HYDROLOGY PROJECT
Technical Assistance

World Bank & Government of The Netherlands funded

Training module # SWDP - 10

How to correct and complete rainfall data

New Delhi, November 1999

CSMRS Building, 4th Floor, Olof Palme Marg, Hauz Khas,
New Delhi – 11 00 16 India
Tel: 68 61 681 / 84 Fax: (+ 91 11) 68 61 685
E-Mail: dhvdelft@del2.vsnl.net.in

DHV Consultants BV & DELFT HYDRAULICS

with
HALCROW, TAHAL, CES, ORG & JPS

Table of contents

1. Module context

While designing a training course, the relationship between this module and the others, would be maintained by keeping them close together in the syllabus and place them in a logical sequence. The actual selection of the topics and the depth of training would, of course, depend on the training needs of the participants, i.e. their knowledge level and skills performance upon the start of the course.

2. Module profile

Title	:	How to correct and complete rainfall data
Target group	:	Assistant Hydrologists, Hydrologists, Data Processing Centre Managers
Duration	:	One Session of 60 minutes
Objectives	:	After the training the participants will be able to*:* • Correct the erroneous rainfall entries • Fill-in the missing data
Key concepts	:	• Use of autographic or digital data for manually observed data • Apportionment of accumulated data • Correcting for systematic shifts • Spatial interpolation by normal ratio and distance power method
Training methods	:	Lecture, exercises, software
Training tools required	:	Board, OHS, computers
Handouts	:	As provided in this module
Further reading and references	:	

3. Session plan

No	Activities	Time	Tools
1	***General*** • Important points (1) • Important points (2) • Correction/completion procedures	2 min	 OHS 1 OHS 2 OHS 3
2	***Use of ARG and SRG data at one or more stations*** • Example 2.1 - SRG record missing or faulty - ARG available • ARG record missing or faulty - SRG available • Example 2.2(a) – Hourly rainfall distribution in the area • Example 2.2(b) – Observed and estimated rainfall • Working with HYMOS	4 min	 OHS 4 OHS 5 OHS 6 OHS 7
3	***Correcting for entries to wrong days*** • Correcting time shifts Working with HYMOS	2 min	 OHS 8
4	***Apportionment for indicated and unindicated accumulations*** • General description • Data correction procedure • Example 4.1(a) – Indentification of accumulation • Example 4.1(b) – Apportioning accumulation	4 min	 OHS 9 OHS 10 OHS 11 OHS 12
5	***Adjusting rainfall data for long term systematic shifts*** • Example 5.1(a) – Double mass curve • Example 5.1(b) – Estimation of correction factor • Example 5.1(c) – Double mass analysis results • Example 5.1(d) – Double mass curve (adjusted data) Working with HYMOS	8 min	 OHS 13 OHS 14 OHS 15 OHS 16
6	***Using spatial interpolation to interpolate erroneous and missing values*** • Arithmetic & normal ratio method • Distance power method – Definition sketch • Example 6.1(a) - Selection of test station • Example 6.1(b) – Computation of station weights • Example 6.1(c) – Estimation of spatial average • Example 6.1(d) – Comparison of observed and estimated values • Correction for heterogeneity Working with HYMOS	10 min	 OHS 17 OHS 18 OHS 19 OHS 20 OHS 21 OHS 22 OHS 23
7	***Exercise*** • Correct the anomalies detected earlier during secondary validation (for KHEDA catchment) using (1) ARG/SRG data and (2) nearest neighbour method • Correct the long term systematic shift detected earlier during secondary validation (for KHEDA catchment) using double mass analysis	15 min each	

4. Overhead/flipchart master

5. Handout

Add copy of Main text in chapter 8, for all participants.

6. Additional handout

These handouts are distributed during delivery and contain test questions, answers to questions, special worksheets, optional information, and other matters you would not like to be seen in the regular handouts.

It is a good practice to pre-punch these additional handouts, so the participants can easily insert them in the main handout folder.

7. Main text

Contents

How to correct and complete rainfall data

1. General

- **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. The process of filling in missing values is generally referred to as 'completion'.**

- **It must be recognised that values estimated from other gauges are inherently less reliable than values properly measured. 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 depend on the type of error and the availability of suitable source records with which to estimate.**

- **Correction and completion will be generally carried out at the State Data Processing Centre on the basis of data validation report from the Divisional Data Processing Centre.**

2. Use of ARG and SRG data at one or more stations

2.1 General description

All observational stations equipped with autographic raingauge (ARG) also have an ordinary or standard raingauge (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 is 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.

.

2.2 Data correction or completion procedure:

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

2.2.1 SRG record missing or faulty - ARG available

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

Example 2.1

Referring back to Example 3.1 of Module 9 wherein 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 value is 165 mm on 23/07/96 and there are a couple of instances (12^{th} & 13^{th} Aug. 96) 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 the equipment. For this, the hourly data series obtained from ARG is used to compile the corresponding daily totals. Then the daily rainfall thus obtained from SRG and ARG are tabulated together for an easy comparison as given in Table 2.1.

Table 2.1: Tabulation result for daily rainfall series obtained from SRG & ARG.

Year	mth	day	ANIOR **MPA** (ARG)	ANIOR **MPS** (SRG)
1996	7	16	11.0	11.0
1996	7	17	20.0	20.0
1996	7	18	8.0	8.0
1996	7	19	.5	.5
1996	7	20	12.0	12.0
1996	7	21	.0	.0
1996	7	22	.0	.0
1996	7	23	**126.0**	**165.0**
1996	7	24	15.5	15.5
1996	7	25	.0	.0
1996	7	26	.0	.0
1996	7	27	42.0	42.0
1996	7	28	190.0	190.0
1996	7	29	17.5	17.5
1996	7	30	.0	.0
1996	7	31	.5	.5
1996	8	1	3.5	3.5
1996	8	2	5.5	6.0
1996	8	3	3.5	3.5
1996	8	4	7.0	.0
1996	8	5	.0	7.0
1996	8	6	63.0	63.0
1996	8	7	55.0	55.0
1996	8	8	26.5	27.0
1996	8	9	.0	.0
1996	8	10	.0	.0
1996	8	11	2.5	2.5
1996	8	12	.0	**4.0**
1996	8	13	**4.0**	**18.0**
1996	8	14	**18.0**	**17.0**
1996	8	15	**17.0**	.0
1996	8	16	.0	.0
1996	8	17	.0	.0
1996	8	18	.0	.0
1996	8	19	.0	.0
1996	8	20	.0	.0
1996	8	21	.0	.0

Both 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 mm 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 according corrected from 165 mm to 126 mm as 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. Thus daily SRG data exhibits shift of one day from two independent comparisons and does not warrant further confirmation from the manuscript. In such a straight forward situation the correction can be made outright. In this case, the SRG data of 12th, 13th & 14th August have to 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.

