The Journal of Open Source Software

agweather-qaqc: An Interactive Python Package for Quality Assurance and Quality Control of Daily Agricultural Weather Data and Calculation of Reference Evapotranspiration

Christian Dunkerly[1¶], **Justin Huntington**[1], **Daniel McEvoy**[1,2], **Asa Morway**[3,4], **and Richard Allen**[5]

1 Desert Research Institute **2** Western Regional Climate Center **3** Huntington Hydrologic **4** Battle Born Homeschool Co-op **5** University of Idaho (ret.) ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Taher Chegini

Reviewers:

- @dostuffthatmatters
- @dlebauer

Submitted: 27 January 2024
Published: unpublished

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Introduction

Agricultural weather stations are necessary for collecting representative observations needed for computing reference evapotranspiration (ET), a measure of atmospheric evaporative demand from a well-watered environment, and a key variable for determining crop water use, water requirements, and irrigation scheduling (Allen et al., 1998). Common weather station measurements include near-surface (2 to 3 m height) measurements for incoming shortwave solar radiation, air temperature, humidity, wind speed, and precipitation. Accurate, continuous, and consistent measurement of these variables from years to decades is often lacking due to sensor malfunction, drift, age, poor calibration, debris, limited station maintenance, communication errors, lack of a well-watered environment, and remote access. As a result, poor quality agricultural weather station data are common, and, unless flagged for removal or corrected, will impact the accuracy of reference ET calculations (Allen, 1996). Ensuring that agricultural station data are high quality is especially important for studies where station data are considered to be 'truth' when comparing to related model predictions and forecasts (Blankenau et al., 2020; McEvoy et al., 2022). Similarly, reference ET station data need to represent well-watered environments if they are to be used to validate or to bias-correct gridded reference ET datasets (Allen et al., 2021; Huntington et al., 2018).

Agricultural weather station data are increasingly being used to support satellite-based remote sensing of agricultural water use. The weather data are often collected from dozens of networks and hundreds of stations and requires large-scale quality assurance and quality control (QAQC) systems to ensure accuracy and representativeness (Huntington et al., 2022; Melton et al., 2022). A common challenge has been the development of QAQC workflows and visualization tools to support implementation of recent advances in recognizing and correcting data error. Reliable, high-quality records of weather observations will become even more important as scarcity in freshwater resources continues to become a more prevalent issue in the world (Kummu et al., 2016; Steduto et al., 2012).

Statement of Need

Having the ability to easily read, visualize, review, flag, and potentially remove, fill, or adjust historical and real time weather data is necessary to advance research and applications focused on evapotranspiration and water use estimation at local to global scales. The development of the `agweather-qaqc` Python package is intended to enable rapid, thorough, and efficient weather data review and QAQC for daily weather data. The `agweather-qaqc` package is a command-line interface (CLI) based, open-source Python package for reading, visualizing, and performing QAQC of daily weather station observations and calculation of reference ET from a wide range of data networks. Many station networks use different storage formats and recorded variables. For example, a network might report humidity data as vapor pressure, dew point temperature, specific humidity, or maximum, minimum, and average relative humidity. As a result, data input variables, units, and conversions are configurable within `agweather-qaqc` to flexibly handle multiple common input variables and formats so that all input data, visualization, QAQC, and calculation of reference ET can be performed in a programmatic, consistent, and easily repeatable manner. Additionally, the CLI-based approach enables researchers and practitioners that are not overly proficient with Python to easily use and understand the software through the use of helpful reminders, prompts, and recommended parameters. While this package was written with a focus on agricultural stations, it can be used to QAQC any source of weather data, but care must be taken when making corrections or drawing conclusions about reference ET.

Design and Features

The `agweather-qaqc` package contains the WeatherQC python class that first reads and standardizes the data, and then allows the user to QAQC the data (Figure 1).

