1.1 PROCESSING AND FILING DATA

3.1 DATA REGISTERS AND FILING

This section sets out the procedures to be followed to provide an adequate, documented record of each step in processing the hydrological data, from the time of its measurement in the field to its final archiving in the Hydrometric Archive.

3.1.1 General

Hydrological data are valuable in that they are relatively expensive to collect, are irreplaceable and have the potential to have very high value following certain events. To realise and maintain their value, there must exist a means of verifying the accuracies and giving assurance that errors are largely absent. Thus the traceability of the data and the methods used to collect and process them must be available in a readily followed form. Many of the provisions of this manual are aimed at achieving this traceability, in conjunction with the efficient processing of the data while preserving and verifying its integrity.

The data processing system includes provisions to

- register the data after collection to confirm its existence and track its processing
- keep backups of the data in its original form
- positively identify the individual batches of data at the various stages of processing
- identify the status of data as to its origin and whether it has been verified as fit for use
- present and store evidence of any modifications to the data
- file all field observations which verify the data
- control the amount and type of editing which can be performed and the authorisation to do this.
- present the data in a number of ways for checking and auditing by trained persons who are independent of the process.

3.1.2 Registers

As soon as data reach the field office (whether by telemetry, as computer files, charts or on hand-written forms) they are entered in one of a set of registers, organised as to station and data type, and in chronological order.

These registers are on hard copy in a folder in the office and are updated daily as batches of data arrive. Initially this involves the noting of the start and finish times of the data batch, and progresses with confirmation of editing, checking and updating to the database. The requirement for these steps to be signed is deliberate and mandatory, in order that staff take responsibility for and "ownership" of their work and its progress.

The registers thus contain a verified chronological record of data processing activities in the field office. There are separate register forms for:

(a) water-level data (Form WL);
(b) automatic raingauge data (Form RF);
(c) flow gaugings (Form GR);

(d) back-up disks (Form BUR);
(e) Updates to the database including checks made at this stage (Form UD).

These registers are to be updated daily, as data progresses through the system. All columns within the registers must be filled in with the required information including the initials of the processor where required.

3.1.3 Identification of Original Records

All data are to be permanently identified with the site numbers and the dates of observation. These can be written directly on chart records and gauging cards, and on original plots of telemetered data (in the event that the plotting process does not do this adequately). Charts require a stamp or label to prompt for the dates and times, staff gauge readings, etc. for both chart on and chart off.

Data files on disk, whether the primary file for updating or a backup file, should have their site identified in the header line(s) and be filed in an appropriate directory. However, if these are binary .dmp files, they can only be identifiable by the file name.

3.1.4 Plots of Original and Edited data

At initial processing on office computer, the first main step is to produce plots of the data. These "original plots" are both evidence of the original state of the batch of data and a backup for all subsequent data processing. They are kept as archived documents in site folders. The date of processing and the initials of the processor shall be hand-written on the plots, along with explanatory notes on any real or apparent anomalies and on any editing done. Because a plot has the potential to illustrate many of the various types of errors possible, and because they are a convenient place to write notes, these original plots together with subsequent ones following editing provide the main evidence of traceable data integrity.
They are filed in the station files (see below).

3.1.5 Station files

A physical file (normally pockets in a filing cabinet drawer) is to be maintained for each water-level and rainfall station. It is intended as a quick reference to the evidence of processing and to station details. As described above, each batch of data is to be plotted and the plots kept together in chronological order, with a printout from the appropriate translation process to show what corrections have been applied to it. The station logbook pages (the tear-out pages) are also to be kept in this file, normally stapled to the plot(s) for that batch.

The file must be stored in a safe environment, preferably fireproof, as many items are original documents.
The following are kept in each station file (or equivalent combination of physical files):

- plots of original and edited data batches as described above, together with printouts from the processes
- the station logbook pages for each batch
- a listing of comments filed on the database

- rating curve data (plots, listings of co-ordinates)
- any cross-section data and other information relating to the extension of rating curves
- station photographs labelled with site number, date, and viewpoint
- station history details on Form SH, containing details of staff gauge locations and zero elevation, benchmarks and survey references (if any). These should be kept up-to-date in the file at all times and photocopied when required to use in the field.

3.1.6 Instrument Register

A file containing the locations, descriptions and serial numbers of all recording equipment is kept. As well as the corporate governance requirement to record the custody and location of all instruments, it should also be useful to provide operational data on ranges, cable/tubing lengths, version numbers, etc. A record of calibration of measuring equipment shall also be part of this.

