HYDROLOGY PROJECT

Technical Assistance

World Bank & Government of The Netherlands funded

Training module # SWDP - 08

How to carry out primary validation for rainfall 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

Page

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 primary validation for rainfall data
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: • Perform primary validation of rainfall data
Key concepts	:	• Tabular and graphical scrutiny of rainfall data • Comparison of data from two equipment at the same site • Validation against various data limits • Qualitative comparison
Training methods	:	Lecture, exercises, 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*** • Primary validation of rainfall data	5 min	 OHS 1
2	***Instruments and observational methods*** • General • Daily rain gauge - SRG • Autographic raingauge - ARG • Tipping bucket raingauge - TBR	10 min	 OHS 2 OHS 3 OHS 4 OHS 5
3	***Comparison of daily time series from independent sources*** • General • Example 1a - Graphical • Example 1b - Tabular • Example 2a - Tabular • Example 2b - Tabular • SRG - ARG: Possible scenarios	10 min	 OHS 6 OHS 7 OHS 8 OHS 9 OHS 10 OHS 11
4	***Checking against maximum and minimum limits*** • Maximum data limit • Example - Graphical	5 min	 OHS 12 OHS 13
5	***Checking against upper warning limits*** • Upper warning level • Example - Graphical	5 min	 OHS 14 OHS 15

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 primary validation for rainfall data

1. General

- **Data validation is the means by which data are checked to ensure that the final figure stored in the HIS is the best possible representation of the true value of the variable at the measurement site at a given time or in a given interval of time.** Validation recognises that values observed or measured in the field are subject to errors which may be random, systematic or spurious.

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

- **Primary validation of rainfall data will be carried out at the Sub-divisional level using Primary module of dedicated data processing software and is concerned with data comparisons at a single station:**

  - ❖ for a single data series, between individual observations and pre-set physical limits
  - ❖ between two measurements of a variable at a single station, e.g. daily rainfall from a daily gauge and an accumulated total from a recording gauge

- **Data entry checks will already have been carried out to ensure that there have been no transcription errors from the field sheets to the SWDES database. Some doubtful values may already have been flagged.**

- **The high degree of spatial and temporal variability of rainfall compared to other climatic variables makes validation of rainfall more difficult.** This is particularly the case on the Indian sub-continent, experiencing a monsoon type of climate involving convective precipitaion.

- **More comprehensive checks can be carried out on daily and longer duration rainfall data by making comparisons with neighbouring stations.** This will be described under secondary validation and carried out at Divisional offices.

2. Instruments and observational methods

- **The method of measurement or observation influences our view of why the data are suspect. To understand the source of errors we must understand the method of measurement or observation in the field and the typical errors of given instruments and techniques.**

- **Data validation must never be considered a purely statistical or mathematical exercise. Staff involved in it must understand the field practice.**

- **Three basic instruments are in use at climatological stations for measurement of daily and shorted duration rainfall:**

  - ❖ standard daily raingauge
  - ❖ syphon gauge with chart recorder
  - ❖ tipping bucket gauge with digital recorder

These will be separately described with respect to the typical errors that occur with each gauge or observation method, and the means by which errors may be detected (if at all).

2.1 Daily rainfall gauge (SRG)

2.1.1 Instrument and procedure

Daily rainfall is measured using the familiar standard gauge (SRG). This consists of:

- **a circular collector funnel** with a brass or gun metal rim and a collection area of either 200 cm^2 (diameter 159.5 mm) or 100 cm^2 (diameter 112.8 mm), leading to a,
- **base unit**, partly embedded in the ground and containing,
- **a polythene collector bottle**

The gauge is read once or twice daily and any rain held in the polythene collector is poured into a **measuring glass** to determine rainfall in millimetres.

