Eddy Covariance Flux Correction Standard Operating Procedure (SOP) for the Long-Term Agro-Ecological Research Stations at Washington State University – Cook Agronomy Farm (CAF-LTAR)

By: Eric S. Russell

Laboratory for Atmospheric Research

Dept. Civil and Environmental Engineering

Washington State University

[eric.s.russell@wsu.edu](mailto:eric.s.russell@wsu.edu)

Thanks to Zhongming Gao, Ian Leslie, and Bryan Carlson for information included in this document.

# List of Updates:

**Updated**: 1/18/2018 (ESR)

**Updated**: 1/28/2018 (ESR)

**Updated**: 2/9/2018 (ESR; added Appendices C and D, Section 4.3)

**Updated**: 2/12/2018 (ESR; updated Appendix C, added text to Section 5 to fill out gap-filling process and procedure)

**Updated**: 2/14/2018 (ESR, added Section 5.2, Appendix E, updates in Section 5)

**Updated**: 2/15/2018 (ESR, added paragraph re. third-party gap filling in Section 5.2)

**Updated**: 2/16/2018 (ESR, added Appendix F)

**Updated**: 2/20/2018 (ESR, added beginnings of Section 6: Soil heat flux, into main body)

**Updated**: 3/5/2018 (ESR, number of edits through the main body to improve clarity and match current practices better)

**Updated**: 3/15/2018 (ESR, Added net Section 5: Meteorology Data QC and Section 6.1 for gap-filling of the QC’ed and missing meteorology data. Minor edits to grammar and format to adjust for new section numbers)

**Updated**: 3/22/2018 (ESR, Added Section 4.4 for AmeriFlux data conversion and updated Section 4.3 to include new column rename idea)

**Updated**: 3/25/2018 (ESR, Added Section 7 to describe ET partitioning).

**Updated**: 3/26/2018 (ESR, Added L3 appendix and descriptions, removed L2 skinny table. Changed combining table to join EPro data with EFlux data to match new program changes. Updates to text and added short paragraph before each appendix table.)

**Updated**: 4/3/2018-4/4/2018 (ESR, added more text to Section 8 to finish initial summary, updated Section 7 for completeness, and general updates and edits to the text. Added procedure for the Python processing script and re-formatted headers into chapters.)

**Updated:** 5/25/2018 (Updated meteorology QC columns to most recent version in Appendix B)

**Updated:** 6/9/2018 (Added Appendix C header, updated column headers for Appendix B and D.)

**Updated:** 12/10/2018 (Added new Section 1.10 to handle post gapfilling processing. Deleted Appendix D as gap-filling code was changed to REddy from REACCH basis. Renamed file to \*\_20181210. Added comments in sections that need to be updated because of the new processing procedures; updates forthcoming.)

**Updated:** 1/4/2021 (ESR, Return to this updating and fixing to match current processing procedures. Fixing typos and other small items as well.)

**Updated:** 3/10/2021 (ESR, Added comments as replies, strikethrough step-by-step processing section since very out of date, added general flow chart as Appendix B.)

# Table of Contents

[List of Updates: 2](#_Toc66261909)

[Table of Contents 3](#_Toc66261910)

[Chapter 1 7](#_Toc66261911)

[1.1 Introduction 7](#_Toc66261912)

[1.2 10Hz Data Collection (Level 0 data) 7](#_Toc66261913)

[1.3 Flux Processing (Level 1 data) 9](#_Toc66261914)

[1.3.1 Flux Data Files: Different Processing Procedures: 9](#_Toc66261915)

[1.3.2 Low and High frequency CO2 data 10](#_Toc66261916)

[1.3.3 Summary 10](#_Toc66261917)

[1.4 Post-Processing (Level 2 data) 11](#_Toc66261918)

[1.4.1 Filter values used 11](#_Toc66261919)

[1.4.2 Flag Identifiers 13](#_Toc66261920)

[1.4.3 Column Dropping 14](#_Toc66261921)

[1.4.4 AmeriFlux Data Processing 14](#_Toc66261922)

[1.5 Meteorology Data Quality Control and Gap-Filling 15](#_Toc66261923)

[1.6 Soil Data QC Checks 17](#_Toc66261924)

