HₒP *Technical Assistance*

HYDROLOGY PROJECT

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

VOLUME 8
DATA PROCESSING AND ANALYSIS

OPERATION MANUAL - PART IV

DATA MANAGEMENT

Table of Contents

1   INTRODUCTION

1.1   GENERAL

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

- Demand driven, i.e. output is tuned to the user needs

- Use of standardised equipment and adequate procedures for data collection and processing

- Computerised, comprehensive and easily accessible database

- Proper infrastructure to ensure sustainability.

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

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

A.   **Design Manual,** which provides information for the design activities to be carried out for the further development of the HIS

B.   **Reference Manual,** including references and additional information on certain topics dealt with in the Design Manual

C.   **Field/Operation Manual,** which is an instruction book describing in detail the activities to be carried out at various levels in the HIS, in the field and at the data processing and data storage centres.

The manual consists of ten volumes, covering:

1.   Hydrological Information System, its structure and data user needs assessment

2.   Sampling Principles

3.   Hydro-meteorology

4.   Hydrometry

5.   Sediment transport measurements

6.   Water Quality sampling

7.   Water Quality analysis

8.   Data processing, and

9.   Data transfer, storage and dissemination.

10. HIS – SW Protocols

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

Part I: Data entry and primary validation

Part II: Secondary validation

Part III: Final processing and analysis

Part IV: Data management

This Part IV deals with data management and describes the data processing plan and activities, the types and forms of data in HIS and the storage and transfer of data between the databases of the data processing centres. The procedures described in the manual have to be applied to ensure uniformity in data processing throughout the Project Area and to arrive at high quality data.

2   SURFACE WATER DATA PROCESSING ACTIVITIES UNDER HIS

2.1   IMPORTANT ASPECTS OF DATA PROCESSING UNDER HIS

The HIS processes, stores and disseminates groundwater as well as surface water data but, as data collection and processing of surface and groundwater is done separately, this module is concerned with surface water data only. There is however, considerable overlap in the principles of data management.

Surface water data entry and processing are carried out almost exclusively by computer. Processing of data is accomplished using dedicated hydrological data entry and processing software SWDES and HYMOS.

Processing of hydrological data is not a single step process. It is carried out in a series of stages, starting with preliminary checking in the field, through receipt of raw data at Sub-divisional offices and successively higher levels of validation, before it is accepted as fully validated data in the State or Regional data storage centre.

The progress of data from field to data storage is not a one-way process. It includes loops and feedbacks. The most important link is between the field station and the lowest processing level at Sub-divisional offices with frequent feedback from both ends but there will also be feedback from State and Divisional offices downward on the identification of faulty or suspect data through validation. Facilities for feedback from data users must also be maintained.

Processing and validation of hydrological data require an understanding of field practices. This includes the principles and methods of observation in the field and the hydrological variable being measured. It must never be considered as a purely statistical exercise. With knowledge of measurement techniques, typical errors can be identified. Similarly knowledge of the regime of a river will facilitate the identification of spurious data. For example for river level or flow, a long period at a constant level followed by an abrupt change to another period of static level would be identified as suspect data in a natural catchment but possible due to dam operation in a regulated river.

2.2   STAGES IN SURFACE WATER DATA PROCESSING

- Receipt of data

- Data entry to computer

- Data validation - primary, secondary and hydrological

- Data completion and correction

- Data compilation

- Data analysis

- Data reporting

- Data transfer

In this introduction an attempt will be made to generalise so that principles apply across all variables but with examples from particular variables.

2.3   RECEIPT OF DATA

Data progresses by stages through the processing system, from field to Sub-divisional office to Division and then hence to State or Regional Data Processing Centres. At each stage in the process target dates for receipt and for onward transmission are prescribed.

A record of receipt and date of receipt for each station record is maintained for each month of the year in suitably formatted registers. Receipt will be recorded on the day of delivery. Such registers will be maintained in each office through the system. These registers have two purposes:

- To provide a means of tracking misplaced data
- To identify the cause of delay beyond a target date whether late from the field or delay at a processing office and hence to follow up with corrective measures.

Data collected in the field are delivered first to a Sub-divisional office or District office in a variety of media, as hand-written forms and notebooks, charts or digital data files on magnetic media.

Arrangements must be made for the storage of raw paper records after entry to computer files. In the case of Sub-divisional offices this will be a temporary storage but permanent storage will be in Divisional offices. Manuscripts must be maintained temporarily but in a well organised manner at the Sub-Divisional Data Processing Centres for a period of three years. Storage of hard-copy data (forms and charts) must also be logical and structured. This is to ensure that the original field data remain accessible for further validation and checking. After a lapse of three years the manuscripts of all the data must be transferred to the respective Divisional Data Processing Centres for the purpose of archival.

Field and processed data on the magnetic media will be sent from Sub-Divisional to Divisional Data Processing Centre and thereafter from Divisional to State/Regional Data Processing Centre.

2.4   DATA ENTRY

The bulk of the surface water data is in the form of time series of hydrological and hydro-meteorological, water quality and quantity data. All such data are entered to computer at the lowest level in the data processing system, i.e. in the Sub-divisional data processing centres. This has the advantage that supervisory field staff share neighbouring offices with data processing staff and can easily be made aware of observer's mistakes or instrumental errors, and feedback given to the field personnel.

One exception to this practice is that data resulting from the analysis of samples in various water quality laboratories is entered at the laboratories themselves.

The Primary Module of dedicated hydrological data processing software, SWDES,, is available for accomplishing data entry from hand-written forms and tabulated autographic chart records. The software incorporates user-friendly data-entry screens which mimic the layout of all standard field forms to simplify direct entry from keyboard. Screens are available to select a station and data type from a list. Once selected, a screen with appropriate date and time labels is displayed against which data are entered.

The software provides automatic checking to reject unacceptable characters, e.g. alpha characters in a numeric field or duplicate decimal point. The validity of entered data is checked against pre-set limits and the program thus detects and rejects gross errors directly, e.g. resulting from misplaced decimal point. A facility is available for assessing the hydrological integrity of the data through the simultaneous display of a graphical plot of the entered data.

Data entry software also allows entry of static and semi-static data.

2.5   DATA VALIDATION

2.5.1   GENERAL

Data validation is the means by which data are checked to ensure that the final figure stored in the HIS is the best possible representation of the true value of the variable at the measurement site at a given time or in a given interval of time. Validation recognises that values observed or measured in the field are subject to errors and that undetected errors may also arise in data entry, in computation and, (hopefully infrequently) from the mistaken 'correction' of good data.

Validation is carried out for three reasons:

- to correct errors in the recorded data where this is possible

- to assess the reliability of a record even where it is not possible to correct errors

- to identify the source of errors and thus to ensure that such errors are not repeated in future.

Measurement errors may be classified as random or systematic or spurious in nature (Figure 2.1).

- Random errors are sometimes referred to as experimental errors and are equally distributed about the mean or 'true' value. The errors of individual readings may be large or small, e.g. the error in a staff gauge reading where the water surface is subject to wave action, but they tend to compensate with time or by taking a sufficient number of measurements.

- Systematic errors or bias is where there is a systematic difference, either positive or negative, between the measured value and the true value and the situation is not improved by increasing the number of observations. Examples are the use of the wrong raingauge measure or the effect on a water level reading of undetected slippage of a staff gauge. Hydrometric field measurements are often subject to a combination of random and systematic errors. Systematic errors are generally the more serious and are what the validation process is designed to detect and if possible to correct.

- Spurious errors are sometimes distinguished from random and systematic errors as due to some abnormal external factor. An example might be an evaporation pan record where animals have been drinking from the pan, or a current meter gauging result using a very bent spindle. Such errors may be readily recognised but cannot so easily be statistically analysed and the measurements must often be discarded.
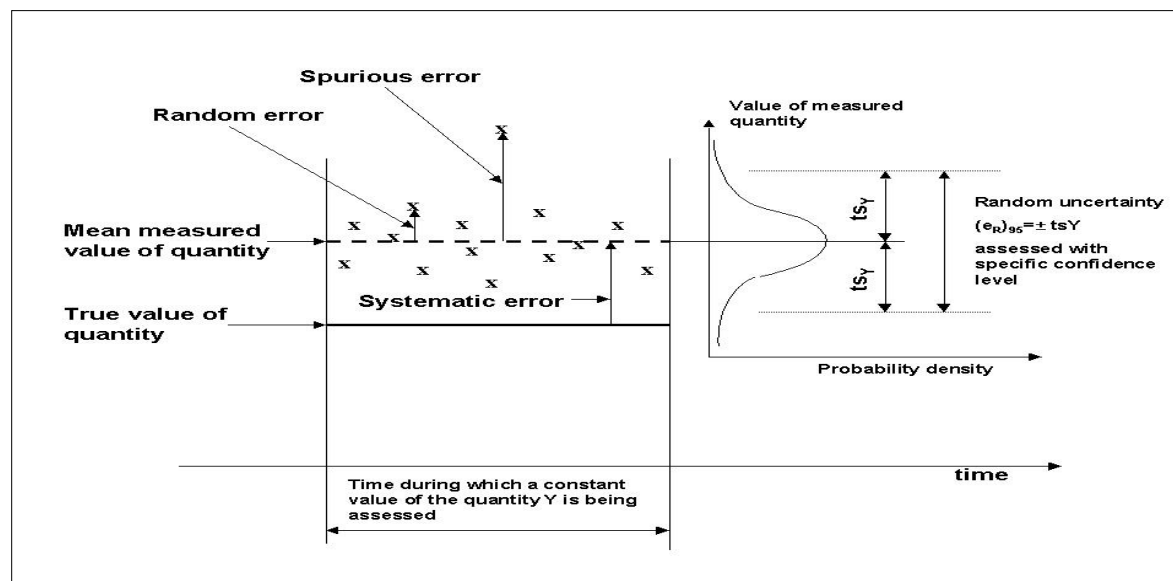
Figure 2.1:   Classification of measurement errors

Typical errors at observation stations are as follows:

- from faulty equipment, e.g. thermometer with air bubble in column

- from malfunction of instrument, e.g. slippage of float tape in level recorder (systematic but not constant)

- from improper instrument setting by observer, e.g. level recorder compared with staff gauge (systematic)

- from exposure conditions, e.g. stilling well blocked so that measured level in well differs from the river (systematic)

- personal observation errors, e.g. gauge misread or value interpolated away from site (random or spurious)

- transcription error in writing the observed reading

- error in field computation, e.g. current meter measurements.

Variables ultimately stored in the HIS may be directly measured (e.g. rainfall) or they may be derived using a relationship with one or more other variables (e.g. discharge). In the latter case the error in the derived value depends both on the field measurements and on the error in the relationship which is both random and systematic, the latter being particularly important if based on a relationship extrapolated beyond the limits of observation.

As a consequence of such errors it is important not only to ensure the use of good equipment and observational procedures but also to monitor the quality of all the data received. Validation procedures must be applied in a rigorous and standardised manner.

Validation involves a process of sequential and complementary comparisons of data and includes:

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

- for a single series between sequential observations to detect unacceptable rates of change and deviations from acceptable hydrological behaviour most readily identified graphically

- between two measurements of a variable at a single station, e.g. daily rainfall from a daily gauge and an accumulated total from a recording gauge

- between two or more measurements at neighbouring stations, e.g. flow at two points along a river

- between measurements between different but related variables, e.g. rainfall and river flow

Improvement in computing facilities now enables such validation to be carried out whereas in the past the volume of data the time required to carry out manual validation was prohibitive.

2.5.2   LEVELS OF VALIDATION

It is preferable to carry out the data validation as soon as the data is observed and as near to the observation station as possible. This ensures that information which may be essential to support the inferences of data validation is fresh in minds of the field staff and supervisors and that interaction between field and processing staff is possible. However, to provide full validation close to observation sites is impractical both in terms of computing equipment and staffing and a compromise must be reached which recognises both the wide geographical spread of observation stations and the staff and equipment available. The sequence of validation steps has therefore been divided so that those steps which primarily require interaction with the observation station are carried out in close proximity (i.e. at Sub-divisional office) whereas the more complex comparisons are carried out at higher levels.

Thus, data validation to be carried out is grouped into three major categories:

- Primary data validation,

- Secondary data validation and

- Hydrological data validation.

Since, every hydrological variable has distinct characteristics it is necessary that data validation techniques be explained for different variables separately for which reference is made to Volume 8, Operation Manual, Part 1 to 3. However, in this section general principles are provided which apply to variables in general.

2.5.3   PRIMARY DATA VALIDATION

Primary validation is carried out at Sub-divisional data processing centres immediately after keying-in or transferring the raw data.

Primary validation is primarily involved with comparisons within a single data series and is concerned with making comparisons between observations and pre-set limits and/or statistical range of a variable or with the expected hydrological behaviour of a hydrological phenomenon. However, information from a few nearby stations within a limited area may also sometimes be available and these may be used while carrying out primary validation for example with respect to daily rainfall data. SWDES provides options to facilitate the primary validation with as little effort and ambiguity as possible.

Primary data validation highlights those data which are not within the expected range or are not hydrologically consistent. These data are then revisited in the data sheets or analogue records to see if there was any error while making computations in the field or during keying-in the data. If it is found that the entered value(s) are different than the recorded ones then such entries are immediately corrected. Where such data values are found to have been correctly entered they are then flagged as doubtful with a remark against the value in the computer file indicating the reason of such a doubt.

Apart from data entry errors, suspect values are identified and flagged but not amended at the Sub-divisional level. However the flag and remarks provide a basis for further consideration of action at the time of secondary and final data validation.

2.5.4   SECONDARY DATA VALIDATION

Secondary data validation is done at Divisional offices after primary validation has been carried out.

Secondary validation consists of comparisons between the same variable at two or more stations and is essentially to test the data against the expected spatial behaviour of the system. Secondary validation is based on the spatial information available from a number of neighbouring observation stations within a comparatively large area. The assumption, while carrying out such comparison, is that the variable under consideration has adequate spatial correlation within the distances under consideration. Such correlation must be confirmed in advance on the basis of historical records and the experience thus gained in the form of various types of statistics is utilised while validating the data. Qualitative evaluation of this relationship is not very difficult to make. For certain hydrological variables like water levels and discharges, which bear a very high degree of dependence or correlation between adjoining stations, the interrelationship can be established with a comparatively higher level of confidence. However, for some variables which lack serial correlation and show great spatial variability (e.g. convectional rainfall), it is difficult to ascertain the behaviour with the desired level of confidence. In such circumstances, it becomes very difficult, if not impossible, to detect errors.

While validating the data on the basis of a group of surrounding stations, the strategy must always be to rely on certain key stations known to be of good quality. If all the observation stations are given the status of being equally reliable then data validation will become comparatively more difficult. This is not done merely to make the data validation faster but on the understanding based on field experience that the quality of data received from certain stations will normally be expected to be better than others. This may be due to physical conditions at the station, quality of instruments or reliability of staff etc. It must always be remembered that these key or reliable stations also can report incorrect data and they do not enjoy the status of being absolutely perfect.

As for the primary data validation, for the secondary data validation the guiding factor is also that none of the test procedures must be considered as absolutely objective on their own. They must always be taken as tools to screen out certain data values which can be considered as suspect. The validity of each of these suspect values is then to be confirmed on the basis of other tests and corroborative facts perhaps based on information received from the station. It is only when it is clear that a certain value is incorrect and an alternative value provides a more reliable indication of the true value of the variable that suitable correction should be applied and the value be flagged as corrected.

If it is not possible to confidently conclude that the suspected value is incorrect then such values will be left as they have been recorded with proper flag indicating them as doubtful. All those data which have been identified as suspicious at the level of primary validation are to be validated again on the basis of additional information available from a larger surrounding area. All such data which are supported by the additional spatial information must be accepted as correct and accordingly the flags indicating them as doubtful must be removed at this stage.

2.5.5   HYDROLOGICAL VALIDATION

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

If a record has been subjected to thorough field checking and primary and secondary validation, soon after the record has been obtained, then hydrological validation should reveal no more than is already known. However, for historical data to which no (or few) such checks have been applied, hydrological validation may become the principal check on the reliability of the record. Where data are to be used for design purposes, hydrological validation is essential. Otherwise hydrological validation may be selective both in terms of the stations and of the tests applied. Thorough hydrological validation requires a high level of professional expertise and can be very time consuming. In the end it may suggest that a particular record is unreliable for particular periods or ranges but it will not always provide the means of correcting a faulty record.

2.6   DATA IN-FILLING (COMPLETION) AND CORRECTION

Observed field data may have missing values or sequences of values due to equipment malfunction, observer absence, etc. these gaps should, where possible, be filled to make the series complete. In addition, all values flagged as doubtful in validation must be reviewed to decide whether they should be replaced by a corrected value or whether doubt remains as to reliability but a more reliable correction is not possible and the original value then remains with a flag.

In-filling or completion of a data series is done in a variety of ways depending on the length of the gap and the nature of the variable. The simplest case is where variables are observed with more than one instrument at the same site (e.g. daily raingauge and recording gauge), the data from one can be used to complete the other. For single value or short gaps in a series with high serial correlation, simple linear interpolation between known values may be acceptable or values filled with reference to the graphical plot of the series. Gaps in series with a high random component and little serial correlation such as rainfall cannot be filled in this way and must be completed with reference to neighbouring stations through spatial interpolation. Longer gaps will be filled through regression analysis or ultimately through rainfall runoff modelling. However, it must be emphasised here that various methods used for in-filling or correction will affect the statistics of the variable unless care is also taken with respect to its randomness. Nevertheless, it is not advisable to use completed or corrected data for the purpose of designing an observational network.

Data correction is to be done using similar procedures as used for completing the data series.

2.7   DATA COMPILATION

Compilation refers primarily to the transformation of data observed at a certain time interval to a different interval. e.g. hourly to daily, daily to monthly, monthly to yearly. This is done by a process of aggregation. Occasionally disaggregation, for example from daily to hourly is also required.

Compilation also refers to computation of areal averages, for example catchment rainfall. Both areal averaging and aggregation are required for validation, for example in rainfall runoff comparisons, but also provide a convenient means of summarising large data volumes.

Derived series can also be created, for example, maximum, minimum and mean in a time interval or a listing of peaks over a threshold, to which a variety of hydrological analyses may be applied.

2.8   DATA ANALYSIS

The HIS is not designed to provide a comprehensive range of hydrological analysis techniques. However, procedures used in data validation and reporting have a wider analytical use. The following are examples of available techniques:

- basic statistics (means standard deviations, etc.)

- statistical tests

- fitting of frequency distributions

- flow duration series

- regression analysis

- rainfall depth-area-duration

- rainfall intensity-frequency-duration

2.9   DATA REPORTING

Past practice has been to publish all available data for a state or river basin. With the larger amount of digital data to be stored in the HIS this is now, neither practical nor desirable as legitimate users can easily be provided with the precise data they need in the format they require.

What is now required is a readily available document indicating what information is available and held in the HIS. This should include the following:

- maps showing observation stations within their catchment and administrative contexts

- lists showing the stations and the period of record available

- summary description of salient facts associated with stations

- summary hydrological information for all stations, e.g. annual and monthly totals

- significant trends in the behavior of the hydrological variables or alarming situations which need immediate attention of planners and designers

- selective listings or graphs(e.g. daily values) to give examples of available formats.

Periodic publication of special reports showing long term statistics of stations or special reports on unusual events may also be prepared. In addition a catalogue of data held in various HIS databases is prepared periodically.

Otherwise specific data can be provided to users on the basis of need. This can be prepared in digital form on magnetic media thus avoiding the need to re-key the data to computer. A wide range of tabular and graphical formats is available, for example showing comparisons of current year values with long-term statistics, thematic maps of variables such as annual and seasonal rainfall, duration and frequency curves, etc. More detailed information such as stage discharge ratings can be provided to meet specific needs.

2.10   DATA TRANSFER

Since data are processed by stages at several locations, rapid and reliable transfer of data from one location to another is essential. The decision on the optimal methods of transport have to be taken on the basis of volume, frequency, speed of transmission and the cost.

Where analogue data must accompany digital data for example from the observation station through to Divisional Data Processing Centres, it is expected that equivalent digital data are prepared on diskettes and CD-ROMs to accompany these data using physical data carriers.

Where there is little accompanying analogue data for example from Division to State Data Processing Centre then electronic transmission may prove more effective. The final decision on the availability of

communication links will be forthcoming only after various data processing centres are functional since the communication technology is evolving at a very fast rate and new services are introduced in rapid succession.

3   SURFACE WATER DATA PROCESSING PLAN

3.1   INTRODUCTION

In Volume 1 the structure of the HIS has been described. In this Chapter, the HIS structure will be examined in more detail along with the underlying logic of operating in this way. Manpower and computing facilities needed to support operations at each level are described.

Processing activities in each state and central surface water agency will be accomplished at three levels. A layered system of this sort may be described as distributed in contrast to a centralised system where all the resources are concentrated in a single or a small number of large centres. Layers are as follows commencing from the lower end:

- At Sub-divisional Data Processing Centres, data are received from field stations and the bulk of data is entered and undergoes primary validation.
- Divisional Data Processing Centres receive data from several Sub-divisional data Processing Centres and secondary validation is carried out.
- State Data Processing Centres receive data from all Divisional Data Processing Centres within the state where final validation, completion and reporting of data is done.
- Regional Data Processing Centres with respect to the Central Water Commission provide the same facilities as the State DPCs, with 5 regions within the project area.

The project area and the network of observation stations to meet the needs of the area is very large as is the resulting volume of data. There is a danger that staff will be completely overwhelmed with the tide of data . To ensure that this does not occur, a workable plan has to be established for efficient management of data. Staff who operate the Hydrological Information System and particularly managers must understand who does what, where, and how the levels are linked.

3.2   DISTRIBUTED DATA PROCESSING – MERITS AND DEMERITS

Data processing activities under HIS will be accomplished in three stages at three levels. There are merits and demerits of this approach and these are discussed here.

Merits of distributed data processing

- logical distribution of huge amount of data processing work for the whole state or region in three stages and at several data processing centres ensure that adequate attention will be paid to all the aspects of processing resulting in the improved data quality. This is also, in general, commensurate with the present staff availability in different offices and leads to an optimal solution.

- data entry operations and subsequent primary validation at Sub-divisional data processing centres is carried out under direct supervision and by the staff who are supervising the observational activities regularly. The equipment and observational conditions together with the feed back from the field staff is fresh in the minds of the sub-divisional staff and will be highly beneficial in carrying out primary validation.

- inconsistencies in the observed data sets can be identified with a short gap of time (maximum of a month) after the observation is made by undertaking primary validation at the sub-divisional data processing centres.

- feed back to the staff at observational stations with respect to any inconsistency found in the data set can be given very quickly and efficiently and subsequent corrective measures can be initiated thereafter.

- staff involved in observation and supervision from the sub-divisional and divisional data processing centres feel associated with the data being produced by getting involved in the data entry, primary and secondary data validation process. This will ultimately help in improving the quality of data.

- the possibility of omissions and mistakes is reduced since in the multi-layered processing plan the data is validated at more than one place.

- an increased level of accountability in the system is ensured, as the activities of one data processing centre will be reviewed at higher levels.

Demerits of distributed data processing are as follows:

- since data processing activities are scattered in a very large areal extent it requires greater resource and effort to operate and maintain the necessary software and hardware.

- the effort required in co-ordinating and fine-tuning various activities at several data processing centres is far more than required for carrying out the entire processing at one place.

3.3   DATA PROCESSING ACTIVITIES AT VARIOUS LEVELS

For the State and Central agencies all the data processing activities before the archival of the fully processed data at the State/Regional Data Storage Centre would take place at Sub-Divisional, Divisional and State Data Processing Centres. The functions performed at these offices are given below.

