Les Warren
03/27/2020
ABE Lab 09 Metadata

This metadata file is for the “program_09.py” script in this repository. The script was created by Les Warren on 03/26/2020. The script uses the input file “DataQualityChecking.txt” also found in this repository. Once the file is read into the script, the script uses four definitions to check for four different types of data quality errors including removing no data values, gross errors, temperature max and min swap, and temperature max to min range. After completing the data quality checks and replacing anything outside the set parameters with NaN, the script makes four figures comparing the original data with the corrected data. Finally, the script outputs two .txt files. The first in the corrected data and the second includes a table of the number of errors fixed for each variable.

The four checks included:

1. Remove No Data Values

    This check replaced any “no data” numbers with NaN. In this case, the no data placeholder was -999. Precipitation, Temp Max, and Temp Min all had two values that were replaced (see Table 1). Wind speed did not have any no data placeholders.

2. Gross Errors

    This check removes values for each variable that are out of an ordinary range for that variable. Any values outside of 0-25 were replaced with NaN for precipitation. For Temp Max and Temp Min, any values outside the range of -25 to 35 were replaced with NaN. Lastly, any values outside the 0-10 range for Wind Speed were replaced with NaN.

    Precipitation had 15 values outside of range, Temp Max had 14 values outside of its range, and Temp Min and Wind Speed both had 2 values replaced (Table 1).

3. Temp Max- Temp Min Swap

    This check was performed to check for any Temp Max values that were lower than the Temp Min value in the same line. If an error was observed, it was assumed that these values were wrongly entered and thus the check swapped the values. There were four instances were the Temp Max was lower than the Temp Min and the data was swapped (Table 1).

4. Temp Max – Temp Min Range

    This check subtracted the Temp Min from the Temp Max to observe the range between the two variables. If the range was >25, degrees, then the Temp Max and Temp Min were replaced with NaN. As shown in Table 1, there were 5 instances were the range was greater than 25 degrees and the values were replaced.

Table 1. Number of data quality corrections made for each variable by correction type.

	Precipitation	Max Temp	Min Temp	Wind Speed
1. No Data	2	2	2	0
2. Gross Error	15	14	2	2
3; Swapped	0	4	4	0
4. Range	0	5	5	0

Plots:

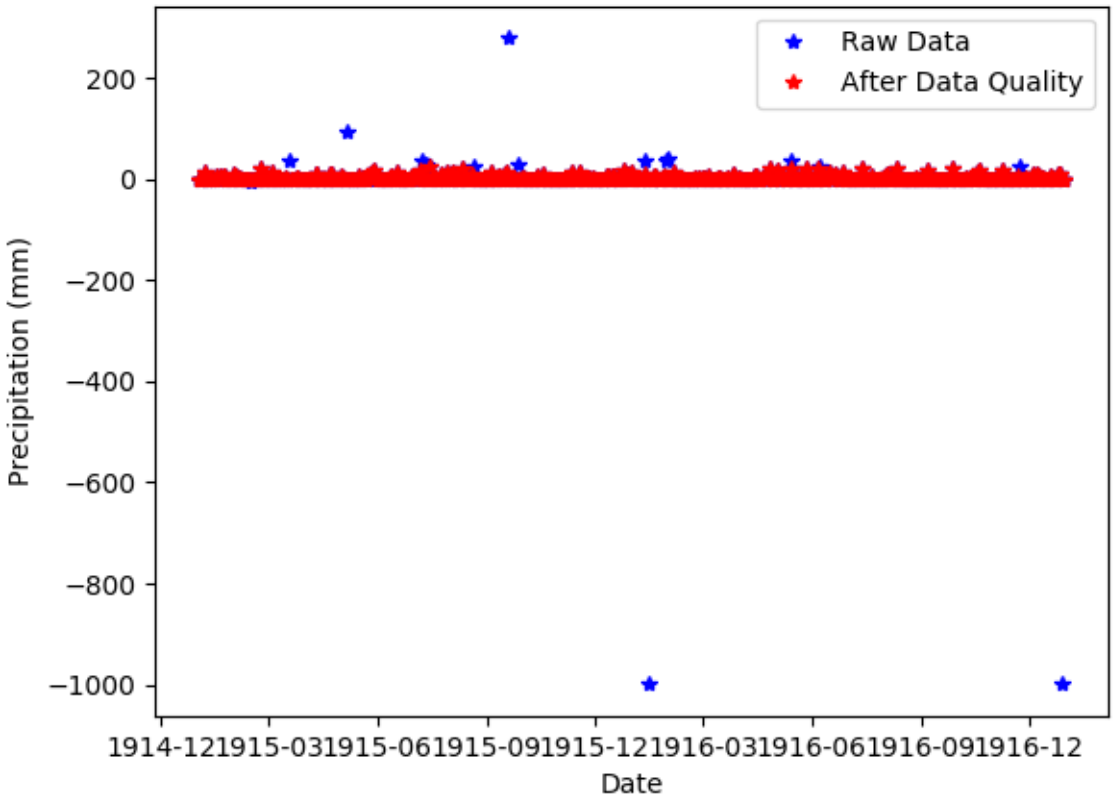

Figure 1. Precipitation scatter plot showing original data (blue) against the corrected data (red). Missing data values (-999) and other out of range data have been removed.

