Government of India & Government of The Netherlands

Technical Assistance
HYDROLOGY PROJECT

DHV CONSULTANTS &
DELFT HYDRAULICS with
HALCROW, TAHAL, CES,
ORG & JPS

VOLUME 8
DATA PROCESSING AND ANALYSIS

OPERATION MANUAL - PART II

SECONDARY VALIDATION

Table of Contents

1 INTRODUCTION 1

1.1 GENERAL 1

2 SECONDARY VALIDATION OF RAINFALL DATA 2

2.1 GENERAL 2
2.2 SCREENING OF DATA SERIES 6
2.3 SCRUTINY BY MULTIPLE TIME SERIES GRAPHS 7
2.4 SCRUTINY BY TABULATIONS OF DAILY RAINFALL SERIES OF MULTIPLE STATIONS 8
2.5 CHECKING AGAINST DATA LIMITS FOR TOTALS AT LONGER DURATIONS 10
2.5.1 GENERAL DESCRIPTION 10
2.5.2 DATA VALIDATION PROCEDURE AND FOLLOW UP ACTIONS 10
2.6 SPATIAL HOMOGENEITY TESTING OF RAINFALL (NEAREST NEIGHBOUR ANALYSIS) 12
2.6.1 GENERAL DESCRIPTION 12
2.6.2 DATA VALIDATION PROCEDURE AND FOLLOW UP ACTIONS 12
2.7 IDENTIFICATION OF COMMON ERRORS 16
2.8 CHECKING FOR ENTRIES ON WRONG DAYS - SHIFTED ENTRIES 16
2.8.1 GENERAL DESCRIPTION 16
2.8.2 DATA VALIDATION PROCEDURE AND FOLLOW UP ACTIONS 17
2.9 ENTRIES MADE AS ACCUMULATIONS 19
2.9.1 GENERAL DESCRIPTION 19
2.9.2 DATA VALIDATION PROCEDURE AND FOLLOW UP ACTIONS 20
2.9.3 SCREENING FOR ACCUMULATIONS ON HOLIDAYS AND WEEKENDS 22
2.10 MISSED ENTRIES 22
2.10.1 GENERAL DESCRIPTION 22
2.10.2 DATA VALIDATION PROCEDURE AND FOLLOW UP ACTIONS 22
2.11 RAINFALL MISSED ON DAYS WITH LOW RAINFALL – RAINY DAYS CHECK 24
2.11.1 GENERAL DESCRIPTION 24
2.11.2 DATA VALIDATION PROCEDURE AND FOLLOW UP ACTIONS 25
2.12 CHECKING FOR SYSTEMATIC SHIFTS USING DOUBLE MASS ANALYSES 25
2.12.1 GENERAL DESCRIPTION 25
2.12.2 DESCRIPTION OF METHOD 26
2.12.3 DATA VALIDATION PROCEDURE AND FOLLOW UP ACTIONS 27

3 CORRECTION AND COMPLETION OF RAINFALL DATA 30

3.1 GENERAL 30
3.2 USE OF ARG AND SRG DATA AT ONE OR MORE STATIONS 30
3.2.1 GENERAL DESCRIPTION 30
3.2.2 DATA CORRECTION OR COMPLETION PROCEDURE 30
3.3 CORRECTING FOR ENTRIES TO WRONG DAYS 35
3.3.1 GENERAL DESCRIPTION 35
3.3.2 DATA CORRECTION PROCEDURE 35
3.4 APPORTIONMENT FOR INDICATED AND UNINDICATED ACCUMULATIONS 37
3.4.1 GENERAL DESCRIPTION 37
3.4.2 DATA CORRECTION PROCEDURE 37
3.5 ADJUSTING RAINFALL DATA FOR LONG TERM SYSTEMATIC SHIFTS 39
3.5.1 GENERAL DESCRIPTION 39
3.5.2 DATA CORRECTION PROCEDURE 40
3.6 USING SPATIAL INTERPOLATION TO INTERPOLATE ERRONEOUS AND MISSING VALUES 43
3.6.1 GENERAL DESCRIPTION 43
3.6.2 ARITHMETIC AVERAGE METHOD 43
3.6.3 NORMAL RATIO METHOD 45
3.6.4 DISTANCE POWER METHOD 47

4 COMPILATION OF RAINFALL DATA 51

4.1 GENERAL 51
4.2 AGGREGATION OF DATA TO LONGER DURATIONS 51
4.2.1 AGGREGATION OF DAILY TO WEEKLY 51
4.2.2 AGGREGATION OF DAILY TO TEN DAILY 51
4.2.3 AGGREGATION FROM DAILY TO MONTHLY 52
4.2.4 HOURLY TO OTHER INTERVALS 52
4.3 ESTIMATION OF AREAL RAINFALL 54
4.3.1 GENERAL DESCRIPTION 54
4.3.2 ARITHMETIC AVERAGE 55
4.3.3 WEIGHTED AVERAGE USING USER DEFINED WEIGHTS 55
4.3.4 THIESSEN POLYGON METHOD 56
4.3.5 ISOHYETAL AND RELATED METHODS 59
4.3.6 KRIGING 63
4.4 TRANSFORMATION OF NON-EQUIDISTANT TO EQUIDISTANT SERIES 75
4.5 COMPILATION OF MINIMUM, MAXIMUM AND MEAN SERIES 75

5 SECONDARY VALIDATION OF CLIMATIC DATA 77

5.1 GENERAL 77
5.2 METHODS OF SECONDARY VALIDATION 78
5.3 SCREENING OF DATA SERIES 78
5.4 MULTIPLE STATION VALIDATION 79
5.4.1 COMPARISON PLOTS 79
5.4.2 BALANCE SERIES 79
5.4.3 DERIVATIVE SERIES 79
5.4.4 REGRESSION ANALYSIS 80
5.4.5 DOUBLE MASS CURVES 80
5.4.6 SPATIAL HOMOGENEITY (NEAREST NEIGHBOUR ANALYSIS) 80
5.5 SINGLE SERIES TESTS OF HOMOGENEITY 81
5.5.1 TREND ANALYSIS (TIME SERIES PLOT) 81
5.5.2 RESIDUAL MASS CURVE 82
5.5.3 A NOTE ON HYPOTHESIS TESTING 82
5.5.4 STUDENT'S T TESTS OF STABILITY OF THE MEAN 82
5.5.5 WILCOXON W-TEST ON THE DIFFERENCE OF MEANS 83
5.5.6 WILCOXON MANN-WHITNEY U-TEST 84

6 CORRECTION AND COMPLETION OF CLIMATIC DATA 85

6.1 GENERAL 85
6.2 TEMPERATURE 85
6.3 HUMIDITY 86
6.4 WIND 87
6.5 ATMOSPHERIC RADIATION 87
6.6 SOLAR RADIATION 87
6.7 PAN EVAPORATION 87

7 SECONDARY VALIDATION OF WATER LEVEL DATA 88

7.1 GENERAL 88
7.2 SCRUTINY OF MULTIPLE HYDROGRAPH PLOTS 88
7.3 COMBINED HYDROGRAPH AND RAINFALL PLOTS 92
7.4 RELATION CURVES FOR WATER LEVEL 93
7.4.1 GENERAL 93
7.4.2 APPLICATION OF RELATION CURVES TO WATER LEVEL 94
7.4.3 DETERMINATION OF TRAVEL TIME 95
7.4.4 FITTING THE RELATION CURVE 97
7.4.5 USING RELATION CURVE FOR DATA VALIDATION 98

8 CORRECTION AND COMPLETION OF WATER LEVEL DATA 100

8.1 GENERAL 100
8.2 CORRECTION USING RIVER LEVEL OR DISCHARGE? 101

8.3 COMPARISON OF STAFF GAUGE AND AUTOGRAPHIC OR DIGITAL RECORDS 102
  8.3.1 OBSERVER ERRORS 102
  8.3.2 RECORDER TIMING ERRORS 102
  8.3.3 PEN LEVEL ERRORS 103
  8.3.4 ERRORS ARISING FROM STILLING WELL AND INTAKE PROBLEMS 103
  8.3.5 MISCELLANEOUS INSTRUMENT FAILURES 103
8.4 LINEAR INTERPOLATION OF SHORT GAPS 105
8.5 USE OF RELATION CURVES WITH ADJACENT STATIONS 106
  8.5.1 GENERAL 106
  8.5.2 INFILLING OF MISSING RECORDS 106
  8.5.3 IDENTIFYING AND CORRECTING MISREADINGS 106
  8.5.4 IDENTIFYING AND CORRECTING SHIFT IN GAUGE ZERO OR CHANGE IN CROSS SECTION 107

9 ESTABLISHMENT OF STAGE DISCHARGE RATING CURVE 107

9.1 GENERAL 107
9.2 THE STATION CONTROL 108
  9.2.1 TYPES OF STATION CONTROL 109
9.3 FITTING OF RATING CURVES 112
  9.3.1 GENERAL 112
  9.3.2 FITTING OF SINGLE CHANNEL SIMPLE RATING CURVE 115
  9.3.3 COMPOUND CHANNEL RATING CURVE 123
  9.3.4 RATING CURVE WITH BACKWATER CORRECTION 124
  9.3.5 RATING CURVE WITH UNSTEADY FLOW CORRECTION 129
  9.3.6 RATING RELATIONSHIPS FOR STATIONS AFFECTED BY SHIFTING CONTROL 133

10 VALIDATION OF RATING CURVE 139

10.1 GENERAL 139
10.2 GRAPHICAL VALIDATION TESTS 139
  10.2.1 GENERAL 139
  10.2.2 STAGE/DISCHARGE PLOT WITH NEW GAUGINGS 140
  10.2.3 PERIOD/FLOW DEVIATION SCATTERGRAM 142
  10.2.4 STAGE/FLOW DEVIATION DIAGRAM 143
  10.2.5 CUMULATIVE DEVIATION PLOT OF GAUGINGS 143
  10.2.6 STAGE DISCHARGE PLOTS WITH GAUGINGS DISTINGUISHED BY SEASON. 144
10.3 NUMERICAL VALIDATION TESTS 144
  10.3.1 USE OF STUDENT'S 'T' TEST TO CHECK GAUGINGS 144
  10.3.2 TEST FOR ABSENCE FROM BIAS IN SIGNS 146
  10.3.3 TEST FOR ABSENCE FROM BIAS IN VALUES 146
  10.3.4 GOODNESS OF FIT TEST 147

11 EXTRAPOLATION OF RATING CURVE 148

11.1 GENERAL 148
11.2 HIGH FLOW EXTRAPOLATION 148
  11.2.1 THE DOUBLE LOG PLOT METHOD 148
  11.2.2 STAGE-AREA / STAGE-VELOCITY METHOD 150
  11.2.3 THE MANNING'S EQUATION METHOD 151
  11.2.4 THE CONVEYANCE SLOPE METHOD 153
11.3 LOW FLOW EXTRAPOLATION 157

12 SECONDARY VALIDATION OF STAGE DISCHARGE DATA 158

12.1 GENERAL 158
12.2 REVIEW OF RATING CURVE ON THE BASIS OF BALANCES 158
12.3 REVIEW OF RATING CURVE ON THE BASIS OF DOUBLE MASS ANALYSIS 162
12.4 REVIEW OF RATING CURVE ON THE BASIS OF RELATION CURVES BETWEEN STAGES AT ADJACENT STATIONS 163

13 COMPUTATION OF DISCHARGE DATA 164

13.1 GENERAL 164
13.2 STATION REVIEW 164
13.3 TRANSFORMATION OF STAGE TO DISCHARGE 164
13.3.1 SINGLE CHANNEL RATING CURVE 165
13.3.2 COMPOUND CHANNEL RATING CURVE 165
13.3.3 RATING CURVE WITH UNSTEADY FLOW CORRECTION 166
13.3.4 RATING CURVE WITH CONSTANT FALL BACKWATER CORRECTION 166
13.3.5 RATING CURVE WITH NORMAL FALL BACKWATER CORRECTION 167

14 SECONDARY VALIDATION OF DISCHARGE DATA 167

14.1 GENERAL 167
14.2 SINGLE STATION VALIDATION 168
14.2.1 VALIDATION AGAINST DATA LIMITS 168
14.2.2 GRAPHICAL VALIDATION 169
14.2.3 VALIDATION OF REGULATED RIVERS 171
14.3 MULTIPLE STATION VALIDATION 172
14.3.1 COMPARISON PLOTS 172
14.3.2 RESIDUAL SERIES 172
14.3.3 DOUBLE MASS CURVES 173
14.4 COMPARISON OF STREAMFLOW AND RAINFALL 174

15 CORRECTION AND COMPLETION OF DISCHARGE DATA 174

15.1 CORRECTION OF DISCHARGE DATA 174
15.2 COMPLETION OF DISCHARGE DATA 175

16 COMPILATION OF DISCHARGE DATA 175

16.1 GENERAL 175
16.2 AGGREGATION OF DATA TO LONGER DURATION 176
16.3 COMPUTATION OF VOLUMES AND RUNOFF DEPTH 176
16.4 COMPILATION OF MAXIMUM AND MINIMUM SERIES 178

17 SECONDARY VALIDATION OF SUSPENDED SEDIMENT CONCENTRATIONS 178

17.1 GENERAL 178
17.2 SINGLE FRACTION, MULTIPLE SEASONS, SINGLE YEAR 178
17.3 SINGLE YEAR, MULTIPLE FRACTIONS 181
17.4 MULTIPLE YEARS, SINGLE FRACTIONS 182

18 COMPILATION OF SEDIMENT LOADS 184

18.1 SEDIMENT TRANSPORT 184
18.2 STEPS FOR SUSPENDED LOAD INCLUDE: 185

19 REFERENCES 188

1 INTRODUCTION

1.1 GENERAL

The prime objective of the Hydrology Project is to develop a sustainable Hydrological Information System for 9 states in Peninsular India, set up by the state Surface Water and Groundwater Departments and by the central agencies (CWC and CGWB) with the following characteristics:

- Demand driven, i.e. output is tuned to the user needs
- Use of standardised equipment and adequate procedures for data collection and processing
- Computerised, comprehensive and easily accessible database
- Proper infrastructure to ensure sustainability.

This Hydrological Information System provides information on the spatial and temporal characteristics of water quantity and quality variables/parameters describing the water resources/water use system in Peninsular India. The information needs to be tuned and regularly be re-tuned to the requirements of the decision/policy makers, designers and researchers to be able to take decisions for long term planning, to design or to study the water resources system at large or its components.

This manual describes the procedures to be used to arrive at a sound operation of the Hydrological Information System as far as hydro-meteorological and surface water quantity and quality data are concerned. A similar manual is available for geo-hydrological data. This manual is divided into three parts:

A. **Design Manual,** which provides information for the design activities to be carried out for the further development of the HIS
B. **Reference Manual,** including references and additional information on certain topics dealt with in the Design Manual
C. **Field/Operation Manual,** which is an instruction book describing in detail the activities to be carried out at various levels in the HIS, in the field and at the data processing and data storage centres.

The manual consists of ten volumes, covering:

1. Hydrological Information System, its structure and data user needs assessment
2. Sampling Principles
3. Hydro-meteorology
4. Hydrometry
5. Sediment transport measurements
6. Water Quality sampling
7. Water Quality analysis
8. Data processing
9. Data transfer, storage and dissemination, and
10. SW-Protocols.

This Volume 8 deals with **data processing** and consists of an Operation Manual and a Reference Manual. The Operation Manual comprises 4 parts, viz:

Part I: Data entry and primary validation

Part II: Secondary validation

Part III: Final processing and analysis

Part IV: Data management

This Part II concerns the second step in data processing, i.e. the secondary data validation, which is executed in the Divisions. The procedures described in the manual have to be applied to ensure uniformity in data processing throughout the Project Area and to arrive at high quality data.

2 SECONDARY VALIDATION OF RAINFALL DATA

2.1 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 2.1:

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 Figure 2.1, Figure 2.2 and Figure 2.3 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.

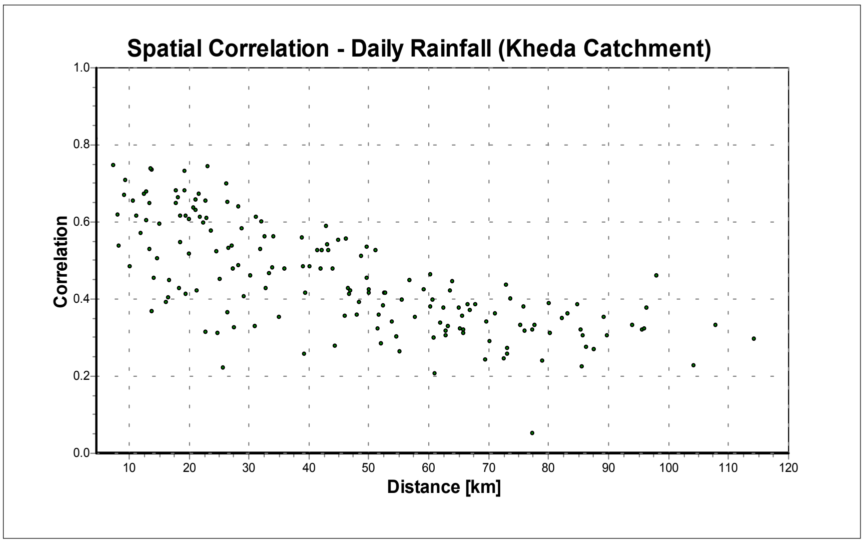


Figure 2.1: Plot of correlation with distance for daily rainfall data

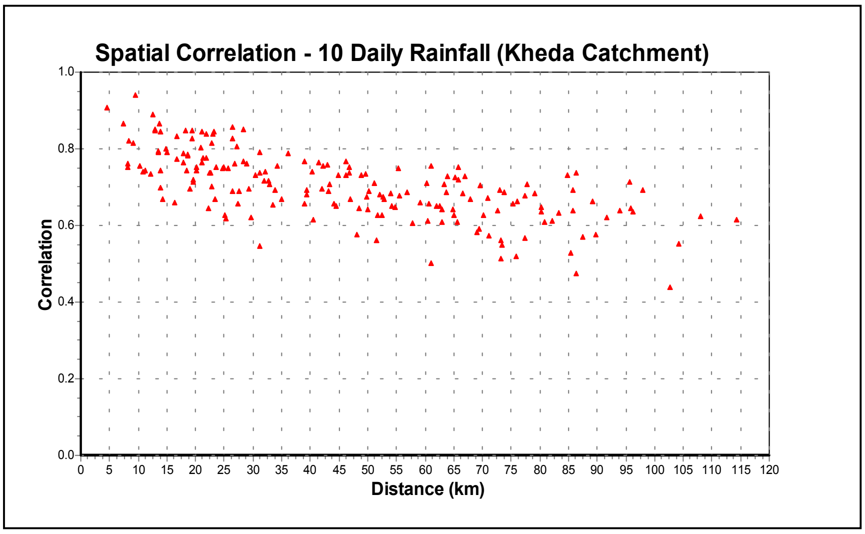


Figure 2.2: Plot of correlation with distance for ten-daily rainfall data

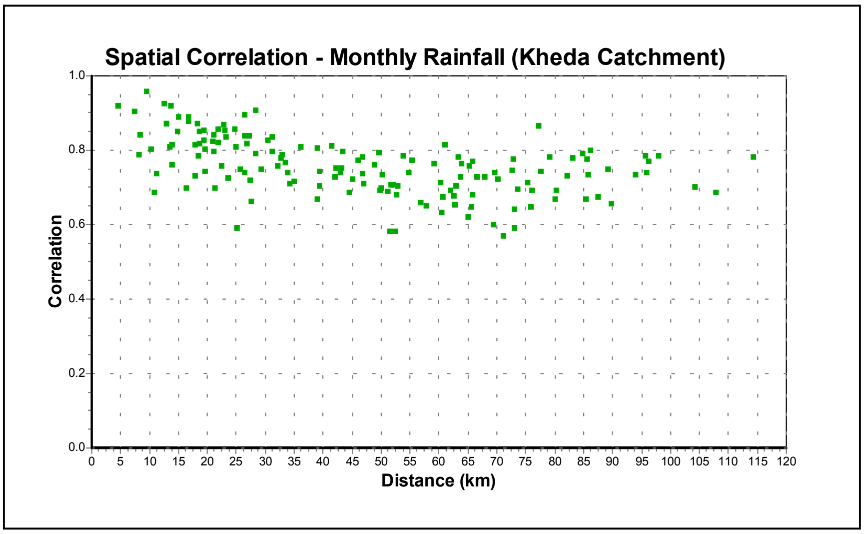


Figure 2.3: Plot of correlation with distance for monthly rainfall data

Example 2.2

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

Figure 2.4 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, Figure 2.5). 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, Figure 2.6 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 Figure 2.7).

It is apparent from Figure 2.6 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 Figure 2.6 in which lot of scatter is shown even for smaller inter station distances.

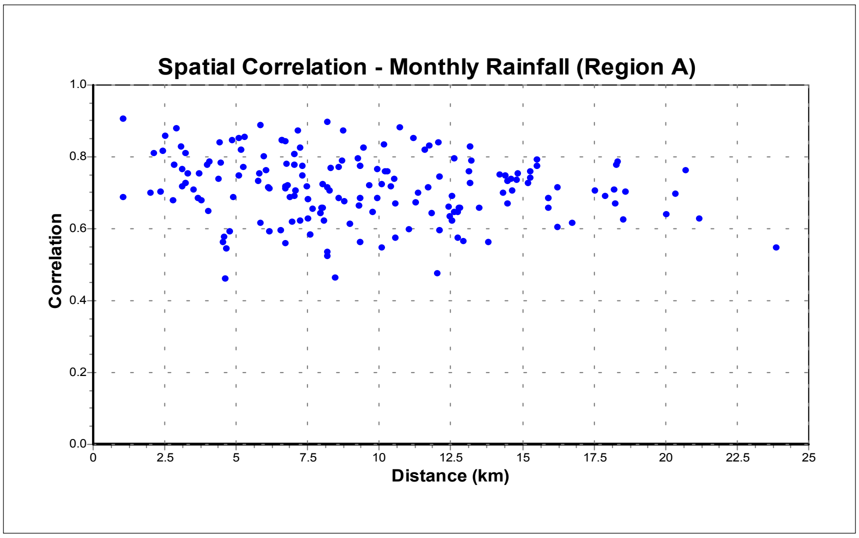


Figure 2.4: Scatter plot of correlation for monthly rainfall in the small hilly region

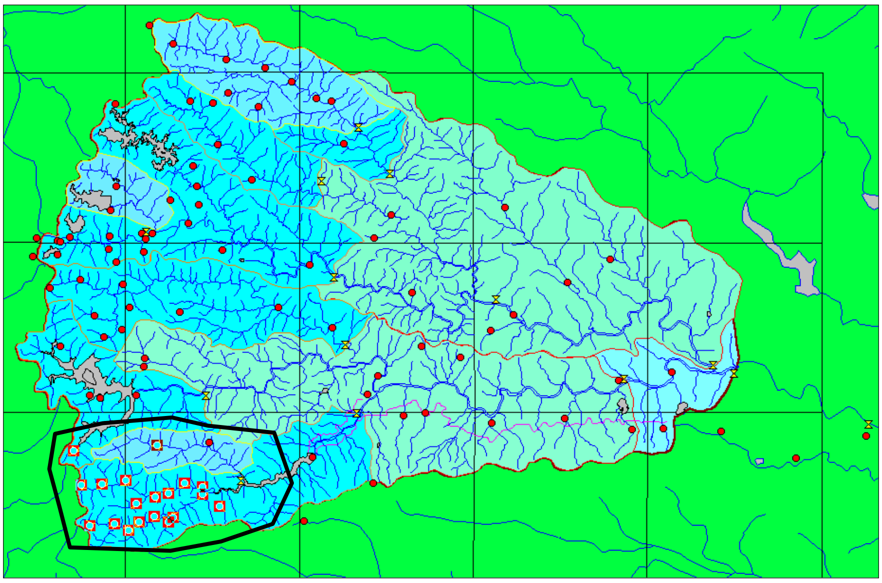


Figure 2.5: Selection of a group of some 20 stations in the hilly region of the catchment.

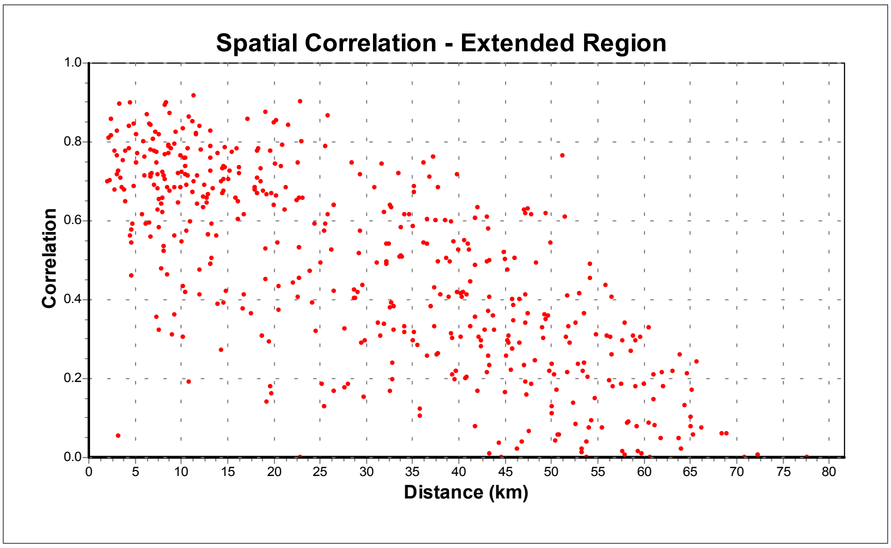


Figure 2.6: Scatter plot of correlation for monthly rainfall in the extended region

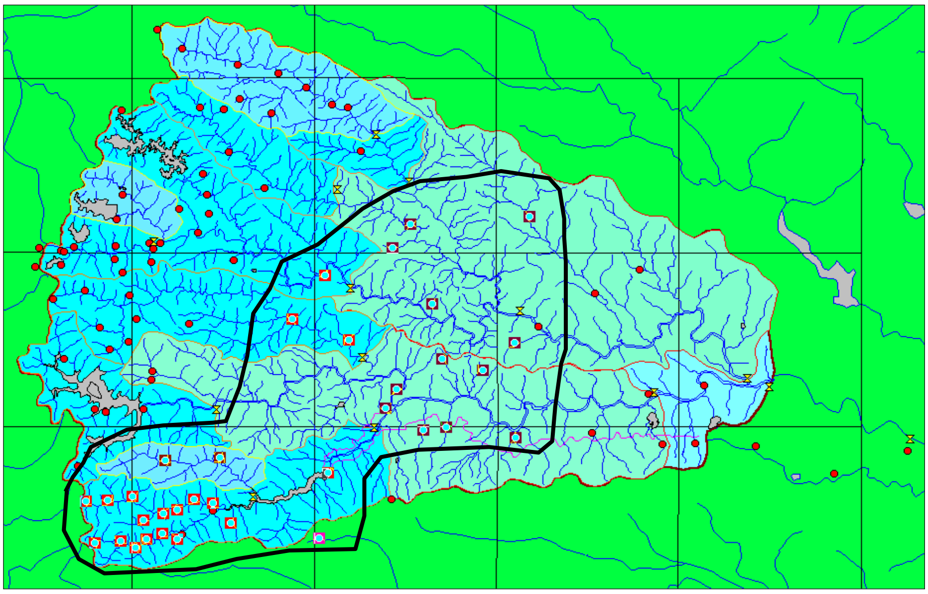


Figure 2.7: Selection of a group of some 34 stations in the extended region of the catchment

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 Part I.

2.2 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.

Example 2.3

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.

```
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 -
```

Table 2.1: Result of the screening process of daily rainfall data for one year

2.3 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.

Example 2.4

Consider that a few of the higher values at ANIOR station of KHEDA catchment during July and August 1996 are suspect. Comparison with adjoining available stations BHEMPODA, RELLAWADA and MEGHARAJ is made for this purpose. Figure 2.8 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.

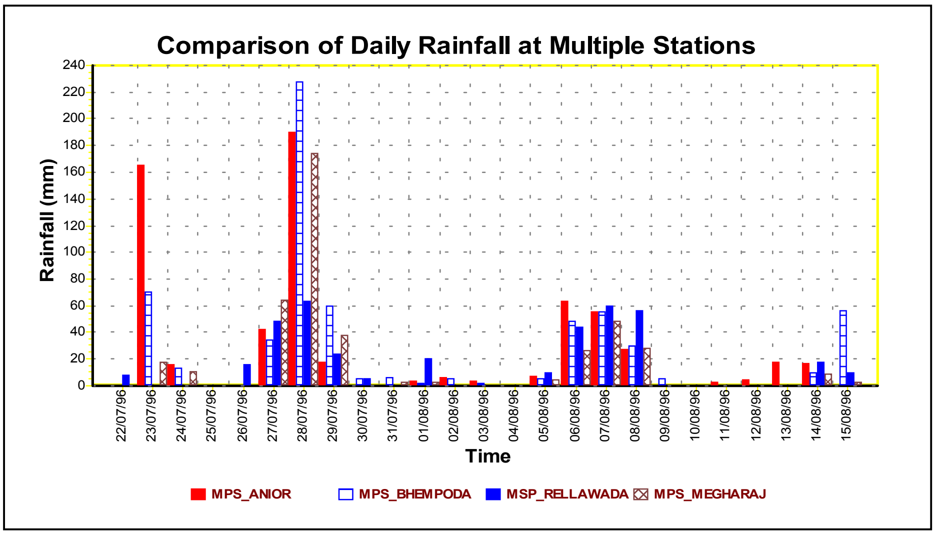


Figure 2.8: Comparison of multiple time series plot of daily rainfall data

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

2.4 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 2.5

As a routine process of scrutinising daily data for a common error of time shift in one or more data series, consider KAPADWANJ, KATHLAL, MAHISA, SAVLITANK and VADOL stations of KHEDA catchment. These stations are within a circle of 25 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 2.2 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.

```
Tabulation of series, Year  1984
                                   ==========Data==========
 Year mth day hr si    KAPADWANJ KATHLAL   MAHISA    SAVLITANK VADOL
                          PH        PH        PH        PH        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
```

Table 2.2: Tabulation for scrutiny of possible error in the timing of daily rainfall data

2.5 CHECKING AGAINST DATA LIMITS FOR TOTALS AT LONGER DURATIONS

2.5.1 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.

2.5.2 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.

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.

Example 2.6

The plot of these yearly values is shown in Figure 2.9. 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.

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 2.3 for an easier comparison.

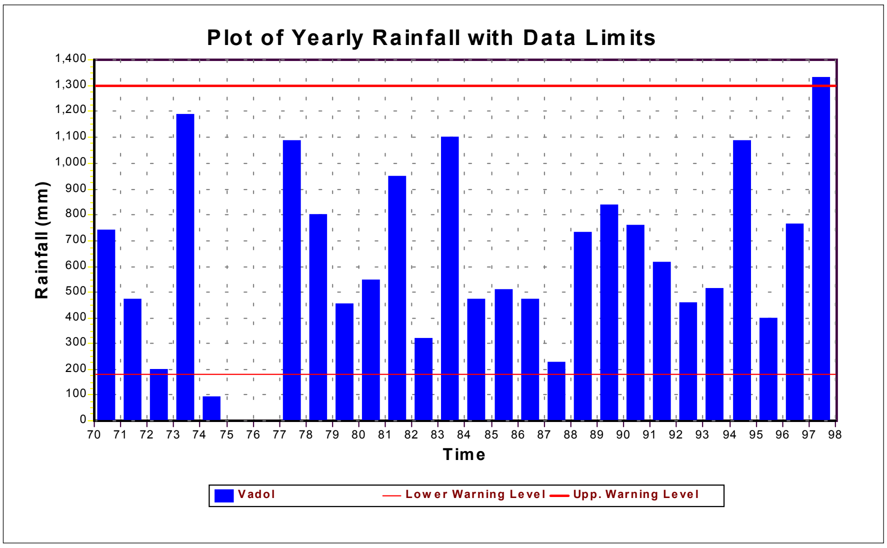


Figure 2.9: Plot of rainfall data compiled at an earlier interval

```
Tabulation of series, Year  1970 - 1997
                                   ==========Data==========
 Year mth day hr si    BALASINOR KAPADWANJ SAVLITANK VADOL     VAGHAROLI
                          MPS       MPS       MPS       MPS       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
```

Table 2.3: Tabulation of yearly rainfall at five neighbouring stations

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 2.4 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.

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

Table 2.4: Tabulation of daily rainfall at VADOL station.

2.6 SPATIAL HOMOGENEITY TESTING OF RAINFALL (NEAREST NEIGHBOUR ANALYSIS)

2.6.1 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.

2.6.2 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 Figure 2.10):

(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:

$$P_{est}(t) = \frac{\sum_{i=1}^{N} P_i(t)/D_i^{\,b}}{\sum_{i=1}^{N} 1/D_i^{\,b}} \qquad (2.1)$$

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

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

D_i = distance between the test and the neighbouring station i

N = number of neighbouring stations taken into account.

b = power of distance D

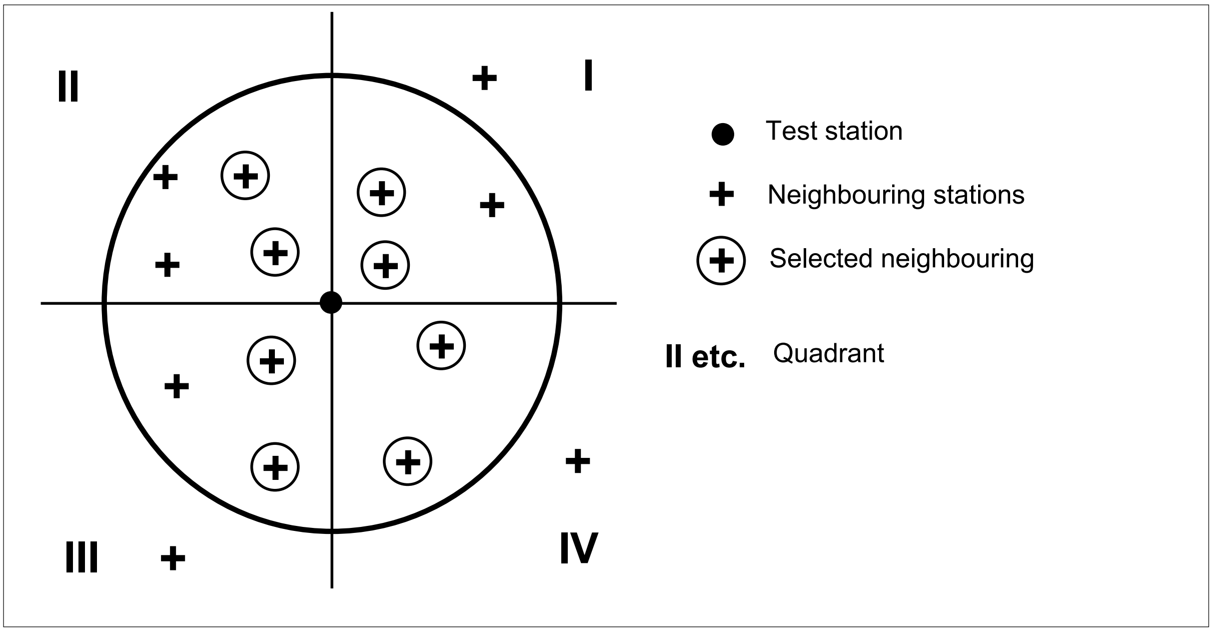


Figure 2.10: Definition sketch of Test and Base (neighbouring) stations

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:

$$\left| P_{obs}(t) - P_{est}(t) \right| \leq X_{abs}$$
$$\left| P_{obs}(t) - P_{est}(t) \right| \leq X_{rel} * S_{Pest}(t) \qquad (2.2)$$

where: X_{abs} = admissible absolute difference

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

X_{rel} = multiplier of standard deviation and

$$S_{P_{est}(t)} = \sqrt{\sum_{i=1}^{N} (P_i(t) - \overline{P}_i(t))^2} \qquad (2.3)$$

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.

Example 2.7

A test is performed for reviewing the spatial homogeneity of the daily rainfall data at SAVLITANK station in KHEDA catchment. An area within a radius of 25 kms. around SAVLITANK station is considered for selecting the base stations (see Figure 2.11). 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 2.5.

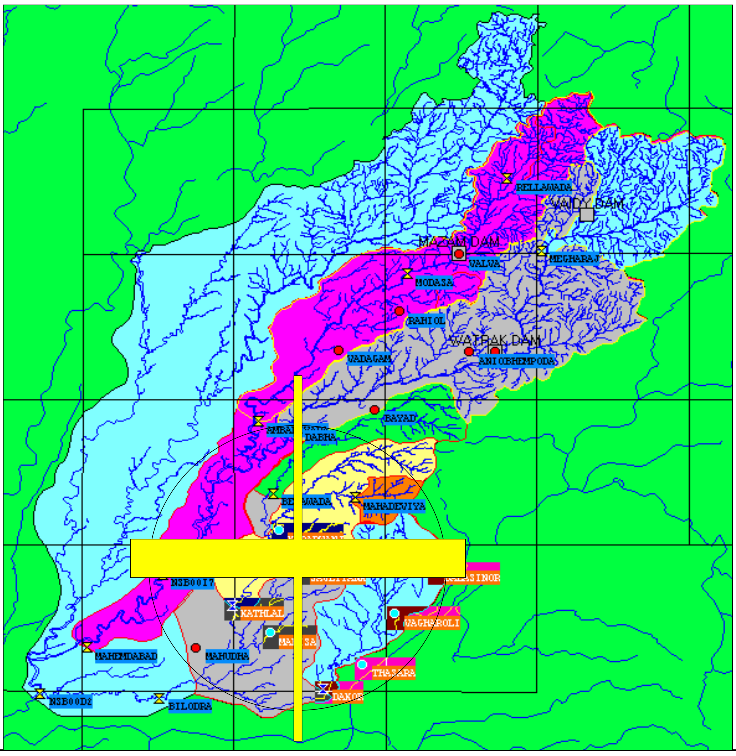


Figure 2.11: Selection of test station SAVLITANK and neighbouring base stations

```
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
```

Table 2.5: Results of the spatial homogeneityest

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.

2.7 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 2.3 and 2.4)
- Spatial homogeneity tests (Section 2.6)

Typical errors are:

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

2.8 CHECKING FOR ENTRIES ON WRONG DAYS - SHIFTED ENTRIES

2.8.1 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.

2.8.2 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.

Example 2.8

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 2.6 below:

```
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
```

Table2.6: Result of the spatial homogeneity test at VADAGAM station.

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 2.7. 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.

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

```
Tabulation of series, Year  1988
                                   ==========Data==========
 Year mth day hr si    ANIOR     BAYAD    MODASA    RAHIOL   VADAGAM
                          PH        PH        PH        PH        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 2.7: Tabulation of daily rainfall at neighbouring stations.

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 2.8):

```
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
```

Table 2.8: Results of the spatial homogeneity test on the corrected series

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

2.9 ENTRIES MADE AS ACCUMULATIONS

2.9.1 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.

2.9.2 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:

$$P_{appor,i} = \frac{P_{est,i} * P_{tot}}{\sum_{i=1}^{N_{acc}} P_{est,i}} \qquad (2.4)$$

where: P_{tot} = accumulated rainfall as recorded

Nacc = number of days of accumulation

Pest,I = estimated daily rainfalls during the period of accumulation on the basis of adjoining stations

Pappor,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 2.9

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 2.9):

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 2.10.

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 Chapter 3)

```
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 2.9: Result of spatial homogeneity test at DAKOR station

```
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
```

Table 2.10: Tabulation of daily rainfall for neighbouring stations

2.9.3 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.

2.10 MISSED ENTRIES

2.10.1 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.

2.10.2 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 2.10

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 2.11:

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 2.12.

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 Chapter 3).

```
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 2.11: Result of spatial homogeneity test at BHEMPODA station

```
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
```

Table 2.12: Tabulation results for daily rainfall at neighbouring stations

2.11 RAINFALL MISSED ON DAYS WITH LOW RAINFALL – RAINY DAYS CHECK

2.11.1 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.

2.11.2 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.

2.12 CHECKING FOR SYSTEMATIC SHIFTS USING DOUBLE MASS ANALYSES

2.12.1 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.

2.12.2 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:

$$rc_i = \frac{\sum_{j=1}^{i} Y_j}{\sum_{j=1}^{i} X_j} \tag{2.5}$$

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

$$pc_i = \frac{\sum_{j=1}^{i} Y_j}{\sum_{j=1}^{N} Y_j} \cdot \frac{\sum_{j=1}^{N} X_j}{\sum_{j=1}^{i} X_j} \tag{2.6}$$

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.

2.12.3 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 Chapter 3 for details).

Example 2.11

Double mass analysis for VADAGAM station (in KHEDA catchment) is carried out considering two stations MEGHARAJ and BAYAD as the base stations for the period from 1968 to 1996. A period of only three months from July to September (92 days) has been taken into consideration while carrying out the analysis. 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 Figure 2.12, 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 2.13 The yearly rainfall and the rainfall accumulated in time for the base and test station is given in columns 2, 3 and 5, 6 respectively. These cumulative rainfall values are then expressed in percent form in columns 4 and 7 respectively. The ratio of these cumulated values in absolute in percent form are given in the last two columns 8 & 9.

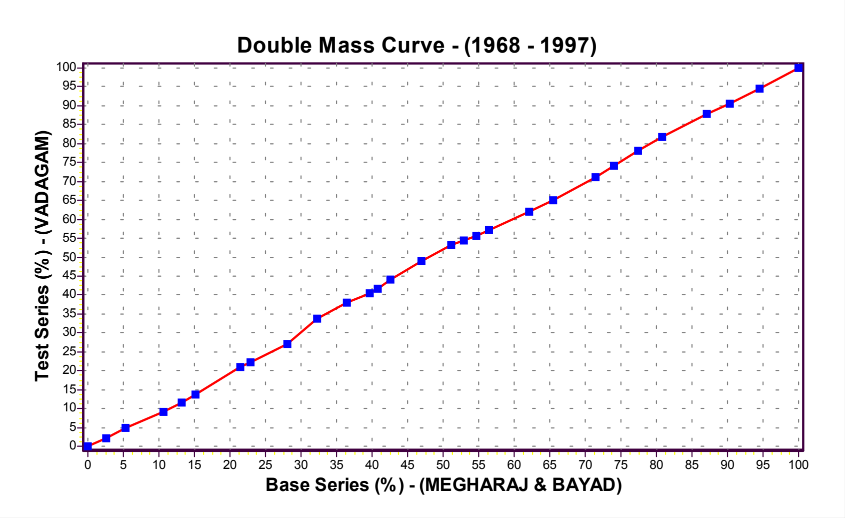


Figure 2.12: Double mass plot showing near consistent trend at test station

```
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
```

Table 2.13: Analysis result of the double mass analysis

Example 2.12

The long term data series of rainfall for the period 1970 to 1996 is considered at VADOL station (in KHEDA catchment) for double mass analysis taking three nearby stations KAPADWANJ, MAHISA and THASARA. Unlike the previous example, which is a case of the test station being homogeneous in time, this example illustrates a case where the test station records 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 Figure 2.13, 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 2.14 showing the results of the test.

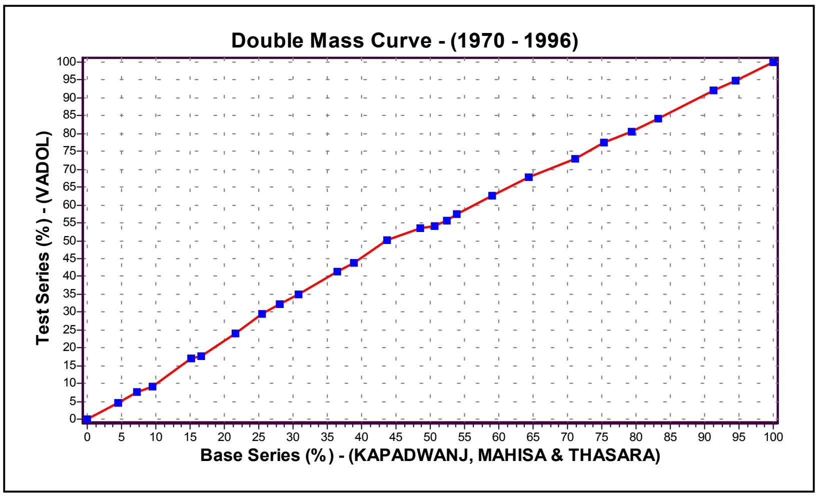


Figure 2.13: Double mass curve for VADOL station showing significant change of slope of the curve after about half the period under consideration.

```
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
```

Table 2.14: Results of the double mass analysis

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 result in such a change can be: (a) a systematic error in the observation of rainfall after the year 1983 or (b) a possible change in the 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.

3 CORRECTION AND COMPLETION OF RAINFALL DATA

3.1 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.

3.2 USE OF ARG AND SRG DATA AT ONE OR MORE STATIONS

3.2.1 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.

3.2.2 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:

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 3.1

Referring back to Example 2.4 of Chapter 2 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 (12th & 13th 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 3.1.

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

Table 3.1: Tabulation result for daily rainfall series obtained from SRG & ARG.

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.

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:

$$H_{test,i} = D_{test} \cdot \frac{H_{base,i}}{\sum_{i=1}^{24} H_{base,i}} \qquad (3.1)$$

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

Example 3.2

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.

Figure 3.1 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.

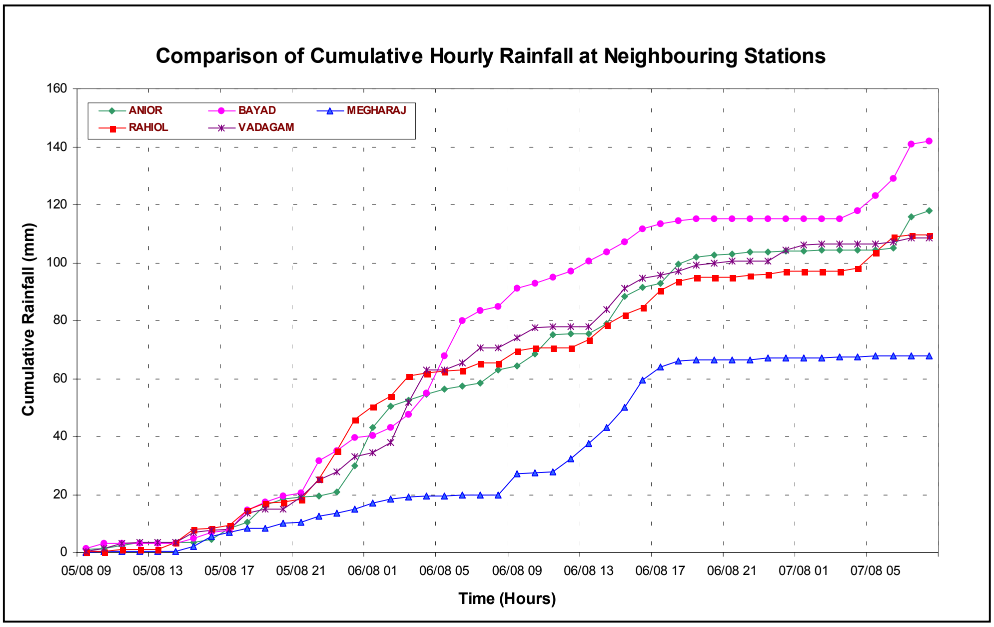


Figure 3.1: Plot of hourly rainfall distribution at RAHIOL and surrounding stations

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 3.2. 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

PRahiol, est,i = PAnior,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 3.2.

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			

Table 3.2: Hourly distribution of observed daily rainfall by SRG on the basis of nearby hourly rainfall by ARG

For judging the efficacy of the procedure, a comparison is made between the observed **(which was not missing actually)** and estimated hourly rainfall values at RAHIOL and is shown in Figure 3.2. 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.

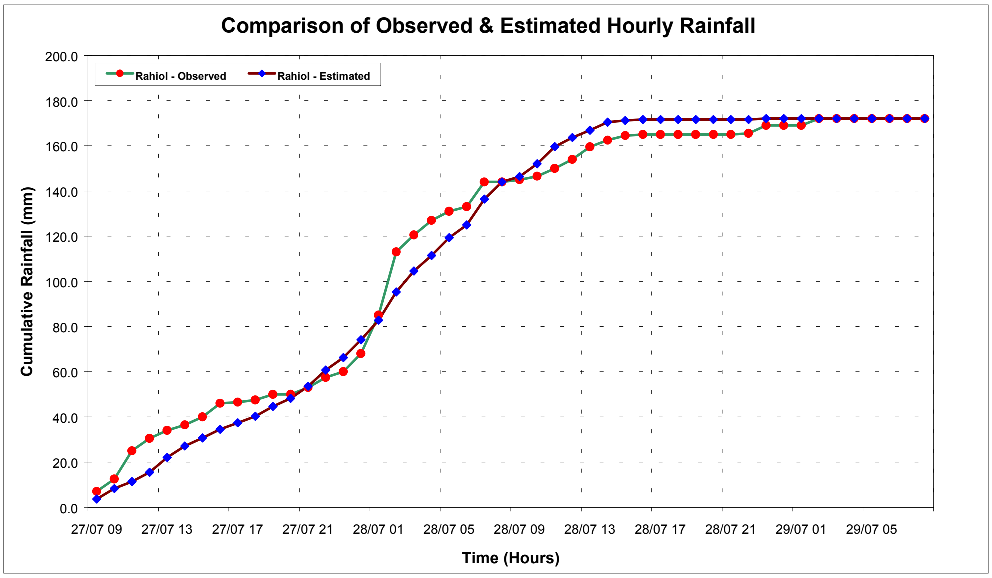


Figure 3.2: Comparison of observed and estimated hourly rainfall at RAHIOL station

3.3 CORRECTING FOR ENTRIES TO WRONG DAYS

3.3.1 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.

3.3.2 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 3.3

Referring back to Example 2.5 in Chapter 2, 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 3.3.

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 3.3. 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.

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

Table 3.3: Correction for shift in time in daily rainfall at SAVLITANK station

3.4 APPORTIONMENT FOR INDICATED AND UNINDICATED ACCUMULATIONS

3.4.1 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.

3.4.2 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:

$$P_{est,j} = \frac{\sum_{i=1}^{N_{base}} (P_{ij} / D_i^b)}{\sum_{i=1}^{N_{base}} (1/D_i^b)} = \sum_{i=1}^{N_{base}} \left((P_{ij}) \frac{(1/D_i^b)}{\sum_{i=1}^{N_{base}} (1/D_i^b)} \right) \qquad (3.2)$$

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:

$$P_{appor,j} = \frac{P_{est,j} * P_{tot}}{\sum_{j=1}^{N_{acc}} P_{est,j}} \qquad \forall \quad j = 1 \text{ to } N_{acc} \qquad (3.3)$$

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 3.4

Referring back to Example 2.9 of Chapter 2, 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 3.4.

After verifying from the field observer it may be possible to know the exact number of days for which accumulated value on 28th July has been reported. Assuming that it has been indicated by the observer that the value of 97.5 mm on 28th July is an accumulation of observations from 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.

```
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
```

Table 3.4: Tabulation of daily rainfall for neighbouring stations

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 3.5

The estimated daily rainfall based on the weighted average of the neighbouring station is computed and is given in Table 3.6. The sum of this estimated daily rainfall for the 8 days of accumulation from 21st to 28th is found to be equal to 104.1 mm. 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 3.6

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

Date	Observed	Weighted Rainfall (for DAKOR) at					Weighted Average	Corrected
	DAKOR	KATHLAL	MAHISA	MAHUDHA	SAVLITANK	THASARA		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 3.6)

Table 3.6: Computation of spatial estimate during period of accumulation and its distribution

3.5 ADJUSTING RAINFALL DATA FOR LONG TERM SYSTEMATIC SHIFTS

3.5.1 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.

3.5.2 DATA CORRECTION PROCEDURE

Consider a double mass plot shown in Figure 3.3 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.

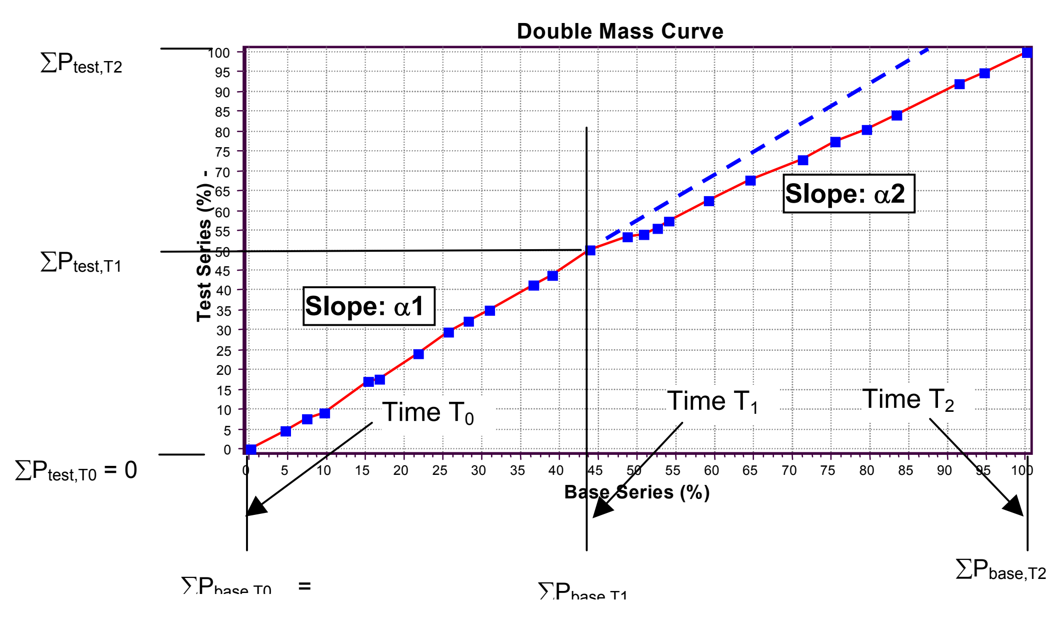


Figure 3.3: Definition sketch for double mass analysis

Considering the double mass plot shown in Figure 3.3, the break points occurs at time T_1 and if the start and end times of the period under consideration are T_0 and T_2 respectively, then the slopes of the curve before and after the break point can be expressed as:

$$\alpha_1 = \frac{\sum_{i=0}^{T_1} P_{test,i}}{\sum_{i=0}^{T_1} P_{base,i}} \tag{3.4}$$

and

$$\alpha_2 = \frac{\sum_{i=T_0}^{T_2} P_{test,i} - \sum_{i=T_0}^{T_1} P_{test,i}}{\sum_{i=T_0}^{T_2} P_{base,i} - \sum_{i=T_0}^{T_1} P_{base,i}} \tag{3.5}$$

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:

$$P_{corr,i} = P_{test,i} \times \frac{\alpha_2}{\alpha_1} \tag{3.6}$$

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 3.5

Referring back to Example 2.12 of Chapter 2, 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 Figure 3.4, that the behaviour of the test station suddenly changes after about half of the time period under consideration.

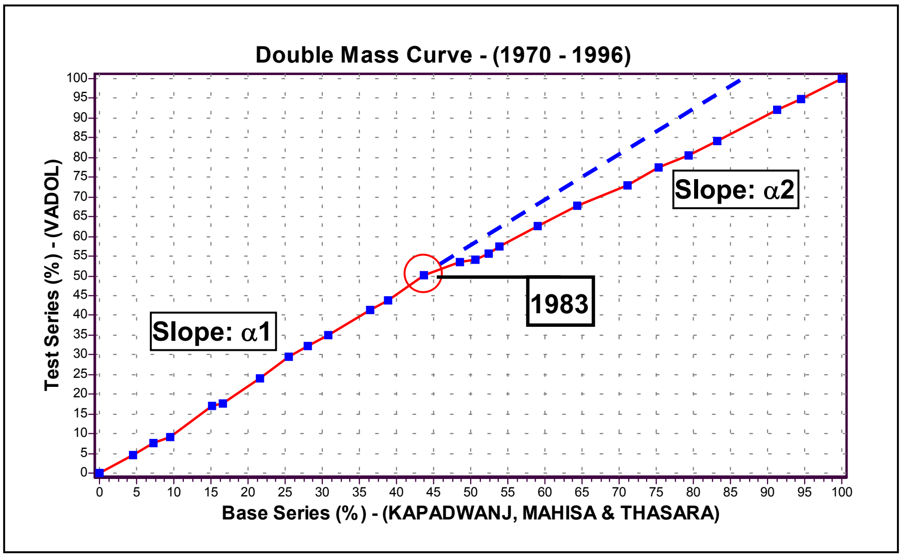


Figure 3.4: Double mass curve for VADOL station showing significant change of slope of the curve after about half the period under consideration.

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 Figure 3.4 and the results of the Double mass analysis given in Table 3.7 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.

