J6.3

Comparison of Manual and Automated Quality Control of Operational Hourly Precipitation Data of the National Weather Service

Dongsoo Kim*
NOAA National Climatic Data Center, Asheville, North Carolina

Edward Tollerud
NOAA-Research Earth System Research Laboratory, Boulder, Colorado

Steve Vasiloff
NOAA-Research National Severe Storm Laboratory, Norman, Oklahoma

Jason Caldwell
NOAA/NWS Lower Mississippi River Forecast Center, Slidell, Louisiana

1. INTRODUCTION

The National Weather Service (NWS) Office of Hydrologic Development operates the collection and dissemination of real-time Hydrometeorological Automated Data System (HADS) and other precipitation gauge data to users at River Forecast Centers (RFCs) and Weather Forecast Offices (WFOs). As most of the data are delivered to the users with minimal quality control (QC) in order to shorten data latency, forecasters at RFCs apply significant effort toward QC to insure proper hydrologic forecasting. Thus there is a great need for automated gauge QC. The NWS Lower Mississippi RFC (LMRFC) has been generating quality codes of hourly precipitation data available within the service area, and made available to WFOs and to the National Climatic Data Center (NCDC). The dataset (called "LOUZE" data) is unique as it represents a refined class of quality control algorithms. As a general rule, the results of manual editing are considered to be of significantly higher value and accuracy than those resulting from automated technique. But manual editing is susceptible to human inconsistencies, biases, and just plain errors that automated processes are not. Three NOAA agencies, NCDC, Earth System Research Laboratory (ESRL) and National Severe Storm Laboratory (NSSL), are engaged in automated QC practices for precipitation data. The NCDC's objective is to apply emulated manual QC algorithms to historical HADS hourly data (Kim et al. 2008); the NSSL's objective is to refine the next generation quantitative precipitation estimate (Q2; Vasiloff et al. 2007) to use radar-based QPE as a tool to improve QC of precipitation data; and the ESRL's Global System Division (GSD) focuses on the quality history of gauge stations used in precipitation assimilation and verification (Tollerud et al. 2005). Even though the objectives of these LOUZE data users are different, from these studies many lessons toward understanding of physical causes of gross errors are possible. Conversely, they can also provide feedback to forecasters who generate LOUZE data.

* *Corresponding author address:* Dongsoo Kim, NOAA/NESDIS/NCDC, 151 Patton Ave, Asheville, NC 28801; e-mail: dongsoo.kim@noaa.gov

2. ANALYSES OF LOUZE DATA

The LOUZE data generated by forecasters at the LMRFC are output of two screening levels of hourly precipitation data. As shown in table 2, the initial "L" is an abbreviation for improperly reported light precipitation (less than 0.1 inch), "O" is over-estimation, "U" is under-estimation, "Z" is inaccurate zero precipitation, and "E" is inaccurate extreme precipitation. The second level of screening is the decision to include the flagged value in hydrologic modeling. Hence, "I" stands for ignored, namely, used in hydrologic modeling, and "R" for rejected. Rejected gauges are monitored for removal from the "R" list.

Figure 1 shows the distribution of flags in each month of 2007. We find type "L" (light precipitation) is the dominant flagged occurrence during 2007. Such errors are typically reported following the period of intense rainfall, or when a gauge is mostly obstructed. The second dominant flag is "Z", namely false zero. Forecasters encode this flag when precipitation is observed on radar and by surrounding reports.

Figure 2 shows a diurnal pattern in the number of flags. The frequencies of total flags are normalized by dividing by the monthly total number of flags. The pattern of the cool season is more pronounced than the warm season. We observed that there is some degree of diurnal variability with increased frequency during the day (13-21 UTC; 7 AM – 3 PM in local time). It is difficult to conclude from the figure that erroneous observations actually do occur more often in daytime than nighttime, or if forecasters on duty are more attentive in decision making during the daytime. Of course, perceptions of individual forecasters are also different.

3. HOURLY PRECIPITATION DATA USED

The total number of rain gauge stations accessible by LMRFC as of October 2008 is 3646 (Fig. 3).