2.2.2 ARG record missing or faulty - SRG available

The autographic record may be missing as a result for example of the failure of the recording mechanism or blockage of the funnel. Records from autographic gauges at neighbouring stations can be used in conjunction with the SRG at the station to complete the record. Essentially this involves hourly distribution of the daily total from the SRG at the station by 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 estimation.

Consider that the daily rainfall (from 0830 hrs. on previous day to 0830 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:

$$H_{test,i} = D_{test} \cdot \frac{H_{base,i}}{\sum_{i=1}^{24} H_{base,i}}$$

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

Example 2.2

Hourly rainfall data at RAHIOL station (KHEDA catchment) is considered for the period of July-August 1996. Though there is no missing data 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.

Fig. 2.1 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.

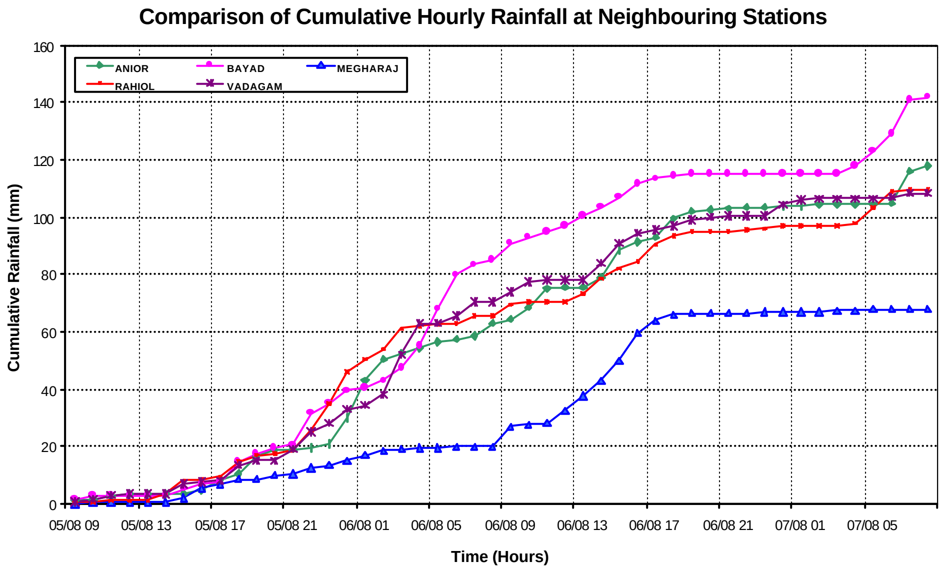

Fig. 2.1: Plot of hourly rainfall distribution at RAHIOL and surrounding stations

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 2.2. 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 **(and 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 that observed at ANIOR station ($P_{Anior,obs,i}$) are worked out as:

$P_{Rahiol,\ est,i}$ = $P_{Anior,obs,i}$ x (152.0) / (190.0) for each ith hour on 28th

and

$P_{Rahiol,\ est,i}$ = $P_{Anior,obs,i}$ x (28.0) / (17.5) for each ith hour on 29th

Similar estimate 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 2.2.

Table 2.2: Hourly distribution of observed daily rainfall by SRG on the basis of nearby hourly rainfall byARG

Date/Time	Observed Hourly rainfall (mm)			Estimated Rainfall at RAHIOL (mm)		
	ANIOR	RAHIOL **(Assumed to be missing)**	VADAGAM	As per rain distribution at		Average
				ANIOR	VADAGAM	
27/07/96 09:30	4.0	*7.0*	5.0	3.2	4.6	**3.9**
27/07/96 10:30	6.5	*5.5*	5.0	5.2	4.6	**4.9**
27/07/96 11:30	3.5	*12.5*	4.0	2.8	3.7	**3.2**
27/07/96 12:30	4.5	*5.5*	5.5	3.6	5.0	**4.3**
27/07/96 13:30	10.0	*3.5*	6.5	8.0	5.9	**7.0**
27/07/96 14:30	6.0	*2.5*	6.5	4.8	5.9	**5.4**
27/07/96 15:30	2.0	*3.5*	6.5	1.6	5.9	**3.8**
27/07/96 16:30	9.5	*6.0*	0.5	7.6	0.5	**4.0**
27/07/96 17:30	6.5	*0.5*	1.0	5.2	0.9	**3.1**
27/07/96 18:30	2.5	*1.0*	4.5	2.0	4.1	**3.1**
27/07/96 19:30	0.5	*2.5*	9.5	0.4	8.7	**4.5**
27/07/96 20:30	1.0	*0.0*	7.5	0.8	6.8	**3.8**
27/07/96 21:30	5.5	*3.0*	7.5	4.4	6.8	**5.6**
27/07/96 22:30	7.0	*4.5*	10.5	5.6	9.6	**7.6**
27/07/96 23:30	2.0	*2.5*	11.0	1.6	10.0	**5.8**
28/07/96 00:30	6.0	*8.0*	13.0	4.8	11.9	**8.3**
28/07/96 01:30	8.5	*17.0*	12.5	6.8	11.4	**9.1**
28/07/96 02:30	24.5	*28.0*	7.5	19.6	6.8	**13.2**
28/07/96 03:30	16.5	*7.5*	7.0	13.2	6.4	**9.8**
28/07/96 04:30	9.0	*6.5*	8.0	7.2	7.3	**7.3**
28/07/96 05:30	15.0	*4.0*	5.0	12.0	4.6	**8.3**
28/07/96 06:30	7.5	*2.0*	6.5	6.0	5.9	**6.0**
28/07/96 07:30	12.0	*11.0*	16.0	9.6	14.6	**12.1**
28/07/96 08:30	20.0	*0.0*	0.0	16.0	0.0	**8.0**
28/07/96 09:30	3.0	*1.0*	0.0	4.8	0.0	**2.4**
28/07/96 10:30	1.5	*1.5*	7.5	2.4	8.8	**5.6**
28/07/96 11:30	3.0	*3.5*	9.0	4.8	10.5	**7.7**
28/07/96 12:30	1.0	*4.0*	5.5	1.6	6.4	**4.0**
28/07/96 13:30	3.0	*5.5*	1.5	4.8	1.8	**3.3**
28/07/96 14:30	4.0	*3.0*	0.5	6.4	0.6	**3.5**
28/07/96 15:30	1.0	*2.0*	0.0	1.6	0.0	**0.8**
28/07/96 16:30	0.5	*0.5*	0.0	0.8	0.0	**0.4**
28/07/96 17:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
28/07/96 18:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
28/07/96 19:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
28/07/96 20:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
28/07/96 21:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
28/07/96 22:30	0.0	*0.5*	0.0	0.0	0.0	**0.0**
28/07/96 23:30	0.5	*3.5*	0.0	0.8	0.0	**0.4**
29/07/96 00:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
29/07/96 01:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
29/07/96 02:30	0.0	*3.0*	0.0	0.0	0.0	**0.0**
29/07/96 03:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
29/07/96 04:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
29/07/96 05:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
29/07/96 06:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
29/07/96 07:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
29/07/96 08:30	0.0	*0.0*	0.0	0.0	0.0	**0.0**
ARG Daily Totals						
28/07/96	190.0	144.0	166.5	152.0	152.0	152.0
29/07/96	17.5	28.0	24.0	28.0	28.0	28.0
Observed Daily Rainfall by SRG						
28/07/96	190.0	152.0	168.0			
29/07/96	17.5	28.0	24.0			

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 Fig. 2.2. 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 not be so good and even then it may be acceptable in view of no other way of estimation.

