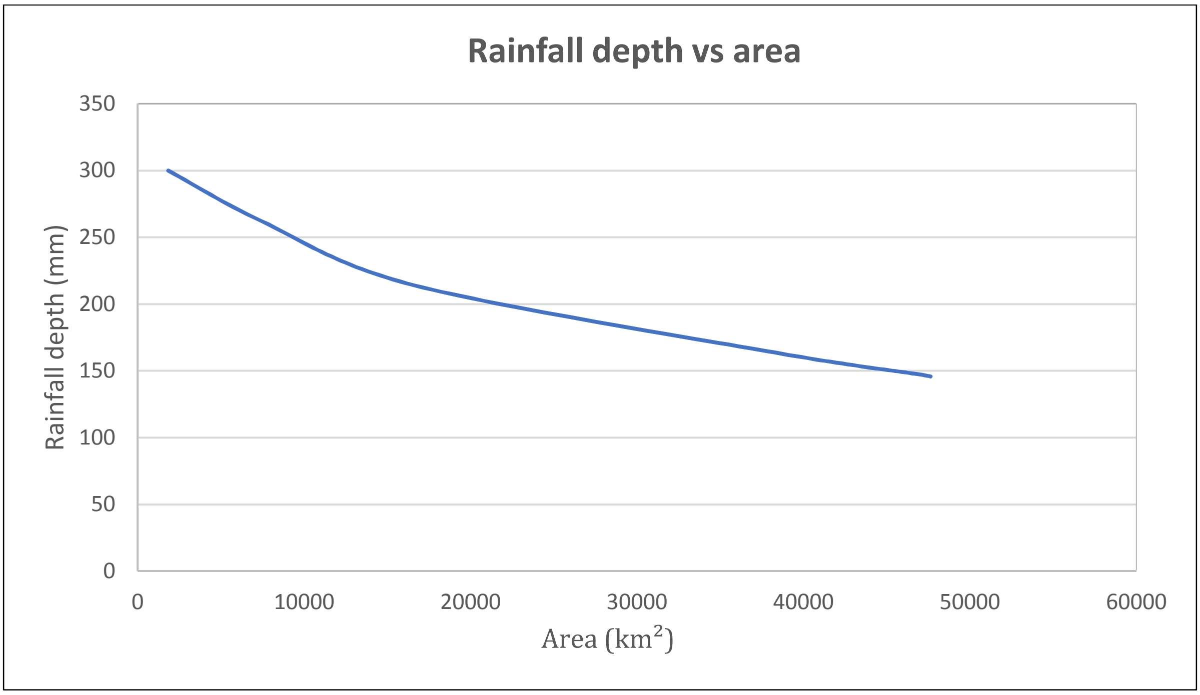


Figure 7.27: Rainfall depth in mm vs area in km²

7.8.3 Areal reduction factor

If the maximum average rainfall depth as a function of area is divided by the maximum point rainfall depth, the ratio is called the Areal Reduction Factor (ARF), which is used to convert point rainfall extremes into areal estimates. ARF-functions are developed for various storm durations. In practice, ARF functions are established based on the average *DAD curves* developed for some selected representative storms.

These ARFs which will vary from region to region are also dependent on the season. Though generally ignored, it would be of interest to investigate whether these ARFs are also dependent on the return period as well. To investigate this, a frequency analysis should be applied to annual maximum depth-durations for different values of area, followed by comparing the curves valid for a particular duration with different return periods.

In a series of Flood Estimation Reports prepared by CWC and IMD, areal reduction curves for rainfall durations of 1 to 24 hrs have been established for various zones in India (see e.g. CWC, Hydrology Division, 1994). An example is presented in Figure 7.28 (zone 1(g)).

