Manitoba Métis Federation
Weather Data Cookbook

CENTRE FOR EARTH OBSERVATION SCIENCE

Document Control

0.1 Version History

Version	Author(s)	Type	Date Modified	Comments
1.0	Friesen, K. L.	Previous version	2022/02/07	Outdated workflow.
1.2	Friesen, K. L.	Working Copy	2022/02/25	Final draft.
1.3	Heppner, K. L.	Working Copy	2023/07/12	Updated standardized variable names and descriptions.

0.2 Document Location

A digital copy of the document can be found in CEOS codebooks repository.

0.3 License

Centre for Earth Observation Science(2022). Weather Data Cookbook, Version 1.3. Centre for Earth Observation Science . University of Manitoba.

Contents

1 Weather Station Data

1.1 Critical Summary

Weather station data is collected from 2 different locations in Northern Manitoba surrounding the Lake Manitoba and Lake Winnipegosis. Currently, 2 stations are setup on Métis property with three more stations planned for deployment in 2023. The first weather station is referred to as St.Laurent and is located on the east shore of Lake Manitoba's south basin. The second station is referred to as Dawson Bay, which collects data on the northwest shore of Lake Winnipegosis. This data collection is a part of the Geoconnections project, where CanWIN represents a third party to support the governance and sovereignty of the Métis. Here we describe the cleaning process of weather station data for archiving on CanWIN's ckan site.

1.1.1 Data Management Tools

1. **Weather_data_QAcompiler.R**: R script that standardizes variables names with controlled vocabulary from CF and BODC conventions as well as associates a result value qualifier metadata field with each variable to identify 'erroneous values'.

2. **CanWIN DataHub Upload**: Upload steps for data shared on CanWIN's data site.

1.2 Weather Station Data

1.2.1 Raw Weather Station Files

Dataset Description

Weather station data is collected via an Iridium satallite. The satallite is used to transmit weather data from the weather stations to a live feed on a computer. Data is logged at 15-minute intervals and a relatively "live" update of the data is provided by DataGarrison and created into a user friendly dashboard. The "live" data is updated once every hour at roughly XX minutes after the hour, depending on when the logger was started or restarted.

DataGarrison Sites:

- St Laurent - live feed
- Dawson Bay - live feed

Dashboard Sites:

- St Laurent Li Taan Aen Staansyoon
- Dawson Bay Li Taan Aen Staansyoon

File Name(s)

1. **MMF_Location_MonthDD_YYYY_MonthDD_YYYY.txt**:
   `MMF_St_Laurent_Oct13_2021_Oct15_2021.txt`

File Source and Location

Raw weather station files are downloaded from the Data Garrison website only when generated by the website, which only occures when the logger box is reset during troubleshooting. The task of downloading will be maintained by CEOS and are currently stored under the Weather Keeper program repository in specific folders separated by station name/location.

You will need to clone the Weather Keeper program repository to your desktop. If you do not know how to clone a repository, consult the GitHub Docs.

Note: Please make sure you are saving and working in the correct raw and processed folders for the weather station you have downloaded the raw files for, either **StLaurent** or **Dawsonbay**.

Raw Dataset Variables

Table 1.1: Variables in raw weather station data downloaded from Data Garrison site.

Header	**Description**	**Data Type**	**Range Values**	**Units**
Date_Time	Date and time of measurement	datetime	N/A	YYYY/MM/DD HH:MM
Pressure_ 20812842_mbar	Air pressure measurement	numerical	660-1070	millibar (one tenth of a bar) mbar
PAR_ 21040570_uE	Photosynthetically Active Radiation measurements for photosynthetic light (400 to 700 nm) wavelentgths in air	numerical	0-2500	microEinsteins per meters squared per second ($\mu E/m^2/s$)
Temperature_ 21048243_deg_C	Temperature of air	numerical	-40 to 75	degrees celsius ($^\circ$C)
RH_21048243_%	Relative humidity	rational number shown numerical	0-100	Percent of actual vapour pressure over saturated vapour pressure (%)
Rain_ 21050258_mm	Quantity of rain collected in a month	numerical	0-127 mm per hour, maximum 4000 tips per logging interval	millimeters (m/s)
Wind Speed_ 21055169_m/s	Movement of air at location	numerical	0-100	meters per second (m/s)
Gust Speed_ 21055169_m/s	Sudden gust or increase in air movement lasting under 20 seconds and in 2-minute intervals	numerical	0-50	meters per second (m/s)
Wind Direction_ 21055169_deg	Angle of air movement	numerical	0-355	degrees
Backup Batteries_20827695_V	Voltage of car battery	numerical	13.200 - 0.001	volts (V)

