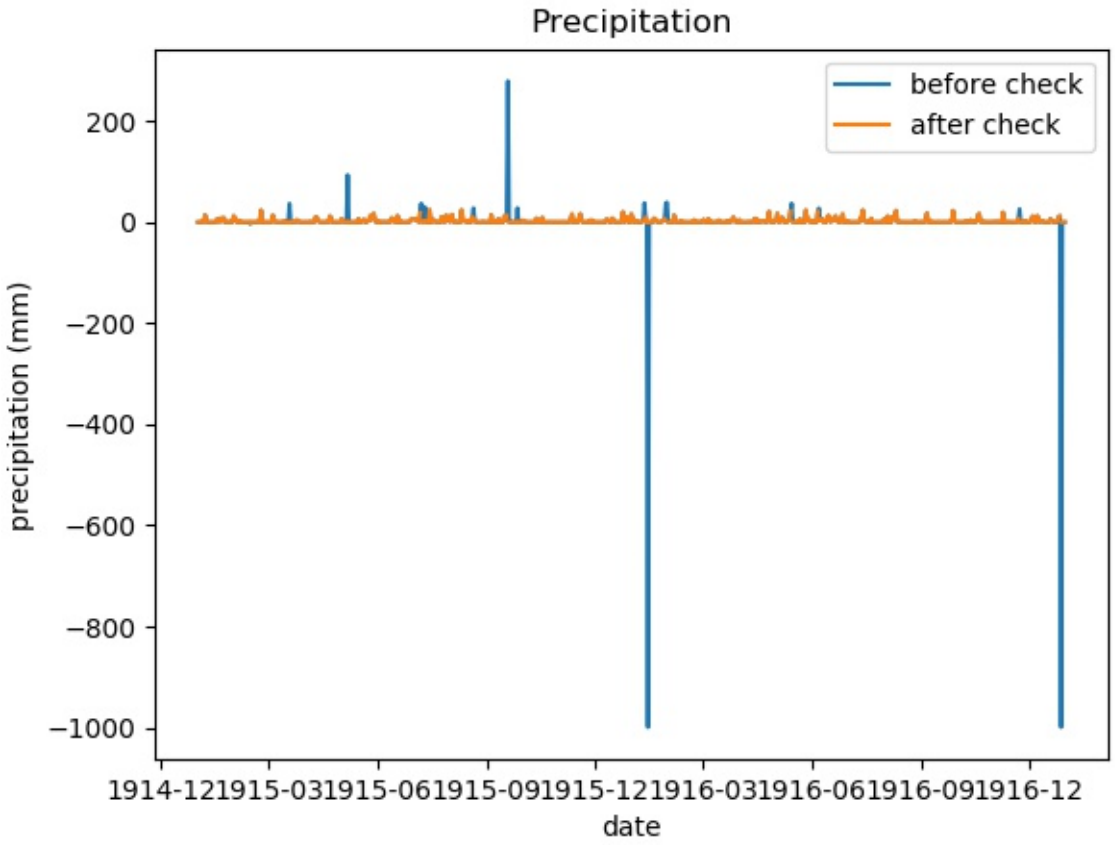
Metadata file
5/05/2020 by Charles Huang

program-09.py is a script which uses python to read the input file, “DataQualityChecking.txt” that is containing data of five variables: Date, Precip, Max Temp, Min Temp, and Wind Speed, which are date, precipitation (mm), maximum air temperature (°C) minimum air temperature (°C), and wind speed (m/s). It then 1. Replace all -999 values in the data to NumPy NaN values; 2. Apply the following error thresholds: 0 <=P <=25, -25<=T<= 35, 0 <=WS <= 10 to the data, and replace all values outside this range to NaN; 3. Swap max and min temp when max temperature is smaller than min temperature; 4. Replace both max and min temp to NaN when max temperature minus min temperature is greater than 25. And each of these four check steps will be recorded in a table with the number of corrections made (See below). According to the table, there were six -999 values replaced by check 1, 33 gross errors filtered by the thresholds of check 2, 4 pairs of max and min temperature swapped by check 3, and 10 range fail max and min temperature captured by check 4.

