Technical Note: AmeriFlux BASE Flux/Met Data QA/QC

Authors

- Documentation
  - Housen Chu - LBL
  - Danielle S. Christianson - LBL
  - You-Wei Cheah - LBL
  - Rachel Hollowgrass - University of California Berkeley, Berkeley
  - Brain Wang - LBL
- Code Development
  - Danielle S. Christianson - LBL
  - You-Wei Cheah - LBL
  - Housen Chu - LBL
  - Gilberto Pastorello - LBL
  - Joshua Geden - LBL
  - Fianna O'Brien - LBL
  - Sy-Toan Ngo - LBL
  - Norman F. Beekwilder - University of Virginia
  - Alessio Ribeca - Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici
  - Carlo Trotta - Division Impacts on Agriculture, Forests and Ecosystem Services (IAFES), Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici
- Code Evaluation
  - Stephen W. Chan - LBL
  - Sigrid Dengel - LBL
  - Dario Papale - University of Tuscia
  - Sébastien C. Biraud - LBL
  - Deborah A. Agarwal - LBL

AmeriFlux MANAGEMENT PROJECT

Table of Contents

Authors
Table of Contents
1. Overview
    1.1. Data QA/QC Workflow
    1.2. Data QA/QC Report
    1.3. Summary Statistics Table
2. Timestamp Alignment Module
    2.1 Module Info
    2.2 Figure Explanation
    2.3 Summary Statistics Table
3. Physical Range Module
    3.1 Module Info
    3.2 Figure Explanation
    3.3 Summary Statistics Table
4. Multivariate Comparison Module
    4.1 Module Info
    4.2 Figure Explanation
    4.3 Summary Statistics Table
5. Diurnal-Seasonal Pattern Module
    5.1 Module Info
    5.2 Figure Explanation
    5.3 Summary Statistics Table
6. USTAR Filtering Module
    6.1 Module Info
    6.2 Figure Explanation
    6.3 Summary Statistics Table
7. Variable Coverage Module
    7.1 Module Info
    7.2 Figure Explanation
8. Appendix
9. Reference

1. Overview

The AmeriFlux Data QA/QC process assesses the quality of flux and meteorological data uploaded to AmeriFlux before publication as an AmeriFlux BASE data product (Chu et al., 2023). It is a secondary data quality assessment that is independent of and complementary to the data quality checks performed by site teams. The AmeriFlux Data QA/QC follows a similar methodology for quality-checking and processing as the FLUXNET2015 dataset (Pastorello et al. 2014, 2020) but includes additional checks based on data user feedback. Additionally, its design takes into account the extensive history of AmeriFlux data repositories and the diverse ecosystems and climates of AmeriFlux sites. Last, the AmeriFlux Data QA/QC uses data visualization and a ticket-tracking system to facilitate communication with site teams. Since the deployment in May 2017, the Data QA/QC process has assessed ~2,900 data submissions till the end of 2024.

1.1. Data QA/QC Workflow

Data QA/QC occurs after the uploaded files pass Format QA/QC. First, the uploaded files are combined with, if any, previously published BASE files (Figure 1). Data QA/QC modules are executed to generate statistics and figures for further review. Six Data QA/QC check modules are implemented currently, including timestamp alignment, physical range, multivariate comparison, diurnal-seasonal pattern, USTAR filtering, and data coverage. Details of each module are explained in Sections 2-7.

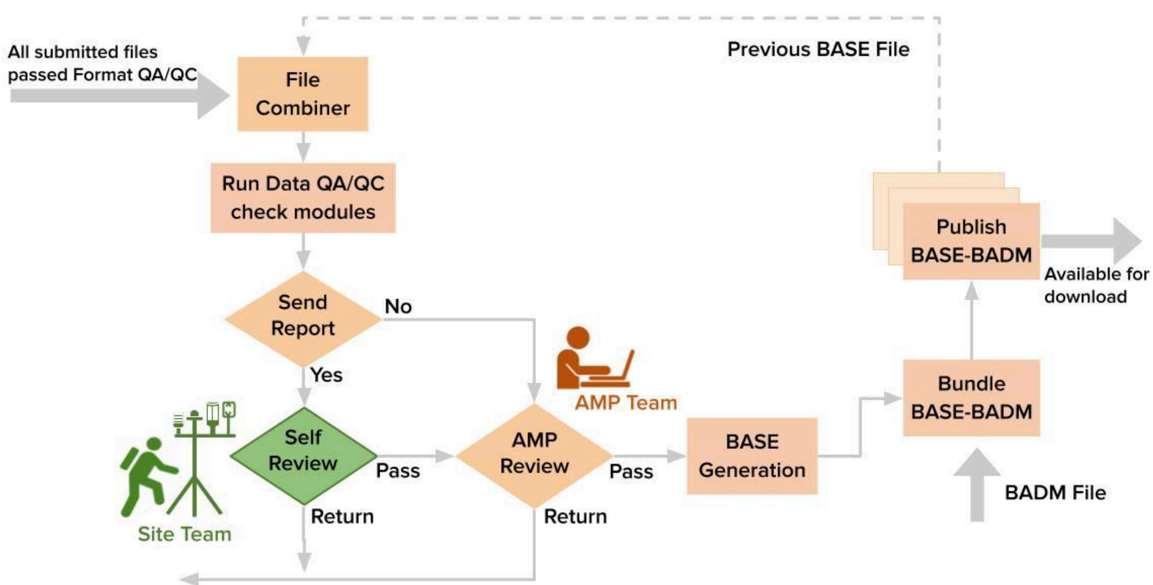


Figure 1. *Data QA/QC workflow begins with submitted files that have passed Format QA/QC and ends with the publication of the BASE-BADM data product. Orange and green colors indicate actions by the AmeriFlux Management Project (AMP) and site teams, respectively. BASE and BADM refer to the continuous flux/meteorological data and the Biological, Ancillary, Disturbance, and Metadata, respectively.*

After generating figures and statistics, the AmeriFlux Management Project (AMP) team reviews the results. The review process is usually done in batches every 1-2 months. If any issues are identified, the AMP team notifies the site team of the necessary corrections. Otherwise, the data are queued for BASE generation and bundled with Biological, Ancillary, Disturbance, and Metadata (BADM) for publication as the BASE-BADM data product. The Data QA/QC step does not include any data filtering and correction. Identified issues must be addressed through resubmission by the site team.

For sites participating in the self-review process, a preliminary report is sent after the statistics and figures are generated (Figure 1). This allows the site teams to review and correct potential issues promptly. If the site team identifies any issues, they can resubmit an updated version. Otherwise, the data are passed to the AMP team for further review.

1.2. Data QA/QC Report

In a typical Data QA/QC review, the AMP team synthesizes the identified issues into a concise and actionable report (Figure 2). The report provides an overview of Data QA/QC and summarizes the identified issues, accompanied by links to explanatory figures. The report also contains all statistics, figures, and the Format QA/QC report associated with the data. The AMP team emails the Data QA/QC report to the site team for clarification or correction.

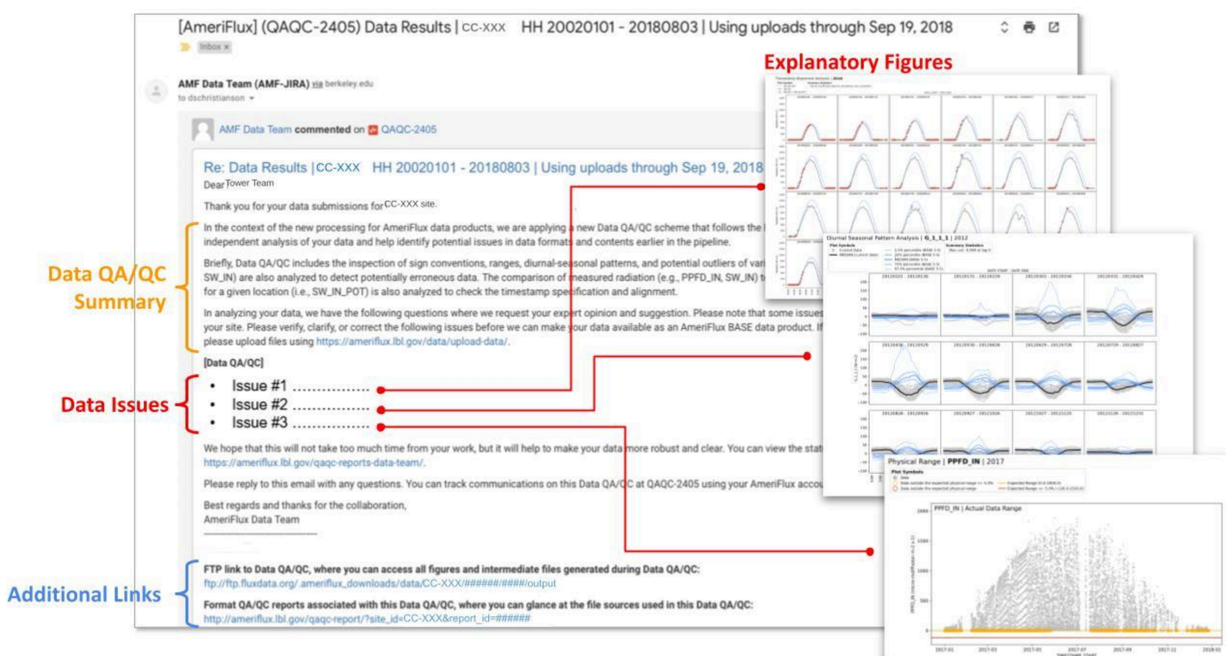


Figure 2. *An example of a typical Data QA/QC report email. The colored text highlights the components of the email.*

For sites participating in the self-review process, the site teams receive a slightly different report email (Figure 3). In particular, the report contains summary statistics generated by the QA/QC modules instead of synthesized issues. The summary statistics provide a quick overview, allowing the site teams to identify and correct any issues promptly. The report also contains the link to this technical note, which helps the site team understand and interpret the QA/QC results.