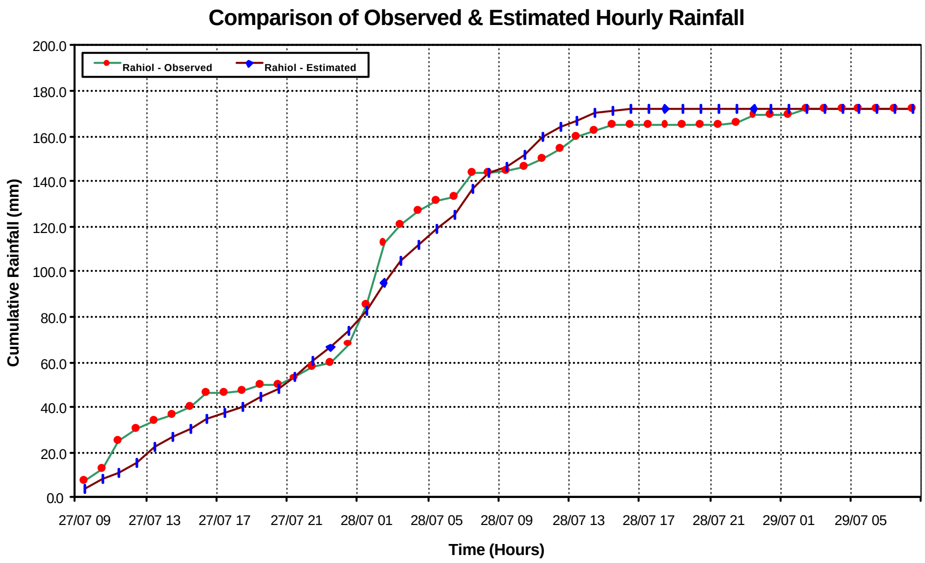

Fig.: 2.2: Comparison of observed and estimated hourly rainfall at RAHIOL station

3. Correcting for entries to wrong days

3.1 General description

Daily rainfall data are commonly 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.

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. The extent of data affected by the shift is the number of data in the group which are affected by the shift. 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 shift may be the full monthly period or a period of days within the month. The data must be corrected by deleting the two unwanted data from the desired location in the month. This deletion must be followed by shifting the affected data backward 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 (normally zero). Where the shift extends to the end of the month then the last 2 data in the month must similarly be filled up with suitable entries. Where the shift continues into the following month, the first two values of the next month are transferred to the last two values of the previous month and the process is continued.

Example 3.1

Referring back to Example 4.1 in Module 9, 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 1994 is given in Table 3.1.

As is clear from the tabulation that there is 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 consequent rainfall events. Exactly same shift is persisting for all 20 days and is confirmed by closely looking at the start and end times of five rainfall events (highlighted) one after another. If the manuscript records does not show any shift then it means that there has been an error while entering or handling the data and must therefore be accordingly corrected. Even if the records also show the same shift at SAVLITANK station, it can be confidently attributed, in such clear cut cases, to the incorrect recording by the observer.

The corrected data series for SAVLITANK station is shown in the last column of Table 3.1. It may be seen that the data from 3rd August to 20th August is advanced by one day using simple copying and pasting option while editing the data series.

Table 3.1: Correction for shift in time in daily rainfall at SAVLITANK station

Date	Daily Rainfall (mm)					
	Observed					**Corrected**
	KAPADWANJ	KATHLAL	MAHISA	**SAVLITANK**	VADOL	SAVLITANK
01/08/84	0	0	0	0	0	0
02/08/84	0	0	0.2	0	0	0
03/08/84	**152.4**	**99.3**	**157.4**	0	**39.3**	150
04/08/84	104.1	50.2	87	**150**	59.2	76
05/08/84	7.7	12	18	76	13.1	16
06/08/84	1.5	35	0	16	0	3
07/08/84	0	0	0	3	0	0
08/08/84	1.3	0	0	0	0	0
09/08/84	0	13	0	0	0	0
10/08/84	**231.2**	**157**	**179**	0	**17.3**	201
11/08/84	43.2	18.3	64	**201**	63.2	26
12/08/84	0	0	0	26	33.3	0
13/08/84	0	0	0	0	13.1	0
14/08/84	0	0	20	0	0	30
15/08/84	0	0	0	30	0	0
16/08/84	**2.6**	**8.3**	**16.5**	0	**16.3**	20
17/08/84	0	0	0	**20**	20.2	0
18/08/84	**32**	**50.3**	**25.6**	0	37.2	27
19/08/84	16.51	8.2	15	**27**	19.3	13
20/08/84	0	0	0	13	0	0

4. Apportionment for indicated and unindicated accumulations

4.1 General description:

Where the daily raingauge has not been read for a period of days and the total recorded 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.

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}

a) 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:

$$P_{est,j} = \frac{\sum_{i=1}^{N_{base}} (P_{ij} / D_i^b)}{\sum_{i=1}^{N_{base}} (1/D_i^b)} = \sum_{i=1}^{N_{base}} \left((P_{ij}) \frac{(1/D_i^b)}{\sum_{i=1}^{N_{base}} (1/D_i^b)} \right)$$

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.

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

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

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 4.1

Referring back to Example 9.1 of Module 9, wherein during validation of data at DAKOR station it is suspected that there has been an accumulation of rainfall during the month of

July 1995 which has not been indicated by the observer. The tabulation of data of DAKOR and other neighbouring stations is given in Table 4.1.

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 21th onwards, it is required to distribute this accumulated value in 8 days. This accumulated value is distributed in proportion of the corresponding estimated values at DAKOR station.

