

flux-data-qaqc: A Python Package for Energy Balance Closure and Post-Processing of Eddy Flux Data

John Volk^{*1}, Justin Huntington¹, Richard Allen², Forrest Melton^{3, 4}, Martha Anderson⁵, and Ayse Kilic⁶

¹ Desert Research Institute ² University of Idaho ³ NASA Ames Research Center ⁴ California State University Monterey Bay ⁵ USDA Agricultural Research Service ⁶ University of Nebraska Lincoln

DOI: [10.21105/joss.03418](https://doi.org/10.21105/joss.03418)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Pierre de Buyl](#) ↗

Reviewers:

- [@ashwinvis](#)
- [@dgketchum](#)

Submitted: 06 May 2021

Published: 16 October 2021

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Introduction

Eddy Covariance (EC) systems provide continuous direct measurements of turbulent and radiative heat and energy fluxes between the atmosphere and land surface ([Baldochi et al., 1988](#)). Evapotranspiration (ET) estimates from EC systems are highly valuable for many scientific disciplines, water policy decisions, and operational applications such as irrigation management. Although the EC technique is considered accurate under certain atmospheric and spatial conditions, inherent theoretical and practical limitations of the technique are magnified by data uncertainty from post-processing methods. Error can result from the well-documented energy balance closure problem resulting from inherent limitations in the technique but also from data processing errors, instrumentation errors, and other sources ([Foken, 2008](#); [Stoy et al., 2013](#)). Several approaches for post-processing and correcting EC-based ET estimates exist, notably those implemented by major flux measurement networks and datasets like FLUXNET2015 ([Pastorello et al., 2020](#)). However, post-processing routines often differ between individual site teams, data networks, and end users, and in some cases post-processing decisions are difficult to reproduce.

Statement of need

The `flux-data-qaqc` open-source and object-oriented Python package is intended for a wide audience including anyone who needs to process EC flux data to estimate daily or monthly ET that has been corrected for energy balance closure. It is also useful for those who wish to perform visual or flag-based quality control and quality assurance (QA/QC) checks on EC flux or meteorological data to filter out or replace any suspect data points before applying the data for any given use. For example, atmospheric and hydrologic modeling and field scientists may find this software useful for simplifying the post-processing of 30-minute EC flux data into daily ET that has been gap-filled and corrected (or not) for energy imbalance using different approaches. The software will also make calculations of reference ET from EC meteorological data and download reference ET from gridded products ([Abatzoglou, 2013](#)). Output daily and monthly ET estimates from `flux-data-qaqc` can then be used by scientists for model validation and applied science, while knowing that the post-processing steps used to generate them are tested and well documented.

Post-processing steps of EC data commonly involve at least some of the following steps: data formatting, unit conversions, gap-filling, time aggregation, ET and other atmospheric calculations, energy imbalance assessment and correction, and data QA/QC (automatic or visual).

*corresponding author

Because `flux-data-qaqc` provides a framework for these common EC post-processing steps using Python objects, it also allows scientists or engineers to document and share their processing decisions and analyses with others via the Python scripts they write, which is critical in any serious study. Although the tools provided by `flux-data-qaqc` are well documented and “high-level,” minimal proficiency in Python is required.

The interactive time series and scatter plots (e.g. tools for hover, pan, zoom, selection, and linked axes) produced by `flux-data-qaqc` are useful for data QA/QC and for anyone interested in understanding or visually documenting the atmospheric and/or hydrologic conditions at an EC tower. Within the same plot document, daily and monthly surface energy balance variables (average fluxes) as well as all commonly measured meteorological variables (e.g. wind speed, temperature, precipitation, vapor pressure deficit, soil moisture) are displayed. In addition, the plots display computed results useful for EC data QA/QC and many ET-oriented analyses, these results include: energy balance closure (daily and monthly), the energy balance ratio initially and after despiking and smoothing, ET as a fraction of gridded reference ET, as well as daily ET gap information.

