TRAINING COURSE

ON

SOFTWARE FOR GROUNDWATER DATA MANAGEMENT

UNDER

WORLD BANK FUNDED HYDROLOGY PROJECT

LECTURE NOTES
ON

HYDROLOGICAL DATA AND ITS MANAGEMENT

BY

S M SETH

ORGANISED BY

NATIONAL INSTITUTE OF HYDROLOGY
ROORKEE - 247 667
INDIA

HYDROLOGICAL DATA AND ITS MANAGEMENT

1.0 INTRODUCTION

The wealth of data on its natural resources is one of the major assets of any nation. This is particularly true of data in hydrology. The efficient and economic planning in water resources affects all aspects of national development. It depends on the availability of proper and good quality data about the existence of water, its quality, location, distribution and variation and other related climatic conditions. These are all important factors having a crucial bearing not only on the planning but also on the design and operation of the water resources projects. With the development and use of computer based modern techniques for modelling of hydrologic systems of river basins, good quality of data are required. But in the absence of a well established information system, that can provide scientific data of required quality and quantity, it will not be possible to use the available information processing technologies for handling the intricate problems in water management.

National Water Policy has clearly laid down that a standardized 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 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. Under the hydrology project also, the creation of national data bank is envisaged.

In this lecture, general data requirement for hydrological studies is discussed. The main features of a hydrological data base management system are described. The computerised hydrological data based management systems developed and being used by various data collecting agencies to manage different types of hydrological data are also described along with the problems and limitations of such data base management systems.

Needs for an integrated hydrological data base management system having the capability of retrieving the different types of data at a time are brought out highlighting the advantages of such systems over the existing systems of data base management.

2.0 HYDROLOGICAL PROBLEMS

Hydrology deals principally with movement, distribution and storage of moisture. Most hydrologic problems are related to either quantity or quality of water or both. Determination of water yield, duration and inter arrival time of flood peaks, dam breach, etc. are some typical water quantity problems. These problems can be addressed in:

(a) Time domain : involving reconstruction of the past (prediction) and construction of the future (forecasting) on different scales, viz. continuous time or descrete time such as hour or less, daily, weekly, ten daily, monthly, seasonally, annual and longer.

(b) Space domain : involving spatial variability and its sampling, regionalization, effect of land use change, etc. on different scales such as channel, field or plot, watershed, river basin consisting of number of watersheds, continental or global.

(c) Frequency domain : involving determining frequency of extremes (high as well as low), volumes, means, hydrologic space time characteristics, etc.

Some examples of hydrological problems as listed by Singh (1989) are reproduced in Tables 1, 2 and 3.

Table 1 : Some Example Problems at Different Time Scales

Scale	Example Problem
Short time or continuous time	a. Flood routing through a reservoir or channel reach. b. Determination of direct runoff due to a specified rainfall event. c. Infiltration during a rainfall episode d. Snow melting e. Transport of contaminants by urban storm water.
Daily	a. Water balance b. Flow forecasting c. Precipitation forecasting d. Reservoir operation e. Evaporation and transpiration
Weekly	a. Water balance b. Evaporation c. Canal operation for farm irrigation. d. Weather forecasting e. Flood damage assessment
Monthly	a. Water balance b. Evaporation c. Water supply d. Weather forecasting e. Effect of land use changes
Seasonal	a. Water balance b. Weather forecasting c. Crop yield d. Agricultural pollution e. Irrigation water supply and scheduling f. Effect of droughts
Yearly	a. Water balance b. Evaporation c. Maximum annual flood values d. Sediment deposition and aggradation e. Effect of land use changes f. Drought modelling
Longer time	a. Effect of land use changes b. Groundwater movement c. Migration of chemicals in surface water d. Migration of chemicals in ground water e. Ecosystem modelling f. River training works g. Drought modelling

Table 2 : Some Example Problems at Different Space Scales

Scale	Example Problem
Short distance or continuous distance	a. Precipitation intensity b. Surficial roughness c. River meandering d. Aggradation and degradation e. Crop irrigation
Channel	a. Open channel flow b. Flood routing through a channel c. Erosion and sediment transport d. Geomorphic features e. Stage-discharge (rating) curve
Plot or field	a. Hydrology of a parking lot b. Overland flow from a field c. Farm irrigation efficiency d. Interrill erosion e. Land use change
Watershed	a. Unit hydrograph b. Unit sediment graph c. Effects of land use change d. Water balance e. Surficial features
River basin	a. Multiobjective development of water resources b. Regional hydrologic analysis c. Interbasin transfer of water d. Droughts e. Effect of land use changes, especially structural
Continental	a. Climatic prediction b. Droughts c. Water balance d. Flood damage e. Atmospheric pollution and its effects
Global	a. Water balance b. Climatic pattern and changes c. Earth-atmosphere-ocean interaction d. Droughts e. Effect of such human activities as unclear explosion

Table 3 : Some Example Problems in the Frequency Domain

Scale	Example Problem
Extremes	a. Extreme rainfall (hourly, every 6 hrs, daily, etc.) b. Annual peak discharge c. Extreme annual low discharge d. Hurricane and tide flooding e. Extreme annual temperature
Volumes	a. Yearly water yeield b. Annual sediment yield c. Annual rainfall d. Annual drop in water table e. Annual flood damage
Means	a. Mean daily temperature b. Mean daily river flow c. Moving average of a time series d. Mean daily rainfall intensity e. Mean reservoir outflow

In general, various approaches to the study of hydrologic problems can be grouped under two categories: (1) physical science approach - also referred to as a basic, pure, casual, dynamic or theoretical approach, and (2) systems approach - also known as an operational, applied, empirical, black box or parametric approach (Fig.1).

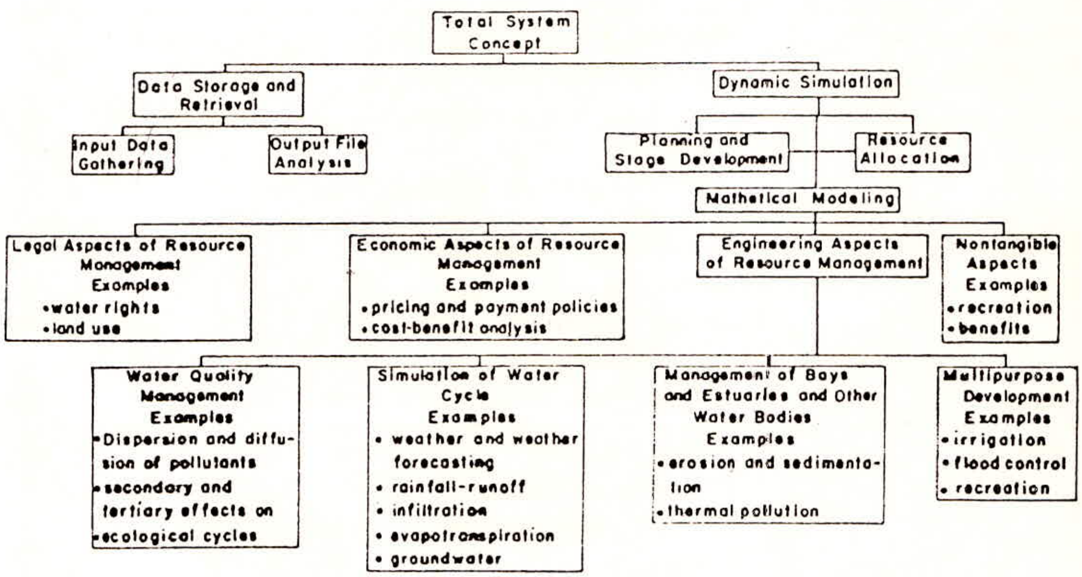

Fig. 1 : Nature of hydrologic problems from a systems point of view (Singh, 1988)

Water management is the application of all available knowledge to the practical development of water resources. One of the fundamental sciences of water management is hydrology. The engineers engaged in design, construction or operation of hydraulic works must solve practical problems. These are of varied nature and in mostcases hydrology is needed for their solution. These include (a) rural water management, (b) river training, (c) municipal water management, (d) structural and hydrualic design or water control structures for differe sâéâes. Some typical questions that the hydrologist is called upon to answer are: (i) Is the flow of a particular stream at a particular site is sufficient to meet the needs of (a) a city or industry seeking a water supply, (b) an irrigation project, (c) a proposed power development, (d) navigation, (e) recreation; (ii) Would a storage reservoir be required in connection with any of the proposed uses and if so, what should be its capacity? (iii) In the design of a flood protection system, a bridge a culvert or a spillway for a dam, what is the maximum flood that may be expected to occur with any specified frequency. (iv) What would be the effect of draining upland area or a swampy region upon the flow of the stream from watershed? (v) How would certain changes in land use or the removal of forests, affect the groundwater level or the stream flow from such an area?

