Technical Assistance
HYDROLOGY PROJECT

World Bank & Government of The Netherlands funded

Training module # SWDP - 23

How to carry out secondary validation of water level data

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	:	How to carry out secondary validation of water level data
Target group	:	Assistant Hydrologists, Hydrologists, Data Processing Centre Managers
Duration	:	One session of 120 minutes
Objectives	:	After the training the participants will be able to: • Carry out secondary validation of water level data
Key concepts	:	• Relation between water levels at adjacent stations • Validation of water level data against established relationship • Qualitative comparison of water level and basin rainfall
Training methods	:	Lecture, Software
Training tools required	:	OHS, computers
Handouts	:	As provided in this module
Further reading and references	:	

3. Session plan

No	Activities	Time	Tools
1	***General***	10 min	
	• Secondary validation of water level data (1)		OHS 1
	• Secondary validation of water level data (2)		OHS 2
2	***Scrutiny of multiple hydrograph plots***	10 min	
	• Scrutiny of multiple hydrograph plots (1)		OHS 3
	• Scrutiny of multiple hydrograph plots (2)		OHS 4
	• Example 2.1 – Text introduction		OHS 5
	• Example 2.1 - Tabular		OHS 6
	• Example 2.1 - Graphical		OHS 7
	• Example 2.1 – Graphical (detail)		OHS 8
	• Example 2.1 – Text		OHS 9
	• Example 2.2 - Graphical		OHS 10
	• Example 2.2 - Graphical (detail)		OHS 11
	• Example 2.2 – Graphical (error explanation)		OHS 12
	• Example 2.2 – Graphical (error explanation)		OHS 13
	• Data validation		OHS 14
3	***Combined hydrograph and rainfall plots***	10 min	
	• Combined hydrograph and rainfall plots		OHS 15
	• Example 3 - Graphical		OHS 16
4	***Relation curves for water level***	15 min	
	• Relation curves (1)		OHS 17
	• Relation curves (2)		OHS 18
	• Example 4.1 – Graphical		OHS 19
	• Example 4.1 – Graphical		OHS 20
	• Relation curves (3)		OHS 21
	• Relation curves (4)		OHS 22
	• Relation curves (5)		OHS 23
	• Relation curves (6)		OHS 24
	• Example 4.1 – Graphical		OHS 25
	• Example 4.1 - Graphical		OHS 26
	• Fitting of relation curve		OHS 27
	• Example 4.1 – Graphical		OHS 28
	• Example 4.1 – Tabular		OHS 29
	• Comparison of relation curves		OHS 30
	• Example 4.2 - Graphical (hydrographs)		OHS 31
	• Example 4.2 - Graphical (relation curve)		OHS 32
	• Example 4.2 – Graphical (comparison of relation curves)		OHS 33
	• Example 4.2 – Graphical (shift of relation curve)		OHS 34

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

How to carry out secondary validation of water level data

1. General

- **Water level data received at Divisional offices have already received primary validation on the basis of knowledge of instrumentation and methods of measurement at the field station** and information contained in Field Record Books. Primary validation has included comparisons between different instruments and methods of observation at the same site.

- **The data processor must continue to be aware of field practice and instrumentation and the associated errors which can arise in the data,**

- **Secondary validation at Division now puts most emphasis on comparisons with neighbouring stations to identify suspect values.** Special attention will be given to records already identified as suspect in primary validation

- **The assumption, while carrying out secondary validation, is that the variable under consideration has adequate spatial correlation. Since the actual value of water level is controlled by specific physical conditions at the station, the amount of secondary validation of level is limited.** Most of the check comparisons with neighbouring stations must await transformation from level to discharge through the use of stage discharge relationships. Only as discharge can volumetric comparisons be made. However validation of level will identify serious timing errors.

- **Secondary validation of level will be used to identify suspect values or sequences of values but not usually to correct the record, except where this involves a simple shift (time or reference level) of a portion of a record.**

- **The main comparisons are between water level series at successive points on the same river channel. Where only two stations are involved, the existence of an anomaly does not necessarily identify which station is at fault. Reference will be made to the historic reliability of the stations.**

Comparisons will also be made between incident rainfall and level hydrographs.

2. Scrutiny of multiple hydrograph plots

Graphical inspection of comparative plots of time series provides a very rapid and effective technique for detecting timing anomalies and shifts in reference level. Such graphical inspection will be the most widely applied validation procedure.

For a given time period several level time series for neighbouring stations are plotted in one graph. **For routine monthly validation, the plot should include the time series of at least the previous month** to ensure that there are no discontinuities between one batch of data received from the station and the next. The time interval of observation rather than averaged values should be displayed. In general, peaks and troughs are expected to be replicated at several stations with earlier occurrence at upstream stations and the lag between peaks, based on the travel time of the flood wave, approximately the same for different events. It should be noted that level values at downstream stations are not necessarily higher than upstream stations - the actual value depends on physical conditions at the stations.

Where peaks occur at one station but not at its neighbour or where the lag time between stations is widely different from the norm, an error at one station may be suspected. However it must be recognised **that the quality of the relationship between neighbouring hydrographs depends not only on the accuracy of the records but on a variety of other factors including:**

- rainfall and inflow into the intervening reach between stations. If the intervening catchment is large or the rainfall high in comparison to that over the upstream basin, a very poor relationship may result.
- river regulation and abstractions between the stations may obscure natural variations, though high flows are usually less affected than low or medium flows.
- An average lag between successive stations can be used in making comparisons but the actual lag is variable, generally diminishing up to bankfull stage and increasing again with overbank flow.
- one station may suffer backwater effects on the stage hydrograph and not another, obscuring the effects of differences in flow. Where such effects are known to occur, comparison should await transformation to discharge.

Anomalies identified from comparative hydrograph plots are flagged for further stage validation or to await validation as discharge.

Example 2.1:
Application of the above described technique is demonstrated for the stations MAHEMDABAD and NSB00I7 on WATRAK river, a tributary of Sabarmati in Gujarat. The stations are distanced some 33 km apart (MAHEMDABAD d/s of NSB00I7) and the lateral inflow in between the sites is small compared to the river flow. The hydrographs of hourly water levels for the months September and October 1998 are shown in Figure 2.1a. From the plot some anomalies are observed. Make sure to have always a tabulated output of the water level observations available when carrying out such analysis to note down possible anomalies.