3.1.7 Telemetry data logging

Daily data retrieval schedules are essential to rapidly identify any stations that have either stopped or have lost radio communication links. This is necessary to rectify faults immediately and reduce the likelihood of missing record. The telemetry log is also useful in determining when telemetry problems were encountered.

Any stations with telemetry as the only recording instrumentation (i.e., without significant on-site data storage) require a registering and back-up filing system to maintain an archive of original data on disk, and to ensure data are not accidentally lost due to hardware or other system failures.

A weekly and monthly back-up filing system is necessary so that in the event of problems occurring, original data can still be retrieved. The procedures for these systems are described in section 6, and files are recorded in the back-up register.

3.1.8 Electronic data

Stations with dataloggers require a system for registering data diskettes and procedures to prevent accidental loss of original data. A diskette register (Form BUR) must be maintained and all diskettes given an individual identification number for data collection and processing tracking purposes.

Files in their original form of *site.dmp* and the derived *site.dat* ASCII files shall be stored in a suitable repository, either discs in fireproof storage, the local network backup or preferably the backup directory of the Hydrometric Archive. If diskettes are used, periodically these should also be copied to CDs. A register of these files shall be maintained as a paper register, using Form BUR.

A diskette is not to be reformatted and re-used until the data is safely stored and backed-up on the Hydrometric Archive, the backup directory and all required registers have been completed.

3.1.9 Chart processing

Charts should be stamped with the prompts for date, time and water-level readings at chart off and chart on. As these indicate the adjustments to the original observations, the need for clear annotation is obvious.

When the digitising is completed, update the processing section of the appropriate register accordingly. At the same time, check that all processing steps for the previous batch or batches of data for that particular site have been completed. Note that all original data should be processed and updated in chronological order for each site.