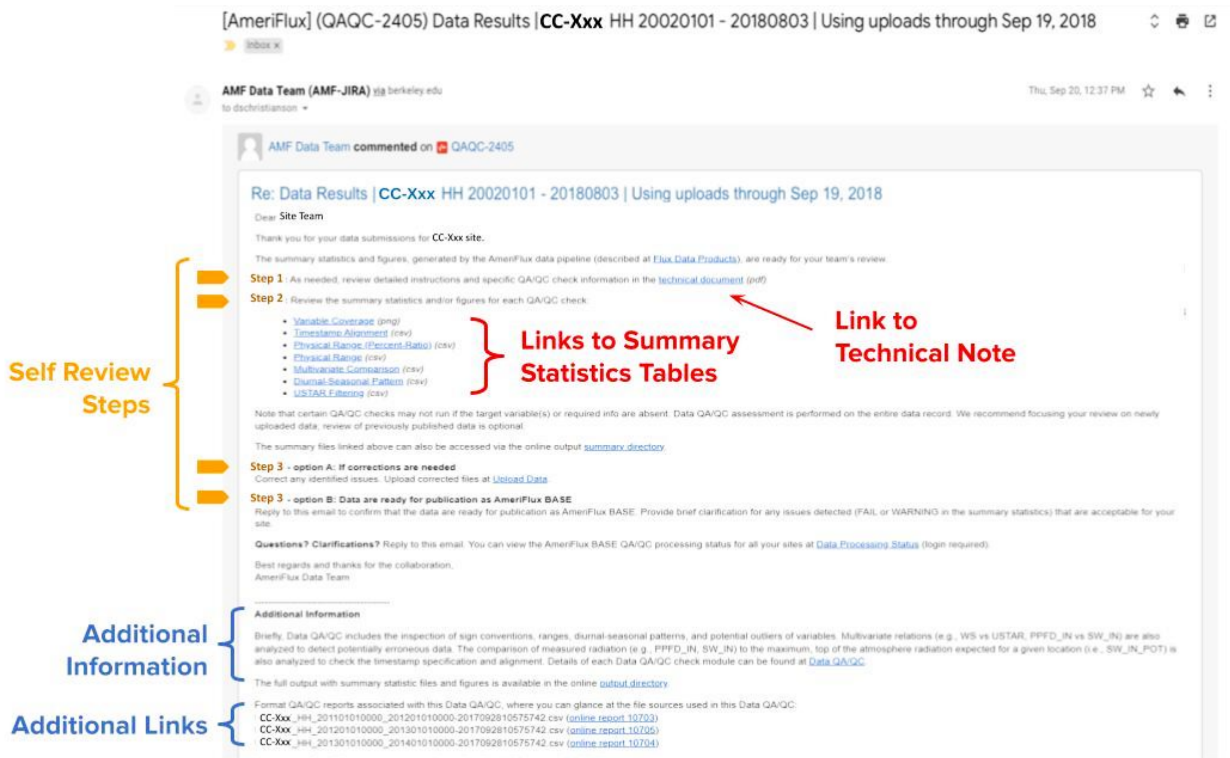


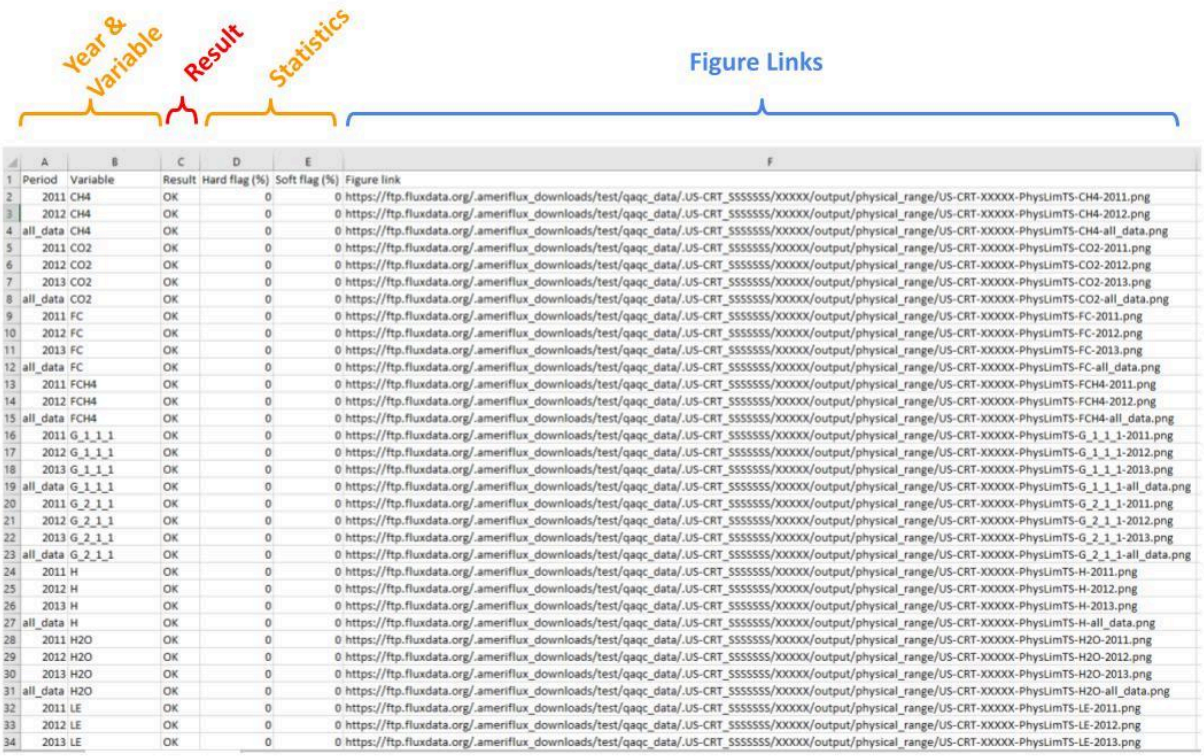
Figure 3. *An example of the Data QA/QC report email sent to sites participating in the self-review process. The colored text highlights the components of the email.*

1.3. Summary Statistics Table

While varying slightly among modules, summary statistics tables consist of columns of variables, periods, results, statistics, and figure links (Figure 4). The result column displays the high-level check results (i.e., OK, WARNING, FAIL) for a variable within a specific period (e.g., a year, all records). See below for general guidance on interpreting the results:

- **OK**: No issue is identified by the module. On some occasions, the module might not be able to generate all statistics (i.e., insufficient data). The determination is based on the partially available information.
- **WARNING**: A potential issue is identified for the variable(s) in a specific period, with a moderate likelihood. We suggest reviewing the associate figure and statistics to determine whether it is an issue or requires any action.
- **FAIL**: An issue is identified for the variable(s) in a specific period, with a high likelihood. We suggest reviewing the associate figure and statistics to determine whether it is an issue or requires any action.

The aforementioned high-level results are determined based on the statistics (e.g., outlier percentage, regression) provided in the statistics columns. Their calculations and rules are described in each module below. The summary statistics table also contains the figure link columns, providing links to corresponding figures on FTP. Details of the summary statistics table for each module are explained in Sections 2-7 (See Summary Statistics Table section in each module).



Period	Variable	Result	Hard flag (%)	Soft flag (%)	Figure link
2011	CH4	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-CH4-2011.png
2012	CH4	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-CH4-2012.png
all_data	CH4	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-CH4-all_data.png
2011	CO2	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-CO2-2011.png
2012	CO2	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-CO2-2012.png
2013	CO2	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-CO2-2013.png
all_data	CO2	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-CO2-all_data.png
2011	FC	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-FC-2011.png
2012	FC	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-FC-2012.png
2013	FC	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-FC-2013.png
all_data	FC	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-FC-all_data.png
2011	FCH4	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-FCH4-2011.png
2012	FCH4	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-FCH4-2012.png
all_data	FCH4	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-FCH4-all_data.png
2011	G_1_1_1	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-G_1_1_1-2011.png
2012	G_1_1_1	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-G_1_1_1-2012.png
2013	G_1_1_1	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-G_1_1_1-2013.png
all_data	G_1_1_1	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-G_1_1_1-all_data.png
2011	G_2_1_1	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-G_2_1_1-2011.png
2012	G_2_1_1	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-G_2_1_1-2012.png
2013	G_2_1_1	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-G_2_1_1-2013.png
all_data	G_2_1_1	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-G_2_1_1-all_data.png
2011	H	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-H-2011.png
2012	H	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-H-2012.png
2013	H	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-H-2013.png
all_data	H	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-H-all_data.png
2011	H2O	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-H2O-2011.png
2012	H2O	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-H2O-2012.png
2013	H2O	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-H2O-2013.png
all_data	H2O	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-H2O-all_data.png
2011	LE	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-LE-2011.png
2012	LE	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-LE-2012.png
2013	LE	OK	0	0	https://ftp.fluxdata.org/.ameriflux_downloads/test/qaqc_data/.US-CRT_SSSSSSS/XXXXX/output/physical_range/US-CRT-XXXXX-PhysLimTS-LE-2013.png

Figure 4. *An example of the summary statistics table (in a comma-separated-values(CSV) format) consisting of variables, periods, results, statistics, and figure links.*

2. Timestamp Alignment Module

2.1 Module Info

- Target Variable: SW_IN, PPFD_IN
- Execution Period: Annual
- Requirement: Top-level SW_IN or PPFD_IN, and a site's latitude, longitude, and UTC offset

The Timestamp Alignment Module examines the alignment between the measured incoming radiation, e.g., photosynthetically active radiation (PPFD_IN), shortwave radiation (SW_IN), and the calculated potential incoming radiation at the top of the atmosphere (SW_IN_POT). The module aims to identify the following issues:

- Wrong timestamp specification

  - Misspecified beginning or ending timestamps
  - Timestamps do not match the time zone specification
  - Use of daylight saving time
  - Data streams not synchronized (clock drift, skipped time steps)
- Radiation measurement issue
  - Sensor not leveled
  - Shaded radiation measurements
  - Higher than expected radiation readings

2.2 Figure Explanation

For each site-year, (half-)hourly SW_IN_POT is calculated based on a site's geolocation and time zone provided in the site's general information. Then, the SW_IN_POT, SW_IN, and PPFD_IN data are aggregated into a "maximum diurnal composite" for each of the 15-day non-overlapping windows (Figure 5). PPFD_IN data are converted into an energy unit (W m^{-2}) by an approximate coefficient of 0.5 J μmol^{-1}. The calculation of the maximum diurnal composite eliminates periods with cloudy conditions, allowing for alignment analysis under mostly clear-sky conditions. The module expects that the diurnal composites align between SW_IN_POT, SW_IN, and PPFD_IN indicated by the cross-correlation, and that SW_IN and PPFD_IN do not exceed SW_IN_POT in the morning or afternoon hours (e.g., Figure 5). Misalignment between measured and calculated radiation indicates possible issues of timestamp alignment or radiation measurements (e.g., Figure 6).

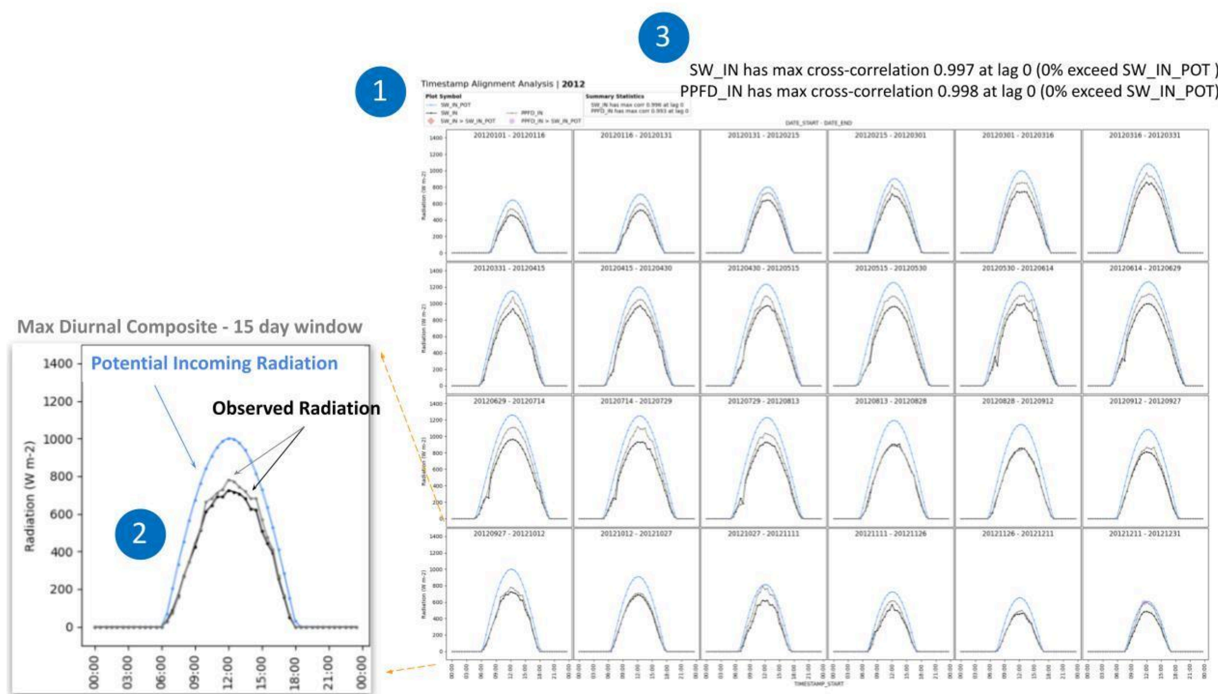


Figure 5. *Example figure with <u>no issues</u> identified by the Timestamp Alignment Module. The "maximum diurnal composite" is calculated for each 15-day non-overlapping window within a year (1). The module expects that the*

diurnal patterns align in time between SW_IN_POT and SW_IN as indicated by cross-correlation (3) and that SW_IN does not exceed SW_IN_POT in most periods (2).

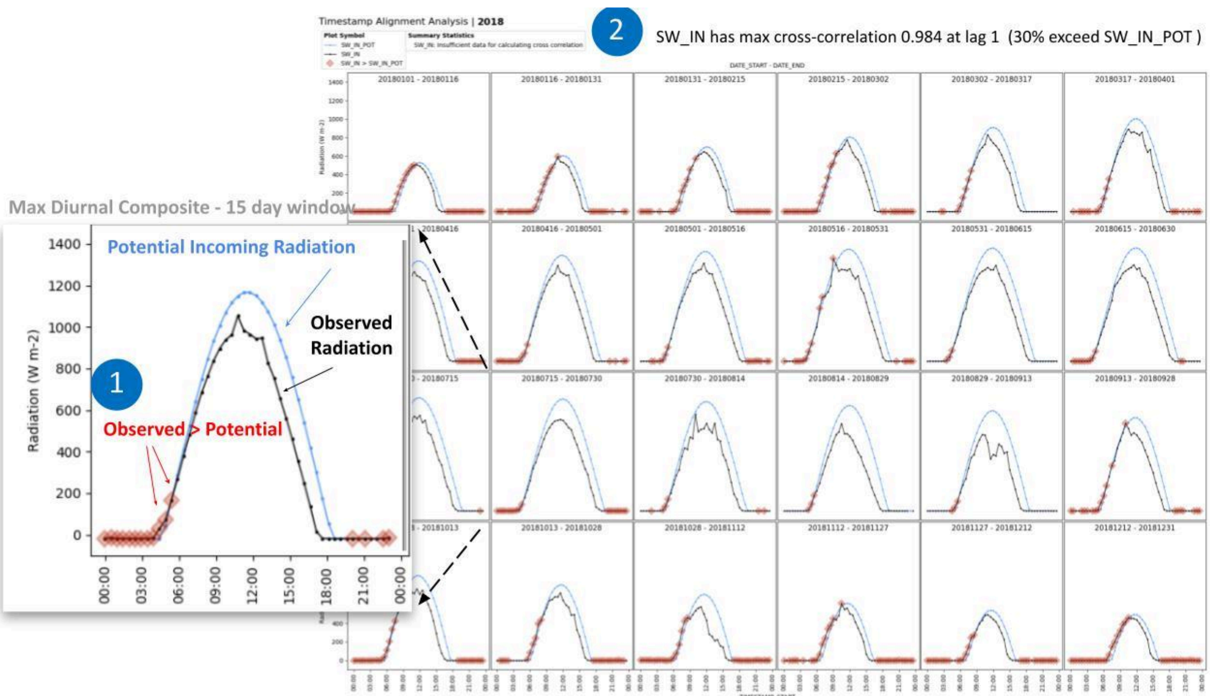


Figure 6. *Example figure with the* *<u>detected issues</u>* *by the Timestamp Alignment Module. This example shows (1) many occasions that observed radiation (SW_IN) exceeds the potential incoming radiation (SW_IN_POT), highlighted by red squares. And (2) the maximum cross-correlation between calculated and observed radiation occurs at a 1-step lag, i.e., a half-hour shift for half-hourly data.*

The module calculates two groups of statistics, i.e., cross-correlations between SW_IN_POT and SW_IN (or PPFD_IN) composite and percentages that SW_IN (or PPFD_IN) exceeds SW_IN_POT (Table 1). The time lag at which the maximum cross-correlation occurs suggests the alignment between the two time series. For example, a time lag of two steps (i.e., one hour for half-hourly data) means a one-hour shift in the timestamps (Figure 6).

