

# Processing of Hydrological Data of Bhakra System Using HYMOS



**NATIONAL INSTITUTE OF HYDROLOGY  
JAL VIGYAN BHAVAN  
ROORKEE – 247 667  
2001-2002**

## PREFACE

Hydrologic data collected from various sources/instruments are generally raw data and the same cannot be used directly in most hydrologic analysis work. The processing of raw data is essential to make them usable for various applications. Most of the hydrological data processing activities are accomplished with the help of computers using various hydrological data processing systems.

Surface Water Data Entry System (*SWDES*) is a dedicated software developed under Hydrology Project for use in WINDOWS 95/98/NT/2000 environment. *SWDES* has been developed using Visual Basic for Access with a Microsoft Access 97 back-end database. The software is, in general, very user friendly as it provides sufficient information and comfort while working. It has been specially designed to fulfil the requirements of the surface water data entry needs. *HYMOS* is another software used for storage, processing and presentation of hydrological and environmental data. It combines an efficient database structure with powerful tools for data entry, validation, completion, analysis, retrieval and presentation. The wide variety of data processing and analysis features make *HYMOS* very suitable for typical project application in water-related studies, research and consultancy.

In the present study, the capabilities of *SWDES* and *HYMOS* software are illustrated for rainfall and discharge data entry, processing, analysis and reporting taking a real case study. The catchment of river Satluj between Namagia to Bhakra gauge-discharge site is taken for this study. Various options of *SWDES* and *HYMOS* are demonstrated with the help of available historical records of the rainfall and discharge for this catchment. The primary and secondary validation techniques have been applied and a few inconsistencies in the data have been brought out and presented in the form of tables and multiple time series graphs. Time series analysis of the runoff data at Rampur and Bhakra has been carried out through a time series model.

The study has been carried out by Sh. M. K. Goel, Sc. "E1", Sh. P. K. Agarwal, PRA, and Dr. S. K. Jain, Sc. "F" of the Water Resources Systems Division of this Institute.

(K. S. Ramasastry)  
Director

## ABSTRACT

Advances in hydrology are dependent on good, reliable and continuous measurements of the hydrological variables. Hydrologic data collected from various sources/instruments are generally raw data and cannot be used directly in most hydrologic analysis work. Thus, the processing of such raw data is the first and major task which has to be undertaken before drawing any hydrological inferences. Processing of hydrological data has two major objectives: one to evaluate the data for its accuracy and the other to prepare the data in a form most valuable to the users. Data processing using high-speed computers has immense potential for handling large volume of hydrologic data in a quick and economic way.

Hydrological and meteorological data are observed manually or digitally using a variety of instruments. *HYMOS* is a software, which is used for storage, processing and retrieval of hydro-meteorological data. This surface water data processing system has been introduced in various agencies in nine southern states of India under the Hydrology Project. A surface water data entry system (*SWDES*) has also been developed under the Hydrology Project and introduced under the project. *SWDES* is primarily used for data entry and primary validation.

In the study reported herein, the capabilities of *SWDES* and *HYMOS* software have been utilized for the processing and analysis of rainfall and discharge data in the Satluj basin. The catchment of river Satluj between Namagia to Bhakra gauging sites is considered for this study. The rainfall data have been validated using various techniques of secondary validation, such as multiple time series plots, spatial homogeneity test, and double mass analysis. Because of higher altitude difference between the stations, the data have been normalized before processing. Using the spatial correlation analysis, groups of stations with significant correlation have been identified. For runoff data between Rampur and Bhakra, the contribution from the intermediate catchment has been compared with the average rainfall. Time series analysis of the runoff data at Rampur and Bhakra has been carried out through a time series model. Few suspect values in the historical record (1977 – 1998) have been identified. Major deviations between observed and expected values were matched with the original record. Some of the suspected values were found to be wrongly entered in the database, which confirms the strength and necessity of data processing.

\* \* \*

# CONTENTS

	<b>Page No.</b>
<b>ABSTRACT</b>	
<b>CHAPTER 1 INTRODUCTION</b>	<b>1</b>
1.1 Hydrological Information System	1
1.1.1 Advantages of HIS	1
1.2 Hydrological Data Processing	2
1.2.1 Stages in Data Processing	2
1.3 Data Validation	3
1.3.1 Type of Measurement Errors	3
1.3.2 Levels of Validation	5
1.4 Scope of the Present Study	6
<b>CHAPTER 2 BRIEF DESCRIPTION OF SWDES &amp; HYMOS</b>	<b>8</b>
2.1 Data Processing Cycle	8
2.2 Surface Water Data Entry System (SWDES)	8
2.2.1 Need and Scope of SWDES	9
2.2.2 Software and Hardware Requirement for SWDES	10
2.2.3 Features of SWDES	10
2.2.4 Menu Items	12
2.3 Hydrological Data Processing System (HYMOS)	14
2.3.1 Software and Hardware Requirements	14
2.3.2 Features of HYMOS	15
<b>CHAPTER 3 THE STUDY AREA</b>	<b>19</b>
3.1 The Satluj Basin	19
3.2 Topography of the Satluj Basin	19
3.3 Climatic Conditions in Satluj Basin	22
3.4 Streamflow Characteristics of Satluj	23
3.5 Data Used in the Study	24
3.5.1 Rainfall data	24
3.5.2 Discharge data	25

<b>CHAPTER 4 DATA PROCESSING AND VALIDATION</b>	<b>26</b>
4.1 Entry of Data	26
4.1.1 Entry of Rainfall Data	26
4.1.2 Entry of Discharge Data	29
4.1.3 Entry of GIS Layers	29
4.2 Processing and Validation of Rainfall Data	31
4.2.1 Normalization of Rainfall	33
4.2.2 Spatial Correlation Analysis	37
4.2.3 Scrutiny by Multiple Time Series Plots	39
4.2.4 Spatial Homogeneity Test	42
4.2.5 Double Mass Analysis	52
4.2.6 Computation of Areal Rainfall	54
4.2.6.1 Computation of Thiessen Weights	54
4.3 Rainfall-Runoff Comparison for Intermediate Basin	55
<b>CHAPTER 5 TIME SERIES MODELLING</b>	<b>58</b>
5.1 Time Series Analysis	58
5.1.1 Components of a Time Series	58
5.1.2 Stationary Stochastic Processes	59
5.2 Time Series Models	60
5.2.1 Autoregressive (AR) Models	60
5.2.2 Moving Average (MA) Models	61
5.2.3 Autoregressive Moving Average (ARMA) Models	61
5.3 Partial Autocorrelation Function	62
5.4 Fitting of ARMA Models	63
5.5 Time Series Modelling of Bhakra Data	65
<b>CHAPTER 6 CONCLUSIONS</b>	<b>73</b>
<b>REFERENCES</b>	<b>75</b>

## **LIST OF TABLES**

<b>Table No.</b>	<b>Title</b>	<b>Page No.</b>
2.1	Hierarchical layout of menu items on main and secondary switchboards	13
2.2	Hierarchical layout of options at primary and secondary level	17
3.1	Altitude of different stations in Satluj basin	22
3.2	Salient features of Satluj catchment	22
3.3	Availability of daily rainfall data in Sutluj basin	24
3.4	Availability of discharge data for Satluj basin	25
4.1	Format for transfer of daily rainfall data at Berthin through MS ACCESS	28
4.2	Basic Statistics of daily rainfall at Bhakra raingauge station	35
4.3	Average annual rainfall for stations in Bhakra basin	36
4.4	Normalization coefficients for Bhakra, Berthin, Kasol, Suni, Kahu	36
4.5	Normalization coefficients for Kalpa, Namagia, Rakchham	37
4.6	Correlation coefficients matrix for daily time steps	39
4.7	Correlation coefficients matrix for monthly time steps	39
4.8	Suspect rainfall values and corresponding dates identified from multiple time series plots of rainfall data	43
4.9	Spatial homogeneity test for Kasol using Bhakra, Berthin, Kahu & Suni	45
4.10	Identification of suspect values using spatial homogeneity test	47
4.11	Verified & corrected suspect values from nodal agency (BBMB)	50
4.12	Results of Thiessen polygons analysis	55
5.1	Serial correlation coefficients of observed data	67
5.2	Serial correlation coefficients of standardized data	68
5.3	Performance of trial models – Bhakra data	70
5.4	Performance of trial models – Rampur data	72

## LIST OF FIGURES

Figure No.	Title	Page No.
1.1	Classification of measurement errors	4
2.1	Data processing cycle	8
2.2	A window showing main switchboard of <i>SWDES</i>	13
2.3	A Typical window of <i>HYMOS</i> data processing software	18
3.1	Index map of the Satluj basin (Indian part only)	20
3.2	DEM elevation map of Satluj basin (Indian part only)	21
4.1	Form for entry or editing of daily rainfall data	28
4.2	A view of graphical presentation of discharge data in <i>HYMOS</i>	30
4.3	Representation of Satluj basin and observation network in <i>HYMOS</i> Netter	31
4.4	Graph showing the derived annual series for Bhakra rain gauge station	34
4.5(a-h)	Multiple time series plots of rainfall at Bhakra, Kahu, Kasol, Berthin, and Suni	42
4.6	Definition sketch of test and base (neighbouring) stations	43
4.7(a-e)	Double mass curve plot for Bhakra, Kahu, Kasol, Berthin, and Suni	53
4.8	Thiessen polygons of rainfall stations for Satluj basin between Rampur & Bhakra	55
4.9	Average monthly rainfall and runoff (mm) in catchment between Bhakra and Rampur	57
5.1	Observed discharge at Bhakra from 1980-1993	66
5.2	Observed discharge at Rampur from 1980-1997	66
5.3	Standardized series for Bhakra	67
5.4	Standardized series for Rampur	68
5.5	Differenced series at Bhakra	69
5.6	Differenced series at Rampur	69
5.7	Correlogram of differenced series for various lags for Bhakra	70
5.8	Correlogram of differenced series for various lags for Rampur	70
5.9	Residuals series for Bhakra	71
5.10	Residuals series for Rampur	71

## **DIRECTOR**

Dr. K. S. Ramasastrri

## **STUDY GROUP**

M. K. Goel                      Scientist 'E1'

P. K. Agarwal                  Principal Research Assistant

Dr. S. K. Jain                  Scientist 'F'

## **TYPING**

Smt. Mahima Gupta          Personal Assistant



# **CHAPTER - 1**

## **INTRODUCTION**

### **1.1 HYDROLOGICAL INFORMATION SYSTEM**

A Hydrological Information System (HIS) comprises the physical infrastructure and human resources to collect, process, store, and disseminate data on hydrological, geo-hydrological and hydro-meteorological variables. The physical infrastructure includes observation networks, laboratories, data communication systems and data storage and processing centers equipped with databases and tools for data entry, validation, analysis, retrieval and dissemination. The human resources refer to the well-trained staff with a variety of skills to observe, validate, process, analyze and disseminate the data. Efficiency requires that all activities in the HIS are well tuned to each other, to provide the required data timely in proper form and at minimum cost.

The primary role of HIS is to provide reliable data sets for long-term planning, design and management of water resource and water use systems and for research activities in the related aspects. The first step is to obtain the information on the temporal and spatial characteristics of the system by having a network of observational stations. The data collected for different hydro-meteorological phenomenon through this observational network is called the raw or observed data. Such observed data have to be processed to ensure the reliability of the resulting information. Both, raw and processed, data sets have to be properly stored. Processed data is used for dissemination of information and raw data is used to permit inspection and revalidation in response to queries from users.

The activities under HIS can be broadly classified as: a) assessing the needs of users, b) establishment of an observational network, c) management of historical data, d) data collection, and, e) data processing, analysis and reporting.

#### **1.1.1 Advantages of HIS**

The rapid advance in computer technology, in hardware speed of operation and data storage capacity as well as the capability of hydrological software has greatly simplified the

management of large quantities of hydrological data that was initially accomplished through time-consuming manual methods. Computer-based systems offer the following advantages:

- They permit and promote the standardisation of processing, validation and reporting procedures,
- Very large amounts of data can easily be handled,
- They greatly speed up the progress from data collection to completion and storage,
- Users can be provided with data in the required tabular or graphical format,
- Their use promotes staff interest by eliminating manual handling and allowing results to emerge quickly.

## **1.2 HYDROLOGICAL DATA PROCESSING**

Hydrologic data collected from various sources/instruments are generally raw and can not be used directly in most hydrologic analysis work. Therefore, processing of such data is essential to make them usable for various studies. It is particularly important to ensure the quality and reliability of data by applying a variety of validation checks before providing the data to the users. The raw data are in a variety of formats such as hand-written records, charts and digital records. Raw data, as observed and recorded, may contain many gaps and inconsistencies. These raw data are passed through a series of operations, such as data entry, validation checks, in-filling of missing values, compilation of data in different forms, and analysis of data for commonly required statistics etc.

Processing and validation of hydrological data require an understanding of field practices. This includes the principles and methods of observation of hydrological variable in field. With knowledge of measurement techniques, typical errors can be identified. Similarly knowledge of the regime of a river will facilitate the identification of spurious data. For example for river level or flow, a long period at a constant level followed by an abrupt change to another period of static level would be identified as suspect data in a natural catchment but possibly due to dam operation in a regulated river.

### **1.2.1 Stages in Data Processing**

Data processing is a broad term covering all activities from receiving records of observed raw data to making them available in a useable form. Processing of hydrological data is not a single step process. It is carried out in a series of stages, starting with preliminary checking in the field, through receipt of raw data at lower level offices and successively

higher levels of validation, before it is accepted as fully validated data in the data storage centre. Various stages of data processing plan are:

- Receipt of data
- Data entry to computer
- Data validation - primary, secondary and hydrological
- Data completion and correction
- Data compilation
- Data analysis
- Data reporting
- Data transfer

Most of the data processing activities are best accomplished with the help of computers using dedicated hydrological data processing software. In the present study, validation of the historical data of Satluj basin up to Bhakhra reservoir has been carried out using two dedicated software, *SWDES* and *HYMOS*. A general description of hydrological data validation is given in the following.

### **1.3 DATA VALIDATION**

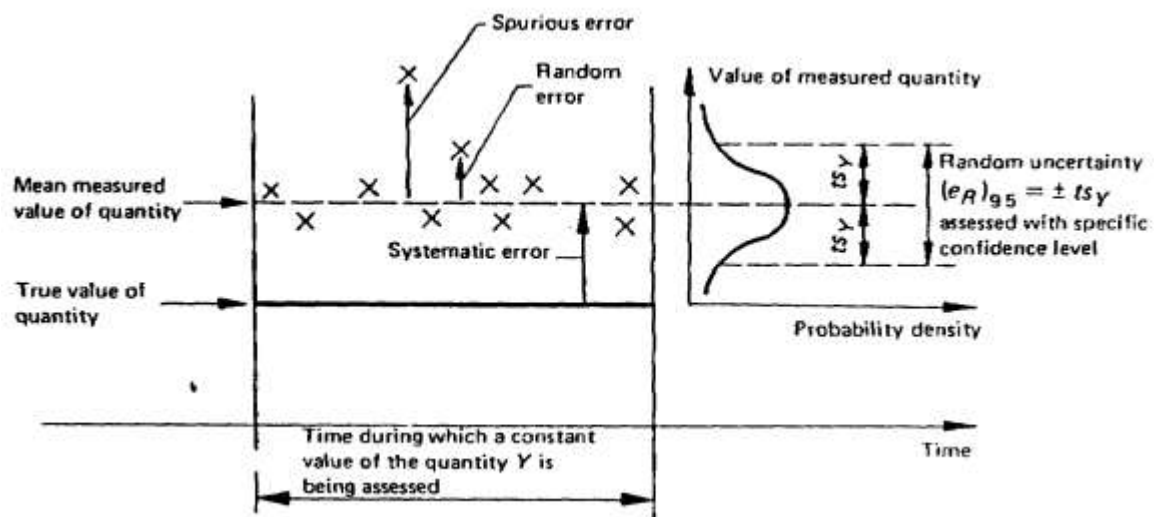
Data validation is the means by which data are checked to ensure that the final figure stored in the HIS is the best possible representation of the true value of the variable at the measurement site at a given time or in a given interval of time. Validation recognizes that values observed or measured in the field are subject to errors and that undetected errors may also arise in data entry and computation. Validation is carried out for three reasons:

- to correct errors in the recorded data where this is possible,
- to assess the reliability of a record even where it is not possible to correct errors,
- to identify the source of errors and thus to ensure that such errors are not repeated in future.

#### **1.3.1 Type of Measurement Errors**

Typical errors at observation stations can be from faulty equipment, from malfunction of instrument (e.g. slippage of float tape in level recorder), from improper instrument setting by observer, from improper exposure conditions (e.g. stilling well blocked), personal observation errors (e.g. gauge misread or value interpolated), transcription error in writing the

observed reading, or error in field computation (e.g. current meter measurements). Measurement errors may also be classified as random or systematic or spurious in nature (Fig. 1.1). Random errors are equally distributed about the mean or 'true' value. The errors of individual readings may be large or small, e.g. the error in a staff gauge reading where the water surface is subject to wave action, but they tend to compensate with time or by taking a sufficient number of measurements. Systematic errors represent a systematic difference, either positive or negative, between the measured value and the true value and the situation is not improved by increasing the number of observations. Examples are the use of the wrong rain gauge measure or the effect on a water level reading of undetected slippage of a staff gauge. Systematic errors are generally the more serious and are what the validation process is designed to detect and if possible to correct. Spurious errors are sometimes distinguished from random and systematic errors as due to some abnormal external factor. An example might be an evaporation pan record where animals have been drinking from the pan. Such errors may be readily recognized and the measurements discarded.



**Figure. 1.1** Classification of measurement errors

As a consequence of such errors it is important not only to ensure the use of good equipment and observational procedures but also to monitor the quality of all the data received. Validation involves a process of sequential and complementary comparisons of data:

- for a single data series, between individual observations and pre-set physical limits,

- for a single series between sequential observations to detect unacceptable rates of change and deviations from acceptable hydrological behaviour most readily identified graphically,
- between two measurements of a variable at a single station, e.g. daily rainfall from a daily gauge and an accumulated total from a recording gauge,
- between two or more measurements at neighbouring stations, e.g. flow at two points along a river,
- between measurements of different but related variables, e.g. rainfall and river flow.

Improvement in computing facilities now enables such validation to be carried out whereas in the past, the volume of data and the time required to carry out manual calculations was prohibitive. This was a major reason why rigorous validation of data was rarely carried out.

### **1.3.2 Levels of Validation**

It is preferable to carry out the data validation as soon as the data is observed and as near to the observation station as possible. However, full validation close to observation sites is impractical both in terms of computing equipment, staff, and wide geographical spread of observation stations. The sequence of validation is, therefore, divided so that those steps, which primarily require interaction with the observation station, are carried out in close proximity whereas more complex comparisons are carried out at higher levels. Thus, data validation to be carried out is grouped into three major categories:

- (a) Primary validation,
- (b) Secondary validation, and
- (c) Hydrological validation.

#### ***a) Primary validation***

Primary data validation highlights those data that are not within the expected range and are not hydrologically consistent. Primary validation is carried out at the lowest level data processing center immediately after entering the raw data to the computer system. Primary validation is mainly involved with comparisons within a single data series and making comparisons between observations and pre-set limits based on expected behavior of a hydrological phenomenon.

### ***b) Secondary validation***

Secondary validation consists of comparisons between the same variable at two or more stations. It tests the data against the expected spatial behavior of the system. Secondary validation is based on the spatial information available from a number of neighboring observation stations within a comparatively large area. The assumption, while carrying out such comparison, is that the variable under consideration has adequate spatial correlation within the distances under consideration. Secondary data validation is done at higher level offices (which have jurisdiction over larger number of gauging stations) after primary validation has been carried out at the lower level offices.

### ***c) Hydrological validation***

Hydrological validation consists of comparing one record with one or more other interrelated variables at the same or adjacent stations and is designed to show up inconsistencies between the time series or their derived statistics. Hydrological validation may be applied to a measured variable (water level) but is more often applied to derived variables (flow, runoff). This is usually done through regression analysis or simulation modeling.

If a record has been subjected to thorough field checking and primary and secondary validation soon after the record has been obtained, hydrological validation should reveal no more than is already known. However, for historical data to which no (or few) such checks have been applied, hydrological validation may become the principal check on the reliability of the record. Thorough hydrological validation requires a high level of professional expertise and can be very time consuming.

## **1.4 SCOPE OF THE PRESENT STUDY**

In the present study, data validation has been carried out for the historical record of the Satlej basin (Indian part only) up to Bhakhra reservoir. Various capabilities of *SWDES* and *HYMOS* software have been used for the processing of rainfall and runoff data of the basin. The utilised software does not include the processing of the snow data. Hence the scope of this report is limited to the processing of rainfall and discharge data only. The primary and secondary validation techniques have been applied and a few inconsistencies in the data have been brought out. The data series from the year 1977 to 1998 have been used in

the study, based on the availability at different stations. Time series analysis of the runoff data at Rampur and Bhakhra has been carried out through a time series model.

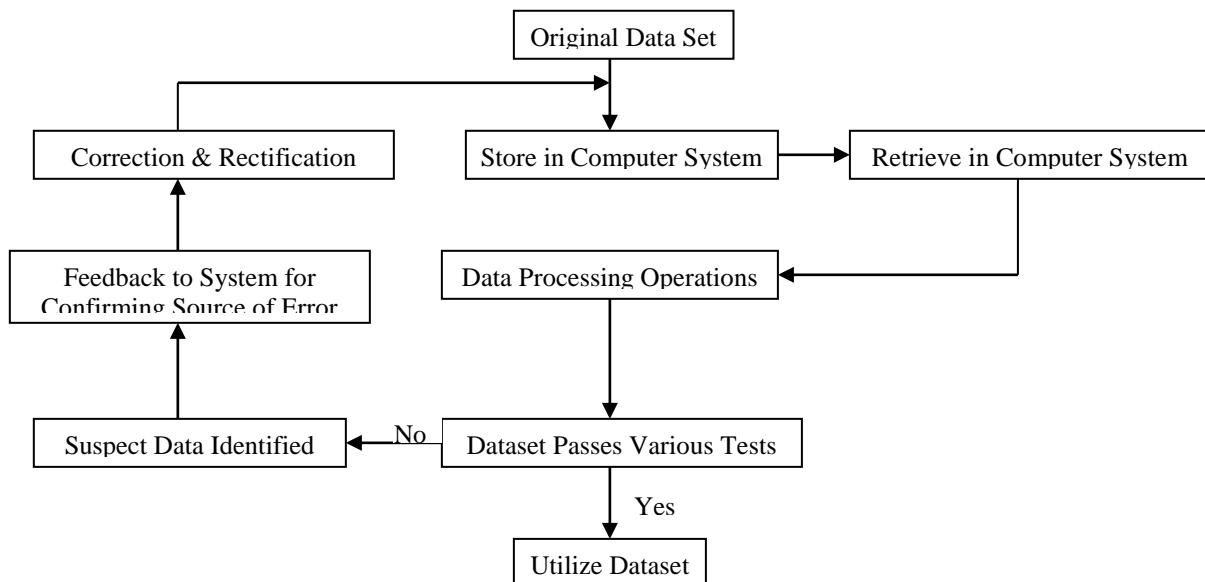
A brief description of the *SWDES* and *HYMOS* software is given in Chapter – 2. General details of the basin, including data availability, is presented in Chapter – 3. Data processing is covered in Chapter – 4 while time series modeling is detailed in Chapter – 5. The conclusions from the study have been presented in Chapter – 6.

