SR—28

STATUS REPORT

ON

PROCESSING AND ANALYSIS OF HYDRO - METEOROLOGICAL DATA

NATIONAL INSTITUTE OF HYDROLOGY

HARD ROCK REGIONAL CENTRE

BELGAUM

March 1992

PREFACE

Hydrology is an applied natural science and therefore hydrologists have to deal with vast amount of data. In recent years, there has been a significant increase in the volume of hydrological data collected. The successful and efficient execution of modern hydrological studies depend on a vast and diverse amount of information. Hence it is essential to organise a systematic body of information in the form of a data base system to fulfill the data needs for the quantitative hydrological investigation. Accuracy of these investigations / studies depend upon the quality of the data used for it. Hence the need of data quality control/processing arises before the actual analysis starts.

The comprehensive study of hydrological processes and data processing techniques is the first step in organising an information base. This includes the rapidly developing area of real time processing which is increasingly used in the operation of hydrological systems.

Prior to the storage of the inflowing data, it is essential that the information be tested, suitable correction be made and thus some measure of its quality established. Quality control, editing, analysis and synthesis of the collected data are collectively called as data processing. Among these the former two processes together called primary processing and the latter, secondary processing.

This report prepared in connection with the ongoing UNDP project, presents the procedures involved in data processing, data processing procedure for different types of hydrological data and the status of data processing techniques in and outside India. This report has been prepared by Sh.Chandramohan T, Scientist B at N I H Regional Centre, Belgaum.

Satish Chandra

(SATISH CHANDRA)
DIRECTOR

CONTENTS

LIST OF FIGURES AND TABLES

1.Ø INTRODUCTION

The concept of the hydrologic cycle forms the basis for the hydrologist's understanding of the source of water at or under the Earth's surface and its consequent movement by various pathways back to the principal storage in the ocean. Two of the greatest problems for the hydrologist are quantifying the amount of water in the different phases in the cycle and evaluating the rate of transfer of water from one phase to another within the cycle. Thus measurement within the cycle is a major function.

Most hydrological variables such as rainfall, stream flow or groundwater have been measured for many years by separate official bodies and private organisations. But the installation of the gauges for rainfall or streamflow has usually been made to serve a single, simple purpose. Now-a-days, with the growth of population and the improvement in communication to serve modern needs, hydrometric schemes are tending to become multipurpose. Moreover, advances in scientific hydrology and the practice of engineering hydrology are dependent on good, reliable and continuous measurements of the hydrological variables. These measurements are recorded by a wide range of methods, from the simple writing down of a number by a single observer to the invisible marking of electronic impulses on a magnetic tape. A hydrologist often finds that his first job is to receive and process information assembled by any of these many different methods.

Once data are collected, the next important and necessary step is their proper storage. In the conventional ways, the data are mostly kept in manuscript form in registers or files. This form of data storage is easiest, does not require any technical skill and independent of such factors like capital investment, availability of sophisticated machines and electricity. Usually no pre-processing is done before the data are stored in the registers/files. With the advancement of technology, computer memories, magnetic tapes, microfilms, etc. have replaced these conventional storage media. In the modern procedure, the data are stored only after a particular level of processing.

Organisation and analysis of hydrological data are essential for development, planning and the associated tasks of project design and operation. The consequent need for comprehensive data processing systems for hydrological data has been widely recognized, and is invariably stated as an objective of all agencies involved in data collection and analysis. Despite these intentions, the successful design and implementation of such systems has been relatively limited, particularly in developing nations.

Most countries have well established and expanding hydrological networks, producing correspondingly increasing quantity of data. However, because of the shortage of suitably trained personnel and the inevitable pressure to perform urgent work at the expense of routine data processing, much of the collected data remains unprocessed. The larger the backlog of raw data, the more difficult is the task of data storage and analysis. This results in manual processing techniques becoming totally impractical, except for limited individual studies.

Such a situation represents a mis-direction of resources into the data collection effort, and leads to a steady deterioration in data reliability as no overall quality control is being performed. This introduces broad limits of uncertainty into the design, leading at best to over investment, and at worst to operational failure. The introduction of automated or electronic data processing is essential where this situation already exists, but is equally valuable in preventing it in places where significant amounts of data are only now becoming available.

The aim of data processing is to manipulate the raw data and to put it in a proper form and extract the required information from it. Basically, all the processing activities are divided into primary and secondary processing. The primary processing consists of data validation, i.e. removing the random and systematic errors if any from the raw data and putting it in a proper format. The secondary processing includes deriving some information from the data such as fixing rating curves using gauge-discharge data or finding the return period of a particular flood using the annual flood peak data.

2.Ø DATA PROCESSING

The past few decades have witnessed tremendous growth of utilisation and management of water resources. The successful and efficient execution of these studies/decisions requires a vast amount of data of interrelated phenomenon. The accuracy of results obtained depends upon the quality and quantity of information used. This information is provided by an information system, which cover, the data operation of the following kind: (1) Data collection and (2) Data processing. Generally a large amount of data is involved, making a proper tuning of all phases in the collection and processing of data a necessity.

Data processing may be defined as any systematic procedure through which basic information is transposed into more accessible or more directly usable forms. It covers all activities after transmission upto and including data dissemination. It comprises editing and organising functions, (detecting and correcting errors, classifying and indexing records, sorting and collating data prior to use, etc.), carrying out necessary calculations and transcribing both data and results in prescribed format to a suitable storage and processing medium, not necessarily in that order. Fig.(1) shows a flow chart of general data processing network.

2.1 Data Processing Cycle

Data processing consists of three basic steps; input, processing, and output.

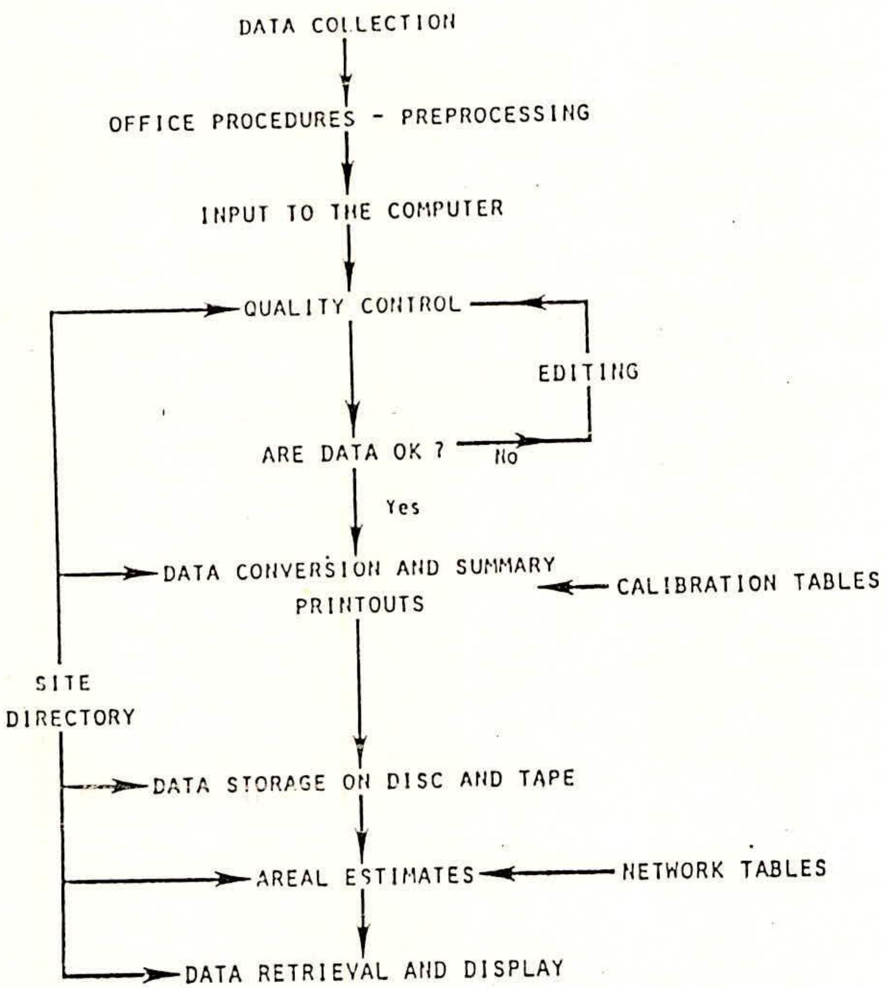


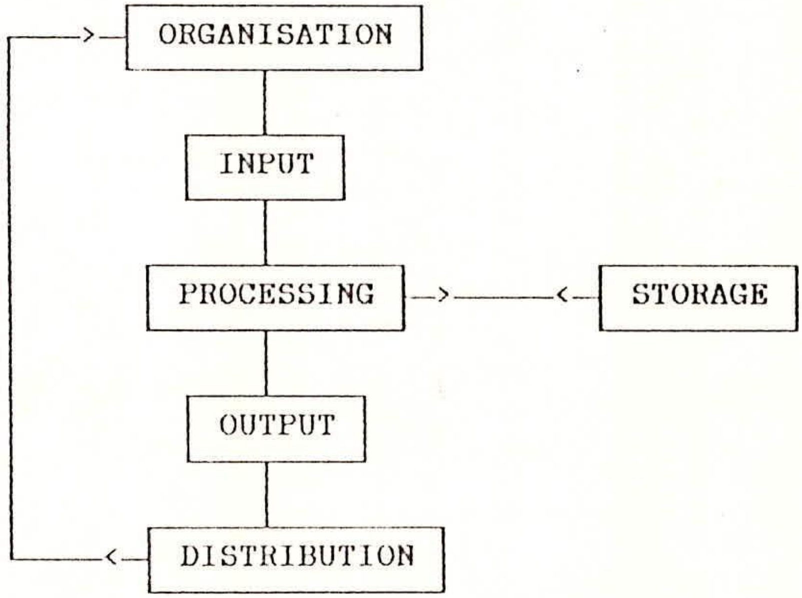
FIG 1 : FLOWCHART SHOWING THE GENERAL DATA PROCESSING FRAMEWORK

Input: In this step, the initial data are prepared in some convenient form for processing. The form will depend on the processing machine. When electro-mechanical devices are used, the input data are punched on cards, but if electronic computers are used, the input data could be recorded on any one of several types of input medium, such as cards, tapes, etc.

Processing: In this step, the input data are changed and usually combined with other information to produce data in a more useful form. This step usually involves a sequence of certain basic processing operations.

Output: Here the results of the processing steps are collected. The particular form of the output depends on the use of the data.

Frequently, three more steps are added to the basic data processing cycle to obtain the expanded data processing cycle.



Organisation: It is the process of collecting the original data. An original recording of data is called a source document.

Distribution: This refers to the distribution of the output data. Recordings of the output data are often called report documents. Sometimes, the report documents may become the source document for future data processing.

Storage: This step is crucial in many data processing procedures. Data processing results are frequently placed in storage to be used as input data for further processing at a later date. A unified set of data in storage is called a file. A collection of related files is called a data base.

2.2 Data Processing Operation

A general data processing procedure normally consists of a number of basic processing operations performed in some order:

Recording: It refers to the transfer of data onto some form or document. It occurs not only during the origination step and during the distribution step but throughout the processing cycle.

Duplicating: Consists of reproducing the data into many forms or documents. It may be done while the data are being recorded manually, or afterwards by some machine.

Validation: Since recording is usually a manual operation, it is important that recorded data be carefully checked for any errors. (using graphical and/or statistical techniques and inter-station comparison).

Correction: Erroneous data are corrected or replaced by missing value indication.

Filling in of missing data: Using interpolation or regression techniques.

Classifying: This operation separates data into various categories.

Sorting: Arranging data in a specific order. The data item which determines the sorting is called the key.

Merging: This operation takes two or more sets of data, all sets having been sorted by the same key, and puts them together to form a single sorted set of data.

Calculating, or performing numerical calculations on the data.

Table look up, searching, retrieving: Refers to finding a specific data item in a sorted collection of data.

Summarising and report writing: A collection of data condensed and certain conclusions from the data are represented in simple, clear forms.

The development of modern computer systems modified the data processing procedure also accordingly. The computer aided data processing procedure can be discussed as follows. Table 1 shows the main components of a computer aided data processing system.

Data preparation: This activity comprises the operations necessary to convert data from the format in which it is received to a format suitable for input to the computer.

Data entry: There has recently been a major shift in methods of data entry. Punched cards and paper tapes have been widely replaced by key-to-tape or key-to-diskette systems. It comprises

TABLE 1

<u>The components of data processing</u>

	DATA PROCESSING						
DATA PREPARATION	DATA ENTRY (INPUT)	VALIDATION	PRIMARY PROCESSING	DATA BASE UPDATING	SECONDARY PROCESSING	RETRIEVAL	OUTPUT
Prepare punching documents by: 1. Transcription Field note book entries Non-standard data formats 2. Coding Reduction/ standardisation of input data	1. Punching document a. Direct keying through VDU b. Keying onto computer compatible media Punched cards Key-to-tape Key-to-disk 2. Charts and maps Direct input by digitiser 3. Computer compatible media a. Tapes/cassette b. Diskettes c. Cards d. Communication lines (telemetred data) e. Mark sense/ optical character readers	1. Range checks 2. Sum checks 3. Inter-station consistency checks	1. Standardisation of units 2. Calculation of derived parameters 3. Further coding of input to reduce storage requirements 4. Arranging data in data base format	1. Add new data sets to existing data base 2. Report any errors	1. Programs for routine reports 2. Statistical summaries 3. Infilling missing data values 4. Interpolation or aggregation of data	1. Selection of data by: a. Parameter type b. Parameter value c. Location d. Period of record e. Time interval of record 2. Selection of output device	1. Printers 2. Plotters 3. VDU 4. Computer storage media 5. Microfilm 6. Telemetry
	ERROR CORRECTION						

two stages; in the first (punching) stage data is keyed by one operator, in the second (verifying) stage the same data is keyed by another operator and compared with the original input. There are other more specialised data entry methods, eg:. the conversion of charts and maps to digital format using manual or automatic.