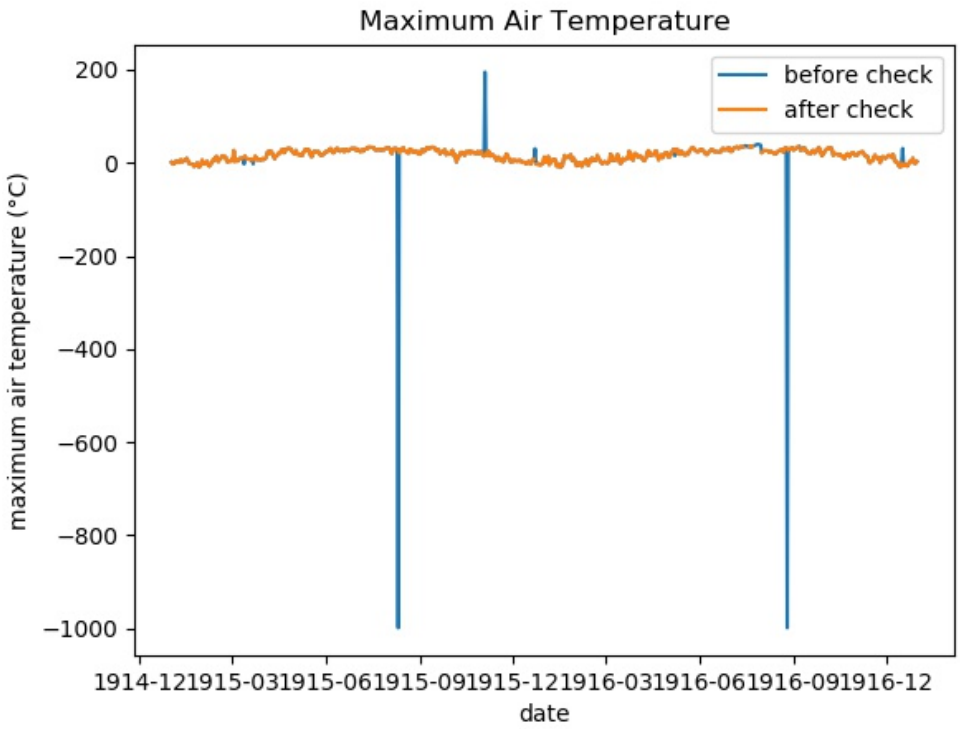
Table with the number of corrections made for all checks:
Final changed values counts.....

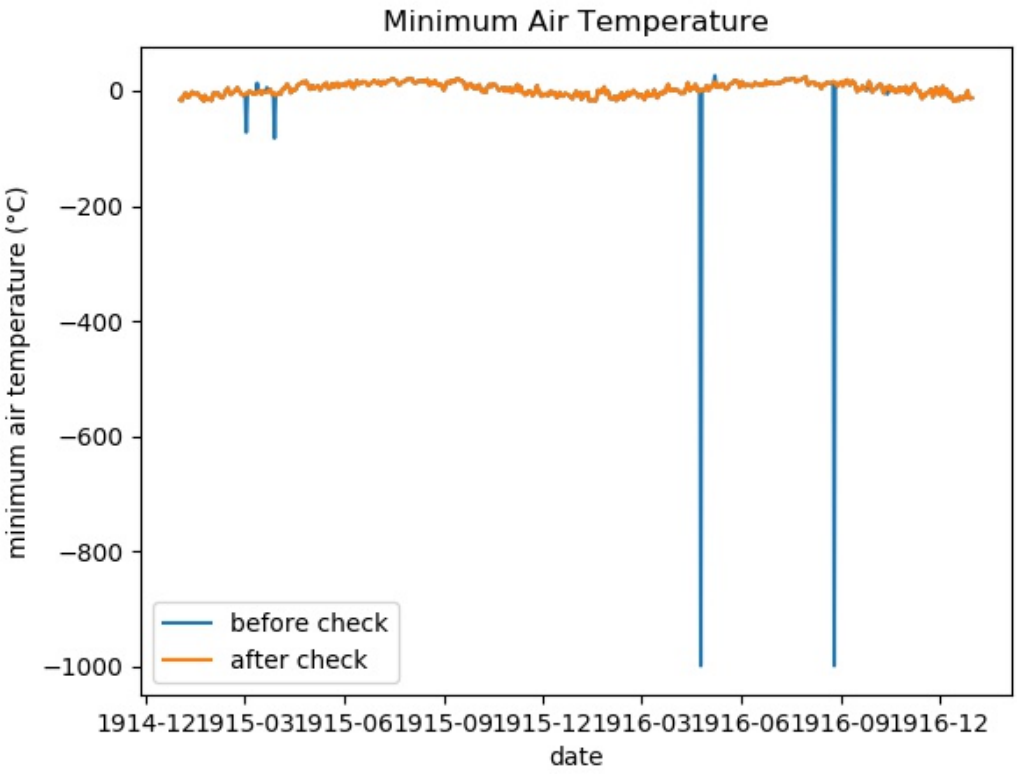
	Precip	Max Temp	Min Temp	Wind Speed
1. No Data	2.0	2.0	2.0	0.0
2. Gross Error	15.0	14.0	2.0	2.0
3. Swapped	0.0	4.0	4.0	0.0
4. Range Fail	0.0	5.0	5.0	0.0