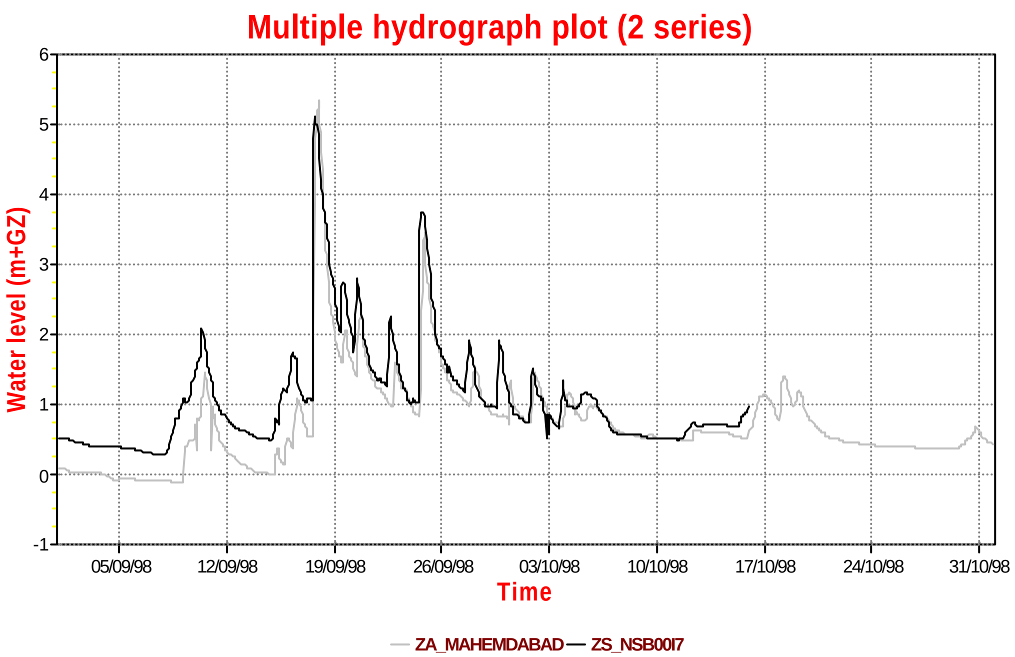

Figure 2.1a Multiple hydrograph plot

To get a better view one should zoom in. A detail of the hydrograph, showing an unnatural regular sequence of peaks is presented in Figure 2.1b. From the Figure two types of errors may be observed:

- spurious errors originating from transcription errors in the field or in the office, and
- errors in the occurrence of peaks

Particular with respect to the last type of error additional information is required to determine which parts of the hydrographs are faulty. Fortunately upstream as well as downstream of the stations hydrometric stations are available. The comparison with the stations upstream of NSB00I7 is presented in the following example.

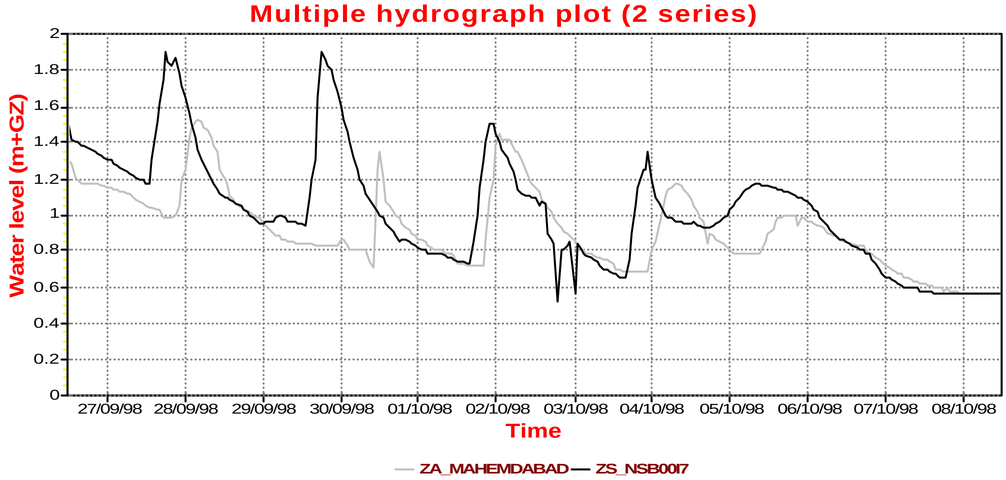

Figure 2.1b Detail of multiple hydrograph plot

<u>Example 2.2</u>:
The hydrograph plot presented in Figures 2.1a and b is now extended with the hydrographs of the stations DHABA on WATRAK and AMBALIYARA on MAZAM (reference is made to the catchment map for their exact locations). The hydrographs are shown in Figure 2.2a. and the detail for the same period of Figure 2.1b in Figure 2.2b.