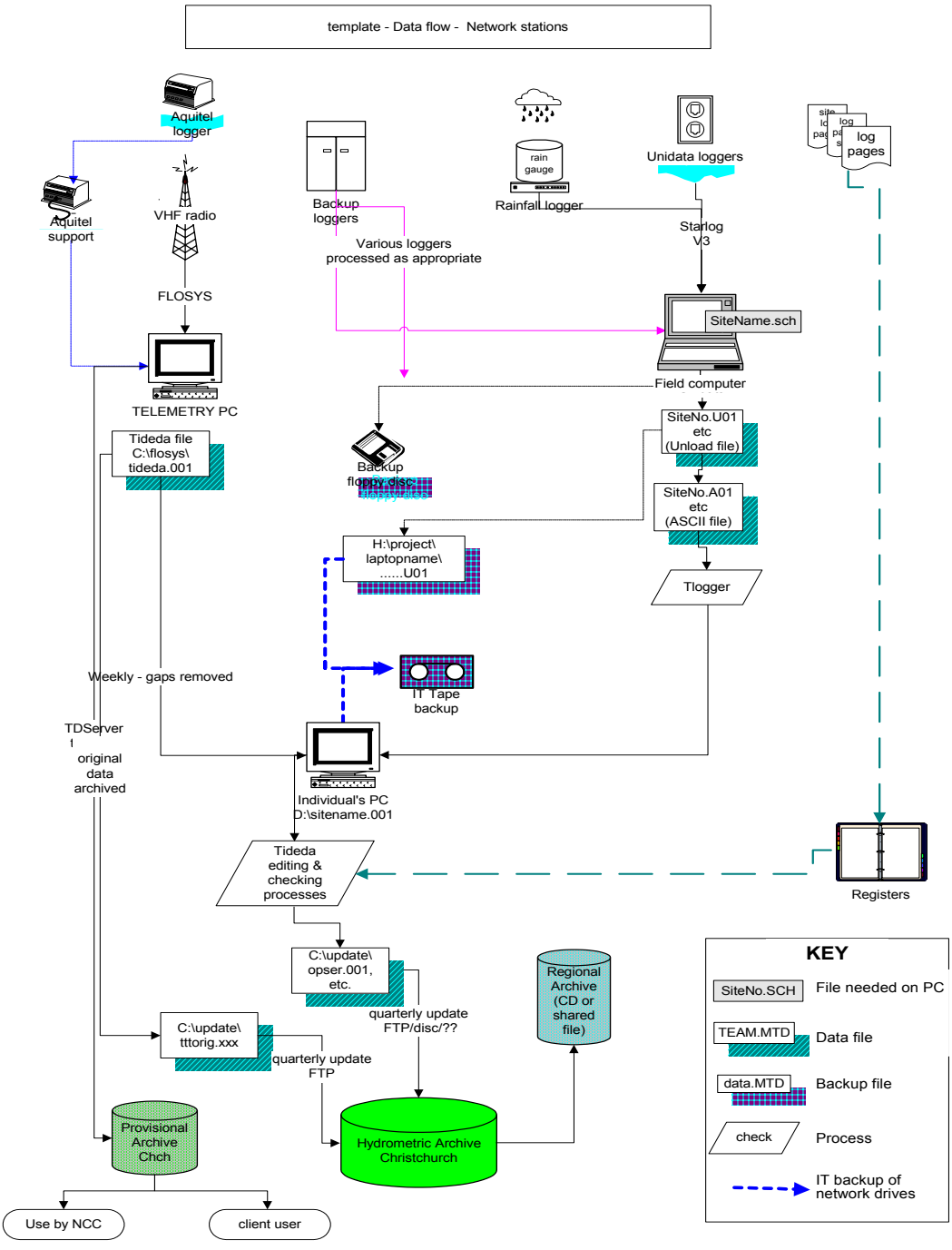

Figure 1. *Data processing flow chart*

3.2 INITIAL DATA PROCESSING

This section sets out the data processing steps to be applied to the different types of data.

Figure 1 presents a flow diagram showing the steps to follow for the translation and subsequent processing of chart and logger data and the updating of these batches to the Hydrometric Archive. This encompasses both water-level and rainfall data.

3.2.1 Adjustments to time and/or water-level

Where adjustments are to be made to either time or water-level throughout the data, the following must be answered before proceeding:

(a) can you provide reasons for the adjustments?
(b) are the adjustments consistent with the previous batch of data?
(c) is any follow-up field work necessary and have the appropriate people been notified?

Answers to (a) and (b) must be fully documented on the original plot.

3.2.2 Checking Water-level Series Data

The first check is that the start date, time and values for the batch match up (within the time increment, as applicable) with the finish values for the previous batch. This is available from the register, and quickly checked.

The initial processing should provide the maximum and minimum gauge height and their relevant times for the batch. Enter these values in the water-level register (Form WL) and check that they are reasonable. (Any unusually high or low values should be a "flag" to check that the values are in context).

The next step is to plot the batch at a suitable scale and examine to detect such problems as

- silting (mainly at stream sites), which will round-off peaks and cause unusually high recessions
- "spikes" or small numbers of values which are obviously out of context and incorrect due to an unnaturally large change in value between adjacent data points. Such occurrences can occur with errors in the digital values between the encoder to datalogger interface
- gaps in the data
- errors induced by the field staff, such as pumping water into a well to flush it
- restrictions to the movement of the float/counterweight system or the encoder (perhaps caused by the cable being the incorrect length
- vandalism
- debris caught in the control structure, the control, or other damming or backwater condition
- and others

Note that such problems, if not detected on the field visit, must be investigated at the earliest opportunity. In cases where the cause has not been positively identified, full processing of the data should be delayed until it can be investigated on-site.

When the plot is complete, attach the printout containing the parameter entries and add the following by hand:

(a) date the data was plotted;
(b) signature of the processor;
(c) notes of all corrections to the data, plus any subsequent actions carried out which change the data in the file from that which is plotted (e.g. removal of spikes, inserting manual data resulting from silting, etc.).

For all sites the first plot becomes the original document. If any corrections are applied, another plot is necessary to show their effects.
Both shall be filed in the station's file as described above.

3.2.3 Checking Rainfall Data

Again, the first check is that the start date, time and values for the batch match with the finish values for the previous batch. This is available from the register, and quickly checked.

The initial processing should provide the recorder total rainfall for the period. Enter this in the rainfall register (Form RF) and check that it matches with the check gauge total within 5%. If they differ by more than this, the rainfall series data are corrected objectively to their associated check gauge total.

The next step is to plot the batch at a suitable scale, preferably over-plotted with a close flow and or rainfall station, and examine to detect such problems as

- unusually high intensities in isolation, suggesting interference by outsiders adding water
- nil or low rainfall compared with other stations, suggesting a malfunction in the gauge or electrical connections
- the rainfall pattern not following the runoff pattern for the catchment or adjacent catchments
- gaps in the data
- errors induced by the field staff, such as extra tips when testing during a visit
- slow response and small peaks, suggesting debris is caught in the gauge blocking the funnel, (or hail, snow, etc.)
- and others

Note that such problems, if not detected on the field visit, must be investigated at the earliest opportunity. In cases where the cause has not been positively identified, full processing of the data should be delayed until it can be investigated on-site.