3.0 GENERAL DATA REQUIREMENTS FOR HYDROLOGICAL STUDIES

In order to treat the land phase of the hydrologic cycle as a closed loop, data are required to define the input and output from this phase. In addition, the analysis and prediction of the response of a catchment to an input requires information on the physical characteristics of the catchment including the data related to the land use changes.

Computer simulation techniques in hydrology range from relatively simple models to the highly complex. Each model requires data in a computer compatible form, sufficient to conduct the analysis to the required level of accuracy. As a rule, the more complex the model, the greater the requirement for detail and accuracy in the data used by the model. Broadly, the required data may be classified in two groups, viz. (i) the time series data, and (ii) spatial data. The time series data are required to be monitored regularly. The network for each type of time series data should be adequate in order to provide the proper spatial representation of the hydrological variable being measured. The time interval of monitoring the variables should be decided considering their temporal variability and catchment characteristics. The measurements of a variable within a time interval are considered to be reasonably uniform. On the other hand, the spatial data are generally fixed and do not vary much with time unless some major changes have taken place in the catchment due to the construction of the water resources projects and the land use changes.

The general requirements of data for hydrological studies are given hereunder:

(a) Hydrometeorological data
- Precipitation
- Evapotranspiration
- Radiation
- Temperature, Humidity and vapour pressure
- Wind speed and wind direction

- Cloud cover
- Sunshine

(b) Surface Water Data

(i) Water quantity data
- Water level in streams, lakes and reservoirs
- Discharge rate

(ii) Reservoir or lake data
- Storage-elevation relationship
- Area-elevation relationship
- Elevation-discharge relationship

(iii)Water quality data - for streams, lakes and reservoirs, etc.
- Physical
- Chemical
- Biological
- Microbial

Here physical water quality parameters include:

* Air temperature, surface water temperature, transparency, stability, pH, salinity and conductivity.

Chemical water quality parameters generally include:

* Carbonate alkalinity, bicarbonate alkalinity, dissolved oxygen, nitrate, phosphate, sulphate, ammonia, organic nitrogen, biological oxygen demand, hydrogen sulphide, chemical oxygen demand.

Biological water quality parameters generally include:

* Chlorophyll, algae, zooplankton, invertebrates, primary production.

Microbial water quality parameters include:

* Total viable count, total coliforms, faecal coliforms, faecal streptococci and heterotrophs.

(iv) Water use data
- Effluent to rivers, lakes and reservoirs. (quantity as well as quality)
- Abstractions from rivers, lakes and reservoirs.

(c) Ground Water Data
- Levels of wells
- Log data
- Spring flows

- Well/aquifer test data
- Pumping and recharge data
- Water quality data of ground water reservoirs

(d) Process based data
- Infiltration
- Soil moisture
- Seepage
- Baseflow

(e) Physical parameters

(i) Land surface physical parameters
- Area
- Elevation (topography)
- Slope
- Overland length and slope
- Vegetation
- Soil

(ii) Natural drainage channel network
- Geomorphology
- Channel cross sections
- Roughness characteristics

(iii) Aquifer systems characteristics
- Thickness
- Porosity
- Storage coefficient
- Transmissivity
- Dispersivity

(f) Ecological Data

Various constituents of the ecological system are of concern in hydrological investigations. Of particular interest are the vegetal cover over a region, microbial organisms in water (such as planktons) and in the biosphere, and small plants and animals that have a direct relation with water quality such as algae, weeds, shrimps and fishes.

(g) Geologic data

Quantitative assessment in these areas is often difficult, and much information exists in 'soft' form i.e. in the form of reports, photographs and file cabinets. The pertinent information consists of sedimentation, lithology, bathymetry, hotogeology and environmental geology.

(h) Land use data

Human activities such as industrialization and agriculture, with their accompanying paraphernalia, affect and alter the hydrological regime in countless ways. The major land use forms are alteration of the vegetal cover (such as forests to grasslands or farmlands), occupancy and use of land for residential and commercial purposes, and changes introduced in water conveyance and storage systems. Urbanisation has also led to rapid changes in water needs and uses for domestic and industrial purposes.

(i) Socio-Economic Data

Hydrological planning and management are largely dependent on socio-economic considerations. These include demography, business and occupational economy, educational, welfare and recreational activites, socio-economic developmental trends and programmes, and government and local rules and regulations.

4.0 CHARACTERISTICS OF HYDROLOGY RELATED DATA

There are several characteristics of hydrology related data which must be considered in order to organise a hydrological data bank. These character-istics may be broadly described as follows:

(a) Data derived from the field

Some hydrological related data are directly measured from the field viz. time series data of hydrometeorological variables, streamflow, ground water levels, etc. Instruments are installed in the field measuring different variables and the raw data are noted down in the prescribed form meant for the respective variables.

(b) Data derived from the lab

Other kinds of hydrological data are those which can not be directly measured in the field. For generating such data, samples are collected from the field and analysed in the laboratory. The data related to water quality and physical properties of soils are examples of the data which belong to this category.

(c) Data derived from the experimental set up

Information derived from the controlled experiments are invariant subject to the experimental constraints. Such data are readily applicable or adaptable for specific purposes provided these constraints are known. Data generated from the experimental set up of a model of a dam and open channel flows in the flumes are examples of such type of data.

(d) Data derived from the maps

Some hydrology related data are derived directly from the maps. Topographic elevations, geomorphology, types of soils and vegetation, etc. are some examples of these types of data. Such data are derived directly from the available maps. Now-a-days digitizer linked with personal computers are being extensively used to create the data base of the maps. Subsequently these data base are being utilised to retrieve the required data from the data base of the maps.

(e) Data derived from the satellite imagery and CCTs

Application of remote sensing techniques provide very useful information about the various hydrology related data. Visual interpretation of the satellite imagery is very much helpful for preparing the soil and land use maps and also for identifying the water bodies on the maps. Image processing software are applied to process the computerised information available on CCTs which may provide useful information for the hydrological studies for the area of interest particularly for remote areas.

(f) Data based on the historical statistics

The data such as socio-economic data are based on the historical statistics which are continuously changing over the time. Hence such information must be compiled and updated periodically.

(g) Time and space coordinates

A hydrological event is the consequence of manifold processes whose domain in space and time are indefinite. However, the relative importance of each factor depends on its spatial and temporal configurations. Hence time and space coordinates must be directly linked with all historic data.

(h) Quality of data

The data as acquired by human efforts depend on human limitations, environmental conditions and the means used to acquire them. Thus an estimation of the reliability and accuracy of data is necessary.

(i) Quantitative and qualitative data

Quantitative data are relatively easy to deal with as they are amenable to mathematical representation and analysis. Qualitative data can not be subjected to mathematical or statistical interpretation without at least some loss of information. In connection with automated data storage and retreval systems distinction is made between 'hard' and 'soft' data. 'Hard' data is the term applied to numerical and narrative data that can be processed by a digital computer. 'Soft' data are those which cannot be processed by digital computers such as information contained in reports, photographs and file cabinets.

(j) Units and format

Quantitative information of the same type may be expressed in various units. For ready use they must be converted to some units, consistent with international conventions, and stored in specific formats.

(k) Use-orientation of data

Hydrological model studies often require the raw data in altered forms. Models may be deterministic or stochastic. For the application of these models different types of data are required from the data bank. Thus, the management of each data item separately may not be an efficient way of data handling. It requires an integrated approach which may provide the required data of different kinds to the users after specifying some specific details about the study area. The management of data also depends on their use-schedule viz. real time and non-real time studies.

The full report of the above characteristics is realised in the efficient design of data libraries.

5.0 DATA BASE MANAGEMENT SYSTEM

The data base management system is evolved with an objective of creating a hydrologic data banks or library. Such system forms the vital link between raw data collected in the field, and their ultimate use in the hydrological studies. Fig. 2 shows the organisation of a typical hydrological data bank. There are several discernible operational procedures for the management of the hydrological data base. These may be summarised as follows:

(i) Data aquisition (information inputting)
(ii) Data processing
(iii) Data storage and retrieval
(iv) Data presentation (information outputting)

5.1 Data Aquisition

Raw data are acquired from various measurement and collection agencies either in manuscript form or on the secondary storage devices of the computers, viz. floppies, tape, microfishes, cartridge, etc. The data collected in the manuscript form are required to be entered on the computer before taking up the data processing task. Thus there are two steps involved in the data aquisition i.e. (i) data collection, and (ii) feeding of the data on computer (if not available on computer compatible secondary storage devices).