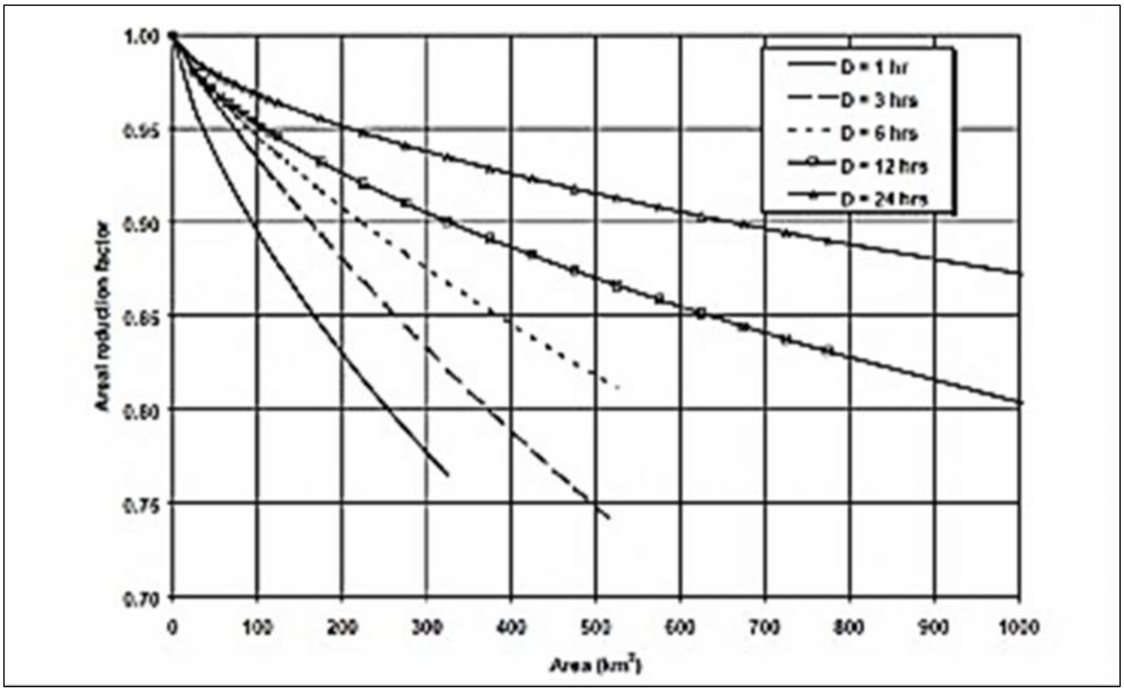


Figure 7.28: Example of areal reduction factors for different rainfall duration

7.8.4 Time distribution of storms

For design purposes once the point rainfall extreme has been converted to an areal extreme with a certain return period, the next step is to prepare the time distribution of the storm. The time distribution is required to provide input to hydrologic/hydraulic modelling. The required distribution can be derived from cumulative storm distributions of selected representative storms by properly normalising the horizontal and vertical scales to percentage duration and percentage cumulative rainfall compared to the total storm duration and rainfall amount respectively. An example for two storm durations is given in Figure 7.29, valid for the Lower Godavari sub-zone – 3 (f).