```
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
```

Table 3.7: Results of the double mass analysis

The average slope of the double mass curve before and after this break can be worked out form the computations shown in Table 3.7 as:

$$\alpha_1 = \frac{\sum_{i=1}^{T_1} P_{test,i}}{\sum_{i=1}^{T_1} P_{base,i}} = \frac{6936}{7302} = 0.9498$$

and

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

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

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

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

Such a correction when employed would make the double mass curve correspond to the dashed line shown after the break point in Figure 3.4. The double mass curve after adjusting the data series is given in Figure 3.5 and the corresponding tabular analysis results in Table 3.8. It may be noted that the double mass curve after the data series is corrected beyond 1983 shows a consistent trend throughout.

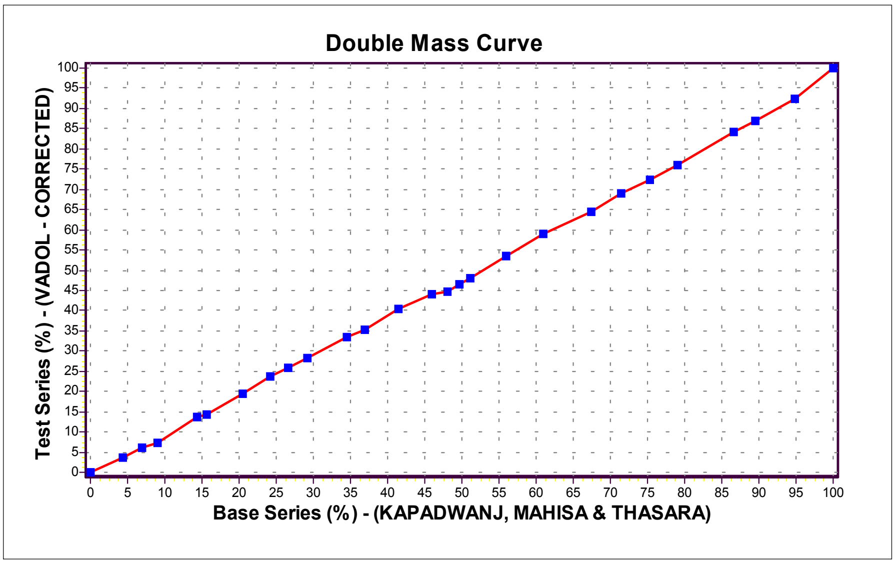


Figure 3.5: Double mass plot after adjusting the rainfall values for the period of inconsistency

3.6 USING SPATIAL INTERPOLATION TO INTERPOLATE ERRONEOUS AND MISSING VALUES

3.6.1 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.

3.6.2 ARITHMETIC AVERAGE METHOD

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

$$P_{test} = \frac{1}{M}(P_{base,1} + P_{base,2} + P_{base,3} + \ldots + P_{base,M}) \tag{3.7}$$

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

```
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
```

Table 3.8: Result of double mass analysis after adjusting the rainfall values for the period of inconsistency

Example 3.6

Consider the station BALASINOR (in KHEDA catchment) at which the daily rainfall record is not available for the year 1988. There are a few stations like MAHISA, & SAVLITANK, 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 3.9 shows the computation of the daily rainfall estimates at BALASINOR station on the basis of above three adjoining (base) stations.

Date	Observed Rainfall (mm)			Estimated Rainfall
	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**

Table 3.9: Estimation of daily rainfall at BALASINOR station by arithmetic average method

3.6.3 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:

$$P_{test} = \frac{1}{M}\left(\frac{N_{test}}{N_{base,1}}P_{1base} + \frac{N_{test}}{N_{base,2}}P_{base,2} + \frac{N_{test}}{N_{base,3}}P_{base,3} + \dots + \frac{N_{test}}{N_{base,M}}P_{base,M}\right) \tag{3.8}$$

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.

Example 3.7

Consider the station BALASINOR (in KHEDA catchment) again at which the daily rainfall record is not available for the year 1988. Assuming that the record for the neighbouring stations like MAHISA, & SAVLITANK, 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:

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

And

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

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 3.10.

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

Table 3.10: Estimation of daily rainfall at BALASINOR station by arithmetic average and normal ration method

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.

3.6.4 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 Figure 3.6. 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.

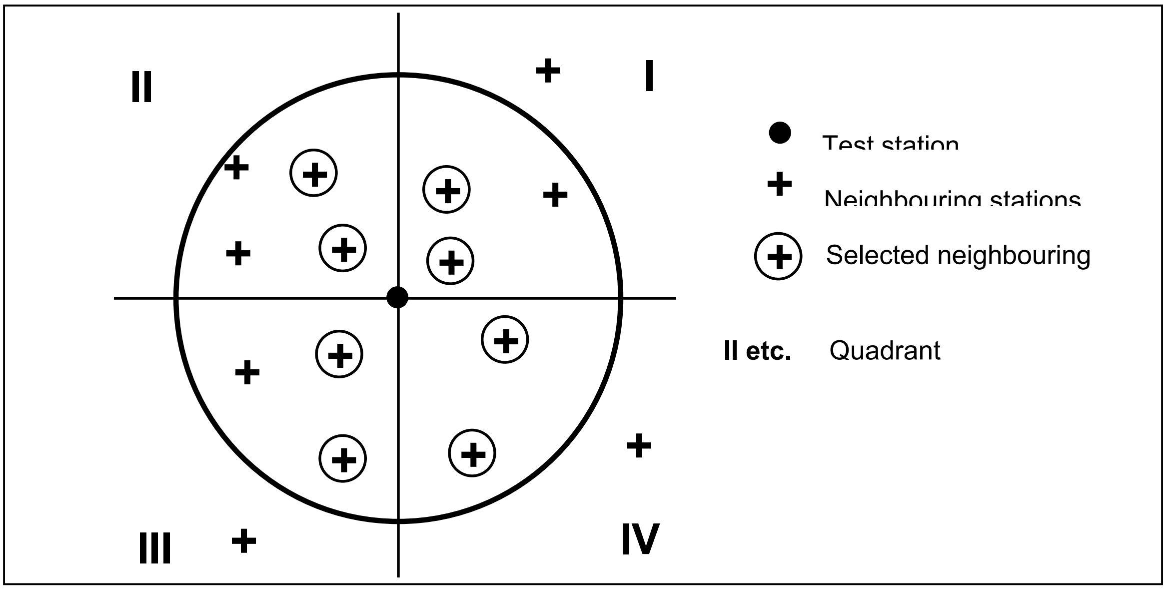


Figure3.6: Definition sketch of Test and Base (neighbouring) stations

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

$$P_{est,j} = \frac{\sum_{i=1}^{M_{base}} P_{i,j}/D_i^{\ b}}{\sum_{i=1}^{M_{base}} 1/D_i^{\ b}} \tag{3.9}$$

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

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:

$$P_{corr,i,j} = (N_{test}/N_{base,i})P_{i,j} \tag{3.10}$$

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:

$$N_i = a_1 + b_1 . H_s \quad \forall \; H_s \leq H_1 \tag{3.11}$$

$$N_i = a_2 + b_2 . H_s \quad \forall \; H_s > H_1 \tag{3.12}$$

Example 3.8

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

For this, the search for neighbouring stations (base stations) is made within a radius of 25 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 Figure 3.7. The nearest two stations are tried to be chosen which fall within the circle of 25 kms. radius. These stations are listed in Table 3.11 along with the quadrant, distances and corresponding normalised weights.

Quadrant	**Station**	**Distance (kms.)**	**Station weights (∝ 1/D²)**	
			(1/D²)	**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	1.000

Table 3.11: Distances and normalised weights of stations adjoining SAVLITANK station

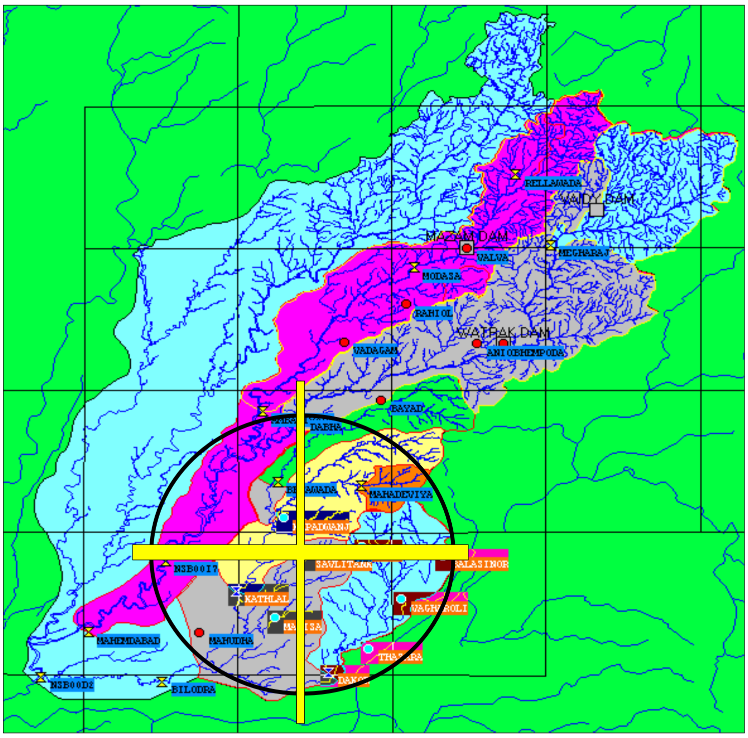


Figure 3.7: Selection of test station SAVLITANK and its adjoining (base) stations

Results of the spatial interpolation are presented in Table 3.12 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 Figure 3.8) of these estimated and observed daily rainfall values indicate that the two matches quite well. There will always be a few small and big deviations expected here and there for the simple reason that the averaging procedure is never expected to yield exactly what would have been the actual rainfall.

It may also be noted however, that by employing such 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.

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

Table 3.12: Observed daily rainfall at base stations and computation of spatial average at SAVLITANK

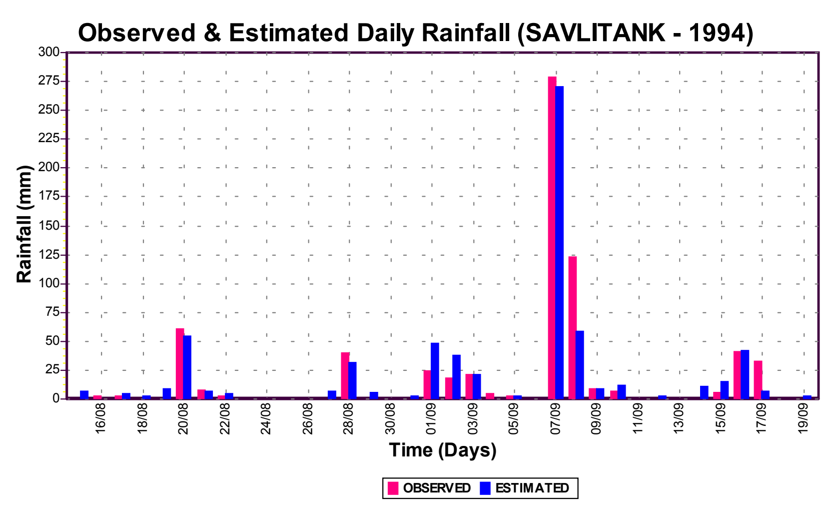


Figure 3.8: Comparison of observed and estimated daily rainfall at SAVLITANK station

4 COMPILATION OF RAINFALL DATA

4.1 GENERAL

Rainfall compilation is the process by which observed rainfall is transformed:

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

Compilation is required for validation, reporting and analysis

Compilation is carried out at the State Data Processing Centre; it is done prior to validation if required, but final compilation is carried out after correction and ‘completion’.

4.2 AGGREGATION OF DATA TO LONGER DURATIONS

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

Aggregation to longer time intervals is required for validation and analysis. For validation small persistent errors may not be detected at the small time interval of observation but may more readily be detected at longer time intervals.

4.2.1 AGGREGATION OF DAILY TO WEEKLY

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

For some application it may be required to get the weekly compilation done for the exact calendar weeks (from Monday to Sunday). In such a case the first week in any year will start from the first Monday in that year and thus there will be 51 or 52 full weeks in the year and one or more days left in the beginning and/or end of the year. The days left out at the end of a year or beginning of the next year could be considered for the 52nd of the year under consideration. There will also be cases of a 53rd week when the 1st day of the year is also the first day of the week (for non-leap years) and 1st or 2nd day of the year is also first day of the week (for leap years).

4.2.2 AGGREGATION OF DAILY TO TEN DAILY

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

4.2.3 AGGREGATION FROM DAILY TO MONTHLY

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

4.2.4 HOURLY TO OTHER INTERVALS

From rainfall data at hourly or lesser time intervals, it may be desired to obtain rainfall data for every 2 hours, 3 hours, 6 hours, 12 hours etc. for any specific requirement. Such compilations are carried out by simply adding up the corresponding rainfall data at available smaller time interval.

Example 4.1

Daily rainfall at ANIOR station (KHEDA catchment) is observed with Standard Raingauge (SRG). An Autographic Raingauge (ARG) is also available at the same station for recording rainfall continuously and hourly rainfall data is obtained by tabulating information from the chart records.

It is required that the hourly data is compiled to the daily interval corresponding to the observations synoptic observations at 0830 hrs. This compilation is done using the aggregation option and choosing to convert from hourly to daily interval. The observed hourly data and compiled daily data is shown if Figure 4.1 and Figure 4.2 respectively.

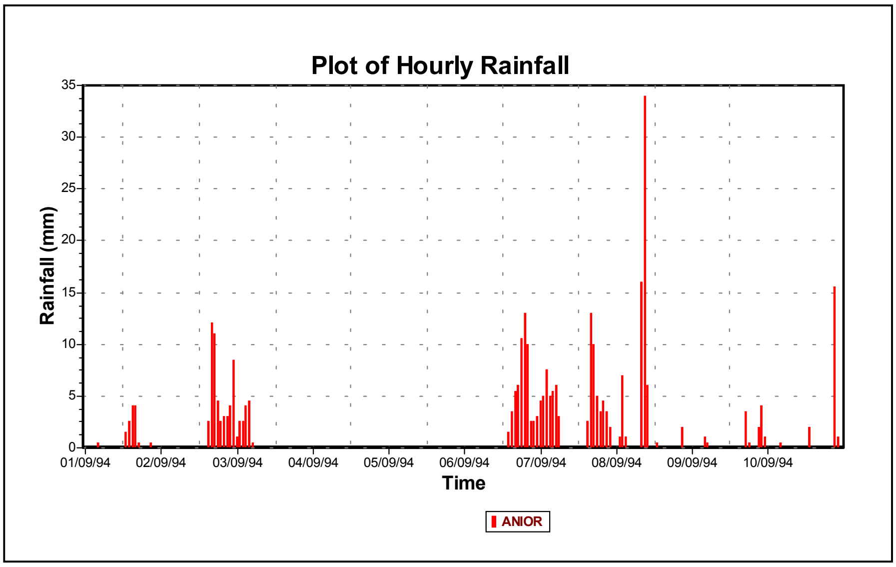


Figure 4.1: Plot of observed hourly rainfall data

Similarly, daily data observed using SRG is required to be compiled at weekly, ten-daily, monthly and/or yearly interval for various application and for the purpose of data validation. For this, the daily data obtained using SRG is taken as the basic data and compilation is done to weekly, ten-daily, monthly and yearly intervals. These are illustrated in Figure 4.3, Figure 4.4, Figure 4.5 and Figure 4.6 respectively.

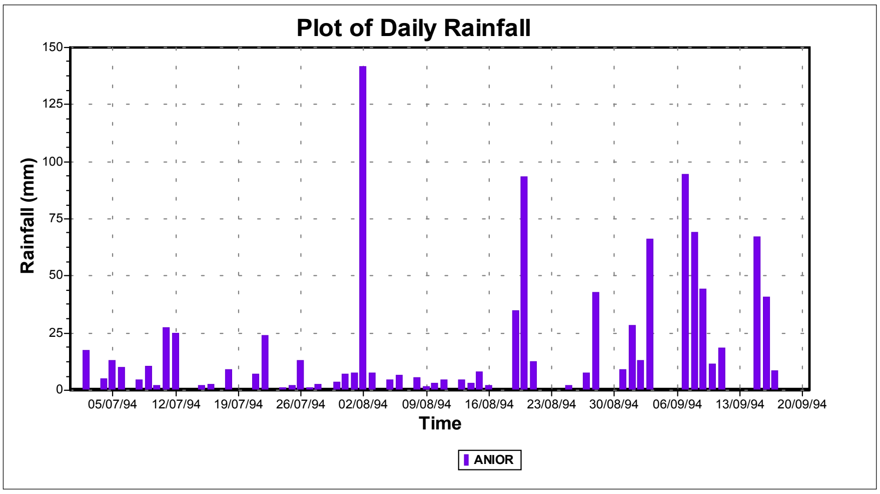


Figure 4.2: Compiled daily rainfall from hourly data tabulated from ARG charts

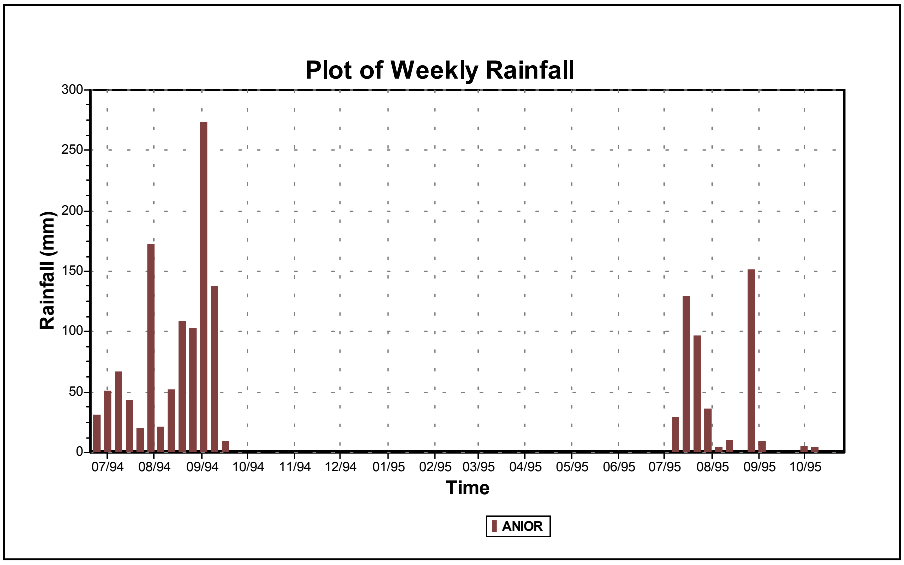


Figure 4.3: Compiled weekly rainfall from hourly data tabulated from ARG charts

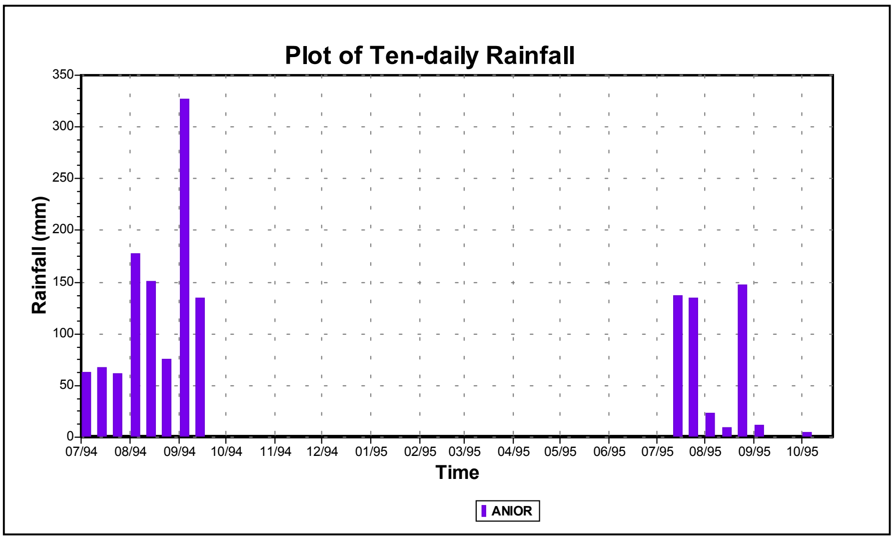


Figure 4.4: Compiled ten-daily data from daily data obtained from SRG records

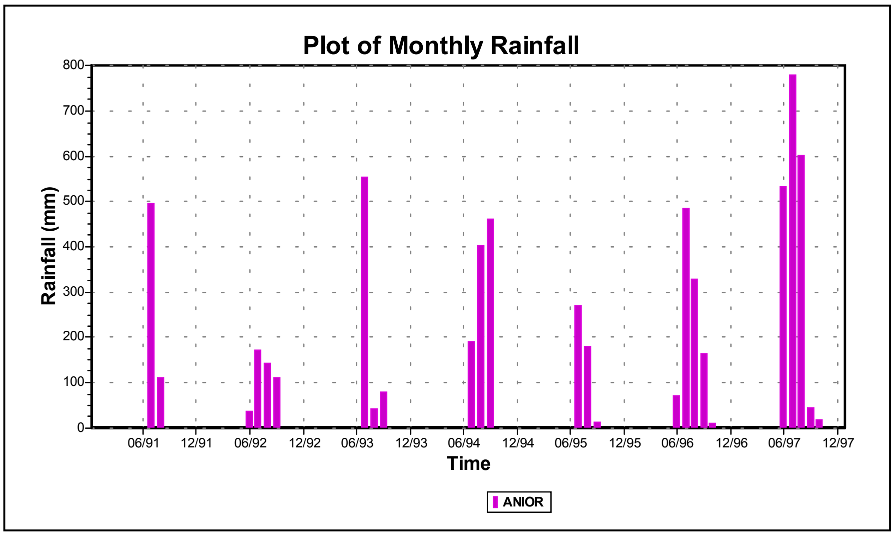


Figure 4.5: Compiled monthly data from daily data obtained from SRG records

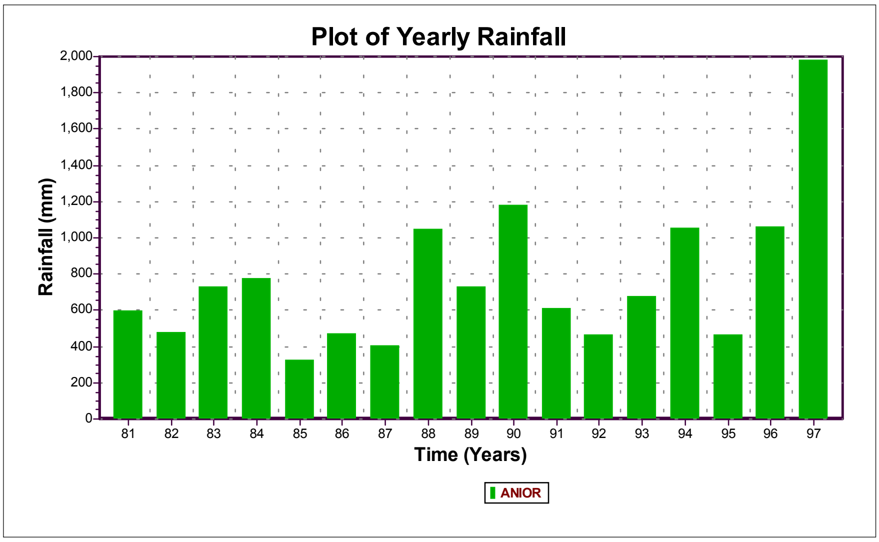


Figure 4.6: Compiled yearly data from daily data obtained from SRG records

4.3 ESTIMATION OF AREAL RAINFALL

4.3.1 GENERAL DESCRIPTION

Raingauges generally measure rainfall at individual points. However, many hydrological applications require the average depth of rainfall occurring over an area which can then be compared directly with runoff from that area. The area under consideration can be a principal river basin or a component sub-basin. Occasionally, average areal rainfall is required for country, state or other administrative unit, and the areal average is obtained within the appropriate political or administrative boundary.

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

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

- The Arithmetic average method,
- Weighted average method,
- Thiessen polygon method,
- Isohyetal and related methods,
- Kriging techniques.

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

4.3.2 ARITHMETIC AVERAGE

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

$$P_{at} = \frac{1}{N}(P_{1t} + P_{2t} + P_{3t} + \cdots + P_{Nt}) = \frac{1}{N}\sum_{i=1}^{N} P_{it} \qquad (4.1)$$

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

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

N = total number of point rainfall stations considered

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

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

4.3.3 WEIGHTED AVERAGE USING USER DEFINED WEIGHTS

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

$$P_{wt} = \frac{1}{N}\left(c_1 P_{1t} + c_2 P_{2t} + c_3 P_{3t} + \dots + c_N P_{Nt}\right) = \frac{1}{N}\sum_{i=1}^{N} c_i P_{it} \qquad (4.2)$$

Where:

c_i = weight assigned to individual raingauge station i (i = 1,N).

To account for under-representation by gauges located in valleys the weights do not necessarily need to add up to 1.

4.3.4 THIESSEN POLYGON METHOD

This widely-used method was proposed by A.M. Thiessen in 1911. The Thiessen polygon method accounts for the variability in spatial distribution of gauges and the consequent variable area which each gauge represents. The areas representing each gauge are defined by drawing lines between adjacent stations on a map. The perpendicular bisectors of these lines form a pattern of polygons (the Thiessen polygons) with one station in each polygon (see Figure 4.7). Stations outside the basin boundary should be included in the analysis as they may have polygons which extend into the basin area. The area of a polygon for an individual station as a proportion of the total basin area represents the Thiessen weight for that station. Areal rainfall is thus estimated by first multiplying individual station totals by their Thiessen weights and then summing the weighted totals as follows:

$$P_{at} = \frac{A_1}{A}P_{1t} + \frac{A_2}{A}P_{2t} + \frac{A_3}{A}P_{3t} + \quad \cdots \quad + \frac{A_N}{A}P_{Nt} = \sum_{i=1}^{N}\left(\frac{A_i}{A}\right)P_{it} \tag{4.3}$$

where:

A_i = the area of Thiessen polygon for station i

A = total area under consideration

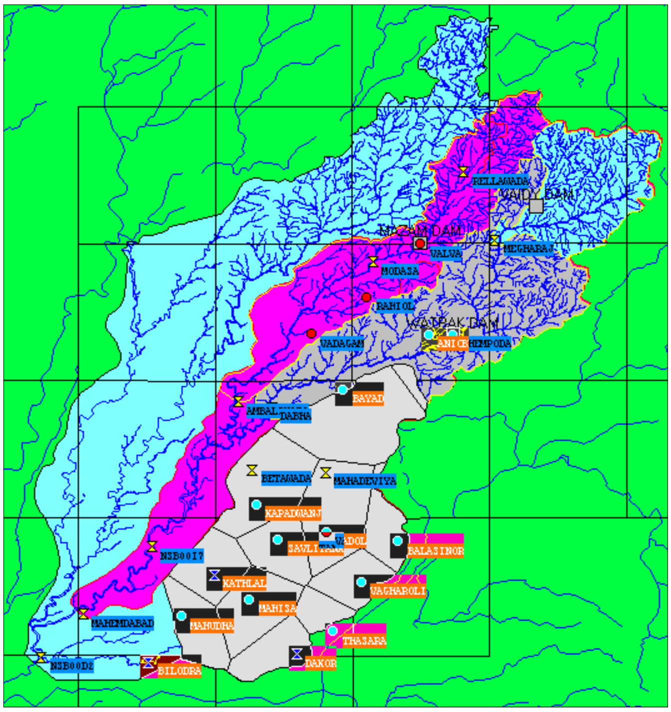


Figure 4.7
Small basin upto BILODRA gauging site (portion shown with Thiessen polygons)

The Thiessen method is objective and readily computerised but is not ideal for mountainous areas where orographic effects are significant or where raingauges are predominantly located at lower elevations of the basin. Altitude weighted polygons (including altitude as well as areal effects) have been devised but are not widely used.

Example 4.2

Areal average rainfall for a small basin upto BILODRA gauging site (shown highlighted in Figure 4.7) in KHEDA catchment is required to be compiled on the basis of daily rainfall data observed at a number of raingauges in and around the region. Areal average is worked out using two methods: (a) Arithmetic average and (b) Thiessen method.

(a) Arithmetic Average

For the arithmetic average method rainfall stations located inside and very nearby to the catchment boundary are considered and equal weights are assigned to all of them. Since there are 11 stations considered the individual station weights work out as 0.0909 and is given in Table 4.1 below. On the basis of these equal station weights daily areal average is computed. The compiled areal daily rainfall worked out using arithmetic average method is shown for the year 1994 in Figure 4.8.

```
Areal computation - Arithmetic Average

 Areal series: BILODRA      MA1

 Station weights
 BALASINOR     =   0.0909
 DAKOR         =   0.0909
 KAPADWANJ     =   0.0909
 BAYAD         =   0.0909
 MAHISA        =   0.0909
 MAHUDHA       =   0.0909
 SAVLITANK     =   0.0909
 THASARA       =   0.0909
 VAGHAROLI     =   0.0909
 VADOL         =   0.0909
 KATHLAL       =   0.0909

 Sum           =    0.999
```

Table 4.1: List of stations and corresponding weights for arithmetic average method

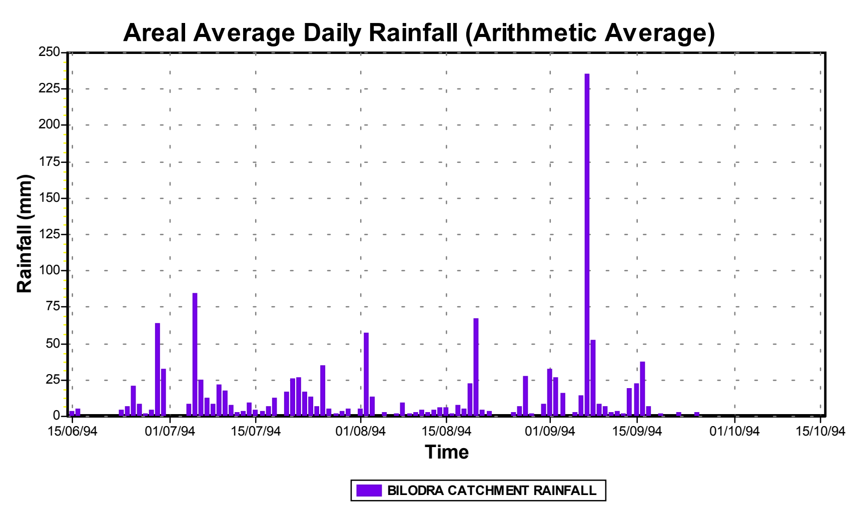


Figure 4.8: Plot of areal daily rainfall for BILODRA catchment using arithmetic average method

(b) Thiessen polygon method

Computation of areal average using Thiessen method is accomplished by first getting the Thiessen polygon layer (defining the boundary of Thiessen polygon for each contributing point rainfall station). The station weights are automatically worked out on the basis of areas of these polygons with respect to the total area of the catchment. The layout of the Thiessen polygons as worked out by the system is graphically shown in Figure 4.7 and the corresponding station weights are as given in Table 4.2. On the basis of these Thiessen polygon weights the areal average of the basin is computed and this is shown in Figure 4.9 for the year 1994. In this case it may be noticed that there is no significant change in the values of the areal rainfall obtained by the two methods primarily on account of lesser variation in rainfall from station to station.

```
Areal computation - Thiessen Polygon Method

Areal series: BILODRA     MA3

Station weights
ANIOR        =   0.0127
BALASINOR    =   0.0556
BAYAD        =   0.1785
DAKOR        =   0.0659
KAPADWANJ    =   0.1369
KATHLAL      =   0.0763
MAHISA       =   0.0969
MAHUDHA      =   0.0755
SAVLITANK    =   0.0724
THASARA      =   0.0348
VADOL        =   0.1329
VAGHAROLI    =   0.0610

Sum          =   1.00
```

Table 4.2: List of stations and corresponding weights as per Thiessen polygon method

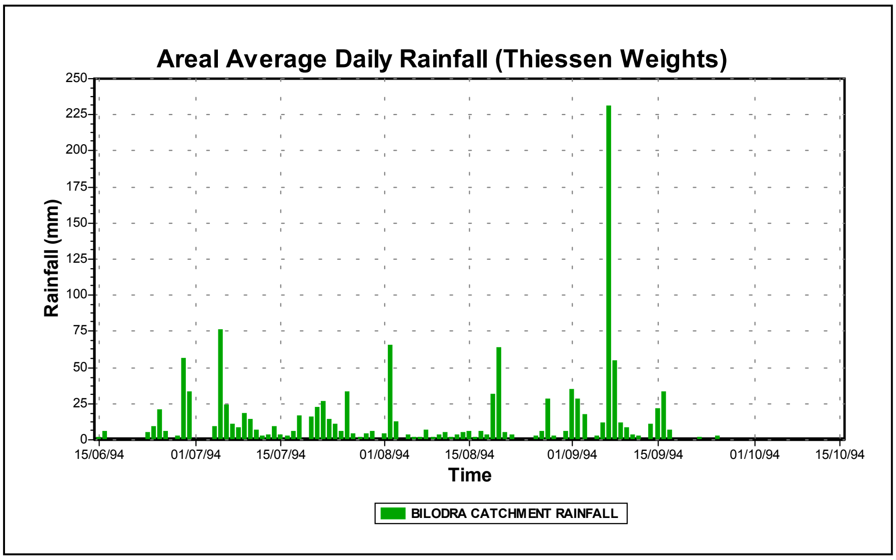


Figure 4.9: Plot of areal daily rainfall for BILODRA catchment using Thiessen polygon method

4.3.5 ISOHYETAL AND RELATED METHODS

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

In flat areas where no orographic effects are present the method simply interpolates linearly between the point rainfall stations. Manually, the procedure is as follows. On a basin map first the locations of the rainfall stations within the basin and outside near the basin boundary are plotted. Next, the stations are connected with their neighbouring stations by straight lines. Dependent on the rain depths for which isohyets are to be shown by linear interpolation between two neighbouring stations the position of the isohyet(s) on these connecting lines are indicated. After having completed this for all connected stations, smooth curves are drawn through the points marked on the straight lines between the stations connecting the concurrent rainfall values for which isohyets are to be shown, see Figure 4.10. In drawing the isohyets personal experience with local conditions and information on storm orientation can be taken into account. Subsequently, the area between two adjacent isohyets and the catchment boundary is planimetered. The average rainfall obtained from the two adjacent isohyets is assumed to have occurred over the entire inter-isohyet area. Hence, if the isohyets are indicated by P_1, P_2, …, P_n with inter-isohyet areas a_1, a_2, …, a_{n-1} the mean precipitation over the catchment is computed from:

$$\overline{P} = \frac{a_1\left(\frac{P_1 + P_2}{2}\right) + a_2\left(\frac{P_2 + P_3}{2}\right) + ... + a_{n-1}\left(\frac{P_{n-1} + P_n}{2}\right)}{A} \qquad (4.4)$$

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

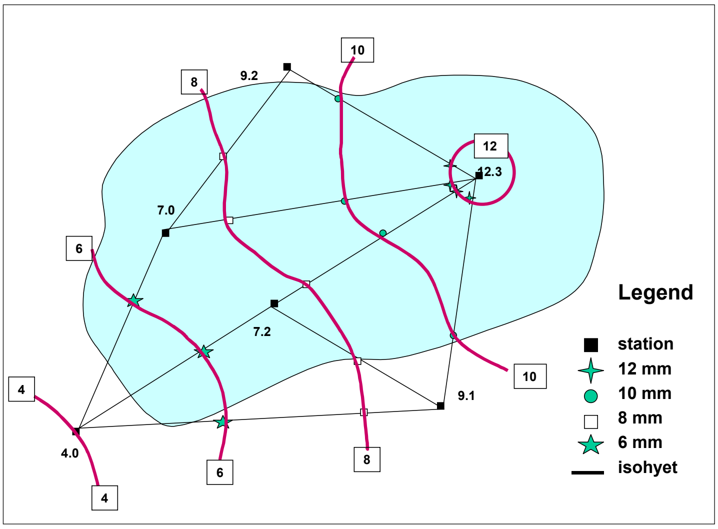


Figure 4.10:
Example of drawing of isohyets using linear interpolation

For flat areas the isohyetal method is superior to the Thiessen method if individual storms are considered as it allows for incorporation of storm features like orientation; for monthly, seasonal or annual values such preference is not available. But its added value is particularly generated when special topographically induced meteorological features like orographic effects are present in the catchment rainfall. In such cases the above procedure is executed with a catchment map overlaying a topographical map to be able to draw the isohyets parallel to the contour lines. Also the extent of rain shadow areas at the leeward side of mountain chains can easily be identified from topographical maps. The computations are again carried out with the aid of equation 4.4. In such situations the isohyetal method is likely to be superior to the Thiessen method.

The **isopercental method** is very well suited to incorporate long term seasonal orographical patterns in drawing isohyets for individual storms or seasons. The assumption is that the long term seasonal orographical effect as displayed in the isohyets of season normals applies for individual storms and seasons as well. The procedure involves the following steps, and is worked out in Example 4.2:

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

Special attention is to be paid to situations where at the higher elevations raingauge stations are non-existing. Then the orographic effect has to be extrapolated from the lower reaches of the mountains by estimating a relation between rainfall and elevation which is assumed to be valid for the higher elevations as well. Using this rainfall-elevation curve a number of points in the ungauged upper reaches are added to the point rainfall data to guide the interpolation process.

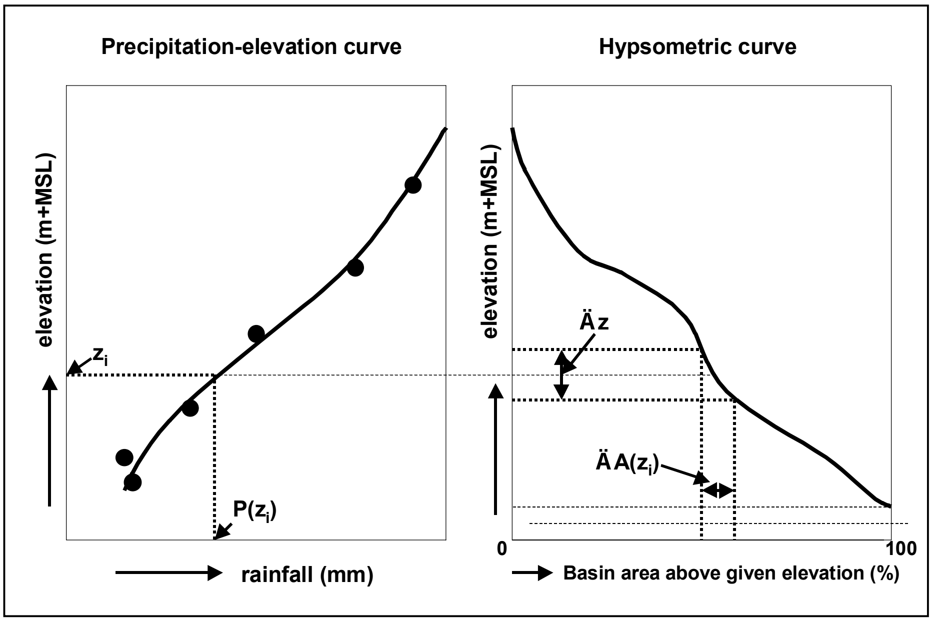


Figure 4.11: Principle of hypsometric method

A simple technique to deal with such situations is the **hypsometric method**, see e.g. Dingman, 2002, where a precipitation-elevation curve is combined with an area-elevation curve (called hypsometric curve) to determine the areal rainfall. The latter method avoids recurrent planimetering of inter-isohyet areas, whereas the results will be similar to the isohyetal method. The precipitation-elevation curve has to be prepared for each storm, month, season or year, but its development will be guided by the rainfall normal-elevation curve also called the orographic equation. Often the orographic equation can be approximated by a simple linear relation of the form:

$$P(z) = a + bz \tag{4.5}$$

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

$$\overline{P} = \sum_{i=1}^{n} P(z_i)\Delta A(z_i) \tag{4.6}$$

where: $\overline{P}$ = areal rainfall

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

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

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

Example 4.3

In this example the application of the isopercental method is demonstrated (NIH, 1988). The areal rainfall for the storm of 30 August 1982 has to be determined for the catchment shown in Figure 4.12a. The total catchment area amounts 5,600 km^2. The observed and normal annual rainfall amounts for the point rainfall stations in the area are given in Table 4.3.

Station	30 August 1982 storm	Normal annual rainfall	Storm rainfall as percentage of annual normal
	(mm)	(mm)	(%)
1. Paikmal	338.0	1728	19.6
2. Padampur	177.0	1302	13.6
3. Bijepur	521.0	1237	42.1
4. Sohela	262.0	1247	21.0
5. Binka	158.0	1493	10.6
6. Bolangir	401.6	1440	27.9

Table 4.3 Storm rainfall and annual normals

For each station the point rainfall as percentage of seasonal normal is displayed in the last column of Table 4.3. Based on this information isopercetals are drawn on a transparent overlay, which is subsequently superimposed on the annual normal isohyetal map. The intersections of the isopercetals and isohyets are identified and for each intersection the isopercental is multiplied with the isohyet to get an estimate of the storm rainfall for that point. These estimates are then added to the point rainfall observations to draw the isohyets, see Figure 4.12b. The inter-isohyet area is planimetered and the areal rainfall is subsequently computed with the aid of equation 4.4 as shown in Table 4.4.

Isohyetal range	Mean rainfall	Area	Volume
(mm)	(mm)	(km^2)	(km^{2}xmm)
110-150	130	80	10400
150-200	175	600	105000
177-200	188.5	600	113100
200-250	225	3370	758250
250-300	275	620	170500
300-400	350	230	80500
400-500	450	90	40500
500-521	510.5	10	5105
	Total	5600	1283355
		Average	1283355/5600= 229.2 mm

Table 4.4: Computation of areal rainfall by isohyetal/isopercental method

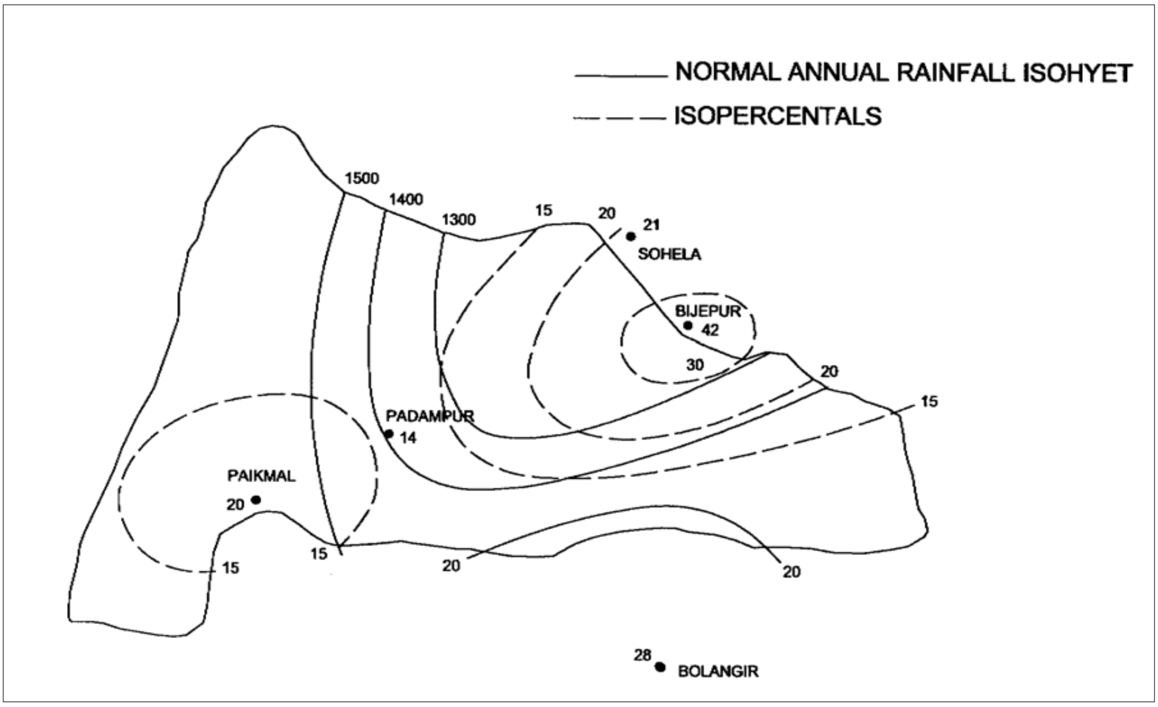


Figure4.12a: Isopercental map

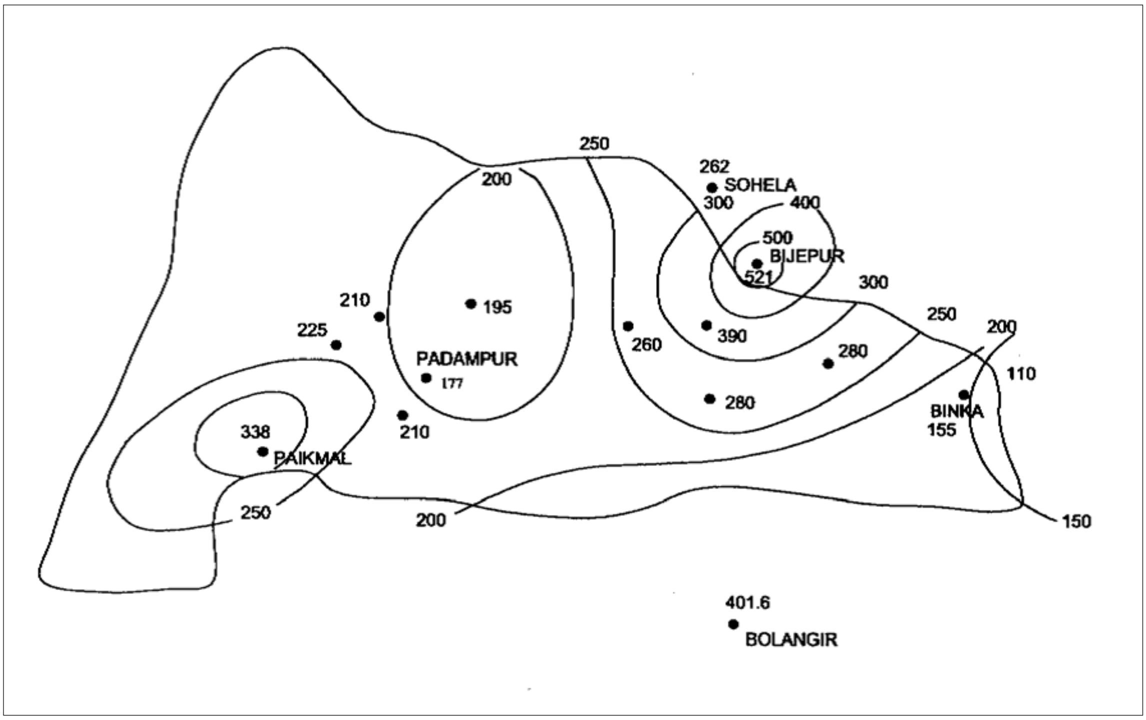


Figure 4.12b: Isohyetal map drawn by isopercental method

4.3.6 KRIGING

General

The Kriging Method is an interpolation method. It provides rainfall estimates (or estimates of any other variable) at points (point-kriging) or blocks (block-kriging) based on a weighted average of observations made at surrounding stations. In this section point-kriging will be discussed. In the application of the kriging method for areal rainfall estimation and drawing of isohyets a dense grid is put over the catchment. By estimating the rainfall for the gridpoints the areal rainfall is simply determined as the average rainfall of all grid points within the catchment. In addition, in view of the dense grid, it is very easy to draw isohyets based on the rainfall values at the grid points.

At each gridpoint the rainfall is estimated from:

$$Pe_0 = \sum_{k=1}^{N} w_{0,k} \cdot P_k \tag{4.7}$$

where: Pe_0 = rainfall estimate at some gridpoint "0"

$w_{0,k}$ = weight of station k in the estimate of the rainfall at point "0"

P_k = rainfall observed at station k

N = number of stations considered in the estimation of Pe_0

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

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

Particularly the error variance minimisation distinguishes the kriging method from other methods like e.g. inverse distance weighting. The advantage of the kriging method above other methods is that it provides besides the best linear estimate of rainfall for a point on the grid also the uncertainty in the estimate. The latter property makes the method useful if locations for additional stations have to be selected when the network is to be upgraded, because then the new locations can be chosen such that overall error variance is reduced most.

Bias elimination and error variance minimisation

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

$$e_0 = Pe_0 – P_0 \tag{4.8}$$

with Pe_0 estimated by (4.7). It is clear from (4.8) that any statement about the mean and variance of the estimation error requires knowledge about the true behaviour of the rainfall at unmeasured locations, which is not known. This problem is solved by hypothesising:

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

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

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

The unbiasedness implies:

$$E[e_0] = 0 \quad \text{so:} \quad E[\sum_{k=1}^{N} w_{0,k}.P_k] - E[P_0] = 0 \quad \text{or:} \quad E[P_0]\left(\sum_{k=1}^{N} w_{0,k} - 1\right) = 0$$

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

$$\sum_{k=1}^{N} w_{0,k} = 1 \tag{4.9}$$

The error variance can be shown to be (see e.g. Isaaks and Srivastava, 1989):

$$\sigma_e^2 = E[(Pe_0 - P_0)^2] = \sigma_P^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,i} w_{0,j} C_{i,j} - 2\sum_{i=1}^{N} w_{0,i} C_{0,i} \tag{4.10}$$

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

$$\mathbf{C\ .\ w = D} \tag{4.11}$$

where:

$$\mathbf{C} = \begin{bmatrix} C_{11}\ldots\ldots\ldots\ldots C_{1N} & 1 \\ . \qquad\qquad\qquad . & . \\ . \qquad\qquad\qquad . & . \\ C_{N1}\ldots\ldots\ldots\ldots C_{NN} & 1 \\ 1\ldots\ldots\ldots\ldots\ldots 1 & 0 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} w_{0,1} \\ . \\ . \\ w_{0,N} \\ \mu \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} C_{0,1} \\ . \\ . \\ C_{0,N} \\ 1 \end{bmatrix}$$

Note that the last column and row in **C** are added because of the introduction of the Lagrangian multiplyer μ in the set of N+1 equations. By inverting the covariance matrix the station weights to estimate the rainfall at location 0 follow from (4.11) as:

$$\mathbf{w = C^{-1}\ .\ D} \tag{4.12}$$

The error variance is then determined from:

$$\sigma_e^2 = \sigma_P^2 - \mathbf{w}^T . \mathbf{D} \qquad (4.13)$$

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

Covariance and variogram models

To actually solve above equations a function is required which describes the covariance of the rainfall field as a function of distance. For this we recall the correlation structure between the rainfall stations discussed in Chapter 2. The spatial correlation structure is usually well described by an exponential relation of the following type:

$$r(d) = r_0 \exp(-d/d_0) \qquad (4.14)$$

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

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

d_0 = characteristic correlation distance.

Two features of this function are of importance:

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

The exponential correlation function is shown in Figure (4.13).

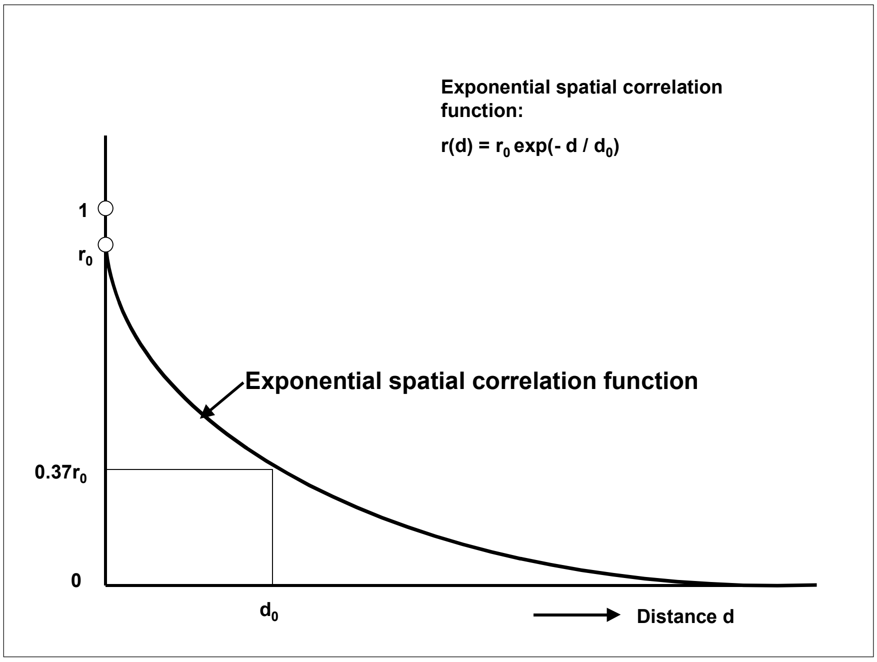


Figure 4.13: Spatial correlation structure of rainfall field

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

$$C(d) = C_0 + C_1 \quad \text{for} \quad d = 0$$
$$C(d) = C_1 \exp(-\frac{3d}{a}) \quad \text{for} \quad d > 0 \tag{4.15}$$

Since according to the definition C(d) = r(d)σ_P^2, the coefficients C_0 and C_1 in (4.15) can be related to those of the exponential correlation model in (4.14) as follows:

$$C_0 = \sigma_P^2(1-r_0) \quad ; \quad C_1 = \sigma_P^2 r_0 \quad \text{and} \quad a = 3d_0 \tag{4.16}$$

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

$$\gamma(d) = \tfrac{1}{2} E[\{P(x_1) - P(x_1-d)\}^2] = \sigma_P^2 - C(d) \tag{4.17}$$

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

$$\gamma(d) = 0 \quad \text{for:} \quad d = 0$$
$$\gamma(d) = C_0 + C_1\left(1 - \exp(-\frac{3d}{a})\right) \quad \text{for:} \quad d > 0 \tag{4.18}$$

The exponential covariance and variogram models are shown in Figures 4.14 and 4.15.

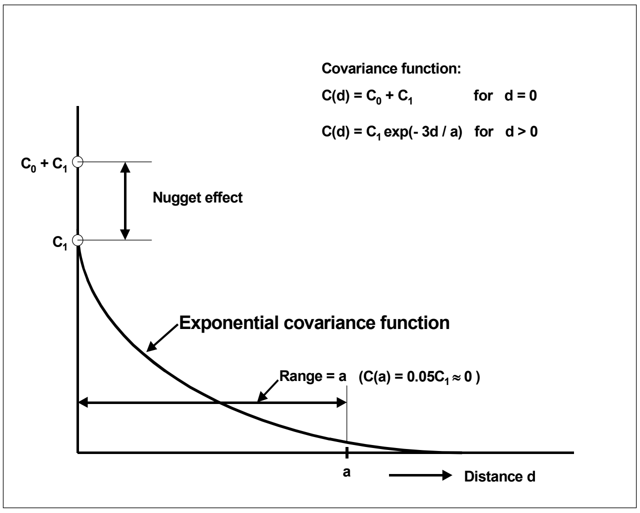


Figure 4.14:
Exponential covariance model

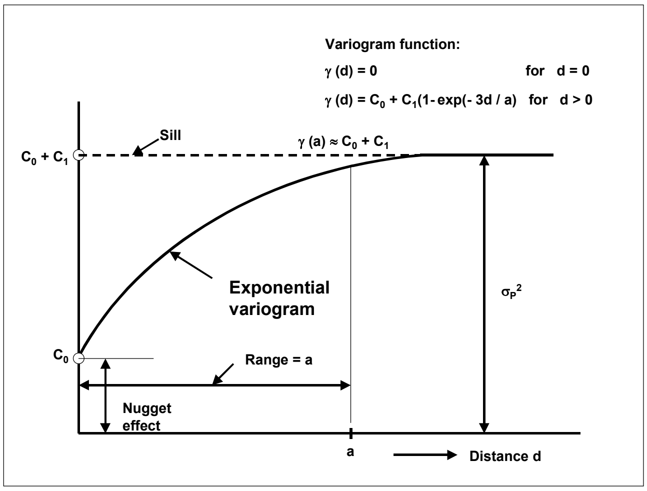


Figure 4.15:
Exponential variogram model

Features of the exponential model are following:

- C_0, which is called the **nugget effect**, provides a discontinuity at the origin; according to (4.16): $C_0 = \sigma_P^2(1-r_0)$, hence in most applications of this model to rainfall data a small nugget effect will always be present
- The distance ‘a’ in the covariance function and variogram is called the **range** and refers to the distance above which the functions are essentially constant; for the exponential model $a = 3d_0$ can be applied
- $C_0 + C_1$ is called the **sill** of the variogram and provides the limiting value for large distance and becomes equal to σ_P^2; it also gives the covariance for $d = 0$.

Other Covariance and semi-variogram models

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

- Spherical model, and
- Gaussian model

These models have the following forms:

Spherical:

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

$$\gamma(d) = 1 \qquad \text{otherwise} \qquad (4.19)$$

Gaussian:

$$\gamma(d) = C_0 + C_1\left(1 - \exp\left(-\frac{3d^2}{a^2}\right)\right) \qquad (4.20)$$

The Spherical and Gaussian models are shown with the Exponential Model in Figure 4.16.

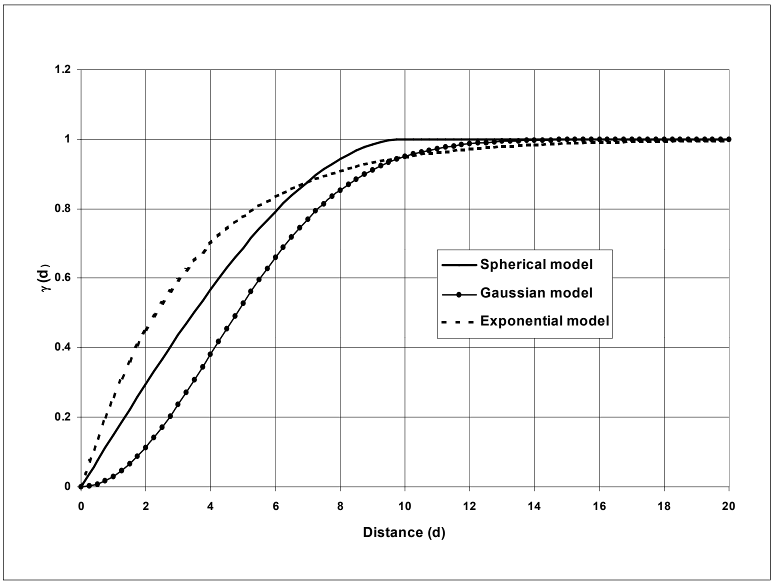


Figure 4.16:
Example of Spherical, Gaussian and Exponential type of variogram models, with C_0=0,C_1=1 and a = 10

The spherical model has a linear behaviour at small separation distances near the origin, with the tangent at the origin intersecting the sill at about 2/3 of the range "a". The model reaches the sill at the range. The gaussian model is fit for extremely continuous phenomena, with only gradually diminishing correlation near the origin, much smoother than the other two models. The range "a" is at a distance the variogram value is 95% of the sill. The exponential model rises sharper than the other two but flattens out more gradually at larger distances; the tangent at the origin reaches the sill at about 1/5 of the range.

Sensitivity analysis of variogram model parameters

To show the effect of variation in the covariance or variogram models on the weights attributed to the observation stations to estimate the value at a grid point an example presented by Isaaks and Srivastava (1989) is presented. Observations made at the stations as shown in Figure 4.17 are used. Some 7 stations are available to estimate the value at point '0' (65,137).

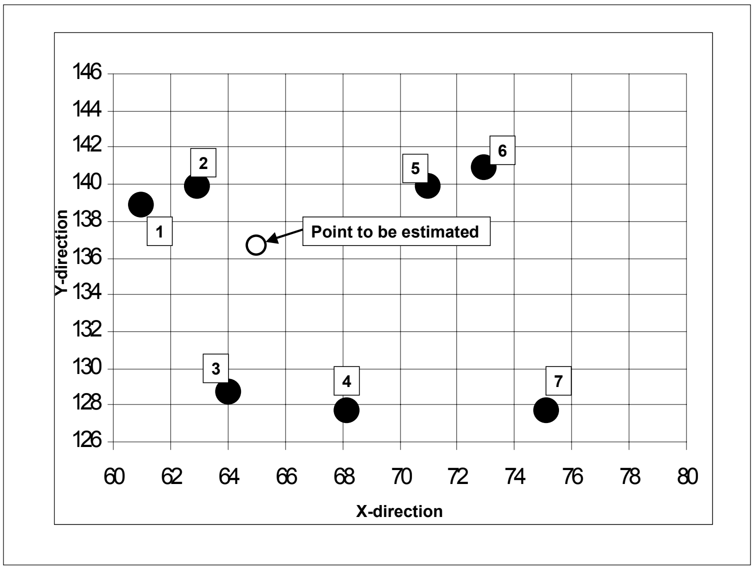


Figure 4.17:
Layout of network with location of stations 1, …., 7

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

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

$$\text{Case 1}: \quad \gamma_1(d) = 10\left(1 - \exp(-\frac{3d}{10})\right) \qquad C_0 = 0 \quad C_1 = 10 \quad a = 10$$

$$\text{Case 2}: \quad \gamma_2(d) = 20\left(1 - \exp(-\frac{3d}{10})\right) = 2\gamma_1(d) \qquad C_0 = 0 \quad C_1 = 20 \quad a = 10$$

$$\text{Case 3}: \quad \gamma_3(d) = 10\left(1 - \exp(-3(\frac{d}{10})^2)\right) \qquad C_0 = 0 \quad C_1 = 10 \quad a = 10 \quad \text{(Gaussian)}$$

$$\text{Case 4}: \quad \gamma_4(d) = 0 \qquad \text{for} : d = 0$$

$$\gamma_4(d) = 5 + 5\left(1 - \exp(-\frac{3d}{10})\right) \quad \text{for} : d > 0 \quad C_0 = 5 \quad C_1 = 5 \quad a = 10$$

$$\text{Case 5}: \quad \gamma_5(d) = 10\left(1 - \exp(-\frac{3d}{20})\right) \qquad C_0 = 0 \quad C_1 = 10 \quad a = 20$$

The covariance functions and variograms for the cases are shown in Figures 4.18 and 4.19.

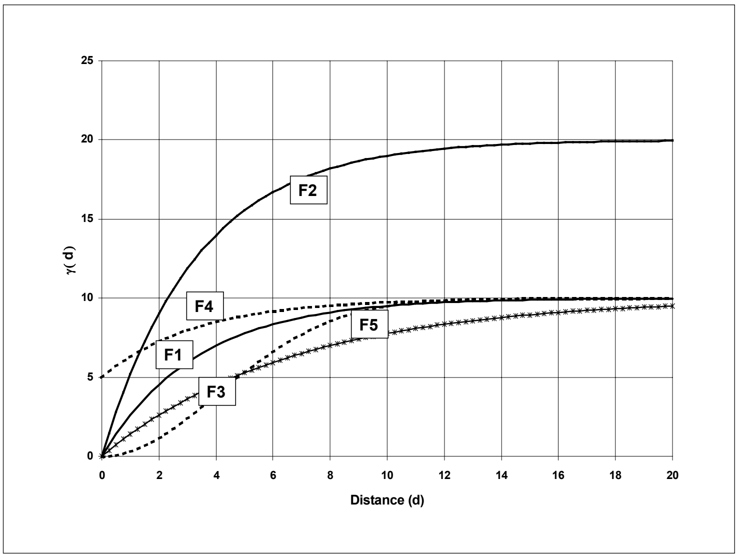


Figure 4.18:
Covariance models for the various cases

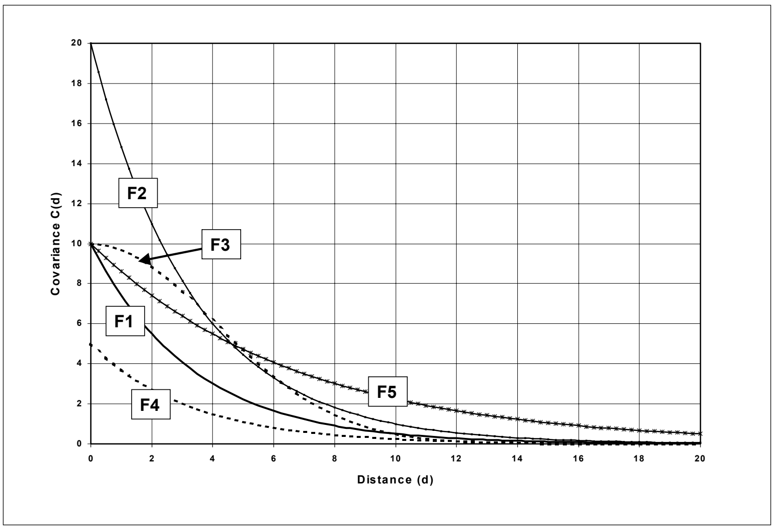


Figure 4.19:
Semi-variograms for the various cases

The results of the estimate and variance at point "0" as well as the weights of the stations computed with the models in estimating point "0" are presented in Table 4.5.

Case	Estimate at "0"	Error variance	Stations	1	2	3	4	5	6	7
			Distance to "0"	4.47	3.61	8.06	9.49	6.71	8.94	13.45
	(mm)	(mm^2)		weights						
1	593	8.86		0.17	0.32	0.13	0.09	0.15	0.06	0.09
2	593	17.91		0.17	0.32	0.13	0.09	0.15	0.06	0.09
3	559	4.78		-0.02	0.68	0.17	-0.01	0.44	-0.29	0.04
4	603	11.23		0.15	0.18	0.14	0.14	0.13	0.13	0.14
5	572	5.76		0.18	0.38	0.14	0.07	0.20	0.00	0.03
ID	590	-		0.44	0.49	0.02	0.01	0.02	0.01	0.01

Table 4.5: Results of computations for Cases 1 to 5 and ID (=Inverse Distance Method) with p = 2

From the results the following can be concluded:

- **Effect of scale**: compare Case 1 with Case 2
  In Case 2 the process variance, i.e. the sill is twice as large as in Case 1. The only effect this has on the result is a doubled error variance at "0". The weights and therefore also the estimate remains unchanged. The result is easily confirmed from equations (4.12) and (4.13) as both **C, D** and σ_P^2 are multiplied with a factor 2 in the second case.
- **Effect of shape**: compare Case 1 with Case 3
  In Case 3 the spatial continuity near the origin is much larger than in Case 1, but the sill is the same in both cases. It is observed that in Case 3 the estimate for "0" is almost entirely determined by the three nearest stations. Note that kriging does cope with clustered stations; even negative weights are generated by stations in the clusters of stations (5, 6) and (1, 2) to reduce the effect of a particular cluster. Note also that the estimate has changed and that the error variance has reduced as more weight is given to stations at small distance. It shows that due attention is to be given to the correlation structure at small distances as it affects the outcome significantly.
- **The nugget effect**: compare Case 1 with Case 4
  In Case 4, which shows a strong nugget effect, the spatial correlation has substantially been reduced near the origin compared to Case 1. As a result the model discriminates less among the stations. This is reflected in the weights given to the stations. It is observed that almost equal weight is given to the stations in Case 4. In case correlation would have been zero the weights would have been exactly equal.

- **Effect of range**: compare Case 1 with Case 5
  The range in Case 5 is twice as large as in Case 1. It means that the spatial correlation is more pronounced than in Case 1. Hence one would expect more weight to the nearest stations and a reduced error variance, which is indeed the case as can be observed from Table 4.5. Cases1 and 5 basically are representative for rainfall at a low and high aggregation level, respectively (e.g. daily data and monthly data).

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

Estimation of the spatial covariance function or variogram.

Generally the spatial correlation (and hence the spatial covariance) as a function of distance will show a huge scatter as shown in Figures 2.1 to 2.4 of Chapter 2. To reduce the scatter the variogram is being estimated from average values per distance interval. The distance intervals are equal and should be selected such that sufficient data points are present in an interval but also that the correct nature of the spatial correlation is reflected in the estimated variogram.

Alternative to kriging

HYMOS offers an alternative to station weight determination by kriging through the inverse distance method. In this method the station weights and estimate are determined by

$$Pe_0 = \sum_{k=1}^{N} w_{0,k}.P_k \qquad w_{0,k} = \frac{1/d_k^p}{\sum_{j=1}^{N} 1/d_j^p} \tag{4.21}$$

It is observed that the weights are proportional to the distance between “0” and station j to some power p. For rainfall estimation often p = 2 is applied.

Different from kriging the **Inverse Distance Method** does not take account of station **clusters**, which is convincingly shown in Table 4.5, last row; the estimate for “0” is seen to be almost entirely determined by the cluster (1, 2) which is nearest to “0”. Hence, this method is to be applied only when the stations are more or less evenly distributed and clusters are not existing.

Example 4.4:

Application of Kriging Method

The kriging method has been applied to monthly rainfall in the BILODRA catchment, i.e. the south-eastern part of the KHEDA basin in Gujarat. Daily rainfall for the years 1960-2000 have been aggregated to monthly totals. The spatial correlation structure of the monthly rainfall for values > 10 mm (to eliminate the dry periods) is shown in Figure 4.20.

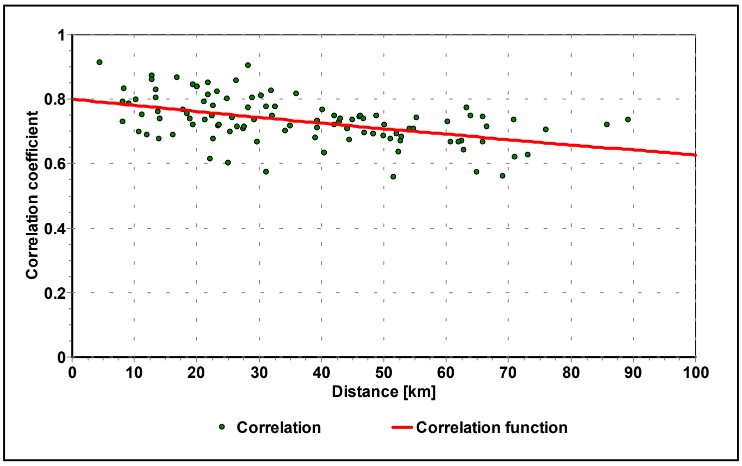
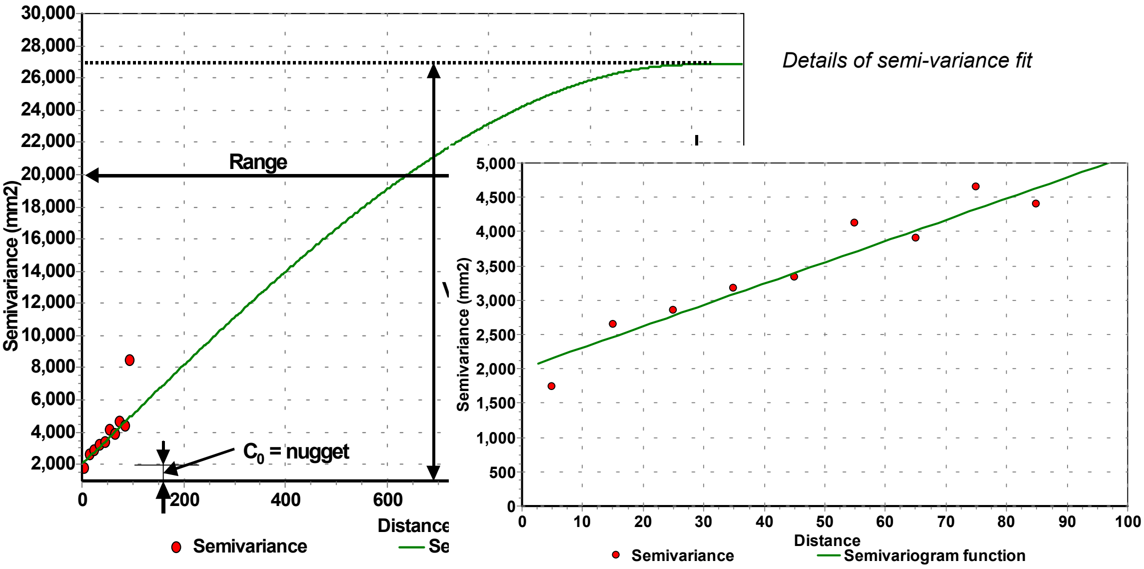


Figure 4.20: Spatial correlation structure of monthly rainfall data in and around Bilodra catchment (values > 10 mm)

From Figure 4.20 it is observed that the correlation only slowly decays. Fitting an exponential correlation model to the monthly data gives: $r_0 \approx 0.8$ and d_0 = 410 km. The average variance of the monthly point rainfall data (>10 mm) amounts approx. 27,000 mm^2. It implies that the sill of the semi-variogram will be 27,000 mm^2 and the range is approximately 1200 km ($\approx 3\ d_0$). The nugget is theoretically $\sigma_P^2(1\text{-}r_0)$, but is practically obtained by fitting the semi-variogram model to the semi-variance versus distance plot. In making this plot a lag-distance is to be applied, i.e. a distance interval for averaging the semi-variances to reduce the spread in the plot. In the example a lag-distance of 10 km has been applied. The results of the fit overall and in detail to a spherical semi-variogram model is shown in Figure 4.21. A nugget effect (C_0) of 2000 mm^2 is observed.



Details of semi-variance fit

Figure 4.21: Fit of spherical model to semi-variance, monthly rainfall Bilodra

Similarly, the semi-variance was modelled by the exponential model, which in Figure 4.22 is seen to fit in this case equally well, with parameters C_0 = 2,000 mm^2, Sill = 27,000 mm^2 and Range = 12,00 km. Note that C_0 is considerably smaller than one would expect based on the spatial correlation function, shown in Figure 4.20. To arrive at the nugget value of 2,000 mm^2 an r_0 value of 0.93 would be needed. Important for fitting the semi-variogram model is to apply an appropriate value for the lag-distance, such that the noise in the semi-variance is substantially reduced.

The results with the spherical model applied to the rainfall of June 1984 in the BILODRA catchment is shown in Figure 4.23. A grid-size of 500 m has been applied. The variance of the estimates is shown in Figure 4.24. It is observed that the estimation variance at the observation points is zero. Further away from the observation stations the variance is seen to increase considerably. Reference is made to Table 4.6 for a tabular output.

For comparison reasons also the isohyets derived by the inverse distance method is shown, see Figure 4.25. The pattern deviates from the kriging results in the sense that the isohyets are more pulled towards the observation stations. As was shown in the sensitivity analysis, the nearest station(s) weigh heavier than in the kriging method.

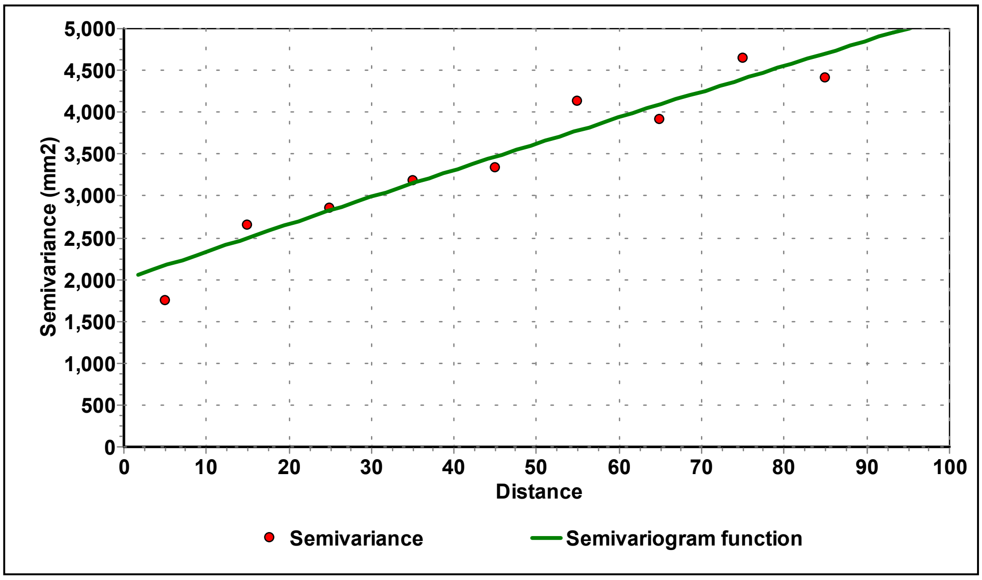


Figure 4.22: Fit of exponential model to semi-variogram, monthly data Bilodra

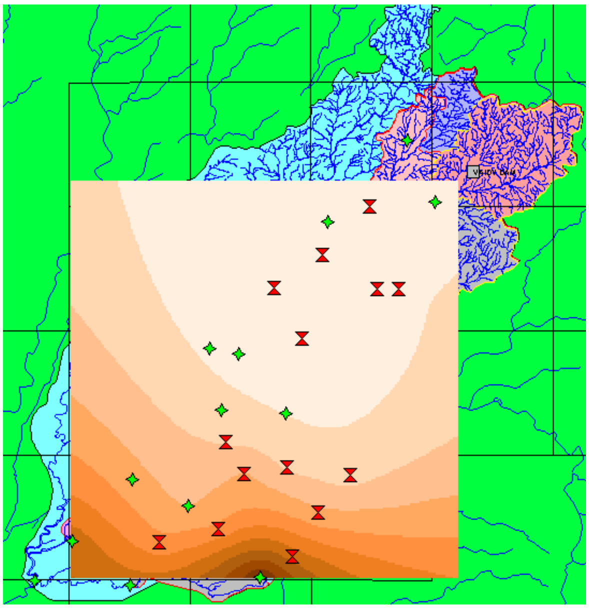


Figure 4.23:

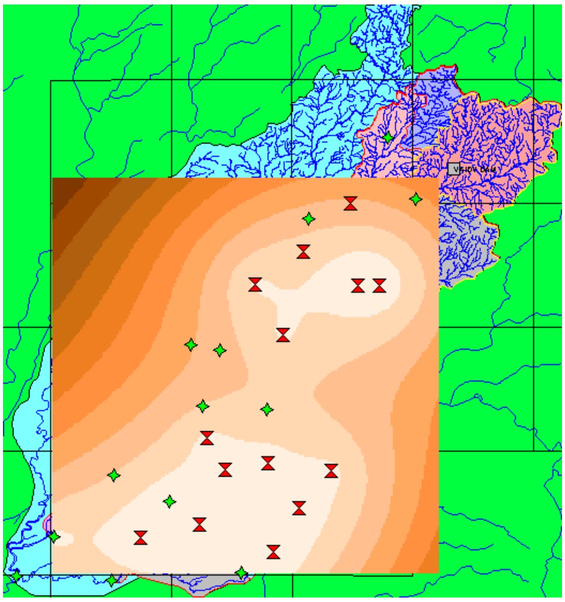


Figure 4.24:

Isohyets of June 1984 rainfall in Bilodra catchment using by spherical semi-variogram model (Figure 4.23) and the variance of the estimates at the grid-points (Figure 4.24).

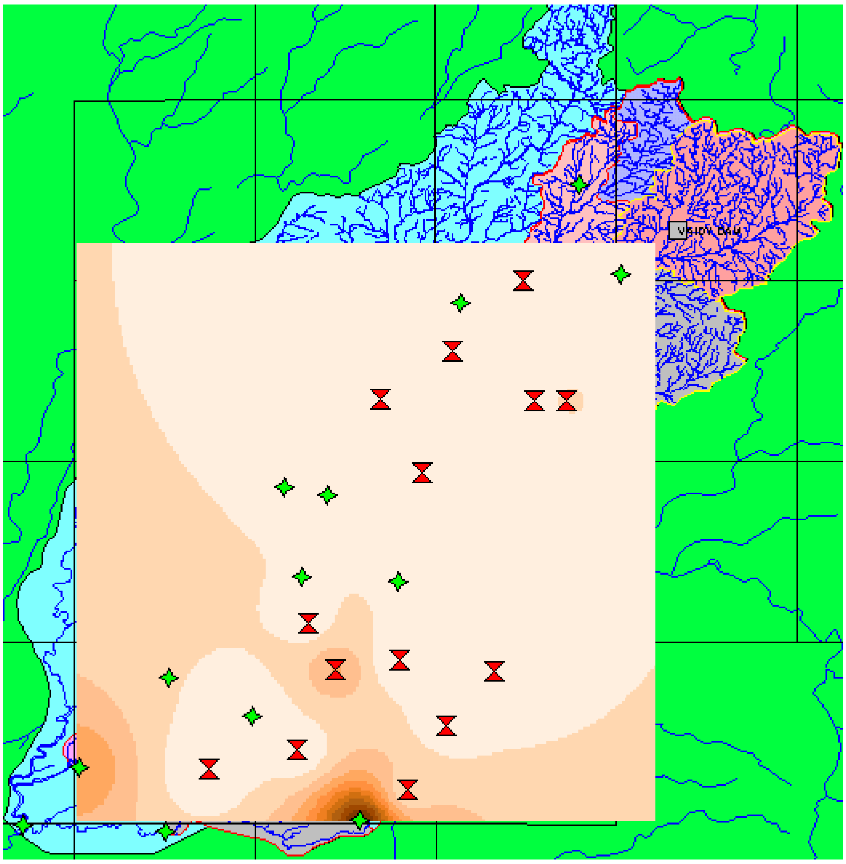


Figure 4.25: Isohyets derived for June 1984 rainfall in Bilodra catchment using inverse distance weighting (compare with Figure 4.23)