Data validation: Once entered into the computer, data should be subjected to a set of checks designed to identify incorrect or unusual data values.

Primary processing: This stage of processing is concerned with preparing and assembling the data in the format necessary for it to be added to the existing data base. This may include standardisation of measurement units, additional levels coding for storage purposes, and estimation of derived parameters.

Data base updating: After primary processing, data may be incorporated into the data base using an updating program. It is usual to update the data base at some fixed time interval, monthly being a typical interval for natural resources planning data.

Secondary processing: This stage covers the reports generated and analyses performed on a routine basis after the data base has been updated.

Retrieval: An important objective of an automated data processing system is to provide a rapid and comprehensive response to adhoc requests for data retrieval and interpretation. Thus a wide range of retrieval options should be available to allow the interrogation of the data base and the selection of ranges and combinations of data.

Output: Modern computers support a wide range of output devices. These include printers, plotters, and cathode ray tubes or visual display units. Devices supporting both graphics and colour options are of increasing use.

2.3 Data Processing Media

Before automatic processing can begin, the data have first to be recorded on one of the available media in the appropriate machine language (code).

In the conventional way, data are mostly kept in registers or files and whenever demanded, they are manually copied and supplied. The punched cards may be used for storing the data but this has a number of limitations like the cards may be damaged, visual checking is difficult and time consuming and proper sequence is to be maintained. There are two main media for mass storage, magnetic disks and magnetic tapes. An important advantage of these mass storage media is that they are exchangeable. Thus, an infinite number of data sets and programmes can be made available for the CPU for processing.

Punched cards, punched paper tape and magnetic tape are used extensively in data processing, and punched cards, magnetic tape and digitally coded microfilms are used for automatic data storage and retrieval. Manuscript records, autographic records, charts and printed paper documents are time honoured media for the storage and retrieval of data. Recently equipments has been developed for automatic retrieval from autographic records and printed documents. To ensure better permanence and more compact, cheaper storage and reproduction, the practice of microfilming paper records has grown in importance.

An ideal medium will be:

a) durable and non-erasable, to ensure permanence of the data,

b) flexible, to allow easy arranging, editing and modifying of the data,

c) versatile, to allow ready use for all purposes and by all means, from visual browsing and manual manipulation, through automatic processing by the entire range of data processing machines, and

d) compact, to allow accumulation and storage over long periods in manageable proportions and at reasonable costs.

2.4 Data Processing Equipment

They may be broadly classified either by function (recorders, sorters, tabulators, calculators, etc.) or by mode of operation (manual, mechanical, electronic).

Manual operations are usually slower and less reliable than mechanical or electro-mechanical, and the latter more so than electronic. But the sort of equipment used and the precise point at which a change is made from one category to another depends on many factors like bulk of the data to be dealt with, the sort and urgency of work to be done from it, the cost of labour, material, office space and machine in the area in question, the cost and availability of maintenance services, the rate of obsolescence of equipment, etc.

1. Punch card machines:

a) Key punch and verifying machines: The transcription from the observational records to cards is done on key board operated machines, one column at a time proceeding from left to right on the card. The power for operating the simplest type is supplied manually by the operator, who also feeds the cards in one at a time. Electrically driven punches are also available with automatic card feed.

Verifiers are similar to key punches in construction and operation but are designed to test for the presence or absence of holes rather than to cut them.

b) Reproducing Machines: These are primarily for punching data from one set of cards to another, usually with automatic checking as the copies are made. All or part of each card can be reproduced and relocating the data fields is also possible if need be. Some are also capable of comparing the punching on separate sets of cards even when their field locations do not correspond.

c) Sorting and collating machines: Sorters are used for rearranging cards in any desired sequence and for segregating or grouping particular categories of data.

Collators are designed to expedite and supplement the work of sorters. They combine the dual function of directing the flow of cards from one or two feeding hoppers into one of several pockets, and determining which is the larger of two numbers, either on the same card, on two successive cards in the same set or on two cards from separate sets.

d) Tabulating machine: Generally they can be programmed to print information in tabular form, add, subtract, or count in the accumulator and in some cases multiply, and classify or select data by means of card-controlled electromagnetic switches. Arithmetic and printing operations can also be carried out on the data under the control of designators punched in the card themselves.

e) Statistical and calculating machines: Statistical machines combine the functions of the sorter, the collator and the tabulator. This is economical compared to tabulator when much counting and little printing need to be done.

Calculating machines are usually limited to arithmetic work, although the larger machines have some logical capabilities as well.

2. Punch tape machine:

a) Key board perforators: The transcription to paper tape in the appropriate code is done on a key punch or key board perforator. This is electrically powered.

b) Tape readers and re-perforators: Automatic tape readers or transmitters are motor driven devices for translating perforations on tape into electrical impulses.

Re-perforators automatically re-convert coded electrical impulses back into holes.

c) Tape verifiers: A tape is prepared using the key punch and is then placed in a reader. The operator then proceeds to punch the data again. The key board, the reader and the perforator are so linked that if the code combination for the depressed key corresponds to that sensed by the reader, it is punched in the new tape. If not, further punching is inhibited and the key is locked down until the operator releases it after deciding what is wrong.

3. Special and Peripheral machines:

a) Analogue-to-digital converters: Digitizers accept the analogue output from observing instruments in the form of voltages, pressures, temperatures, shaft positions, rate counts, measurements of weight or length, etc. This output is converted to digital form for recording on cards or tapes.

b) Digital-to-analogue converters: There are several automatic line drawing or graph plotting (like frequency curves, time or space cross-sections, contour charts, etc.) machines commercially available which accept digital input from cards or tape, or directly from computers.

c) Digital-to-digital converters: Converting digital information from one medium to another (including microfilms).

4. Digital Electronics Computers and Electronic Data Processing (EDP) Systems:

The most fundamental advantages of computerised processing systems, and perhaps still the most significant are:

- Data can be stored in a compact, organized manner
- Data may be rapidly retrieved in a wide variety of formats and aggregation.

With respect to subsequent data application, there are other important benefits. The computer´s ability to perform systematic consistency checks offers a powerful aid for improving and maintaining the quality of data and data analysis techniques may be much more sophisticated than those found in any routine manual processing systems.

The first computer capable of automatically performing a long sequence of arithmetical and logical operations came into existence in 1944. It was an electromechanical (rely) device. Subsequently, first generation computers with vacuum tube in place of relays and capable of storing programs were developed. In second generation computers (1959-65), solid state devices (transistors) were used in place of vacuum tubes and it was having magnetic core storage. In this period computers became much smaller in size, faster, more reliable and much greater in processing capacity.

Introduction of integrated solid state circuitry (I.C.) gave improved secondary storage devices and new input/output devices (visual display terminals, magnetic ink readers, high speed printers, etc.) for the third generation computers (1965-70). The primary storage unit, or memory, augmented by secondary storage device located outside the computer proper, together with faster input and output devices made possible multiprocessing and multiprogramming, whereby a number of data processing problems from different sources could be run virtually at the same time on a single centrally located computer.

In fourth generation computers (since 1970), a microprocessor, which is a tiny solid state device which in itself is a small computer capable of performing arithmetic and logical operations, is introduced. It resulted in further improvement of secondary storage and input/output devices. Among the advanced input/output devices employed in fourth generation computers are optical readers by which an operator can visually introduce data or instructions and graphic display terminals by which an operator can feed pictures into the computer.

2.5 Data Processing Systems and Centres

A few or many of the large range of data processing devices and machines available can be variously combined so as to constitute a system designed to carry out one or more specific tasks, simultaneously or in succession. By merely multiplying machines and operators, a large elaborate system capable of a heavy anticipated schedule or work can be built up. But then, apart from consideration of space economy, the problems of management and control tend to grow disproportionately. Modern electronic data processing (EDP) systems are designated to overcome these difficulties by concentrating a number of high speed ancillaries (printers, tapes, card readers, etc.) round a single computer, capable of carrying out several functions and several programs concurrently.

If a country is large and climatically diverse and/or there are severe communication difficulties, it may be better to maintain several smaller data processing centres rather than one main centre. But it will generally be more economical to set up a single centre with the possibilities of fully integrated working and for more efficient use of staff and equipment that this offers, by concentrating several activities there.

2.6 Quality Control

Data quality has always been a limiting factor in hydrology. The advent of digital computers has made possible the development of complex hydrological models which have significantly improved the ability to forecast stream flow. This has also led to a growth in the volume of data through improvement in the gauge network and increase in the frequency of observations. However, often the data collected are not accurate enough for efficient use. With the modern computers having the capability to receive data directly from sensors in the field, manual quality control has become impracticable.

The term quality control is used here to denote the steps that can be taken to ensure that good quality basic data emerge from the observation, collection and processing procedures. Such procedures are an essential part of any data collection programme. These main steps include inspection of stations, preliminary checking of data and detection of errors by internal consistency checks. The quality control of data for one element of the hydrologic cycle can frequently be improved by comparisons with data on an allied element.

1. Inspection of stations: It is essential, in maintaining good quality observation, that stations be inspected periodically. WMO technical regulations state that members should arrange for inspection of principal climatological stations at least once every two years and ordinary climatological stations at least once every four years. For a stream gauging station such inspections are made to check the stability of the rating curve,

and to check the relations between the gauges and permanent level reference points to verify that no movements of the gauges have taken place. In any event every stream gauging station should be inspected at least twice a year.

2. Preliminary checking of the data: The difference between preliminary checking and error detection is rather arbitrary. Procedures included under preliminary checking in one country may be thought of as error detection in another, and vice versa. For data collected manually and latter transferred to machine readable form, the term preliminary checking will be used to cover those procedures performed prior to transcribing the data into machine readable form. For data collected directly in digital machine readable form, there may be little possible checking prior to the first machine processing. Under these conditions, the quality control of data by means of computer is the only possible method of control.

For data collected manually, preliminary checking should generally include:

a) Logging of date of receipt of report form;
b) Ensuring completeness and correctness of indicative information;
c) Ensuring completeness of data for station, including totals, means and extremes where required;
d) Checking observer's arithmetic in computing totals and means.

In many countries this last step may be done by machine methods. Corrections should be entered legibly and in ink of a different colour from that used in completion of the original form, making sure that the original entry is not erased or made illegible. At times data will be missing from observer's report, and in certain circumstances the missing observations may be estimated or interpolated.

3. Error detection: The efficiency of quality control procedure depends greatly on whether or not machine processing facilities are available, and if so whether only simple punched card tabulating equipment or more comprehensive digital computer facilities are available.

Techniques for quality control of data differ for various elements. The basis of most quality control procedures of manually collected temperatures and precipitation data are machine tabulated arrays of daily data by district or region. From such arrays, it is easy to detect, by eye, stations at which the data are consistently credited to the wrong day or gross errors in temperature or precipitation measurements.

Obvious discrepancies should be followed up by a study of the original report from the stations, by a check on the history of the station as to its quality of record, and by an appraisal of whether the synoptic weather situation which produced heavy rain at most of the stations on one day could have reasonably been expected to miss the station in question. All of these factors must be evaluated before an apparent error is corrected. In some countries computer programmes have been prepared for quality control of precipitation data. The plotting on a map of data can also be of assistance in detecting errors, in a station with a serious error in precipitation, temperature, etc. can be readily detected.

The quality of computed stream flow records may be checked by plotting daily values of discharge against time and comparing this plot with similar graphs for nearby streams and with observed values of other parameters such as precipitation and temperature. The routing of flood hydrograph is a useful technique in checking flood discharge where several stations are operated on the same stream.

2.7 Hydrological Analysis

Hydrological analysis is related to well established principles of hydrodynamics and thermodynamics. The central problem is the application of these principles in a natural environment which is irregular, sparsely samples, and only partially known. The event samples are usually unplanned and uncontrolled. Analyses are performed to obtain average aerial values of certain elements, regional generalization, frequency distributions, and relationships among variables. Often the pertinent elements are not or cannot be measured directly.

Evaporation from a lake is an element which cannot be measured directly. It cannot be caught in gauges, nor concentrated in a channel as with streamflow. The problem of aerial sampling is exemplified by estimating rainfall over a drainage area, where only a small portion is caught in gauges and actually measured. The problem is to interpolate the rainfall between gauges.

In relating rainfall to runoff, the hydrograph of runoff is divided into components, so that the portion associated with a particular rain event is separated from the rest of the hydrograph. This separation is achieved by computation, based on analytical models, rather than by a physical measurement.

Analyses include case studies and statistical examination of masses of data. Statistical analyses include fitting of data to frequency distributions and to theoretical models by regression, or time series analysis methods. The validity of derived relationships is often tested on independent data; hydrograph reconstitution is a typical hydrological test.