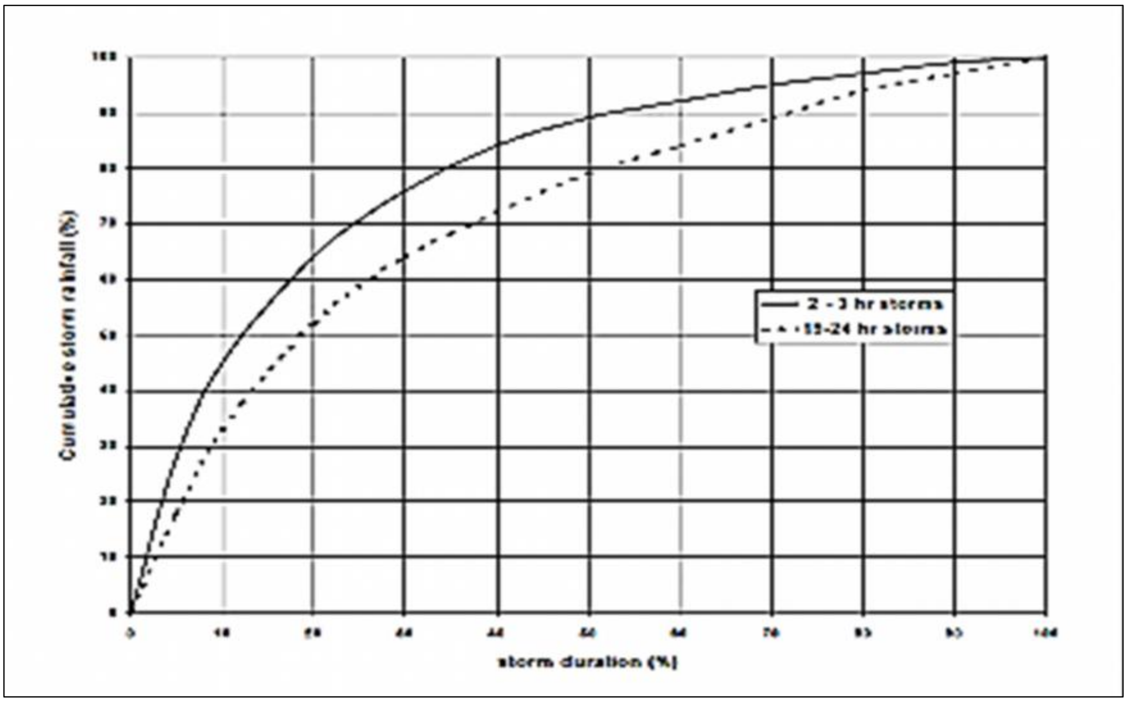


Figure 7.29: Time distributions of storms in Lower Godavari area for 2-3 and 19-24 hrs storm durations

From Figure 7.29 it is observed that the highest intensities are occurring in the first part of the storm (about 50% within 15% of the total storm duration). Though this type of storm may be characteristic for the coastal zone further inland different patterns may be determining. A problem with high intensities in the beginning of the design storm is that it may not lead to most critical situations, as the highest rainfall abstractions in a basin will be at the beginning of the storm. Therefore, one should carefully select representative storms for a civil engineering design and keep in mind the objective of the design study. There may not be one design storm distribution but rather a variety, each suited for a particular use.

8 GENERATION OF RAINFALL DATA REPORTS

8.1 General

There is a wide range of potential users in water resources sector whose requirements for hydro meteorological data may vary significantly. For instance, the standard period of data required for 'reliable' estimates of mean annual rainfall is 30 years. However, for long observed periodicities in rainfall, a longer data length may be required. In addition, for a clear evidence of global warming and associated climate change, scientists and engineers require long-term records to be able to detect and monitor trends in rainfall. In general, the longer the series and the higher the data quality, the more valuable the data. Published reports are the primary visible output of proper data management.

India has been implementing a web enabled Water Resources Information System (WRIS) with the objective of developing a nationwide water resource related database which includes storage, access and dissemination of water related data. Report generation in the context of rainfall data under WRIS has several purposes, some of which are briefly outlined below:

i. To provide information on the availability of data for use in planning and design.

Rainfall data are used for a variety of purposes and are required at a range of time scales. For example, near real-time rainfall data are required for flood forecasting, hydropower and reservoir operation. Summaries of storm rainfall event data are required for assessment of the severity of events on a weekly or monthly time scales. Rainfall bulletins for agricultural and irrigation operation are needed at similar time scales.

ii. To promote the value of the Water Resource Information System and its capability and to create interest and awareness amongst potential users.

It is conceivable that most requests for data could be met by querying the database. In India, the availability of rainfall data may not be well known even in related government departments; the annual report of rainfall available for download from WRIS may help change this perception.

iii. To provide feedback to data management agencies and acknowledge their contribution

The annual report shows how statistical summaries of observations at individual stations of rainfall data for stations that were selected for processing and validation.

The traditional annual report of daily rainfall is often not the most convenient format of rainfall data for users. For design purposes, the user often requires long term records for a single station or a group of stations - i.e., data by station rather than by year. Rather than keying of the data into the computer for the required analysis, it is now more efficient and cost effective download the entire data series for all available years of data for a selected station directly from WRIS.

