TRAINING COURSE

ON

SOFTWARE FOR SURFACE WATER DATA MANAGEMENT

UNDER

WORLD BANK FUNDED HYDROLOGY PROJECT

ROORKEE

19-30 AUGUST 1996

HYPRO - HYDROLOGICAL DATA STORAGE, RETRIEVAL AND PROCESSING SYSTEM

BY

SHARAD KUMAR JAIN
DEEPA CHALISGAONKAR

ORGANISED BY:

NATIONAL INSTITUTE OF HYDROLOGY
ROORKEE - 247 667 (U.P.)

HYPRO - HYDROLOGICAL DATA STORAGE, RETRIEVAL & PROCESSING SYSTEM

1.0 INTRODUCTION

Most of the nations are concerned about their natural resources but the concern for water is invariably the most prominent. Water is already a scarce commodity today and it is going to be much more scarce in future. Hydrology deals with the occurrence and movement of water on and above the surface of the earth. The hydrological appraisal of water resources projects is an essential element in planning, design, operation, maintenance and management of water resources projects. A successful and efficient completion of hydrological studies depends on timely availability of vast and diverse amount of data. Reliable and representative data are necessary for better management of water resources projects. Where sufficient data are not available, quantification of hydrological processes is possible only with limited accuracy and plans for hydrologic control and development need to make compensatory precision, usually such provisions are not perfect surrogates of the relevant information. The insufficient and inaccurate data may lead to uneconomical design, inefficient operation and ineffective management of the water resources projects.

The hydrological data are collected over the years with enormous efforts and expenditure. Conventionally, the data so collected are stored in registers or files scattered over many offices. Usually no proper inventory is maintained and in very few cases such an inventory is maintained at a central place. Often very little effort is put-in for the proper documentation of data in a scientific manner, resulting in loss/damage of data nonavailability in time, and their limited use.

The data in unprocessed form are of limited use. Besides, every user has to do preliminary processing/consistency checks. The processed data in the desired form, known as information, are of real importance to the end-user.

The traditional practice of information collection, compilation and inference not only takes considerable amount of time but also does not have any mechanism of correcting the data. This is due to the nonstandardization of collection and compilation procedures, which play an instrumental role in enormous delays. With the introduction of digital computers, data handling can be done more rapidly, easily and economically.

1.1 USE OF COMPUTER MEDIA FOR DATA STORAGE

Up to 1980, punch cards were mainly used for storing the data in computer compatible form. This type of data storage had a number of limitations like the cards may be damaged, they were inconvenient to carry, visual checking of the data was difficult and time consuming and proper sequence of the cards was to be maintained.

To avoid the problems associated with the cards, from the 80's, some organisations started storing their data on computer compatible magnetic storage media. In the simplest way, data are organized in sequential files which are stored on magnetic media for example tapes. Whenever needed, the data are picked up from the media and are used. To access a specific data record from a sequential file, it is required to check each record in sequence until the desired one is encountered. Clearly, sequential access holds a distinct disadvantage for a user interested only in specific records. The contents of a record on a magnetic tape cannot be edited, nor can the records be added or deleted in between.

To overcome the drawbacks of magnetic tapes, magnetic disks are used as the storage devices. A magnetic disk is the direct access medium on which data are recorded in concentric circles and the disk drive mechanism directly locates the place where the read/write operation is to be performed.

Now a days computer based database management techniques are being applied for efficient handling of hydrological data. They consist of creating a database, updating it frequently as and when new data are available and after carrying out validation checks, retrieving the data from the database when desired, The data

file is stored on either on-line storage or off-line storage. In the on-line storage mode, the data files are available for operations at all the times. However, this mode of storage is slightly expensive. In the off-line mode the data files are stored on magnetic tapes. This mode of operation is cheaper than on-line and is mostly used to store the data. It may be pointed out that the magnetic tapes require careful handling and storage to maintain data integrity. The tape should be preferably kept in a controlled environment and should be periodically used else the magnetic material tends to come-off the tape, resulting in data loss. In the recent past, cost of disk drive of higher storage capacity, say, 1 GB or more, has come down substantially and these are gradually replacing the magnetic tapes. The recently developed magnetic devices like CD-ROM and DAT are also becoming more popular.

