Metadata

Name of Script: program_09.py
Creator: Kush Paliwal
Created On: 20th April 2020

Program Objective:

- Read meteorological data file (DataQualityChecking.txt) containing daily precipitation, daily maximum and minimum air temperature, and daily wind speed.
- Perform four data quality checks on the given data
- Plot and save each dataset before and after correction has been made
- Save the data that has passed the quality check into a new file (Modified data.txt)
- output a summary of the failed checks to a separate Tab delimited file (Error Check.txt).

Data Quality Checks Performed

• Check 1: Removes No Data values.

Replace all values of -999 (no data values) in this file with the NumPy NaN values. Record the number of values replaced for each data type in the dataframe “ReplacedValuesDF” with the index “1. No Data”.

• Check 2: Check for gross errors

Identify all values outside the thresholds 0 <= P <= 25; -25 <= T <= 35, 0 <= WS <= 10 and replace with NaN. Record the number of values replaced for each data type in the dataframe ReplacedValuesDF with the index "2. Gross Error"

• Check 3: Swap Max Temp and Min Temp when Max Temp is less than Min Temp.

Check that all values of Max Temp are greater than for Min Temp for the current day's observations. Where they are not, swap the values. Record the number of values replaced for each data type in the dataframe ReplacedValuesDF with the index "3. Swapped"

• Check 4: Check for daily temperature range exceedance.

Identify days with temperature range (Max Temp minus Min Temp) greater than 25°C. When range the value is exceeded replace both the maximum and the minimum values with NaN. Record the number of values replaced for each data type in the dataframe ReplacedValuesDF with the index "4. Range Fail"

Failed Checks

Errors present in each dataset.

	Precip	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 Fail	0	5	5	0

Plots after performing quality check

Precipitation

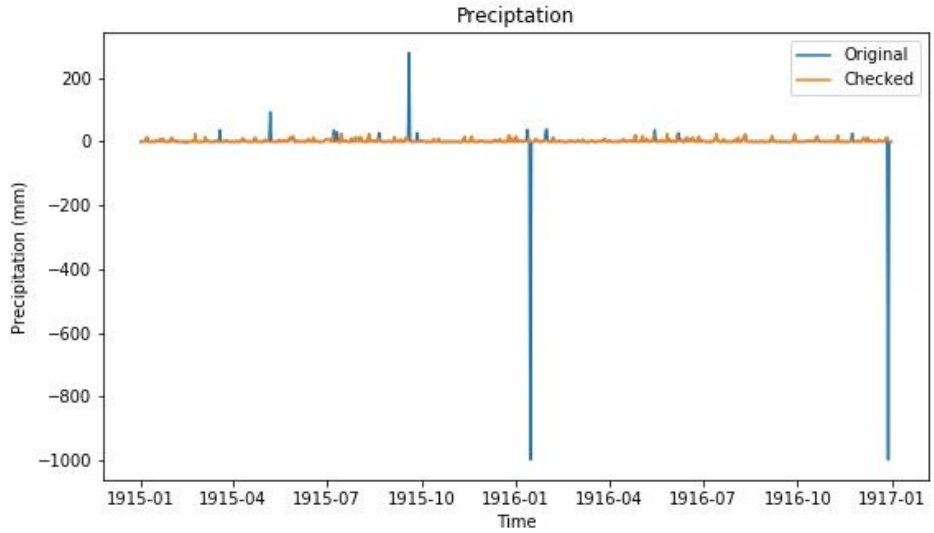

Figure 1 Precipitation Data before and after quality check.

Figure 1 shows the comparison between original and the quality checked data of daily precipitation. Most part of the plot overlaps each other. Two distinct spikes can be seen reaching the value of -999. They represent the no data values in this data, while the other shorter spikes represent the 15 gross errors located outside the range of 0 and 25 mm.

Maximum Air Temperature

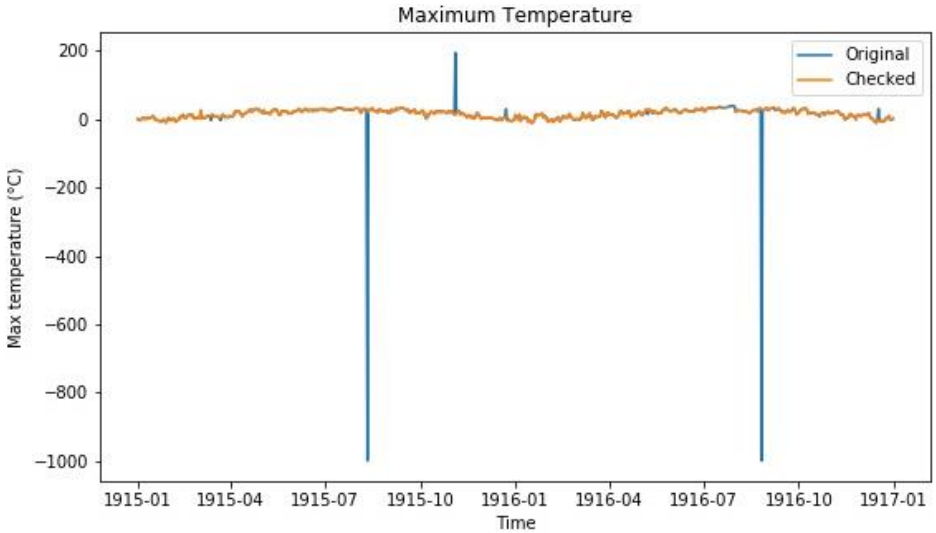

Figure 2 Maximum Air Temperature Data before and after quality check

Figure 2 shows the comparison between original and the checked data of daily maximum air temperature. Most part of the plot overlaps each other. Two distinct spikes can be seen reaching the value of -999. They represent the no data values in this data, while the other shorter spikes represent the 2 gross errors (outside the range, $-25 \leq T \leq 25$), 4 swapped errors ($T_{max} < T_{min}$) and 5 range exceedance errors ($T_{max} - T_{min} > 25$).