1.3 Scripts and Analytical Processes

1.3.1 Step 1: Weather Station Standardizing and Metadata

Script Type: **R Script**
File(s) In: raw weather station data in .txt format
File(s) Out: processed weather station data in .csv format

User Instructions

This section will describe how to download historical, or archived, weather data from the Data Garrison site and when you should download from this site.

CAUTION: Please make sure you are saving and working in the correct raw and processed folders for the weather station, either **StLaurent** or **Dawsonbay**.

Checking for archived weather station data:

1. Navigate to the Data Garrison site and log into the user service using the serial number of the weather station and password associated with the account.
2. This information can also be found on our private Gitlab Weather Keeper Program README.md file (at the bottom of the page).
3. Once you have logged in, select a weather station from the list on the current page.
4. You should now see a live view of the data plotted on various graphs.
5. Select **Control Panel** from the menu ribbon at the top of the page right-hand side of your browser screen.
6. Scroll down to the **Download data** section and see if there are any files that have been archived.
   (a) In another tab on your browswer check the raw data (Data/raw) folder in the Weather Keeper Program repository of the specific weather station.
      i. St Laurent Li Taan Aen Staansyoon
      ii. Dawson Bay Li Taan Aen Staansyoon
7. If the file you see is not included in the raw folder then you can proceed to download the file.

Note: the last file in the **Download data** table represents the current data range period to date and also lacks a red x in the last column of the table. Date that has been archived will show a red x, which signifies you are able to delete the record.

Downloading archived data:

1. In the **Download data** table, select the **Description** of the file you wish to download.

2. Select **tab delimited** beside **Choose desired file type**.
3. Right-click on the .txt file name and select **Save link as...**.
4. Navigate to your cloned **geoconnect_grant** repository and place the file under the data/raw folder.
5. Click save.

To run the R script:

To run **Weather_data_QAcompiler** script, you must have both **R** and **RStudio** installed on your computer. Please refer to this video for downloading R on a Windows OS or Mac OS. Additionally, you can refer to this guide for more information.

1. Open RStudio either from your program start menu or by opening the script in the **weather keeper program** repository in the data/scripts.
2. Make sure the raw weather station file(s) (.txt) are all located in the same folder (data/raw).
3. Place cursor at indicated line in the script and click the **Run** button at the top right-hand corner of the script window.
4. Follow dialogue prompts to direct script to the appropriate folder location where the raw data is stored. Note: you wont see raw data file(s) in the folder in this step.
5. Follow prompts to direct script to save processed file(s) in the data/proccessed folder of the **geoconnect_grant** repository.
6. Sit back and enjoy, the script should run in less than a minute.
7. The location to where the "..._QA-process_.." file(s) is saved should be printed in the **Console** window.
8. If you have any issues email portalco@umanitoba.ca.

Analytical Process

This is a single R script that performs the following tasks:

1. Requests you to select the directory to where the raw weather station file(s) are saved.
2. Requests you to select the directory to where the processed weather station file(s) will be saved.
3. The script will then standardize the variable names of the raw file(s) and add a **variable_result_value_qualifier** field beside each variable.
4. The script will then go through each line of data checking to see if the measurement is within the calibrated range of the sensor or if the measurement is likely erroneous due to specific weather conditions.

5. For specific sensors issues, that are recorded in the Deployment Details for each station, specific filters are applied for a given time range if the sensor failed, rebooted, or was operating within a specific season that it shouldn't.

6. A code is then recorded in the **variable_result_value_qualifier** field to indicate an issue with the measurement.

7. These codes are provided below in Table 1.3.

Table 1.3: Varaible Result Variable Qualifier.

Code	**Description**
Above Detection Limit	ADL
Below Detection Limit	or BDL
Probably bad	prob_bad
Not collected	NC
Field Equipment Failure	FEF

1.4 Intermediate Data Files

In order to upload the processed weather station data to the CanWIN DataHub we will need to compile all the archived files into a single file. This next section is completed with one script.