2.0 FEATURES OF HYPRO

HYPRO is a Data Storage, Retrieval and Processing System for hydro-meteorological data designed for use on IBM compatible PCs in DOS environment. It arranges the data in a convenient form and provides an extensive set of tools for data entry, validation, analysis, retrieval and reporting.

HYPRO is comprehensive, well-tuned and easy-to-use via full-screen menus with on-line help to guide the user. The package includes many tabular and graphical options facilitating efficient reporting. It has been developed to streamline the storage and processing of (geo-)hydrological and meteorological data. HYPRO data are time-oriented. Duplication of data manipulation has been avoided; for which many powerful packages are available. Instead more effort has been put to create fast data retrieval and entry options for easy data transfer to various applications.

HYPRO integrates the distinctive phases in the processing of hydrological data. The activities are carried out in specific processing modules, each module consisting of a number of programs for a particular computation. The modules are structured according to a logical sequence of activities in data processing. All the modules are linked to the HYPRO database. In HYPRO, the related data items are stored in a database in the form of records. The records have further been divided into fields.

HYPRO is a menu driven and user friendly package. The system automatically selects the available graphics mode. Help is available at the various stages and the help messages are displayed on the bottom most line on the screen. At the beginning, the main menu is displayed on the screen. The options from the main menu can be selected by moving the cursor bar using the arrows keys available on the keyboard and pressing the Enter key. Next, submenu will appear on the screen, depending on the option selected. This process will continue till the objective is achieved. Switching from the menu to its submenu is performed by moving the cursor to the desired option and pressing the ENTER key while switching from a submenu to its master menu can be done by pressing the F10 key.

2.1 MODULES OF HYPRO

The HYPRO comprises of the following modules :

- Database Management module
- A Data Storage and Retrieval module, covering data entry, editing, reporting in tabular and graphical form as well as transfer and retrieval of data.
- Data Processing module, which includes data validation and various Analysis.

The flow diagram of HYPRO is shown in Fig 1.

2.2 HARDWARE AND SOFTWARE REQUIREMENTS

The HYPRO package has been written in the programming languages FORTRAN-77, C and dBase. The hardware and software requirements for using HYPRO are:

- Minimum 640 KB memory
- Hard Disk of 20 MB or more (required storage will depend upon the database size)
- Graphic Card
- DOS with version 5.0 or later
- A compatible printer with graphical capabilities.

For fast and pleasant operation, at least a PC-386 with colour monitor is recommended

2.3 DATA ENTRY & EDITING

Data Entry and Editing module includes the following options :

- Entry and Editing of catchment data
- Creation of station and data series, and entry and editing of station particulars like, station name, district, country, river basin, geographical latitude and longitude, altitude, catchment area and name of data collecting agency, series name, unit etc. A Full Screen editor for editing and display of time series with graphical display is also available.
- Loading of time-series data from user files or manually.
- Full screen editor for editing and display of time-series with on-line graphical display.

2.3.1 Station and Series Definition

The hydrological stations for which data series are to be entered as well as the data series themselves must be defined before actual data can be stored in the database. Before creating a database of hydrological data for a region, it is necessary to evolve a mechanism for assigning unique code to the stations. The coding of stations is necessary due to inherent problems while working with station names. It is better to assign station codes by following a logical order based on, for example, river basins, sub-basins or latitude and longitude. This makes it easier to retrieve the data for geographic entity, say a sub-basin or an area bounded by given latitude and longitude range.

Ministry of Agriculture, Government of India, has published an Atlas wherein codes have been assigned to all basins of India, their sub-basins and sub-sub-basins. Similarly, UNESCO has already advanced a scheme for stations based on geographic conditions. It is necessary that one such scheme is formulated for the entire country.

For station as well as series definition, the following options are available :

- **Add**: Add station/series code with other details if station/series code does not exist in the database.
- **Edit**: Modify other details of an already existing station/series code. However, station/series code cannot be modified.
- **Review**: View the details of the existing station/series.
- **Delete**:Remove the station/series code from the HYPRO database.

The station definition includes the following information:

- **Station Code** : A unique set of 1 to 8 characters to identify the station, for example NIHR. This code forms the first part of the file in which the corresponding data is stored.

- **Station Name** :The actual name of station. It has no operational consequences.