(a) Data collection

The collection of hydrological data forms an important and most difficult part of the data base management system. Normally the data collection programs must provide the following information for each area under consideration.

(i) Organisation or agencies maintaining the data

(ii) Representative data on the areal and temporal variability of hydrometeorologic process.

(iii) Representative data on the physical catchment characteristics

(iv) Key stations permanently operated, providing continuous information on dominant input output such as rainfall, radiation, temperature, evaporation and streamflow. Telemetry should be possible at these stations.

(v) A routine review of the degree of representation of the existing data collection program with a view to improvement (network design).

(vi) Automatic abstraction, updating and publication of all data.

(vii) The routine filing in computer form of all data collected by different organisations with a central regional agency for use by designers, planners, and researchers.

The organisations presently collecting the different types of hydrology related data in India are given in Appendix-I. However, a questionaire may be circulated to different data collecting agencies asking for the types of data being collected by them along with the location of the observing stations on the map being maintained by the respective organisations. Such efforts would be helpful for identifying the various agencies, collecting the different types of data. The duplication, if any, in the data collection works would also be known through this effort. Based on this, further action may be taken to avoid the duplication in the data collection works in future.

(b) Data abstractions and feeding on computer

Three steps are involved in taking raw data and entering them into the computer memory. These are data abstraction, data conversion to a computer compatible form, and finally data input and storage. Data abstraction entails the preparation of data from the original forms on which they were measured into a form convenient for processing. Hydrometeorological data involves the greatest effort in abstraction and input organisation. Consider as an example the abstraction of rainfall. Measurements of rainfall and other hydrometeorologic data are normally recorded on strip charts or are manually tabulated, if the readings are taken by an observer. When a manual reading or strip chart record is taken, it is returned to central office of the recording agency where it is checked for errors and corrected. River stage records require particular attention in checking for errors due to timing or gauge malfunction.

Following this initial check, the data are abstracted from the original charts and forms onto standard listings called as coding sheets. The purpose of the computer coding forms is to arrange the data in order, within a specified spacing or format corresponding to the program statement which reads them into the computer. They also allow rapid data preparation by computer key punch or paper tape punch operators.

The abstraction procedure is similar for all types of hydrometeorologic data. Different formats can be designed for different requirements. Once abstraction has taken place, the data are then ready for conversion to a computer compatible form. Three primary methods exist for feeding information into a digital computer, card input, paper tape input, and direct terminal input. These input devices dictate the form of the data input. Third form of the input, i.e. computer terminals are becoming much popular in comparison to the other two forms of the input. The computer terminal is a device which resembles a standard electronic office typewriter, with electronic conversion facilities which translate the typed character into a pulse to be transmitted to the computer over standard telephone lines. The typewriter simultaneously prints the typed characters onto a paper in the normal fashion. Other computer terminals replace the paper copy with a cathode ray tube (CRT) output. When the terminal is 'signed on' to the computer the typed information is transferred directly to a buffer in the computer memory and can be stored on further input instructions.

5.2 Data Processing

Prior to the storage of inflowing data it is imperative that the information be tests, suitable rectifications be made, and some measure of its quality established. The first concern arises with respect to the time base. The data collected from various sources is not likely to have a common time base. A computer based control and data logging system is, therefore, necessary, not only for precise time correlations, but also for real time data used in water resources systems operation and management. Apart from real time considerations, data processing is concerned with quality control, editing, analysis and synthesis of the collected information. The former two processes are sometimes collectively called primary processing, and the latter secondary processing.

5.2.1 Quality control

Users of data must be aware of the potential problems arising in hydrological studies as a result of the quality of data. The measurement of data is affected not merely by the method employed, but also the field conditions at the time of measurement. With reference to the former a list of standard measurement/observation procedures are recommended for use. For the field situation specific factors need to be recorded which represent the environmental condition.

Errors and inconsistencies in raw data exist due to a variety of causes. The main types of error are:

(i) Procedural errors: Arising due to improper techniques, measuring devices, base information or calibration charts.

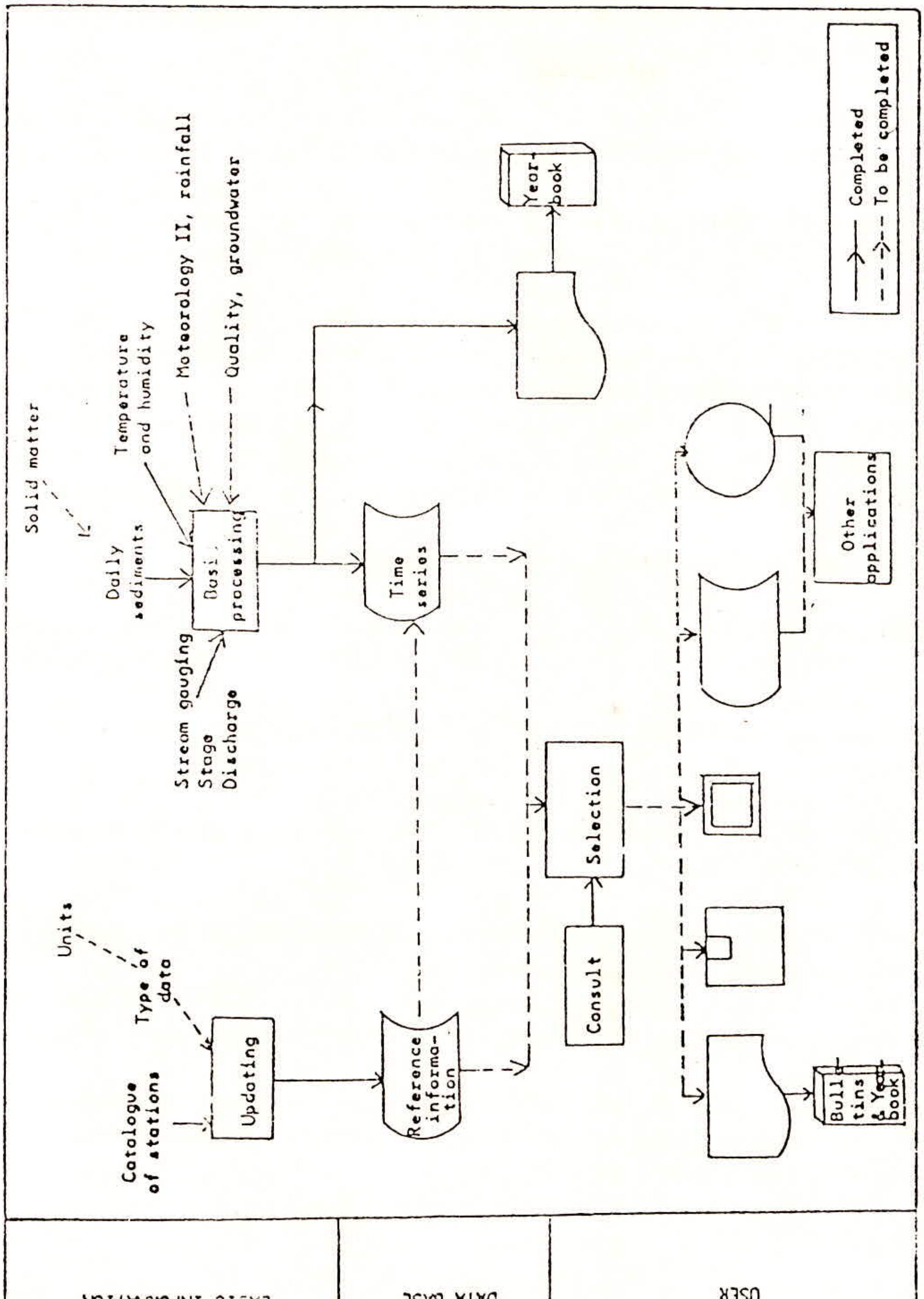

Figure 2 : A Typical Hydrological Data Bank

(ii) Operational errors: caused by the faulty operation of the adopted mechanism or procedure, or the result of human limitations.

(iii) Peripheral errors: are independent of the procedure adopted,or their execution, such as unknown or unpredictable environmental factors.