The activities at the Sub-divisional Data Processing Centres include:

- receipt of field data in manuscript and/or digital form and maintaining a record of its receipt,

- entry of field and digitised data in computer files and carrying out primary data validation,

- feed back to the field stations in case of discrepancies found during checking and for delays in receiving the field data,

- transfer of data to the Divisional Data Centre and maintaining a record of the transfer, and

- archiving of original field registers (pertaining to current three years) with proper documentation.

The activities at the Divisional Data Processing Centres include:

- collection of data from sub-divisional data processing centres and maintaining a record of its receipt,

- additional automatic digitisation of analogue records from strip and drum charts,

- entry of additional field and digitised data in computer files,

- carrying out secondary data validation,

- feed back to the sub-divisional data processing centres if necessary as a follow up of validation exercise,

- transfer of data to the state/regional data processing centre and maintaining a record of the transfer,

- backing up the necessary data records and archiving on proper magnetic media, and

- final archiving all field records with proper documentation.

The activities at the State Data Processing Centres include:

- collection of digitised data from the Divisional Data Processing Centres,

- loading of data in the state or agency database within the dedicated hydrological surface water data processing software,

- transfer and retrieval of field data to/from the Data Storage Centre,

- validation, correction, processing and compilation of field data relating to the surface water component of the hydrological cycle, including, precipitation, evaporation, evapotranspiration (and the climatic variables required for their computation), streamflow, sediment transport and water quality parameters

- hydrological analysis as is required for the thorough validation of the data and for preparation of yearbooks, reports and documents.

- preparation of yearbooks, reports and documents in tabular and graphical format.

- transfer and retrieval of processed data to/from the Data Storage Centre.

- exchange of information from within the state between state and central organisations through the State Data Storage Centre.

3.4   COMPUTING FACILITIES AT VARIOUS LEVELS

The type and amount of data processing activity to be carried at various types of data processing centres require varying configuration of hardware and software.

HYMOS software will be available as the dedicated software for the processing of surface water data but, based upon the activities mentioned above, the software will be made available in three distinct modules for the three types of offices as follows:

- for Sub-Divisional offices   "Primary Module (i.e. SWDES)",

- for Divisional offices   "Secondary Module"

- for State/Regional offices   "Full package"

A complete checklist of the features available in each of these modules is given in Table 3.1. The important requirement is that the secondary package would include all the features available in the primary module and similarly the full package will include all the features of the secondary and primary modules.

The following hardware will be available at different types of data processing centres:

- at the sub-divisional data processing centres two personal computers together with an ink-jet and a laser printer (6 ppm) and a CD Writer.

- at the divisional data processing centres three personal computers along with an ink-jet and a laser printer (12 ppm) and a CD Writer .

- at the state data processing centres 4-6 personal computers with a network server together with laser printers, plotter (A3), digitiser (A0), scanner (A3), a CD Writer, suitable back-up and other essential peripherals.

Features	Sub-features	Primary Module	Secondary Module	Full Package
Data Entry and Editing	Space Oriented Data			
	• maps of basin features	No	No	Yes
	• basin descriptive data	No	No	Yes
	• hydraulic infrastructure	No	No	Yes
	Location Oriented Data			
	• Observation stations	No	No	Yes
	• Hydraulic structures	No	No	Yes
	Time Oriented Data			
	• Equidistant time series	Yes	Yes	Yes
	• Non-equidistant time series	Yes	Yes	Yes
	Relation Oriented Data			
	• Profile measurement data	Yes	Yes	Yes
	• Concurrent Observations	Yes	Yes	Yes
	• Relationship parameters	No	Yes	Yes
Primary Validation	• Listing of Data	Yes	Yes	Yes
	• Test on Extremes	Yes	Yes	Yes
	• Test on Timing Errors	Yes	Yes	Yes
	• Inspection of Temporal Variation	Yes	Yes	Yes
	• Inspection of Cross-sectional Variations	Yes	Yes	Yes
Secondary Validation	• Checks on Physical & Chemical Consistency	No	Yes	Yes
		No	Yes	Yes
	• Inspection of Longitudinal/Spatial Variation	No	Yes	Yes
		No	Yes	Yes
	• Test on Relations	No	Yes	Yes
	• Double Mass Analysis			
	• Nearest Neighbour Check			
Hydrological Validation	• Rainfall-runoff Simulation	No	No	Yes
Data Correction & Completion	• Time Shifting of Data	No	Yes	Yes
	• Interpolation	No	Yes	Yes
	• Regression	No	No	Yes
	• For ARG Data	No	No	Yes
Flow Measurements	• Discharge Computations	Yes	Yes	Yes
	• Fitting of Rating Curve	Yes	Yes	Yes
	• Shift Adjustment	No	Yes	Yes
	• Validation of Rating Curve	No	Yes	Yes
	• Extrapolation of Rating Curve	No	No	Yes
	• Stage-Discharge Transformation	No	No	Yes
	• Hydraulic Computations	No	No	Yes
Sediment Data	• Sediment Load Computations	No	No	Yes
	• Fitting of Sediment Rating	No	No	Yes
	• Processing of Reservoir Sediment Data	No	No	Yes
	• Validation of Sediment Rating	No	No	Yes
	• Extrapolation of Sediment Rating	No	No	Yes
	• Sediment Transport Computations	No	No	Yes
Data Compilation	• (Dis-) Aggregation of Series	No	Yes	Yes
	• Series Transformation	No	Yes	Yes
	• Creation of Derived Series	No	No	Yes
	• Computation of Areal rainfall	No	No	Yes
	• Computation of Evapotranspiration	No	No	Yes
Statistical Analysis	• Statistical Tests	No	Yes	Yes
	• Basic Statistics	Yes	Yes	Yes
	• Fitting Frequency Distributions	No	Yes	Yes
	• Correlogram Analysis	No	Yes	Yes
	• Spectral Analysis	No	Yes	Yes
	• Range and Run Analysis	No	Yes	Yes
	• Flow Duration Curves	No	No	Yes
	• Frequency Curves	No	No	Yes
	• DAD and IDF Curves	No	No	Yes
Data Reporting	• Standard outputs primary used for validation purposes	Yes	Yes	Yes
	• Customised tabular and graphical Outputs for preparing reports	No	No	Yes
Data Transfer	• Data transfer utilities	Yes	Yes	Yes

Table 3.1:   Availability of options in different types of packages.

3.5   AVAILABILITY OF STAFF AT VARIOUS LEVELS

The following staff will be available for carrying out various data processing activities at various data processing centres:

- at sub-divisional data processing centre

    − One data processing centre assistant for data entry and assistance job

    − One assistant hydrologist for carrying out primary data validation will be available.

    − The functioning of the sub-divisional data processing centre will be ensured by the sub-divisional data processing centre manager.

- at the divisional data processing centre

    − One data processing centre assistant for data entry and assistance job

    − One assistant hydrologist for carrying out secondary data validation will be available.

    − The functioning of divisional data processing centre will be ensured by the divisional data processing centre manager.

- at the state data processing centre

    − Two data processing centre assistants for data entry and assistance job

    − 4-6 hydrologists (as per the amount of work) for accomplishing organisation, final data validation, compilation and reporting activities will be available.

    − Support of water quality, database and information technology expert will also be available at the centre.

    − The overall functioning of the state data processing centre will be ensured by the state data processing centre manager.

3.6   TIME SCHEDULE FOR DATA PROCESSING AT VARIOUS LEVELS

Maintenance of strict time schedule for all the data processing activities at various data processing centres is of utmost importance. Since the data from different sub-divisional data processing centres will be used in conjunction at the divisional data processing centre and similarly the data from different divisional data processing centres will be simultaneously used at the state data processing centre it is all the more important that activities at all the data processing centres are carried out in time.

The time schedule for the completion of activities at various data processing centres is as given below:

- at the Sub-divisional Data Processing Centres
    − The data of any month from all the observational stations falling under its jurisdiction are required to be entered and primary validation completed by the 10th of the following month.

    − The field and processed data sets along with the primary validation report for each preceding month must leave for divisional data centre by the 10th of every month. That is to say that the data set of June must be finalised and dispatched from the sub-divisional data centres by 10th July.

    − To maintain such a schedule, it is appropriate that all the field data for the preceding month is received at the sub-divisional data processing centre by the 4th working day of every month. However to ensure that data processing work is distributed evenly over the whole month, data will be forwarded from the field three times per month in ten day periods. This will also ensure that the entry and primary validation activities will not be rushed through at the last moment.

- At the Divisional Data Processing Centres

    – The data of any month from all sub-divisional data processing centres under its jurisdiction must be available by the 15th on the following month. That is to say that the data sets of June must be available at the divisional data processing centres by the 15th of July.

    – The secondary data validation and all other activities required to be completed at the divisional data processing centres must be completed by the end of this month. The field and processed data sets along with the primary and secondary validation reports for each preceding month must leave the divisional data processing by the end date of every month

- At the State Data Processing Centre

    The data of any month from all the divisional data processing centres of the state must be available by the 5th of next to next month. That is to say that the data sets of June must reach the State Data Processing Centre by 5th August.

    – By 15th August the field data set must be transferred to the Data Storage Centre.

    – At the data processing centre all the required actions must be completed on the incremental data sets by the last date of the month in which the data has been received. That is to say that the processing of the observed data of June is completed at the state data processing centre by 31 August. These data will be held as a provisional processed data-set until the end of the water year when they will be forwarded to the State Storage Centre as a confirmed data-set for general dissemination to users.

- At the State Data Storage Centre

    – Both field and fully processed and validated data will be held at the State/Regional Storage Centre

    – Field data sets will be received one and a half months after the month in which the data have been collected.

    – Fully processed and validated data will be received at the end of the water year and made available for general dissemination. Only under exceptional circumstances will validated data from the State Storage Centre be retrieved for correction at the State Data Processing Centre. An example would be where gauging in an exceptional flood shows that previous extrapolation of a rating curve has been incorrect, thus requiring reprocessing of some extreme flood discharges.

- Inter-agency/inter-level meetings

There must also be at least two meetings every year in which different agencies who operate in the region discuss the consistency of the data sets between the agencies and finalise them. Such meetings must be concluded by the end of February and August each year for the data pertaining to June-December and Jan-May respectively. More preparatory meetings may be held by the concerned agencies as and when desirable to make the final meetings more effective. Before such finalisation the processed data sets must be considered as of provisional nature only. It must be emphasised here that there is an urgent need to formalise such inter agency interactions so that regular interaction is ensured between the agencies on a sustained basis.

With only the experience of manual data processing systems, the time schedule of processing from field to archive may seem very tight. In contrast, with experience of computerised data processing systems the schedule and timings are well within the limits of what is achievable.

4   TYPES AND FORMS OF DATA IN HIS

4.1   INTRODUCTION

A Hydrological Information System is a means of managing current and historical hydrological data and related information in an organised form.

As hydrologists we are accustomed to classify data in terms of hydrological variables - water level, rainfall etc. However data types may be grouped or divided in quite different ways to simplify computer processing and management. Hydrological variables can then be considered within this new framework.

Hydrological data in HIS can thus be classified three major categories

- space-oriented data: Space-oriented data comprise all the information related to physical characteristics of catchments, rivers, lakes and reservoirs. They also include the characteristics of observational stations and data series and various attributes associated with

- time oriented data: Time-oriented data comprise all the hydro-meteorological, quality and quantity data for which observations are periodically made in time at various observational stations. Time-oriented data can be equidistant, cyclic or non-equidistant in nature according to whether the observations are made at intervals which are equal, unequal but at defined intervals, or at unequal intervals. Most surface water data is equidistant or cyclic.

- relation oriented data: Relation-oriented data comprise information about the relationships established between two or more variables. Stage-discharge data and the calibration ratings of various instruments can also be considered under this category of relation-oriented data.

- An alternative way of classifying data in the Hydrological Information System is to categorise them as:

- static or semi-static data: These include most space-oriented data including catchment boundaries, topographic features, station location etc., which may be said to be static. Some features such as physical characteristics of rivers, lakes or reservoirs etc. do change in time but often very gradually and can be considered as semi-static in nature.

- time oriented data: On the other hand, all hydro-meteorological quality and quantity data including relationships between them can be considered as time oriented data as they change regularly in time. Time-oriented data are grouped as meteorological data, hydrological data, water quality data and sediment data.

A brief description of each type of data available in HIS is presented here.

4.2   SPACE-ORIENTED DATA

Space-oriented data comprise:

- Catchment data: physical and morphological characteristics

- River data: cross-sections, profile, bed characteristics

- Lake/reservoir data: elevation-area

- Station data: characteristics, history etc.

4.2.1   CATCHMENT DATA: PHYSICAL AND MORPHOLOGICAL CHARACTERISTICS

The physical characteristics of the catchment, of hydrological relevance, which can be stored in the HIS include its boundary, its geographical area, the layout of the river network and topographical features. River network characteristics in terms of number, length and area for different stream orders can be associated with any catchment. Such geographical information is generally available in the form of maps from which it may be digitised, or may be derived from the remotely sensed data. Each element of the space-oriented data is referenced by its position using a co-ordinate system referred to latitude and longitude. Such geographical data can be organised in different map layers so that it is possible to use one or more such layers at any time.

4.2.2   RIVER DATA: CROSS-SECTIONS, PROFILE, BED CHARACTERISTICS

River channel cross sections, longitudinal profiles and bed characteristics are needed for many hydrological applications. This type of data can be considered as semi-static and therefore must be obtained for each observation station at adequate intervals of time. River cross sections at gauging stations are of prime importance in interpreting stage-discharge data. Each cross section data set must therefore be associated with a period of applicability. Similarly, longitudinal profile data at each gauging station is to be associated with a time period. Bed characteristics in the vicinity of gauging stations including river course stability, river type (meandering or braiding), nature of the river bed and bank material, nature of mobile bed sediments and bed forms, presence of rapids up- or downstream from the station, state of vegetation, etc. must also be available.

4.2.3   LAKE/RESERVOIR DATA: ELEVATION-AREA

Such data based on topographic survey of a lake or reservoir might also be considered as physical or morphological data but, because of the structure of the data, are considered separately. Data are structured as a matrix of elevation and area for a lake or reservoir that is represented by an observational station. Such data are used in estimation of its capacity and evaporation corresponding to different levels. Each such matrix again has to be associated with the period of applicability.

4.2.4   STATION DATA: CHARACTERISTICS, HISTORY ETC.

A wide variety of information can be associated with each gauging station. This includes names and codes which are used for identification, attributes which are used to categorise and descriptive information which is used for station management or for assessing the quality of data processing.

Every observational station is given a unique station code. Other station characteristics include:

- station name,
- type of station,
- administrative units (state, district and tehsil/taluka),
- drainage boundaries (major river basin/zone, tributary to independent river and local river/basin),
- location (latitude and longitude),
- altitude,
- catchment area (for river gauging stations),
- reference toposheet number (w.r.t. 1:50,000 scale maps of Survey of India),
- agency in charge of the station, controlling offices (regional/state, circle, division, sub-division and section office)
- photographs, with dates, pertaining to the stations and various equipment installed at the stations. These can be scanned to a digital data file.

A record of the historical background of the station must also be maintained. Such a record must include the details of setting up of the station and of any major activity or change in its location, equipment installation or observational procedures. Special mention is made of records of the reduced level of the gauge zero for which a historical record of changes must be maintained.

4.3   TIME-ORIENTED DATA

4.3.1   CATEGORIES OF TIME SERIES DATA

Time series data include all those measurements which have an associated observation time, whether the measurement is of an instantaneous value (e.g. water level), an accumulative value (e.g. daily rainfall), a constant value (e.g. a gate overflow level setting), an averaged value (e.g. mean monthly discharge) or a statistic from a specified time period (e.g. daily maximum temperature, annual maximum flow). The distinction between instantaneous and accumulative values is of importance in determining whether in aggregation to another time interval, they should be averaged (e.g. flow) or accumulated (e.g. rainfall)

A further important classification of time series data is as follows:

* Equidistant time series are all those measurements which are made at regular intervals of time (hourly, 3-hourly, daily) whether by manual observation, by abstraction of values at regular intervals from a chart, or digitally. Values may be instantaneous, accumulative or averaged. In such a series the associated time of measurement can be generated automatically by the computer and does not require entry with each entered value of the variable.

* Cyclic time series are those measurements which are made at irregular intervals of time but where the irregular time sequence is repeated from day to day. An example is the observation of temperature measured twice daily at 0830 and 1730 hrs. Again, once the times of observation are specified, the associated time of measurement does not require entry with each entered value of the variable.

* Non-equidistant time series are those measurements which are made at irregular intervals of time in which the time is not pre-specified. Such values are generally instantaneous. An example is the recording of tipping bucket rainfall where each tip of the bucket recording the occurrence of a depth (e.g. 0.5 mm) has an associated registered time. Another example is the irregular manual dipping of well level. Constant series may also be non-equidistant. Such series require entry of the time of occurrence with each entered value of the variable. In the case of digitally recorded data (e.g. tipping bucket rainfall) this entry is made by transfer from a logger file.

* Note - It is possible to convert a non-equidistant time series to an equidistant time series. This is often done for example with rainfall measured by tipping bucket, where the transformation may either be carried out by the logger or in HYMOS.

Hydrological variables may then be considered in their more conventional groupings of meteorological, hydrological and water quality. Within these groups the treatment of individual variables is outlined with reference to the above classification.

4.3.2   METEOROLOGICAL DATA

Meteorological information has a variety of purposes but, with respect to hydrology, the meteorological network is primarily concerned with storage and transfer of water between the land and atmospheric portions of the hydrological cycle. Meteorological data that are strongly related to the hydrological processes are thus required in HIS. Precipitation being the primary source of fresh water is essential for most hydrological studies. The other major variable is evapo-transpiration for which data are needed for hydrological budgeting. As it is difficult to measure evapo-transpiration directly, indirect methods are generally used to estimate it. Such indirect estimation requires measurement of a range of meteorological variables which control the evaporative processes. Alternatively estimates may be made from directly measured evaporation in an evaporation pan

Meteorological and climatic variables required in the HIS are categorised and described below:

- precipitation

- pan evaporation

- temperature

- air pressure

- atmospheric humidity

- wind speed and direction

- sunshine

- derived meteorological variables

Precipitation data

Observations on point precipitation are made regularly at a number of observational stations suitably distributed in space. Precipitation experienced throughout the project area is in the form of rainfall, and raingauges are installed for its observation. Presently there are three types of raingauges in use. They are:

- Standard or ordinary raingauges (SRG) which are read manually. The standard raingauge data is recorded once or twice a day at fixed hours at 0830 and 1730 hrs. (some agencies like CWC record at 0800 hrs and 2000 hrs.). Daily observations are taken in the morning hours. Once-daily readings are thus equidistant and accumulative; twice-daily readings are cyclic and accumulative.

- Autographic or self-recording raingauges (ARG). Data from autographic raingauges is available in the form of a continuous plot from which the data are read manually and tabulated at hourly intervals corresponding to standard timings of the daily rainfall (again equidistant).

- Automatic raingauges having data logger. Data from automatic raingauges is available in digital form directly through its data logger and is recorded either as rainfall at fixed interval (usually half an hour) (equidistant) or as timings for each event of rainfall of fixed quantity (usually 1 mm or 0.5 mm) (non-equidistant).

Pan evaporation data

Evaporation records for pans are frequently used to estimate evaporation form lakes and reservoirs and evapo-transpiration from an area. US Class A pans are employed in India for measuring pan evaporation. Evaporation pan readings are taken once (equidistant and accumulative) or twice a day (cyclic) at standard times at 0830 and 1730 hrs. Daily observations are taken in the morning hours.

Temperature

Temperature is primarily of interest to hydrology as a controlling variable in the evaporative processes. The temperature of air, of the soil and of water bodies is all of interest. Periodic observations of air temperature are made using thermometers whereas continuous record is obtained using thermographs. Four types of thermometers, dry bulb, wet bulb, maximum and minimum thermometers are in use. Dry-bulb thermometer is used to measure temperature of the surrounding air. Wet-bulb thermometer is used to measure the temperature of the saturated air for determining relative humidity and dew point of the surrounding air. Maximum and minimum thermometers indicate the maximum and minimum temperatures of the surrounding air achieved since the last measurement. Observations for these four temperatures are made once or twice a day at standard times at 0830 and 1730 hrs (equidistant and instantaneous for dry and wet bulb and their derivatives; equidistant and statistic for maximum and minimum temperatures). Data from the thermograph is tabulated at hourly intervals (equidistant and instantaneous).

Water temperature of a water body may be considered a water quality parameter, influencing the rate of chemical and biological activity in the water. Water temperature measured at an evaporation pan is a factor in determining evaporation.

Atmospheric Pressure

Atmospheric pressure influences the rate of evapo-transpiration and is a useful though not a critical variable in its estimation. At any point it is the weight of the air column that lies vertically above the unit area. It is usually observed using a mercury barometer for instantaneous data and can also be continuously recorded using a barograph. The observations from barometer are made daily (equidistant and instantaneous) or twice daily (cyclic) at standard times at 0830 and 1730 hrs. The thermograph record is tabulated at hourly intervals corresponding to the standard timing of the daily observations (equidistant).

Humidity

Atmospheric humidity is a most important influence on evapotranspiration. Direct measurement of relative humidity is accomplished using a hygrograph. Indirect estimation of relative humidity is accomplished using dry and wet bulb temperature records. The estimated humidity from the dry and wet bulb temperatures are obtained daily (equidistant and instantaneous) or twice daily at standard times at 0830 and 1730 hrs. The continuous record obtained from the hygrograph is tabulated at hourly intervals.

Wind speed and direction

Wind speed and direction is of importance while estimating the evapo-transpiration. Wind speed and wind direction is measured using anemometer and wind vane respectively. Observations are made daily or twice daily at standard times at 0830 and 1730 hrs. Wind speed measurements may be instantaneous or, if wind run over a time interval is observed, then it is accumulative. Wind direction may influence measured evaporation totals if the surrounding environment in terms of wetness differs in different directions It is not normally used in calculations but if required, it can be recorded on a 16-point scale.

Sunshine duration

The duration of sunshine is a useful variable in estimation of evapotranspiration. The instrument commonly employed for observation of sunshine duration is the Campbell Stokes sunshine recorder. The lengths of burnt traces on the sunshine card indicate the sunshine duration. Sunshine duration data is tabulated for each hourly duration in the day and is considered as of equidistant and accumulative nature.

Derived meteorological quantities

Several meteorological quantities are only estimated from other directly measured meteorological variables. Such quantities are relative humidity, dew point temperature, estimated lake or reservoir evaporation and evapo-transpiration. Some are manually-derived before entry to computer whilst others (e.g. evapotranspiration) are computed using HIS software from other variables previously entered.

4.3.3   HYDROLOGICAL DATA

Quantitative records of time series of level (stage) and flow in surface water bodies constitute the bulk of hydrological data. Direct measurement of water level data is made using a variety of equipment.

Measurement of flows is comparatively difficult to accomplish and is therefore estimated indirectly by using the stage-discharge relationship. However, intermittent measurements of flow are made for the establishment of the stage-discharge relationship (held as relation-oriented data). A brief description of time series hydrological data in the HIS is given below.

Water level data

Water level is observed in three ways:

- from staff gauges

- from autographic water level (chart) recorders

- from digital type water level recorders.

From the staff gauges the observations are generally made once in a day in lean season (equidistant and instantaneous) and at multiple times a day (equidistant or cyclic and instantaneous) during flood. For a flashy river staff gauges may even be read at hourly intervals during the flood season. However, standard timings are generally followed for the observations during the day by different agencies.