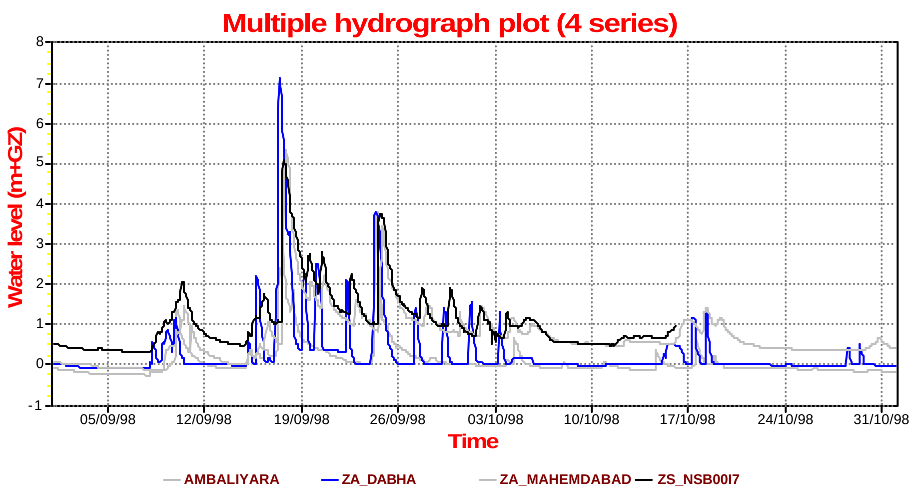

Figure 2.2a Multiple hydrograph plot, 4 stations

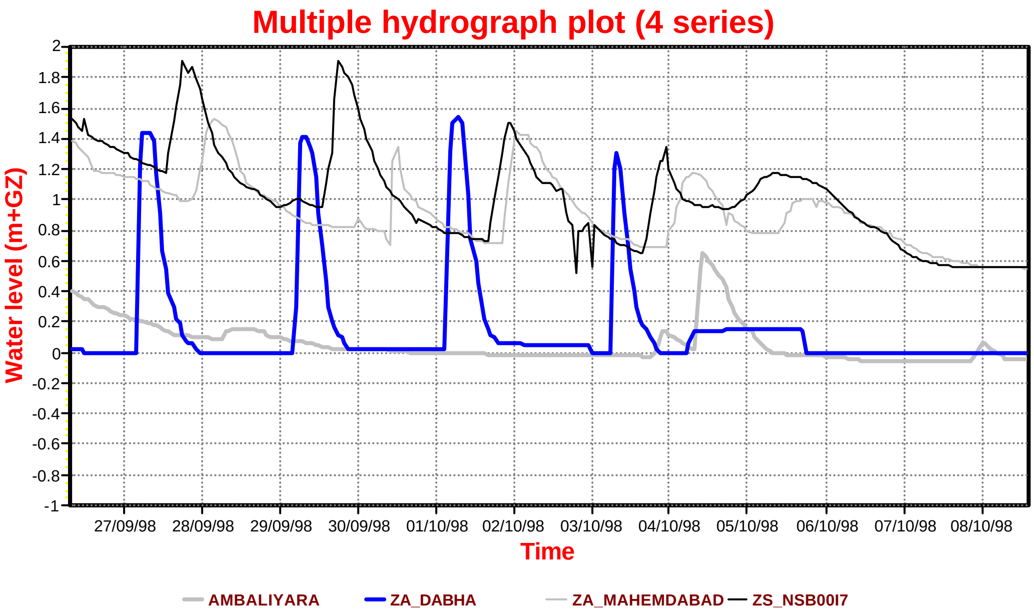

Figure 2.2b Detail of Figure 2.2a.

It is observed from Figure 2.2b that the first 4 peaks originate from upper WATRAK. These peaks are due to releases from an upstream reservoir. The last peak is recorded at AMBALIYARA. The records of DHABA and AMBALIYARA do not show distinct anomalies and hence can act as a guide to detect anomalies in the records of MAHEMDABAD and NSB00I7. To proceed, first the suspect parts of the hydrographs of MAHEMDABAD and NSB00I7 are noted down, based on resemblance of patterns. Possible anomalies (strange patterns, sharp rises and falls) are indicated in Figure 2.2c.

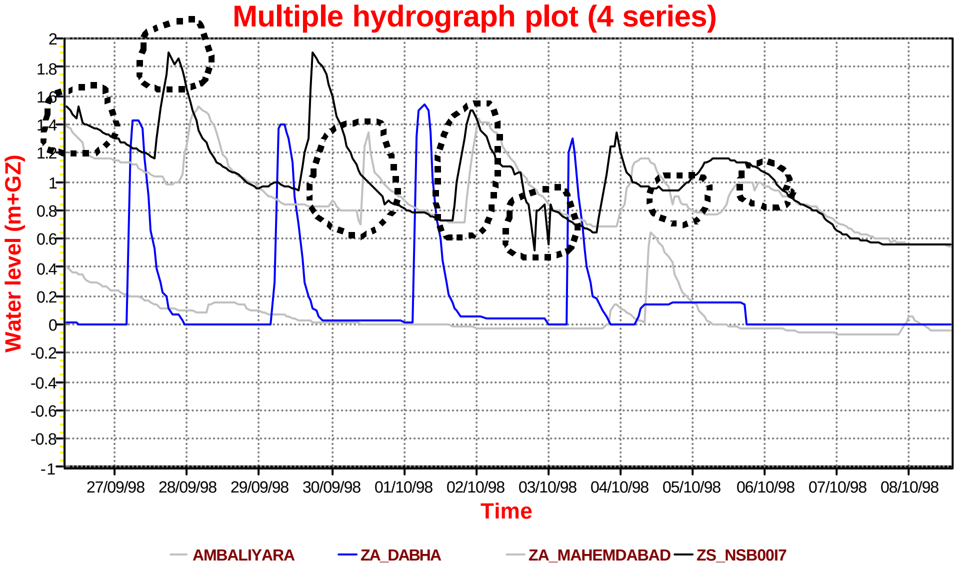

Figure 2.2c Identification of suspect values based on resemblance of hydrograph patterns

Secondly, anomalies in the time lag between the peaks are investigated, see Figure 2.2d.

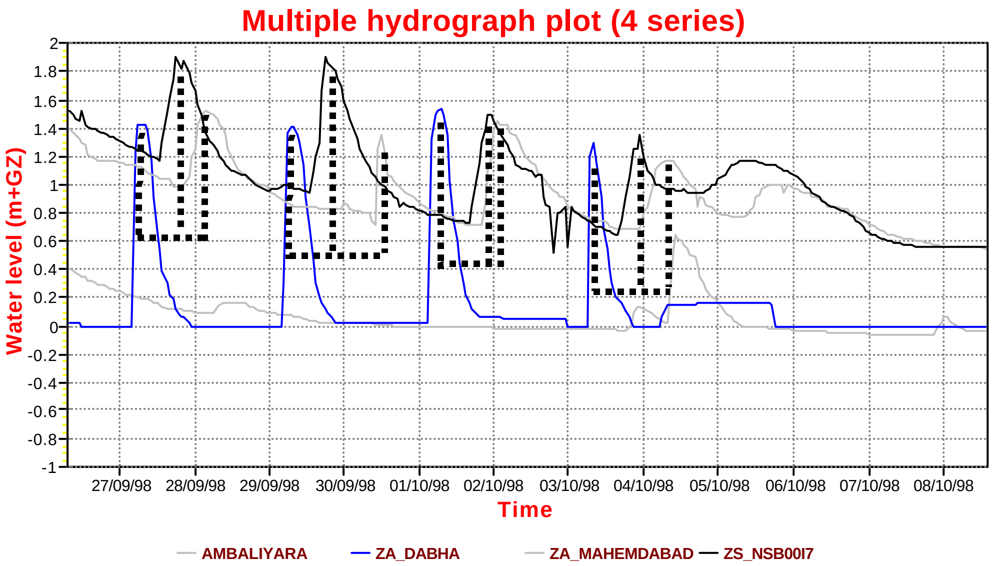

Figure 2.2d Identification Anomalies

The time lags between the first set of peaks are taken as a guide, as no distinct anomalies are present. Next, these time lags are applied to the other sets of peaks as well. From the comparison it is observed that the second peak of MAHEMDABAD and the third peak of NSB00I7 are suspect. The time lags between the peaks in the fourth set resembles the time lags of the first set. Note that this comparison is only allowed for similar flow sizes; if the ranges are different, wave celerities may be different and so will be the time lags, see below.