2.1.2 Typical measurement errors

- Observer reads measuring glass incorrectly
- Observer enters amount incorrectly in the field sheet
- Observer reads gauge at the wrong time (the correct amount may thus be allocated to the wrong day
- Observer enters amount to the wrong day
- Observer uses wrong measuring glass (i.e., 200 cm^2 glass for 100 cm^2 gauge, giving half the true rainfall or 100 cm^2 glass for 200 cm^2 gauge giving twice the true rainfall.
- Observed total exceeds the capacity of the gauge.
- Instrument fault - gauge rim damaged so that collection area is affected
- Instrument fault - blockage in raingauge funnel so that water does not reach collection bottle and may overflow or be affected by evaporation
- Instrument fault - damaged or broken collector bottle and leakage from gauge

It may readily be perceived that errors from most of these sources will be very difficult to detect from the single record of the standard raingauge, unless there has been a gross error in reading or transcribing the value. These are described below for upper warning and maximum limits.

Errors at a station are more readily detected if there is a concurrent record from an autographic raingauge (ARG) or from a digital record obtained from a tipping bucket raingauge (TBRG). As these too are subject to errors (of a different type), comparisons with the daily raingauge will follow the descriptions of errors for these gauges.

The final check by comparison with raingauges at neighbouring stations will show up further anomalies, especially for those stations which do not have an autographic or digital raingauge at the site. This is carried out under secondary validation at the Divisional office where more gauges are available for comparison.

2.2 Autographic raingauge (natural syphon)

2.2.1 Instrument and procedure

In the past short period rainfall has been measured almost universally using the natural syphon raingauge. **The natural syphon raingauge consists of the following parts:**

- **a circular collector funnel** with a gun metal rim, 324 cm^2 in area and 200 mm in diameter and set at 750 mm above ground level, leading to ,
- **a float chamber** in which is located a float which rises with rainfall entering the chamber,

- **a syphon chamber** is attached to the float chamber and syphon action is initiated when the float rises to a given level. The float travel from syphon action to the next represents 10 mm rainfall
- **a float spindle** projects from the top of the float to which is attached,
- **a pen** which records on,
- **a chart** placed on
- **a clock drum** with a mechanical clock.

The chart is changed daily at the principal recording hour. During periods of dry weather the rainfall traces a horizontal line on the chart; during rainfall it produces a sloping line, the steepness of which defines the intensity of rainfall. The chart is graduated in hours and the observer extracts the hourly totals from the chart and enters it in a register and computes the daily total.

2.2.2 Typical measurement errors

Potential measurement faults are now primarily instrumental rather than caused by the observer and include the following:

- Funnel is blocked or partly blocked so that water enters the float chamber at a different rate from the rate of rainfall
- Float is imperfectly adjusted so that it syphons at a rainfall volume different from 10 mm.
- In very heavy rainfall the float rises and syphons so frequently that individual pen traces cannot be distinguished.
- Clock stops; rainfall not recorded or clock is either slow or fast and thus timings are incorrect.
- Float sticks in float chamber; rainfall not recorded or recorded incorrectly.
- Observer extracts information incorrectly from the pen trace.

In addition differences may arise from the daily raingauge due to different exposure conditions arising from the effect of different level of the rim and larger diameter of collector. It has been traditional to give priority to the daily SRG where there is a discrepancy between the two.

2.3 Tipping bucket raingauge

2.3.1 Instrument and procedures

Short period rainfall is more readily digitised using a tipping bucket raingauge. It consists of the following components.

- **A circular collector funnel** with a brass or gunmetal rim of differing diameters, leading to a
- **Tipping bucket arrangement** which sits on a knife edge. It fills on one side, then tips filling the second side and so on.
- **A reed switch** actuated by a magnet registers the occurrence of each tip
- **A logger** records the occurrence of each tip and places a time stamp with the occurrence

The logger stores the rainfall record over an extended period and may be downloaded as required. The logger may rearrange the record from a non-equidistant series of tip times to an equidistant series with amounts at selected intervals. The digital record thus does not require the intervention of the field observer. For field calibration, a known amount of rainfall is periodically poured into the collector funnel and checked against the number of tips registered by the instrument.