Table 4.1: Tabulation of daily rainfall for neighbouring stations

Tabulation of series, Year 1995

Year	mth	day	DAKOR	KATHLAL	MAHISA	MAHUDHA	SAVLITANK	THASARA
1995	7	11	.0	7.0	10.0	1.5	27.0	9.0
1995	7	12	.0	.0	3.0	2.0	3.0	17.0
1995	7	13	**.0**	45.0	.0	.0	.0	.0
1995	7	14	**.0**	10.0	20.0	7.5	.0	7.0
1995	7	15	**.0**	14.0	50.0	33.5	24.0	77.0
1995	7	16	.0	.0	8.0	9.5	25.0	8.0
1995	7	17	**.0**	20.0	4.0	1.0	.0	22.0
1995	7	18	**.0**	10.0	8.0	1.0	6.0	11.0
1995	7	19	**.0**	23.0	20.0	43.0	27.0	16.0
1995	7	20	.0	.0	35.0	32.5	14.0	48.0
1995	7	21	**.0**	57.0	27.0	23.0	14.0	56.0
1995	7	22	.0	.0	6.0	7.0	4.0	.0
1995	7	23	.0	.0	4.0	12.0	2.0	27.0
1995	7	24	**.0**	10.0	.0	.0	.0	.0
1995	7	25	**.0**	11.0	10.0	3.0	6.0	3.0
1995	7	26	**.0**	25.0	.0	10.0	5.0	8.0
1995	7	27	**.0**	18.0	3.0	4.0	25.0	9.0
1995	7	28	**97.5**	25.0	24.0	46.0	3.0	12.0
1995	7	29	16.7	40.0	4.0	6.0	.0	.0
1995	7	30	6.8	45.0	34.0	22.0	62.0	52.0
1995	7	31	.0	10.0	3.0	13.0	39.0	9.0

Use is made of the estimation procedure outlined in the description above and assuming the value of the exponent as 2.0. The distances and computation of weights of the neighbouring stations is computed as given in Table 4.2:

The estimated daily rainfall based on the weighted average of the neighbouring station is computed and is given in Table 4.3. 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. Now, 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 the Table 4.3.

Table 4.2: Computation of normalised weights for neighbouring stations on the basis of distance power method

Name of Neighbouring station	Distance from DAKOR	Factor	Station weight
	D_i	$(1/D_i)^{**}2$	$\{(1/D_i)^{**}2\}/\sum\{(1/D_i)^{**}2\}$
THASARA	8.25	0.0020	0.082
MAHISA	13.95	0.0051	0.208
KATHLAL	22.12	0.0019	0.078
MAHUDHA	22.70	0.0018	0.074
SAVLITANK	23.40	0.0138	0.558
	SUM	0.0247	1.0

Table 4.3: Computation of spatial estimate during period of accumulation and its distribution

Date	Observed	Weighted Rainfall (for DAKOR) at					Weighted	Corrected
	DAKOR	KATHLAL	MAHISA	MAHUDHA	SAVLITANK	THASARA	Average	DAKOR
		Station weight						
		0.0819	0.2079	0.0785	0.0739	0.5575	$R_{est,j}$	$R_{est,j}$*97.5/104.1
07/07/95	0	0.000	0.000	0.000	0.000	0.000	0.0	0
08/07/95	0	0.000	0.000	0.000	0.000	0.000	0.0	0
09/07/95	0	0.000	0.000	0.000	0.000	0.000	0.0	0
10/07/95	0	0.000	0.000	0.000	0.000	0.000	0.0	0
11/07/95	0	0.574	2.080	0.118	1.996	5.018	9.8	*
12/07/95	0	0.000	0.624	0.157	0.222	9.479	10.5	*
13/07/95	0	3.689	0.000	0.000	0.000	0.000	3.7	*
14/07/95	0	0.820	4.160	0.589	0.000	3.903	9.5	*
15/07/95	0	1.148	10.399	2.631	1.774	42.933	58.9	*
16/07/95	0	0.000	1.664	0.746	1.848	4.461	8.7	*
17/07/95	0	1.640	0.832	0.079	0.000	12.267	14.8	*
18/07/95	0	0.820	1.664	0.079	0.444	6.133	9.1	*
19/07/95	0	1.885	4.160	3.378	1.996	8.921	20.3	*
20/07/95	0	0.000	7.279	2.553	1.035	26.764	37.6	*
21/07/95	0	4.673	5.616	1.807	1.035	31.224	**44.4**	**41.5**
22/07/95	0	0.000	1.248	0.550	0.296	0.000	**2.1**	**2.0**
23/07/95	0	0.000	0.832	0.943	0.148	15.054	**17.0**	**15.9**
24/07/95	0	0.820	0.000	0.000	0.000	0.000	**0.8**	**0.8**
25/07/95	0	0.902	2.080	0.236	0.444	1.673	**5.3**	**5.0**
26/07/95	0	2.049	0.000	0.785	0.370	4.461	**7.7**	**7.2**
27/07/95	0	1.476	0.624	0.314	1.848	5.018	**9.3**	**8.7**
28/07/95	**97.5**	2.049	4.992	3.613	0.222	6.691	**17.6**	**16.5**
29/07/95	16.7	3.279	0.832	0.471	0.000	0.000	4.6	16.7
30/07/95	6.8	3.689	7.071	1.728	4.583	28.994	46.1	6.8
31/07/95	0	0.820	0.624	1.021	2.883	5.018	10.4	*

* Error on these days are not due to accumulation but due to either non-observation or incorrect recording and is to be corrected using appropriate spatial interpolation method (See section 6)

5. Adjusting rainfall data for long term systematic shifts

5.1 General description

The double mass analysis technique is used in validation to detect significant long-term systematic shift in rainfall data. The same technique is 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 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 later part, where an erroneous source of the inconsistency is known or where the record has been discontinued. The correction procedure is described below.

5.2 Data correction procedure

Consider a double mass plot shown in Fig. 5.1. 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 for the period of adjustment to be considered 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 are 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.

