**Functional Specifications for**

**Quality Control and Spatial Interpolation of Precipitation Observations (DailyQC)**

**Version 1.0, RKHCS**

This document provides the functional specifications and requirements for a program that will serve the purposes of quality controlling ground-based precipitation observations for durations of 1, 6, and 24 hours and the spatial distribution of those observations into a 4km grid.

**Data Resources**

* Precipitation data packages are created (separate program) and are stored in zipped for specific 24-hour periods. Each package contains:
	+ The date (mm/dd/yyyy). This is the standard time date at the end of the 24-hour period ending at 12 GMT.
	+ The latitude/longitude of the QC domain.
	+ Precipitation units (inches or mm).
	+ Observing station list (id, latitude, longitude, elevation).
	+ Raw data matrix
		- Requires the pre-processing of shorter timestep data into the 1, 6, and 24 amounts by a separate program.
		- 2 rows per station
			* 31 QC flags (Z=obs/calculated, M=missing (missing value=0.0).
			* 31 values (missing values=0.0)
				+ 24 1-hour amounts
				+ 4 6-hour amounts are 12Z-18Z, 18Z-00Z, 00Z-06Z, and 06Z-12Z.
				+ 1 24-hour amounts are 12Z-12Z.
	+ QC’d data matrix
		- 3 rows per station
			* 31 QC flags (S=screened, V=verified, Q=questionable, B=bad, M=missing.
			* 31 values estimated from nearest neighbors (initially 0.0)
				+ 24 -1hour, 4 6-hour, and 1 24-hour values
			* 31 values of normalized deviation from the estimate (initially 0.0)
				+ 24 -1hour, 4 6-hour, and 1 24-hour values
	+ QC process matrix (flags: “ “=not complete, “C”=complete, initially “ “)
		- 24 -1hour, 4 6-hour, and 1 24-hour flag
	+ QC’d gridded 4km fields (initially “0.0”)
		- 24 1-hour, 4 6-hour, and 1 24-hour
* PRISM rasters of mean monthly precipitation at 4km resolution (1981-2010)
	+ Can be combined into seasonal and annual fields if desired.
* Background GIS data layers of
	+ Rasters
		- Elevation
		- WestWRF analysis (date/time appropriate)
	+ Vectors
		- Roads
		- Streamflow network
		- Basin boundaries
	+ Points
		- Gaging stations (hydromet)

**QC Process**

The program is launched for a specific latitude-longitude domain and date. This selection will define the precipitation data package file. The file is unzipped and pulled into the processing interface.

The monthly PRISM rasters are used to define the background climatology. The user has the option of using the month associated with the specific period being reviewed, a seasonal average of the users choosing (e.g. November – March), or the annual average.

There are four (4) user selectable parameters for the QC process that can be modulated from the display.

1. The PRISM period selected for normalization and estimation. The default is monthly. Other options are seasonal (user selected group of months), or annual.
2. Number of standard deviation threshold for identifying an observation as “questionable.” (default = 2.0)
3. Number of nearest (non-missing, non-bad) observations used to compute the estimated value for each location.
4. Radial distance in km over which the sample of normalized observations is taken to compute a standard deviation. (default = 100)

These can be changed by the user at any time but are always set to the default when the program is launched. (We may want to create a process for establishing and storing updated defaults).

The QC matrix provides a picture of what periods have been QC’d. It also provides the user’s navigation to the QC process for each period. This matrix is part of the QC package for the date so that a user can complete a portion of the review and come back and complete it later.





When the QC process is completed for a specific period, a “C” will appear in the matrix associated that period. Non-QC’d periods remain blank.

Review can take place in any order for any duration. Over time, users will learn the most effective way to work through the QC. Amounts QC’s for one period do not impact the QC process for another. Logic for a different paradigm could be explored in a subsequent version of the software.