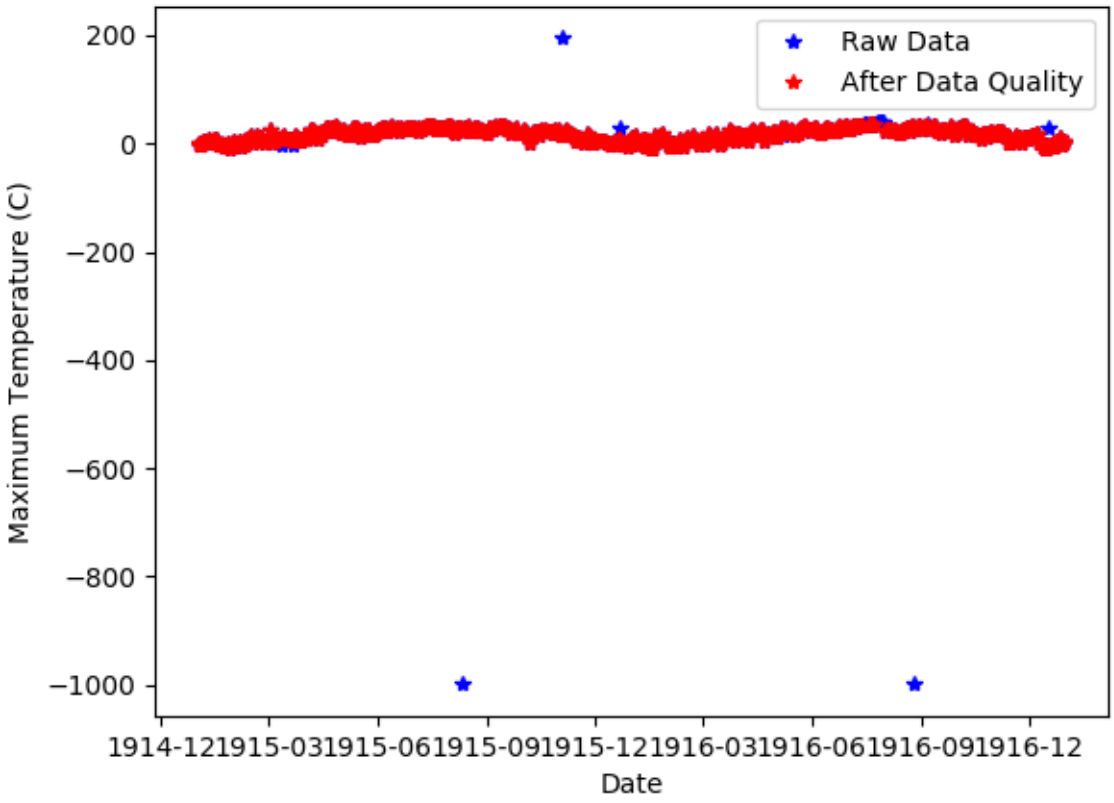

Figure 2. Max Temp scatter plot showing original data (blue) against the corrected data (red). Missing data values (-999) and other out of range data have been removed.