Ultimately, obtaining final ET estimates from EC data involves multiple human-driven processing decisions and this presents an uncertainty issue when comparing data that are processed differently. `flux-data-qaqc` does not address issues regarding the deployment configuration of the EC system or the initial processing of high frequency EC data, for which several software exist. Instead, this software provides a framework and provenance for EC data post-processing (e.g., gap-filling, temporal aggregation), energy balance closure analyses, and ET estimates (Wilson et al., 2002).

The [online documentation](#) for `flux-data-qaqc` provides more detail regarding these applications, including a comprehensive tutorial with usage examples. Also included are easy-to-follow installation instructions via the Python Package Index and a virtual environment for automatic third party software dependency management.

Design and features

The `flux-data-qaqc` package follows a simple hierarchical structure with a few low-level classes that are then inherited in a linear fashion by two classes that are more specific to EC post-processing (Figure 1). The software is also modular to facilitate incorporation of additional classes and methods. Adding new scientific routines to the package is also facilitated by its standardized and intuitive internal naming schemes and data formats shared by the core classes and their attributes.

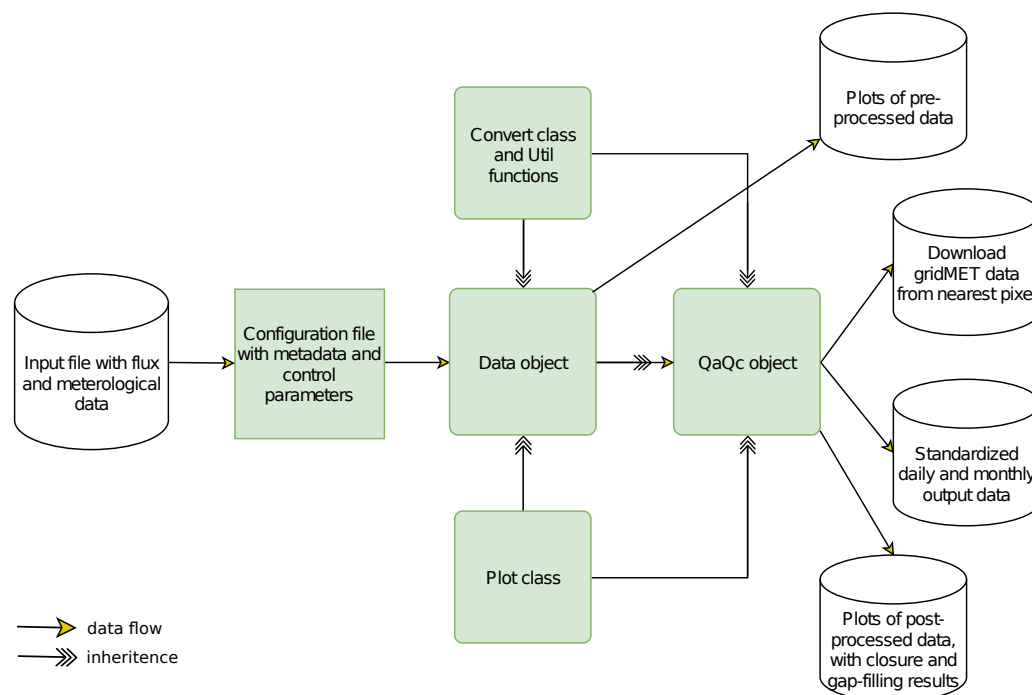


Figure 1: Diagram showing the relationship between software components, green highlighted boxes indicate software whereas non-highlighted indicate input or output files of the software.