2.3.2 Typical measurement errors

- Funnel is blocked or partly blocked so that water enters the tipping buckets at a different rate from the rate of rainfall
- buckets are damaged or out of balance so that they do not record their specified tip volume
- reed switch fails to register tips
- reed switch double registers rainfall tips as bucket bounces after tip. (better equipment includes a debounce filter to eliminate double registration.
- failure of electronics due to lightning strike etc. (though lightning protection usually provided)
- incorrect set up of measurement parameters by the observer or field supervisor

Differences may arise from the daily raingauge (SRG) for reasons of different exposure conditions in the same way as the autographic raingauge.

3. Comparison of daily time series for manual and autographic or digital data

3.1 General description:

At stations where rainfall is measured at short durations using an autographic or a digital recorder, a standard raingauge is always also available. Thus, at such observation stations rainfall data at daily time interval is available from two independent sources. The rainfall data at hourly or smaller interval is aggregated at the daily level and then a comparison is made between the two. Differences which are less than 5% can be attributed to exposure, instrument accuracy and precision in tabulating the analogue records and are ignored. Any appreciable difference (more than 5%) between the two values must be probed further. The observation made using standard raingauge has generally been taken as comparatively more reliable. This is based on the assumption that there is higher degree of possibility of malfunctioning of autographic or digital recorders owing to their mechanical and electromechanical systems. However, significant systematic or random errors are also possible in the daily raingauge as shown above.

If the error is in the autographic or digital records then it must be possible to relate it either to instrumental or observational errors. Moreover, such errors tend to repeat under similar circumstances.

3.2 Data validation procedure and follow up actions:

This type of validation can be carried out in tabular or graphical form. For both approaches, the values of hourly data are aggregated to daily values to correspond to those observed using a standard raingauge. A comparison is made between the daily rainfall observed using standard and automatic gauges. Percent discrepancy can be shown by having a second axis on the plot. Tabular output for those days for which the discrepancy is more than 5% can be obtained. A visual inspection of such a tabulated output will ensure screening of all the suspect data with respect to this type of discrepancy.

The following provides a diagnosis of the likely sources of error with discrepancies of different sorts:

(a) Where the recording gauge gives a consistently higher or lower total than the daily gauge, then the recording gauge could be out of calibration and either tipping buckets (TBRG) or floats (ARG) need recalibration.

Accept SRG and adjust ARG or TBRG

(b) Where agreement is generally good but difference increases in high intensity rainfall suggests that for the ARG:

- the syphon is working imperfectly in high rainfall, or
- the chart trace is too close to distinguish each 10 mm trace (underestimate by multiples of 10 mm)

For the TBRG:

- gauge is affected by bounce sometimes giving double tips

Accept SRG and adjust ARG or TBRG

(c) Where a day of positive discrepancy is followed by a negative discrepancy and rainfall at the recording gauge was occurring at the observation hour, then it is probable that the observer read the SRG at a different time from the ARG. The sum of SRG readings for successive days should equal the two-day total for the TBRG or ARG

Accept TBRG or ARG and adjust SRG

(d) Where the agreement is generally good but isolated days have significant differences, then the entered hourly data should be checked against the manuscript values received from the field. Entries resulting from incorrect entry are corrected. Check that water added to the TBRG for calibration is not included in rainfall total. Otherwise there is probable error in the SRG observation.

Accept ARG or TBRG and adjust SRG

However, it will be very interesting to note that in certain cases the values reported for the daily rainfall by SRG and ARG matches one to one on all days for considerable period not withstanding the higher rainfall values etc. It is very easy to infer in those situations that there has been at attempt by the observer to match these values forcefully by manipulating one or both data series. It is not expected that both these data series should exactly match in magnitude. Since such variation must be there due to variance in the catch and instrumental and observation variations. And it is therefore highly undesirable that such matching is effected by manipulation.

Example: 3.1:
Consider the daily totals of hourly rainfall (observed by an autographic raingauge - ARG) and the daily rainfall observed by the standard raingauge (SRG) at station ASKHEDA of PARGAON catchment. The graphical and tabular comparison of these two data series for the period from 1/9/1996 to 31/10/99 is given in Fig. 3.1 and Table 3.1 respectively.