```
Variogram parameters

   Nugget (C0):                 2000.000000
   Sill (C1)                   25000.000000
   Range (a):                   1200.000000

 Grid characteristics:
   Number of cells in X, Y:         200          200
   Origin of X and Y Blocks:    0.000000E+00     0.000000E+00
   Size of X and Y Blocks:      5.000000E-01     5.000000E-01

 Search Radius:                 1.000000E+10
 Minimum number of samples:            4
 Maximum number of samples:           15

Data: ANIOR      MP2 1 at       65.473      63.440 value:     10.40000
Data: BALASINOR MP2 1 at        59.317      21.981 value:       .00000
Data: BAYAD      MP2 1 at       49.430      52.552 value:      1.00000
Data: BHEMPODA   MP2 1 at       70.017      63.390 value:     18.70000
Data: DAKOR      MP2 1 at       39.945       -.552 value:    176.00000
Data: KAPADWANJ MP2 1 at        32.921      29.687 value:     11.00000
Data: KATHLAL    MP2 1 at       24.756      15.644 value:       .00000
Data: MAHEMDABADMP2 1 at          .122       7.998 value:     68.20000
Data: MAHISA     MP2 1 at       31.242      10.327 value:       .00000
```

Contd.

```
Data: MAHUDHA   MP2 1 at       18.654       7.421 value:       .00000
Data: SAVLITANK MP2 1 at       36.817      22.560 value:     54.00000
Data: THASARA   MP2 1 at       46.848       3.977 value:     22.00000
Data: VADAGAM   MP2 1 at       43.604      64.007 value:       .00000
Data: VADOL     MP2 1 at       45.951      23.984 value:       .00000
Data: VAGHAROLI MP2 1 at       52.382      13.755 value:      5.00000

 Estimated        40000 blocks
 average        17.581280
 variance      101.393300

 Column        Row        Estimate       Variance
    1           1       45.806480     2685.862000
    1           2       45.719250     2680.906000
    1           3       45.625660     2676.289000
    1           4       45.525860     2672.001000
    1           5       45.420020     2668.018000