The degree of detail and of precision of the analysis should be consistent with the quality and sampling adequacy of the data used, and with the degree of precision required by the application of the analysis. Consideration should be given to the relationship between the cost and time devoted to an analysis, and the benefits expected and time required. In many instances, graphical and relatively simple computational methods are more economical than more sophisticated methods, and are sufficiently precise for the data and purposes involved.

Analyses may serve one or more of several purposes, such as network design, evaluation of control measures, forecasting of streamflow, and planning of programmes for managing water resources.

3.Ø PROCESSING AND ANALYSIS OF HYDROLOGICAL DATA

Satisfactory solutions to many water resources problems require readily accessible, reliable observational data on the elements of the hydrological cycle and related factors. The collection,processing and storage of such data is essential in order to obtain a long term picture of events. Typical examples of data handled in this way are rainfall, evaporation, river stage and discharge, and ground water levels. Such data have many applications for drought and river flood investigation and for application for the design of reservoirs and flood protection schemes.

It is important to establish the various observational networks on an integrated basis, particularly the precipitation and stream flow networks. The measurement stations have to be so located that the data collected will be useful in developing relation between the hydrological and meteorological factors and the significant physical parameters such as slope, elevation, morphology, geology, land use and soil types.

It is not only necessary to observe the data, but also equally necessary to ensure that good quality basic data emerge from the observation, collection and processing procedures. Such procedures form an essential part of any data collection program and include inspection of stations, preliminary checking of data and detection of errors by internal consistency checks. In many cases it is desirable to subject basic observational data to some analysis before or immediately after publication to put the data in a form most useful to the users. Statistical procedures are used for the purpose.

3.1 Precipitation Data

It is common experience that the precipitation data in its raw form would contain many gaps and inconsistent values. As such preliminary processing of precipitation data is essential before it is put to further use in analysis.

The processing system consists of a series of steps and procedures. The efficiency, economy, and speed of the system would depend upon the type of storage devices, the quality of machine and software. The methodology of executing the various steps involved in the processing system are briefly described.

1. Preliminary Scrutiny: Before the precipitation data is stored on computer compatible devices for computer processing, it becomes necessary to carry out preliminary checks, manual sorting, etc. The reports received from manually observed stations by telephone or other communication channels are checked by a repeat back system. Improper registering of data includes entering data against wrong time and date, alteration of figures, etc. The official at receiving station could check the reasonableness of report by judging the report based on past experience and statistics of the station and region to which the station belongs.

Some of the climatological parameters used for checking are the values of normal rainfall, highest observed rainfall, or value of rainfall corresponding to 25, 50 or 100 year return period.

2. Quality Control: Quality control is a pre-requisite before the precipitation data are used either in an operational system for flood forecasting or archived for climatological purposes. The basic objective of the quality control procedure is to detect and if possible correct errors in observational data at the earliest stage possible in the flow of data from local data source to the centralised data base.

Most of the errors could be classified as systematic errors and random errors. Systematic errors are essentially due to malfunctioning of instrument, wrong exposure conditions and/or lack of knowledge of observer. Random errors could arise due to spilling of the water when transferring it to the measuring jar, leakage into or out of the receiver, observational error, etc.

3. Estimation of missing data: While retrieving data for climatological purpose or inputting data in real time, one often comes across missing data situations. Since blank in a data set is read as zero by computer, necessary software for identifying the blanks and marking them appropriately, need to be developed.

Data for the period of missing rainfall data could be filled using estimation technique. The length of period upto which the data could be filled is independent on individual judgment. Rainfall for the missing period is estimated either by using the normal ratio method or the distance power method.

a) Normal ratio method - In this method, the rainfall at a station is estimated as a function of the normal monthly or annual rainfall of the station under question and those of the neighbouring stations for the period of missing data at the station under question.

b) Distance power method - In this method, the rainfall at a station is estimated as a weighted average of observed rainfall at the neighbouring stations. The weights are taken as the reciprocal of the distance or some power of distance of the estimator stations.

4. Internal consistency check: The internal consistency or self consistency checks are applied by using statistical information based on historical data of the station and current data in case of short duration rainfall.

5. Spatial consistency check: Spatial consistency checks for precipitation data are carried out by relating the observations from surrounding stations for the same duration with the rainfall observed at the station. This is achieved by interpolating the rainfall at the station under question with rainfall data of neighbouring stations.

6. Adjustment of data: To obtain homogeneity among and within measurements of precipitation, adjustment of data becomes necessary. It has two principal objectives. First is to make the record homogeneous with a given environment and the second is to eliminate or reduce extraneous influences by correcting for change in gauge location or exposure. Adjustments for these errors is made by 'Double Mass Analysis'.

It is a graphical method for identifying and adjusting inconsistencies in a station's data by comparing with the trend of reference stations data. In a double mass curve, the accumulated seasonal or annual precipitation values of reference station of stations is taken as abscissa and those of the station under test as ordinate.

A change in the regime of the raingauge such as change in exposure, change in location, etc. is revealed by a change in the slope of the straight line fit. The other records are adjusted by multiplying the precipitation values by the ratio of the slope of the later period to the slope of the earlier period.

7. Data conversion: For hydrologic analysis, rainfall data of shorter duration is required. The network of recording raingauges in India being small in comparison to that of daily (non-recording) raingauges, it becomes necessary to convert the daily rainfall into shorter period intervals either manually or by using appropriate computer routines.

A comparison of the mass curves of the recording raingauge stations with those of the non-recording stations would help in deciding which recording raingauges or group of gauges could be considered as representative of which of the non-recording raingauge for the purpose of distributing daily rainfall into hourly rainfall. A mass curve is a graphical representation of accumulated rainfall vs. time.

For converting the daily rainfall into hourly rainfall, the hourly rainfall from 0800 hr. for consecutive hours is cumulated and the rainfall during hour is expressed as a ratio of the total rainfall during 24 hrs (Ø8ØØ to Ø8ØØ). These ratios are used to distribute the daily rainfall for the corresponding duration at those raingauge stations which are represented by the SRRG.

8. Estimation of Mean Arial Precipitation: Precipitation observations from gauges are point measurements and is characteristics of the precipitation process, exhibits appreciable spatial variation over relatively short distance. An accurate assessment of mean aerial precipitation is a pre-requisite and basic input in the hydrological analysis.

Numerous methods of computing aerial rainfall from point raingauge measurements have been proposed. The most commonly used methods are:

a) Arithmetic average

b) Thiessen polygon method, and

c) Isohyetal method.

The choice of the method is dependent on:

a) quality and nature of data

b) importance of use and required precision

c) availability of time, and

d) availability of computer.

3.2 Climatological Data

When considering general quality control procedure it was noted that validation of climatological data by methods of inter-station comparison was questionable in many cases because of the sparsity of the climatological station network. Thus, the basic validation techniques applied are range checks, rate of range checks, and consistency checks between related parameters observed at the same site. For example, all reported data should be checked or re-computed to see that the dry bulb temperature exceeds or equals a reported wet bulb or dew point temperature and/or relative humidity should be computed and checked against the reported value. Similarly empirical relationships between evaporation pan or lysimeter data and other observed parameters could give broad indications of suspect data at the validation stage.

For all climatological data, station and parameter codes should be tested for validity and, where relevant, sensor calibration values and ranges should be output with suspect values.

A first step in the processing of climatological data is to derive average values, usually on a daily basis, for temperature, relative humidity, vapour pressure and the slope of the vapour pressure curve. An important derived factor is the degree day index, the accumulated departure of temperature from a standard reference temperature. Since this factor is most commonly used in monitoring snow and ice conditions, the reference temperature is normally zero.

There are several climatological parameters which need to be transformed to standard conditions for storage and/or application. For example, wind speed measured at non-standard heights may need to be transformed to a standard 2m height using the wind speed power law. Similarly, pressure measurements may be corrected to correspond to a mean sea level value.

Where direct measurement techniques are used, the computer may be used to verify evaporation estimates by checking the water levels and the water addition/subtraction recorded. In order to compute lake evaporation from pan data, the relevant pan coefficient needs to be applied. In some cases the coefficient is not a fixed value, but must be computed by an algorithm involving other climatological parameters. Pan coefficient or their algorithms must be provided in the station description file.

All methods of calculation and determination of evaporation present difficult problems in interpretation both, for pre and post publication analysis. The physical nature of evaporation from the point of view of mass transfer, energy balance, and water budget must be kept clearly in mind if serious errors are to be avoided.

3.3 Snow and Ice

Even if the water equivalent of falling snow caught in raingauges may be validated along with rainfall data, other snow and ice parameters are more difficult to handle.

Data on the extent of snow cover may only be validated by a time consuming manual synthesis of field observations, aerial survey data and satellite imagery. Techniques to perform automated interpretation of satellite imagery for snow extent are being developed. But there are still problems of differentiating between snow and cloud cover, and of insufficient image resolution.

Data on snow depth and water equivalent again demand much manual validation and verification, integrating from snow courses, snow gauges and conventional precipitation gauges. The large spatial variation in snow cover makes inter-station comparisons difficult. However, there are techniques to estimate the statistical reliability of snow course observations. Under conditions of melting snow, degree day factors are widely used for correlation purposes and, where snowmelt represents a significant proportion of river flow, established relationships between runoff and snow water equivalents may be used. Air and water temperature relationships are valuable not only for the computation of degree day factors, but are also used for assisting in the validation of ice cover and thickness data, and in the forecasting of ice formation and break-up data.

3.4 Streamflow Data

Design, planning and hydrological modelling are some of the important aspects of the water resources projects where the streamflow data are utilized in one form or other. Streamflow data collected from the field as such cannot be utilized in hydrological studies. Therefore, processing of the raw data is extremely essential in order to reduce in a manner to suit it for an analysis.

1. Data Requirements: Generally the streamflow data are required

in the following forms for different hydrological studies:

a) Instantaneous discharge every day, hour or at smaller units,

b) 3 days, 10 days, monthly seasonal and yearly mean discharges,

c) Annual maximum flow,

d) Annual minimum flow,

e) 1 day, 7 days, 10 days, 30 days, seasonal low flows & volumes.

The length of data for use in hydrological simulation studies vary from 10 years to 40 years depending upon the type of project and their use.

2. Compilation and Processing of Streamflow Data:

a) Preliminary processing and Scrutiny - Preliminary processing and scrutiny of the data are essential before the observed data is stored on computer. The preliminary processing includes:

Verification - The reports received from manually observed stations by telephones or other communication channels like wireless need to be checked by a repeat back system.

Valid Status - The station reporting should form part of a standard network accompanied by proper identification with respect to its location.

Reasonable Report - Improper registering of data includes entering against wrong time and dates, alteration of figures, etc. Also transmission errors occur while sending the data either through telegram or wireless.

Quality of Data -

i) Methods of measurement/observation of hydrological data, standards followed, instruments used, frequency of observation, etc. shall be discussed item wise.

ii) Details of history of station, shift in the location, and shift in the rating curves should be identified. Sample calculation for discharge should be furnished. Mention shall be made as to whether discharge data is observed or estimated from the rating curve.

iii)Discuss development of stage-discharge curves at discharge site, the extrapolations shall be verified by other methods such as hydraulic calculations, etc.

Filling up of short data gaps - The following techniques can be used for gap filling.

i) Random choice from values observed for that period.

ii) Interpolation from adjoining values by plotting a smooth hydrograph

iii)Double mass curve techniques

iv) Correlation with adjoining station either of the same hydrologic element or different hydrologic elements

v) Autocorrelation with earlier period at the same station.

Consistency of Data -

i) Internal consistency check: The consistency of the observed data at specific control points and corrections if any made shall be checked and discussed. The check can be done by study of stage-discharge relationship for different periods. Large variation, if any, should be investigated, corrected and explained if required.

Trend analysis should be performed:

- To detect a slow continuous variation of meteorological conditions or a long periodic variation of the climate.
- To observe the modification of catchment physiography especially through human activity.

ii) External consistency check : The consistency of the observed streamflow data should be discussed with reference to the rainfall in the project catchment and observed data in adjacent locations.

Quality control procedures - Some of the methods for quality control are:

- Testing the stage or discharge of a given day within a year against the highest and lowest values of the same data in all the previous years.
- Apply the same test on the difference between the value on the day and the day before.
- Comparing observed data with estimates based on data from adjacent stations. The estimates may be based on regressions. By plotting the estimates, possible errors are easily identified.
- Comparing the observed data with estimates based on a precipitation-runoff relationship.
- Checking for negative values during the computation of inflow to a reservoir when the stage-storage relationship and the outflow are known.
- Comparing the runoff at a station with runoff at upstream stations.
- Applying double mass curve analysis to identify shift in control.

- Applying time series analysis to detect changes in the homogeneity in time series. This is a valuable supplement to double mass analysis.
- Plotting a graph of the points at which measurements are made and comparison with the original cross section.
- Plotting the graph of the annual regime of specific discharges and regional comparison.
- Regional comparison of monthly and annual streamflow deficits.

Adjustment of Records - The adjustment of flows to natural and virgin conditions for historical uses in the upper reaches and the manner in which this has been done should be discussed duly supported by the withdrawal data, reservoir operation data and irrigation statistics. Where adjustments due to upstream storage are made, such storage changes and evaporation losses are to be properly accounted for. Apart from adding upstream withdrawals, return flows have to be subtracted. Adjustment with the flow records is required in those cases where appreciable changes in land use have taken place.

b) Secondary data processing - Specific tasks in secondary data processing include:

- Calculation of mean velocity and discharge based on stream gauging.
- Analytical fitting of stage-discharge relations
- Conversion of stages to discharges
- Preparation of regular time series containing monthly tables of hourly values with means and extremes, annual tables of daily values with means and extremes and

miscellaneous graphs showing variations with time.

