HYDROLOGY PROJECT
Technical Assistance

World Bank & Government of The Netherlands funded

Training module # SWDP - 09

How to carry out secondary validation of 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 carry out secondary validation of rainfall data
Target group	:	Assistant Hydrologists, Hydrologists, Data Processing Centre Managers
Duration	:	Three sessions of 60 minutes each
Objectives	:	After the training the participants will be able to: • Perform secondary validation of rainfall data
Key concepts	:	• Spatial correlation structure of rainfall for various durations • Spatial homogeneity • Entries at wrong days • Accumulated rainfall • Transposed entries • Number of rainy days • Double mass analysis • Auto correlation and spectral density functions
Training methods	:	Lecture, exercises, software
Training tools required	:	OHS, Computers
Handouts	:	As provided in this module
Further reading and references	:	

3. Session plan

No	Activities	Time	Tools
1	***General*** • Secondary validation – General points • Spatial correlation – Influencing factors • KHEDA Catchment • Spatial correlation – Daily interval • Spatial correlation – Ten-daily interval • Spatial correlation – Monthly interval • Spatial correlation – Region A • Catchment map (Region A) • Spatial correlation – Extended region • Catchment map (Extended region) • Use of spatial correlation • Validation procedures	15 min	 OHS 1 OHS 2 OHS 3 OHS 4 OHS 5 OHS 6 OHS 7 OHS 8 OHS 9 OHS 10 OHS 11 OHS 12
2	***Screening of data*** • Screening • Example 2.1- Tabular	10 min	 OHS 13 OHS 14
3	***Scrutiny of multiple time series graphs*** • Example 3.1 - Multiple plot (stacking side-by-side) • Example 3.1 - Multiple plot (shifted vertically)	10 min	 OHS 15 OHS 16
4	***Scrutiny of tabulations of daily rainfall at multiple stations*** • Example 4.1 - Scrutiny of tabulation	10 min	 OHS 17
5	***Checking against data limits for totals at longer durations*** • Example 4.1 - Scrutiny by tabulation	5 min	 OHS 18
6	***Spatial homogeneity testing of rainfall (nearest neighbour analysis)*** • General • Nearest neighbour test - Definition • Spatial estimate • Example 6.1(a) – Selection of test station • Example 6.1(b) – Tabular results	15 min	 OHS 19 OHS 20 OHS 21 OHS 22 OHS 23
7	***Identification of common errors*** • Example 7.1 - Shift in entries (test results) • Tabulation of suspect period • General – Unidentified accumulation • Example 7.2 – Test results • Tabulation of suspect period • Example 7.3 – Missed entries (test results) • Tabulation of suspect period	10 min	 OHS 24 OHS 25 OHS 26 OHS 27 OHS 28 OHS 29 OHS 30

No	Activities	Time	Tools
8	***Checking for systematic shifts using double mass curves*** • General principle • Detection of in-homogeneity • Test procedure • Double mass curve - Consistent data set • Example 12.1 – Inconsistent data set		
9	***Exercise:*** • For KHEDA data set detect at least one suspect situation each by employing four methods: (1) Screening, (2) Multiple graphs, (3) Tabulation, and (4) Limits for totals on longer duration • Employ spatial homogeneity test for a station in KHEDA catchment and infer the test results • Apply double mass analysis and detect in-homogeneous data set at one or more stations in KHEDA catchment	30 min 30 min 30 min	Computer

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 carry out secondary validation of rainfall data

1. General

- **Rainfall data received at Divisional offices have already received primary validation on the basis of knowledge of instrumentation and conditions at the field station** and information contained in Field Record Books.

- **Secondary validation now puts most emphasis on comparisons with neighbouring stations to identify suspect values.** Some of the checks which can be made are oriented towards specific types of error 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 at Division. However since comparison with neighbouring stations is limited by Divisional boundaries, the validation of some stations near the Divisional boundaries will have to await assemblage 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 whilst another may remain dry only a few miles away. Over a month or monsoon season such spatial unevennesss tends to be smoothed out and aggregated totals are much more closely correlated.

  Spatial correlation in rainfall thus depends on:
  duration (smaller at shorter durations),
  distance (decreasing with distance),
  type of precipitation, and
  physiographic characteristics of a region.

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

Example 1.1:
The effect of aggregation of data to different time interval and that of the inter-station distances on the correlation structure is illustrated here.

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 Fig. 1.1, Fig. 1.2 and Fig. 1.3 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 becomes flatter as the aggregation is done. This demonstrates that the correlation distance for monthly interval is much more than that for ten daily interval.

And similarly the correlation, which sharply reduces with increase in distance for the case of daily time interval, does maintain its significance over quite longer distances.