\* \* \*

## CHAPTER - 2 BRIEF DESCRIPTION OF SWDES & HYMOS

### 2.1 DATA PROCESSING CYCLE

Data is defined as "things known" and information as "items of knowledge". Data is the raw material used to produce information. Data is selected and organized with respect to user, problem, time, place, and the function. A data processing cycle represents the data processing activities executed in a cyclic way. In the data processing cycle, we first gather the data for processing. This is known as origination and collection. The collected data is then processed and stored for subsequent retrieval. The result and information obtained after data processing is used to identify any suspect values, which is then confirmed from the existing setup. If the error sources are confirmed, corrective measures can be taken. Data processing life cycle activities are illustrated in Figure 2.1.



**Figure – 2.1** Data processing cycle

### 2.2 SURFACE WATER DATA ENTRY SYSTEM (SWDES)

The first activity in data processing is to make the data available for processing on the magnetic media, either by loading the digital data obtained from data loggers directly or keying-in the data available from the data sheets. It is very essential to use a software which is customised to facilitate data punching operations, carrying out necessary checks while



entering the data and maintaining the punched data meticulously in well organised databases. The requirements for such a data entry scheme for surface water data is fulfilled by “**Surface Water Data Entry System**” called **SWDES** in short.

### **2.2.1 Need and Scope of SWDES**

The bulk of meteorological and hydrologic quantity and quality data is observed manually from a variety of instruments and equipment. Some of the data is also observed and stored automatically on magnetic media using electronic dataloggers. There are scores of hydrometric and meteorological observation stations within a State/Region generating large quantity of data on a continuous basis. The data thus collected has to be properly entered on to the magnetic media for further processing. There are number of general purpose software readily available which, though can be employed for entering the hydrological data but do not satisfy all the requirements specific to entry, scrutiny and easy organisation of surface water data. The raw data as observed in the field is recorded on forms having layouts suitable to the type of data. These data forms are used while keying-in the data. For faster data entry it is essential to have data entry screens of the software resembling these data forms ensuring minimum mistakes at the time of data entry operations. Also, there must be enough scope for scrutinising the data as recorded in the forms for obvious mistakes which may have been committed while making computations or transcribing the forms. At the same time, the data entry system must provide a user-friendly environment so that it does not require special skills to operate for entering, organising and maintaining the data sets.

Surface Water Data Entry System (**SWDES**) is a dedicated software developed by the Hydrology Project for use on personal computers in WINDOWS 95/98/NT/2000 environment. The software has been developed using Visual Basic for Access with a Microsoft Access 97 back-end database. The software is, in general, very user friendly as it provides sufficient information and comfort while working. It has been specifically designed to fulfil the requirements of the surface water data entry needs. **SWDES** provides facilities for entry of rainfall, water level, stage-discharge, water quality, sediment and climatic data related to hydrology. The entered data are made to pass through essential data entry checks so that the possibility of committing mistakes is minimized. The graphical facility gives added advantage of visual means of validating the entered data. In addition to data entry, **SWDES** is useful for primary validation of the data. For higher level of validation and processing, another dedicated hydrological surface water data processing software (**HYMOS**) is used. The

database of *SWDES* can be easily exported to *HYMOS* for higher level of data validation. Hence, *SWDES* provides a kind of front end for the data entry personnel for entry and primary validation of meteorological and surface water quality and quantity data.

*SWDES* is primarily designed for the entry of time series data but it also incorporates space-oriented data sufficient to locate and catalogue the stations under the control of a particular state or agency. Stations and series can then be accessed from typical Windows Menus and Toolbars by clicking on appropriate buttons. For all equidistant and cyclic time series data, the software provides entry screens automatically with date and time labels against which the variable values are entered. This simplifies data entry and avoids the potential errors of date/time entry. It also provides data entry checking capability, rejecting clearly spurious values and flagging suspect ones for inspection. For example, it will reject entry of an alpha character in a numeric field or duplicate decimal point and will highlight for inspection values above or below a pre-set limit. In addition, options for plotting time series graphs at time of entry are available in most cases.

### **2.2.2 Software and Hardware Requirement for *SWDES***

The minimum recommended hardware and software support necessary for running this software program is:

- Pentium with 166 MHz + (preferably – Pentium 200/Pentium II)
- Colour monitor with 840 x 600 resolution
- 32 Mb RAM+
- 1.2 GB+ Hard disk with minimum 200 MB free space
- Operating System – Windows 95/98/NT
- Software – MS Office 97 with full Access 97 (Run-time version not requiring Access 97 also available)
- 3.5 inch floppy drive/CD-ROM drive

### **2.2.3 Features of *SWDES***

*SWDES* is primarily oriented towards easy and reliable data entry options for various hydro-meteorological quality and quantity variables. It organizes the data into well-defined

databases using concept of relational database systems. The main features of the software are given below.

***a) User authorization***

Adequate facility for user authorization and identification is available. Number of users can be authorized by the system manager for working with the software with varying level of authority.

***b) Organized databases***

The data pertaining to different sub-basins or offices or periods can be organized in separate databases (referred to as Workareas in the software) in a well-organized and methodical way.

***c) Extended data types***

A comprehensive set of variables is available in the program. The characteristics of the variables, like description, unit and type of measurement etc. are also maintained. Adequate flexibility is provided for adding new data variables in the program.

***d) Master information***

The software maintains a set of important hierarchical information on administrative and drainage boundaries and that on the offices controlling various observation stations. This master information helps in avoiding duplicate entries and wrong spellings for the same item in the database and at the same time the user is not required to waste time every time in keying-in the same information. Adequate facility is available in the system to extend or modify this type of information.

***e) Static/semi-static data***

The software provides for the entry of necessary characteristics associated with the observation stations. The data pertaining to various variables are stored under well-defined data series that also have useful characteristics associated with them. The data on current meters and setting of the zero of the gauge for different validity periods, cross-sectional profiles etc. are also stored in an organized manner.

#### ***f) Data entry screens***

A number of user-friendly data entry screen layouts are available that are appropriate for data pertaining to different types of variables and time intervals.

#### ***g) Data entry checks***

Adequate facilities are built into the system for providing a number of data entry checks so that the amount of data entry errors can be reduced to a very low level. Sufficient flexibility is available in the system for making these data entry checks more effective.

#### ***h) Graphical options***

Facility is provided in most of the data entry options to make plots of the data being entered. This provides the users a very convenient way to graphically visualize the entered data, which thereby will help in reduction of errors in data entry.

#### ***i) User friendliness***

The software provides sufficient level of user friendliness while working. Most of the work is accomplished by choosing an item by clicking it from the available list. Only the actual data to be entered are required to be keyed-in using the keyboard. Most of the actions are performed by choosing an option by clicking appropriate buttons. Suitable error and help messages are displayed for guiding the user to work with the software.

#### ***j) On-line users' manual***

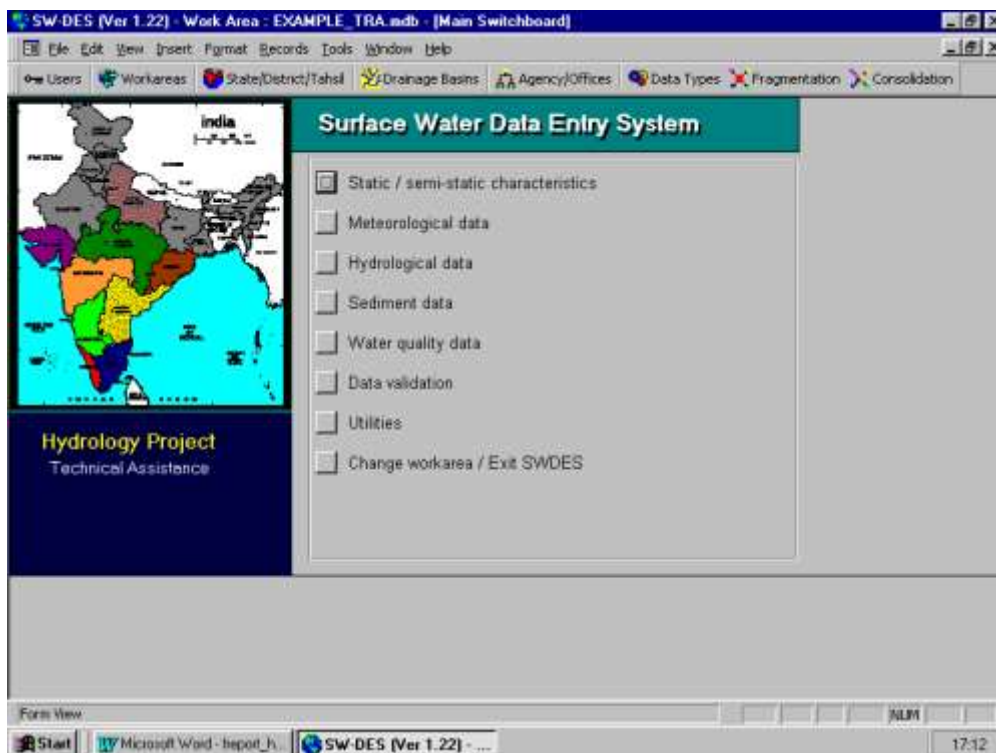
The software provides the contents of the user manual on-line. Appropriate portions of the manual are displayed on the basis of screens from which the help is invoked. Flexible and faster navigation facilities are provided to browse through the manual conveniently.

### **2.2.4 Menu Items**

The major options available in the program are listed as different menu items in the program window (also called as *main switchboard*). Any menu item can be invoked by clicking on the button associated with it. The options on the main switchboard lead to another menu screen, called the *secondary switchboard*, for selection of the desired sub-option. The function of various menu items is self-evident and therefore does not need any explanation. The hierarchical layout of these menu items is given in Table - 2.1. The layout of the main menu screen (main switchboard) is shown in Figure - 2.2.

**Table 2.1** Hierarchical layout of menu items on main and secondary switchboards

Main switchboard items	Corresponding items in secondary switchboards
1. Static/semi-static characteristics	1. Station characteristics
	2. Series characteristics
	3. Current meter information
	4. Current meter ratings
	5. R L of gauge zero
	6. X-section data
	7. Salient features of reservoir/diversion scheme
	8. Elevation-Area-Capacity data
2. Meteorological data	1. Rainfall – Daily
	2. Rainfall – Twice daily
	3. Rainfall – Hourly
	4. Climatic – Daily
	5. Climatic – Twice-daily
	6. Pressure/Temperature/Humidity – Hourly
	7. Sunshine duration – Hourly
3. Hydrological data	1. Water level – Multiple times a day & temperatures
	2. Water level – Multiple times a day
	3. Water level – Hourly
	4. Flow measurement data
	5. Summary stage-discharge data
	6. Reservoir/diversion scheme data
4. Sediment data	1. Summary suspended sediment data
5. Water quality data	1. Laboratory information
	2. Parameter information
	3. Sample collection information
	4. Sample data entry & validation register



**Figure 2.2** A window showing main switchboard of SWDES

## **2.3 HYDROLOGICAL DATA PROCESSING SYSTEM (*HYMOS*)**

*HYMOS* is an information system for storage, processing and presentation of hydrological and environmental data. It combines an efficient database structure with powerful tools for data entry, validation, completion, analysis, retrieval and presentation. The wide variety of data processing and analysis features make *HYMOS* very suitable for typical project application in water-related studies, research and consultancy. These projects often require processing of large amounts of data in a limited time frame.

Through continuous feedback from users, *HYMOS* has been developed into a flexible and multi-functional hydrological information system. The information system facilitates processing of a wide range of water-related data. Fields of application of *HYMOS* include:

- surface and ground water hydrology
- meteorology
- water quality and ecology
- water resources management

*HYMOS* makes use of a data structure, part of which is open, which ensures convenient interaction and data transfer with other existing databases and modeling systems. Being part of DELFT TOOLS, *HYMOS* is fully integrated with other software packages of WL | DELFT HYDRAULICS like SOBEK, DELWAQ and RIBASIM. The *HYMOS* database is time series oriented with common facilities for spatial analysis. In combination with a Geographical Information System (GIS) for comprehensive geographical data analysis, it covers all data storage and processing requirements for planning, design and operation of water management systems. Exchange of data between *HYMOS* and most common GIS systems is easy and user friendly.

### **2.3.1 Software and Hardware requirements**

*HYMOS* works on a personal computer equipped with at least:

- a Pentium processor
- 32 MB internal memory
- 100 MB available memory on the hard disk
- a CD drive

- MS-Windows 95 or MS-Windows NT;
- a SVGA-graphics adapter with 1 MB video memory and corresponding monitor
- a printer.

### **2.3.2 Features of *HYMOS***

*HYMOS* is comprehensive, well tuned and easy to use. The graphical and map-based user interface offers a pleasant and efficient working environment. Various features of *HYMOS* are briefly described in the following:

#### ***a) Entry and editing***

Monitoring network information and time series data can be entered manually, exchanged with spreadsheets and databases by using the Clipboard or imported from file with help of the data editor. Data from automatic monitoring networks can be downloaded automatically.

#### ***b) Validation and completion***

Validation and completion is an important first step in data processing. *HYMOS* offers a wide range of tabular, graphical, computational and statistical validation techniques such as tabulation of series and flagging of outliers, plotting of time series, relation curves and double mass analysis, and near-neighbor checks and series homogeneity tests.

Various options are available for series completion. Among these are interpolation techniques that use series relations derived from different methods such as regression techniques or spatial relations. Much emphasis has been given to facilities for administration of the data processing steps and data qualification features. The latter include assignment of labels to individual data indicating the source, e.g. original or corrected and reliability of data.

#### ***c) Analysis of meteorological data***

Meteorological data like rainfall, snow depth, wind speed and radiation can be stored and analysed with *HYMOS*. Typical analyses are the computation of catchment rainfall - e.g. based on Thiessen polygons and Kriging - and evapotranspiration. The latter module provides a range of methods including Penman with or without FAO-correction, pan evaporation, Christiansen, radiation, Makkink, Jensen-Haise, Blaney-Criddle and mass transfer method.

#### ***d) Flow measurements and stage-discharge analysis***

The stage-discharge analysis options include the elaboration and validation of current metering data, the computation and validation of discharge rating curves and the extension of rating curves. Rating curves can be represented by parabolic or power type equations, with up to five different water level ranges at a time. The Jones correction for unsteady flow can be applied and a procedure for effectively eliminating backwater effects while modeling the rating curve is available.

Water levels can be converted into discharges based on rating curves or pre-programmed and user defined structure equations.

#### ***e) Rainfall-runoff simulation***

*HYMOS* includes the Sacramento rainfall-runoff model, which is a well known and widely used conceptual hydrological model. Other hydrological, hydrodynamic or environmental models can be easily linked to the system when required.

#### ***f) Statistical and time series analysis***

An extensive statistical package is available under *HYMOS* covering amongst others:

- computation of basic statistical parameters,
- fitting distributions including normal, log-normal, various Pearson types, log-Pearson, Rayleigh, exponential, various Goodrich extreme types, Pareto, & P.O.T. distributions
- data generation,
- correlogram and spectral analysis,
- range and run analysis,
- computation of frequency and duration curves,
- polynomial, multiple and step-wise regression analysis, and
- time series analysis, e.g. cross and autocorrelation and spectrum.

#### ***g) Reports and graphs***

*HYMOS* has a powerful graph server and standard templates for the graphical output of all analysis modules. The flexible graph server allows the user to easily tailor graph layouts to individual requirements. A wide range of reports - e.g. tables ready for presentation in yearbooks - can be generated by *HYMOS*. For final editing, when required, the output

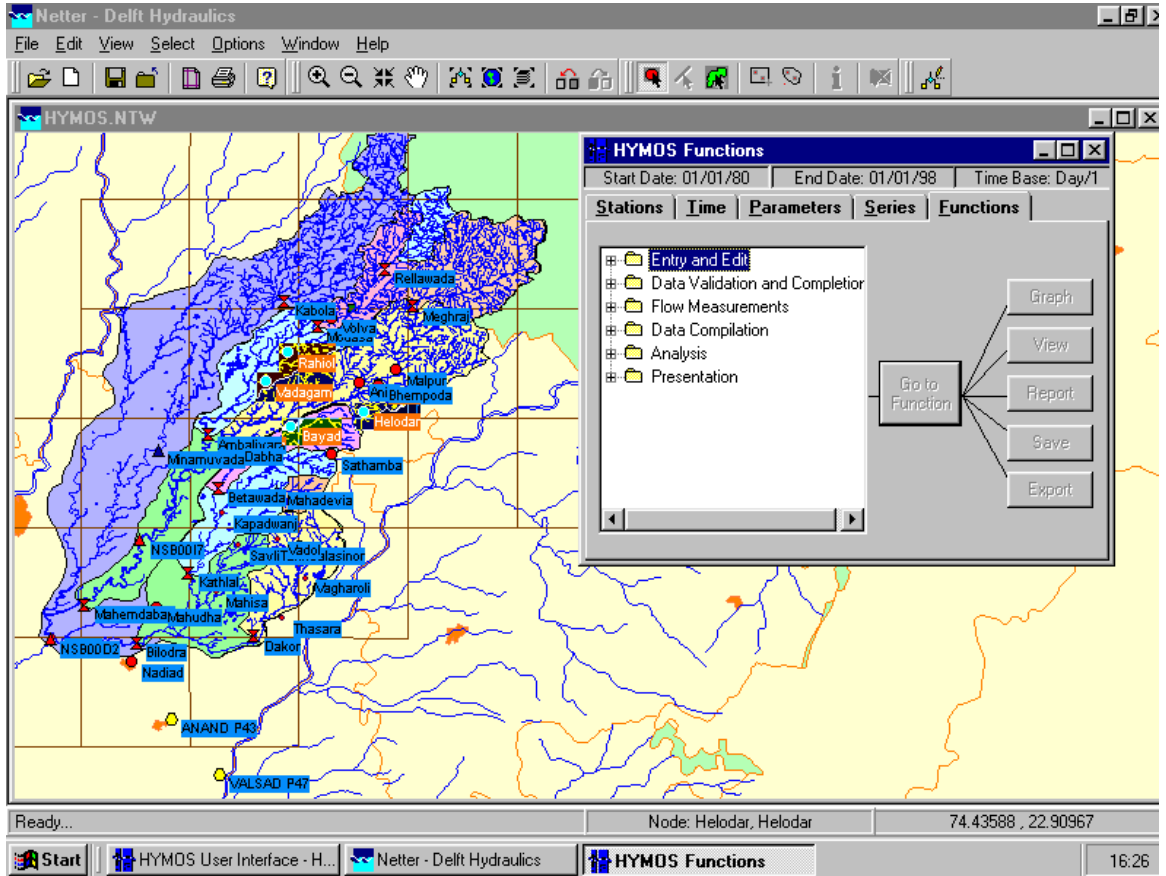


graphs and reports can be easily transferred to other WINDOWS programs such as the MS Office components - EXCEL and WORD.

Data processing functions of HYMOS are activated once the required observation stations are selected from the GIS background map loaded on the active window. The hierarchical layout of options at the primary and secondary level are given in Table - 2.2. Figure 2.3 shows a typical window of HYMOS data processing software.

**Table 2.2** Hierarchical layout of options at primary and secondary level

<b>Primary Level Options</b>	<b>Secondary Level Options</b>
1. Entry and Edit	1. Station
	2. Series
	3. Stage Discharge
	4. Relations
2. Data Validation and Completion	1. Time Series Graphs
	2. Screening of Data
	3. Conditional Checks
	4. Compute Relation Curve
	5. Compare Relation Curve
	6. Double Mass Analysis
	7. Series Homogeneity Tests
	8. Spatial Homogeneity Tests
	9. Spatial Correlation
	10. Interpolation
	11. Regression Models
	12. Sacramento Models
3. Flow Measurements	1. Processing of Flow Measurements
	2. Fitting of Rating Curves
	3. Validation of Rating Curves
	4. Extrapolation of Rating Curves
	5. Shift Adjustment
	6. Stage-Discharge Transformation
	7. Stage-Discharge Structure
	8. Measurement Structures
4. Data Compilation	1. Aggregation of Series
	2. Dis-Aggregation of Series
	3. Series Transformation
	4. Areal Weights
	5. Areal Theissen
	6. Areal Computation
	7. Min-Mean-Max Series
	8. Evaporation
	9. Spatial Interpolation
	10. Special Compilation Functions
5. Analysis	1. Basic Statistics
	2. Fitting Distributions
	3. Statistical Tables
	4. Frequency and Distribution Curves
	5. Time Series Analysis
	6. IDF Relations
6. Presentations	1. Print Report



**Figure 2.3.** A Typical window of HYMOS data processing software

\* \* \*

## **CHAPTER - 3**

### **THE STUDY AREA**

#### **3.1 THE SATLUJ BASIN**

The Satluj river rises in the lakes of Mansarover and Rakastal in the Tibetan Plateau at an elevation of about 4570 m above mean sea level (msl) and forms one of the main tributaries of Indus river. It travels about 322 km in the Tibetan province of Nari-Khorsam forming a plateau by successive deposits of boulders, gravel, clay and mud. The flow of Satluj, obtained mainly from snow and glaciers, has cut a valley about 914 m deep through these deposits. After flowing in north-westerly direction, it changes direction towards south-west and covers another 322 km up to Bhakra gorge, where a 225 m high straight gravity dam (Bhakra/Govind Sagar) has been constructed. The lower catchment largely drains directly into the reservoir and the higher slopes drain through tributaries. The important tributaries of the Satluj river are the Soel khad, Alseed khad, Ali khad, Gamrola khad, Ghambhar khad, Seer khad, Sukhar khad, Sarhali khad and Lunkar khad. The map of the Satluj basin is shown in Figure - 3.1.

#### **3.2 TOPOGRAPHY OF THE SATLUJ BASIN**

The total geographical area of Satluj river up to Bhakra dam is about 56,980 km<sup>2</sup> of which about 37,153 km<sup>2</sup> lies in Tibet. The rest about 19,827 km<sup>2</sup> area lies in the Indian territory. Indian part of the Satluj basin is elongated in shape. The shape and location of this basin is such that major part of the basin area lies in the greater Himalayas where heavy snowfall is experienced during winters. The catchment area lies between longitudes 76° 22' E to 78° 42' E and latitudes 31° 13' N to 32° 23' N.

Major portion of the Satluj basin lies in the greater Himalayan range. The elevation of the catchment varies widely from about 500 m to 7000 m above msl, although only a very small area exists above 6000 m above msl. Mean elevation of the basin is about 3600 m above msl. The gradient is very steep at its source and gradually reduces downstream. Digital elevation map of area is shown in Figure – 3.2. Owing to large differences in seasonal temperatures and great range of elevation in the catchment, the snowline is highly variable, descending to an elevation of about 2000 m during winter. In present study, the analysis has been focused for the area lying in India territory. Altitude of the rainfall/discharge stations, considered in the study, varies from around 500 m to about 3640 m above msl as presented in Table - 3.1. Salient features of the whole Satluj catchment are summarised in Table 3.2.