[1.7 Flux Data Gap-Filling 17](#_Toc66261925)

[1.7.3 Formatting Level 3 Data 18](#_Toc66261926)

[*1.7.4 Level 3 Data Naming Convention* 18](#_Toc66261927)

[*1.8* *Transpiration and Evaporation Partitioning* 18](#_Toc66261928)

[*1.9* *Soil heat flux and soil storage* 19](#_Toc66261929)

[1.10 Contact Information 19](#_Toc66261930)

[Chapter 2: 20](#_Toc66261931)

[2.1 Introduction 20](#_Toc66261932)

[2.2 Step-by-step processing for EddyPro 21](#_Toc66261933)

[~~2.3 Python L3 Processing and Gap-Filling Script~~ 25](#_Toc66261934)

[~~2.3.1 Import Libraries~~ 25](#_Toc66261935)

[~~2.3.2 Variable Definitions~~ 25](#_Toc66261936)

[~~2.3.3 Order of Function Calls and Definitions~~ 27](#_Toc66261937)

[~~2.4 Step-by-step Processing:~~ 30](#_Toc66261938)

[2.5. Author comments on Processing Procedure: 31](#_Toc66261939)

[Chapter 3. Summary of Flux Corrections for CAF LTAR at Washington State University 32](#_Toc66261940)

[3.1. Introduction 32](#_Toc66261941)

[3.2. Data filtering and Pre-Processing 32](#_Toc66261942)

[3.2.1 Despiking 33](#_Toc66261943)

[3.2.2 Diagnostic Flags 33](#_Toc66261944)

[3.2.3 Hard-limit filter 33](#_Toc66261945)

[3.3. Coordinate Rotation 34](#_Toc66261946)

[3.3.1. Double Rotation 34](#_Toc66261947)

[3.3.2. Planar Fit Method 35](#_Toc66261948)

[3.4. Spatial Sensor Separation 36](#_Toc66261949)

[3.5. Frequency corrections 36](#_Toc66261950)

[3.5.1 High Frequency Loss 37](#_Toc66261951)

[3.5.2 Low Frequency Loss 37](#_Toc66261952)

[3.6. SND Buoyancy Correction 38](#_Toc66261953)

[3.7. WPL Density Fluctuation Correction 39](#_Toc66261954)

[3.8. Data quality grades 39](#_Toc66261955)

[3.8.1 Steady State 41](#_Toc66261956)

[3.8.2 Surface layer turbulence characteristics 41](#_Toc66261957)

[3.8.3 Wind directions sectors 42](#_Toc66261958)

[3.9 Final Comments 42](#_Toc66261959)

[3.10. CAF-Specific Post-Processing 43](#_Toc66261960)

[Chapter 4: EddyPro – EasyFlux Comparison 44](#_Toc66261961)

[References 55](#_Toc66261962)

[Appendix A: Added columns from EddyPro Output to EasyFlux data 59](#_Toc66261963)

[Appendix B: General workflow for CAF EC data processing. 60](#_Toc66261964)

This Page Left Intentionally Blank

This Page Left Intentionally Blank

# Chapter 1

## Introduction

This document describes the procedure in which the raw data measured at the eddy covariance (EC) flux towers at the CAF LTAR research sites are processed. Updates to this will be made as appropriate. There are many corrections and calculations required to produce high-quality EC data from the initial data and here is defined their usage specific to CAF data and at what stage they are applied. Multiple levels of data are produced for tracking purposes and as an indication of what processing has been applied in its generation. The higher the data level, the further processed the data. As of this writing there are 4 data levels (0-3) with Level 3 data being the final, high-quality, 30-minute ecosystem flux data set. Level 0 (Section 1.2) data is the raw 10Hz data. Level 1 (Section 1.3) is either the online or offline flux data prior to post-processing. Level 2 (Section 1.4) data has been post-processed to remove poor quality data but not gap-filled. Level 3 (Section 1.5) data is Level 2 data that has been gap-filled and is that last processing step accomplished here. Discussion of the soil heat-flux correction will be in Section 1.6. Meteorology data processing is described in Section 1.5. Column descriptions are separate data dictionaries.