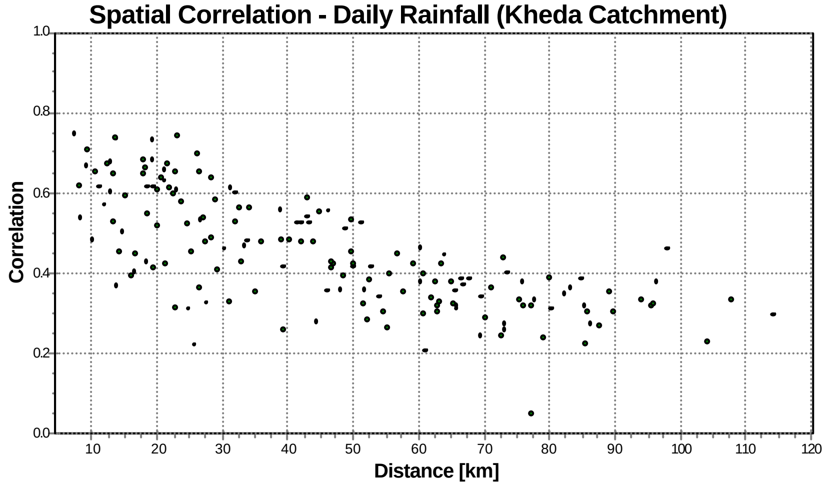

Fig. 1.1: Plot of correlation with distance for daily rainfall data

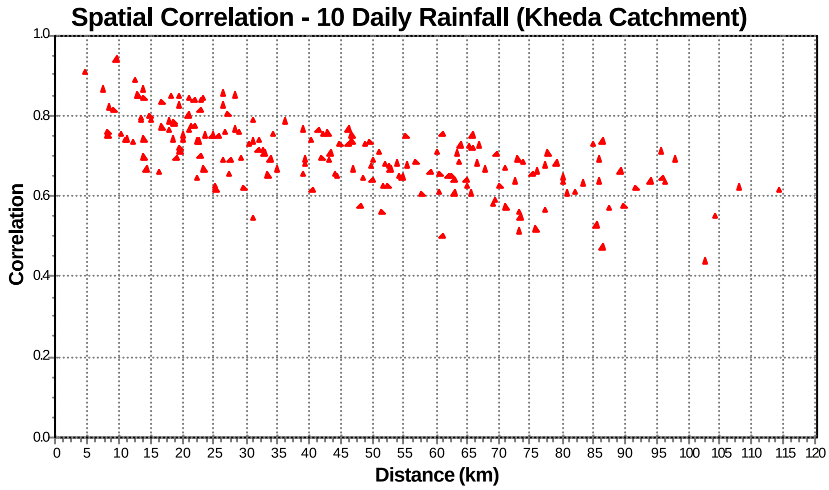

Fig. 1.2: Plot of correlation with distance for ten-daily rainfall data

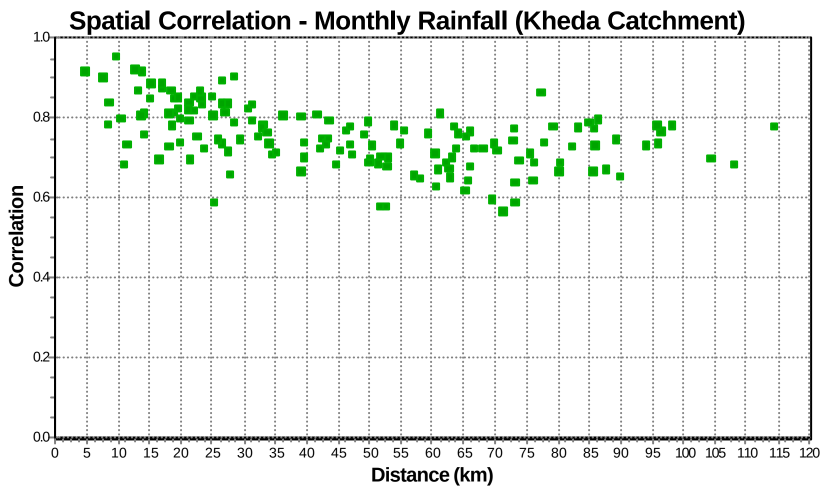

Fig. 1.3: Plot of correlation with distance for monthly rainfall data

Example 1.2:
Effect of physiographic characteristics over the correlation structure is illustrated by considering monthly rainfall for two groups of stations in the PARGAON catchment.

Fig. 1.4 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, Fig.1.5). This small region can be considered as homogeneous in itself and which is also substantiated by the scatter plot of the correlation. Monthly rainfall data has been considered for this case and as is clear from the plot there is a very high level of correlation among stations and the general slope of the scatter diagram indicates a high value of the correlation distance.

However, Fig. 1.6 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 Fig. 1.7).

It is apparent from Fig. 1.6 that in case such a combination of stations, in which there are a few stations from the hilly region and another lot from the adjoining plain region, is taken then the resulting correlation shows a weaker correlation structure. The correlation decays very fast against distance and even for shorter distances it is very much 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 Fig. 1.6 in which lot of scatter is shown even for smaller inter station distances.

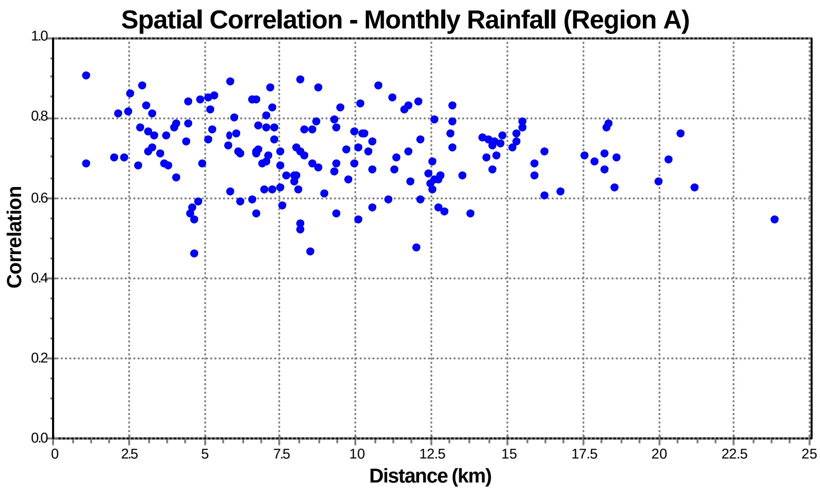

Fig. 1.4: Scatter plot of correlation for monthly rainfall in the small hilly region

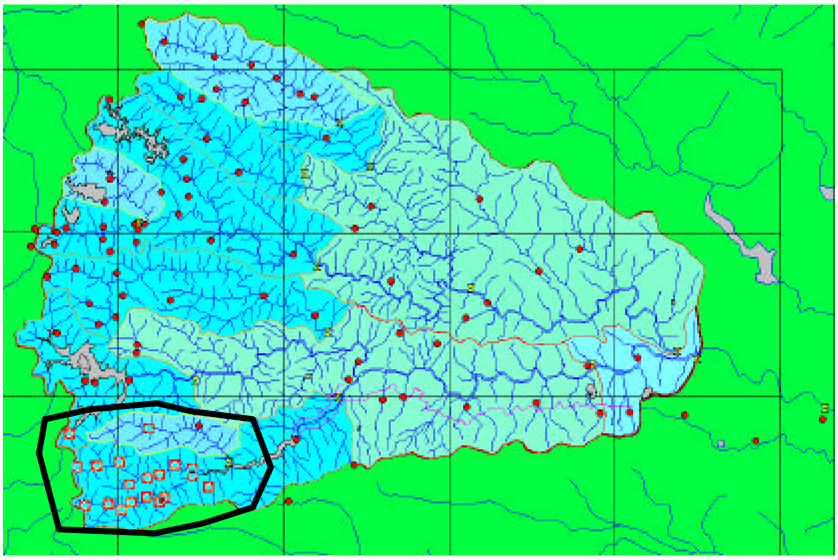

Fig. 1.5: Selection of a group of some 20 stations in the hilly region of the catchment.

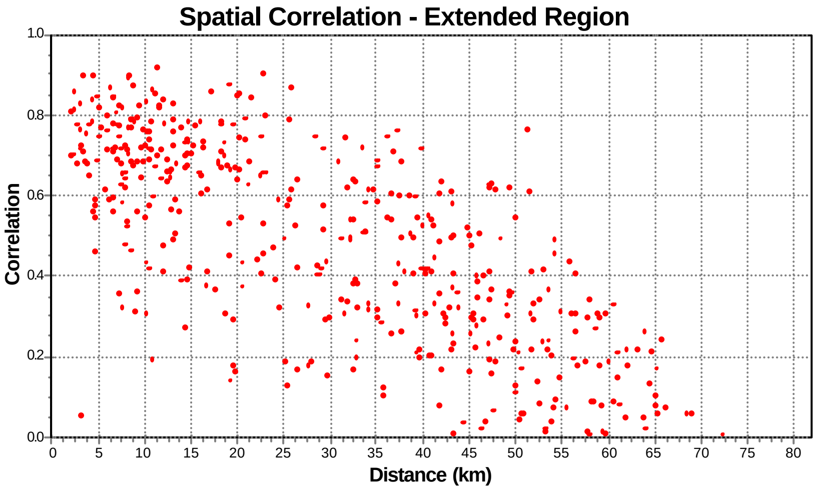

Fig. 1.6: Scatter plot of correlation for monthly rainfall in the extended region

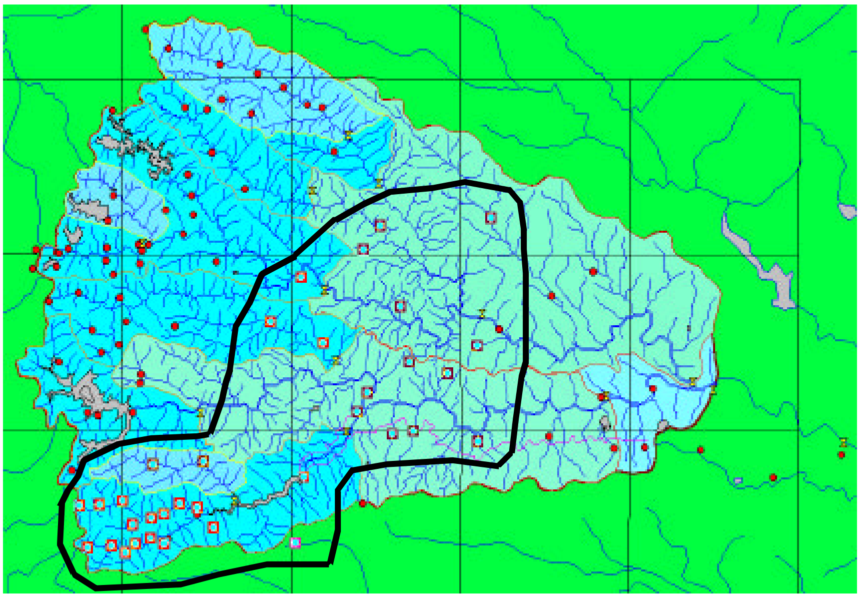

Fig. 1.7: Selection of a group of some 34 stations in the extended region of the catchment.