1.4.1 Step 2: Compiling Standardized Weather Station Data

Dataset Description

Weather station files that have been processed are intermediate output files and required compiling in order to appropriately present them on our CanWIN site.

File Name(s)

- **MMF_Location_MonthDD_YYYY_MonthDD_YYYY_QA-process_YYYY-MM-DD.csv**: `MMF_St_Laurent_Oct13_2021_Oct15_2021_QA-process_YYYY-MM-DD.csv`

File Source and Location

- weather station processed folder cloned from the Weather Keeper program repository.

QA Dataset Variables

Table 1.5: Variables of QA/QC weather station data.

Header	Description	Data Type	Range Values	Units
Date_and_Time	Date and time of measurement	datetime	N/A	YYYY/MM/DD HH:MM
air_pressure	Air pressure measurement	numerical	660-1070	millibar (one tenth of a bar) mbar
air_pressure_result_value_qualifier	Air pressure metadata field	text	"ADL" or "BDL"	N/A
Photosynthetically_Active_Radiation	Photosynthetically Active Radiation (PAR) measurements for photosynthetic light (400 to 700 nm) wavelentgths in air	numerical	0-2500	microEinsteins per meters squared per second ($\mu E/m^2/s$)
Photosynthetically_Active_Radiation_result_value_qualifier	PAR metadata field	text	"ADL" or "BDL"	N/A
air_temperature	Temperature of air	numerical	-40 to 75	degrees celsius ($^\circ$C)
air_temperature_result_value_qualifier	Air temperature metadata field	text	"ADL" or "BDL"	N/A
relative_humidity	Relative humidity (RH)	rational number shown numerical	0-100	Percent of actual vapour pressure over saturated vapour pressure (%)
realtive_humidity_result_value_qualifier	RH metadata field	text	"ADL" or "BDL"	N/A
Precip	Quantity of rain collected in a month	numerical	0-127 mm per hour, maximum 4000 tips per logging interval	millimeters (m/s)
Precip_result_value_qualifier	Air pressure metadata field	text	"ADL", "BDL", or "NC"	N/A
wind_speed	Movement of air at location	numerical	0-100	meters per second (m/s)

wind_speed_result_value_qualifier	Wind speed metadata field	text	"ADL", "BDL", "prob_bad", or FEF	N/A
wind_speed_gust	Sudden gust or increase in air movement lasting under 20 seconds and in 2-minute intervals	numerical	0-50	meters per second (m/s)
Wind__speed_gust_result_value_qualifier	Wind speed metadata field	text	"ADL", "BDL","prob_bad", or FEF	N/A
wind_from_direction	Angle of air movement	numerical	0-355	degrees
wind_from_direction_result_value_qualifier	Wind speed metadata field	text	"prob_bad"	N/A
battery_output	Voltage of car battery	numerical	13.200 - 0.001	volts (V)
battery_output_result_value_qualifier	Battery output metadata field	text	"FEF"	N/A

User Instructions

To run the R script:

To run Weather_station_compiled.R script, you must have both **R** and **RStudio** installed on your computer.

1. Open RStudio either from your program start menu or by opening the script in the **weather keeper program** repository in the data/scripts.
2. Make sure the QA weather station file(s) (.csv) are all located in the same folder (data/processed).
3. Place cursor at indicated line in the script and click the **Run** button at the top right-hand corner of the script window.
4. Follow dialogue prompts to direct script to the appropriate folder location where the QA data is stored. Note: you wont see the QA data file(s) in the folder in this step.
5. Sit back and enjoy, the script should run in less than a minute.
6. The location to where the "..._compiled_.." file(s) is saved should be printed in the **Console** window.
7. If you have any issues email portalco@umanitoba.ca.

Analytical Process

This is a single R script that performs the following tasks:

1. Requests you to select the directory to where the QA processsed weather station file(s) are saved.
2. The script will take the first file in the processed folder and insert all the lines of data from the other file into the first file.
3. The script will create a new dataset and save as a single compiled data file.

Note: this process is intended for creating the first intial compiled data file, as well as updating previous compiled files when more data is archived on the DataGarrison site. Once you have created the compiled file from the QA files, move the QA files into the **Archive-QA** folder. There should only be one compiled file in the **data**/**processed** folder of the **geoconnect_grant** repository at all times.