           etc.
```

Table 4.6: Example output of interpolation by kriging

4.4 TRANSFORMATION OF NON-EQUIDISTANT TO EQUIDISTANT SERIES

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

HYMOS provides a means of transforming such non-equidistant series to equidistant series by accumulating each unit tip measurement to the corresponding time interval. All those time interval for which no tip has been recorded are filled with zero values.

4.5 COMPILATION OF MINIMUM, MAXIMUM AND MEAN SERIES

The annual, seasonal or monthly maximum series of rainfall is frequently required for flood analysis, whilst minimum series may be required for drought analysis. Options are available in HYMOS for the extraction of minimum, maximum, mean, median and any two user-defined percentile values (at a time) for any defined period within the year or for the complete year.

For example if the selected time period is ‘monsoon months’ (say July to October) and the time interval of the series to be analysed is ‘ten daily’, then the above statistics are extracted for every monsoon period between a specified start and end date.

Example 4.5

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

These statistics are shown graphically in Figure 4.26 and are listed in tabular form in Table 4.7. Data of one of the year (1975) is not available and is thus missing. Lot of inferences may be derived from plot of such statistics. Different pattern of variation between 25 %ile and 90 %ile values for similar ranges of values in a year may be noticed. Median value is always lower than the mean value suggesting higher positive skew in the ten daily data (which is obvious owing to many zero or low values). A few extreme values have been highlighted in the table for general observation.

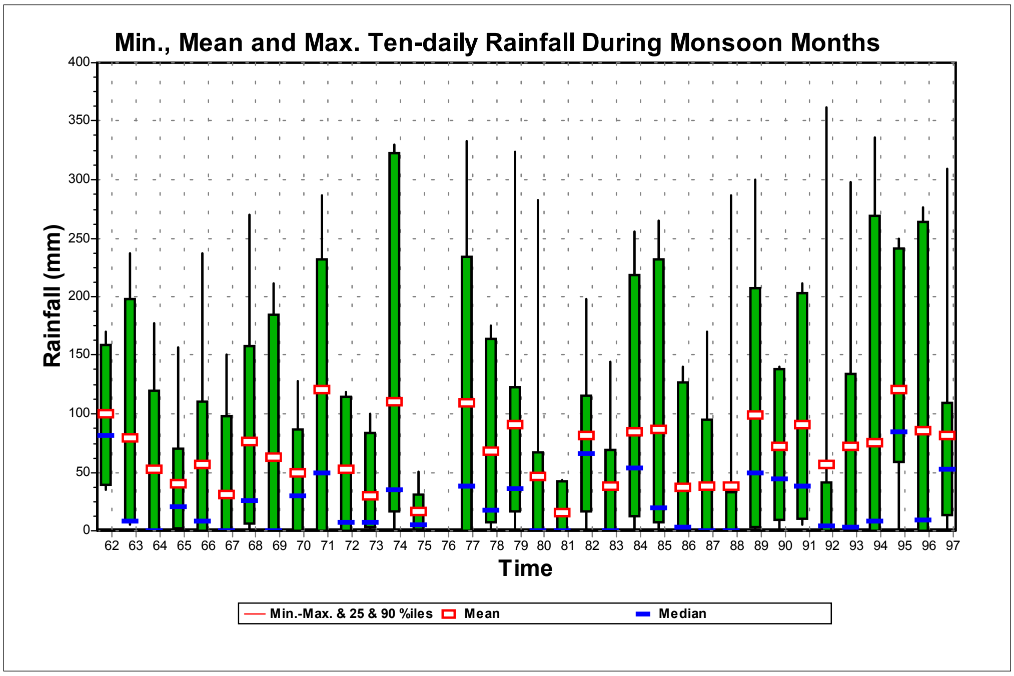


Figure 4.10: Plot of statistics of ten-daily rainfall series at MEGHARAJ station

year	Min.	Max.	Mean	Median	25 %ile	90 %ile
1961	34.54	170.39	99.6	81.03	39.36	158.47
1962	5.6	237.6	78.9	8.6	8.4	197.5
1963	0	177.44	53.0	0	0	119.1
1964	0	157.2	39.7	20.7	1.7	69.6
1965	0	237	56.3	8	0	110.6
1966	0	151	31.4	0	0	98
1967	0	270	75.9	26	6	158
1968	0	211	63.0	0	0	185
1969	0	128	49.2	30	0	87
1970	0	287	**120.7**	50	0	232
1971	0	118.5	53.1	7	0	114
1972	0	99.6	29.9	7	2.6	83.3
1973	0	**330.4**	110.8	34.8	17	**322.6**
1974	0	**51**	16.5	5	1.5	**31.2**
1976	0	**333.4**	108.8	38.2	0	234.2
1977	0	175.4	67.6	18	7	164
1978	0	324	90.3	36	16	123
1979	0	282	46.0	0	0	67
1980	0	**43**	**15.3**	0	0	**42**
1981	0	198	81.0	65.5	16	115.5
1982	0	144	38.5	0	0	69
1983	0	256	84.7	54	12	219
1984	0	265	87.0	19.5	7.5	231.5
1985	0	140.5	36.9	3	0	127
1986	0	170	38.4	0	0	94.5
1987	0	287	38.5	0	0	**33**

year	Min.	Max.	Mean	Median	25 %ile	90 %ile
1988	0	300	99.0	50	3	207
1989	0	140	72.3	44.5	9	138.5
1990	5	211.5	91.1	38.5	10	203.5
1991	0	**361.5**	56.7	4	0	41.5
1992	0	298	72.2	3	0	134
1993	0	336.5	75.7	8	0	269
1994	0	249	**121.1**	85	**58.5**	241.5
1995	0	276.5	85.9	9.5	0	264
1996	0	309	81.9	52.5	13.5	109
1997	0	**391**	105.7	23	10	242.5
Full Period	0	**391**	68.7			

Table 4.7: Ten-daily statistics for MEGHARAJ station between 1st July and 30st Sept.

5 SECONDARY VALIDATION OF CLIMATIC DATA

5.1 GENERAL

Secondary validation will be carried out at Divisional offices. Data which has been primary validated will be forwarded from the Sub-divisional office to the Divisional office in files generated by the primary module, and identifying and annotating all flagged values. Any hard copy field data relating to suspect data will also be forwarded.

Secondary validation is mainly concerned with spatial comparisons between neighbouring stations to identify anomalies in recording at the station. Some secondary validation more appropriately carried out at Divisional level is concerned with investigation of a single series but with long data series rather than simply with current data. Such testing may not be possible until a significant amount of historical data has been added to the database.

Spatial validation of climatic data is not so much concerned with individual values as with the generality of values received from a station. This is often best illustrated by the use of aggregated data

Procedures used in secondary validation apply over a range of variables. However their applicability may vary with the spatial variability of the variable. Some of the methods have already been described for rainfall in Module 9. They will be mentioned but not described in full again.

In interpreting the results of secondary validation, it is essential to be aware of the physical properties, limits and controls of a variable and the method of measurement. This has been outlined in more detail in Module 16 on primary validation.

Secondary validation is designed to detect anomalies in time series such as trends or changes in spatial relationships. The user should be warned that these may result from real environmental changes as well as data error. Data processors should be careful not to adjust data unless they are confident that the data or a portion of the data are incorrect rather than due to a changed microclimate. The existence of trend should be noted in the station record and supplied to data users with the data. For some analytical purposes data users may wish to adjust for the trend, in others to retain it.

5.2 METHODS OF SECONDARY VALIDATION

- Multiple station validation
  - Comparison plots of stations
  - balance series
  - regression analysis
  - double mass curve
  - test of spatial homogeneity (nearest neighbour analysis)
- Single station validation tests for homogeneity
  - mass curves
  - residual mass curves
  - a note on hypothesis testing
  - Student's 't' test of difference of means
  - Wilcoxon W-test on the difference of means
  - Wilcoxon-Mann-Whitney U-test to determine whether series are from the same population

The section on multiple station validation is placed first in the text as it is generally chronologically first to be carried out at the Divisional office.

5.3 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 pan evaporation, minimum or maximum temperature 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 and maximum pan evaporation 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.

5.4 MULTIPLE STATION VALIDATION

5.4.1 COMPARISON PLOTS

The simplest and often the most helpful means of identifying anomalies between stations is in the plotting of comparative time series. This should generally be carried out first, before other tests. For climate variables the series will usually be displayed as line graphs of a variable at two or more stations where measurements have been taken at the same time interval, such as 0830 dry bulb temperature, atmospheric pressure or daily pan evaporation.

In examining current data, the plot should include the time series of at least the previous month to ensure that there are no discontinuities between one batch of data received from the station and the next - a possible indication that the wrong data have been allocated to that station.

For climatic variables which have strong spatial correlation, such as temperature, the series will generally run along closely parallel, with the mean separation representing some locational factor such as altitude. Abrupt or progressive straying from this pattern will be evident from the comparative plot which would not necessarily have been perceived at primary validation from the inspection of the single station. An example might be the use of a faulty thermometer, in which there might be an abrupt change in the plot in relation to other stations. An evaporation pan affected by leakage may show a progressive shift as the leak develops. This would permit the data processor to delimit the period over which suspect values should be corrected.

Comparison of series may also permit the acceptance of values flagged as suspect in primary validation because they fell outside the warning range. Where two or more stations display the same behaviour there is strong evidence to suggest that the values are correct. An example might be the occurrence of an anomalous atmospheric pressure in the vicinity of a tropical cyclone.

Comparison plots provide a simple means of identifying anomalies but not of correcting them. This may be done through regression analysis, spatial homogeneity testing (nearest neighbour analysis) or double mass analysis.

5.4.2 BALANCE SERIES

An alternative method of displaying comparative time series is to plot the differences. This procedure is often applied to river flows along a channel to detect anomalies in the water balance but it may equally be applied to climatic variables to detect anomalies and to flag suspect values or sequences. Considering Z_i as the balance series of the two series X_i and Y_i , the computations can be simply done as:

$$Z_i = X_i - Y_I \qquad (5.1)$$

HYMOS provides this option under “Balances”. Both the original time series and their balances can be plotted on the same figure. Anomalous values are displayed as departures from the mean difference line.

5.4.3 DERIVATIVE SERIES

For scrutinising the temporal consistency it is very useful to check on the rate of change of magnitude at the consecutive time instants. This can be done by working out a derivative series. The derivative of a series is defined as the difference in magnitude between two time steps. The series Z_I of a series X_I is simply the difference between the consecutive values calculated as:

$Z_i = X_i - X_{i-1}$ (5.2)

Together with the original series the derivative series can be plotted against the limits of maximum rate of rise and maximum rate of fall. This gives a quick idea of where the rate of rise or fall is going beyond the expected values.

5.4.4 REGRESSION ANALYSIS

Regression analysis is a very commonly used statistical method. In the case of climatic variables where individual or short sequences of anomalous values are present in a spatially conservative series, a simple linear relationship with a neighbouring station of the form:

$Y_i = a X_i + c$ (5.3)

may well provide a sufficient basis for interpolation.

In a plot of the relationship, the suspect values will generally show up as outliers but, in contrast to the comparison plots, the graphical relationship provides no indication of the time sequencing of the suspect values and whether the outliers were scattered or contained in one block.

The relationship should be derived for a period within the same season as the suspect values. (The relationship may change between seasons). The suspect values previously identified should be removed before deriving the relationship, which may then be applied to compute corrected values to replace the suspect ones.

In HYMOS the validation section provides a section on “Relation Curves” which gives a number of variants of regression analysis including polynomial regression and introduction of time shifts, generally more applicable to river flow than to climate. A more comprehensive description of regression analysis is given in the Chapter on “Series completion and regression”

5.4.5 DOUBLE MASS CURVES

Double mass curve analysis has already been described in the secondary validation of rainfall (Section 2.12) and a full description will not be repeated here. It may also be used to show trends or inhomogeneities between climate stations but it is usually used with longer, aggregated series. However in the case of a leaking evaporation pan, described above, the display of a mass curve of daily values for a period commencing some time before leakage commenced, the anomaly will show up as a curvature in the mass curve plot.

This procedure may only be used to correct or replace suspect values where there has been a systematic but constant shift in the variable at the station in question, i.e., where the plot shows two straight lines separated by a break of slope. In this case the correction factor is the ratio of the slope of the adjusted mass curve to the slope of the unadjusted mass curve. Where there has been progressive departure from previous behaviour, the slope is not constant as in the case of the leaking evaporation pan, and the method should not be used.

5.4.6 SPATIAL HOMOGENEITY (NEAREST NEIGHBOUR ANALYSIS)

This procedure has already also been described in Section 2.6 for rainfall for which it is most commonly used and will not be covered fully again. Its advantage for rainfall in comparison to climate is that there are generally more rainfall stations in the vicinity of the target station than there are climate stations. The advantage for some climate variables is that there is less spatial variability and the area over which comparison is permitted (the maximum correlation distance R_{max}) may be increased.

Although there is strong spatial correlation, there may be a systematic difference due, for example to altitude for temperature. In these cases normalised rather than actual values should be used. This implies that the observations at the neighbour stations are multiplied by the ratio of the test station normal and the neighbour station normal:

$$T_{ci} = (N_{test} / N_i).\ T_I \qquad (5.4)$$

where: T_{ci} = Variable corrected for difference of mean at neighbour station

N_{test} = Mean of test station

N_i = Mean of neighbour station i

The analysis provides an estimate of a variable at a target station on the basis of a weighted mean of the neighbour stations, weighted as a function of the distance from the target station. It provides a list of those values (flagged values + or -) which are outside a specified range (mean + standard deviation times a multiplier), and provides the option of replacing the 'observed' values with the 'estimated' values.

5.5 SINGLE SERIES TESTS OF HOMOGENEITY

Single series testing for homogeneity will normally only be used with long data sets and therefore will have to await the data entry of historical data. Once these are in place it will be necessary to inspect them for homogeneity and especially for trend. Even here it is expected that spatial comparison of two or more series will be more commonly, but not exclusively used.

It is not expected that these methods will be widely used for current data.

Series may be inspected graphically for evidence of trend and this may often be a starting point. However statistical hypothesis testing can be more discriminative in distinguishing between expected variation in a random series and real trend or more abrupt changes in the characteristics of the series with time.

5.5.1 TREND ANALYSIS (TIME SERIES PLOT)

A series can be considered homogeneous if there is no significant linear or curvilinear trend in the time series of the climatic element. The presence of trend in the time series can be examined by graphical display and/or by using simple statistical tests. The data are plotted on a linear or semi-logarithmic scale with the climatic variable on the Y-axis and time on the X-axis. The presence or absence of trend may be seen by examination of the time series plot. Mathematically one may fit a linear regression and test the regression coefficients for statistical significance.

Trend generally does not become evident for a number of years and so the tests must be carried out on long data series, often aggregated into monthly or annual series. Trend may result from a wide variety of factors including:

- change of instrumentation
- change of observation practice or observer
- local shift in the site of the station
- growth of vegetation or nearby new buildings affecting exposure of the station

- effects of new irrigation in the vicinity of the station (affecting humidity, temperature and pan evaporation
- effects of the urban heat island with growing urbanisation
- global climatic change

The presence of trend does not necessarily mean that part of the data are faulty but that the environmental conditions have changed. Unless there is reason to believe that the trend is due to instrumentation or observation practices or observer, the data should not generally be altered but the existence of trend noted in the station record.

5.5.2 RESIDUAL MASS CURVE

A residual mass curve represents accumulative departures from the mean. It is a very effective visual method of detecting climatic variabilities or other inhomogeneities. The residual mass curve is derived as:

$$Y_i = Y_{I-1} + (X_i - m_x) = \sum_{j=1}^{i} (X_j - 1/N \sum_{k=1}^{N} X_k) \qquad (5.5)$$

where: N = number of elements in the series

m_x = mean value of X_i , i=1,N

The curve can be interpreted as follows:

- an upward curve indicates an above average sequence
- a horizontal curve indicates an about average sequence
- a downward curve indicates a below average sequence

5.5.3 A NOTE ON HYPOTHESIS TESTING

Hypothesis testing forms a framework for many statistical tests; it is concerned with an assumption about the distribution of a statistical parameter. The assumption is stated in the null-hypothesis H_0 and is tested against and alternative formulated in the H_1 hypothesis. The parameter under investigation is presented as a *standardised variate* called a *test statistic.* Under the null-hypothesis the test statistic has some standardised sampling distribution, e.g. a standard normal or a Student's t-distribution. For the null hypothesis to be true the value of the test statistic should be within the acceptance region of the sampling distribution of the parameters under the null-hypothesis. If the test statistic does not lie in the acceptance region, expressed as a significance level, the null-hypothesis is rejected and the alternative is assumed to be true. Some risk is involved however in being wrong in accepting or rejecting the hypothesis. For a brief note on Type I and Type II errors refer to the HYMOS manual

5.5.4 STUDENT'S T TESTS OF STABILITY OF THE MEAN

The series may be tested for stability of variance and mean. Considering only the simpler case where the variance has not changed, the test for stability of the mean requires computing and then comparing the means of two or three sub-sets of the time series. The 't' value is computed as follows and is compared with the tabulated Student's 't' which is the standardised test statistic, for corresponding degrees of freedom, for testing its significance:

$$|t| = |x_1 - x_2| / S_{12} \qquad (5.6)$$

and

$$S_{1,2} = \sqrt{\left\{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} * \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right\}} \quad (5.7)$$

where n_1 and n_2 are the number of data points in each sub-set, x_1 and x_2 are the corresponding means and S_1^2 and S_2^2 are the corresponding variance values. The number of values in each sub-set are usually taken as equal but if the graphical analysis indicates presence of trend in a particular part of the time series, the sub-sets can be decided accordingly.

Examples 5.1

Mean annual temperature is recorded at a station from 1978 to 1993 as follows:

Year	78 79 80 81 82 83 84 85 86	87 88 89 90 91 92 93
Annual mean temp. °C	18.5 16.5 17 19 18.5 18 18.5 17.5 19.5 Mean 18.1	17 15.7 16.5 17.3 16.5 15.5 17.5 Mean 16.5

By applying the above equation for stationarity it can be demonstrated that the data at station A is not-homogeneous.

A more robust test, can be carried out by comparison to homogeneity at adjacent stations which can in turn be used for adjustment of the non-homogeneous series if this seems appropriate.

5.5.5 WILCOXON W-TEST ON THE DIFFERENCE OF MEANS

The Wilcoxon test again tests under the null-hypothesis that the means of two series A_i (i = 1,m) and B_j (j = 1,n) are the same. All values of A_i are compared with all B_j defining a value of $w_{i,j}$ as follows:

$$\begin{aligned} \text{where: } A_i &< B_j \quad \text{then} \quad w_{i,j} = 2 \\ A_i &= B_j \quad \qquad\quad w_{i,j} = 1 \\ A_i &> B_j \quad \qquad\quad w_{i,j} = 0 \end{aligned} \quad (5.8)$$

The Wilcoxon statistic W is formed by:

$$W = \sum_{l=1}^{m} \sum_{j=1}^{n} w_{i,j} \quad (5.9)$$

Where the means are the same the W-statistic is asymtotically *normally* distributed with N (μ_w , σ_w)

where: $\mu_w = mn$

$\sigma_w^2 = mn\,(N+1)/3$

$N = m + n$

The absolute value of the following standardised test statistic is computed

$$|\,u\,| = |\,W - \mu_w\,| / \sigma_w \quad (5.10)$$

and comparison is made against a tabulation of the Normal distribution to test the validity of the null-hypothesis at the significance level.

5.5.6 WILCOXON MANN-WHITNEY U-TEST

The Wilcoxon-Mann-Whitney U-test investigates whether two series of length m and n are from the same population. The series may again may be split samples from a single series or series from paired instruments on the same site but as we would hardly expect series from different sites to be from the same population, this comparison is not recommended for such comparisons.

The data of both series are ranked together in ascending order. Tied observations are assigned the average of the tied ranks. The sum of ranks in each of the series, S_m and S_n, is then calculated. The U statistic is then computed as follows:

$$U_m = mn + m(m+1)/2 - S_m$$

$$U_n = mn - U_m$$

$$U = \min(U_m, U_n) \qquad (5.11)$$

A correction for tied values is incorporated as follows:

(a) tied observations are given the average rank, and

(b) the variance of U is corrected by a factor F_1

$$F_1 = 1 - \Sigma(t^3 - t)/(N^3 - N) \qquad (5.12)$$

where: t = number of observations tied for a given rank.

If the elements of the two series belong to the same population then U is approximately *normally* distributed with $N(\mu_U, \sigma_u)$:

$$\mu_u = mn/2$$

$$\sigma_U^2 = F_1 \, . \, mn(N+1)/12$$

where: $N = m + n$

For the null hypothesis that the series are from the same population, the absolute value of the following standardised test statistic is computed:

$$|u| = |U + c - \mu_u| / \sigma_u \qquad (5.13)$$

where: c = a continuity correction; $c = 0.5$ if $U < \mu_u$, and $c = -0.5$ if $U > \mu_u$

and comparison is made against a tabulation of the Normal distribution to test the validity of the null-hypothesis at the significance level.

6 CORRECTION AND COMPLETION OF CLIMATIC DATA

6.1 GENERAL

Procedures for correction and completion of climatic data are generally based on regression analysis, as the spatial correlation of climatic variables is considerable in many instances. The correlation structure may vary though with the seasons. Hence different relations will apply from one season to another. The regression estimate may either be used to replace an erroneous entry or to fill in a gap, but in case of a clear transcription error it is merely used to hint on the size of the correction to be applied. In the latter case the regression value is not used as such.

It has always to be verified whether good correlation exists with nearby stations. This may vary from location to location and from season to season.

In all cases it is **mandatory** to revalidate the series after correction and/or completion, to ensure that the series remain consistent with the other climatic variables observed at the site and with the same variable measured at nearby stations.

In this chapter the procedures are presented for the correction and completion of the following climatic variables:

- Temperature
- Humidity
- Wind
- Atmospheric pressure
- Solar radiation, and
- Pan evaporation

6.2 TEMPERATURE

The type of data entered/received for temperature include:

- Dry bulb temperature - sampling frequency daily or twice daily
- Wet bulb temperature - sampling frequency daily or twice daily
- Maximum thermometer - sampling frequency daily or twice daily
- Minimum thermometer - sampling frequency daily or twice daily
- Autographic recording of temperature - tabulated values on chart

The correction of temperature data depends on the type of error.

- In case of a misreading of a thermometer, generally by a full number of ^oC (1 or 5 or 10 ^oC) the correction to be applied is simply deduced considering:
- time series graph by observing surrounding values in the graph
- time series graph of related temperature parameters, and/or
- time series graph of the same variable observed at nearby stations.
- If the thermograph has been out of calibration for some time, the correction can be obtained from a comparison with the minimum and maximum thermometer readings. A series of daily minimum and maximum temperatures from the thermograph is established. These time series are first graphically compared with the thermometer readings to detect any sudden deviation and then subjected to a double mass analysis with the same to establish the start date of malfunctioning

and the type and size of error, similar to the procedure outlined for rainfall data. Corrections can be applied in a similar fashion (step-trend or linear trend will be most common). Check also whether the deviations are a function of temperature. If so, this has to be incorporated in the correction equation. Similarly, a comparison can be made with the thermograph/temperature records at nearby stations to detect/correct the records if the measurements at your station are doubtful.

If data are missing, the gaps may be filled in:

- In case one or a few data is/are missing from the thermograph record: by interpolation within the series, if nearby stations do not show abnormal change with time.
- In case a number of data is missing from the thermograph record: apply regression analysis on nearby stations, using a regression relation applicable for that season, and adjust the equation to match with the observed minimum and maximum temperatures measured at the thermograph location.
- In case a number of data is missing for maximum, minimum, dry and/or wet bulb temperature and no thermograph record is available: apply regression analysis on nearby stations, using a regression relation applicable for that season. The equation is adjusted to match with the last value before the gap and the first value after.
- In case a number of data is missing from the wet bulb temperature record, regression can be applied on relative humidity with a nearby station(s). Then the wet bulb temperature is obtained either from psychrometric tables (mind the pressure level!) or from the psychrometer relation and the relation between saturation vapour pressure and temperature (see Volume 3, Design Manual, Hydro-meteorology) from which an expression can be obtained for the wet bulb temperature as a function of dry bulb temperature, relative humidity and psychrometric constant. The equation has to be solved iteratively.

Note: after correction or infilling it is **mandatory** to execute consistency checks on the results, including:

- Dry bulb temperature should be greater than or (rarely) equal to the wet bulb temperature, but on wet days their difference should be small as the humidity is likely to be high.
- Maximum temperature should be greater than minimum temperature.

6.3 HUMIDITY

The type of data entered/received for humidity include:

- Relative humidity - historical data, previously computed
- Autographic recording of relative humidity - tabulated values on chart

The historical records of relative humidity can be corrected, first of all by recomputing the variable from the dry and wet bulb temperatures if available. Else, regression relations may be established with nearby stations to correct or fill in the missing values. Note that if the erroneous value is due to a transcription error, the regression equation may hint on the size of the wrong entry and the error is adjusted accordingly, rather than to use the regression estimate.

Hygrometer records can be corrected or filled in by means of regression on nearby stations, adjusted by local instantaneous values from dry and wet bulb temperatures when available. The filled in record should always be rechecked on consistency and continuity of the record at the start and end of the gap.

6.4 WIND

The type of data entered for wind speed and direction include:

- Instantaneous wind speed - sampling frequency daily or twice daily
- Daily wind run - sampling frequency daily
- Wind direction - sampling frequency daily or twice daily

It is likely that when the FSC are properly located, that a good correlation regarding wind speed and direction exist between nearby stations. Hence, regression analysis will be an appropriate means to fill in missing data, or to correct erroneous data on wind speed and similarly for wind direction. Note that if the erroneous value is due to a transcription error, the regression equation may hint on the size of the wrong entry and the error is adjusted accordingly, rather than to use the regression estimate.

6.5 ATMOSPHERIC RADIATION

The type of data entered for atmospheric pressure include:

- Atmospheric pressure - sampling frequency daily or twice daily
- Autographic recording of atmospheric pressure - tabulated values on chart

As for the other climatic variables, a good correlation between atmospheric pressure between nearby FCS's may be expected, and hence regression analysis will be the appropriate tool to correct erroneous values or to fill in missing ones. Note that if the erroneous value is due to a transcription error, the regression equation may hint on the size of the wrong entry and the error is adjusted accordingly, rather than to use the regression estimate.

6.6 SOLAR RADIATION

The type of data entered for solar radiation include:

- Sunshine hours - from Campbell Stokes sunshine recorder card

It is to be seen from a comparison of records at nearby stations whether the number of sunshine hours show good correlation or not. If a good correlation is available, regression analysis may be applied for replacing erroneous data or to fill in missing ones. A correction/completion thus obtained should **always** be checked and adjusted against the rainfall record at the same station or nearby stations. Particularly the records of recording raingauges are of importance here, as they can put an upper limit to the number of sunshine hours.

6.7 PAN EVAPORATION

The type of data entered for pan evaporation include:

- Pan evaporation - sampling frequency daily or twice daily
- Pan water temperature - sampling frequency daily or twice daily

Pan evaporation data can be corrected or filled in based on evaporation estimates obtained from the climatic variables through the Penman method. It implies that the pan coefficient for the season has to be derived first, using the available reliable records on climate data and pan evaporation.

7 SECONDARY VALIDATION OF WATER LEVEL DATA

7.1 GENERAL

Water level data received at Divisional offices have already received primary validation on the basis of knowledge of instrumentation and methods of measurement at the field station and information contained in Field Record Books. Primary validation has included comparisons between different instruments and methods of observation at the same site.

The data processor must continue to be aware of field practice and instrumentation and the associated errors which can arise in the data,

Secondary validation at Division now puts most emphasis on comparisons with neighbouring stations to identify suspect values. Special attention will be given to records already identified as suspect in primary validation

The assumption, while carrying out secondary validation, is that the variable under consideration has adequate spatial correlation. Since the actual value of water level is controlled by specific physical conditions at the station, the amount of secondary validation of level is limited. Most of the check comparisons with neighbouring stations must await transformation from level to discharge through the use of stage discharge relationships. Only as discharge can volumetric comparisons be made. However validation of level will identify serious timing errors.

Secondary validation of level will be used to identify suspect values or sequences of values but not usually to correct the record, except where this involves a simple shift (time or reference level) of a portion of a record.

The main comparisons are between water level series at successive points on the same river channel. Where only two stations are involved, the existence of an anomaly does not necessarily identify which station is at fault. Reference will be made to the historic reliability of the stations.

Comparisons will also be made between incident rainfall and level hydrographs.

7.2 SCRUTINY OF MULTIPLE HYDROGRAPH PLOTS

Graphical inspection of comparative plots of time series provides a very rapid and effective technique for detecting timing anomalies and shifts in reference level. Such graphical inspection will be the most widely applied validation procedure.

For a given time period several level time series for neighbouring stations are plotted in one graph. For routine monthly validation, the plot should include the time series of at least the previous month to ensure that there are no discontinuities between one batch of data received from the station and the next. The time interval of observation rather than averaged values should be displayed. In general, peaks and troughs are expected to be replicated at several stations with earlier occurrence at upstream stations and the lag between peaks, based on the travel time of the flood wave, approximately the same for different events. It should be noted that level values at downstream stations are not necessarily higher than upstream stations - the actual value depends on physical conditions at the stations.

Where peaks occur at one station but not at its neighbour or where the lag time between stations is widely different from the norm, an error at one station may be suspected. However it must be recognised that the quality of the relationship between neighbouring hydrographs depends not only on the accuracy of the records but on a variety of other factors including:

- rainfall and inflow into the intervening reach between stations. If the intervening catchment is large or the rainfall high in comparison to that over the upstream basin, a very poor relationship may result.
- river regulation and abstractions between the stations may obscure natural variations, though high flows are usually less affected than low or medium flows.
- An average lag between successive stations can be used in making comparisons but the actual lag is variable, generally diminishing up to bankfull stage and increasing again with overbank flow.
- one station may suffer backwater effects on the stage hydrograph and not another, obscuring the effects of differences in flow. Where such effects are known to occur, comparison should await transformation to discharge.

Anomalies identified from comparative hydrograph plots are flagged for further stage validation or to await validation as discharge.

Example 7.1

Application of the above described technique is demonstrated for the stations MAHEMDABAD and NSB00I7 on WATRAK river, a tributary of Sabarmati in Gujarat. The stations are distanced some 33 km apart (MAHEMDABAD d/s of NSB00I7) and the lateral inflow in between the sites is small compared to the river flow. The hydrographs of hourly water levels for the months September and October 1998 are shown in Figure 7.1. From the plot some anomalies are observed. Make sure to have always a tabulated output of the water level observations available when carrying out such analysis to note down possible anomalies.

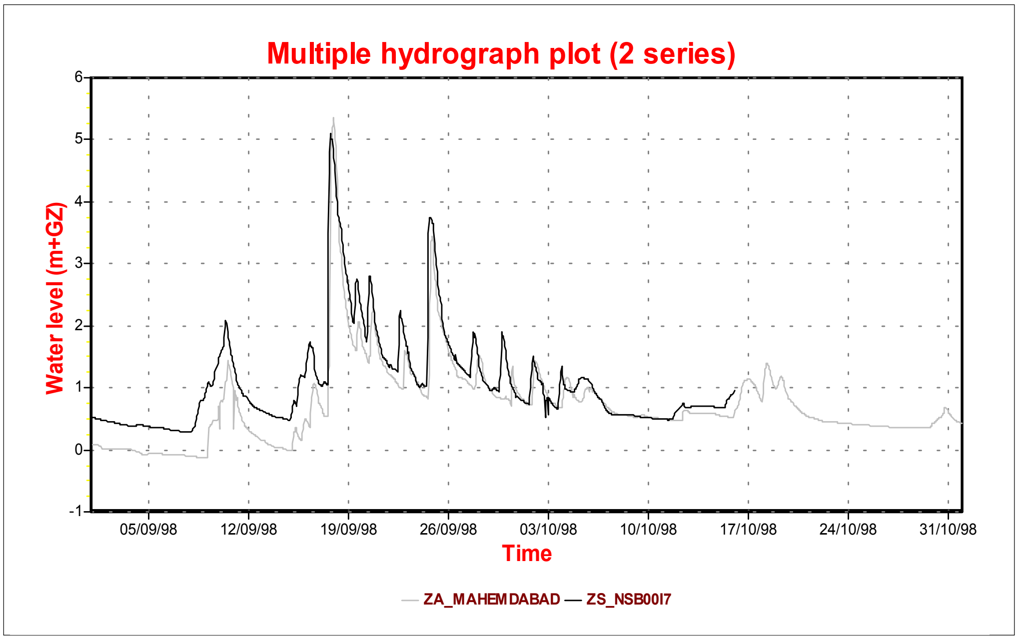


Figure 7.1 a: Multiple hydrograph plot

To get a better view one should zoom in. A detail of the hydrograph, showing an unnatural regular sequence of peaks is presented in Figure 7.1b. From the Figure two types of errors may be observed:

- spurious errors originating from transcription errors in the field or in the office, and
- errors in the occurrence of peaks

Particular with respect to the last type of error additional information is required to determine which parts of the hydrographs are faulty. Fortunately upstream as well as downstream of the stations hydrometric stations are available. The comparison with the stations upstream of NSB00I7 is presented in the following example.

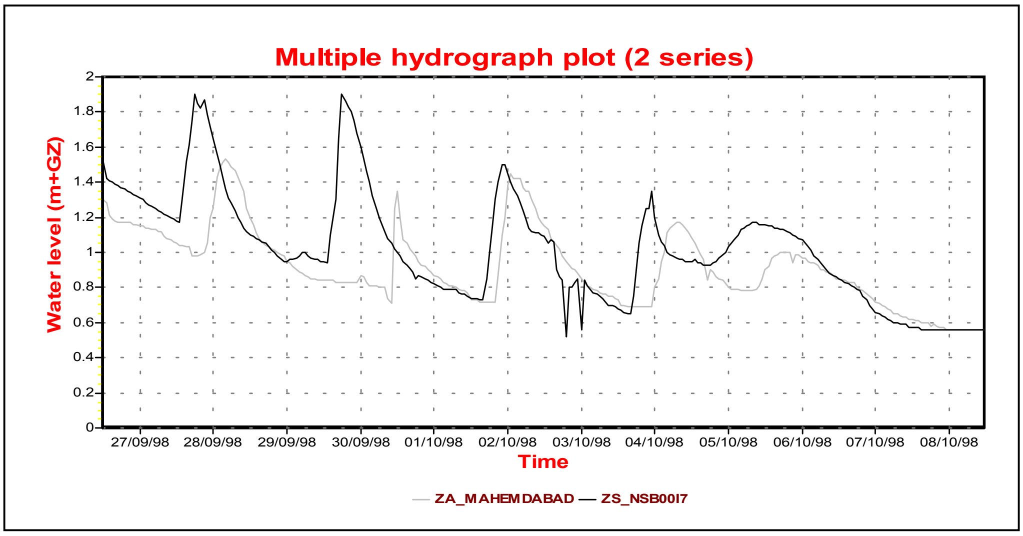


Figure 7.1b Detail of multiple hydrograph plot

Example 7.2

The hydrograph plot presented in Figures 7.1a and b is now extended with the hydrographs of the stations DHABA on WATRAK (upstream feeder river with reservoir) and AMBALIYARA on MAZAM (another feeder river). The hydrographs are shown in Figure 7.2a. and the detail for the same period of Figure 7.1b in Figure 7.2b.

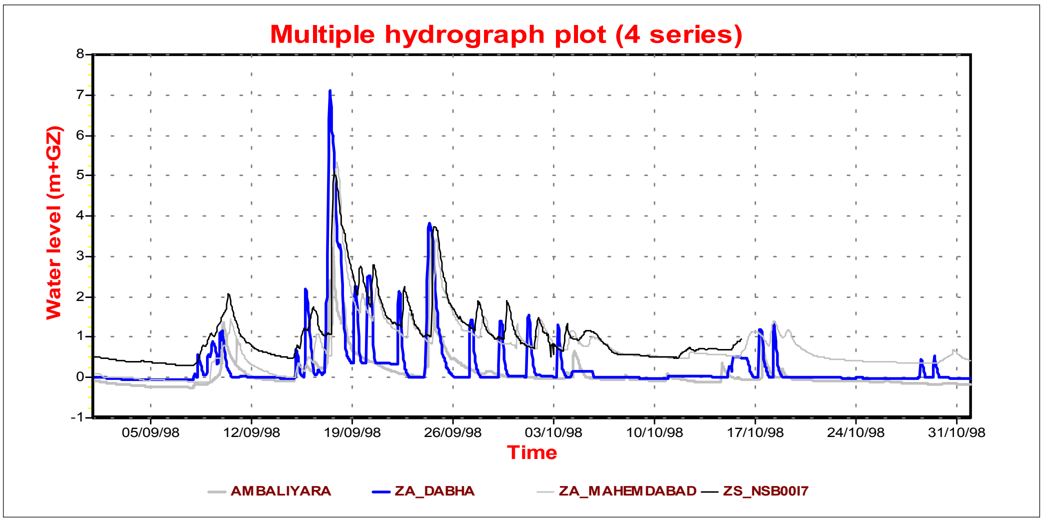


Figure 7.2a: Multiple hydrograph plot, 4 stations

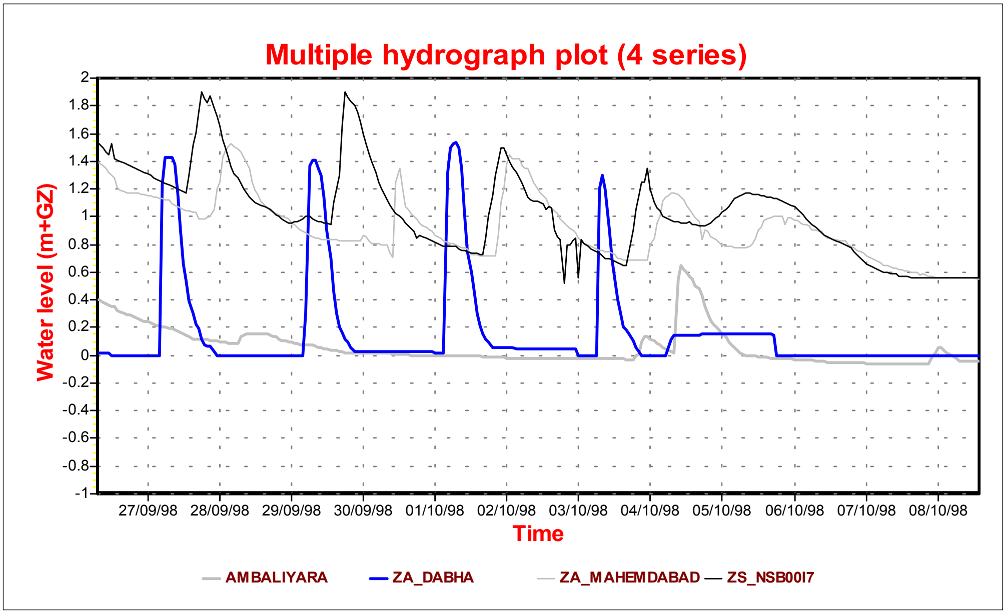


Figure 7.2b: Detail of Figure 7.2a

It is observed from Figure 7.2b that the first 4 peaks originate from upper WATRAK. These peaks are due to releases from an upstream reservoir. The last peak is recorded at AMBALIYARA. The records of DHABA and AMBALIYARA do not show distinct anomalies and hence can act as a guide to detect anomalies in the records of MAHEMDABAD and NSB00I7. To proceed, first the suspect parts of the hydrographs of MAHEMDABAD and NSB00I7 are noted down, based on resemblance of patterns. Possible anomalies (strange patterns, sharp rises and falls) are indicated in Figure 7.2c.

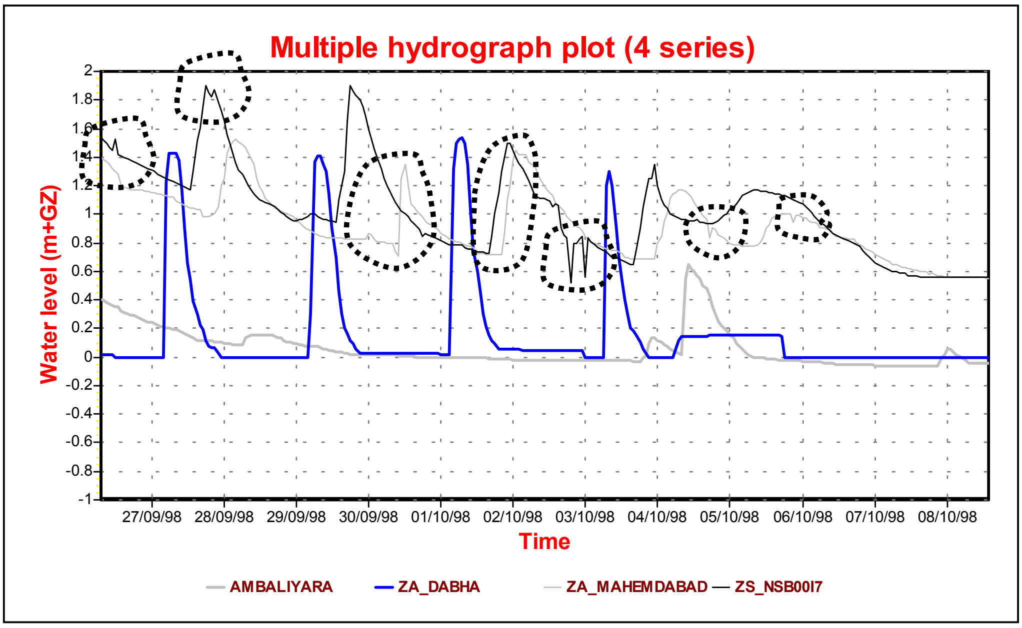


Figure 7.2c Identification of suspect values based on resemblance of hydrograph patterns

Secondly, anomalies in the time lag between the peaks are investigated, see Figure 7.2d.

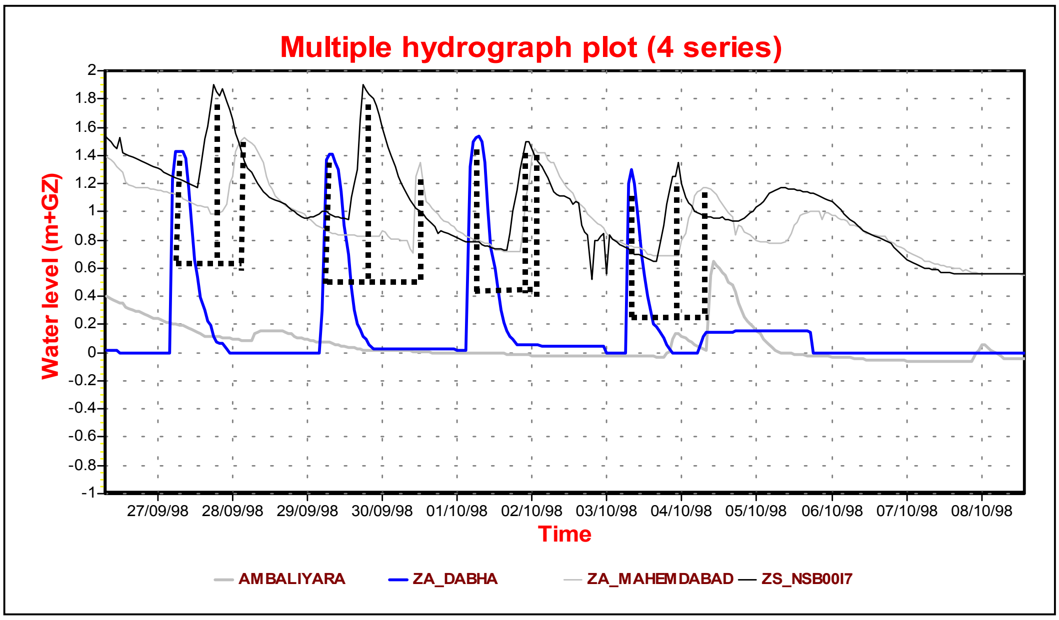


Figure 7.2d: Identification Anomalies

The time lags between the first set of peaks are taken as a guide, as no distinct anomalies are present. Next, these time lags are applied to the other sets of peaks as well. From the comparison it is observed that the second peak of MAHEMDABAD and the third peak of NSB00I7 are suspect. The time lags between the peaks in the fourth set resembles the time lags of the first set. Note that this comparison is only allowed for similar flow sizes; if the ranges are different, wave celerities may be different and so will be the time lags, see below.

7.3 COMBINED HYDROGRAPH AND RAINFALL PLOTS

The addition of rainfall to the comparative plots provides a means of assessing timing errors and of investigating the effects of inflow into the intervening reach between stations. Comparison may be made using an average rainfall determined using Thiessen polygons or other methods over the entire basin or for the intervening sub-basin corresponding to various gauging stations. Where the basin is small or the number of rainfall stations limited, individual rainfall records may be plotted.

In general a rise in river level must be preceded by a rainfall event in the basin and conversely it is expected that rainfall over the basin will be followed by rise in level. There must be a time lag between the occurrence of rainfall and the rise in level. Where these conditions are violated, an error in rainfall or in the level hydrograph may be suspected. However the above conditions do not apply universally and the assumption of an error is not always justified especially for isolated storms in arid areas. For example:

- An isolated storm recorded at a single raingauge may be unrepresentative and much higher than the basin rainfall. The resulting runoff may be negligible or even absent.
- Where storm rainfall is spatially variable, it may be heavy and widespread but miss all the raingauges, thus resulting in a rise in river level without preceding measured rainfall.
- The amount of runoff resulting from a given rainfall varies with the antecedent catchment conditions. Rainfall at the onset of the monsoon on a very dry catchment may be largely absorbed in soil storage and thus little reaches the river channel.

- The use of comparative plots of rainfall and level is therefore qualitative but it provides valuable ancillary information with the multiple hydrograph plots.

Example 7.3

An example of a combined hydrograph and rainfall plot is presented in Figure 7.3, which displays the water level record of station AMBALIYARA on MAZAM river together with the rainfall records of stations RAHOL and VADAGAM.

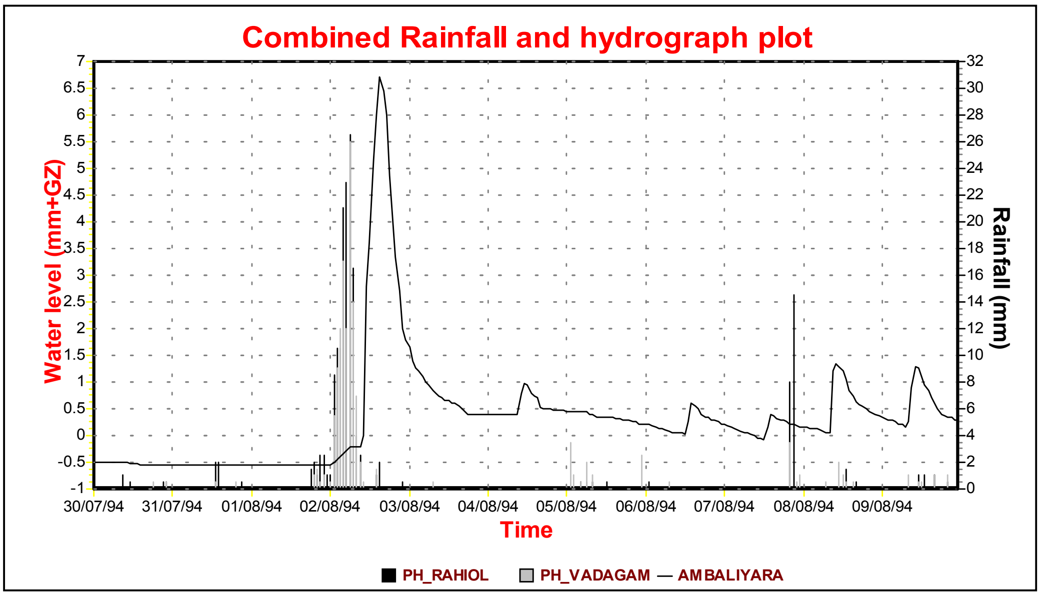


Figure 7.3: Combined hydrograph and rainfall plot

From the graph it is observed that the first peak is preceded by substantial rainfall. The remainder, however, shows suspect combinations of hydrographs of stage and rainfall, where hydrograph appears to occur before the rainfall and where the response to the rainfall is delayed. One may be tempted to doubt the rest of the record. However, in this case the peaks may as well be caused by releases from an upstream reservoir. Therefore, in this case additional information is required about the reservoir releases to conclude on the series. Nevertheless, the delayed response to the second rainfall peak remains suspicious.

7.4 RELATION CURVES FOR WATER LEVEL

7.4.1 GENERAL

A relation curve gives a functional relationship between two series of the form:

$$Y_t = F(X_{t+t1}) \qquad (7.1)$$

To account for the lag between level changes at one station and the next downstream, it may be necessary to introduce a time shift (t1) between the two time series.

Relation curves will normally be applied to water level data rather than discharge. However it may be appropriate on occasions to use them for discharge data, especially where one or both stations are affected by backwater conditions.

If there is a distinct one to one relationship between two series, random errors will be shown in a relation curve plot as outliers.

By comparing two relation curves, or data of one period with the curve of another period, shifts in the relationship, e.g., in the water level series due to changes in the gauge zero can be detected.

7.4.2 APPLICATION OF RELATION CURVES TO WATER LEVEL

If two water level stations are located on the same river and no major branch joins the main stream between the two locations, a relation can be expected between the recordings at the two locations. With the help of this relation, the stage at a particular station can be derived from the available data series of the adjacent station. A sample plot of such relationship between the two stations is shown in example Figure 7.4.

Two important conditions need to be satisfied to obtain a high degree of relationship between the stage data of adjacent stations. These are:

- No major tributary joins the main stream in between the two adjacent stations.
- Time of travel of the flood wave between the two stations is taken into consideration.
- Time of travel of the flood wave between the two stations is taken into consideration.

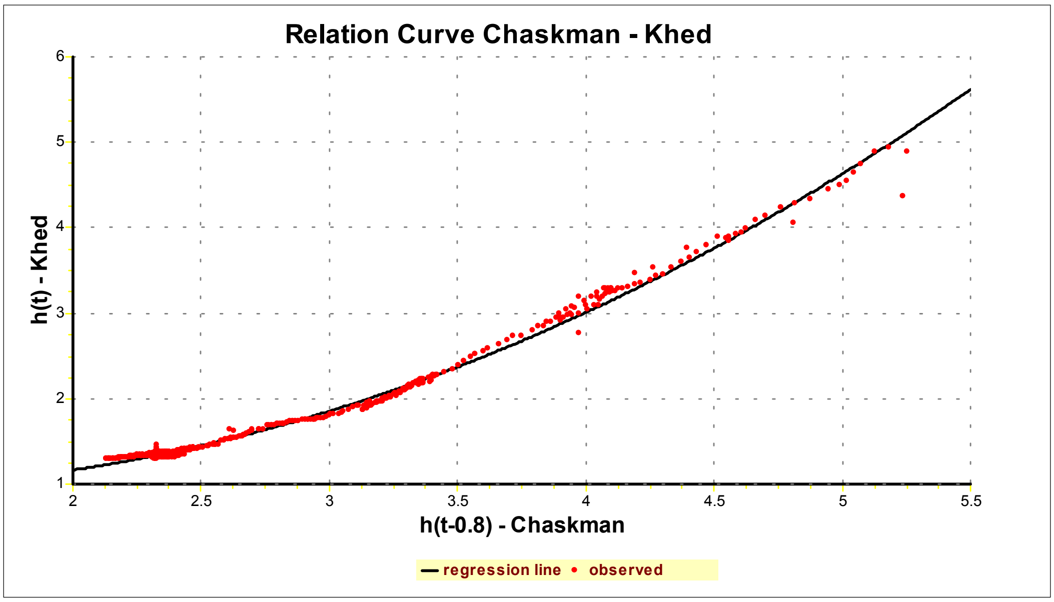


Figure 7.4: Example of relation curve

As noted above for comparative hydrograph plots, the occurrence of lateral inflow between stations limits the quality of the relationship between neighbouring stations. The lateral inflow may occur as a main tributary inflow or distributed over the reach as surface and groundwater inflows. In either case if it is a significant proportion of the downstream flow or variable, then good correlation may not be obtained.

7.4.3 DETERMINATION OF TRAVEL TIME

With respect to the second condition, the relationship between the two station time series must incorporate a time shift, representing the mean travel time of a flood wave between the stations. Plotting without a time shift results in a looped relationship as shown in Example 7.4. The time shift may be assessed using:

- physical reasoning, or
- from an analysis of the time series.

Example 7.4

A relation curve is to be established between the stations MAHEMDABAD and NSB00I7. The hydrographs for the period for which the relation curve is to be established is presented in Figure 7.5a. It is observed that the hydrograph at the downstream site is lagging behind. Plotting the observations at the two sites without a correction for travel time of the flood wave results in a looped scatter plot vas displayed in Figure 7.5b.

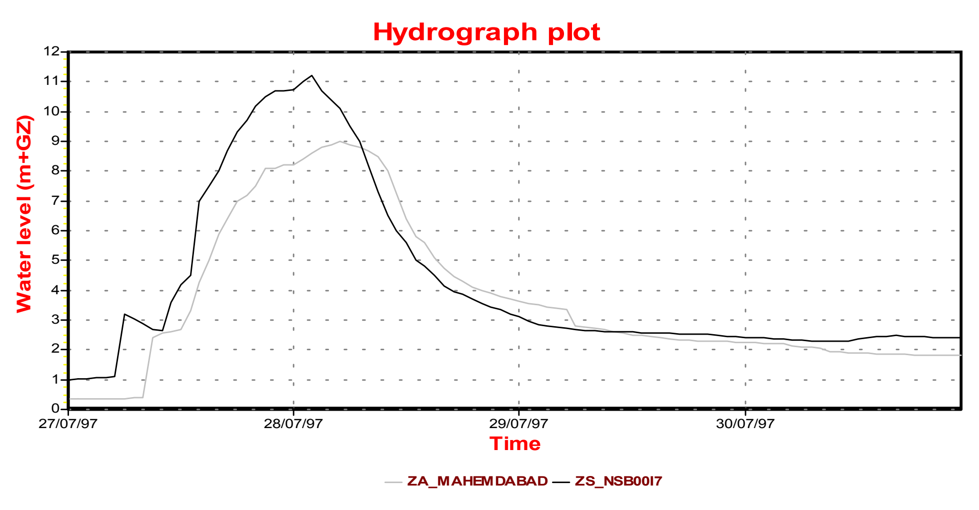


Figure 7.5a: Hydrographs along WATRAK

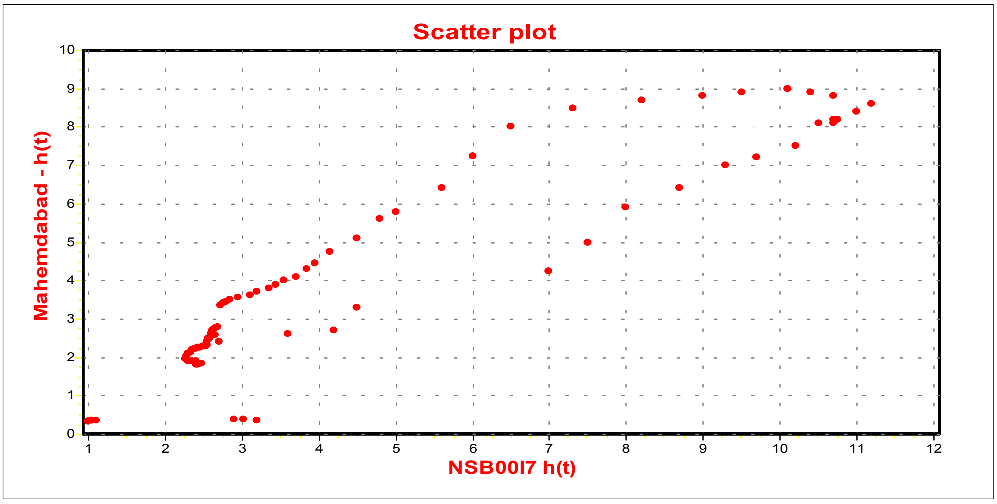


Figure 7.5b Scatter plot of h(t) Mahemdabad versus h(t) NSB00I7

From physical reasoning

The time of travel of a flood wave can be approximately determined by the division of the interstation distance by the estimated flood wave velocity. Consider a relation curve of daily water levels of stations X and Y which are spaced at a distance s km from each other along the same river, X upstream of Y. Let the average flow velocity (assessed from current meter gaugings at the stations) be u m/s, then the propagation velocity or celerity of the flood wave c is approximately equal to 1.7 (B_r/B_s).u, where B_r is the river width and B_s is the storage width (river+flood plain). So, if the river is in-bank, celerity becomes equal to 1.7u, and the time shift to be applied between X and Y for a one to one relationship amounts to distance/celerity = -s * 1000/(1.7 u) sec. When the river expands in the flood plain, the celerity is different from 1.7.u and should be multiplied with B_r/B_s ; consequently, a different time shift will result.

From cross correlation analysis

Another computational procedure to derive the time shift between the two related series is based on cross-correlation analysis. The estimate of the cross-correlation function (R_{xy} t) between the time series X(t) and Y(t) of two stations is computed for different lags τ. The lag time corresponding to the maximum of the cross correlation function indicates the travel time between the two stations.

A plot is made between the lag time and the related cross correlation coefficient . The lag time corresponding to the maximum of the cross-correlation function indicates the travel time between the two cross sections. After calculating the time shift, it can be applied by entering the data of station X at time T and corresponding it with the data of station Y at time T + τ. It is then advisable to plot the resulting X-Y relationship:

- to inspect for the absence of looping,
- to detect outliers
- to determine the existence of breakpoints in the relationship
- to assess the functional form of the relationship.

Example 7.4 (continued)

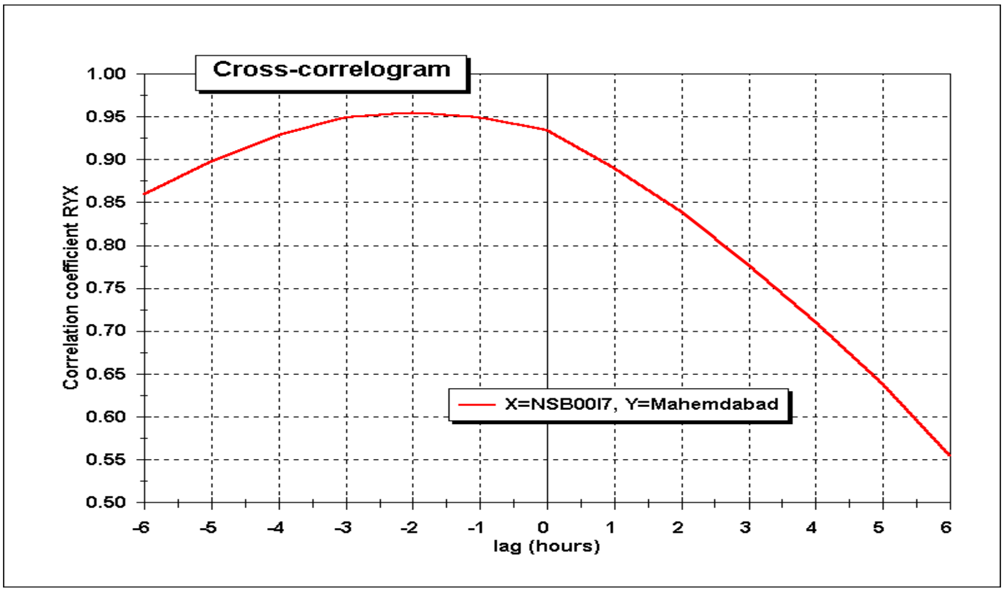


Figure 7.5c: Cross-correlation function to determine optimum time lag

Application of cross-correlation analysis to the series of MAHEMDABAD (Y-series) and NSB00I7 (X-series) reveals an optimum time shift of –2.8 hrs. From hydraulic calculations given a Manning roughness K_m = 33, a bed slope S = 3.6×10^{-4}, a flow depth h = 5 m, $B_r = B_s$ and L = 11 km one obtains roughly a travel time of 3 hrs (execute the calculation!) By applying such a time shift (i.e. h(t) MAHEMDABAD versus h(t-2.8) NSB00I7 the scatter shows only a limited looping left, see Figure 7.5 d (exact elimination is not possible as the optimum time shift varies with the celerity, i.e. with the flow rate and varying lateral inflow also distorts the relationship).

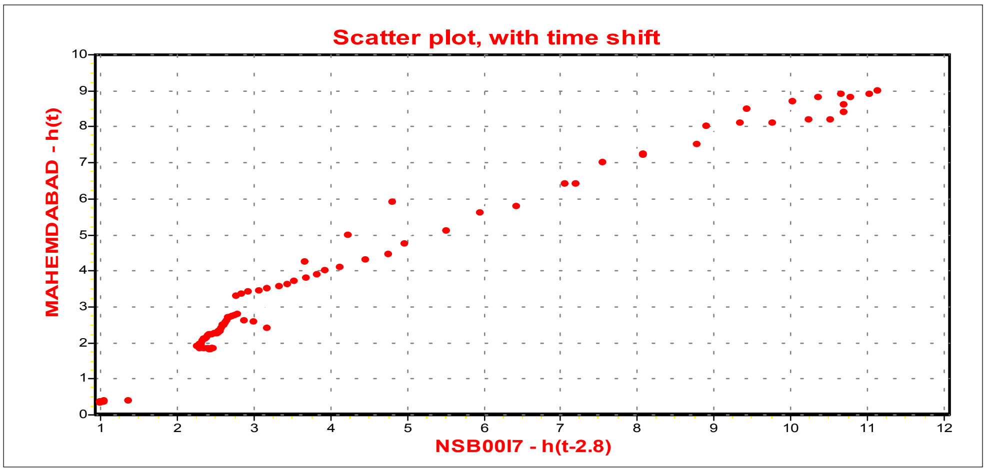


Figure 7.5d: Scatter plot with corrected for time lag

7.4.4 FITTING THE RELATION CURVE

The X-Y data can be fitted in HYMOS using a relation equation of polynomial type:

$$Y_t = c_0 + c_1 X_{t+t1} + c_2 X_{t+t1}^2 + c_3 X_{t+t1}^3 + \ldots\ldots\ldots \qquad (7.2)$$

After inspection of the scatter plot the form of the relationship (the degree of the polynomial) can be decided. It is recommended to use a polynomial not greater than order 2 or 3; in many instances a simple linear relationship will be acceptable. The least squares principle is applied to estimate the coefficients.

Where inspection of the scatter plot indicates the presence of breakpoints, then separate relationships may be established for different ranges of X (analogous to different ranges of the stage discharge relationship) with a maximum of 3 intervals of X.

Example 7.4 (continued)

The scatter plot shown in Figure 7.4d is fitted by a second order polynomial and is displayed in Figure 7.5e.

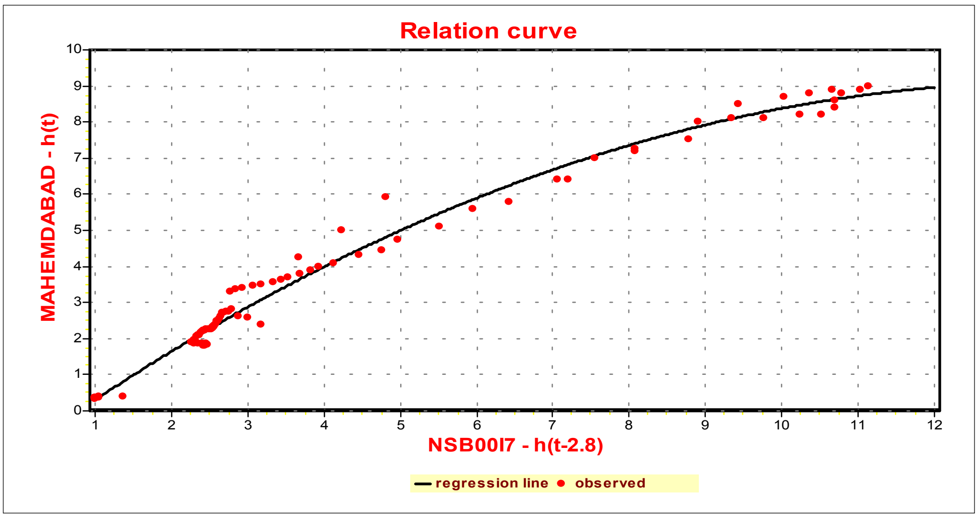


Figure 7.5e: Fitting of relation curve

7.4.5 USING RELATION CURVE FOR DATA VALIDATION

Relation curves can be displayed and plotted and relation equations valid for different periods can be compared. Shifts in the relationship between the two series indicate a physical change at one of the stations, such as shifts in gauge zero, changes in cross section or even relocation of station.

Where such shifts in relation curve are accompanied by changes in the stage discharge relationship at one station, the changed relation curve is acceptable. However, where no such accompanying change in stage discharge has been notified, an explanation should be sought from Sub-divisional staff or the field supervisor.

Example 7.5

To investigate any change in the relationship between the observations at MAHEMDABAD and NSB00I7 a relation curve was established for the next flood wave on WATRAK river. The hydrographs and relation curve for this second flood are shown in the Figures 7.6a and 7.6b. Next the this relation curve is compared with the one established above. Note that for the establishment of the second relation curve the same time shift was applied as before. The two relation curves are shown in Figure 7.6c. It is observed that the match between the two curves is almost perfect, indicating that the no gauge shifts nor changes in the control sections did take place during the passage of the floods. A shift in the gauge at one of the sites would have resulted in a shifted relation as displayed in Figure 7.6d

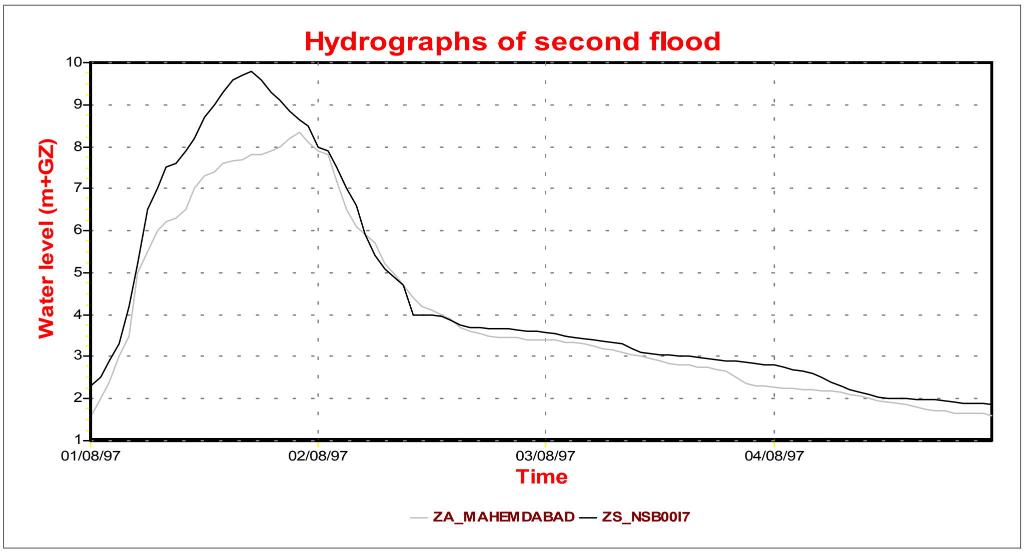


Figure 7.6a: Hydrograph of second floor

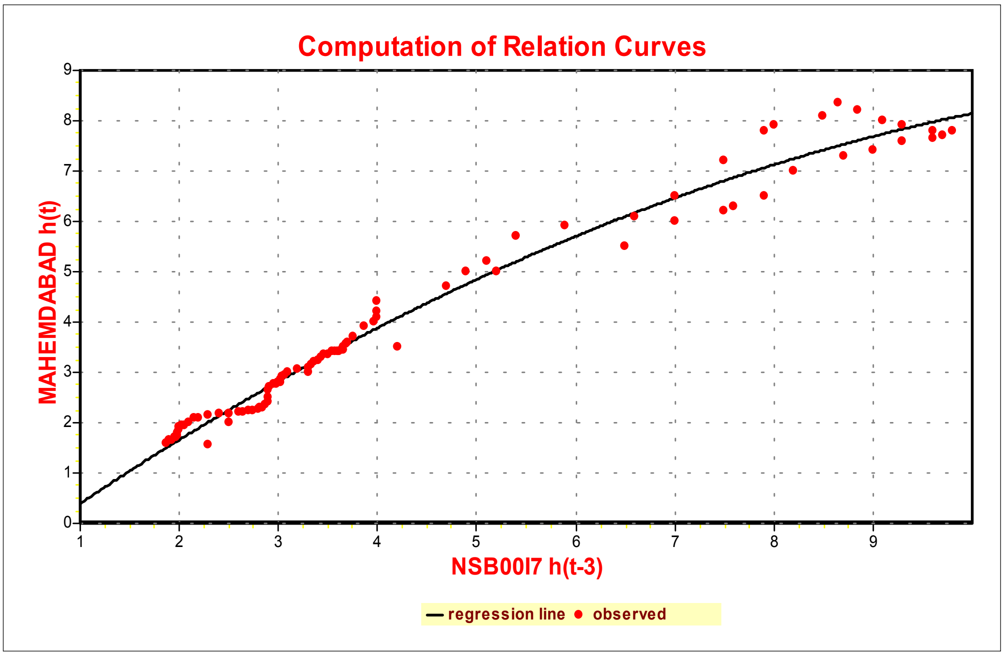


Figure 7.6 b: *Relation curve of second flood*

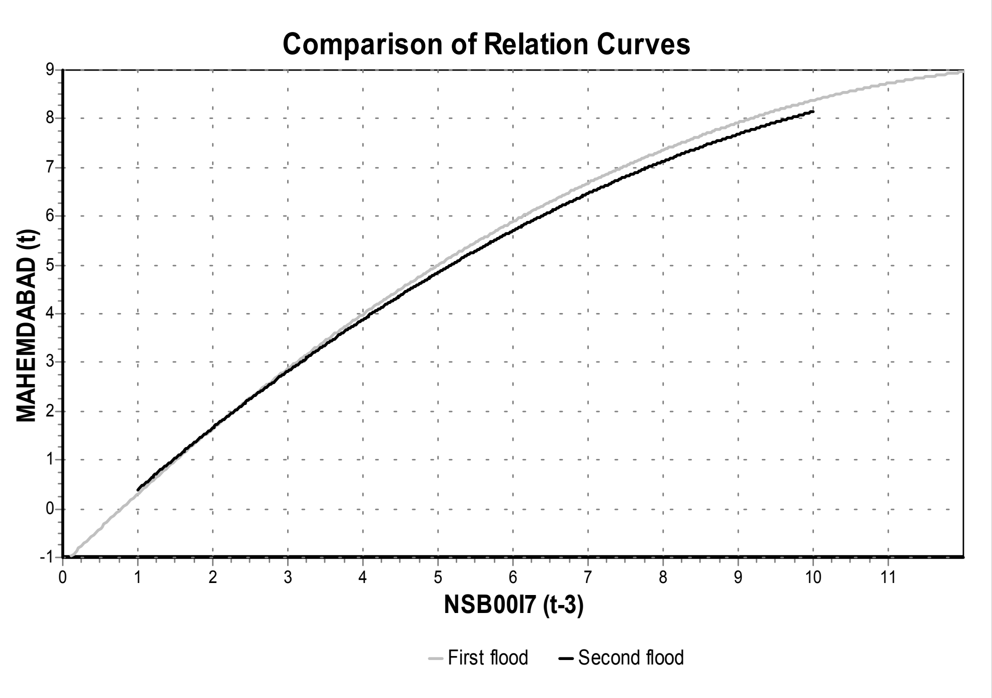


Figure 7.6c: *Comparison of relation curve established for different periods*

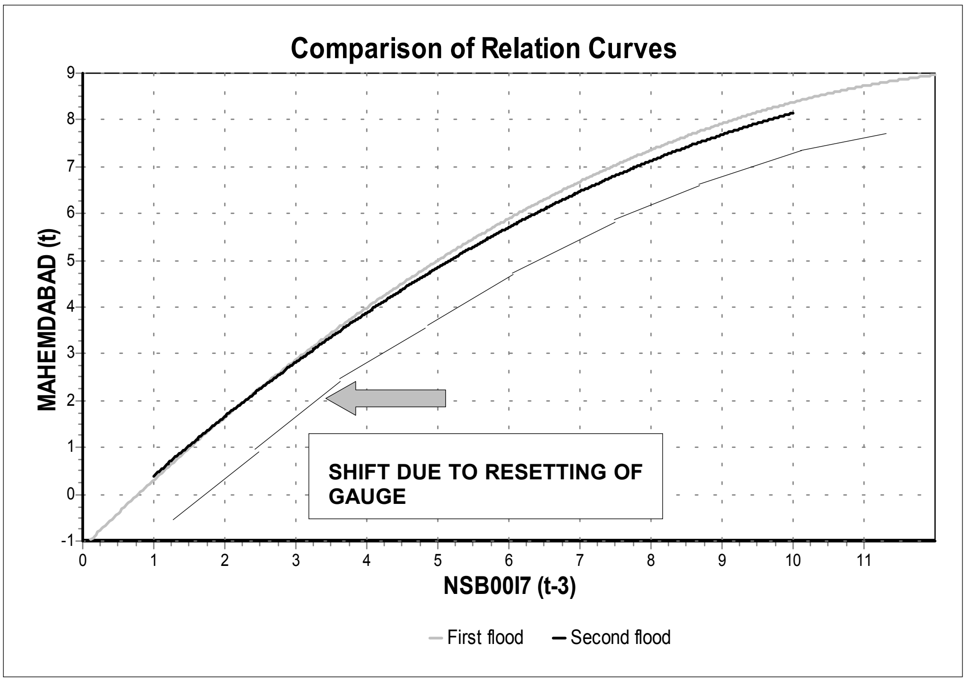


Figure 7.6d: Typical behaviour of relation curve in case of shifts in gauge zero

All steps in secondary validation can be illustrated for data of stations CHASKMAN and KHED in BHIMA catchment with HYMOS.

- Print water level data of Chaskman and Khed, Period 1/8/97 – 31/8/97
- Plot the water level series of both stations
- Plot the hydrograph of Chaskman with rainfall observed at Chaskman
- Edit data (i.e. entry errors)
- Make a scatter plot of period 31/7-21/8/97 (shift – 0.8 hrs)
- Fit a relation curve and save the relation
- Apply the relation to Period 22/8 – 13/9 (force a time shift of –0.8 hrs)

8 CORRECTION AND COMPLETION OF WATER LEVEL DATA

8.1 GENERAL

Correction and completion of water level will be carried out at Divisional offices.

Although separated from validation in these Modules, correction and completion will generally be done as a continuous process with validation

After 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 interpolation or other observations at the same station or neighbouring stations. The process of filling in missing values is generally referred to as ‘completion’.

Values identified as suspect by numerical validation tests will be inspected and corrected if necessary or the flag removed if they are found to be acceptable. Numerical test of records with respect to maximum, minimum and warning limits and rates of rise will have identified suspect values (and flagged them) during primary validation. Unless these were due to entry error, they will not have been corrected and will thus require further inspection and correction and completion if necessary.

Where multiple level records at the same station are thus flagged, but the observations agree, then the records may be assumed to be correct. Other suspect values outside warning limits are inspected for violations of typical hydrological behaviour but are also checked against neighbouring stations before correction or acceptance.

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’

8.2 CORRECTION USING RIVER LEVEL OR DISCHARGE?

Correction and completion may be carried out with respect to the water level series or it may await transformation to discharge using a stage discharge relationship. The choice of water level or discharge for correction depends on the type of error, the duration of missing or faulty records and the availability of suitable records with which to estimate. Correction as level has the advantage that it is the primary measurement whereas error in the discharge may result either from error in the level record or in the stage discharge relationship; it has the disadvantage that it provides no volumetric water balance checks.

Conditions where correction and completion will usually be carried out as level include the following:

- where the level record is complete but the recorder has gone out of adjustment and periodic check observations are available
- where the level record is correct but shifted in time
- where the primary record (e.g., from a digital water level recorder) is missing but an alternative level record of acceptable quality is available at the same station
- where the record is missing but the duration is short during a period of low flow or recession.

Correction and completion may be carried out as level include:

- where a record is available from a neighbouring station with little lateral inflow or abstraction between the stations

Correction and completion will normally be carried out as discharge:

- where a record is available only from a neighbouring station with much lateral inflow or abstraction
- where one or both stations are affected by variable backwater
- where the only available means of infilling is from catchment rainfall and the use of a rainfall runoff model.

Records completed as stage will receive further validation as discharge and may require further correction.

8.3 COMPARISON OF STAFF GAUGE AND AUTOGRAPHIC OR DIGITAL RECORDS

Where two or more measurements of the same variable are made at a station, one record may be used to correct or replace the other where one is missing. Where more than one record exists but they differ, the problem in the first instance is to determine which record is at fault. Typical measurement errors from each source are described under ‘primary validation’ (See Part 1) and guidelines are provided for identifying which record is at fault. Suspect values are flagged during validation. Errors and their correction may be classified as follows:

- observer errors
- recorder timing errors
- pen level errors
- errors arising from stilling well and intake problems
- miscellaneous instrument failures

8.3.1 OBSERVER ERRORS

Staff gauge and autographic or digital records can be displayed together graphically as multiple time series plots. Differences can also be displayed. Simple and isolated errors in reading and transcription by the observer (e.g., 6.57 for 5.67) can be identified and replaced by the concurrent measurement at the recording gauge. Persistent and erratic differences from the recording gauge (negative and positive) indicate a problem with the observer’s ability or record fabrication. They should be notified to the Sub-division for corrective action; the full staff gauge record for the period should be flagged as doubtful, left uncorrected and the recording gauge record adopted as the true stage record for the station.

8.3.2 RECORDER TIMING ERRORS

When the clock of the recording gauge runs fast or slow, the rate at which the recorder chart moves in the time direction under the pen will also be fast or slow. This can be detected by comparing with staff gauge readings, e.g. if observations are taken daily at 0800 and the clock of the recording instrument is running slower, then the observer’s stage record at 0800 will correspond to the same observation in the recording gauge before 0800, say 0700 Clock times and recorder times annotated on the chart or recorded in the Field Record book at the time of putting on or taking off the chart can be used to determine the time slippage during the record period.

Correction Procedure

For time corrections, it is assumed that a clock runs fast or slow at a constant rate. Where a digital record is produced from an analogue record using a pen-follower digitiser, the annotated clock and recorder time and level can be fed into the digitising program and the level record expanded or contracted as required to match the clock duration.

Where a digital record is extracted manually at a fixed interval from a chart, it will result in extra records for a fast clock and deficient records for a slow clock. This can be expediently corrected by removing or inserting (interpolating) records at appropriate intervals, e.g. if the clock runs 4 hours fast in eight days, and hourly data have been extracted, then one data point should be removed at 2 day intervals.

8.3.3 PEN LEVEL ERRORS

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

Correction Procedure

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

HYMOS provides such facility for correcting the gradual spread of error in digital records extracted from a chart recorder, with a growing adjustment from the commencement of the error until error detection. Let the error be ΔX observed at time t = i+k and assumed to have commenced at k intervals before, then the applied correction reads:

$$X_{corr,j} = X_{meas,j} - ((j - i)/k)\Delta X \qquad \text{for } j = i, i+1, \ldots\ldots., i+k \qquad (8.1)$$

Prepare the time-series plot of deviation of staff gauge observations from the recording gauge observations. If the deviation is static with time, then the difference must be settled (increased or decreased) directly from the analogue gauge observations. However, if the deviation increases gradually with time, then corrections for the difference between the pen observation and the staff gauge reading are made in the same way as time corrections. For example, assume that the pen trace record gradually drifted 0.08 m away (recording lower levels) from the corresponding staff gauge record in a period of 10 days. This shows that the pen readings has an error which is increasing gradually from 0 to 8 cms in 10 days period. Now error in such a data can be compensated by adding a proportionate amount of 8 mm per day from the starting point of the error.

8.3.4 ERRORS ARISING FROM STILLING WELL AND INTAKE PROBLEMS

Problems with stilling well or intake pipe may be intermittent or persistent and can be serious. In extreme floods, the hydrograph may be truncated due to inadequate height of the well, restricting the travel of the float, or counterweight reaching the well bottom. Blockage of the intake pipe with silt will result in a lag between river level (as recorded by the staff gauge) and well level, or a flat trace.

Correction procedure

The recorder trace is replaced by the observer’s staff gauge record if the time interval is sufficiently small in relation to the changes in the water levels. If the staff gauge record is intermittent or frequent changes in the levels are expected to be present then use of relation curves, as described in subsequent sections, is to be preferred for correcting the water level record.

8.3.5 MISCELLANEOUS INSTRUMENT FAILURES

Unacceptable recorder traces may result from a wide variety of instrument problems. These are often displayed as stepped or flat traces and may be corrected by interpolating a smooth curve on the hydrograph plot.

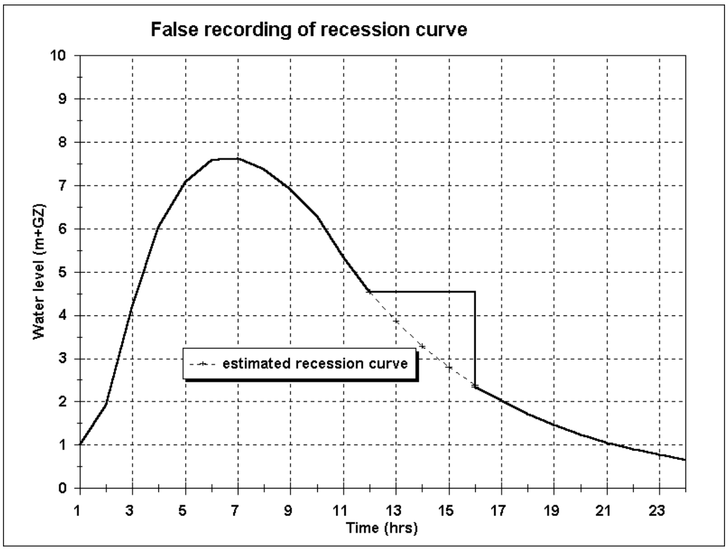


Figure 8.1:
False recording of recession curve

Figure 8.1 represents false recording of the recession curve because of: a) silting of stilling well; or b) blocking of intakes; or c) some obstruction causing the float to remain hung. The figure also shows the time when the obstruction is cleared. The correct curve can be estimated by reading the smooth curve that joins the first and last reading during the period of obstruction.

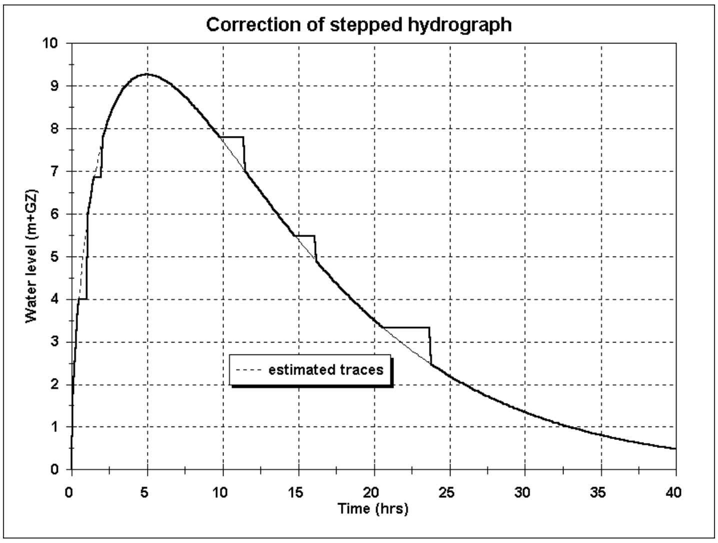


Figure 8.2:
Correction of stepped hydrograph

Figure 8.2 shows small steps in the stage records because of the temporary hanging of the float tape or counterweight, or kinks in the float tape. Such deviations can be easily identified and true values can be interpreted by reading the smooth curve in the same way as for recession curve.

8.4 LINEAR INTERPOLATION OF SHORT GAPS

Where only a single record is available at a station, gaps may occur due to instrument failure, observer sickness, station maintenance, etc. Gaps may be infilled by simple linear interpolation where they occur during periods of low flow or during recession and the difference between the level at the beginning and end of the gap is small. During periods of low flow, gaps of one to several days may be infilled in this way but it is recommended that infilling by linear interpolation during the monsoon or on a heavily regulated river should not exceed 6 hours.

For longer periods of missing data during a recession when the runoff is result only of outflow from a groundwater reservoir, the flow shows an exponential decay, which, when plotted as discharge on a sem-logarithmic scale, plots a a straight line. Using the stage discharge relationship it is possible to infill the series as water level rather than flow, but infilling as flow is conceptually simpler (Chapter 5). Gaps of a month or more may be filled in this way.

8.5 USE OF RELATION CURVES WITH ADJACENT STATIONS

8.5.1 GENERAL

The use of relation curves for water level data validation has been described in Chapter 7. It is also an effective way of infilling missing records and of correcting suspect ones especially for sequential stations on a river with little lateral inflow between. The following are typical uses.

- infilling of missing records
- identifying and correcting misreadings in one series
- identifying and correcting shift in gauge zero or change in cross section

8.5.2 INFILLING OF MISSING RECORDS

A relation curve based on the data of two series can be used to infill the missing data in the dependent variable of the relationship. The relation curve is used to calculate the missing value(s) at the station corresponding to the observed values at the adjacent station An example is given in Figure 8.3.

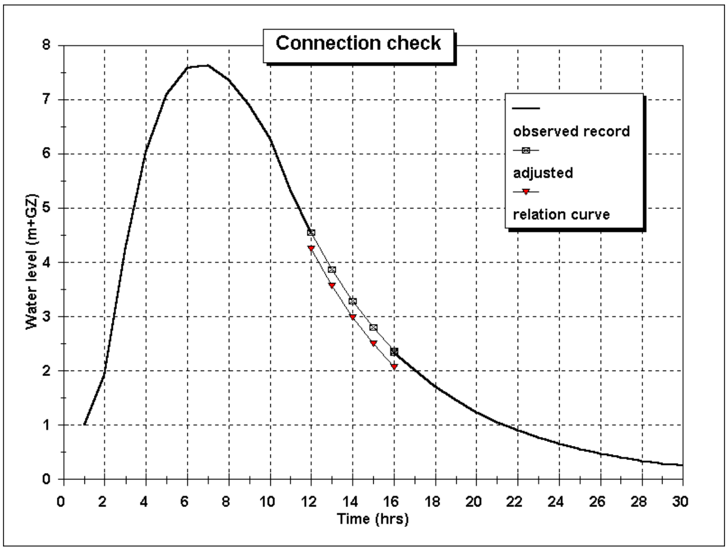


Figure 8.3:
Infilling of missing data with relation curve

Figure 8.3 shows that the relation curve did not provide a good connection between the existing record and the infilled part. These situations do sometimes happen, particularly, when the standard error is more than a few centimetres. Therefore, one should always verify the correctness of the infilled part.

8.5.3 IDENTIFYING AND CORRECTING MISREADINGS

If, after taking the lag between stations into account, there is a strong relationship between the two series, incidental misreading or incorrect booking will show up as outliers in the relation curve plot. Having identified its occurrence it is then required to determine in which series the problem has arisen and the actual value at fault, taking into account the lag time between stations. A corrected value is estimated using the relation curve or relation equation and substituted in the time series.

8.5.4 IDENTIFYING AND CORRECTING SHIFT IN GAUGE ZERO OR CHANGE IN CROSS SECTION

Shifts in water level observations due to change in gauge zero or changes in cross section conditions can be detected by comparing two relation curves or the plot of one period with that of another. For routine validation and completion, the comparison will be between data for the current period and an established curve for the station. If the new relation differs and there is a new stable relationship between the records and the deviation from the previous relation is constant, then a shift in the reference gauge is suspected. The time of its occurrence can be identified from the comparative plots. If there is a change in slope of the relation curve compared with the standard curve, then a change in cross section at one of the stations may be suspected.

On the identification of such changes, consultation should be made with sub-divisional staff and the Field Record Book inspected. If the conditions of change had been previously recognised in the field and adjustments made to the rating curve to account for the shift in gauge zero (or change in station location) or altered cross section, then no further action need be taken. If the change had not been recognised in the field then, since the analysis does not indicate which station is in error, then further action is necessary on following lines:

- Where additional stations are available for comparison, further relation curves may be developed and the station in error identified.
- Field staff are requested to re-survey gauges and the cross section at both stations
- If, after survey the gauge zero at one station is found to have inadvertently altered, then it should be reset to its former level. The stage level during the period between gauge shift and resetting should be corrected by the deviation shown by survey (and confirmed by the constant difference in relation curves).
- If no change in gauge zero is found but the cross section at one station has altered, then field staff are requested to intensify current meter gauging to establish a new stage discharge relationship. Usually the stage record will not be changed but the revised rating curve applied over the period from the occurrence of the change in cross section (usually during a flood).

9 ESTABLISHMENT OF STAGE DISCHARGE RATING CURVE

9.1 GENERAL

Flow is the variable usually required for hydrological analysis but, continuous measurement of flow past a river section is usually impractical or prohibitively expensive. However, stage can be observed continuously or at regular short time intervals with comparative ease and economy. Fortunately, a relation exists between stage and the corresponding discharge at river section. This relation is termed a stage-discharge relationship or stage-discharge rating curve or simply, rating curve.

A rating curve is established by making a number of concurrent observations of stage and discharge over a period of time covering the expected range of stages at the river gauging section.

At many locations, the discharge is not a unique function of stage; variables such as surface slope or rate of change of stage with respect to time must also be known to obtain the complete relationship in such circumstances.

The rating relationship thus established is used to transform the observed stages into the corresponding discharges. In its simplest form, a rating curve can be illustrated graphically, as shown in Figure 9.1, by the average curve fitting the scatter plot between water level (as ordinate) and discharge (as abscissa) at any river section. If Q and h are discharge and water level, then the relationship can be analytically expressed as:

$$Q = f(h) \tag{9.1}$$

Where; f(h) is an algebraic function of water level. A graphical stage discharge curve helps in visualising the relationship and to transform stages manually to discharges whereas an algebraic relationship can be advantageously used for analytical transformation.

Because it is difficult to measure flow at very high and low stages due to their infrequent occurrence and also to the inherent difficulty of such measurements, extrapolation is required to cover the full range of flows. Methods of extrapolation are described in a later module.

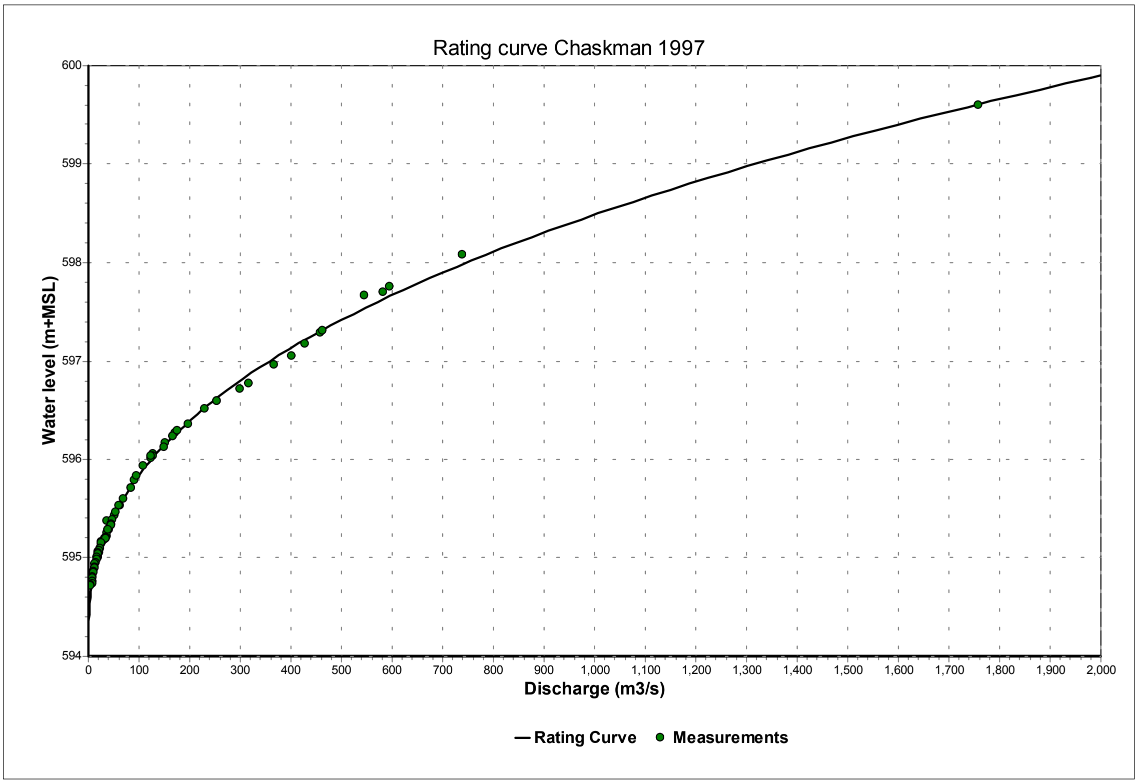


Figure 9.1: Example of stage-discharge rating curve

9.2 THE STATION CONTROL

The shape, reliability and stability of the stage-discharge relation are controlled by a section or reach of channel at or downstream from the gauging station and known as the station control. The establishment and interpretation of stage discharge relationships requires an understanding of the nature of controls and the types of control at a particular station.

Fitting of stage discharge relationships must not be considered simply a mathematical exercise in curve fitting. Staff involved in fitting stage discharge relationships should have familiarity with and experience of field hydrometrics.

The channel characteristics forming the control include the cross-sectional area and shape of the stream channel, expansions and restrictions in the channel, channel sinuosity, the stability and roughness of the streambed, and the vegetation cover all of which collectively constitute the factors determining the channel conveyance.

9.2.1 TYPES OF STATION CONTROL

The character of the rating curve depends on the type of control which in turn is governed by the geometry of the cross section and by the physical features of the river downstream of the section. Station controls are classified in a number of ways as:

- section and channel controls
- natural and artificial controls
- complete, compound and partial controls
- permanent and shifting controls

Section and channel controls

When the control is such that any change in the physical characteristics of the channel downstream to it has no effect on the flow at the gauging section itself then such control is termed as section control. In other words, any disturbance downstream the control will not be able to pass the control in the upstream direction. Natural or artificial local narrowing of the cross-section (waterfalls, rock bar, gravel bar) creating a zone of acceleration are some examples of section controls (Figures 9.2 and 9.3). The section control necessarily has a critical flow section at a short distance downstream.

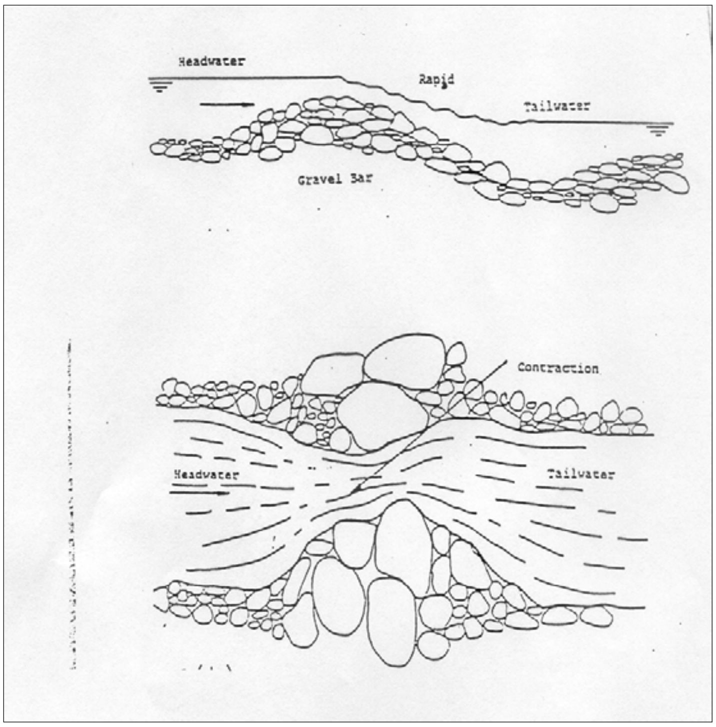


Figure 9.2:
Example of section control

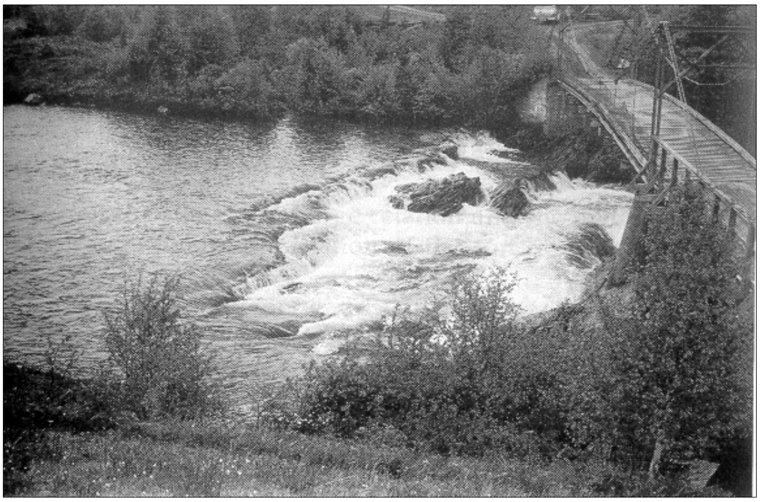


Figure 9.3:
Example of section control. (low-water part is sensitive, while high-water part is non-sensitive)

A cross section where no acceleration of flow occurs or where the acceleration is not sufficient enough to prevent passage of disturbances from the downstream to the upstream direction then such a location is called as a channel control. The rating curve in such case depends upon the geometry and the roughness of the river downstream of the control (Figure 9.4). The length of the downstream reach of the river affecting the rating curve depends on the normal or equilibrium depth h_e and on the energy slope S ($L \propto h_e/S$, where h_e follows from Manning $Q=K_mBh_e^{5/3}S^{1/2}$ (wide rectangular channel) so $h_e = (Q/K_mS^{1/2})^{3/5}$). The length of channel effective as a control increases with discharge. Generally, the flatter the stream gradient, the longer the reach of channel control.

Figure 9.4:
Example of channel control

Artificial and natural controls

An artificial section control or structure control is one which has been specifically constructed to stabilise the relationship between stage and discharge and for which a theoretical relationship is available based on physical modelling. These include weirs and flumes, discharging under free flow conditions (Figure 9.5). Natural section controls include a ledge of rock across a channel, the brink of a waterfall, or a local constriction in width (including bridge openings). All channel controls are 'natural'.

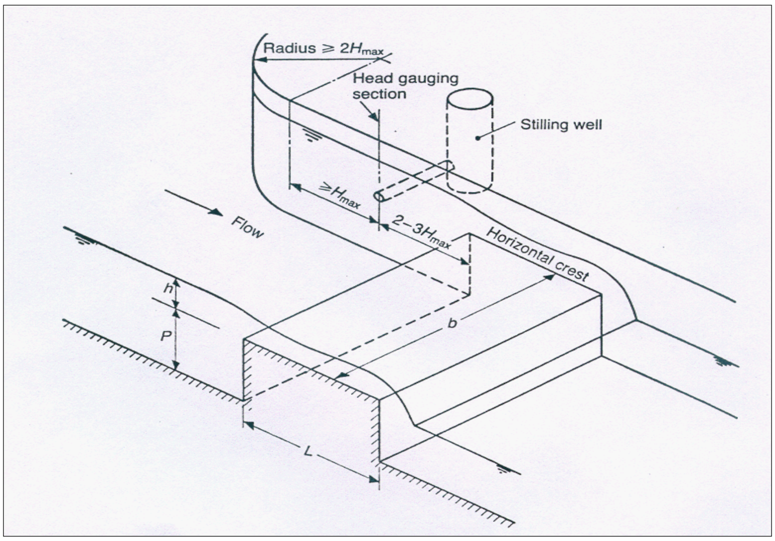


Figure 9.5:
Example of an artificial control

Complete, compound and partial controls

Natural controls vary widely in geometry and stability. Some consist of a single topographical feature such as a rock ledge across the channel at the crest of a rapid or waterfall so forming a complete control. Such a complete control is one which governs the stage-discharge relation throughout the entire range of stage experienced. However, in many cases, station controls are a combination of section control at low stages and a channel control at high stages and are thus called compound or complex controls. A partial control cases, station controls are a combination of section control at low stages and a is one which operates over a limited range of stage when a compound control is present, in the transition between section and channel control. The section control begins to drown out with rising tailwater levels so that over a transitional range of stage the flow is dependent both on the elevation and shape of the control and on the tailwater level.

Permanent and shifting controls

Where the geometry of a section and the resulting stage-discharge relationship does not change with time, it is described as a stable or permanent control. Shifting controls change with time and may be section controls such as boulder, gravel or sand riffles which undergo periodic or near continuous scour and deposition, or they may be channel controls with erodible bed and banks. Shifting controls thus typically result from:

- scour and fill in an unstable channel
- growth and decay of aquatic weeds
- overspilling and ponding in areas adjoining the stream channel.

The amount of gauging effort and maintenance cost to obtain a record of adequate quality is much greater for shifting controls than for permanent controls. Since rating curves for the unstable controls must be updated and/or validated at frequent intervals, regular and frequent current meter measurements are required. In contrast, for stable controls, the rating curve can be established once and needs validation only occasionally. Since stage discharge observations require significant effort and money, it is always preferred to select a gauging site with a section or structure control. However, this is not practicable in many cases and one has to be content with either channel control or a compound control.

9.3 FITTING OF RATING CURVES

9.3.1 GENERAL

A simple stage discharge relation is one where discharge depends upon stage only. A complex rating curve occurs where additional variables such as the slope of the energy line or the rate of change of stage with respect to time are required to define the relationship. The need for a particular type of rating curve can be ascertained by first plotting the observed stage and discharge data on a simple orthogonal plot. The scatter in the plot gives a fairly good assessment of the type of stage-discharge relationship required for the cross section. Examples of the scatter plots obtained for various conditions are illustrated below.

If there is negligible scatter in the plotted points and it is possible to draw a smooth single valued curve through the plotted points then a simple rating curve is required. This is shown in Figure 9.6.

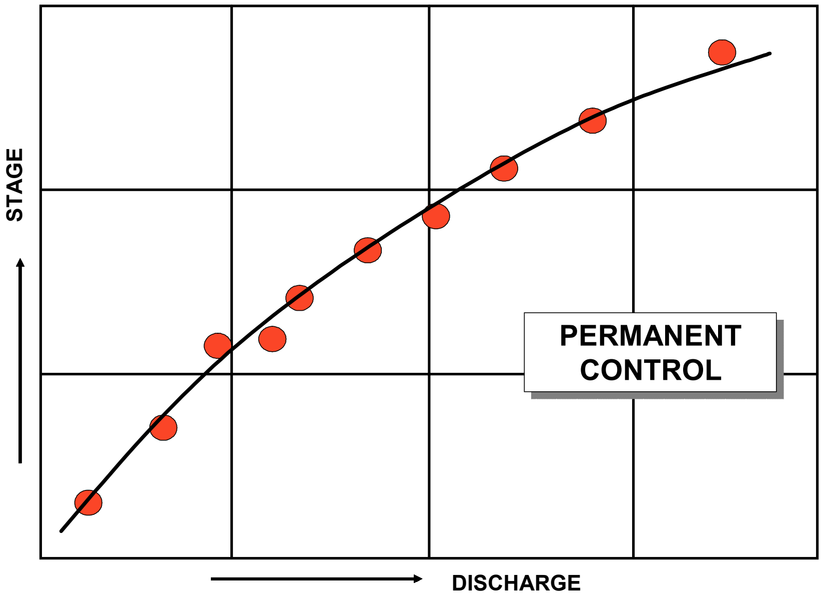


Figure 9.6: Permanent control

However, if scatter is not negligible then it requires further probing to determine the cause of such higher scatter. There are four distinct possibilities:

- The station is affected by the variable backwater conditions arising due for example to tidal influences or to high flows in a tributary joining downstream. In such cases, if the plotted points are annotated with the corresponding slope of energy line (≈surface slope for uniform flows) then a definite pattern can be observed. A smooth curve passing through those points having normal slopes at various depths is drawn first. It can then be seen that the points with greater variation in slopes from the corresponding normal slopes are located farther from the curve. This is as shown in Figures 9.7a and b.

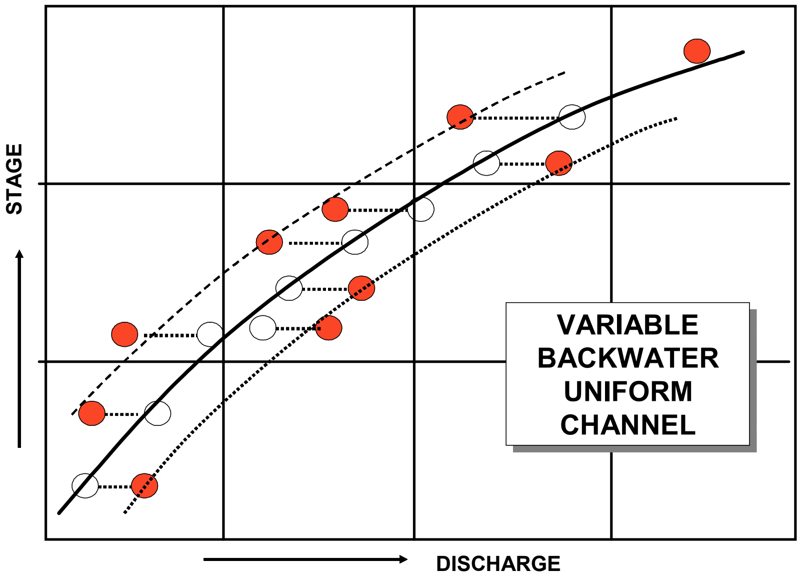


Figure 9.7 a: Rating curve affected by variable backwater (uniform channel)

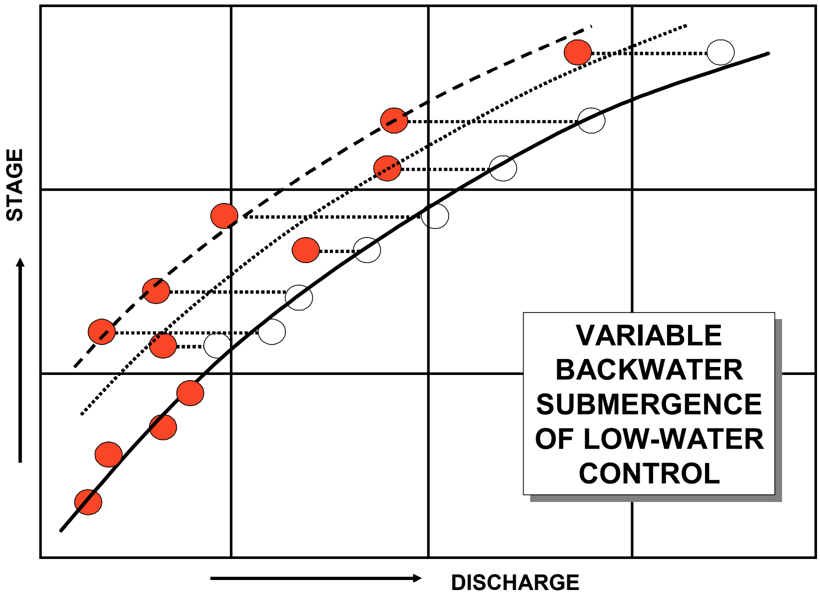


Figure 9.7 b: Rating curve affected by variable backwater (submergence of low-water control)

- The stage discharge rating is affected by the variation in the local acceleration due to unsteady flow. In such case, the plotted points can be annotated with the corresponding rate of change of slope with respect to time. A smooth curve (steady state curve) passing through those points having the least values of rate of change of stage is drawn first. It can then be seen that all those points having positive values of rate of change of stage are towards the right side of the curve and those with negative values are towards the left of it. Also, the distance from the steady curve increases with the increase in the magnitude of the rate of change of stage. This is as shown in Figure 9.8.

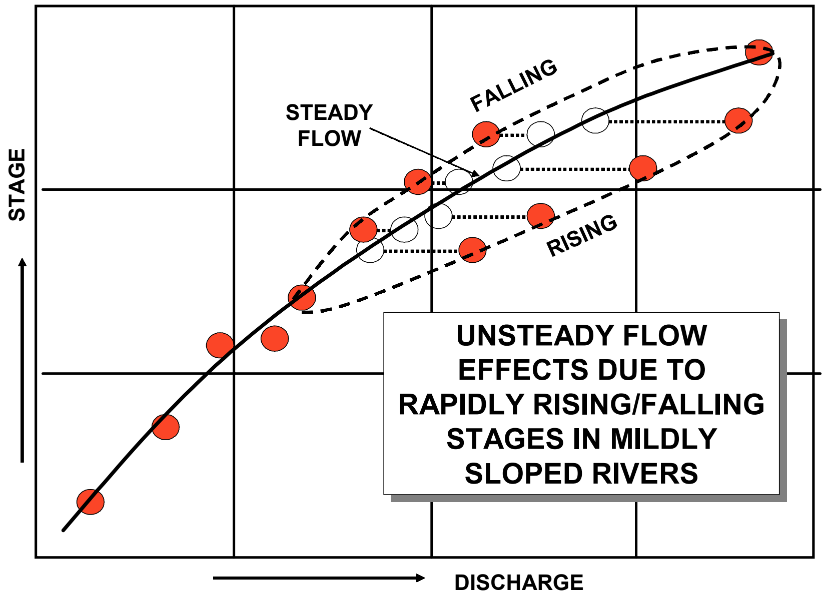


Figure 9.8: Rating curve affected by unsteady flow

- The stage discharge rating is affected by scouring of the bed or changes in vegetation characteristics. A shifting bed results in a wide scatter of points on the graph. The changes are erratic and may be progressive or may fluctuate from scour in one event and deposition in another. Examples are shown in Figures 9.9 and 9.10.

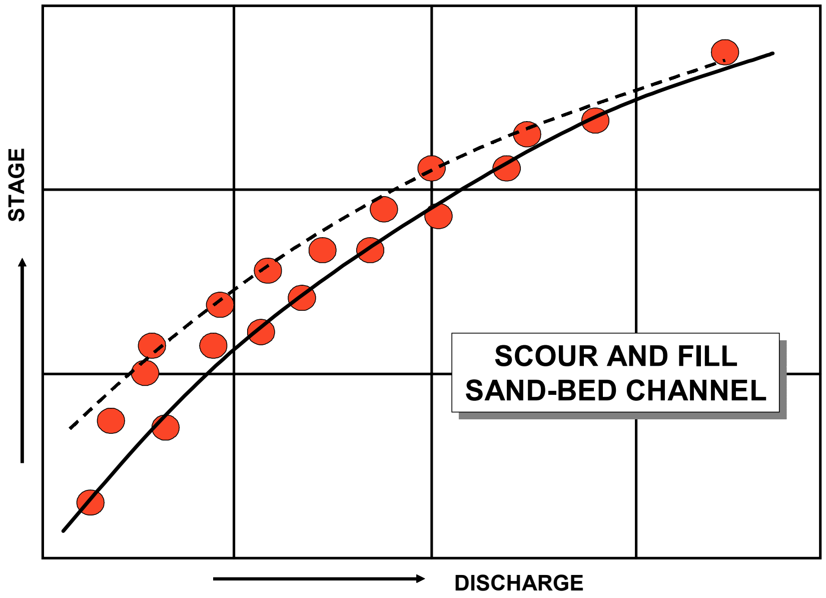


Figure 9.9: Stage-discharge relation affected by scour and fill

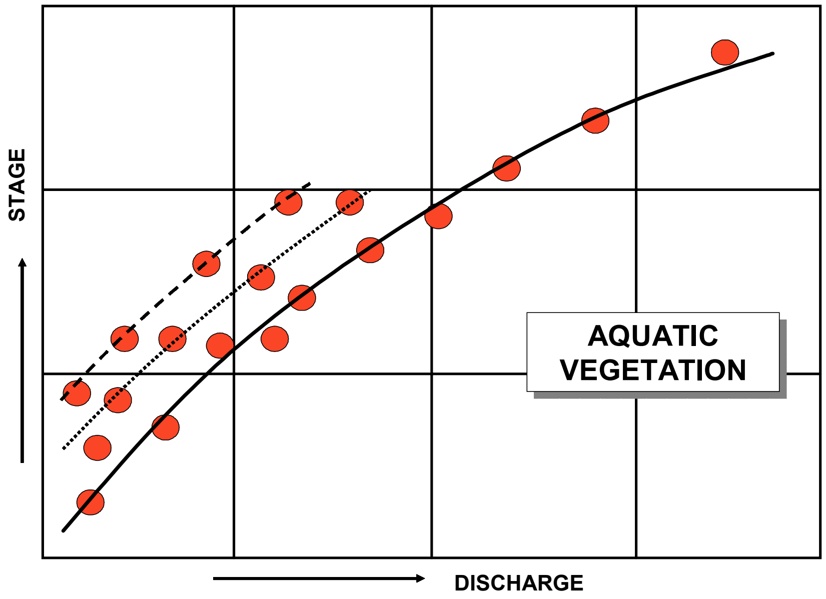


Figure 9.10: Stage-discharge relation affected by vegetation growth

- If no suitable explanation can be given for the amount of scatter present in the plot, then it can perhaps be attributed to the observational errors. Such errors can occur due to non-standard procedures for stage discharge observations.

Thus, based on the interpretation of scatter of the stage discharge data, the appropriate type of rating curve is fitted. There are four main cases:

- Simple rating curve: If simple stage discharge rating is warranted then either single channel or compound channel rating curve is fitted according to whether the flow occur essentially in the main channel or also extends to the flood plains.
- Rating curve with backwater corrections: If the stage discharge data is affected by the backwater effect then the rating curve incorporating the backwater effects is to be established. This requires additional information on the fall of stage with respect to an auxiliary stage gauging station.
- Rating curve with unsteady flow correction: If the flows are affected by the unsteadiness in the flow then the rating curve incorporating the unsteady flow effects is established. This requires information on the rate of change of stage with respect to time corresponding to each stage discharge data.
- Rating curve with shift adjustment: A rating curve with shift adjustment is warranted in case the flows are affected by scouring and variable vegetation effects.

9.3.2 FITTING OF SINGLE CHANNEL SIMPLE RATING CURVE

Single channel simple rating curve is fitted in those circumstances when the flow is contained the main channel section and can be assumed to be fairly steady. There is no indication of any backwater affecting the relationship. The bed of the river also does not significantly change so as create any shifts in the stage discharge relationship. The scatter plot of the stage and discharge data shows a very little scatter if the observational errors are not significant. The scatter plot of stage discharge data in such situations, typically is as shown in Figure 9.1. The fitting of simple rating curves can conveniently be considered under the following headings:

- equations used and their physical basis
- determination of datum correction(s)
- number and range of rating curve segments
- determination of rating curve coefficients
- estimation of uncertainty in the stage discharge relationship

Equations used and their physical basis

Two types of algebraic equations are commonly fitted to stage discharge data are:

1. **Power type equation** which is most commonly used:

$$Q = c\,(h + a)^b \tag{9.2}$$

2. **Parabolic type of equation**

$$Q = c_2\,(h_w + a)^2 + c_1\,(h_w + a) + c_0 \tag{9.3}$$

where: Q = discharge (m^3/sec)
h = measured water level (m)
a = water level (m) corresponding to $Q = 0$
c_i = coefficients derived for the relationship corresponding to the station characteristics

It is anticipated that the power type equation is most frequently used in India and is recommended. Taking logarithms of the power type equation results in a straight line relationship of the form:

$$\log(Q) = \log(c) + b\log(h + a) \tag{9.4}$$

or

$$Y = A + B\,X \tag{9.5}$$

That is, if sets of discharge (Q) and the effective stage (h + a) are plotted on the double log scale, they will represent a straight line. Coefficients A and B of the straight line fit are functions of a and b. Since values of a and b can vary at different depths owing to changes in physical characteristics (effective roughness and geometry) at different depths, one or more straight lines will fit the data on double log plot. This is illustrated in Figure 9.11, which shows a distinct break in the nature of fit in two water level ranges. A plot of the cross section at the gauging section is also often helpful to interpret the changes in the characteristics at different levels.

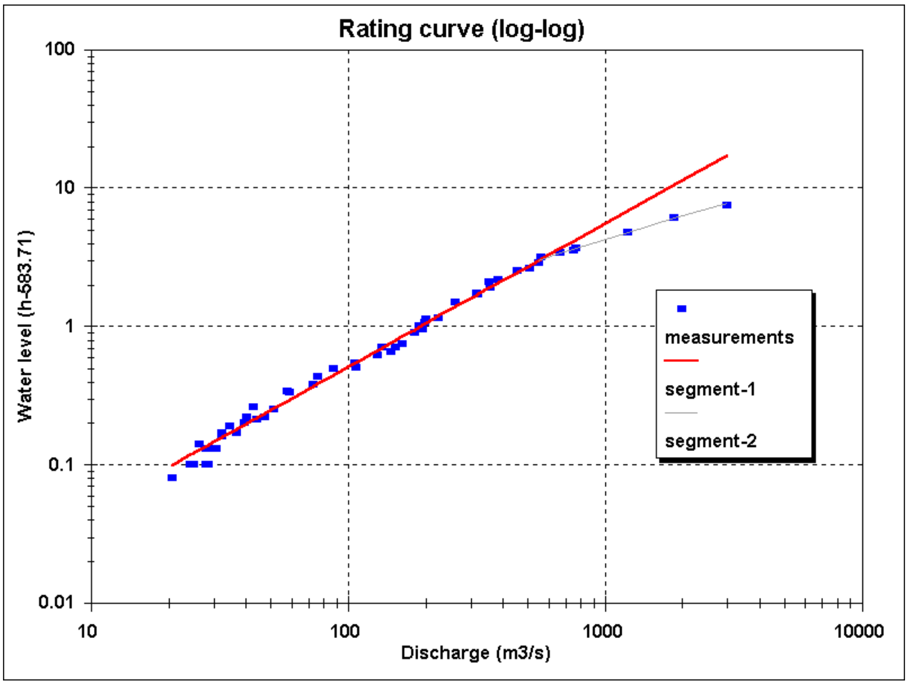


Figure 9.11: Double logarithmic plot of rating curve showing a distinct break

- The relationship between rating curve parameters and physical conditions is also evident if the power and parabolic equations are compared with Manning's equation for determining discharges in steady flow situations. The Manning's equation can be given as:

$$Q_m = \frac{1}{n} AR^{2/3} S^{1/2} = \left(\frac{1}{n} S^{1/2}\right)(AR^{2/3}) \tag{9.6}$$

= function of (roughness & slope) & (depth & geometry)

Hence, the coefficients a, c and d are some measures of roughness and geometry of the control and b is a measure of the geometry of the section at various depths. The value of coefficient b for various geometrical shapes are as follows:

For rectangular shape	:	about 1.6
For triangular shape	:	about 2.5
For parabolic shape	:	about 2.0
For irregular shape	:	1.6 to 1.9

Changes in the channel resistance and slope with stage, however, will affect the exponent b. The net result of these factors is that the exponent for relatively wide rivers with channel control will vary from about 1.3 to 1.8. For relatively deep narrow rivers with section control, the exponent will commonly be greater than 2 and sometimes exceed a value of 3. Note that for compound channels with flow over the floodplain or braided channels over a limited range of level, very high values of the exponent are sometimes found (>5).

Determination of datum correction (a)

The datum correction (a) corresponds to that value of water level for which the flow is zero. From equation (9.2) it can be seen that for Q = 0, (h + a) = 0 which means: a = -h.

Physically, this level corresponds to the zero flow condition at the control effective at the measuring section. The exact location of the effective control is easily determined for artificial controls or where the control is well defined by a rock ledge forming a section control. For the channel controlled gauging station, the level of deepest point opposite the gauge may give a reasonable indication of datum correction. In some cases identification of the datum correction may be impractical especially where the control is compound and channel control shifts progressively downstream at higher flows. Note that the datum correction may change between different controls and different segments of the rating curve. For upper segments the datum correction is effectively the level of zero flow had that control applied down to zero flow; it is thus a nominal value and not physically ascertainable. Alternative analytical methods of assessing “a” are therefore commonly used and methods for estimating the datum correction are as follows:

- trial and error procedure
- arithmetic procedure
- computer-based optimisation

However, where possible, the estimates should be verified during field visits and inspection of longitudinal and cross sectional profiles at the measuring section:

<u>Trial and error procedure</u>

This was the method most commonly used before the advent of computer-based methods. The stage discharge observations are plotted on double log plot and a median line fitted through them. This fitted line usually is a curved line. However, as explained above, if the stages are adjusted for zero flow condition, i.e. datum correction a, then this line should be a straight line. This is achieved by taking a trial value of “a” and plotting (h + a), the adjusted stage, and discharge data on the same double log plot. It can be seen that if the unadjusted stage discharge plot is concave downwards then a positive trial value of “a” is needed to make it a straight line. And conversely, a negative trial value is needed to make the line straight if the curve is concave upwards. A few values of “a” can be tried to attain a straight line fit for the plotted points of adjusted stage discharge data. The procedure is illustrated in Figure 9.12. This procedure was slow but quite effective when done earlier manually. However, making use of general spreadsheet software (having graphical provision) for such trial and error procedure can be very convenient and faster now.

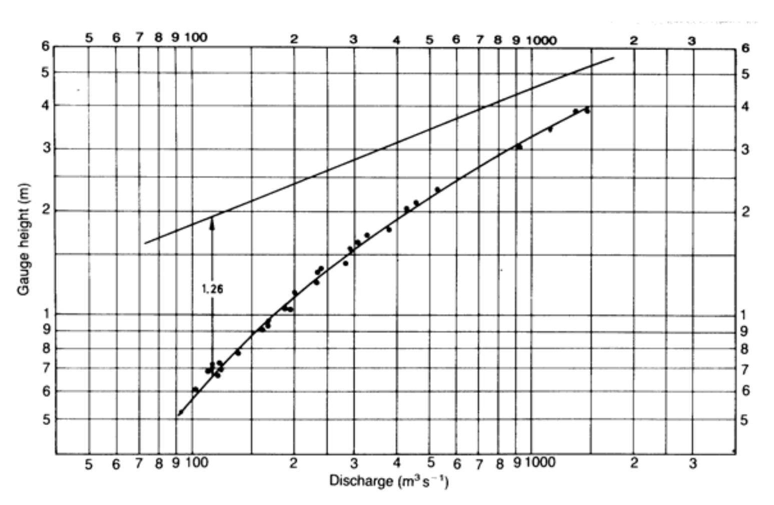


Figure 9.12: Determination of datum correction (a) by trial and error

Arithmetic procedure:

This procedure is based on expressing the datum correction “a” in terms of observed water levels. This is possible by way elimination of coefficients b and c from the power type equation between gauge and discharge using simple mathematical manipulation. From the median curve fitting the stage discharge observations, two points are selected in the lower and upper range (Q_1 and Q_3) whereas the third point Q_2 is computed from $Q_2^2 = Q_1.Q_3$, such that:

$$\frac{Q_1}{Q_2} = \frac{Q_2}{Q_3} \tag{9.7}$$

If the corresponding gauge heights for these discharges read from the plot are h_1, h_2 and h_3 then using the power type, we obtain:

$$\frac{c\,(h_1 + a)^b}{c\,(h_2 + a)^b} = \frac{c\,(h_2 + a)^b}{c\,(h_3 + a)^b} \tag{9.8}$$

Which yields:

$$a = \frac{h_2^2 - h_1 h_3}{h_1 + h_3 - 2h_2} \tag{9.9}$$

From this equation an estimated value of “a” can be obtained directly. This procedure is known as Johnson method, which is described in the WMO Operational Hydrology manual on stream gauging (Report No. 13, 1980).

Optimisation procedure:

This procedure is suitable for automatic data processing using computer and “a” is obtained by optimisation. The first trial value of the datum correction “a” is either input by the user based on the field survey or from the computerised Johnson method described above. Next, this first estimate of “a” is varied within 2 m so as to obtain a minimum mean square error in the fit. This is a purely mathematical procedure and probably gives the best results on the basis of observed stage discharge data but it is important to make sure that the result is confirmed where possible by physical explanation of the control at the gauging location. The procedure is repeated for each segment of the rating curve.

Number and ranges of rating curve segments:

After the datum correction “a” has been established, the next step is to determine if the rating curve is composed of one or more segments. This is normally selected by the user rather than done automatically by computer. It is done by plotting the adjusted stage, (h-a) or simply “h” where there are multiple segments, and discharge data on the double log scale. This scatter plot can be drawn manually or by computer and the plot is inspected for breaking points. Since for (h-a), on double log scale the plotted points will align as straight lines, breaks are readily identified. The value of “h” at the breaking points give the first estimate of the water levels at which changes in the nature of the rating curve are expected. The number and water level ranges for which different rating curves are to be established is thus noted. For example, Fig. 3.6 shows that two separate rating curves are required for the two ranges of water level – one up to level “h_1” and second from “h_1” onwards. The rating equation for each of these segments is then established and the breaking points between segments are checked by computer analysis (See below).

Determination of rating curve coefficients:

A least square method is normally employed for estimating the rating curve coefficients. For example, for the power type equation, taking α and β as the estimates of the constants of the straight line fitted to the scatter of points in double log scale, the estimated value of the logarithm of the discharge can be obtained as:

$$\hat{Y} = \alpha + \beta\, X \tag{9.10}$$

The least square method minimises the sum of square of deviations between the logarithms of measured discharges and the estimated discharges obtained from the fitted rating curve. Considering the sum of square the error as E, we can write:

$$E = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{N} (Y_i - \alpha - \beta\, X_i)^2 \tag{9.11}$$

Here i denotes the individual observed point and N is the total number of observed stage discharge data.

Since this error is to be minimum, the slope of partial derivatives of this error with respect to the constants must be zero. In other words:

$$\frac{\partial E}{\partial \alpha} = \frac{\partial \{\sum_{i=1}^{N} (Y_i - \alpha - \beta\, X_i)^2\}}{\partial \alpha} = 0 \tag{9.12}$$

and

$$\frac{\partial E}{\partial \beta} = \frac{\partial \{\sum_{i=1}^{N} (Y_i - \alpha - \beta\, X_i)^2\}}{\partial \beta} = 0 \tag{9.13}$$

This results in two algebraic equations of the form:

$$\sum_{i=1}^{N} Y_i - \alpha\, N - \beta \sum_{i=1}^{N} X_i = 0 \tag{9.14}$$

and

$$\sum_{i=1}^{N} (X_i\ Y_i) - \alpha \sum_{i=1}^{N} X_i - \beta \sum_{i=1}^{N} (X_i)^2 = 0 \tag{9.15}$$

All the quantities in the above equations are known except α and β. Solving the two equations yield:

$$\beta = \frac{N \sum_{i=1}^{N} (X_i\ Y_i) - (\sum_{i=1}^{N} X_i)(\sum_{i=1}^{N} Y_i)}{N \sum_{i=1}^{N} (X_i)^2 - (\sum_{i=1}^{N} X_i)^2} \tag{9.16}$$

and

$$\alpha = \frac{\sum_{i=1}^{N} Y_i - \beta \sum_{i=1}^{N} X_i}{N} \tag{9.17}$$

The value of coefficients c and b of power type equation can then be finally obtained as:

$$b = \beta \quad \text{and} \quad c = 10^{\alpha} \tag{9.18}$$

Reassessment of breaking points

The first estimate of the water level ranges for different segments of the rating curve is obtained by visual examination of the cross-section changes and the double log plot. However, exact limits of water levels for various segments are obtained by computer from the intersection of the fitted curves in adjoining the segments.

Considering the rating equations for two successive water level ranges be given as $Q = f_{i-1}(h)$ and $Q = f_i (h)$ respectively and let the upper boundary used for the estimation of f_{i-1} be denoted by hu_{i-1} and the lower boundary used for the estimation of f_i by hl_i . To force the intersection between f_{i-1} and f_i to fall within certain limits it is necessary to choose: $hu_{i-1} > hl_i$. That is, the intersection of the rating curves of the adjoining segments should be found numerically within this overlap. This is illustrated in Figure 9.13 and Table 9.1. If the intersection falls outside the selected overlap, then the intersection is estimated for the least difference between $Q = f_{i-1}(h)$ and $Q = f_i (h)$. Preferably the boundary width between hu_{i-1} and hl_i is widened and the curves refitted.

It is essential that a graphical plot of the fit of the derived equations to the data is inspected before accepting them.

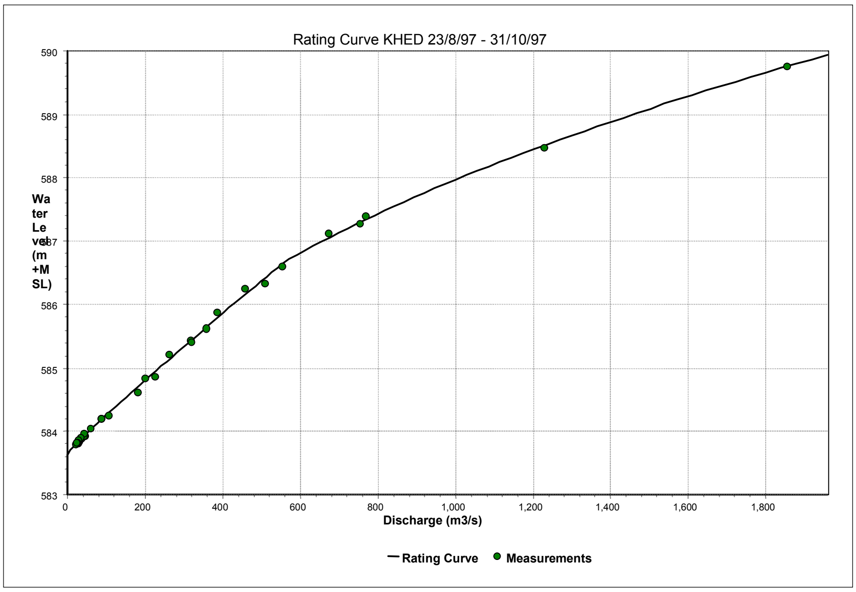


Figure 9.13: Fitted rating curve using 2 segments

Estimation of uncertainty in the stage discharge relationship:

With respect to the stage discharge relationship the standard error of estimate (Se) is a measure of the dispersion of observations about the mean relationship. The standard error is expressed as:

$$S_e = \sqrt{\frac{\sum(\Delta Q_i - \overline{\Delta Q})^2}{N-2}} \qquad (9.19)$$

Here, ΔQ_i is the measure of difference between the observed (Q_i) and computed (Q_c) discharges and can be expressed in absolute and relative (percentage) terms respectively as:

$$\Delta Q_i = Q_i - Q_c \qquad (9.20)$$

or

$$\Delta Q_i = \frac{Q_i - Q_c}{Q_i} \times 100\% \tag{9.21}$$