- **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 well to give weight to the previous performance of the station and the observer.

One must particularly be wary of rejecting extreme values, as true extreme values are for design purposes the most interesting and useful ones in the data series. True extreme values (like 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 Module 8.

2. Screening of data series

After the data from various Sub-Divisional offices has been received at the respective Divisional office, it is organised and imported into the temporary databases of secondary module of dedicated data processing software. The first step towards data validation is making the listing of data thus for various stations in the form of a dedicated format. Such listing of data is taken for **two main objectives: (a) to review the primary validation** exercise by getting the data values screened against desired data limits and **(b) to get the hard copy of the data on which any remarks or observation about the data validation can be maintained and communicated** subsequently to the State/Regional data processing centre.

Moreover, for the case of validation of historical data for period ranging from 10 to 40 years this listing of the screening process is all the more important. This screening procedure involves, for example 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. Below this matrix of data the monthly and yearly basic statistics like total rainfall, maximum daily rainfall, number of rainy days etc. are listed. Also, the number of instances where the data is missing or has violated the data limits is also given.

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

<u>Example: 2.1</u>:
An example of the listing of screening process for MEGHARAJ station of KHEDA catchment for the year 1991 is given in Table 2.1. 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 2.1: Result of the screening process of daily rainfall data for one year

Daily data and statistics of series MEGHARAJ   MPS   Year = 1997

Day	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
1	.0	.0	192.5*	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
2	.0	.0	15.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
3	.0	.0	1.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
4	.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	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
6	.0	.0	.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
7	.0	.0	1.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
8	.0	.0	32.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
9	.0	.0	1.0	25.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
10	.0	.0	.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
11	.0	.0	.0	14.5	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
12	.0	.0	7.0	1.5	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
13	.0	.0	1.0	4.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
14	.0	.0	.5	.5	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
15	.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	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
17	.0	.0	.5	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
18	.0	.0	.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
19	.0	10.0	12.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
20	.0	.0	1.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
21	.0	2.0	6.5	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
22	.0	1.0	.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
23	12.0	.0	9.5	2.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
24	9.0	.0	125.5	27.5	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
25	138.0*	1.0	11.0	.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	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
27	38.0	312.0*	1.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
28	54.0	32.5	.0	.0	.0	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*	-99.0*
29	.0	4.5	.5	.0	.0	-99.0*	-99.0*	-99.0	**********	-99.0*	-99.0*	-99.0*
30	.0	12.0	.5	.0	.0	-99.0*	-99.0*	-99.0	**********	-99.0*	-99.0*	-99.0*
31	*********	22.0	.0	*********	.0	*********	-99.0*	-99.0	**********	-99.0	**********	-99.0*
Data	30	31	31	30	31	30	31	31	28	31	30	31
Eff.	30	31	31	30	31	0	0	0	0	0	0	0
Miss	0	0	0	0	0	30	31	31	28	31	30	31
Sum	397.0	401.0	474.5	76.0	5.5	-99.0	-99.0	-99.0	-99.0	-99.0	-99.0	-99.0
Mean	13.2	12.9	15.3	2.5	.2	-99.0	-99.0	-99.0	-99.0	-99.0	-99.0	-99.0
Min.	.0	.0	.0	.0	.0	-99.0	-99.0	-99.0	-99.0	-99.0	-99.0	-99.0
Max.	138.0	312.0	192.5	27.5	5.5	-99.0	-99.0	-99.0	-99.0	-99.0	-99.0	-99.0
High	130.0	130.0	130.0	130.0	130.0	.0	.0	.0	.0	.0	.0	.0
Numb	2	1	1	0	0	0	0	0	0	0	0	0
Low	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
Numb	0	0	0	0	0	0	0	0	0	0	0	0

Annual values:

Data	365 *	Sum	1354.0 *	Minimum	.0 *	Too low	0
Effective	153 *	Mean	8.8 *	Maximum	312.0 *	Too high	4
Missing	212						

Exceedance of:
- Lower bound (.00) marked with *
- Upper bound (130.00) marked with *
- Rate of rise (320.00) marked with +
- Rate of fall (320.00) marked with -

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:

(a) Choose a set of stations within a small area with an expectation of spatial correlation.

(b) Include, if possible, in the set one or more stations which historically have been more reliable.

(c) 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. Stacking side-side is presently possible with the software.

(d) After inspection for anomalies and comparing with climate, all remaining suspect values are flagged, and comment inserted as to the reason for suspicion.

<u>Example 3.1</u>:

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. Fig. 3.1 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 not more than 5 kms. apart. Though it is not that such variation could not be possible but at least such deviations are sufficient for one to cross check with other information. On checking with the hourly observations available at ANIOR station it is noticed that the compiled daily rainfall is only 126 mm. This substantiates the earlier suspicion of it being comparatively larger.

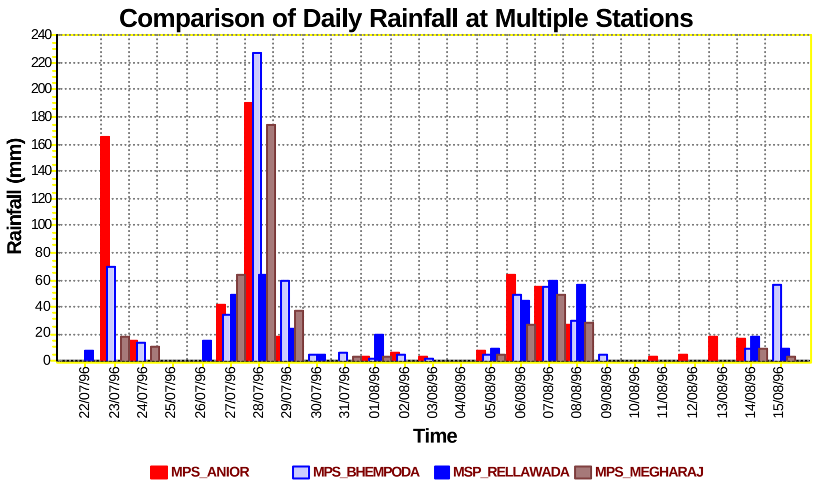

Fig. 3.1: Comparison of multiple time series plot of daily rainfall data

Further it may be noticed from the plot that the daily rainfall for 12th and 13th August at ANIOR seems to 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 observation can be helpful in getting more insight into the probable reasons of such departures.

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 below), plotted as histograms. Scanning such tabular series will often be the first step in secondary data validation. Anomalies to look out for are:

- Do the daily blocks of rainydays 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 4.1:
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 kms. diameter and thus are expected to experience similar rainfall on an average.

For an 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 4.1 for a period of 1st August to 20th August 1984. A very casual look at this tabulation reveal 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 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 by underlining) one after another.

Such a finding then must be followed by first a closer look at the manuscript record and to see if the shift has been during entering or managing the data series. If it is found that the this 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 feed back from the observer will help in settling this type of discrepancy and also will encourage observer to be careful subsequently.