**Figure – 3.1:** Index map of the Satluj River basin up to Bhakra dam

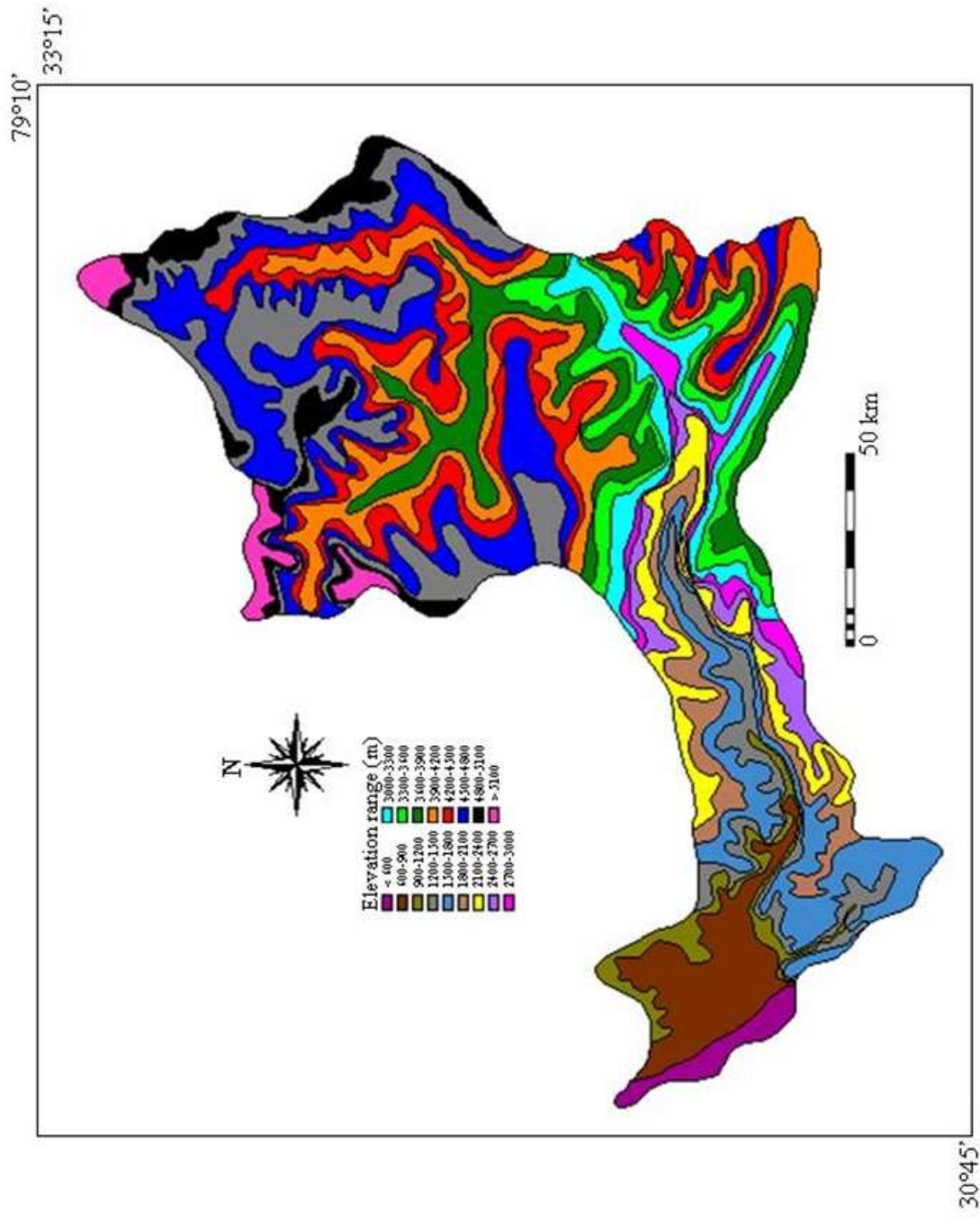


Figure – 3.2: Digital elevation map of Satluj Basin up to Bhakra dam (Indian part only)

**Table - 3.1** Altitude of different stations in Satluj basin

S. No.	Name of station	Altitude (m)
1	Berthin	500 (approx)
2	Bhakra Dam	518
3	Kaza	3639
4	Namagia	2910
5	Rakchham	3130
6	Rampur	1066
7	Suni	625
8	Kalpa	2439
9	Kasol	661
10	Kahu	649

**Table - 3.2** Salient features of Satluj catchment

Reach	Catchment area (km <sup>2</sup> )	Elevation range (m)	Average annual Rainfall (mm)	Major contribution to streamflow
Tibetan plateau	37050	4000-6000	Nil	Snow and glacier
Spiti Valley	7084	3300-5300	Scarce	Snow and glacier
Namgia to Rampur	6490	3000-4800	Little	Snow and rainfall
Rampur to Suni	2068	1200-3000	1000-1500	Rainfall
Suni to Kasol	700	900-2000	910-1630	Rainfall
Kasol to Bhakhra	3108	600-2000	1520	Rainfall

### 3.3 CLIMATIC CONDITIONS IN SATLUJ BASIN

The great contrast in the topographical relief results in variety of climate in the Himalayas. Principal controls producing such differences are those of altitude, local relief and mountain barrier effect. The most important factors controlling the weather and climate in the Himalayas are the altitude and aspect. Largely due to variations in altitude, the climate varies from hot and moist tropical climate in lower valleys, to cool temperate climate at about 2000 m and tends towards polar climate as the altitude increases beyond 2000 m. Depending upon broad climatic conditions the following four seasons prevail over the basin:

***a) Winter season (December - March)***

The precipitation during this season is caused by extratropical weather system of mid-latitude region (also called western disturbances) originating from Caspian sea and moving eastward through Iran, Afghanistan and Pakistan. The precipitation during this season is generally in the form of snow in the greater Himalayas, snow and rain in the middle Himalayas, and light to moderate rain over outer Himalayas and north Indian plains.

***b) Pre-monsoon season (April - June)***

Generally, pre-monsoon season lasts for about 3 months from April to June and is considered as transit period between winter and south-west monsoon. Light to moderate rains are essentially caused by air mass convective storms. Convection increases because of increasing trend of temperature in the Himalayan region in this season.

***c) Monsoon season (July - September)***

The moist air currents from the Bay of Bengal cause precipitation over the Himalayas in this season. This is the season of abundant rain and rivers are generally flooded. Snow and glaciers at very high altitudes continue melting during this season. The monsoon normally starts withdrawing from this region towards the end of September.

***d) Post-monsoon season (October- November)***

During this season, clear autumn weather sets in and there is generally little rainfall. This is the driest season in the entire Himalayas as well as in the plain areas.

Altitude controls not only temperature and climate in the basin but also the rainfall. As shown in Table 3.2, some parts of the catchment receive heavy rainfall during the monsoon season from July to mid September. Sometimes, monsoon season extends up to late September. The average rainfall in the catchment is about 1140 mm. The Satluj runoff basically consists of two parts, one part is derived from the melting of the snow and the other results from the rainfall in the catchment. The monsoon is generally marked by high river flows and occasional floods in the Satluj river.

### **3.4 STREAMFLOW CHARACTERISTICS OF SATLUJ**

The streamflow of the Satluj river consists of the contribution from rain, snow and glaciers and respective contribution of each component varies with time of the year.

Generally, snowmelt contribution starts from March and lasts until June/July depending upon the snowpack water equivalent accumulated in the preceding winter season and prevailing temperatures in the summer season. As the summer season progresses, the snowmelt contribution increases continuously and exceeds the rainfall component. Thus, in the pre-monsoon season (April-June), a major part of streamflow is generated from seasonal snow. During monsoon season (July-September), monsoon rains producing higher discharges in the river augment flow. Generally, high discharges/floods are observed in the months of July and August and these are essentially due to heavy rain in the lower part of the basin. Usually, the end of May/June ablates seasonal snow accumulated on glaciers during winter season and glaciers start contributing to streamflow thereafter. Glaciers contribute to their maximum in the months of July and August. As such, glacier-melt runoff contribution lasts till September/October. In the post-monsoon season, streamflow is believed to be partly from the glaciers and some occasional rain events. Minimum streamflow is observed during winters because no melting takes place due to lower temperature regime.

### 3.5 DATA USED IN THE STUDY

#### 3.5.1 Rainfall data

The daily rainfall data for the Satluj basin have been collected for ten raingauge stations, i.e. Berthin, Bhakra dam, Kahu, Kasol, Kaza, Namagia, Rakchham, Rampur, Suni, and Kalpa. However, it is observed that Kaza station in the Spiti sub-basin lies at very high altitude as compared to the other raingauge stations with most of the precipitation occurring as snow. Because of the lack of other nearby station in the vicinity of Kaza station, this station was not included for further study. The period of the availability of the data at each raingauge station is given in Table 3.3.

**Table - 3.3** Availability of daily rainfall data in Sutluj basin

<b>S. No.</b>	<b>Name of raingauge station</b>	<b>Period of Data Availability</b>
1	Berthin	01.01.1977-30.06.1998
2	Bhakra Dam	01.01.1977-30.06.1998
3	Kaza	01.01.1984-31.12.2000
4	Namagia	01.01.1983-30.06.1998
5	Rakchham	01.01.1985-31.12.2000
6	Rampur	01.01.1962-31.12.2000
7	Suni	01.10.1976-30.06.1998
8	Kalpa	01.01.1984-31-12-2000
9	Kasol	01.01.1977-30.06.1998
10	Kahu	01.01.1977-30.06.1998



### 3.5.2 Discharge data

Daily discharge data at Bhakra, Namagia and Rampur have been collected for the period as mentioned in Table 3.4. Keeping in view, that the Namagia gauging site lies at very high altitude (2910 m above msl) and there is appreciable snowmelt contribution between Namagia and Rampur gauging stations, the discharge data of Rampur and Bhakra have been used and runoff generation in the intermediate catchment between Rampur and Bhakra have been compared with the rainfall record.

**Table - 3.4** Availability of discharge data for Satluj basin

<b>S. No.</b>	<b>Name of gauging station</b>	<b>Period of data available</b>
1	Namagia	01.01.1977-31.12.2000
2	Bhakra	01.01.1963-31.12.2000
3	Rampur	01.01.1980-31.12.2000

\* \* \*

## CHAPTER - 4

# DATA PROCESSING AND VALIDATION

### 4.1 ENTRY OF DATA

Data entry is the first task before the processing of data on computer using any software. For the Satluj basin, three types of data have been input to the computer: a) historical rainfall data, b) historical discharge data, and c) basin boundary of the intermediate catchment between Rampur and Bhakra gauging sites. The rainfall data has been entered in the *SWDES* system while the discharge data and the GIS related basin map have been entered in the *HYMOS* system. The entry of data for each hydrological variable is described below.

#### 4.1.1 Entry of rainfall data

To facilitate the entry of rainfall data, the *SWDES* software developed for the purpose has been utilized. First, a new workarea (Bhakra) is created and registered in the software. After creation, the workarea is opened from the initial 'Login' screen and the data entry for the Bhakra basin is started.

Historical rainfall data of ten raingauge stations of Satluj basin, namely Berthin, Bhakra, Kahu, Kasol, Kaza, Namagia, Rakchham, Rampur, Suni, and Kalpa, were entered in the system. Before the entry of the time series rainfall data, it is first required to specify the static and semi-static characteristics of the stations and the associated time series. Attributes associated with the observational stations or equipment which do not change with time are considered as static type of data. Some of these attributes change but very infrequently and are thus taken to be of semi-static nature. Entry of static/semi-static characteristics is explained below:

##### *a) Entry of station characteristics*

Before entry of time series data for any station, it is required to create the station and enter its static characteristics. Many important attributes, such as name, code, drainage unit, administrative unit, location details, establishment details, and station categorization are assigned to the observational station. The station characteristics are entered by choosing the option of "Static/Semi-static Characteristics" from the main switchboard and then choosing the menu item "Station Characteristics". Ten stations were created in the *SWDES* system

specifying only the name, code (same as name), latitude, longitude, altitude, establishment details, and station category. This information was sufficient enter the time series data for the station.

#### ***b) Entry of series characteristics***

The bulk of hydrological and hydro-meteorological data is time series data. At every station a number of variables are observed and sometimes at varying time intervals. Thus, the time series data is required to be organised in different series at every station for each combination of the required variables and time intervals of observation. After the creation of stations, characteristics of the time series, such as type of measurement, time interval, equidistant or non-equidistant observations, time of observation, various data limits etc. are specified. These attributed are useful for identification and providing necessary information about the series and in validation of the elements of the series. Any time series is specified by its series identification code which comprise of three parts: station code, data type and time interval code. The combination of these three entities is unique and defines a specific series.

For the ten rain gauge stations selected in this study, series characteristics were defined. The data type for each station was taken as MPS (Meteorological-Precipitation-Standard rain gauge) and the time interval was selected as ‘daily’.

#### ***c) Import of time series data***

After establishing the station and series characteristics for different rain gauges, the time series data for each station were imported in *SWDES*. The daily rainfall data for ten raingauge stations in the Satluj basin (collected from the Bhakra Beas Management Board by the Mountain Hydrology Division of NIH) were used. The data were available in computer in some specified format. Using a small computer program, the data were converted in the format required by the *SWDES* system for direct import (a small sample shown in Table – 4.1). After converting the data in required format, the data series of particular stations were pasted in the database of corresponding stations in MS ACCESS.

After importing the data of a station for the whole period in *SWDES* database, the same is displayed in the monthly forms. In these forms, column showing cumulative rainfall is automatically filled up. At the bottom of the form, the number of rainy days, total and maximum rainfall for the month as available in the manuscript has to be entered. The

computed values of these quantities as per the entries by the user are automatically filled. Layout of a monthly form for entry or editing of data is shown in Figure – 4.1.

**Table 4.1** Format for transfer of daily rainfall data at Berthin through MS ACCESS

Station	Data Type	Year	Month	Day	RF	Remarks
Berthin	MPA	1977	1	1	0.00	
Berthin	MPA	1977	1	2	0.00	
Berthin	MPA	1977	1	3	0.00	
Berthin	MPA	1977	1	4	0.00	
Berthin	MPA	1977	1	5	0.00	
Berthin	MPA	1977	1	6	0.00	

The screenshot displays the 'Rainfall - Daily' form in the SW-DES application. The form is for station 'Bhakra' in January 1985. It features a data table with the following columns: Dt., Rainfall at 08:00 Hrs., Cum. Rainfall, and Remarks. The data table shows rainfall values for 18 days, with a total of 57.1 mm. Summary statistics at the bottom indicate 5 rainy days and a maximum daily rainfall of 41.5 mm. The interface includes a menu bar (File, Edit, View, Insert, Format, Records, Tools, Window, Help) and a toolbar with various icons for data manipulation and reporting.

Dt.	Rainfall at 08:00 Hrs.	Cum. Rainfall	Remarks
1	41.5	41.5	
2	0.0	41.5	
3	0.0	41.5	
4	0.0	41.5	
5	6.3	47.8	
6	0.0	47.8	
7	0.0	47.8	
8	1.4	49.2	
9	0.0	49.2	
10	0.0	49.2	
11	0.0	49.2	
12	0.0	49.2	
13	0.0	49.2	
14	0.0	49.2	
15	0.0	49.2	
16	0.0	49.2	
17	0.0	49.2	
18	0.0	49.2	

As entered	As in form	Remarks in case of mismatch
Total Rainfall (Mth)	57.1	0.0
Rainy Days (Mth)	5	0
Max. Daily Rainfall	41.5	0.0

**Figure 4.1** Form for entry or editing of daily rainfall data

The number of rainy days, maximum rainfall, and total rainfall in the month were checked with the values available in the original manuscript. At the places of mismatch, the required corrections were made according to the data in original manuscript.

#### **4.1.2 Entry of discharge data**

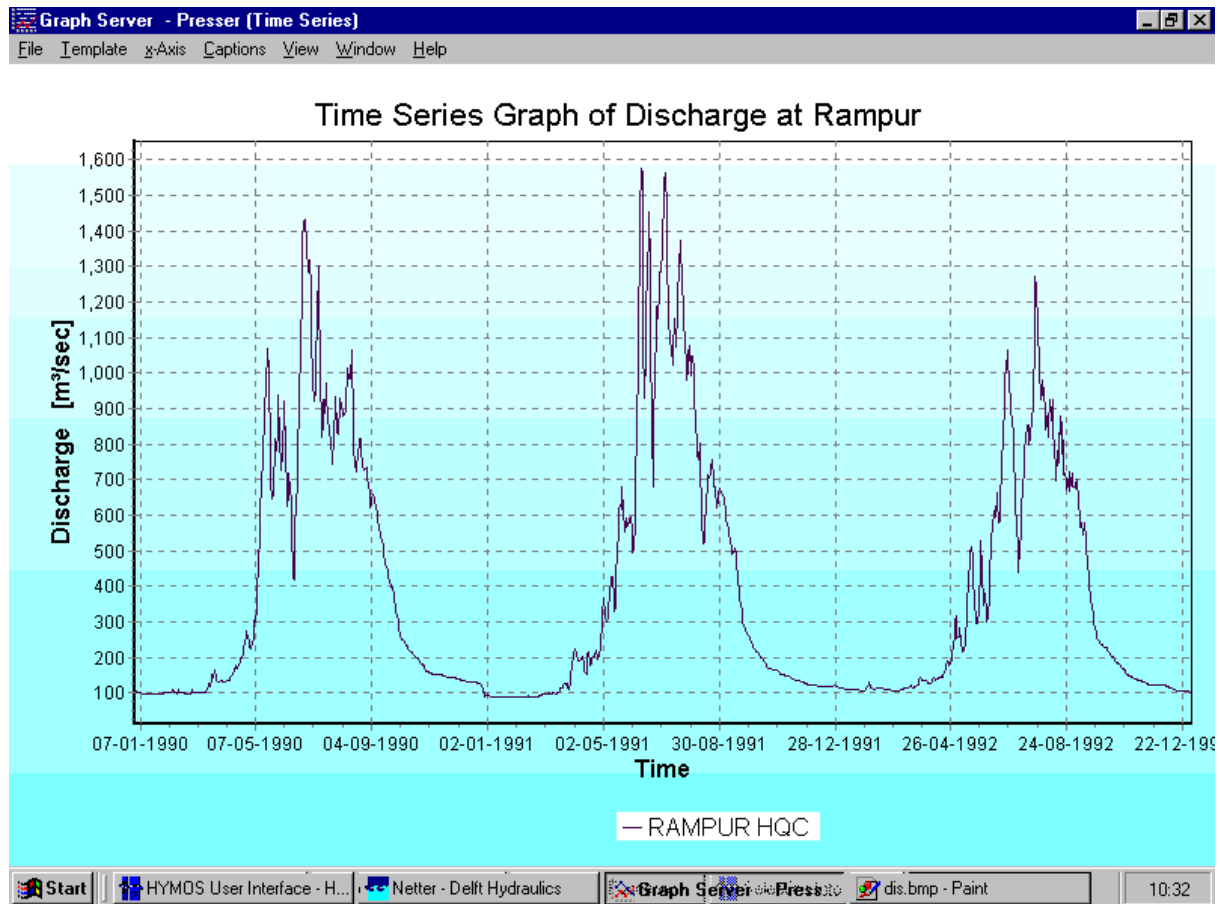
Daily discharge data at Namagia, Rampur and Bhakra were available. However, Namagia gauging site is located at a very high altitude and there is no other site located in its vicinity which can be used for the validation of the discharge data at Namagia. Further, Rampur gauging site is located quite far away from Namagia and there is appreciable unknown intermediate flow (from rain and snowmelt) between Rampur and Namagia. Hence, the Namagia data were not included in the analysis. The discharge data of Bhakra and Rampur were used and the runoff from the intermediate catchment between Rampur and Bhakra were correlated with the average rainfall in the intermediate basin. Discharge data at Rampur and Bhakra were used for the period 1980-1997.

*HYMOS* database has an option to include the time series data of all those instruments which have an associated observation time, whether the measurement is of an instantaneous value (e.g. water level), an accumulative value (e.g. daily rainfall), a constant value (e.g. a gate overflow level setting), an averaged value (e.g. mean monthly discharge) or a statistic from a specified time period (e.g. daily maximum temperature, annual maximum flow). The discharge data were directly imported in the *HYMOS* system as there was no separate module for the entry of discharge data directly in the *SWDES*. The gauging stations were first created in the *SWDES* by defining the station characteristics and series characteristics. As these characteristics are exported from *SWDES* and imported in *HYMOS*, the location of station in the map layer in *HYMOS* appears. The time series data of the station is first prepared in the required format using a computer program as is done for the rainfall data. Using the “Entry and Edit” option in *HYMOS*, the time series data of a station is opened and the data is pasted in the *HYMOS* form and saved. The entered time series can be projected in graphical form for visualization and inspection. A view of the graphical presentation of discharge data in *HYMOS* is shown in Figure – 4.2.

#### **4.1.3 Entry of GIS layers**

The physical characteristics of a catchment that are of hydrological relevance and which can be derived from *GIS* and stored in *HYMOS* include its boundary, its geographical area, the layout of the river network and topographical features. Such geographical information is generally available in the toposheets from which they can be digitised. Each element of the space-oriented data is referenced by its position using a coordinated system

referred to latitude and longitude. Such geographical data are organised in different map layers so that it is possible to use one or more such layers at any time.



**Figure 4.2** A view of graphical presentation of discharge data in *HYMOS*

In the present case, the river network of the Satluj basin was obtained from the Digital Chart of the World (DCW Clipper). The basin boundary for the intermediate catchment between Rampur and Bhakra gauging sites was digitised and imported as a map layer in the *HYMOS* Netter. The basin boundary is used for generating the Thiessen polygon map of the intermediate catchment and finding the weights of different rainfall stations for calculating the average rainfall. A view showing the map layers of the Satluj basin (basin boundary, intermediate catchment, drainage network) and the location of different stations is presented in Figure 4.3.



***a) Comparison of time series for manual and autographic data***

At stations where rainfall is measured at short durations using an autographic or a digital recorder, a standard raingauge is also always available. Thus, at such observation stations rainfall data at daily time interval is available from two independent sources. The rainfall data at hourly or smaller interval is aggregated at the daily level and then a comparison is made between the two. Differences less than 5% can be attributed to exposure, instrument accuracy and precision in tabulating the analogue records and are ignored. Any appreciable difference (more than 5%) between the two values is probed further. The observation made using standard raingauge is generally taken as comparatively more reliable.

***b) Checking against maximum and minimum data limits***

Rainfall data must be validated against limits within which it is expected to physically occur. Such limits are required to be quite wide to avoid the possibility of rejecting true extreme values. For rainfall data, it is obvious that no data can be less than zero which perfectly serves as the limiting minimum value. However, it is quite difficult to assign an absolute maximum limit for the rainfall data in a given duration occurring at a particular station. Nevertheless, on the basis of past experience and physical laws governing the process of rainfall it is possible to arrive at such maximum limits which in all probability will not be exceeded. For 1-day duration, the India Meteorological Department and Indian Institute of Tropical Meteorology have prepared atlas for 1-day Probable Maximum Precipitation (PMP) which gives the expected maximum amount that can physically occur in a given duration at a given location. Values extracted from this map can be applied or else could be derived as the historical maximum value from the long term records.