As the cross-correlations are calculated only when both time-series composites have no gaps, the module computes the percentage of time steps that SW_IN (or PPFD_IN) exceeds SW_IN_POT to help detect timestamp alignment issues. For example, excessive radiation consistently across all windows in the early morning suggests a potential shift in the timestamps (Figure 5). The check results (i.e., OK, WARNING, FAIL) are then determined based on a combination of both groups (Result column in Table 1). A summary table of commonly seen issues is provided to aid the interpretation and correction of identified issues (Table 2).

2.3 Summary Statistics Table

Table 1. Summary statistics table for the Timestamp Alignment Module

Period	Variable	Result	Time lag	Cross correlation	Excessive radiation daytime (%)	Excessive radiation nighttime (%)	Figure
YEAR[1]	VAR[1]	RESULT[2]	t_{max}[3]	$max(abs(R_{xy}))$[3]	P_{day}[4]	P_{night}[4]	LINK

Abbreviation: t_{max}: the lag at which the maximum cross-correlation is found, $max(abs(R_{xy}))$: max cross-correlation between the time series X and Y, P_{day}: percentage of flagged points in the daytime (i.e., SW_IN or PPFD_IN > SW_IN_POT), P_{night}: percentage of flagged points in the nighttime (i.e., SW_IN or PPFD_IN > 10 W m^{-2})

[1] This check is performed on an annual scale, requires top-level SW_IN or PPFD_IN present, and a site's latitude, longitude, and UTC offset

[2] Criteria:

- **FAIL** if either below
    - $max(abs(R_{xy})) > 0.4$ AND $abs(t_{max}) > 0$ AND ($0 < P_{day} < 4.8\%$ OR $0 < P_{night} < 4.8\%$)
    - $P_{day} > 4.8\%$ OR $P_{night} > 4.8\%$; A different threshold of 11.1% for P_{day} and P_{night} is used if the data are provided in an hourly resolution (HR)
- **WARNING** if either below
    - $max(abs(R_{xy})) > 0.4$ AND $abs(t_{max}) > 0$
    - $0 < P_{day} < 4.8\%$ OR $0 < P_{night} < 4.8\%$; A different threshold of 11.1% for P_{day} and P_{night} is used if the data are provided in an hourly resolution (HR)
- **OK** if none above

[3] Cross-correlation and time lag are calculated only when both composite time series have no gaps.

[4] The time steps at which the sunrise and sunset occur and the adjacent time steps are excluded when counting the percentages of excessive radiation points.

Table 2. Summary of issues and criteria for issue detection for the Timestamp Alignment Module

Issue types	Issues	Criteria
Timestamp specification	Misspecified beginning or ending timestamps	• $max(abs(R_{xy})) > 0.4$ & $abs(t_{max}) = 1$ • $P_{day} > 0$ or $P_{night} > 0$
	Timestamps not matched with time zone	• HH: $max(abs(R_{xy})) > 0.4$ & $abs(t_{max}) = 2n$ (n: natural number) • HR: $max(abs(R_{xy})) > 0.4$ & $abs(t_{max}) = n$ (n: natural number) • $P_{day} > 0$ or $P_{night} > 0$
	Use of daylight saving time	• HH: $max(abs(R_{xy})) > 0.4$ & $abs(t_{max}) = 2$ • HR: $max(abs(R_{xy})) > 0.4$ & $abs(t_{max}) = 1$ • $P_{day} > 0$ or $P_{night} > 0$, particularly in April-October windows for north-hemisphere sites; in September-April windows for south-hemisphere sites
	Datastream not synchronized	• $max(abs(R_{xy})) > 0.4$ & $abs(t_{max})_{PPFD\_IN} \neq abs(t_{max})_{SW\_IN}$
Radiation measurement issue	Radiation sensor not leveled	• $max(abs(R_{xy})) > 0.4$ & $abs(t_{max}) \neq 0$ • $P_{day} = 0$ & $P_{night} = 0$ OR asymmetric shape in diurnal composites

	Higher than expected radiation readings	• $max(abs(R_{xy})) > 0.4$ & $abs(t_{max}) = 0$ • $P_{day} > 0$ or $P_{night} > 0$

3. Physical Range Module

3.1 Module Info

❖ Target Variable: All variables (Variables with a % unit for Percent-Ratio Sub-Module)
❖ Execution Period: Annual or entire record
❖ Requirement: Either the upper or lower bound of a variable's physical range is defined

The Physical Range Module examines the full range of the target variable. The module assesses all variables. The module can be used to identify the following issues:

- Plausibility check
  - Outlier (i.e., out-of-range) points
  - Percent-ratio check (i.e., percentages provided as ratios)
- Variability check
  - Trend
  - Step change
  - Repeating patterns or filled constants
  - Measurement or processing cut-off
  - Other unrecognized patterns

3.2 Figure Explanation

For each variable, the accepted range is defined based on its physically plausible range (Table A1), for example, 0-100 for percentage variables, and the distribution of the published data across the AmeriFlux sites. A ±5% buffer is applied to account for possible edge values near the lower and upper bounds, commonly observed for radiation variables, relative humidity, and snow depth. A data point is soft-flagged outside the expected physical range but within the buffer range (±5% of the physical range) and hard-flagged outside the buffer range. The percentage of flagged points each year and in the entire record is used to determine if a variable has excessive out-of-range data points (Table 3, Figures 7-8). The module also assesses if a percentage variable (e.g., RH, SWC, 0-100) is provided in ratios (i.e., 0-1) (Table 4).

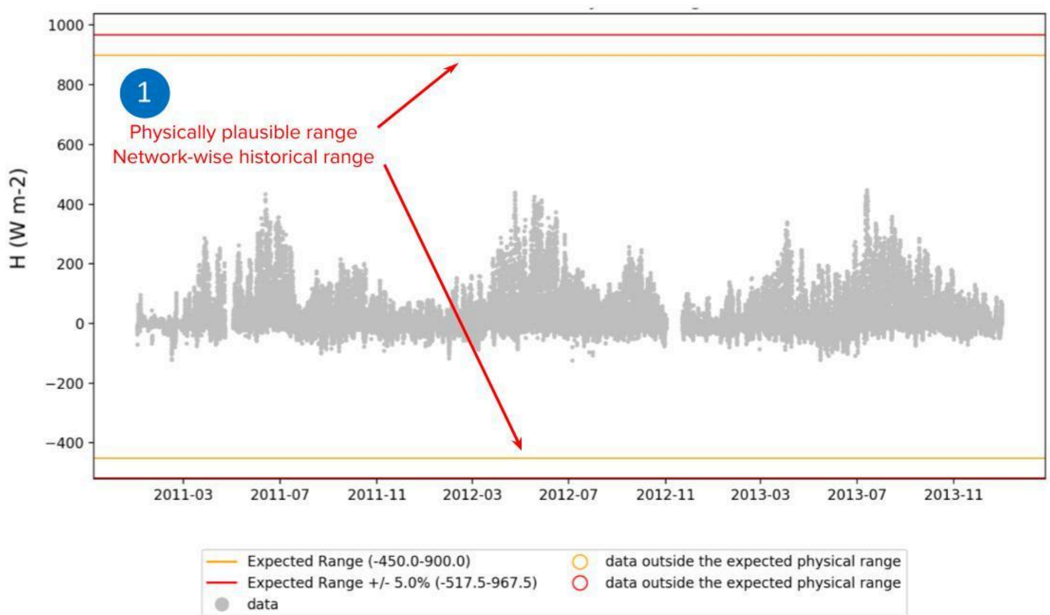


Figure 7. *Example figure with <u>no issues</u> detected by the Physical Range Module. (1) The yellow lines indicate the accepted range defined based on the physically plausible range and the network-wide historical range. The red lines indicate the accepted range plus a ±5% buffer range. In this case, no data points were beyond the plausible ranges.*

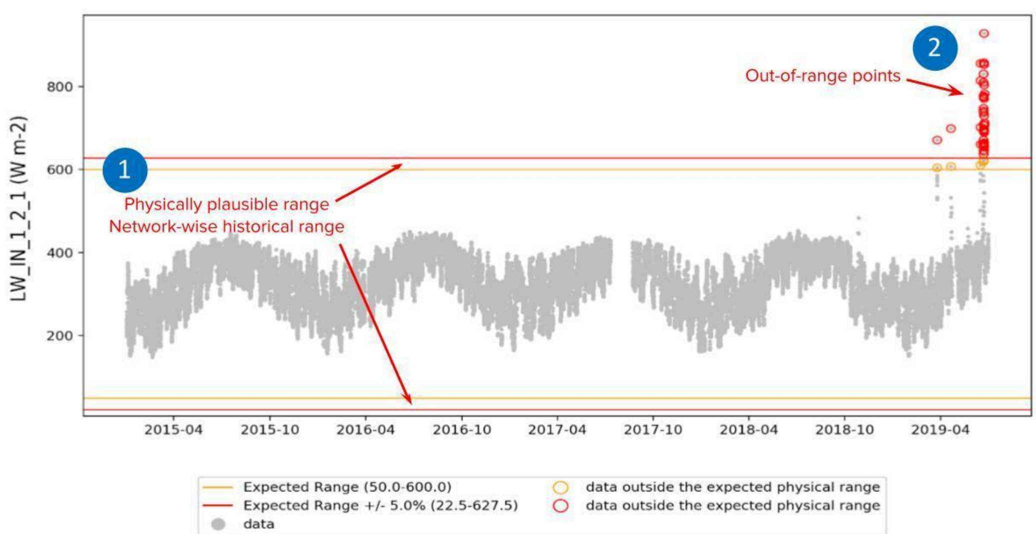


Figure 8. *Example figure with <u>identified issues</u> by the Physical Range Module. (1) The yellow lines indicate the accepted range defined based on the physically plausible range and the network-wide historical range. The red lines indicate the accepted range plus a ±5% buffer range. (2) The out-of-range data points (highlighted by red or yellow circles) are detected based on the accepted range. This case has ~0.2% of data points beyond the accepted range plus a ±5% buffer.*

While the module is mainly designed to detect out-of-range data points, additional issues, like trends and step changes, may be identified through manual inspection of the multi-year figures (Figure 9). For sites with previously published BASE datasets, the Diurnal-Seasonal Pattern Module may provide additional quantitative information on the potential changes of full ranges (e.g., trends, step change, cut-off) (see section 5). In addition, the Multivariate Comparison Module also assesses the potential changes of full ranges over years if a pair of associated variables (e.g., SW_IN, PPFD_IN) are both measured (see section 4).