An autographic (chart) type of water level recorder on the other hand gives a continuous record of water level in time from which levels must be extracted manually, usually at an hourly interval, before entry to the computer.

Digital data from a digital water level recorder can either be at equal intervals of time, usually at quarter or half hour interval, or can be reported only for those instants when there is a change in water level which is more than a pre-set amount (non-equidistant and instantaneous). Such a non-equidistant record can be converted to an equidistant record by interpolation.

Derived hydrological quantities

As mentioned above the direct observation of discharge is made only for the purpose of establishing the stage-discharge relationship. Once the relationship is established, the same can be used to transform observed stages into a derived time series of discharge.

4.3.4   WATER QUALITY DATA

Water quality parameters can rarely be measured continuously and are therefore based on sampling which may be at regular intervals (equidistant) or more often at irregular intervals (non-equidistant) and the recorded values represent the state of the water at a particular time (instantaneous). More important water quality parameters or those which change more rapidly with time are measured with greater frequency than those which are conservative in character.

Water quality analysis is carried out for certain field parameters at the observation station and are forwarded to Sub-division. Many analyses require more sophisticated equipment and are therefore determined at a divisional laboratory where they are entered.

Water quality parameters take many forms and there is a wide range of measurement techniques and units used. The following groups may be distinguished:

- organic matter

- dissolved oxygen

- major and minor ions

- toxic metals and organic compounds

- nutrients

- biological properties

Organic matter

This is important for the health of a water body because the decomposition of organic matter draws upon the oxygen resources in the water and may render it unsuitable for aquatic life. Common parameters to characterise it are: BOD (biochemical oxygen demand), COD (chemical oxygen demand) and volatile solids. The widely-used BOD test measures the oxygen equivalence of organic matter and is the most important parameter in assessing pollution by organic matter.

Dissolved oxygen

Similarly direct measurement of dissolved oxygen is an important indication of the health of the water. Absence of DO or a low level indicates pollution by organic matter. It is recorded as a percentage of saturation.

Major and minor ions

Natural waters contain a variety of salts in solution originating from rain or soil and rock with which they have been in contact. Analysis is most frequently carried out and recorded for major cations, calcium (Ca^{++}), magnesium (Mg^{++}), sodium (N^{+}) and potassium (K^{+}) and associated anions include sulphate ($SO_4^{=}$), bicarbonate (HCO_3^{-}) and chloride (Cl^{-}) and reported as concentration   (mg/l). Aggregate salts are derived as total dissolved solids (TDS) (mg/l).

Toxic metals and organic compounds

Toxic metals and other elements may exist naturally in waters but may be seriously increased through man's activity, including copper (Cu), chromium (Cr), mercury (Hg), lead (Pb), Nickel (Ni), Cadmium (Cd) and arsenic (As). Toxic organic substances will also be sampled and recorded but infrequently and at few sites as they require advanced instrumentation.

Nutrients

Nutrient elements such as P and K may be enhanced by agricultural and industrial activity. An excess may lead to deteriorating water quality and regular sampling is carried out.

Biological properties

Surface water polluted by domestic wastewater may contain a variety of pathogenic organisms including viruses, bacteria, protozoa and helminths. The cost of testing for all these organisms is prohibitive and the most common test is for *Escherichia coli* whose presence is an indicator of the potential for other pathogenic organisms. Results of testing for *E coli* are recorded as the most probable number (MPN)/100 ml)

4.4   RELATION ORIENTED DATA

Any kind of relationship between two or more variables is classified as relation oriented data. The relationship can be of any mathematical form which is the result of regression or a calibration exercise. The variables themselves may form a time series but it is their relationship rather than their occurrence in time which is the principal focus of the data and their storage and management in the HIS.

Relation-oriented data include the following:

- General

- Flow Data
    - Stage discharge data
    - Stage discharge rating parameters
- Sediment data
    - Discharge sediment concentration/load data
    - Discharge concentration/load rating parameters

4.4.1   GENERAL

A mathematical relationship between two or more hydrological variables can be established for the purpose of validation and in-filling of the data for another period. Relationships between stages at two adjacent gauging stations, between the average rainfall in a catchment and the resulting flow are typical examples. In some instances relationships may be established between water quality parameters and discharge to determine chemical loads. The parameters of the relationship along with the ranges of independent variables, error statistics and the period of applicability are required to be available for any relationship established.

4.4.2   FLOW DATA

Stage discharge data

Stage discharge data are the most common example of this type of data, although in this case special provision must be made, for the inclusion of ancillary gauging station information and for various types of correction. Stage-discharge observations are the primary data for establishing the relationship between the stage and discharge, called the rating curve, at any river gauging station.

Individual current meter measurement results in the computation of paired values of stage and discharge and ancillary information for its interpretation. Data from many gaugings results in a series of such paired data - the summary stage discharge data. The HYMOS software provides a means of deriving rating relationships based on these paired values and ancillary information.

Stage discharge rating parameters

Analysis of stage discharge observations results in the production of parametric relationships. For simple relationships this may be expressed in parabolic or in power form. For a given station and time, more than one equation may be required to characterise the relationship and the relationships may change with time. There is thus a need for a means of storing coefficients of equations, range of level applicability and period of validity

4.4.3   SEDIMENT DATA

Movement of solids transported in any way by a flowing liquid is termed as the sediment transport. From the aspect of transport it is the sum of the suspended load transported and the bed load transported. From the aspect of origin it is the sum of the bed material load and the wash load. As the discharges are related with the corresponding stages, discharge-sediment data also fit the classification of relation-oriented data. However, the relationship exhibited is often not so strong as that between stages and discharges for obvious reasons.

Sediment may be divided into two types of material:

• bed material forming the bed or transported as bed load and as suspended load

• wash load material, only transported as suspended load

The traditional classification as per ISO (ISO 4363: 1993) is shown in Figure 4.1 as below:
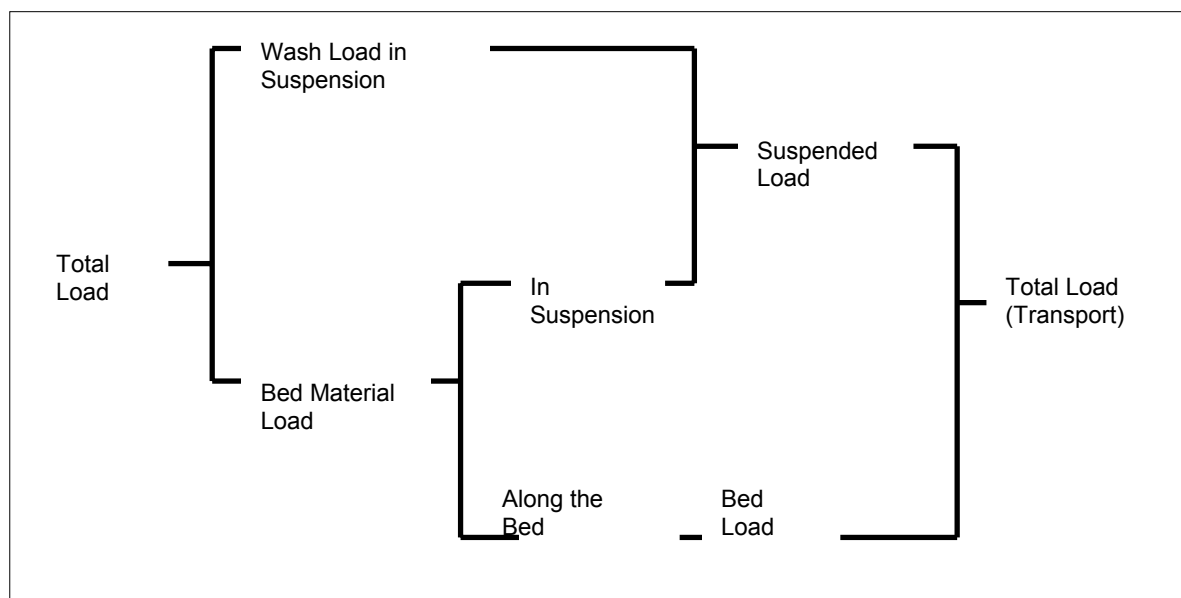
Figure 4.1:   Definition of sediment load and transport (ISO 4363: 1993)

In India suspended load and bed material are being sampled till recently. Bed load measurements are likely to be started at some stations. Near bed load transport measurements might become needed in future in relation with special problems.

The information on sediment includes the sediment size and the sediment concentration for the transported load: bed load and suspended load. The relationships which might be established between discharges and various sediment concentrations would also be available in the databases.

Flow/ sediment concentration load summary

When stable flow discharge - sediment concentration relationships exist, observations of flow and sediment in a station in a given period may be stored as relation oriented data and result in the production of a parametric relationship. This may be the case for stations with predominantly wash load transported in suspension.

Flow sediment relationships - a word of warning

Sediment originates from various sources, including river basin soil erosion, mass wasting or river bed and bank erosion. The sediment may be temporarily stored and mobilised again, depending on its source and on the flood events. Consequently sediment transport rates will depend on many factors, and may differ from the sediment transport capacity because of sediment availability, and sediment concentration is not necessarily closely linked to the flow discharge. In many rivers, unstable relationships between flow discharge and sediment concentration display loop effects, the intensity of which may vary from one flood to another.

Flow sediment rating parameters

Where parametric relationships are possible, they may be expressed by an exponential, parabolic or power form. As for the stage discharge rating (Section 4.2), more than one equation may be required to characterise the relationship; the breaking points may or may not be the same. The coefficients need to be stored with their domain of applicability (range of stage) and period of validity.

Sediment size data

Distinction has to be made between information on bed material and on transported sediment load. Bed material data may be considered as space-oriented data only when the river bottom is bedrock. River bed sediment sizes may change during the flood events by selective erosion and have to be considered as time-oriented. In the absence of direct bed load transport measurements, size distribution of bed material is required for their calculation. The changes with time of bed sediment sizes must be available. Data on size distribution must also be available for sediment transported as bed load.

For sediment transported as suspended load, data on sediment size distribution are not available because of the small sizes of the samples. Only the percentage of that load is given for three fractions:

- the coarse fraction (particles above 0.2 mm diameter)

- the medium fraction (particles between 0.075 and 0.2 mm diameter)

- the fine fraction (particles below 0.075 mm diameter)

5   RECEIPT OF DATA AT DIFFERENT LEVELS

5.1   GENERAL

Data are transferred by stages through the processing system, from field to Sub-divisional office to Division and hence to State or Regional Data Processing Centres. In order that routine processing proceeds smoothly the following features of data administration are essential:

- targets
- auditing
- feedback
- filing

At each stage in the process target dates for receipt and for onward transmission are prescribed.

A record of receipt and date of receipt for each station record is maintained for each month of the year in suitably formatted registers.

Where records are not received by the target date at each office or are received in an incomplete or unsatisfactory form, follow-up action is taken with the next lowest level to identify the source of the problem.

Data collected in the field are delivered to a Sub-divisional office in a variety of media, as hand-written forms and notebooks, charts or digital data files on magnetic media. Arrangements must be made for the storage of raw paper records after entry to computer files.

From the Sub-divisional office field and processed data on magnetic media will be sent to Divisional data processing centres with the accompanying primary validation report. Wherever considered appropriate for conveying comments, listing of data with data validation remarks have to be sent along with the data on magnetic media. Similarly, from Divisional to State/Regional data processing centre data will be sent on magnetic media along with comments/report on secondary data validation.

5.2   TARGETS

As already noted in Module 3 target dates are prescribed for the receipt and despatch of field and processed data from each office. These are shown in Table 1 where the bracketed figure refers to the month following the month in which the data have been observed or measured. Thus data measured in June must be received in the Sub-divisional office by 4 July and processing completed at the State/Regional Data Processing Centre by 31 August.

Stage	Receipt	Entry/ Validation, etc.	Dispatch
Subdivision	4 (1)	10 (1)	10 (1)
Division	15 (1)	31 (1)	31 (1)
Data Processing Centre	5 (2)	31 (2)	Field data: 15 (2) Processed data: End of water year
Data Storage Centre	31 (2)	-	As and when requested and available

Table 5.1:   Target dates for data processing and archiving

5.3   AUDITING OF RECEIPT

It is important to maintain a record of receipt of data so that problems of data collection or transmission can be identified at an early stage. Receipt is recorded on the day of delivery. Such records are maintained in each office through the system and an example for stage data is shown in Fig. 1. These records of receipt have two purposes:

- To provide a means of tracking misplaced data

- To identify the cause of delay beyond a target date whether late from the field or delay at a processing office and hence to follow up with corrective measures.

Failure to audit receipt may result in a much longer period of lost data. It is recommended that a date is set for follow up action should data not be received, say a further 5 days after the target receipt from the field station.

5.4   FEEDBACK

There is a requirement for action to be taken when data fail to arrive by a target date. A few additional days (no more than 5) should be given to allow for postal delays or transport problems. The station or office must then be contacted by telephone or post and an explanation requested for the delay. The date of feedback contact should also be noted on the record of receipt, e.g., Figure 5.1.

5.5   FILING

Whilst the purpose of computer processing and archiving hydrological data is to store field records and derived data in digital form, the original paper records must be filed in a way that they are readily retrieved. Permanent storage of hard-copy records will be in Divisional offices. Facilities for storage of hard-copy forms and charts must also be logical and structured and it is recommended that all time series data are stored by station rather than by year as it is most often in this format that the data are subsequently required.

HYDROLOGICAL INFORMATION SYSTEM, W R DEPTT., MAHARASTRA

RECORD OF RECEIPT OF WATER LEVEL DATA

Name of Data Processing Centre: …………………………………………...

Name of Basin/Sub-basin: ………………… **Year: ………………..**

STATION	Time interval of data	Staff/ Chart/ DWLR	Date of Receipt of Data for the Observation made during the Month of											
			Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Station 1	Thrice daily	Staff	3/2	2/3	2/4	1/5	7/6	10/7	1/8	2/9	2/10	6/11	3/12	1/1/81
	Hourly	Chart	N.A.	N.A.	N.A.	N.A.	N.A.	10/7	1/8	10/9	2/10	6/11	3/12	N.A.
Station 2	Thrice daily	Staff	N.A.	N.A.	N.A.	N.A.	N.A.	2/7	5/8	2/9	3/10	4/11	2/12	N.A.
Station 3	Thrice daily	Staff	N.A.	N.A.	N.A.	N.A.	N.A.	5/8	5/8	3/9	4/10	1/11	5/12	N.A.
Station 4	Daily	Staff												
	Non-equal	DWLR												
etc.														

Remarks..

....

Figure 5.1: *Example of record of receipt form for water level data*

6   ORGANISATION OF DATA INTO TEMPORARY DATABASES

6.1   SEPARATION BETWEEN DATA PROCESSING AND DATA STORAGE FUNCTIONS

Under HIS the data processing and the data storage functions are separated; data processing is carried out in various Data Processing Centres in a distributed manner at three distinct levels, whereas the data archival is accomplished in the Data Storage Centres. This separation between the data processing and the data storage functions is illustrated in Figure 6.1.
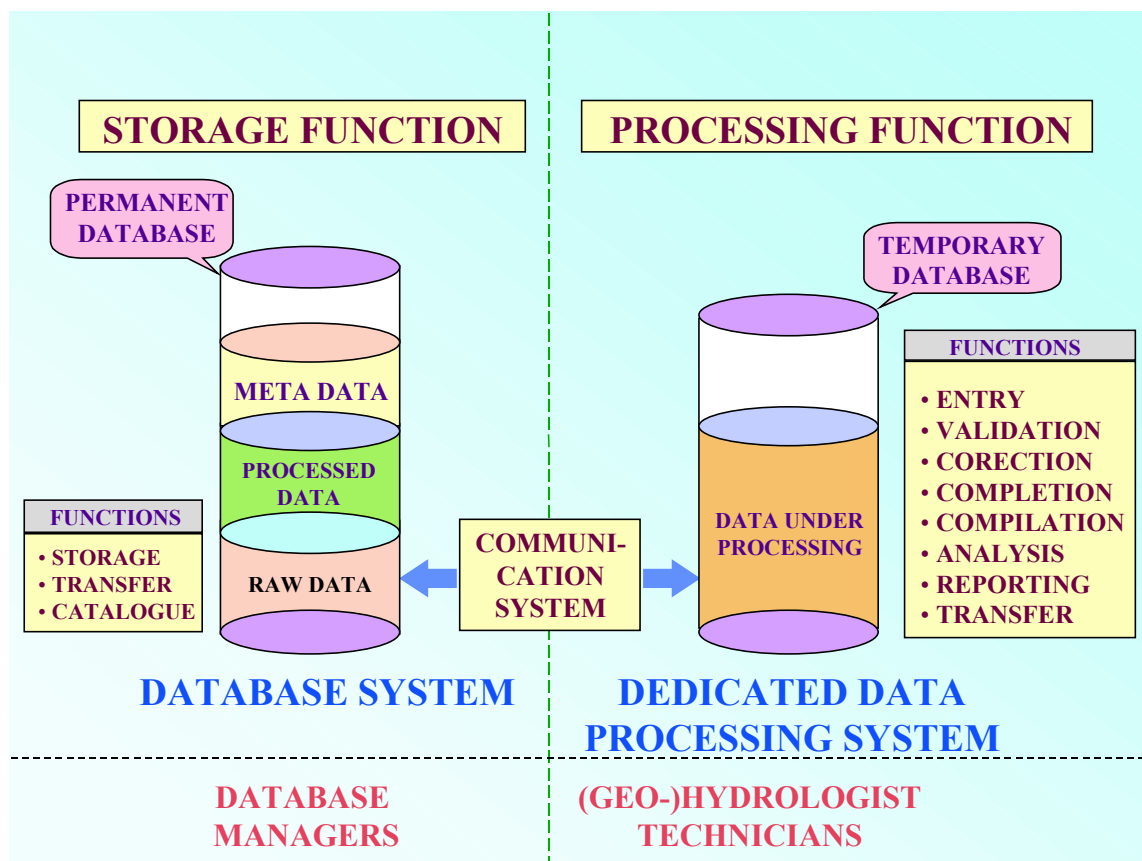

Figure 6.1:   Illustration of clear separation between data processing and data storage functions

As can be seen in the illustration above, functions like data entry, validation, correction, completion, compilation, analysis and reporting are required to be undertaken under data processing activities whereas functions like storage, retrieval and cataloguing are required for data storage activities. Data in the databases under data processing systems are processed hydrologically for the purpose of validation, completion, correction and compilation and hence may get revised. On the other hand, data in the data storage systems is not required to be revised at all. The functions needed in the data processing system are the ones needed for a comprehensive hydrological data processing whereas those needed in the data storage system are purely for an efficient storage and management of data. It is also to be realised that the personnel required for the data processing activities are those knowing the hydrological data processing concepts whereas those required for the data storage function are those who are specialists on data management and information technology aspects.

Such a distinct separation will be very helpful in making an unambiguous hydrological data archival which is essential for its sustainability and would also incidentally provide an adequate scope for any upgradation which might be required for either of the two systems independently.

6.2   WHAT ARE TEMPORARY, PERMANENT AND TRANSFER DATABASES?

As described above, the data processing centres are expected to provide excellent hydrological data processing opportunities whereas data storage centres are to function as effective electronic hydrological data libraries. Since the data under data processing centres are under processing and are liable for a change the character of data sets held are called as temporary and hence the name *temporary databases*. On the other hand, the data sets in the data storage centres are not to be modified at all and hence the name *permanent databases*.

Thus all the databases which are initiated by and for the dedicated hydrological data processing systems (SWDES and HYMOS) at the three levels; Sub-Divisional, Divisional and State Data Processing Centres, would be called as temporary databases. Similarly, the databases in which all the data is archived at the data storage centres are called as permanent databases.

All other databases which are used for transferring/exchanging data between two data centres would be termed as *transfer databases*. These transfer databases may be required for transfer/exchange of data from a Sub-Divisional Data Processing Centre to a Divisional Data Processing Centre or from a Divisional Data Processing Centre to State/Regional Data Processing Centre or from one Data Storage Centre to another Data Storage Centre. This is as illustrated in Figure 6.2 below.

6.3   WHAT ARE FIELD AND PROCESSED DATABASES?

The data as observed in the field by the observers is entered in the field note books and the data entry forms. The same data is keyed-in in the computer using SWDES. During the data entry a few mistakes may occur and the data entered and available on computer may not be exactly same as in the field forms. Upon detection of such mistakes, by employing simple data entry checks and graphical viewing, these errors are corrected such that the data in the computer database is the same as is available in the field forms. However, sometimes the data available on the field form itself bear some error that is very clear and self-evident. These are due to calculation mistakes or slippage in the decimal places while scribbling on the forms by the observer. Highly self-evident errors are corrected in the computer database and an equivalent correction and remark made on the manuscript. Thus the data as observed in the field by the observers and finally available in the computerised databases is termed as the *field database*.

This field data as available in the databases may have gaps or inconsistencies. Such gaps and inconsistencies are looked into at the time of processing the data. Wherever sufficient related data is available for filling-up these data gaps and correcting the inconsistencies, appropriate data in-filling and data correction procedures are employed. In this process of in-filling and correction, some of the data values get filled up or modified. This modification is carried out on the copy of the field data set and not on the original itself. Such a of modified set of data is termed a *processed database*.

A set of *filed data* is always to be stored for the purpose of future reference and archival. This availability of field data set is maintained at all DPCs, as *temporary databases*, for the requirement of day-to-day reference to the field data at the time of data processing. The availability of field data at the Data Storage Centre (DSC), as *permanent databases*, is purely for the purpose of long term archival and for limited dissemination for specific purposes, if required.

Processed data sets in the DPCs are the *temporary databases* that are worked upon during validation and processing whereas those in the DSCs are the final data sets as *permanent databases*.

Other Data Storage Centre(s)
(Permanent Databases)
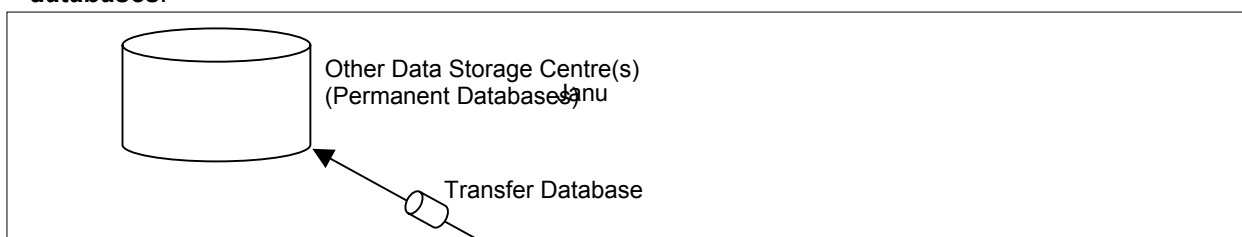
Transfer Database

Figure 6.2:   Illustration of permanent, temporary and transfer databases

6.4   ORGANISING TEMPORARY DATABASES AT SUB-DIVISIONAL DATA PROCESSING CENTRES

Surface Water Data Entry System (SWDES) is to be employed at the Sub-Divisional Data Processing Centres of all surface water agencies for entering the required observed meteorological and hydrological data. SWDES has been designed with a view to have a software using a standard, inexpensive and commonly available database software. Microsoft ACCESS software, being in the family of Microsoft Office suite, is expected to be available on all the computers being used by the

agencies. This Microsoft ACCESS database, therefore, has been used as the back-end database in SWDES.