- **River** : Name of the river (basin) in which the station is located.

- **District** : District in which the station is located.

- **Country** : Country in which the station is located.

- **Latitude** : Latitude in degrees, minutes and seconds, North or south.

- **Longitude**: Longitude in degrees, minutes and seconds East or west.

- **Altitude** : Altitude in metres relative to mean sea level or any other datum (0 < altitude < 8000).

- **Catchment Area**: Area upstream of the station, in sq. km. (0 < area < 10,00,000).

- **Agency**: Name of the agency who operates the station.

The series definition includes the following information:

- **Series Code**: A unique set of 2 alphabets to identify the series. This code forms the second part of the file in which the corresponding data is stored. Valid examples are: RF, QH and DEP

- **Time Unit**: It represents the time distance between successive elements of data stored. It may be year, month, day or hour or a part thereof. The valid codes are as follows :

Year = 1, Month = 2, Day = 3, Hour = 4.

- **Divider**: It represents the division factor applied to the relevant Time Unit. Its value may be in the range 1 to 99 (integer). The annual data will have time unit = 1 & divider = 1; six monthly data will have time unit = 1 & divider = 2 and 6 hourly data will have time unit = 3 & divider = 4.

- **Description** : Description of the series. It has no operational consequence.

- **Unit**: Unit of the data, for example mm, cumec.

- **Missing Value**: For each data set a value is stored in the database to indicate that a data point is missing. Although the choice of this value is free it should always be less than the lowest possible value.

For example, -9 can be used to denote missing rainfall.

- **Minimum Value**: For validation purpose the likely minimum value of a data set is stored as series characteristic. In data validation module, the data values, which are less than this value will be flagged.

- **Maximum value**: For validation purpose the likely maximum value of a data set is stored as series characteristic. In data validation module the data value, which is greater than this value will be flagged.

Using Time Unit and Divider as indicated above, time intervals can be specified in a very flexible way. Practically all commonly used intervals between one minute and one year fit the definition.

2.3.2 Entry & Editing of Time-Series Data

Data can be entered in HYPRO database either by reading from ASCII file or via the data entry menus from the keyboard. It can also be modified using the keyboard. The details are as follows :

○ **Read Data From File** : Data are read from data files. In the file all data should be separated by a blank. The layouts of the ASCII file for transfer of data to the HYPRO databases are as follows:

Yearly Data :

Year, Data1, Data2,.....Data30

Monthly Data :

Year, Data1, Data2,.....Data12

Daily Data :

Year, Month, Data1, Data2,.....Data31

Hourly Data :

Year, Month, Day, Data1, Data2,.....Data24

While entering the day or month the values from 1 to 9 should be entered as 01, 02,.....09.

After selection of this option the station code is requested. If the station code is already defined, the series code is requested. If the series code is already defined, the data file name with path is requested. If the data file exists, the data stored in it will be entered in HYPRO database. In case of any error, appropriate message is displayed and system exits from this option.

○ **Entry of Constants** : In many hydrological series, sometimes a particular value is repeated many times. In this situation, for speedy entry of such data through the keyboard, this option may be used. After selection of this option, the station code is requested. If the station code is already defined, the series code is requested. If the series code is already defined, the starting date is requested. If the starting date is between 01/01/1901 and the date of execution, the ending date is requested. If the ending date is equal to or after the starting date, the value of constant to be filled is requested and the value is entered in the database for the given time period. In case of any error, the appropriate message is displayed and the system will exit from this option.

○ **Edit Time-Series Data** : Using this option the data can be edited with the help of a keyboard. After selection of this option the station code is requested. If the station code is already defined, the series code is requested. If the series code is already defined, the starting date is requested. If the starting date is between 01/01/1901 and the date of execution, the ending date is requested. If the ending date is equal to or after the starting date, the data editing screen appears on the monitor. In case of any error, the corresponding message will be displayed and system exits from this option. Otherwise, the data will be edited in the database for the given time period.

2.4 DATA VALIDATION

Before the data can be used for various analysis, they should be validated which is often a labourious job. HYPRO offers several standard techniques for quick data validation. Tabular, graphical and computational procedures are available for validation of various types of data. Data validation module includes the following options:

○ **Data screening**

HYPRO offers an option for indication of unlikely values by marking the data outside defined boundaries, i.e. by checking for less than and greater than the defined limits.