Table 4.1: Tabulation for scrutiny of possible error in the timing of daily rainfall data

Tabulation of series, Year 1984

Year	mth	day	hr	si	KAPADWANJ PH	KATHLAL PH	MAHISA PH	SAVLITANK PH	VADOL PH
1984	8	1			.0	.0	.0	.0	.0
1984	8	2			.0	.0	.2	.0	.0
1984	8	3			152.4	99.3	157.4	.0	39.3
1984	8	4			104.1	50.2	87.0	150.0	59.2
1984	8	5			7.7	12.0	18.0	76.0	13.1
1984	8	6			1.5	35.0	.0	16.0	.0
1984	8	7			.0	.0	.0	3.0	.0
1984	8	8			1.3	.0	.0	.0	.0
1984	8	9			.0	13.0	.0	.0	.0
1984	8	10			231.2	157.0	179.0	.0	17.3
1984	8	11			43.2	18.3	64.0	201.0	63.2
1984	8	12			.0	.0	.0	26.0	33.3
1984	8	13			.0	.0	.0	.0	13.1
1984	8	14			.0	.0	20.0	.0	.0
1984	8	15			.0	.0	.0	30.0	.0
1984	8	16			2.6	8.3	16.5	.0	16.3
1984	8	17			.0	.0	.0	20.0	20.2
1984	8	18			32.0	50.3	25.6	.0	37.2
1984	8	19			16.5	8.2	15.0	27.0	19.3
1984	8	20			.0	.0	.0	13.0	.0
1984	8	21			.0	.0	.0	.0	.0

5. Checking against data limits for totals at longer durations

5.1 General description:

Many systematic errors are individually so small that they can not 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) contains 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 also the resulting series compiled at larger time interval show the possible incorrectness more visibly. Accordingly, if the observed data are accumulated for longer time intervals, then the resulting time series can again be checked against corresponding expected limits. This check applies primarily to daily rainfall at stations at which there is no recording gauge.

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 is consistent with the prescribed data limits for such time intervals.

Together with the upper warning level or maximum limit, for monsoon months and yearly values use of lower warning level data limit can also be made to see if certain values are unexpectedly low and thus warrants a closer look. Aggregated values violating the prescribed limits for monthly or annual duration are flagged as suspect and appropriate remarks made in the data validation report stating the reasons for such flagging. These flagged values must then validated on the basis of data from adjoining stations.

Example 5.1:

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

The plot of these yearly values is shown in Fig. 5.1. In this case of yearly rainfall data the values can be validated against two data limits as upper and lower warning levels. The values of such limits can be drawn from the experience of the distribution of the yearly rainfall in the region. In this case, the mean of the 26 yearly values is about 660 mm with an 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 – 1.5 x (standard deviation) = 660 – 1.5 x 320 = 180 mm

and

Upper warning level = mean + 2.0 x (standard deviation) = 660 + 2.0 x 320 = 1300 mm

The multipliers to the standard deviation for the lower and upper warning levels have been taken differently in view of the data being positively skewed with a finite lower bound. 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.

These limits have been shown in the plot of the 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 the year 1997 a large value of yearly rainfall more than 1329 mm is reported and similarly for year 1974 the reported rainfall is as low as 92.6 mm.

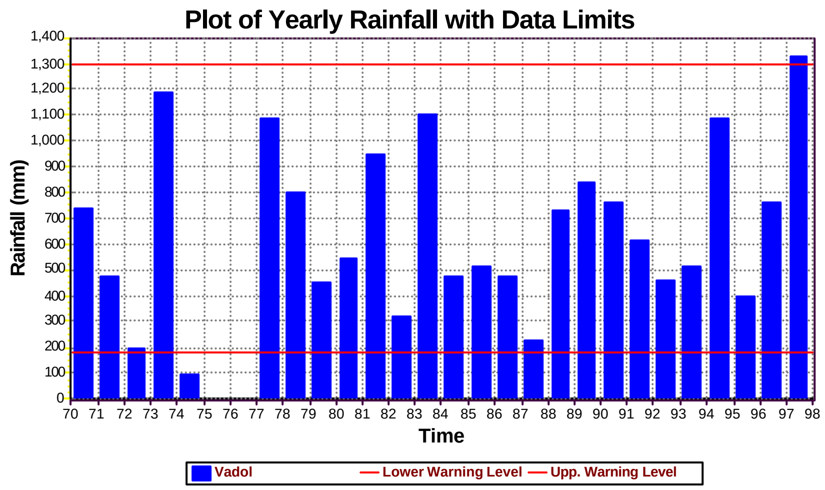

Fig. 5.1: Plot of rainfall data compiled at an yearly interval

After screening such instances of extreme values in the data series compiled at longer time intervals, it is then 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 Table 5.1 for an easier comparison.

Table 5.1: Tabulation of yearly rainfall at five neighbouring stations

Tabulation of series, Year 1970 - 1997

	=========Data=========				
Year mth day hr si	**BALASINOR MPS**	**KAPADWANJ MPS**	**SAVLITANK MPS**	**VADOL MPS**	**VAGHAROLI MPS**
1970	802.8	927.2	-99.0	739.8	-99.0
1971	546.7	569.5	-99.0	475.0	-99.0
1972	338.2	291.0	-99.0	198.2	-99.0
1973	1061.2	1305.0	1226.0	1186.4	1297.4
1974	**338.1**	**421.0**	**268.5**	**92.6**	**-99.0**
1975	-99.0	-99.0	-99.0	-99.0	-99.0
1976	-99.0	-99.0	-99.0	-99.0	-99.0
1977	1267.2	1217.5	1168.9	1083.5	1575.8
1978	672.8	507.5	517.0	801.4	1347.0
1979	437.5	428.5	525.5	455.6	1197.0
1980	551.3	661.6	378.0	545.7	892.0
1981	917.7	1273.6	1004.0	950.7	722.0
1982	302.1	540.2	376.0	320.1	267.0
1983	1028.0	1088.5	1020.0	1099.1	1110.0
1984	523.1	882.9	888.0	475.1	649.6
1985	438.9	661.5	1101.0	510.8	1173.0
1986	526.9	474.9	256.0	470.7	505.0
1987	257.0	256.0	209.0	227.5	232.0
1988	-99.0	1133.0	826.0	734.5	849.4
1989	1088.0	1064.0	787.0	840.8	-99.0
1990	1028.1	971.0	1042.0	761.0	1174.0
1991	451.0	815.0	523.0	618.1	628.0
1992	421.1	1028.0	469.0	459.6	606.0
1993	531.0	410.5	781.0	512.8	781.0
1994	1085.0	1263.0	1039.0	1083.3	1332.0
1995	590.0	528.0	422.0	399.6	525.0
1996	1397.0	968.0	760.0	762.6	1050.0
1997	1272.0	1876.0	1336.2	1329.0	950.0

It may be seen from this table that for the year 1997 at most of the neighbouring stations the reported rainfall is very high and is even about 1875 mm for KAPADWANJ station. At two other stations also it is in the range of 1200 to 1300 mm except that for VAGHAROLI 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 similar higher values reported nearby. Comparison for the year 1974 shows that though all the stations seems 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.

For looking at the daily data for the year 1974 a tabulation is again obtained as given in Table 5.2 for the neighbouring stations. Only a portion of the year for a brief period in May is given the Table.

Though, there are comparatively more zeros reported for the VADOL station then other stations for many rain events during the season but looking at the variability in the neighbouring stations it 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 have an error due to non-observation or incorrect reporting. It is necessary to refer 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 5.2: Tabulation of daily rainfall at VADOL station.

1974	5	23	.0	.0	.0	.0	-99.0
1974	5	24	.0	.0	.0	.0	-99.0
1974	5	25	.0	.0	.0	.0	-99.0
1974	5	26	**4.2**	**75.0**	**73.0**	**.0**	-99.0
1974	5	27	**23.0**	**30.0**	**19.0**	**.0**	-99.0
1974	5	28	.0	.0	.0	.0	-99.0
1974	5	29	12.0	.0	.0	.0	-99.0
1974	5	30	.0	.0	.0	.0	-99.0
1974	5	31	.0	.0	.0	.0	-99.0

6. Spatial homogeneity testing of rainfall (Nearest neighbour analysis)

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 the weighted average of rainfall observed at the surrounding stations. Whenever the difference between the observed and the estimated values exceed the expected limiting value then such values are considered as suspect values. Such values are then flagged for further investigation and ascertaining the possible causes of the departures.