- Preparation of chronological tables with elementary statistical parameters, daily data tables for spatial comparison, multi-annual summary tables of monthly and annual value with elementary statistical parameters, discharge classified into ranges and probability envelope curves, characteristic discharges and probability envelope curves, etc.

c) Analysis of processed data - The following analysis are normally performed with the processed data;

- Computation of flow duration curves
- Computation of summation and regulation curves
- Computation of natural runoff from a regulated reservoir
- Computation of the inflow to a reservoir
- Routing of flood through reservoir or river channels
- Unit hydrograph analysis
- Flood forecasting
- Low flow frequency analysis
- Computation of flow frequency curves
- Low flow frequency analysis
- Analysis of flood or low water volume
- Multiple linear regression analysis
- Time series analysis

3.5 Water quality Data

Validation procedures for water quality data are absolute check of analysis codes, relative checks of expected ranges and physical/chemical checks of determined relationships.

If some parameter values have been determined in the laboratory and all the relevant associated data are available to the computer, they may be re-computed for verification purposes.

There are four main areas of activity in the primary processing of water quality data.

1. Verification of laboratory derived values
2. Conversion of measurement units and adjustment of values to some standard reference scale.
3. Computation of water quality indices
4. Mass balance calculations

The standardisation of units is important in obtaining consistency of values stored in the data base. The operations necessary comprise the conversion of measurement units, conversion of the chemical units used, or correction of values to match a reference standard.

Water quality indices are generally based on empirical relationships which attempt to classify pertinent characteristics of water quality for a specified purpose. Thus indices exist for 'suitability for drinking' 'treatability', 'toxicity', 'hardness' etc.

Mass balance calculations are performed to monitor pollution leadings and as a further check on the reliability of water quality data. Leadings are calculated as the product of concentrations and flow.

3.6 Sediment data

Many of the cross-section plotting techniques for river gauging are directly relevant to suspended sediment data validation. The two sets of data may be validated together if both sediment and velocity observations were made.

As with water quality data, mass balance calculations may be performed if sufficient quantities of data exist. A useful check where catchments are reservoired is to test that river sediment loads within a reasonable distance downstream of the reservoir are less than those upstream. If a sediment rating curve exists for section sampled, the departure of the sampled value from the curve may be estimated for its statistical significance and/or plotted for manual sorting.

4.0 STATUS OF DATA PROCESSING IN INDIA

In India, a number of organisations dealing with hydrological data have attempted to develop some computer based systems for data storage and retrieval. These attempts are so far scattered and a consolidated picture is yet to emerge. Moreover, these attempts are so far limited to only Central government departments. Regarding State government departments, who collects bulk of the data, a perceptible beginning is yet to be made. Most of the hydrological data collection activity, especially in states, is still based on manual observations. Usually these data are transferred to various users and they in turn will use them for their purpose, either research or design, without proper processing. Very few users are having proper methodology to detect abnormal values, to fill in missing data, to convert stage to discharge with the use of rating curves, etc.

At present all the precipitation data is collected through India Meteorological Department (IMD), State Departments, Central Water Commission (CWC) and other agencies. More than 6000 raingauge stations are maintained by State governments for collection of rainfall data. The IMD, some of the state irrigation departments, central agencies, universities and agricultural agencies, collect meteorological data through observatories maintained by the respective agencies. The data relating to streamflow and sediment is collected by state irrigation departments, CWC, various river valley authorities, railways, and highway design organisations. There are about 2000 stage discharge sites maintained by various states. The Central Ground Water Board (CGWB) and state ground water departments are the main organisations involved in ground water data collection.

4.1India Meteorological Department (IMD)

IMD has been storing climatological data on punch cards for last four decades. It has assigned a unique 3 digit catchment number to different catchments in India. The daily rainfall data were being punched in 31 cards format until 1970. These are being punched in 24 cards format since 1971 onwards. In 31 cards format, each card contains a catchment No., subdivision No., latitude and longitude of station along with station No. For storage of data in 24 card format, 2 records are needed for each month. The fields in each record are catchment No. latitude, station No, year, month and 16/15 rainfall values.

For hourly rainfall data, the format includes element code, index No of raingauge station, year, month, date, card No (either 1 or 2) and hourly rainfall values. The second card also has filled for amount and duration of maximum one hour precipitation during the 24 hr. period.

The formats for recording daily and hourly rainfall records areas shown in fig.(2) and fig.(3).

In earlier times, processing was generally done manually. The need for use of faster methods has been felt with increase in the number of data records for processing and this led to the use of computers. In 1964 IMD went in for machine processing of

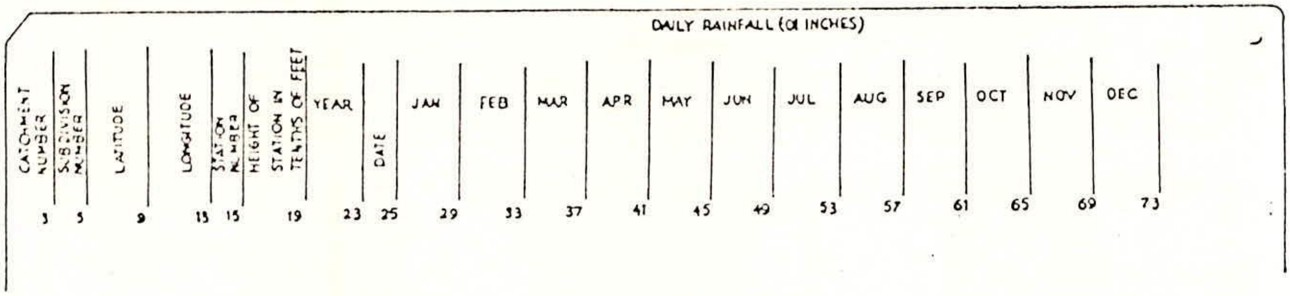


FIG 2a : FORMAT FOR RECORDING DAILY RAINFALL RECORDS (By IMD) - 31 CARD FORMAT

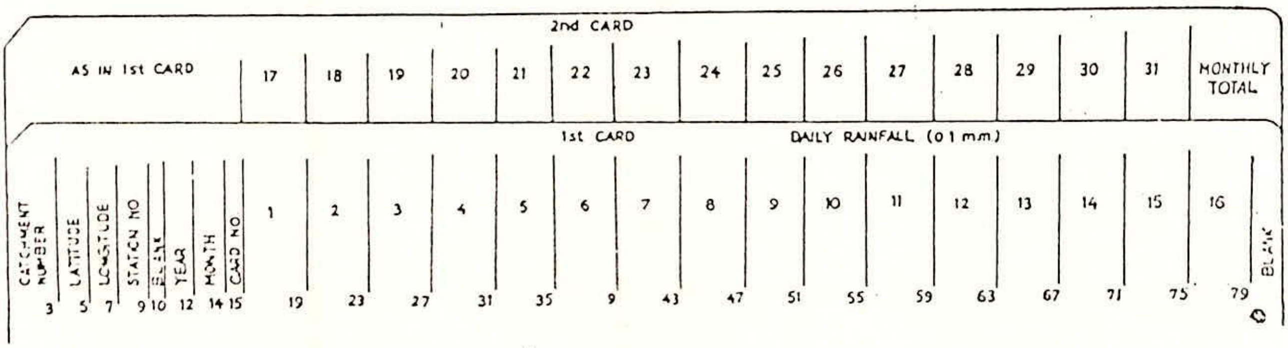


FIG 2b : FORMAT FOR RECORDING DAILY RAINFALL RECORDS (By IMD) - 24 CARD FORMAT

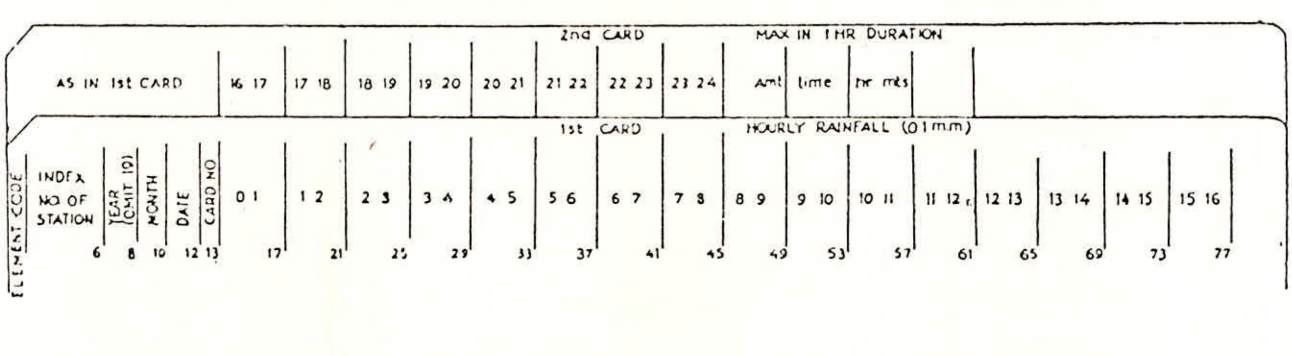


FIG 3 : FORMAT FOR RECORDING HOURLY RAINFALL RECORDS (By IMD) rainfall data using Holeritu machines which can sort out, make tables and give print outs in desired format. When IMD, Delhi acquired an IBM 360 in 1970, some rainfall data from punched card was transferred on to tapes. The transfer of data on to tapes became easier with the acquisition of a EC1040 computer system at IMD, Pune equipped with tape drives and key to tape punch units.

Currently IMD is using two computer systems namely ROBOTRON EC-1040 mainframe computer system and VAX 11/730 digital computer system for the storage and processing of the collected data. Types of storage media used are magnetic tapes, disks and floppy diskettes. Different types of formats are used for the storage of various types of data as devised by IMD/WMO, with record length varying from 80 to 125 bytes. After keying in and verification of data on data entry machines (floppies), they are transferred onto magnetic tapes for electronic data processing on computer. Various types of checks viz. multipunch check, duplicate check, validity check, absurd value check and climatological check etc. are applied to make the data as free of errors as possible before final archival on magnetic tapes. The special programmes required for data processing, quality control checks etc. have been developed in-house.

4.2 Central Water Commission (CWC)

CWC presently operates a national hydrological network comprising about 540 gauge-discharge observations spread on principal rivers of India. Amongst these stations, apart from gauge and discharge observations, sediment observations are also carried out at 254 selected stations. CWC also archives stage - discharge data collected by States for some important sites.

The Central Water Commission is engaged in computerised storage and retrieval of hydrological data for quite some time. The various hydrologic data collected are transmitted from the observation stations to sub divisions, divisions and circle offices of the CWC in the formats devised for the purpose. The data are subjected to scrutiny at various levels and finally authenticated by the circle office. From circle/division these are transmitted to the central data bank for storage and retrieval.

River data directorate and statistics directorate in the headquarters of CWC are primarily responsible for establishing a data bank for systematic collection, processing, storage and analysis of hydrologic data. For rapid and accurate processing and retrieval of large volumes of data, the electronic data processing facilities of the National Informatic Centre, Planning Commission, Govt. of India, are utilised. Data on computer is initially processed mechanically with two fold objectives:

a) Storing of hydrological data on magnetic tapes, and
b) Processing of data on computers for various hydrological computations.

One of these software packages processes the gauge and discharge data of the various rivers flowing in the country and produces the reports usually known as 'water year books'. The daily gauge and discharge data for all the months for particular site on a river is printed in a tabular format. Computed discharge is substituted in case of missing data. The maximum and the minimum water level/discharge and its dates of occurrence are presented monthwise. Extract of water levels and discharge is further carried out in a stratified spectrum such as monsoon season, non-monsoon season and annual. Flow duration of the discharge is further analysed to depict the flowing water. Ten daily, monthly, seasonal and annual analysis for water levels and discharges are carried out on long term basis for the data of 5 years and in multiples of it.

Data received from the field units are subjected to thorough checking to avoid incompleteness, duplication, and inconsistencies. Thereafter, the concerned data collection units are asked to sort out missing data and discrepancies detected in the process. The checked data are then transformed into standard format (8Ø column coding sheets) which facilitates transfer of data to magnetic tapes for long term use by using key-to-tape machines. It is also possible to make use of these data on tapes directly as input to computer programme by suitably modifying the input procedure of the programme. Alongwith data on magnetic tapes, a complete inventory of the observation stations giving catchment particulars, location and other relevant details are also maintained on tapes.

The data stored on magnetic tapes are then processed on CDC-CYBER 17Ø/72Ø and NEC S-1ØØØ computer systems in National Informatics Centre of Planning Commission, Govt. of India to bring out the printouts in the form of Water Availabilities and Sediment Year Books etc.

Recently, a new computerised technique known as HYDAC SYSTEM (Hydrographic Data Aquisition System) has been developed which can carry out sedimentation survey in 2-3 months and hence can be used for more number of reservoirs at regular intervals. In order to improve the technique of sedimentation surveys in the reservoir in india, presently a scheme has been started in CWC under the assistance of UNDP for the induction of sophisticated technology which includes HYDAC SYSTEM. The data aquisition and analysis are carried out using two softwares programmes specially designed for the purpose. The system is extremely useful in deciding suitable measures for sedimentation control as well as efficient operation of reservoirs enabling to derive the maximum benefit. This will also enable to formulate regional siltation indices which can serve as an indicator in planning the future reservoirs.