Input to `flux-data-qaqc` is tabular timeseries EC/meteorological data and site metadata which are parsed using a configuration file that can be written by the user or using a tool provided with the package, and full [example files](#) are provided. The `Data` class parses the data and has tools for unit conversion, quality-based filtering, visualization, and some atmospheric calculations such as dew point temperature and potential solar radiation. From here, the `QaQc` class is used for gap-filling, daily and monthly aggregation, energy balance closure correction, writing data using a standard format, and creating interactive diagnostic plots. The `Data` and `QaQc` classes leverage the Pandas Python library, giving users access to pre- and post-processed data in a Pandas dataframe which facilitates a myriad of other Python-based scientific analyses via Numpy, SciPy and many other packages. The `Plot` class contains routines for creating standardized diagnostic plots of pre- and post-processed data, it also provides a robust API for generating interactive (via the [Bokeh package](#)) scatter and time series plots of arbitrary time series data. The `Plot` and `Convert` classes along with the `Util` module are used or inherited throughout the package; these low-level tools are modular enough to easily be used in other software or custom data pipelines.

Here are some capabilities and tools provided by `flux-data-qaqc`:

- ability to read a variety of tabular data formats using a configuration file
- tools for batch processing via writing configuration files programmatically
- automatic unit conversions
- filtering of poor quality data (flag-based and numerical)
- gap-filling of hourly or higher frequency EC data with day and night time treated separately
- temporal aggregation of high frequency data to daily and monthly
- interactive visual inspection tools at pre-processed data frequency and for daily and monthly post-processed data, for both EC and meteorological data
- management and archiving of pre- and post-processed data and metadata, including standardized naming

- multiple energy balance closure correction routines: an Energy Balance Ratio method, the Bowen Ratio method, and a multiple linear regression method
- atmospheric calculations, including the hourly and daily forms of the American Society of Civil Engineers standardized reference ET equation (Allen et al., 2005) using the RefET python package
- ability to download gridded weather data and reference ET that corresponds with the EC tower (Abatzoglou, 2013)
- daily gap-filling of ET using gridded reference ET and site-based fraction of reference ET

State of the field

`flux-data-qaqc` is most similar to the ONEFlux pipeline (Pastorello et al., 2020) in that it includes the same energy balance closure-based corrections with slight modifications. However, there are many differences and improvements for processing of micrometeorological measurements to calculate daily water and energy fluxes. For one, ONEFlux is a command line application as opposed to an object-oriented framework. ONEFlux has many other uses outside of the scope `flux-data-qaqc`, which is focused on water vapor flux. Similarly, several tools in `flux-data-qaqc` are not included in ONEFlux, for example: reading of data in different formats via a configuration file, multiple hourly gap-filling methods, a daily gap-filling method using gridded reference ET, visualization tools, atmospheric calculations such as standardized reference ET, monthly aggregation, and ability for the user to select from multiple alternative energy balance closure correction routines.

Other EC data processing and QA/QC software can be classified into one or more of these groups: 1) processing of high frequency EC data; 2) flux partitioning, e.g. partitioning water vapor flux into transpiration and direct evaporation components; 3) friction velocity (u^*) filtering of CO₂; 4) post-processed gap-filling and filtering; and 5) flux footprint generation. `flux-data-qaqc` is dedicated to tasks in group 4 and has methods and scope unique to other open-source post-processing software known to the authors. Besides ONEFlux, here is a list of all other known and similar EC post-processing software, their purposes and some major differences:

- [MDI Meteo gap-filling tool](#). Online application that performs gap-filling and uses a different methodology.
- [GaFir](#). R package that performs gap-filling and uses a different methodology.
- [ReddyProc](#) (Wutzler et al., 2018). R package that performs u^* filtering, gap-filling, and flux partitioning. Differences include: CO₂ focused, no closure corrections, different gap-filling methods for half-hourly/hourly, no daily gap-filling, no monthly aggregation or plots, strict input formatting, missing ancillary atmospheric calculations.
- [openeddy](#). R package that performs data handling, summarizing, and plotting. Differences include: no closure corrections, different QA/QC procedures, no gap-filling, strict input formatting, missing ancillary atmospheric calculations.
- [PyFluxPro](#). Python GUI application that performs standardized post-processing and QA/QC routines, gap-filling, u^* filtering, flux partitioning, and plotting. Differences include: different gap-filling methods, no closure corrections, no daily gap-filling, missing ancillary atmospheric calculations.
- [hesseflux](#) (Cuntz, 2020). Python package similar to ReddyProc with similar differences, namely lack of energy balance closure corrections.