***c) Checking against upper warning limit***

Validation of rainfall data against an absolute maximum value does not discriminate those comparatively frequently occurring erroneous data which are less than the prescribed maximum limit. In view of this, it is advantageous to consider one more limit, called the upper warning level, which can be employed to see if any of the data value has violated it. This limit is assigned a value with an intention of screening out those high data values which are not expected to occur frequently. The underlying purpose of carrying out such a test is to consider a few high data values with suspicion and subsequently scrutinize them.



For the daily rainfall data this limit can be set statistically, for example, equal to 99 percentile value of the actual rainfall values excluding zero rainfall values. The central idea while setting these upper warning levels is that the higher rainfall data is screened adequately, that is the limits must be such that it results in neither too many nor too less data values being flagged for validation.

Primary validation of data is carried out at the lower level office and this activity, in general, goes parallel with the entry of data. In the present study, the data was available on the magnetic media and the original manuscripts of data were available for a very short span of time (2-3 years). Hence, more emphasis has been given to the application of secondary validation techniques. The secondary validation puts most emphasis on comparisons with neighboring stations.

Several Specific tasks, which have been carried out for the daily rainfall data at different stations in the Bhakhra basin, are discussed below:

#### **4.2.1 Normalization of rainfall**

The topography of the Satluj basin is hilly. The major portion of the Satluj lies in the greater Himalayan range with the elevation variation from 500 m to 7000 m. The altitudes of different raingauge stations also vary greatly as specified in Table 3.1. Such an altitude difference can have orographic effect on the occurrence of rainfall and its distribution in the catchment. To correct for sources of heterogeneity, such as orographic effects, normalized rather than actual values may be used. So, for making comparisons of the rainfall data of neighboring stations, it is necessary to normalize the data. To normalize the data at a neighbor stations, its data are multiplied by a coefficient, obtained by the ratio of the test station normal and the neighbor station normal:

$$P_{ci} = (N_{test}/N_i).P_i \quad \dots(4.1)$$

where,

$P_{ci}$  = Corrected normal rainfall at neighbor station

$N_{test}$  = Normal rainfall (annual average rainfall) at test station

$N_i$  = Normal rainfall (annual average rainfall) at neighbor station

$P_i$  = Observed rainfall at the neighbor station

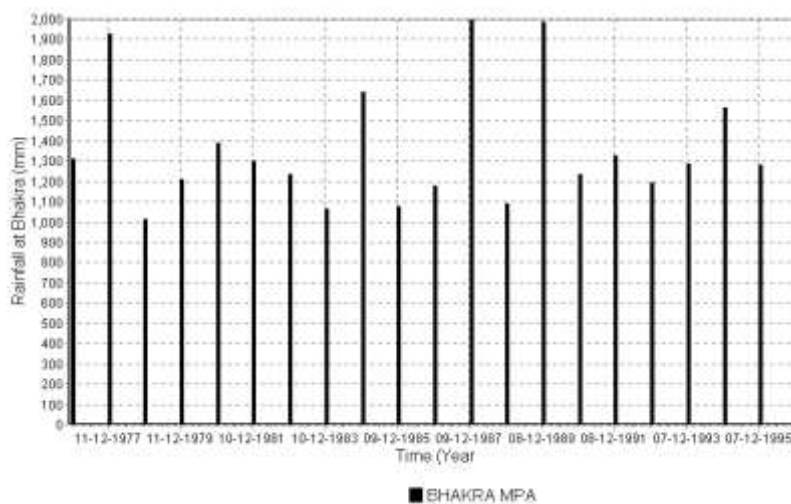
To find the normalise values at any raingauge station (known as test station), in comparision to the other neighboring stations (also known as base stations), it is necessary to get the normal rainfall (average annual rainfall) of the test/base stations. Average annual rainfall at different stations were obtained by aggregating the daily rainfall to annual values and then by finding the average of the number of years of annual values. The data series were aggregated from daily to monthly and from monthly to annual using the “Aggregation of series” option of *HYMOS* and the average of all annual values was obtained by using the “Basic statistics” option of *HYMOS*.

**a) Aggregation**

Aggregation of series implies the creation of a series with a larger time interval by adding or averaging data of the series with the smaller interval. For example, aggregation of series with an interval of one day to a series with an interval of one month, or from month to year. In aggregating series two cases are distinguished:

1. for instantaneous observations, the aggregated data are average of the originals, e.g. discharges in m<sup>3</sup>/s.
2. for accumulative observations, the aggregated data are the sum of the originals, e.g. rainfall or runoff in mm.

Since the rainfall is an accumulative observation, aggregated values are find by summing the individual day’s values. A graph showing the derived annual series for Bhakra rain gauge station is shown in Figure – 4.4.



**Figure 4.4** Graph showing the derived annual series for Bhakra rain gauge station

**b) Basic statistics**

The following statistics and distributions are computed in this option: minimum, maximum, mean, median, mode, standard deviation, skewness, kurtosis, empirical frequency distribution, and empirical cumulative frequency distribution. An output of the basic statistics for the daily data of Bhakra rain gauge station is shown in Table – 4.2. The average annual rainfall for the various raingauge stations in the Satluj basin, which are used for normalizing the data for different stations, are shown in Table – 4.3.

**Table 4.2** Basic Statistics of daily rainfall at Bhakra raingauge station

Series code	=	<b>Bhakra</b>	<b>MPS</b>
First year	=	1977	
Last year	=	1996	
Actual values are used			
<b>Annual period considered: from 1-1-1977 to 1-1-1997</b>			
Lower Threshold	=	0.010	
Upper Threshold	=	2009.321	
<b>Basic Statistics of series Bhakra MPS</b>			
Mean	=	1365.1840	Median = 1282.4540
Mode of classes	=	1260.2980	Standard deviation = 303.9843
Variance	=	92406.4500	Skewness = 1.1934
Kurtosis	=	3.4198	Range = 1010.6240 to 2009.3210
25% Quantile	=	1182.1830	75% Quantile = 1520.9800
Number of elements	=	20	
<b>95% Confidence Intervals</b>			
1222.9160 < Mean < 1507.4530			
27433.1600 < Variance < 109732.7000			
<b>Decile</b>	<b>Value</b>	<b>Chegadayevev formula</b>	
1	1043.914000	1055.233000	
2	1110.494000	1123.144000	
3	1177.074000	1191.055000	
4	1230.337000	1239.526000	
5	1270.285000	1280.272000	
6	1310.233000	1328.210000	
7	1376.813000	1396.121000	
8	1609.842000	1671.761000	
9	1942.741000	1964.712000	
<b>Cumulative frequency distribution and histogram</b>			
<b>Upper class limit</b>	<b>Cum. Probability</b>	<b>Probability</b>	<b>Number of elements</b>
1010.624000	.050000	.034314	1.
1110.494000	.200000	.181373	3.
1210.364000	.350000	.328431	3.
1310.233000	.600000	.573529	5.
1410.103000	.750000	.720588	3.
1509.972000	.750000	.720588	0.
1609.842000	.800000	.769608	1.
1709.712000	.850000	.818627	1.
1809.581000	.850000	.818627	0.
1909.451000	.850000	.818627	0.
2009.321000	1.000000	.965686	3.

**Table 4.3** Average annual rainfall for stations in Bhakra basin

S. No.	Name of Station	Average annual rainfall (mm)
1	Berthin	1224.7
2	Bhakra Dam	1365.2
3	Kaza	132
4	Namagia	133.2
5	Rakchham	212.8
6	Rampur	810.1
7	Suni	985.2
8	Kalpa	417.4
9	Kasol	1331.1
10	Kahu	1267.1

After calculating average annual rainfall for all stations in the basin, coefficients for finding normal rainfall for each base stations with different test stations were obtained seperately. The coefficients have been obtained using the following formula:

$$\text{Coefficient} = \frac{\text{Normal Rainfall for Test Station}}{\text{Normal Rainfall for Base Station}}$$

It is well known fact that as the distance between the rainfall stations increases, their spatial correlation decreases. Based on the spatial correlation analysis (explained in the next section), the different rainfall stations in the Satluj basin have been clustered in three groups: i) Berthin, Bhakra, Kahu, Kasol and Suni, ii) Kalpa, Namagia, Rakchham, and iii) Rampur and Suni. The coefficients obtained for each base station in set number (i) and (ii) for different neighboring stations are shown in Table - 4.4 and Table – 4.5 respectively. For the third set, the normalization coefficient for Suni station, considering Rampur as test station, comes out to be 0.8222.

**Table 4.4** Normalization coefficients for Bhakra, Berthin, Kasol, Suni, Kahu

Test Station	Base Station				
	Berthin	Bhakra	Suni	Kasol	Kahu
Berthin	1.0000	0.8971	1.2431	0.9200	0.9665
Bhakra	1.1147	1.0000	1.3876	1.0256	1.0774
Suni	0.8044	0.7216	1.0000	0.7401	0.7775
Kasol	1.0869	0.9750	1.3512	1.0000	1.0505
Kahu	1.0347	0.9282	1.2862	0.9519	1.0000

**Table 4.5** Normalization coefficients for Kalpa, Namagia, Rakchham

Test Station	Base Stations		
	Kalpa	Namagia	Rakchham
Kalpa	1.0000	3.1325	1.9618
Namagia	0.3192	1.0000	0.6263
Rakchham	0.5097	1.5968	1.0000

After getting the coefficients for each station corresponding to other test station, the normal rainfall for each station have been obtained. To obtain the daily normal values of rainfall for the raingauge stations the actual daily rainfall of each station have been multiplied with the coefficient of each station.

$$\text{Normal daily Rainfall} = \text{coefficient of normalization} \times \text{daily rainfall of the Station}$$

#### 4.2.2 Spatial correlation analysis

Rainfall poses special problems for spatial comparisons because of limited or uneven correlation between stations. When rainfall is convective 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 unevenness tends to be smoothed out and aggregated totals are much more closely correlated. Spatial correlation in rainfall thus depends on duration (smaller at shorter duration), distance (decrease with distance), type of precipitation, and physiographic characteristics of a region.

Before applying different validation checks, it is necessary to identify various cluster of stations that show good spatial correlation based on the historical rainfall data. To find such combinations of stations, spatial correlation analysis is carried out. The two correlation functions, the correlation and the semivariance, give an indication of similarities in mechanisms causing a phenomenon to exhibit. Both functions use the co-ordinates of the selected series in their calculation together with the measured values.

Functions that describe the relation between inter-station correlation and inter-station distance, the so-called cross-correlation functions, can be determined empirically from a graph or calculated by least squares fitting. The objective is to find the relationship between inter-station correlation and inter-station distance. There are many types of functions, i.e.

higher degree functions, power functions, exponential functions or Bessel functions. *HYMOS* uses the power function for calculation of the correlation distance:

$$r(h) = r(0) * \exp(-h/h_0) + a \quad \dots(4.2)$$

where  $h$  is the distance between two stations,  $h_0$  is the correlation distance,  $a$  is the threshold in function, and  $r(0)$  is the correlation at distance zero. Spatial correlation analysis can be carried out in *HYMOS* using the “Spatial correlation” option.

In this study, spatial correlation matrix has been derived for different stations. The matrix shows the correlation coefficient among the stations. For deriving the matrix, original data of one test station and normalized data of other station has been considered and correlation coefficients have been determined. Analysis has been carried out for the daily and monthly time steps. Different stations have been taken as test stations one by one and corresponding correlation coefficients have been found out. Five different combinations of stations have been tried on the basis of their altitude above mean sea level and intermediate distance. These combinations are:

1. Berthin, Bhakra, Kahu, Kasol and Suni
2. Kalpa, Rakchham, Rampur, and Suni
3. Kalpa Namagia, Rakchham, and Rampur
4. Kahu, Kasol, Rampur and Suni
5. Kalpa, Kaza, Namagia, Rakchham and Rampur

It is found that the correlation of Kaza is very poor ( $<0.50$ ) with all other stations, due to different topographical/physiographic conditions of the area near the station as compared to other rainfall stations. Because of poor spatial correlation of Kaza with all other available stations, the same has not been considered for further validation processes in this study. From the results, it was found that the following sets exhibit good spatial correlation:

1. Berthin, Bhakra, Kahu, Kasol, Suni
2. Kalpa, Namagia, Rakchham,
3. Rampur and Suni

The correlation matrices for these combination of stations for daily and monthly time steps are presented in Table – 4.6 and Table – 4.7 respectively.

**Table 4.6** Correlation coefficients matrix for daily time steps

Test station having original rainfall data series	Correspondingly base stations having normalized daily rainfall data								
	Berthin	Bhakra	Kahu	Kasol	Suni	Rampur	Namagia	Rakchham	Kalpa
Berthin	1.000	0.709	0.656	0.648	0.569	NA	NA	NA	NA
Bhakra	0.709	1.000	0.576	0.549	0.501	NA	NA	NA	NA
Kahu	0.656	0.576	1.000	0.806	0.589	NA	NA	NA	NA
Kasol	0.648	0.549	0.806	1.000	0.642	NA	NA	NA	NA
Suni	0.569	0.501	0.589	0.642	1.000	0.578	NA	NA	NA
Rampur	NA	NA	NA	NA	0.578	1.000	NA	NA	NA
Namagia	NA	NA	NA	NA	NA	NA	1.000	0.236	0.498
Rakchham	NA	NA	NA	NA	NA	NA	0.236	1.000	0.535
Kalpa	NA	NA	NA	NA	NA	NA	0.498	0.535	1.000

**Table 4.7** Correlation coefficients matrix for monthly time steps

Test station having original rainfall data series	Correspondingly base stations having normalized monthly rainfall data								
	Berthin	Bhakra	Kahu	Kasol	Suni	Rampur	Namagia	Rakchham	Kalpa
Berthin	1.000	0.922	0.872	0.896	0.834	NA	NA	NA	NA
Bhakra	0.922	1.000	0.855	0.867	0.801	NA	NA	NA	NA
Kahu	0.872	0.855	1.000	0.919	0.837	NA	NA	NA	NA
Kasol	0.896	0.867	0.919	1.000	0.836	NA	NA	NA	NA
Suni	0.834	0.801	0.834	0.836	1.000	0.804	NA	NA	NA
Rampur	NA	NA	NA	NA	0.804	1.000	NA	NA	NA
Namagia	NA	NA	NA	NA	NA	NA	1.000	0.477	0.699
Rakchham	NA	NA	NA	NA	NA	NA	0.477	1.000	0.656
Kalpa	NA	NA	NA	NA	NA	NA	0.699	0.656	1.000

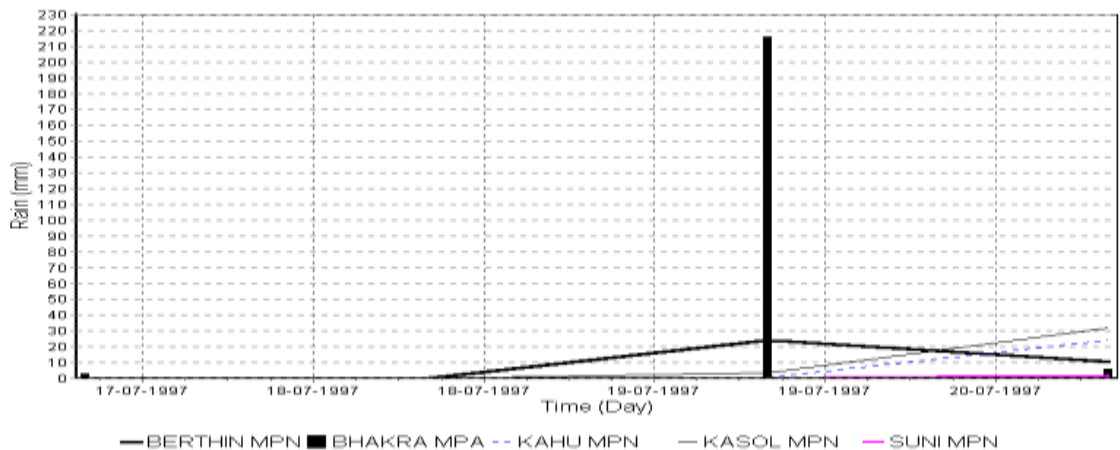
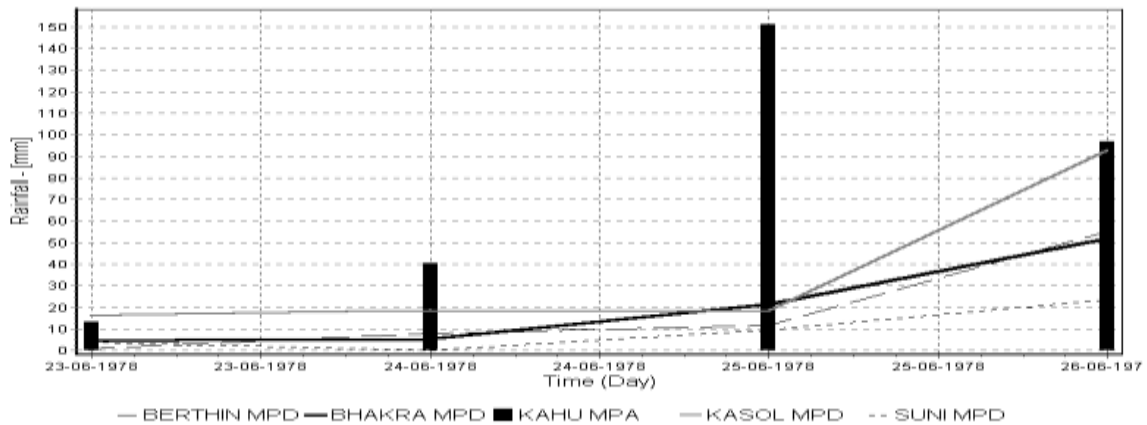
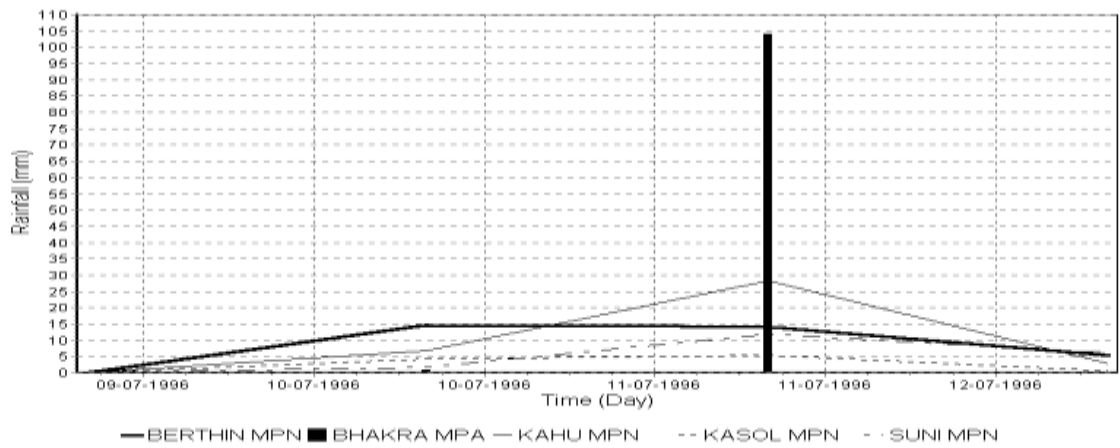
### 4.2.3 Scrutiny by multiple time series plots

Inspection of multiple time series graphs is used as an alternative to inspection of tabular data. Visual inspection is found to be 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 duration.

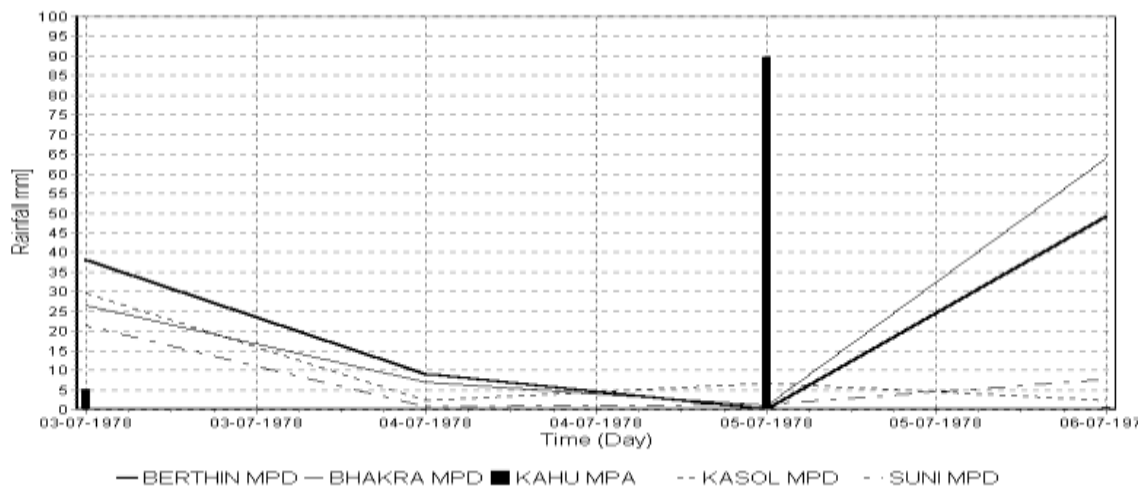
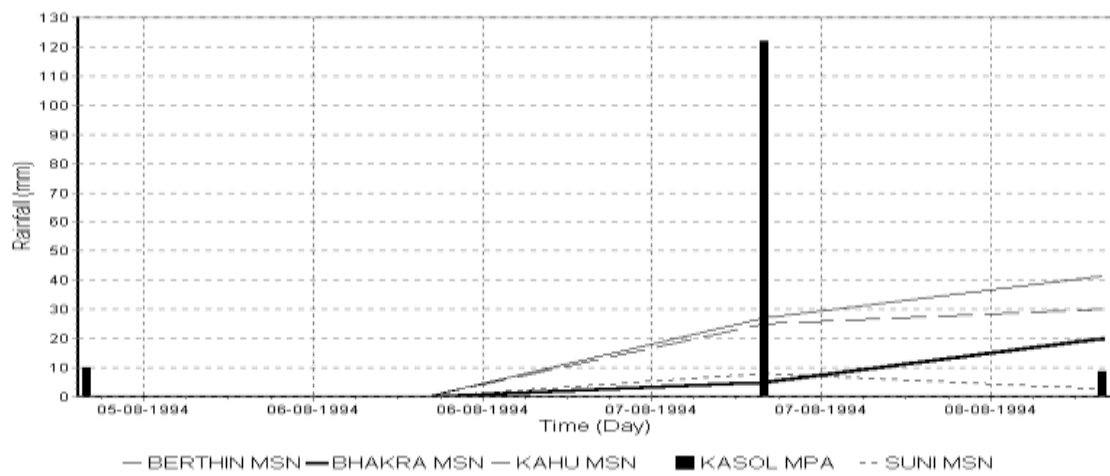
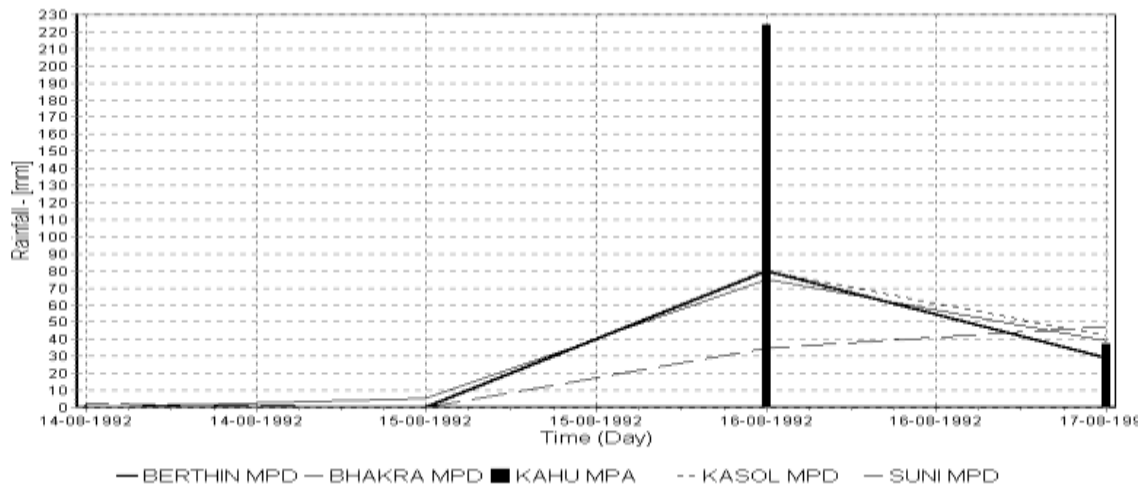
Under the process, a set of stations within a small area with adequate spatial correlation is chosen. Rainfall series of stations are plotted as bars stacked side by side and preferably in different color for each station. Efficient comparison on the magnitudes of rainfall at different stations is possible 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.