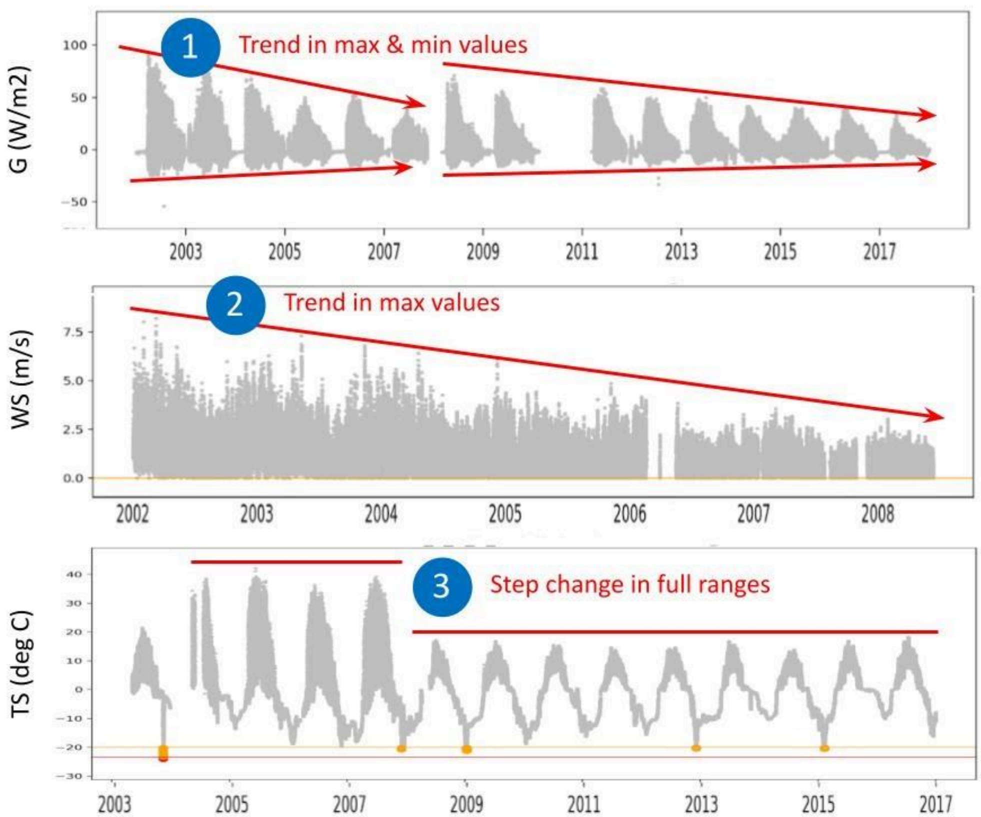


Figure 9. *Example figure with detected issues via manual inspection: (1) trends in both the maximum and minimum values; (2) trends in the maximum values; and (3) a step-change in the full ranges.*

3.3 Summary Statistics Table

Table 3. Summary statistics table for the Physical Range Module

Period	Variable	Result	Hard flag (%)	Soft flag (%)	Figure
YEAR[1]	VAR	RESULT[2]	P_{hard_flag}[3]	P_{soft_flag}[3]	LINK
ALL[1]	VAR	RESULT[2]	P_{hard_flag}[3]	P_{soft_flag}[3]	LINK

Abbreviation: P_{soft_flag}: percentage of data points outside the expected physical range but within the buffer range (±5% of physical range), P_{hard_flag}: percentage of data points outside the expected physical range plus the buffer range.

[1] This check is performed on an annual scale and on the entire record.

[2] Criteria:

- **FAIL** if
  - $P_{hard\_flag} > 0.1\%$
- **WARNING** if both
  - $P_{soft\_flag} > 1\%$
  - VAR is not one of D_SNOW, PPFD_IN, PPFD_OUT, PPFD_BC_IN, PPFD_BC_OUT, PPFD_DIF, PPFD_DIR, PPFD_UW_IN, SW_IN, SW_OUT, SW_BC_IN, SW_BC_OUT, SW_DIF, SW_DIR, and variables with unit %. These variables are known for commonly observed values near or slightly beyond the lower and upper bounds, so the WARNING for these variables is turned off.
- **OK** if none above

[3] Both hard- and soft-flagged percentages are calculated when a variable's upper and lower physical ranges are defined.

Table 4. Summary statistics table for the Percent-Ratio Sub-Module.

Period	Variable	Result	Is percent	Figure link
YEAR[1]	VAR	RESULT[2]	TRUE/FALSE	LINK

Abbreviation: Is percent: TRUE if all data points are within the range of 0-1, plus a ±5% buffer range.

[1] This check is performed on an annual basis and applies to variables with percentage units.

[2] Criteria:

- **FAIL** if return FALSE
- **OK** if return TRUE

4. Multivariate Comparison Module

4.1 Module Info

- Target Variable: SW_IN-PPFD_IN, TA-T_SONIC, WS-USTAR, and TA profile
- Execution Period: Annual and entire record
- Requirement: Both variables of a target variable pair present

Multivariate Comparison Module examines the relationship between a pair of associated variables that measure different but physically related quantities, e.g., SW_IN vs. PPFD_IN, USTAR vs. WS, TA vs. T_SONIC. In addition, the module also compares variables that measure the same quantity at different locations or using different sensors, e.g., vertical air temperature (TA) profiles and replicates (not yet implemented). The module assumes a consistent or predictable relationship between associated variables over time and uses that to identify the following potential issues:

- Short-term mismatch
  - Outlier (sporadically flagged points)
  - Short-term mismatch (flagged points for a specific period)

- Shaded radiation (periodically flagged)
- Unexpected relationship
    - Variables not synchronized in time (excessive scattering)
    - Derived one from another (perfectly fit)
- Change of slope
    - Trend (systematic change in the regression slope)
    - Step change in full range (change in the regression slope)

4.2 Figure Explanation

The module first fits a linear regression model (Model II) between the two targeted variables for each year or using the entire data record (Figures 10-12, Table 5). The module then calculates the orthogonal distance of each point to the regression line. Data points with relatively large deviations from the regression line are flagged as possible outliers. The percentages of flagged points in each year or the entire record are used to determine if a variable has excessive out-of-range data points for that period. A summary table of commonly seen issues is provided to aid the interpretation and correction of identified issues (Table 6).

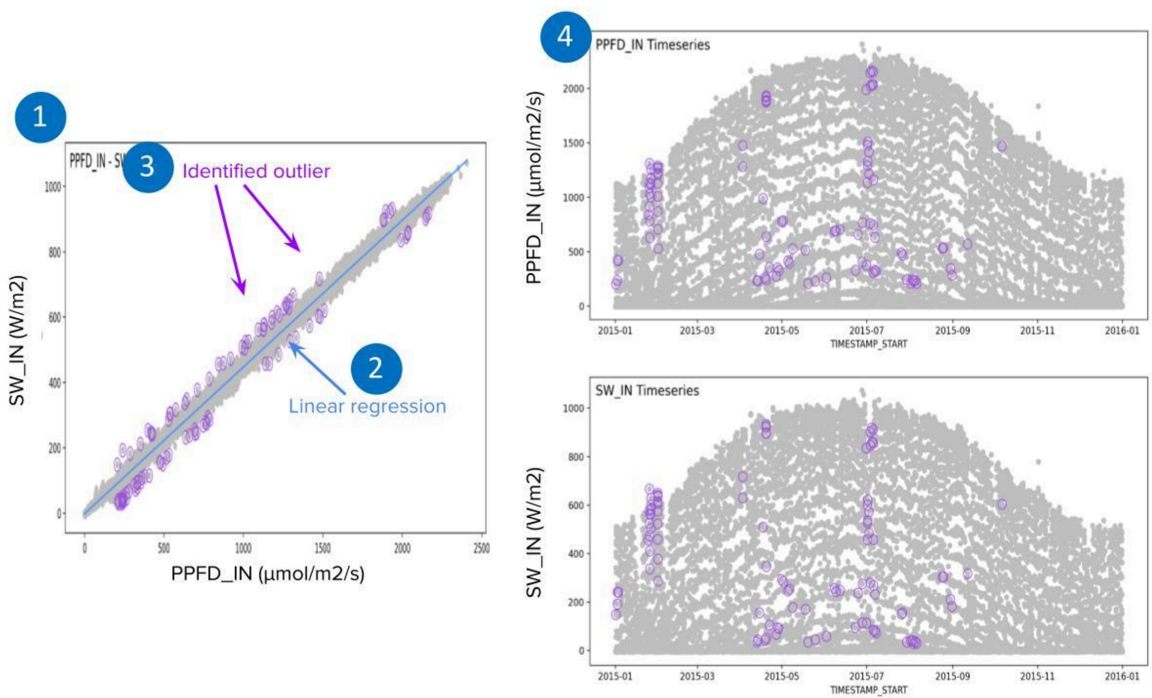


Figure 10. *Example figure with <u>no issues</u> identified by the Multivariate Comparison Module. The right panel (4) shows a one-year time series of PPFD_IN and SW_IN. The left panel (1) shows the scatter plot of SW_IN and PPFD_IN from the same one-year period. The blue line represents the orthogonal linear regression generated from all data (2), while the purple, highlighted circles denote data points flagged as potential outliers based on their orthogonal distance from the regression line (3). It is common for a few data points to be flagged as potential*

outliers, considering the random measurement errors and stochastic nature. A WARNING is triggered when > 1% of data points are flagged.

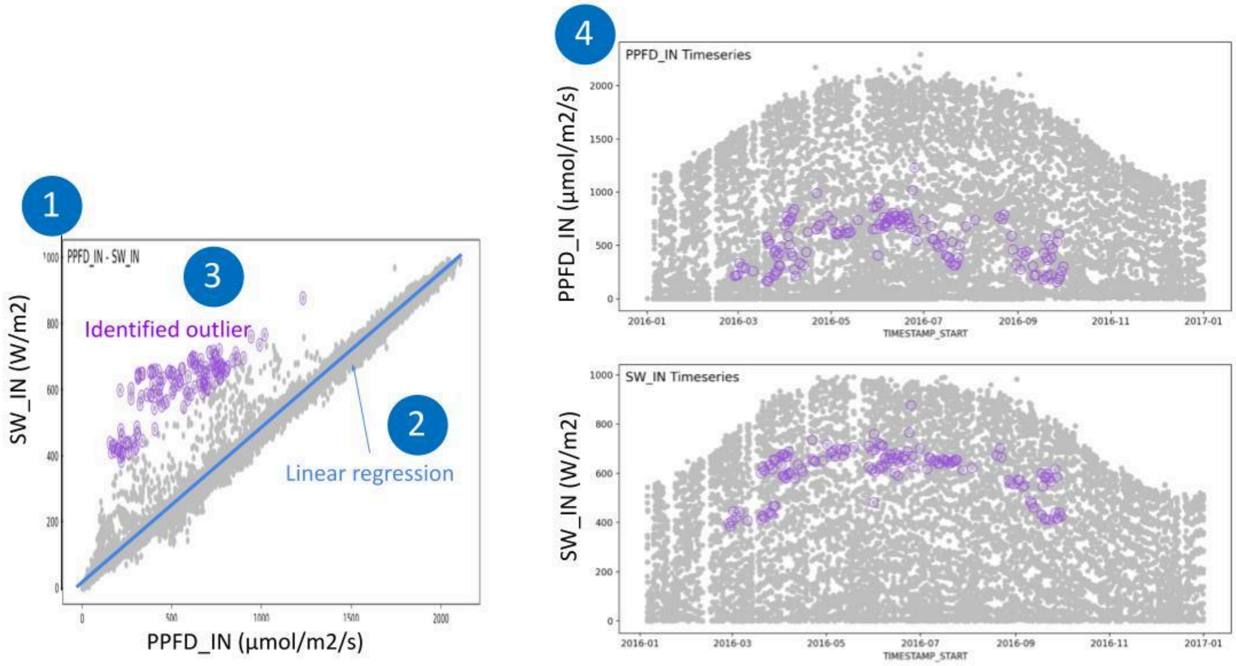


Figure 11. *Example figure with* <u>*detected issues*</u> *by the Multivariate Comparison Module. The right panel (4) shows a one-year time series of PPFD_IN and SW_IN. The left panel (1) shows the scatter plot of SW_IN and PPFD_IN from the same one-year period. The blue line denotes the orthogonal linear regression generated from all data (2), while the purple, highlighted circles denote data points flagged as outliers based on the orthogonal distance from the regression line (3). This case has a slightly higher percentage (> 1%) of flagged data points. And the periodic occurrence of flagged outliers suggests one of the radiation sensors, PPFD_IN in this case, is shaded periodically when the other sensor is not.*

For variables that are provided for more than one year, the module also examines the year-to-year changes in the annual regression slopes. We anticipate the slopes to be relatively stable over the years (Figure 12). The module calculates the deviation of an annual slope from the overall mean and returns a WARNING or FAIL if the relative deviation is higher than 10% and 20%, respectively. Potentially, the change in regression slopes over the years could indicate a trend or a step change in the full range of a variable (Figure 13, Table 5).