4.3 Central Ground Water Board (CGWB)

The data collection of hydrological parameters on regional basis commenced in 1969 with establishment of hydrograph network stations all over the country. The data collection was limited to 410 hydrograph network stations in the country where 5 measurements were being taken during a year. The chemical data was being taken once a year.

The data processing including data acquisition and application related to the CGWB activities can be enumerated as follows.

The data base system establishes a linkage between data acquisition and application program. It includes:

a) data collection - i.e. hydrograph network measurements and water sample collection and chemical analysis,

b) data entry - through computer terminal,

c) storage and retrieval of the data using software package dbase III plus and FoxBASE plus under Xenix operating system on Super AT which can also act as a terminal to NEC-1ØØØ.

d) report presentation, and

e) analysis presentation - trend calculation, rise and fall maps, time series analysis, and chemical results presentation.

The PC/AT has been linked with NEC 1ØØØ and multi-user possibilities of the system is being explored. Provision for plotter and digitizer is also being made. The DBMS in dbase III plus software consists of 18 programs interlinked to store data in main and data files. The program is menu driven with following choice:

a) data entry,

b) data modify,

c) data delete,

d) data listing.

The program is further linked to software 'GRAPHER' to produce well hydrograph and to calculate the general trend of water levels.

Presently, the DBMS package being used is FoxBASE plus which is having its own programming language with a number of simple and powerful commands. It allows to create a data base file tailored to the needs and even amend the structure of a database without any loss of data. Data is stored in tables which are collection of records and attributes. Each records/attributes have same group of data items. The data items thus form the columns of the table and each row is a distinct record.

The entire system is menu driven. Most of the screens or menus from which a module or a variable is to be selected by pressing a single key according to user´s choice. The data entry screens are blank formats with field names and the user has to simply fill up the blanks. Provision is made to store the processed values in files to avoid repetitive processing for use at later period.

The data base system contains a main file storing details of the hydrograph network stations and a data file for water level data of stations. These two files are interlinked with the main key, well number, and data in these files can be stored and retrieved using this key.

Type of storage media used are Hard Disk, Cartridge Tapes (6Ø MB,Ø.25") and (5.25",1.2 MB) floppies.

4.4 National Informatic Centre (NIC)

In order to improve the data acquisition and transmission system, National Informatcs Centre of Planning Commission, Government of India, in the first phase established an intra city computer network in Delhi called NIC NET consisting of 15 computer system and of interactive terminals located in various areas of Delhi and connected to a large computer system CYBER 170/730. In its second phase, NIC proposes to extend the NICNET to cover the entire country. The 4 super computers, NEC S-1000, installed at its headquarters in New Delhi and the three regional centres at Bhubaneswar, Pune and Hyderabad form the national grid of this informatics network. Super-mini computer system ND-550 in some of the State capitals and minis and super-micro systems in the remaining States and Union Territories will be connected to the national grid. The PC/AT compatible system installed in the district headquarters will form the third level of NICNET. Micro-earth stations will be installed in all the district and State capitals, with a mother earth station at New Delhi to provide a satellite based data communication network.

NIC has launched its district information system (DISNIC) under which selected information from the districts will be transmitted to the state computers. Similar information from all the districts will be shifted through by the computers in the state capitals under the supervision of the state govt. before passing it on to the regional centres and the national centre. DISNIC through NICNET will provide information for planning agencies and will facilitate decision making at various levels of govt. of India. In this particular direction, DISNIC has already evolved a data base with the parameters of the water use data at the district level. Water use related statistics and water level profiles of the various sources are maintained. The work is in progress and would soon be implemented with the coordination of the state Governments.

Remotely sensed data obtained from satellites and aircraft surveys are used to study and monitor natural resources, natural phenomena and the dynamic effects of human activities with them. Rapid, synoptic and accurate studies on crop, soil resources including growth, water potential, forests, environment and pollution, waste land and land use, desertification and urbanisation, geology, mineral resources etc., are now feasible with satellite remote sensing and digital image analysis. DISNIC is utilising NNRMS (National Natural Resources Management System) for analysis and use at the district level.

The DISNIC software division has undertaken the Irrigation system study and analysis in DISNIC-IRRIGATION sector at the District as well as State level. Under this, the following aspects of the irrigation sector has been studied in detail.

- Minor Irrigation
- Medium Irrigation
- Major Irrigation
- Flood and Drought Relief Schemes
- Anti-Sea Erosion
- Land Navigation

The project GISNIC (Geographic Information System of NIC) is under development in NIC using NICNET facilities to bring the concept of Geographic Information System for decision support in various sectors including irrigation sector.

4.5 National Institute of Hydrology (NIH)

N.I.H. is basically a research institute under Ministry of Water Resources. It collects hydrological data such as rainfall, stage, discharge, ground water levels etc. from various agencies and related state government departments. But so far there is no centralised storage and retrieval system for these collected data. Recently a proposal is there for setting up a hydrological data bank.

A large number of programmes are developed in FORTRAN IV language and are implemented on the VAX-11/780 computer system available with N.I.H for reading rainfall data from I.M.D tapes, to check the accuracy of the collected data,to fill the gaps in data series and to convert the collected data into desired form and format

The computer programme TAPE.FOR is meant to read daily rainfall data from I.M.D tapes and identify the missing data periods. The daily data is read from either the 24 card or 31 card format as written back on to an output file replacing the missing data by -999. For identifying the leap years the programme uses a sub-routine LEAPYR.

A program DAILY reads the daily rainfall data (after identifying the card format and checking for leap year) and writes it into a 12 x 31 matrix. It also produces long term series of ten daily, monthly, seasonal and annual (calendar or water year) totals.

Another program HOURLY reads the hourly rainfall data and stores the data in a 31x24 / 28x24 / 29x24 / 30x24 matrix as the case may be. The total of 24 hrs. is computed from the hourly data for each day and the maximum hourly rainfall is determined.

The computer programme GAPF.FOR is used for estimating the missing station rainfall data using normal ratio method. Another programme DISPOW.FOR is meant for estimation of missing rainfall data using the distance power method. The programme uses the latitude and longitude as provided by the user and computes the distance of each estimator station from the station whose data is estimated.

The computer programme DOUBLE.FOR check the consistency of a particular record using double mass curve analysis.

Programme TENDAY.FOR uses the output file from the tape.for to compute the ten daily rainfall totals and the average rainfall during the ten day period. If the data for any day or days is missing during the ten day period, the total is computed for the available number of days and the average is computed on the basis of the number of days for which data is available.

MAX.FOR is meant for the computation of the maximum 1 day, 2 days, 3 days, 4 days and 5 days rainfall from the daily rainfall data processed for missing rainfall amounts.

CATCH.FOR computes the weighted average catchment rainfall using the Theissan weights supplied by user. ISO.FOR computes the variation of depth with area over the catchment using iso-hyetal method.

The programme DAILY.FOR converts the daily rainfall data of O.R.G stations into hourly rainfall data in the ratio of the hourly rainfall values of an appropriate S.R.R.G station for the day. The choice of the S.R.R.G stations for each O.R.G station has to be made by the user. Further more, the programme computes the average hourly rainfall values in the catchment during the storm using Thiessen polygon method.

The computer program RAIN has been developed to distribute the rainfall recorded at the totaling gauges during an observational day (Ø8.3Ø hrs to Ø8.3Ø hrs) into hourly, using an average pattern hyetograph derived from a set of representative recording raingauges in the vicinity of the totaling gauges as per a given set of weightages. The hourly rainfall pattern thus obtained at each of the totaling gauges is then used along with those observed at the recording gauges to determine the average hourly rainfall which eventually could be used as an input to the rainfall runoff models.

4.6 Central Water & Power Research Station (CWPRS)

The CWPRS, Pune, is primarily a sophisticated user of hydrometeorological data. As such, the institution carries out project/problem-oriented collection, processing and analysis of water data; with the basic data generally obtaining from other organisations such as IMD, CWC, State Government Departments etc. which are involved in collection, initial processing and publication of water related data.

Hydrometeorological data, collected by CWPRS are rainfall data, streamflow data, evaporation and evapo-transpiration data, drainage basin characteristics such as elevation, stream lengths, area, soil types, vegetation. etc, and details about water resources projects/hydraulic structures in the region, etc.

These collected data are stored in a Cyber 180/840 A system with a large network of terminals which can also work as individual PCs when required. In addition, a large number of PCs, both XT and AT are available for use.

Those data that are actively used in any specific analysis are kept in a data base system on hard disk. The rest of the data are compiled and stored in computer compatible form such as magnetic disks, floppy diskettes and data cartridges.

Pre-processing/quality control is being done mainly to rainfall and stream flow data which are being used as inputs to the parametric hydrological models.

4.7 Institute of Hydraulics and Hydrology (IHH)

IHH, Poondi is a research institute under Government of Tamil Nadu Public Works Department. The types of data available with IHH are rainfall data, discharge data on river gauging stations, silt survey data for selected reservoirs and meteorological data of Poondi and Vaigai stations. They are using VAX/VMS Version V4.5 computer mainframe system available at IRS Anna University, Madras for data storage. Magnetic tapes are being used as storage media and there is no uniform format is used for storing the data

4.8 Andhra Pradesh Engineering Research Laboratory (APERL)

APERL, Hyderabad is a research institute under the Andhra Pradesh Government. They are doing studies in various fields in Engineering and Technology.

Development of data base management system in respect of hydrological data was taken up as research study in this institute. This study was initiated by Central Board of Irrigation & Power and sponsored by the Ministry of Water Resources under the guidance of National Advisory Committee for Research and Development Pertaining to hydrology. The APERL have used dBase III and developed the system called Hydrometeorological data creation and retrieval system (HYDCARS) considering parameters like rain fall, flow data, evaporation, temperature, sediment data, wind velocity and humidity.

Another CBIP project has been taken up by the APERL namely 'Development of software for preliminary processing of hydrological data using micro computers' and the same is in progress. Following are the features of this software:

- conversion of daily rainfall into weekly, fortnightly, monthly and seasonal.
- to find maximum one hourly and six hourly rainfalls.
- computation of daily rainfall with the use of automatic rain fall data.
- estimating the missing rainfall data by normal ratio method.
- estimation of weighted rainfall using the weights of the polygons.

-estimation of absolute, maximum and minimum rainfall.

-given the coordinates of the station computing the Theissan polygons and finding out the weight of each.

-conversion of daily inflows into weekly, fortnightly, monthly and seasonal.

-conversion of hourly gauge data into different periods.

-estimation of peak flows in the reservoir for different return periods.

-estimation of various statistical parameters of rainfall and runoff data.

-to check the internal consistency of data.

-stage discharge relations.

-computation of sediment rating curve for rising and falling floods.

-geometric elements of open channel systems.

-reservoir routing.

-estimation of dependable yields by different methods.

4.9 Water Resources Development Organisation (WRDO)

The WRDO is a research organisation under Irrigation Department of Karnataka Government. They are mainly engaged in collection analysis of stage - discharge data for different river basins in Karnataka State.

Data are collected and stored in floppies with the use of a PC/XT (21 MB - IBM Compatible - 768 K Ram - 16 Bit). Manual checks are being performed on collected data before its storage and before use a batch processing is being done. Various application softwares are developed for data analysis such as max. value, min. value, mean, standard deviation, ten daily value, monthly average etc. Using all these derived quantities they are producing Water Year Books for the entire river basins of the state.

5.STATUS OF DATA PROCESSING IN OTHER COUNTRIES

A survey of hydrological services conducted in 1977 showed that 47 countries had developed computer based data banks and 14 more were considering doing so in near future. These data banks covered various fields- the ideal being made up of statistics covering all hydrological parameters from all parts of the country assembled in a form that would be easy to access. The survey showed that such a centralised data bank existed in 12 countries. Coordinated data banks operated by two or more bodies existed in 16 more, and in a number of countries there were separate data banks spread between a number of different bodies but uncoordinated. Most of the countries took part in the survey indicated that automation in water level and flow data collection and processing was further developed than in other fields. The next was precipitation data collection and processing. A smaller number of countries indicated that they had established a centralised data bank rather than a coordinated one, but over half the countries in the survey showed they had no data bank at all.

The establishment and maintenance of a national data bank for hydrological observations, including derived statistical and analytical informations, is the basic duty of a national hydrologic service or equivalent agencies responsible for operational hydrology. Many countries are presently using or may soon be using computers to process and store their hydrological data. It is, therefore, useful to make known appropriate procedures and methodologies facilitating the establishment of computer based centralised or co-ordinated hydrological data banks, which permits easy access by water resources planners and other users of hydrological data.

With the advent of high speed digital computers, a number of computer programs have been developed the world over for processing, estimating and filling-in missing data records. Most of the developed countries have devoted considerable resources for the improvement of their hydrologic information systems. This improvement is directed towards the increased use of automation of data collection and processing systems which includes the use of automatic sensors and devices which transfer data from the field stations to the processing centers and prepares the data in a form suitable for computer compilation and processing.

A complete national hydrological data bank would contain data and statistics covering all hydrological variables for all parts of a country and would permit any potential user to easily extract and use these data without extensive manipulation.