3. Combined hydrograph and rainfall plots

The addition of rainfall to the comparative plots provides a means of assessing timing errors and of investigating the effects of inflow into the intervening reach between stations. Comparison may be made using an average rainfall determined using Thiessen polygons or other methods over the entire basin or for the intervening sub-basin corresponding to various gauging stations. Where the basin is small or the number of rainfall stations limited, individual rainfall records may be plotted.

In general a rise in river level must be preceded by a rainfall event in the basin and conversely it is expected that rainfall over the basin will be followed by rise in level. There must be a time lag between the occurrence of rainfall and the rise in level. Where these conditions are violated, an error in rainfall or in the level hydrograph may be suspected. **However the above conditions do not apply universally** and the assumption of an error is not always justified especially for isolated storms in arid areas. For example:

- An isolated storm recorded at a single raingauge may be unrepresentative and much higher than the basin rainfall. The resulting runoff may be negligible or even absent.
- Where storm rainfall is spatially variable, it may be heavy and widespread but miss all the raingauges, thus resulting in a rise in river level without preceding measured rainfall.
- The amount of runoff resulting from a given rainfall varies with the antecedent catchment conditions. Rainfall at the onset of the monsoon on a very dry catchment may be largely absorbed in soil storage and thus little reaches the river channel.

The use of comparative plots of rainfall and level is therefore qualitative but it provides valuable ancillary information with the multiple hydrograph plots.

<u>**Example 3.1**</u>:
An example of a combined hydrograph and rainfall plot is presented in Figure 3.1, which displays the water level record of station AMBALIYARA on MAZAM river together with the rainfall records of stations RAHOL and VADAGAM.

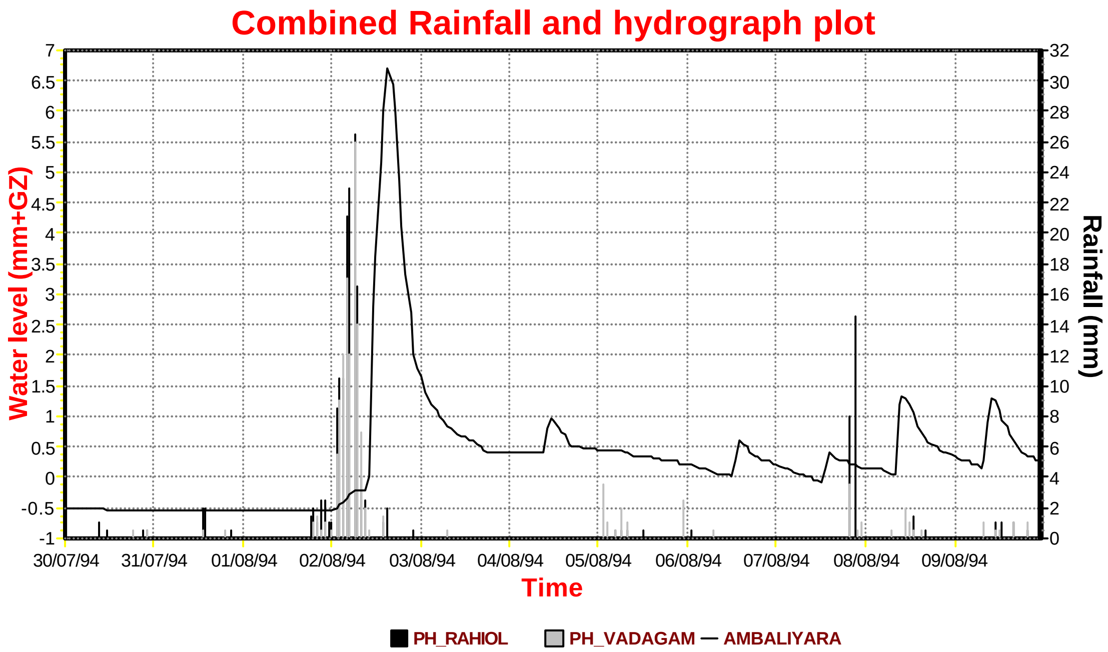

Figure 3.1 Combined hydrograph and rainfall plot

From the graph it is observed that the first peak is preceded by substantial rainfall. The remainder, however, shows suspect combinations of hydrographs of stage and rainfall, where hydrograph appears to occur before the rainfall and where the response to the rainfall is delayed. One may be tempted to doubt the rest of the record. However, in this case the peaks may as well be caused by releases from an upstream reservoir. Therefore, in this case additional information is required about the reservoir releases to conclude on the series. Nevertheless, the delayed response to the second rainfall peak remains suspicious.

4. Relation curves for water level

4.1 General

- **A relation curve gives a functional relationship between two series of the form:**

$$Y_t = F(X_{t+t1})$$

To account for the lag between level changes at one station and the next downstream, it may be necessary to introduce a time shift (t1) between the two time series.

- **Relation curves will normally be applied to water level data rather than discharge. However it may be appropriate on occasions to use them for discharge data, especially where one or both stations are affected by backwater conditions.**

- **If there is a distinct one to one relationship between two series, random errors will be shown in a relation curve plot as outliers.**

- **By comparing two relation curves, or data of one period with the curve of another period, shifts in the relationship, e.g., in the water level series due to changes in the gauge zero can be detected.**

4.2 Application of relation curves to water level

If two water level stations are located on the same river and no major branch joins the main stream between the two locations, a relation can be expected between the recordings at the two locations. With the help of this relation, the stage at a particular station can be derived from the available data series of the adjacent station. A sample plot of such relationship between the two stations is shown in example Fig. 4.