(iv) Non-chargable errors: occur after measurement, i.e. while recording or transferring information. They may also arise due to misinterpretation of procedural instructions.

Due to the varied and uncertain nature of errors many of the errors may not be identifiable or rectifiable. Thus quantitative checking cannot completely root out the errors, though values can be labelled suspect where appropriate. Routine error corrections are useful where a definite trend or correlation is expected, and are usually made by computerized methods. However, automated quality control programs are limited in scope, and may not be completely reliable. Hence they must be supplemented by manual intervention.

The commonly employed quality and consistency checks are:

(i) Testing for absolute or physical limits of data (such as negative inflows to reservoirs, measurements outside range of instrument capacity), i.e. physical consistency.

(ii) Checking for mutually conflicting sets of data values within a set of data, i.e. internal consistency.

(iii) Double mass curve analysis (for rainfall, runoff, etc.) of contemporary sets of data of neighbouring stations.

(iv) Time series analysis to detect changes in the homogeneity in time series. This is a valuable supplement to double mass analysis.

(v) Cross corelation and regression analyses to compare between related types of data.

From the above tests conflicts between data values are resolved, and missing data filled up where possible. Correlation coefficients and regression equations are also computed where useful and physcially meaningful.

5.2.2 Editng

Editing consists of conversion of the semi-processed data into those with consistent units and formats. Usually the following international conventions are used:

(i) Universal Decimal Classification (UDC) : sponsored by the Federation of Internationale de Documentation (FID) for subject and miscellaneous classification.

(ii) I.S.O. recommendation for common elements and their units.

The above recommendations do not encompass the whole spectrum of hydrology related data. Additional nomenclature, notations and units are sometimes required. Such specifications have been compiled by WMO, NAWDEX and other large scale organisations. In addition to the symbols and units, formats are standardized according to the type of data and economy of storage.

5.2.3 Analysis and forecasting

Many of the hydrological variables recorded are not directly used for modelling purposes. Often a set of data is reduced to, represented by or augmented with additional parameters (statistical or otherwise) in actual computational studies. Since these parameters are frequently used, it is convenient to develop routine programs for computing and storing them in the data banks. Such computations include areal estimates of data (precipitation, runoff, evaporation, etc.) statistical moments, and correlation and regression analyses. Typical simulation and optimization models are also run on the data depending on their expected needs.

Information available in analogue form often need to be digitized for purposes of processing storage and ready reproducibility. Spatial information may be stored in two or three dimensional grid systems and the necessary quality control tests performed on them. The process of digitization may be achieved manually by digitizing tables, or by automated or semi-automated digitizing instruments. The conversion process can be checked by computer plotting of the data into overlays and comparing with the original traces.

The acquiring of fresh data necessitates the updating of the existing data file by comparison with the new data values, and recomputing parameters if necessary. Thus time series and auto-correlation analyses are re-invoked to ensure consistency of data in time, and to establish long-term trends and statistical moments. Moreover, if the files in which data are to be stored are of predefined lengths, it is necessary to check for storage problems, and reallocation of storage space is necessary.

For economy of storage as well as for ease of data handling, it is necessary to reduce data to a minimum. While some redundant data can be eliminated by processing, quality control and corrective measures may actually increase the volume of information. Further, due to uncertainty in the nature of data itself, it is desirable to preserve the original information in addition to the corrected values. One way of reducing the amount is by wiping out data values which, within a certain tolerance interval, can be interpolated between neighbouring values. This procedure can significantly compress the body of information for continuously recorded data.

Water resources and water quality regulating services like reservoir operation, toxicity warning and flood forecasting, may necessitate real time data handling. In such cases measurement stations must be equipped with suitable devices (such as telemetering equipment, on-line connections to computer, etc.) for immediate transmission of data to the data bank. Real time data processing involving quality control is difficult since manual checks may not be possible, and automated quality control can misinterpret and distort data resulting in crucial errors. Real time data handling is highly complex and depends largely on the specific use of data and the available computer software.

5.3 Data Storage and Retireval

Traditionally, the hydrological data like other types of data are being stored on registers or files using manual methods. This involves storing of data in registers as and when it is received from the field observation sites. This type of storage has many problems associated with it. Most pronounce being that data inventory is not available, data security is doubtful and retireval of the data is very cumbersome. Since no inventory is available many times it is not known whether data of a particular case is available or not. Whenever data is required for some analysis it is either manually copied or lately photocopied. To avoid this problem, it is necessary to have computerised storage and retrieval of hydrological data. The computerised storage and retrieval consists of creating a data base, updating it frequently as and when new data are available and retrieving the data from this data base when desired. A software system is used as an inter-face between the user and data base.

Data base mangement concepts are used in storing information with relevant site/instrument specifications and quality indices. Data are stored in graps in direct, sequential or random access files depending on the data base structure.

The format of data base with regard to time has several different but classifiable structures. In fact, most natural resource data needs two storage formats for each sampling point. The first format is required for fixed (or stationary, or non-time series) data which does not vary, or varies only slowly, with time. The second format is for the time series: the series of data values needed to represent physical parameters having significant temporal variation. Raingauge location and elevation are items of fixed data, whilst the daily rainfall values recorded at the gauge represent a time series. Time series data may be categorised as continuous, regular, or irregular in time. Examples of these three series are water level tracings on charts, daily rainfall totals, and flood gaugings. Fig. 3 shows the format of time series records.

The above distinctions are made because the different cateogires need handling in different ways. Regular time series require only the start date, the interval between observations, and the number of observations stored in each physical computer record, i.e. whether one record contains ten days, one month or one year or daily data. For irrgular time series, each observation must be accompanied by its relevant time reference. Continuous series need to be converted into digital form for storage, and this digitization process will create a regular or irregular series. When a digitizer is used to abstract chart data, an irregular time series format is produced, even though the points may have been abstracted at regular time intervals. This is because the digitizer generates pairs of time value (X, Y) coordinates. It is possible to convert this to a regular series for storage purposes.

A special type of time series common in water data is the multiple parameter time series. These are series where several observations are made at the same site. Examples are climate stations and water quality sampling points. Much of this data normally comes from the field on a single data sheet, and is input in this format. Whilst the data could also be stored in this way, there are advantages in an alternative format which separates all the variables into several single parameter time series files.

A variation of the multiple parameter series is encountered when data are sampled at various depths at the same location, e.g. water quality and certain sediment samples. These types of data require space, time and depth references. In an exactly analogous way to time, the regular or irregular spacing of sampling depths will determine the type of depth reference required.

Typical spatial data structures are shown in Fig. 4. These include gridded or vectoral data structures, and it must be recognized that in future, the collection of some types of water data by satellite remote sensing will require such geocoded data base systems. However, very specialized software such as Geographical Information System (GIS) is required to handle data of this type, a major problem for remote sensing data being compression of the vast quantities of data to an amount which may be included in the usual scale of most data bases.

After carrying out validation checks and primary data processing, the resulting data is used to update the data base. This updating may be done at some pre-determined frequency usually monthly or fortnightly. The data base file may be kept on online storage. In the on-line storage mode, the data files are available for operations at all the times. However, this mode of storage is slightly expensive. In the off-line mode the data files are stored on magnetic tapes. This mode of operation is cheaper compared to online and is mostly used to store the data.

Data retrieval system must be able to rapidly retrieve selected data items from the data base. It must be possible to retrieve the data by river basins, sub-basin, station, time period and the parameter. The retrieval package must also be capable of doing simple statistical analysis like computation of mean, standard deviation, correlation, etc. Now-a-days very efficient and interactive software packages are available for data retrieval which can also perform the above required statistical functions.

Effective retrieval of data required not merely their storage in an organised structure (data base) but also familiarity of the user with the data base components and their inter-relationships. It is necessary to provide instructions and software devices to search, display and retrieve data in the desired forms. The instructions for locating data are usually aided by a query facility to serve the common user.

The basic operational links of a computerized data bank are illustrated at Fig. 5. As evident from the figure, the operations of a data library are closely and sequentially connected. As the magnitude and diversity of the information that need to be stored increase the complexities of the processing, storage and retireval procedures increase proportionately. This calls for great skill in the functioning of vast and growing data libraries.