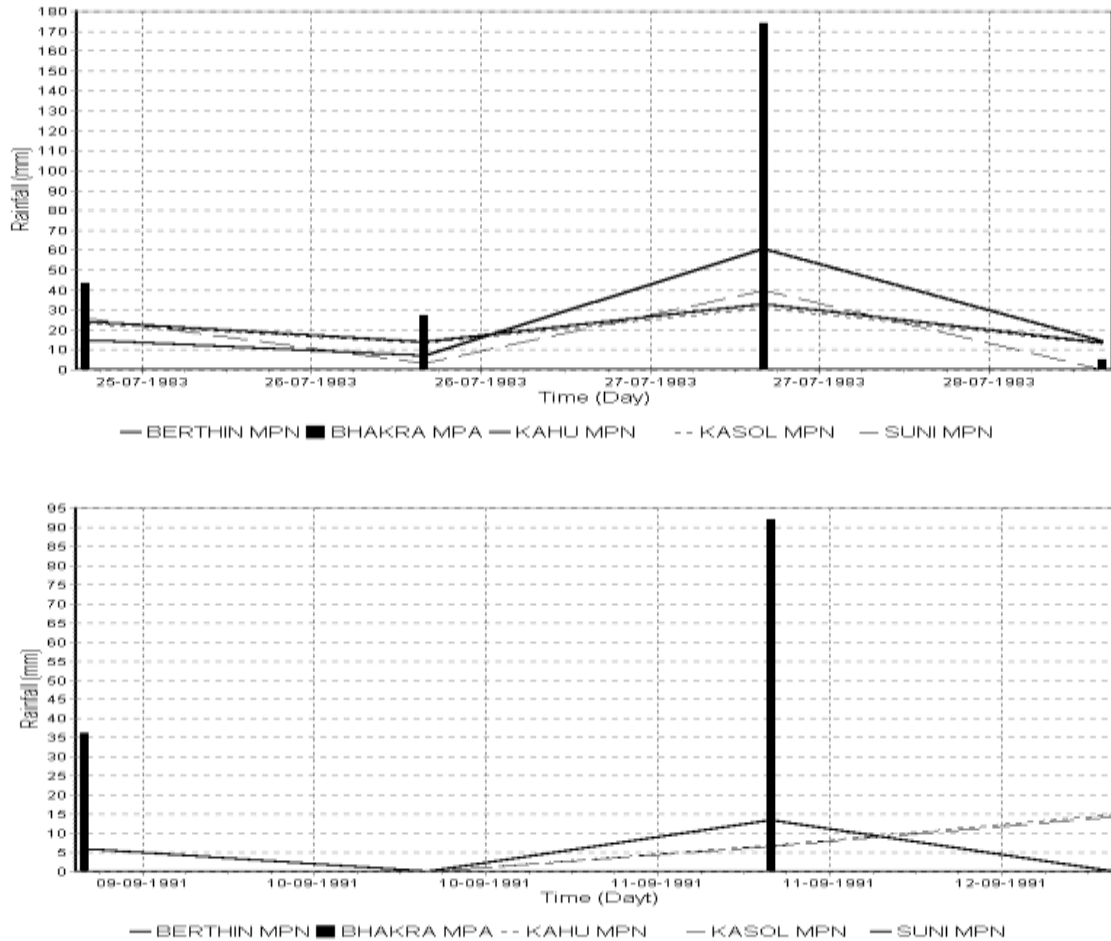
Using the “Time series graphs” option of *HYMOS*, plots of multiple stations of a set were made and some inconsistencies were observed. These have been presented in Figure –

4.5 (a-h). The corresponding station name and dates have also been mentioned in Table – 4.8. The inspection of these graphs shows that very high rainfall values have been observed at one station though other surrounding stations show moderate rainfall. Though, in a mountainous terrain, it may be possible, as the rainfall amounts on the leeward and windward side vary appreciably, yet there is a need to scrutinize these values. A scrutiny of some suspect values with the original record (presented later) confirms some inconsistent entries in the original dataset which demonstrates the necessity and usefulness of data validation techniques.









**Figure 4.5(a-h)** Multiple time series plots of rainfall at Bhakra, Kahu, Kasol, Berthin, and Suni

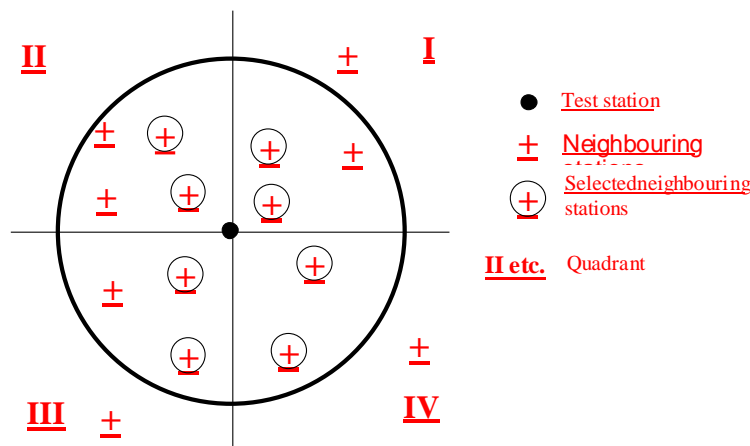
#### 4.2.4 Spatial homogeneity test

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.

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 its neighboring stations. Ideally, the stations selected as neighbors 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 (Figure 4.6).

- (a) the distance between the test and the neighboring station must be less than the specified maximum correlation distance.
- (b) a maximum of 8 neighboring stations can be considered for interpolation.
- (c) to reduce the spatial bias, it is appropriate to consider a maximum of only two stations within each quadrant.



**Figure 4.6** Definition sketch of test and base (neighbouring) stations

**Table 4.8** Suspect rainfall values and corresponding dates identified from multiple time series plots of rainfall data

Date	Rainfall (mm)				
	Berthin	Bhakra	Kahu	Kasol	Suni
11.07.96	12.60	<b>104.00</b>	26.30	5.20	11.00
25.06.78	11.10	23.20	<b>151.20</b>	19.00	7.50
19.07.97	21.90	<b>216.00</b>	0.50	3.50	0.00
16.08.92	77.00	80.70	<b>224.00</b>	84.00	27.00
07.08.94	24.80	5.00	24.00	<b>121.90</b>	6.00
05.07.78	0.00	1.20	<b>89.60</b>	7.00	1.00
27.07.83	54.40	<b>174.00</b>	30.40	30.40	37.00
11.09.91	12.00	<b>92.00</b>	6.40	6.40	0.00

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

$$P_{est}(t) = \frac{\sum_{i=1}^N P_i(t) / D_i^b}{\sum_{i=1}^N 1 / D_i^b} \quad \dots(4.3)$$

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

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

$$|P_{obs}(t) - P_{est}(t)| \leq X_{abs} \quad \dots(4.4)$$

$$|P_{obs}(t) - P_{est}(t)| \leq X_{rel} * S_{P_{est}(t)} \quad \dots(4.5)$$

where:

- $X_{abs}$  = admissible absolute difference
- $S_{P_{est}(t)}$  = standard deviation of neighbouring values
- $X_{rel}$  = multiplier of standard deviation given by the following formula:

$$S_{P_{est}(t)} = \sqrt{\sum_{i=1}^N (P_i(t) - \bar{P}_i(t))^2} \quad \dots(4.6)$$

When the departures between the estimated and observed values 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. This value may also 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  $X_{rel} \times S$  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 present study, spatial homogeneity test was carried out for the available rainfall data of different stations. Five rain gauge stations, namely Bhakra, Kahu, Berthin, Kasol, and Suni, that have good spatial correlation, were selected. The Spatial homogeneity test for each rain gauge station, considering as test station has been applied. Table 4.9 shows the results of the spatial homogeneity test taking Kasol as the test station and Bhakra, Kahu, Berthin, and Suni as the surrounding stations. Using the spatial homogeneity test for different stations, periods have been identified when the rainfall occurrence at a single station is high as compared to other stations. These are presented in Table – 4.10. Though, it is quite possible to have uneven spatial pattern of rainfall especially in a mountainous region like Himalayas, there is a need to scrutinize the corresponding data.

**Table 4.9** Spatial homogeneity test for Kasol using Bhakra, Berthin, Kahu & Suni

Test station Kasol		MPS							
Start date:	1977 1 1 0 1								
End date:	1997 12 31 0 1								
Radius of circle of influence	:	50.000	(km)						
Station weights proportional to	:	1/D <sup>2.00</sup>							
Admissible absolute error	:	50.000							
Multiplier to stdv of neighbours:		2.000							
Selected neighbour stations:									
Quadrant	Station	Distance (km)							
2	Berthin	MPS	33.171						
2	Bhakra	MPS	42.957						
3	Kahu	MPS	21.605						
4	sun i	MPS	27.626						
Year	mth	day	hr	si	P_obs	flag	P_est	Stdv	n
1977	5	10	0	1	48.00	+	17.52	9.53	4
1977	6	24	0	1	27.20	+	.07	.17	4
1977	7	28	0	1	45.80	+	5.27	1.75	4
1977	8	4	0	1	117.50	+	66.12	23.61	4
1977	8	10	0	1	38.50	+	5.31	8.83	4
1977	8	31	0	1	40.00	+	1.93	.98	4
1977	9	3	0	1	22.00	+	3.22	2.53	4
1977	9	5	0	1	33.00	+	3.40	3.19	4
1977	9	10	0	1	42.00	+	13.51	7.92	4
1977	9	18	0	1	36.50	+	12.03	6.74	4
1978	6	25	0	1	19.00	-	72.69	59.72	4
1978	6	28	0	1	118.50	+	12.60	20.34	4
1978	7	20	0	1	31.00	+	4.74	9.78	4
1978	8	2	0	1	33.40	-	92.94	41.95	4
1978	8	3	0	1	12.50	-	61.48	18.23	4
1978	9	14	0	1	30.90	+	4.29	2.70	4
1979	8	4	0	1	50.10	+	17.02	15.23	4
1979	8	9	0	1	5.00	-	61.62	46.66	4
1979	8	14	0	1	51.00	+	12.98	10.35	4

1979	8	15	0	1	29.10	+	3.04	2.72	4
1980	6	14	0	1	6.10	-	36.09	10.48	4
1980	6	15	0	1	34.50	+	2.77	1.94	4
1980	6	21	0	1	38.00	+	3.27	2.51	4
1980	6	28	0	1	35.50	+	8.37	4.47	4
1980	7	18	0	1	52.00	+	.00	.00	4
1980	7	26	0	1	24.00	+	.53	.87	4
1980	7	27	0	1	65.00	+	22.31	5.44	4
1980	8	3	0	1	107.00	+	24.13	8.95	4
1980	8	4	0	1	72.00	+	5.83	5.68	4
1980	8	10	0	1	22.50	+	2.09	1.34	4
1980	8	15	0	1	42.00	+	8.14	5.98	4
1980	9	2	0	1	42.00	+	13.58	10.47	4
1981	7	13	0	1	114.40	+	3.68	3.71	4
1981	7	17	0	1	58.60	+	4.32	4.14	4
1981	7	22	0	1	24.50	+	.45	.67	4
1981	7	23	0	1	60.00	+	1.06	1.43	4
1981	7	26	0	1	38.00	+	6.39	9.67	4
1981	7	29	0	1	66.50	+	22.30	16.93	4
1981	7	30	0	1	70.00	+	15.84	20.84	4
1986	7	7	0	1	205.70	+	137.79	64.03	4
1992	8	8	0	1	40.10	+	12.31	6.70	4
1992	8	13	0	1	24.10	+	1.93	1.64	4
1992	8	28	0	1	102.70	+	35.47	7.95	4
1992	9	7	0	1	40.00	+	8.19	8.20	4
1993	7	12	0	1	16.00	-	57.33	14.37	4
1993	7	17	0	1	56.90	+	22.71	16.88	4
1993	8	6	0	1	40.70	+	8.77	9.17	4
1993	8	7	0	1	50.00	+	11.70	1.78	4
1993	8	23	0	1	42.50	+	4.26	2.91	4
1994	2	22	0	1	36.00	+	8.15	13.21	4
1994	7	20	0	1	65.70	-	135.05	54.15	4
1994	8	7	0	1	121.90	+	17.24	9.46	4
1994	8	14	0	1	49.30	+	14.65	10.39	4
1994	8	28	0	1	34.90	+	2.18	2.17	4
1995	6	19	0	1	8.20	-	53.32	22.50	4
1995	6	20	0	1	3.50	-	29.25	8.98	4
1995	7	16	0	1	37.90	+	2.62	5.53	4
1995	7	30	0	1	46.00	+	12.79	9.88	4
1995	7	31	0	1	141.00	+	36.38	20.60	4
1995	8	3	0	1	48.00	+	5.29	5.79	4
1995	8	11	0	1	46.60	+	13.04	10.53	4
1995	8	19	0	1	52.00	+	.87	1.85	4
1996	6	18	0	1	39.40	+	8.77	7.09	4
1996	6	27	0	1	54.40	+	21.70	13.46	4
1996	7	5	0	1	43.10	+	20.25	10.29	4
1996	7	24	0	1	29.20	+	3.63	3.35	4
1996	7	27	0	1	35.40	+	.43	.65	4
1996	9	14	0	1	24.00	+	4.93	8.10	4
1996	9	15	0	1	25.00	+	1.29	2.79	4
1997	7	28	0	1	50.60	+	9.88	6.09	4
1997	9	7	0	1	124.20	+	25.88	9.76	4

**Table 4.10** Identification of suspect values using spatial homogeneity test

Year	mth	day	Berthin	Bhakra	Kahu	Kasol	Suni
1977	5	12	<b>22.00</b>	0.00	0.00	0.00	0.00
1977	5	26	2.00	0.00	2.20	0.00	<b>28.00</b>
1977	6	15	8.20	0.00	0.00	0.00	<b>36.00</b>
1977	6	20	0.00	<b>24.50</b>	0.00	0.00	0.00
1977	6	24	0.40	0.00	0.00	<b>27.20</b>	0.00
1977	6	30	0.60	0.00	0.00	0.00	<b>51.50</b>
1977	7	7	2.00	<b>35.20</b>	0.00	3.50	0.50
1977	7	15	41.00	43.00	74.00	67.00	<b>6.50</b>
1977	7	17	<b>60.00</b>	0.20	0.00	2.00	0.00
1977	7	19	0.00	1.00	<b>46.00</b>	6.00	0.00
1977	7	28	2.30	4.50	5.60	<b>45.80</b>	7.10
1977	8	27	2.50	8.00	<b>65.40</b>	17.00	11.50
1977	8	31	2.90	0.50	1.40	<b>40.00</b>	2.70
1977	9	1	<b>35.00</b>	2.00	0.00	3.00	0.70
1977	9	5	0.00	0.00	3.00	<b>33.00</b>	7.80
1978	4	15	0.00	0.00	0.00	1.00	<b>90.00</b>
1978	6	20	0.00	0.00	0.00	0.00	<b>40.00</b>
1978	6	25	11.10	23.20	<b>151.20</b>	19.00	7.50
1978	6	28	50.30	1.80	<b>0.00</b>	<b>118.50</b>	11.50
1978	7	1	0.00	0.00	<b>43.10</b>	0.00	2.00
1978	7	5	0.00	1.20	<b>89.60</b>	7.00	1.00
1978	7	8	0.00	0.00	<b>31.50</b>	0.00	0.00
1978	7	9	43.30	<b>112.00</b>	6.80	34.50	34.50
1978	7	13	56.50	<b>0.00</b>	31.40	21.00	21.50
1978	7	28	2.00	0.30	<b>39.20</b>	3.00	7.00
1978	8	12	50.50	41.50	<b>9.00</b>	34.00	52.00
1978	9	2	1.40	0.00	3.80	3.50	<b>31.00</b>
1978	11	12	0.00	<b>52.60</b>	5.20	0.00	4.50
1978	11	13	52.70	<b>0.00</b>	40.00	47.00	40.20
1979	2	19	5.20	4.70	<b>72.40</b>	5.00	1.00
1979	2	20	0.00	<b>53.50</b>	0.50	0.00	0.00
1979	6	26	6.00	0.00	<b>34.00</b>	0.50	0.00
1979	6	29	1.00	0.00	1.50	12.50	<b>42.50</b>
1979	7	19	1.50	<b>70.50</b>	6.60	7.50	3.40
1979	8	9	67.40	2.50	<b>111.50</b>	5.00	0.50
1979	9	15	0.00	2.80	<b>40.00</b>	0.00	0.00
1980	6	15	0.00	0.00	4.20	<b>34.50</b>	3.50
1980	6	20	50.00	30.00	<b>4.00</b>	44.00	53.50
1980	6	25	<b>45.50</b>	0.30	4.00	0.00	0.00
1980	7	7	27.30	25.80	55.80	42.00	<b>3.90</b>
1980	7	17	0.00	0.00	<b>32.50</b>	2.00	3.00
1980	7	18	0.00	0.00	0.00	<b>52.00</b>	0.00
1980	7	19	0.00	1.00	<b>36.40</b>	4.40	0.00
1980	8	3	22.60	7.50	24.00	<b>107.00</b>	32.30
1980	8	4	14.80	12.30	2.80	<b>72.00</b>	1.90
1980	8	6	5.80	<b>65.30</b>	5.50	0.50	0.00
1980	9	19	0.00	<b>34.00</b>	0.00	0.00	2.60
1981	6	28	59.20	<b>0.00</b>	<b>0.00</b>	42.00	<b>130.00</b>
1981	7	13	7.20	0.90	0.00	<b>114.40</b>	8.40
1981	7	14	55.60	<b>102.50</b>	<b>0.00</b>	64.80	32.00
1981	7	15	<b>158.20</b>	48.00	0.00	20.00	25.00
1981	7	17	11.40	7.60	0.00	<b>58.60</b>	5.10
1981	7	20	2.60	<b>62.00</b>	0.00	10.00	2.50
1981	7	23	0.00	1.10	0.00	<b>60.00</b>	3.50
1981	7	29	37.60	36.20	<b>0.00</b>	66.50	42.40
1982	7	20	18.00	<b>58.80</b>	2.00	2.00	11.30
1982	8	16	<b>80.00</b>	5.20	30.40	30.40	1.40
1983	7	2	0.00	<b>41.00</b>	0.00	0.00	0.00
1983	7	14	4.00	<b>39.50</b>	6.50	6.50	13.00

1983	7	27	54.40	<b>174.00</b>	30.40	30.40	37.00
1983	8	7	0.00	<b>52.00</b>	2.00	2.00	15.00
1983	8	10	17.20	<b>48.50</b>	0.00	0.00	0.00
1983	8	19	6.00	<b>38.00</b>	4.00	4.00	2.30
1983	8	24	17.20	<b>86.20</b>	22.00	22.00	<b>0.00</b>
1983	8	26	6.40	<b>64.20</b>	4.00	4.00	4.40
1983	8	30	0.00	<b>25.50</b>	0.00	0.00	0.00
1983	9	9	44.00	<b>86.50</b>	3.00	3.00	1.00
1983	9	10	52.00	<b>147.00</b>	40.80	40.80	18.00
1984	6	11	2.00	<b>43.30</b>	5.00	5.00	3.70
1984	7	3	0.00	<b>43.40</b>	0.00	0.00	3.00
1984	7	5	<b>63.60</b>	0.00	6.00	6.00	0.80
1984	7	17	<b>30.00</b>	0.30	0.00	0.00	1.00
1984	8	2	<b>30.00</b>	0.00	0.00	0.00	3.50
1984	8	7	<b>48.00</b>	6.20	8.00	8.00	0.30
1984	8	8	<b>92.00</b>	2.30	10.00	10.00	0.00
1984	8	25	18.40	<b>91.00</b>	20.00	20.00	3.00
1984	8	27	8.40	<b>83.20</b>	11.40	11.40	1.00
1984	9	1	8.00	<b>43.50</b>	0.00	0.00	2.20
1984	9	5	9.50	<b>62.80</b>	1.40	1.40	0.00
1984	9	19	<b>30.00</b>	0.00	1.00	1.00	2.80
1985	7	18	<b>65.00</b>	10.00	4.00	4.00	4.90
1985	8	3	0.00	0.50	0.00	0.00	<b>21.60</b>
1985	8	17	<b>32.00</b>	0.80	0.00	0.00	0.00
1985	8	21	<b>91.80</b>	<b>0.80</b>	48.40	48.40	32.00
1985	8	22	11.60	7.30	45.50	45.50	<b>124.00</b>
1985	8	25	26.00	<b>98.00</b>	26.00	26.00	<b>2.00</b>
1985	8	30	<b>51.00</b>	0.00	2.00	2.00	0.00
1985	9	4	25.00	<b>0.00</b>	27.30	27.30	10.60
1986	8	9	5.50	1.70	9.30	9.30	<b>42.00</b>
1986	8	15	<b>44.00</b>	5.00	0.00	0.00	0.00
1986	8	22	0.00	<b>45.50</b>	0.00	0.00	2.50
1986	9	28	3.00	<b>64.20</b>	6.80	6.80	4.00
1986	12	12	<b>32.00</b>	0.50	0.00	0.00	0.00
1987	5	18	4.50	11.20	1.00	1.00	<b>39.60</b>
1987	6	9	<b>41.00</b>	12.40	20.00	20.00	9.00
1987	6	10	<b>0.00</b>	61.00	18.40	18.40	56.50
1987	7	23	5.00	<b>40.00</b>	0.00	0.00	0.00
1987	8	14	25.00	<b>91.50</b>	38.20	38.20	14.40
1987	8	19	1.00	<b>25.30</b>	0.00	0.00	0.00
1987	8	27	7.70	<b>64.00</b>	4.00	4.00	0.00
1987	9	1	8.70	<b>33.20</b>	0.00	0.00	0.00
1988	7	15	2.60	<b>31.00</b>	3.60	3.60	2.20
1988	8	11	<b>36.80</b>	0.00	0.00	0.00	0.00
1988	9	26	<b>80.00</b>	22.50	30.50	30.50	10.00
1989	5	31	8.20	17.20	6.30	6.30	<b>50.00</b>
1989	8	1	8.50	<b>55.00</b>	10.50	10.50	3.00
1989	9	19	0.00	<b>60.40</b>	9.50	9.50	0.00
1990	5	12	<b>41.00</b>	1.00	3.00	3.00	0.00
1990	7	15	<b>3.40</b>	47.00	56.00	56.00	57.00
1990	7	25	2.40	<b>30.20</b>	1.50	1.50	0.00
1990	7	28	31.80	<b>113.00</b>	36.00	36.00	8.50
1990	7	29	<b>70.80</b>	2.00	18.50	18.50	5.00
1990	8	15	0.00	<b>44.50</b>	0.00	0.00	0.00
1990	8	17	<b>2.20</b>	24.40	56.00	56.00	48.50
1990	8	28	15.90	<b>80.00</b>	8.50	8.50	0.00
1990	9	4	40.60	<b>123.60</b>	24.80	24.80	0.00
1990	9	19	<b>112.00</b>	13.80	55.20	55.20	20.20
1991	2	27	<b>60.20</b>	22.40	16.50	16.50	6.50
1991	7	15	7.40	<b>36.40</b>	1.00	1.00	0.00
1991	8	8	0.00	<b>58.00</b>	0.00	0.00	0.00