6.2 Data validation procedure and follow up actions

First of all, 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 Fig. 6.1):

(a) the distance between the test and the neighbouring station must be less than a specified maximum correlation distance, say R_{max} 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:

$$P_{est}(t) = \frac{\sum_{i=1}^{N} P_i(t) / D_i^{\,b}}{\sum_{i=1}^{N} 1 / D_i^{\,b}}$$

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 taken into account.
b = power of distance D

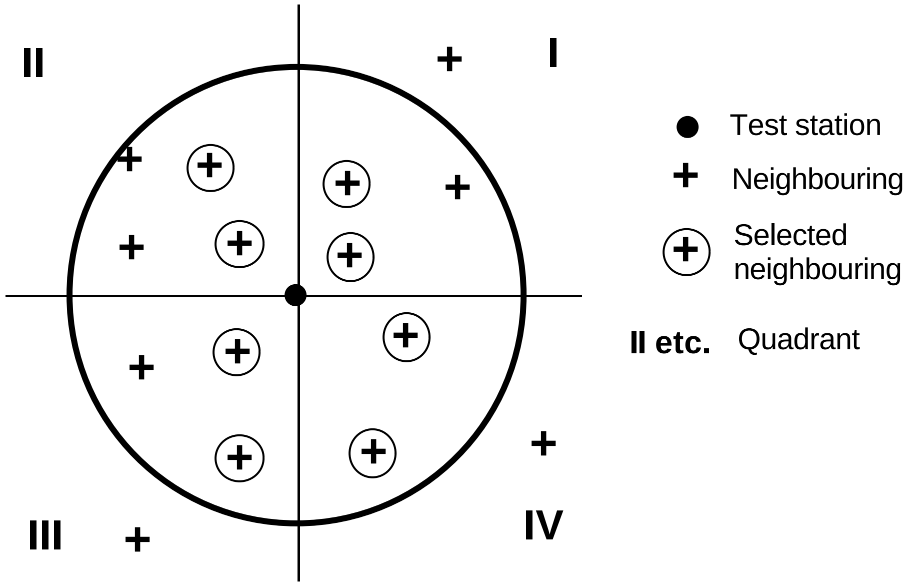

Fig. 6.1: Definition sketch of Test and Base (neighbouring) stations

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:

$$|P_{obs}(t) - P_{est}(t)| \leq X_{abs}$$

$$|P_{obs}(t) - P_{est}(t)| \leq X_{rel} * S_{Pest}(t)$$

where:

X_{abs} = admissible absolute difference
$S_{Pest}(t)$ = standard deviation of neighbouring values
X_{rel} = multiplier of standard deviation and

$$S_{P_{est}(t)} = \sqrt{\sum_{i=1}^{N} (P_i(t) - \overline{P}_i(t))^2}$$

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 values (say 2-3%) as suspect values.

It is customary to select a reasonably high value of X_{abs} to avoid having to deal with a large number of difference values in the lower range. In the example, illustrated below, X_{abs} = 25 mm. This value may be altered seasonally. 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 rainfalls where X_{rel} x S has a small absolute 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.

<u>Example 6.1</u>:
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 kms. around SAVLITANK station is considered for selecting the base stations (see Fig. 6.2). 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 6.1.

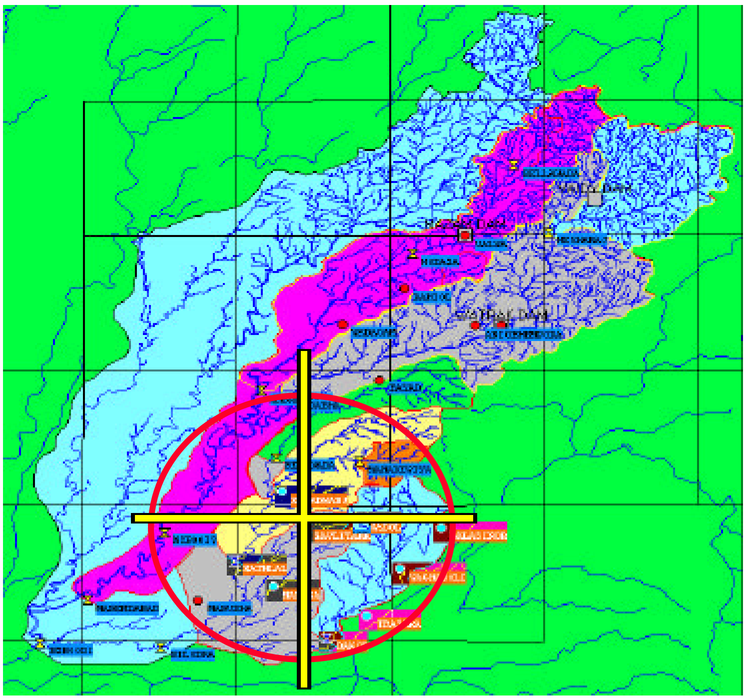

Fig. 6.2: Selection of test station SAVITANK and neighbouring base stations

Table 6.1: Results of the spatial homogeneity test.

Spatial homogeneity check
===================================
Test station SAVLITANK PH

Start date: 1984 6 1 0 1 End date: 1985 10 31 0 1

Radius of circle of influence : 25.000 (km)

Station weights proportional to : 1/D^2.00
Admissible absolute error : 50.000
Multiplier to stdv of neighbours: 2.000

Selected neighbour stations:

Quadrant	Station		Distance (km)
1	VADOL	PH	9.225
2	KAPADWANJ	PH	8.139
3	MAHISA	PH	13.480
3	KATHLAL	PH	13.895
4	VAGHAROLI	PH	17.872
4	THASARA	PH	21.168

Year	mth	day	hr	si	P_obs	flag	P_est	Stdv	n
1984	6	14	0	1	9.00	+	.00	.00	6
1984	6	15	0	1	14.00	+	.00	.00	6
1984	6	16	0	1	23.00	+	.00	.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	+	.32	1.21	6
1984	**8**	**3**	**0**	**1**	**.00**	**-**	**96.59**	**65.70**	**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**	**.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	+	.50	1.89	6
1984	8	19	0	1	27.00	+	16.81	4.91	6
1984	8	28	0	1	8.00	+	.00	.00	6
1985	6	13	0	1	9.00	+	.00	.00	6
1985	6	14	0	1	14.00	+	.00	.00	6
1985	6	16	0	1	8.00	+	.00	.00	6
1985	7	2	0	1	21.00	+	.07	.37	6
1985	7	6	0	1	47.00	+	.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	+	.15	.75	6
1985	7	25	0	1	25.00	+	3.15	3.78	6
1985	8	1	0	1	10.00	+	.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	+	.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	+	.00	.00	6
1985	9	14	0	1	17.00	+	.00	.00	6
1985	9	15	0	1	3.00	+	.00	.00	6
1985	**10**	**8**	**0**	**1**	**145.00**	**+**	**70.17**	**67.38**	**6**
1985	**10**	**9**	**0**	**1**	**.00**	**-**	**86.03**	**116.43**	**6**

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.

Following can be easily deduced from the above listing:

(a) There are quite a few very large differences in the observed and the estimated values e.g. those on 3rd, 4th, 10th, 11th August 1984 and 4th, 11th August 1985 and 8th, 9th October 1985 (highlighted in the table). Such large differences warrant a closer look at the observed values in conjunction of the rainfall at the neighbouring stations.