```
Analysis of stage-discharge data

 Station name :  KHED
 Data from 1997  1  1 to 1997 12 31

 Single channel

 Gauge Zero on  1997  7 30 =       .000 m

 Number of data =  36
 h minimum =    583.79 meas.nr =  44
 h maximum =    589.75 meas.nr =  49
 q minimum =    20.630 meas.nr =  44
 q maximum =  1854.496 meas.nr =  49

 Given boundaries for computation of rating curve(s)
 interval lower bound upper bound  nr. of data
        1     583.000     587.000          31
        2     586.500     590.000           6

 Power type of equation q=c*(h+a)**b is used

 Boundaries / coefficients
 lower bound upper bound        a         b         c

      583.00      586.68  -583.650     1.074   .1709E+03
      586.68      590.00  -581.964     2.381   .1401E+02

 Number   W level    Q meas    Q comp      DIFf  Rel.dIFf      Semr
                M      M3/S      M3/S      M3/S       0/0       0/0

2   584.610   181.500   163.549    17.951     10.98      3.08
      3   587.390   768.820   785.226   -16.406     -2.09      3.54
      5   586.240   457.050   475.121   -18.071     -3.80      4.93
      6   585.870   386.020   402.597   -16.577     -4.12      4.60
      7   585.440   316.290   319.452    -3.162      -.99      4.16
      9   585.210   261.650   275.570   -13.920     -5.05      3.89
     10   584.840   200.850   206.013    -5.163     -2.51      3.41
    ...............................................................
     48   583.870    32.490    33.574    -1.084     -3.23      3.35
     49   589.750  1854.496  1854.988     -.492      -.03      6.02
     50   588.470  1228.290  1209.632    18.658      1.54      3.48
     51   587.270   753.580   744.518     9.062      1.22      3.79
     52   587.120   673.660   695.381   -21.721     -3.12      4.14
     53   586.600   553.930   546.431     7.499      1.37      5.21
     54   586.320   509.230   490.911    18.319      3.73      4.99
     55   585.620   357.030   354.093     2.937       .83      4.35
     56   585.410   319.860   313.697     6.163      1.96      4.12
     57   584.870   226.080   211.593    14.487      6.85      3.45
     62   584.040    59.020    62.120    -3.100     -4.99      2.69