1991	8	18	<b>59.80</b>	15.00	0.00	0.00	0.00
1991	8	23	27.00	<b>62.20</b>	15.40	15.40	21.40
1991	9	2	<b>40.20</b>	8.80	0.00	0.00	0.00
1991	9	9	5.40	<b>36.20</b>	0.00	0.00	0.00
1991	9	11	12.00	<b>92.00</b>	6.40	6.40	0.00
1992	7	6	40.00	30.30	46.00	46.00	<b>4.40</b>
1992	7	29	64.00	29.40	33.20	33.20	<b>0.00</b>
1992	8	2	24.80	<b>70.70</b>	32.80	32.80	<b>0.00</b>
1992	8	16	77.00	80.70	<b>224.00</b>	84.00	27.00
1992	8	18	20.20	<b>62.40</b>	0.00	0.00	1.20
1992	8	26	46.40	<b>117.30</b>	38.00	51.20	7.00
1992	8	28	44.60	22.30	34.60	<b>102.70</b>	36.00
1992	8	29	<b>55.90</b>	9.00	2.50	10.00	2.00
1992	9	6	0.20	<b>22.50</b>	0.00	0.00	0.00
1993	7	21	0.00	7.00	<b>72.50</b>	0.00	0.00
1993	7	24	18.00	<b>66.00</b>	8.50	6.20	2.40
1993	8	3	4.00	6.00	<b>42.20</b>	0.20	6.00
1993	8	7	9.50	13.80	11.00	<b>50.00</b>	13.50
1993	8	8	1.50	<b>37.80</b>	0.00	0.90	0.00
1993	8	23	2.50	2.70	8.00	<b>42.50</b>	0.00
1993	8	25	0.00	0.00	0.00	0.00	<b>26.20</b>
1994	8	7	24.80	5.00	24.00	<b>121.90</b>	6.00
1994	8	10	20.40	<b>67.50</b>	15.60	7.90	0.00
1994	8	28	0.00	0.00	5.00	<b>34.90</b>	0.00
1995	4	11	3.30	4.20	10.00	11.00	<b>44.00</b>
1995	5	11	<b>26.00</b>	0.00	1.50	0.00	0.00
1995	6	28	0.60	<b>55.20</b>	0.00	14.90	1.40
1995	7	7	25.00	49.30	26.80	25.00	<b>0.00</b>
1995	7	8	11.00	<b>60.00</b>	22.40	3.70	2.20
1995	7	31	0.60	18.50	56.00	<b>141.00</b>	36.50
1995	8	3	1.00	2.00	2.00	<b>48.00</b>	15.00
1995	8	9	1.80	<b>55.00</b>	8.00	6.50	19.20
1995	8	13	<b>135.00</b>	53.50	29.00	7.50	12.60
1995	8	17	<b>68.40</b>	25.00	0.00	7.00	0.00
1995	8	19	2.00	4.50	0.00	<b>52.00</b>	0.00
1995	8	22	10.00	<b>34.70</b>	9.20	2.40	6.60
1995	9	1	0.00	0.00	0.00	0.20	<b>27.50</b>
1996	4	24	6.50	4.50	3.20	0.00	<b>38.80</b>
1996	7	4	9.20	<b>55.50</b>	14.00	2.50	0.00
1996	7	11	12.60	<b>104.00</b>	26.30	5.20	11.00
1996	7	27	1.60	0.00	0.00	<b>35.40</b>	0.50
1996	7	31	47.90	<b>121.70</b>	62.40	2.50	2.20
1996	8	11	9.00	<b>72.00</b>	14.50	2.60	0.00
1996	8	15	10.60	14.00	<b>61.40</b>	37.00	6.70
1996	9	18	0.00	<b>42.00</b>	0.00	0.00	0.00
1997	6	20	0.30	2.00	13.00	4.30	<b>60.60</b>
1997	7	8	20.00	7.00	<b>93.50</b>	34.40	0.00
1997	7	19	21.90	<b>216.00</b>	0.50	3.50	0.00
1997	9	6	0.00	23.40	<b>51.50</b>	2.50	0.00
1997	9	7	42.00	36.00	16.40	<b>124.20</b>	26.00
1997	9	8	16.80	15.00	<b>89.00</b>	18.60	3.00

Some of the major suspected values (as mentioned above) were sent to the concerned department for scrutiny from the original records (Table-4.11). Some of the suspected values were found to be wrongly entered in the database (showing the usefulness of data validation exercise) while in other cases, values of stations other than suspect were found different.

**Table - 4.11 Verified & Corrected suspect values from nodal agency (BBMB)**

Year	mth	day	Berthin	Bhakra	Kahu	Kasol	Suni
1977	6	30	<b>6.00</b>	0.00	0.00	0.00	51.50
1977	7	17	60.00	0.20	0.00	2.00	<b>8.20</b>
1977	7	19	0.00	1.00	46.00	6.00	<b>13.10</b>
1977	7	28	2.30	4.50	5.60	45.80	<b>0.10</b>
1977	8	27	2.50	8.00	65.40	17.00	11.50
1977	8	31	2.90	0.50	1.40	40.00	2.70
1978	4	15	0.00	0.00	0.00	1.00	90.00
1978	6	20	0.00	0.00	0.00	0.00	40.00
1978	6	25	11.10	23.20	<b>40.50</b>	19.00	7.50
1978	6	28	50.30	1.80	<b>97.00</b>	118.50	11.50
1978	7	1	0.00	0.00	<b>0.00</b>	0.00	2.00
1978	7	5	0.00	1.20	<b>0.00</b>	7.00	1.00
1978	7	9	43.30	112.00	<b>31.50</b>	34.50	34.50
1978	11	12	0.00	<b>0.00</b>	5.20	0.00	4.50
1979	2	19	5.20	4.70	<b>5.80</b>	5.00	1.00
1979	2	20	0.00	53.50	<b>0.00</b>	0.00	0.00
1979	6	29	1.00	0.00	1.50	12.50	42.50
1979	7	19	1.50	70.50	6.60	7.50	3.40
1979	8	9	67.40	2.50	111.50	5.00	0.50
1979	9	15	0.00	2.80	40.00	0.00	0.00
1980	6	25	45.50	0.30	4.00	0.00	0.00
1980	7	7	27.30	25.80	55.80	42.00	3.90
1980	7	17	0.00	0.00	32.50	2.00	3.00
1980	7	18	0.00	0.00	0.00	52.00	0.00
1980	7	19	0.00	1.00	36.40	4.40	0.00
1980	8	3	22.60	7.50	24.00	107.00	32.30
1980	8	4	14.80	12.30	2.80	72.00	1.90
1980	8	6	5.80	65.30	5.50	0.50	0.00
1981	6	28	59.20	0.00	0.00	42.00	<b>26.00</b>
1981	7	13	7.20	0.90	<b>4.50</b>	114.40	8.40
1981	7	15	158.20	48.00	<b>64.20</b>	20.00	25.00
1981	7	17	11.40	7.60	<b>11.40</b>	58.60	5.10
1981	7	20	2.60	62.00	<b>29.00</b>	10.00	2.50
1981	7	23	0.00	1.10	<b>26.50</b>	60.00	3.50
1983	7	2	0.00	41.00	0.40	0.00	0.00
1983	7	27	54.40	174.00	<b>13.00</b>	30.40	37.00
1983	8	7	0.00	52.00	<b>0.50</b>	2.00	<b>0.00</b>
1983	8	24	17.20	86.20	<b>20.00</b>	22.00	0.00
1983	8	26	6.40	64.20	<b>13.50</b>	4.00	4.40
1983	8	30	0.00	25.50	<b>10.50</b>	0.00	0.00
1984	6	11	2.00	43.30	<b>3.80</b>	5.00	3.70
1984	7	3	0.00	43.40	<b>1.80</b>	0.00	3.00
1984	7	5	63.60	0.00	<b>36.00</b>	6.00	0.80
1984	8	7	48.00	6.20	<b>38.10</b>	8.00	0.30
1984	8	8	92.00	2.30	<b>6.80</b>	10.00	0.00
1984	8	25	18.40	91.00	<b>18.00</b>	20.00	3.00
1984	8	27	8.40	83.20	<b>0.50</b>	11.40	1.00
1984	9	1	8.00	43.50	<b>15.70</b>	0.00	2.20
1984	9	5	9.50	62.80	<b>0.70</b>	1.40	0.00
1985	7	18	65.00	10.00	<b>19.80</b>	4.00	4.90
1985	8	3	0.00	0.50	0.00	0.00	21.60
1985	8	17	32.00	0.80	<b>0.50</b>	0.00	0.00
1985	8	21	91.80	0.80	<b>11.30</b>	48.40	32.00
1985	8	22	11.60	7.30	<b>11.00</b>	45.50	124.00
1985	8	25	26.00	98.00	<b>44.70</b>	26.00	2.00
1985	8	30	51.00	0.00	<b>20.00</b>	2.00	0.00
1985	9	4	25.00	0.00	<b>10.60</b>	27.30	10.60
1986	8	9	5.50	1.70	<b>15.90</b>	9.30	42.00
1986	8	15	44.00	5.00	<b>18.50</b>	0.00	0.00

1986	8	22	0.00	45.50	3.50	0.00	2.50
1986	9	28	3.00	64.20	<u>2.00</u>	6.80	4.00
1986	12	12	<u>0.00</u>	0.50	0.00	0.00	0.00
1987	5	18	4.50	11.20	<u>6.40</u>	1.00	39.60
1987	6	10	<u>41.00</u>	61.00	<u>44.20</u>	18.40	56.50
1987	7	23	5.00	40.00	0.00	0.00	0.00
1987	8	14	25.00	91.50	<u>71.00</u>	38.20	14.40
1987	8	19	1.00	25.30	<u>14.00</u>	0.00	0.00
1987	8	27	7.70	64.00	<u>24.00</u>	4.00	0.00
1987	9	1	8.70	33.20	0.00	0.00	0.00
1988	8	11	36.80	<u>3.00</u>	0.00	0.00	0.00
1988	9	26	80.00	<u>122.50</u>	<u>32.60</u>	30.50	10.00
1989	5	31	8.20	17.20	<u>9.00</u>	6.30	50.00
1989	8	1	8.50	55.00	<u>60.00</u>	10.50	3.00
1989	9	19	0.00	60.40	<u>10.20</u>	9.50	0.00
1990	5	12	41.00	1.00	<u>20.60</u>	3.00	0.00
1990	7	15	3.40	47.00	<u>0.00</u>	56.00	57.00
1990	7	28	31.80	113.00	<u>35.20</u>	36.00	8.50
1990	7	29	70.80	2.00	<u>66.60</u>	18.50	5.00
1990	8	15	0.00	44.50	0.00	0.00	0.00
1990	8	28	15.90	80.00	<u>0.00</u>	8.50	0.00
1990	9	4	40.60	123.60	<u>30.20</u>	24.80	0.00
1990	9	19	112.00	13.80	<u>132.60</u>	55.20	20.20
1991	2	27	60.20	22.40	<u>16.00</u>	16.50	6.50
1991	7	15	7.40	36.40	<u>14.00</u>	1.00	0.00
1991	8	8	0.00	58.00	0.00	0.00	0.00
1991	8	18	59.80	15.00	<u>20.00</u>	0.00	0.00
1991	8	23	27.00	62.20	<u>6.80</u>	15.40	21.40
1991	9	2	40.20	8.80	0.00	0.00	0.00
1991	9	9	5.40	36.20	0.00	0.00	0.00
1991	9	11	12.00	92.00	<u>0.00</u>	6.40	0.00
1992	8	16	77.00	80.70	224.00	84.00	27.00
1992	8	18	20.20	62.40	0.00	0.00	1.20
1992	8	26	46.40	117.30	38.00	51.20	7.00
1992	8	28	44.60	22.30	34.60	102.70	36.00
1992	8	29	55.90	9.00	2.50	10.00	2.00
1992	9	6	0.20	22.50	<u>0.00</u>	0.00	0.00
1993	7	21	0.00	7.00	<u>2.00</u>	0.00	0.00
1993	7	24	18.00	66.00	3.50	6.20	2.40
1993	8	7	9.50	13.80	11.00	50.00	13.50
1994	8	10	20.40	67.50	15.60	7.90	0.00
1994	8	28	0.00	0.00	5.00	34.90	0.00
1995	7	8	11.00	60.00	22.40	3.70	2.20
1995	7	31	0.60	18.50	56.00	141.00	36.50
1995	8	3	1.00	2.00	2.00	48.00	15.00
1995	8	9	1.80	55.00	8.00	6.50	19.20
1995	8	13	135.00	53.50	29.00	7.50	12.60
1995	8	19	2.00	4.50	0.00	52.00	0.00
1995	8	22	10.00	34.70	9.20	2.40	6.60
1996	4	24	6.50	4.50	3.20	0.00	38.80
1996	7	4	9.20	55.50	14.00	2.50	0.00
1996	7	11	12.60	104.00	26.30	5.20	11.00
1996	8	11	9.00	72.00	14.50	2.60	0.00
1996	8	15	10.60	14.00	61.40	37.00	6.70
1996	9	18	0.00	42.00	0.00	0.00	0.00
1997	6	20	0.30	2.00	13.00	4.30	60.60
1997	7	8	20.00	7.00	93.50	34.40	0.00
1997	7	19	21.90	216.00	0.50	3.50	0.00
1997	9	7	42.00	36.00	16.40	124.20	26.00
1997	9	8	16.80	15.00	89.00	18.60	3.00

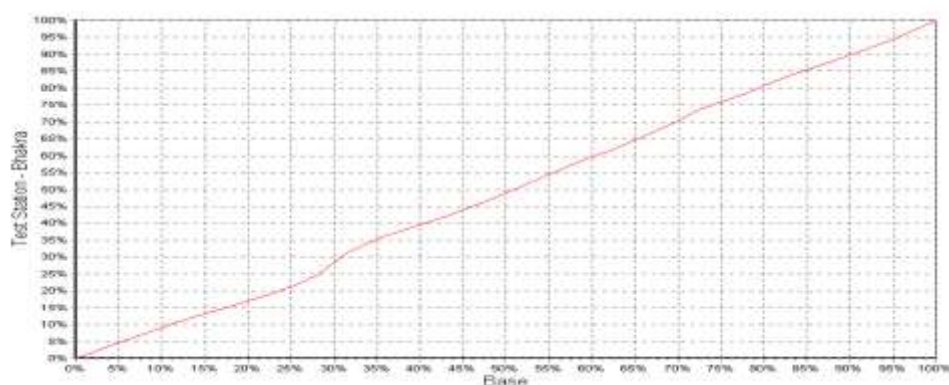
#### 4.2.5 Double mass analysis

Double mass analysis is a technique that is effective in detecting systematic shifts, 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 that may exist for a considerable period of time. Inconsistencies present in the rainfall data of a station can occur for various reasons such as follows:

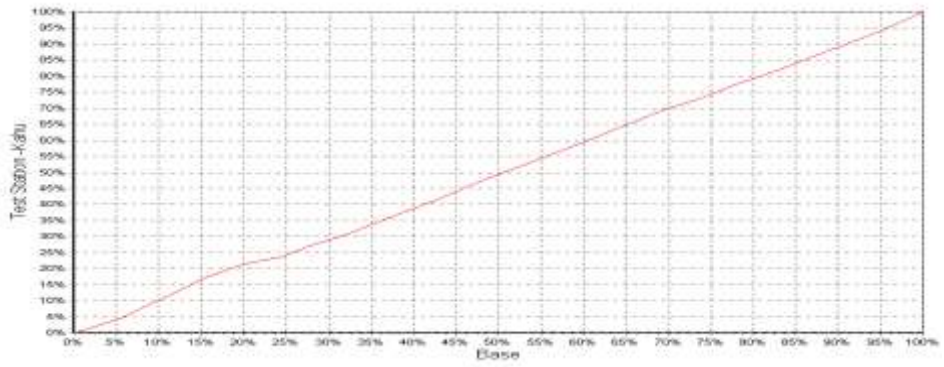
- The raingauge might have been installed at different sites in the past,
- The exposure conditions of the gauge may have undergone a significant change due to the growth of trees or construction of buildings in its proximity,
- There might have been unaccounted change in the raingauge from 125 mm to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

This technique is used to detect possible non-homogeneity in series by investigating the ratio of accumulated values of two series, viz the series to be tested, and the base series. The result of the analysis is presented in a table or in a plot known as the double mass curve. This curve shows a straight line if the test series is homogeneous. A jump in the test-series will create a break in the double mass curve, whereas a trend will create a curved line.

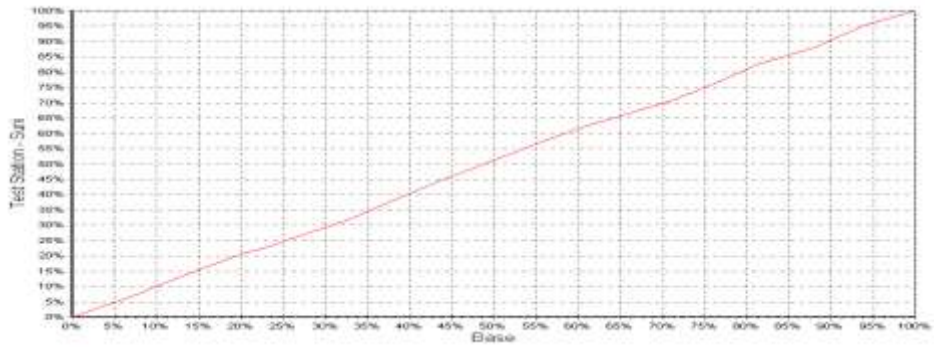
In the present study, data series of all raingauge stations have been tested for non-homogeneity using surrounding stations as the base. The double mass curves for Bhakra, Berthin, Kahu, Kasol, and Suni have been presented in Figure – 4.7 (a-e). From the figures, it is clear that Berthin, Kahu, Kasol, and Suni show homogeneous series while inconsistency is observed in the rainfall of Bhakra station from the year 1983 onwards and in the Kasol data from the year 1981 onwards.



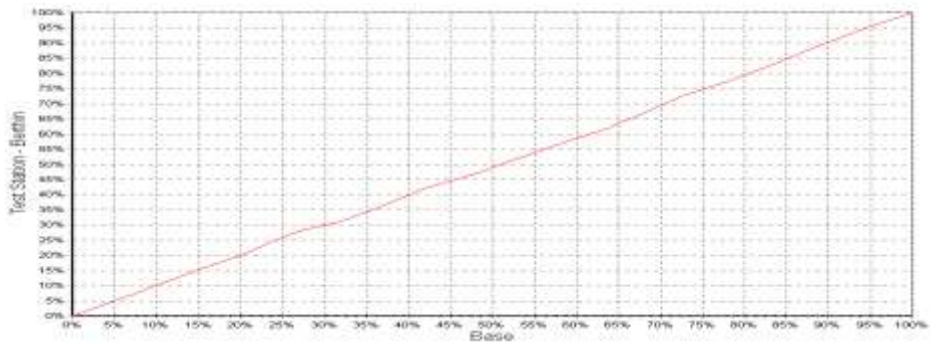
(a)



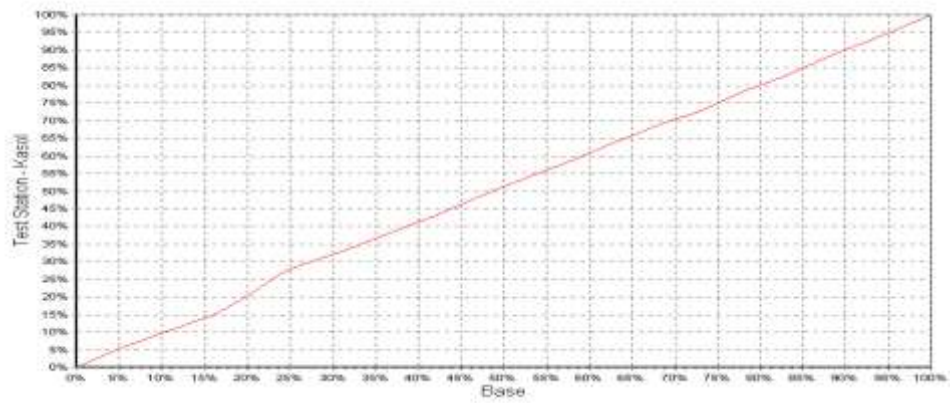
(b)



(c)



(d)



(e)

Figure 4.7(a-e) Double mass curve plot for Bhakra, Kahu, Kasol, Berthin, and Suni

The inconsistency observed in the Bhakra and Kasol data using double mass analysis was referred to BBMB. It was informed that Kasol site was shifted in the year 1990 rather than in the year 1981. However, it is mentioned that exposure conditions might have undergone some change due to growth of trees or construction activity in the proximity of these stations.

#### 4.2.6 Computation of areal rainfall

*HYMOS* offers a number of options to compute catchment rainfall from point rainfall data. There are various methods to find the weights given to the point rainfall stations. These include: equal station weights, user specified station weights, and weights according to Thiessen polygon method. In the present study, weights according to the Thiessen polygon method have been calculated and used to find the areal rainfall.