The WRIS is an integrated system in which rainfall and other data are transferred in stages from the field to local and regional offices for data entry, processing and validation. It provides opportunities for storage, retrieval and reporting on electronic media, making data reporting and use more efficient by:

- reducing the amount of published data and cost of annual reports
- providing statistical summaries in tabular and graphical from which are more accessible and interesting to the user
- avoiding duplication of effort by users in keying in of data by provision on electronic media

Since the hydrological year corresponds to a complete cycle of replenishment and depletion, it is more appropriate to report on the basis of a hydrologic year rather than with respect to the calendar year. Annual reports are produced with respect to rainfall over the hydrological year from the 1st June to the 31st of May in the subsequent year. Such reports incorporate:

- Statistical summaries of information on the pattern of rainfall over the year in question
- Information on the long-term spatial and temporal pattern of rainfall in the region and how the recent year compares with past statistics.

Reports of long-term statistics of rainfall are be prepared and published at 5- or 10-year intervals. These incorporate spatial as well as temporal analysis.

Annual and other reports will be produced at the State Data Processing centre. Annual reports are produced in draft form within six months from the end of the year covered by the publication, and the final report is published within twelve months.

8.2 Annual/yearly reports

The annual report provides a summary of the rainfall for the reported year in terms of distribution of rainfall in time and space, and it makes comparisons with the long term statistics. It includes the details of the observational network and data availability. A summary of the hydrological impacts of rainfall is provided with particular reference to floods and droughts. The following are typical contents of the annual report:

(a) Introduction

(b) The Observational Network Maps and listing

(c) A descriptive account of rainfall occurrence during the reported year

(d) Thematic maps of monthly, seasonal and annual rainfall

(e) Graphical and mapped comparisons with average patterns

(f) Basic rainfall statistics

(g) Description and statistical summaries of major storms

(h) Summary of results of applying data validation tests

(i) Bibliography

8.2.1 Introduction

The report introduction, which may change little from year to year, will describe the administrative organization of the rainfall network and the steps involved in the collection, data entry, processing, validation, analysis and storage. It will list those agencies contributing to the included data. It will describe how the work is linked with other agencies collecting or using rainfall data including the India Meteorological Department and operational departments in hydropower and irrigation sectors. It will describe how additional data may be requested and under what terms and conditions they are supplied.

8.2.2 The observational network

The salient features of the observational network are summarized in map and tabular form. The rainfall station map must also show major rivers and basin boundaries and distinguish each site by symbol between daily, autographic and digital recorder and whether rainfall alone is observed or the gauge is sited at a climatologically station.

Tabulations of current stations are listed by named basin and sub-basin. Also listed are latitude, longitude, altitude, responsible agency, the full period of observational record and the period of observation which is available in digital format. A similar listing of closed stations, (or a selection of closed stations with long records) may be provided. All additions and closures of stations must be highlighted in the yearly report. Similarly, station upgrading and the nature of the upgrading should be reported.

8.2.3 Descriptive account of rainfall during the report year

An account of the rainfall occurrence in the region in the year can be concisely given in the form of a commentary for each month, placed in its meteorological context. Significant stretches of dry or wet periods in the parts of the region under reporting can be highlighted.

8.2.4 Maps of monthly, seasonal and yearly areal rainfall

Thematic maps showing spatial distribution of average rainfall over the region for monthly, seasonal or yearly periods provide a convenient summary of the rainfall pattern in space and time. Basin or administrative boundaries may also be shown to illustrate variations between districts or basins. The rainfall may be mapped as the actual value at each station for the specified period or by the drawing of isohyets of equal rainfall over the region. For such interpolations the rainfall is first interpolated on a very fine grid laid over the region using manual or computer-based techniques. Grid point values are then used to draw isohyets at suitable intervals.

8.2.5 Graphical and mapped comparisons with average patterns

Maps will also be provided to show relative rainfall - the amount as a percentage of long-term average. The period over which the long-term average is taken must be noted.