     Overall standard error =     6.603

     Statistics per interval
     Interval Lower bound Upper bound Nr.of data Standard error
        1     583.000     586.680         31          7.11
        2     586.680     590.000          5          2.45
```

Table 9.1: Results of stage-discharge computation for example presented in Figure 9.13

Standard error expressed in relative terms helps in comparing the extent of fit between the rating curves for different ranges of discharges. The standard error for the rating curve can be derived for each segment separately as well as for the full range of data.

Thus 95% of all observed stage discharge data are expected to be within t x S_e from the fitted line where:

Student's $t \cong 2$ where $n > 20$, but increasingly large for smaller samples.

The stage discharge relationship, being a line of best fit provides a better estimate of discharge than any of the individual observations, but the position of the line is also subject to uncertainty, expressed as the Standard error of the mean relationship (S_{mr}) which is given by:

$$S_{mr} = S_e \sqrt{\frac{1}{n} + \frac{(P_i - \bar{P})^2}{S_P^2}} \quad \text{and} \quad CL_{95\%} = \pm tS_{mr} \tag{9.22}$$

where: t = Student t-value at 95% probability

P_i = ln $(h_i + a)$

S^2_P = variance of P

$CL_{95\%}$ = 95% confidence limits

The S_e equation gives a single value for the standard error of the logarithmic relation and the 95% confidence limits can thus be displayed as two parallel straight lines on either side of the mean relationship. By contrast S_{mr} is calculated for each observation of (h + a). The limits are therefore curved on each side of the stage discharge relationship and are at a minimum at the mean value of ln (h + a) where the S_{mr} relationship reduces to:

$$S_{mr} = \pm Se / n^{1/2} \tag{9.23}$$

Thus with $n = 25$, S_{mr}, the standard error of the mean relationship is approximately 20% of Se indicating the advantage of a fitted relationship over the use of individual gaugings.

9.3.3 COMPOUND CHANNEL RATING CURVE

If the flood plains carry flow over the full cross section, the discharge (for very wide channels) consists of two parts:

$$Q_{river} = (h\, B_r)\, (K_{mr}\; h^{2/3}\; S^{1/2}) \tag{9.24}$$

and

$$Q_{floodplain} = (h - h_1)\, (B - B_r)\, [K_{mf}\; (h - h_1)^{2/3}\; S^{1/2}] \tag{9.25}$$

assuming that the floodplain has the same slope as the river bed, the total discharge becomes:

$$Q_{total} = h\, B_r\; (K_{mr}\; h^{2/3}\; S^{1/2}) + (h - h_1)\, (B - B_r)\, [K_{mf}\; (h - h_1)^{2/3}\; S^{1/2}] \tag{9.26}$$

This is illustrated in Figure 9.14. The rating curve changes significantly as soon as the flood plain at level h-h_1 is flooded, especially if the ratio of the storage width B to the width of the river bed B_r is large. The rating curve for this situation of a compound channel is determined by considering the flow through the floodplain portion separately. This is done to avoid large values of the exponent b and extremely low values for the parameter c in the power equation for the rating curve in the main channel portion.

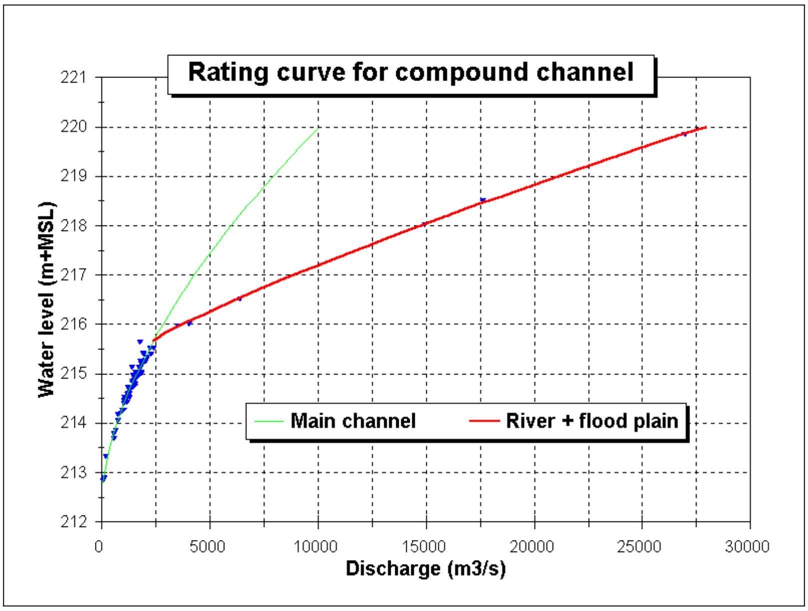


Figure 9.14: Example of rating curve for compound cross-section

The last water level range considered for fitting rating curve is treated for the flood plain water levels. First, the river discharge Q_r will be computed for this last interval by using the parameters computed for the one but last interval. Then a temporary flood plain discharge Q_f is computed by subtracting Q_r from the observed discharge (O_{obs}) for the last water level interval, i.e.

$$Q_f = Q_{obs} - Q_r. \tag{9.27}$$

This discharge Q_f will then be separately used to fit a rating curve for the water levels corresponding to the flood plains. The total discharge in the flood plain is then calculated as the sum of discharges given by the rating curve of the one but last segment applied for water levels in the flood plains and the rating curve established separately for the flood plains.

The rating curve presented in Figure 9.14 for Jhelum river at Rasul reads:

For $h < 215.67$ m + MSL: $Q = 315.2(h-212.38)^{1.706}$

For $h > 215.67$ m + MSL: $Q = 315.2(h-212.38)^{1.706} + 3337.4(h-215.67)^{1.145}$

Hence the last part in the second equation is the contribution of the flood plain to the total river flow.

9.3.4 RATING CURVE WITH BACKWATER CORRECTION

When the control at the gauging station is influenced by other controls downstream, then the unique relationship between stage and discharge at the gauging station is not maintained. Backwater is an important consideration in streamflow site selection and sites having backwater effects should be avoided if possible. However, many existing stations in India are subject to variable backwater effects and require special methods of discharge determination. Typical examples of backwater effects on gauging stations and the rating curve are as follows:

- by regulation of water course downstream.
- level of water in the main river at the confluence downstream
- level of water in a reservoir downstream
- variable tidal effect occurring downstream of a gauging station
- downstream constriction with a variable capacity at any level due to weed growth etc.
- rivers with return of overbank flow

Backwater from variable controls downstream from the station influences the water surface slope at the station for given stage. When the backwater from the downstream control results in lowering the water surface slope, a smaller discharge passes through the gauging station for the same stage. On the other hand, if the surface slope increases, as in the case of sudden drawdown through a regulator downstream, a greater discharge passes for the same stage. The presence of backwater does not allow the use of a simple unique rating curve. Variable backwater causes a variable energy slope for the same stage. Discharge is thus a function of both stage and slope and the relation is termed as slope-stage-discharge relation.

The stage is measured continuously at the main gauging station. The slope is estimated by continuously observing the stage at an additional gauge station, called the auxiliary gauge station. The auxiliary gauge station is established some distance downstream of the main station. Time synchronisation in the observations at the gauges is necessary for precise estimation of slope. The distance between these gauges is kept such that it gives an adequate representation of the slope at the main station and at the same time the uncertainty in the estimation is also smaller. When both main and auxiliary gauges are set to the same datum, the difference between the two stages directly gives the fall in the water surface. Thus, the fall between the main and the auxiliary stations is taken as the measure of surface slope. This fall is taken as the third parameter in the relationship and the rating is therefore also called stage-fall-discharge relation.

Discharge using Manning's equation can be expressed as:

$$Q = K_m R^{2/3} S^{1/2} A \tag{9.28)}$$

Energy slope represented by the surface water slope can be represented by the fall in level between the main gauge and the auxiliary gauge. The slope-stage-discharge or stage-fall-discharge method is represented by:

$$\frac{Q_m}{Q_r} = \left(\frac{S_m}{S_r}\right)^p = \left(\frac{F_m}{F_r}\right)^p \tag{9.29}$$

where: Q_m = the measured (backwater affected) discharge

Q_r = a reference discharge

F_m = the measured fall

F_r = a reference fall

p = a power parameter between 0.4 and 0.6

From the Manning's equation given above, the exponent "p" would be expected to be ½. The fall (F) or the slope (S = F/L) is obtained by the observing the water levels at the main and auxiliary gauge. Since, there is no assurance that the water surface profile between these gauges is a straight line, the effective value of the exponent can be different from ½ and must be determined empirically.

An initial plot of the stage discharge relationship (either manually or by computer) with values of fall against each observation, will show whether the relationship is affected by variable slope, and whether this occurs at all stages or is affected only when the fall reduces below a particular value. In the absence of any channel control, the discharge would be affected by variable fall at all times and the correction is applied by the constant fall method. When the discharge is affected only when the fall reduces below a given value the normal (or limiting) fall method is used.

Constant fall method:

The constant fall method is applied when the stage-discharge relation is affected by variable fall at all times and for all stages. The fall applicable to each discharge measurement is determined and plotted with each stage discharge observation on the plot. If the observed falls do not vary greatly, an average value (reference fall or constant fall) F_r is selected.

Manual computation

For manual computation an iterative graphical procedure is used. Two curves are used (Figures 9.15 and 9.16):

- All measurements with fall of about F_r are fitted with a curve as a simple stage discharge relation (Figure 9.15). This gives a relation between the measured stage h and the reference discharge Q_r.
- A second relation, called the adjustment curve, either between the measured fall, F_m, or the ratio of the measured fall for each gauging and the constant fall (F_m / F_r), and the discharge ratio (Q_m / Q_r) (Figure 9.16)
- This second curve is then used to refine the stage discharge relationship by calculating Q_r from known values of Q_m and F_m/F_r and then replotting h against Q_r .
- A few iterations may be done to refine the two curves.

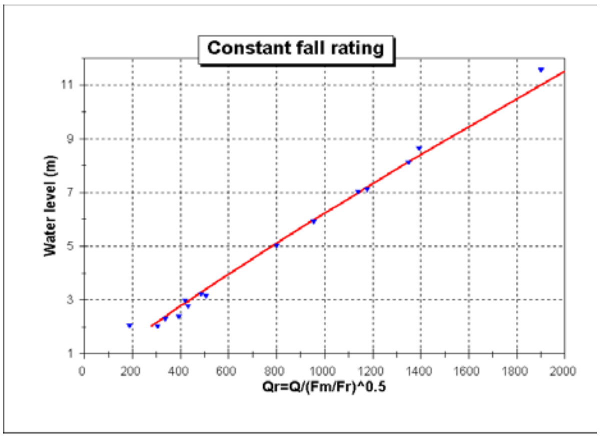


Figure 9.15: Q_r=f(h) in constant fall rating

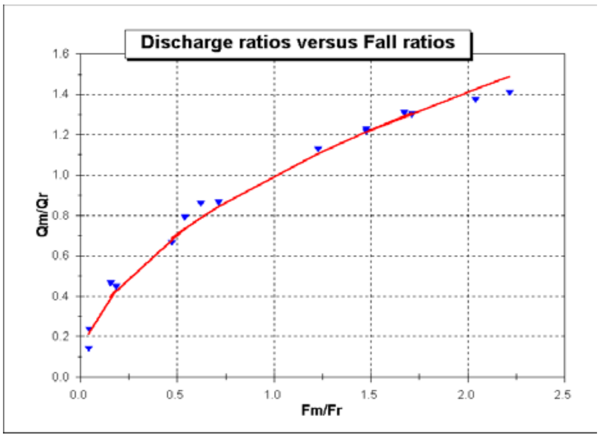


Figure 9.16: $Q_m/Q_r = f(F_m/F_r)$

The discharge at any time can be then be computed as follows:

- For the observed fall (F_m) calculate the ratio (F_m/F_r)
- read the ratio (Q_m / Q_r) from the adjustment curve against the calculated value of (F_m/F_r)
- multiply the ratio (Q_m / Q_r) with the reference discharge Q_r obtained for the measured stage h from the curve between stage h and reference discharge Q_r.

Computer computation

For computer computation, this procedure is simplified by mathematical fitting and optimisation. First, as before, a reference (or constant) fall (*F_r*) is selected from amongst the most frequently observed falls.

A rating curve, between stage *h* and the reference discharge (*Q_r*), is then fitted directly by estimating:

$$Q_r = Q_m \left(\frac{F_r}{F_m}\right)^p \tag{9.30}$$

where p is optimised between 0.4 and 0.6 based on minimisation of standard errors.

The discharge at any time, corresponding to the measured stage h and fall F_m , is then calculated by first obtaining Q_r from the above relationship and then calculating discharge as:

$$Q = Q_r \left(\frac{F_m}{F_r}\right)^p \tag{9.31}$$

A special case of constant fall method is the unit fall method in which the reference fall is assumed to be equal to unity. This simplifies the calculations and thus is suitable for manual method.

Normal Fall Method:

The normal or limiting fall method is used when there are times when backwater is not present at the station. Examples are when a downstream reservoir is drawn down or where there is low water in a downstream tributary or main river.

Manual procedure

The manual procedure is as follows:

- Plot stage against discharge, noting the fall at each point. The points at which backwater has no effect are identified first. These points normally group at the extreme right of the plotted points. This is equivalent to the simple rating curve for which a *Q_r -h* relationship may be fitted (where *Q_r* in this case is the reference or normal discharge) (Figure 9.17).

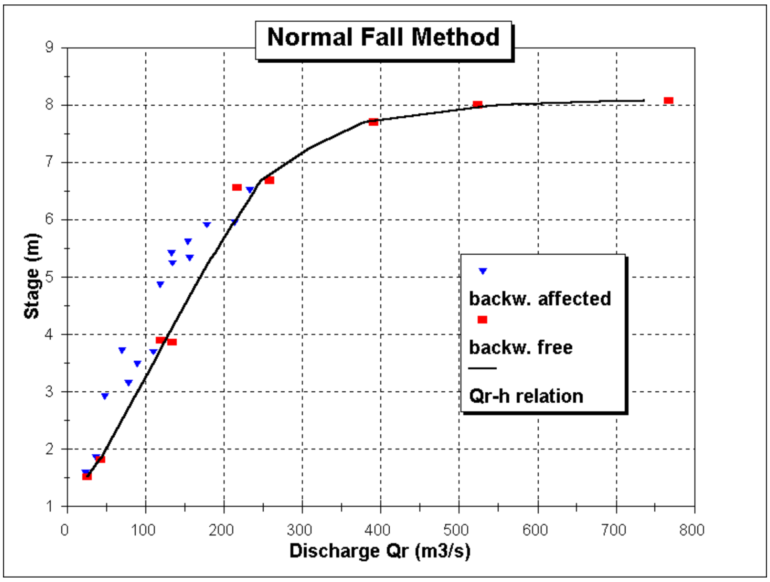


Figure 9.17:
Q_r-h relationship for Normal Fall Method

- Plot the measured fall against stage for each gauging and draw a line through those observations representing the minimum fall, but which are backwater free. This represents the normal or limiting fall F_r (Figure 9.18). It is observed from Figure 9.18 that the line separates the backwater affected and backwater free falls.

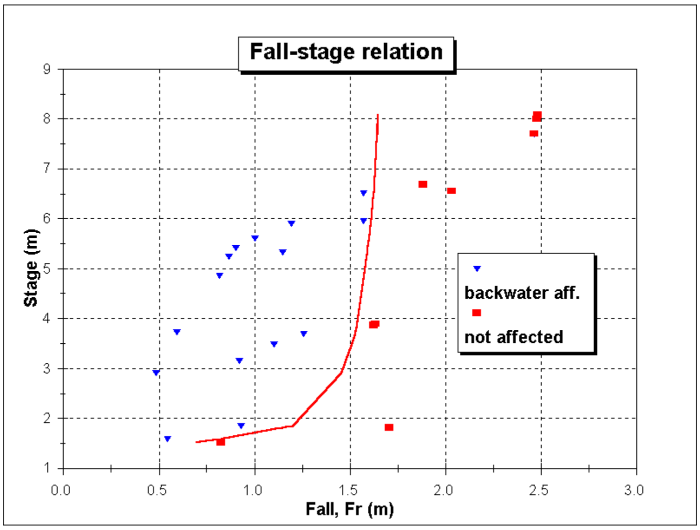


Figure 9.18:
F_r-h relationship

- For each discharge measurement derive Q_r using the discharge rating and F_r , the normal fall from the fall rating.
- For each discharge measurement compute Q_m/Q_r and F_m/F_r and draw an average curve (Figure 9.19).

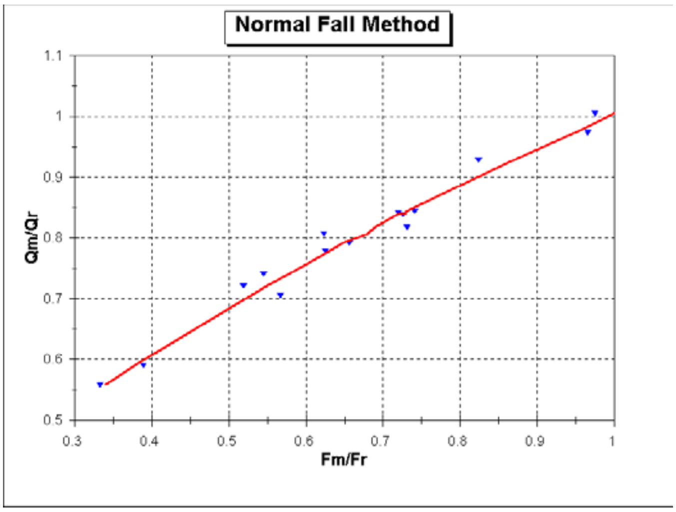


Figure 9.19:
Q_m/Q_r - F_m/F_r relationship

- As for the constant fall method, the curves may be successively adjusted by holding two graphs constant and re-computing and plotting the third. No more than two or three iterations are usually required.

The discharge at any time can be then be computed as follows:

- From the plot between stage and the normal (or limiting) fall (F_r), find the value of F_r for the observed stage h
- For the observed fall (F_m), calculate the ratio (F_m/F_r)

- Read the ratio (Q_m / Q_r) from the adjustment curve against the calculated value of (F_m/F_r)
- Obtain discharge by multiplying the ratio (Q_m / Q_r) with the reference discharge Q_r obtained for the measured stage h from the curve between stage h and reference discharge Q_r.

Computer procedure

The computer procedure considerably simplifies computation and is as follows:

- Compute the backwater-free rating curve using selected current meter gaugings (the Q_r -h relationship).
- Using values of Q_r derived from (1) and F_r derived from:

$$F_r = F_m \left(\frac{Q_r}{Q_m} \right)^{1/p} \tag{9.32}$$

a parabola is fitted to the reference fall in relation to stage (h) as:

$$F_r = a + b\,h + c\,h^2 \tag{9.33}$$

The parameter p is optimised between 0.4 and 0.6.

The discharge at any time, corresponding to the measured stage h and fall F_m , is then calculated by:

- obtaining F_r for the observed h from the parabolic relation between h and F_r
- obtaining Q_r from the backwater free relationship established between h and Q_r
- then calculating discharge corresponding to measured stage h as:

$$Q = Q_r \left(\frac{F_m}{F_r} \right)^{p} \tag{9.34}$$

9.3.5 RATING CURVE WITH UNSTEADY FLOW CORRECTION

Gauging stations not subjected to variable slope because of backwater may still be affected by variations in the water surface slope due to high rates of change in stage. This occurs when the flow is highly unsteady and the water level is changing rapidly. At stream gauging stations located in a reach where the slope is very flat, the stage-discharge relation is frequently affected by the superimposed slope of the rising and falling limb of the passing flood wave. During the rising stage, the velocity and discharge are normally greater than they would be for the same stage under steady flow conditions. Similarly, during the falling stage the discharge is normally less for any given gauge height than it is when the stage is constant. This is due to the fact that the approaching velocities in the advancing portion of the wave are larger than in a steady uniform flow at the corresponding stages. In the receding phase of the flood wave the converse situation occurs with reduced approach velocities giving lower discharges than in equivalent steady state case.

Thus, the stage discharge relationship for an unsteady flow will not be a single-valued relationship as in steady flow but it will be a looped curve as shown in the example below. The looping in the stage discharge curve is also called hysteresis in the stage-discharge relationship. From the curve it can be easily seen that at the same stage, more discharge passes through the river during rising stages than in the falling ones.

Application

- For practical purposes the discharge rating must be developed by the application of adjustment factors that relate unsteady flow to steady flow. Omitting the acceleration terms in the dynamic flow equation the relation between the unsteady and steady discharge is expressed in the form:

$$Q_m = Q_r \sqrt{\left(1 + \frac{1}{c\,S_0}\frac{dh}{dt}\right)} \tag{9.35}$$

where: Q_m = measured discharge

Q_r = estimated steady state discharge from the rating curve

c = wave velocity (celerity)

S_0 = energy slope for steady state flow

dh/dt = rate of change of stage derived from the difference in gauge height at the beginning and end of a gauging (+ for rising ; - for falling)

Q_r is the steady state discharge and is obtained by establishing a rating curve as a median curve through the uncorrected stage discharge observations or using those observations for which the rate of change of stage had been negligible. Care is taken to see that there are sufficient number of gaugings on rising and falling limbs if the unsteady state observations are considered while establishing the steady state rating curve.

Rearranging the above equation gives:

$$\frac{1}{c\,S_0} = \frac{(Q_m / Q_r)^2 - 1}{dh/dt} \tag{9.36}$$

The quantity (dh/dt) is obtained by knowing the stage at the beginning and end of the stage discharge observation or from the continuous stage record. Thus the value of factor ($1/cS_0$) can be obtained by the above relationship for every observed stage. The factor ($1/cS_0$) varies with stage and a parabola is fitted to its estimated values and stage as:

$$\frac{1}{c\,S_0} = a + b\,h + ch^2 \tag{9.37}$$

A minimum stage h_{min} is specified beyond which the above relation is valid. A maximum value of factor ($1/cS_0$) is also specified so that unacceptably high value can be avoided from taking part in the fitting of the parabola.

Thus unsteady flow corrections can be estimated by the following steps:

- Measured discharge is plotted against stage and beside each plotted point is noted the value of dh/dt for the measurement (+ or -)
- A trial Q_s rating curve representing the steady flow condition where dh/dt equals zero is fitted to the plotted discharge measurements.
- A steady state discharge Q_r is then estimated from the curve for each discharge measurement and Q_m , Q_r and dh/dt are together used in the Equation 9.35 to compute corresponding values of the adjustment factor $1/cS_0$
- Computed values of $1/cS_0$ are then plotted against stage and a smooth (parabolic) curve is fitted to the plotted points

For obtaining unsteady flow discharge from the steady rating curve the following steps are followed:

- obtain the steady state flow Q_r for the measured stage h
- obtain factor ($1/cS_0$) by substituting stage h in the parabolic relation between the two
- obtain (dh/dt) from stage discharge observation timings or continuous stage records
- substitute the above three quantities in the Equation 35 to obtain the true unsteady flow discharge

The computer method of analysis using HYMOS mirrors the manual method described above.

It is apparent from the above discussions and relationships that the effects of unsteady flow on the rating are mainly observed in larger rivers with very flat bed slopes (with channel control extending far downstream) together with significant rate change in the flow rates. For rivers with steep slopes, the looping effect is rarely of practical consequence. Although there will be variations depending on the catchment climate and topography, the potential effects of rapidly changing discharge on the rating should be investigated in rivers with a slope of 1 metre/ km or less. Possibility of a significant unsteady effect (say more than 8–10%) can be judged easily by making a rough estimate of ratio of unsteady flow value with that of the steady flow value.

Example 9.1

The steps to correct the rating curve for unsteady flow effects is elaborated for station MAHEMDABAD on WAZAK river. The scatter plot of stage discharge data for 1997 is shown in Figure 9.20a. From the curve is apparent that some shift has taken place, see also Figure 9.20. The shift took place around 24 August.

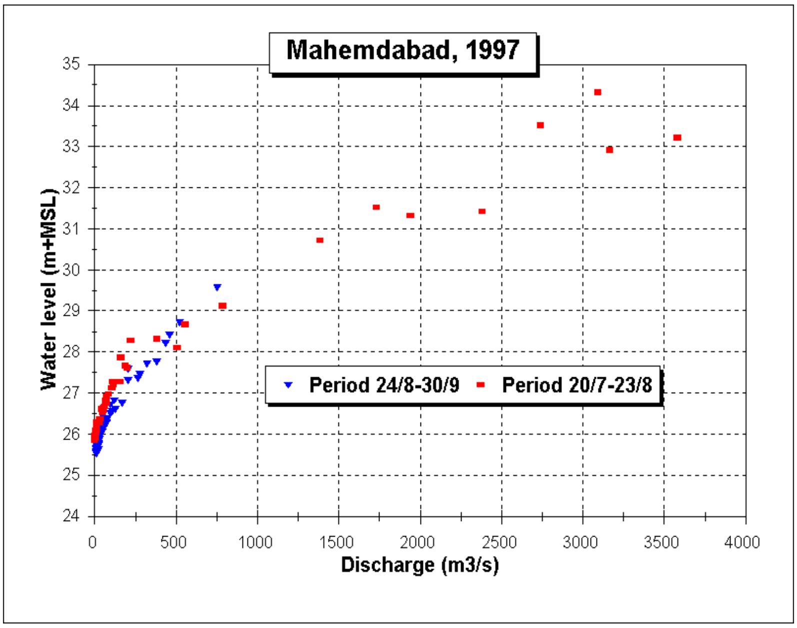


Figure 9.20a: Stage-discharge data of station Mahemabad on Wazak river, 1997

In the analysis therefore only the data prior that date were considered. The scatter plots clearly show a looping for the higher flows. To a large extent, this looping can be attributed to unsteady flow phenomenon. The Jones method is therefore applied. The first fit to the scatter plot, before any correction, is shown in Figure 9.20c.

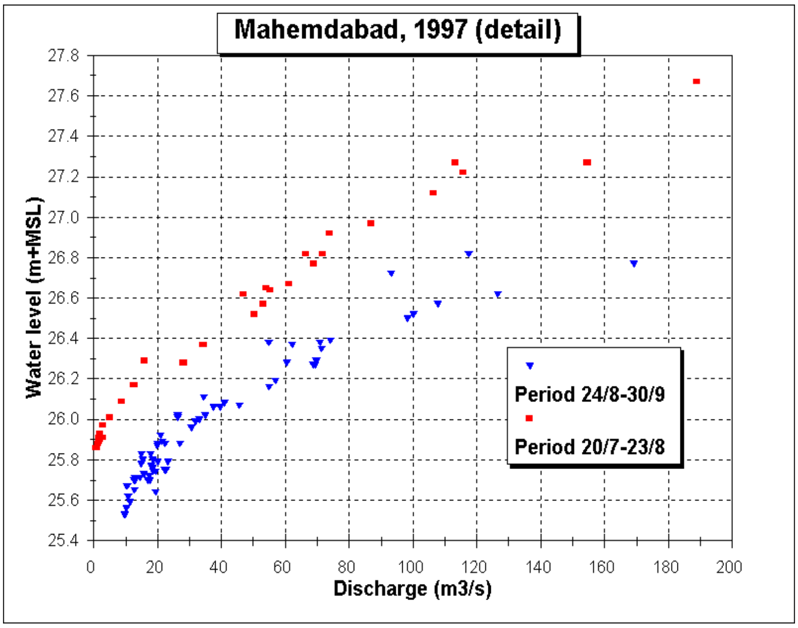


Figure 9.20b: Detail of stage discharge data for low flows, clearly showing the shief

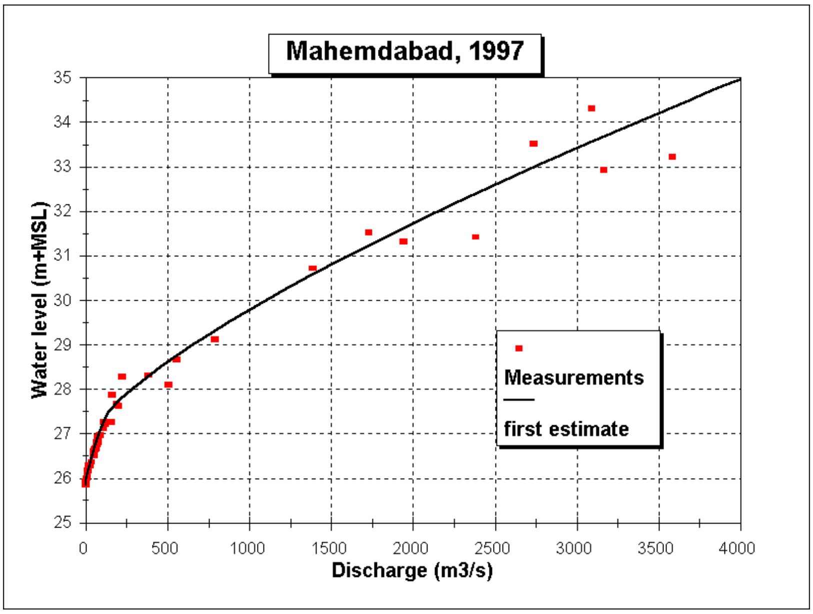


Figure 9.20c: First fit to stage-discharge data, prior to adjustment

Based on this relation and the observed discharges and water level changes values for $1/cS_0$ were obtained. These data are depicted in Figure 9.20d. The scatter in the latter plot is seen to be considerable. An approximate relationship between $1/cS_0$ and h is shown in the graph.

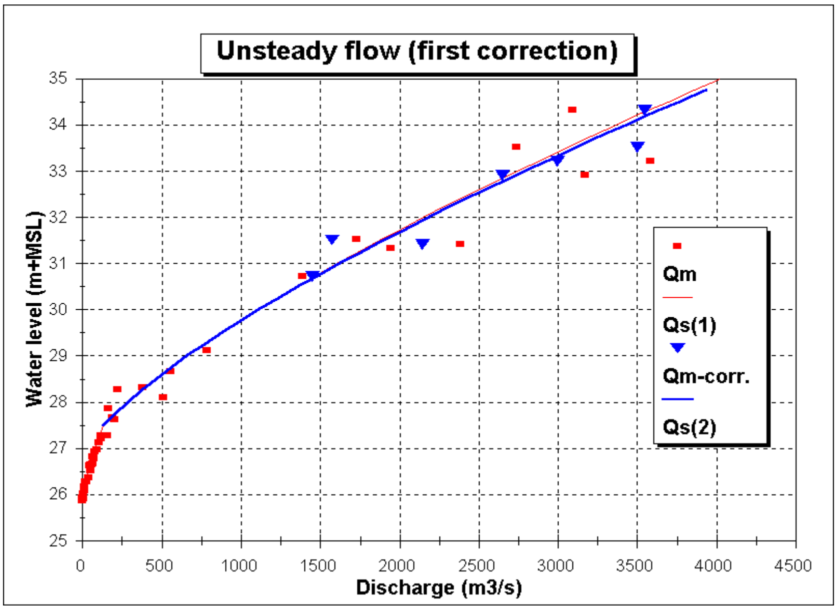


Figure 9.20d: Scatter plot of 1/cS$_0$ as function of stage, with approximate relation.

With the values for 1/cS$_0$ taken from graph the unsteady flow correction factor is computed and steady state discharges are computed. These are shown in Figure 9.20e, together with the uncorrected discharges. It is observed that part of the earlier variation is indeed removed. A slightly adjusted curve is subsequently fit to the stage and corrected flows.

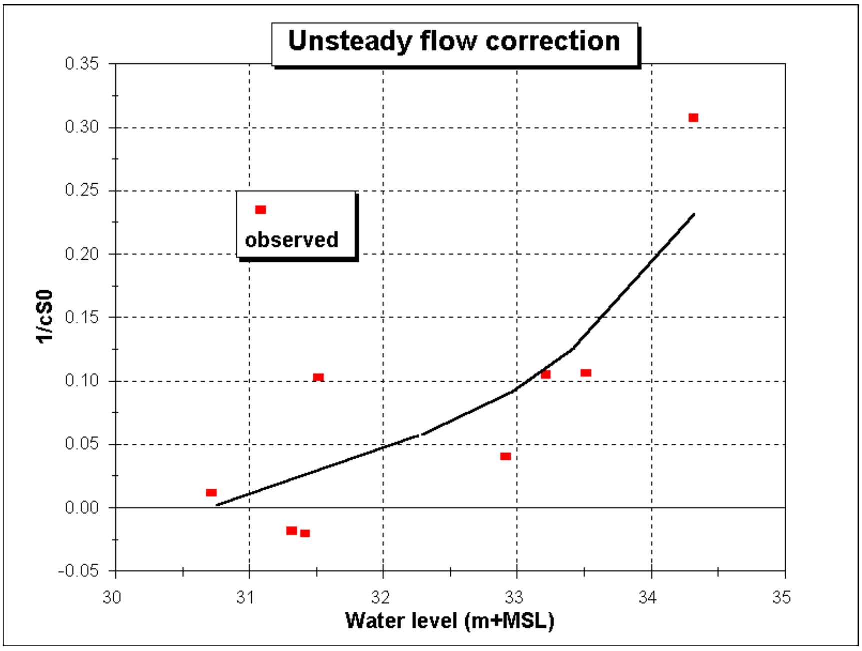


Figure 9.20e: Rating curve after first adjustment trial for unsteady flow.

Note that the looping has been eliminated, though still some scatter is apparent.

9.3.6 RATING RELATIONSHIPS FOR STATIONS AFFECTED BY SHIFTING CONTROL

For site selection it is a desirable property of a gauging station to have a control which is stable, but no such conditions may exist in the reach for which flow measurement is required, and the selected gauging station may be subject to shifting control. Shifts in the control occur especially in alluvial sand-bed streams. However, even in stable stream channels shift will occur, particularly at low flow because of weed growth in the channel, or as a result of debris caught in the control section.

In alluvial sand-bed streams, the stage-discharge relation usually changes with time, either gradually or abruptly, due to scour and silting in the channel and because of moving sand dunes and bars. The extent and frequency with which changes occur depends on typical bed material size at the control and velocities typically occurring at the station. In the case of controls consisting of cobble or boulder sized alluvium, the control and hence the rating may change only during the highest floods. In contrast, in sand bed rivers the control may shift gradually even in low to moderate flows. Intermediate conditions are common where the bed and rating change frequently during the monsoon but remain stable for long periods of seasonal recession.

For sand bed channels the stage-discharge relationship varies not only because of the changing cross section due to scouring or deposition but also because of changing roughness with different bed forms. Bed configurations occurring with increasing discharge are ripples, dunes, plane bed, standing waves, antidunes and chute and pool (Figure 9.21). The resistance of flow is greatest in the dunes range. When the dunes are washed out and the sand is rearranged to form a plane bed, there is a marked decrease in bed roughness and resistance to the flow causing an abrupt discontinuity in the stage-discharge relation. Fine sediment present in water also influences the configuration of sand-bed and thus the resistance to flow. Changes in water temperature may also alter bed form, and hence roughness and resistance to flow in sand bed channels. The viscosity of water will increase with lower temperature and thereby mobility of the sand will increase.

For alluvial streams where neither bottom nor sides are stable, a plot of stage against discharge will very often scatter widely and thus be indeterminate (Figure 9.22) However, the hydraulic relationship becomes apparent by changing the variables. The effect of variation in bottom elevation and width is eliminated by replacing stage by mean depth (hydraulic radius) and discharge by mean velocity respectively. Plots of mean depth against mean velocity are useful in the analysis of stage-discharge relations, provided the measurements are referred to the same cross-section.

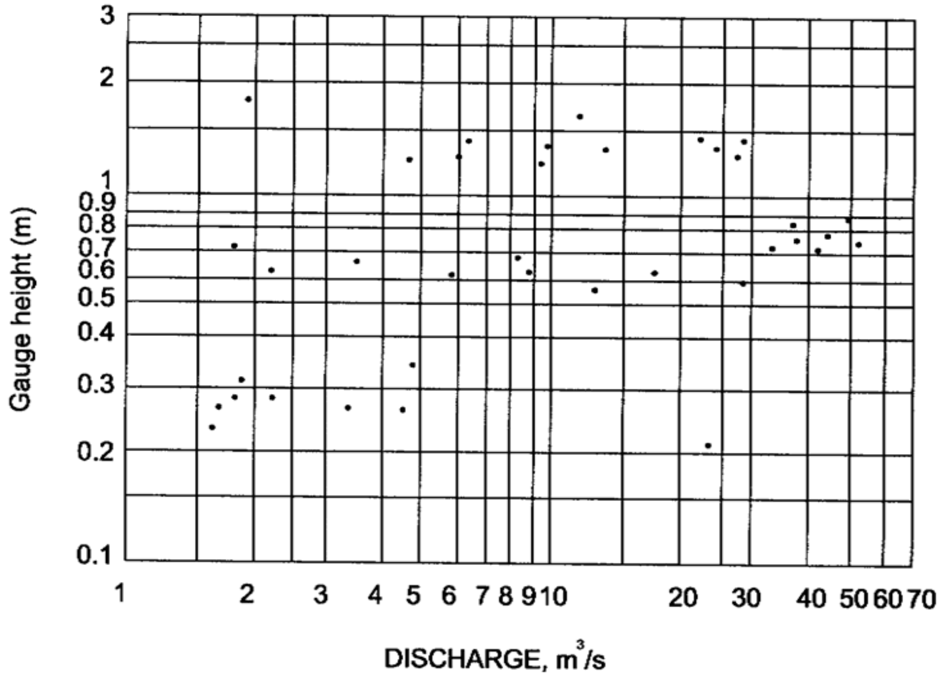
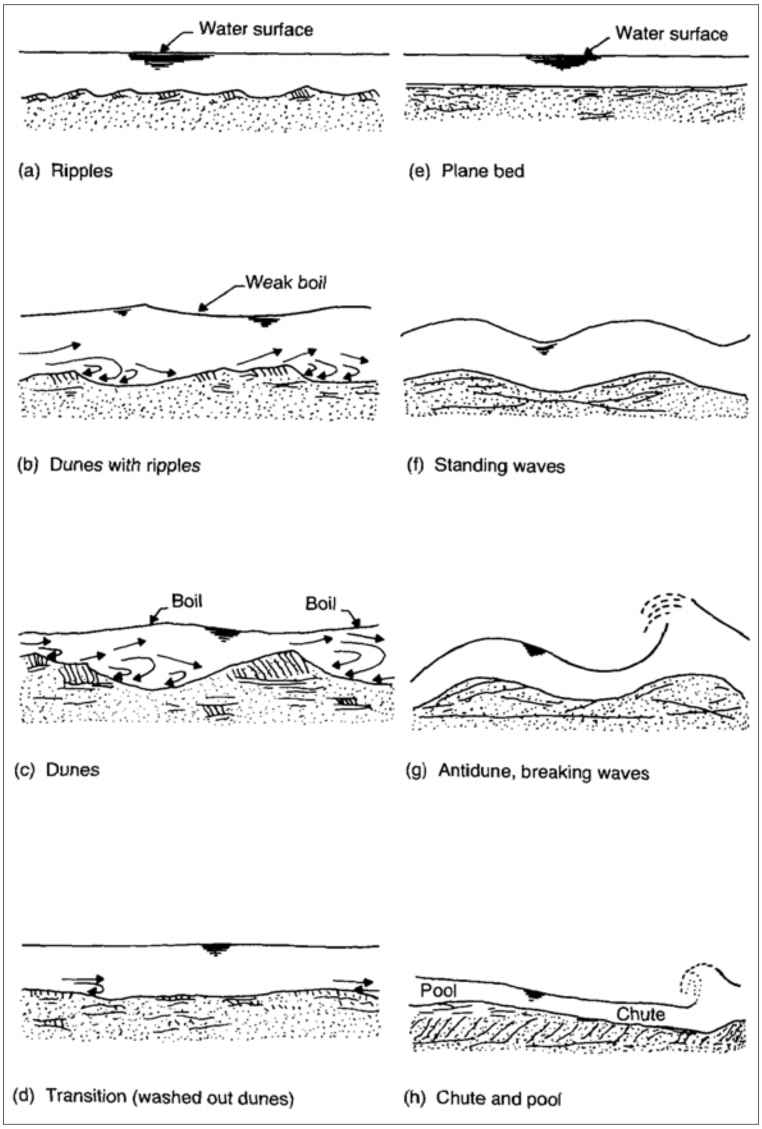


Figure 9.21: Plot of discharge against stage for a sand-bed channel with indeterminate stage-discharge relation



(a) Ripples
(e) Plane bed
(b) Dunes with ripples
(f) Standing waves
(c) Dunes
(g) Antidune, breaking waves
(d) Transition (washed out dunes)
(h) Chute and pool

Figure 9.22:
Bed and surface configurations configurations in sand-bed channels

These plots will identify the bed-form regime associated with each individual discharge measurement. (Figure 9.23). Thus measurements associated with respective flow regimes, upper or lower, are considered for establishing separate rating curves. Information about bed-forms may be obtained by visual observation of water surfaces and noted for reference for developing discharge ratings.

There are four possible approaches depending on the severity of scour and on the frequency of gauging:

- Fitting a simple rating curve between scour events
- Varying the zero or shift parameter
- Application of Stout’s shift method
- Flow determined from daily gauging

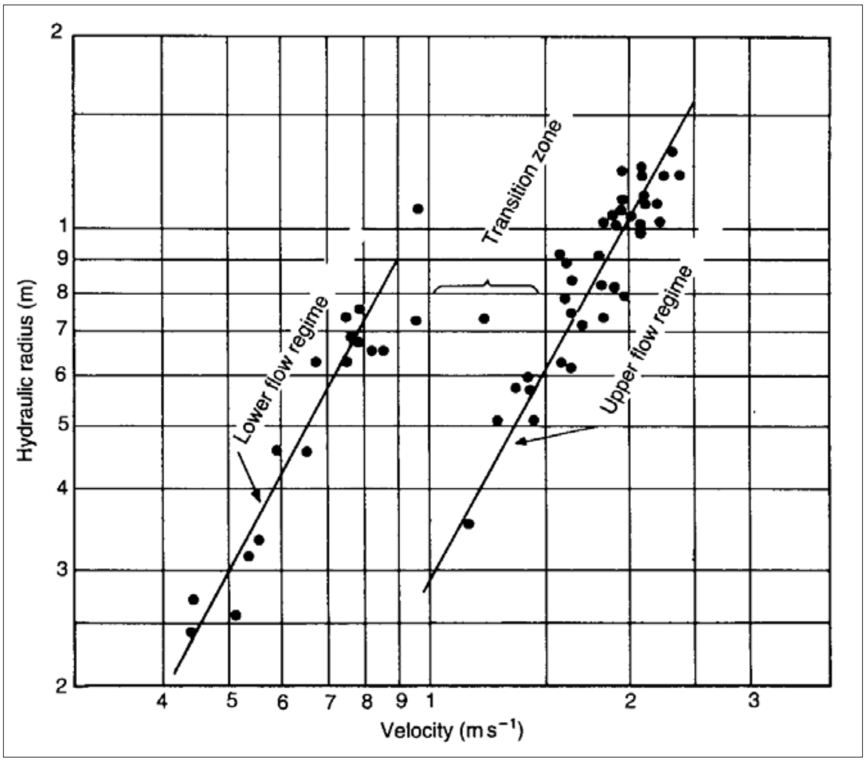


Figure 9.23: Relation of mean velocity to hydraulic radius of channel in Figure 9.22

Fitting a simple rating curve between scour events

Where the plotted rating curve shows long periods of stability punctuated by infrequent flood events which cause channel adjustments, the standard procedure of fitting a simple logarithmic equation of the form $Q = c_1(h + a_1)^{b1}$ should be applied to each stable period. This is possible only if there are sufficient gaugings in each period throughout the range of stage.

To identify the date of change from one rating to the next, the gaugings are plotted with their date or number sequence. The interval in which the change occurred is where the position of sequential plotted gaugings moves from one line to the next. The processor should then inspect the gauge observation record for a flood event during the period and apply the next rating from that date.

Notes from the Field Record book or station log must be available whilst inspection and stage discharge processing is carried out. This provides further information on the nature and timing of the event and confirms that the change was due to shifting control rather than to damage or adjustment to the staff gauge.

Varying the zero or shift parameter

Where the plotted rating curve shows periods of stability but the number of gaugings is insufficient to define the new relationship over all or part of the range, then the parameter ‘a’ in the standard relationship $Q = c_1(h + a_1)^{b1}$ may be adjusted. The parameter ‘a_1’ represents the datum correction between the zero of the gauges and the stage at zero flow. Scour or deposition causes a shift in this zero flow stage and hence a change in the value ‘a_1’..

The shift adjustment required can be determined by taking the average difference (Δa) between the rated stage (h_r) for measured flow (Q_m) and measured stage (h_m) using the previous rating. i.e.

$$\Delta a = \sum_{i=1}^{n} (h_r - h_m)/n \tag{9.38}$$

The new rating over the specified range then becomes:

$$Q = c_1(h + a_1 + \Delta a)^{b1} \tag{9.39}$$

The judgement of the processor is required as to whether to apply the value of Δa over the full range of stage (given that the % effect will diminish with increasing stage) or only in the lower range for which current meter gauging is available. If there is evidence that the rating changes from unstable section control at low flows to more stable channel control at higher flows, then the existing upper rating should continue to apply.

New stage ranges and limits between rating segments will require to be determined. The method assumes that the channel hydraulic properties remain unchanged except for the level of the datum. Significant variation from this assumption will result in wide variation in (h_r - h_m) between included gaugings. If this is the case then Stout's shift method should be used as an alternative.

Stout's shift method (Not Recommended to be Applied)

For controls which are shifting continually or progressively, Stout's method is used. In such instances the plotted current meter measurements show a very wide spread from the mean line and show an insufficient number of sequential gaugings with the same trend to split the simple rating into several periods. The procedure is as follows:

- Fit a mean relationship to (all) the points for the period in question.
- Determine h_r (the rated stage) from measured Q_m by reversing the power type rating curve or from the plot:

$$h_r = (Q_m / c)^{1/b} - a \tag{9.40}$$

- Individual rating shifts (Δh), as shown in Figure 9.24, are then:

$$\Delta h = h_r - h_m \tag{9.41}$$

- These Δh stage shifts are then plotted chronologically and interpolated between successive gaugings (Figure 9.24) for any instant of time t as Δh_t.
- These shifts, Δh_t ,are used as a correction to observed gauge height readings and the original rating applied to the corrected stages, i.e.

$$Q_t = c_1(h_t + \Delta h_t + a_1)^{b1} \tag{9.42}$$

The Stout method will only result in worthwhile results where:

a) Gauging is frequent, perhaps daily in moderate to high flows and even more frequently during floods.

b) The mean rating is revised periodically - at least yearly.

The basic assumption in applying the Stout's method is that the deviations of the measured discharges from the established stage-discharge curve are due only to a change or shift in the station control, and that the corrections applied to the observed gauge heights vary gradually and systematically between the days on which the check measurements are taken. However, the deviation of a discharge measurement from an established rating curve may be due to:

a) gradual and systematic shifts in the control,

b) abrupt random shifts in the control, and

c) error of observation and systematic errors of both instrumental and a personnel nature.

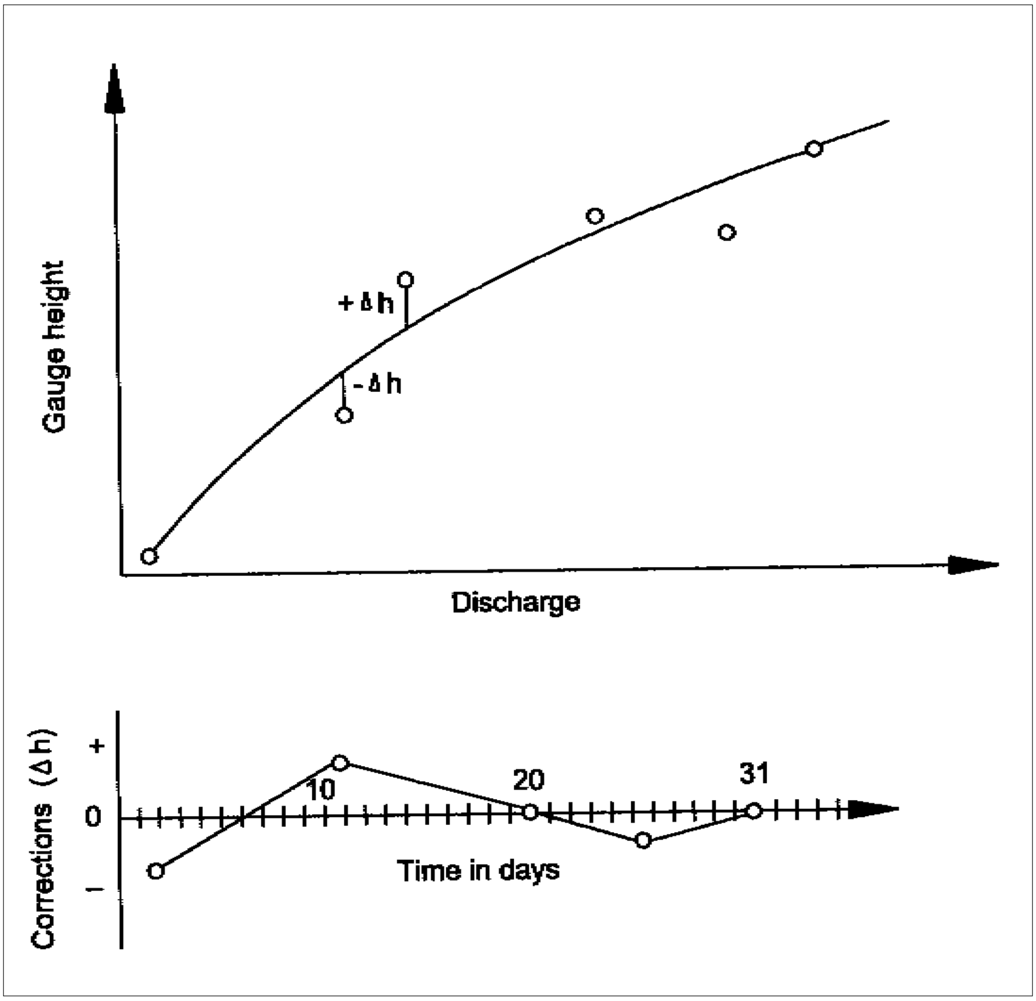


Figure 9.24: Stout's method for correcting stage readings when control is shifting

Stout's method is strictly appropriate for making adjustments for the first type of error only. If the check measurements are taken frequently enough, fair adjustments may be made for the second type of error also. The drawback of the Stout's method is that all the errors in current meter observation are mixed with the errors due to shift in control and are thus incorporated in the final estimates of discharge. The Stout method must therefore never be used where the rating is stable, or at least sufficiently stable to develop satisfactory rating curves between major shift events; use of the Stout method in such circumstances loses the advantage of a fitted line where the standard error of the line S_{mr} is 20% of the standard error of individual gaugings (S_e). Also, when significant observational errors are expected to be present it is strongly recommended not to apply this method for establishing the rating curve.

Flow determined from daily gauging

Stations occur where there is a very broad scatter in the rating relationship, which appears neither to result from backwater or scour and where the calculated shift is erratic. A cause may be the irregular opening and closure of a valve or gate in a downstream structure. Unless there is a desperate need for such data, it is recommended that the station be moved or closed. If the station is continued, a daily measured discharge may be adopted as the daily mean flow. This practice however eliminates the daily variations and peaks in the record.

It is emphasised that even using the recommended methods, the accuracy of flow determination at stations which are affected by shifting control will be much less than at stations not so affected. Additionally, the cost of obtaining worthwhile data will be considerably higher. At many such stations uncertainties of ± 20 to 30% are the best that can be achieved and consideration should be given to whether such accuracy meets the functional needs of the station.

10 VALIDATION OF RATING CURVE

10.1 GENERAL

Validation of a rating curve is required both after the relationship has first been fitted and subsequently when new gaugings have been carried out, to assess whether these indicate a change in rating. Validation is also used to assess the reliability of historical ratings.

Current meter gauging is carried out with variable frequency depending on previous experience of the stability of the control and of the rating curve. As a minimum it is recommended that six gaugings per year are carried out even with a station with a stable section and previously gauged over the full range of level. At unstable sections many more gaugings are required. The deviation of such check gaugings from the previously established relationship is computed and any bias assessed to determine whether they belong to the same population as the previous stage-discharge relationship.

Graphical and numerical tests are designed to show whether gaugings fit the current relationship equally and without bias over the full range of flow and over the full time period to which it has been applied. If they do not, then a new rating should be developed as described in Chapter 9, but taking into account the deficiencies noted in validation.

Validation will be carried out at Divisional offices or at the State Data Processing Centre.

10.2 GRAPHICAL VALIDATION TESTS

10.2.1 GENERAL

Graphical tests are often the most effective method of validation. These include the following:

- Stage/discharge plot with the new gaugings
- Period/flow deviation scattergram
- Stage/flow deviation scattergram
- Cumulative deviation plot of gaugings.
- Stage/discharge plots with gaugings distinguished by season

Judgements based on graphical displays are often indicative rather than prescriptive - a judgement on the part of the data processor is still required.

10.2.2 STAGE/DISCHARGE PLOT WITH NEW GAUGINGS

The simplest means of validating the rating curve with respect to subsequent gaugings is to plot the existing rating curve with the new check gaugings. This is shown in the example for Station Pargaon. A rating curve is established for the period 30/6 – 3/8/97, see Figure 10.1. It shows a proper fit of the data to the existing rating curve, of which the numerical results are shown in Table 10.1. New data are available for the period 4-23/8/97. The new data with the existing rating curve are shown in Figure 10.2. From this plot it is observed that the new gaugings do not match with the existing curve. In Figure 10.3 the new gaugings are shown with the rating curve and its the 95% confidence limits (derived as t-times the standard error S_e). From this plot it can be judged if most check gaugings lie inside the confidence limits and thus whether they can be judged acceptable with respect to deviation. It is expected that 19 out of 20 observations will lie inside the limits if the standard error is considered at 5% significance level. However, except insofar as one can see whether all the new points lie

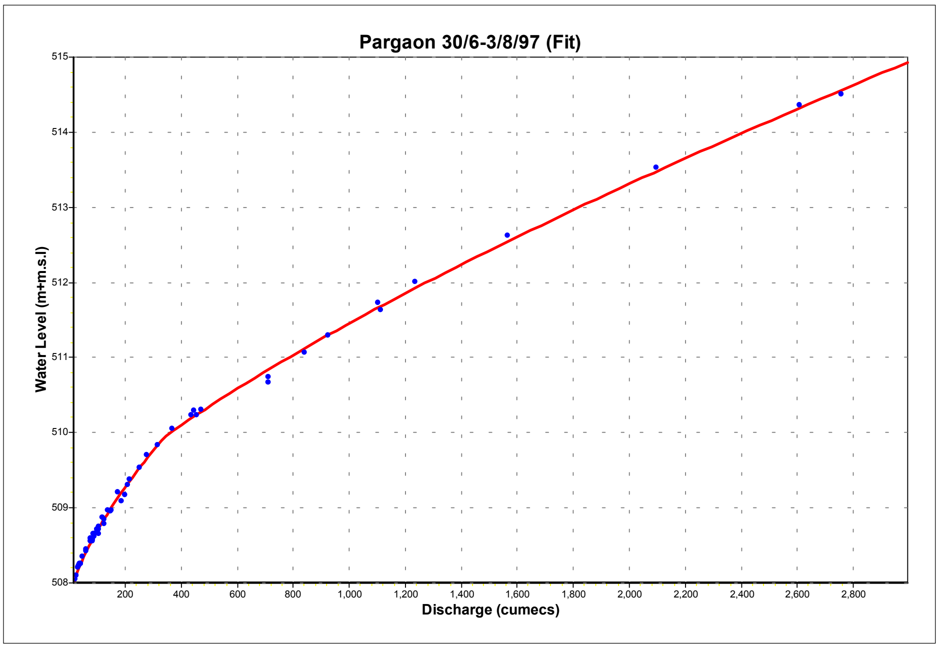


Figure 10.1: Rating curve for station PARGAON established based on data for the period 30/6-3/8/97

```
Analysis of stage-discharge data
 Station name :  PARGAON
 Data from 1997  6 30 to 1997  8  2
 Single channel

 Given boundaries for computation of rating curve(s)
 interval lower bound upper bound  nr. Of data
        1     508.000     510.200           49
        2     509.900     515.000           12

 Power type of equation q=c*(h+a)**b is used

 Boundaries / coefficients
 lower bound upper bound         a         b         c

      508.00      509.98  -507.730     1.489   .1043E+03
      509.98      515.00  -508.490     1.475   .1936E+03

 Number    W level     Q meas     Q comp       DIFf  Rel.dIFf       Semr
                 M       M3/S       M3/S       M3/S       0/0        0/0

     54    508.050     22.830     19.116      3.714     19.43       4.44
     56    511.240    820.010    861.213    -41.203     -4.78       2.58
     57    510.740    711.230    640.582     70.648     11.03       2.96
     62    512.620   1566.740   1568.961     -2.221      -.14       3.12
     63    514.510   2757.370   2735.225     22.145       .81       4.60
     64    514.360   2609.830   2635.279    -25.448      -.97       4.49
     65    513.530   2098.120   2104.625     -6.505      -.31       3.84
     66    512.010   1235.470   1239.485     -4.015      -.32       2.71
     67    511.730   1103.470   1096.850      6.620       .60       2.59

 Overall standard error =     6.061

 Statistics per interval
 Interval Lower bound Upper bound Nr.of data Standard error
        1     508.000     509.981         48           6.55
        2     509.981     515.000         12           4.01
```

Table 10.1: Results of rating curve fitting

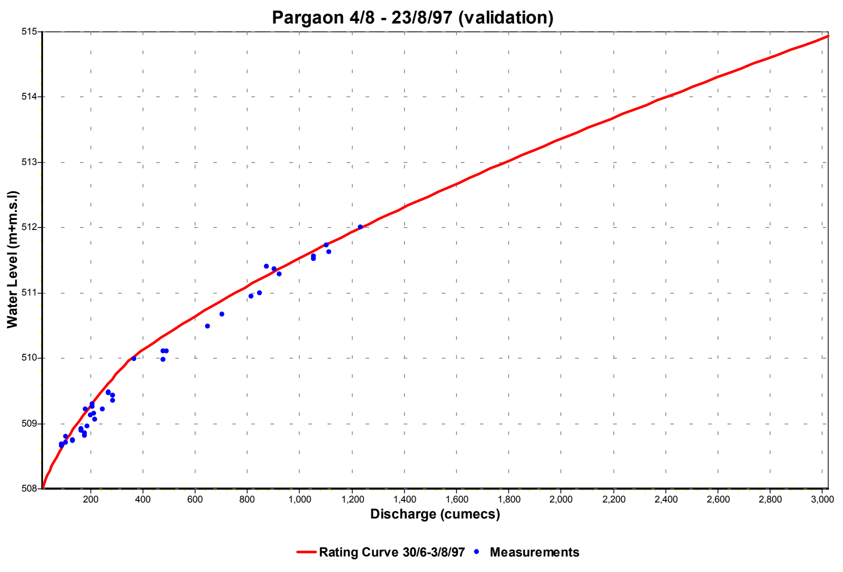


Figure 10.2: New gaugings at Pargaon station plotted against the existing rating curve

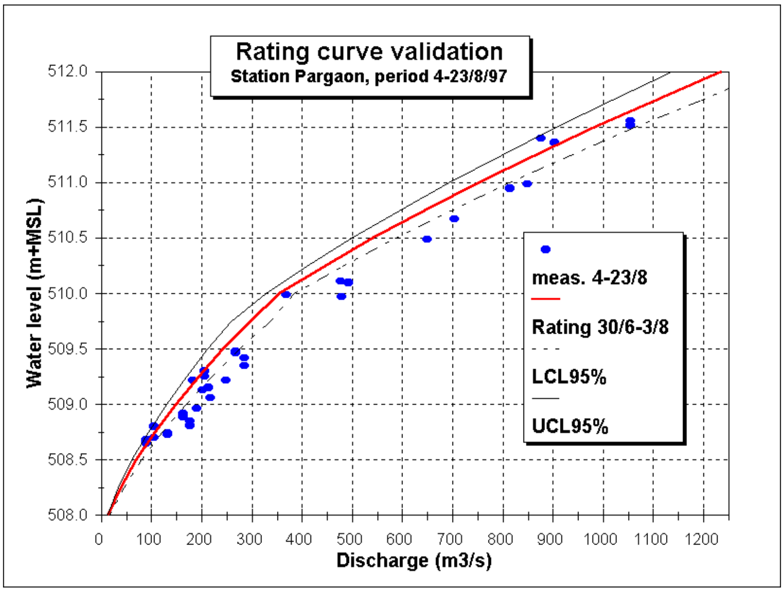


Figure 10.3:
New gaugings at Pargaon station plotted against existing rating curve with 95% confidence limits

above or below the previous regression line, the graph does not specifically address the problem of bias. For example, if some 25 new gaugings may all lie scattered within 95% confidence limits, it does not show any significant change in behaviour. However, if these points are plotted and sequence of each observation is also considered and if upon that a certain pattern of deviation (with respect to time) is perceivable and significant then such situation may warrant new ratings for different periods of distinct behaviour. For the Pargaon-case the plot confirms earlier observations that the new gaugings significantly differ from the existing rating.

10.2.3 PERIOD/FLOW DEVIATION SCATTERGRAM

A period/flow deviation scattergram (Figure 10.4) is a means of illustrating the negative and positive deviation of each current meter gauging from the present rating curve and whether there has been a gradual or sudden shift in the direction of deviations within the period to which the rating has been applied. It also shows whether recent additional gaugings show deviation from previous experience. In the example shown in Figure 10.4, percentage deviations are very high; there are far more gaugings with positive than with negative deviations. The rating is therefore biased and a revision of the rating is strongly recommended.

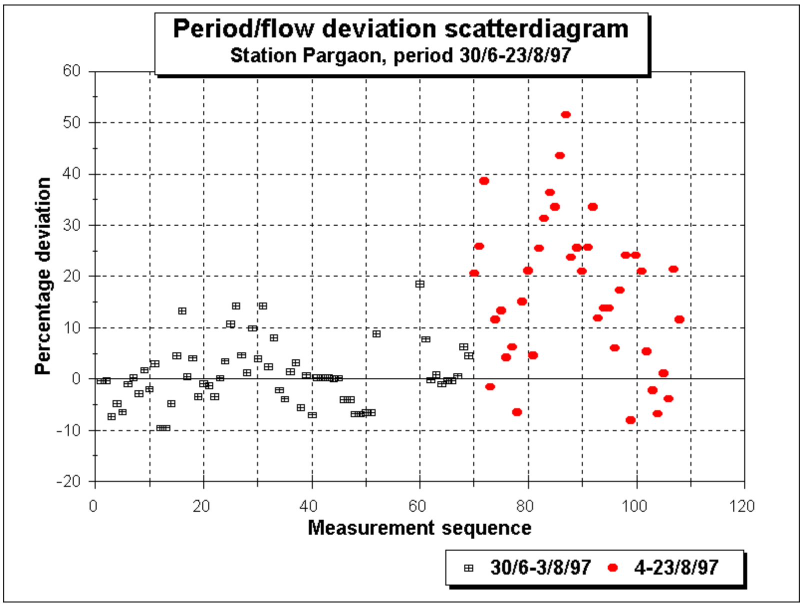


Figure 10.4:
Period-flow deviation scatterdiagram for Pargaon rating curve data and new gaugings

10.2.4 STAGE/FLOW DEVIATION DIAGRAM

A similar scattergram plot shows the percentage deviation with stage (Figure 10.5) and is a means of illustrating whether over certain ranges of stage the relationship is biased. Most recent gaugings can also be placed within this context.

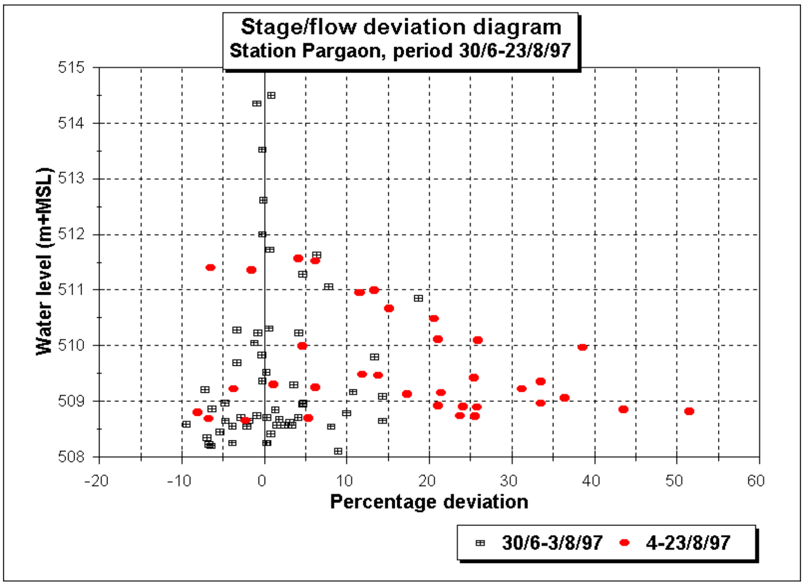


Figure 10.5:
Stage-flow deviation scatterdiagram for Pargaon rating curve data and new gaugings

In the example shown in Figure 10.5, there is some difference in deviation at different stages; particularly at the lower stages the differences are substantial. This plot also confirms the necessity for revision of the rating curve.

10.2.5 CUMULATIVE DEVIATION PLOT OF GAUGINGS

A plot of the cumulative deviation of gaugings from the rating curve give another indication of bias and whether that bias changes with time. Figure 10.6 shows such a plot for the example of station Pargaon. From the upward trend of the line for the new gaugings it is concluded that the new gaugings produce consistently higher flow values for the same stages than before.

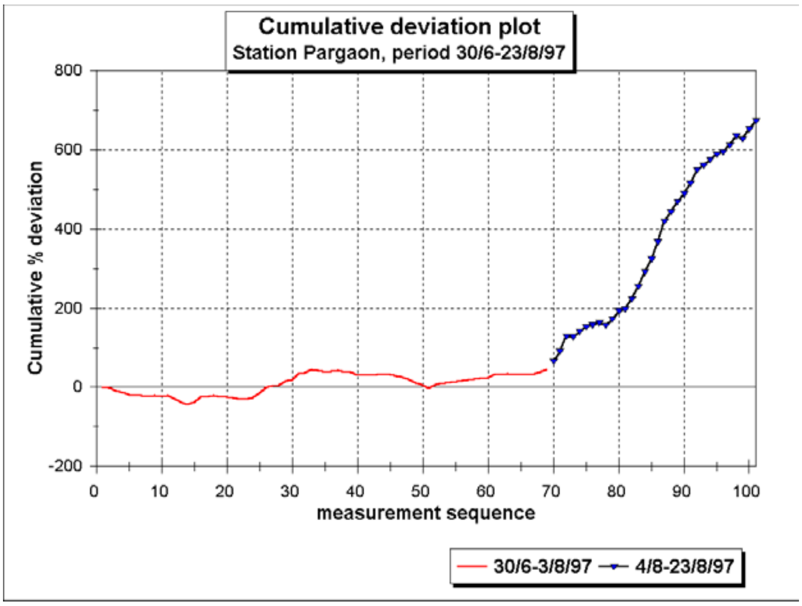


Figure 10.6:
Cummulative deviation plot for Paragon

10.2.6 STAGE DISCHARGE PLOTS WITH GAUGINGS DISTINGUISHED BY SEASON.

It is sometimes helpful to separate gaugings between seasons to demonstrate the effect of varying weed growth or other seasonal factors on the stage discharge relationship. The effects of weed growth may be expected to be at a maximum in low flows before the onset of the monsoon; monsoon high flows wash out the weed which increases progressively from the end of the rains. The discharge for given level may thus differ from one month to another. This shows up more clearly in rivers where winter low flows are little affected by weed growth than summer low flows and thus show much smaller spread. Where an auxiliary gauge is available, a backwater rating curve (normal fall method) may be used. Otherwise a simple rating curve may be used in weed absent periods and Stout's shift method during periods of variability.

10.3 NUMERICAL VALIDATION TESTS

10.3.1 USE OF STUDENT'S 'T' TEST TO CHECK GAUGINGS

A test such as Student's "t" test may be used to decide whether check gaugings can be accepted as part of the homogeneous sample of observations making up the existing stage-discharge curve. Such a test will indicate whether or not the stage-discharge relation requires re-calculation or the section requires recalibration.

In this test, the 't' statistic is calculated as the ratio of the mean deviation and the standard error of the difference of the means as:

$$t = \bar{d}_1 / S \qquad (10.1)$$

where: $\bar{d}_1$ = the mean deviation of the new gaugings (in percent) from the existing curve

and S = the standard error of the difference in the means expressed as:

$$S = a\sqrt{(N+N_1)/NN_1} \qquad (10.2)$$

where: N = the number of gaugings used to derive the existing rating

and N_1 = the number of new gaugings

a = given by the following expression:

$$a = \sqrt{\frac{\sum(d^2)+\sum(d_1-\bar{d}_1)^2}{N+N_1-2}} \qquad (10.3)$$

where: $\Sigma(d)^2$ = the sum of the squares of the percent differences for the old gaugings from the existing rating.

If this computed value of 't' = $\bar{d}$ / S is greater than the critical value of 't' for ($N + N_1 - 2$) degrees of freedom at 95% probability level then further action must be considered. Either the development of a new rating or a request to field staff for additional check gaugings. The critical values of Students 't' statistic at the 95% confidence level can be obtained from the standard tables available for the Student's 't' distribution. It should be noted that rating changes are more frequent and more noticeable in the low flow range. Review and validation is therefore done with respect to each range and, unless there is evidence to the contrary, unaffected ranges should retain the old rating but with the range limits adjusted for the new intersection.

As an example, the validation of the new gaugings at station Pargaon is shown in Table 10.2. The results of the ‘t’-test is seen to support earlier observation of significant deviation.

```
Validation stage-discharge data
 Station name :  PARGAON
 Data from 1997  8  4 to 1997  8 22

 Procedure    : Standard
 Equation type: Power
 Interval  Boundaries            Parameters:
      1       508.000   509.981  -507.730     1.489   104.324
      2       509.981   515.000  -508.490     1.475   193.614

 Data used to estimate parameters:
 Interval  St. error of est.  Number of data
     1            6.546             48
     2            4.013             12

 Number   W level    Q meas    Q comp      DIFf  Rel.dIFf      Semr
                M      M3/S      M3/S      M3/S       0/0       0/0

     81   509.990   368.450   352.261    16.189      4.60     15.48
     82   509.420   285.790   227.843    57.947     25.43     15.51
     83   509.220   247.940   188.880    59.060     31.27     11.81
     84   509.060   217.610   159.488    58.122     36.44     10.23
     85   508.960   189.660   141.964    47.696     33.60     10.57
     86   508.850   177.270   123.482    53.788     43.56     12.38
     87   508.810   177.270   116.972    60.298     51.55     13.39
     88   508.740   130.970   105.863    25.107     23.72     15.53
     89   508.730   130.970   104.309    26.661     25.56     15.88
     90   508.920   163.570   135.149    28.421     21.03     11.05
     91   508.890   163.570   130.107    33.463     25.72     11.55
     92   509.350   285.790   213.934    71.856     33.59     14.11
     93   509.480   268.610   239.991    28.619     11.93     16.73
     94   509.460   268.610   235.915    32.695     13.86     16.32


 Overall standard error =    14.729

 Statistics per interval
 Interval Lower bound Upper bound Nr.of data Standard error
        1     508.000     509.981         27          25.78
        2     509.981     515.000         15          13.18

 Results of student T-test on absence of bias
 Interval Degrees of freedom 95% T-value Actual T-value Result
        1                 73       1.993          7.673 Reject
        2                 25       2.060          2.920 Reject
```

Table 10.2: Results of validation

10.3.2 TEST FOR ABSENCE FROM BIAS IN SIGNS

A well-balanced rating curve must ensure that the number of positive and negative deviations of the observed values from the rating curve is evenly distributed. That is, the difference in number between the two should not be more than can be explained by chance fluctuations. The test is employed to see if the curve has been established in a balanced manner so that the two sets of discharge values, observed and estimated (from the curve), may be reasonably supposed to represent the same population.

This test is performed by counting observed points falling on either side of the curve. If Q_i is the observed value and Q_c the estimated value, then the expression, $Q_i - Q_c$, should have an equal chance of being positive or negative. In other words, the probability of $Q_i - Q_c$ being positive or negative is ½. Hence, assuming the successive signs to be independent of each other, the sequence of the differences may be considered as distributed according to the binomial law (p+q) N, where N is the number of observations, and p and q, are the probabilities of occurrence of positive and negative values are ½ each. The expected number of positive signs is positive signs is Np. Its standard deviation is √(Npq). The "t" statistic is then found by dividing the difference between the actual number of positive signs N_1 and expected number of positive signs Np by its standard deviation√(Npq):

$$t = \frac{|N_1 - Np| - 0.5}{\sqrt{Npq}} \qquad (10.4)$$

The resulting value is compared with the critical value of "t" statistic for 5% significance level for the degrees of freedom equal to the total number of stage discharge data. If the value of the critical "t" statistic is more than that obtained for the observed data then it can be considered that the data does not show any bias with respect to sign of the deviations between observed and computed discharges.

10.3.3 TEST FOR ABSENCE FROM BIAS IN VALUES

This test is designed to see if a particular stage discharge curve, on average, yields significant under estimates or over estimates as compared to the actual observations on which it is based. (Compare the graphical test using the period/flow deviation and stage /flow deviation scattergrams) The percentage differences are first worked out as:

$$P = 100\,(Q_i - Q_c)\,/\,Q_c \qquad (10.5)$$

If there are N observations and P_1, P_2, P_3, …, P_N are the percentage differences and P_{av} is the average of these differences, the standard error of P_{av} is given by:

$$Se = \sqrt{\frac{(\sum P_i - P_{av})^2)}{N(N-1)}} \qquad (10.6)$$

The average percent P_{av} is tested against its standard error to see if it is significantly different from zero. The "t" statistic for in this case is computed as:

$$t = (P_{av} - 0)\,/\,S_e \qquad (10.7)$$

If the critical value of "t" statistic for 5% significance level and N degrees of freedom is greater than the value computed above then it may be considered that there is no statistical bias in the observed values with respect to their magnitudes as compared with that obtained by the rating curve.

The percentage differences have been taken as they are comparatively independent of the discharge volume and are approximately normally distributed about zero mean value for an unbiased curve.

10.3.4 GOODNESS OF FIT TEST

Due to changes in the flow regime, it is possible that long runs of positive and/or negative deviations are obtained at various stages. This may also be due to inappropriate fitting of the rating curve. This test is carried out for long runs of positive and negative deviations of the observed values from the stage-discharge curve. The test is designed to ensure a balanced fit in reference to the deviations over different stages. (Compare the graphical tests using stage/flow deviation scattergram and cumulative deviation plot of gaugings)

The test is based on the number of changes of sign in the series of deviations (observed value minus expected or computed value). First of all, the signs of deviations, positive or negative, in discharge measurements in ascending order of stage are recorded. Then starting from the second sign of the series, “0” or “1” is placed under sign if the sign agrees or does not agree respectively with the sign immediately preceding it. For example,