				Repeated pattern of parameter values
Station number	Parameter type code	Time of 1st value	No. of cycles	

(a) Fixed interval

				Observation pairs			
				First observation		Second observation	
Station number	Parameter type code	Time of 1st value	No. of value pairs	Time value	Parameter value	Time value	Parameter value

(b) Irregular interval

				Values for each parameter in the set		
Station number	Parameter set code	Time of observation		Parameter 1 value	Parameter 2 value	

(c) Multiple parameter series with fixed parameter sets

				Pairs of values for each parameter				
				Parameter 1		Parameter 2		
Station number		Time of observation	No. of values	code	value	code	value	

(d) Multiple parameter series with variable parameter sets

Figure 3 : Format of time series records

AREA IDENTIFIER	DATA CATEGORY I	------------	CATEGORY N

DATA CATEGORY 1		
ITEM 1	----------------	ITEM N

ITEM 1			
VALUE 1	VALUE 2	VALUE 3	VALUE 4
24	0	0	76

EACH ENTRY GIVES THE PROPORTION OF THE SUB-AREA WITH THE GIVEN ATTRIBUTE VALUE

or

ITEM 1			
VALUE CODE	AMOUNT	VALUE CODE	AMOUNT
1	24	4	76

ALTERNATE PRESENTATION OF SUB AREA GIVEN OVER TO PARTICULAR ATTRIBUTE VALUE

Fig 4 : Table representation of area and sub-area data

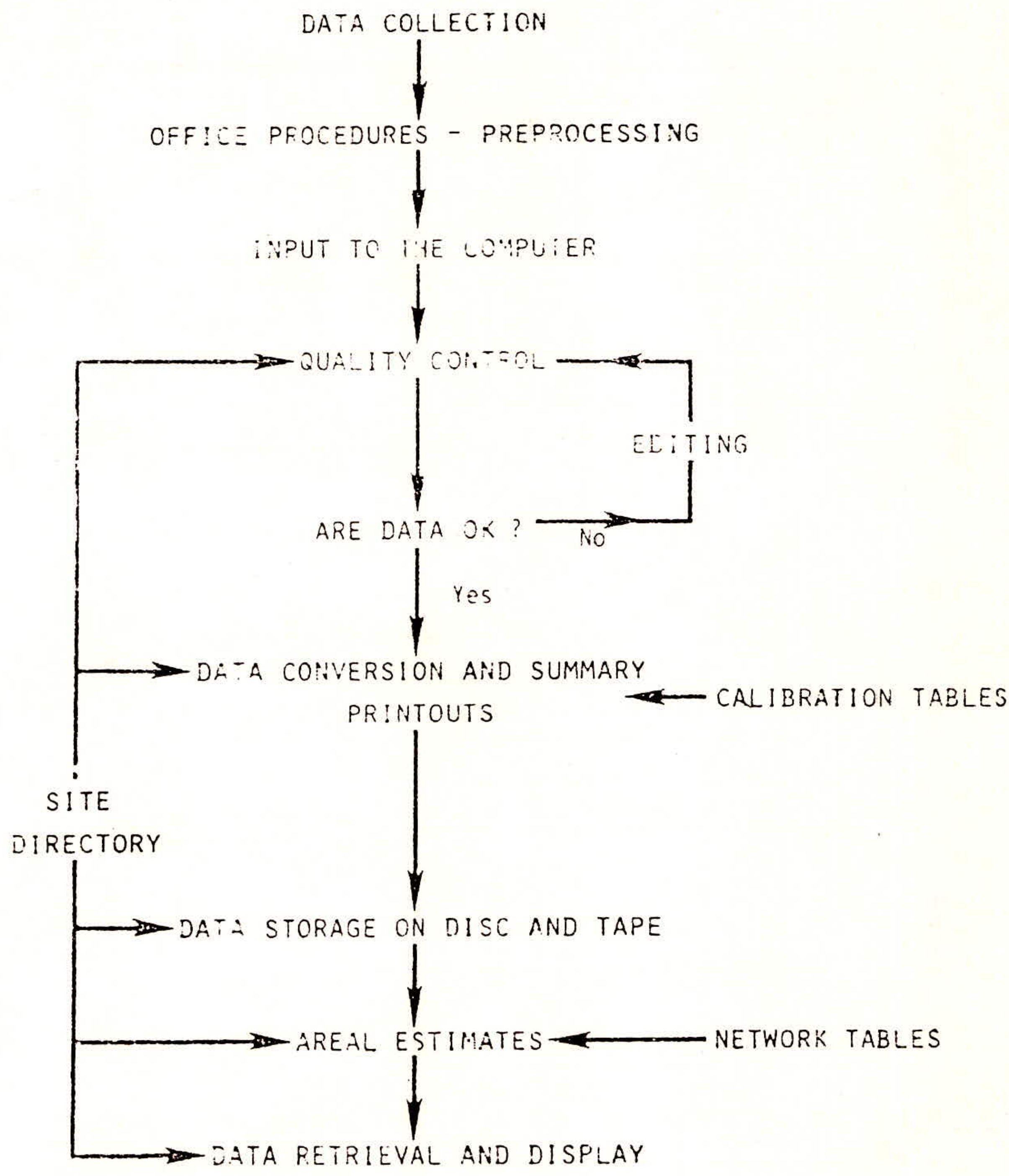

Figure 5 : Hydrological Data Library Framework

Depending on the way in which data are collected and processed, data banks may be centralized, or distributed (co-ordinated). A fully centralized hydrological data bank is physically located at one site and comprises data covering all hydrological variables covering the entire country. Such organisations permit substantial economy of processing and storage, software and management. However, due to the complexity and largeness of structure they may be impractical in big countries with a variety of source agencies and user requirements.

A distributed (co-ordinated) data bank is a co-ordinated or co-operative arrangement between different data banks, or their operating agencies. Such banks are more reliable, managable, and adaptable to changing technologies or users' needs. They are typically used in real time applications where the end user must have easy access to interactive use and control of data in the data bank.

Each data bank needs to be organized according to its specific needs and limitations. Individual data libraries incapable of meeting all user needs directly often maintain two parallel modes of functioning, viz., (a) storage of data as a data base, wherefrom data can be retrieved per se, and (b) indexing of data to provide requisite information on the types and brief description of data available from alternative sources.

5.4 Data Presentation and Publication

The retrieved data may be presented in the most useful form and published in the form of water year book in order to provide an opportunity to the users who are not having access to the computerised data base.

At present the daily rainfall collected by the State Government agencies are published by the respective State Govts. Many State Govts. have not published the data for the last 20 years. While the difficulties of the State agencies in terms of manpower and lack of funds is understandable, the users are deprived of valuable information. The data from IMD is available on magnetic tapes in coded form.

The CWC publishes data year books containing data of daily gauge and discharge, monthly runoff and related data observed at gauges maintained by CWC. These water year books are prepared by the various circle offices and zonal offices of CWC. The Statistics Directorate of Central Water Commission is storing the hydrological data of gauge sites of CWC and States in various river basins in computer compatible storage media. There is no standard format for water data as yet. However, this is being attempted at present.

Very few States like Karnataka, Maharashtra, Gujarat, Bihar and Andhra Pradesh bring out water year books. Although it makes available all the data at one place, these water year books do not follow a standard format nor provide adequate information in respect of the gauges sites, streamflow behaviour, etc.

6.0 COMPUTERISED STORAGE AND RETRIEVAL SYSTEMS IN INDIA - A STATUS

The computer based system for storage and retrieval of hydrological data are now in vogue for more than a decade in many countries. In India also, a number of organisations dealing with hydrological data are attempting to develop such systems. 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. As regard to State Government Departments, where bulk of the data lies, a perceptible beginning is yet to be made.

The major central agencies involved in development of computer based storage and retireval system are CWC, CGWB, IMD and NRDMS under DST. A brief description follows:

6.1 India Meteorological Department (IMD)

The India Meteorological Department (IMD) has been storing climatological data on punch cards for last four decades. Regarding rainfall data, 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 number.

The IMD has assigned a unique 3 digit catchment number to different catchments in India. In the 31 cards format, the catchment number, sub-division number and station number were recorded in each card. For each year, 31 cards were required and each card contains data for a specific date of each months.

For storage of data in 24 card format, 2 records are needed for each month. The fields in each record are catchment number, latitude, longitude, station number, year, month and 16/15 rainfall values.

A different scheme is used for storage of hourly rainfall data of the self-recording raingauges. This format includes element code, index number of raingauge station, year, month, card number (either 1 or 2) and hourly rainfall values. The second card also has field for amount and duration of maximum one hour precipitation during the 24 hr. period. These formats are shown in Fig. 6.