All the data entered using SWDES is organised in a **well-defined database** which is also called as *workarea*. The software can initiate and maintain one or more workareas as per the requirement of the user. However, only one workarea is active at any point of time while working with the software. There are no restrictions put on the size of these workareas other than the one of the ACCESS databases themselves which is not restrictive in anyway for such small front-end applications. Though all the data from all the stations under a Sub-Divisional Data Processing Centre can be stored in one workarea itself but it is preferable to use smaller workareas for the purpose of better management, maintaining adequate speed while working and to categorise data in well-defined units. The suggested approach to be practised at the Sub-Divisional Data Processing Centres is as given here.

A **Sub-Divisional Data Processing Centre** (SDDPC) of the state SW agencies and the Central Water Commission (CWC) normally has a few river gauging stations, a few rainfall stations and a few full climatic stations under its jurisdiction. For the CWC, there are a maximum of 10 river gauging stations, on an average, in any SDDPC. The SDDPCs of CWC does not have any other type of station besides a few flood forecasting gauge stations. The number of observation stations under the SDDPCs of the state agencies is relatively larger. On an average, the state SDDPCs may have a maximum of 10 river gauging stations, 20 rainfall stations and 3 HP full climate stations.

From the hydrological data processing point of view it is useful to keep the data from all the stations together in one workarea so that the data of any of the station can be viewed immediately. But from the management and clearness sake it is preferable to have multiple databases categorised by the type of station; river gauging station, rainfall station and HP full climate station. Though the prevailing number of stations under a SDDPC can be organised in a single workarea, it is preferable to use multiple workareas whenever the numbers are very high. As a guideline, when there are upto about 10 river gauging stations, 10 raingauge stations and 4 HP Full Climatic stations, data from all the stations can be organised in one single workarea.

If the numbers are exceedingly high as compared to the ones given above then data can be organised in more than one workarea. While categorising data in multiple workareas one underlining principle has to be followed and that no station must appear in two different workareas. As a guideline, such multiple workareas can be based on the type of the station; river gauging station or rainfall station or HP full climate station. It is obvious that some of the river gauging stations would have observations on rainfall or even climatic variables and if multiple workareas are maintained at that SDDPC then all such rainfall and climatic data have to be entered with the concerned station which has been created and maintained in the workarea pertaining to river gauging stations.

Workareas for entering and managing historical data must not be combined with the current data (to be considered from 1/1/1999 onwards). For historical data, the organisation of stations in the workarea(s) may be different than as mentioned above, if it is so desired. This might be needed due to historical data being organised district wise and also in view of large number of years of historical data available with most of the stations. In such cases of historical data organisation, stations can be group on the basis of districts and also for the three types of stations separately as the volume of data would be very high. For the current data, workareas can be envisaged for a period of 5 years so that the same workarea remains operational for a considerable time and the burden of maintaining a huge number of workareas is avoided.

Every workarea or database is a single file containing all types of data, of all the stations within it. These files bear filenames by which the workareas are identified. All these workarea files are to compulsorily have the extension as ".MDB" in the filenames. Though there is no restrictions in naming the workareas but it is good practice to follow certain guidelines. This would bring uniformity in naming the workareas and also be helpful for an easy identification and recognition of the workarea. As a guideline, the name of the workarea may have reference, in the form of acronyms, to (a) the name of DDPC, (b) the name of SDDPC, (c) category of stations and (d) the period for which the data is maintained in that workarea. It would be a good practice to intersperse all these four acronyms with

an "underscore" ("_") for making the name a contiguous and unambiguous string of characters. Though this guideline is not a rigid instruction to be always followed but if done in this manner it would make it standardised, uniform and self-explanatory all over.

The acronyms for the DDPCs and the SDDPCs are to be either the names of the offices or the names of the places whichever is more commonly used for other purposes also. In some agencies, the DDPCs and SDDPCs are based on distinct sub-basins/basins and they are named on the basis of the names of these sub-basins/basins and that is why are better recognised by such names as compared to the name of the place. In other cases, the name of the cities/towns can be more common in use for identifying a certain DDPC or SDDPC. The basic idea is to use some acronym which can be easily recognised and related with the DDPC or SDDPC rather than using some abstract numbering as DDPC 1, DDPC 2 or SDDPC 1, SDDPC 2 etc. .

For the portion of the name on the category of the station a common guideline of using "GD", "CLIM" and "RAIN" can be followed for the three types viz. river gauging (gauge-discharge), HP full climatic and rainfall stations (both SRGs and ARGs included) respectively.

The period for which the data is maintained in a particular workarea can be simply given by stating the start and end years in the form of two digit numbers in succession. For example, if a certain workarea is to maintain data for a period of hydrological year 1999-2000 to year 2003-2004, it can be simply represented in the name as "9903".

Three examples below illustrate how the data for a Sub-Division in different situations can be maintained.

Case 1: Consider that the Lower Godawari Sub-Division No. 2 (Rajamundry) of Lower Godawari Division (Hyderabad) of CWC has 9 river gauging stations in all. Since the number of stations is not very large (even less than 10 as mentioned above) it is appropriate to have only one workarea at the Sub-Division for storing data of all these 9 stations together. If it is thought to have the workarea for a period of five years from 2000-2001 to 2004-2005 then the name of the file can be given as "LGD_RJY_0004.MDB". It is important to note here that for the DDPC and SDDPC "LGD" and "RJY" have been used which implies that DDPC is Lower Godawari Division and SDDPC is at Rajamundry. Both of them are though different in the way that one indicates the name of the DDPC and other the place of the SDDPC but since these are the more commonly understandable names for these DPCs they are used.

Case 2: Consider that there are 27 rainfall stations (including both, SRGs & ARGs), 8 HP full climatic stations and 17 river gauging stations under Kolhapur Sub-Division of Pune Division of Maharastra state. Since the number of stations are comparatively more it is good to organise these stations in three different workareas as per the category of stations. Thus three workareas can be organised for the three types of stations. If is required to organise 5 years of data from 1999-2000 to 2003-2004 then the names of these workareas can be given as: "PUNE_KOLH_RAIN_9903.MDB", "PUNE_KOLH_CLIM_9903.MDB" and "PUNE_KOLH_GD_9903.MDB" respectively.

Case 3: Consider that there are 17 rainfall stations (including both, SRGs & ARGs) and 8 river gauging stations under Nellore Sub-Division of Guntur Division of Andhra Pradesh. Since the total number of stations in each category is not very large it is better to organise all the stations in one workarea itself. If 5 years of data from 1999-2000 to 2003-04 is required to be organised then the name of workarea filename can be "GUNT_NELL_9903.MDB".

Normally, all the stations under the SDDPC belong to the same sub-basin of the independent river basin. It is unlikely that a SDDPC's jurisdiction is spread over parts of more than one sub-basin. If in case the jurisdiction is indeed spread over more than one sub-basin and at the DDPC also these sub-basins are to be organised in separate workareas then the stations of each sub-basin are to be kept in separate workareas at the SDDPC, on the lines similar to as mentioned above.

Thus at the SDDPCs there will be ***only one workarea*** if the number of stations are not excessive as discussed above or ***at the most three*** if there are very many stations under a Sub-Division. Whatever is the number of workareas at the SDDPCs, these are to be taken as ***temporary databases*** which contains ***field data***. At the level of SDDPCs there is no requirement to keep the processed databases as no data is expected to be modified at this level. Only the suggested values in case of any inconsistency or gaps in the data, whenever and wherever required, are to be recorded in the slots for remarks or in writing on the reports taken from SWDES and communicated to the DDPCs. This communication of hardcopy is made as soon as the transfer database is sent from SDDPC to the DDPC at monthly interval.

6.5   ORGANISING TEMPORARY DATABASES AT DIVISIONAL DATA PROCESSING CENTRES

At the DDPCs, both, the primary and the secondary modules of hydrological data processing system would be operational. The primary module (SWDES) would be used to consolidate the field data arriving from various SDDPCs at regular intervals and transfer data from primary to secondary module. At the DDPCs, it is very essential to have an exact replica of the contents of the databases available at SDDPCs. For this purpose, it is required to have a copy of all workareas which are used at the SDDPCs available at DDPC. The names of these workareas must also be exactly same as used at the SDDPCs except for a suffix "_DDPC" to indicate that it is maintained at DDPC level. When the incremental data from SDDPCs is regularly transferred to the DDPCs, it is consolidated in the respective workarea. This would also ensure a full backup at DDPCs, of the works as carried out at various SDDPCs. Such workareas can keep working for a considerable period (say 5 years).

In case there is only one workarea at each SDDPC then the number of workareas at DDPC would be small and manageable. However, if there are separate workareas for different categories then the number would be larger and less compact. To make the transfer of data from primary to secondary module at the DDPCs with adequate brevity and comfort, it is required to consolidate individual SDDPC workareas into one common workarea for each of the SDDPC. That is, the contents of the multiple workareas, if available at any SDDPC, are to be consolidated into a single workarea for that SDDPC. However, in case an SDDPC is having separate workareas on account of having jurisdiction spread to two or more distinct sub-basins (which is a highly unlikely case) similar distinction can be retained while consolidating into common workareas.

The acronyms for the workareas in primary module at the DDPCs can be on similar lines as those at the SDDPCs except that sometimes there can be separate workareas at the SDDPCs for different types of stations whereas at DDPCs they would all be combined into one workarea. Thus the acronym for the type of station would not be required at the DDPCs. The name of the workareas for the primary module at the DDPCs would be based on (a) name of DDPC, (b) name of SDDPC, (c) period for which the data is to be maintained at the DDPC and (d) the suffix "_DDPC". Normally, a period of 5 years can be a suitable length of time for which this organisation can be done.

Thus, at every DDPC, there will be exact replicas of all the workareas being used at different SDDPCs (with only the suffix "_DDPC" added to the file names) and if there are multiple workareas in use at one or more SDDPC then one combined workarea for each SDDPC would be required to be established additionally. A full view of each SDDPC's workarea can be obtained from these combined workareas for each SDDPC.

In most cases the jurisdiction of the DDPCs would be within an independent river basin. In only rare case it may cover more than one distinct sub-basins or part thereof of the independent river basin. Organisation of the work and the databases in the secondary module at the DDPCs is to be done such that all the data (rainfall, climate and hydrological) pertaining to each distinct sub-basin or group of sub-basins resides in a separate workarea. That is to say that the tributary to the independent river (i.e. the sub-basin) is to be taken as the smallest unit for organising the database at the DDPCs and thereby all stations within must be together. On the other side, if the DDPC's jurisdiction is extending upto a location on the independent river itself or covers the whole independent river then the workarea

is to be for that extent of basin coverage. It is needless to say that, if a DDPC covers more than one independent river basin then the sub-basins with these basins must be organised in different workareas separately.

Thus at DDPCs there would be usually one workarea in the secondary module on the basis of drainage area under its jurisdiction. Obviously, all the data from one or more SDDPCs pertaining to this drainage area (of the independent river) is to be pooled together in this every time when incremental data from SDDPC is available. Only in rare cases, it may have two or more workareas for distinct sub-basins/zones of independent river basin/zone.

The name for the workareas in the secondary module at the DDPCs can be based on the name of the DDPC and on the name of the drainage area covered. Since these workareas in the secondary module are to serve for considerably long period of time there is no specific purpose solved by including the period also in the name of the workarea. However, any extra qualifier required for greater distinction from any other workareas available at the DDPC can always be included. Thus the name of the workareas in the secondary module at the DDPCs would be based on the (a) name of DDPC and (b) name of drainage basin/area.

The examples below illustrate how the data in the secondary module, at the DDPC, could be maintained.

Case 1: Consider the case of Pune Division of Maharastra state which has six SDDPCs and there are two distinct drainage areas, Bhima and Krishna upto ???, under its jurisdiction. Since these forms two distinct portions for the independent river Krishna, it is appropriate to organise all the data of the DDPC in two parts: one for Bhima sub-basin and another for sub-basin of Krishna at ???. The names of the workarea and its directory can be given as: "PUNE_BHIMA" and "PUNE_KRISHNA".

Case 2: Consider the case of Dowalaiswaram Division of AP state which has lower east Godawari region and a sizable region in northern Andhra Pradesh called "North Circars" draining directly into the sea. At DDPC, workareas can be maintained for these distinct regions. The names of the directory and workareas for these regions at the DDPC can be DOWL_LOWEAST_GODAWARI and DOWRM_NORTHCIRCARS.

6.6   ORGANISING TEMPORARY DATABASES AT THE WATER QUALITY LABORATORIES

All the WQ data would be entered in well-defined SWDES databases maintained by the water quality laboratories II / II+. Typically, a state or regional office would have 2-3 water quality level II / II+ laboratories covering jurisdiction of one or two Divisions. About 50–100 surface water observation stations would normally be supported by a laboratory on an average. All the data from all the stations under a laboratory can be stored in one workarea (database) itself but in case it is so required, for maintaining adequate speed while working and/or for sub-dividing data into smaller sets, more than one databases could be maintained.

Workareas for entering and managing historical data must not normally be combined with the workareas for current data. For historical data, the organisation of stations in the workarea(s) may be different than as mentioned above. This might be needed due to historical data being organised district wise and also in view of large number of years of historical data available with most of the stations. In such cases of historical data organisation, stations can be grouped on the basis of districts and on the basis of periods of long duration (say 20-30 years) as per the convenience.

For the current data, workareas can be easily kept for periods of say about 5 years so that the same workarea remains operational for a considerable time and the burden of maintaining a huge number of workareas is avoided.

Every workarea is a single file containing all types of data, of all the stations within it. These files bear filenames by which the workareas are identified. All these workarea files are to compulsorily have the extension as ".MDB" in the filenames. There is no restriction in naming the workareas but it is good practice to follow certain guidelines. This would bring uniformity in naming the workareas and also be helpful for an easy identification and recognition of the workarea. As a guideline, the name of the workareas maintained at the WQ laboratories may have reference, in the form of acronyms, to (a) the name of the laboratory, (b) the name of the place (city) where the laboratory is located, (c) the period for which the data is maintained in that workarea. It would be a good practice to intersperse all these three acronyms with an "_ (underscore)" or a "- (hyphen)" for making the name a contiguous and unambiguous string of characters. For the first two acronyms, i.e. the name of the laboratory and its place, the laboratory codes as already available and used in SWDES are very well suited. Though this guideline is not a rigid instruction to be always followed but if done in this manner it would make it standardised, uniform and self-explanatory all over.

The period for which the data is maintained in a particular workarea can be simply given by stating the start and end years in the form of two digit numbers in succession. For example, if a certain workarea is to maintain data for a period of hydrological year 1999-2000 to year 2003-2004, it can be simply represented in the name as "9903".

Two examples below illustrate how the data at water quality laboratory of central and state agencies can be maintained.

Case 1: Consider that Dowleishwaram SW-WQ laboratory of A.P. state has to make a database for all the stations under its jurisdiction and the database is intended to have data for the 1999-2000 to 2003-2004. Since this laboratory has the code as "SAP-DOW1", the name of the database can be "SAP-DOW1-9903.mdb".

Case 2: Similarly, for a CWC laboratory at Coimbatore, the data for the period 2000-2001 to 2005-2006 may be organised in a database with the name "CTN-COIM-0005".

6.7   ORGANISING TEMPORARY DATABASES AT STATE/REGIONAL DATA PROCESSING CENTRES

At the SDPCs/RDPCs, the whole state or a very large drainage region is under the jurisdiction. The state would include parts or full of one or more independent river basins whereas the regions of CWC being based on the river basins would include one or more complete independent river basins. Both primary and full package would be operational at the SDPCs/RDPCs.

The primary module would be employed to consolidate the field data emanating from all the DDPCs. The DDPCs will have regular consolidation all data of every SDDPC under it in one single workarea. Similar workarea is to be available at the SDPC/RDPC for the purpose of reference and most importantly for transfer of field data to the Data Storage Centre. This is done by regular consolidation of incremental data sets being sent by the DDPCs to SDPC/RDPC. The names of these workareas have to be exactly similar to what is used at the DDPCs with the exception that instead of "_DDPC" as the suffix "_SDPC/_RDPC" is to be used.

The full package at the SDPCs/RDPCs would be employed to aim at hydrological data validation and reporting for the complete river basin(s) or part thereof within the state. The SDPCs will also require to pool data from all the river basin(s) within the state boundary together to get an overall view for the whole state or smaller administrative units, especially for rainfall information. The contents of such unified workarea may not include all the data to the maximum available details but will largely have finalised summary data on daily/monthly/yearly levels for various variables and as per the requirement of the SDPCs.

Thus, for the organisation of data in the full package at the SDPCs/RDPCs, two type of workareas have to be maintained. The first type of workareas will be for the individual independent river basins or part thereof within the state. Thus as many workareas would have to be established as the number of independent river basins within the SDPC/RDPC's jurisdiction. In special circumstances, a group of very small independent rivers (as will be the case with the coastal rivers which directly drain into the sea) may be clubbed together in one workarea. The second type of workarea is required at the SDPCs to pool up all the data for the whole state together.

The names of the workareas of the first type have to be based on the name of the independent river basin or the group of smaller basins taken together with the indication of name of the concerned state, if it is including only a part of it within its boundary. Thus the name for the workarea would comprise of acronyms for (a) name of the independent river basin or group of smaller basins and (b) name of the state. The name of the workarea of second type is obviously to be based on name of the state itself since this would contain data for the whole of the state.

6.8   ORGANISING TEMPORARY DATABASES AT THE NATIONAL DATA PROCESSING CENTRES

At the national level the Central Water Commission will have a National Data Processing Centre (NDPC) together with a National Data Storage Centre (NDSC). At this NDPC, the required data of all the CWC observation stations for different river basins/zones of the HP area will be organised. Apart from the CWC observation stations, data of the selected observation stations of various states could also be available at the NDPC depending on the requirement and protocol between various states and the CWC.

The NDPC will require to organise the authenticated data from the respective RDPCs . The contents of the workareas at the NDPC may not include all the data to the maximum available details but will largely have finalised summary data on daily/monthly/yearly levels for various variables. Such an

organisation is essential at the national level for providing an integrated view on a macro level of the peninsular region as a whole. Also, NDPC will be required cater to any requirement which may come to it for providing hydrological information on any river basin in the peninsular region of the country.

The name of the workareas for various river basins at the NDPC is to be exactly similar to those used at the RDPCs except for the fact that a suffix "_NDPC" can be put at the end so as to make a distinction from the similar workareas operative at the RDPCs.

7   MAP LAYERS AND NETWROK LAYERS

7.1   WHAT ARE MAP AND NETWORK LAYERS IN HYMOS

HYMOS provides for a graphical user interface, component called NETTER, in the form of maps showing various topographic features. Various features like rivers, canal, lakes, reservoirs, observation stations, elevation contours, roads and other transport/communication lines, cities, districts/province and other administrative boundaries etc. are normally of interest to the hydrologists while working for hydrological data processing. Information on all these topographical features is required in digital form for making use in various software systems. This digital information is produced in various forms as vector or raster by various digitising procedures in different formats. Such information is generally used in any typical Geographical Information System (GIS) for purpose of reference and analysis.

In NETTER, the observation stations are depicted by **nodes** of the monitoring network and are combinedly taken as a **network layer**. A distinction is made between the map information on the location of observation stations (to be associated with the database) and that on all other map features. Locations of observation stations are considered as nodes and lot of data for the same is maintained in the associated databases. These location and other attributes of these observation stations (called as nodes in NETTER) are kept in separate file called "HYMOS.NTW". The important attributes of the observation stations being kept in the network layers are the latitude, longitude and the type of the station (i.e. the node type). These network layers are created for a database by either adding observation stations one-by-one or by importing stations from a transfer database.

The **map layers**, in HYMOS, on the other hand are those features which are required to be used for the purpose of reference. Such layers are though not strictly related with the data in the database but may sometimes be used during certain hydrological computation. The basic building blocks of any network or map layer are the geometric entities as point(s), line(s) and polygon(s). Any line or polygon feature also, in fact, is a composition of several points only. And that any map layer is a collection of blocks of the required point(s), line(s) and/or polygon(s). Usually, point, line or polygon type of features are kept in separate layers with due distinction but may also be combined in certain cases is so required. Similarly, one or more topographical features as river, lakes, spot levels, elevation contours etc. can be in separate or combined files as per the requirement of the user. Also, for a certain feature like river, the information can be organised for individual rivers separately or in combined form as per the requirement.

7.2   HOW TO OBTAIN MAP LAYERS

Map layers are obtainable from various types and sources of information like topographical maps and remotely sensed data in digital form etc. Any topographic feature can be considered as comprised of point(s), line(s) or polygon(s). From the base topographic maps such information is to be digitised using a digitiser. The digitising operation can be accelerated by scanning the topographic maps and then digitising directly on the computer screen, in a semi-automatic manner, on the basis of scanned image.

While digitising any information from the topographic maps, certain aspects like the scale of the base map (which affect the accuracy), map projection system and format of the digital output required etc. must be taken into account. There are standard software, like Didger, specifically for digitising the information from a paper map and from scanned images on computer screen directly. Also, most of the standard GIS software like ARC-INFO, ILWIS etc. come with such digitising components. It is also necessary and possible to check on the consistency of the digitised information using certain tools embedded in the digitising software. Various formats used by different digitisation software are:

- *.BNA   (Atlas GIS, but also PC-Arc/Info can export this format)
- *.MIF   (Map Info Format)
- *.MOS   (MOSS format, a grid based GIS)
- *.SHP   (Universal Shape Format)

Under HP the desired information is tried to be prepared State-wise by digitising topographical themes from 1:50,000 scale toposheets of Survey of India. Some of the additional themes like soil type, land use and geology would be prepared from information available on 1:250,000 scale with related agencies like National Bureau of Soil Survey & Land Use (NBSSLU), Nagpur and Geological Survey of India (GSI), Calcutta.

Since the above process of digitisation by the agencies will take considerable time, it is planned that for the time being readily available digitised information on 1:1000,000 from other sources like Digital Chart of the World (DCW) is used. Such information has been distributed to all the users in the form of a CD containing digital data on themes like rivers, lakes, canals, contours, spot heights, roads, country, cities, villages, etc. on 1:1000,000 scale for around the globe. Users can obtain information on any of these themes from this CD for any desired area.