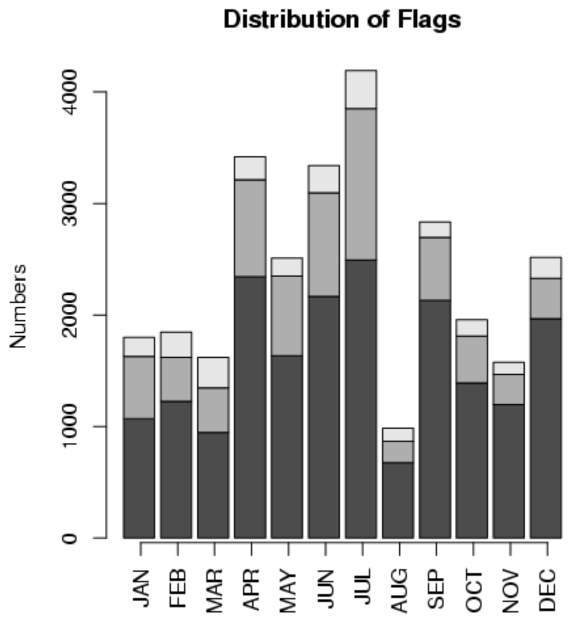

Fig.1. Monthly bar graph of frequencies with respect to flag types in 2007. Only two are dominant types (“L” for dark color, “Z” for medium gray, and light gray represent all of “O”, “U” and “E”).

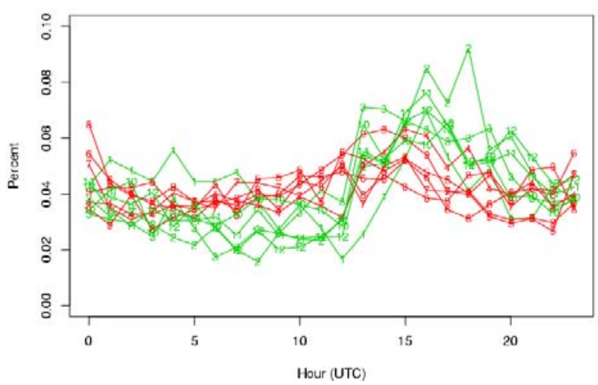

Fig.2. The distribution of frequencies by hour (UTC) and month as a percent. The green colors are cool season months (Oct – Mar) and the red colors are warm season months (Apr – Sep). The total number of flags at the given hour is divided by the monthly total number of flags, so that diurnal patterns can be compared with other months.

However, not all of them are used in the quality flagging process. Many are daily COOP stations not available in timely fashion, while the ingest of other gauge observations is delayed. Overall, we determined that 1662 stations were flagged at least once in 2007, and 1438 stations during the first 10 months of 2008.

The hourly precipitation dataset for which NCDC plans to encode quality flags is reprocessed HADS hourly data which excludes ASOS and Mesonet data. As explained by Kim et al. (2008), reprocessed HADS hourly precipitation data are not necessarily the same as real-time HADS data due to characteristics of the processing environment. The real-time HADS hourly

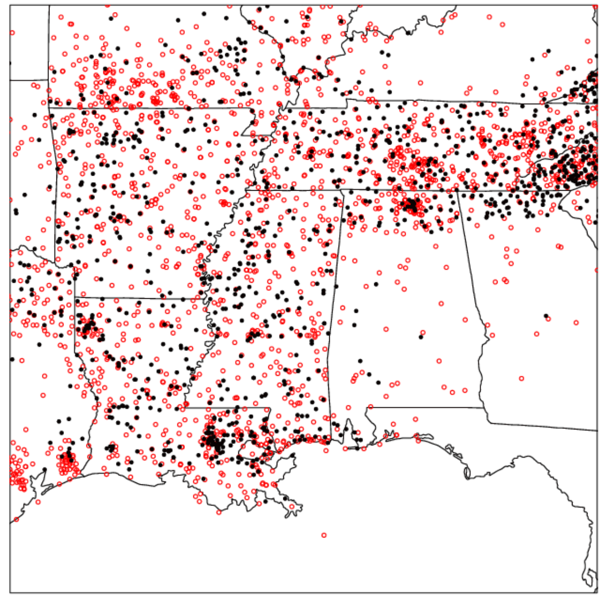

Fig.3. The gauge locations used by LMRFC as of October 2008 numbers 3646. The black dots designate the 1438 stations that received a QC flag at least once during the first 10 months in 2008. Many are daily COOP stations that do not come into the LOUZE process.

precipitation data along with ASOS data are served by National Center for Environmental Prediction (NCEP). Both NSSL and ESRL use hourly precipitation data from NCEP. Further, NSSL’s QVS limits use of off-the-top-of-the-hour gauge data to avoid inconsistency in observation time.