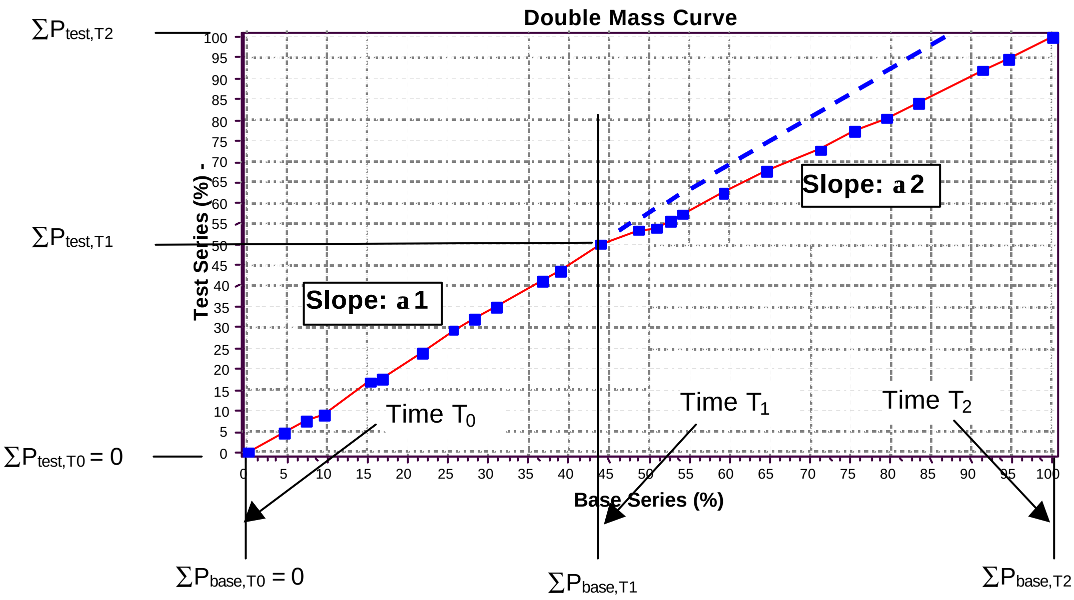

Fig. 5.1: Definition sketch for double mass analysis

Considering the double mass plot shown in Fig. 5.1, 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:

$$a_1 = \frac{\sum_{i=0}^{T_1} P_{test,i}}{\sum_{i=0}^{T_1} P_{base,i}}$$

and

$$a_2 = \frac{\sum_{i=T_0}^{T_2} P_{test,i} - \sum_{i=T_0}^{T_1} P_{test,i}}{\sum_{i=T_0}^{T_2} P_{base,i} - \sum_{i=T_0}^{T_1} P_{base,i}}$$

In case the earlier portion between T_0 and T_1 is desired to be corrected for then the correction factor and the corrected observations at the test station can be expressed respectively as:

$$P_{corr,i} = P_{test,i} \times \frac{a_2}{a_1}$$

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 monthly (rather than daily) data and carried out annually. 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 data intervals

Example 5.1

Referring back to Example 12.2 of Module 9, wherein 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 Fig. 5.2, that the behaviour of the test station suddenly changes after about half of the time period under consideration.

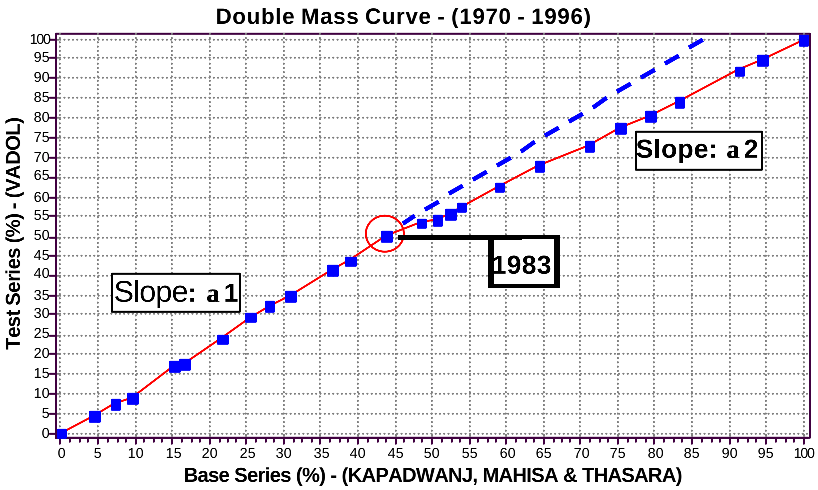

Fig. 5.2: Double mass curve for VADOL station showing significant change of slope of the curve after about half the period under consideration.

Assuming that, on the basis of a visit to the station and feed back from the observer, it has been found that the exposure conditions at the raingauge site have not been upto the desired standards. If the lower rainfall catch 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 were existing. This is done carrying out following computations:

As is apparent from Fig. 5.2 and the results of the Double mass analysis given in Table 5.1 that from the year 1984 onwards, the rainfall received at VADOL station is comparatively lesser then previous 13 year period in relation to the base stations KAPADWANJ, MAHISA & THASARA around it.

Table 5.1: Results of the double mass analysis

```
Double mass analysis
 Test series: VADOL       PH
                               Weight
 Base series: KAPADWANJ   PH      .33
              MAHISA      PH      .33
              THASARA     PH      .33
```

1	2	3	4	5	6	7	8	9
Period	**BASE**			**TEST**			**Ratios**	
	Amount	**Cum**	**Perc**	**Amount**	**Cum**	**Perc**	**(6)/(3)**	**(7)/(4)**
	MM	**MM**		**MM**	**MM**		**-**	**-**
1970	767.4	767.	4.6	624.4	624.	4.5	.81	.98
1971	454.0	1221.	7.3	426.0	1050.	7.6	.86	1.04
1972	372.5	1594.	9.5	197.9	1248.	9.0	.78	.94
1973	935.3	2529.	15.1	1114.2	2363.	17.0	.93	1.13
1974	240.3	2769.	16.6	72.8	2435.	17.6	.88	1.06
1977	843.8	3613.	21.6	882.8	3318.	23.9	.92	1.11
1978	646.4	4260.	25.5	758.8	4077.	29.4	.96	1.15
1979	436.7	4696.	28.1	370.2	4447.	32.1	.95	1.14
1980	450.2	5147.	30.8	388.9	4836.	34.9	.94	1.13
1981	950.0	6097.	36.5	898.1	5734.	41.4	.94	1.13
1982	403.6	6500.	38.9	320.1	6054.	43.7	.93	1.12
1983	801.4	**7302.**	43.7	882.1	**6936.**	50.0	.95	1.15
1984	**806.0**	8108.	48.5	**475.1**	7411.	53.5	**.91**	**1.10**
1985	364.2	8472.	50.7	82.8	7494.	54.1	.88	1.07
1986	281.5	8753.	52.3	234.0	7728.	55.7	.88	1.06
1987	257.7	9011.	53.9	227.5	7956.	57.4	.88	1.06
1988	866.1	9877.	59.1	734.5	8690.	62.7	.88	1.06
1989	877.0	10754.	64.3	693.3	9384.	67.7	.87	1.05
1990	1145.0	11899.	71.2	746.0	10130.	73.1	.85	1.03
1991	682.7	12582.	75.2	618.1	10748.	77.5	.85	1.03
1992	697.7	13279.	79.4	422.2	11170.	80.6	.84	1.01
1993	639.8	13919.	83.2	512.8	11683.	84.3	.84	1.01
1994	1350.0	15269.	91.3	1083.3	12766.	92.1	.84	1.01
1995	525.0	15794.	94.5	371.6	13137.	94.8	.83	1.00
1996	926.7	**16721.**	100.0	725.0	**13862.**	100.0	.83	1.00