A daily rainfall tabulation can be a useful checking tool.
For all sites the first plot becomes the original document. A second plot is necessary to show the effect of correction to the check gauge. Any other editing should also be shown.

When the plots have been fully checked, attach the printout containing the parameter entries and add the following by hand:

(a) date the data was plotted;
(b) signature of the processor;
(c) notes of all corrections to the data, plus any subsequent actions carried out which change the data in the file from that which is plotted (e.g. removal of tips)

Both plots shall be filed in the station's file as described above.
Finally, make sure all relevant parts of the raingauge register have been completed.

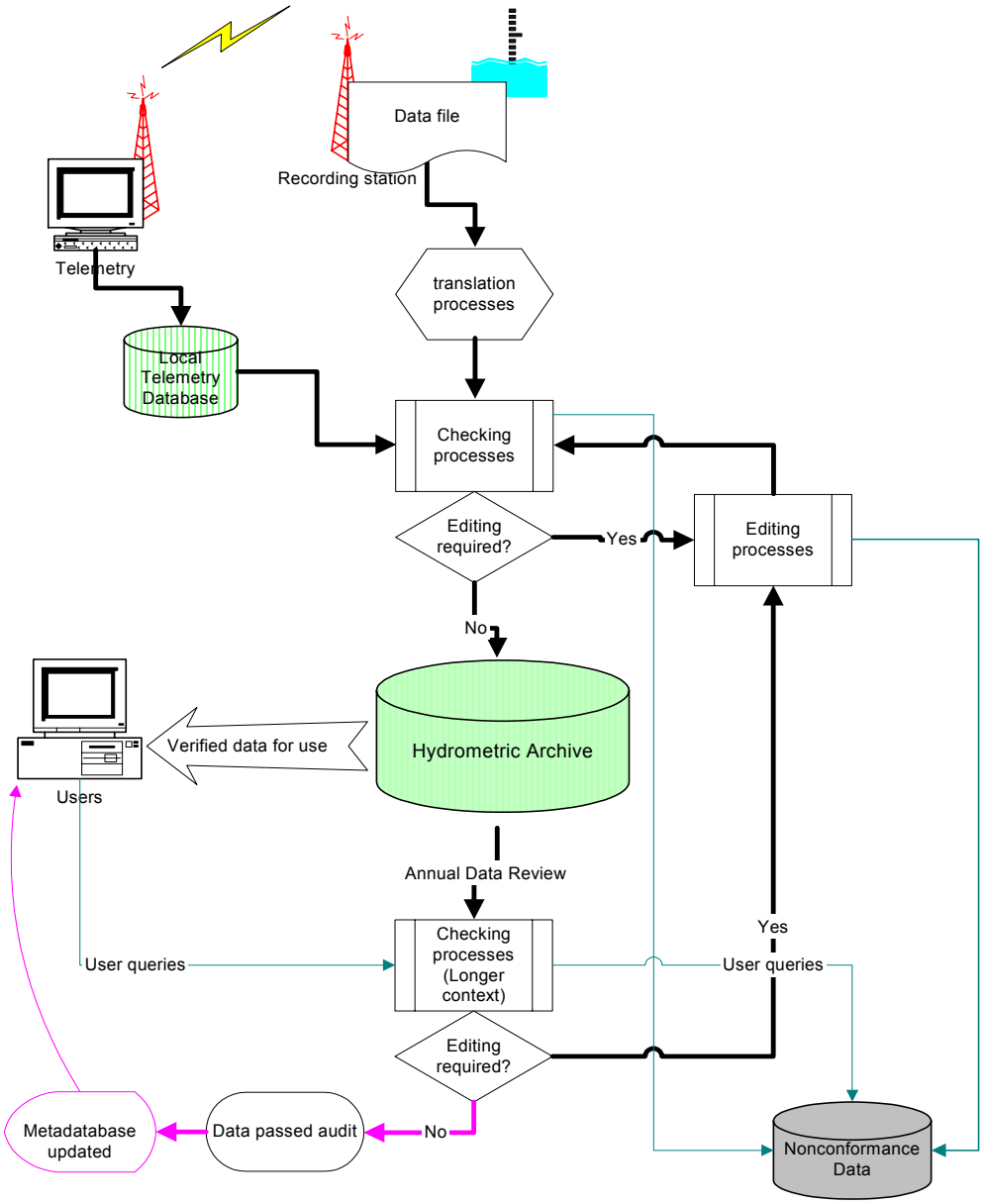

Figure 2. *Quality control system for hydrological data*

Note: The purpose of an automatic raingauge is to measure rainfall intensity. Where a considerable discrepancy occurs between the recorder and the check gauge totals, file only those data from the recorder that are reliable.