##### 4.2.6.1 Computation of Thiessen weights

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. Stations outside the basin boundary are also included in the analysis as they may have polygons extending 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 rainfall by their Thiessen weights and then summing the weighted rainfall of all stations as follows:

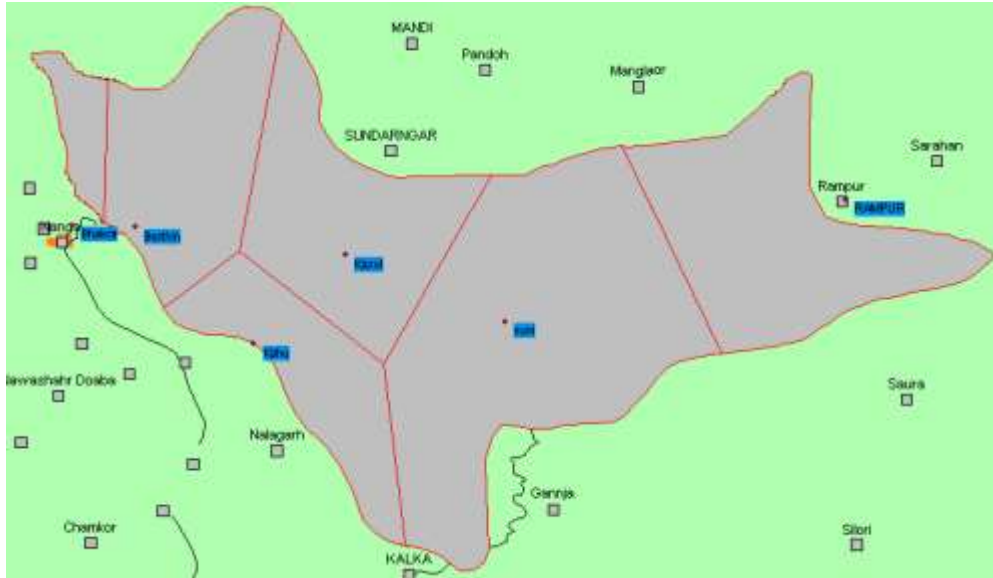
$$P_{at} = \frac{A_1}{A} P_{1t} + \frac{A_2}{A} P_{2t} + \frac{A_3}{A} P_{3t} + \dots + \frac{A_N}{A} P_{Nt} = \sum_{i=1}^N \left( \frac{A_i}{A} \right) P_{it}$$

where:

$A_i$  = the area of Thiessen polygon for station  $i$

$A$  = total area under consideration

The station weights are worked out in the *HYMOS* using the “Areal Thiessen” option. The polygons can be plotted as GIS layer also. The layout of the Thiessen polygons as worked out by the system for the basin between Rampur to Bhakra is shown in Figure 4.8. Corresponding station weights are given in Table 4.12.



**Figure 4.8** Thiessen polygons of rainfall stations for Satluj basin between Rampur & Bhakra

**Table 4.12** Results of Thiessen polygons analysis

<b>Raingauge Station</b>	<b>Thiessen Weight</b>
Berthin	0.150006
Bhakra	0.029276
Kahu	0.107034
Kasol	0.168517
Rampur	0.203142
Suni	0.342025

These weights were saved as a relation in the *HYMOS* database. These weights were convoluted with the daily rainfall data of the different stations involved in the Thiessen polygon derivation and the resultant average rainfall series for the basin in-between Bhakra and Rampur gauging sites was derived. This average daily rainfall for the intermediate basin between Bhakra and Rampur were compared with the discharge from the intermediate catchment as discussed below.

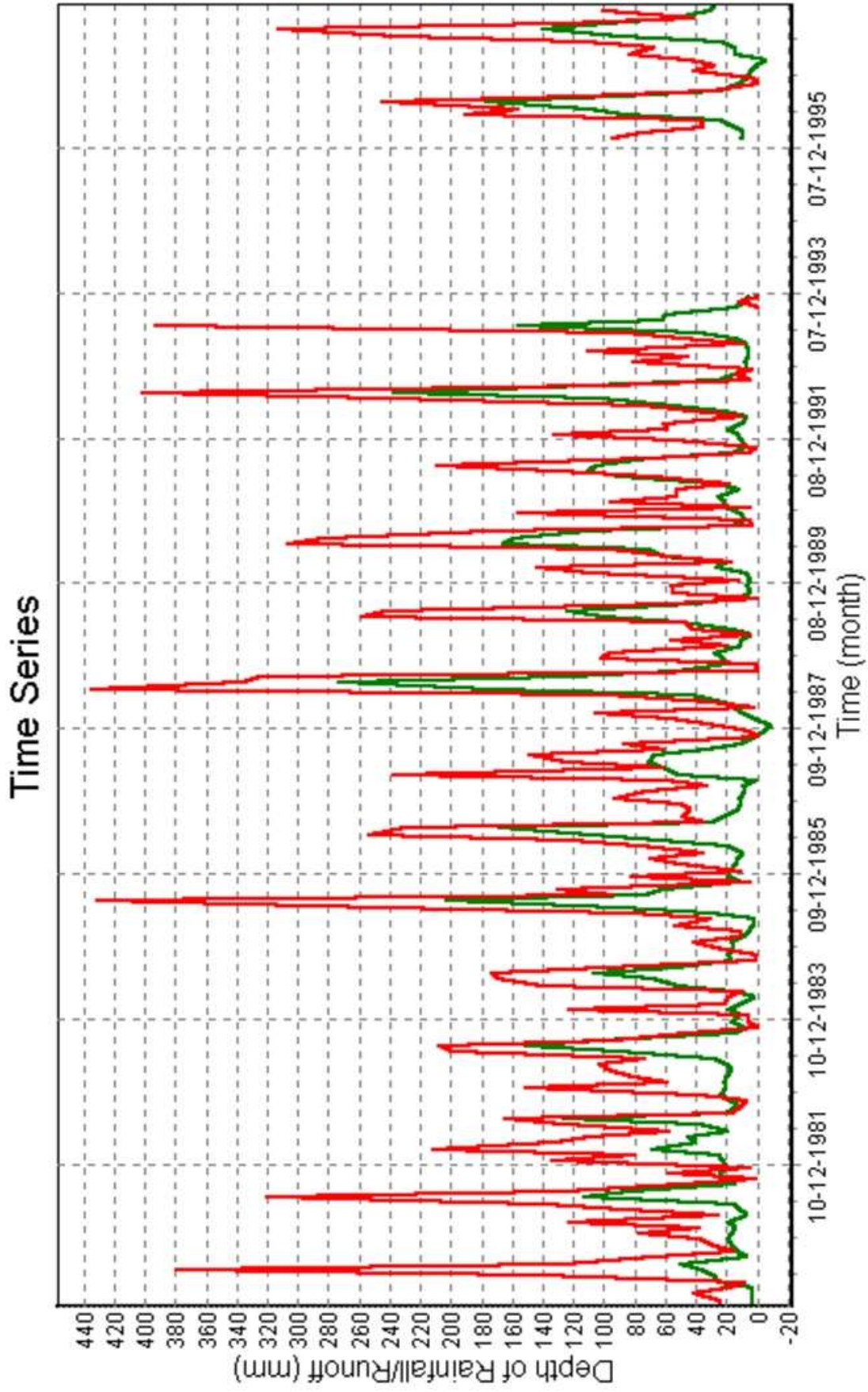
#### **4.3 RAINFALL-RUNOFF COMPARISON FOR INTERMEDIATE BASIN**

For checking the runoff generation from the intermediate catchment in-between Rampur and Bhakra gauging stations, the runoff was compared with the average rainfall in the intermediate catchment. Since the area of the intermediate catchment is large with

appreciable lag time between Rampur and Bhakra, monthly time step was adopted for making the comparison. The computed aerial rainfall and the observed discharge record were analyzed for the period from 1.1.1980 to 31.01.98 except for the intervening period from 1.1.1994 to 31.01.1996 for which runoff data at Bhakra were not available. The aerial rainfall (mm) for the intermediate basin obtained by Thiessen polygon method on daily time step was converted to the monthly values through aggregation.

The daily runoff data at Rampur and Bhakra were available in cumec, the same was obtained for the monthly time step by finding the average of all the days in the month (runoff being instantaneous values). The average runoff in cumec for the month was converted in the terms of depth over the intermediate catchment by dividing with the total quantity of water by the catchment area. The average monthly rainfall and runoff (both in mm) were plotted on the same graph and are shown in Figure – 4.9. From the graph, it is seen that the runoff series lies under the rainfall series and follows the trend quite significantly.





**Figure 4.9** The average monthly rainfall and runoff (both in mm)

## CHAPTER-5 TIME SERIES MODELLING

### 5.1 TIME SERIES ANALYSIS

A *time series* is a set of observations generated sequentially in time. If the set is continuous, the series is said to be *continuous*; if it is discrete, the time series is said to be *discrete*. Here only discrete time series shall be considered. The observations in a discrete series made at equidistant time intervals  $\tau_0 + h, \tau_0 + 2h, \dots, \tau_0 + th, \dots, \tau_0 + Nh$  may be denoted by  $z(\tau_1), z(\tau_2), \dots, z(\tau_i), \dots, z(\tau_N)$ . For many purposes, the value of  $\tau_0$  and  $h$  are unimportant. If  $\tau_0$  is adopted as the origin and  $h$  as the unit of time,  $z_t$  can be regarded as the observation at time  $t$ .

A discrete time series may arise in two ways:

- (1) By sampling a continuous time series; for example, the continuous river flow at a station may be sampled at hourly intervals.
- (2) By accumulating a variable over a period of time; for example rainfall may be accumulated over a period of a day.

A hydrologic time series can be divided in two basic groups: 1) univariate (single) time series, e.g., monthly streamflow at a point, and 2) multivariate (multiple) series of different kinds at one point. The examples of the second type are series of flow and water quality variables at a station. If a time-series, e.g., daily precipitation, is composed of nonzero and zero values, it is known as intermittent series. A time series whose values have been observed at regular intervals, such as each day or each hour, is termed as regularly spaced time series.

Time series analysis is useful for many applications, such as forecasting, detecting trends in records, filling-in missing data, and generation of synthetic data. A brief introduction of time-series analysis is given in the following.

#### 5.1.1 Components of a time series

A time series can be divided into a number of components. The main components of a hydrologic time series are: trends and other deterministic changes, cycles or periodic changes and autocorrelation, the almost periodic changes such as tidal effects, and components

representing stochastic or random variations. Trend and jumps are introduced in a time series due to gradual or sudden changes in the major factors of the process that are responsible for the time series. For example, a runoff time series will have a trend as a result of major land use changes in the upstream catchment because land use change is a relatively slow process. On the other hand, a water quality time series may show jump if a new factory upstream begins to discharge its effluent in the river. Similarly, the closure of a diversion dam will lead to a jump in the series because the flow will be reduced due to diversion.

The time series of hydrologic variables which are measured or accumulated at sub-annual time intervals normally have periodic patterns. Such patterns can be seen in time series, for example, monthly rainfall, daily runoff, daily volumes of urban water demands, and these series are said to have seasonal or periodic patterns. Due to seasonality, the statistical properties of time series vary with time.

### 5.1.2 Stationary stochastic processes

A special class of stochastic processes, called *stationary process*, is based on the assumption that the process is in a particular state of statistical equilibrium. A stochastic process is said to be strictly stationary if its properties are unaffected by a change of time origin. Thus, for a discrete process to be strictly stationary, the joint distribution of any set of observations must be unaffected by shifting all the times of observation forward or backward by any integer amount  $k$ . The statistical properties of a non-stationary time-series are time dependent.

Assuming that the stationarity assumption holds good, the joint probability distribution  $p(z_{t_1}, z_{t_2})$  is the same for all times  $t_1, t_2$ , which are a constant interval apart. Therefore, the nature of the joint distribution can be inferred by plotting a scatter diagram using pairs of values  $(z_t, z_{t+k})$ , of the time series, separated by constant interval or lag  $k$ . The covariance between  $z_t$  and  $z_{t+k}$  is called the autocovariance at lag  $k$  and is calculated by

$$\gamma_k = \text{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)] \quad \dots(5.1)$$

For a stationary process, the variance at time  $(t + k)$  is the same as at time  $t$ . The estimate of the  $k^{\text{th}}$  lag autocovariance  $\gamma_k$  is

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}), \quad k = 0, 1, 2, \dots, K \quad \dots(5.2)$$

The autocorrelation at lag k is given by

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E[(z_t - \mu)^2]E[(z_{t+k} - \mu)^2]}} = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2} \quad \dots(5.3)$$

$$\text{or} \quad \rho_k = \gamma_k / \gamma_0 \quad \dots(5.4)$$

The estimate of lag k autocorrelation is obtained by

$$r_k = c_k / c_0 \quad \dots(5.5)$$

## 5.2 TIME SERIES MODELS

A mathematical model representing a time series or stochastic process is called a *time series model*. The model has a certain structure and a set of parameters. The important categories of models are as follows.

### 5.2.1 Autoregressive (AR) models

AR models are extremely useful to represent certain practical series. Let the values of a process at equally spaced times t, t-1, t-2, ... be  $Y_t, Y_{t-1}, Y_{t-2}, \dots$  and let  $z_t, z_{t-1}, z_{t-2}, \dots$  be the deviations from the mean  $\mu$ ; for example,  $z_t = Y_t - \mu$ . In an AR model, the current value of the process is expressed as a finite, linear aggregate of previous values of the process and a shock  $a_t$ . Thus

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + a_t \quad \dots(5.6)$$

is called an autoregressive process of order p and is denoted by AR(p). Introducing a *backward shift operator* B, defined by  $Bz_t = z_{t-1}$ , eq. (5.6) can be written as:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) z_t = a_t \quad \dots(5.7)$$

$$\text{or} \quad \phi(B) z_t = a_t \quad \dots(5.8)$$

Here  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  is termed as an autoregressive operator of order  $p$ . The AR models have been extensively used in water resources because this form has an intuitive type of time dependence and the AR models are simple to use.

### 5.2.2 Moving average (MA) models

Another kind of model, of great practical importance in the representation of observed time series, is the finite moving average process. Here  $z_t$  is linearly dependent on a finite number  $q$  of previous  $a$ 's. Thus,

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad \dots(5.9)$$

is called a moving average (MA) process of order  $q$  and is denoted by MA( $q$ ). Similar to the autoregressive operator, a moving average operator of order  $q$  can be written as:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad \dots(5.10)$$

and the MA( $q$ ) process can be written as

$$z_t = \theta(B) a_t \quad \dots(5.11)$$

### 5.2.3 Autoregressive moving average (ARMA) models

Greater flexibility in fitting time series models is achieved by including both autoregressive and moving average terms in the model. This leads to the mixed autoregressive-moving average ARMA( $p,q$ ) model:

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad \dots(5.12)$$

$$\text{or } z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad \dots(5.13)$$

$$\text{or } \phi(B) z_t = \theta(B) a_t \quad \dots(5.14)$$

which employs  $p+q+2$  unknown parameters  $\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2$ , that are estimated from the data. The ARMA(1,1) model can be written as

$$z_t - \phi_1 z_{t-1} = a_t - \theta_1 a_{t-1} \quad \dots(5.15)$$

The ARMA models are suitable for stationary hydrologic series. In case of nonstationary series, the periodic or seasonal fluctuation can be removed by taking the differences and the ARMA model can be applied to the resultant series. The resultant model is termed as Autoregressive Integrated Moving Average (ARIMA) model. A time series that is homogeneous, except in level, can be adequately represented by a model of the form

$$\phi(B)\nabla z_t = \theta(B) a_t \quad \dots(5.16)$$

where  $\nabla$  is the *backward difference operator* defined as

$$\nabla z_t = z_t - z_{t-1} = (1-B)z_{t-1} \quad \dots(5.17)$$

Thus, ARIMA(p,d,q) is ARMA model that is fit after taking the d-th difference of the series:

$$\phi(B)\nabla^d z_t = \theta(B) a_t \quad \dots(5.18)$$

The combination of AR and MA models makes it possible to simulate many hydrologic processes by using a small number of parameters. An adequate representation of actually occurring stationary time series can be frequently obtained with autoregressive, moving average, or mixed model, in which p & q are not greater than 2 and often less than 2.

### 5.3 PARTIAL AUTOCORRELATION FUNCTION

The partial autocorrelation function is a device which exploits the fact that whereas an AR(p) process has an autocorrelation function which is infinite in extent, it can by its very nature be described in terms of p non-zero functions of the autocorrelations. Thus the partial autocorrelation function is another way of representing the time dependence structure of a series. It is useful in identification of the type and order of the model of a given time series. Let  $\phi_{kj}$  denote the jth coefficient in an autoregressive process of order k, so that  $\phi_{kk}$  is the last coefficient. Now  $\phi_{kk}$  satisfies the set of equations

$$\rho_j = \phi_{k1} \rho_{j-1} + \phi_{k(k-1)} \rho_{j-k+1} + \phi_{kk} \rho_{j-k}, \quad j=1,2, \dots k \quad \dots(5.19)$$

leading to the Yule-Walker equations which may be written as

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & & \rho_{k-2} \\ \vdots & & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad \dots(5.20)$$

or

$$P_k \phi_k = \rho_k \quad \dots(5.21)$$

Solving these equations for  $k = 1, 2, 3, \dots$ , successively, the values of  $\phi_{11}, \phi_{22}, \dots$  are obtained as a function of  $\rho$ . The quantity  $\phi_{kk}$ , regarded as a function of the lag  $k$ , is called the *partial autocorrelation* function. For an autoregressive process of order  $p$ , the partial autocorrelation function  $\phi_{kk}$  will be nonzero for  $k$  less than or equal to  $p$  and zero for  $k$  greater than  $p$ . Therefore, the partial autocorrelation function of a  $p^{\text{th}}$  order autoregressive process has a cutoff after lag  $p$ . Conversely, the autocorrelation function of an MA( $q$ ) has a cutoff after lag  $q$ , while its partial autocorrelation tails off. If both autocorrelations and partial autocorrelations tail off, a mixed process is suggested. Furthermore, the autocorrelation function of a mixed ARMA( $p, q$ ) process is a mixture of exponentials and damped sine waves after the first  $(q - p)$  lags. Conversely, the partial autocorrelation function for a mixed process is dominated by a mixture of exponentials and damped sine waves after the first  $(p - q)$  lags.

Partial autocorrelations may be estimated by successively fitting autoregressive processes of orders 1, 2, 3, ... by least squares and picking out the estimates  $\phi_{11}, \phi_{22}, \dots$  of the last coefficient fitted at each stage. Alternatively, approximate Yule-Walker estimates of the successive autoregressive processes may be employed. The estimated partial autocorrelations can then be obtained by substituting estimates  $r_j$  for the theoretical autocorrelations in eq. (5.19) to yield

$$r_j = \hat{\phi}_{k1} r_{j-1} + \hat{\phi}_{k2} r_{j-2} + \dots + \hat{\phi}_{k(k-1)} r_{j-k+1} + \hat{\phi}_{kk} r_{j-k}, \quad j = 1, 2, \dots, k \quad \dots(5.22)$$

#### 5.4 FITTING OF ARMA MODELS

Initially, one does not know the order of ARMA process for fitting to an observed time series. The problem is analogous to deciding on the number of independent variables to be included in multiple regression. Therefore, the model is built iteratively, i.e., a set of models

is identified using the characteristics of the data and its adequacy is tested. Depending on the results, the model may be adopted or another candidate model is identified. While fitting a time series model, at least 50 observations should be used. If a sufficient number of observations are not available, one proceeds by using experience to yield a preliminary model. This model may be updated as more data becomes available. While applying the ARMA models, the main stages are:

- *Model Identification* which involves the use of the data and any information on how the series was generated to identify a subclass of parsimonious models worthy to be entertained.
- *Parameter Estimation* which involves an efficient use of the data to make inferences about parameters conditional on the adequacy of the entertained model.
- *Diagnostic Checking* involves checking the fitted model in its relation to the data with the intent to reveal model inadequacies and to achieve model improvement.

In any time series modelling, it is worthwhile to first plot the data. A visual inspection of the data always gives useful information about the behaviour of the process. Having tentatively decided the order of differencing  $d$ , the general appearance of the estimated autocorrelation and partial autocorrelation functions of the appropriately differenced series are studied. This provides clues about the choice of the orders  $p$  and  $q$  for the AR and MA operators. In doing so, the characteristic behaviour of the theoretical autocorrelation function and of the theoretical partial autocorrelation function for AR, MA, and mixed processes is used. For example, if  $\phi_1$  of an AR(1) model is +ve,  $\rho_k$  decays exponentially to zero while for  $\phi_1$  -ve, it oscillates in sign. The autocorrelation function of AR(2) model has different forms depending on the values that the parameters take.

The common techniques for estimation of the parameters of a time series model are the method of moments, the method of least squares, and the method of maximum likelihood. Mostly the parameters obtained by the method of moments are used as the first approximation and are refined by other methods.

After a model has been chosen and the parameters are estimated, it is necessary to apply statistical tests to check its adequacy and suitability. The tests which are used for this



purpose include *The Porte Manteau Lack of Fit Test*, the *Akaike Information Criterion (AIC)*, and the test of correlogram. The statistic AIC is calculated by

$$AIC(p, q) = N \ln \left( \hat{\sigma}_\epsilon^2 \right) + 2(p + q) \quad \dots(5.23)$$

where, N is the sample size, and  $\hat{\sigma}_\epsilon^2$  is the maximum likelihood estimate of the residual variance. The model which gives the minimum AIC is selected. It is always helpful to examine the residuals of a model. For a sufficiently long data series, the residuals resemble white noise. A test to check whether the residuals of a model are independent or not is the *Portemanteau Lack of fit* test whose Q statistic is determined by

$$Q = N \sum_{k=1}^L r_k^2 \quad \dots(5.24)$$

where  $r_k$  denotes autocorrelation of residuals at lag k and L is the number of lags considered. Q approximately follows a chi-square distribution with (L-p-q) degrees of freedom. The technique of overfitting in which a more elaborate model is fitted to the data and then the results are compared, has also been recommended. Box and Jenkins (1976), Salas et al. (1980), and Salas (1992) are good references for time series models.

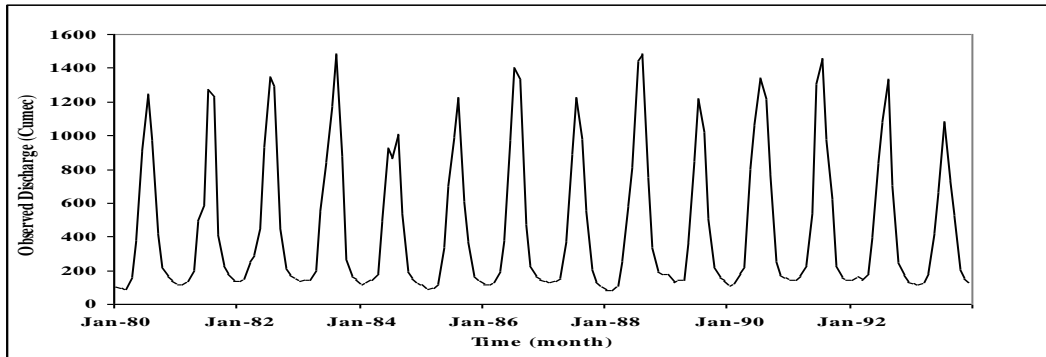
## 5.5 TIME SERIES MODELLING OF BHAKRA DATA

The observed monthly discharged data at two stations in the study area were available. These were Bhakra and Rampur. There were some missing values in the Bhakra data series near the end of series. The segment of series that was selected for analysis at each station was the data without any missing values. At station Bhakra, the input data consisted of 168 values for the period January 1980 to December 1993 while for the Rampur station, the series consisted of 216 values from January 1980 to December 1997.