The selected period amounts are geographically plotted. The background of the display can contain any number of GIS layers such as topography, roads, rivers, lakes, etc. for orientation. These can be selected or deselected by the user. The color key for plotting is as follows:

1. White = screened “S” or unevaluated “Z”
2. Yellow = questionable “Q” (failed screening test)
3. Green = validated “V” (changed from yellow, or red by user)
4. Pink = missing “M”
5. Red = bad (changed by user from white, yellow, or green)

Initially all non-missing amounts are plotted in white with missing amounts are plotted as pink “M’s.”

The QC process is initiated by a button click while looking at a geographical display of the observations associated with the period selected. The computational steps associated with the QC process are:

1. Compute statistics for each station with a “non-missing” and “non-bad” observation
	1. Compute the fraction of normal ($F\_{X}$) by dividing the observation by the PRISM mean associated with the 4km grid cell within which the station’s latitude and longitude falls. ( $F\_{x}=X/\overset{‾}{PRISM}\_{x}$)
	2. Compute the standard deviation ($σ\_{y}$) from the sample of non-missing and non-bad $F\_{X}$ values from all stations within 100 km (distance can be modulated as desired).
2. Compute the estimated $F\_{X}$ for each station ($F\_{y}$)
	1. Select the nearest n=5 stations with non-missing and non-bad observations (n can be modulated if desired).
	2. Establish the weight for each station as $W\_{i}=1/d^{2}$ (d can be modulated if desired)
	3. $F\_{y}=\sum\_{i=1}^{n}w\_{i}F\_{xi}/\sum\_{i=1}^{n}w\_{i}$
3. Compute the normalized deviate for each station as $ND\_{y}=\left|F\_{x}-F\_{y}\right|σ\_{y}$
4. If $ND\_{y}\geq 2$.0 the observation is marked as “questionable”, otherwise the observation is marked as “screened.” (2.0 can be modulated if desired. For 1-hour, a greater number may be needed to avoid an excessive number of questionable observations. We might want to consider having defaults for each period length (1, 6, and 24)).
5. Compute the estimated observation for each location as $E\_{y}=F\_{y}\*\overset{‾}{PRISM}\_{y}$
6. Update the QC’d data matrix

Once the “QC” button is selected, the computations are made, and the observations are redisplayed with colors associated with the new status (as defined above).

The user can then enter an interactive process of reviewing individual stations and re-initiating the QC process. By clicking on an individual station, a pop-up will appear that allows the user to see

1. Raw data value
2. QC flag currently assigned
3. Estimate of the station based on nearest neighbors
4. Normalized deviation of the observation
5. Options changing the QC to “Validated” or “Bad” if the current QC flag is “Q” or “B.” A “Screened” value can also be set to “B” if desired.

At any point, the user can select to re-perform the QC process. This repeats the process shown but the stations included in the analysis will change based on the user’s re-designation of individual station flags. Because “screened”, “validated”, and “questionable” values are used to screen others, the setting of a single “questionable” observation to “bad” can move observations from previously “questionable” to “screened.”

At any point the user can choose to “render” the grid.” This process spatial distributes the $F\_{x}$ values for all “screened” and “validated” observations into the 4km grid using $1/d^{2}$ distance weighting. The $F\_{x}$ values for each grid cell are then converted to units of precipitation using the underlying PRISM mean. If selected, the interface will show the gridded field beneath the observations.

When satisfied with the results, the user then selects to SAVE the analysis. The grid for this field is then rendered, the data package updated and saved to disk, and the QC matrix updated with a “C” for the period saved. The user then has the option to move to another period to perform the QC process.

**Features of Underlying Display and Analysis Program**

1. Display georeferenced rasters of any resolution
2. Display georeferenced vectors (e.g. roads, streams, basin boundaries, etc.)
3. Display color-coded georeferenced point values of precipitation
4. Pop-ups to provide station specific information and the re-designation of the QC flag
5. Spatial distribution of selected points into a 4 km grid using a distance weighted normalization process