The main elements of a data bank are the computer software (programs) and hardware required to handle the hydrological data. The operation of a data bank consists in the organisation, description, processing, storage, management and updating of data. It normally also includes access to the data series by users and data exchange. Data banks differ in many details, for example as regards the way in which a data base is organised, the way in which the data series are addressed, and utilisation procedure. In what follows, there are brief descriptions about the hydrological information systems (hydrological data bank) adopted by different countries and institutions.

A schematic representation of a system of data storing and processing developed by the Lund Institute of Technology, University of Lund, Sweden(1984) is presented in fig.(4). Another schematic representation which was suggested by the Institute Royal Meteorologique De Belgique (1969) is shown in fig.(5).

5.1 Australia

Australia has a co-ordinated national hydrological data bank consisting of various separate data banks covering hydrological data for a region or state. The Australian Water Resources Council (AWRC) coordinates different hydrological data banks in different regions of Australia.

Altogether there are at least 26 organisations collecting records from 2800 stream gauging stations. In addition, there are over 800 stations at which other surface data are collected. The majority of stream gauging stations are equipped with either conventional float-well installations or pressure sensing devices attached to continuous analogue type recorders providing a record in chart form. Digital recorders that provide data for direct input into computers have been developed but are currently used only at few stations. Only very few stations provide manually read daily records. A limited number of stations are provided with telemetering equipment for use with flood forecasting systems.

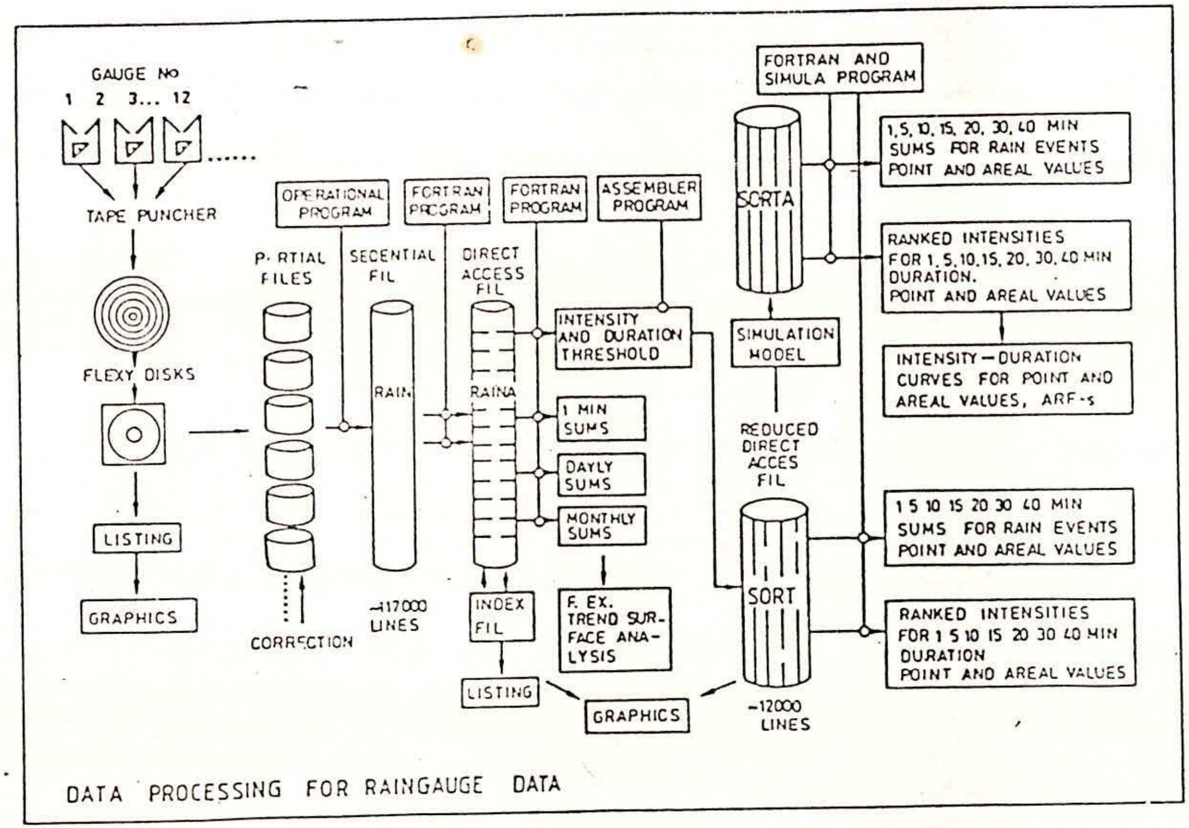
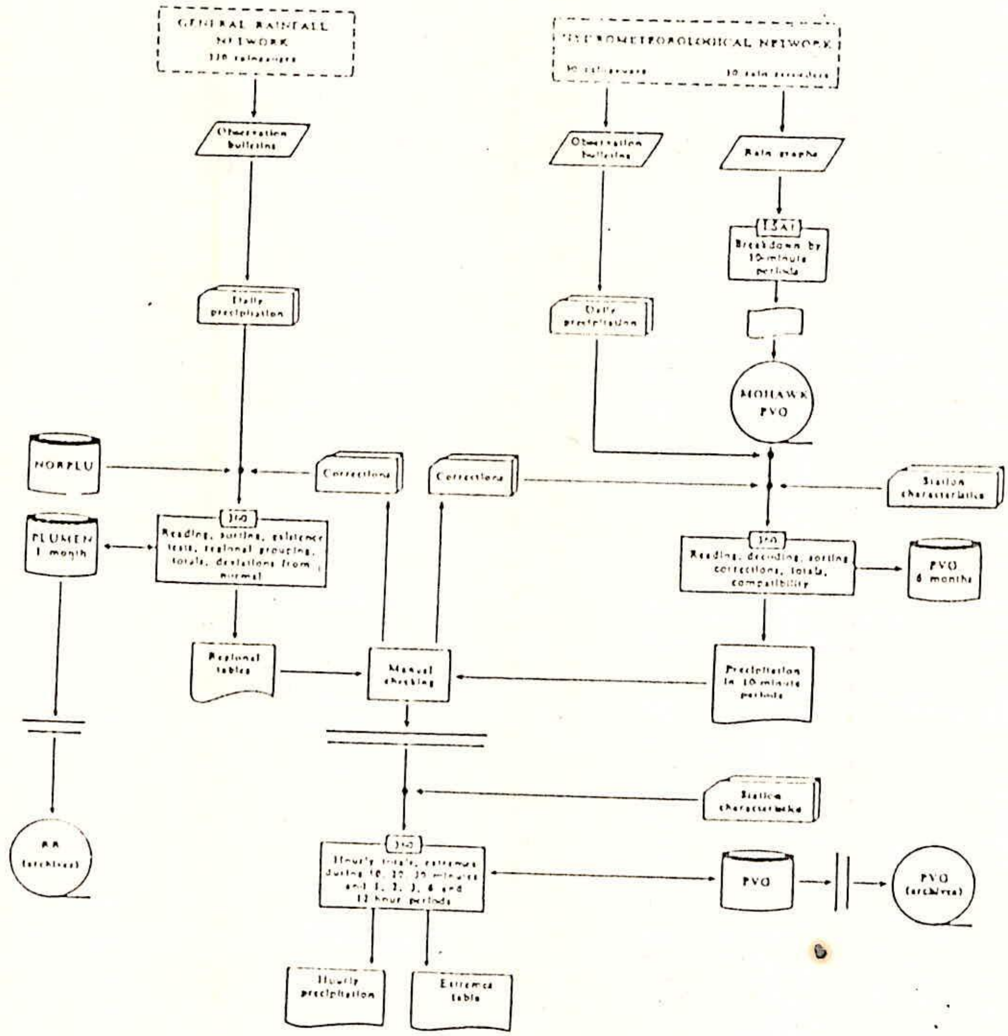


FIG 4 : DATA PROCESSING SYSTEM FOR RAINFALL RECORDS
(Developed by Lund Institute of Technology, Sweden)



NORPLU: Monthly rainfall norms and station characteristics.
PLUMEN: Daily raingauge readings (1 month).
PVO: Results of rain-graph analysis.
RR: Daily raingauge readings (archives).

FIG 5 : DATA PROCESSING SYSTEM FOR RAINFALL RECORDS
(Developed by Institute of Royal Meteorologique De Belgique)

As with the data processing system, there is a wide diversity in the systems used for the storage of data although magnetic tape is the favoured medium. There is an increasing use of microfiche for the storage and transfer of records.

In the case of continuous data, the quality of it is maintained by a program designed to check the data thoroughly. The analogue-to-digital conversion process is checked by occasional computer plotting of the data onto overlays, which are then placed over the original chart.

Rating curves are developed from current metering for each gauging stations and checked at least annually for stability. These rating curves are stored as a set of X and Y co-ordinates in a disk resident data base for ready access by other programs.

Daily flow volumes and peak flow rates are calculated from the water level data and the rating curve routinely. Shorter time interval data are calculated on request. Annual and average flow duration curves are also calculated. Real time data for flood forecasting or day-to-day operational decisions are not collected or processed.

Non-continuous data (water quality analysis and ground water levels) are entered into the machine by a key-to-disk operation directly from the field forms. All water quality informations pass through the same series of checking and updating programs.

The correctness of the data for the station number is first checked to see that it exists. To further verify the station number, a check digit is also entered alongside it. The check digit is calculated on the Modulus-II system.

The checking of the data for correctness and internal consistency is thorough. Calculations carried out as part of the laboratory analyses are checked again by the computer on data entry. where parameters are correlated, such as electrical conductivity and chloride ion concentration, general relationships are used to ensure that the data values are within an acceptable tolerance. Range tests are applied to all parameters to ensure results are at least in the right order of magnitudes.

Retrieval of data can be affected by interactive application programmes mostly written in Fortran IV language. For data analysis and processing, a standardised working file structure has been developed for easy data manipulation. Current developmental activities are in the field of hardware reconfiguration, the information systems, and more efficient software for data processing and retrieval.

5.2 **Belgium**

The Belgian national hydrological data bank consists of a group of five data banks, i.e. one central bank managed by the Hydrology Section of the Royal Meteorological Institute, and four specialised data banks set up by various public administrations.

Primary processing comprises the following conventional phases: data collection, decoding,identification and composition of data in conventional units, validation of both the identification and the data themselves and addressing and storing in appropriate files.

Secondary processing of the data consists of

- intercomparison of recorded values and direct readings,
- plotting hydrographs for manual comparison with the original charts,
- regional comparisons based on the ratio of reduced values of stages,
- for stream gaugings, plotting a graph of the points at which measurements are made and comparison with the original cross section,
- plotting the graph of the annual regime of specific discharges and regional comparisons,
- regional comparisons of monthly and annual streamflow deficits.

Specific tasks in secondary data processing include:

- calculation of the average aerial precipitation over river basins,
- calculation of daily mean values of evapotranspiration from river basins,
- calculation of mean velocity and discharge based on stream gaugings,
- analytical fitting of stage discharge relations,
- conversion of stage to discharge,
- drawing ground water level maps,
- evaluation of the indices of the water quality.

5.3 Canada

The Inland Water Directorate, Environment Canada, maintains two hydrological data banks: one operated by the Water Survey of Canada(WSC), contain data on water levels, discharge and sediment; and the other, operated by the Water Quality Branch, contains data on water quality.

The Water Survey of Canada has been collecting and publishing streamflow and water level data since 1908 and sediment data since 1961. Here data are forwarded annually to Ottawa by the regions either on punched cards or magnetic tape, processed on a CDC CYBER 74computer, and printouts return to regions for verification of the data. These data are then processed at Ottawa to various types of publication using computer programs, manuscripts are produced automatically by photocomposition which gives high quality output suitable for direct printing.

The Water Quality Branch, Inland Waters Directorate, Environment Canada, is responsible for monitoring the conditions of surface water in Canada. The need for flexibility and speed in storage and retrieving information led to the establishment of NAQUADAT, a computerised national water quality data system. Since then, the system, while fully operational, has been modified to utilize advances in computer technology and to better respond to user needs.

NAQUADAT was designed to store the results of water quality analyses carried out in departmental laboratories and elsewhere.

It was intended to provide informations required for monitoring and identifying trends and anomalies in water quality.

NAQUADAT operates commercially with time sharing facilities on IBM computers and with interactive softwares in COBOL, MARK IV and FORTRAN languages. The software devices are used to update, search and retrieve information. Considerable flexibility is allowed in specifying the information to be retrieved. Browsing and interactive reporting are facilitated by two time sharing utility programmes STS: WYLBUR and STS: INFO. Limited statistics are generated in the reporting programmes. Elaborate statistical analysis is achieved via integrated statistical packages such as SPSS or BMDP. They can be retrieved as computer outputs, microfiche, digital plots and standard magnetic tape output. Current developmental efforts are devoted to interactive graphical presentation and overall efficiency of operation.

5.4 U.S.A

The U.S.Geological Survey investigates the occurrence, quality, quantity, distribution and movement of the surface and underground waters that comprises the water resources of the United States.

The data collected during field activities are entered into the national Water Data Storage and Retrieval system (WATSTORE). The WATSTORE system consists of several files in which data are grouped and stored by common characteristics and data collection frequencies.