6.2 Central Water Commission (CWC)

The Central Water Commission is engaged in computerised storage and retrieval of hydrological data for quite some time. Recently, they completed a major exercise of coding the data. In their scheme, the entire country has been divided into 20 sub-basins. These have been further sub-divided into sub-basins. Codes have been devised for these basins and sub-basins. The coding for individual stations is under progress. It is understood that the past data is being converted into the new devised format. The format which has been in use in CWC is given as Fig. 7.

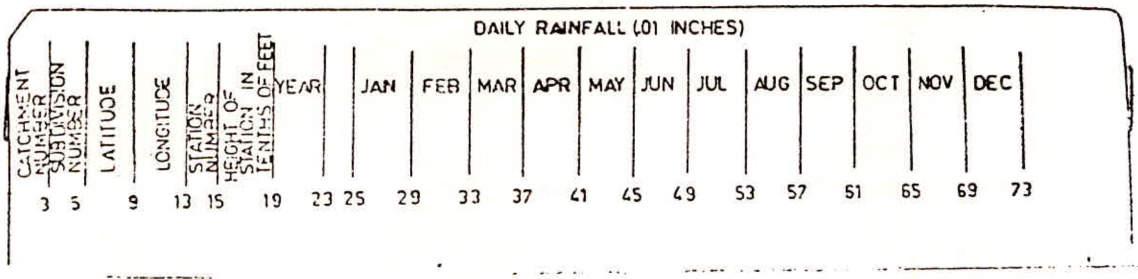

Fig 6A : Daily rainfall - 31 card format

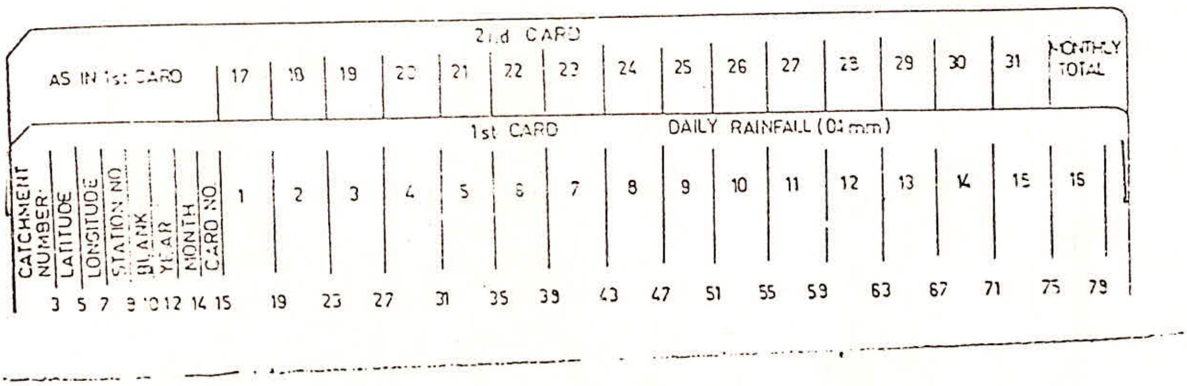

Fig. 6B. Daily rainfall - 24 card format

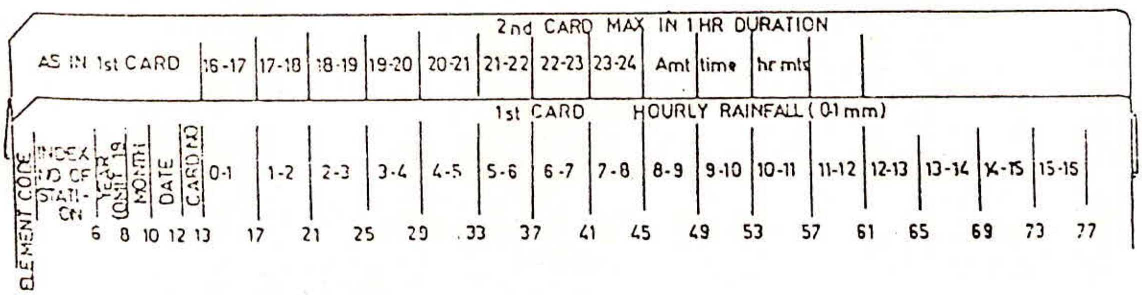

Fig. 6C : Hourly rainfall format

1-5	6-12	13-14	15-16	17	18-22	23-29	30-34	35-41	42-46	47-53	54-58	59-65	66-70	71-76
Blank	Site code	Year	Month	Card No.	Gauge Day-1	Discha-rge Day-1	Gauge Day-2	Disch-arge Day-2	Gauge Day-3	Disch-arge Day-3	Gauge Day-4	Disch-arge Day-4	Gauge Day-5	Disch-arge Day-5

Note: In card No 7; Col. 30-34 contain zero of R.L., and
Col. 42-46 contain temporary R.L.

Fig. 7 : Stage discharge data format adopted by CWC

CENTRAL GROUND WATER BOARD
HYDROGEOLOGICAL DATA FILE
AND UPDATE FILE

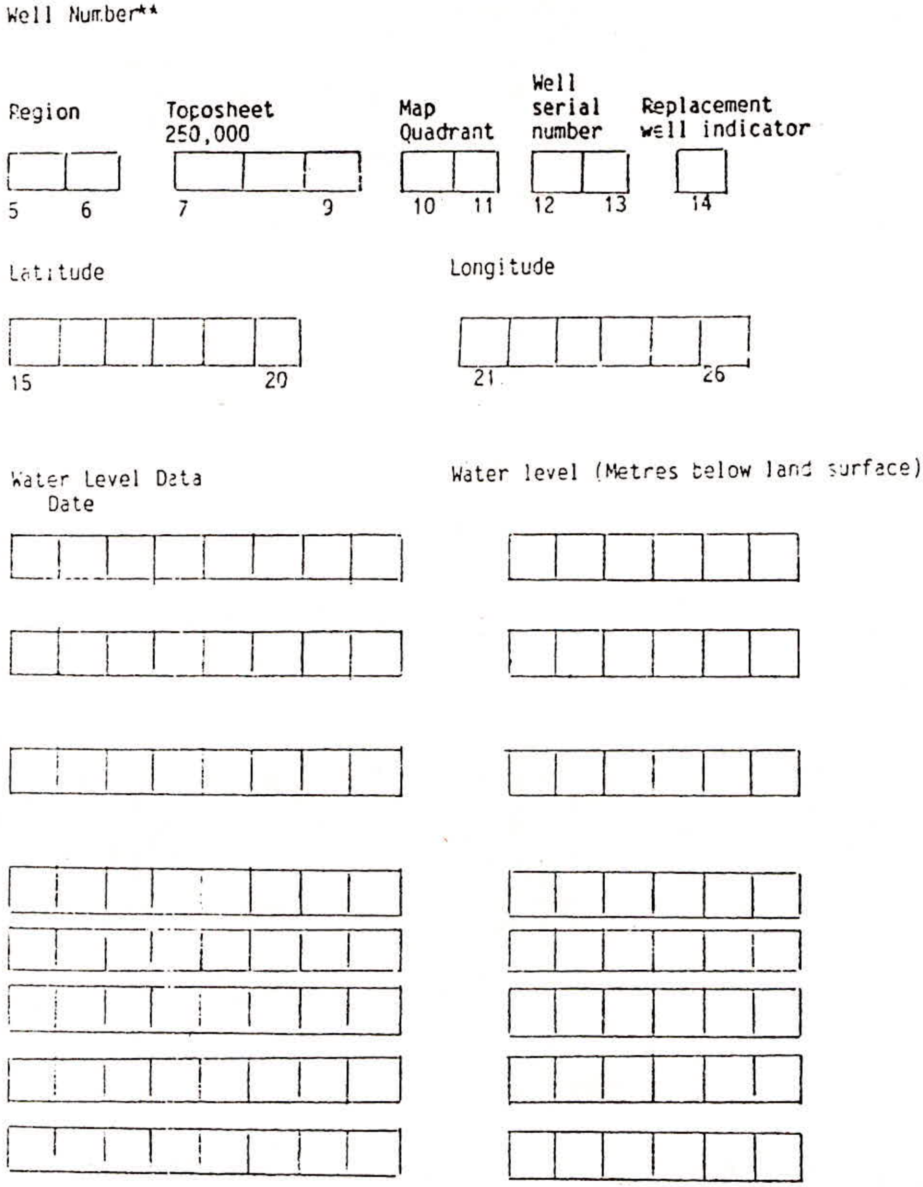

Fig. 8: Format for Ground Water Data Storage adopted by CGWB.