Minimum Air Temperature

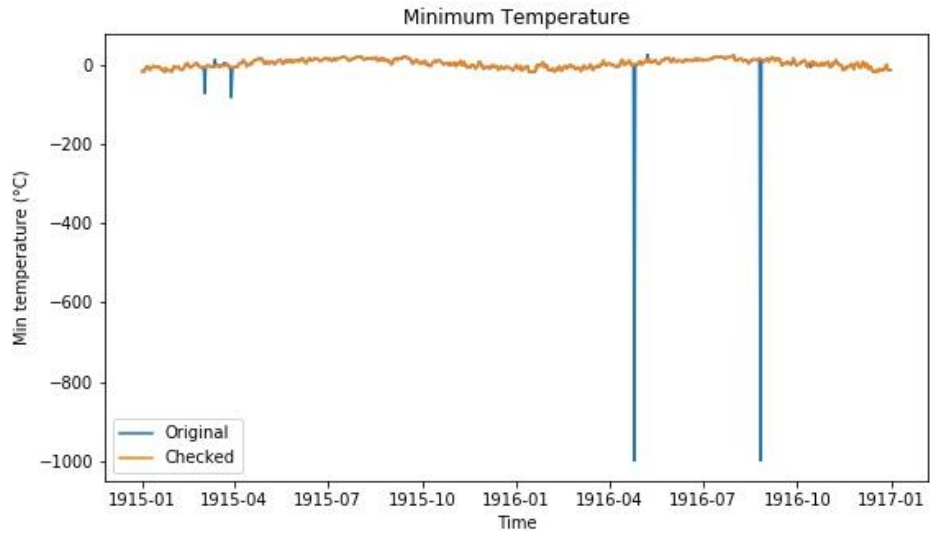

Figure 3 Minimum Air Temperature Data before and after quality check

Figure 3 shows the comparison between original and the checked data of daily minimum air temperature. Most part of the plot overlaps each other. Two distinct spikes can be seen reaching the value of -999. They represent the no data values in this data, while the other shorter spikes represent the 14 gross errors (outside the range, $-25 \leq T \leq 25$), 4 swapped errors ($T_{max} < T_{min}$) and 5 range exceedance errors ($T_{max} - T_{min} > 25$).

Wind Speed

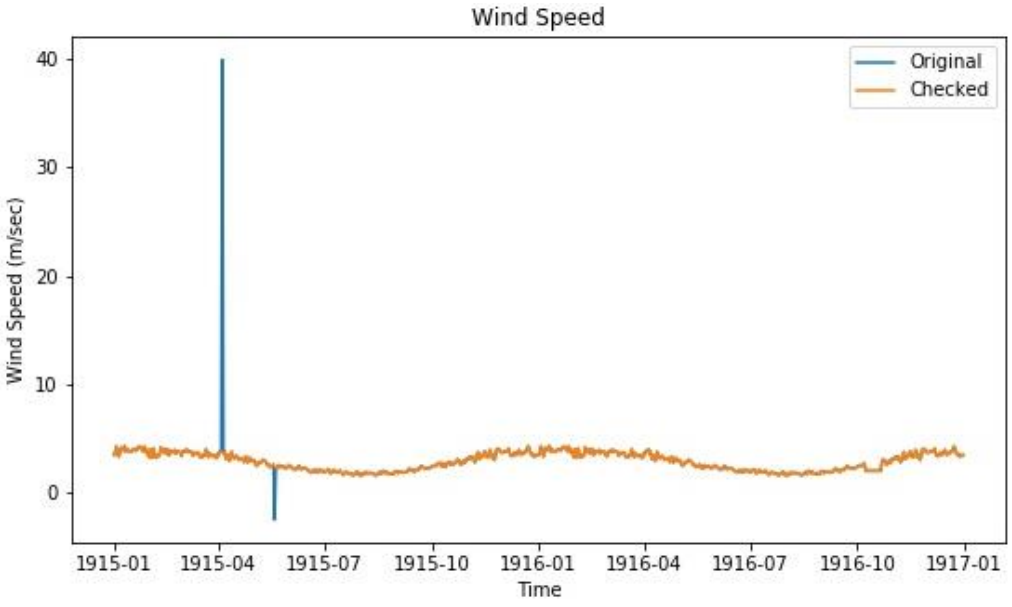

Figure 4 Wind Speed Data before and after quality check

Figure 4 shows the comparison between original and the checked data of daily wind speed. Most part of the plot overlaps each other. The two spikes represent the two gross errors (values outside the range, $0 \leq WS \leq 10$ m/s).