For extracting information from "DCW" CD, a software called "DCW-Clipper" which is available on the same CD, is to be installed. This Installation provides two components: (a) DCW-Mapper and (b) DCW-Clipper. First, DCW-Mapper software is to be run which brings the map of the globe as shown in Figure 7.1. User can then zoom-in to any portion of interest on the map and use the option of "Clip" to input data for extracting desired map information as shown in Figure 7.2. The digital map information
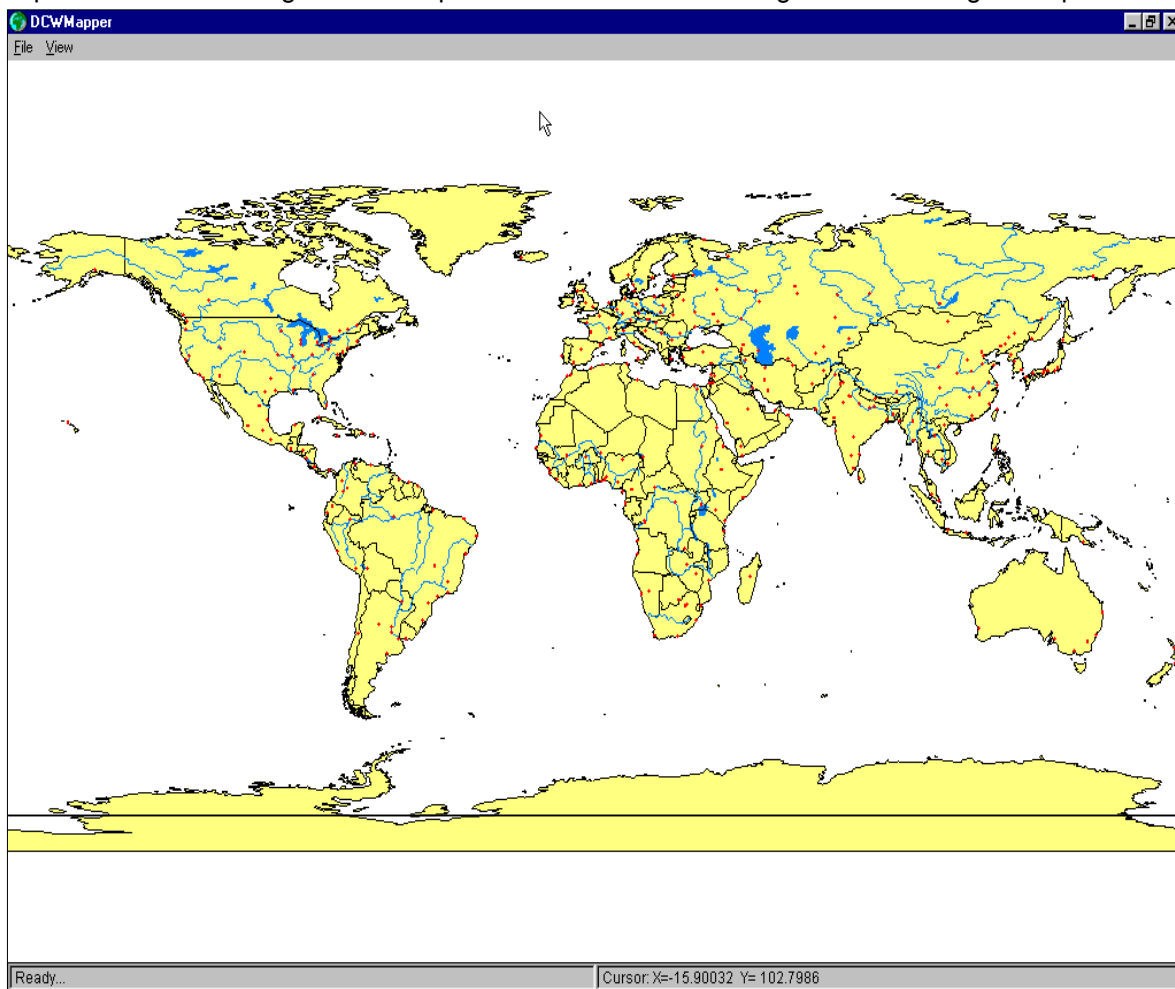
can be extracted in any of the formats like "Atlas*GIS (*.BNA), MapInFo (*.MIF), Mapper (*.MPL) and ArcView Shape (*.SHP). The extracted information is thus exported at the user specified location.

Figure 7.1:   DCW-Mapper screen showing map of the World for defining area of interest to be clipped
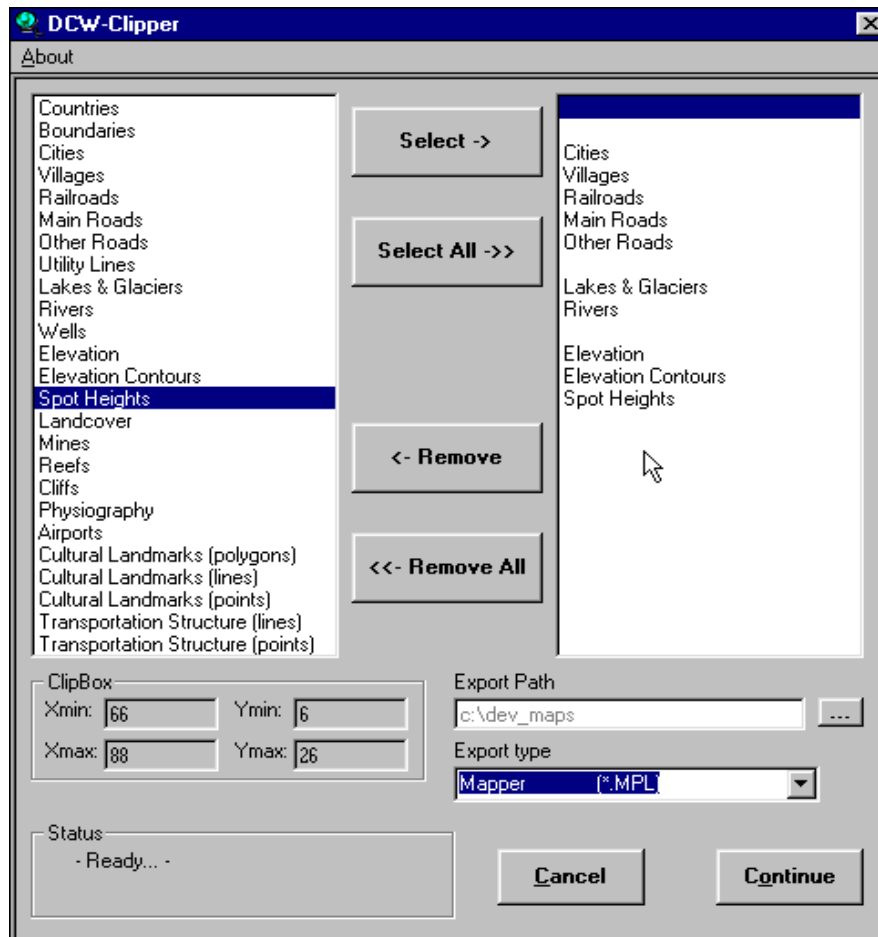

Figure 7.2:   DCW-Clipper screen showing input box for selection of layers to be clipped

Though NETTER can work with any of the formats mentioned above, a conversion program called "MAPLINK" is also provided. This program helps converting digital map layer information in one format to another. In fact, this program was necessary with previous versions of NETTER which use to accept map layers in "*.MPL" format only. For converting information in one format to another the MAPLINK program is used as shown in Figure 7.3. User has to specify the path at which the files to be converted are located together with the input and output file types. After the desired input is given, conversion is done by the program and user obtains information in the desired format.
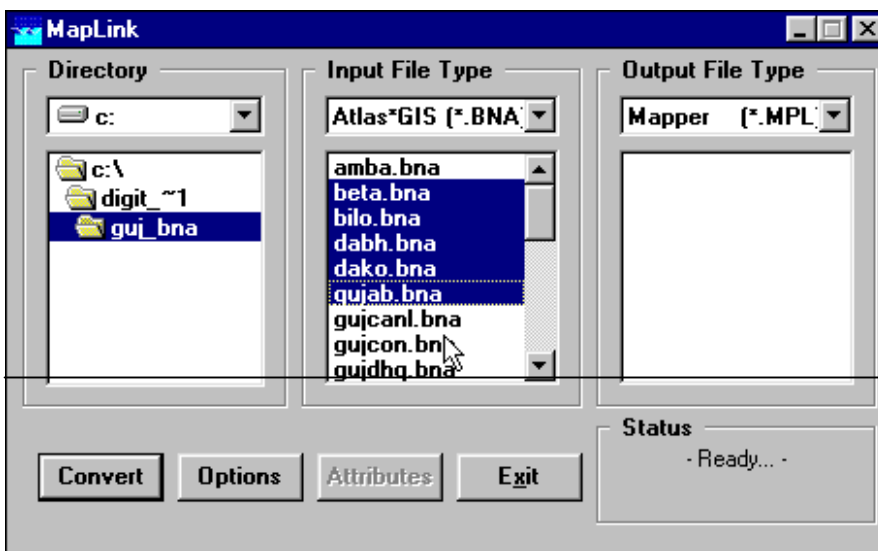
Figure 7.3: MAPLINK program for

conversion of GIS layers from one format into another.

7.3   HOW TO MAKE NETWORK LAYERS

Network layer comprise of observation stations of different types. Any number of types of observation stations can be defined in the system. Some of the pre-defined types of stations are:

- water level

- discharge

- groundwater

- water quality

- meteorological

- rainfall

- structure

- spatial average etc.

In HYMOS, the observation stations are depicted by **nodes** of the monitoring network and are combinedly taken as a **network layer**.

In NETTER, a distinction is made between the map information on the location of observation stations (to be associated with the database) and that on all other map features. Locations of observation stations are considered as nodes and lot of data for the same is maintained in the associated databases. These location and other attributes of these observation stations (called as nodes in NETTER) are kept in separate file called **Network layer**. The important attributes of the observation stations being kept in the network layers are the latitude, longitude and the type of the station (i.e. the node type). These network layers are created for a database by either adding observation stations one-by-one or by importing stations from a SWDES transfer database.

Observation stations (network nodes) can be manually added in HYMOS using "Edit Network option" of Netter. Various options like "Add", "Move", "Delete", "Node Type", "Rename" and "Properties" are available as editing options (see Figure 7.4).
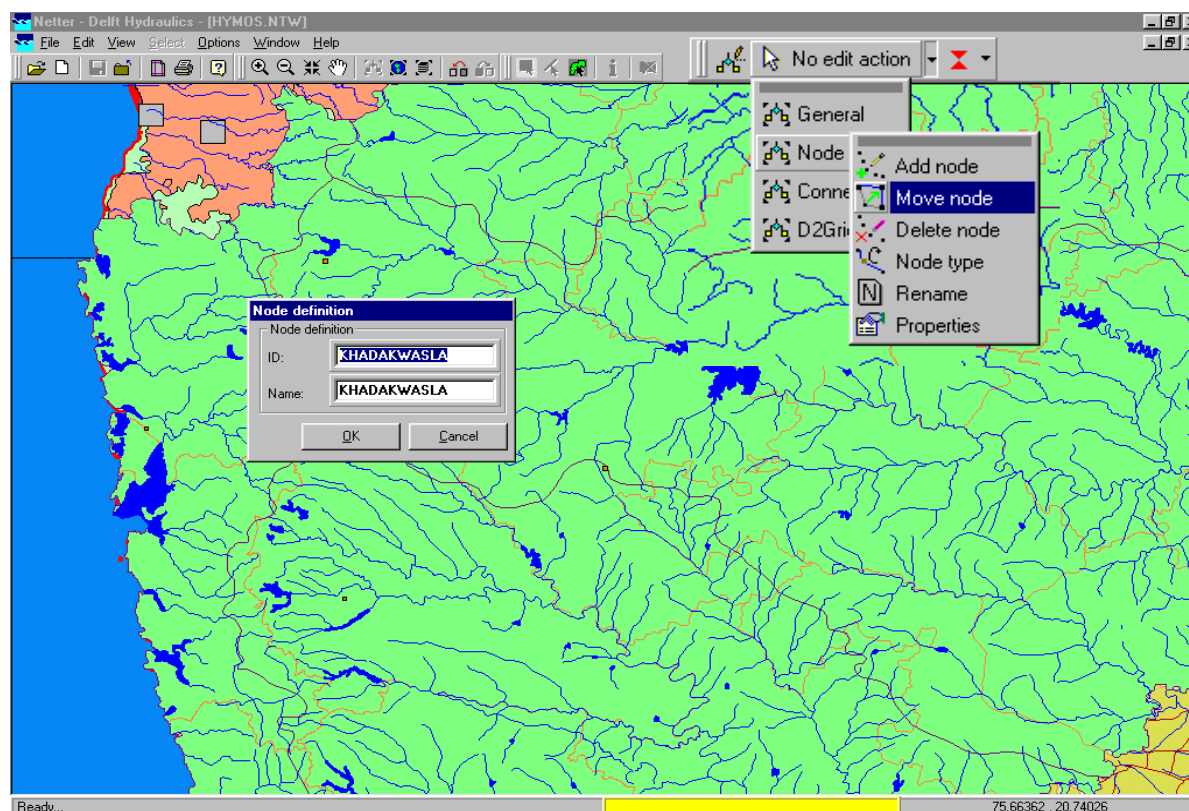
Figure 7.4:   Editing option for management of network nodes.

By using "Add" option an observation station can be created on the map. User is to give details on the station code and name and the station has to be located by the mouse pointer. Similarly, the created station can be moved elsewhere, if required, by choosing the option of "Move" or can be deleted by the "Delete" option. The station type is the category of the station which it belongs to and can be edited by using "Node Type" option. In case the name of a station is to be edited then "Rename" option can be used. The option on "Properties" is used for relocating the station by numerically changing the co-ordinates. The station automatically takes the new position as per the changes made. Normally, the "station codes" are displayed adjacent to the location of the station but in case it is needed to display "station names" or not to display anything then suitable changes can be made using "options" from the main menu and "options …" from the list of sub-menus. In this manner the network nodes can be created and maintained.

In case the network is to be established by directly importing the stations from the SWDES transfer database, the procedure is very simple. From SWDES the required stations are exported using the "Export to HYMOS" option and then the transfer database thus created is imported in HYMOS database. As soon as the import is over all the stations are created together with the associated characteristics.

7.4   HOW TO WORK WITH MAP LAYERS

Once the map layers are available they can be copied in the "Maps" directory in a particular database. Though these layers can be accessed from anywhere without placing them in a particular directory but it is always better for the sake of convenience that they are available in the folder of the concerned database itself. Maps layers are managed using the sub-option of "Map Options" under "Options" item on the menu bar of Netter program. Upon selecting this "Map Options" item, an input box as shown in Figure 7.5 appears on the screen. There are few entities on this input box through which the display and use of any map layer can be effected.
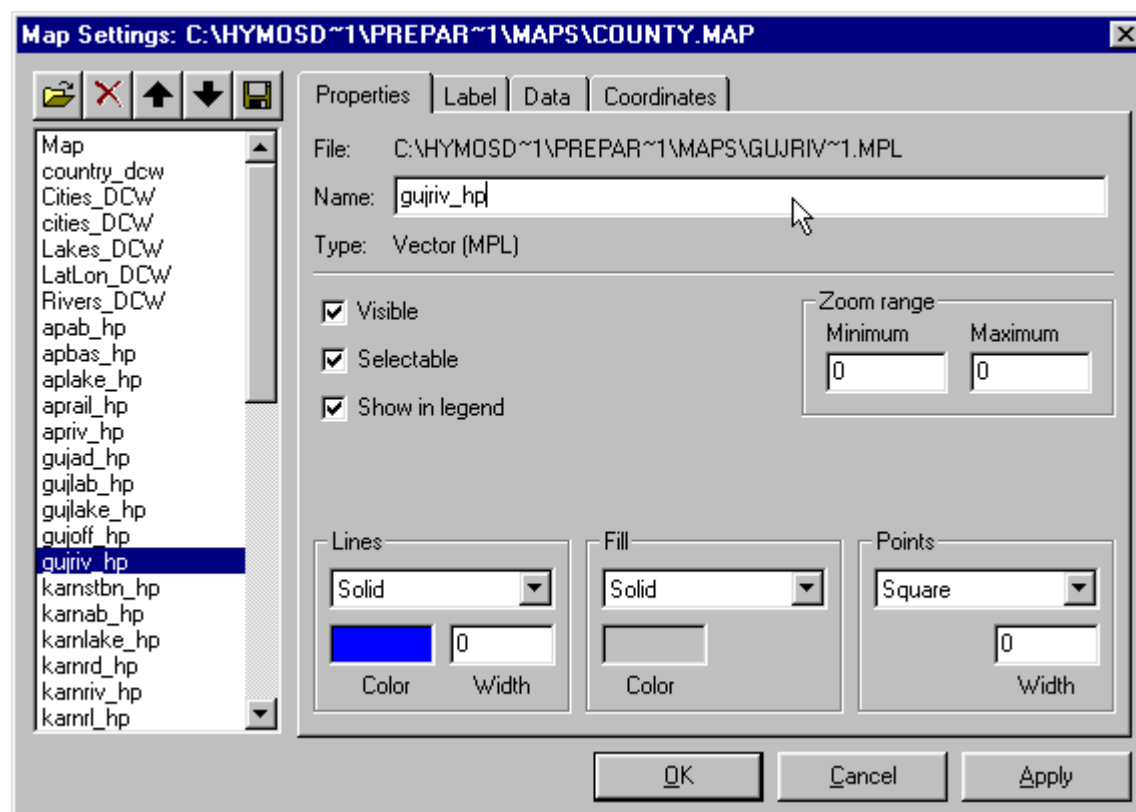
Figure 7.5:   Input dialog box for management of map layers.

First of all, there is a list box in the left-hand side showing all those map layers which are registered with a particular database. There is a button on the top for locating and opening up any new map layer. Few more buttons are available for deleting, advancing or retreating any map layer. Further, there are some more options for setting the properties of any layer. Using the option of "Properties" any map layer can be made visible or selectable or if it is to be reflected in the legend. The colour and thickness of the lines or the colour of the fill of a polygon or the size and type of the points can all be settable from this "Properties" option. A very useful option is to set the minimum and maximum zoom range for any map layer so that the layer ceases to show up when the scale of the map becomes less than or more than the specified limits respectively by zooming out or zooming in. The minimum limit is very useful for avoiding display of certain details which otherwise may make the map cluttered.

Secondly, provision is available for setting the properties for the point items on the map by using "Lable" option. Settings like name and identification of the point and location of the label, its size, fonts etc.. Also, the option on "Coordinates" shows the areal extent of the map layer towards the four sides.

Most of the map layers are used for the purpose of reference by displaying on the screen. It may also be possible to look at the properties of these layers such as length of the river, area or perimeter of a catchment, value of the contour provided that the same is available in the database. However, few layers such as the catchment boundaries are also used for calculation of areal estimates of rainfall etc.

7.5   HOW TO WORK WITH NETWORK LAYERS

Network layers are basically the nodes and their inter-connections denoting observation stations and hydrological links as river or channel reaches. After the network is established as mentioned in previous section, these nodes can be used for selection of stations and thereupon processing the data available at any of the selected stations. Selection of stations can be made in two ways: (a) by individually selecting stations by clicking with mouse (multiple stations by keeping the SHIFT key

pressed) or (b) by selecting all the stations within a map item like a basin or sub-basin. For the later case it is required to choose the option of "Select by map item" under the "Select" option in the menu bar of Netter.

8   DATA TRANSFER BETWEEN TEMPORARY DATASETS

8.1   TRANSFERRING FIELD DATA FROM SDDPCs TO DDPCs OR FROM WQ LABORATORIES TO SDPCs

The field data is to be archived at the SDSCs/RDSCs for long term reference and it is also to be available at all the DPCs/SDPCs/RDPCs for immediate reference during the process of validation of data. The field data sets are originally prepared at the SDDPCs or at the WQ laboratories by keying-in the observed/analysed data into the SWDES workareas and removing all data entry mistakes and highly self-evident errors. From any SDDPC, this field data is transferred to the controlling DDPC by making transfer workarea of incremental data readied beyond the last such transfer to the DDPC. In case of water quality data, the primary validated data from WQ laboratories would be sent directly to SDPCs/RDPCs.

8.2   TRANSFERRING RAW DATA FROM SDDPCs TO DDPCs

The raw data is to be archived at the SDSCs/RDSCs for long term reference and it is also to be available at all the DPCs/SDPCs/RDPCs for immediate reference during the process of validation of data. The raw data sets are originally prepared at the SDDPCs by keying-in the observed data into the SWDES workareas and removing all data entry mistakes and highly self-evident errors. From any SDDPC, this raw data is transferred to the controlling DDPC by making transfer workarea of incremental data readied beyond the last such transfer to the DDPC.

Use of "Fragmentation" option available in SWDES is made for making such incremental data sets at the SDDPCs. Normally, as a regular routine monthly transfer of incremental data sets from SDDPCs to DDPCs (by 10th of each month), the data entered for the last month is fragmented and sent to the DDPC. Fragmentation option is available as a button on the tool bar of the main switch board. The form for fragmentation option is as shown in Figure 8.1. In this, all the entered data for any user-defined set of stations from any start date to any end date can be copied from an existing database and put into a new transfer workarea. The selected stations are highlighted and the chosen start and end dates are displayed for the purpose of reference at the time of making the transfer workarea. The user has to specify the name of the file to be created by the program and receive all the data being transferred. Such transfer workareas are made separately for each of the workareas available at the SDDPCs.

8.3   TRANSFERRING RAW DATA FROM SDDPCs TO DDPCs

The raw data is to be archived at the SDSCs/RDSCs for long term reference and it is also to be available at all the DPCs/SDPCs/RDPCs for immediate reference during the process of validation of data. The raw data sets are originally prepared at the SDDPCs by keying-in the observed data into the SWDES workareas and removing all data entry mistakes and highly self-evident errors. From any SDDPC, this raw data is transferred to the controlling DDPC by making transfer workarea of incremental data readied beyond the last such transfer to the DDPC.

Use of "Fragmentation" option available in SWDES is made for making such incremental data sets at the SDDPCs. Normally, as a regular routine monthly transfer of incremental data sets from SDDPCs to DDPCs (by 10th of each month), the data entered for the last month is fragmented and sent to the DDPC. Fragmentation option is available as a button on the tool bar of the main switch board. The form for fragmentation option is as shown in Figure 8.1. In this, all the entered data for any user-defined set of stations from any start date to any end date can be copied from an existing database

and put into a new transfer workarea. The selected stations are highlighted and the chosen start and end dates are displayed for the purpose of reference at the time of making the transfer workarea. The user has to specify the name of the file to be created by the program and receive all the data being transferred. Such transfer workareas are made separately for each of the workareas available at the SDDPCs.

Figure 8.1:   Layout of the screen for fragmentation of the data

A uniform guideline may be followed for assigning file names to such transfer workareas being sent from the SDDPCs to the controlling DDPCs. This will be highly beneficial in recognising the transfer workarea by merely knowing its filename. The filenames to be used for these transfer workarea are recommended to be based on: (a) a string "TRANnn", (b) yymmdd and (c) name of the parent workarea. Here "nn" in the string is any number used for adequately distinguishing different transfer workareas of the same date and "parent" workarea is the one from which the transfer is being made.

Thus the data required to be transferred from a SDDPC to the DDPC can be prepared as explained above. The actual communication of these transfer workarea files to the DDPCs can be through any of the links prescribed under the Hydrology Project such as using physical media (preferably on CD), data transfer networks, telephone line etc. For the purpose of copying transfer workarea file(s) on to CD every DPC will be supplied with a CD writer individually. Practice for such transfer using physical media must be ensured so that data can be communicated (though with a lag of couple of days) even when other automatic communication link may not be working perfectly.