4. AUTOMATED QC ALGORITHMS

4.1 *NCDC HADS-QC Algorithm*

The NCDC’s HADS-QC algorithm is designed to complement reprocessed HADS data by providing quality guidance flags. The flags will help user’s decisions about data quality. The algorithm employs hierarchical quality assessment methods from spatial (+/- 0.5 deg) and temporal (+/- 1-hr) samples with exclusion of values at target stations. But we do include values at other time-levels at target stations. Let P be a precipitation value in inches. (1) For $P < 0.0$, $P \geq 4.0$ and any missing value, the quality code will be “9”. (2) For $P = 0$, we compute wet fraction of the samples, F_w. If $F_w > 0.25$ and the nearest raining station is within 0.2 degrees distance away, then quality code is set to “4”, but the nearest station is farther than 0.2 degree distance, the quality code is “3”. The code “4” will allow user to set the P as a missing value. Hence, the quality codes “3” and “4” refines LOUZE flag “Z”. If $F_w \leq 0.25$, then the quality code is set to “0” and the P value remains the same. (3) For $0 < P < 0.1$, if mean value of samples is zero, the quality code is set to “0”, but $P \geq 0.1$ the quality code will be set to “1”. The quality code

"0" will help user to reset the non-zero precipitation value to 0.0 with minimum risk. (4) For $0.1 \leq P < 4.0$, we follow the RFC operational QC algorithm (Kondragunta and Shrestha, 2006). We compute the mean absolute difference (MAD) between target values and median, and the difference between the 3rd and 1st quartiles (DIF). If the ratio of MAD and DIF is greater than a threshold value of 10, then we assign a quality code "2" to the target. The quality code "2" emulates LOUZE's "O", "U", and "E".

The number of neighbors is capped at 20 of the nearest stations, which provides a maximum of 62 samples. In data sparse areas or dry periods the quality assessment is less robust. Hence, the threshold values will have to be adjusted locally.

4.2 *Automated Daily QC at GSD/ESRL*

The GSD Quality control system (GSD/QC) is based on 24h total gage precipitation accumulations, between 1200-1200 UTC. It thus has the advantage of longer time periods and hence more meaningful neighbor comparisons. The accompanying disadvantage is that it cannot distinguish shorter-period data quality problems, and is usually more conservative than shorter-period QC methodologies. In the context of this paper, it is thus distinctively different from the NCDC and NSSL methods, but somewhat similar to the second-level LOUZE screening (either R or I, reject or accept). It consists of several individual algorithms that are customized to characteristics of the principal input HADS data stream, including checks for extreme observations, excessive missing hourly reports, stuck gauges manifested by repeating hourly observations patterns or zero-nonzero anomalies, and apparent faulty-zero reports. Verification for the 24h totals built from the hourly observations is provided by an independent set of daily observations that are of high quality and vetted daily at individual River Forecast Centers. A portion of these daily values are HADS or ASOS station totals that have also been screened at the RFCs. Prior to use as verification stations, these daily observations have been screened by internal neighbor checking and by some of the same extreme value checks that are subsequently applied to the hourly stations. A full description of these checks is available at (http://www-frd.fsl.noaa.gov/mab/sdb/readme.txt).