It is amply clear from this graphical and tabular outputs that there has been a marked difference between the reported daily rainfall as observed from the standard raingauge and that obtained by compiling the hourly values, tabulated from autographic chart, to daily level.

Table: 3.1: Tabular comparison of daily rainfall obtained from SRG and ARG at the same station

Year	Month	Day	ASKHEDA		
			SRG	ARG	% Diff.
1996	9	1	0	0	-
1996	9	2	0	0	-
1996	9	3	0	0	-
1996	9	4	0	0	-
1996	9	5	0	0	-
1996	9	6	18.7	18.5	-1.1
1996	9	7	0	0	-
1996	9	8	3.7	4	8.1
1996	9	9	0	0.2	-
1996	9	10	0	0	-
1996	9	11	0	0	-
1996	9	12	0	0	-
1996	9	13	5	0	-100.0
1996	9	14	0	4.8	-
1996	9	15	3.9	0	-100.0
1996	9	16	3.8	4.8	26.3
1996	9	17	7.2	3.5	-51.4
1996	9	18	0	6.9	-
1996	9	19	2	0	-100.0
1996	9	20	0	2	-
1996	9	21	0	0	-
1996	9	22	0	0	-
1996	9	23	14	0	-100.0
1996	9	24	13.2	14.8	12.1
1996	9	25	3.8	13.5	255.3
1996	9	26	6.8	3.5	-48.5
1996	9	27	3	7.2	140.0
1996	9	28	0	3	-
1996	9	29	2.7	0	-100.0
1996	9	30	0	2.4	-
1996	10	1	0	0	-
1996	10	2	19	18.3	-3.7
1996	10	3	75.8	0.5	-99.3
1996	10	4	2.8	1.5	-46.4
1996	10	5	4	4.5	12.5
1996	10	6	0	0	-
1996	10	7	0	0	-
1996	10	8	0	0	-
1996	10	9	0	0	-
1996	10	10	0	0	-
1996	10	11	0	0	-
1996	10	12	0	0	-
1996	10	13	0	0	-
1996	10	14	0	0	-
1996	10	15	0	0	-
1996	10	16	0	0	-
1996	10	17	0	0	-
1996	10	18	0	0	-
1996	10	19	0	0	-
1996	10	20	2	1.8	-10.0
1996	10	21	50.3	50.1	-0.4
1996	10	22	0.7	1.5	114.3
1996	10	23	70	70.5	0.7
1996	10	24	9	8.9	-1.1
1996	10	25	0	0	-
1996	10	26	0	0	-
1996	10	27	0	0	-
1996	10	28	0	0	-
1996	10	29	6	6	0.0
1996	10	30	0	0	-
1996	10	31	0	0	-