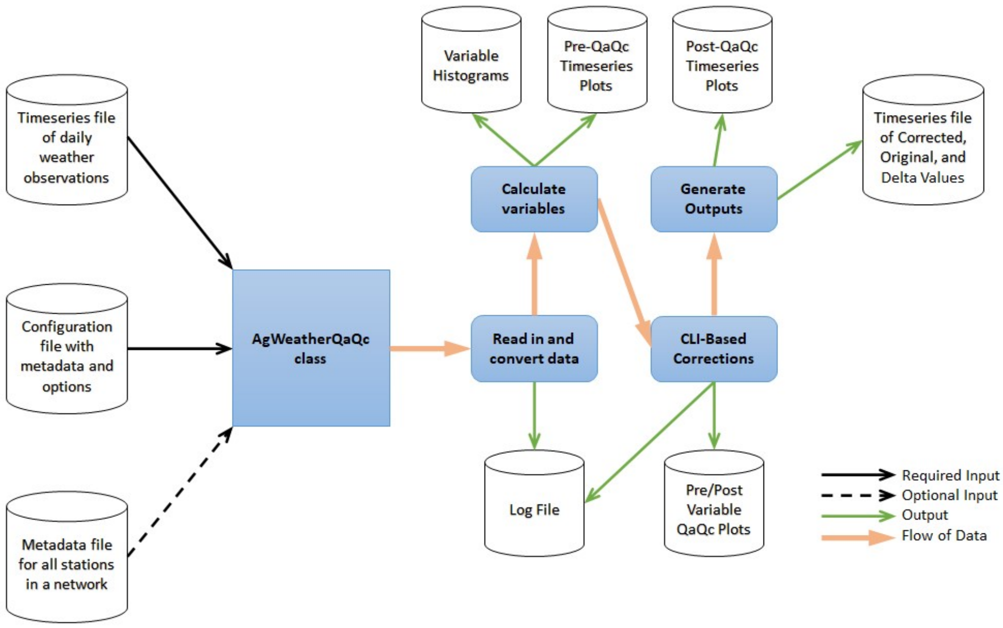

Figure 1: Diagram showing the required inputs, flow of data, and outputs produced from the software.

Inputs to `agweather-qaqc` are a tabular data file of daily data and a configuration file that details where variables are arranged within the data file. The configuration file also provides metadata about the station, such as latitude, longitude, elevation, and units of the variables. An optional input of a metadata file for multiple stations may also be provided, enabling a single configuration file to be used for any number of data sources that share a common format.

The `agweather-qaqc` package contains example files of each input.

Once given data, the WeatherQC class performs unit conversions and removal of unreasonable values according to limits and ranges specified in the documentation, (e.g., negative wind speed readings). This is followed by calculating variables, including ASCE reference ET, as well as diagnostic variables of theoretical clear-sky solar radiation (R_{so}) and dew point depression (i.e., daily minimum temperature minus dew point temperature). Command-line interface (CLI) prompts facilitate corrections on desired variables, and includes recommendations as to best practices. All downstream dependent variables are recalculated once each correction is performed to assemble the quality-controlled version of the data. Once the user is finished making adjustments, output files are generated that include records of all changes through both a human-readable log file and interactive time series plots of pre- and post-processed observations. The generated output files form a complete record of original and processed observations with correction flags and amounts and station metadata so that a documented data archive is available for later use.

Two optional functionalities can be enabled by the configuration file: The first automates some of the interactive prompts to use the suggested QC procedures and parameters at each step. The second provides statistical gap-filling of any missing or removed observations to enable weather and reference ET summaries for any period of record.

One of the most useful features of the software is the interactive visualization of weather variable time series that facilitate rapid human data assessment and pattern recognition.The observation data are displayed in Bokeh (2018) time series plots both before and after adjustment via the QAQC process, enabling users to readily visualize and assess data patterns and trends related to sensor drift, probable miscalibration, data outliers, sensor malfunction, etc. These plots are accompanied by interactive tools (ex. pan and zoom controls, display info on hover), and feature linked axes, allowing the user to readily visualize and assess how variables vary and covary over time and at different time scales (e.g. daily, monthly, annual). These graphs are saved as stand-alone HTML files to allow for easy sharing of results.