In terms of the number of gauges failed by this set of algorithms, those involving neighbor checks are by far the most important. A distinguishing characteristic of these checks is that they are binary in nature; that is, they are designed to identify stations that anomalously report zero precipitation in a field of non-zero reports, or vice versa. In some respects, this technique is more flexible than comparing amount categories, especially in fields of light precipitation when category differences are small and swamped by spatial variability, and thus difficult to interpret. Also, these binary checks are compatible with many verification tests that compare reports in something like this binary fashion. However, it does make these checks function differently than the other automated QC methodologies compared in this paper, and under some conditions when a gauge reads consistently well below or above neighbors but is not obviously in error in other respects, it will miss important observational differences. We will make further reference to this aspect in later sections.

We note that the GSD/QC system as presently designed uses a neighbor check to screen the scenarios described in the previous discussion of manual LOUZE screening. However, without knowledge of the time history of a gauge (which an RFC forecaster make use of), clusters of 'incorrect' small or zero values present a situation whereby such a station may be given a 'good' classification by GSD/QC because it is compared with neighbors that are 'good' in the same way. Since the GSD/QC uses daily values and independent gauges for comparison, this situation is somewhat alleviated. Indeed, the second level of RFC screening (performed with 24h totals) often passes the station for this longer accumulation period by assigning it an ignore ('I') flag instead a remove ('R') flag.

4.3 *NSSL's Q2 Verification System (QVS)*

The NSSL uses National Mosaic and Multi-Sensor QPE (NMQ – "Q2") products as the basic tool to quality control hourly gauge data. Because of high spatial resolution (1km), frequency (5 min) and improved algorithms such as dynamically allocated Z-R relationship at each pixel, Q2 offers distinct advantages over the WSR-88D Digital Precipitation Array (DPA). Additive radar-gauge biases are interpolated onto Q2 grids by using inverse distance weighting (IDW) methods. For each rain gauge location, all error estimates at radar pixels within a radius of ~10 km are compared with the error value at the gauge. If 25% or less of those error estimates are within 5 mm (a difference threshold, named "diffval") of the error at the gauge, then the rain gauge is considered problematic. The rain gauges that meet these criteria are removed and cross-validation is rerun, which often results in a new set of IDW parameter values. The same procedure is repeated using a smaller difference threshold of 4 mm and two more iterations using difference values of 3 mm and 2 mm follow, with cross-validation running between iterations. To ensure that a large number of rain gauges are not eliminated, the procedure is terminated before the 2 mm iteration if more than 10% of the total number of gauges is omitted. Failed hourly data by QVS receive codes 2-5 which corresponds to "diffvals" of 2 mm – 5 mm.

5. ALGORITHM RESULTS

The results of the three independent algorithms are compared against LOUZE data collected during August 2008. During this month, 662 gauge stations were flagged by LOUZE. Table 1 shows a list of the 12 stations that received the most LOUZE flags (the "dirty dozen"). Most of the flags in the dirty dozen ended up being rejected (the third column). Columns 4-6 indicate availability of data in each QC system. The NA* denote gauge data are not available because of mis-match of observation time, namely, off-the-top-of-the-hour values. Gauge values at BDTV2 and BVE were not available. The station WKXA1 is excluded in NCDC's reprocessed HADS database.

Because of the lack of the top-of-the-hour observations, the number of stations in QVS is reduced relative to NCDC and GSD samples. RFC manual processing also applies a narrow observation time window. However, ESRL's GSD/QC system does not screen for such shifted observation time since its effects are minimized in 24-hr summations. The reprocessed HADS does not include stations whose original format report is made in incremental instead of cumulative precipitation value.

Table 1. Dirty dozen LOUZE stations during August 2008, and their availability for QC flag comparison.

Station	"I" &"R"	"R"	NCDC	ESRL	NSSL
GREL1	108	105	Y	Y	Y
WKXA1	69	67	NA	Y	NA*
MOCM7	67	66	Y	Y	Y
MASA1	61	59	Y	Y	NA*
JSIM7	46	45	Y	Y	NA*
COLL1	43	42	Y	Y	Y
BLBL1	41	37	Y	Y	NA*
EPGM7	40	40	Y	Y	NA*
BDTV2	37	36	NA	NA	NA
XIIA1	36	36	Y	Y	NA*
BVE	32	31	NA	NA	NA
LGRT1	31	29	Y	Y	NA*

Although, a universal quantitative performance metric is difficult to define, we can qualitatively discuss relative strengths of QC system. We compared three "dirty dozen" stations (GREL1, MOCM7 and COLL1) in the side-by-side comparisons with LOUZE flags in figs. 5-7.