As the data collection activities of U S federal agencies increased, concerns were expressed about possible duplication of effort and about the availability of the hydrological data collected to all agencies. As a result, the office of Water data coordination of the U S Department of Interior was established in 1964 to coordinate the national hydrological data collection effort. Several important programmes have arisen from this including: the National Stream Quality Accounting Network (NASQAN) which provides a data base for monitoring national surface water quality, a common system of hydrological unit classification, a series of river quality assessment intended to evaluate the accuracy of existing water quality data, a national programme for coordinating the collection of water use data, a programme for water data exchange and the development of the standardised, recommended methods for water data collection.

An important aid to the coordinated data collection is a national index to the data collected. For the U S, this is provided by the National Water Data Exchange (NAWDEX) operated by the U S G S. NAWDEX does not store actual water data, but maintains two linked data bases: the Water Data Source Directory and the Master Water data Index.

STORET (The Storage and Retrieval system), was established in early sixties by the U S Environmental Protection Agency (EPA) to handle the water quality data. It currently contains data from over 300 000 sites. As with WATSTORE, STORET provides a wide variety of data outputs.

The data measurement and collection procedure is highly systematized. Digital recorders are used at many locations to record river stages, conductivity, water temperature, turbidity, wind direction and chemical contents. Field data are stored on paper tapes and transmitted to the data bank via telephone lines. Water quality data are sampled and analysed with uniform techniques by NASQUAN at 525 stations. Two central laboratories analysing about 150000 water samples per year also contribute information to WATSTORE.

Quality control of data involves routine checking by computer programmes from an auxilary file to ensure an acceptable level of consistency. Secondary processing involves computation of aerial estimates, parametric models, statistical analysis and graphical plots. Real time data processing has been in operation as part of a satellite data collection testing programmes since 1972.

In order to reduce high costs involving on line disk files, only current data are held on disks while historical data are maintained on magnetic tapes. The ground water site inventory file uses the SYSTEM 2000 database management system. Application programmes allow for retrieval of data from current and historical records. Data can be output as computer printouts, digital plots, or on punch cards or magnetic tapes.

5.5 Germany

The water resources data bank of Bavaria, Germany was formed in 1976 by the Bavarian Ministry of Interior whose State Water Resource Office operates the hydrologic network through its local offices. Hydrometric data are obtained through the Federal Weather Service.

The information Management Systems (IMS) serves as a data base management of the Bavarian Data Bank. Most of the data are available in analogue form. These are digitized, recorded on magnetic tapes and transferred from mini computers to the main IBM/3Ø33 computer for processing and storage using PL-1 programming language.

Quality control of data is considered very important. So all incoming data are automatically tested on plausibility. As only crude errors and mistakes can be detected by this method, manual checks by well-trained staff are necessary. Further controls are made by means of automatic regression computations or similar methods.

The operational sequences for surface water and ground water data processing have been established at the Federal level by a German Association for Water Resources and Reclamation working group. These sequences are also valid for the data bank.

The data base of the IMS storage and retrieval system is a hierachical structure of data elements related by their logical dependencies. The data base is supported by the data communications service which allows direct access to data files through appropriate on-line terminals.

5.6 New Zealand

In New Zealand, hydrological data are stored in two national data banks. The main bank is controlled by the Ministry of Works and Development (MWD) and contains data collected by the MWD and the various regional catchment authorities. The second bank is controlled by the Meteorological Service and contains rainfall and climatological information collected by the Service and its agents.

In 1971, with an expanding recorder network and a greater demand for data in non-standard forms, a computerised storage system, known as TIDEDA, was developed in the MWD. Now it is used extensively throughout New Zealand for storing,editing and retrieving time dependent data. The data are stored in TIDEDA in a non-interpretive form, so that data values can be retrieved in a variety of ways without any loss of accuracy.

Once the field document and coded data have been translated into a suitable form on magnetic disk, the respective data processing sections check the data for translation errors. With water level data, the Tektronix terminal and the Calcomp plotter are used to display the data and expose spikes and other anomalies that may have arisen from improper translation and can be corrected.

With rainfall and event water level digital paper tapes, the translated data are entered directly into a TIDEDA disk file.

These give a diagnostic listing of possible translation errors. Another verification for water level is the preparation of a flow report. The flow report is concerned with the corresponding flow. The water levels have been transformed to discharge using stage-discharge rating curves. In checking the discharges, the report is concerned with rainfall, the corresponding discharges at nearby recording stations and the accuracy of the rating curves, and it comments on the quality of the data.

5.7 Norway

The hydrological data bank at the Norwagian Water Resources and Electricity Board (NVE), located in Oslo, is a centralized bank for surface water and ground water quality. It covers hydrological data from all parts of the country. Co-ordination exists with other institutions handling precipitation, evaporation and water quality data.

Data are processed on a CONTROL DATA CYBER 171,132K 64 bits word computer, with disk, magnetic tape card reader and printer equipment. The daily administration of the machine as well as primary punching of incoming data is taken care of by a separate electronic data-processing department. The bulk of programs operating the data bank runs in remote batch mode. Retrieval and analysis are mostly interactive.

When the data from one year are corrected, the extremes and monthly means are compared with values from previous years. If these statistics and plots or listings of daily values seem reasonable, the series are accepted and merged into the historical archives of annual records.

Quality control and data processing involve manual correction for ice damming, spatio-temporal correlation and regression analyses,time-series analyses, precipitation-runoff models, stage-discharge rating curves and simulation runs.

In 1977 a Nordic working group on methods of quality control was created. Some of the recommended methods are :

- testing the stage or discharge of a given day within a year against the highest and lowest value of the same date in all the previous years observed at the station,
- comparing observed data with estimates based on data from adjacent stations,
- comparing the observed data with estimates based on a precipitation-runoff model,
- checking for negative values during the computation of inflow to a reservoir when the stage storage relationship and the outflow are known,
- comparing the runoff at a station with runoff at upstream stations,
- applying time series analysis to detect changes in the homogeneity in time series.

5.8 United Kingdom

The United Kingdom data bank can be classified as a co-ordinated national hydrological data bank. The Meteorological Office holds the rainfall and evapotranspiration data and the surface water, ground water and certain water quality data are held by the Department of the Environment (water data unit).

The Meteorological Office receives, archives and publishes all rainfall data of acceptable quality supplied to it. Data bank consists of powerful and versatile computers includes IBM 360/195 and an IBM 370/158. Water Data Unit consists of an ICL1904S with 256K words of main memory, 10 exchangeable disk drives each holding 60 million characters, 8 magnetic tape decks and a high speed drum.

Quality control of rainfall data is based on comparison with near neighbor data. Initial quality control , ie. the elimination of obvious errors in individual records, is effected regionally by water authorities and similar bodies. Final quality control provides a best-estimate data set, but original values subsequently replaced by estimates are also preserved. Quality control of surface water involves testing of successive values of stage to check if difference exceeds a chosen value.

Primary data processing of streamflow data consists of the conversion of stage to discharge by means of a computer program. Secondary data processing of streamflow data consists of analysis of the data to meet user requirements.

6.0 A TYPICAL PROCESSING SYSTEM - HYMOS

In this, an example of a well planned and developed data processing system is discussed in detail. This software (HYMOS) is developed by Delft Hydraulics, Netherlands. It is developed to streamline the storage and process of (geo)hydrological and meteorological data. It is tailored for use by hydrological and meteorological data processing branches, water resources management authorities, water boards, water engineering consultants, and hydrological advisers. HYMOS 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 runs on stand-alone computers, but can also be used in a network system. Securities have been built in to restrict the access for certain activities to qualified staff only.

HYMOS is a data base management and processing system for hydrometeorological quality and quantity data, designed for use on PC's. It arranges a convenient structuring of data in a database and provides an extensive set of tools for data entry, validation, completion, analysis, retrieval and reporting. The HYMOS software package is written in the programming languages FORTRAN 77 and C. It runs on computer systems of the following specifications:

IBM XT, AT or compatible personal computer, with 640 kb core memory, 20 mb hard disk, EGA or VGA graphics card, printer, HP or

Calcomp compatible plotter, etc.

This system integrates the distinctive phases in processing of hydrological data. The activities are carried out in specific processing modules; each module consists of a number of programs enclosing particular compilations/computations. The modules are structured according to a logical sequence of activities in data processing. All modules are linked to the HYMOS data base, structured and controlled by a data base management system. Different data bases can be operational under HYMOS.

The types of data, handled by HYMOS, can be categorised in the following groups:

a) Space oriented data - catchment characteristics station particulars, station histories geohydrological profiles.

b) Time oriented data - equidistant time series non - equidistant time series

c) Relation oriented data - stage discharge data relation or rating curves parameters valid for certain time period.

The structure of the HYMOS is as shown in the Fig.(6).

6.1 Data Storage and Retrieval System

Data storage and retrieval activities are comprised in two modules: The 'entry and editing' module and the module 'reporting and retrieval'.

1. Entry and editing: The entry and editing module includes following options:

a) entry and editing of catchment data

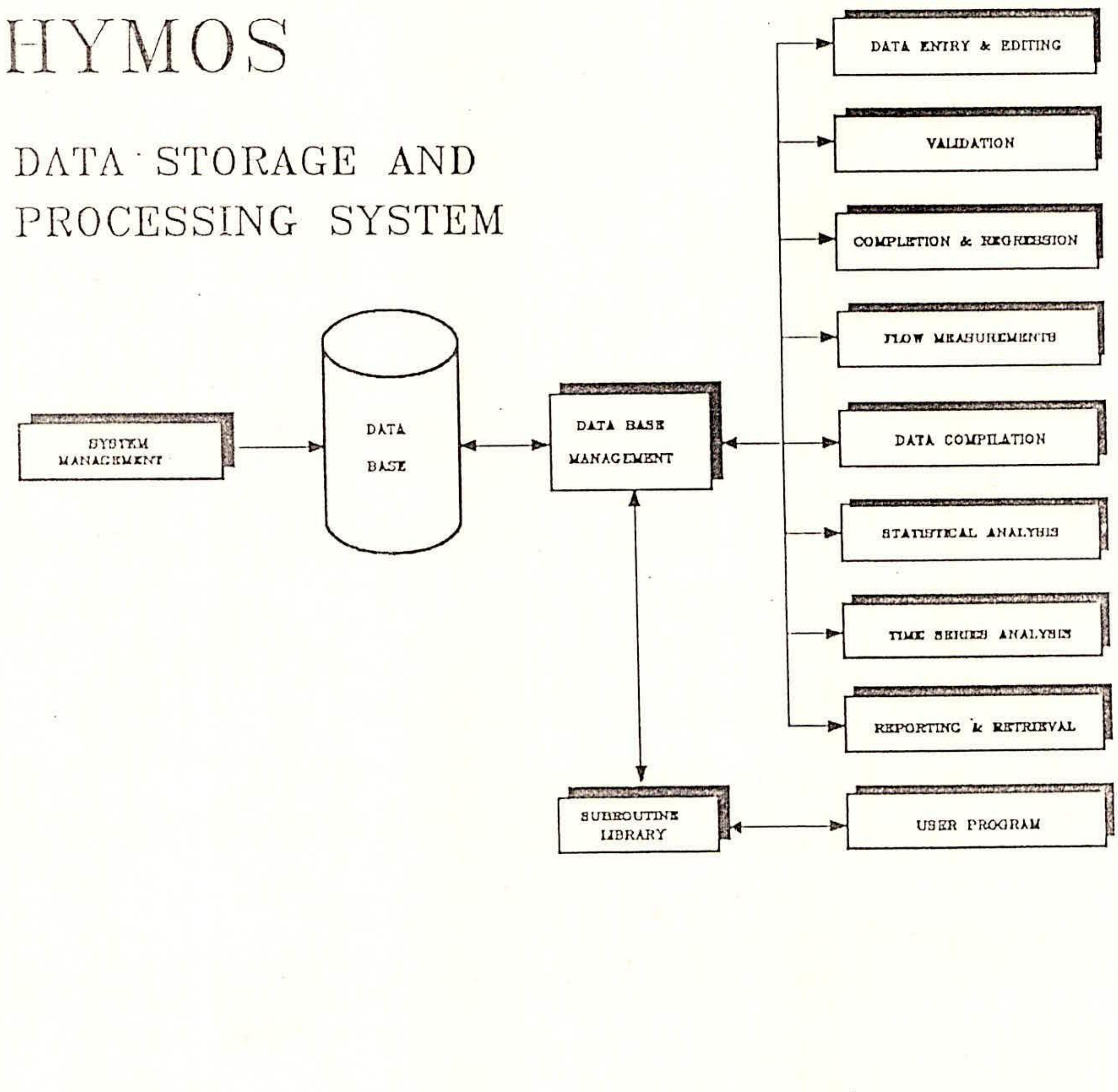


FIG 6 : STRUCTURE OF HYMOS

b) creation of stations and series, and entry and editing of station particulars and geohydrological profile data.

c) loading of equidistant and non-equidistant time series data from user files or manually; direct links with telemetering systems can be established, while also fast data transfer from EPROMs to the data base is possible.

d) full screen editors for editing and display of equidistant and non-equidistant time series, with on line graphical display.

e) entry and editing of current metering data and discharge and sediment transport rating data

f) entry and editing of data files for particular analysis (statistical analysis, regression, etc.)

2. Reporting and retrieval: The reporting and retrieval module comprises of:

a) preparation of reports on station and series characteristics and time series

b) preparation of mixed tables of data base quantities

c) plotting data base quantities in user specified layout

d) retrieval data stored in the data base

e) transfer of data from one data base to another.

