

## CoCoRaHS Data Quality Assurance and Quality Control

A stated goal of the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS) “is to provide the highest quality data for natural resource, education and research applications.” Training and education are a large component of the effort to ensure quality data. Well-trained observers are less prone to errors and generally know how to begin to deal with unusual or unexpected situations that arise. However, even the best-trained and experienced observers make mistakes. In order to maintain a high standard of data quality we utilize a number of procedures and techniques to prevent, intercept, and to monitor for, follow up on, and if needed correct errors that make it into the CoCoRaHS database.

QC at the CoCoRaHS Headquarters in Fort Collins is a mix of automated and manual processes. Some data checks have been built in to the web-based forms used by observers to enter their observations and catch the easiest to make and obvious errors. In addition, daily CoCoRaHS observations ingested into the National Centers for Environmental Information (NCEI) Global Historical Climate Network-Daily (GHCN-D) are processed through NCEI QA/QC routines (see last section). Below are the categorized automated checks included on the web-based data entry forms.

### Daily Precip Report Form

#### Observation Date

- Cannot be after current date
- Format of MM/DD/YYYY

#### Observation Time

- On Daily report form, current local time cannot be more than 59 minutes before observation time. (This buffer exists as a convenience to observers so they are not blocked from entering data a few minutes ahead of their default observation time.)
- Time must be entered in 12 hour time (HH:MM AM/PM)

#### Daily Precip Amount

- Must have value to nearest hundredth, T, or NA.
- Any value up to and including 30.00 is allowed.
- If New Snowfall amount was entered, a precip value greater than 0.00, T, or NA must be entered.

### Snowfall

- Snow depth (new snow and total snow) must be entered to nearest tenth (0.0), hundredth not accepted.
- T, NA is allowed.
- New snow water equivalent cannot be more than 90 percent of the new snow depth.
- The water content (SWE) of new snow must be zero or NA if the depth of the total snow and ice is zero.
- If the depth of new snow is a trace, then the SWE should be a trace or NA.

### Total Snow and Ice on the Ground

- The snow water equivalent (SWE) of the total snow and ice on the ground cannot be more than 90 percent of the depth.
- The water content (SWE) of the total snow must be zero or NA if the depth of the total snow and ice is zero.
- If the total depth of ice and snow is a trace, then the SWE should be a trace or NA.

### Overall

- Cannot submit two daily report forms for the same station on the same date.

### Hail Report Form/Significant Weather Report:

- Storm date required on hail form
- Time not required, but if entered must be in 12-hour format
- Observation date and time are required on Significant Weather Report.

### Multi-day Accumulation Report form:

- Start date and end date must be at least 1 day apart. (e.g., Start: 4/20/2011, End must be at least: 4/21/2011)
- Time must be entered
- Precip amount must be to nearest hundredth inch, T or NA. Amount is required.
- Snow depth, nearest tenth, T or NA (not required).
- Cannot submit a multi-day accumulation for a period more than 60 days.
- If a daily report with an amount greater than or equal to zero exists within the range of the multi-day report, attempted submission of the multi-day report will generate an error message.
- A multi-day report cannot overlap any date within an existing multi-day report.

## CoCoRaHS QC Manual Process

CoCoRaHS headquarters staff and a number of our volunteer coordinators review the data each day, and typically look back a week or more at late-arriving reports. The national map is examined to look for very high values that don't seem to belong. False zeros may also be identified here. This is done for the precipitation, new snow, and total snow depth maps.

We employ a web tool that compares CoCoRaHS observations to surrounding reports which helps to identify potentially erroneous high reports and false zeros.

When questionable values are identified, the person performing QC will examine the observation in context of surrounding observations. We utilize the Multi-sensor Precipitation Estimate (MPE) maps generated by the Advanced Hydrologic Prediction Service to help determine the potential validity of an observation as well as radar maps. Questionable reports may be changed to "NA". Notes are added in the comments section of the report, and a "QC ticket" is submitted. This logs the observation in a QC database, and then reports the potential error or verification to the regional coordinator and state coordinator for that station. The coordinator should then follow up with the observer to resolve the suspected error.

## Common Types of Errors

Most errors we find in CoCoRaHS data are *reporting* errors, not measurement errors. Here are some of the most common types of errors:

### Typo/Decimal

Observer mistypes the amount into the precipitation field, or misplaces the decimal point.

### False Zero

Observer entered zero but evidence suggests value should be >0

### Incorrect date

Observer enters the incorrect date for the observation. These usually occur when the observations are entered one or more days after the observations were made, and often occur in pairs, i.e. ob1 is entered for date2, and ob2 is entered for date1.

### Multi-day Accumulation Entered as a Daily Amount

Observers often forget that a different entry form is needed to enter accumulations occurring over a period of 2 or more days, and enter the amount using the Daily Report Form.

### Entering Snowfall in the Daily Precipitation Field

Precipitation (melted from the snow and ice in the gauge) and the depth of new snow (measured on a snowboard or other flat surface) are two separate measurements. Observers will sometimes enter their depth of new snow in the daily precipitation field instead of or in addition to the New Snowfall field.

### Overflowing Rain Gauges

When the gauge overflows, observers typically report 11 inches of rain (though sometimes 10 or 12). One particular issue we've had with these reports is that observers usually don't include any comments with these high rainfall observations. At this point we don't know if this is a real measurement or an estimate. We want observers to be able to report their observation of these extreme rainfall events, even though an 11-inch gauge overflow is usually at best a low estimate.

