HYDROLOGY PROJECT

Technical Assistance

World Bank & Government of The Netherlands funded

Training module # SWDP - 04

Understanding different types and forms of data in HIS

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

Page

1.	**Module context**	**2**
2.	**Module profile**	**3**
3.	**Session plan**	**4**
4.	**Overhead/flipchart master**	**5**
5.	**Handout**	**6**
6.	**Additional handout**	**8**
7.	**Main text**	**9**

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 different types and forms of data in HIS
Target group	:	Assistant Hydrologists, Hydrologists, Data Processing Centre Managers
Duration	:	One session of 60 minutes
Objectives	:	After the training the participants will be able to: • Know all the types and forms of data in HIS • Know the different variables in HIS
Key concepts	:	• Types of data • Forms of data • Different hydro-meteorological variables
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***: • Forms of data in HIS and its classification	5 min	 OHS 1
2	***Space-oriented data*** • Space-oriented data • Illustration - Catchment map • Illustration - cross sections	5 min	 OHS 2 OHS 3 OHS 4
3	***3.1 Time-oriented data*** • Classification of time-oriented data • Illustration - Equidistant time series • Illustration - cyclic time series • Illustration - non-equidistant time series	5 min	 OHS 5 OHS 6 OHS 7 OHS 8
	3.2 Meteorological data • Meteorological data 1a • Meteorological data 1b • Meteorological data 1c	10 min	 OHS 9 OHS 10 OHS 11
	3.3 Hydrological data • Water level and derived hydrological quantities • Illustration - Observed WL and computed dischare	10 min	 OHS 12 OHS 13
	3.4 Water Quality data • WQ data	5 min	OHS 14
4	***Relation-oriented data*** • Relation oriented data-General and flow data • Illustration - Stage-discharge relationship • Relation oriented data - Sediment data • Sediment data - Types • Illustration - Suspended sediment & corresponding discharge	5 min	 OHS 15 OHS 16 OHS 17 OHS 18 OHS 19
5	***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

1.	**Introduction**	**1**
2.	**Space-oriented data**	**2**
3.	**Time-oriented data**	**3**
4.	**Relation oriented data**	**8**

Understanding different types and forms of data in HIS

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.

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.

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.

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.

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.

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.

3. Time-oriented data

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.

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

3.2.1 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).

3.2.2 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.

3.2.3 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.

3.2.4 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).

3.2.5 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.

3.2.6 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.

3.2.7 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.

3.2.8 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.

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.

3.3.1 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.

3.3.2 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.

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

3.4.1 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.

3.4.2 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.

3.4.3 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).

3.4.4 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.

3.4.5 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.

3.4.6 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. 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

4.1 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.

4.2 Flow data

4.2.1 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.

4.2.2 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

4.3 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 Fig. 4.1 as below:

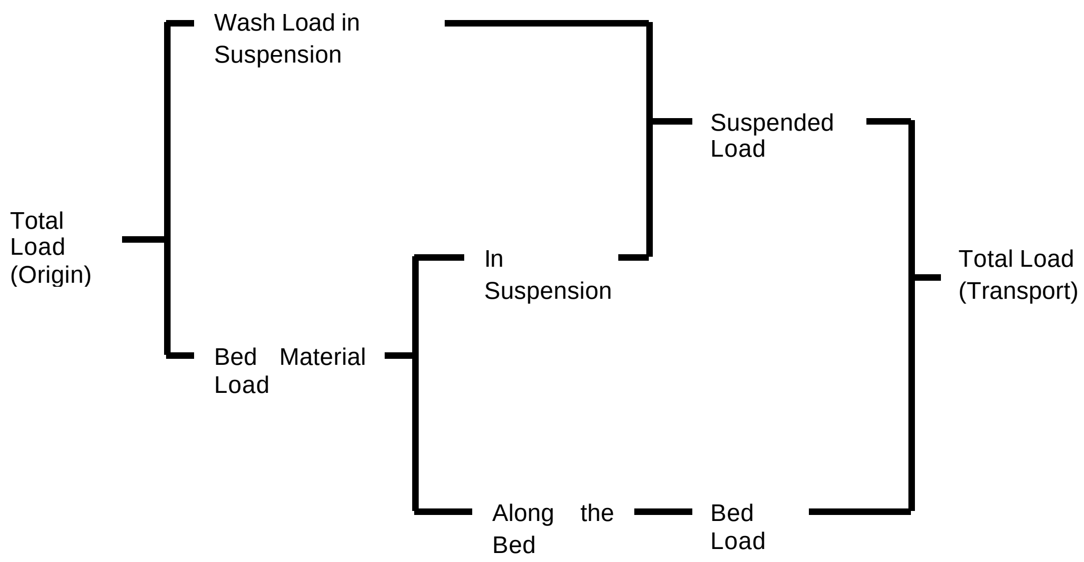

Fig. 4.1: Definition of sediment load and transport (ISO 4363: 1993)

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.

4.3.1 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.

4.3.2 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.

4.3.3 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.

4.3.4 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)