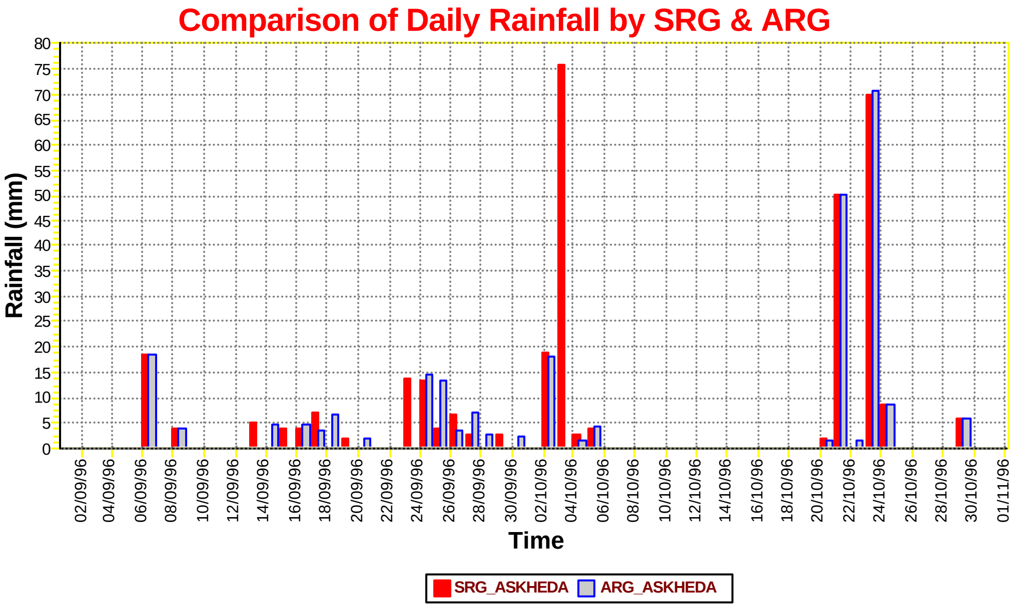

Fig. 3.1: Graphical comparison of daily rainfall obtained from SRG and ARG at the same station.

Following points can be noticed:

a) The difference in daily values from SRG and ARG varies considerably; from a very reasonable deviation like 1.1, 0.4, 0.7 % (on 6/9/96, 21/10/96 and 23/10/96 respectively) to unacceptably high values like 51.4, 255.3, 99.3 % (on 17/9/96, 25/9/96 and 3/10/96 respectively).

b) In this example, the resulting errors can be categorised in three types major classes:

- There are many instances where a larger difference is caused by the shifting of one of the data series by one day. From 13/9/96 to upto 31/9/96 a shift of one day in one of the series can be very clearly noticed. This shift is not present before and after this period. Such errors are not exactly due the differences in the two observations but are the result of recording or entering one of the data series inappropriately against wrong date. However, even if this time shift were not present, then also there would have been substantial differences in the corresponding values as can be easily inferred from the tabulated values.
- There are a few instances where the difference is due to mistake in recording or entering or might even be due to failure of ARG. Such differences like the one on 3/10/96 where SRG record shows 75.8 mm whereas ARG data shows 0.5 are clear cases of mistakes. Such errors are very easy to be detected also.
- There are lot of instances where the percent difference is moderate to high which can be attributed to observational errors, instrumental errors and the variation in the catch in the two raingauge. Most of these high percentage differences are for the very low rainfall values which also highlights the variation in the catch or the accuracy of equipment at such low rainfall events.

Following actions must be taken as a follow-up of data validation:

a) The cause of the shift in one of the data series can be very easily detected and removed after looking at the dates of the ARG charts and corresponding tabulated data.

b) Cause of mistake like that on 3/10/96 can be removed if ARG chart also shows comparable rainfall. If ARG data is found correct according to the chart and there does not seem to be any reason to believe instrumental failure etc. then the daily rainfall as reported by the SRG can be corrected to correspond to the ARG value. Else, if there is any scope of ambiguity then the daily data has to be flagged and it has to be reviewed at the time of secondary validation on the basis on rainfall recorded at the adjoining stations.

c) Moderate to large differences (more than 5%) in the two data series are to be probed in detail by looking at the ARG chart and corresponding tabulations. Any errors in tabulations are to be corrected for. Inspection of the differences in this case shows that there is no particular systematic error involved. Sometime the SRG value is more by a few units and sometimes ARG is more by similar magnitude. Sometimes this might be due to observation SRG at non-standard times or incorrect tabulation of the ARG chart. At low rainfall these differences can also be due to variation in the catch or due to inaccuracy of the equipment. In both circumstances, it must be ensured whether standard equipment and exposure conditions are maintained at the station.

4. Checking against maximum and minimum data limits

4.1 General description:

Rainfall data, whether daily or hourly must be validated against limits within which it is expected to physically occur. Such limits are required to be quite wide to avoid the possibility of rejecting true extreme values. For rainfall data, it is obvious that no data can be less than zero which perfectly serves as the limiting minimum value. However, it is quite difficult to assign an absolute maximum limit for the rainfall data in a given duration occurring at a particular station. Nevertheless, on the basis of past experience and physical laws governing the process of rainfall it is possible to arrive at such maximum limits which in all probability will not be exceeded. The limit may be set as the maximum capacity of the raingauge, but care should be taken in rejecting values on this basis where the gauge observer has read the gauge several times to ensure the gauge capacity was not exceeded.