CENTRAL GROUND WATER BOARD

MASTER FILE

FINDEX

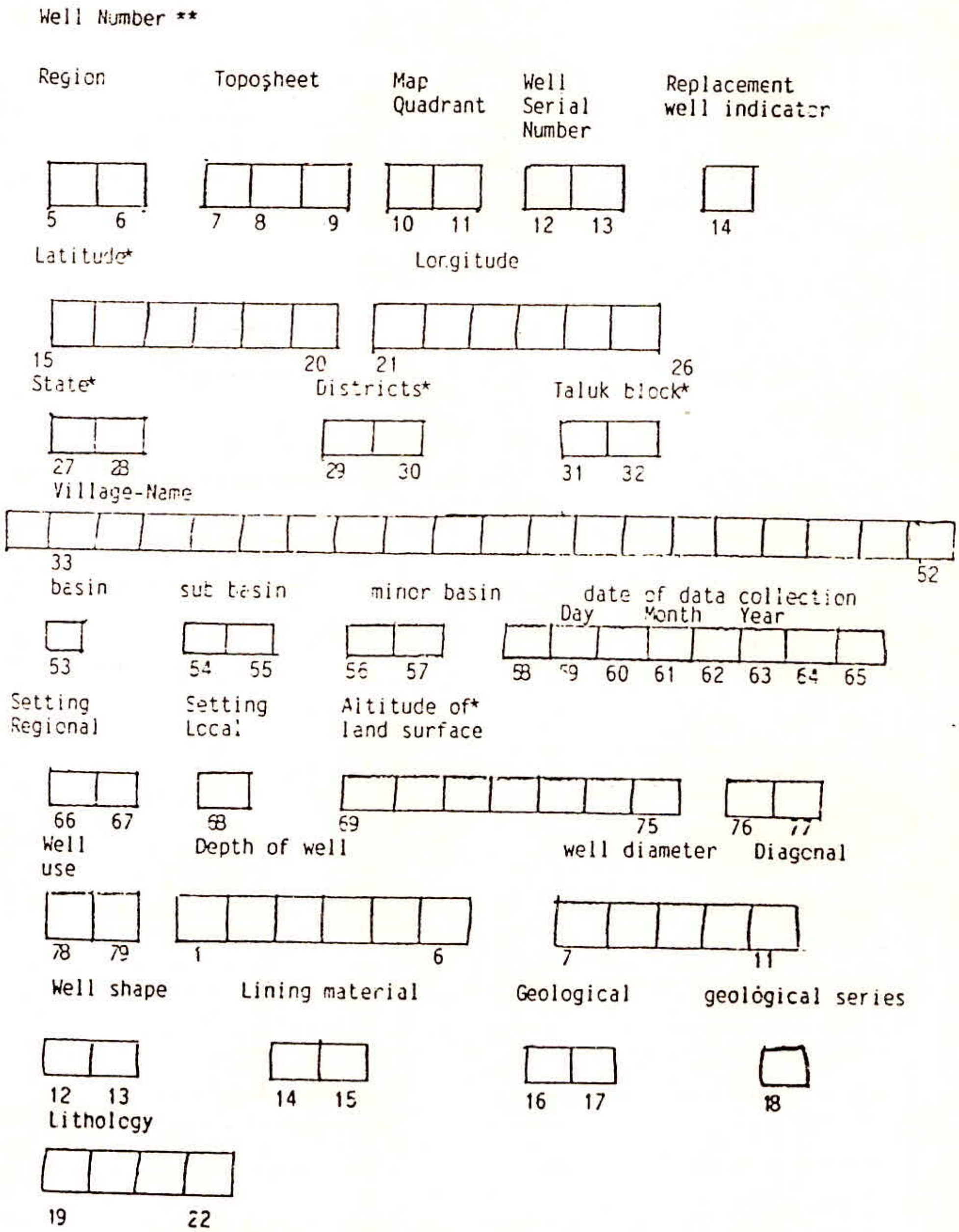

Fig. 9 : Format for Ground Water Data Storage adopted by CGWB.

6.3 Central Ground Water Board (CGWB)

About five years ago, Central Ground Water Board started the work for computerised storage and retrieval of ground water data. As first part of this work, they took up computerisation of data of National Hydrograph Stations only. An elaborate coding mechanism has been developed by CGWB for this purpose. Codes have been assigned to each State, district and river basin in which the station lies. The information stored also includes the latitude, longitude of a well and a well number. Three character codes have been chosen for the geology of the well site and water quality parameters. Further, the lithology of the geological units have also been codified. The data format adopted is shown in Fig. 8 and 9. The data has been computerised on IBM-PC compatible computers. The software package dBASE III Plus has been used for storage and retrieval of data.

6.4 System for Uttar Pradesh Ground Water Investigation Organisation Developed by National Institute of Hydrology

A data storage and retireval system for ground water has been developed at National Institute of Hydrology, Roorkee for the data of Ground Water Investigation Organisation (GWIO), Uttar Pradesh Irrigation Department. In the present practice, the GWIO maintains their data district-wise. A district is further subdivided into blocks and a block into stations. Thus to specify a particular well location, the name of the well, station name, block name and district name have to be specified. To be consistent with the practice being followed by them, this data storage and retireval system has been developed using the rational data base model. Since the spellings of the district, block and station names differ sometimes and also for achieving saving in the computer storage space, two digit codes have been given to the district name, block name and station name. The structure which has been adopted is shown in Fig. 10.

```
Structure for database : a:district1.dbf
Number of data records :        319
Date of last update    : 04/27/83
Field  Field name  Type          Width   Dec
    1  DISTRICT    Character         2
    2  BLOCK       Character         2
    3  STATION     Character         2
    4  WELLNO      Character         6
    5  LATITUDE    Character         6
    6  LONGITUDE   Character         6
    7  RLEVEL      Numeric           7     3
    8  YEAR        Numeric           2
    9  WLJAN       Numeric           6     2
   10  WLFEB       Numeric           6     2
   11  WLMAR       Numeric           6     2
   12  WLAPR       Numeric           6     2
   13  WLMAY       Numeric           6     2
   14  WLJUN       Numeric           6     2
   15  WLJUL       Numeric           6     2
   16  WLAUG       Numeric           6     2
   17  WLSEP       Numeric           6     2
   18  WLOCT       Numeric           6     2
   19  WLNOV       Numeric           6     2
   20  WLDEC       Numeric           6     2
** Total **                        106
```

Fig. 10 : Structure of Database for GWIO Data

The package dBASE III plus has been used for the development of this data storage and retireval system. The addition of new data to the existing data can be achieved using the simple commands. A menu driven interface has also been developed for ease in operation of the system.