Two important conditions need to be satisfied to obtain a high degree of relationship between the stage data of adjacent stations. These are:

- **No major tributary joins the main stream in between the two adjacent stations.**
- **Time of travel of the flood wave between the two stations is taken into consideration.**

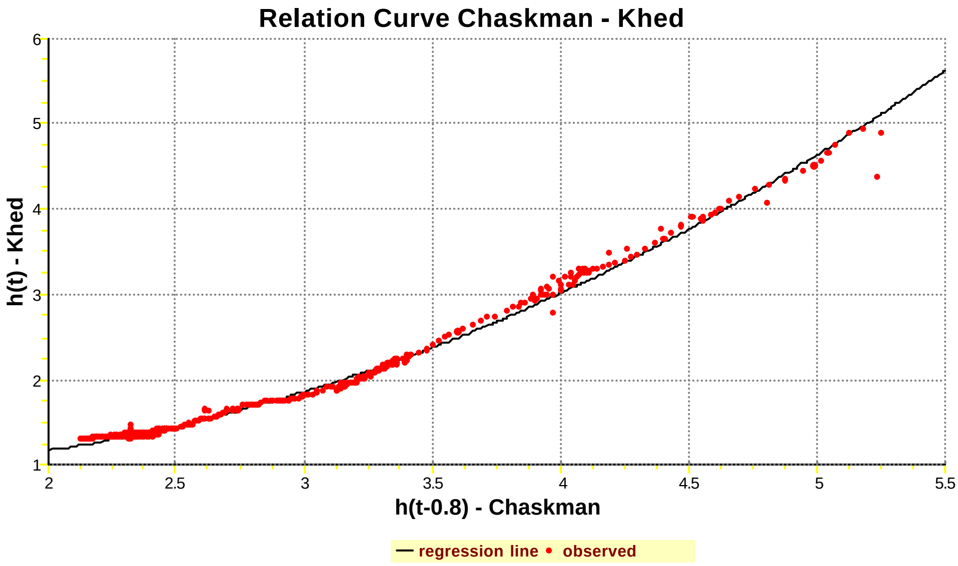

Figure 4 Example of relation curve

As noted above for comparative hydrograph plots, the occurrence of lateral inflow between stations limits the quality of the relationship between neighbouring stations. The lateral inflow may occur as a main tributary inflow or distributed over the reach as surface and groundwater inflows. In either case if it is a significant proportion of the downstream flow or variable, then good correlation may not be obtained.

4.3 Determination of travel time

With respect to the second condition, **the relationship between the two station time series must incorporate a time shift, representing the mean travel time of a flood wave between the stations**. Plotting without a time shift results in a looped relationship as shown in Example 4.1. The time shift may be assessed using:

- physical reasoning, or
- from an analysis of the time series.

<u>Example 4.1</u>

A relation curve is to be established between the stations MAHEMDABAD and NSB00I7. The hydrographs for the period for which the relation curve is to be established is presented in Figure 4.1a. It is observed that the hydrograph at the downstream site is lagging behind. Plotting the observations at the two sites without a correction for travel time of the flood wave results in a looped scatter plot vas displayed in Figure 4.1b

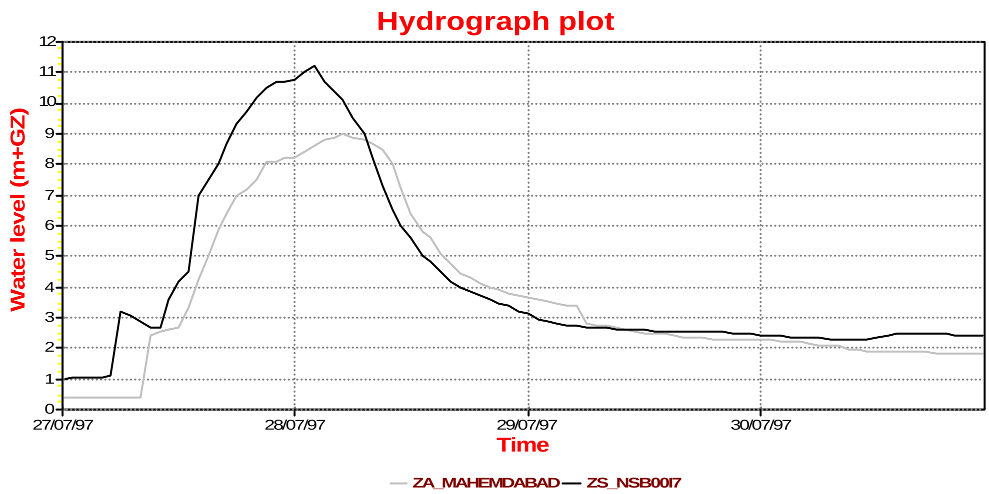

Figure 4.1a Hydrographs along WATRAK

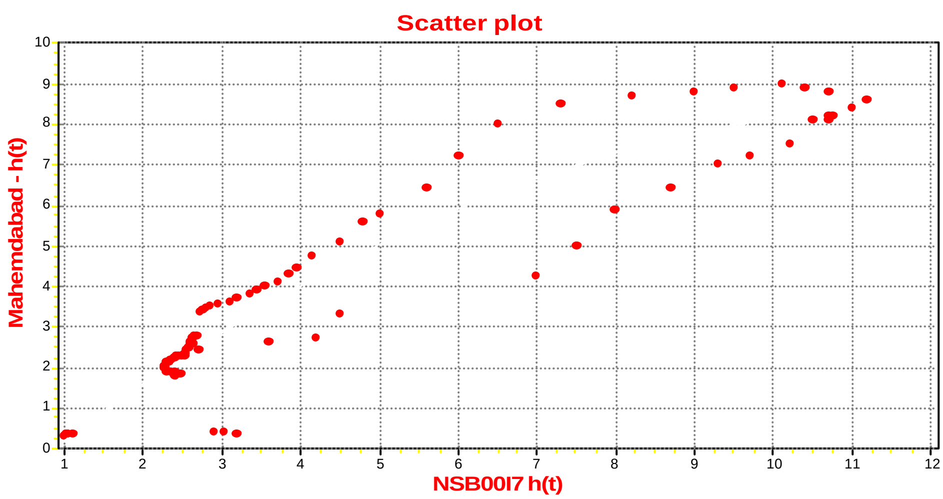

Figure 4.1b Scatter plot of h(t) Mahemdabad versus h(t) NSB00I7

4.3.1 From physical reasoning

The time of travel of a flood wave can be approximately determined by the division of the interstation distance by the estimated flood wave velocity. Consider a relation curve of daily water levels of stations X and Y which are spaced at a distance s km from each other along the same river, X upstream of Y. Let the average flow velocity (assessed from current meter gaugings at the stations) be u m/s, then the propagation velocity or celerity of the flood wave c is approximately equal to 1.7 (B_r/B_s).u, where B_r is the river width and B_s is the storage width (river+flood plain). So, if the river is in-bank, celerity becomes equal to 1.7u, and the time shift to be applied between X and Y for a one to one relationship amounts to distance/celerity = -s * 1000/(1.7 u) sec. When the river expands in the flood plain, the celerity is different from 1.7.u and should be multiplied with B_r/B_s ; consequently, a different time shift will result.