Figure 4 shows the location of GREL1 in QVS valid at 0400 UTC 13 August taken from http://nmq.ou.edu. The gauge value 0.07 inch at this hour passed QVS/QC. But, QVS/QC flagged at 0100 and 0200 UTC as seen in figure 5. The rest of the day, QVS/QC passed all light rain observations for the following 18 hours. The LOUZE assigned "R" flags from 0600 UTC through each of the following 12 hours except for 1300 UTC. The NCDC/QC flagged gauge value 0.13 inch at 0200 UTC and the following 9 hours except for 0700 UTC. The NCDC/QC's distance-based neighborhood samples over-flagged in this case. The forecasters had the advantage of using reflectivity data to QC the gauges, an example of the strength of manual QC. The LOUZE flagged most light rain observations during days 21 – 24, but QVS/QC flagged none. We speculate that QVS/QC is less robust toward light rain in data sparse areas (Fig. 4). The GSD/QC rejections during the period 10 – 12 August were caused by one apparently erroneous 1-hour report of 0.01 inch precipitation on each of these days (captured also by LOUZE). For GSD/QC, this flagging resulted from comparison with neighboring stations that observed no precipitation during these 24h periods.

The station MOCM7 (Fig. 6) has apparently performed very poorly, and has been systematically flagged on most of the days of non-zero precipitation by all the QC algorithms except NSSL. This case illustrates the potential usefulness of an overall quality metric that could be implemented even before QC on hourly data is performed. The NCDC/QC's code "9" for missing value is indicated by the tick mark of a negative value.

The station COLL1 (Fig. 7) is very challenging for the automated QC systems in real-time. An investigation of the original data showed cumulative precipitation amounts fluctuating over time, resulting in hourly precipitation values that became negative and were then followed by an extreme value of 0.6 inch (see http://www.ncdc.noaa.gov/hads/). The pattern of numerous missing values (19 in this month) in the historical data set is an indication of gauge maintenance problems.

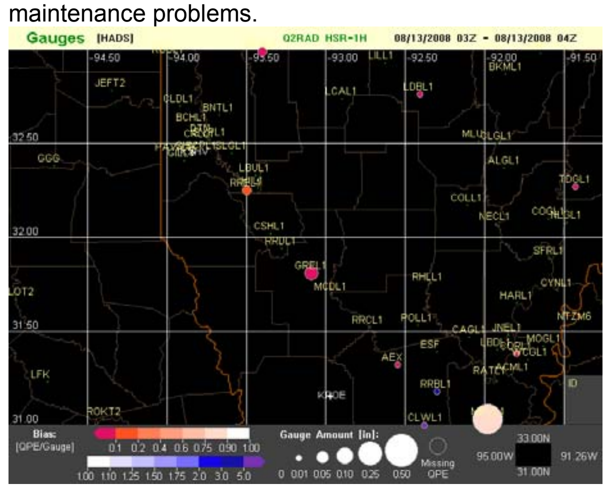

Fig.4. The screen capture of QVS valid at 0400 UTC August (http://nmq.ou.edu). The GREL1 is in the center of domain in red. Red color indicates over-estimation against Q2.

6. SUMMARY

Rain gauge data from August 2008 were used to compare automated QC algorithms at NCDC, ESRL and NSSL with manually QC'ed LOUZE data from the LMRFC. Out of 662 possible comparisons, the top 12 offenders were identified. Of those 12 only three were common to all three automated QC systems. Yet, we found valuable lessons;

1) The low gauge network density and anisotropic structure of precipitation systems limit QC algorithm performance.

2) Station quality history can be made and should be a part of metadata.

3) Additional tools must be developed to provide access of these data sets to researchers and operational decision makers.

4) There are hints of inconsistent QC practices at the LMRFC. This warrants continued collaboration between research and operations.