1.5 Final Output Data Files

Uploading the data can be completed by MMF technician to CanWIN's DataHUB site. All you need to do is create and account, which will allow you to edit MMF and Weather Keeper Program content on the site. If you are having any problems with editing content or if you have any questions, please feel free to contact portalco@umanitoba.ca.

1.5.1 Uploading Data Files

Dataset Description

The compiled weather station data is the final output file that will need to be uploaded to the corresponding weather station dataset page on the CanWIN site and available for open-access. Archived raw data on the DataGarrison site will go through the same procedure and appended to the compiled file once undergoing the QA process.

Currently, raw weather station data are downloaded and curated by CanWIN staff. Once data is ready for upload, the files are uploaded to their respective dataset pages (below) under the Weather Keeper program on the CanWIN DataHub. Additionally, the archived file (if not recorded) will be updated on the deployment details info section of the page. This process will be transitioned to MMF technicians in the future.

CanWIN DataHub pages:

- St Laurent Li Taan Aen Staansyoon dataset page
- Dawson Bay Li Taan Aen Staansyoon dataset page
- St Laurent Li Taan Aen Staansyoon Deployment Details
- Dawson Bay Li Taan Aen Staansyoon Deployment Details

File Name(s)

1. **MMF_Location_MonthDD_YYYY_MonthDD_YYYY_compiled_YYYY-MM-DD.csv**:
   `MMF_St_Laurent_compiled_2022-02-23.csv`

File Source and Location

After compiling the QA-process file(s) into a single .csv file, you can now transfer the QA-process files into the **Archive-QA** folder in the **Data** folder of the **weather keeper program** repository. This ensures that, if need be, you will be able to recompile the QA file(s) into a single file if the compiled file is accidentally deleted or lost. Once you have moved the files into the appropriate folder you should only have the single compiled .csv file in the data/processed folder of the repository.

CAUTION: Please make sure you are saving and working in the correct raw and processed folders for the weather station, either **StLaurent** or **Dawsonbay**.

Compiled Dataset Variables

Compiled data variables are the same as the "...QA-process..." variables, only the individual files can now all be found in one compiled .csv file.

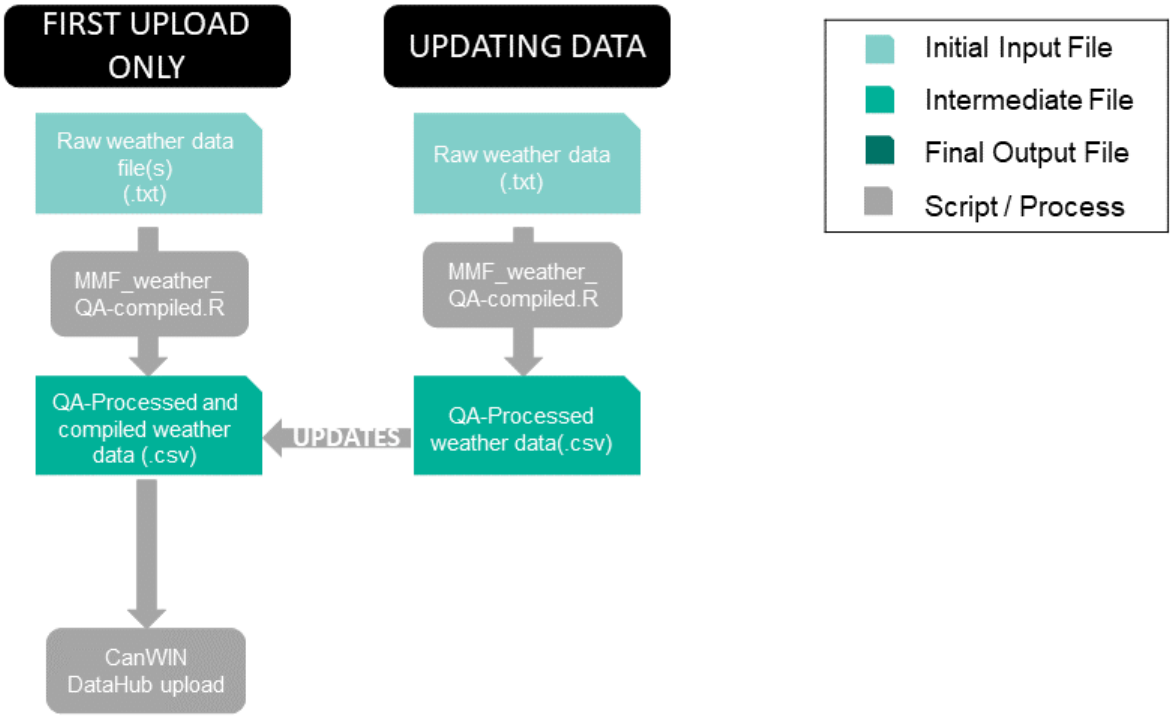

Figure 1.1: Flow chart of Weather Station Data Management.