This document does not cover the QA/QC procedures or processing of the PhenoCam data. The operating procedure for the Python scripts for the gap-filling happen are contained in a separate read-me file specific to the operation of the code. More detailed instructions and documentation for the scripts is in a separate read-me with the code.

## 1.2 10Hz Data Collection (Level 0 data)

The raw 10Hz data is stored on compact flash cards at each tower site and collected approximately once a month or in conjunction with site maintenance. The 10Hz data are not currently collected through the cellular modem but eventually will be. This data is stored in G-Drive managed by the USDA ARS (Bryan Carlson). This initial data is referred to as ‘Level 0’ data because it has had no processing done to it. The filename system for the flat, uncompressed, comma-separated, time series files is:

TOA5\_\*\*\*\*.TimeSeries\_\*\_yyyy\_mm\_dd\_HHMM.dat

Where \*\*\*\* indicates the site ID number, the second wildcard (\*) is a counting number assigned by the logger based on the number of time-series files since the logger was last shut-off, yyyy is the year, mm is a two-digit month, dd is the two-digit date, and HHMM is the hour and minute (0000 indicating midnight). The site ID values assigned by the data loggers are:

Boyd North (BN): 6506

Boyd South (BS): 6034

Cook West (CW): 6505

Cook East (CE): 6503

The raw card data (compressed) data is archived in the G-Drive. A folder exists for each site. After each card collection, a new folder is created and labeled as the date of the card collection in the format: “yyyymmdd”. Within that folder all the data on the card is copied in the uncompressed form. The file format for these files is:

\*\*\*\*.Time\_Series\_\*.dat Daily 10Hz data

\*\*\*\*.Flux\_\*.dat 30-minute online flux data calculated via EasyFlux (Section 3)

\*\*\*\*.LTAR\_Met\_\*.dat 15-minute LTAR meteorology data

With the wildcards representing the same information as in the uncompressed files. The uncompressed data is not presently stored within the G-Drive, so it is up to the end-user to decompress files as needed. For data processing (EddyPro V7.X, LiCor Biogeosciences, Lincoln, NE) and other work, the 10Hz data is downloaded, uncompressed and stored locally. Conversion from the raw, uncompressed data off the data logger is done using the Card Convert function with from Campbell Scientific (Logan, UT).

Prior to processing to Level 1 data; the time series data is checked using TS\_Check.py to fill in any gaps in the 10Hz time series due to skipped records, power issues, or split data files (card collect times). EddyPro *does not* do any time-checks within the processing and assumes the files are complete based off the start time in the filename.

## 1.3 Flux Processing (Level 1 data)

Two flux calculation programs have been used to process the 10Hz data. The settings for the processing steps for each were set to be identical per the respective program’s settings as an initial comparison (See Chapter X). The table below represents the source of the corrections and tests that are used for each flux processing program. Further explanation of each processing step and grading system is presented in Chapter X. The long-term goal is to eliminate the EddyPro step except as an occasional baseline comparison with the EasyFlux data, which is to be the primary data source for 30-minute data, reducing the amount overall data processing.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **EddyPro V7.0.4** | **EasyFlux** | **Notes** |
| **Despike** | Vickers & Mahrt (1997) | Filtered from sensor diagnostics, signal strength, and measurement range thresholds | EasyFlux did not cite a reference for the despike process. |
| **Density Corrections** | Schotanus et al., (1983) and Webb et al., (1980) | Schotanus et al., (1983), Van Dijk, (2001), and Webb et al., (1980) | EasyFlux followed van Dijk, (2001) implementation of Schotanus el al. (1983) |
| **Spectral Correction** | Massman (2000) and (Moncrieff et al. (1997) | Based on Moore, (1986) | Massman, (2000) is based on Moore, (1986) and is used in EasyFlux. |
| **Lag/Sensor separation** | Covariance maximization (Horst and Lenschow, 2009) | Covariance maximization (Aubinet et al., 2012; Horst & Lenschow, 2009) | EasyFlux manual indicates additional constraints for physically possible lag times |
| **Grading System** | 0-9; Aubinet et al. (2012) | 0-9; Aubinet et al. (2012) | Differences in grading with respect to wind direction grading system |
| **Coordinate Rotation** | Double Rotation; Wilczak et al. (2001) | Double Rotation; Tanner & Thurtell (1969) | Planar fit not utilized currently. |