For a few selected rainfall stations, a graphical comparison of the monthly rainfall amounts for the whole year can be made with the long-term statistics. The actual monthly distribution can be plotted against the long-term average for minimum, maximum and average monthly amounts. This kind of plot also makes it easy to comprehend the type of temporal distribution of rainfall.

8.2.6 Basic statistics for various durations

This forms the core of the report. As noted above, the full reporting of daily or hourly data is no longer required though sample tabulations of daily and hourly data may be provided for selected stations to illustrate the format of information available. Instead, summary statistics of monthly rainfall for the reported year provide a ready means of making comparisons between stations and between months and will satisfy the needs of general data users.

Stations are listed by basin and sub-basin order (rather than alphabetical or numerical order). In addition to monthly rainfall totals, the maximum daily amount in the year and the date of its occurrence is noted. Any daily, monthly or annual totals which exceed previous maxima of record are shown in bold type.

For stations with digital or autographic records a similar tabulation is provided by basin giving the maximum observed amount for selected durations including 1 hour, 2, 3, 6, 12 and 24 hours with dates of occurrence.

8.2.7 Description and statistical summaries of major storms

Major storms which are known to have caused flooding are described in more detail. Selection of events for this list may be made in terms of the impacts or on the assessed areal amount and distribution. For rainfall regimes of arid and semi-arid regions a lower value is adopted whereas for high rainfall regimes a higher threshold value is adopted. Usually, a threshold of about 10% of the seasonal normal rainfall may be taken for the most frequent storm duration over the region. The threshold value also depends upon the size of the catchment area. For smaller catchment a higher threshold and for larger catchments smaller threshold value may be adopted. An average precipitation depth of 50 mm per day over a catchment of medium size (say 10,000 – 15,000 sq. kms.) would be appropriate. The peripheral isohyet for one day storm must be at least 50 mm in the moderate rainfall regime whereas it must be about 10 to 20 mm for arid or semi-arid regions with low seasonal rainfall.

Storms should be described with respect to their meteorological context, centre of concentration, movement across the river basins and also the characteristics of the time distribution of rainfall within the storm.

8.2.8 Data validation and quality

The limitations of data should be made known to users. The validation process not only provides a means of checking the quality of the raw data but also a means of reporting. The number of values corrected or in-filled as a total or a percentage may be noted for individual stations, by basin or by agency. The types of anomaly typically detected by data validation and remedial actions should be described.

8.2.9 Bibliography

Data users may be interested in other sources of rainfall data or in the related climatic or hydrological data. The following should be included:

- Concurrent annual reports from the HIS of climate or hydrological data
- Previous annual rainfall reports (with dates) from the WRIS
- Previous annual rainfall reports (with dates) published by each agency and division within the state
- Special summary reports of rainfall statistics produced by the WRIS or other agencies.
- A brief note on the administrative context of previous reports, methods of data compilation, and previous report formats would be helpful.

8.3 Periodic reports - long term statistics

The long-term point and areal statistics are important for planning, management and design of water resources systems. They also play an important role in validation and analysis. These statistics must be updated regularly and an interval of 10 years is recommended. The following will be typical contents of such reports:

i. Introduction
ii. Data availability - maps and tables
iii. Descriptive account of annual rainfall since the last report
iv. Thematic maps of mean monthly and seasonal rainfall
v. Basic rainfall statistics - monthly and annual means, maxima and minima for period of records.
vi. Additional point rainfall statistics for example, daily maximum rainfall, persistence of dry or wet spells during the monsoon, dates of onset or termination of the monsoon.
vii. Additional mean areal rainfall statistics for administrative or drainage areas for periods of a month or year
viii. Analysis of temporal variability using moving averages or residual mass curves to identify major wet and dry periods for a number of representative stations.
ix. Frequency analysis of rainfall data

8.3.1 Frequency analysis of rainfall data

The frequency of occurrence of rainfall of various magnitudes is important in the application of mathematical models for synthesising hydrological data. Estimates of design runoff from small areas are often based on rainfall-runoff relations and rainfall frequency data due to sparse stream flow measurements and limitation in transposing such data among small areas. Generalised estimates of rainfall frequencies for a few durations up to 72 hours and up to a few hundred years are useful if are readily available. Some such maps are available at country level for specified duration of rainfall and frequency of occurrence (or return periods). These maps must be revised after having collected a significant amount of additional data. Standard methods recommended by India Meteorological Department must be followed for the derivation of such maps. Though the primary responsibility for making such maps lies with the India Meteorological Department, it is appropriate to include such maps in the reports with the permission of the IMD.