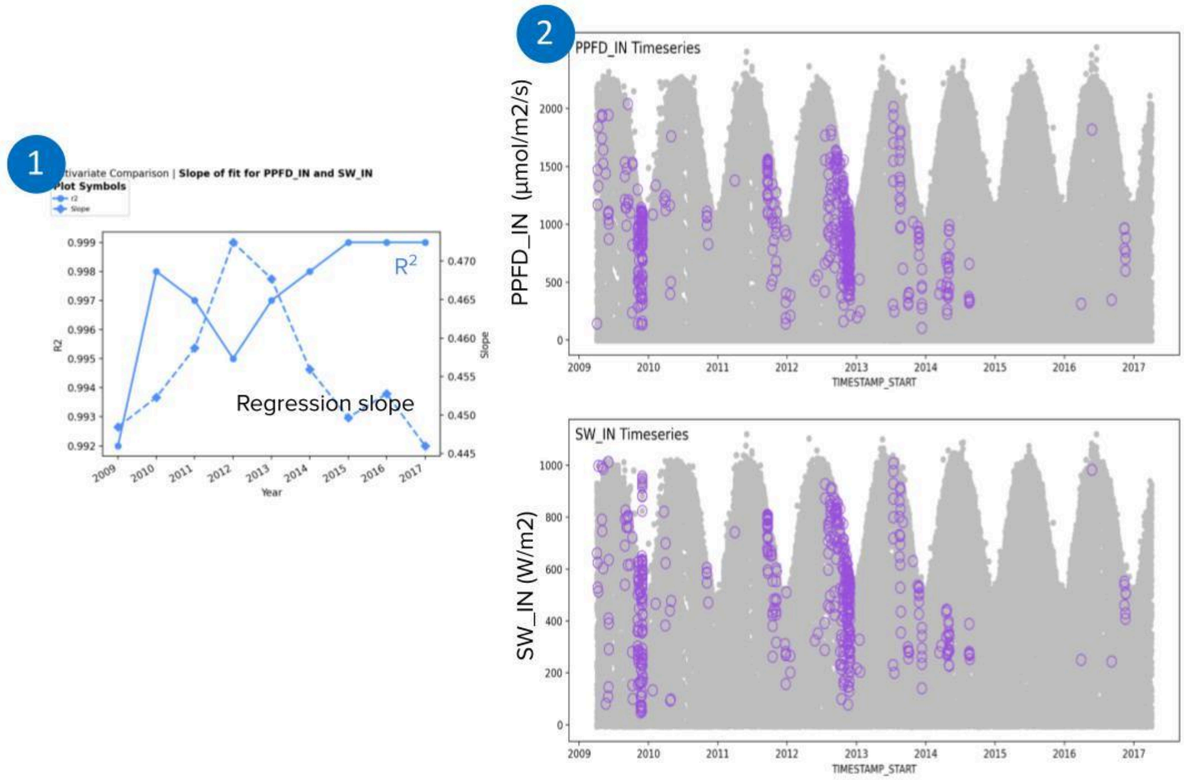


Figure 12. *Example figure with <u>no issues</u> detected by the Multivariate Comparison Module over multiple years. The right panel (2) shows an 8-year time series of PPFD_IN and SW_IN. The left panel (1) shows the time series of regression slopes and R^2, calculated between PPFD_IN and SW_IN each year (as shown in Figure 10-11). The regression slopes were relatively stable (±3%) over the years, suggesting no evident shift or trend in either sensor.*

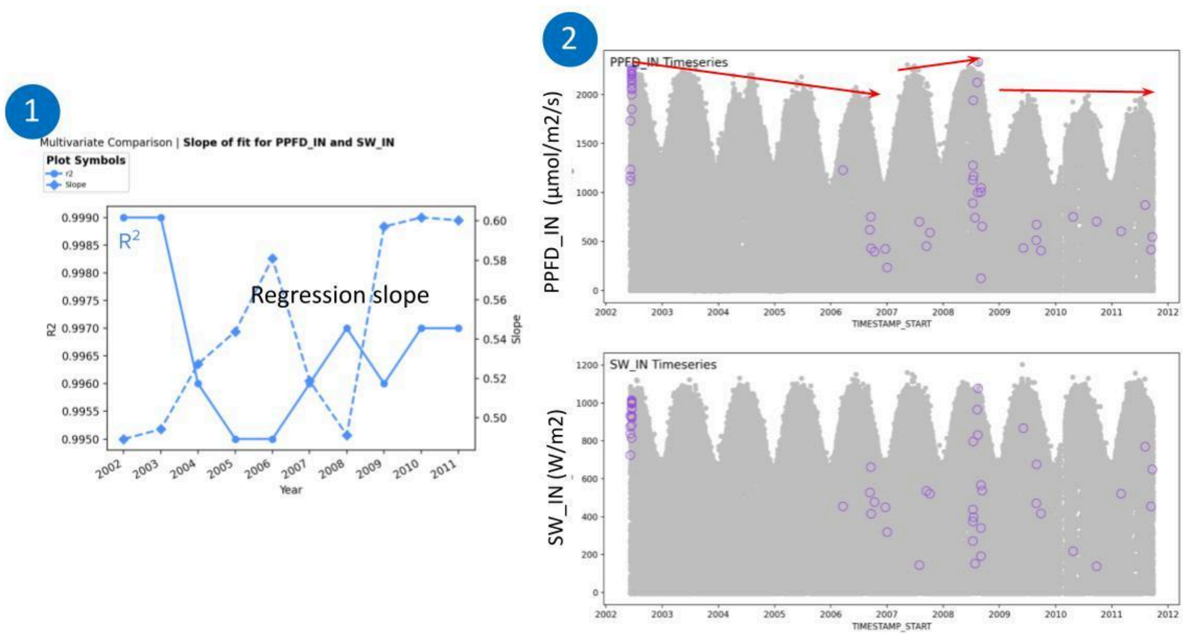


Figure 13. *Example figure with <u>detected issues</u> by the Multivariate Comparison Module over multiple years. The right panel (2) shows a 10-year time series of PPFD_IN and SW_IN. The left panel (1) shows the time series of*

regression slopes and R^2, calculated between PPFD_IN and SW_IN each year (as shown in Figure 10-11). The changes in regression slopes over the years (±11%) suggest one of the radiation sensors, PPFD_IN in this case, has a shifted full range over the years (red arrows) as compared with the other sensor.

4.3 Summary Statistics Table

Table 5. Summary statistics table for the Multivariate Comparison Module

Period	Variable 1	Variable 2	Result	Regression slope	Regression R^2	Slope deviation (%)	Outlier (%)	Figure 1	Figure 2
YEAR[1]	VAR1	VAR2	RESULT[2]	S_{xy}	R^2	ΔS_{xy}[3]	P_{xy}[4]	LINK[5]	LINK[6]
ALL[1]	VAR1	VAR2	RESULT[2]	S_{xy}	R^2	NA	P_{xy}[4]	LINK[5]	LINK[6]

Abbreviation: S_{xy}: orthogonal linear regression slope between the time series X and Y. R^2: coefficient of determination of the linear regression. ΔS_{xy}: relative deviation (%) of a year's S_{xy} to the mean S_{xy}. P_{xy}: percentage (%) of points deviated from the regression line based on the orthogonal distance.

[1] This check is performed on the annual scale and the entire record and requires target variable pairs present in the data. Current target variable pairs: SW_IN-PPFD_IN, TA-T_SONIC, WS-USTAR, and TA profile.

[2] Criteria:

- **FAIL** if ANY of the below
  - $R^2 = 1$
  - $abs(\Delta S_{xy}) > 20\%$
- **WARNING** if ANY of the below
  - $R^2 < 0.7$ for all variable pairs, 0.5 for WS-USTAR
  - $P_{xy} > 1\%$
  - $20\% > abs(\Delta S_{xy}) > 10\%$
- **OK** if none above

[3] $\Delta S_{xy} = (S_{xy} - mean(S_{xy})) / mean(S_{xy}) * 100$; $mean(S_{xy})$ is calculated using all years' S_{xy}, for those years with $R^2 > 0.7$ for all variable pairs, 0.5 for WS-USTAR.

[4] The percentage (%) of data points flagged based on the orthogonal distance to the regression line (threshold * standard deviation of all points' orthogonal distances)

- Variable-specific thresholds: 4.5 for all pairs, 6 for TA profile

[5] Links to multivariate time series and scatter plot

[6] Links to multiyear regression slope and R^2 plot

Table 6. Summary of issues and criteria for issue detection for the Multivariate Comparison Module

Issue types	Issues	Criteria
Short-term mismatch	Erroneous data for a specific period	● $P_{xy} > 1\%$ & $R^2 > 0.7$ for all variables, flagged points for a specific period
	Excessive outlier in either variable	● $P_{xy} > 1\%$ & $R^2 > 0.7$ for all variables, sporadically flagged points

	Shaded radiation	• P_{xy} > 1% & R^2 > 0.7 for PPFD_IN-SW_IN, periodically flagged points
Relationship	Variables not synchronized or poor relationship	• R^2 < 0.5 for WS-USTAR; R^2 < 0.7 for all other variables
	One variable derived from the other	• R^2 == 1 for all variables
Change of slope	Step change in either variable's full range	• abs(ΔS_{xy}) > 10% (warning) or 20% (error) for any year
	Trend in either variable's full range	• abs(ΔS_{xy}) > 10% (warning) or 20% (error) for any year, systematic change in the regression slope

5. Diurnal-Seasonal Pattern Module

5.1 Module Info

- Target Variable: Variables present in the previously published BASE data
- Execution Period: Annual
- Requirement: A site's historical ranges present (i.e., a site has previously published BASE data with 3+ years of record).

The module examines the diurnal-seasonal pattern of a target variable against the historical records at a site and determines if the (newly submitted) data are within the expected ranges. In particular, the module relies on the pronounced temporal variations at the diurnal and seasonal scales of most micrometeorological variables. This module is considered a companion to the Physical Range Module and uses more constrained expected ranges. The check only performs at sites that have previously published BASE data versions and have data records for at least three years. The check could be used to identify the following issues:

- Misalignment between the median diurnal composite
  - Change of the sign convention
  - Shift in timestamps
- Unexpected data ranges
  - Physically unlikely values
  - Outlier
  - Step change in the full range

5.2 Figure Explanation

The module compares the diurnal-seasonal pattern of the newly submitted data against the historical records at the site (e.g., Figure 14). The historical ranges (i.e., 2.5%, 25%, 50%, 75%, and 97.5%) were generated for each variable based on the last version of published BASE data at the site. Both the newly submitted data and historical ranges are organized by the month of the year (i.e., 12 windows per year) and (half-)hour of the day in each month (i.e., 48 steps for a half-hourly resolution). The newly submitted data are also aggregated into a "median diurnal composite" for each monthly window. The module expects the median diurnal composites to align in time between the newly submitted data and the historical records (cross-correlation) (Figure 14). For example, a negative cross-correlation suggests that the submitted data may have opposite sign convection (Figure 15).

The module also checks the percentages of newly submitted data within the 25%-75% and 2.5%-97.5% ranges of the historical ranges. We anticipate that the newly submitted data will mostly fall within the historical range (e.g., Figure 14). If new data have fewer than the expected data percentages within the corresponding ranges, then the module returns a WARNING or FAIL result (Table 7). Figure 16 shows an example year with relatively higher LW_OUT readings (more grey points beyond the historical ranges), indicating a possible shift of the full range. A summary table of commonly seen issues is provided to aid the interpretation and correction of identified issues (Table 8).

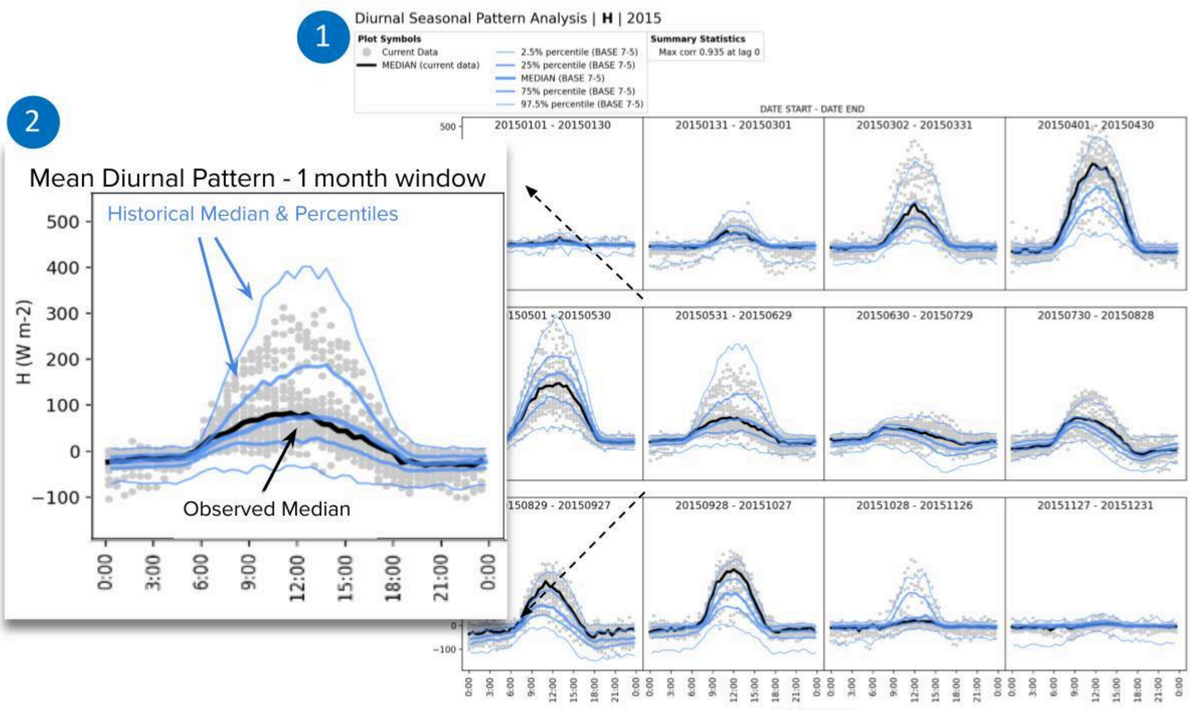


Figure 14. *Example figure with <u>no issues</u> detected by the Diurnal-Seasonal Pattern Module. (1) The right panel shows the monthly diurnal plots for a one-year time series of sensible heat flux (H), including newly submitted data (gray data points and the black median line) and historical ranges (2.5th, 25th, 50th, 75th, 97.5th percentiles in blue lines). Panel (2) shows a monthly example. This case shows that most new data points are within the historical 2.5th-97.5th percentile range and have similar median diurnal patterns to the historical records.*