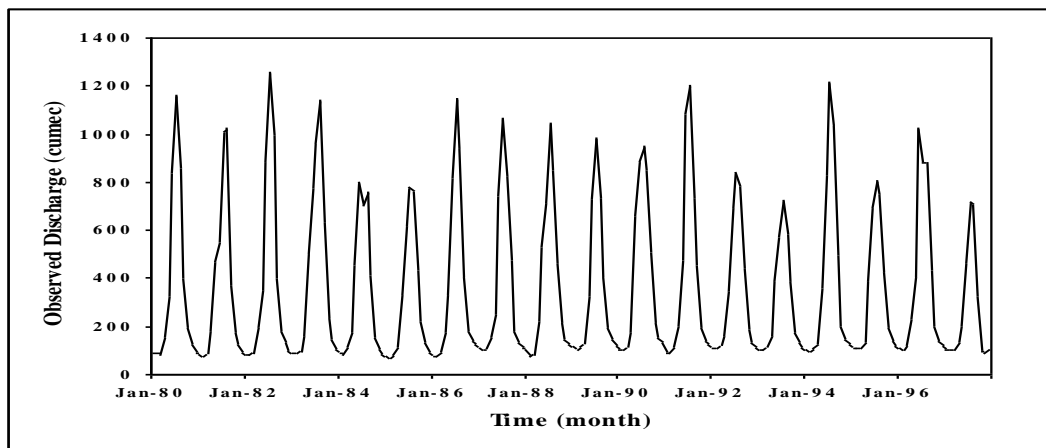
Although the present report is related to use of HYMOS software, it was thought appropriate to apply time series modelling for these two sites. A software developed at NIH was used for this purpose.

The observed data for these two stations is plotted in Figure 5.1 and 5.2 clearly showing a cyclic pattern in the data. For both the sites, lag 1 serial correlation for lags 1 to 20

for the observed data were computed and are given in Table 5.1. As can be seen from this table, there was significant correlation at various lags, high correlation at lag 12 may be noted. Since the input data consists of monthly values, a high correlation at lag 12 is expected. This high correlation indicates that some processing is necessary before applying ARMA models.



**Figure 5.1** Observed discharge at Bhakra from 1980-1993



**Figure 5.2** Observed discharge at Rampur from 1980-1997

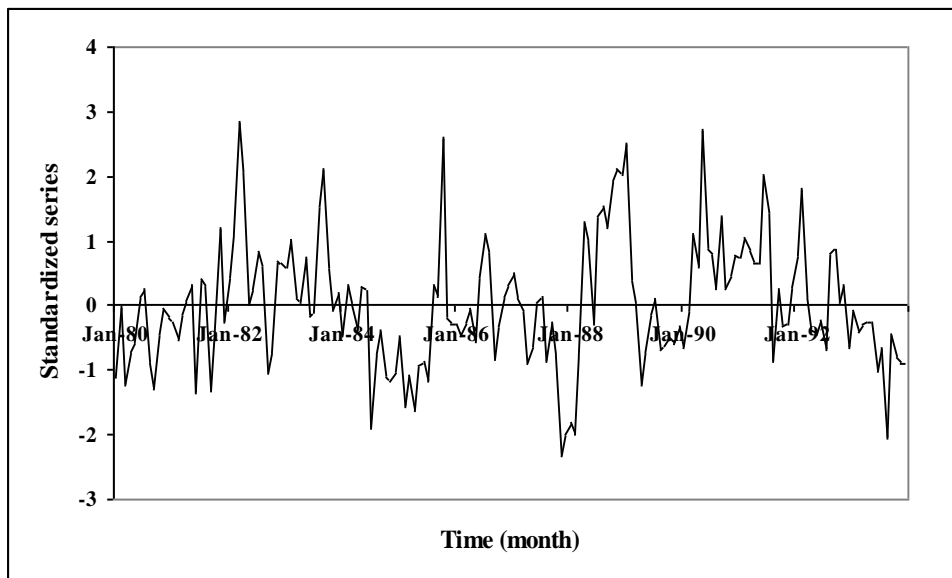
The input data series were standardized by subtracting from each value, the mean for the corresponding month and dividing by the corresponding standard deviation.

$$Z_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j} \quad \dots(5.25)$$

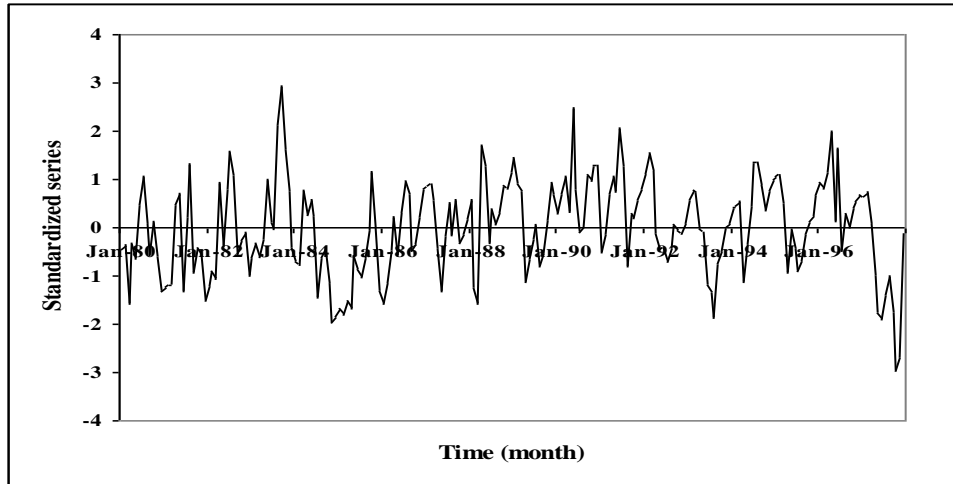
where,  $x_{i,j}$  = flow for year  $i$ , month  $j$ ,  $\mu_j$  &  $\sigma_j$  = Mean and standard deviation for month  $j$  and,  $Z_{i,j}$  = standardized series. The standardized series for Bhakra and Rampur are shown in Fig. 5.3 and 5.4.

**Table 5.1** Serial correlation coefficients of observed data

Lag	Correlation	
	Bhakra	Rampur
1	0.7514	0.7667
2	0.2779	0.2954
3	-0.1614	-0.1563
4	-0.4433	-0.4510
5	-0.5670	-0.5803
6	-0.5877	-0.6021
7	-0.5381	-0.5547
8	-0.4097	-0.4187
9	-0.1417	-0.1389
10	0.2671	0.2805
11	0.6828	0.6904
12	0.8748	0.8748
13	0.6885	0.6999
14	0.2588	0.2754
15	-0.1633	-0.1567
16	-0.4275	-0.4359
17	-0.5319	-0.5542
18	-0.5502	-0.5763
19	-0.5048	-0.5272
20	-0.3838	-0.3949



**Figure 5.3** Standardized series for Bhakra



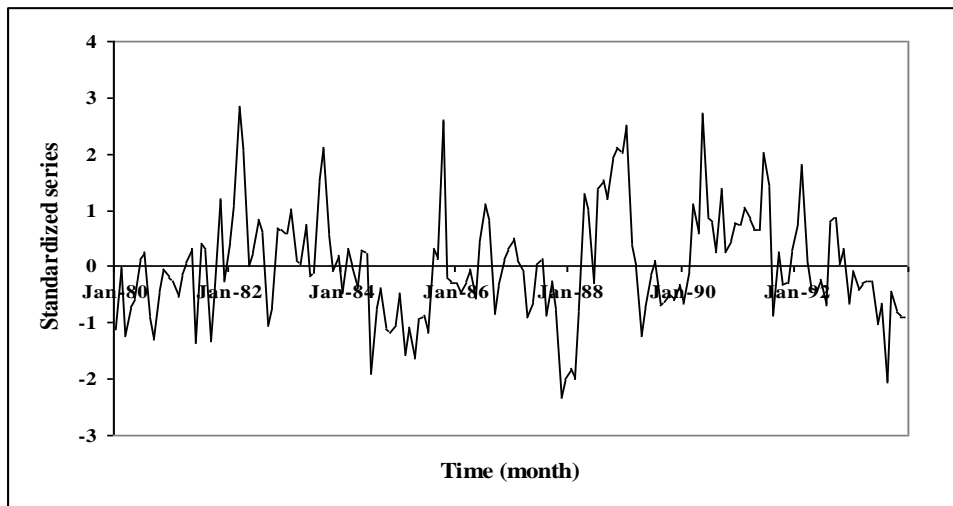
**Figure 5.4** Standardized series for Rampur

Next, correlations for different lags of the standardized series for the two stations were computed and these are given in Table 5.2. This table shows that still there is some correlation in both the series. In order to remove the small correlations, the series were differenced with a lag one.

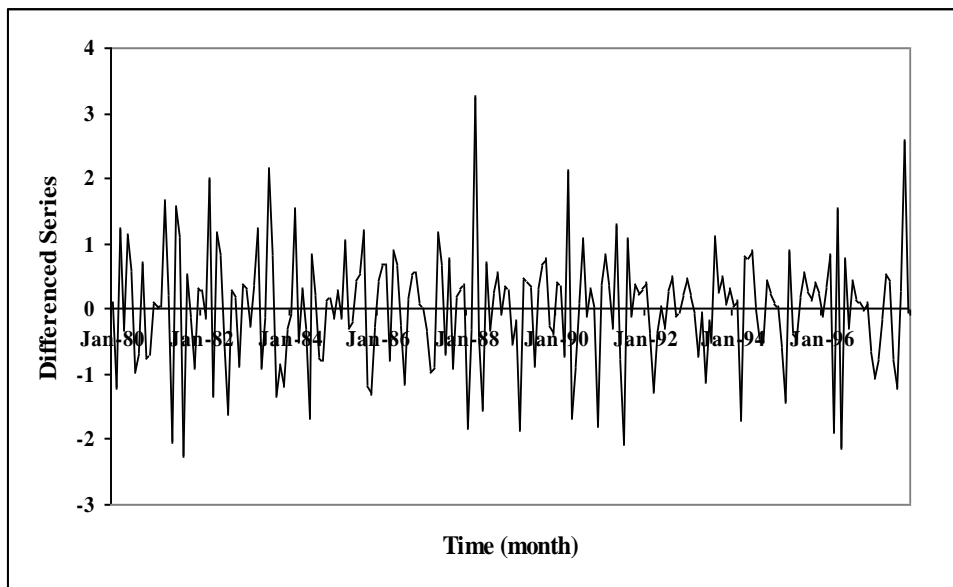
**Table 5.2** Serial correlation coefficients of standardized data

Lag	Correlation	
	Bhakra	Rampur
1	0.5616	0.6048
2	0.3581	0.3348
3	0.2945	0.2454
4	0.2286	0.1488
5	0.0974	0.0692
6	0.0622	0.0979
7	0.0707	0.0392
8	0.0504	0.0059
9	0.0157	0.0604
10	-0.1042	0.0685
11	-0.2087	0.0167
12	-0.1939	0.0390
13	-0.1338	0.0678
14	-0.0837	0.0558
15	-0.0530	0.0235
16	0.0155	-0.0365
17	0.0005	-0.1243
18	0.0265	-0.1441
19	0.0308	-0.1792
20	0.0027	-0.1184

The differenced series are plotted in Fig. 5.5 and 5.6; their correlograms are shown in Fig. 5.7 and 5.8. This figure shows that now the correlations are substantially small.



**Figure 5.5** Differenced series at Bhakra



**Figure 5.6** Differenced series at Rampur

As a first trial, the ARMA(1,1) model was applied on the standardized series and its parameters were estimated. The residuals, i.e., the difference between the observed data and those computed using ARMA model were determined and the statistical properties of these residuals, mean, variance and lag 1 correlation were determined. A number of ARMA models were tried on differenced series by varying the order of AR and MA parts. Table 5.3 gives the statistical properties of residuals as well as AIC for each of these models.

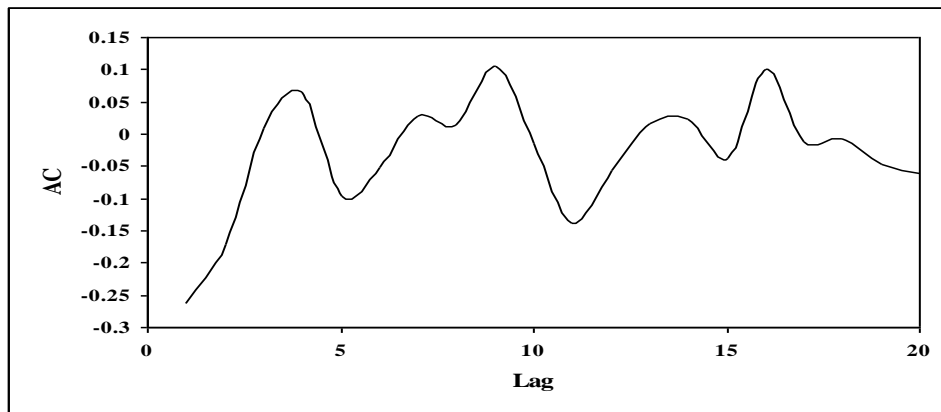


Figure 5.7 Correlogram of differenced series for various lags for Bhakra

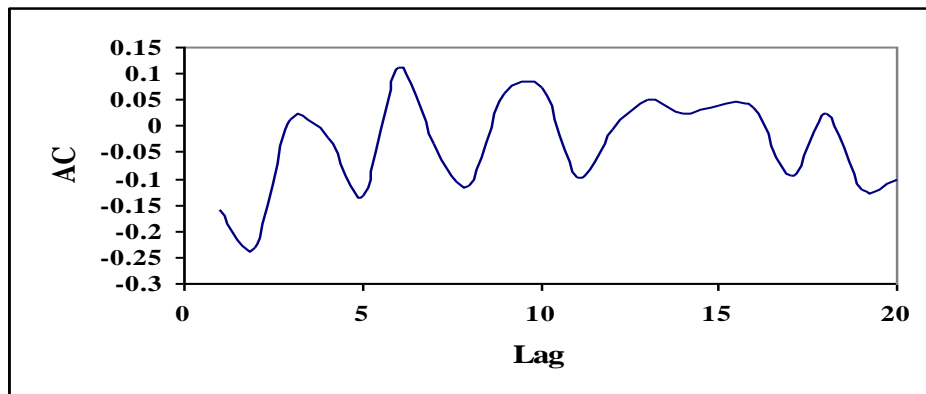


Figure 5.8 Correlogram of differenced series for various lags for Rampur

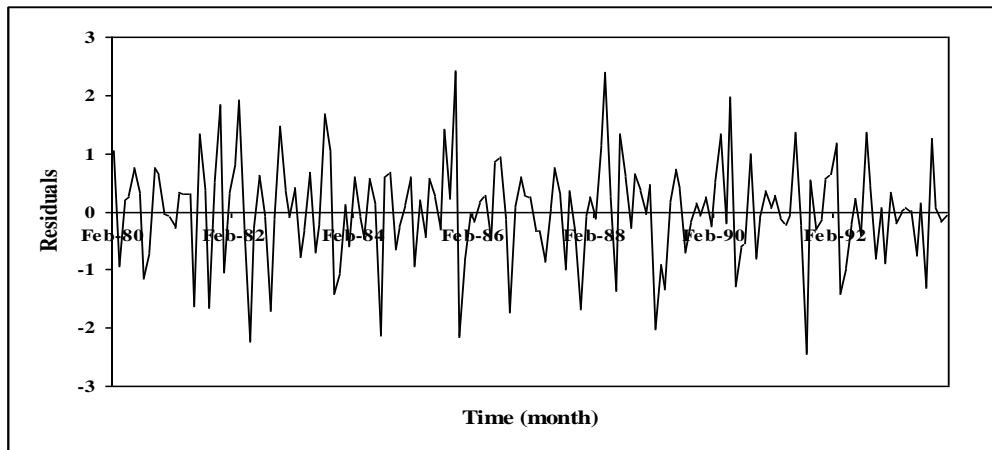
Table 5.3 Performance of trial models – Bhakra data

Model	AIC	Residual mean	Lag-1 correlation of residuals	Residual variance
ARMA(1,1)	-72.079	-0.0012	0.0088	0.636
Diff-ARMA(1,1)	-66.183	-0.0056	0.0264	0.657
<b>Diff-ARMA(1,0)</b>	<b>-43.868</b>	<b>-0.0001</b>	<b>-0.0661</b>	<b>0.760</b>
Diff-ARMA(0,1)	-54.617	-0.0012	0.0963	0.712
Diff-ARMA(2,1)	-64.791	0.0061	-0.0030	0.654
Diff-ARMA(2,2)	-63.952	0.0369	-0.0338	0.650

It can be seen from this table that for the ARMA(1,0) or AR(1) model, the AIC is the minimum; the residual properties also suggest this to be the best model for Bhakra data. The value of the parameter  $\phi_1$  for this model was  $-0.265$ . Further, for this model the residual mean is  $-0.0001$  and Lag 1 correlation is  $-0.0661$ . This indicates that the residuals are close to white noise. The residual series for Bhakra is plotted in Fig. 5.9. Based on these results,

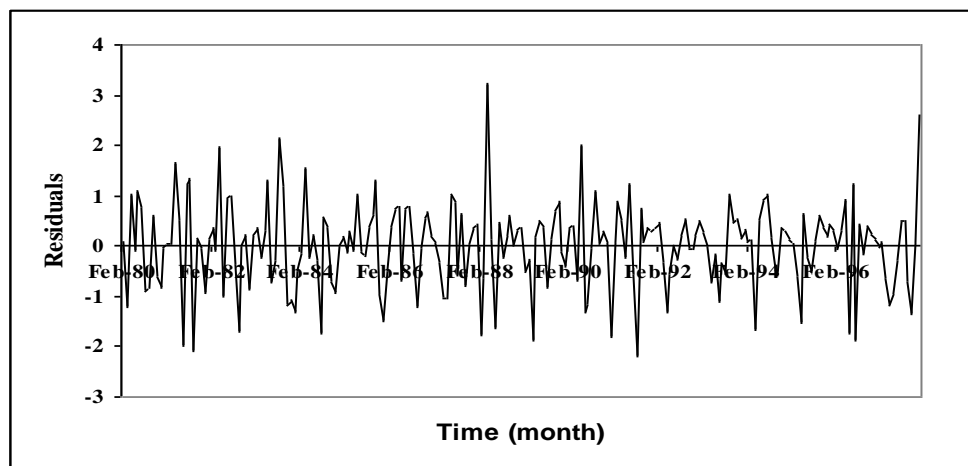
AR(1) model applied to the differenced series can be considered to be appropriate for the data of Bhakra site. Mathematically this model can be written as

$$z_t - z_{t-1} = -0.265(z_t - z_{t-1}) + a_t \quad \dots(5.26)$$



**Figure 5.9** Residuals series for Bhakra

The results of analysis on the same line for Rampur site are also given in Table 5.4 and for this site also, AR(1) model appears to be the best based on the properties of residuals as well as AIC. . The value of the parameter  $\phi_1$  for Rampur data was  $-0.167$ . The residual series for Rampur is plotted in Fig. 5.10.



**Figure 5.10** Residuals series for Rampur

**Table 5.4** Performance of trial models – Rampur data

Model	AIC	Residual mean	Lag-1 correlation of residuals	Residual variance
ARMA (1,1)	-101.767	-0.0023	-0.404	0.611
<b>Diff-ARMA (1,0)</b>	<b>-65.947</b>	<b>-0.0001</b>	<b>-0.0445</b>	<b>0.729</b>
Diff-ARMA (0,1)	-73.302	-0.0040	-0.1050	0.705
Diff-ARMA (2,2)	-101.144	0.0246	-0.0086	0.602

Mathematically, the model for Rampur can be written as

$$z_t - z_{t-1} = -0.167(z_t - z_{t-1}) + a_t \quad \dots(5.27)$$

It may be noted that in view of the differenced series, these models cannot be used for forecasting or synthetic data generation.



## CHAPTER - 6 CONCLUSIONS

Advances in hydrology are dependent on good, reliable and continuous measurements of the hydrological variables. Hydrologic data collected from various sources/instruments are generally raw data and can not be used directly in most hydrologic analysis work. Thus, the processing of such raw data is the first and major task that has to be undertaken before taking up any hydrological study. Data validation, classified as primary and secondary, is essential before putting the data for further use in analysis. Processing of hydrological data has two major objectives: i) to evaluate the data for its accuracy, and ii) to prepare the data in a form most valuable to the users. Data processing software using high-speed computers have immense potential for handling large volume of data in a quick and economic way.

In this study, the *SWDES* and *HYMOS* software have been used for the processing and validation of rainfall and discharge data in the Satluj basin. The processing is mainly concentrated on the part of the basin that does not receive snowfall. The historical data, as collected from the Bhakra Beas Management Board, were available on the magnetic media and the same were imported in the processing software database after their conversion in the desired format. The data were available from 1977 to 1998. Since the topography of the Satluj basin is quite irregular with elevation variation from 500 to 7000 m above mean sea level, the rainfall data of different stations were first normalized. Various checks of data validation, such as multiple time series plots, spatial homogeneity test, double mass curve analysis were then applied on the data. For the screening of discharge data at Rampur and Bhakra, the monthly flow from the intermediate catchment (between Bhakra and Rampur) were compared with the average aerial rainfall derived using the Thiessen polygon method.

Suspect values have been identified and listed in the report. Multiple time series plot of rainfall at neighboring stations brings out some instances when the rainfall at a station is abnormally high as compared to other stations. Though this could be possible in a mountainous area because of the orographic effect on the windward and leeward sides, there is also possibility of wrong observation or transcription error. Similarly, using the spatial homogeneity test, some instances of larger variation among neighboring stations have been identified. Using the double mass analysis, it is observed that the trend of the double mass

plot shifts appreciably at Bhakra station from the year 1983 onwards and for Kasol station from 1981 onwards. It could be because of the change in characteristics of the raingauge station or its surroundings. The graphical comparison of rainfall and the resulting runoff shows satisfactory observation of the discharge. Figures have been plotted for showing the results of various tests performed. Some major deviations from the expected values (found through validation) were referred to BBMB for their comparison with the original records. In a number of cases, errors in the original database entered in the computer were observed which justifies the usefulness and necessity of validation of hydrological observations before their utilization.

The time series analysis of the discharge observations at Bhakra and Rampur has also been carried out using a software developed at NIH. The data series were standardized and then first order differencing was taken. The AR(1) model was found to be the best for the resultant series of both sides.

To sum up, hydrological data processing software such as *SWDES* and *HYMOS* are very useful and efficient for the effective processing and validation of hydrological data. Data can be checked through a number of validation checks in a short span of time and the results can be presented in graphical form for easy interpretation. There is a need to process and validate the data as soon as possible so that the anomalies and the suspect values, if any, can be verified from the field records before any analysis is carried out and important conclusions are drawn from the original database.

\* \* \*

## REFERENCES

- DHV Consultants and Delft Hydraulics, (1999). “Surface Water Data Entry System (*SWDES*) – Users’ Manual”, Version 1.02, Technical assistance under Hydrology Project, November 1999.
- Jain, S. K., (2001). “Snowmelt runoff modelling and sediment studies in Satluj basin using remote sensing and GIS”, unpublished Ph.D. thesis, University of Roorkee, Roorkee.
- DHV Consultants and Delft Hydraulics, (1999). Basic surface water data processing (*HYMOS*) – Training Modules”, November 1999.
- Singh, Pratap and Jain, S. K., (2002). “Snow and glacier melt contribution in the Satluj river at Bhakra dam in the western Himalayan region”, *Hydrological Sciences Journal*, Vol 47(1), February 2002, pp. 93-106.
- WL/Delft Hydraulics, (1999). “*HYMOS* Manual”, Version 4.01, Technical assistance under Hydrology Project, July 1999.

\* \* \*