A Conceptual Description of the CPC Gauge Analysis Procedure

This document provides a conceptual description of how the CPC gauge based analysis of daily precipitation is defined and how the quality control (QC) is conducted for the station reports of daily precipitation over the Contiguous United States (CONUS).

1. *General description*

CPC gauge-based analysis of daily precipitation is constructed through interpolating station reports of gauge daily precipitation. The analysis is produced day by day on a 0.125 (1/8) degree latitude / longitude grid over the global land areas. The analysis created on this native grid of 0.125^{o}lat/lon is then averaged onto grids of coarser resolutions (0.25^{o}lat/lon, 0.5^{o}lat/lon) and released to general public.

On average, daily precipitation reports from 8,000-9,000 stations are available for the construction of gauge analysis over the CONUS. The average distance between two reporting stations is about 30km, or a little bit less than 20 miles. The station network, however, is highly inhomogeneous, with sparse station densities over western mountainous regions.

2. *How analysis on a 0.125^{o}lat/lon grid point is defined*

Daily precipitation value at a 0.125^{o}lat/lon grid point is defined in ***two big steps***:

Step 1: ***Climatology (long-term mean) of daily precipitation at the grid point is first defined.***

This is done through desegregating the PRISM monthly climatology to daily (365-day) resolution. This climatology reflects the seasonal evolution of long-term mean precipitation, with orographic effects accounted for.

Step 2: ***Grid fields of daily precipitation departure from the long-term mean are computed. This is accomplished as follows:***

Step 2.1: *Selecting reporting stations from regions adjacent to the target grid point:*

Step 2.1.1: First, an initial search distance R_0 (km) is determined based on the number of reporting stations received that day. The R_0 is determined in such a way that on average, there are 7 reporting stations available within that distance from a grid point;

Step 2.1.2: While performing grid analysis for each grid point, the number of reporting stations within that initial search distance is counted. If the number is between 4 - 10, then all reporting stations within the initial searching distance is selected.

Step 2.1.3: If the number of reporting stations within the initial search distance is less than 4. The searching distance is expanded until at least 4 reporting stations are available. These 4 stations are then selected;

Step 2.1.4: If the number of reporting stations within the initial searching distance is more than 10, then the searching distance is narrowed down until only 10 reporting stations are included. These 10 stations are selected;

Step 2.2: Computing the departure of daily precipitation for each station selected by subtracting the daily climatology at the station location calculated in Step 1 from the daily total precipitation reported;

Step 2.3: Interpolating the daily departures from the selected stations through the Optimal Interpolation (OI). Generally speaking, in the OI, the weight for each station is a function of station to grid point distance (so that station closer to a grid point receives more weight) and also influenced by the correlation between the station reports (so that two stations close to each other will have less weights than two stations located in totally different directions from the grid point).

Step 3: Grid fields of TOTAL daily precipitation is finally defined as the summation of the daily climatology and the daily departure defined respectively in Steps 1 and 2.

Technical details of the OI-based interpolation algorithm described above is peer reviewed and published in the American Meteorology Society (AMS) Journal of Hydrometeorology (JHM) (https://journals.ametsoc.org/jhm/article/8/3/607/5721/A-Gauge-Based-Analysis-of-Daily-Precipitation-over). A separate work conducted by CPC assessed the performance of several objective analysis techniques for gauge precipitation interpolation and concluded that the OI presents superior performance (https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2007JD009132)

3. *How analysis on a 0.25°lat/lon grid box is defined*

The procedures described in Section2 are intended to define values of daily precipitation representative of grid POINTS in 0.125°lat/lon intervals. These values are converted into area mean precipitation representative of a grid BOX of 0.25°lat/lon. This is done as follows:

Step 1: ***Define area mean precipitation for each grid BOX of 0.125°lat/lon***

For each 0.125°lat/lon grid BOX, areal mean precipitation is defined as the mean of the grid POINT values at the four corners calculated in Section 2;

Step 2: ***Define rea mean precipitation for each grid BOX of 0.25°lat/lon***

Inside each 0.25°lat/lon grid BOX, there are 4 0.125°lat/lon grid BOXES. Areal mean precipitation for the 0.25°lat/lon grid BOX is defined as the weighted mean of the precipitation at those 4 0.125°lat/lon boxes. The weight is set as the area of the 0.125°lat/lon box (cosine of the latitude).