Maximum limits also vary spatially over India with climatic region and orography. Also, this maximum limit has a strong non-linear relationship with rainfall duration. For example, for any place, the maximum limit for daily rainfall is not equal to 24 times the maximum limit for hourly rainfall. It is certainly much lower than this amount. For this reason, it is essential to set maximum limits for rainfall for durations other than 1-day. Limits for 1-day and 1-hour should be set and this will generally be sufficient to identify gross errors over the intervening range of duration.

For 1-day duration, the India Meteorological Department and Indian Institute of Tropical Meteorology have prepared atlas for 1 day Probable Maximum Precipitation (PMP) which gives the expected maximum amount that can physically occur in a given duration at a given location. Values extracted from this map should be applied or else could be derived as the historical maximum value from the long term records now available for most of the regions. Though there might be some variation in the values obtained from both these atlases but for the purpose of prescribing the maximum limit here such differences may be ignored. Alternatively, the derived information on observed maximum 1-Day point rainfall, which is available for scores of stations across the country from long term records, can be used as a reasonably good estimate of the maximum limit of rainfall.

Similarly, use can be made of the maps showing 50 year – 1 hour maximum rainfall, as developed by India Meteorological Department, for presecribing the maximum data limit for the case of hourly rainfall data. Such initial estimates can be adjusted on the basis of local judgement adjusted on the basis of experience or of local research studies based on either:

- storm maximisation by considering precipitable moisture and inflow of moist air
- statistical analysis of observed extreme values for shorter durations.

4.2 Data validation procedure and follow up actions:

Setting minimum and maximum limits in SWDES ensures filtering of values outside the specified limits. Such values are considered suspect. They are first checked against manuscript entries and corrected if necessary. If manuscript and entry agree and fall outside prescribed limits, the value is flagged as doubtful. Where there are some other corroborative facts about such incidents, available in manuscript or notes of the observer or supervisor then they must be incorporated with the primary data validation report. This value then has to be probed further at the time of secondary data validation when more data from adjoining stations become available.

However, when the current data is being entered in subsequent month and there happens a entries which are more than the prescribed maximum limit and such heavier rainfall events are also very fresh in the minds of data processing staff then it implies that the earlier maximum has been crossed in reality. In such circumstances the maximum limit is reset to a suitable higher value. In such cases there will surely be a few adjoining stations recording similar higher rainfall and will support such inferences. However, if no such basis is available for justification of such very high values then the value is reported in the form of a remark which can be reviewed at the secondary validation stage.

Example: 4.1:

Consider the long term daily rainfall at MEGHARAJ station in KHEDA catchment as shown in Fig. 4.1.

This is a long term rainfall data of the station (37 years) and it can be seen that the maximum daily rainfall in any year has usually been about 100 mm on an average. A few times the daily rainfall has been more than 150 mm and only thrice in a period of 37 years the value has exceeded 200 mm. Only once the value has exceeded 250 mm. So, if value of about 320 – 325 mm is taken for setting the maximum limit for daily rainfall at this station then it will help in validating if certain data values are beyond this value.

Such values can be derived from the isolines of maps giving observed maximum 1-Day rainfall or 1-Day PMP values.