In summary, if the gauge overflows observers are instructed to:

- 1) Report 11.00 inches
- 2) Include comments that gauge overflowed or anything pertinent to the observation.

11 or 12 inch amounts submitted without comments will be set to NA in the QC process until follow up can be made with the observer.

### Quality Controls Checks by NCEI for GHCN-D Precipitation Data

All data submitted into CoCoRaHS since 1998 are continually streamed to NCDC and stations where at least 100 individual data values are present for at least one data element are included in the CoCoRaHS data source for GHCN-Daily. New data values received from CoCoRaHS are retrieved by GHCN-Daily on a daily basis for existing stations and updated periodically to include any newly commissioned CoCoRaHS stations.

CoCoRaHS raw data values are put through a carefully evaluated set of fully automated Quality Assurance (QA) procedures to detect as many errors as possible for Precipitation, Snowfall and Snow Depth during a station's period of record. The QA system consists of checks which detect instances of duplicate data, climatological outliers and various inconsistencies (internal, temporal and spatial). The checks are arranged in a deliberate sequence in which the performance of the later checks is enhanced by the error detection of the earlier checks. When a value fails a particular check, it is assigned an associated data quality flag.

Here's a further breakdown of the checks that CoCoRaHS Data are tested by GHCN-Daily's QA procedures:

## 1. Precipitation Checks

- a. *Trace flag consistency check* – Checks for days on which the data measurement flag indicates a trace of precipitation yet the precipitation amount is nonzero.
- b. *Duplicate checks* – Check for duplication of the data between entire years, different years within the same calendar months, and different months within the same year, using only years and months with at least three nonzero totals.
- c. *Bounds check* – Identifies precipitation totals that exceed the world extreme for the highest 24-hour total ever observed.
- d. *Streak check* – Checks for sequences of 10 or more consecutive identical values in Time series of nonzero precipitation observations.
- e. *Frequent-value check* – Checks for clusters of 5-9 identical moderate to heavy totals in time series of nonzero precipitation observations.
- f. *Gap check* – Identifies values that are at least 300 mm (or 11.81 in.) larger than all other precipitation totals for a given station and calendar month.
- g. *Snowfall to precipitation ratio check* – Checks for cases in which snowfall is excessively large compared to precipitation. I.e., if the current day's snowfall is more than 100 times larger than both the current+previous and current+following days' precipitation sums, then the current day's precipitation and snowfall totals fail the check.
- h. *Snow depth increase to precipitation ratio check* – Checks for cases in which a day-to-day increase in snowed depth is excessively large compared to precipitation. For example, if the snow depth increase between the previous and current days is more than 100 times larger than both the current+previous and current+following days' precipitation sums, the current day's precipitation total and the current and previous day's snow depths fail the check.
- i. *Spatial consistency check* – Checks for precipitation totals that differ significantly from totals reported at neighboring stations on the preceding, current, and following days.

## 2. Snowfall Checks:

- a. *Trace flag consistency check* - Checks for days on which the data measurement flag indicates a trace of snowfall yet the snowfall amount is nonzero.
- b. *Duplicate checks* - Check for duplication of the data between entire years, different years within the same calendar months, and different months within the same year, using only years and months with at least three nonzero totals.
- c. *Bounds check* - Identifies snowfall totals that exceed the world extreme for the highest 24-hour total ever observed.
- d. *Streak check* - Checks for runs of 10 or more consecutive days on which the same nonzero snowfall total is reported.
- e. *Snowfall to snow depth increase consistency check* - Checks for days on which the increase in snow depth from the previous day to the current day exceeds the current+previous and current+following days' snowfall sums by more than 25 mm (or 1.0 in.). In such cases, the current day's snowfall and snow depth as well as the previous day's snow depth fail the check.

- f. *Snowfall to precipitation ratio check* - Checks for cases in which snowfall is excessively large compared to precipitation. See the same entry under the Precipitation Checks above.
  - g. *Climatological outlier check* - Identifies daily nonzero snowfall totals on dates outside the likely snow season as calculated from the snowfall records at the station and its suitable neighbors.
3. Snow Depth Checks :
- a. *Trace flag consistency check* - Checks for days on which the data measurement flag indicates a trace of snow depth yet the snow depth amount is nonzero.
  - b. *Bounds check* - Identifies snow depths that exceed the world extreme for the highest snow depth ever observed.
  - c. *Temporal consistency check* - Checks for day-to-day increases in snow depths that exceed the world record 24-hour snowfall total.
  - d. *Streak check* - Checks for runs of 90 or more consecutive days on which the same nonzero snow depth is reported.
  - e. *Gap check* - Identifies values that are at least 35 cm (14 in.) larger than all other reported snow depths for a given station and calendar month.
  - f. *Snowfall to snow depth increase consistency check* - Checks for days on which the increase in snow depth from the previous day to the current day exceeds the current+previous and current+following days' snowfall total by more than 25 mm (or 1 in.). See the same entry under Snowfall consistency check.
  - g. *Snow depth increase to precipitation ratio check* - Checks for cases in which a day-to-day increase in snow depth is excessively large compared to precipitation. See the same entry under Precipitation Checks above.
  - h. *Climatological outlier check* - Checks for day-to-day increases in snow depth on dates that fall outside the likely snow season as calculated from the snow depth records at the station and its suitable neighbors.

For more information on GHCN-Daily please visit <https://www.ncdc.noaa.gov/ghcn-daily-description>