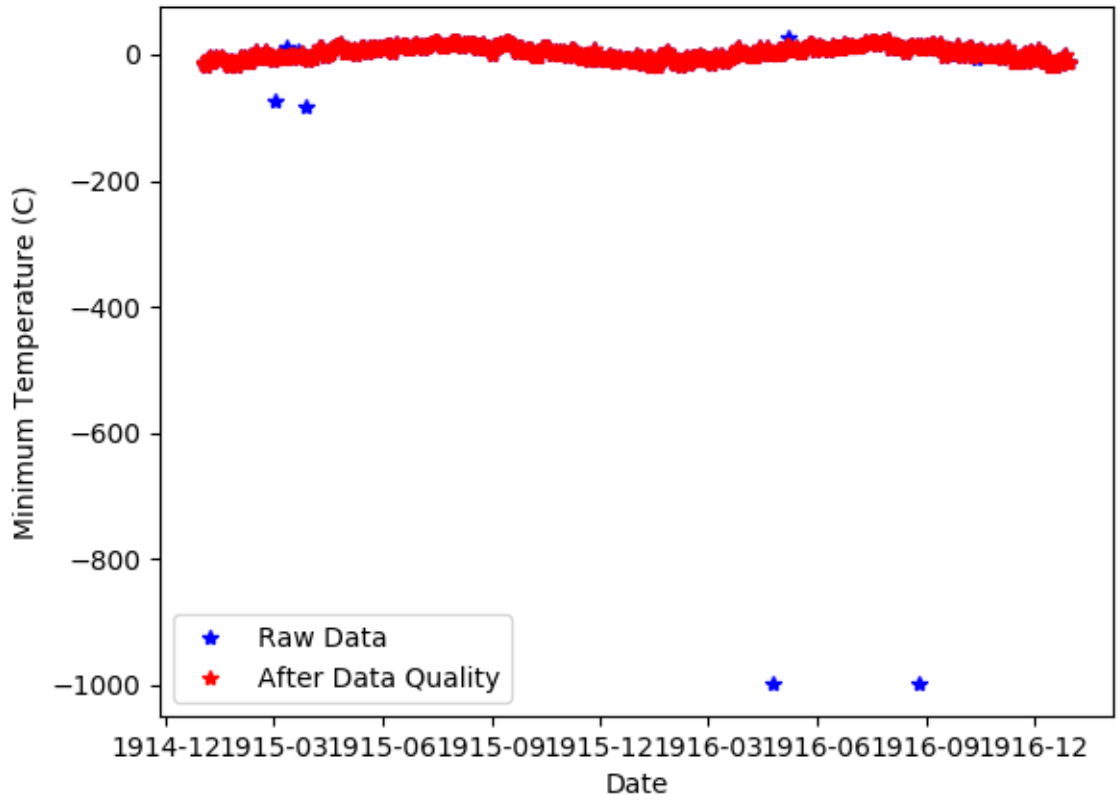

Figure 3. Min Temp scatter plot showing original data (blue) against the corrected data (red). Missing data values (-999) and other out of range data have been removed.

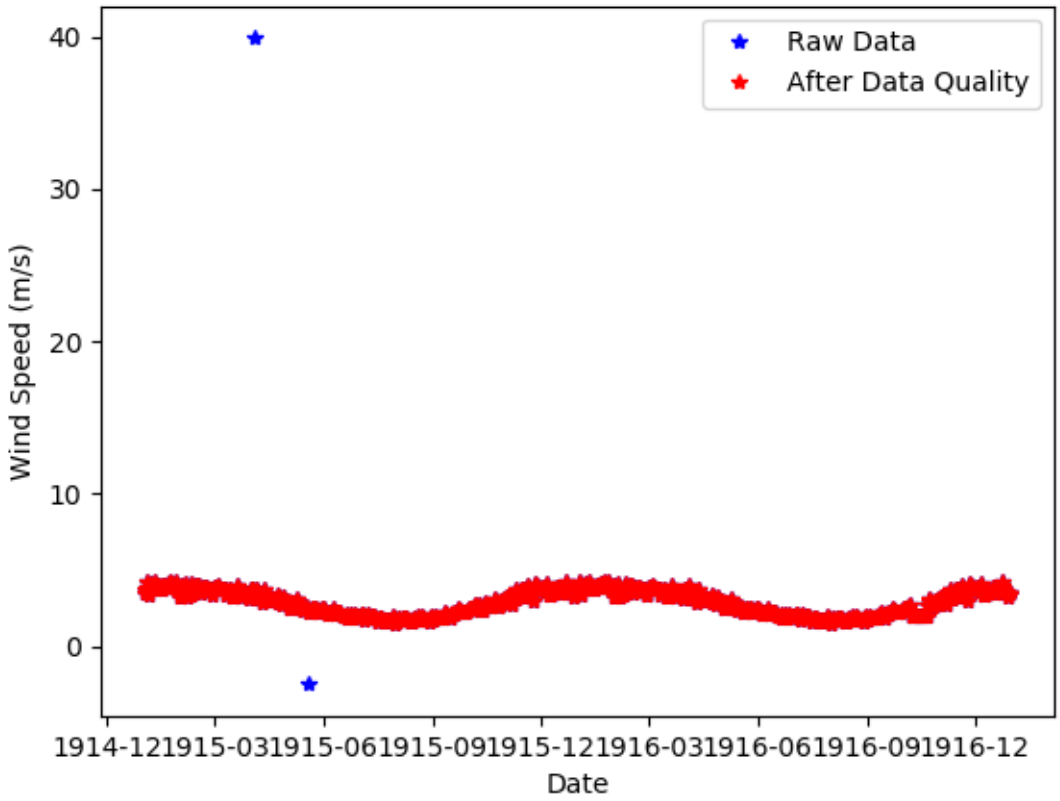

Figure 4. Wind Speed scatter plot showing original data (blue) against the corrected data (red). The data had no -999 values and removed several points out of range.