Use of "Fragmentation" option available in SWDES is made for making such incremental data sets at the SDDPCs/WQ laboratories. Normally, as a regular routine monthly transfer of incremental data sets from SDDPCs/WQ Labs to DDPCs / SDPCs respectively (by 10th of each month), the data entered for the last month is fragmented and sent to the DDPC/SDPC. Fragmentation option is available as a button on the tool bar of the main switch board. The forms for fragmentation option are as shown in Figure 8.1 and 8.2. From the first form, all the stations for which some data is required to be put in the fragmented subset are selected. In the second form the required types of data can be selected individually along with the required period for which the subset is to be made. In this, any type of entered data for any user-defined set of stations from any start date to any end date can be copied from an existing database and put into a new transfer workarea. The selected stations are highlighted and the chosen start and end dates are displayed for the purpose of reference at the time of making the transfer workarea. The user has to specify the name of the file to be created by the program and receive all the data being transferred. Such transfer workareas are made separately for each of the workareas available at the SDDPCs.
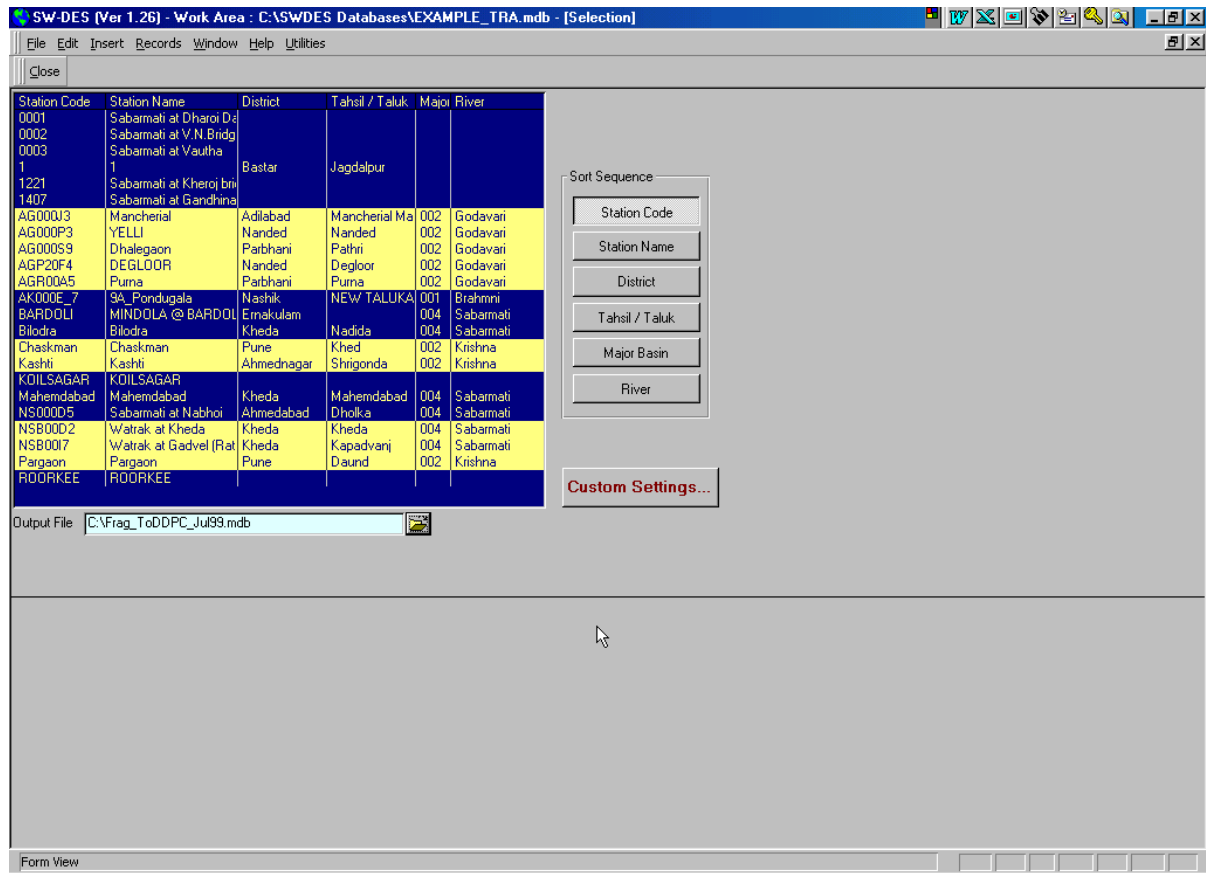
Figure 8.1:   Layout of the screen for selecting stations for fragmentation of the data
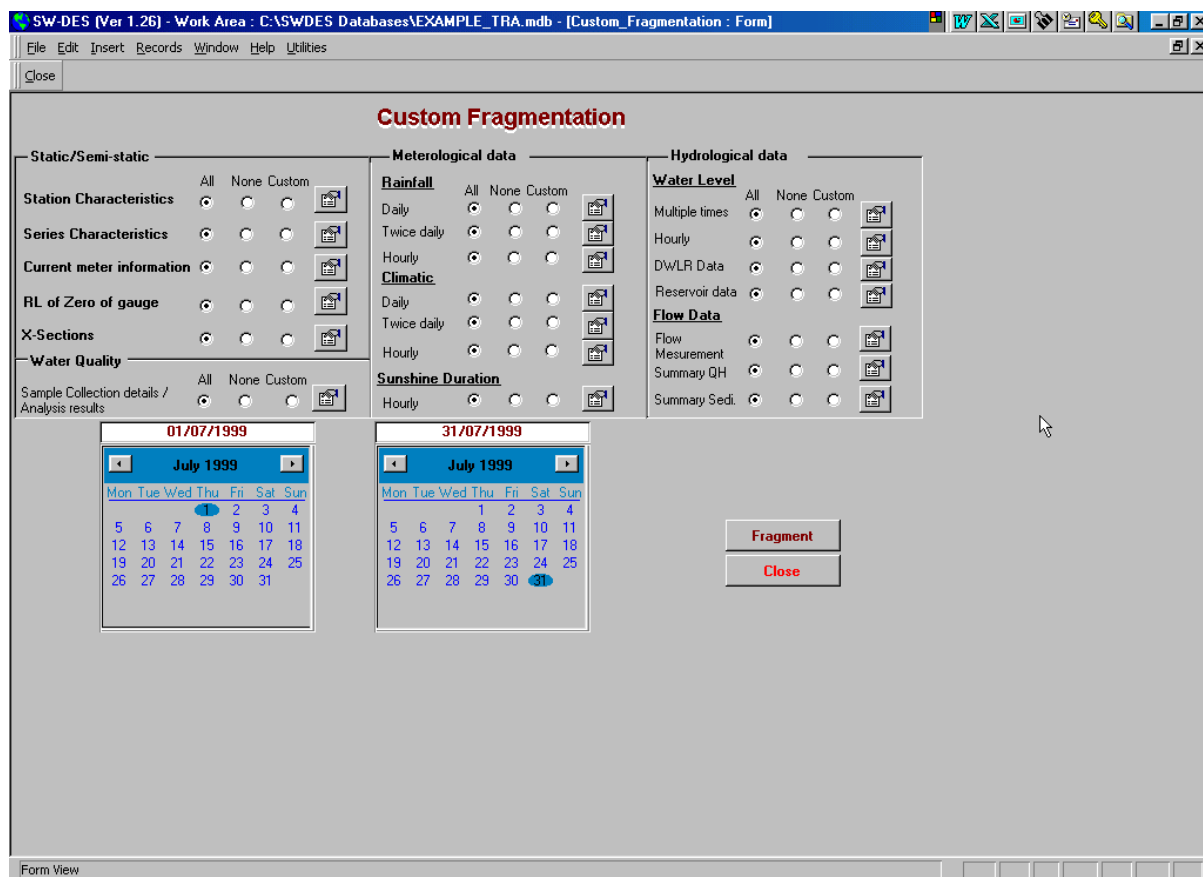
Figure 8.2:   Layout of the screen for selection of data types and period for fragmentation

A uniform guideline may be followed for assigning file names to such transfer workareas being sent from the SDDPCs to the controlling DDPCs. This will be highly beneficial in recognising the transfer workarea by merely knowing its filename. The filenames to be used for these transfer workarea are recommended to be based on: (a) a string "TRANnn", (b) yymmdd and (c) name of the parent workarea. Here "nn" in the string is any number used for adequately distinguishing different transfer workareas of the same date and "parent" workarea is the one from which the transfer is being made.

Thus the data required to be transferred from a SDDPC to the DDPC can be prepared as explained above. The actual communication of these transfer workarea files to the DDPCs can be through any of the links prescribed under the Hydrology Project such as using physical media (preferably on CD), data transfer networks, telephone line etc. For the purpose of copying transfer workarea file(s) on to CD every DPC will be supplied with a CD writer individually. Practice for such transfer using physical media must be ensured so that data can be communicated (though with a lag of couple of days) even when other automatic communication link may not be working perfectly.

8.4   CONSOLIDATING FIELD DATASETS AT THE DDPCs

At the DDPCs there are SWDES workareas which are exact replicas of the workareas operative at SDDPCs. Once the field data sets are received at the DDPCs (sent from the SDDPCs), they are consolidated in these respective workareas using the "Consolidation" option available in SWDES. Each incremental dataset in the form of transfer workarea when consolidated to the existing workarea at DDPCs makes the replicas up-to-date vis-à-vis the status of the respective workarea at the SDDPCs.

Consolidation option is available as a button on the tool bar of the main switch board. The form for consolidation option is as shown in Figure 8.2. For consolidation the receiving workarea (into which the transfer workarea is being merged) must be active. Then the transfer workarea (containing incremental data set) as received from the SDDPC is selected for merger (consolidation) into the corresponding existing workarea at the DDPC. With mere one instruction as "Consolidate" the content of the transfer workarea is copied in the receiving workarea.
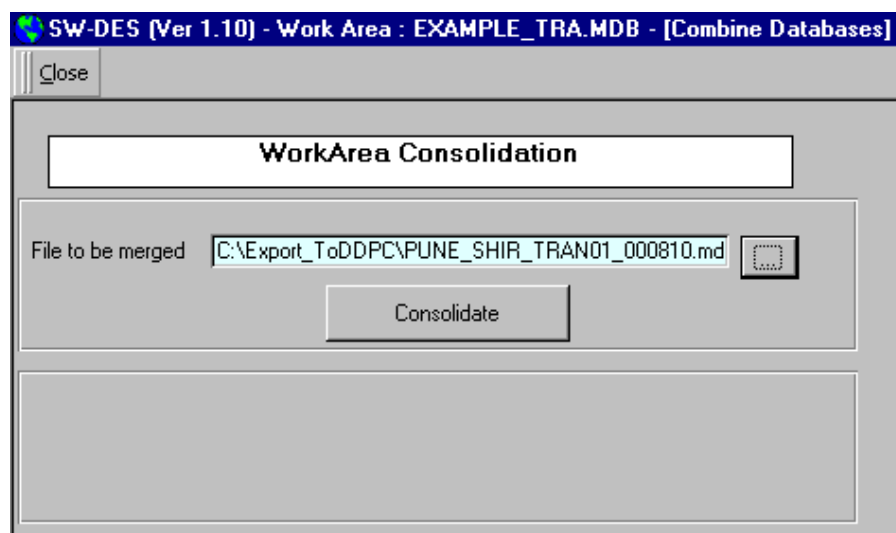

Figure 8.2:   Layout of the screen for consolidation of transfer work area into existing work area

In case there is only one workarea containing data of all the stations under the SDDPC then at the DDPC also there would be only one corresponding workarea. However, if at the SDDPC there are more than one workareas for different categories of stations (on account of a large number of stations under the SDDPC) then at the controlling DDPC also there would be equal number of workareas as replicas of the ones available at the SDDPC. In such a case there will also be one more workarea at the DDPC for combining individual workarea among the multiple workareas of the SDDPC. Such a combined workarea at the DDPC can also be maintained by similarly consolidating all individual incremental transfer workareas arriving from the SDDPCs. Thus in such cases consolidation of these incremental transfer workareas is required to be done twice: (a) once into the individual replica workareas and (b) into the combined single workarea for each SDDPC.

8.5   TRANSFERRING FIELD DATA FROM SWDES WORKAREAS TO THAT OF HYMOS

The incremental field data sets arriving every month from the SDDPCs are consolidated promptly at the DDPCs in the workareas for each SDDPC separately. These workareas contain all types of data for all the stations under the SDDPC belonging to the same distinct sub-basin of the independent river. From these workareas, data is regularly exported to an intermediate "transfer database" which is then used for importing into workareas of secondary module. Such transfer of data is regularly done on a monthly basis for incoming monthly incremental data. Otherwise such transfer of data can be made as and when there is a requirement to do so.

As mentioned earlier, there would usually be one workarea in the secondary module, at the DDPCs, pertaining to the same drainage area (of the independent river). Only in rare cases, it may have two or more workareas for distinct sub-basins/zones of independent river basin/zone. In these workareas, data is regularly imported from the intermediate "transfer databases", created by export of field data from SWDES workareas. This data would then undergo secondary validation and other necessary hydrological processing.

8.5.1   EXPORTING DATA FROM SWDES WORKAREA TO TRANSFER DATABASE

"Data Export to HYMOS - (New version)" option as available in SWDES is used for exporting the required data to transfer database. Any set of stations and data series can be chosen by the user in SWDES and data for any period (defined by setting the start and end dates) can then be exported to transfer database. The screen for the export of data to HYMOS is as shown in Figure 8.3. There are two types of data – static/semi-static and time-oriented data. These different types of data are grouped under suitable headings and also for various time intervals on the form. For every item there are three possible options: (a) to export "all" the data, (b) to export "none" of the data and (c) to choose a few ("custom") from the available list. Together with the selection on various data series the time period for which the data has to be transferred is also to be specified. All the exported data is put in one single data file called the transfer database. This transfer database is an MS Access database file with a fixed extension of ".mdb".
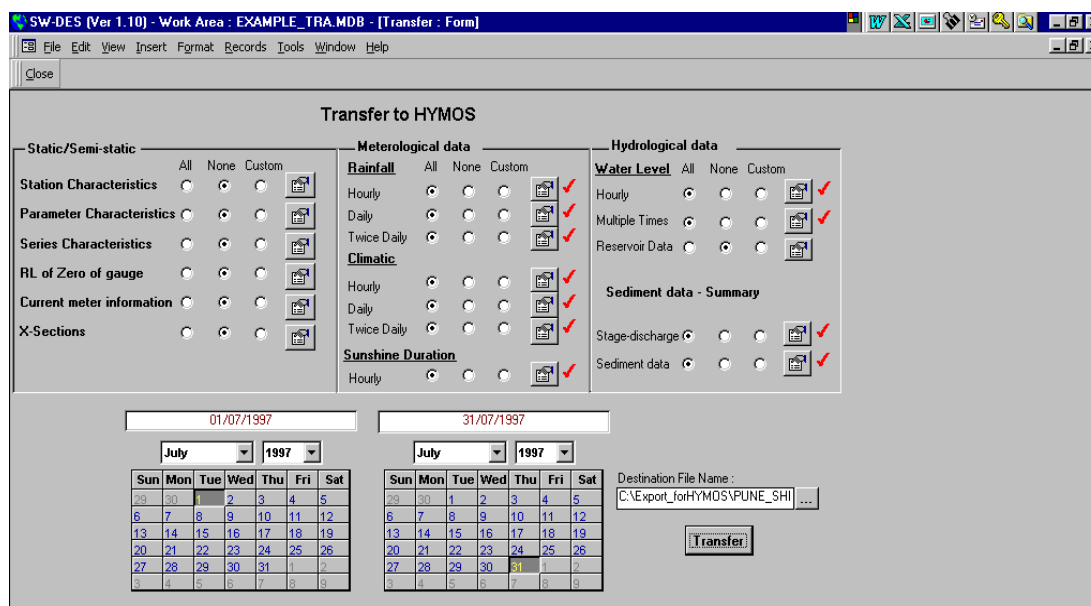

Figure 8.3:   Layout of the screen for exporting data from SWDES to HYMOS

As a practice these transfer databases would be made at the DDPCs (at SDPCs for WQ data) and would be imported in HYMOS databases at the same place and time. In such circumstances the names for these transfer databases can be any strings which the user may find suitable for recognition. However a uniform guideline if followed would make such filenames in a standard manner and can be beneficial in recognising any transfer database even after a lapse of considerable period. The filenames to be used for these transfer databases are recommended to be based on: (a) a string "EXPOnn", (b) yymmdd and (c) name of the parent workarea. Here "nn" in the string is any number used for adequately distinguishing different transfer databases of the same date and "parent" workarea is the one from which the transfer is being made.

8.5.2   IMPORTING DATA FROM TRANSFER DATABASE TO HYMOS DATABASE AT THE DDPCs/SDPCs

Import of data from the transfer database into the HYMOS databases is accomplished using one of the utilities "Import Transfer Database" available in HYMOS under "Managers" option. The opening window of this utility, illustrating import of transfer database to HYMOS database, is as shown in Fig. 4. The next screen shown as Figure 8.5 helps in identifying the transfer database which is tried to be imported by the user. Next a HYMOS database has to be specified, as shown in Figure 8.6, which is to receive data upon this import. For being able to import a transfer database it is required to specify the "import template", as shown in Figure 8.7, on the basis of which the import would be carried out. A default import template copied as "...|HYMOS 4\System\ImportTemplate.mdb" at the time of installing HYMOS can be used for import of data from SWDES exported transfer database to HYMOS database. It is also possible to use any other user-defined template in case the transfer database is not in the SWDES transfer database format. At this stage, a comparison is to be made of the contents of the source and the target databases. The reports of this comparison is also available for reference to the user. This is as shown in Figure 8.8. After reviewing the report on this comparison, the import is finally made, as shown in Figure 8.9, by appropriately taking care of the common data in the source and the target databases. The final report of the import operation is available at this stage to see what actions have been carried out during import.
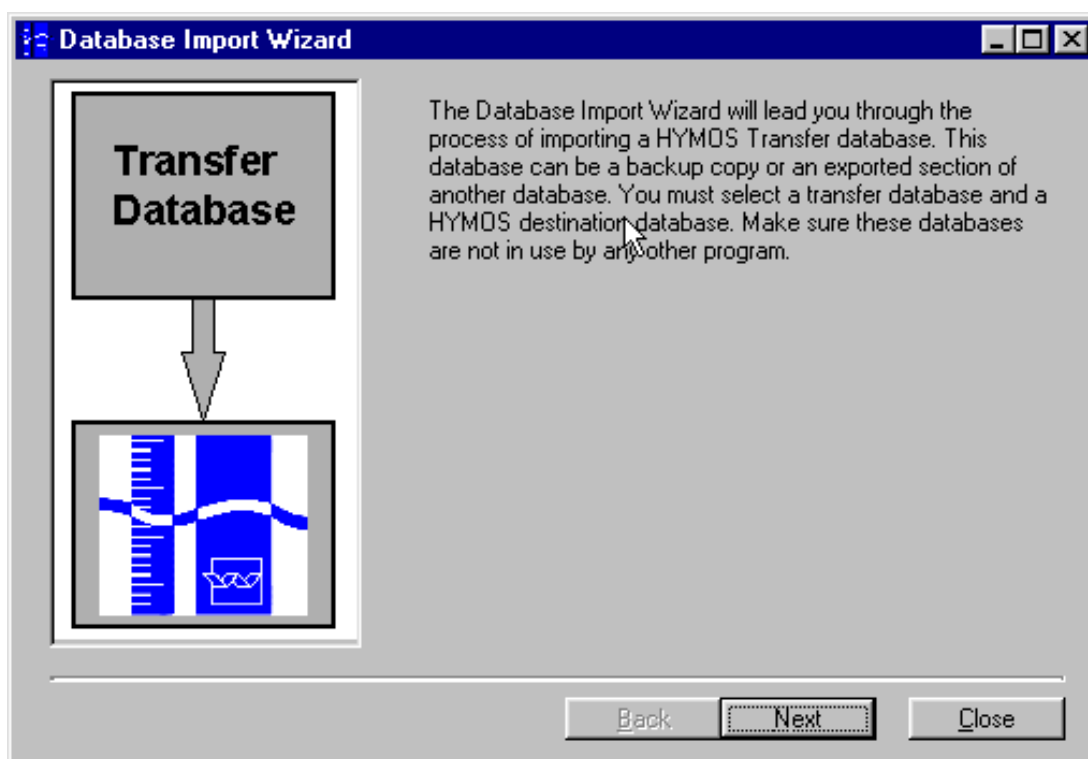

Figure 8.4:   Layout of the opening screen for import of data from transfer database to HYMOS
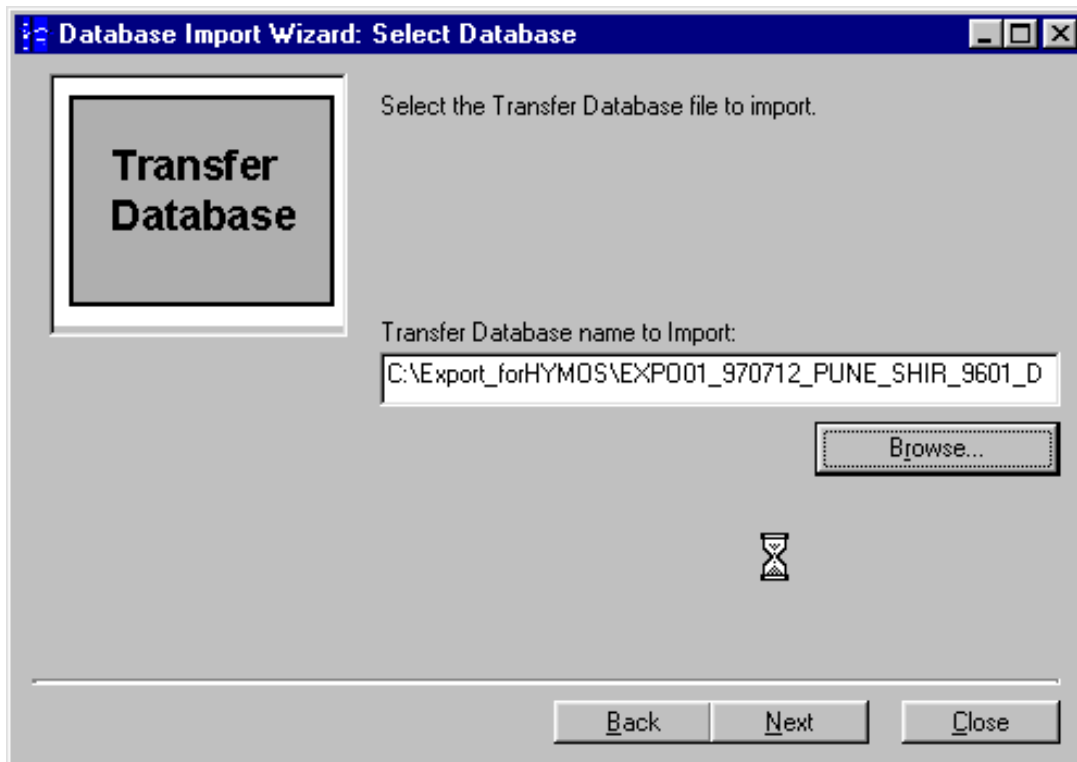
*Figure 8.5:   Layout of the screen for specifying transfer database during
                   import of data into HYMOS*
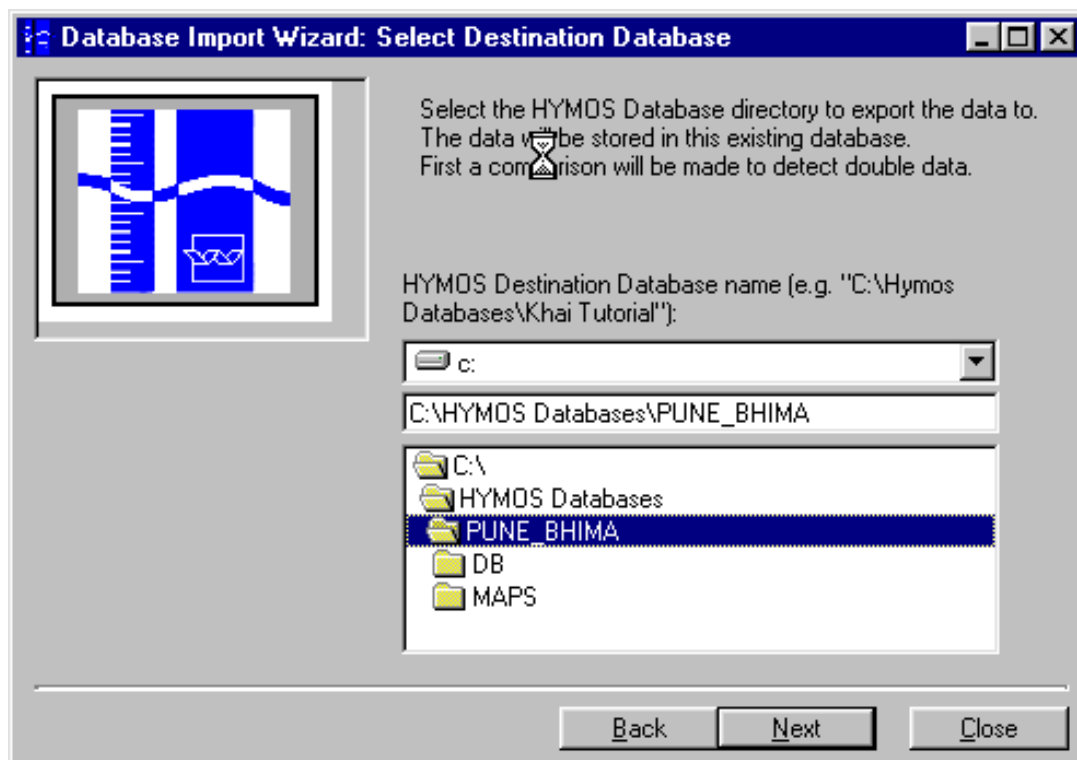