○ **Plotting of time-series graphs**

This enables the user to obtain a quick visual presentation of the measured values as well as possible measurement errors. For example, abrupt change in an otherwise smooth hydrograph may indicate a possible measurement/ recording error.

2.5 HYDROLOGICAL ANALYSIS

The main objective of creation of a hydrological database is to carry out an analysis using the data. Therefore, if the database management software is capable of carrying out the analysis of data, it may result in saving of time and efforts. With this view in mind, a few modules have been provided in HYPRO to carry out some tasks which are frequently performed by the hydrologists. The limited modules provided at present may also be considered more as demonstrators; indeed the list can be readily extended. It is planned to add more such modules for basic analysis. Currently, hydrological analysis includes the following options:

- Statistical Summary
- Time Series Analysis
- Frequency Analysis

2.5.1 Statistical Summary

This option provides values of several statistical properties of selected data series. The following parameters are computed : Mean Standard Deviation, Skewness Coefficient, Kurtosis, Series Correlation Coefficient, and Maximum and Minimum of the selected data series.

2.5.2 Time Series Analysis

A time series is a set of observations generated sequentially in time. Typical examples include the time series of discharge, rainfall, and groundwater levels. These data are frequently analyzed using the time series techniques.

2.5.3 Frequency Analysis

Frequency analysis consists of fitting various probability distributions to the hydrological data. This analysis is usually carried out to determine design flood corresponding to various return periods. This option has been provided in HYPRO to assist the designer in obtaining necessary information from the observed annual peak floods. The following features have been provided in HYPRO.

- Computation of basic statistical summary of data either in natural or log domain or both.
- The user has the option to go through the analysis for outlier/inlier, view the modifications and either accept or reject the analysis before going to the next step.
- Analysis for checking the persistence structure of data is available to the user.
- It allows the user to choose from amongst potential distribution with different parameter estimation techniques.
- Quantile estimates together with their standard error at specified return periods, with facility to compute quantiles to additionally specified return periods, are available to the user.
- Options are also available to the user to graphically plot the quantiles.

2.6 REPORTING & RETRIEVAL

This module includes the following facilities :

- Preparation of reports on station and series characteristics and time series.
- Retrieval of desired data stored in the database.
- Transfer of data.

The options are :

Station Details : Using this option, the details of all the stations can be obtained, on the basis of following:

- Station Code
- Station Name
- River
- District
- Catchment Area
- Latitude
- Longitude
- Altitude
- Country
- Agency

If the value of the selected item is of character type than the system will ask for only one value. Otherwise, it will ask for a range for the value. The details of the stations which will satisfy the criterion will be displayed.

Data Series : The system will prompt for the station code. If the station code has already been defined, it will ask for the series code. If the first character of the station/series code is a blank character or a carriage return, the control will come out of the option. Otherwise, the HYPRO will ask for the output filename, starting date and ending date. Here the ending date should not be less than the starting date. After specifying all these options, the data for the desired station, for the selected period will be stored in the desired file in ASCII format.

2.7 DOS COMMANDS

Sometimes it is required to execute a DOS command with in HYPRO package. Using this option the user can execute a DOS command within HYPRO and continue his processing. However, no TSR routines of DOS should be loaded as this may result in conflict in memory usage and the computer may hang.

3.0 USING THE SPECIFIC KEYS WITHIN HYPRO

A number of special, hot keys can be used with in HYPRO. The details are as follows :

3.1 SPECIAL KEYS FOR THE MENU SCREEN

- **Cursor Keys** : These keys enable the user to walk through the choices to be made.

- **Return Key** : This key confirms the choice. With this key the system goes to the fetch screen or the corresponding submenu, depending upon the case.

- **F10 Key** : This key can be used if the user wants to go back from a submenu to the corresponding parent menu. If the user is in main menu, then this key moves the cursor to the first choice in the main menu.

3.2 SPECIAL KEYS FOR THE 'FILLING-IN' SCREEN

- **Cursor Keys** : The left and right cursor keys can be used to shift the cursor within a certain field. The up and down cursor keys can be used to jump from one input field to another input field. This applies only to the input fields which can be seen in position one of the top of the other.