(b) A few of these instances of large differences are preceded or followed by 0 rainfall values at the test station 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. However, another possibility is that there have 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 aspect can be ruled out. Tabulation of data at these base stations in fact reveal possibility of such shiftings.

(c) 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 (specially 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. However, in the present example, another possibility is indicated at least for those in the month of June. It can be noticed that on all the instances of June, the estimated rainfall is 0 implying that there has been zero rainfall reported at all the six neighbouring stations. And since the resulting standard deviation is also zero all these instances have been short listed. In fact, 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 and inadvertently all these missing data on and before 16th June has been reported as 0 mm. Further, SAVLITANK station being on a reservoir site might have an arrangement of having the observation throughout the year and thus the reported rainfall values may be correct.

(d) As explained above, for the listed inconsistencies possible scenarios are required to be probed further and only then a judicious corrective measure can be forthcoming. In case, none of the corroborative facts substantiates the suspicion further then either the value can be left as suspect or if the variability of the process is considered very high such suspect values can be cleared of subsequently.

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 3 and 4)
- Spatial homogeneity tests (Section 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.

8. Checking for entries on wrong days - shifted entries

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. This results 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 is being entered.

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.

<u>Example 8.1</u>:
Spatial homogeneity test for daily rainfall series of VADAGAM station in KHEDA catchment is carried out with neigbouring stations MODASA, RAHIOL, BAYAD and ANIOR as base stations. The result of this test is reported as given in Table 8.1 below:

It may be noticed from above listing that a –ve flag together with 0 mm observed rainfall followed by a +ve 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 VADAGAM. Other instances listed in the test report are primarily due to very small standard deviation among base stations during low rainfall days and may be overlooked.

This suspicion is confirmed after looking at the tabulation of this station data alongwith the other four base stations as given in Table 8.2. It may be seen that except for the event starting on 5th 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 –ve flag and 0 observed rainfall and followed later with a +ve flag in the recession phase of the event.

Table 8.1: Result of the spatial homogeneity test at VADAGAM station.

```
Spatial homogeneity check
====================================

Test station VADAGAM     PH

Start date: 1988  7  1  0  1
End   date: 1988  9 30  0  1

Radius of circle of influence   :     25.000 (km)

Station weights proportional to : 1/D^2.00

Admissible absolute error       :     50.000
Multiplier to stdv of neighbours:      2.000

Selected neighbour stations:

Quadrant      Station      Distance (km)

    1          RAHIOL      PH     12.606
    1          MODASA      PH     18.689
    4          BAYAD       PH     12.882
    4          ANIOR       PH     21.829


    Year mth day  hr  si    P_obs  flag   P_est    Stdv   n

    1988   8   1   0   1      .50    -     8.32    3.83   4
    1988   8   5   0   1      .00    -   181.97   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    +      .59    1.43   4
    1988   8  25   0   1    32.00    +     1.97    4.34   4
    1988   9   6   0   1     9.50    +      .00     .00   4
    1988   9  29   0   1    12.00    +     1.09    1.30   4

                                       Legend
                                       n = number of neighbour stations
                                       + = P_obs - P_est > 0
                                       - = P_obs - P_est < 0
                                       * = P_est is missing
```

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 following results were obtained (Table 8.3):

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 flag 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 8.2: Tabulation of daily rainfall at neighbouring stations.

Tabulation of series, Year 1988

Year	mth	day	hr	si	ANIOR PH	BAYAD PH	MODASA PH	RAHIOL PH	VADAGAM PH
1988	7	12			.0	.0	.0	.0	.0
1988	7	13			.0	.0	70.0	.0	.0
1988	7	14			33.0	65.0	75.0	30.0	14.0
1988	7	15			8.0	17.8	12.5	5.0	3.0
1988	7	16			**26.8**	**14.0**	**31.0**	**60.6**	**40.0**
1988	7	17			5.4	1.2	10.0	2.0	1.0
1988	7	18			.0	2.0	.0	.0	1.0
1988	7	19			40.0	57.8	2.5	50.8	35.0
1988	7	20			54.2	46.0	60.0	32.8	46.0
1988	7	21			7.0	17.0	4.0	4.0	19.0
1988	7	22			**113.0**	**78.4**	**124.0**	**91.8**	**82.0**
1988	7	23			.0	11.2	15.0	6.8	16.3
1988	7	24			13.0	.0	29.0	7.4	.0
1988	7	25			8.0	14.0	43.5	35.8	23.1
1988	7	26			18.0	27.0	1.0	.0	4.2
1988	7	27			31.0	1.0	.0	3.4	1.2
1988	7	28			29.0	42.0	7.0	10.0	23.0
1988	7	29			.0	14.0	15.0	4.0	10.0
1988	7	30			13.4	.0	43.0	2.0	.0
1988	7	31			4.2	17.0	6.0	.0	.0
1988	8	1			8.0	3.0	13.0	11.4	.5
1988	8	2			4.0	.0	2.0	.0	.0
1988	8	3			.0	.0	17.0	22.0	4.0
1988	8	4			.0	1.0	1.0	.0	.0
1988	8	5			**253.0**	**135.0**	**161.0**	**212.8**	.0
1988	8	6			139.0	94.0	112.0	110.6	**140.0**
1988	8	7			20.0	24.0	4.0	7.6	161.0
1988	8	8			11.2	8.0	11.0	16.5	4.0
1988	8	9			9.0	8.0	9.0	4.8	18.0
1988	8	10			2.6	3.0	8.0	1.0	1.2
1988	8	11			3.5	.0	1.0	.0	4.2
1988	8	12			.0	.0	3.0	.0	3.0
1988	8	13			.0	.0	.0	.0	.0

Table 8.3: Results of the spatial homogeneity test on the corrected series

```
Spatial homogeneity check
====================================

Test station VADAGAM     TMA

Start date: 1988  7  1  0  1
End   date: 1988  9 30  0  1

Radius of circle of influence   :      25.000 (km)

Station weights proportional to : 1/D^2.00

Admissible absolute error       :      50.000
Multiplier to stdv of neighbours:       2.000

Selected neighbour stations:

Quadrant       Station       Distance (km)

    1           RAHIOL       PH      12.606
    1           MODASA       PH      18.689
    4           BAYAD        PH      12.882
    4           ANIOR        PH      21.829
```

Year	mth	day	hr	si	P_obs	flag	P_est	Stdv	n
1988	8	1	0	1	.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	+	.00	.00	4
1988	9	29	0	1	12.00	+	1.09	1.30	4

Legend
n = number of neighbour stations
+ = P_obs - P_est > 0
- = P_obs - P_est < 0
* = P_est is missing

9. Entries made as accumulations

9.1 General description:

The rainfall observer is expected to take rainfall observations every day at the stipulated time, without discontinuity for either holidays, weekends or sickness. Nevertheless, it is likely that on occasions the raingauge 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 also occurs. 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.

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 unindicated 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 –ve flag on days on which there was significant rain during the period of accumulation and a +ve 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, 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 the 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 the respective days as:

$$P_{appor,i} = \frac{P_{est,i} * P_{tot}}{\sum_{i=1}^{N_{acc}} P_{est,i}}$$

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 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 9.1:
As a routine secondary validation, spatial homogeneity test for station DAKOR (KHEDA catchment) for the year 95 is carried out considering a few neighbouring stations. The test results are as given below (Table 9.1):

On examining the above results, it can be apparent that there are a few "–ve" flags having nil observed rainfall which is followed by a "+ve" flag having a very high rainfall value. Such combination indicate 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 the daily rainfall for neighbouring stations considered for the above spatial homogeneity test is as given in Table 9.2.

Upon careful examination it can be seen that at DAKOR station the rainfall recorded for 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 that is 28 July there has been a very large value recorded for DAKOR station whereas the other nearby stations are not experiencing that 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 the 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 feed back 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 Module 10).