3.3 Editing of Water-level and Rainfall Series Data

The software packages available provide many ways of editing or modifying data. It is often impossible to know absolutely the source and effect of any data errors after the event and, although people may postulate theories, these may not be the only explanation. Therefore it is imperative

that any modification of data intended for the database is done conservatively, within strict guidelines, and by or under the guidance of experienced personnel.

3.3.1 Conditions for Editing Data

Editing or altering original data carries with it the responsibilities of ensuring traceability and data integrity. The following guidelines shall be observed in all cases:

(a) No alterations shall be made unless justification for the assumptions made is scientifically defensible, and recorded, as below.

(b) Such alterations must have the explanation recorded either on, or attached to, the plot of the original data in the site file or as a Comment on the database.

(c) **As a general guideline, gaps due to missing record will <u>not</u> be filled with synthetic data or interpolated.** Any approximate data available shall be made available to users by inclusion or reference to it in the Comment for that gap. Exceptions to the non-use of synthetic data and interpolation are in (d) to (e) below.

(d) A gap in a water-level record may be filled with a straight line or curve as applicable, if <u>all</u> of the following conditions are fulfilled:

(i) The river is in a natural recession with the water-level lower (or the same) at the end of the period, and
(ii) It has been ascertained that no significant rain fell in the catchment during the time of concentration which would relate to the gap period, and
(iii) The catchment is known to be free of abstractions and discharges which modify the natural flow regime (e.g., power station , irrigation scheme), and
(iv) The resulting plot of the data shows consistency with the data on either side
(v) In some situations, (e.g. power stations) an adjacent station may measure the same data or almost the same data. In the former case, the record can be filled in as if it were a backup recorder. In the latter, the data may be filled in if the uncertainty is less than that in the standard or if the correlation between stations for that parameter and that range can be shown to be 0.99 or greater. A comment containing the details of the relationship must be filed.
(vi) The station is not a lake which normally has seiche or wind tilt. (These are often studied, and synthetic record will not be able to re-create the phenomena.)

(e) Where the conditions do not meet these criteria, but trained personnel were on-site for the whole period (e.g. flushing) and recorded manual observations, the gap may be filled with these values and interpolated accordingly.

(f) Filling a gap in the original data record with synthetic data derived by correlation is not permissible. It can be used, however, for supplying data requests where the user is informed of the uncertainty involved. Such data must be carefully controlled to ensure it is not erroneously filed on either the local or central archives.

(g) A gap in a rainfall record may be "interpolated" only if it can be established that no rain fell during the period, by means of correlation with other gauges within or without the catchment for which there is an established correlation and with a correlation coefficient of 0.99 or higher.

3.3.2 Telemetry Data

As telemetry data arrive on the computer system and are available for use immediately without checking (albeit as "unverified" data), they are subject to the same potential errors as other data. Users who are provided with data at this stage need to be made aware of its status.

Despite the speedy arrival, the data's verification is still subject to later on-site checking of the sensor's operation by hydrological staff. Therefore telemetry data are to be checked, processed and verified from site visits in the same manner as other data retrieved via the monthly site visit regime.

The telemetry log may be kept as evidence of communications between the remote stations and the base, until the data have been quality checked and filed on the Archive.

For stations with telemetry as the only recording instrument, a regular (daily) retrieval and checking schedule is required for both water-level and rainfall recorders, with a weekly back-up procedure to eliminate loss of recorded data.
The following steps are carried out to achieve this:

(a) Daily data is collected via telemetry and filed on a telemetry file
(b) Check all station data have been received. The series should be checked for large gaps. A record should be kept of all stations not responding or showing problems (e.g. a log book or weekly plots with notes attached.
(c) Resolve all problems by either checking through the log for likely problems, trying to retrieve the data again and scheduling an urgent trip to the affected station.
(d) A regular back-up procedure shall be in place to ensure against the loss of telemetry data.
(e) The back-up period must be based on the remote memory capacity and shall be done at least weekly. The back-up disks are to be stored in a safe environment and not to be touched until the data has been copied to the monthly back-up disk. These disks must be individually labelled and registered in the Back-up Register and a listing of the disk contents stored with them.
(f) When the field visits have been made, those data can be retrieved from the telemetry file to undergo the normal data processing regime of plotting, editing if required, transfer, etc.