Research enabled by flux-data-qaqc

This package was designed for and is being used to generate a benchmark ET dataset for ground validation and intercomparison with remote sensing ET models that are part of OpenET (Melton et al., 2021). OpenET is a large collaborative effort to provide satellite-based ET estimates at the field scale across the western United States, which will greatly improve water management practices among other uses. This package is also being used for processing ET from GRAPEX vineyard EC sites to evaluate spectral-based ET approaches (Kustas et al., 2018).

Acknowledgements

We gratefully acknowledge support for this work from the Walton Family Foundation; Lyda Hill Philanthropies; the S.D. Bechtel, Jr. Foundation; the Gordon and Betty Moore Foundation; the Windward Fund; the Water Funder Initiative; the NASA Applied Science Program and the NASA Western Water Applications Office; the USGS Landsat Science Team; the California State University Agricultural Research Institute; and the Idaho Agricultural Experiment Station and Nebraska Agricultural Experiment Station.

References

- Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1), 121–131. <https://doi.org/10.1002/joc.3413>
- Allen, R. G., Walter, I. A., Elliott, R. L., Howell, T. A., Itenfisu, D., Jensen, M. E., & Snyder, R. L. (2005). *The ASCE standardized reference evapotranspiration equation*. American Society of Civil Engineers. <https://doi.org/10.1061/9780784408056>
- Baldocchi, D. D., Hincks, B. B., & Meyers, T. P. (1988). Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology*, 69(5), 1331–1340. <https://doi.org/10.2307/1941631>
- Cuntz, M. (2020). *hesseflux: a Python library to process and post-process Eddy covariance data* (Version v2.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.3831489>
- Foken, T. (2008). The energy balance closure problem: An overview. *Ecological Applications*, 18(6), 1351–1367. <https://doi.org/10.1890/06-0922.1>
- Kustas, W. P., Anderson, M. C., Alfieri, J. G., Knipper, K., Torres-Rua, A., Parry, C. K., Nieto, H., Agam, N., White, W. A., Gao, F., McKee, L., Prueger, J. H., Hipps, L. E., Los, S., Alsina, M. M., Sanchez, L., Sams, B., Dokoozlian, N., McKee, M., ... Hain, C. (2018). The grape remote sensing atmospheric profile and evapotranspiration experiment. *Bulletin of the American Meteorological Society*, 99(9), 1791–1812. <https://doi.org/10.1175/BAMS-D-16-0244.1>
- Melton, F., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Senay, G., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., ... Anderson, R. (2021). OpenET: Filling a critical data gap in water management for the western united states. *Journal of the American Water Resources Association*, In review.
- Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca,

- A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., ... Papale, D. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, 7(1), 225. <https://doi.org/10.1038/s41597-020-0534-3>
- Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M. A., Arneth, A., Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Gianelle, D., van Gorsel, E., Kiely, G., Knohl, A., Margolis, H., McCaughey, H., ... Varlagin, A. (2013). A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. *Agricultural and Forest Meteorology*, 171-172, 137–152. <https://doi.org/10.1016/j.agrformet.2012.11.004>
- Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B. E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., ... Verma, S. (2002). Energy balance closure at FLUXNET sites. *Agricultural and Forest Meteorology*, 113(1), 223–243. [https://doi.org/10.1016/S0168-1923\(02\)00109-0](https://doi.org/10.1016/S0168-1923(02)00109-0)
- Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., & Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, 15(16), 5015–5030. <https://doi.org/10.5194/bg-15-5015-2018>