As described above, no spatial smoothing is applied in defining the precipitation fields on a 0.25°lat.lon grid resolution.

The latitude / longitude included in the data set indicate the position of the CENTER of each 0.25°lat/lon grid box.

4. Quality control procedures for the input station reports of daily precipitation

All station reports are subject to quality control procedures before they are infused into the analysis procedures described above. In the real-time production, an automatic procedure is executed at a NOAA/NWS/NCEP operational IT infrastructure called Compute Farm (CF) to perform the quality control for the input station reports. This automatic QC procedure is carried out in two steps, i.e.

Step 1: We maintain a list of stations that are known to have various problems (e.g. lat/lon errors, reporting same values every day, etc.). The daily input station data is checked against this list and any reporting station found on the list will be removed from the input file; and

Step 2: Standard QC procedures composed of the following:

a) A "duplicate station check" which eliminates duplicates and key punch errors from the rain-gauge reports;
b) A "buddy check" to eliminate extreme values from the dataset. The "buddy check" examines the absolute value of the difference between the current station and all stations within a one-degree grid box. If more than 50% exceed a specified threshold, then the current station is tossed; and
c) a standard deviation check, which compares the daily rain-gauge data against a gridded daily climatology. For the standard deviation check we use a daily climatology derived from the historical CPC daily unified analysis. The observations are compared to the nearest grid-point value from the climatology. The current observation must be within 5 standard deviations (10 for hurricane events) of the daily climatology.

Quality control for station reports of daily precipitation is highly challenging due to the extreme temporal and spatial variation of precipitation and the complicated sources of various errors in the station reports. Once the gauge analysis is created using the station reports passed the automatic QC procedures described above, a set of manual QC procedures are applied once a month at the beginning of each month by experienced meteorologists. The manual QC is conducted as follows:

Step 1: Gridded analysis of daily precipitation over the CONUS is inspected manually day by day to identify any suspicious spots / regions with unusual precipitation values or presenting unnatural spatial patterns;

Step 2: Gridded analysis **of** daily precipitation over the CONUS is compared with concurrent MRMS radar precipitation for suspicious spots / regions showing large and unnatural differences;

Step 3: Suspicious spots / regions identified in steps 1 and 2 are marked and recorded for further examinations as detailed in steps 4, 5, and 6;

Step 4: Analysis value at each suspicious grid point for each day marked in step 3 is traced back to identify all station reports utilized in the calculations;

Step 5: Reports of daily precipitation identified in Step 4 are examined as follows:

a) examining the CPC station data archive (SHEF) for the station reports updating and archiving status; and

b) comparing against records of daily precipitation for the same stations in the archives maintained by the NOAA National Centers for Environmental Information (NCEI) and the Iowa State University for reported precipitation as well as station location information;

Step 6: A discussion will be held between / among the meteorologists involved in the above mentioned manual QC procedures to determine if or not suspicious station reports should be thrown away;

Step 7: Gauge analysis for each day of the target month will be re-run with the station reports passed manual QC;

Step 8: Manual QC for steps 1 to 7 will be repeated until satisfactory.

The Manual QC procedures described in steps 1 to 8 started from January 2019. No such manual QC is conducted for periods before.

5. CPC gauge analysis release

CPC gauge based analysis over CONUS is released to the public through ftp at:

https://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_CONUS/

The real-time version analysis, constructed using station reports that are only quality controlled through objective procedures, are available one day after the stamped date at the sub-directory named 'RT' underneath the location above. They are named:

yyyy/PRCP_CU_GAUGE_V1.0CONUS_0.25deg.lnx.20200114.RT

where 'yyyy','mm', and 'dd' indicate the year, month, and date, respectively.

The updated version of the analysis, constructed using manually quality controlled station reports, are uploaded in a sub-directory named 'UPDATED' underneath the above mentioned ftp site, with file names:

Yyyy/PRCP_CU_GAUGE_V1.0CONUS_0.25deg.lnx.20200105.UPDATED

where 'yyyy','mm', and 'dd' indicate the year, month, and date, respectively.