```
+ - + + + +  - -  - -  + + + + -
  1 1 0 0 0 1 0 0 0 1 0 0 0 1
```

If there are N numbers in the original series, then there will be (N – 1) numbers in the derived series 11000100010001

If the observed values are regarded as arising from random fluctuations about the values estimated from the curve, the probability of a change in sign could be taken to be ½. However, this assumes that the estimated value is the median rather than the mean. If N is fairly large, a practical criterion may be obtained by assuming successive signs to be independent (i.e. by assuming that they arise only from random fluctuations), so that the number of “1”s (or “0”s) in the derived sequence of (N – 1) members may be judged as a binomial variable with parameters (N – 1) and ½.

From the above derived series, the actual number of changes of sign is noted. The expected number of changes of sign is computed by multiplying total possible numbers (i.e. N –1) with the probability of change of sign (i.e. ½). The statistical significance of the departure of the actual number of change of signs from the expected number is known by finding the “t” statistic as follows:

$$t = \frac{\left|N' - (N-1)p\right| - 0.5}{\sqrt{(N-1)\,pq}} \qquad (10.8)$$

where N’ denotes the actual number changes of sign.

If the critical value of “t” statistic, for (N – 1) degrees of freedom, is more than that computed above then it can be considered to be having adequate goodness of fit. Otherwise, the results will indicate that there is significant bias in the fitted curve with respect to long runs of positive or negative deviations.

11 EXTRAPOLATION OF RATING CURVE

11.1 GENERAL

Extrapolation of rating curves is required because the range of level over which gauging has been carried out does not cover the full range of observed levels. The rating curve may fall short at both the lower and the upper end. Extreme flows are often the most important for design and planning and it is important that the best possible estimates are made.

Calibration at very high instantaneous flows is particularly difficult as they occur infrequently and are of short duration. They may occur at night. Peak flow gauging requires the gauging team to be on site when the flood arrives - which may not be possible. It also requires that facilities are available for flood gauging in safety. In practice, the gauging site may be inaccessible, the gauging facilities no longer serviceable and the river may have spread from a confined channel to an inaccessible flood plain.

Extrapolation is not simply a question of extending the rating from existing gaugings to extreme levels (although in some cases this may be acceptable); a different control may apply, the channel geometry may change, flow may occur over the floodplain and form and vegetation roughness coefficients may change.

Applicable methods of extrapolation depend on the physical condition of the channel, whether inbank or overbank and whether it has fixed or shifting controls. Consideration must also be given to the phenomenon of the kinematic effect of open channel flow when there may be reduction in the mean velocity in the main channel during inundation of the flood plain. Methods given below are suitable for rivers with defined banks and fixed controls, as well as for a channel with spill.

Extrapolation of stage discharge relationships will be carried out at the State Data Processing Centre.

11.2 HIGH FLOW EXTRAPOLATION

The following methods are considered below:

- double log plot method
- stage-area / stage-velocity method
- the Manning’s equation method
- the conveyance slope method

11.2.1 THE DOUBLE LOG PLOT METHOD

Where the hydraulic characteristics of the channel do not change much beyond the measured range, then simple extrapolation of the logarithmic stage discharge relationship may be applied. Graphically, the relationship in this case can simply be extended beyond the measured range by projecting the last segment of the straight line relationship in log-log domain. Such an extrapolation is illustrated by the dashed straight line in Figure 11.2 for the cross-sectional profile shown in Figure 11.1.

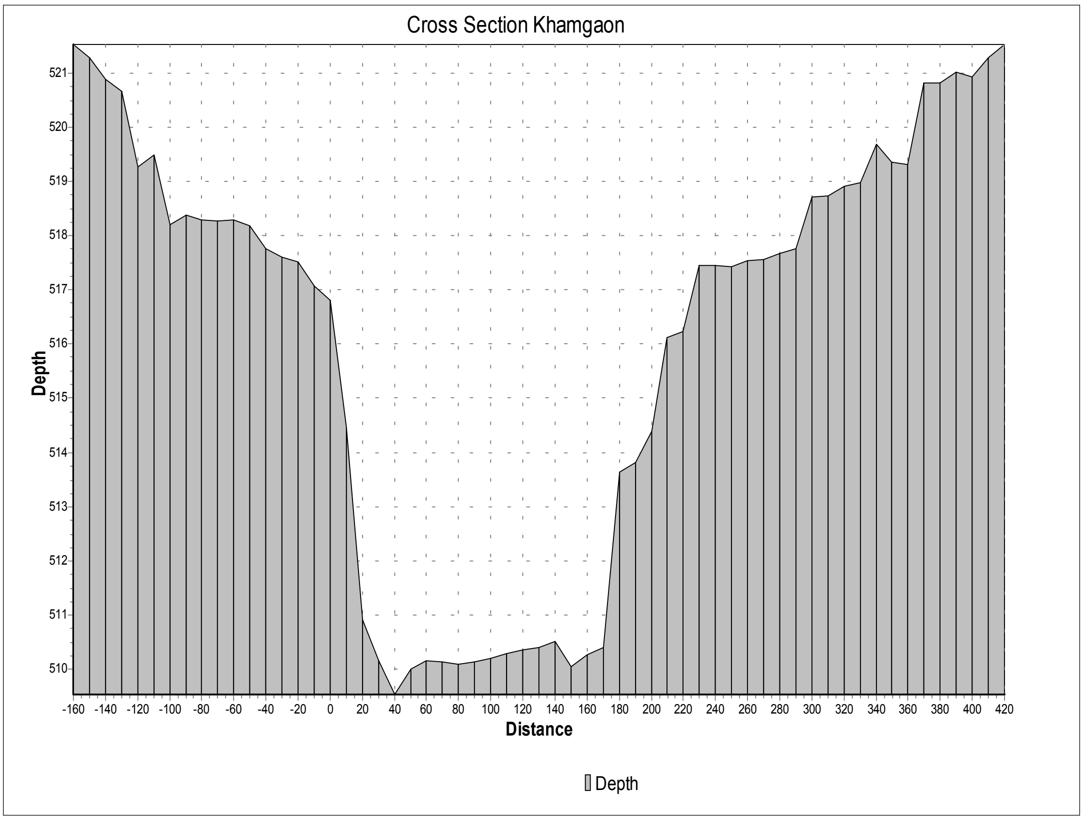


Figure 11.1: Cross-section of river at Khamgaon used in examples in this module

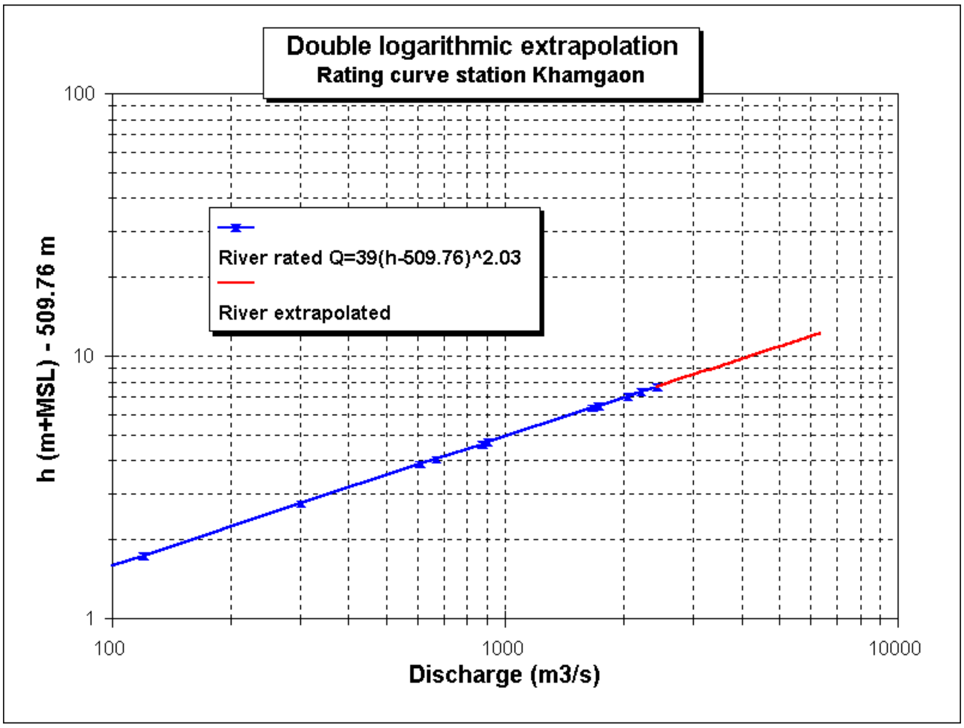


Figure 11.2: Example of double logarithmic extrapolation of rating curve

In the example presented in Figure 11.2 a rating curve has been established for the river flows up to flood plain level. This curve had to be is extended to cover the highest observed water level, which was about 4 m above flood plain level. Double logarithmic technique was applied for this extrapolation. Double-logarithmic extrapolation implies that the same power type equation is used for the higher stages as well. The correctness of the use of this technique for the cross-section shown in Figure 11.1, which shows the existance of a flood plain, is doubtful. One of the basic conditions for the application of the double logarithmic method, namely no change in the hydraulic characteristics at the higher stages, is not fulfilled. It is likely that this method will lead to an underestimation of the discharge, since the contribution of the floodplain flows to the total river flow is not taken into consideration.

11.2.2 STAGE-AREA / STAGE-VELOCITY METHOD

Where extrapolation is needed either well beyond the measured range, or there are known changes in the hydraulic characteristics of the control section, then a combination of stage-area and stage-velocity curves may be used. Stage-area and stage-mean velocity curves are extended separately. For stable channels the stage-area relationship is fixed and is determined by survey up to the highest required stage. The stage-velocity curve is based on current meter gaugings within the measured range and, since the rate of increase in velocity at higher stages diminishes rapidly this curve can be extended without much error for in-bank flows. Discharge for a given (extended) stage is then obtained by the product of area and mean velocity read using extrapolated stage-area and stage-mean velocity curves (Figure 11.3). This method may be used for extrapolation at both the upper and lower end of the rating.

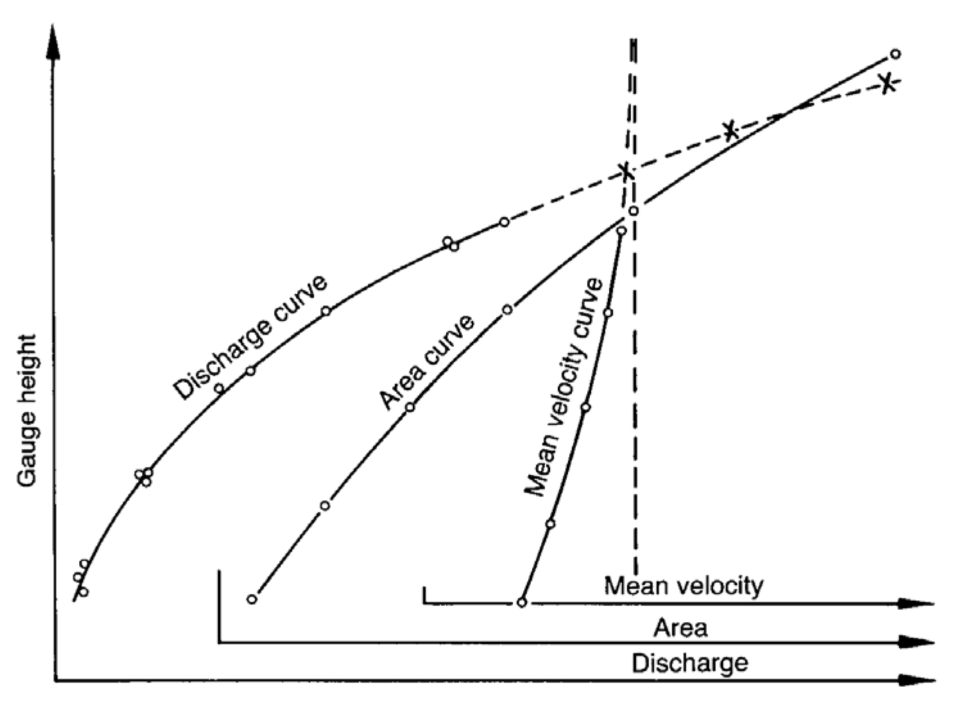


Figure 11.3: Extrapolation based on stage-area/stage-velocity technique

The mean velocity curve can also be extrapolated by the use of a logarithmic plot of mean velocity against hydraulic radius. The hydraulic radius can be found for all stages from the cross section by survey. The logarithmic plot of mean velocity and hydraulic radius generally shows a linear relationship and thus can be extended linearly beyond the extent of measurements. Mean velocity in the extrapolated range can be obtained from this curve. Extrapolated discharge as before is obtained as the product of mean velocity thus estimated and the corresponding area from the stage-area curve.

11.2.3 THE MANNING'S EQUATION METHOD

A slight variation of the stage-area-velocity method is the use of Manning's equation for steady flow. In terms of the mean velocity the Manning equation may be written:

$$v = K_m R^{2/3} S^{1/2} \tag{11.1}$$

Since for higher stages the value of $K_m S^{1/2}$ becomes nearly constant, the equation can be rewritten:

$$v = K^* R^{2/3} \tag{11.2}$$

or $$K^* = v / R^{2/3} \tag{11.3}$$

The relationship of stage (h) to K^* is plotted from discharge measurements. This curve often approaches a constant value of K^* at higher stages (Figure 11.4). This value of K^* may then be used in conjunction with extrapolated relationships between h and A and, h and $R^{2/3}$ based on survey. Discharge for extrapolated stage is then obtained by applying the Manning equation with K^* and extrapolated values of A and $R^{2/3}$

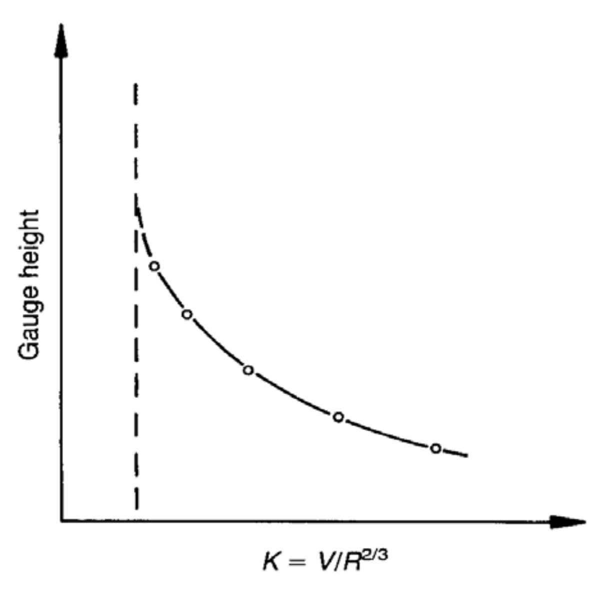
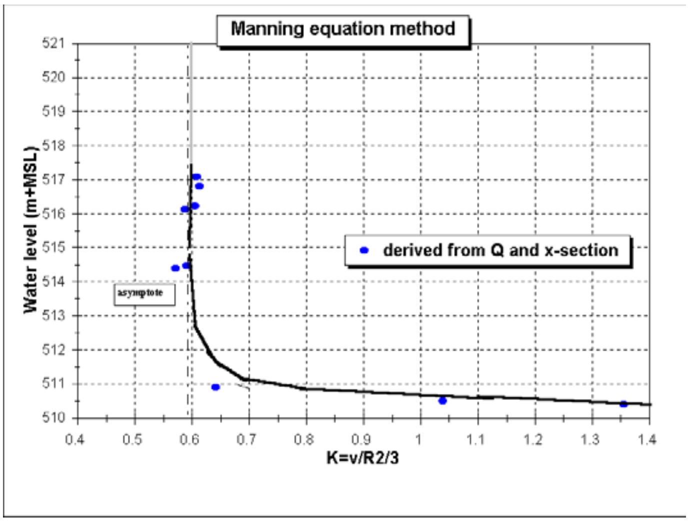


Figure 11.4:
K^* versus gauge height

Above bankfull stage the discharge on the floodplain must be determined separately by assuming an appropriate K_m value as done using the conveyance slope method.

This method was applied to the Khamgaon river cross-sectional data shown in Figure 11.1 and observed discharges. The results are shown in the Figures 11.5 to 11.7. From Figure 11.5 it is observed that K^* indeed tends to an approximately constant value for the higher stages, which was subsequently applied in the extrapolation. Together with the cross-sectional area and the hydraulic radius the river the flow through the main section was computed. For the flood plain the Manning equation was applied separately. The result is shown in Figure 11.7. In this Figure also the result of the double logarithmic extrapolation technique is shown for reference purposes. It is observed that the flow through the main river is approximately the same between the two methods, however the total flow with the Manning technique is larger since in this method due account is given to the flood plain flow.



Figures 11.5:
K versus gauge height for Kangaon example*

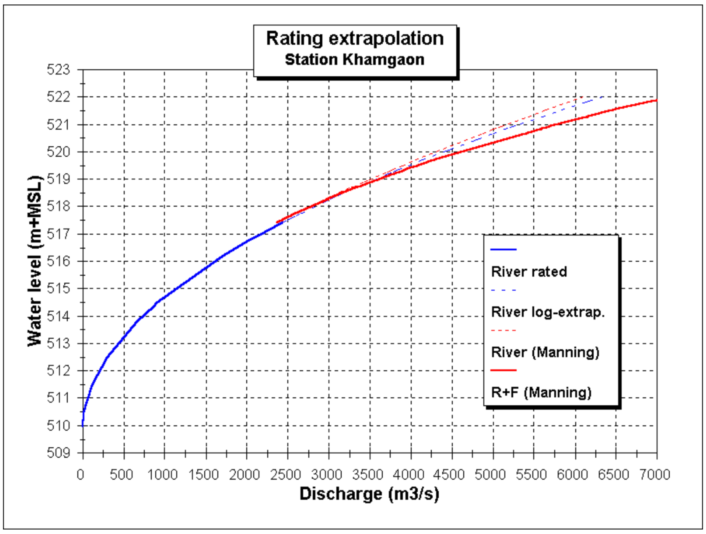


Figure 11.6:
Cross-sectional areas of river and flood plain in Khamgaon example

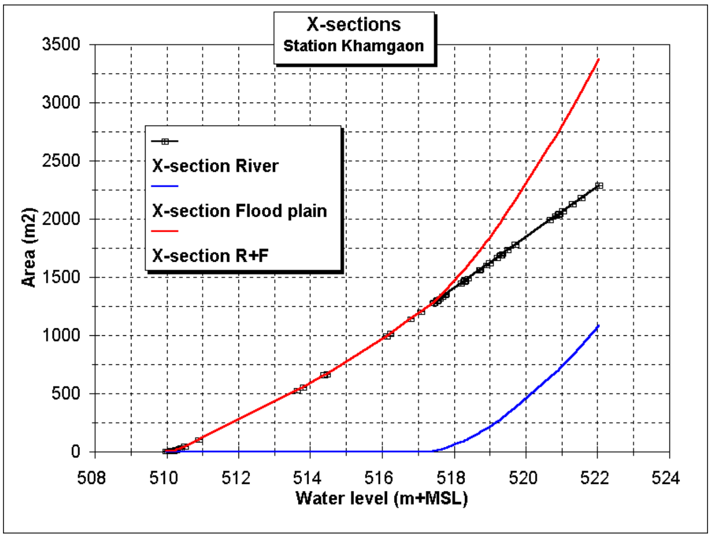


Figure 11.7:
Extrapolation based on Manning equation method compared with double-logarithmic extrapolation

11.2.4 THE CONVEYANCE SLOPE METHOD

In the conveyance slope method, the conveyance and the energy slope are extrapolated separately. It has greater versatility than the methods described above and can be applied in sections with overbank flow. It is therefore recommended for use. It is normally, again, based on the Manning equation:

$$Q = K_m R^{2/3} S^{1/2} A \tag{11.4}$$

or: $$Q = K S^{1/2} \tag{11.5}$$

where the conveyance is

$$K = K_m A R^{2/3} \tag{11.6}$$

For the assessment of *K* for given stage, *A* and *R* are obtained from field survey of the discharge measurement section and values of n are estimated in the field. Values of *K* are then plotted against stage up to the maximum required level (usually on natural graph paper) (Figure 11.8)

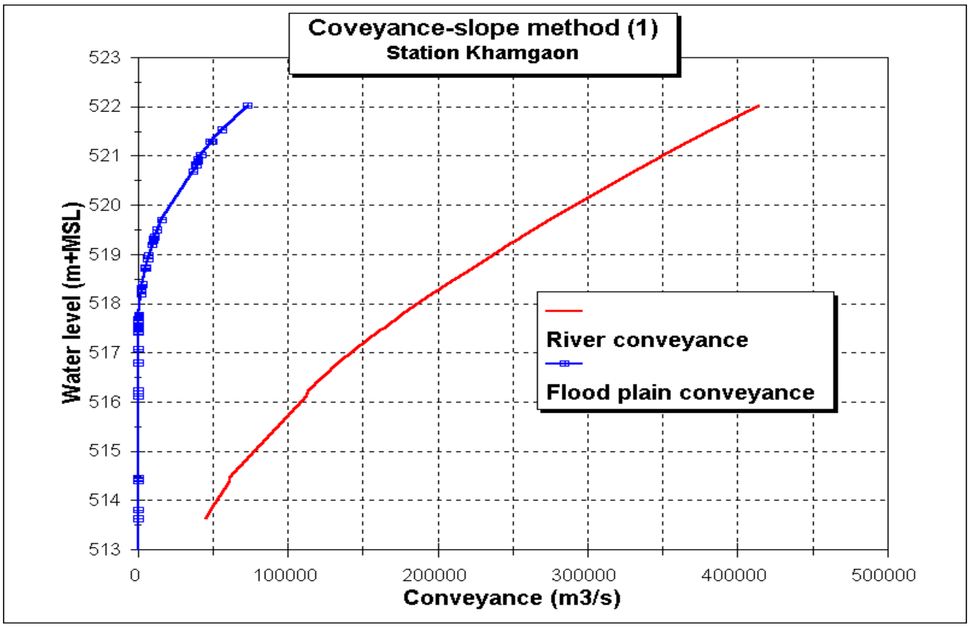


Figure 11.8: Conveyance as f(h)

Values of S, which is the energy gradient are usually not available but, for measured discharges, $S^{1/2}$ can be computed by dividing the measured discharge by its corresponding K value. S is then calculated and plotted against stage on natural graph paper and extrapolated to the required peak gauge height, in the knowledge that S tends to become constant at higher stages at the limiting slope of the stream-bed, (see Figure 11.9 for the Khamgaon case).

The discharge for given gauge height is obtained by multiplying the corresponding value of K from the K curve by the corresponding value of $S^{1/2}$ from the S curve. It should be noted that in this method, errors in estimating K_m have a minor effect, because the resulting percentage error in computing K is compensated by a similar percentage error in the opposite direction in computing $S^{1/2}$.

The whole procedure can be accomplished in various stages as given below:

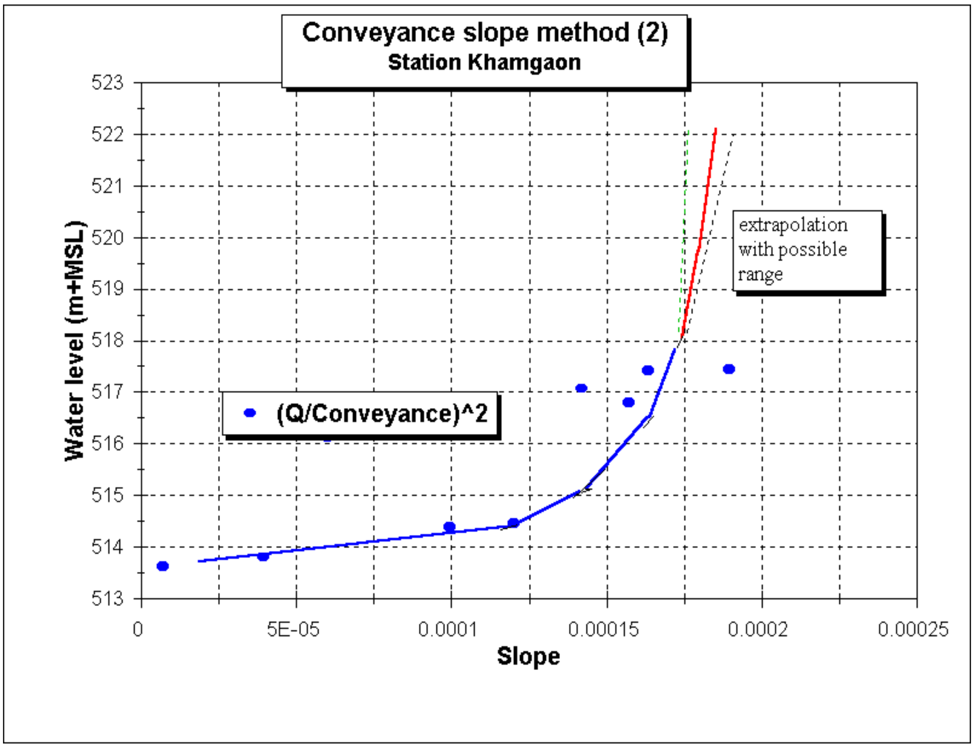


Figure 11.9: Slope extrapolation

Computation of cross-sectional data

First of all the cross-sectional contour is obtained from:

- distance (x) from an initial point
- depth (y) and
- depth correction (y_c)

The depth correction y_c may be introduced to evaluate quickly the effects of changes in the cross-section on geometric and hydraulic parameters.

The actual depth y_a is computed from:

$$y_a = y + y_c \tag{11.7}$$

A plot can be made of the cross section and for levels at fixed interval the following quantities are computed (see Table 11.1):

- surface width, (B)
- wetted perimeter, (P)
- cross-sectional area, (A)
- hydraulic radius, (R): R = A/P
- factor {area x (hydraulic radius)$^{2/3}$}, (AR$^{2/3}$)

```
Computation of cross-sectional data

     Computation of cross-sectional parameters

     Station    : KHAMGAON
     Date       : 1997 1  1

     Coordinates of profile

     Dist. from initial point       Level
                 m                      m
           -160.00                1521.54
           -150.00                521.30
           -140.00                520.90
           -130.00                520.68
               .                     .
            390.00                521.03
            400.00                520.93
            410.00                521.30
            420.00                522.03

       Section  1

     Left bound -160.00 m, right bound     .00 m from initial point
     Water boundaries included

     K-Manning: 30.0, Sqrt(S).K-Manning:     .4135

     Stage      Width     Wetted       Area   Hydraulic  A*R**2/3      Q
                          Perimeter              radius
         m          m          m       sq-m          m   m**(8/3)     m3/s

    512.00       .000       .000       .000       .000       .00       .00
    513.00       .000       .000       .000       .000       .00       .00
    514.00       .000       .000       .000       .000       .00       .00
    515.00       .000       .000       .000       .000       .00       .00
    516.00       .000       .000       .000       .000       .00       .00
    517.00      7.143      8.738       .715       .082       .13       .06
    518.00     45.568     47.233     25.858       .547     17.30      7.16
    519.00    106.139    107.976    113.077      1.047    116.61     48.22
    520.00    125.143    127.195    230.063      1.809    341.53    141.23
    521.00    142.500    144.755    360.636      2.491    662.77    274.07

     Section  2

     Left bound     .00 m, right bound  230.00 m from initial point
     Water boundaries included

     K-Manning: 40.0, Sqrt(S).K-Manning:     .5514

     Stage      Width     Wetted       Area   Hydraulic  A*R**2/3      Q
                          Perimeter              radius
         m          m          m       sq-m          m   m**(8/3)     m3/s

    512.00    158.042    158.482    277.671      1.752    403.55    222.50
    513.00    163.946    164.716    438.664      2.663    842.81    464.69
    514.00    181.982    182.984    608.462      3.325   1355.54    747.39
    515.00    195.834    196.996    798.552      4.054   2030.17   1119.36
    516.00    205.888    207.247    999.412      4.822   2852.65   1572.84
    517.00    226.337    227.860   1218.693      5.348   3727.16   2055.02
    518.00    230.000    232.731   1447.878      6.221   4897.61   2700.35
    519.00    230.000    234.731   1677.878      7.148   6226.16   3432.87
    520.00    230.000    236.731   1907.878      8.059   7669.16   4228.48
    521.00    230.000    238.731   2137.882      8.955   9219.35   5083.20

     Section  3

     Left bound  230.00 m, right bound  420.00 m from initial point
```

```
Water boundaries included

     K-Manning: 30.0, Sqrt(S).K-Manning:     .4135

     Stage      Width     Wetted       Area    Hydraulic  A*R**2/3      Q
                          Perimeter              radius
         m          m          m       sq-m          m   m**(8/3)     m3/s

    512.00       .000       .000       .000       .000        .00       .00
    513.00       .000       .000       .000       .000        .00       .00
    514.00       .000       .000       .000       .000        .00       .00
    515.00       .000       .000       .000       .000        .00       .00
    516.00       .000       .000       .000       .000        .00       .00
    517.00       .000       .000       .000       .000        .00       .00
    518.00     62.393     63.274     27.705       .438      15.98      6.61
    519.00    100.209    101.231     99.981       .988      99.16     41.00
    520.00    134.521    135.689    219.425      1.617     302.31    125.01
    521.00    167.389    168.803    359.942      2.132     596.30    246.58

    Stage   Discharge/section  >>> Total Discharge
   512.00       .00   222.50      .00   222.50
   513.00       .00   464.69      .00   464.69
   514.00       .00   747.39      .00   747.39
   515.00       .00  1119.36      .00  1119.36
   516.00       .00  1572.84      .00  1572.84
   517.00       .06  2055.02      .00  2055.07
   518.00      7.16  2700.35     6.61  2714.12
   519.00     48.22  3432.87    41.00  3522.09
   520.00    141.23  4228.48   125.01  4494.72
   521.00    274.07  5083.20   246.58  5603.85
```

Table 11.1: Example of HYMOS Report on stage-discharge extrapolation

These parameters may be determined for the whole cross-section or for parts of cross-section, e.g. for main river and flood plain separately.

It must be noted that when the cross-section is divided, the wetted perimeter for each part may be determined in two ways:

- the water boundary not considered:
- for flood plain : $P_{floodplain}$ = ABC
- for the main river: P_{river} = CEFG
- the water boundary is treated as a wall:
- for the flood plain: $P_{floodplain}$ = ABCD
- for the river: P_{river} = DCEFG

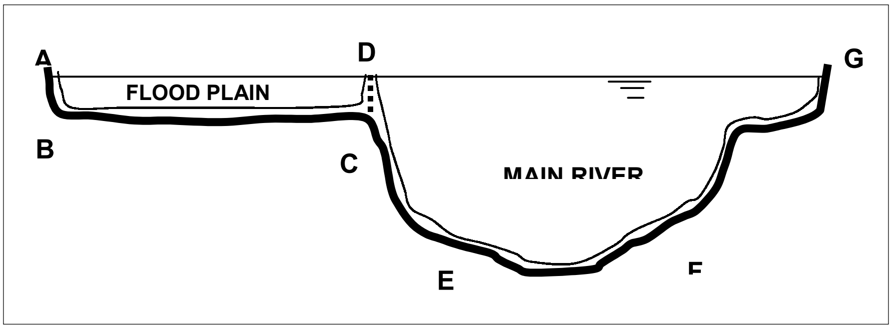


Figure 11.10: River flow in main channel and flood plain during high stages

To account for the lateral transport of momentum between river and flood plain the latter option appears to be more realistic. It reduces generally the discharge capacity of the main channel. However, to obtain consistency with hydraulic computations, where generally the first approach is used, both options are included.

Computation of hydraulic quantities in the measured range

Next, the geometric and hydraulic quantities are obtained in one of the following ways:

- from stage-discharge database, provided that the cross-sectional parameters are also given
- from a combination of the cross-section profile and the rating curve

The following parameters are obtained for various depths:

- surface width, (B)
- wetted perimeter, (P)
- cross-sectional area, (A)
- hydraulic radius, (R): R = A/P
- factor {area x (hydraulic radius)$^{2/3}$}, (AR$^{2/3}$)
- discharge, (Q)
- average velocity. (v): v = Q/A
- Conveyance, (K): K = (1/n)(AR$^{2/3}$) ; where n is an estimated value
- Slope (S): S = (v/K)^2

Estimation of discharge in the extrapolated range

The estimated values of slope (S) in the measured range is plotted against stages and since this curve is asymptotic to the bed slope at higher stages extrapolation is accordingly. The conveyance curve is also plotted making use of the estimated values of K in the full range of stages. Now, for any stage in the extrapolated range the value of K and S are read from the two curves and the product of these two quantities and the area of cross section (A) yield the estimated value of discharge (Q).

After synthetic data stage-discharge data have been obtained for the extrapolated range, these data are incorporated in the set of stage-discharge data. Subsequently, new attempts can be made to fit rating equation to the measured and estimated stage-discharge data.

11.3 LOW FLOW EXTRAPOLATION

Manual low flow extrapolation is best performed on natural graph paper rather than on logarithmic graph paper because the co-ordinates of zero flow can not be plotted on such paper. An eye-guided curve is drawn between the lowest point of the known rating to the known point of zero flow, obtained by observation or by survey of the low point of the control. There is no assurance that the extrapolation is precise but improvement can only come from further low flow discharge measurements. However low flows persist for a sufficient period for gaugings to be carried out and there is little physical difficulty in obtaining such measurements.

In HYMOS the power type equation used for the lowest segment with stage-discharge observations is assumed applicable also below the measured range. The shift parameter ‘a’ is either determined based on the measurements by the system or is introduced based on cross-sectional information.

12 SECONDARY VALIDATION OF STAGE DISCHARGE DATA

12.1 GENERAL

Rating curves are normally developed and validated with respect to current metering observations at an individual station. It is often necessary to extrapolate the relationship beyond the measured range.

One means of providing a further check on the reliability of the extrapolated rating curve is to make comparisons of discharges computed using the stage discharge relationships between neighbouring stations. The secondary validation of discharge, as will be described in Chapter 14 Part 3, thus also provides a basis for secondary validation of stage discharge relationships. If there is an inconsistency or an abrupt change in the relationship between discharge time series at sequential stations on a river or around a confluence, then the most likely source is the stage discharge relationship at one or more of the compared stations. Where such inconsistencies are observed, rating curves and their extrapolations must be reviewed.

12.2 REVIEW OF RATING CURVE ON THE BASIS OF BALANCES

After finalising rating curves, observed stage time series are converted to discharge. Discharge time series are then aggregated and compiled to successively longer time intervals - from hourly to daily to ten-daily and monthly. Discharge time series at consecutive stations on a river should then show a consistent pattern of relationship and water balance, taking into consideration the intervening catchment area, major tributary inflows and abstractions. The balance of flows can be checked using the average discharge or flow volumes during a time interval. Generally better and less variable relationships are obtained using longer time intervals. Comparison plots of discharge time series provide a helpful means of identifying anomalies.

In addition a residual series can be plotted (Figures 12.1a - h) alongside the comparison plots as the difference between discharges at the two stations.

Residual series generally provide a better means of detecting anomalies. Where inconsistencies occur, the station at fault may not be immediately evident. A potential source, which should be investigated, are periods when rating curve extrapolation has been used at one or both stations.

In the Figures 12.1a to 12.1h an application of the technique is outlined. In Figure .1a the hourly water level time series of the stations Chaskman and Khed in the Bhima basin are show. Both stations are located along the same river, Khed d/s of Chaskman. The rating equations fitted to the stage-discharge data available for 1997 are shown in Figures 12.1b and c. Next, the hourly water level time series have been transformed into hourly discharge time series using the rating curves presented in Figures 12.1b and c. The results are shown in Figures 12.1d and e, where the latter is a detail. Since Chaskman is upstream of Khed, and lateral inflow may occur, one should expect that the discharge at Khed exceeds the discharge at Chaskman. From the comparison of the two series it is observed that this is generally the case, except for a short duration prior to the second peak. The differences are far better exposed if the difference between the series are plotted, see Figure 12.1f. It is noted that particularly with sharp rises of the hydrograph and little inflow in between the stations the peak at the upstream station advances the downstream one, hence creating negative values in the balance, which is apparent from the first peak. Large positive values as is observed for the second peak is likely due to lateral inflow (provided that timing errors in the water level hydrograph do not exist).

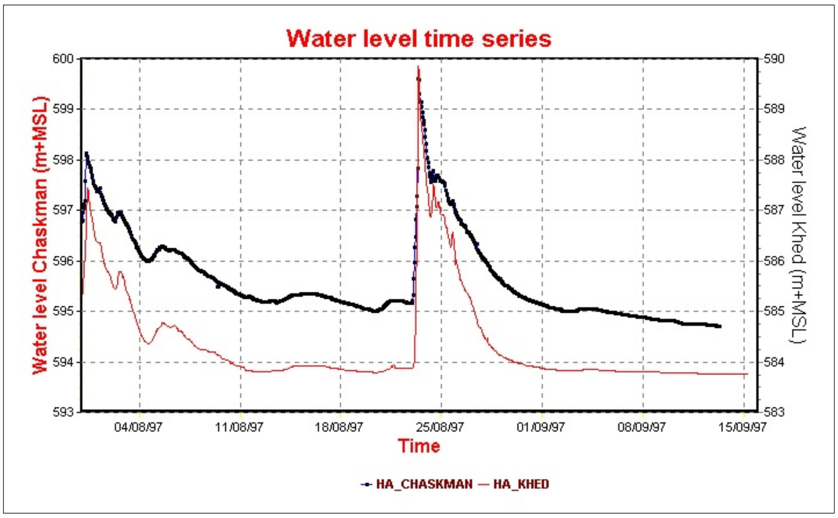


Figure 12.1a:
Hourly water level time series

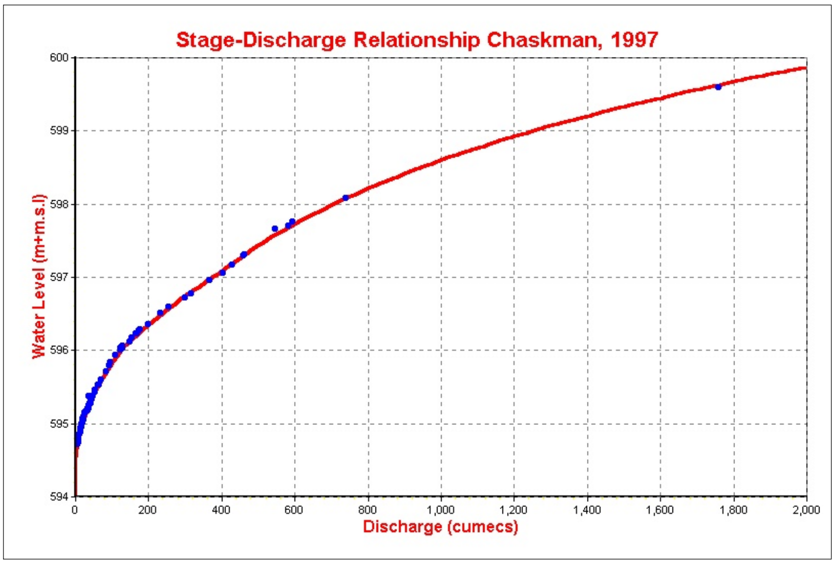


Figure 2.1b:
Stage-discharge rating curve Chaskman, 1997

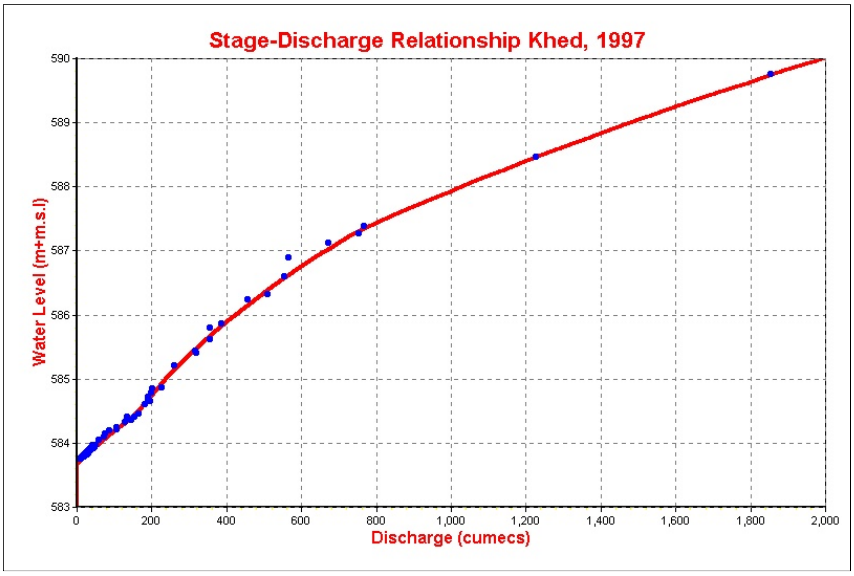


Figure 12.1c:
Stage-discharge rating curve Khed, 1997

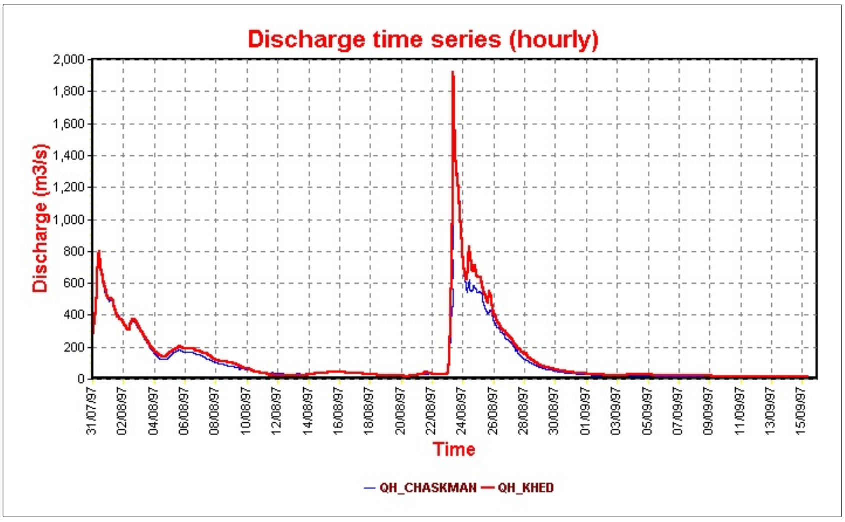


Figure 12.1d:
Hourly discharge time series Chaskman and Khed

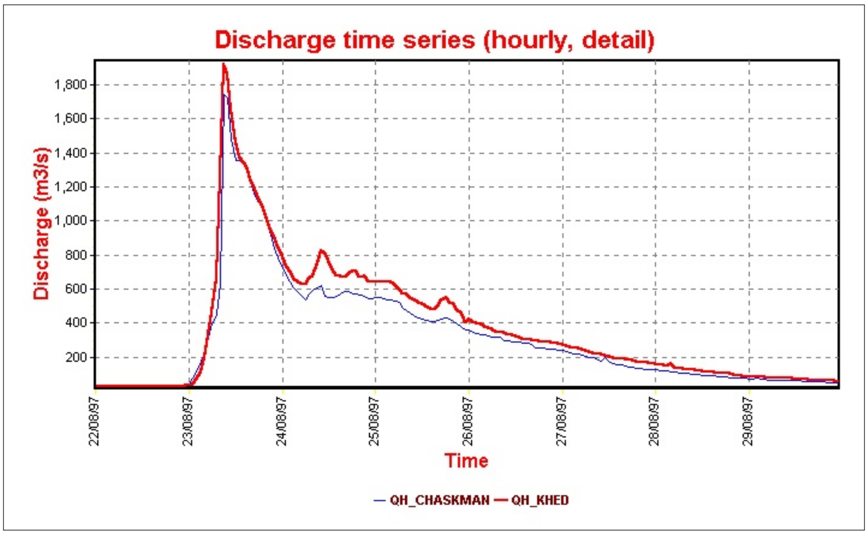


Figure 12.1e:
Hourly discharge time series Chaskman and Khed (detail)

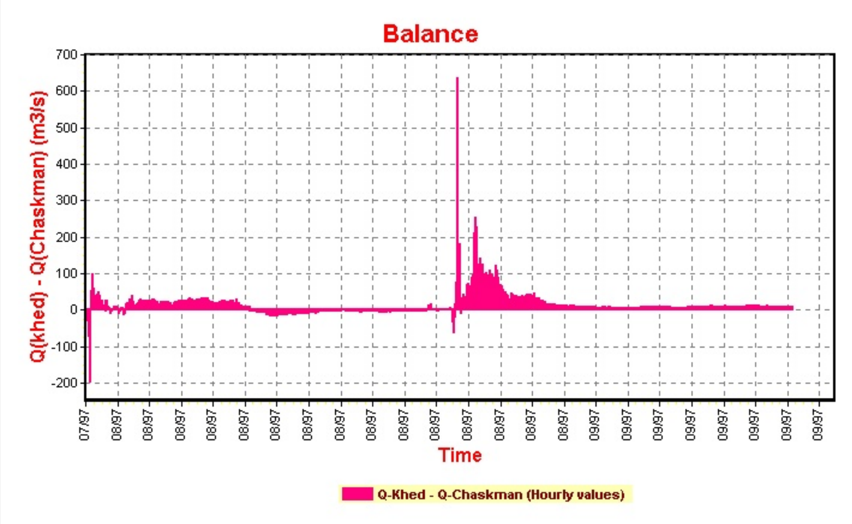


Figure 12.1f:
Q(Khed) – Q(Chaskman) hourly discharge series

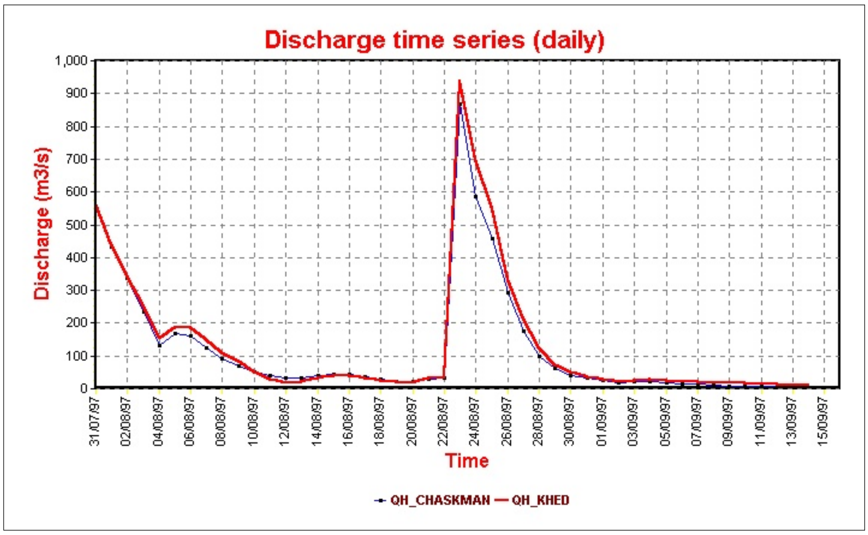


Figure 12.1g:
Daily discharge time series Chaskman and Khed

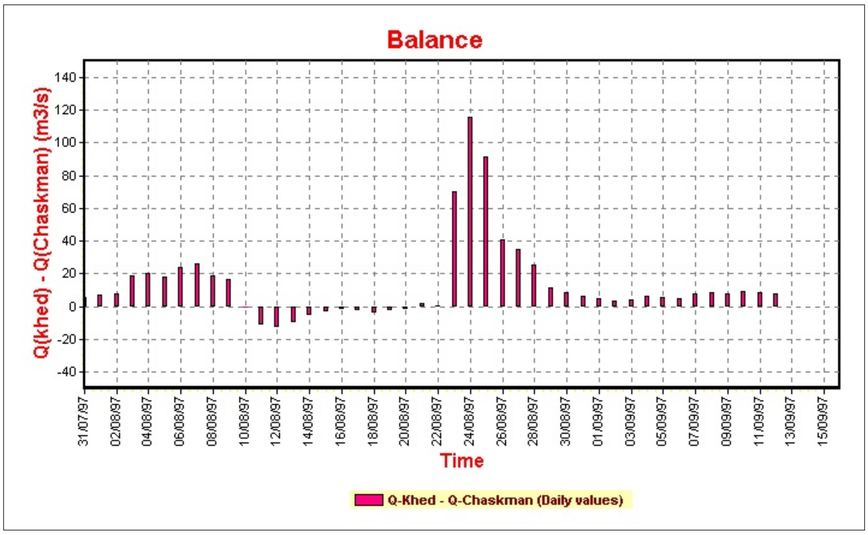


Figure 12.1h:
Q(Khed) – Q(Chaskman) daily discharge series

The eliminate these travel time problems in the water balance the water balance for the discharge time series should be executed at a higher aggregation level. In view of the travel time of the flood wave between Chaskman and Khed of approximately 1 hour an aggregation up to daily intervals will do. For larger rivers with larger distances between the stations a higher aggregation level may be more appropriate. The daily discharge time series and the water balance are shown in Figures 12.1g and h. It is observed that between 10 and 20 August 1997 the daily discharges as computed for Chaskman are exceeding those of Khed. Provided that no water is being abstracted from the river, the reasons could be:

- Either the water series at one or at both sites are erroneous, or
- The rating curves established for one or both sites are biased for that period, or
- Both water level series and rating curves are incorrect.

These possibilities then have to be investigated in detail. If the anomaly is due to one or both rating curves, more segments have to be incorporated in the rating curves or the rating curves for shorter periods of time have to be developed.

It is noted here that for the peaks some negative values in the water balance may occur due to damping of the flood wave. The damping per unit length is approximately:

$$\frac{dQ_{max}}{dx} \approx \frac{(3/5)^3}{2}\frac{(B_s/B_r)^2}{K_m^2 h_{max}^{1/3} S_0^2}\frac{\partial^2 Q}{\partial t^2} \tag{12.1}$$

where : B_s = total width of river and flood plain

B_r = width of the main river

K_m = K-Manning (=1/n)

h_{max} = flow depth for the flood peak

S_0 = slope of the river bed

To get an impression of the magnitude assume that the shape of the flood wave can be approximated by the following cosinus function:

$$Q(t) = Q_0 + a_0\left(1 - \cos\left(\frac{2\pi t}{T}\right)\right) \quad \text{hence :} \quad \left(\frac{\partial^2 Q}{\partial t^2}\right)_{Q\,max} = \frac{-a_0 4\pi^2}{T^2} \tag{12.2}$$

With the above, the damping per unit of length becomes:

$$\frac{dQ_{max}}{dx} \approx -\frac{(3/5)^3 4\pi^2}{2}\frac{(B_s/B_r)^2}{K_m^2 h_{max}^{1/3} S_0^2}\frac{a_0}{T^2} \tag{12.3}$$

The equation shows that the damping is large if:

- B_s/B_r is large, i.e if the a wide flood plain is present
- K_m is small, i.e. Manning's n is large that is a hydraulically rough river bed
- S_0 is small; in steep rivers the attenuation is small
- The amplitude a_0 of the wave is large and its period T (duration) is short, i.e. rapidly rising and falling hydrographs.

Using this for the flood peak in Figure 12.1e with:

- $Q_0 = 400\ m^3/s$, $a_0 = 750\ m^3/s$ and T = 36 hrs
- $B_s/B_r = 1$ (no flood plain), $K_m = 40$, $h_{max} = 5$ m and $S_0 = 8x10^{-4}$

Then $dQ_{max}/dx = 1.1x10^{-4}\ m^3/s/m$ and the damping over a distance of 11 km is approximately 1.2 m^3/s, which is negligible. For a bed slope of 10^{-4}, the damping would have been 64 times as large, whereas a flood plain would have increased the damping further.

12.3 REVIEW OF RATING CURVE ON THE BASIS OF DOUBLE MASS ANALYSIS

Double mass curve analysis has already been described in the secondary validation of rainfall (Chapter 2) and climate (Chapter 5). It can also be used to show trends or inhomogeneities between (a) flow records at neighbouring stations or (b) observed flow records and flows computed on the basis of regression relationships with rainfall and is normally used with aggregated series (usually monthly). It can again be used to identify potential problems in the rating curve of one or more stations.

A distinct break of slope in the double mass curve between neighbouring stations suggests inhomogeneity in one of the records. Inspection for rating changes at the time of the break of slope will help to identify the source. It should be noted however that inhomogeneities also arise from artificial changes in the catchment, for example the commencement of abstraction for an irrigation scheme.

12.4 REVIEW OF RATING CURVE ON THE BASIS OF RELATION CURVES BETWEEN STAGES AT ADJACENT STATIONS

Relationship between stages at adjoining stations for steady state conditions can be established. At both such stations relation between stages and corresponding discharges would have been also established. It can then be possible to combine these three relationships together in the following way. Consider the stations Cheskman and Khed, which are two adjoining stations on a river reach. Figure 12.2 shows the relationship between stages h-Chaskman and h-Khed. Figures 12.1b and c show the rating curves for stations Chaskman and Khed (i.e. relation between discharge $Q_{Chaskman}$ & $h_{Chaskman}$ and Q_{Khed} & h_{Khed}) respectively.

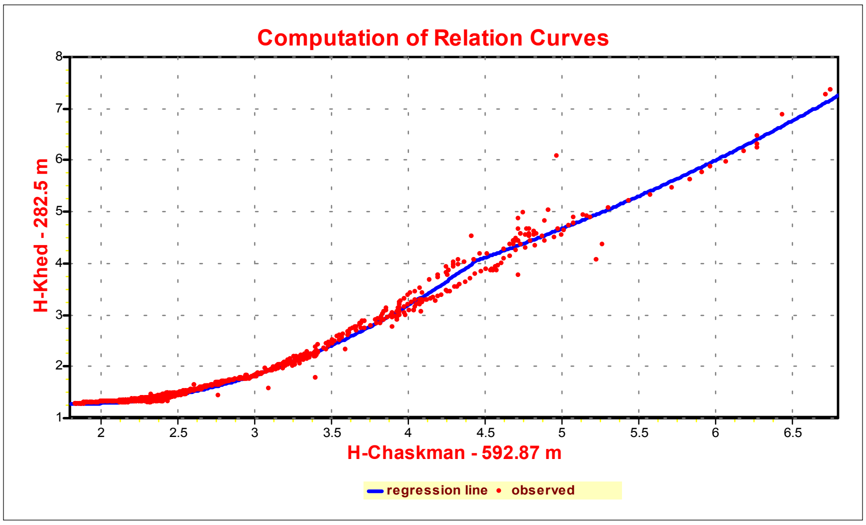


Figure 12.2: Stage-relation curve h(Khed) = f(h(Chaskman))

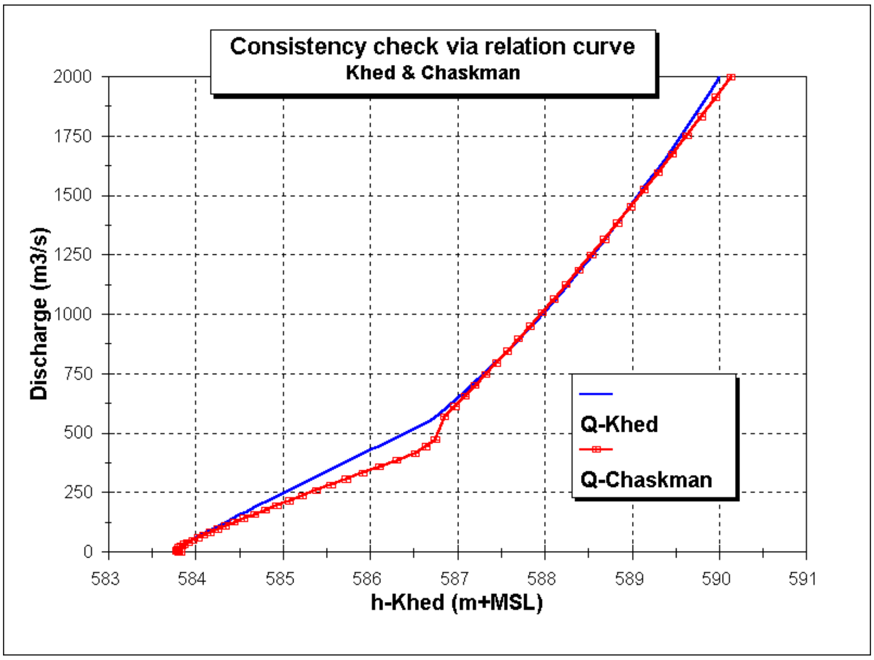


Figure 12.3: Discharge at Chaksman and Khed displayed versus h-Khed

Now, using Figure 12.2 and Figure12.1b the relationship between $Q_{Chaskman}$ & $h_{Chaskman}$ can be translated into a relationship between $Q_{Chaskman}$ & h_{Khed}. The relationship between Q_{Khed} & h_{Khed} is also superimposed on the same plot depicting relationship between $Q_{Chaskman}$ & h_{Khed}. These superposed relationships are depicted in Figure 12.3.

Now a comparison can be made for $Q_{Chaskman}$ & Q_{Khed} for the same h_{Khed}. The range of discharge variation for the same stage in this plot depends upon whether the condition of flow is steady state or if it unsteady with or without lateral contributions from the intervening catchment between the two stations. It is observed from Figure 12.3 that for the same stage at Khed the flow in Chaskman is less than at Khed, which confirms the consistency of thye two curves.

13 COMPUTATION OF DISCHARGE DATA

13.1 GENERAL

With limited exceptions, discharge cannot be measured both directly and continuously. Instead measurements of stage (or level) are made continuously or at specified intervals at a gauging station and these are converted to discharge by the use of stage discharge relationships.

Computation of discharge normally be carried out monthly on the stage data from the previous month but will always be reviewed annually before transferring to the archive.

Computation of discharge will be carried out at Divisional offices and reviewed at the State Data Management Centre.

13.2 STATION REVIEW

Before computing discharge, it is essential to have available a summary of all the relevant information for the station, including:

- the stage record - to ensure that it is complete and without abrupt discontinuities.
- a listing of stage discharge relationships to check that periods of application do not overlap or have gaps between ratings.
- Ancillary information based on field records (Field Record Book) or on information from validation of stage or stage discharge relationships. In particular field information on datum changes, scour and deposition, blockage and backwater effects should be assembled along with any adjustments or corrections applied during validation.

13.3 TRANSFORMATION OF STAGE TO DISCHARGE

The procedure used to transform stage to discharge depends on physical conditions at the station and in the river reach downstream The following alternatives are considered:

- single channel rating curve
- compound channel rating curve
- rating curves with unsteady flow correction
- rating curves with constant fall backwater correction
- rating curves with normal fall backwater correction

13.3.1 SINGLE CHANNEL RATING CURVE

When unsteady flow and backwater effects are negligibly small the stage discharge data are fitted by a single channel relationship, valid for a given time period and water level range. Rating equations will have previously been derived either as parabolic or power law equations; it is assumed that in the vast majority of cases the recommended power law relationship will have been applied. Equations for standard and non-standard gauging structures may also be re-computed in this form with little loss of accuracy.

The basic equations are as follows:

(a) For the power type equation used for curve fitting.

$$Q_t = c_{1,i}\,(h_t + a_{1,i})^{b_{1,i}} \tag{13.1}$$

(b) For the parabolic type of equation used for curve fitting

$$Q_t = a_{2,i} + b_{2,i}\,h_t + c_{2,t}\,h_t^2 \tag{13.2}$$

where: Q_t = discharge at time t (m^3/sec)

h_t = measured water level at time t(m)

a_1, b_1, c_1 = parameters of the power equation

a_2, b_2, c_2 = parameters of the parabolic equation

i = index for water level interval for which the parameters are valid ($1 \le i \le 4$)

The parabolic form of rating equation is however not recommended for use while establishing the rating curves.

HYMOS permits a maximum of 5 equations applicable over different level ranges in a single stage discharge relationship; normally three will be sufficient. One equation may require to be used for the situation where water is ponded at the gauge and a non-zero level has been measured but the flow is zero. In this case an equation may be introduced for the range from h = 0.0 to h at minimum flow, taking the power law form with c = 0.0, a = 0.0 and b = 1.0

13.3.2 COMPOUND CHANNEL RATING CURVE

The compound channel rating curve is used to avoid large values of the parameter b and very low values of the c-parameter in the power equation at levels where the river begins to spill over from its channel to the floodplain.

When a compound channel rating has been applied, the discharge for the flood plain interval will be computed by adding the discharge computed for the river section up to the maximum flood plain level using the parameters for the one but last interval, and the discharge computed for the flood plain section for the last interval. That is:

$$Q_{tot} = Q_r + Q_{fp} \tag{13.3}$$

where: Q_{tot} = total discharge

Q_r = discharge flowing through the main river section up to the maximum water level

Q_{fp} = discharge flowing through the flood plain section.

13.3.3 RATING CURVE WITH UNSTEADY FLOW CORRECTION

Where an unsteady flow correction is required, the application of the simple rating curve first yields a discharge for steady flow which must then be multiplied by the unsteady flow correction to give the discharge for the required rate of change of water level. The stage-discharge transformation used for this case is:

$$Q_t = Q_{st}\sqrt{\left(1+\frac{1}{S_0\,c}\frac{dh_t}{dt}\right)} \qquad (13.4)$$

where: Q_t = the required discharge corresponding to the observed stage h_t and rate of change of stage (dh_t/dt)

Q_{st} = the steady state discharge as obtained from the available steady state rating curve.

The expression ($1/S_0c$) is expressed in the form of parabolic equation as:

$$\frac{1}{S_0 c} = a_3 + b_3\,h_t + c_3 h_t^2 \qquad \text{and} \quad h_t > h_{min} \qquad (13.5)$$

where: a_3, b_3, c_3 = parameters of the equation

h_{min} = the lowest water level below which the correction is not to be applied.

The parameters of the above parabolic equation and that of the steady state equation are available from the established rating curve.

The rate of change of stage with respect to time (dh_t/dt) at time t can be obtained from the stage time series as:

$$\frac{dh_t}{dt} = \frac{h_{t+1} - h_{t-1}}{2\Delta t} \qquad (13.6)$$

where: Δt = time interval between two successive observations. If h_{t+1} or h_{t-1} does not exist, its value is replaced by h_t and the denominator by Δt.

13.3.4 RATING CURVE WITH CONSTANT FALL BACKWATER CORRECTION

Where the station is affected by backwater and the rating curve with constant fall type of backwater correction has been established for it then the stage-discharge transformation is carried out using the following equation:

$$Q_t = Q_{rt}\left(\frac{F_t}{F_r}\right)^p \qquad (13.7)$$

where: Q_{rt} = reference discharge computed using the established rating curve with h_t replaced with h_{1t}

F_r and p = reference fall and exponent in the equation and are available as parameters of the established rating curve

F_t = $h_{1t} - h_{2t}$

= is the measured fall between the stages at the station under consideration (h_{1t}) and the reference station (h_{2t}). The stages used for calculating the fall have to be precisely synchronised in time.

13.3.5 RATING CURVE WITH NORMAL FALL BACKWATER CORRECTION

Where the station is affected by backwater and the rating curve with normal fall type of backwater correction has been established for it, then the stage-discharge transformation is carried out using the equation:

$$Q_t = Q_{rt} \left(\frac{F_t}{F_{rt}} \right)^p \tag{13.8}$$

where: Q_{rt} = backwater free discharge computed using established rating curve with h_t replaced with h_{1t}

p = is the exponent in the above equation and is available as the parameter of the established rating curve

F_t = $h_{1t} - h_{2t}$

= is the measured fall between the stages at the station under consideration (h_{1t}) and the reference station (h_{2t}). The stages used for calculating the fall have to be precisely synchronised in time.

The reference fall, F_{rt} in this case is expressed as:

$$F_{rt} = a_4 + b_4\, h_t + c_4\, h_t^2 \quad \text{and} \quad h_t > h_{min} \tag{13.9}$$

The parameters a_4, b_4 and c_4 are available from the established rating curve and the reference fall is evaluated for any

14 SECONDARY VALIDATION OF DISCHARGE DATA

14.1 GENERAL

Secondary validation of discharge will be carried out at Divisional offices on completion of transformation of stage to discharge at the same office. The discharge series will contain flagged values whose accuracy is suspect (based on assessment of stage), corrected, or missing. These will require to be reviewed, corrected, or inserted.

The quality and reliability of a discharge series depends primarily on the quality of the stage measurements and the stage discharge relationship from which it has been derived. In spite of their validation, errors may still occur which show up in discharge validation. Validation flags which have been inserted in the validation of the stage record are transferred through to the discharge time series. These include the data quality flags of ‘good’, ‘doubtful’ and ‘poor’ and the origin flags of ‘original’, ‘corrected’ and ‘completed’. This transfer of flags is necessary so that stage values recognised as doubtful or poor can be corrected as discharge.

Discharge errors may also arise from the use of the wrong stage discharge relationship, causing discontinuities in the discharge series, or in the use of the wrong stage series.

Validation of discharge is designed to identify such problems. The principal emphasis is in the comparison of the time series with neighbouring stations but preliminary validation of a single series is also carried out against data limits and expected hydrological behaviour.

Validation using regression analysis and hydrological modelling is specifically excluded from this module. They are considered separately in later modules.

14.2 SINGLE STATION VALIDATION

Single station validation will be carried out by the inspection of the data in tabular and graphical form. The displays will illustrate the status of the data with respect to quality and origin, which may have been inserted at the stage validation stage or identified at discharge validation. Validation provides a means of identifying errors and, following investigation, for correcting and completing the series.

14.2.1 VALIDATION AGAINST DATA LIMITS

Data will be checked numerically against, absolute boundaries, relative boundaries and acceptable rates of change, and individual values in the time series will be flagged for inspection.

- Absolute boundaries:

  Values may be flagged which exceed a maximum specified by the user or fall below a specified minimum. The specified values may be the absolute values of the historic series. The object is to screen out spurious extremes, but care must be taken not to remove or correct true extreme values as these may be the most important values in the series.

- Relative boundaries:

  A larger number of values may be flagged by specifying boundaries in relation to departures (α and β) from the mean of the series (Q_{mean}) by some multiple of the standard deviation (s_x), i.e.

  $$\text{Upper boundary} \quad Q_u = Q_{mean} + \alpha\, s_x$$

  $$\text{Lower boundary} \quad Q_l = Q_{mean} - \beta\, s_x \qquad (14.1)$$

  Whilst Q_{mean} and s_x are computed by the program, the multipliers α and β are inserted by the user with values appropriate to the river basin being validated. The object is to set limits which will screen a manageable number of outliers for inspection whilst giving reasonable confidence that all suspect values are flagged. This test is normally only used with respect to aggregated data of a month or greater.

- Rates of change.

  Values will be flagged where the difference between successive observations exceeds a value specified by the user. The specified value will be greater for large basins in arid zones than for small basins in humid zones. Acceptable rates of rise and fall may be specified separately, generally allowable rates of rise will be greater than allowable rates of fall.

For looking at the possible inconsistencies, it is very convenient if a listing of only those data points which are beyond certain boundaries is obtained. An example is given in Table 14.1