6.2 Data Processing System

The data processing system of HYMOS includes following modules:

1. Data validation: For data validation, use can be made of the following procedure:

a) data screening by flagging, printing and tabular comparison of time series and computation of basic statistics

b) graphical evaluation of time series (lines or bars), including:

- plotting of time series (max. 5 per graph)
- residual time series graphs
- residual mass curves
- moving averages
- water balances
- graphical presentation of series availability in the database.

c) relation curve analysis, to establish relation equations and to investigate shifts in the relationships; it includes:

- graphical display of relations
- fitting of relation curves by polynomials, divided in 1, 2, or 3 intervals to account for effects of irregular or compound cross sections.
- storage of relation curve parameters in the data base
- comparison of relation curves of different time periods

d) investigation of series homogeneity by means of double mass analysis, presented in graphs and tables

e) statistical tests on data homogeneity and randomness

f) spatial homogeneity tests, where data at a base station are compared with weighted averages of neighbouring stations, selected on distance and orientation.

2. Series completion and regression: The completion and regression module comprises:

a) interpolation techniques for filling in missing data based on time and space interpolations

b) regression models to establish relationships, or to fill in missing data; the models may vary seasonally and can be of the following types:

- polynomial equation
- power equation
- logarithmic equation
- hyperbolic equation
- exponential equation
- simple and multiple linear regression equations
- stepwise regression

c) Physically based lumped parameter rainfall runoff model SAMO for filling in missing runoff data.

3. Flow measurements: The module comprises a number of techniques for validation and elaboration of flow measurements and rating curves, including:

a) entry and editing of flow velocity measurements, stage discharge data and rating curve parameters with cross-sectional parameters.

b) processing of flow velocity measurements by profile and moving boat methods, allowing:

- various methods for measurements in the vertical
- wet and airline corrections
- mean and mid section method to compute the discharge

- graphical and computational validation of measurements
- transfer of condensed results to the data base

c) computation of stage discharge relations given as parabolic and power type equations, with:
- coefficient for upto 3 water level ranges per relation
- corrections for backwater effects
- corrections for unsteady flow
- detailed error analysis
- transfer of coefficients with validity period to the date base

d) validation of rating equations for different periods and new measurements

e) extrapolation of rating curves:
- computation of cross section capacities and parameters
- graphs of cross sectional parameters versus stage
- computation of synthetic stage discharge data beyond the measured range

f) stage discharge transformation using:
- rating curves stored in the data base
- rating equations of measuring control structure for critical and sub-critical flow conditions and variable sill level.

4. Data compilation: The data compilation module comprises:

a) aggregation and dis-aggregation of time series, where accumulative and instantaneous data are treated differently.

b) series transformation with various arithmetic transformation options

c) minimum, mean and maximum series computation for selected time periods and transfer to the data base.

d) computation of aerial rainfall by:

- (weighted) average of point rainfall data
- Thiessen method
- Kriging method

e) interpolation and computation of best linear estimates of and uncertainties in aerial quantities by point and block kriging method

f) computation of potential evapotranspiration using:

- Penman method
- Pan evaporation method
- Christiansen method
- Radiation method
- Makkink method
- Jensen-Haise method
- Blaney-Criddle method
- Mass Transfer Method

5. Statistical analysis: The module for statistical analysis include:

a) computation of basic statistics and histograms

b) fitting of distribution functions of the following type:

- Normal distribution (with Box-Cox transformation to normality)
- Log normal distribution
- Exponential distribution
- Pearson 3 and general Pearson distribution
- Log Pearson distribution

- Raleigh distribution
- Extreme type 1 (Gumbel), 2 and 3 distribution
- Goodrich distribution
- Pareto distribution for peaks over threshold.

c) statistical tables, i.e. computation of probabilities and varieties for the various probability distributions.

d) generation of normal and gamma distributed random numbers

e) computation of IDF curves (Intensity Duration Frequency Curves) from monthly maximum precipitation in short intervals

f) computation and plotting of:
- frequency curves
- duration curves
- average duration curves

g) time series analysis
- auto and cross-correlation function computation
- spectral analysis
- run analysis: computation of up and down crossings, run lengths and run sum
- range analysis: computation of range of cumulative departures from the mean
- analysis of storage requirements by the sequent peak algorithm.

6.3 Special Features

To facilitate the use of HYMOS special function keys are available for on line help, display of temporary output files, use of text editors, change of data base and packing of data base files.

The station and series selection is enhanced by selection from displayed tables or catchment maps on the screen.

Keys are available to set the screen colours and hardware configuration and to switch to DOS during a HYMOS section.

7.0 CONCLUSIONS

Automated collection and processing techniques can lead to significant improvements in hydrological data collection, quality control, storage, accessibility, preparation of publications, and the use of data for analytical studies. However, the realisation of these improvements requires specialised training of the personnel involved in all phases of the work.

Great care is needed in the design of data processing systems, not only to achieve the desired technical goals, but also to reduce the complexity and costs of the associated non computer tasks. The most important design considerations are the scope and objectives of the tasks to be performed by the proposed system. These provide the framework for determining the types, volumes and formats of data to be handled which in turn define hardware and software requirements. In developing countries it is essential that the overall design reflects the availability of hardware, the ability to maintain that hardware, and the availability and capability of data processing personnel.

It is important that the use of automated techniques does not necessarily imply the use of highly sophisticated computer installations. The role of microcomputers as an economical, powerful, and readily available way to introduce computer based technologies and to develop the necessary human skills is very much important now-a-days. It has also been shown that distributed data processing systems tend to be more flexible in meeting changing needs, but that a centralized approach may be necessary where expertise and/or funds are limited. Distributed systems may also present some problems of compatibility, particularly if several different types of microcomputers are to be linked to the system.

In our National Water Policy it has been laid down that a standardised national information system should be established with a network of data banks and data bases, integrating and strengthening the existing central and state level agencies and improving the quality of data and the processing capabilities. It also lays down that there should be free exchange of data among the various agencies and duplication in data collection should be avoided. The CWC in collaboration with NIC, INCOH, Karnataka Irrigation Dept. and the Institution of Engineers (India-Karnataka State Centre) have made a beginning in the direction of evolving a National Hydrological Data Bank by organising a seminar on 19-20 May 1988 at Bangalore. The various recommendations which emerged in the seminar are namely, the desirability of distributed system of National Water Data Bank, a national programme for training of field staff, quality improvement in hydrological data collection, real time data transmission, use of software available for data processing through international cooperation, regular publication of hydrological data, etc.

A national hydrologic information system should be established with a network of data banks and data bases, integrating and strengthening the existing central and state level agencies and improving the quality of data and processing capabilities. In order to remove the drawbacks in the present hydrologic data base structure, there should be free exchange of hydrologic data among various collecting and processing agencies and duplication in data collection and processing should be avoided. Improvements in the existing hardware and software facilities and innovation of new techniques will be needed to improve and enhance the available capabilities for efficient processing of hydrological data.

ACKNOWLEDGMENTS

The author is grateful to Dr.S.M.Seth, Coordinator, UNDP project and Sh.K.S.Ramasastri, Technical Coordinator, Regional Centre, Belgaum, for their valuable guidance and useful discussions. Special thanks to Dr.J.S.Rawat, Head Regional Centre Belgaum, for his keen interest and assistance during the preparation of this report. He expresses his gratitude to Dr.Sathish Chandra, Director, National Institute of Hydrology, for extending this opportunity to attend the training programme as part of UNDP project.

His sincere thanks are also due to Additional Director General of Meteorology (Research), IMD Pune, Sh.S.K.Shama, Scientist C, CGWB New Delhi, Joint Director, CWPRS Pune. Sh.Gurumurthy, Research Officer, APERL Hyderabad, Director, IHH Poondi and Capt. Raja Rao, Chief Engineer, WRDO Bangalore, for the valuable informations given by them for completing the chapter on 'Status of Data Processing in India'.

Finally, the assistance rendered by Sh.T.P.Panicker in typing this report and cooperation extended by his colleagues at Regional Centre, Belgaum is gratefully acknowledged.

CHANDRAMOHAN.T

BIBLIOGRAPHY

1. Chow,V.T., 'Handbook of Hydrology', Mc Graw Hill Book Co., New York.

2. Delft Hydraulics, Lecture Notes on 'HYMOS', Netherlands, April 1991.

3. Dutt,D.K., 'Status of Data Base Management System in Central Ground Water Board', Seminar on National Hydrological Data Bank, Bangalore, 1988.

4. Dodge, 'Analysis of Experiments with Missing Data'

5. Elizabeth. M. shaw, 'Hydrology in Practice', Van Nostrand Reinhold (UK) Co. Ltd.,1983.

6. Futaki,K. and Vargay,Z., 'Primary Processing of Hydrological Data on Microcomputers', Microprocessors in operational Hydrology, WMO Publication.

7. Herschy,R.W., 'Hydrometry- Principles and Practices', A Wiley Interscience Publication, John Wiley & Sons, 1978.

8. Herschy,R.W., 'Hydrological Data Collection by Satellite',T.N-275, Proc. Instn. Civ. Engrs., Part I, 1980.

9. I.H (1981), 'The Processing of Hydrological Data', Institute of Hydrology, Wallingford,UK.

10. Inland Waters Directorate (1975), 'Manual of Hydrometric Data Computation and Publication Procedures', Water Resources Branch, Ottawa, Canada.

11. Isherwood, W.L., 'Use of Digital Computers in Processing, Publishing and Analysing Hydrological Data', Nordic Hydrology-1, 1970.

12. James Martin, 'Principles of Data Management'

13. John C Rodda, 'Facets of Hydrology II', Wiley International Publication, New York, 1985.

14. Lipschutz, 'Theory and Problems of Data Processing'.

15. Mehta,N.K., 'NICNET in Computerisation of Hydrological Data Bank', Seminar on National Hydrological Data Bank, Bangalore, 1988.

16. Molnau,M., 'A Guide to the Use of HISARS, A Hydrologic Information Storage and Retrieval System'

17. National Institute of Hydrology (1983-84), 'Data Systems and Library', status Report, SR-4.

18. National Institute of Hydrology (1984-85), 'Processing of Precipitation Data', Report M-2.

19. National Institute of Hydrology (1987-88), 'Processing and Analysis of Rainfall Data', Report UM-36.

20. National Institute of Hydrology (1988-89), 'Workshop on Flood Estimation by Unit Hydrograph Techniques', Chapter 3 & 4.

21. Parashar, R.K., 'Sediment Observation in River Basins of india by CWC', Lecture Notes, Regional Training Course on Reservoir Sedimentation and control, Dec. 9-22, 1991, New Delhi.

22. Paulhus, J.L.H. and Kohler, M.A., 'Interpolation of Missing Precipitation Records', Monthly Weather Review, Vol.80, 1952.

23. Plinston, D.T. and Hill, A., (1974), 'A System for the Quality Control and Processing of Streamflow, Rainfall, and Evaporation Data', NERC Institute of Hydrology Report No.15.

24. Raisinghani, R.J., 'Hydrological Data Bank in Hydrological Small Catchments Directorate, CWC', Seminar on National Hydrological Data Bank, Bangalore, 1988.

25. Ramasastri, K.S., Prathap Singh and Seth, S.M., 'Computerised Processing and Analysis of Raifall Data', Hydrology Journal of IAH, Vol.VIII No.3, 1985.

26. Ramasastri, K.S., 'Quality Control Procedure for Precipitation Data in an Operational System', Report submitted to National Weather Service, USA.

27. Reddy, M.S., Structure of National Hydrological Data Bank', Seminar on National Hydrological Data Bank, Bangalore, 1988.

28. Sherman, R.J., 'Computer Quality Control of Daily and Monthly Rainfall Data', Meteorological Magazine, 104, 1975.

29. Sathish Chandra, 'Current Status of Hydrological Data Observation, Recording, Processing, Storage and Retrieval', Seminar on National Hydrological Data Bank, Bangalore, 1988.

30. Taylor, M.E.V., Thompson, S.N. and Rodgers, M.W., 'MICROTIDEDA', Microprocessors in Operational Hydrology, WMO Publications.

31. Venkata Rao, K., 'Introduction to Quantitative Techniques and Data and Processing', S.Chand & Company (Pvt) Ltd., New Delhi.

32. Whetstone, G.W. and Grigoriev, V.J., 'Hydrological Information Systems', UNESCO-WMO, 1972.

33. WMO, 'Data Processing by Machine Methods', Tech. Note No. 74

34. WMO, 'Machine Processing of Hydrometeorological Data' Tech. Note No. 115, WMO-275, 1971

35.WMO, 'Guide to Hydrological Practices- Vol.I, Data Acquisition and Processing', No.168, 1981.

36.WMO, 'Guide on the Global Data Processing Systems', No.305, 1981.

37.WMO, 'Case Studies of National Hydrological Data Bank', Operational Hydrology Report No.17, WMO No.576, 1981.

38.WMO, 'Methods of Correction for Systematic Error in Point precipitation Measurement for Operational use', Operational Hydrology Report No.21, WMO No.589, 1982

39.WMO, 'Guidelines for Computerised Data Processing in Operational Hydrology and Land and Water Management', WMO No.634, 1988.

PROJECT DIRECTOR	:	DR. SATISH CHANDRA
PROJECT COORDINATOR	:	DR. S M SETH
STUDY GROUP	:	CHANDRAMOHAN, T. SCIENTIST 'B'
DOCUMENTATION/OFFICE STAFF	:	T.P. PANICKER