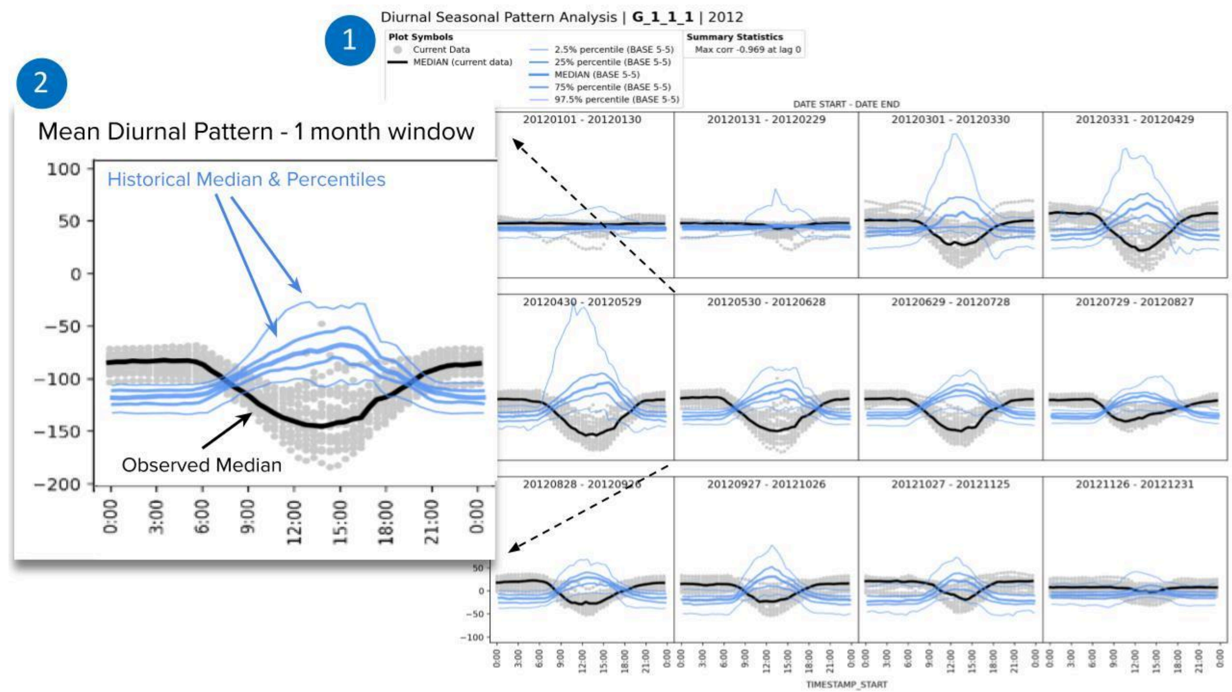


Figure 15. *Example figure with* *<u>detected issues</u>* *by the Diurnal-Seasonal Pattern Module. (1) The right panel shows the monthly diurnal plots for a one-year time series of soil heat flux (G), including newly submitted data (gray data points and the black median line) and historical ranges (2.5th, 25th, 50th, 75th, 97.5th percentiles in blue lines). Panel (2) shows a monthly example. The opposite mean diurnal patterns between the two data versions suggest that the newly submitted data have an opposite sign convention.*

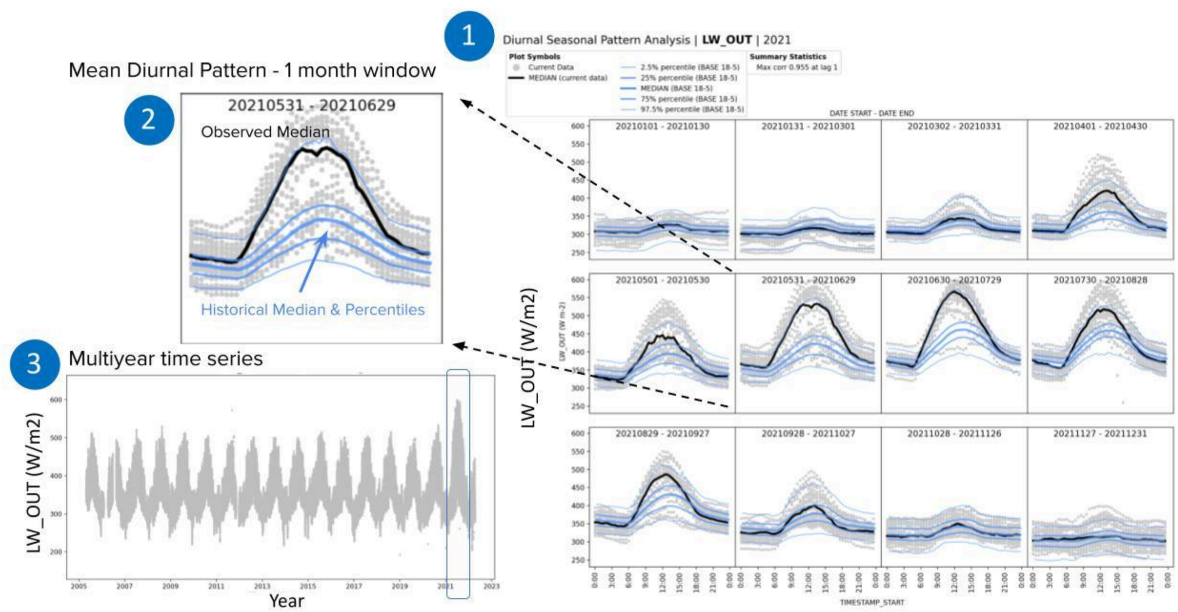


Figure 16. *Example figure with* *<u>detected issues</u>* *by the Diurnal-Seasonal Pattern Module. (1) The right panel shows the monthly diurnal plots for a one-year time series of outgoing longwave radiation (LW_OUT), including newly submitted data (gray data points and the black median line) and historical ranges (2.5th, 25th, 50th, 75th, 97.5th percentiles in blue lines). Panel (2) shows a monthly example. Panel (3) shows the multi-year time series, with the*

box highlighting the one year in the right panel. The example year has relatively higher LW_OUT readings than previous years (more gray points beyond the historical ranges in (1) and (2)).

5.3 Summary Statistics Table

Table 7. Summary statistics table for Diurnal-Seasonal Pattern Module

Period	Variable	Result	Time lag	Cross correlation	Percentage within the historical interquartile range (%)	Percentage outside the historical 95% range (%)	Figure
YEAR[1]	VAR	RESULT[2]	t_{max}[3]	R_{xy}|max(abs(R_{xy}))[3]	P_{iqr}	$P_{95\%}$	LINK

Abbreviation: R_{xy}|max(abs(R_{xy})): cross-correlation (R_{xy}) when maximum absolute R_{xy} was found between the time series X and Y. t_{max} : the timestep shift at which the R_{xy}|max(abs(R_{xy})) is found. P_{iqr} : percentage of occasions that the new data are within the 25%-75% ranges of historical records. $P_{95\%}$: percentage of occasions that the new data are outside the 2.5%-97.5% range of historical records.

[1] This check only performs on an annual scale and requires a site's historical ranges to be present (i.e., a site has previously published BASE data with 3+ years of record). The check is performed on variables that have matched variable names within the historical ranges.

[2] Criteria:

- **FAIL** if ANY of the below
    - R_{xy}|max(abs(R_{xy})) < -0.4
    - P_{iqr} < 15% OR $P_{95\%}$ > 30%
- **WARNING** if ANY of the below
    - abs(R_{xy}|max(abs(R_{xy}))) > 0.4 AND abs(t_{max}) > 0
    - P_{iqr} < 30% OR $P_{95\%}$ > 15%
- **OK** if none above

[3] Cross-correlation and time lag are calculated when both composite time series have no gaps.

Table 8. Summary of issues and criteria for issue detection for the Diurnal-Seasonal Pattern module

Issue types	Issues	Criteria
Misalignment between the median diurnal composite	Shift in timestamps	• abs(R_{xy}|max(abs(R_{xy}))) > 0.4 & abs(t_{max}) > 0
	Opposite sign convention	• R_{xy}|max(abs(R_{xy})) < -0.4
Unexpected data ranges	Slight shift in data ranges	• P_{iqr} < 30% & $P_{95\%}$ > 15%
	Evident shift in data ranges	• P_{iqr} < 15% & $P_{95\%}$ > 30%

6. USTAR Filtering Module

6.1 Module Info

- ❖ Target Variable: USTAR-FC
- ❖ Execution Period: Annual
- ❖ Requirement: Both FC and USTAR present

The USTAR (friction velocity) Filtering Module examines whether the CO_2 flux (FC) is filtered by using USTAR thresholds. Flux data submitted to AmeriFlux should not be USTAR-filtered, as the further ONEFlux processing step implements a standard procedure in determining the USTAR thresholds and filtering FC data. The module is used to identify the following issues:

- Filtered FC by USTAR threshold
- Filtered USTAR

6.2 Figure Explanation

For each pair of USTAR and FC (i.e., values for the same year), the module finds the lower bound of USTAR when the concurrent FC is not missing and from all USTAR data in a year, respectively. The module expects these two lower bounds to match and be close to the expected lower range of USTAR (Figure 17, Table 9). The check returns a WARNING or FAIL if the two lower bounds differ by more than 0.02 or 0.1 m s^{-1} (Figure 18), indicating FC is potentially or very likely filtered using a USTAR threshold.

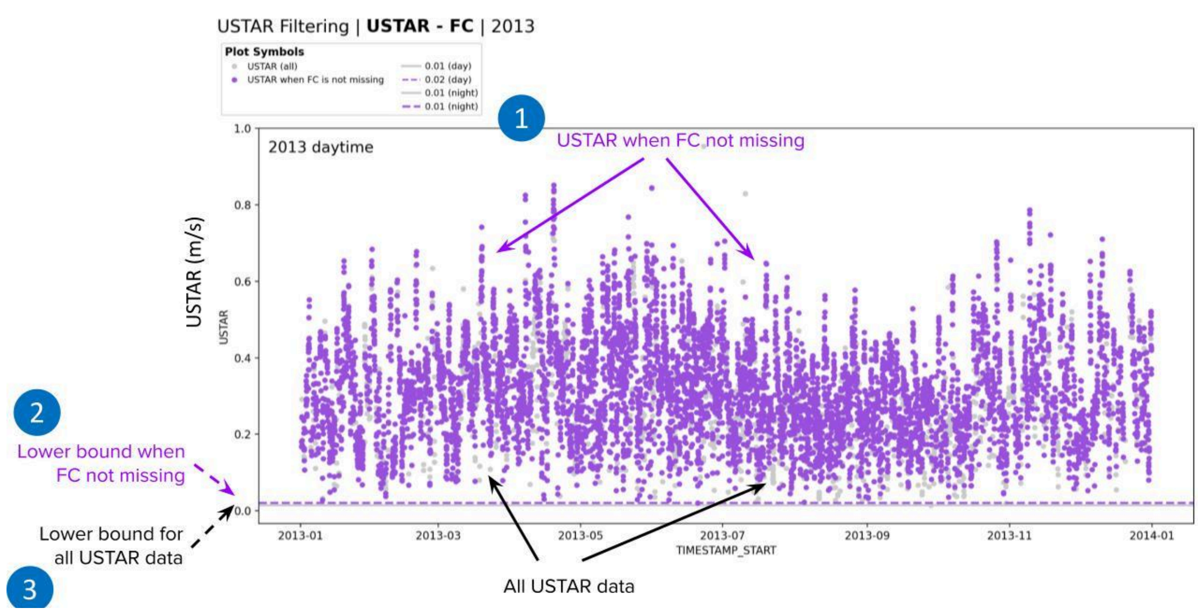


Figure 17. *Example figure with <u>no issues</u> detected by the USTAR Filtering Module. This example shows that the lower bound of USTAR when concurrent FC is not missing (purple dashed line) is equal to or close to the lower bound from all USTAR data (solid black line).*