Only data processed and validated by field check visits can be transferred to the Hydrometric Archive files.

3.3.3 Chart Data

Chart data can be digitised using TIDEDA process RSTRIP to give a *batch.dat* file and then TSTRIP into a Tideda file. Water-level heights, or rainfall accumulation, are automatically adjusted within the translation processes, and care must be taken to ensure that the corrections being applied are correct.

3.3.4 Manually Entered Series Data

Manual input of series data can be used to:

- supply corrections to existing data;
- input data for which there is no alternative method (e.g., manual staff-gauge readings, storage raingauge data or flow gaugings);
- input data read manually from chart recorders;
- input synthetic data to cover gaps in records. Note that there are strict conditions for filing synthetic data; see above.

3.3.5 Comment Data

Comments covering the accuracy of data and gaps in records should be informative, coherent and identify the period(s) for which the data are suspect or missing. Comments covering gaps in record should be filed one time interval (usually 15 minutes) before the record recommences.

All recording stations require an initial comment detailing station information and parameters measured. Flow stations require a "rating comment", to provide information on the ranges covered by flow gaugings, and "recorder comments" are required for both water-level and rainfall stations, giving instrument and sensor types and their resolutions and accuracy. Comments relating to gaps, missing and synthetic records are also required in a standard format, and all of these types of comment formats are given in xxxxxxx Appendix A.

3.3.6 Flow Ratings

Most flow stations have reasonably well established ratings. These are to be checked periodically by gaugings which are immediately plotted on the master rating curves held in the station file. Changes in the hydraulic conditions at the station (such as aggradation, degradation, channel shift, and consequent changes in channel slope and velocity distribution) may shift the rating and require that it be re-defined.
The construction and identification of changes in stage discharge ratings are covered in detail in NIWA's "Procedure for rating a flow station" (McMillan, Thomas, Stead, 1994).

Sufficient co-ordinates should be entered to accurately represent the shape of the rating curve, with emphasis around the parts of the curve identifying changes in control and/or cross-section shape (e.g., change to berm flow). Normally between 10 and 20 sets of co-ordinates are required to define the rating shape.

3.3.7 Flow Gaugings

Gauging calculation is to be done using an approved program, e.g., the GAUGE program (Thomas, 1994). A listing of approved software is in Appendix 3 xxxm
This computer program calculates the gauging, saves a gauging card image, and produces a 15-item transfer file for filing on the database. The transfer file has the following 15 items:

- *Item 1* - Gauge height (mm)
- *Item 2* - Discharge (litres or millilitres/second)
- *Item 3* - Area (cm^2)
- *Item 4* - Mean velocity (mm/second)

Item 5 - Maximum depth (mm)
- *Item 6* - Slope (mm/1000 m)
- *Item 7* - Width (mm)
- *Item 8* - Hydraulic radius (mm)
- *Item 9* - Wetted perimeter (mm)
- *Item 10* - Sediment concentration (mg/litre)
- *Item 11* - Water temperature (degrees C x 1000)
- *Item 12* - Water-level change (+\- mm/hour)
- *Item 13* - Method code (ccnn) where cc = method and nn = no. verticals
- *Item 14* - Number of verticals and sampling points used
- *Item 15* - Gauging number

There are several processes which can be used to display gauging data, with either ratings or series data stored on a file.

Bed plots can be used to check the fit of gaugings against flow rating curves:

- Time ordered bed plot - Checks each gauging within the test period against the appropriate rating. If the percentage deviation of the measured flow from the rated flow is greater then the acceptable limit (usually +/_ 8%), then the gauging's accuracy needs investigation and a filed Comment.
- Sorted bed plot – as above, plus look at the number of runs of positive gauging deviations to negative ones. There are statistical rules which can be used to determine whether or not the errors are random. The test fails if they are not random, as the implication is that they are systematic and could indicate a faulty or biased rating.

If a systematic trend is identified (i.e. gaugings plot on one side) even though it may be less than the acceptable maximum percentage (usually +/- 8%), and covering a period that will affect long-term discharge results, **a new rating has to be constructed and filed.**

Gaugings plotting more than +/- 8% from the rated flow shall have an explanatory Comment filed.