6.5 National Resources Data Management Systems (NRDMS)

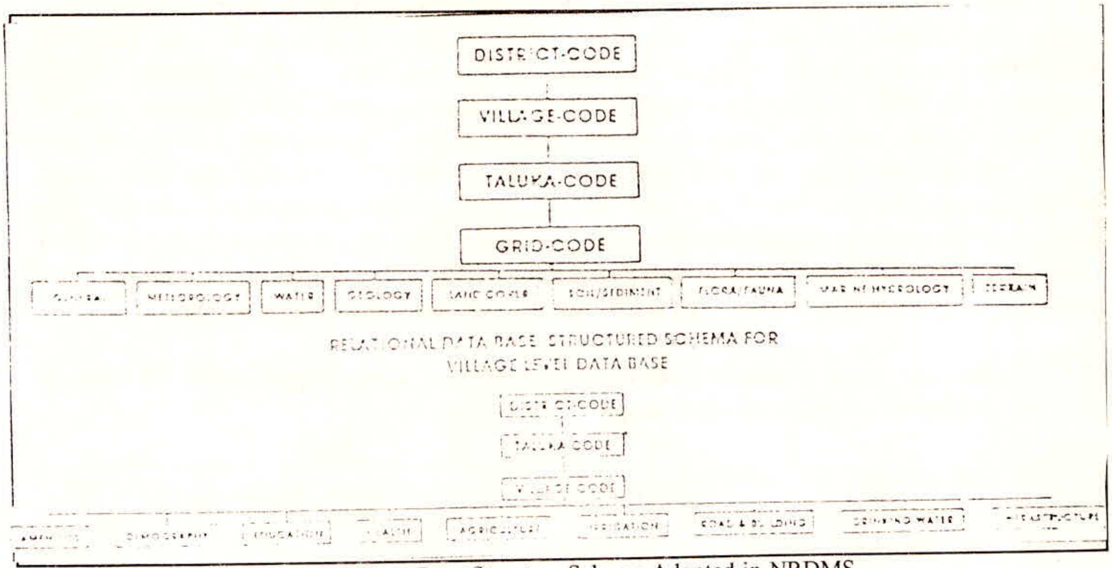

Fig. 11 : Data Base Structure Scheme Adopted in NRDMS

It has been pointed out at various forums that the planning by Govt. of India can be made more effective if it is decentralized. Accordingly, now-a-days more emphasis is being placed on district level planning. As an aid in this planning, a district-wise data base is being created under the project NRDMS of Deptt. of Science and Technology. The main purpose of NRDMS are to make an assessment of natural resources, to identify the information required for various category of users, to evolve methodologies for computer based storage, processing and retrieval of information and to use it for the purpose of planning and development. The NRDMS project is working as a multi-agency team. The approach adopted so far has been to assign various districts to various agencies. Water resources form an important part of the natural resources for which the data are being compiled and collected. The data being collected includes for each block/village, rainfall, number of dug/tube wells, depth of water table, water quality in the wells, number of tanks, rivers and streams in the block/village, canals and minors and their discharge. The other data which are used in hydrological analysis includes the cropping pattern and irrigated area, and temperature. The data base structure scheme for the natural resources data base is shown in Fig. 11 and the flow chart for the data management is shown in Fig. 12.

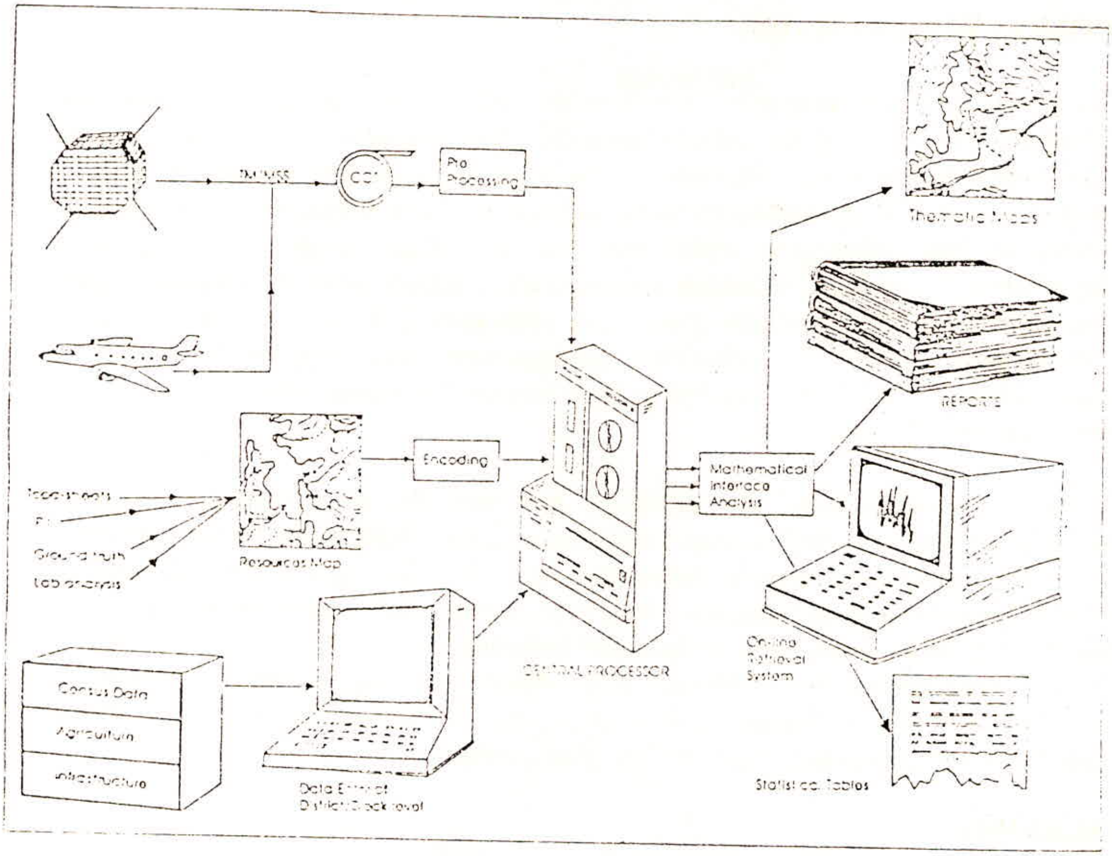

Fig. 12 : Flow Chart for Data Management System for NRDMS

7.0 INTEGRATED DATA BASE MANAGEMENT SYSTEM

The major disadvantage of the existing data base management system is that each organisation has developed its own system for the data being collected by that organisation. Consequently the same types of data (say rainfall) are being stored in different formats by different organisations. A user faces a great difficulty when he wants to use those data in some hydrological studies. In this regard, there is no co-ordination among the various organisations.

In order to avoid the problems in the existing data base management system, it is suggested that an integrated data base management system should be developed. This system should have the capability to read the data base files created for different types of data, without altering the contents of the files. Depending upon the user requirement, different types of data should be retrieved in the desired format for the hydrological studies. In this regard a National Data Bank linked with an integrated data base management system software is required to be developed. There should be a provision of free exchange of data among the various user agencies without any administrative hindrance. Hydrology project envisages the creation of National Data Bank as one of its major objectives.

8.0 CONCLUDING REMARKS

To achieve a desired degree of growth in the field of investigation and development of water resources for meeting the various demands, there is a need to improve the system of data collection, infrastructural facilities at the State, regional and national level. The present network is inadequate and needs to be augmented. That is why the establishment of the new sites for the hydrological observations has been taken up on priority under the Hydrology Project. To maintain standards and quality of measurements the personal engaged in the measurement need to be educated and made responsible. To achieve these objectives it would be necessary to train the technicians and observers. For an optimum utilisation of the new instruments, it would be required to have service and maintenance facilities for the new instruments and devices.

It may be however, noted that no matter how well data are collected, it cannot be made use of by the analyst unless a proper and effective data storage and retrieval system is developed. The information need to be exchanged among the various user agencies and through coordinated efforts, an integrated data base management system should be evolved at National level to store and retrieve all kinds of hydrology related data for the hydrological studies of any geographical area of the country. There is an urgent need for building up trained and well qualified manpower at different levels. It is expected that the Hydrology Project would provide necessary input for the above aspects.

BIBLIOGRAPHY

Central Ground Water Board. *Computerisation of Ground Water Data.* Ministry of Irrigation, Govt. of India, New Delhi.

National Institute of Hydrology, 1987. *Proposal for National Hydrology Project (Suface Water).*

National Institute of Hydrology, 1988. *Proposal for National Hydrology Project (Ground Water).*

National Institute of Hydrology, 1988. *Proposal for National Hydrology Project (Water Quality).*

Fleming,G., 1975. *Computer Simulation Techniques in Hydrology.* Elsevier Environmental Science Series.

Proc. of the Seminar on: *National Hydrological Data Bank.* 19-20 May 1988, Bangalore.

WMO, 1985. *Guidelines for Computerised Data Processing in Operational Hydrology and Land and Water Management.* Joint FAO/WMO publication No 634.

APPENDIX - I

ORGANISATIONS COLLECTING HYDROLOGY RELATED DATA IN INDIA

A. Hydrometeorological Data
- India Meteorological Department (IMD)
- State Irrigation and Water Resources Departments
- Central Water Commission (CWC)
- State Revenue Departments
- Universities, Technical, Academic Institutions and Research organisations.

B. Surface Water Data
- Central Water Commission
- State Irrigation and Water Resources Departments
- Some Universities, Technical, Academic Institutions and research organisations (limited to experimental watersheds)

C. Ground Water Data
- Central Ground Water Board (CGWB)
- State Ground Water Boards
- Some other organisations (limited to pump test data for their own pumping stations).

D. Water Quality Data
- Central Board for Prevention and Control of Water Pollution
- Central Water Commission
- State Pollution Control Boards
- State Irrigation and Water Resources Departments
- Central and State Ground Water Boards
- Universities and research organisation
- Ganga project directorate
- Non-Governmental agencies

E. Process Based Data
- Technical Universities and Institutions (limited to experimental plots)
- Research organisations (limited to experimental watersheds)

F. Maps and Atlas
- Survey of India
- Land Use & Soil Survey Departments
- State Irrigation and Water Resources Departments

G. Satellite Imageries and CCTS
- National Remote Sensing Agency

** *** **