*Figure 8.6:   Layout of the screen for specifying the HYMOS database
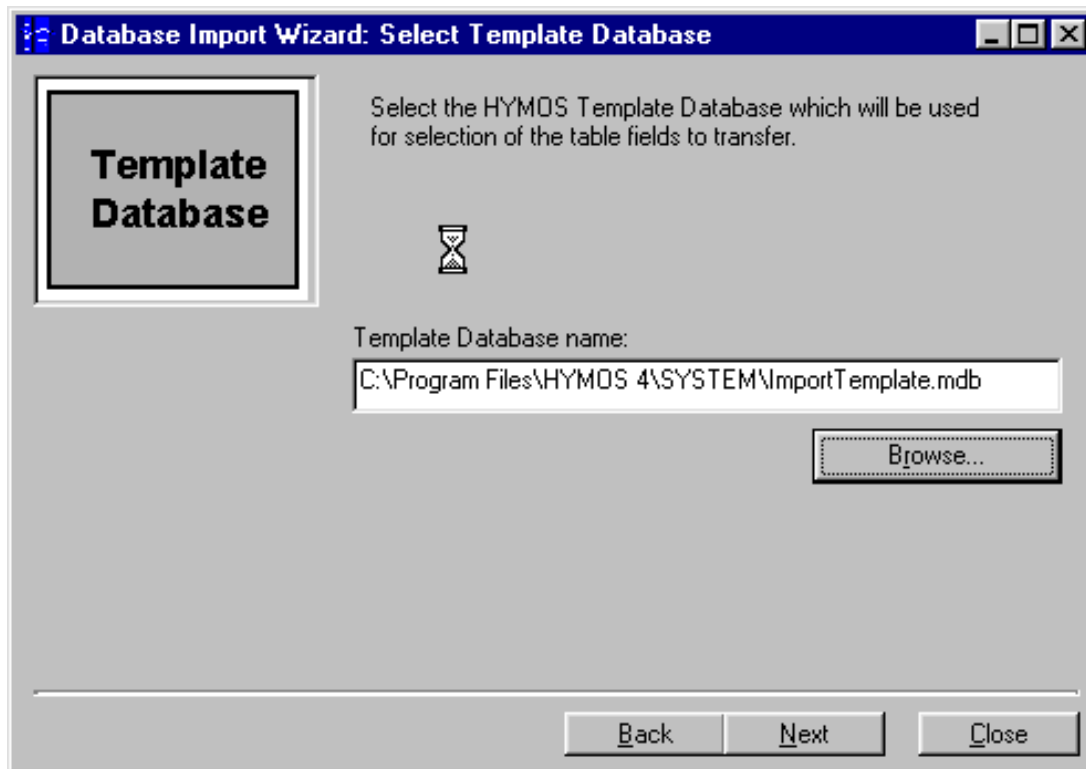   receiving data during import*

Figure 8.7:　　Layout of the screen for specifying the template to be used while importing data in HYMOS
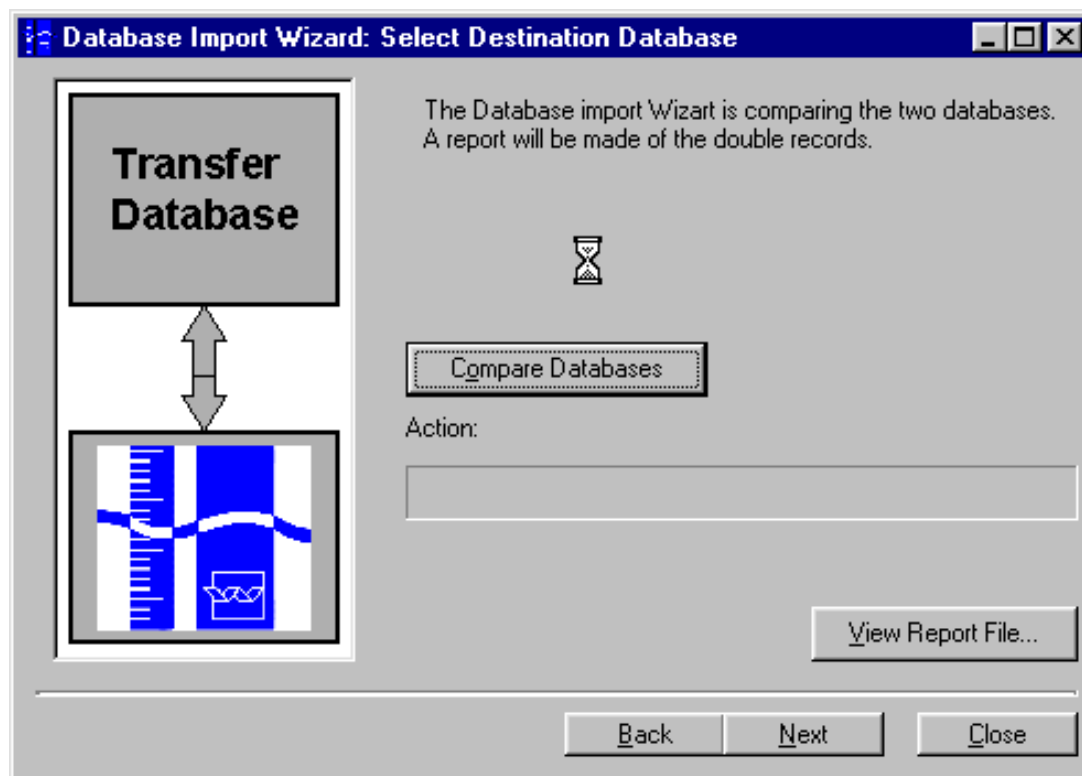

Figure 8.8:　　Layout of the screen for comparing the contents of transfer and HYMOS database
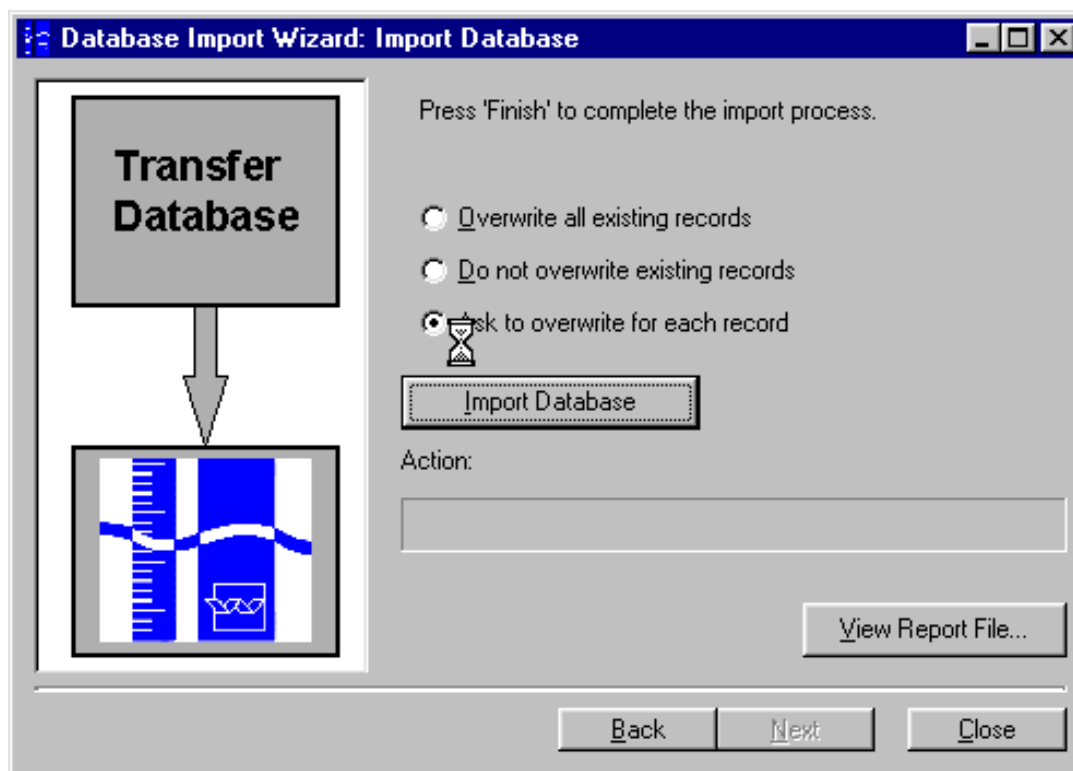
Figure 8.9:   Layout of the screen for finally importing data in the HYMOS database

8.6   TRANSFERRING FIELD DATA FROM DDPCS TO SDPCs/RDPCs

The field data as transferred and available at the DDPCs are transferred to the SDPCs/RDPCs for reference during hydrological validation and more importantly for onward transmission to the SDSCs/RDSCs for the purpose of archival. Field data from all the SDDPCs are first sent to the corresponding DDPCs and at each of the DDPCs these incremental data sets are consolidated in one single workarea for each SDDPC separately, as discussed earlier. From any DDPC, the field data from these unified workareas (for each SDDPC) is transferred to the controlling SDPC/RDPC by making a transfer workarea of incremental data available beyond the last such transfer.

Use of "Fragmentation" option available in SWDES is made for making such incremental transfer data sets at the DDPCs. Normally, as a regular routine monthly transfer of incremental data sets from SDDPCs would be available around 13th of each month. After consolidation of such incremental data in the unified SWDES workareas, this data is transferred to the corresponding SDPCs/RDPCs (by 15th of that month). Fragmentation option of SWDES is used for making such transfer workareas. The making of such transfer workarea is explained earlier in the text. Such transfer workareas are made separately for the workareas available for each of the SDDPCs.

A uniform guideline may be followed for assigning file names to such transfer workareas being sent from the DDPCs to the controlling SDPCs/RDPCs. This will be highly beneficial in recognising the transfer workarea by merely knowing its filename. The filenames to be used for these transfer workarea are recommended to be based on: (a) a string "TRANnn", (b) yymmdd and (c) name of the parent workarea as used at DDPC. Here "nn" in the string is any number used for adequately distinguishing different transfer workareas of the same date and "parent" workarea is the one from which the transfer is being made.

Thus the data required to be transferred from a DDPC to the SDPC/RDPC can be prepared as explained above. As mentioned earlier, the actual communication of these transfer workarea files to

the SDPCs/RDPCs can be through any of the links prescribed under the Hydrology Project such as using physical media (preferably on CD), data transfer networks, telephone line etc.

At SDPCs/RDPCs, SWDES will be used for consolidating such incremental transfer field datasets into respective SWDES workareas for each SDDPC of the state or region separately. "Consolidation" option is to be used to simply merge the contents of the in-coming incremental transfer dataset into the respective existing SWDES workarea for the each SDDPC.

8.7 TRANSFERRING PROCESSED DATA FROM DDPCs TO SDPCs/RDPCs

The incremental processed data sets validated every month at the DDPCs are promptly transferred (by 30th of every month) to the corresponding SDPC/RDPC for further hydrological and inter-agency validation and thereafter for submission of authenticated data to the SDSC/RDSC for dissemination and archival. As mentioned earlier, there would usually be one workarea in the secondary module, at the DDPCs, pertaining to the same drainage area (of the independent river). Only in rare cases, it may have two or more workareas for distinct sub-basins/zones of independent river basin/zone. From these workareas, data after necessary secondary validation, is regularly exported to an intermediate "transfer database" which is then sent to the SDPC/RDPC. Such "transfer databases" would then be used at SDPCs/RDPCs for importing the data in the corresponding HYMOS workareas.

8.7.1 EXPORTING DATA FROM HYMOS WORKAREA TO TRANSFER DATABASE AT THE DDPCs

Export of data from HYMOS databases to the intermediate transfer database is accomplished using one of the utilities "Export Transfer Database" available in HYMOS under "Managers" option. The opening window of this utility, illustrating export of transfer database from HYMOS database, is as shown in Figure 9.10. The next screen, as shown as Figure 8.11, helps in identifying a HYMOS database from which the export is to be made. Next the transfer database which is tried to be created by the user during this export is specified, as shown in Figure 8.12. For being able to export data to a transfer database it is required to specify the "export template", as shown in Figure 8.13, on the basis of which the export would be carried out. A default export template copied as "...|HYMOS 4\System\ExportTemplate.mdb" at the time of installing HYMOS can be used for export of data from a HYMOS database to a transfer database. It is also possible to use any other user-defined template in case the transfer database need to be in a different format. At this stage, the user can select the type of data which are to be exported, as shown in Figure 8.14. And following this step, the user is to specific the stations, from the total list of stations as shown in 8.15, for which the data is to be exported. Together with the stations to be exported, the time period between which the data is to be exported is specified, as shown in Figure 8.16, by selecting the required start and end dates. In the last, the export is actually carried out by pressing "Export Database", as shown in Figure 8.17. The report of the export available on the same screen, stating what is exported and what is not exported, can be reviewed at this stage for reference. This completes the export of desired data from a HYMOS database to a transfer database.
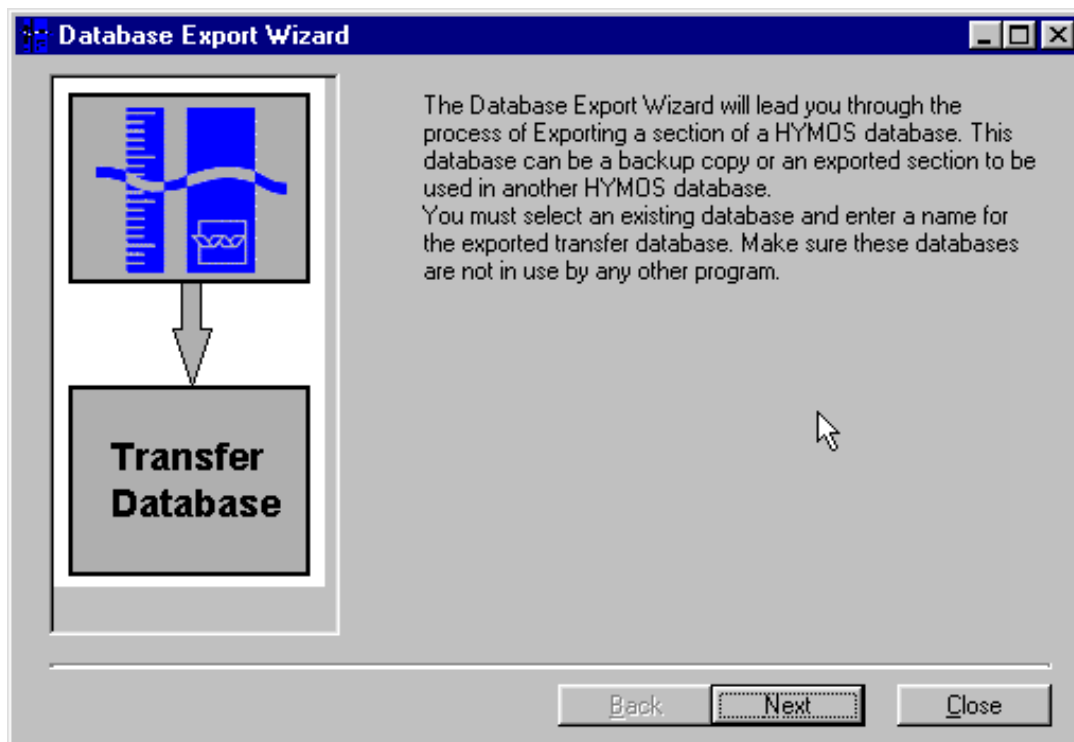
Figure 8.10   Layout of the opening screen for export of data from HYMOS to transfer database
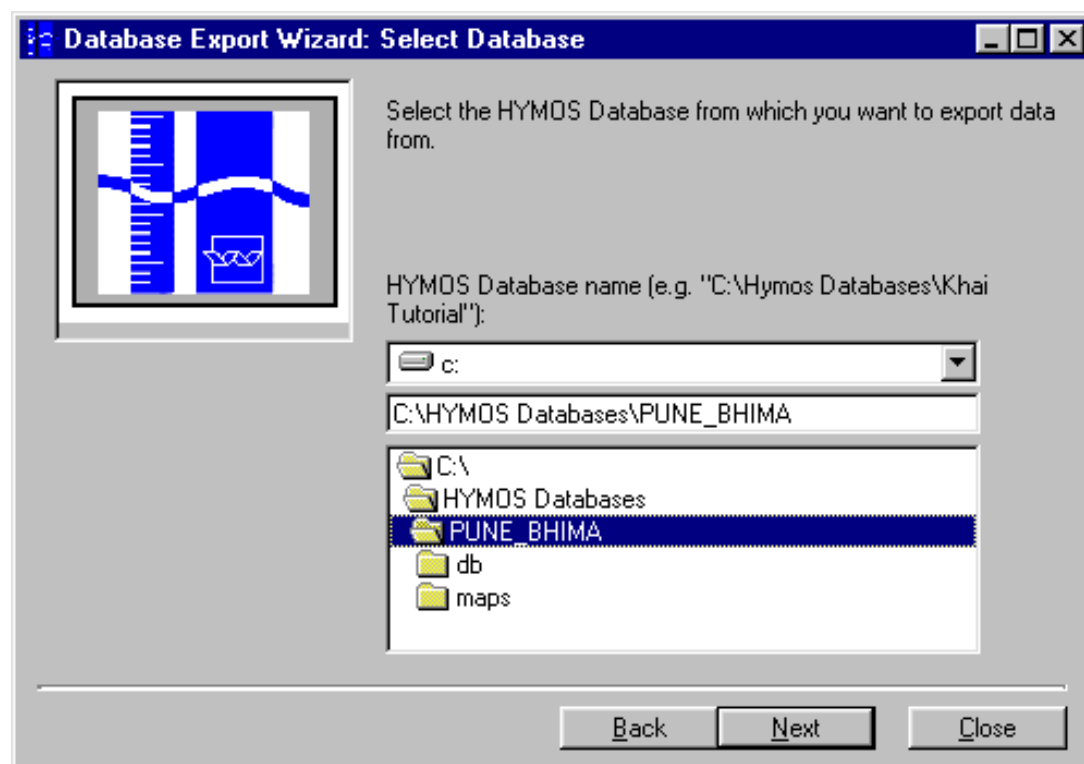
*Figure 8.11:   Layout of the screen for specifying HYMOS database during
                    export of data into transfer database*

Figure 8.12:   Layout of the screen for specifying the transfer database to be created data during export

Figure 8.13:   Layout of the screen for specifying the template to be used while exporting data from HYMOS

*Figure 8.14:   Layout of the screen for choosing the data types to be exported
                    to the transfer database*

Figure 8.15:   Layout of the screen for selecting the stations for which the data is to be exported

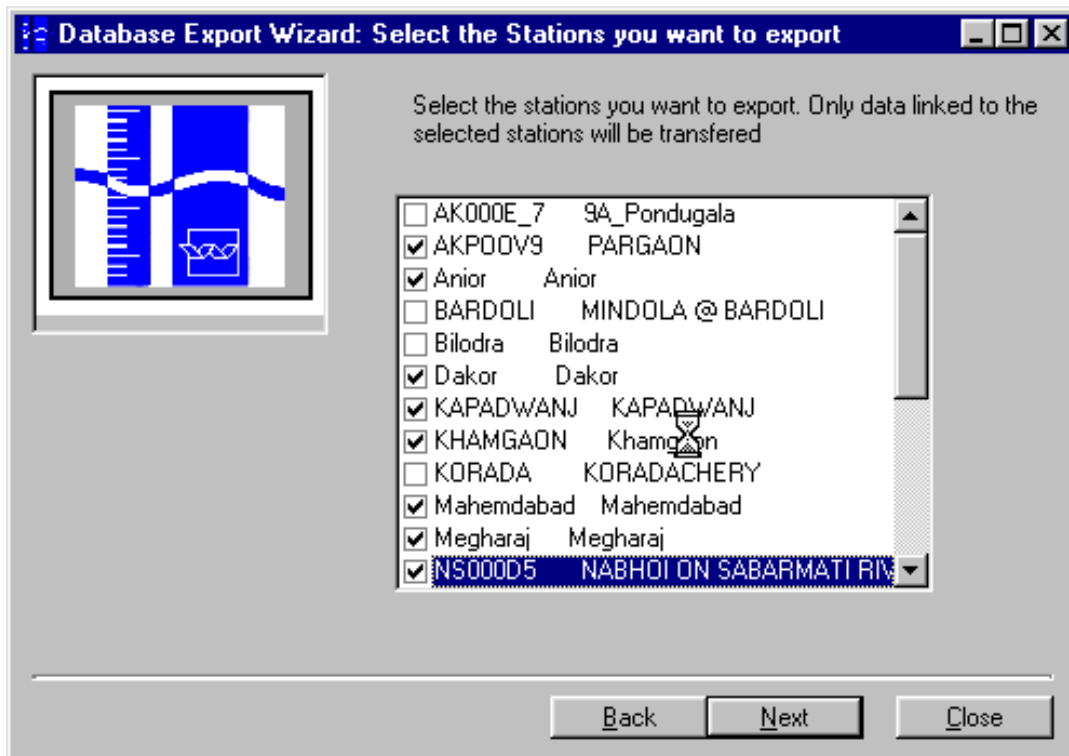*Figure 8.16:   Layout of the screen for specified the period for export of data to the transfer database*
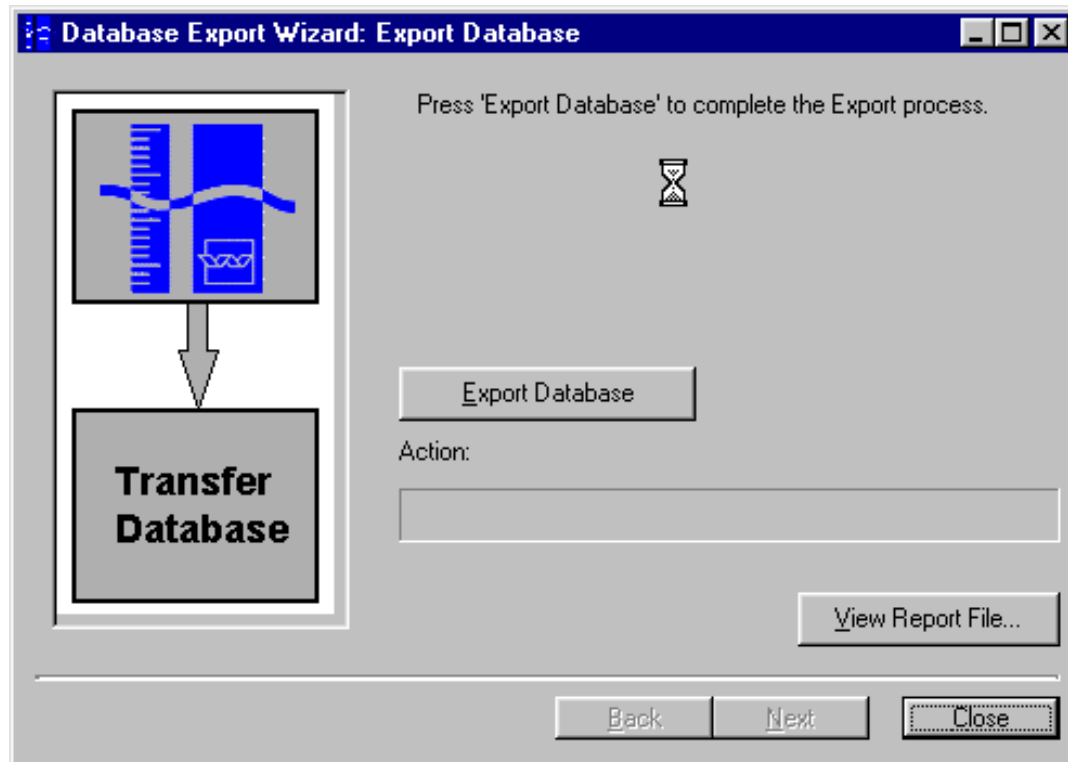

Figure 8.17:   Layout of the screen for finally exporting the selected data to the transfer database

As a practice these transfer databases would be made at the DDPCs and would be sent to the SDPC/RDPC. At the SDPC such incremental datasets would be appropriately imported in the relevant basin's workarea and the dataset would be further validated and finalised. A uniform guideline is to be followed for naming these transfer databases for an easy recognition. The filenames to be used for these transfer databases are recommended to be based on: (a) a string "EXPOnn", (b) yymmdd and (c) name of the parent workarea. Here "nn" in the string is any number used for adequately distinguishing different transfer databases of the same date and "parent" workarea is the one from which the transfer is being made.