```
Series KHED_QH  less than 0 or greater than 1,500
 Month 8

                              ==========Data==========
   Year month   day  hour sub.i          Value

   1997     8    23     9     1          1918. +
   1997     8    23    10     1          1865. +
   1997     8    23    11     1          1658. +
```

Table 14.1: Example of listing of data exceeding given limits

14.2.2 GRAPHICAL VALIDATION

- Graphical inspection of the plot of a time series provides a very rapid and effective technique for detecting anomalies. Such graphical inspection will be the most widely applied validation procedure and will be carried out for all discharge data sets.
- The discharge may be displayed alone or with the associated stage measurement (Figure 14.1). Note that in this example the plot covers 2 months to reveal any discontinuities which may appear between successive monthly updates updates of the data series.
- The discharge plots may be displayed in the observed units or the values may be log-transformed where the data cover several orders of magnitude. this enables values near the maximum and minimum to be displayed with the same level of precision. Log-transformation is also a useful means of identifying anomalies in dry season recessions. Whereas the exponential decay of flow based on releases from natural storage are curved in natural units, they show as straight lines in log-transformed data.

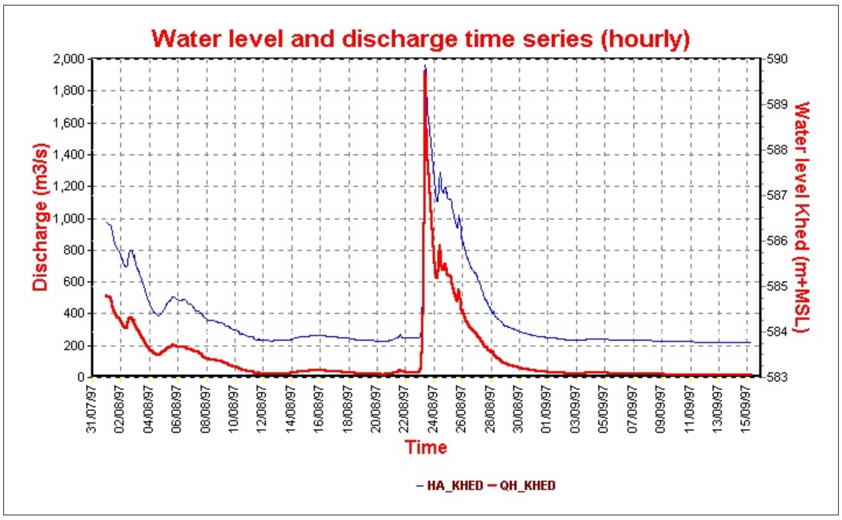


Figure 14.1:
Q(t) and h(t) of station Khed for consecutive months

- The graphical displays will also show the absolute and relative limits. The plots provide a better guide than tabulations to the likely reliability of such observations.
- The main purpose of graphical inspection is to identify any abrupt discontinuities in the data or the existence of positive or negative ‘spikes’ which do not conform with expected hydrological behaviour. It is very convenient to apply this test graphically wherein the rate of change of flow together with the flow values are plotted against the expected limits of rate of rise and fall in the flows. Examples are:
  - Use of the wrong stage discharge relationship (Figure 14.2). Note that in this example, discharge has been plotted at a logarithmic scale
  - Use of incorrect units (Figure 14.3)
  - Abrupt discontinuity in a recession (Figure 14.4).
  - Isolated highs and lows of unknown source (Figure 14.5) but may be due to recorder malfunction with respect to stage readings.

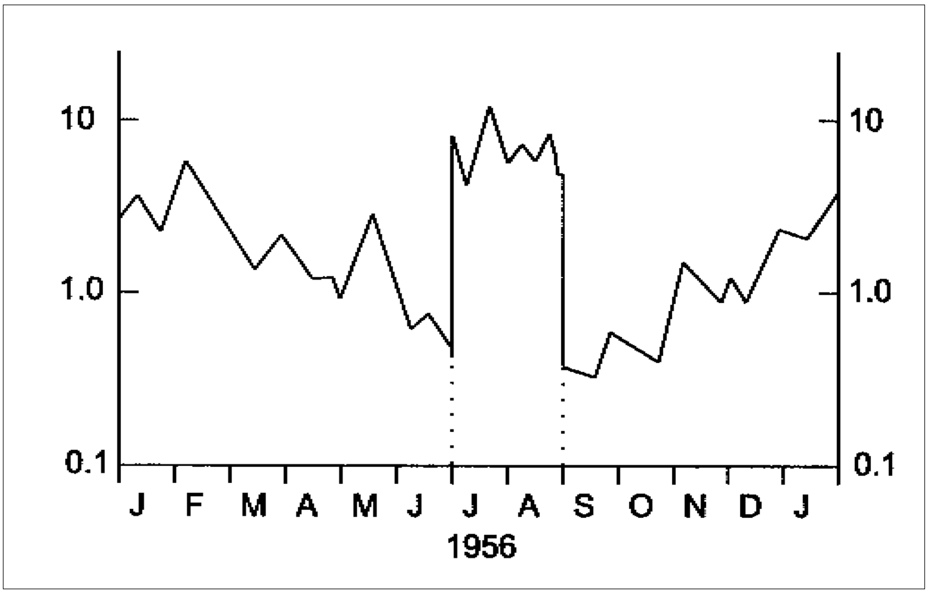


Figure 14.2: Use of incorrect rating for part of the year

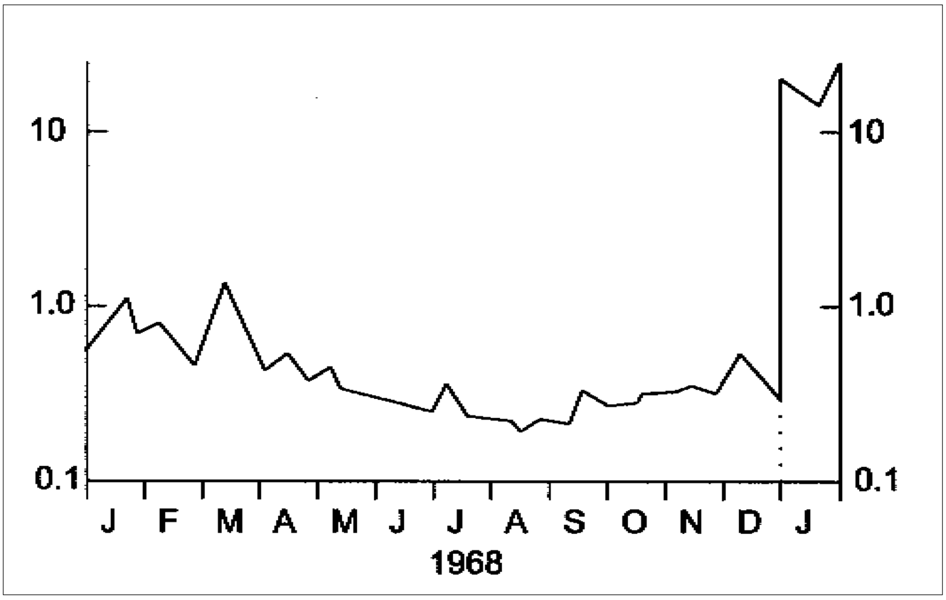


Figure 14.3: Use of incorrect units

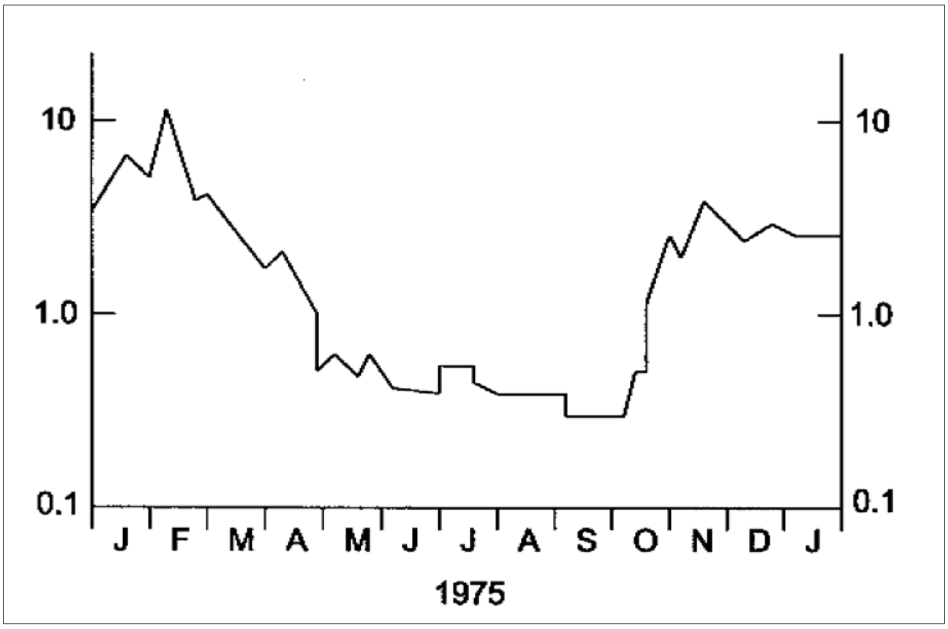


Figure 14.4: Unrealistic recession

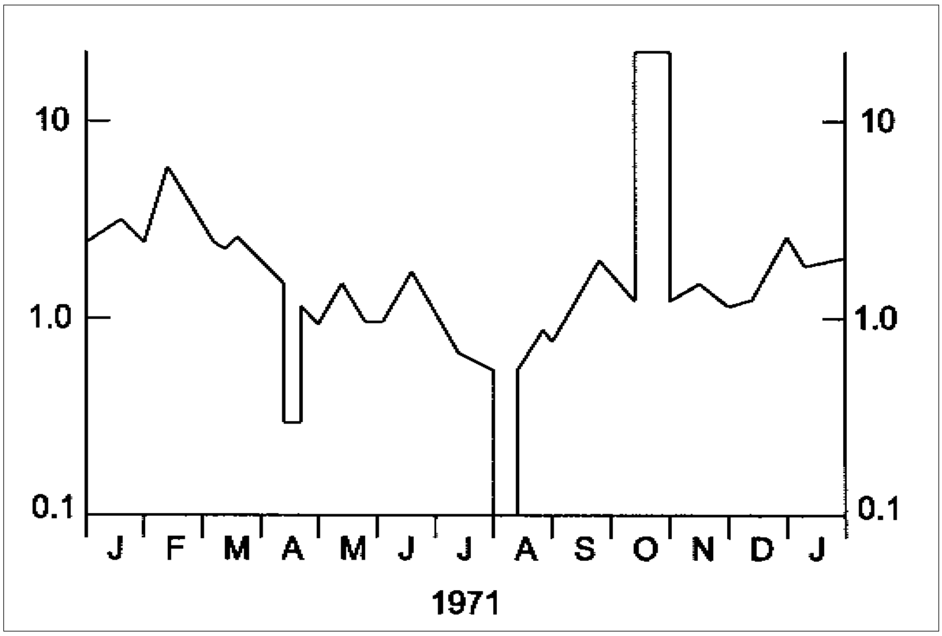


Figure 14.5: Isolated 'highs' and 'lows'

14.2.3 VALIDATION OF REGULATED RIVERS

The problems of validating regulated rivers has already been mentioned with respect to stage data and should also be borne in mind in validating discharge data. Natural rivers are not common in India; they are influenced artificially to a greater or lesser extent. The natural pattern is disrupted by reservoir releases which may have abrupt onset and termination, combined with multiple abstractions and return flows. The influences are most clearly seen in low to medium flows where in some rivers the hydrograph appears entirely artificial; high flows may still observe a natural pattern. Officers performing validation should be aware of the principal artificial influences within the basin, the location of those influences, their magnitude, their frequency and seasonal timing, to provide a better basis for identifying values or sequences of values which are suspect.

14.3 MULTIPLE STATION VALIDATION

14.3.1 COMPARISON PLOTS

The simplest and often the most helpful means of identifying anomalies between stations is in the plotting of comparative time series. HYMOS permits the plotting of multiple for a given period in one graph. There will of course be differences in the plots depending on the contributing catchment area, differing rainfall over the basins and differing response to rainfall. However, gross differences between plots can be identified.

The most helpful comparisons are between sequential stations on the same river. The series may be shifted relative to each other with respect to time to take into account the different lag times from rainfall to runoff or the wave travel time in a channel.

In examining current data, the plot should include the time series of at least the previous month to ensure that there are no discontinuities between one batch of data received from the station and the next - a possible indication that the wrong data have been allocated to the station.

Comparison of series may permit the acceptance of values flagged as suspect because they fell outside the warning ranges, when viewed as stage or when validated as a single station. When two or more stations display the same behaviour there is strong evidence to suggest that the values are correct, e.g. an extreme flood peak.

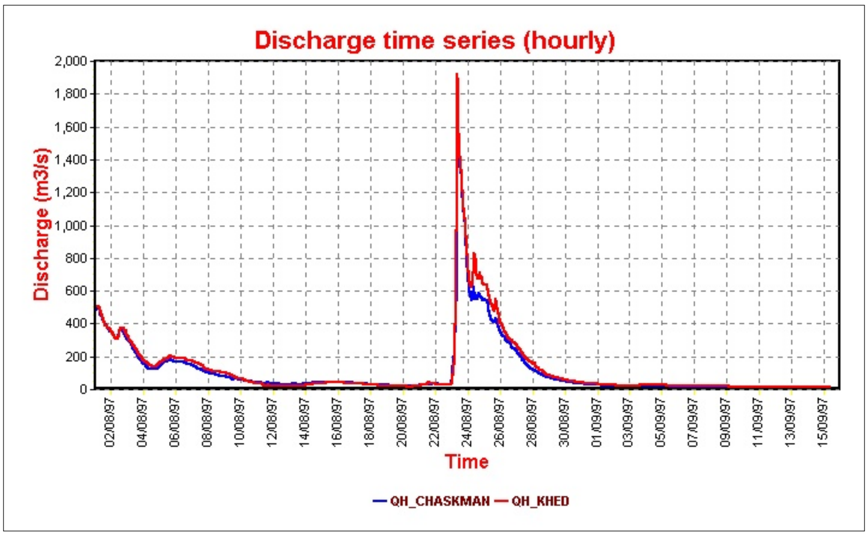


Figure 14.6:
Plot of multiple discharge series of adjacent stations

Comparison plots provide a simple means of identifying anomalies but not necessarily of correcting them. This may best be done through regression analysis, double mass analysis or hydrological modelling.

14.3.2 RESIDUAL SERIES

An alternative way of displaying comparative time series is to plot their differences. This procedure may be applied to river flows along a channel to detect anomalies in the water balance. HYMOS provides a means of displaying residual series under the option ‘Balance’. Both the original time series and their residuals can be plotted in the same figure.

Water balances are made of discharge series of successive stations along a river or of stations around a junction, where there should be a surplus, balance or deficit depending on whether water is added or lost. The basic equation is expressed as:

$$Y_i = \pm a.X_{1,i} \pm b.X_{2,i} \pm c.X_{3,i} \pm d.S_{4,I} \qquad (14.2)$$

where: a, b, c, d = multipliers entered by the user (default = 1)

± = sign entered by user (default = +)

A maximum of four series is permitted. An example for the series presented in Figure 14.6 is shown in Figure 14.7 where the comparison is simply between two stations, upstream and a downstream. Reference is also made to Chapter.10 Any anomalous behaviour should be further investigated. Sharp negative peaks may be eliminated from the plot by applying the appropriate time shift between the stations or to carry out the analysis at a higher aggregation level.

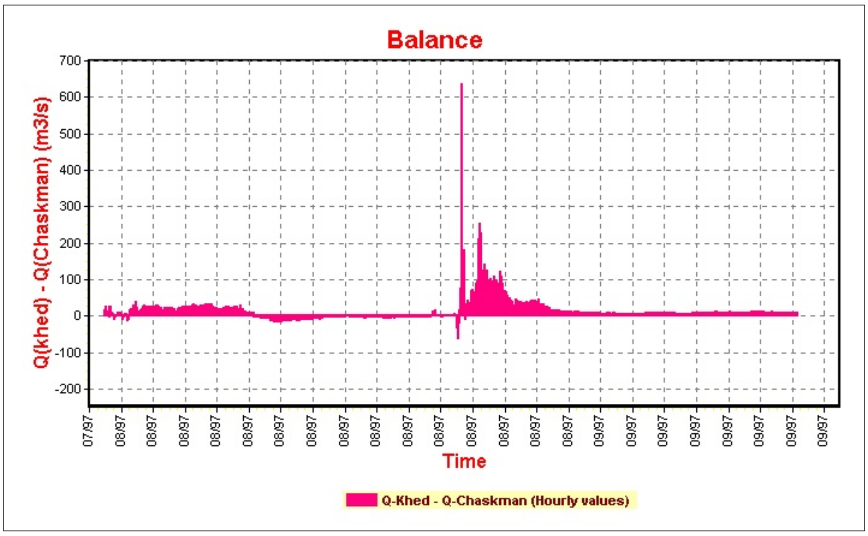


Figure 14.7: Example of water balance between two adjacent stations

14.3.3 DOUBLE MASS CURVES

Double mass curve analysis has already been described in the secondary validation of rainfall (Chapter 2) and climate (Chapter 5). It can also be used to show trends or inhomogeneities between flow records at neighbouring stations and is normally used with aggregated series.

A difficulty in double mass curves with streamflow is in the identification of which if any station is at fault; this may require intercomparisons of several neighbouring stations. There may also be a legitimate physical reason for the inhomogeneity, for example, the construction of a major irrigation abstraction above one of them. In the latter case no correction should be applied unless one is attempting to ‘naturalise’ the flow. (Naturalisation is the process of estimating the flow that would have occurred if one, several or all abstractions, releases or land use changes had not occurred).

14.4 COMPARISON OF STREAMFLOW AND RAINFALL

The principal comparison of streamflow and rainfall is done through hydrological modelling. However, a quick insight into the consistency of the data can be made by graphical and tabular comparison of areal rainfall and runoff. The computation of areal rainfall is described in Chapter 4; it will be realised that areal rainfall is also subject to error which depends upon the density of stations within the basin and the spatial variability of rainfall. Basically the basin rainfall over an extended period such as a month or year should exceed the runoff (in mm) over the same period by the amount of evaporation and changes in storage in soil and groundwater. Tabular comparisons should be consistent with such physical changes. For example an excess of runoff over rainfall either on an annual basis or for monthly periods during the monsoon will be considered suspect.

Graphical comparison on a shorter time scale can be made by plotting rainfall and streamflow on the same axis. In general the occurrence of rainfall and its timing should be followed by the occurrence of runoff separated by a time lag but precise correspondence should not be expected owing to the imperfect assessment of areal rainfall and to the variable proportion of rainfall that enters storage.

15 CORRECTION AND COMPLETION OF DISCHARGE DATA

15.1 CORRECTION OF DISCHARGE DATA

Erroneous discharge data, which have been transformed from water levels using a stage discharge relation, can be corrected in various ways. Reference is made to Section 8.2 where an overview is given of the various options. Consult this Section 8.2 for making a proper assessment as to which option is most appropriate. Basically two procedures apply:

- Either the discharge is corrected by correction or completion of the water level data and/or the stage-discharge relation at your station, and application of the stage-discharge transformation,
- Or the discharge is corrected by means of water balance considerations and estimates of lateral inflow and/or regression analysis using nearby stations.

The first option, when feasible, is to be preferred in order to establish an internally consistent database. Reference is made to the previous chapters on procedures to correct or fill in the water levels at your station and/or correct the stage-discharge relationship.

When the latter option has to be applied, i.e. that no use is made of water levels, make sure that clear notes are made on the procedure used for these corrections, and why the first option has not been used.

It is noted that when use is made of the discharge computed for (an) upstream station(s) and the lateral inflow in between the stations due account is to be given to the travel time of the flow in between the stations, so as to properly route the upstream and lateral inflow to your discharge station. Often, translation of the contributions is sufficient. If, however, large flood plain storage is present in between the sites a Muskingum routing procedure, which accounts for damping, can be applied. This procedure is outlined in Chapter 3 of Part III of this Volume 8. The procedure can conveniently be carried out by the option in the Sacramento model in HYMOS, which allows the application of a layered Muskingum approach.

15.2 COMPLETION OF DISCHARGE DATA

Gaps in the discharge record may be due to absence of water level data and/or a stage-discharge relationship. These gaps may be filled as follows:

1. In case water levels are missing and a stage- discharge relation exists: filling in of the missing water level by making use of a relation curve with water level data at an upstream or downstream site and subsequently transformation of these levels to discharge using the existing stage-discharge relation. This only applies when the lateral inflow between the sites in the relationship is comparatively small. One has to be sure that the stage-discharge relation is still applicable. This will be so if the control section is stable.
2. In case water levels are missing and a stage-discharge relation is also missing: filling in of the missing water level by making use of a relation curve with water level data at an upstream or downstream site. Creation of a stage-discharge relation by means of the Manning equation and making use of river cross-sections in the control section of the station. Subsequently transformation of these levels to discharge using the estimated stage-discharge relation. This only applies when the lateral inflow between the sites in the relationship is comparatively small.
3. In case water levels are available and a stage-discharge relation is missing: creation of a stage-discharge relation by means of the Manning equation and making use of river cross-sections in the control section of the station. Subsequently transformation of these levels to discharge using the estimated stage-discharge relation.
4. In case water levels are missing and a stage-discharge relation is also missing and no appropriate stage relation curves can be developed to reliably estimate the water level at the station:
    - Apply a water balance approach to estimate the discharge from flow records of upstream or downstream stations, adjusted for lateral inflow and due account of routing effects
    - Apply the Sacramento model, to estimate streamflow from observed rainfall and potential evapotranspiration. To make such a procedure acceptable, a record of concurrent rainfall and runoff data has to be available to calibrate and verify the model.
    - If only monthly discharge series have to be completed a regression approach may be used on rainfall in the same (and one or more previous) months, where the regression coefficients vary monthly or seasonally and/or with the rainfall depth.

16 COMPILATION OF DISCHARGE DATA

16.1 GENERAL

Discharge compilation is the process by which discharge at its observational or recorded time interval and units is transformed:

- to another time interval
- from one unit of measurement and especially from discharge (a rate of flow) to volume or runoff (a depth over the catchment)

Computations for aggregation of data from one time interval to another depends on the data type. If the data is of instantaneous nature then the aggregation is effected by computing the arithmetic average of the individual constituent data values. Whereas when the data type is of accumulative nature then the constituents values are arithmetically summed up for obtaining the aggregated value.

Compilation is required for validation, analysis and reporting

Compilation is carried out at Divisional offices; it is done prior to validation if required, but final compilation is carried out after correction and 'completion'.

16.2 AGGREGATION OF DATA TO LONGER DURATION

Discharge and its precursor water level is observed at different time intervals, but these are generally one day or less. Manual observation may be daily, hourly for part of the day during selected seasons, or some other multiple of an hour. For automatic water level recorders a continuous trace is produced from which hourly level and hence discharge is extracted. For digital water level recorders level is usually recorded at hourly intervals though for some small basins the selected interval may be 15 or 30 minutes. Sub-hourly, hourly and sub-daily discharges, computed from levels, are typically aggregated to daily mean. For example, the daily mean discharge (Q_d) is computed from hourly values (Q_i) by:

$$Q_d = \frac{1}{24} \sum_{i=1}^{24} Q_i \tag{16.1}$$

For a given day the mean is normally calculated for hours commencing 0100 and finishing 2400. For some purposes daily discharge averages are calculated over the day from 0800 to 0800 (i.e. for hourly measurements the average of observations from 0900 to 0800) to enable direct comparison to be made with daily rainfall.

Daily data are typically averaged over weekly, ten daily, 15 daily, monthly, seasonally or yearly time intervals. In general,

$$Q_{Nd} = \frac{1}{Nd} \sum_{i=1}^{Nd} Q_i \tag{16.2}$$

where: Q_{Nd} = discharge for Nd days duration,

Q_i = discharge of ith day in duration of Nd days.

Time intervals used while aggregating the data generally corresponds to the month or year ending. For example a ten daily data series corresponds to three parts of every month in which the first two parts are the 1-10 and 11-20 days of the month and the third part is the remaining part of the month. Thus every third value in the series corresponds to 8,9,10 or 11 days (the last part of the month) depending on the total days in the month. Similarly, weekly data depending on its objective is the taken in two ways: (a) as four parts of the months where first three parts are of seven days each and the fourth part is of 7, 8, 9 or 10 days period (as per the total days in the month) or (b) as 52 parts of a year where first 51 weeks are of 7 days each and the last week is of 8 or 9 days depending upon whether the year is a leap or a non-leap year. Such culmination are often desirable for the operational purpose as the time interval is reset to the 1st of a month or year every time.

Averaging over longer time intervals is required for validation and analysis. For validation small persistent errors may not be detected at the small time interval of observation but may more readily be detected at longer time intervals.

16.3 COMPUTATION OF VOLUMES AND RUNOFF DEPTH

To facilitate comparisons between rainfall and runoff it is usual to express values of rainfall and flow in similar terms. Both may be expressed as a total volume over a specified period (in m^3, thousand m^3, (Tcm) or million m^3 (Mcm)). Alternatively, discharge may be expressed as a depth in millimetres over the catchment.

Volume is simply the rate in m^3/sec (cumecs) multiplied by the duration of the specified period in secs., i.e. for daily volumes in cubic metres with respect to daily mean flow Q_d in cumecs following equation may be used:

$$V_d\ (m^3) = (24 \times 60 \times 60 \text{ seconds})\ Q_d\ (\text{cumecs}) = 86400\ Q_d\ (m^3) \qquad (16.3)$$

Runoff depth is the volume expressed as depth over the specified catchment area with a constant to adjust units to millimetres; i.e. for daily runoff:

$$R_d(\text{mm}) = \frac{V_d\ (m^3) \times 10^3}{\text{Area}\ (km^2) \times 10^6} = \frac{V_d\ (m^3)}{\text{Area}\ (km^2) \times 10^3} = \frac{86.4\ Q_d}{\text{Area}\ (km^2)} \qquad (16.4)$$

Runoff depths provide not only a ready comparison with rainfalls; they also provide a comparison with other catchments standardised by area. Such comparions may be made for monthly, seasonal and annual totals but are not generally helpful for daily or shorter duration depths, where basins respond at different time scales to incident rainfall.

For the purposes of annual reporting it is usual to compare the monthly and annual runoff from a station with the long term average, maximum and minimum monthly runoff derived from the previous record. This requires the annual updating of runoff statistics with the concatenation of the previous year with earlier statistics

Volumes and runoff depths may also be required for irregular periods to compare with rainfall depths over a storm period. Providing sufficient measurements are available over the period, the runoff over the storm period can be expressed simply as:

$$R(\text{mm}) = \frac{0.001}{\text{Area}} \sum_{i-1}^{N} (Q_i \Delta t) \qquad (16.5)$$

where: N = Number of observations in the period

Δt = Time step in seconds

A = Catchment area in km^2

Q_i = Discharge at time i in m^3 / sec

It is not generally necessary to use more complex procedures such as Simpson's rule, to account for the non-linear variation of flow between observations.

For the purposes of storm analysis, it is also generally necessary to separate the storm flow, resulting from the incident rainfall, and the base flow originating from continuing groundwater sources. Various methods have been suggested for such separation; they are described in standard texts and not discussed further here.

Another unit which is sometimes used to standardise with respect to area is specific discharge which may be computed with respect to instantaneous discharges or the mean discharge over any specified duration as discharge over area (m^3/sec per km^2).

Imperial and other units are regarded as **obsolete** and should not be used; these include Mgd (million gallons per day), acre-feet and ft^3 /sec (cusecs).

16.4 COMPILATION OF MAXIMUM AND MINIMUM SERIES

The annual, seasonal or monthly maximum series of discharge is frequently required for flood analysis, whilst minimum series may be required for drought analysis. Options are available for the extraction of maximum and minimum values for the following time periods:

- day
- month
- year
- period within a year

For example if the selected time period is 'month' and the time interval of the series to be analysed is 'day', then the minimum and maximum daily value is extracted for each month between a start and end date.

17 SECONDARY VALIDATION OF SUSPENDED SEDIMENT CONCENTRATIONS

17.1 GENERAL

The secondary validation of suspended sediment concentrations includes the following steps:

1. Single fraction, multiple season year by year
    - Comparison of the concentration of the coarse fraction for various seasons or months
    - Comparison of the concentration of the medium fraction for various seasons or months
    - Comparison of the concentration of the fine fraction for various seasons or months
2. Single year, multiple fractions
    - Comparison of the concentration of the coarse and medium fractions
    - Comparison of the concentration of the C+M and fine fractions
3. Multiple years, single fractions
    - Comparison of the concentration of C+M fractions
    - Comparison of the concentration of the fine fractions
    - Comparison of the total suspended sediment concentrations.

These steps with examples are elaborated in the next sections.

17.2 SINGLE FRACTION, MULTIPLE SEASONS, SINGLE YEAR

The objective of the validation of a single fraction, multiple season in a particular year is to detect anomalies. Generally, for coarse and medium sized fractions there should be a distinct relation with the flow velocity and hence with the discharge in the main stream. It implies that the relationship is expected not differ much from season to season, unless bank erosion or river bed mining does takes place just upstream of the measuring location. For the fine fraction a seasonal dependency is expected, in a manner that the concentrations will be highest at the start of the wet season, due to the supply from the catchment. The behaviour of the various fractions is shown in the following figures.

Coarse fraction

For the coarse fraction linear and semi-logarithmic scales are applied of concentration of coarse suspended solids versus discharge. Different symbols are applied for each month. From Figure 17.1, which shows the concentrations for a test site on a linear scale with the river discharge, it is observed that the relation does not change from month to month.

It is also observed that for the highest discharges the concentrations seem to reduce. This may be due to measurement errors by sampling with a bottle sampler when the velocities are high. Another reason could be that a considerable part of the discharge is conveyed by the flood plain. It is the flow velocity in the main stream that brings the bed material in suspension. Hence, the discharge through the main stream, rather than the total discharge should be compared with the sediment concentration.

Another check to be carried out is the occurrence of zero concentrations. There should be a distinct discharge range for zero concentrations irrespective of the season. This can best be observed from a semi-logarithmic plot as shown in Figure 17.2.

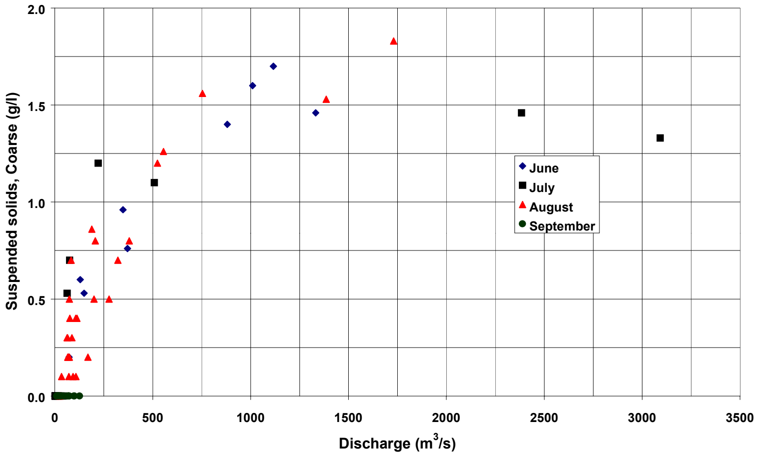


Figure 7.1:
Coarse suspended sediment concentration as a function of discharge at test site for monsoon months of the year 1997

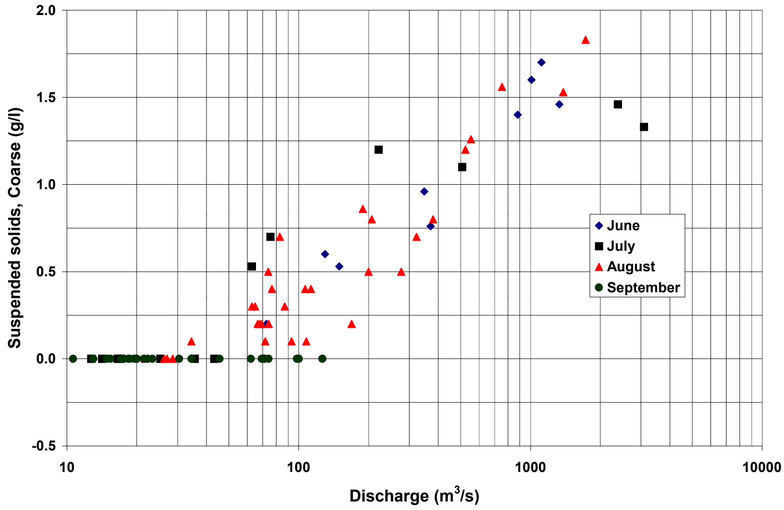


Figure 17.2:
Coarse suspended sediment concentration as a function of discharge on logarithmic scale at test site for monsoon months of the year 1997

From Figure 17.2 it is observed that a very distinct unique relationship between concentration and flow like for stage and discharge is not to be expected.

Medium fraction

In a similar manner as for the coarse fraction the medium fraction of various months in a particular hydrological year is investigated, hence the Figures 17.1 and 17.2 are repeated for the medium fraction, using the same reasoning. An example is shown in Figure 17.3.

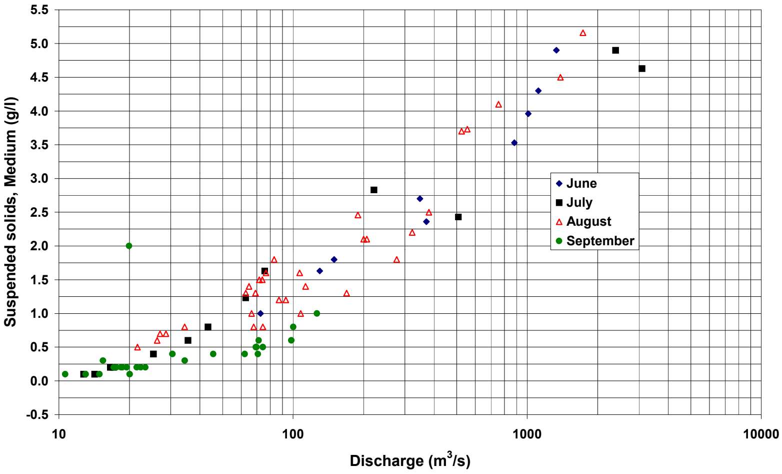


Figure 17.3:
Suspended sediment concentration of the medium fraction as a function of discharge on logarithmic scale at test site for monsoon months of the year 1997

From Figure 17.3 it is observed that the concentration of the medium fraction as a function of discharge is fairly independent of the month in the season. Only the concentrations for September, at the end of the monsoon season, appear to be less than earlier in the season. This may be due to the fact that in the medium fraction part of the material is wash load, which concentration is highest at the beginning of the monsoon season and less thereafter.

Note also in comparison with Figure 17.2, that the threshold discharge to get the medium bed material fraction in suspension is less than for the coarse fraction.

Fine fraction

The concentration of the fine fraction is highest at the beginning of the wet season, as its concentration is to a large extent determined by the supply of the fine fraction from the basin rather than by the flow velocity alone. This is clearly depicted in Figure 17.4 for Mahemdabad, where the concentration for low discharges at the start of the monsoon is much larger than at the end. From the Figure it is also observed that the concentration increases with the discharge, which implies that part of the fine fraction is bed material, brought in suspension by the flow; in case of wash load such a relationship with the discharge is generally not to be expected as it is the supply of sediment rather than the hydraulic conditions, which determine the concentration of the wash load as mentioned above.

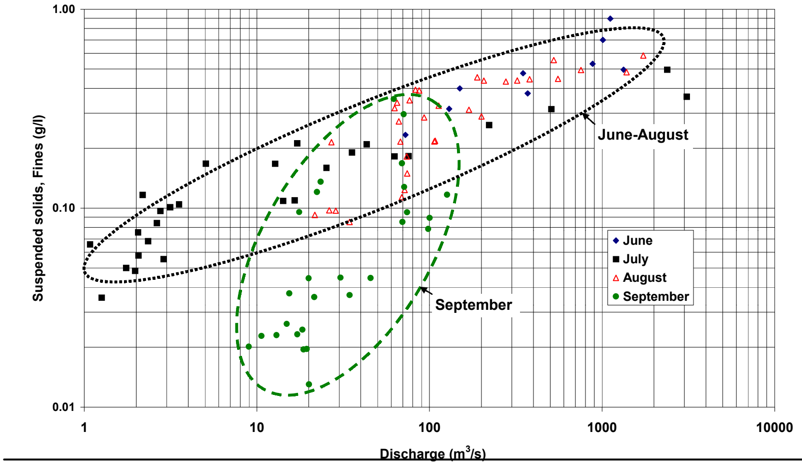


Figure 17.4:
Suspended sediment concentration of the fine fraction as a function of discharge at test site for monsoon months of the year 1997

If the results strongly deviate from the above sketched behaviour, the whole sampling and analysis procedure should be checked on possible errors.

17.3 SINGLE YEAR, MULTIPLE FRACTIONS

The next step in the secondary validation of suspended sediment concentrations comprises the comparison of the distinguished fractions for the hydrological year.

Coarse and medium fractions

The concentration of the coarse and medium fractions as a function of the discharge will show a similar pattern, with the observation that the threshold flow value for non-zero concentrations is less for the medium fraction. An example is shown in Figure 17.5.

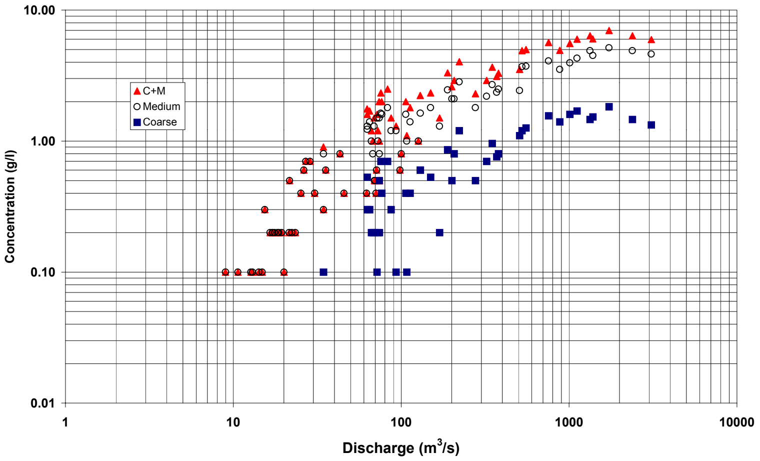


Figure 17.5:
Combined coarse and medium fraction at test site, year 1997

Coarse, medium and fine fractions

In a second type of graph, showing the sand/silt fraction (C+M) and the fine fraction, the relative importance of the various fractions in the total load is determined, see Figure 17.6. This is important for later use, when the concentrations have to be transformed into loads. If the concentration of fines is much smaller than of the coarse and medium fractions, then one sediment-discharge relationship valid for the whole hydrological year will be sufficient to derive the sediment loads. In case the fines constitute an important part of the sediment concentration separate curves for the C+M fraction and for the fine fractions will be required as no single relation between the concentration of fines and the discharge in a single year will exist.

In Figure 17.7 the concentration of the total suspended material as a function of the discharge is presented. Since the fines in this case play a minor role in the total concentration of the material in suspension, a fairly constant pattern is observed throughout the years with little variation from month to month as discussed above.

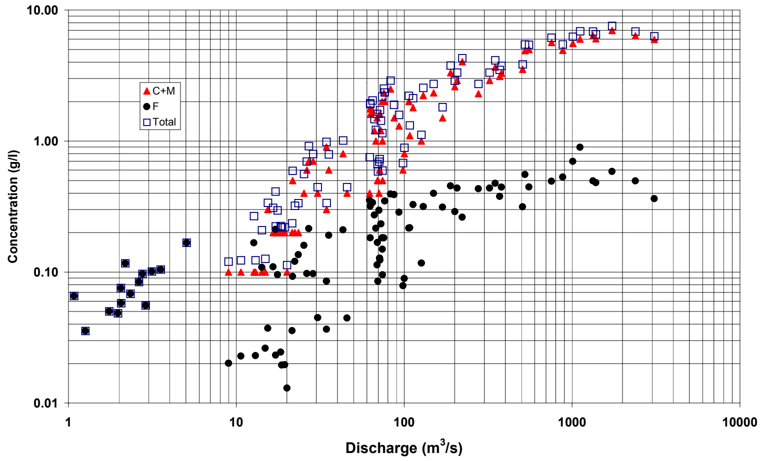


Figure 17.6:
C+M and fine fraction and total suspended sediment concentration at test site, year 1997

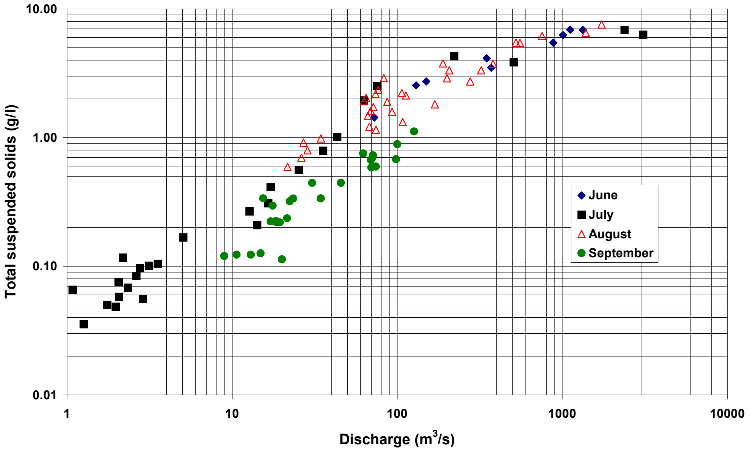


Figure 17.7:
Total suspended concentration at test site, year 1997

17.4 MULTIPLE YEARS, SINGLE FRACTIONS

A final check on the suspended sediment concentrations will be to compare the relationships between concentration and discharge for distinct fractions or totals from year to year.

Comparison of the concentration of C+M fractions

In Figure 17.8 the C+M concentrations for the test site for the years 1990 to 1994 are displayed. It is observed that a fairly constant relation between concentration and discharge exists during these years, though at times considerable deviations do occur.

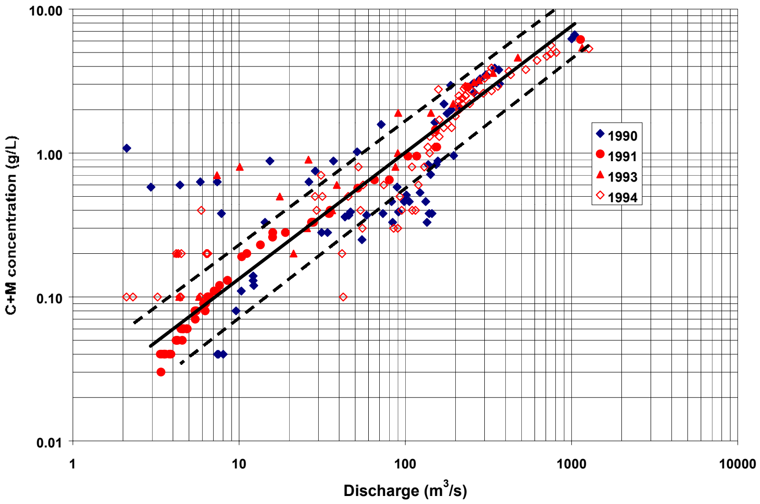


Figure 17.8:
C+M concentrations at test site for the years 1990-1994

In Figure 17.9 the C+M concentrations are presented for the years 1995 to 2000. In this graph also the range of values found in the period 1990-1994 are displayed. It is observed that particularly the years with the higher discharges fit well into the previous range. Years with moderate discharges deviate from this pattern. At this station, for a given discharge in the moderate years the concentration appears to be higher than in the high flow years. A reason could river bed mining just upstream of the site; during moderate years the mining will continue during a longer period of time than during a wet year. As a consequence, the sediment concentration discharge relation in this case might not be the same for sequential years.

A disturbing factor can also be large scale bank erosion, upstream of the measuring station. It might add material with a grain size distribution different from the material found in the river bed.

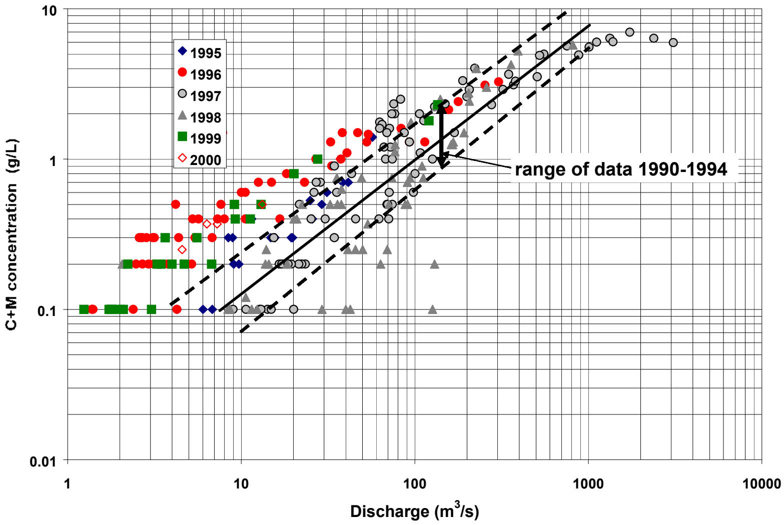


Figure 17.9:
C+M concentrations at test site for the years 1995-2000, compared to data range for the years 1990-1994

Comparison of the concentration of the fine fractions

Comparison of concentrations of the fine fractions from year to year will show a very cloudy picture due to varying supplies from the catchment and mining. Concentrations are likely to be high after a long dry period succeeded by heavy flows. When making comparisons for successive years, make sure that the same seasons are being compared.

Comparison of the total suspended sediment concentrations

Comparison of total suspended sediment concentrations for successive years makes sense when the concentration of the fines is small compared to the coarse and medium concentrations. In that case a pattern similar to the C+M fraction will occur.

18 COMPILATION OF SEDIMENT LOADS

18.1 SEDIMENT TRANSPORT

The sediment transport per unit width as a function of the flow velocity is often represented by:

$$s = av^b \tag{18.1}$$

where: s = sediment transport per unit width of river

v = flow velocity

a, b = coefficients

For the power generally values b > 3 are found. E.g. if the Engelund-Hansen transport equation is used, then b = 5. The total sediment transport S in a wide river with width B then becomes:

$$S = aBv^b = aB\left(\frac{S_e^{1/2}}{n}\right)^b h^{2b/3} = a_S Q^{2b/5} \text{ with } a_S = a\left(\frac{B^{1-2b/5} S_e^{3b/10}}{n^{3b/5}}\right) \tag{18.2}$$

where: a_S = coefficient

S_e = energy slope

n = hydraulic roughness coefficient

It is observed that for b = 5 it follows $S :: Q^2$. For coarse material it may well be that (18.2) has to be extended with a threshold flow value, as transport will only take place as the critical bed shear stress is exceeded (as in Meyer-Peter and Muller formula). The latter can be derived from the Shields curve, see Sub-section 2.6.2 of the Design Manual Sediment Transport Measurements Volume 5. Then the sediment-discharge relation gets the following general form:

$$S = a_S (Q - Q_0)^p \quad \text{with} \quad p = 2b/5 \tag{18.3}$$

This equation is identical to the power type stage-discharge relation, and the same procedure can be applied.

In practice sediment concentrations c in g/L are measured. The total sediment load in T/day then becomes:

$$S = c\left[\frac{g}{L}\right]Q\left[\frac{m^3}{s}\right] = \left[\frac{10^{-6}T}{10^{-3}m^3}\frac{m^3}{s}\right]cQ = 10^{-3}\ cQ\left[\frac{T}{s}\right] = 86.4\ cQ\left[\frac{T}{day}\right] \tag{18.4}$$

For coarse and medium sized material equations (18.1) and (18.2) or (18.3) apply. From (18.2) one gets:

$$c = a_S Q^{2b/5-1} \tag{18.5}$$

E.g. for b = 5 (Engelund -Hansen) it follows: c :: Q. Generally, by plotting the concentrations of coarse and medium sized material against discharge for each fraction some relationship should become visible either or not with a threshold value.

For wash load, the load is usually not dependent on the discharge but on the supply from the catchment. This load will be high after a prolonged dry period and falls back thereafter. For distinct periods in the year the load will be constant or c inversely proportional with the discharge. In Figure 17.4 however, the concentration, at least for some period, was shown to be a function of the discharge. Hence, here also a plot of c versus discharge is necessary to decide on the type of relationship to be established.

So plots of c versus Q for different fractions separately and for totals are very useful to identify the type of relationship between the sediment concentration and the discharge. For the coarse and medium fractions the relationship is likely to be reasonably stable through the seasons as discussed in Chapter 17, whereas for the fines a strong seasonality may be available. To see whether the concentration-discharge relation varies with time, it is required that in the c-Q plot different periods can be represented by different symbols and colours. Outliers are easily detected by the same procedure as used for the stage-discharge relation computation.

When the flood plain carries a large discharge a proper relationship between c or S and Q may not exists. In that case c or S versus $Q_{\text{main stream}}$ should give a better relationship, as the latter is then a better indicator of the carrying capacity of the main stream.

18.2 STEPS FOR SUSPENDED LOAD INCLUDE:

1. compute the discharge,
2. plot c versus discharge for the 3 distinguished fractions, with different symbols and colours for selected periods, (see Chapter 17),
3. identify clear outliers and eliminate wrong entries,
4. repeat the c-Q plot for aggregated fractions,
5. transform the concentrations into loads and determine for which fractions or aggregation of fractions an S-Q relationship can be established,
6. fit a power type curve of the type equation (18.3) through the S-Q plot,
7. if a considerable amount of wash load is available add a time variable load to the S-Q relation, derived from fitted relations for short periods of time,
8. create an S(t) series using the S-Q relation(s) and the Q(t) series,
9. enter monthly total suspended load values in the yearbooks,
10. all sediment entries in SWDES are to be transferred to the database in the Data Storage Centre; only the relevant data, needed to execute above steps, are to be transferred to HYMOS.

Note

Prior to the Hydrology Project, sediment concentrations were taken at 0.6 of the flow depth only. It depends on the value of u_*/W (i.e. shear velocity/fall velocity) whether the concentration at this depth will give an unbiased estimate of the average concentration in the vertical. As shown in Sub-section 2.6.3 of the Design Manual of Volume 5 for low values of u_*/W a strongy biased result may be obtained. It is therefore **required** that for those sites comparisons are being made between concentrations obtained from concentration measurements over the full depth and the single point measurements at 0.6 of the depth. To improve the historical data, a correction factor may be established for the various fractions, to be applied to the old measurements.

Example 18.1

The total suspended concentrations the test site data used in Chapter 17 for the year 1997 are displayed in Figure 17.7. From the analysis of the data presented in Chapter 17 it is concluded that:

- a distinct invariable relationship between C+M fractions and discharge exists for the entire year
- a fairly distinct relationship between the F fraction and discharge is applicable till September, which drops thereafter. However, in all, the concentration of fines appears to be small in comparison with the C+M fraction.

Therefore, in view of the small concentration of fines, one relation between total suspended load and discharge will be established for the year. Hence, the concentration of TSS (Total Suspended Sediments) is first transformed to total suspended load. Subsequently, a power type relation is established between discharge and total suspended load for different reaches of the discharge. The result is shown in Figure 18.1.

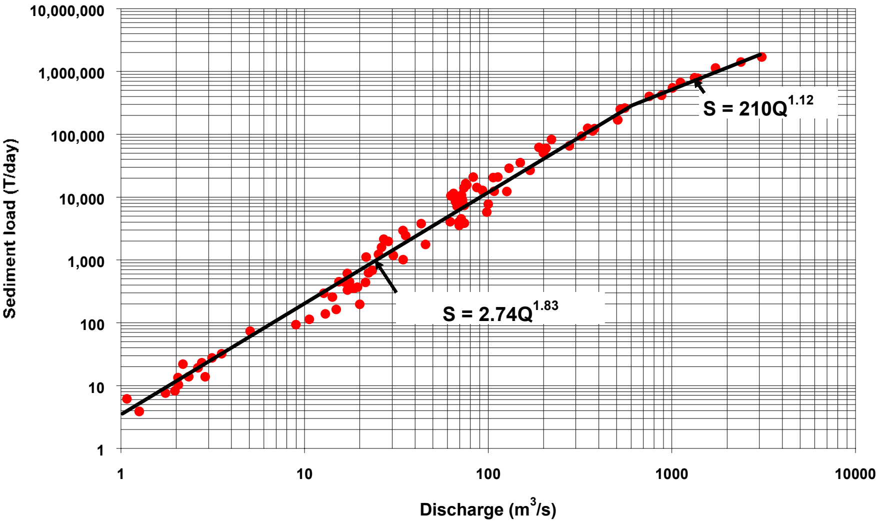


Figure 18.1: Fitting of S-Q relation to the total suspended load data for test site for the year 1997

From the results shown in Figure 18.1, it is seen, that the observed data are well fitted by two equations:

For Q < 538 m^3/s: $S = 2.74\ Q^{1.83}$ [T/d]

For Q □ 538 m^3/s: $S\ 210\ Q^{1.14}$ [T/d]

Next, the relationships are used to derive the suspended sediment load time series by transformation of the discharge time series. Still, it has to be determined whether a correction factor has to be applied to arrive at an unbiased estimate for the sediment load.

Subsequent activities

If the river is entering a reservoir, which is regularly surveyed, a comparison should be made with the sedimentation rate in the reservoir. For this a percentage has to be added to the suspended load to account for bed load transport. The match will further be dependent on the trap efficiency of the reservoir.

Finally, if more sediment transport stations are available on the same river, a sediment balance should be made for river stretches to estimate the sedimentation or erosion rate in the reach. Such information can be compared with data on bed levels for the reach for the balance period, when available.

19 REFERENCES

Dingman L.S. (2002)
Physical Hydrology
Second Edition, Prentice-Hall, Inc. New Jersey, USA

Isaaks, E.H,. and R.M. Srivastava (1989)
Applied Geostatistics
Oxford University Press, New York

Upadhya, D.S. et. al. (1990)
Space correlation structure of rainfall over India
Mausam 41, 4. pp 523-530.

WL|Delft Hydraulics (2002)
HYMOS 4, Processing System for Hydrological Data, User Manual
WL|Delft Hydraulics, Delft, Netherlands

WMO (1980)
Manual on Stream Gauging
World Meteorological Organisation, Operational Hydrology Report No 13, Geneva