Total number of periods analysis: 25

The average slope of the double mass curve before and after this break can be worked out form the computations shown in Table 5.1 as:

$$a_1 = \frac{\sum_{i=1}^{T_1} P_{test,i}}{\sum_{i=1}^{T_1} P_{base,i}} = \frac{6936}{7302} = 0.9498$$

and

$$a_2 = \frac{\sum_{i=T_0}^{T_2} P_{test,i} - \sum_{i=T_0}^{T_1} P_{test,i}}{\sum_{i=T_0}^{T_2} P_{base,i} - \sum_{i=T_0}^{T_1} P_{base,i}} = \frac{13862 - 6936}{16721 - 7302} = 0.7353$$

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

$$Correction \quad Factor = \frac{a_2}{a_1} = \frac{0.9498}{0.7353} = 1.2916$$

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 Fig. 5.2. The double mass curve after adjusting the data series is given in Fig. 5.3 and the corresponding tabular analysis results in Table 5.2. It may be noted that the double mass curve after the data series is corrected beyond 1983 shows a consistent trend throughout.

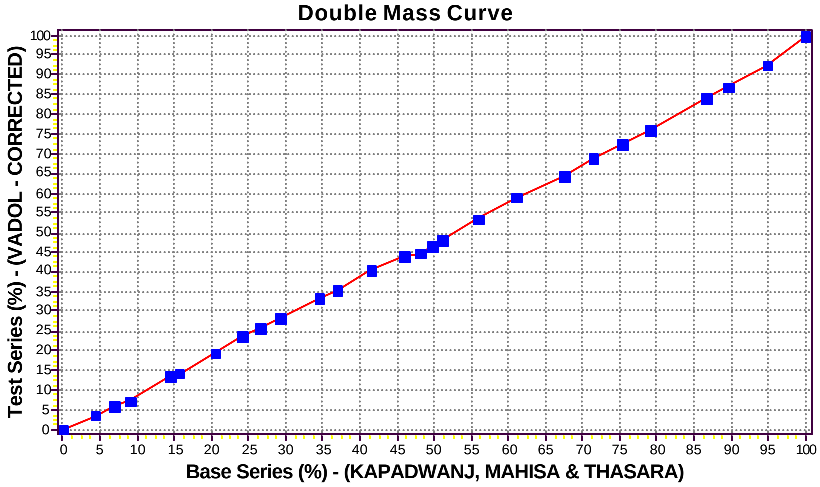

Fig. 5.3: Double mass plot after adjusting the rainfall values for the period of inconsistency

Table 5.2: Result of double mass analysis after adjusting the rainfall values for the period of inconsistency

```
Double mass analysis
 Test series: VADOL       TMA
                               Weight
 Base series: KAPADWANJ   PH      .33
              MAHISA      PH      .33
              THASARA     PH      .33
```

1	2	3	4	5	6	7	8	9
Period	BASE			TEST			Ratios	
	Amount	Cum	Perc	Amount	Cum	Perc	(6)/(3)	(7)/(4)
	MM	MM		MM	MM		-	-
1970	767.4	767.	4.4	624.4	624.	3.6	.81	.84
1971	454.0	1221.	6.9	426.0	1050.	6.1	.86	.88
1972	372.5	1594.	9.0	197.9	1248.	7.3	.78	.80
1973	935.3	2529.	14.3	1114.2	2363.	13.8	.93	.96
1974	240.3	2769.	15.7	72.8	2435.	14.2	.88	.90
1977	843.8	3613.	20.5	882.8	3318.	19.3	.92	.94
1978	646.4	4260.	24.2	758.8	4077.	23.7	.96	.98
1979	436.7	4696.	26.6	370.2	4447.	25.9	.95	.97
1980	450.2	5147.	29.2	388.9	4836.	28.2	.94	.96
1981	950.0	6097.	34.6	898.1	5734.	33.4	.94	.97
1982	403.6	6500.	36.9	320.1	6054.	35.3	.93	.96
1983	801.4	7302.	41.4	882.1	6936.	40.4	.95	.98
1984	806.0	8108.	46.0	613.6	7550.	44.0	.93	.96
1985	364.2	8472.	48.1	106.9	7657.	44.6	.90	.93
1986	281.5	8753.	49.7	302.2	7959.	46.4	.91	.93
1987	257.7	9011.	51.1	293.8	8253.	48.1	.92	.94
1988	866.1	9877.	56.0	948.7	9202.	53.6	.93	.96
1989	877.0	10754.	61.0	895.5	10097.	58.8	.94	.96
1990	1145.0	11899.	67.5	963.5	11061.	64.4	.93	.95
1991	682.7	12582.	71.4	798.3	11859.	69.1	.94	.97
1992	697.7	13279.	75.3	545.3	12404.	72.3	.93	.96
1993	639.8	13919.	79.0	662.3	13067.	76.1	.94	.96
1994	1350.0	15269.	86.6	1399.2	14466.	84.3	.95	.97
1995	525.0	15794.	89.6	480.0	14946.	87.1	.95	.97
1996	926.7	16721.	94.9	936.4	15882.	92.5	.95	.98
1997	907.7	17628.	100.0	1283.9	17166.	100.0	.97	1.00

Total number of periods analysis: 26

6. Using spatial interpolation to interpolate erroneous and missing values

6.1 General description

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 usually only a limited number of data values will require to be estimated at a stretch. Three analytical procedures for estimating rainfall using such spatial interpolation methods are described below.

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:

$$P_{test} = \frac{1}{M}(P_{base,1} + P_{base,2} + P_{base,3} + ... + P_{base,M})$$

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

Example 6.1

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, 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 are considered to get an idea of variability. The station normal annual rainfall at these stations are obtained as under:

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 difference in the normal annual rainfall at the three base stations is about 5.5, 1.3 and 7.8 % only 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 mulitplying them with an equal weight of 0.333. Table 5.3 shows the computation of the daily rainfall estimates at BALASINOR station on the basis of above three adjoining (base) stations.