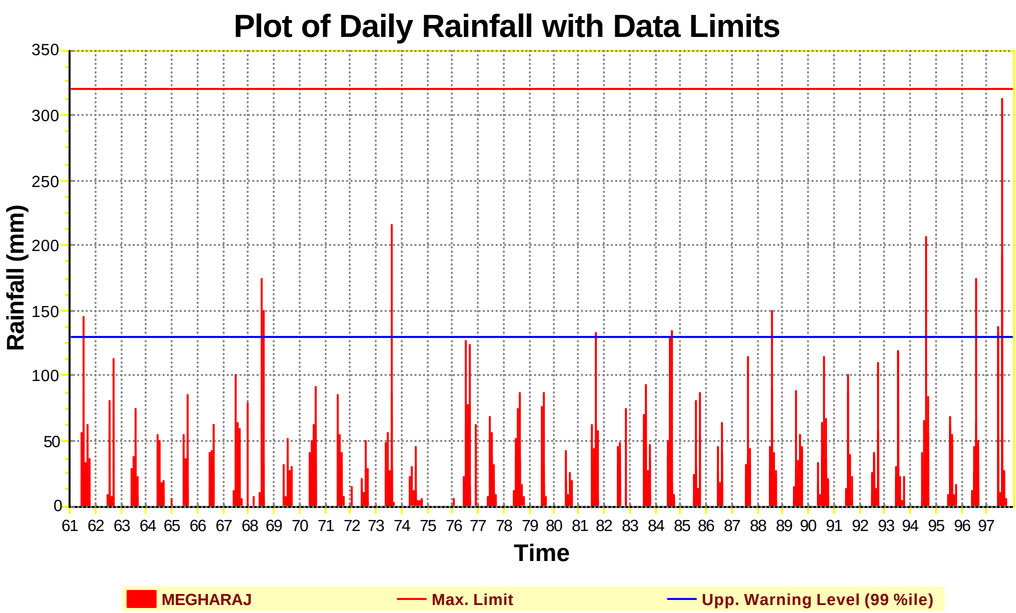

Fig. 4.1: Graphical plot showing physical significance of maximum limit and upper warning level

5. Checking against upper warning level

5.1 General description:

Validation of rainfall data against an absolute maximum value does not discriminate those comparatively frequently occurring erroneous data which are less than the prescribed maximum limit. In view of this, it is advantageous to consider one more limit, called the upper warning level, which can be employed to see if any of the data value has violated it. This limit is assigned a value with an intention of screening out those high data values which are not expected to occur frequently. The underlying purpose of carrying out such a test is to consider a few high data values with suspicion and subsequently scrutinising them.

For the daily rainfall data this limit can be set statistically, for example, equal to 99 percentile value of the actual rainfall values excluding zero rainfall values. In other words, it leads to screening out those values which are higher than that daily rainfall value which is exceeded only once in 100 rainfall events on an average. Say, if the data for 30 years is available and there are 3456 non-zero daily rainfall values then 99 percentile value will be about 34^{th} highest in the lot. It indicates that rainfall value which is equalled or exceeded, on an average, once in every 100 rainfall events.

Similar statistic can be employed for obtaining suitable value for upper warning level for hourly rainfall data. The central idea while setting these upper warning levels is that the higher rainfall data is screened adequately, that is the limits must be such that it results in not too many and not too less data values being flagged for validation.

5.2 Data validation procedure and follow up actions:

Setting warning limits in the Primary module (like maximum limits) results in filtering values outside specified limits. Values are checked against manuscript entries and corrected if necessary. Remaining values are flagged as doubtful, and any associated field notes or corroborative facts are incorporated with the primary validation report and forwarded to the Divisional Data Processing Centre for secondary validation.

<u>Example: 5.1</u>:
Same example as mentioned for example 2 can be considered for illustrating the meaning of another data limits as Upper warning level (see Fig. 4.1)

As can be seen that a value of 130 mm is taken as the Upper warning level and that this is crossed in as many as 9 years and on about 12-13 instances. Now this frequency of violation of data limit is such that on an average once in 3-4 years certain higher values will be checked for its validity. The limit of 130 mm is worked out here on the basis of 99 percentile value of the set of non-zero rainfall value in this long period of 37 years. Such a statistic provides a value which has been exceeded only once in 100 occasions of rainfall event. Normally, for this station there are about 50-60 rainy days in any year and 99 %ile means that, on an average such value will be exceeded about once in 2 years time. Flagging of all those instances is done which are larger than this value of 130 mm so that cross checking on such large values can be ensured.