The interactive plots also facilitate viewing data patterns at different timescales, such as showing how minimum temperature and dew point temperature vary at daily and monthly timesteps, as well as how the difference in these variables, known as dew point depression, varies throughout the year. Dew point depression gives insight to how well-watered the area around the station may be: transpiration from plants and evaporation from soil both increases humidity, which results in increased dew point and reduced dew point depression. Higher than expected or abrupt changes in dew point depression values, might indicate non-ideal reference ET station siting, land cover changes, or erroneous data.

While ease-of-use for a non-technical user was one of the principal goals, the software workflow can be automated with the inclusion of libraries such as PyAutoGUI.

The `agweather-qaqc` package features include:

- Importing daily observational data without having to convert it to a standardized format, with unit conversions based on a user-specified configuration file.
- Converting multiple input formats from separate sources or networks into a single, uniform format for easier downstream analysis.
- Visualizing data before and after processing with interactive plots, as daily time series and as mean monthly averages.
- Filtering and removal of data, both manual and automatic, with statistics-based approaches to identify and correct issues such as sensor miscalibration.
- Calculation of theoretical clear-sky solar radiation using date, location and elevation information along with humidity data based on ASCE standardizations (Allen et al., 2005).
- Calculation of expected solar radiation using the empirical Thornton-Running approach (Thornton & Running, 1999) with Monte-Carlo optimized empirical parameters based on observed solar radiation data.
- Calculation of the daily dew point depression (i.e., daily minimum temperature minus daily average dew point temperature) used to assess whether the data collection environment included a well-watered surface having expected feedbacks on near-surface humidity and

air temperature.

- Calculation of grass and alfalfa reference evapotranspiration according to the American Society of Civil Engineers (ASCE) Standardized Reference Evapotranspiration equation (Allen et al., 2005) via the RefET (Morton, 2022) library.
- Evaluating station aridity through the visualization of both relative humidity and dew point depression plots, with the option to adjust relative humidity if required.
- Optional gap-filling of data using station climatologies, empirical approaches (e.g. Thornton-Running solar), or random sampling from distributions based on previous observations, resulting in a complete record of daily reference evapotranspiration for cumulative monthly and annual totals.
- Creating archival charts and log files that record and flag how each variable was changed during the QAQC process.

The documentation provides more detail on these features and includes a tutorial with usage examples. An environment.yml is included for installation and third-party package management.

Selected Example

Inaccuracies in measured incoming shortwave solar radiation (R_s) are common due to pyranometer calibration drift, non-level baseplate, sensor degradation, and sensor obstructions or debris covering the sensor (e.g. accumulated dust or residue).

`agweather-qaqc` can be used to visualize measurements of R_s with respect to theoretical clear-sky R_s (R_{so}). A common use of `agweather-qaqc` is to perform corrections of measured R_s. Corrections are performed by dividing the record into specified periods of time (e.g., 60 days), and comparing R_s measurements for specified periods against R_{so}, which is a function of day of year, atmospheric water vapor, elevation, and latitude (Allen, 1996). R_s should approach R_{so} when a cloud-free day occurs, and potentially slightly surpass it due to incident reflected R_s (e.g., distant cloud that doesn't obstruct sunlight) (Allen et al., 2005).

It is expected that a cloud-free day should occur with some regularity for areas with significant agriculture, especially in semi-arid and arid areas. In the example time series shown in Figure 2.A (Figure 2), observed R_s from before July 2013 frequently approaches R_{so}, and after this time period, R_s rarely approaches R_{so}, especially during summer and winter months, indicating that the pyranometer is likely reporting inaccurate values and that observed data should be adjusted.