Table 5.3: Estimation of daily rainfall at BALASINOR station by arithmetic average method

Date	Observed Rainfall (mm)			Estimated Rainfall (mm)
	MAHISA	SAVLITANK	VADOL	BALASINOR
	Station Weights			
	0.333	**0.333**	**0.333**	
12/07/88	0.0	0.0	0.0	**0.0**
13/07/88	13.0	0.0	2.0	**5.0**
14/07/88	25.0	50.0	37.2	**37.4**
15/07/88	46.0	30.0	42.0	**39.3**
16/07/88	97.0	50.0	17.0	**54.7**
17/07/88	4.0	3.0	5.0	**4.0**
18/07/88	8.0	3.0	14.0	**8.3**
19/07/88	7.0	15.0	16.0	**12.7**
20/07/88	21.0	28.0	18.5	**22.5**

21/07/88	6.0	6.0	3.0	**5.0**
22/07/88	62.0	45.0	28.0	**45.0**
23/07/88	15.0	18.0	38.0	**23.7**
24/07/88	5.0	8.0	4.0	**5.7**
25/07/88	18.0	10.0	4.8	**10.9**
26/07/88	6.0	15.0	20.0	**13.7**
27/07/88	43.0	0.0	12.0	**18.3**
28/07/88	40.0	125.0	47.4	**70.8**
29/07/88	11.0	21.0	17.6	**16.5**
30/07/88	0.0	5.0	6.6	**3.9**
31/07/88	11.0	11.0	5.2	**9.1**

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:

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

Where:
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)

A minimum of three adjoining stations must be generally used for obtaining good estimates using normal ratio method.

<u>Example 6.2</u>
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, VADOL around this station is also not available. However, records for two stations KAPAWANJ and THASARA which are at comparatively farther distance from BALASINOR station is 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 are considered to get an idea of variability. The station normal annual rainfall at these stations are obtained from 20-25 years of data between 1970 to 1997 as under:

For BALASINOR = N_{test} = 715 mm
For KAPADWANJ = $N_{base,1}$ = 830 mm
For 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 is more than 10% criterion and thus the normal ratio method for obtaining the estimates of daily rainfall at BALASINOR station is tried.

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

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

and

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

The normalised averaging can be carried out by employing the process of algebraic series transformation on the two base series taken together and mulitplying 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 the Table 6.1.

Table 6.1: Estimation of daily rainfall at BALASINOR station by arithmetic average and normal ration method

Date	Observed Rainfall (mm)		Rainfall at BALASINOR (mm)		
	KAPADWANJ	THASARA	Estimated		Observed
			Arithmetic	Normal Ratio	
			Weights		
			0.5 & 0.5	0.431 & 0.450	
25/08/73	0.0	0.0	0.0	0.0	8.0
26/08/73	0.0	4.4	2.2	2.0	2.0
27/08/73	0.0	4.0	2.0	1.8	2.0
28/08/73	0.0	0.0	0.0	0.0	2.0
29/08/73	35.0	8.6	21.8	19.0	24.0
30/08/73	86.0	33.0	59.5	51.9	54.0
31/08/73	119.0	170.8	144.9	128.1	130.0
01/09/73	36.0	107.0	71.5	63.7	71.8
02/09/73	25.0	6.0	15.5	13.5	20.0
03/09/73	35.0	21.0	28.0	24.5	20.0
04/09/73	12.0	34.0	23.0	20.5	30.0
05/09/73	17.0	21.0	19.0	16.8	15.0
06/09/73	8.0	3.0	5.5	4.8	5.6
07/09/73	71.0	54.0	62.5	54.9	58.0
08/09/73	113.0	43.8	78.4	68.4	66.0
09/09/73	4.0	0.0	2.0	1.7	0.0
10/09/73	0.0	0.0	0.0	0.0	2.0

It may be seen from the above estimation 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 method, 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.

6.4 Distance power method

This method weights neighbouring stations on the basis of their distance from the station under consideration, on the assumption that 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. Normally, an exponent of 2 is used with the distances to obtain the weighted average.

In this method four quadrants are delineated by north-south and east-west lines passing through the raingauge station under consideration, as shown in Fig. 6.1. 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 following:

- The adjoining stations must lie within the specified radius having significant spatial correlation with one another.
- A maximum number of 8 adjoining 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 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.

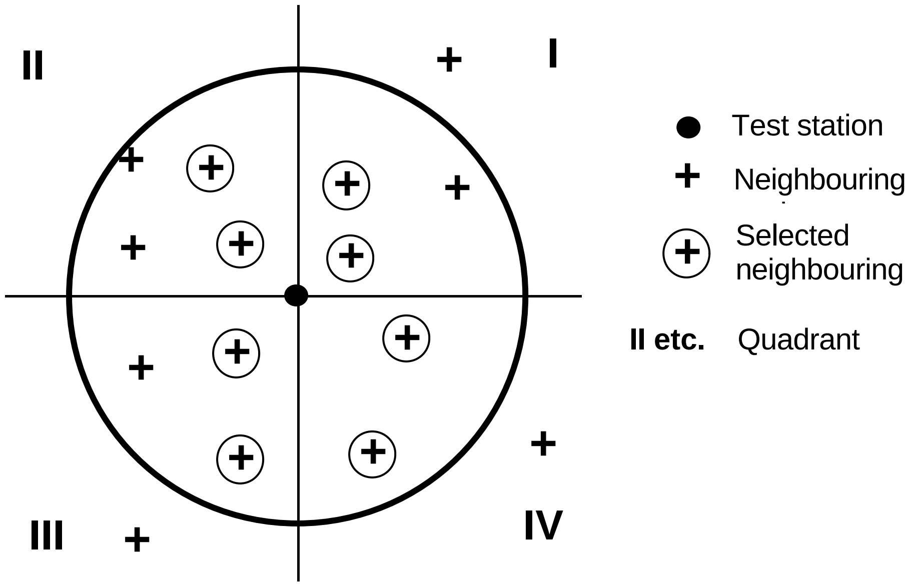

Fig. 6.1: Definition sketch of Test and Base (neighbouring) stations

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

$$P_{est,j} = \frac{\sum_{i=1}^{M_{base}} P_{i,j} / D_i^{\,b}}{\sum_{i=1}^{M_{base}} 1/D_i^{\,b}}$$

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

6.4.1 Correction for heterogeneity

To correct for the sources of heterogeneity, e.g. orographic effects, normalised values must be used in place of actual rainfall values at the adjoining stations. This implies that the observed rainfall values at the adjoining 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). That is:

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

Where:
$P_{corr,i,j}$ = for heterogeneity corrected rainfall value at 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 normals are either found from the historical records and are readily available. Otherwise, they may be computed from established relationships, as a function of altitude, if sufficient data is not available at all stations for estimating station normals. The relationship for station normals as a function of the station altitude (H) is of the form:

$$N_i = a_1 + b_1 \, . \, H_s \qquad \forall \quad H_s \le H_1$$

$$N_i = a_2 + b_2 \, . \, H_s \qquad \forall \quad H_s > H_1$$

Example 6.3
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 kms. and using the option of "Spatial Interpolation" and six such stations are identified. Selection of the test and base stations is also shown in Fig. 6.2. The nearest two stations are tried to be chosen which fall within the circle of 25 kms. radius. These stations are listed in Table 6.2 along with the quadrant, distances and corresponding normalised weights.