Information on the frequency of rainfall is a vital input for planning domestic or industrial water supply, agricultural planning, hydropower and other water use sectors. Inferences on various time intervals such as daily, weekly, ten-daily, fortnightly and monthly are usually required for planning in various sectors.

8.4 Periodic reports on unusual rainfall events

Special reports should also be prepared on the occurrence of unusual rainfall events. As these will also have unusual hydrological consequences, the reports will normally be combined with reports of the resulting stream flow and flooding within the affected area.

The rainfall component of such reports will include the following:

- Tabulations of hourly or daily point rainfall within the affected area
- Isohyet maps of total storm rainfall
- Hyetograph plots of rainfall time distribution based on recording rain gauges
- Assessment of event return periods for selected durations based on historic point rainfall data on the same stations where which recorded the unusual event
- Areal storm rainfall totals over the affected basin.

BIBLIOGRAPHY

Chow, V. T. (Ed.) (1964). *Handbook of Applied Hydrology*, Mc Graw Hill, New York.

Clarke, I (1979). *Practical Geostatistics*, 151pp.

e-Surface water Information System - *Secondary Validation Manual*.

Grubbs, F. E., and G. Beck (1972). *Extension of sample sizes and percentage points for significance tests of outlying observations*, Technometrics, 14(4), 847–854.

Hydrology Project I, *Training module SWDP module no 10: How to correct and complete rainfall data*.

Hydrology Project I, *Training module SWDP module no 11: How to compile rainfall data*.

Hydrology Project I, *Training module SWDP module no 12: How to analyse rainfall data*.

Hydrology Project I, *Training module SWDP module no 13: How to report rainfall data*.

Hydrology Project I, *Training module SWDP module no 8: How to carry out primary validation for rainfall data*.

Hydrology Project I, *Training module SWDP module no 9: How to carry out secondary validation of rainfall data*.

Isaaks, E.H. and Srivastava, R.M. (1989). *Applied Geostatistics*, Oxford University Press, New York, 561pp.

Lloyd, C.D. (2010). *Spatial Data Analysis: An Introduction for GIS Users*, Oxford University Press, New York, 206pp.

MoWR RD & GR (2018). *An Introduction to Real-time Hydrological Information Systems*, National Hydrology Project, Ministry of Water Resources, River Development and Ganga Rejuvenation, Government of India, January, 2018.

Patra, K.C. (2008). *Hydrology and Water Resources Engineering* (2nd Edition), Narosa Publishing House, New Delhi.

Ram Babu, Tejwani,K. G. Agrawal,M. C. andBhushan. L. S. (1979). *Rainfall intensity duration–return period equations and nomographs of India*. Central Soil and Water.

Raudkivi, A. J. (1979). *An Advance Introduction to Hydrological Processes and Modelling*, Pargamon Press, London.

Upadhyay, D. S., Kaur, S., Mishra, M. S. and Gupta, M. K., (1990). *Mausam*, 41, 4, 523-530.

WMO – 168 (2008). *Guide to Hydrological Practices*, Volume I & II (Hydrology – From Measurement to Hydrological Information), Weather, Climate, Water.

SUBJECT INDEX

NATIONAL HYDROLOGY PROJECT
Ministry of Water Resources, River Development and Ganga Rejuvenation
2nd Floor, Rear wing, MDSS (MTNL) Building, 9, CGO Complex,
Lodhi Road, New Delhi 110003
Email: nhp-mowr@nic.in | Website: http://nhp.mowr.gov.in