- **Ins key** : This key (de-)activates or toggles the 'insert character function', within an input field.

- **Del Key** : This key acts as 'delete character' within an input field.

- **Ctrl-Home Key:** This key jumps the cursor directly to the first input field.

- **Ctrl-End** : This key jumps the cursor directly to the last input field.

- **End** : This key jumps the cursor to the last character of the current field.

- **Home** : This key jumps the cursor to the beginning of the current field.

- **Backspace Key:** This key deletes the character to the left of the cursor. If the cursor is on the first character in a field, this key deletes the last character of the previous field.

- **Return Key** : This key finishes the operation in an input screen and the cursor will automatically jump to the next input screen.

- **PgUp Key** : This key can be used to leave a 'fill-in' screen. Data that has already been typed is stored. The user will now return to last selection screen.

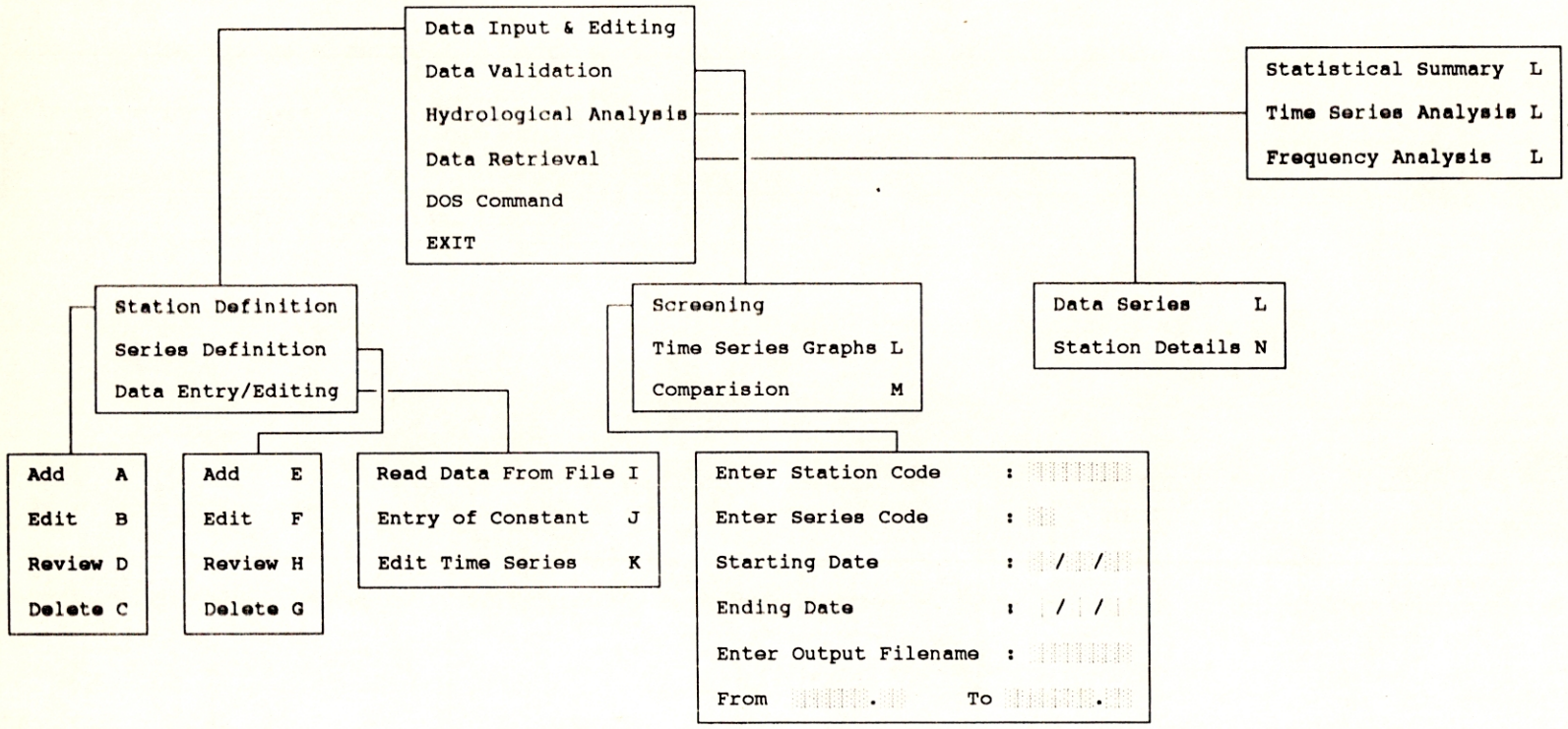
- **PgDn Key** : This key has same function as PgUp screen.