Table 9.1: Result of spatial homogeneity test at DAKOR station

Spatial homogeneity check
====================================

Test station DAKOR PH
Start date: 1995 6 1 0 1
End date: 1995 9 30 0 1

Radius of circle of influence : 25.000 (km)
Station weights proportional to : 1/D^2.00

Admissible absolute error : 50.000
Multiplier to stdv of neighbours: 2.000

Selected neighbour stations:

Quadrant	Station		Distance (km)
1	THASARA	PH	8.252
1	VAGHAROLI	PH	18.976
2	MAHISA	PH	13.948
2	KATHLAL	PH	22.216
2	MAHUDHA	PH	22.694
2	SAVLITANK	PH	23.403

Year	mth	day	hr	si	P_obs	flag	P_est	Stdv	n
1995	7	15	0	1	**.00**	-	56.64	20.50	6
1995	7	18	0	1	**.00**	-	8.79	3.34	6
1995	7	19	0	1	**.00**	-	21.24	8.73	6
1995	7	20	0	1	**.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 9.2: Tabulation of daily rainfall for neighbouring stations

Tabulation of series, Year 1995

Year	mth	day	hr	si	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

9.3 Screening for accumulations on holidays and weekends

To screen for accumulated values on holidays and weekends it may be appropriate to prepare a list of all holidays and weekends. Then 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.

10. Missed entries

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 mis-entry. A zero may have been inserted for the day (or days). Similarly, some longer periods may have missed readings without an accumulated value at the end, for example resulting from breakage of the measuring cylinder.

10.2 Data validation procedure and follow up actions

For rainy periods such missed values will be anomalous in the multiple station tabulation and plot and will be indicated by a series of "-ve" 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 10.1:
The spatial homogeneity test for BHEMPODA station (KHEDA catchment) for the year 1997 is carried out. The results of the test are as given below in Table 10.1:

On examining the above tabular result of the test it can be noticed that there are very many instances in succession which are flagged "-ve" and also have nil (0 mm) observed rainfall. At the same time, on these days of "-ve" 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 BHEMPODA the rainfall has either not been observed and wrongly reported as 0 mm or has been observed but has been wrongly reported/entered.

The above suspicion is very strongly corroborated after looking at the tabulation of these neighbouring stations as given in Table 10.2.

It is almost certain that the rainfall at BHEMPODA station has been reported/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 checked with the records of the data at BHEMPODA and if the values available in the records are different then those available in the database then the same must be corrected. Instead, if the manuscript also shows same values then these have to be flagged for necessary correction subsequently using spatial interpolation (see Module 10).

Table 10.1: Result of spatial homogeneity test at BHEMPODA station

```
Spatial homogeneity check
====================================

Test station BHEMPODA    PH

Start date: 1997  6  1  0  1
End   date: 1997  9 30  0  1

Radius of circle of influence   :     25.000 (km)

Station weights proportional to : 1/D^2.00

Admissible absolute error       :     40.000
Multiplier to stdv of neighbours:      2.000

Selected neighbour stations:
Quadrant      Station      Distance (km)

    1          MEGHARAJ    PH     20.897
    2          RAHIOL      PH     17.898
    3          ANIOR       PH      4.535
    3          BAYAD       PH     23.253
```

Year	mth	day	hr	si	P_obs	flag	P_est	Stdv	n
1997	6	9	0	1	9.00	+	.00	.00	4
1997	6	14	0	1	3.00	+	.00	.00	4
1997	6	22	0	1	20.00	+	4.79	2.38	4
1997	6	23	0	1	17.00	+	2.11	4.20	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	.00	-	1.31	.65	4
1997	7	20	0	1	3.00	+	1.34	.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	**.00**	-	20.98	5.32	4
1997	8	9	0	1	.00	-	2.37	.56	4
1997	8	11	0	1	.00	-	.44	.22	4
1997	8	14	0	1	.00	-	2.66	1.14	4
1997	8	19	0	1	**.00**	-	48.96	18.63	4
1997	8	24	0	1	**.00**	-	87.56	42.17	4
1997	9	11	0	1	**.00**	-	18.50	6.03	4
1997	9	13	0	1	**.00**	-	15.36	5.79	4

Table 10.2: Tabulation results for daily rainfall at neighbouring stations

Tabulation of series, Year 1997

Year	mth	day	hr	si	ANIOR	BAYAD	BHEMPODA	MEGHARAJ	RAHIOL
1997	7	25			.0	.0	.0	1.0	.0
1997	7	26			13.0	11.0	1.0	4.0	16.0
1997	7	27			225.0	147.5	125.0	312.0	209.5
1997	7	28			420.5	194.5	280.0	32.5	60.0
1997	7	29			4.0	1.5	3.0	4.5	5.5
1997	7	30			16.5	9.0	13.0	12.0	7.0
1997	7	31			3.0	4.0	3.0	22.0	1.5
1997	8	1			290.0	257.0	275.0	192.5	129.5
1997	8	2			38.5	57.5	94.0	15.0	2.5
1997	8	3			11.5	28.5	24.0	1.0	11.5
1997	8	4			.0	.0	.0	.0	.0
1997	8	5			15.0	.0	23.0	.0	.0
1997	8	6			.0	.0	.0	.0	.0
1997	8	7			1.0	1.5	.0	1.0	.0
1997	8	8			20.5	25.0	**.0**	32.0	18.0
1997	8	9			2.5	2.0	.0	1.0	1.5
1997	8	10			.0	.0	.0	.0	.0
1997	8	11			.5	.0	.0	.0	.0
1997	8	12			4.0	1.0	.0	7.0	8.0
1997	8	13			2.5	6.0	.0	1.0	.0
1997	8	14			3.0	1.0	.0	.5	.0
1997	8	15			.0	.0	.0	1.0	.0
1997	8	16			.0	2.0	.0	.0	.0
1997	8	17			.0	.0	.0	.5	1.0
1997	8	18			.0	.0	.0	.0	3.0
1997	8	19			54.0	33.0	**.0**	12.0	7.0
1997	8	20			.0	7.0	.0	1.0	30.0
1997	8	21			1.0	.0	.0	6.5	.0
1997	8	22			.0	.0	.0	.0	.0
1997	8	23			3.0	.0	.0	9.5	19.5
1997	8	24			91.0	13.5	**.0**	125.5	50.0
1997	8	25			16.5	33.0	**.0**	11.0	31.0
1997	8	26			29.0	19.0	**.0**	54.5	.0
1997	8	27			2.5	5.5	.0	1.0	.0
1997	8	28			.0	.0	.0	.0	.0
1997	8	29			.0	.0	.0	.5	.0
1997	8	30			15.5	33.0	.0	.5	31.5
1997	8	31			.0	.0	.0	.0	.0

11. Rainfall missed on days with low rainfall - rainy days check

11.1 General description:

Whilst it is required that observers inspect the raingauge for rain each day, the practice of some observers may be to visit the gauge only when they know that rain has occurred. This will result in zeros on a number of days on which a small amount of rain has occurred. 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 occurrence of rainfall within the day, 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 later 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 utilise this fact to see if the observed data follow such characteristic. A graphical or tabular comparison of the difference 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 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.

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 is 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 the monthly basis then each tabular or graphical matrix can accommodate a period of one year.

Any glaring 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 much different from 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 further probed. 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 then comparison with other related data like temperature and humidity at the station, if available, can be made. Together with 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 neighbouring stations. Such action implies that there are unreported errors remaining in the time series, which it has 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 observer.

12. Checking for systematic shifts using double mass analyses

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 record contains such inconsistencies which may exist for a considerable period of time. **Inconsistencies present in the rainfall data of a station can occur for various reasons:**

-
- The raingauge 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 raingauge
- The raingauge may have been faulty for a considerable period etc.

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 have been entered to the database.