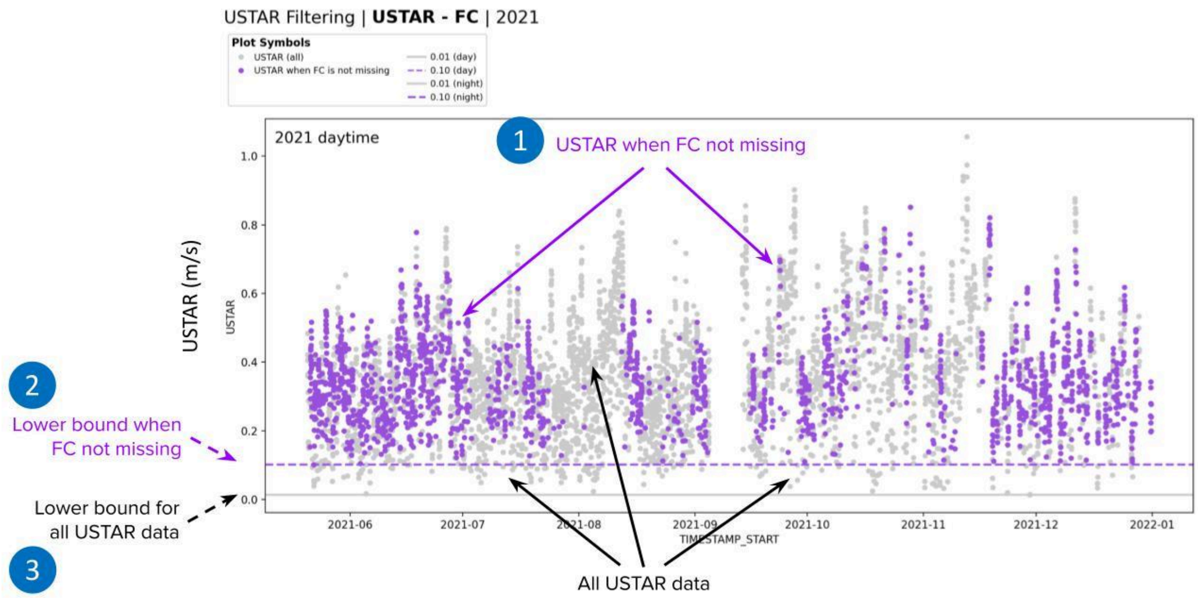


Figure 18. *Example figure with detected issues by the USTAR Filtering Module. This example shows that the lower bound of USTAR when concurrent FC is not missing (purple dashed line) is higher than the lower bound from all USTAR data (solid gray line). The difference indicates that the FC data are filtered using a USTAR threshold of 0.1 m s^{-1}.*

6.3 Summary Statistics Table

Table 9. Summary statistics table for USTAR Filtering Module

Period	FC variable	USTAR variable	Result	Daytime min USTAR	Daytime min USTAR with FC	Nighttime min USTAR	Nighttime min USTAR with FC	Figure
YEAR[1]	VAR1	VAR2	RESULT[2]	$min(u^*)_{day}$	$min(u^*)_{night}$	$min(u^*\|FC)_{day}$	$min(u^*\|FC)_{night}$	LINK

Abbreviation: $min(u^*)_{day}$: Annual minimum daytime USTAR, $min(u^*)_{night}$: Annual minimum nighttime USTAR, $min(u^*|FC)_{day}$: Annual minimum daytime USTAR when FC is not missing, $min(u^*|FC)_{night}$: Annual minimum nighttime USTAR when FC is not missing

[1] This check is performed on an annual basis and requires the presence of both FC and USTAR.

[2] Criteria:

- **FAIL** if ANY of the below
  - $min(u^*)_{day} > 0.1$
  - $min(u^*)_{night} > 0.1$
  - $min(u^*|FC)_{day} > 0.1$
  - $min(u^*|FC)_{night} > 0.1$
  - $abs(min(u^*|FC)_{day} - min(USTAR)_{day}) > 0.1$
  - $abs(min(u^*|FC)_{night} - min(USTAR)_{night}) > 0.1$
- **WARNING** if ANY of the below
  - $min(u^*)_{day} > 0.02$

  - $min(u^*)_{night} > 0.02$
  - $min(u^*|FC)_{day} > 0.02$
  - $min(u^*|FC)_{night} > 0.02$
  - $abs(min(u^*|FC)_{day} - min(USTAR)_{day}) > 0.02$
  - $abs(min(u^*|FC)_{night} - min(USTAR)_{night}) > 0.02$
- **OK** if none above

7. Variable Coverage Module

7.1 Module Info

The variable coverage module examines the presence and coverage of all variables at a site. The generated figure provides a quick overview of the available variables and their data coverage for each year of the entire record (Figures 19-20). The figure can be used to examine whether data in certain years are entirely missing (e.g., inactive years), whether specific variables are missing for certain periods (e.g., not measured or submitted), or whether certain variables are entirely missing (e.g., all empty columns). For long-running and heavily instrumented sites, the figure can be used to verify the presence and continuity of variables across the entire record.

7.2 Figure Explanation

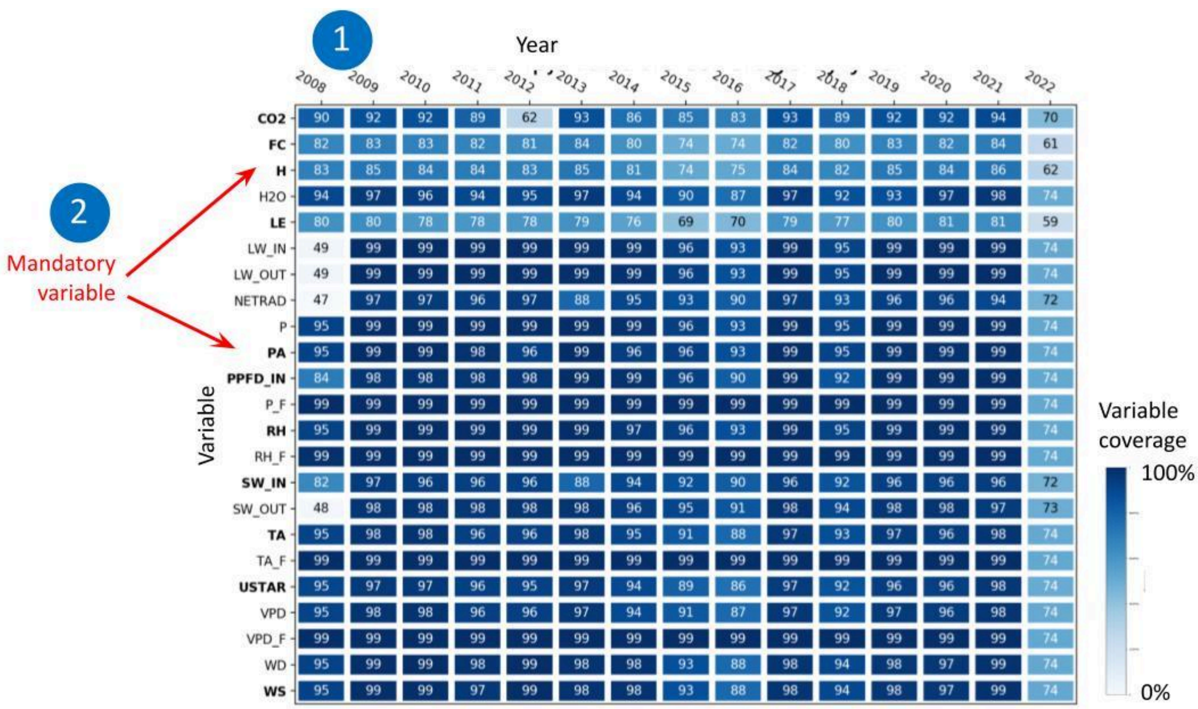


Figure 19. *Example figure with* <u>*no issues*</u> *detected by the Variable Coverage Module. The figure shows the variable coverage (color gradient) by year. The example demonstrates that all measurements are present in every year (1). Bolded texts in variable names denote mandatory variables required for ONEFlux processing.*

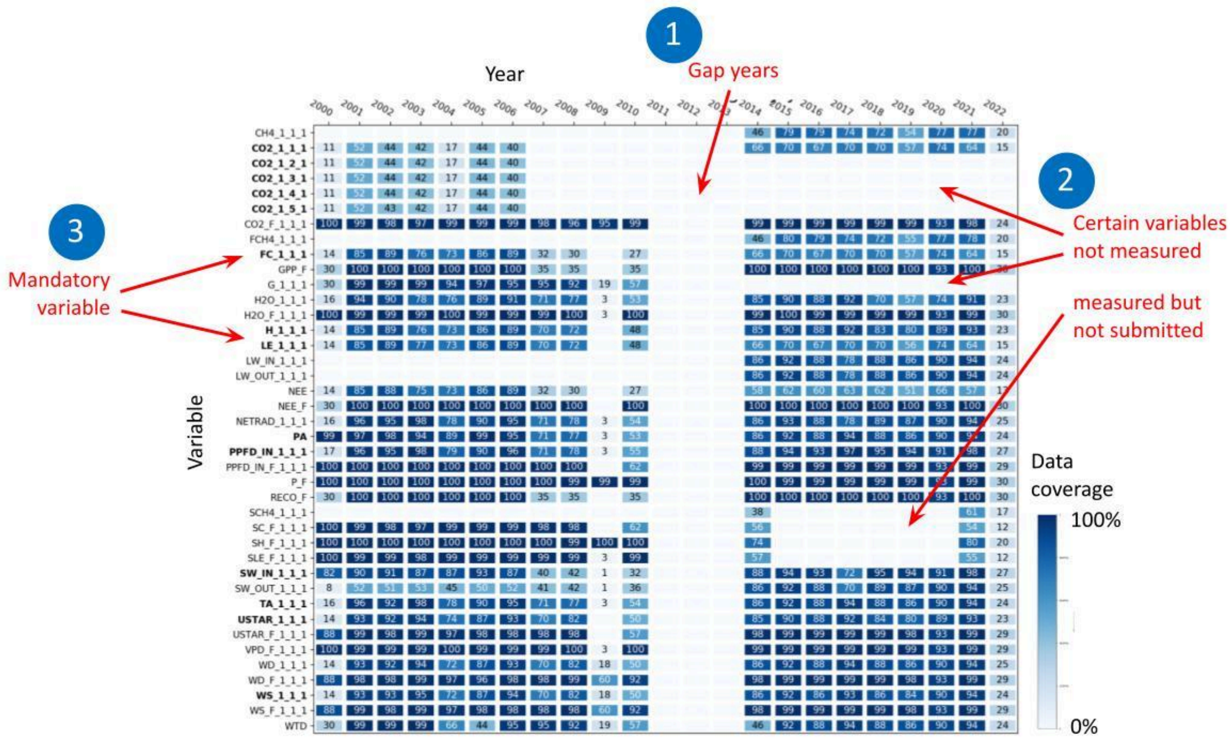


Figure 20. *Example figure with potential* <u>*issues detected*</u> *by the Variable Coverage Module. The figure shows the variable coverage (color gradient) by year. The example shows three gap years when all measurements are inactive (1), and the periods when certain variables are not measured or submitted (2). Bolded texts in variable names denote mandatory variables required for ONEFlux processing.*

8. Appendix

Table A1. List of Flux-Processing (FP) Standard variables and their defined upper and lower thresholds in the Physical Range Module.

Name	Description	Unit	Lower	Upper
COND_WATER	Conductivity (i.e., electrical conductivity) of water	µS cm-1	0	10000
DO	Dissolved oxygen in water	µmol L-1	0	*NA*
PCH4	Dissolved methane (CH4) in water	nmolCH4 mol-1	0	*NA*
PCO2	Dissolved carbon dioxide (CO2) in water	µmolCO2 mol-1	0	10000
PN2O	Dissolved nitrous oxide (N2O) in water	nmolN2O mol-1	0	*NA*
PPFD_UW_IN	Photosynthetic photon flux density, underwater, incoming	µmolPhotons m-2 s-1	0	2400
TW	Water temperature	deg C	-20	50
DBH	Diameter of tree measured at breast height (1.3m) with continuous dendrometers	cm	0	500

LEAF_WET	Leaf wetness, range 0-100	%	0	100
SAP_DT	Difference of probes temperature for sapflow measurements	deg C	-10	10
SAP_FLOW	Sap flow	mmolH2O m-2 s-1	*NA*	*NA*
T_BOLE	Bole temperature	deg C	-50	70
T_CANOPY	Temperature of the canopy and/or surface underneath the sensor	deg C	-50	70
FETCH_70	Distance at which cross-wind integrated footprint cumulative probability is 70%	m	0	10000
FETCH_80	Distance at which cross-wind integrated footprint cumulative probability is 80%	m	0	12000
FETCH_90	Distance at which cross-wind integrated footprint cumulative probability is 90%	m	0	15000
FETCH_FILTER	Footprint quality flag (i.e., 0, 1): 0 and 1 indicate data measured when wind coming from direction that should be discarded and kept, respectively	nondimensional	0	1
FETCH_MAX	Distance at which footprint contribution is maximum	m	0	5000
CH4	Methane (CH4) mole fraction in wet air	nmolCH4 mol-1	0	15000
CH4_MIXING_RATIO	Methane (CH4) in mole fraction of dry air	nmolCH4 mol-1	0	15000
CO	Carbon Monoxide (CO) mole fraction in wet air	nmolCO mol-1	0	*NA*
CO2	Carbon Dioxide (CO2) mole fraction in wet air	µmolCO2 mol-1	150	1200
CO2_MIXING_RATIO	Carbon Dioxide (CO2) in mole fraction of dry air	µmolCO2 mol-1	150	1200
CO2_SIGMA	Standard deviation of carbon dioxide mole fraction in wet air	µmolCO2 mol-1	0	150
CO2C13	Stable isotopic composition of CO2 - C13 (i.e., d13C of CO2)	‰ (permil)	*NA*	-6
FC	Carbon Dioxide (CO2) turbulent flux (no storage correction)	µmolCO2 m-2 s-1	-100	100
FCH4	Methane (CH4) turbulent flux (no storage correction)	nmolCH4 m-2 s-1	-500	4000
FN2O	Nitrous oxide (N2O) turbulent flux (no storage correction)	nmolN2O m-2 s-1	*NA*	*NA*