7. ACKNOWLEDGMENT

The authors acknowledge the Precipitation Gauge Working Group of the Multi-sensor Precipitation Reanalysis (MPR) project assigned to this task. In addition to the authors, this group includes Henry Fuelberg, Judi Bradberry, and Gerry Creager.

Table 2 LOUZE codes descriptions

QUALITY CONTROL FLAGS

L	Light precipitation amounts, generally in regions with no precipitation reported. Also applies to dripping gauges following rainfall events and snowmelt.
O	Overestimated gauge reports. The gauge is over-estimating the amount of precipitation observed on radar and surrounding reports. This is rare but accounts for problems such as double tipping.
U	Underestimated gauge reports. The gauge is under-estimating the amount of precipitation observed on radar and surrounding reports. This is common for clogged gauges.
Z	Zero reports. The gauge reports zero when precipitation is observed on radar and by surrounding reports. This is common for clogged gauges as well.
E	Enormous/extreme reports. The gauge reports a value larger than 0.10" and typically is due to equipment malfunction or decoding issues.

DATABASE REMOVAL FLAGS

I	Ignored gauges. Upon further quality control these gauges were found to have 24-hour totals representative of the radar estimates and surrounding gauge network. The data is not removed from the database and is used in operations for hydrologic modeling.
R	Removed gauges. Upon further quality control, these gauges were found to be unrepresentative of the radar estimated and/or reported significant and notable issues with quality. The data is removed from the database and not used in operations.

8. REFERENCES

Kim, D., B. Nelson and D. J. Seo, 2008: Reprocessed hydrometeorological automated data system (HADS) hourly precipitation data, *Submitted to Weather and Forecasting*.

Kondragunta, C. and K. Shrestha, 2006: Automated real-time operational rain gauge quality controls in NWS hydrologic operations. *Preprints, 20th AMS Conf. on Hydrology*, 29 Jan. – 2 Feb. 2006, Atlanta, GA.

Tollerud E, R. Collander, Y. Lin and A. Loughe, 2005: On the performance , impact, and liabilities of automated precipitation gauge screening algorithms. *Preprints, 21st AMS Conf. on Weather Analysis and Forecasting,* Washington D.C.

Vasiloff, S. V., and coauthors, 2007: Improving QPE and very short term QPF: An initiative for a community-wide integrated approach. *Bull. Amer. Meteor. Soc.*, **60**, 1048-1058.