- **Esc Key** : This key has same function as PgUp Key.

EPILOGUE:

The HYPRO presents a framework and a number of modules for data entry, storage and processing of hydrological data. The capabilities of version 1.0 of this package have been described above. The system is under continuous development and it is planned to add/ enhance the modules from time-to-time.

Fig.1 Flow Diagram of HYPRO



A

```
Station Code    :
Station Name    :
River           :
District        :
Country         :
Latitude        :   °   '   " N
Longitude       :   °   '   " E
Altitude        :       .     (m)
Catchment Area  :       .     sq.km
Agency          :

 Do You Want to Add More :
```

B

```
Station Code    : NIH-AHS
Station Name    : N.I.H.,Roorkee
River           : Solani
District        : Hardwar
Country         : India
Latitude        : 29°52'0 " N
Longitude       : 77°53'52" E
Altitude        :  265.00 (m)
Catchment Area  :    1000.00 sq.km
Agency          : Self

   Do You Want to Edit it :
```

C

```
Station Code    : NIH-AHS
Station Name    : N.I.H.,Roorkee
River           : Solani
District        : Hardwar
Country         : India
Latitude        : 29°52'0 " N
Longitude       : 77°53'52" E
Altitude        :  265.00 (m)
Catchment Area  :    1000.00 sq.km
Agency          : Self

   Do You Want to Delete it :
```

D

Review of Available Staions

Station Code	Station Name	Available Series	Start Date	Ending Date
NIH-AHS	N.I.H.,Roorkee	G1	01/01/80	31/05/85
		G2	01/01/80	31/12/81
UOR-1	Univ.of Roorkee	PR	01/03/90	31/12/95

E

```
Series Code     :
Time Unit       :
Divider         :
Description     :
Unit            :
Missing Value   :
Minimum Value   :     .
Maximum Value   :     .

 Do You Want to Add More :
```

F

```
Series Code     : G1
Time Unit       : 1
Divider         : 1
Description     : Annual GW Data
Unit            : m
Missing Value   : -99
Minimum Value   : 0.00
Maximum Value   : 1000.00

 Do You Want to Edit it :
```

G

```
Series Code     : G1
Time Unit       : 1
Divider         : 1
Description     : Annual GW Data
Unit            : m
Missing Value   : -99
Minimum Value   : 0.00
Maximum Value   : 1000.00

 Do You Want to Delete it :
```

H

Review of Available Series				
Series Code	Description	Unit	Time Unit	Divider
G1	Annual GW Data	m	1	1
RH	Daily Humidity	%	3	1

I

```
Enter Station Code :
Enter Series Code  :
Enter Filename     :
```

J

```
Enter Station Code :
Enter Series Code  :
Starting Date      :  /  /
Ending Date        :  /  /
Constant           :        .

IF YOU WANT TO QUIT PRESS<ESC>
```