.*4.3.2 From cross correlation analysis*

Another computational procedure to derive the time shift between the two related series is based on cross-correlation analysis. The estimate of the cross-correlation function (R_{xy} t) between the time series X(t) and Y(t) of two stations is computed for different lags τ. The lag time corresponding to the maximum of the cross correlation function indicates the travel time between the two stations.

A plot is made between the lag time and the related cross correlation coefficient . The lag time corresponding to the maximum of the cross-correlation function indicates the travel time between the two cross sections. After calculating the time shift, it can be applied by entering the data of station X at time T and corresponding it with the data of station Y at time T + τ. It is then advisable to plot the resulting X-Y relationship:

- to inspect for the absence of looping,
- to detect outliers
- to determine the existence of breakpoints in the relationship
- to assess the functional form of the relationship.

Example 4.1 (continued)

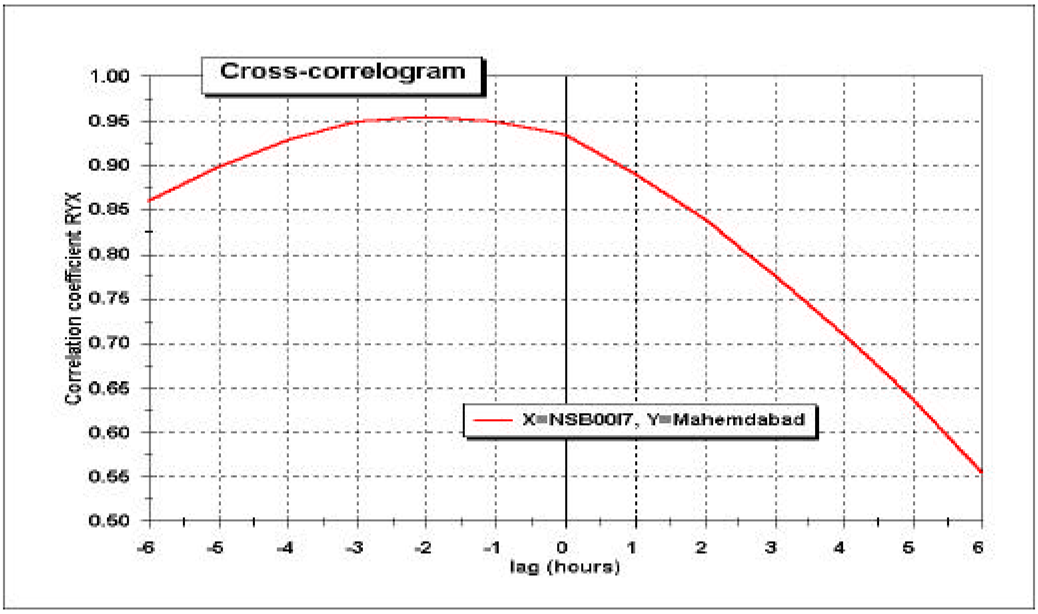

Figure 4.1c Cross-correlation function to determine optimum time lag

Application of cross-correlation analysis to the series of MAHEMDABAD (Y-series) and NSB00I7 (X-series) reveals an optimum time shift of –2.8 hrs. From hydraulic calculations given a Manning roughness K_m = 33, a bed slope S = $3.6x10^{-4}$, a flow depth h = 5 m, $B_r = B_s$ and L = 11 km one obtains roughly a travel time of 3 hrs (execute the calculation!) By applying such a time shift (i.e. h(t) MAHEMDABAD versus h(t-2.8) NSB00I7 the scatter shows only a limited looping left, see Figure 4.1d (exact elimination is not possible as the optimum time shift varies with the celerity, i.e. with the flow rate and varying lateral inflow also distorts the relationship).

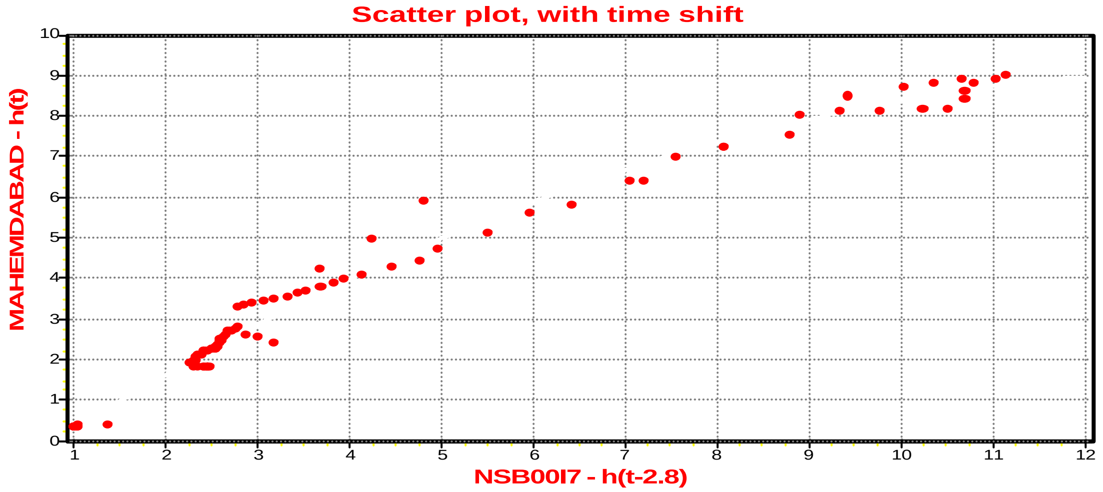

Figure 4.1d Scatter plot with corrected for time lag

4.4 Fitting the relation curve

The X-Y data can be fitted in HYMOS using a relation equation of polynomial type:

$$Y_t = c_0 + c_1\, X_{t+t1} + c_2\, X_{t+t1}^{\;2} + c_3\, X_{t+t1}^{\;3} + ...$$

After inspection of the scatter plot the form of the relationship (the degree of the polynomial) can be decided. It is recommended to use a polynomial not greater than order 2 or 3; in many instances a simple linear relationship will be acceptable. The least squares principle is applied to estimate the coefficients.

