Metadata File for ABE 65100 Lab 09

Created March 22, 2020
by Miriam Stevens

Program name: program_09.py
Program template: program_09_template.py
Input file name: DataQualityChecking.txt
Output files: Checked-data.txt, Fail-checks-summary.txt, Precipitation-Before-After.pdf, Wind-speed_Before-After.pdf

All files related to Lab 09 can be found in the GitHub repository accessible at the following link: https://github.com/Environmental-Informatics/09-data-quality-checking-steve276.git

Summary of Script

The script *program_09.py* performs four data quality checks on the dataset *DataQualityChecking.txt,* which is a time series dataset of daily measurements of precipitation (P), minimum and maximum temperature (T), and wind speed (WS) between 1915 and 1916. The program reads the data into a pandas module data frame, records how many values do not meet the data quality conditions, and currently performs two of the following four checks:

Check 1 - Remove No Data values
This quality check replaces No Data values, defined as -999, with the numpy module null value, NaN. The result of this check can be clearly seen in Figure 1. When values of -999 are removed, the daily precipitation values are never negative and reach a maximum of about 20mm.

Check 2 - Check for gross errors
This quality check replaces values that fall outside of an expected range with NaN. The acceptable ranges for each variable are: $0 \leq P \leq 25$; $-25 \leq T \leq 35$, $0 \leq WS \leq 10$.
The results of this check is clearly shown in Figure 2 of wind speed. Only WS values between 0 and 10 are kept in the dataset. Once values outside the error threshold are removed, an annual wind speed pattern can be observed.

Check 3 - Swap Max Temp and Min Temp when Max Temp < Min Temp
This quality check swaps the values for Max Temp and Min Temp if Max Temp < Min Temp.

Check 4 - Check for daily temperature range exceedence
This check is to replace both the Max and Min Temp values for a given day if the difference between the two temperatures is greater than 25 °C.

For all checks, the occurrences of failure are counted and added to a data frame that summarizes the number of values that fail by data type. See Table 1. for the final fail check summary.

	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

Table 1. Summary of data that failed quality checks.

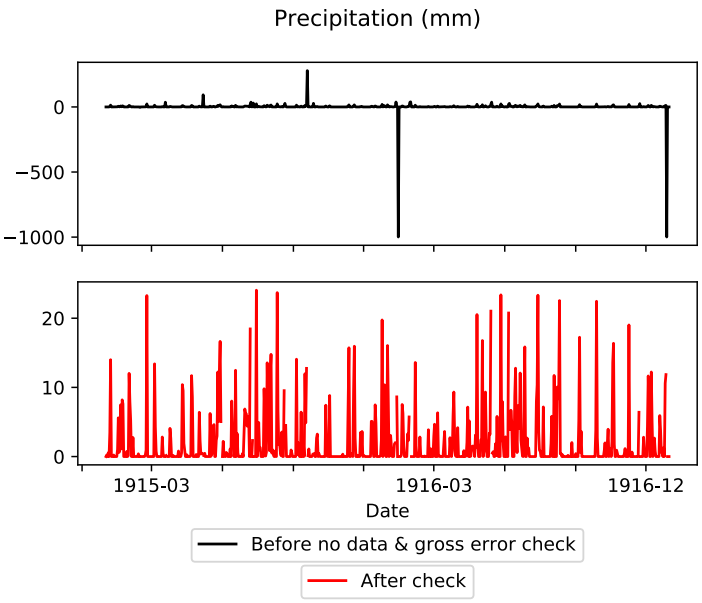

Figure 1. Precipitation data before and after checks 1 & 2.

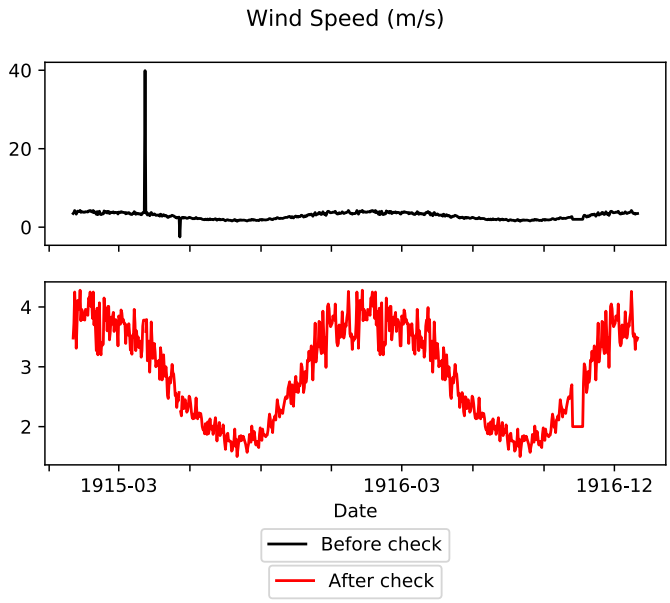

Figure 2. Wind Speed data before and after check 2.