The following are plots for each variable before and after quality checking:

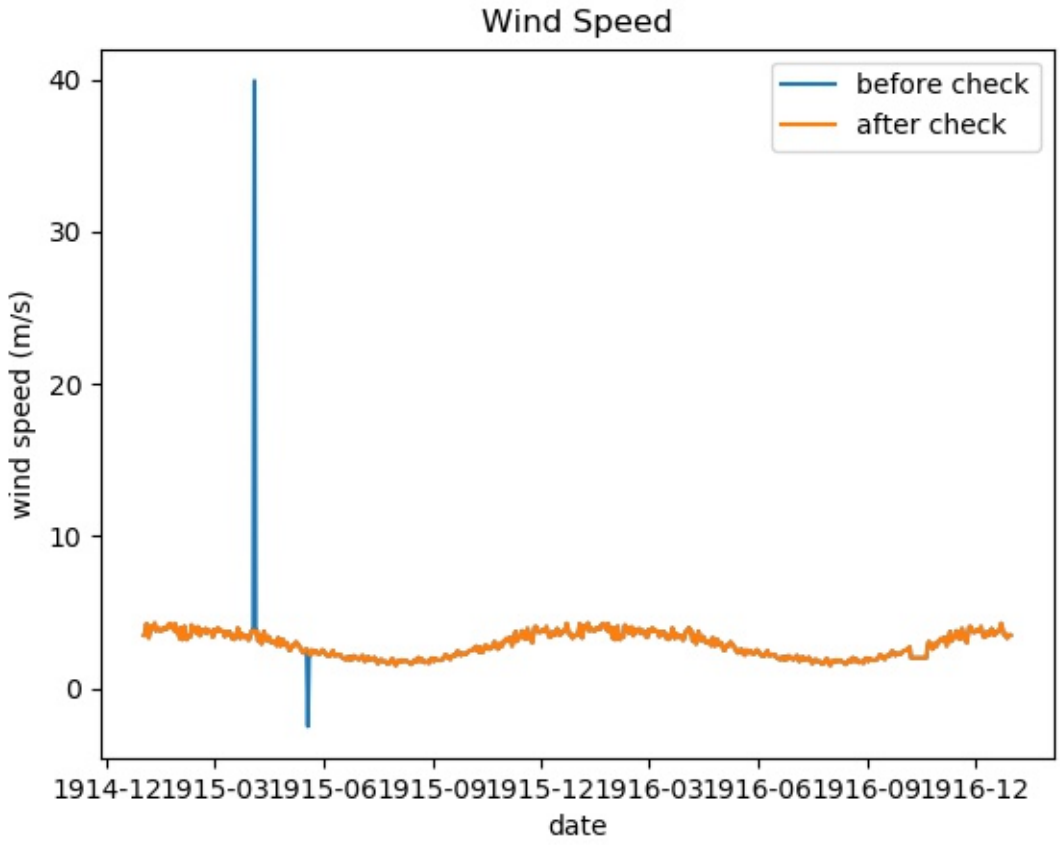


According to the plot, check 1 excluded the -999 values for precipitation data, and check 2 filtered out the data that is out of (0,25).





According to both plots, the max and min temperature fall into the appropriate range after check 1, 2 and 4, and there are swapped data which can be seen at the left side.



According to the plot, check 1 excluded the -999 values for wind speed data, and check 2 filtered out the data that is out of (0,10).