FNO	Nitric oxide (NO) turbulent flux (no storage correction)	nmolNO m-2 s-1	*NA*	*NA*
FNO2	Nitrogen dioxide (NO2) turbulent flux (no storage correction)	nmolNO2 m-2 s-1	*NA*	*NA*
FO3	Ozone (O3) turbulent flux (no storage correction)	nmolO3 m-2 s-1	*NA*	*NA*
H2O	Water (H2O) vapor in mole fraction of wet air	mmolH2O mol-1	0	100
H2O_MIXING_RATIO	Water (H2O) vapor in mole fraction of dry air	mmolH2O mol-1	0	100
H2O_SIGMA	Standard deviation of water vapor mole fraction	mmolH2O mol-1	0	15
N2O	Nitrous Oxide (N2O) mole fraction in wet air	nmolN2O mol-1	0	*NA*
N2O_MIXING_RATIO	Nitrous Oxide (N2O) in mole fraction of dry air	nmolN2O mol-1	0	*NA*
NO	Nitric oxide (NO) mole fraction in wet air	nmolNO mol-1	0	*NA*
NO2	Nitrogen dioxide (NO2) mole fraction in wet air	nmolNO2 mol-1	0	*NA*
O3	Ozone (O3) mole fraction in wet air	nmolO3 mol-1	0	*NA*
SC	Carbon Dioxide (CO2) storage flux	µmolCO2 m-2 s-1	-100	100
SCH4	Methane (CH4) storage flux	nmolCH4 m-2 s-1	*NA*	*NA*
SN2O	Nitrous oxide (N2O) storage flux	nmolN2O m-2 s-1	*NA*	*NA*
SNO	Nitric oxide (NO) storage flux	nmolNO m-2 s-1	*NA*	*NA*
SNO2	Nitrogen dioxide (NO2) storage flux	nmolNO2 m-2 s-1	*NA*	*NA*
SO2	Sulfur Dioxide (SO2) mole fraction in wet air	nmolSO2 mol-1	0	*NA*
SO3	Ozone (O3) storage flux	nmolO3 m-2 s-1	*NA*	*NA*
FH2O	Water vapor (H2O) turbulent flux (no storage correction)	mmolH2O m-2 s-1	-10	20
G	Soil heat flux	W m-2	-250	400
H	Sensible heat turbulent flux (no storage correction)	W m-2	-450	900
LE	Latent heat turbulent flux (no storage correction)	W m-2	-450	900
SB	Heat storage flux in biomass	W m-2	*NA*	*NA*

SG	Heat storage flux in the soil above the soil heat fluxes measurement	W m-2	-100	250
SH	Sensible heat (H) storage flux	W m-2	-150	150
SLE	Latent heat (LE) storage flux	W m-2	-150	150
PA	Atmospheric pressure	kPa	60	105
PBLH	Planetary boundary layer height	m	0	3000
RH	Relative humidity, range 0-100	%	0	100
T_SONIC	Sonic temperature	deg C	-50	50
T_SONIC_SIGMA	Standard deviation of sonic temperature	deg C	0	5
TA	Air temperature	deg C	-50	50
VPD	Vapor Pressure Deficit	hPa	0	80
D_SNOW	Snow depth	cm	0	500
P	Precipitation	mm	0	50
P_RAIN	Rainfall	mm	0	50
P_SNOW	Snowfall	mm	0	50
RUNOFF	Run off	mm	0	200
STEMFLOW	Excess precipitation that drains from outlying branches and leaves and is channeled through the stems to the ground	mm	0	200
THROUGHFALL	Excess precipitation that passes directly through a canopy or drips from wet leaves to the ground	mm	0	20
ALB	Albedo, range 0-100	%	0	100
APAR	Absorbed PAR	µmolPhoton m-2 s-1	0	2300
EVI	Enhanced Vegetation Index	nondimensional	-1	1
FAPAR	Fraction of absorbed PAR, range 0-100	%	0	100
FIPAR	Fraction of intercepted PAR, range 0-100	%	0	100
LW_BC_IN	Longwave radiation, below canopy incoming	W m-2	50	600
LW_BC_OUT	Longwave radiation, below canopy outgoing	W m-2	100	750
LW_IN	Longwave radiation, incoming	W m-2	50	600
LW_OUT	Longwave radiation, outgoing	W m-2	100	750

MCRI	Carotenoid Reflectance Index (Gitelson et al., 2002)	nondimensional	0	10
MTCI	Meris Terrestrial Chlorophyll Index (Dash and Curran, 2004)	nondimensional	0	10
NDVI	Normalized Difference Vegetation Index	nondimensional	-1	1
NETRAD	Net radiation	W m-2	-200	1100
NIRV	Near Infrared Vegetation Index (Badgley et al., 2017)	W m-2 sr-1 nm-1	0	2
PPFD_BC_IN	Photosynthetic photon flux density, below canopy incoming	µmolPhoton m-2 s-1	0	2400
PPFD_BC_OUT	Photosynthetic photon flux density, below canopy outgoing	µmolPhoton m-2 s-1	0	2000
PPFD_DIF	Photosynthetic photon flux density, diffuse incoming	µmolPhoton m-2 s-1	0	1400
PPFD_DIR	Photosynthetic photon flux density, direct incoming	µmolPhoton m-2 s-1	0	2400
PPFD_IN	Photosynthetic photon flux density, incoming	µmolPhoton m-2 s-1	0	2400
PPFD_OUT	Photosynthetic photon flux density, outgoing	µmolPhoton m-2 s-1	0	2000
PRI	Photochemical Reflectance Index	nondimensional	-1	1
R_UVA	UVA radiation, incoming	W m-2	0	85
R_UVB	UVB radiation, incoming	W m-2	0	20
REDCI	Red Edge Chlorophyll Index	nondimensional	0	10
REP	Red Edge Position (Dash and Curran, 2004)	nm	400	800
SPEC_NIR_IN	Radiation (near infra-red band), incoming (hemispherical)	W m-2 nm-1	0	2
SPEC_NIR_OUT	Radiation (near infra-red band), outgoing	W m-2 sr-1 nm-1	0	2
SPEC_NIR_REFL	Reflectance (near infra-red band)	nondimensional	0	1
SPEC_PRI_REF_IN	Radiation for PRI reference band (e.g., 570 nm), incoming (hemispherical)	W m-2 nm-1	0	2
SPEC_PRI_REF_OUT	Radiation for PRI reference band (e.g., 570 nm), outgoing	W m-2 sr-1 nm-1	0	2
SPEC_PRI_REF_REFL	Reflectance for PRI reference band (e.g., 570 nm)	nondimensional	0	1
SPEC_PRI_TGT_IN	Radiation for PRI target band (e.g., 531 nm), incoming (hemispherical)	W m-2 nm-1	0	2

SPEC_PRI_TGT_OUT	Radiation for PRI target band (e.g., 531 nm), outgoing	W m-2 sr-1 nm-1	0	2
SPEC_PRI_TGT_REFL	Reflectance for PRI target band (e.g., 531 nm)	nondimensional	0	1
SPEC_RED_IN	Radiation (red band), incoming (hemispherical)	W m-2 nm-1	0	2
SPEC_RED_OUT	Radiation (red band), outgoing	W m-2 sr-1 nm-1	0	2
SPEC_RED_REFL	Reflectance (red band)	nondimensional	0	1
SR	Simple Ratio	nondimensional	0	10
SW_BC_IN	Shortwave radiation, below canopy incoming	W m-2	0	1300
SW_BC_OUT	Shortwave radiation, below canopy outgoing	W m-2	0	800
SW_DIF	Shortwave radiation, diffuse incoming	W m-2	0	750
SW_DIR	Shortwave radiation, direct incoming	W m-2	0	1300
SW_IN	Shortwave radiation, incoming	W m-2	0	1300
SW_OUT	Shortwave radiation, outgoing	W m-2	0	800
TCARI	Transformed Chlorophyll Absorption in Reflectance Index	nondimensional	0	10
SWC	Soil water content (volumetric), range 0-100	%	0	100
SWP	Soil water potential	kPa	-750	0
TS	Soil temperature	deg C	-40	65
TSN	Snow temperature	deg C	-40	4
WTD	Water table depth	m	-10	10
MO_LENGTH	Monin-Obukhov length	m	*NA*	*NA*
TAU	Momentum flux	kg m-1 s-2	-10	2
U_SIGMA	Standard deviation of velocity fluctuations (towards main-wind direction after coordinates rotation)	m s-1	0	12
USTAR	Friction velocity	m s-1	0	8
V_SIGMA	Standard deviation of lateral velocity fluctuations (cross main-wind direction after coordinates rotation)	m s-1	0	10
W_SIGMA	Standard deviation of vertical velocity fluctuations	m s-1	0	5
WD	Wind direction	Decimal degrees	0	360

WD_SIGMA	Standard deviation of wind direction (Yamartino, 1984)	decimal degree	0	180
WS	Wind speed	m s-1	0	40
WS_MAX	maximum WS in the averaging period	m s-1	0	50
ZL	Monin-Obukhov Stability parameter	nondimensional	*NA*	*NA*
GPP	Gross Primary Productivity	µmolCO2 m-2 s-1	-30	100
NEE	Net Ecosystem Exchange	µmolCO2 m-2 s-1	-100	100
RECO	Ecosystem Respiration	µmolCO2 m-2 s-1	-20	50
FC_SSITC_TEST	Results of the quality flagging for FC according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests by Foken and Wichura (1996) (i.e., 0, 1, 2)	nondimensional	0	2
FCH4_SSITC_TEST	Results of the quality flagging for FCH4 according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests by Foken and Wichura (1996) (i.e., 0, 1, 2)	nondimensional	0	2
FN2O_SSITC_TEST	Results of the quality flagging for FN2O according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests by Foken and Wichura (1996) (i.e., 0, 1, 2)	nondimensional	0	2
FNO_SSITC_TEST	Results of the quality flagging for FNO according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests by Foken and Wichura (1996) (i.e., 0, 1, 2)	nondimensional	0	2
FNO2_SSITC_TEST	Results of the quality flagging for FNO2 according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests by Foken and Wichura (1996) (i.e., 0, 1, 2)	nondimensional	0	2
FO3_SSITC_TEST	Results of the quality flagging for FO3 according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests	nondimensional	0	2

	by Foken and Wichura (1996) (i.e., 0, 1, 2)			
H_SSITC_TEST	Results of the quality flagging for H according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests by Foken and Wichura (1996) (i.e., 0, 1, 2)	nondimensional	0	2
LE_SSITC_TEST	Results of the quality flagging for LE according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests by Foken and Wichura (1996) (i.e., 0, 1, 2)	nondimensional	0	2
TAU_SSITC_TEST	Results of the quality flagging for TAU according to Foken et al 2004, based on a combination of Steady State and Integral Turbulence Characteristics tests by Foken and Wichura (1996) (i.e., 0, 1, 2)	nondimensional	0	2

9. Reference

- Badgley, G., Field, C.B. and Berry, J.A., 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. Science Advances, 3(3): e1602244. 10.1126/sciadv.1602244
- Chu, H., Christianson, D.S., Cheah, YW. et al. AmeriFlux BASE data pipeline to support network growth and data sharing. Sci Data 10, 614 (2023). https://doi.org/10.1038/s41597-023-02531-2
- Dash, J. and Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. International Journal of Remote Sensing, 25(23): 5403-5413. 10.1080/0143116042000274015
- Foken, T. et al., 2004. Post-field data quality control. In: X. Lee, W. Massman and B. Law (Editors), Handbook of Micrometeorology: A Guide for Surface Flux Measurement and Analysis. Kluwer Academic, Dordrecht, Netherlands, pp. 181-208.
- Foken, T. and Wichura, B., 1996. Tools for quality assessment of surface-based flux measurements. Agric For Meteorol, 78(1-2): 83-105.
- Gitelson, A.A., Kaufman, Y.J., Stark, R. and Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote sensing of Environment, 80(1): 76-87.
- Pastorello, G., et al. (2014), Observational data patterns for time series data quality assessment, paper presented at e-Science (e-Science), 2014 IEEE 10th International Conference on, 20-24 Oct. 2014. DOI:10.1109/eScience.2014.45
- Pastorello, G., et al. (2020), The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific Data, 7(1), 225, DOI:10.1038/s41597-020-0534-3.

- Yamartino, R.J., 1984. A Comparison of Several “Single-Pass” Estimators of the Standard Deviation of Wind Direction. Journal of Applied Meteorology and Climatology, 23(9): 1362-1366. 10.1175/1520-0450(1984)023<1362:Acospe>2.0.Co;2