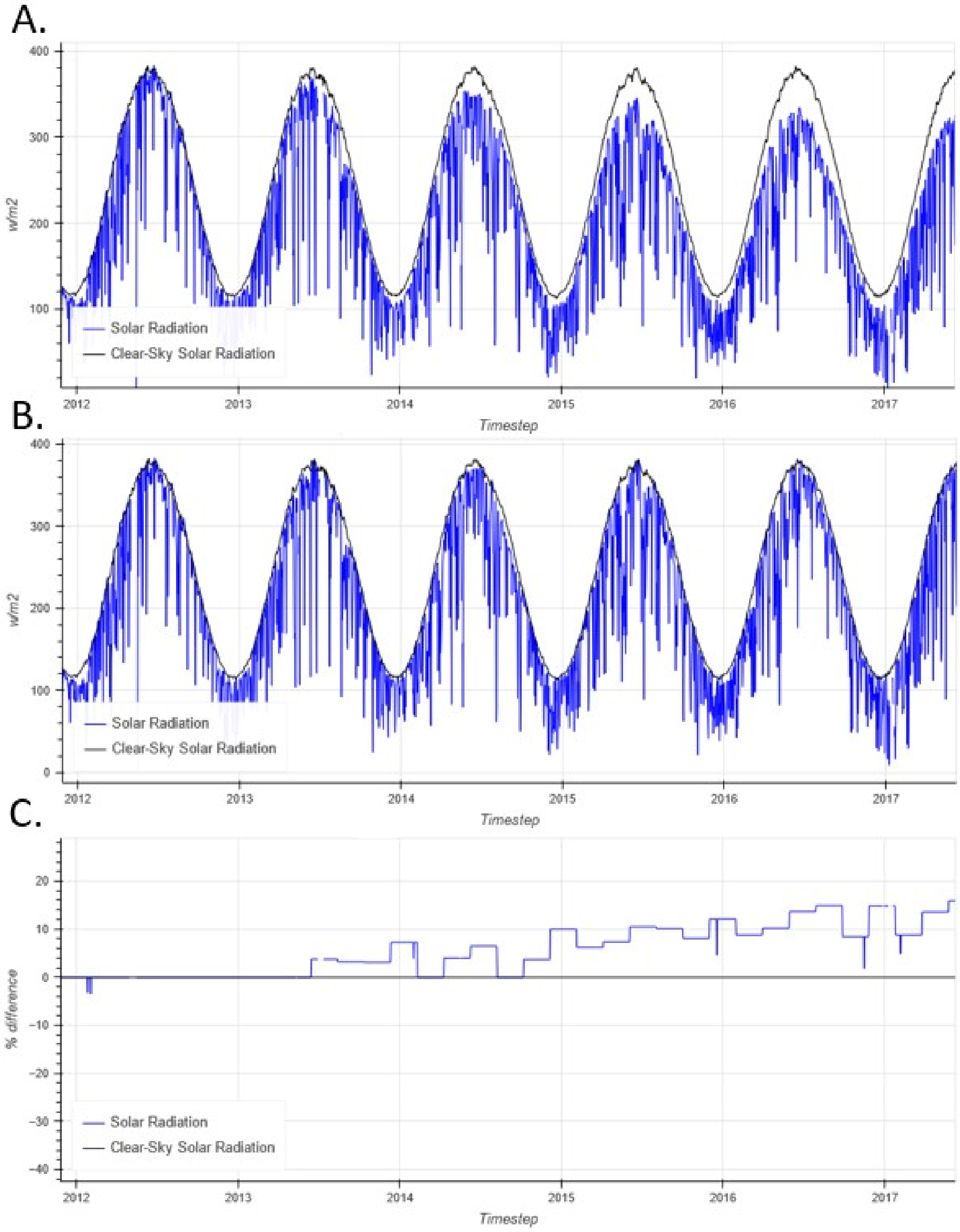

Figure 2: Daily shortwave radiation compared against expected clear-sky solar radiation for a weather station in Reno, Nevada. (A) The downward drift in the sensor readings over the four year period is made obvious by comparing against the theoretical clear-sky values. (B) Post-QC daily shortwave radiation compared against clear-sky solar radiation, the downward drift has been adjusted back to the expected behavior of R_s. (C) The percent difference between pre- and post-QC R_s values as a result of applying the $\frac{R\ so}{R\ s}$ correction factor ratio to each sixty-day period..