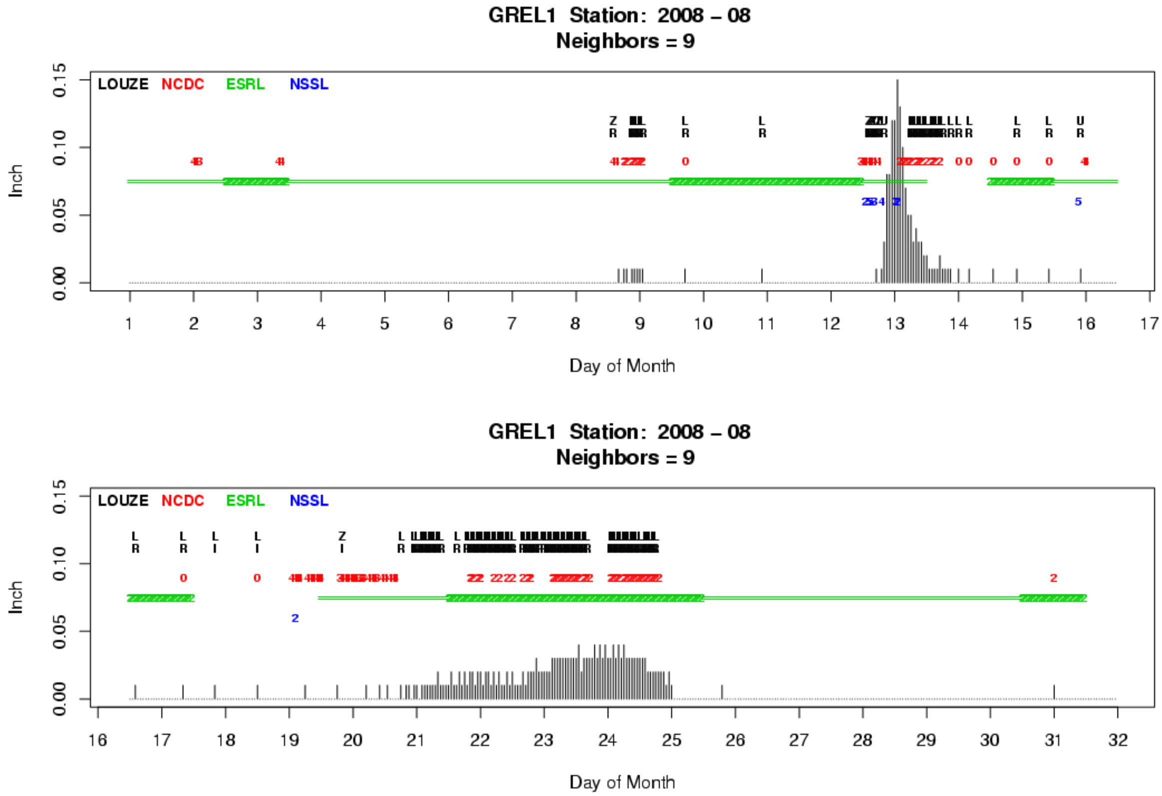

Fig.5. Time-series of hourly precipitation and QC results for August 2008 at station GREL1. The LOUZE codes in black display L,O,U,Z,E and I, R code below. NCDC's codes 0-4 are written in red, GSD/QC code is in green color "=" (narrow) for pass and "2"(wide) for failure in 24 hour period. Discontinued codes indicate no GSD/QC was performed for the corresponding day. The NSSL's QVS code is shown in blue.