12.2 Description of method

Double mass analysis is a technique to detect possible inhomogeneities in series by investigating the ratio of accumulated values of two series, viz.:

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

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

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

$$rc_i = \frac{\sum_{j=1}^{i} Y_j}{\sum_{j=1}^{i} X_j}$$

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

$$pc_i = \frac{\sum_{j=1}^{i} Y_j .}{\sum_{j=1}^{N} Y_j} \cdot \frac{\sum_{j=1}^{N} X_j}{\sum_{j=1}^{i} X_j}$$

These ratios in absolute and percent form gives the overall slope of the double mass plot from origin to each consequent 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. 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 had 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.

It must also be emphasised here that the double mass technique is based on the presumption that only a part of the data under consideration is subjected to systematic error. Where the whole length of the data being considered has such an error then the double mass analysis will fail to detect any error.

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 test and 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 Module 10 for details).

Example 12.1:

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. Though the reliability of records and the homogeneity of these base stations have to be ascertained before considering them for the analysis but here it has been assumed that they are reliable stations.

It can be seen from double mass plot of this analysis, as shown in Fig. 12.1, that the data of VADAGAM station is 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°), the plot shows a similar trend throughout the period.

The result of this analysis on yearly basis is given in Table 12.1. 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 are given in the last two columns 8 & 9.

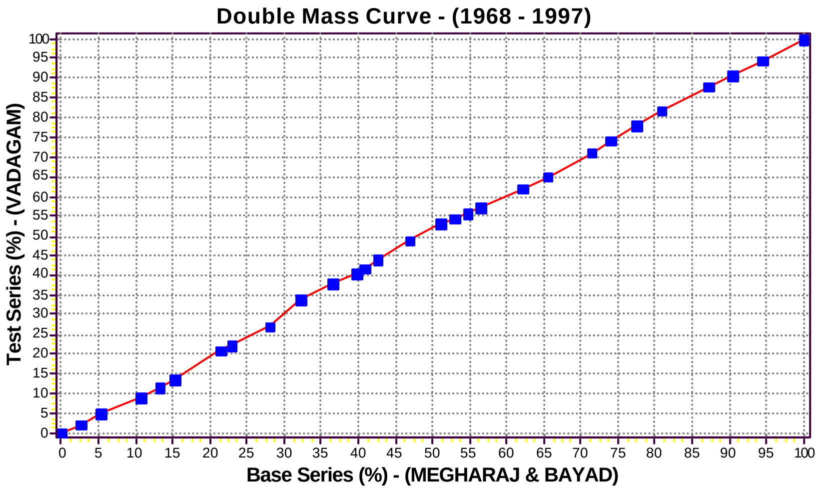

Fig. 12.1: Double mass plot showing near consistent trend at test station

Table 12.1: Analysis result of the double mass analysis

```
Test series: VADAGAM      PH
                                Weight
 Base series: MEGHARAJ     PH      .50
              BAYAD        PH      .50
```

1	2	3	4	5	6	7	8	9
Period	BASE			TEST			Ratios	
	Amount	Cum	Perc	Amount	Cum	Perc	(6)/(3)	(7)/(4)
1968	451.5	452.	2.5	382.4	382.	2.2	.85	.88
1969	487.5	939.	5.3	437.0	819.	4.8	.87	.90
1970	957.4	1896.	10.7	743.1	1563.	9.1	.82	.85
1971	462.3	2359.	13.3	443.4	2006.	11.7	.85	.88
1972	332.1	2691.	15.2	339.1	2345.	13.7	.87	.90
1973	1124.8	3816.	21.5	1266.3	3611.	21.0	.95	.98
1974	247.8	4063.	22.9	214.9	3826.	22.3	.94	.97
1976	910.2	4974.	28.0	831.6	4658.	27.1	.94	.97
1977	751.0	5725.	32.3	1124.1	5782.	33.7	1.01	1.04
1978	735.0	6460.	36.4	748.2	6530.	38.0	1.01	1.05
1979	576.0	7036.	39.6	389.1	6919.	40.3	.98	1.02
1980	205.3	7241.	40.8	234.3	7154.	41.7	.99	1.02
1982	323.6	7565.	42.6	417.7	7571.	44.1	1.00	1.03
1983	766.3	8331.	46.9	817.4	8389.	48.9	1.01	1.04
1984	737.8	9069.	51.1	737.0	9126.	53.2	1.01	1.04
1985	312.4	9381.	52.8	198.4	9324.	54.3	.99	1.03
1986	313.8	9695.	54.6	229.6	9554.	55.7	.99	1.02
1987	337.3	10032.	56.5	261.9	9816.	57.2	.98	1.01
1988	986.0	11018.	62.1	837.7	10653.	62.1	.97	1.00
1989	605.8	11624.	65.5	493.0	11146.	64.9	.96	.99
1990	1047.8	12672.	71.4	1065.5	12212.	71.1	.96	1.00
1991	481.0	13153.	74.1	508.5	12720.	74.1	.97	1.00
1992	596.8	13749.	77.5	697.0	13417.	78.2	.98	1.01
1993	598.0	14347.	80.8	599.0	14016.	81.7	.98	1.01
1994	1101.0	15448.	87.0	1079.5	15096.	87.9	.98	1.01
1995	592.5	16041.	90.4	478.5	15574.	90.7	.97	1.00
1996	746.8	16788.	94.6	647.6	16222.	94.5	.97	1.00
1997	963.0	17751.	100.0	944.0	17166.	100.0	.97	1.00

Total number of periods analysis: 28

Example 12.2:

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 shows 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 Fig. 12.2, 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 12.2 showing the results of the test.

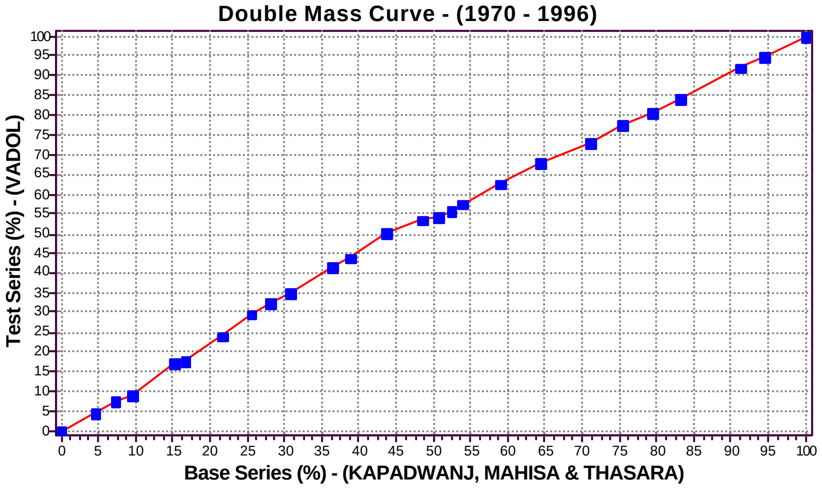

Fig. 12.2: Double mass curve for VADOL station showing significant change of slope of the curve after about half the period under consideration.

Table 12.2: 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 MM	Cum MM	Perc	Amount MM	Cum MM	Perc	(6)/(3) -	(7)/(4) -
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

It is amply clear that from the year 1985 onwards the test station, i.e. VADOL, started receiving rainfall which is comparatively lesser than what it use to receive before that time. And this change in behaviour is not short lived but is continuos thereafter. The reasons for such variations are required to be ascertained. Various factors which could results 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 meteorologic factors around the test station (which is **very unlikely** since any meteorologic changes would generally be spread wide enough to cover all and more neighbouring stations). For both the possibilities the reasons have to be identified beyond doubt before any corrective measure can be taken. A visit to the station and looking at the exposure conditions etc. and taking the history from the observer will be very useful in trying to establish the reasons of this change in the behaviour.