To perform the adjustment, measured R_s data are parsed into 60-day periods. Each period is first checked for bad or suspect values, and then the ratio of R_{so} to R_s is calculated from the average of the six largest daily R_s observations and corresponding R_{so} values (i.e. top 10 percentile for the 60-day period). This ratio is multiplied by all R_s measurements during the respective 60-day period to correct measured R_s data (Figure 2.B). If the R_{so} to R_s ratio is between 0.97 and 1.03, no adjustment is applied. The length of period and number of largest R_s/R_{so} ratios (e.g. percentile threshold) to compute average adjustment ratios for the respective period are configurable through interactive prompts.

After adjustment of R_s is performed, a time series plot of percent difference between pre- and post-corrected R_s is produced for visualization and archiving adjustments made (Figure 2.C). If faulty R_s observations had not been adjusted, reference ET would be underestimated in this case by approximately 5%. This example highlights how agweather-qaqc can be used to visualize and correct weather variables to improve evaporative demand estimates.

Limitations:

- Recommendations are made for procedures and default parameters to use, but the end user is ultimately responsible for specifying parameters and corrections made.
- Knowledge about the station metadata is required, such as latitude, elevation, anemometer measurement height, and any field conditions potentially compromising data quality (e.g. tree located next to station obstructing wind speed and R_s measurements, dryness of the weather station environment).
- If the weather station is not located over a well-watered surface according to ASCE guidelines, reference ET will likely be overestimated (Singh et al., 2023).

State of the Field

agweather-qaqc builds on the concepts and functionality of the Ref-ET Software (Allen, 2000), which performs QAQC of data and calculates ASCE reference ET. agweather-qaqc builds on these capabilities, is open source, streamlines the processing of data, and increases accessibility and batch processing by operating outside the Windows platform. Open source access also allows for the inclusion of additional methods and options. Other packages that have been developed, such as the R package CrowdQC+ and the python library titanlib, feature statistics-based QAQC of data, however they focus on spatial variation between multiple sources to detect outliers. agweather-qaqc differs from these libraries by focusing on the temporal variation within a single source.

Research enabled by agweather-qaqc

This software was used to generate a benchmark agricultural weather station reference ET dataset for intercomparison and bias correction of a gridded weather dataset, gridMET (Abatzoglou, 2013) as part of OpenET (Melton et al., 2022), an online platform that provides satellite-based actual ET estimates at field scale across the western United States. This software has also been used to generate datasets of reference ET for weather stations in the Upper Colorado River Basin (UCRB) as part of a yearly Bureau of Reclamation publication of agricultural consumptive uses and losses (Pearson et al., 2019, 2020, 2021), intercomparison of OpenET models for the UCRB (Huntington et al., 2022), and was used in a skill analysis of NOAA's Forecast Reference Evapotranspiration (McEvoy et al., 2022).

Acknowledgements

We would like to thank the Bureau of Reclamation, NASA Applied Sciences Program, and the Western States Water Use Program at the Desert Research Institute for providing funding for the development of this software. We would like to thank the ASCE-EWRI Technical Committee on Evapotranspiration in Irrigation and Hydrology for promoting the development and application of QAQC procedures and standardization of reference ET calculation. We would also like to thank John Volk, David Ketchum, and Charles Morton for their technical expertise.

References

Abatzoglou, J. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*(1), 121–131.

Allen, R. (1996). Assessing integrity of weather data for reference evapotranspiration estimation. *Journal of Irrigation and Drainage Engineering*, *122*(2), 97–106.