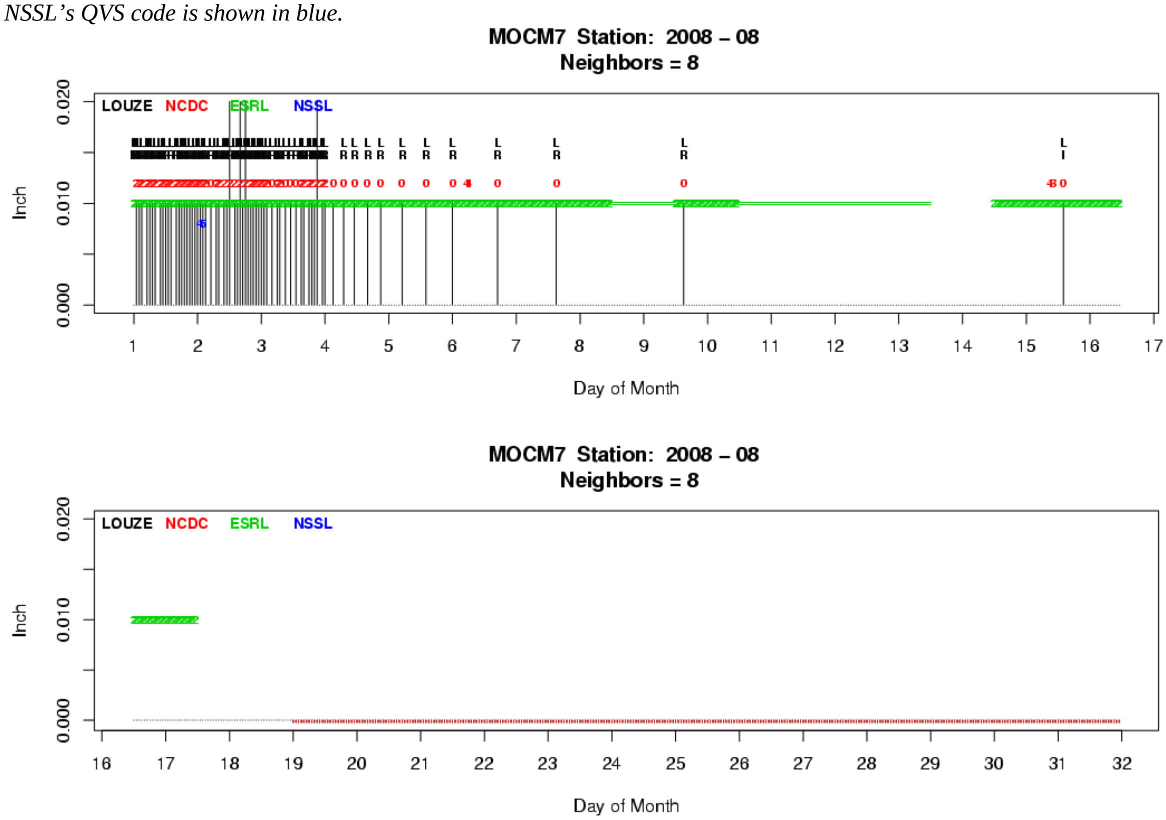

Fig.6. As in Fig.5 except for the station MOCM7. This station ceased to provide data since 00 UTC 19 August.

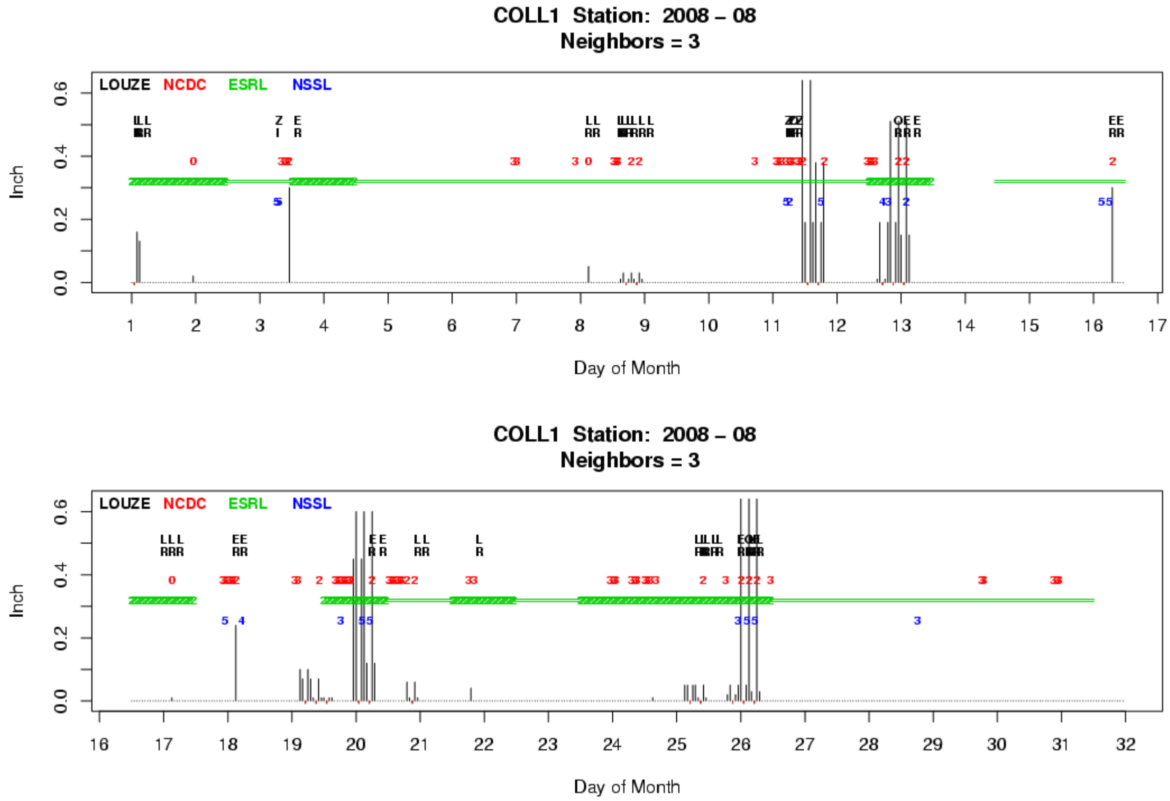

Fig.7. As in Fig.5 except for the station COLL1. The 3 neighbor stations give maximum 11 samples in NCDC/QC system. Some missing values in 11 samples further limit the reliable assignment of flag when gauge functions erratically.