Where inspection of the scatter plot indicates the presence of breakpoints, then separate relationships may be established for different ranges of X (analogous to different ranges of the stage discharge relationship) with a maximum of 3 intervals of X.

Example 4.1 (continued)

The scatter plot shown in Figure 4.1d is fitted by a second order polynomial and is displayed in Figure 4.1el

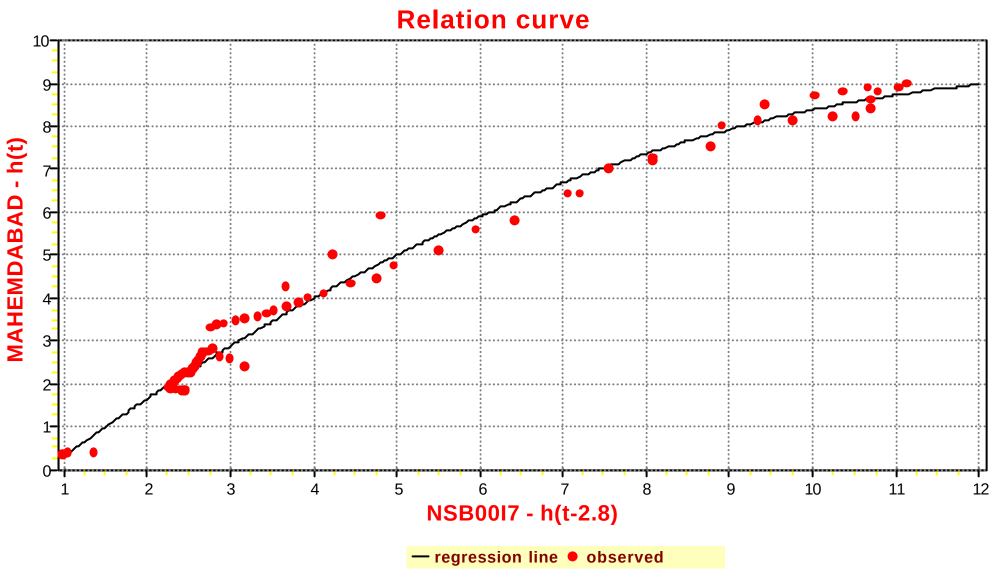

Figure 4.1e Fitting of relation curve

4.5 Using relation curve for data validation

Relation curves can be displayed and plotted and relation equations valid for different periods can be compared. Shifts in the relationship between the two series indicate a physical change at one of the stations, such as shifts in gauge zero, changes in cross section or even relocation of station.

Where such shifts in relation curve are accompanied by changes in the stage discharge relationship at one station, the changed relation curve is acceptable. However, where no such accompanying change in stage discharge has been notified, an explanation should be sought from Sub-divisional staff or the field supervisor.

Example 4.2

To investigate any change in the relationship between the observations at MAHEMDABAD and NSB00I7 a relation curve was established for the next flood wave on WATRAK river. The hydrographs and relation curve for this second flood are shown in the Figures 4.2a and 4.2b. Next the this relation curve is compared with the one established above. Note that for the establishment of the second relation curve the same time shift was applied as before. The two relation curves are shown in Figure 4.2c. It is observed that the match between the two curves is almost perfect, indicating that the no gauge shifts nor changes in the control sections did take place during the passage of the floods. A shift in the gauge at one of the sites would have resulted in a shifted relation as displayed in Figure 4.2d