Allen, R. (2000). REF-ET: Reference evapotranspiration calculation software for FAO and ASCE standardized equations. In *University of Idaho*. https://www.uidaho.edu/cals/kimberly-research-and-extension-center/research/water-resources/ref-et-software

Allen, R., Dhungel, R., Dhungana, B., Huntington, J., Kilic, A., & Morton, C. (2021). Conditioning point and gridded weather data under aridity conditions for calculation of reference evapotranspiration. *Agricultural Water Management*, *245*, 106531.

Allen, R., Pereira, L., Raes, D., Smith, M., & others. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-FAO irrigation and drainage paper 56. *Fao, Rome*, *300*(9), D05109.

Allen, R., Walter, I., Elliott, R., Howell, T., Itenfisu, T., Jensen, M., & Snyder, R. (2005). *The ASCE standardized reference evapotranspiration equation*. American Society of Civil Engineers. https://doi.org/10.1061/9780784408056

Blankenau, P., Kilic, A., & Allen, R. (2020). An evaluation of gridded weather data sets for the purpose of estimating reference evapotranspiration in the united states. *Agricultural Water Management*, *242*, 106376.

Huntington, J., Bromley, M., Morton, C., & Minor, T. (2018). *Remote sensing estimates of evapotranspiration from irrigated agriculture, northwestern nevada and northeastern california*. Desert Research Institute.

Huntington, J., Pearson, C., Minor, B., Volk, J., Morton, C., Melton, F., & Allen, R. (2022). *Appendix g: Upper colorado river basin OpenET intercomparison summary*.

Kummu, M., Guillaume, J. H., Moel, H. de, Eisner, S., Flörke, M., Porkka, M., Siebert, S., Veldkamp, T. I., & Ward, P. (2016). The world's road to water scarcity: Shortage and stress in the 20th century and pathways towards sustainability. *Scientific Reports*, *6*(1), 38495.

McEvoy, D., Roj, S., Dunkerly, C., McGraw, D., Huntington, J. L., Hobbins, M., & Ott, T. (2022). Validation and bias correction of forecast reference evapotranspiration for agricultural applications in nevada. *Journal of Water Resources Planning and Management*, *148*(11), 04022057.

Melton, F., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J., & others. (2022). OpenET: Filling a critical data gap in water management for the western united states. *JAWRA Journal of the American Water Resources Association*, *58*(6), 971–994.

Morton, C. (2022). ASCE standardized reference evapotranspiration (ET). In *GitHub repository*. GitHub. https://github.com/WSWUP/RefET

Pearson, C., Huntington, J., Volk, J., & Dunkerly, C. (2019). *Agricultural evapotranspiration and net irrigation water requirements for the upper colorado river basin: 2018 ET demands model summary*.

DRAFT

Pearson, C., Huntington, J., Volk, J., & Dunkerly, C. (2020). *Agricultural evapotranspiration and net irrigation water requirements for the upper colorado river basin: 2019 ET demands model summary*.

Pearson, C., Huntington, J., Volk, J., & Dunkerly, C. (2021). *Agricultural evapotranspiration and net irrigation water requirements for the upper colorado river basin: 2020 ET demands model summary*.

Singh, A., Taghvaeian, S., Mirchi, A., & Moriasi, D. (2023). Station aridity in weather monitoring networks: Evidence from the oklahoma mesonet. *Applied Engineering in Agriculture*, *39*(2), 167–177.

Steduto, P., Faurès, J.-M., Hoogeveen, J., Winpenny, J., & Burke, J. (2012). Coping with water scarcity: An action framework for agriculture and food security. *Food and Agriculture Organization of the United Nations*.

Team, B. D. (2018). *Bokeh: Python library for interactive visualization*. https://bokeh.pydata.org/en/latest/

Thornton, P., & Running, S. (1999). An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology*, *93*(4), 211–228.

DRAFT