Glossary

R Script - The most basic form of R code. An R Script is a single .R file that usually requires modification to be run by different users on different computers. 4

A Reference Tables

A.1 Data Levels

Level 0 – Raw data: unprocessed data and data products that have not undergone quality control. Depending on the data type and data transmission system, raw data may be available within seconds or minutes after real-time. Examples include real-time precipitation, streamflow, and water quality measurements

Level 0.1 – First pass QC: A first quality control pass has been performed to remove out of range and obviously erroneous values. These values are deleted from the record. E.g: Online Environment Canada stream-flow data, laboratory data

Level 1 – Quality Controlled Data: Data that have passed quality assurance procedures such as Level 0.1 and have been further quality controlled by data provider before being submitted to CanWIN (e.g. Idronaut data with only downwelling (upwelling data removed) data included.

Level 1.5 – Advanced Quality Controlled Data: Data have undergone complete data provenance (i.e. standardized) in CanWIN. Metadata includes links to protocols and methods, sample collection details, incorporates CanWIN's or another standardized vocabulary, and has analytical units standardized. Note: Process still under development in CanWIN (as of May 13, 2020).

Level 2 – Derived Products: Derived products require scientific and technical interpretation and can include multiple data types. E.g.: watershed average stream runoff derived from stream-flow gauges using an interpolation procedure.

Level 3 – Interpreted Products: These products require researcher (PI) driven analysis and interpretation and/or model-based interpretation using other data and/or strong prior assumptions. E.g.: watershed average stream runoff and flow using streamflow gauges and radarsat imagery

Level 4 – Knowledge Products: These products require researcher (PI) driven scientific interpretation and multidisciplinary data integration and include model-based interpretation using other data and/or strong prior assumptions. E.g.: watershed average nutrient runoff concentrations derived from the combination of stream-flow gauges and nutrient values.

Content retrieved from https://lwbin.cc.umanitoba.ca on July 06, 2020.

A.2 Result Value Qualifiers

ADL	Above Detection Limit
BDL	Below Detection Limit
FD	Field Duplicate
LD	Lab Duplicate
$	Incorrect sample container
EFAI	Equipment failure, sample lost
FEF	Field equipment failed
FEQ	Field Equipment Questionable
FFB	Failed. Field blank not acceptable
FFD	Failed. Field Duplicate
FFS	Failed. Field spike not acceptable
H	Holding time exceeded
ISP	Improper sample preservation
ITNA	Incubation time not attained
ITNM	Incubation temperature not maintained
JCW	Sample container damaged, sample lost
NaN	Value is missing and reason is not known
NC	Not collected
ND	Not detected
NR	Sample taken/measured on site but information in this field not recorded
NS	Sample collected but not submitted
OC	Master Coordinate List Used
P	Analysis requested and result pending

prob_good	probably good value. Data value that is probably consistent with real phenomena but this is unconfirmed or data value forming part of a malfunction that is considered too small to affect the overall quality of the data object of which it is a part
prob_bad	probably bad value. Data value recognised as unusual during quality control that forms part of a feature that is probably inconsistent with real phenomena
Interpolated	This value has been derived by interpolation from other values in the data object
Q	Below limit of quantification (LOQ). The value was below the LOQ of the analytical method. The value in the result field is the limit of quantification (limit of detection) for the method

B Glossary of Options and Packages

B.1 R Packages

Visit https://cran.r-project.org/web/packages/available_packages_by_name.html to learn more about R packages

- **Package 1** - Description
- **Package 2** - Description

B.2 Python

B.2.1 Python Script-Specific Options

- **Option 1** - Description
- **Option 2** - Description

B.2.2 Python Packages

Visit https://docs.python.org/3/library/ to learn more about python packages

- **Package 1** - Description
- **Package 2** - Description

Example: Section 2.1 from Victory's semi-hemi codebook