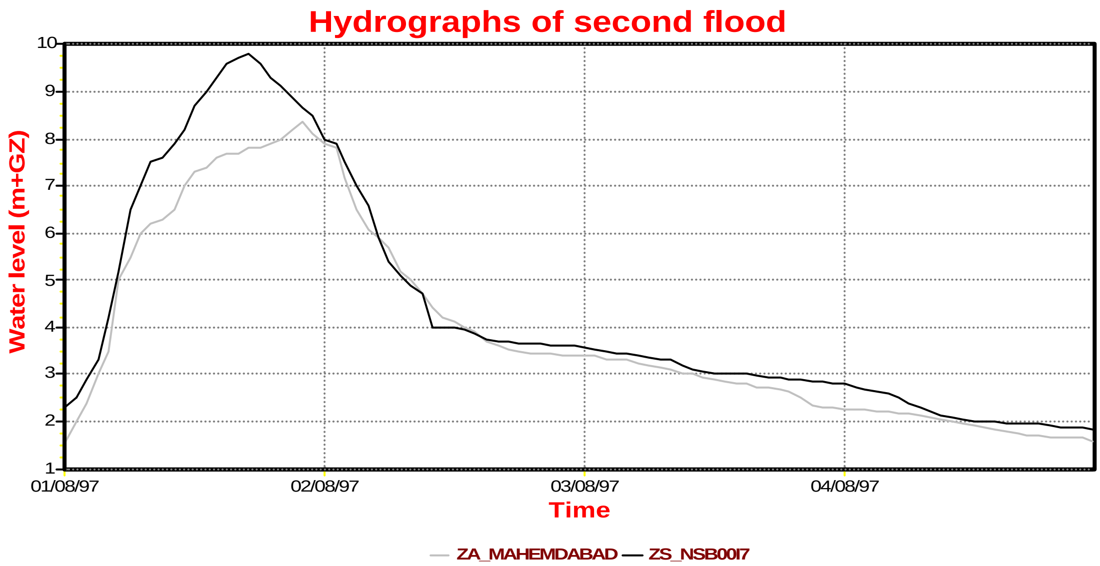

Figure 4.2a Hydrograph of second flood

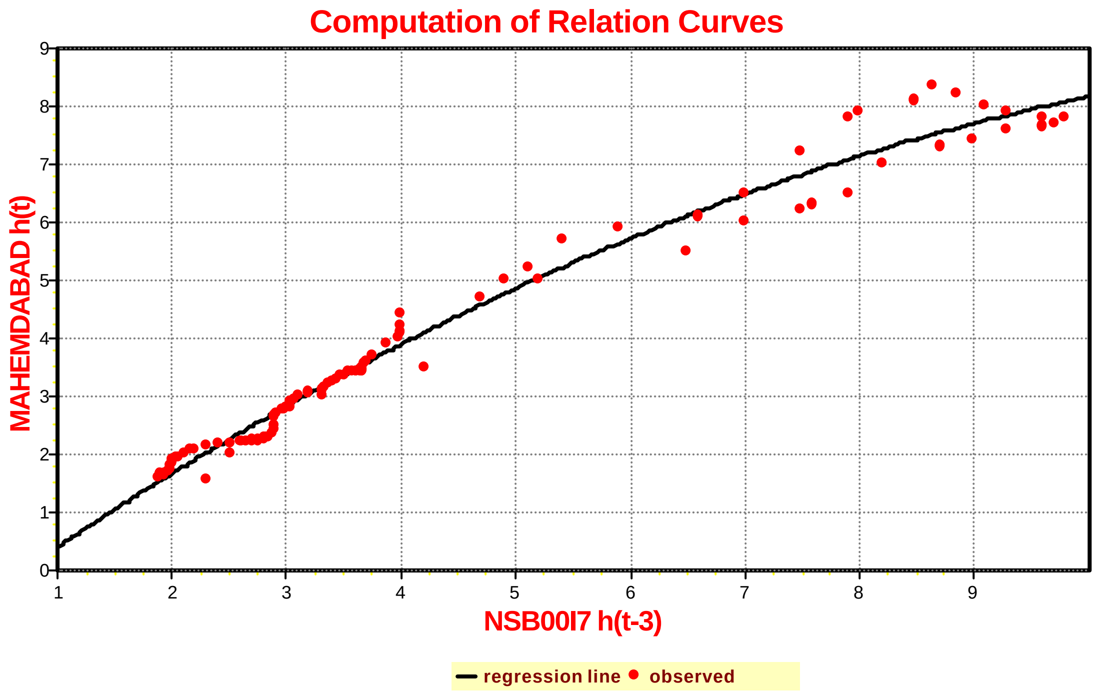

Figure 4.2b Relation curve of second flood

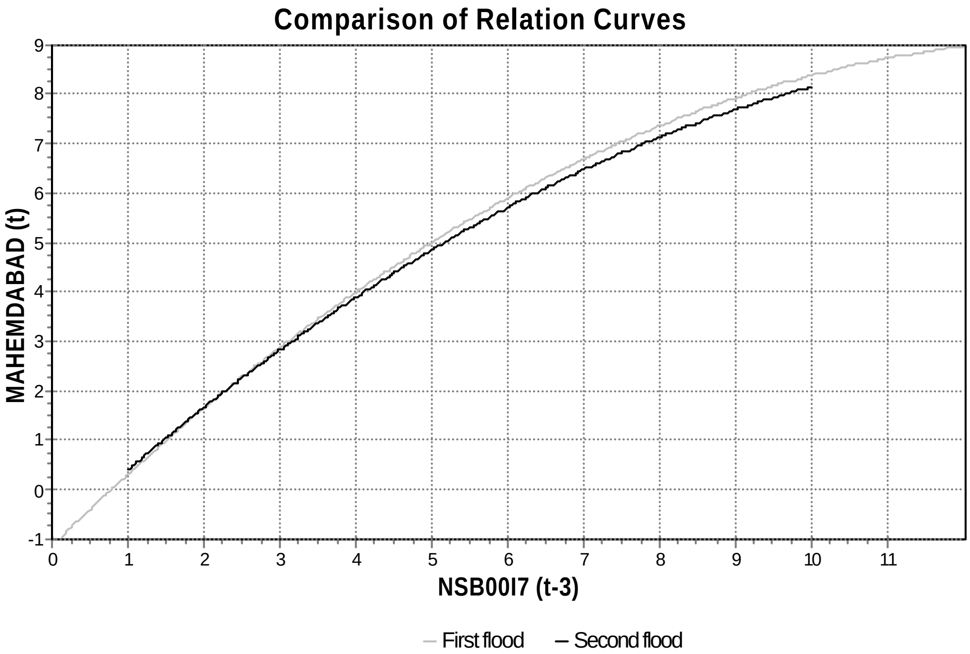

Figure 4.2c Comparison of relation curve established for different periods

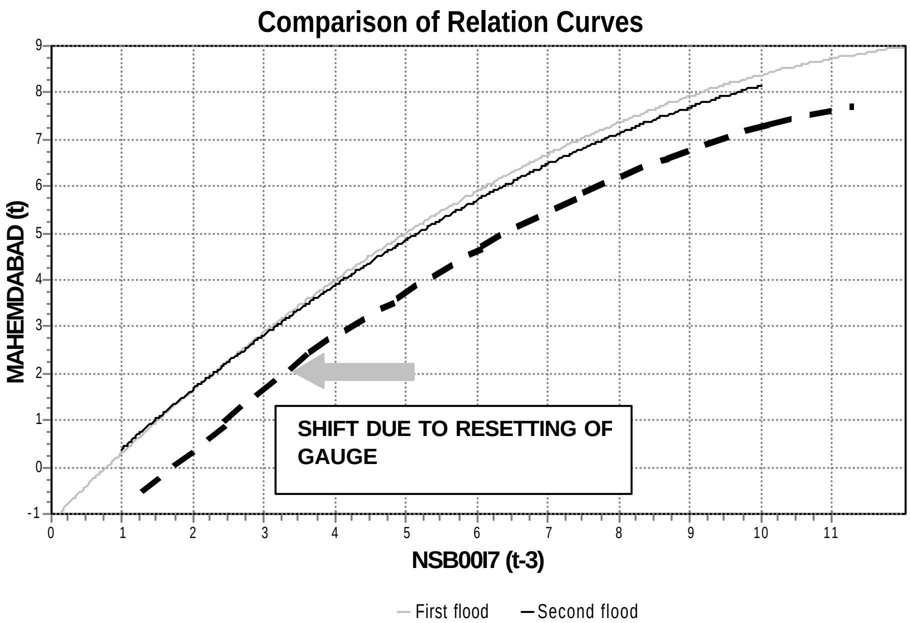

Figure 4.2.d Typical behaviour of relation curve in case of shifts in gauge zero

Summing up:

All steps in secondary validation are illustrated for data of stations CHASKMAN and KHED in BHIMA catchment with HYMOS.

- Print water level data of Chaskman and Khed, Period 1/8/97 – 31/8/97
- Plot the water level series of both stations
- Plot the hydrograph of Chaskman with rainfall observed at Chaskman
- Edit data (i.e. entry errors)
- Make a scatter plot of period 31/7-21/8/97 (shift – 0.8 hrs)
- Fit a relation curve and save the relation
- Apply the relation to Period 22/8 – 13/9 (force a time shift of –0.8 hrs)