HYDROLOGY PROJECT

Technical Assistance

World Bank & Government of The Netherlands funded

Training module # SWDP - 03

Understanding Surface Water data processing plan

New Delhi, November 1999

CSMRS Building, 4th Floor, Olof Palme Marg, Hauz Khas,
New Delhi – 11 00 16 India
Tel: 68 61 681 / 84 Fax: (+ 91 11) 68 61 685
E-Mail: dhvdelft@del2.vsnl.net.in

DHV Consultants BV & DELFT HYDRAULICS

with
HALCROW, TAHAL, CES, ORG & JPS

Table of contents

1. Module context

While designing a training course, the relationship between this module and the others, would be maintained by keeping them close together in the syllabus and place them in a logical sequence. The actual selection of the topics and the depth of training would, of course, depend on the training needs of the participants, i.e. their knowledge level and skills performance upon the start of the course.

2. Module profile

Title	:	Understanding Surface Water data processing plan
Target group	:	Assistant Hydrologists, Hydrologists, Data Processing Centre Managers
Duration	:	One session of 60 min
Objectives	:	After the training the participants will be able to*:* • Understand the data processing plan under HIS • Appreciate the concept of data processing at various levels • Know the computing facilities to be used at various levels • Know the type of staff required to carry out different activities
Key concepts	:	• Distributed and centralised hydrological data processing • Computing facilities and staff • Various levels of data processing • Time plan for processing of incremental data
Training methods	:	Lecture
Training tools required	:	OHS
Handouts	:	As provided in this module
Further reading and references	:	

3. Session plan

No	Activities	Time	Tools
1	***Introduction***: • Highlighted text and bullets - data processing levels • HIS structure at State/Regional level • Cartoon - Staff overwhelmed in a tide of data	5 min	 OHS 1 OHS 2 OHS 3
2	Distributed data processing - merits and demerits • Merits • Demerits	5 min	 OHS 4 OHS 5
3	***Data processing activities at various levels*** • At SDDPCs • At DDPCs • At SDPC/RDPCs	10 min	 OHS 6 OHS 7 OHS 8
4	***Computing facilities at various levels*** • Computing facilities - software • Computing options at various levels 1 • Computing options at various levels 2 • Computing facilities - Hardware	15 min	 OHS 9 OHS 10 OHS 11 OHS 12
5	***Availability of staff*** • At SDDPC & DDPC • At SDPC/RDPC	5 min	 OHS 13 OHS 14
6	***Time schedule for completion of activities*** • Data processing plan	5 min	 OHS 15
7	***Wrap up***	15 min	

4. Overhead/flipchart master

5. Handout

Add copy of Main text in chapter 8, for all participants.

6. Additional handout

These handouts are distributed during delivery and contain test questions, answers to questions, special worksheets, optional information, and other matters you would not like to be seen in the regular handouts.

It is a good practice to pre-punch these additional handouts, so the participants can easily insert them in the main handout folder.

7. Main text

Contents

Understanding Surface Water data processing plan

1. Introduction

In Modules 1 and 2 the structure of the HIS has been briefly described. In this Module, 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.

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. 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.

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 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.

Table 4.1: Availability of options in different types of packages.

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
	• Inspection of Longitudinal/Spatial Variation	No	Yes	Yes
	• Test on Relations	No	Yes	Yes
	• Double Mass Analysis	No	Yes	Yes
	• Nearest Neighbour Check	No	Yes	Yes
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

Features	Sub-features	Primary Module	Secondary Module	Full Package
	• 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

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.

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 raw 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 raw 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 raw 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 raw and fully processed and validated data will be held at the State/Regional Storage Centre
  - ❖ Raw 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.