K

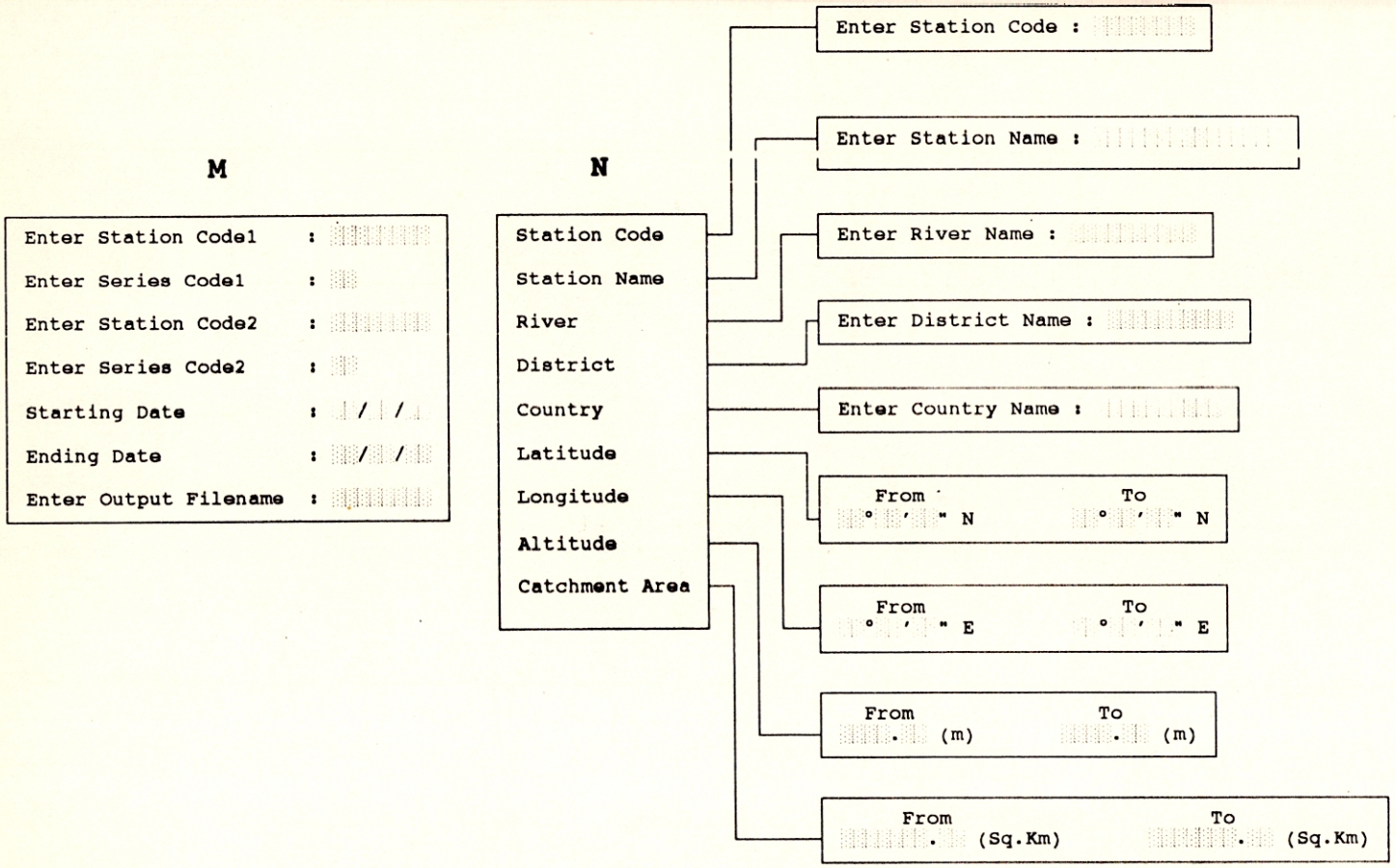
```
Enter Station Code :
Enter Series Code  :
Starting Date      :  /  /
Ending Date        :  /  /

IF YOU WANT TO QUIT PRESS<ESC>
```

L

```
Enter Station Code  :
Enter Series Code   :
Starting Date       :  /  /
Ending Date         :  /  /
Enter Output Filename :
```

Station NIH	Equidistant Time Series				Year - 1995	
Value1	######.##	######.##	######.##	######.##	######.##	######.##
Value7	######.##	######.##	######.##	######.##	######.##	######.##
Value13	######.##	######.##	######.##	######.##	######.##	######.##
Value19	######.##	######.##	######.##	######.##	######.##	######.##
Value25	######.##	######.##	######.##	######.##	######.##	######.##


M
Enter Station Code1 :
Enter Series Code1 :
Enter Station Code2 :
Enter Series Code2 :
Starting Date : / /
Ending Date : / /
Enter Output Filename :
N
Station Code
Station Name
River
District
Country
Latitude
Longitude
Altitude
Catchment Area
Enter Station Code :
Enter Station Name :
Enter River Name :
Enter District Name :
Enter Country Name :
From ° ' " N
To ° ' " N
From ° ' " E
To ° ' " E
From . (m)
To . (m)
From . (Sq.Km)
To . (Sq.Km)