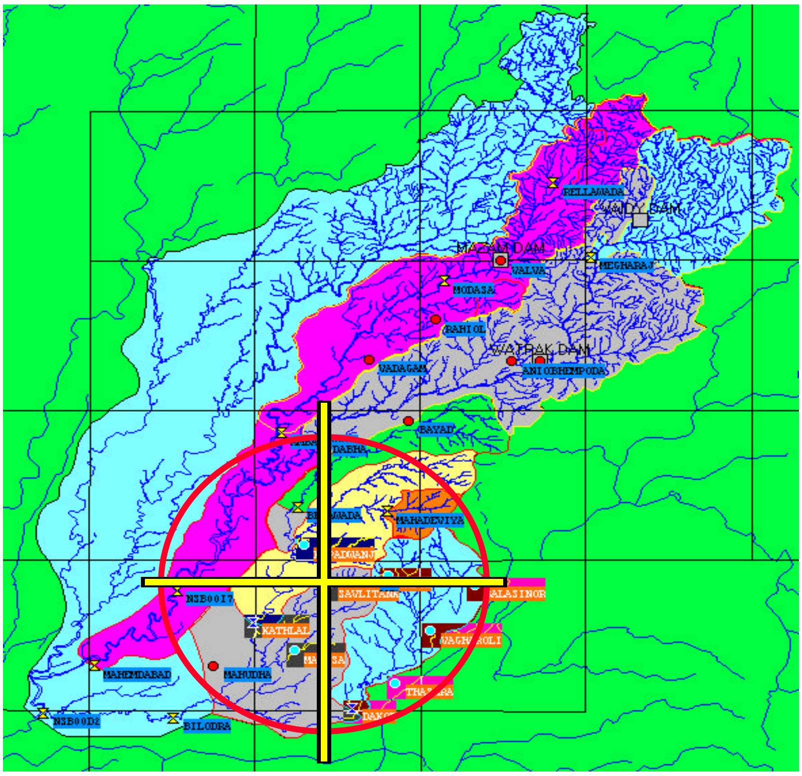

Fig. 6.2: Selection of test station SAVLITANK and its adjoining (base) stations

Table 6.2: Distances and normalised weights of stations adjoining SAVLITANK station

Quadrant	Station	Distance (kms.)	Station weights (µ 1/D^2)	
			(1/D^2)	Normalised weights
I	VADOL	9.225	0.011751	0.274
II	KAPADWANJ	8.139	0.015096	0.352
III	MAHISA	13.480	0.005503	0.128
III	KATHLAL	13.895	0.005179	0.121
IV	VAGHAROLI	17.872	0.003131	0.073
IV	THASARA	21.168	0.002232	0.052
		Sum =	0.042892	1.0

Results of the spatial interpolation are presented in Table 6.3 for July-August 1994 wherein the observed rainfall at all the six base stations is listed followed with 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 Fig. 6.3) 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 an 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 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 station 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 6.3: Observed daily rainfall at base stations and computation of spatial average at SAVLITANK

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	13	0	0	9	20	**6.3**	0
16/08/94	0	3	0	0	3	0	**1.3**	2
17/08/94	8	0	0	6	15	8	**4.4**	2
18/08/94	0	2	0	0	2	22	**2.0**	0
19/08/94	18	4	0	10	6	0	**8.0**	0
20/08/94	68	50	0	15	120	132	**53.7**	60
21/08/94	0	14	5	3	0	5	**6.2**	7
22/08/94	14	0	0	0	5	0	**4.2**	2
23/08/94	0	0	0	0	0	0	**0.0**	0
24/08/94	0	0	0	0	0	0	**0.0**	0
25/08/94	0	0	2	0	0	0	**0.3**	0
26/08/94	0	0	0	5	0	0	**0.6**	0
27/08/94	9	4	6	5	5	7	**6.0**	0
28/08/94	40	43	0	0	43	43	**31.5**	39
29/08/94	0	14	0	0	0	0	**4.9**	0
30/08/94	0	0	0	7	0	0	**0.8**	0
31/08/94	0	0	0	0	0	40	**2.1**	0
01/09/94	50	74	30	10	30	15	**47.8**	24
02/09/94	27	60	25	8	25	45	**36.9**	18
03/09/94	0	48	0	5	18	41	**20.9**	21
04/09/94	0	0	6	0	0	0	**0.8**	4
05/09/94	0	4	3	0	10	0	**2.5**	2
06/09/94	0	0	0	7	0	0	**0.8**	0
07/09/94	220	336	315	100	305	312	**269.5**	278
08/09/94	61	60	65	50	45	42	**57.7**	122
09/09/94	0	19	8	0	12	0	**8.6**	8
10/09/94	15	15	5	10	0	7	**11.6**	6
11/09/94	0	0	0	0	0	4	**0.2**	0
12/09/94	8	0	0	0	0	0	**2.2**	0
13/09/94	0	0	0	0	0	0	**0.0**	0
14/09/94	15	0	0	0	0	115	**10.1**	0
15/09/94	0	0	80	18	0	40	**14.5**	5
16/09/94	40	44	16	33	45	112	**41.6**	40
17/09/94	0	13	0	10	12	0	**6.7**	32
18/09/94	0	0	0	12	0	0	**1.4**	0
19/09/94	0	0	0	15	0	0	**1.8**	0
20/09/94	0	0	0	0	0	0	**0.0**	0

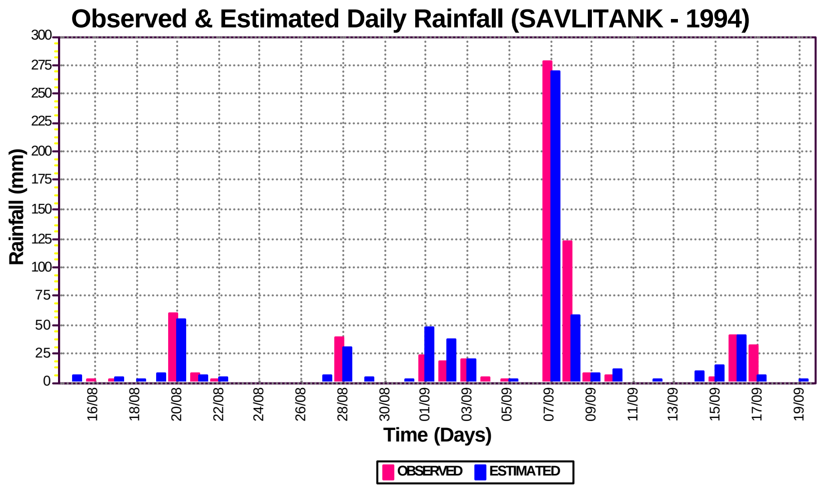

Fig. 6.3: Comparison of observed and estimated daily rainfall at SAVLITANK station