8.7.2   IMPORTING DATA FROM A TRANSFER DATABASE TO HYMOS DATABASE AT THE SDPCs/RDPCs

Since HYMOS databases at DDPCs and SDPCs/RDPCs are based on the basin/zone or sub-basin concept and the former is generally a subset of the later, the transfer dataset(s) received from the DDPCs are to be just imported in the relevant HYMOS database. More often one or more DDPCs transfer databases would be the part of the same HYMOS database encompassing the basin area covered by those DDPCs. The procedure for import of these transfer databases into the HYMOS databases is exactly same as explained previously in section 3.2 except for the fact that these transfer databases have originated from HYMOS databases at DDPCs instead of from SWDES workareas at SDDPCs.

9   DATA TRANSFER FROM TEMPORARY DATASETS TO PERMANENT DATASETS

9.1   TRANSFERRING FIELD DATA FROM SDPCs/RDPCs TO SDSCs/RDSCs

Under HIS the data processing and the data storage functions are separated. The field and processed data sets are to be archived at various data storage centres (DSCs) of state and central agencies. The availability of field data at the DSCs, as **permanent databases**, is purely for the purpose of long term archival and for limited dissemination for specific purposes, if required. The field data as prepared in SWDES workareas at various SDDPCs is routed through the controlling DDPC and is finally available at the SDPC/RDPC. A copy of the same is also kept at the DDPCs for ready reference at the time of data validation. At DDPCs and SDPCs/RDPCs all the field data is available in distinct SWDES workareas for each SDDPC separately in one single file.

The field data available at SDPCs is in the temporary databases of SWDES. This data is to be transferred to the permanent database(s) of the respective DSCs using dedicated data storage software (under procurement process). The databases in the DSCs will make use of the uniform data structure in an Oracle database environment to be compatible with the selected database system for the dedicated groundwater data processing system (presently under procurement) & surface water data processing system (HYMOS). This data storage application software will be developed to carry out all required database functions in all SDSCs/RDSCs/NDSCs. The data storage software will be standard, uniform and generic in nature having all necessary auxiliary features like user authentication and management, access control, data integrity and encryption etc. Though the detailed specifications are under development and its subsequent procurement and implementation will take some more time from now (i.e. January 2001), it is attempted here, in the following paragraphs, as how the transfer of the field data from the side of the SDPCs/RDPCs would be made. Though the exact procedure of import of these transfer data into the permanent database(s) by the data storage system would be known only after the system is fully defined. However, such import functionality would obviously have all the standard features required for import of new data streams like sender's authentication, data comparison, data integrity etc. No attempt, except for a brief mention, is therefore being made here to further elaborate on this aspect.

9.1.1   EXPORT OF FIELD DATA FROM SDPCs/RDPCs TO TRANSFER DATABASES

The incremental SWDES transfer datasets, containing field data, arriving at SDPCs/RDPCs from the DDPCs are used for two purposes: (a) transferring field data from SDPCs/RDPCs to respective SDSCs/RDSCs and (b) regularly consolidating in respective SWDES workareas for use at SDPCs/RDPCs for reference.

Normally, as a regular routine monthly transfer of incremental data sets from SDDPCs would be available at DDPCs by around 13th of each month. From the DDPCs, similar incremental databases are transferred to the corresponding SDPCs/RDPCs (by 15th of the same month). After the receipt of these incremental transfer databases, they are transferred to the SDSC/RDSC by around 20th of every month. By the same time consolidation of these transfer workareas is also made in the respective SWDES workarea for each SDDPC separately.

For transfer of incremental field datasets as received from the DDPCs to the DSCs, first the workareas are registered in SWDES and then using the option of "Export of data to HYMOS (in ".mdb" format)" is used to make an intermediate transfer dataset. This transfer dataset will be in MS Access database structure and would be used for import into the permanent databases by the data storage system.

A uniform guideline may be followed for assigning file names to such transfer workareas being sent from the SDPCs/RDPCs to the respective SDSCs/RDSCs. This will be highly beneficial in recognising the transfer workarea by merely knowing its filename. The filenames to be used for such transfer workareas are recommended to be based on: (a) a string "TRANnn", (b) yymmdd and (c) name of the parent workarea as used at SDPC/RDPC. Here "nn" in the string is any number used for adequately distinguishing different transfer workareas of the same date and "parent" workarea is the one from which the transfer is being made.

9.1.2   IMPORT OF FIELD DATA FROM TRANSFER DATABASES INTO SDSCS/RDSCS

The transfer databases prepared by the SDPCs/RDPCs pertaining to each SDDPC separately, for the period beyond the last such transfer, are imported by the data storage system available at DSCs. The details of the import procedure would be available only when the design of the system being procured is fully known and hence is not further elaborated here.

9.2   TRANSFERRING PROCESSED DATA FROM SDPCs/RDPCs TO SDSCs/RDSCs

At the SDPCs/RDPCs, the whole state or a very large drainage region is under the jurisdiction. The state would include parts or full of one or more independent river basins whereas the regions of CWC being based on the river basins would include one or more complete independent river basins. There are two types of workareas for processed data at the SDPCs/RDPCs: (a) workareas for the individual independent river basins or part thereof within the state and (b) unified workarea for data of the whole state/region together (having only finalised/summary data).

Transfer of data between SDPCs/RDPCs and the respective SDSCs/RDSCs will be accomplished, as a routine, from the workareas maintaining data of individual independent river basins or part thereof.. Since these workareas would have all the detailed authenticated data, these will be used as the nodal databases for exchange of authenticated data from SDPCs/RDPCs to the respective SDSCs/RDSCs. The processed data from DDPCs are available at SDPCs/RDPCs by the 1st of the second month after the month of observation. At SDPCs/RDPCs hydrological validation will be carried out and also data would be compiled in various forms as authenticated data. Such processing would be over by the end of the month and at that time (i.e. by 30th of the second month after the month of observation) these finalised data are transferred to SDSCs/RDSCs.

For this, transfer databases from each of the available individual workareas in HYMOS will be made. Such transfer data shall also be prepared by using the "Export transfer database" option in exactly same manner as explained earlier in the sections on "transfer of processed data between temporary databases".

A uniform guideline can be followed for naming these transfer databases for an easy recognition. The filenames to be used for these transfer databases are recommended to be based on: (a) a string "EXPOnn", (b) yymmdd and (c) name of the parent workarea. Here "nn" in the string is any number used for adequately distinguishing different transfer databases of the same date and "parent" workarea is the one from which the transfer is being made. Obviously the details of naming of such transfer databases is dependent on the type of protocol with the data storage system and therefore may differ from what is prescribed here on the basis on actual details of the system. However, idea of a uniform guideline for naming such transfer file is generic and that the basis may differ due to actual requirements.

As soon as these transfer databases are made available to the SDSCs/RDSCs, these will be imported in the permanent databases after all necessary requirements of protocol like sender's identification, authorisation, data comparison, integrity etc. are fulfilled. The details of such import options are not discussed here as the system is yet to be fully defined and developed.

10   WORKING WITH HYMOS IN A NETWORK ENVIRONMENT

10.1   NETWORK ENVIRONMENT AT THE SPDCs/RDPCs

All data processing activities under Hydrology Project is being carried out on computers with Microsoft Windows 95/98/2000/NT as the operating system depending on the type of Data Processing Centres and the time at which the systems were bought. All the SDPCs/RDPCs will essentially have Windows NT or equivalent environment as about 4 to 6 individual computers would be operational at these centres together with attached GW Data Processing Centre and/or the State Data Storage Centre. Since this would be a substantial set-up it is appropriate to have all the computers and peripherals connected with the network so that the operations can be smooth and efficient and very importantly to optimise the resources used for printing, backing up etc.

In such a situation of networked environment it is very important to establish the protocols and procedures for working in general and for processing the data using HYMOS in particular. If such procedures are not evolved and adhered to then it may result in confusion, duplication of data files and in all an inefficient working atmosphere. Few things which are extremely important are: (a) availability of well-defined databases on the network, (b) specified user profiles and controlled accessibility to various databases, (c) rights to various users on read, write, deletion activities, (d) transfer of data from temporary databases to permanent databases and vice-versa and (e) maintenance of fragmented SWDES databases & HYMOS transfer databases received from the Divisional Data Processing Centres. Following sections outlines these important aspects.

10.2   CONTROLLED ACCESSIBILITY TO HYMOS DATABASES

Since there will be more than one officer designated for processing the data at the SDPCs/RDPCs, it is appropriate that every basin in the state or region is earmarked to one officer (Hydrologist). In case there are more basins, one officer may look after the processing of data of more than one basin. It may however be required that one more person is attached with the main person to ensure adequate support in case of absence of the main person. That is to say that there would be one First Officer and one Second Officer responsible for all the data processing within any database. The Second Officers may be working as First Officer for another database, if so required. The primary responsibility of processing of data in any database would be of the First Officer but in periods of long absence or if additional support is required the Second Officer will also work on that database. The idea is that the

First Officer will generally be take the primarily responsibility to finalise the data assigned and the Second Officer would be available for additional support, if required.

Apart from the individual river basin databases, there will also be one common database for the whole State in the case of the SDPCs. This unified database is simply the aggregation of various individual basin's databases, within the State. Responsibility for maintaining such database is also to be given to one of the First Officer (the Hydrologist) working at the SDPC. This database is meant primarily to consolidate the data finalised in different individual basin databases at one place. This is basically to enable compilation of information for the whole state or region and for

As such all the Officers designated for particular databases will have well-defined User Profiles authenticating them to work with various databases. The databases will also be protected against deletion or editing from outside the HYMOS

10.3   DATABASE RIGHTS TO VARIOUS USERS

Individual river basin HYMOS databases will normally be under the controlled of the designated First officer and if required also the Second Officer. The databases must therefore be fully accessible by these two hydrologists for all data organisation, validation, correction, completion, compilation and reporting features. All the data pertaining to the individual basin databases would be finalised by the First Officer with help from Second Officer, if required. Any other user should not be given any editing rights unless specifically required. The other hydrologists working on neighbouring river basins can be given rights to view or export data for the purpose of reference and use in their own databases.

For maintaining the combined or unified database for the state, the responsibility of exporting the authenticated data from the individual river basin databases must lie with the respective First Officers while its import into the unified database must be done by the Officer maintaining it. The rights for editing and processing of data in the unified database must only be with the Officer who maintains it. Other hydrologists working on individual river basin databases must have the rights to only view or export the data from it.

10.4   TRANSFER OF DATA FROM TEMPORARY DATABASES TO PERMANENT DATABASES AND VICE-VERSA

It is required that the responsibility of transfer of all the authenticated (finalised) data from the temporary databases of individual river basin databases to permanent databases (in Data Storage Centre) is given to the First Officer of the respective databases. Thus they would be directly responsible for any data being sent to the data storage centre for permanent archiving. Similarly, standing permission to retrieve data from the data storage centre for the purpose of reference must be available to the First officer.

10.5   MAINTENANCE OF FRAGMENTED SWDES DATABASES & HYMOS TRANSFER DATABASES

Fragmented SWDES databases & HYMOS transfer databases will arrive every month from various DDPCs. The monthly incremental field data in the form of fragmented SWDES databases will have to be consolidated in the respective SWDES database of each SDDPC. Similarly, the monthly incremental processed data in the form of HYMOS transfer database pertaining to individual river basin or part thereof would have to be imported in the respective river basin's database. However, it is very essential to keep a strict procedure to store all these incremental data streams arriving from various DDPCs at a pre-defined location. Proper directory structure has to be maintained for storing all the data files received every month in a categorised manner. Two main categories of folders which may be used for this purpose can be (a) FRAG_DB and (B) TRANS_DB. Under these folders, one more level can be for the year and the month to which the databases belong, like – 2001_JAN,

2001_FEB etc.. The fragmented and transfer databases can then be arranged under these folders. Since the fragmented and transfer database files from every SDDPC would bear proper identification, there will not be any problem to reach to any desired file at moment of time. Contents of all such folders must also be adequately protected from any tempering from an external user.

11   BACKUP OF HYMOS DATABASES

11.1   BACKUP AND ITS ADVANTAGES

Any information stored on the computer or on a disk is vulnerable to damage or loss for a variety of reasons like theft, fire, wear and tear, computer viruses, power failure, magnetic fields etc. One of the major threats is the user himself – a single wrong command can destroy months of effort. Therefore, backup of data becomes critical and has to be performed regularly. A backup routine has not been provided as in integral part of either the SWDES or the HYMOS system because of the number of options available in the WINDOWS environment and the extremely easy availability of general purpose software in the public domain.

Various levels of backup usually performed are:

Local backup: The first level of backup can be kept on the same computer, preferably in a different partition.

On-line backup: The second level of backup can be kept on the hard disk of another computer, if connected on a network.

Off-line backup: This is the most important backup and serves as the permanent archive of data. This could be on floppies, CD's, magnetic tapes or other devices, depending on the hardware configuration of the computer and the facilities available in the data centre.

Off-line incremental backup: This option is useful for very large volumes of periodic data. In this procedure only that information which has changed after the last backup is backed up and thus saves lot of space.

Off-line backup can be taken on floppies, CDs., cartridge tape or DAT and JAZZ drives etc.. Typically, the capacity of floppies is 1.44 MB, of CDs is 650 MB and of DAT cartridges is between 2 and 24GB. Backup of a large file(s) (say upto about 6-8 MB) can be spanned on multiple floppies, if required. However, since floppies normally are more prone to corruption and become unreadable frequently, backup of information on a set of multiple floppies is difficult to work with and thus is not preferred. It becomes quick and easy though to backup smaller file(s) on one or two floppies.

CDs and tapes on the other hand have substantial storage space and are preferred methods for backing up large amount of information. The availability of such specialised backup tools and the higher technical skill required for operation and maintenance may limit their application in some cases.

Backup on CD's, cartridge tapes etc. will be done using the software made available with the backup device. All procedure for CD-ROM drives, cartridge tapes, ZIP drives, DAT and JAZZ drives are different and equipment specific, and thus would not be conducive to a system based back-up protocol.

It is self defeating to take backups on the same piece of physical removable media cycle after cycle. This leads to the following problems:

- Corruption of data because of media failure : Loss of previous backup also
- Loss of backup in case undetected virus in current backup
- All corruption, errors in previous data made in the current cycle transmitted to the backup

To safeguard against these problems, it is suggested that the Grandfather..father..son technique is used, whereby backups are taken cyclically on three different sets of media. This ensures that at any point of time, at least two different secure backups are always available. It also provides for redundancy in the backup system and for checking on data sanctity and validity.

Restoring Data: The backup system followed will always have a corresponding restore utility. In case a restore is required it must be ensured that it is done such that no existing valid data is over-written.

Procedure to be followed for backing up surface water data as available in SWDES and HYMOS databases at various Data Processing Centres is discussed in the following sections. However, similar policy can be followed for any other important data being worked with at these offices which obviously also needs backing up at suitable time interval.

11.2   EMPLOYING COMPACTION AND COMPRESSION BEFORE BACKING UP OF DATA

The databases grow each time data is added to them. However, it is important to note that when data is deleted, they do not become smaller automatically. This is done as a standard practice in most **RDBMS** and other databases to provide faster response times. Thus, it becomes important to 'compact' the database periodically as the database keeps deleting its own temporary and work data. Compaction reduces the size of the database and will therefore help in creating smaller backups. It is therefore recommended that a Microsoft Access database be compacted frequently, especially prior to a backup process.

Another tool commonly available is data compression. Compression reduces file sizes for the purpose of storage by "packing" the data in a smaller space. Since a compression process reduces the size of storage required for backup it is always useful when the amount of data to backed up is larger and at the same time the receiving media is having relatively lesser space. Such is case when the backup of few moderately sized (say 6 to 8 MB) files is attempted to taken on the floppies. There are several utilities available for compressing the files such as WINZIP, Microsoft backup and other compression utilities like LHA, RAR etc. Most of these utilities also support simultaneous process of compression and backup functions (with backup spanning on multiple floppies).

While taking backups on floppies it is very important to ensure that all backup on floppies are tested by copying the files back into a temporary directory after ejecting and reinserting the floppy diskettes!!!

11.3   BACKUP POLICY AT SDDPCs

SDDPCs will routinely work with a maximum of three SWDES workareas (in many cases there will be only one workarea) only. Each SWDES workarea has only one physical file (with ".MDB" as the extension) that contains all the data; i.e.; each work-area file is complete in itself. Only where the DWLR data would also be dealt, there will be an additional file of the same name as the workarea itself but with ".MDD" as the extension.

Thus the regular work at the SDDPCs will be on few SWDES workarea files. Since these files would contain data for about 4 to 5 years, they will be about 8 –10 MB in size. Compaction and compression can bring them to the size of one or two floppies.

Off-line back up at SDDPCs: Since the data files will be of comparatively smaller size, it might be a simpler and quicker option to take an off-line back up the data on floppies. Such off-line back up is recommended to be taken every 10 days. It is appropriate to fix the dates of such back ups as 1st, 11th and 21st of every month or the next working day in case it happens to be a holiday. In case the staff at any SDDPC feels well acquainted with taking back up on CDs then it can be the preferred medium as there will be less chances of corruption etc.

On-line back up at SDDPCs: As a second level of safety, it is appropriate to keep the copies of content of datafiles on the second computer available at the SDDPCs. This can be easily done by using the same set of back up taken on floppies or CD, every 10 days, to restore on the second computer. This would incidentally also check if the floppies are working properly.

11.4   BACKUP POLICY AT DDPCs

At DDPCs, both SWDES and HYMOS databases will be in use. SWDES databases are as received from SDDPCs and thereafter consolidated into unified databases for each SDDPC. HYMOS database(s) will be operational of each of the river basin(s) or part thereof within the jurisdiction of the DDPCs. Normally, a DDPC will have one HYMOS database but in some cases it may be 2-3 as well. There will also be availability of "HYMOS transfer databases" as data from DDPCs HYMOS database will be transferred regularly to the SDPCs/RDPCs.

Thus four types of databases available at DDPCs, on a regular basis, are: (a) SWDES fragmented databases as received from SDDPCs for incremental data for each month, (b) unified SWDES databases for each SDDPCs in which all incremental SWDES databases are consolidated, (c) HYMOS databases and (d) incremental processed "HYMOS transfer databases" as sent to SDPC/RDPC every month. Databases of the first and the last types are not important to be backed up as the information is already available in the unified SWDES databases or the HYMOS databases.

It is desired that all the four types of databases be organised in well-defined folders. For example, all the SWDES fragmented databases received from various SDDPCs every month can be organised in a folder called "FRAG_SWDESDB". Under such folder the databases can be organised under separate directories for each SDDPC separately. Similarly, all the SWDES unified databases for each SDDPC would available in the SWDES program directory. All HYMOS databases will similarly be available under the "HYMOS Databases" directory with one more sub-directory level to identify different databases. All the HYMOS transfer databases, exported from DDPC and to be sent to SDPC/RDPC, can be stored in a prescribed directory as "TRAN_HYMOSDB". Databases of the first and the last types are not important to be backed up as the information is already available in the unified SWDES databases or the HYMOS databases respectively.

Off-line back up at DDPCs: It would be very convenient to back up the contents of the SWDES and HYMOS databases from the two directories on the CDs as the amount of data would be difficult to handle by using floppies. Regular off-line backups of these databases at the DDPCs is recommended to be taken regularly at a ten daily interval as in the case of SDDPCs.

On-line back up at DDPCs: In case the computers available at DDPC are interconnected with the help of network or other link, it would be useful and easy to copy the contents of the four folders on to the other computer. If SWDES and HYMOS work is carried on 2-3 computers available at the DDPC then such copying from one computer to another computer must be done for all computers so that every computer keeps a back up of useful data of other computers.

11.5   BACKUP POLICY AT SDPCs/RDPCs

At SDPCs/RDPCs there will be availability of full-fledged network and also 4-6 computer nodes working with comparatively larger amounts of data.

At SDPCs/RDPCs also, both SWDES and HYMOS databases will be in use. SWDES databases are as received from DDPCs and thereafter consolidated into respective databases for each SDDPC. HYMOS database(s) will be operational of each of the river basin(s) or part thereof within the State. There will also be availability of "HYMOS transfer databases" as data from DDPCs will be received regularly at the SDPCs/RDPCs.

Four types of databases available at SDPCs/RDPCs, on a regular basis, are: (a) SWDES fragmented databases as received from DDPCs for incremental data for each month, (b) unified SWDES databases for each SDDPCs in which all incremental SWDES databases are consolidated, (c) HYMOS databases and (d) incremental processed as "HYMOS transfer databases" as received from DDPCs every month. Databases of the first and the last types are not important to be backed up as the information is already available in the consolidated SWDES databases or the HYMOS databases.

On-line back up at SDPCs/RDPCs: As all the computers at SDPC/RDPC will be connected with the network it would be useful and easy to take a weekly backup of important data on all the computers on the server computer.

Off-line back up at SDPCs/RDPCs: As the data at SDPC/RDPC will be very important, it is essential to keep a off-line back up on CDs or DAT drives as well. As a regular activity after the on-line back up is taken it is appropriate to ensure the off-line backups as well. The frequency of the off-line backups must be the same as on-line backup i.e. weekly.