### 1.3.1 Flux Data Files: Different Processing Procedures:

To facilitate research efforts regarding data quality and measurement biases, multiple processing runs of the data have been completed on data from each site for the crop year 2017 data. Three different processing schemes have been used for Level 1 data to determine their relative consistency and quality. These are:

1. EddyProV6.2.0 using Moncrieff spectral correction
2. EddyProV6.2.0 using Massman spectral correction
3. EasyFlux-dl (CR3000) run in real time on the data logger

The EasyFlux (“online”) processed data is transferred directly to the archive server as well as collected with each card collection with the 10Hz data. All the EasyFlux processing is done online for each site with site-specific values and post-processed after collection. The EddyPro (“offline”) processing was done separately every time the data cards are collected and pooled into one comma-separated file. The EasyFlux data is the same as the EddyPro data when using the Massman (2000) spectral correction within the EddyPro program (See Chapter X).

### 1.3.2 Low and High frequency CO2 data

The low frequency-based CO2 fluxes (Fclf) are being used as the “default” value for the Level 2 data within the EasyFlux code until the program is changed to the high-frequency CO2 values. The high-frequency-based CO2 fluxes are calculated using EddyPro direct from the 10Hz data. This will be updated once this switch has been made. See Russell et al. (2019) and Helbig et al., (2016) for details on the impact of the high and low frequency flux values.

### 1.3.3 Summary

In total; four 30-minute flux data sets were initially produced from each LTAR site to cover the range of processing and data calculation procedures. These include 3 through EddyPro processing and one from the online EasyFlux processing:

1. EddyPro using Moncrieff spectral correction – *low* frequency CO2
2. EddyPro using Massman spectral correction – *low* frequency CO2
3. EddyPro using Massman spectral correction – *high* frequency CO2
4. EasyFlux-DL run in real time on the data logger – *low* frequency CO2

Only data sets 3 and 4 are actively maintained. Dataset 3 is the set of record and used for full processing; dataset 4 acts as the backup and used to check towers are working correctly until the program updates occur. EddyPro set-up details are described in Chapter X. Version numbers of EddyPro are generally left off due to relatively consistent updates that do not impact the underlying correction procedures. If this is the case, the version numbers will be added as needed.

## Post-Processing (Level 2 data)

After the initial flux processing and calculations are complete (Level 0 -> Level 1), 7 QA/QC checks are performed on the flux data to further filter and flag data quality. Each of these can trigger the filtering of a turbulent flux data point individually. Each time the flags are triggered, they are recorded and indexed in time and for which flux along with tripping the Boolean for that time and flux to 'False'. Each 30-minute flux receives a bit flag to indicate the results of the QC checks. The Boolean value is returned as ‘True’ only under '0000000' conditions (See Section 1.4.2). If any of the flags are tripped, the \*\_graded data has a NaN value instead of data and needs to be gap-filled. There is a bit value for each QA/QC flag and a combined Boolean to use as a filter or mask for the data. These masks are added to the end of the datafile along with the gapped flux data. Data output after all these processes occur will be referred to as 'Level 2' data in future documentation.

### 1.4.1 Filter values used

The initial QA/QC checks are:

1. ***Bounds Check:***
   1. Fluxes (H, LE, Fc, and tau), but all data eligible. CAF bounds used:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Upper** | **Lower** |
| Sensible Heat Flux (H) | 750 W m-2 | -150 W m-2 |
| Latent Heat Flux (LE) | 750 W m-2 | -150 W m-2 |
| CO2 Flux (Fc) | 20 µmol m-2 s-1 | -60 µmol m-2 s-1 |
| Momentum Flux (Tau) | 10 | -10 |

1. ***IRGA Signal Strength:*** 
   1. Diagnostic if the IRGA window is obstructed. The reader is referred to the manual for their specific IRGA for more information. The minimum value should be >= 0.7 (or 70%). Below this, the IRGA is obstructed enough to start causing issues with the data and can bias the gas fluxes and densities (Fratini and Mauder, 2014). CAF is using 0.7 as the minimum value; EasyFlux uses 0.8 by default.
2. ***Precipitation Check:*** 
   1. Filter specifically for fluxes; eddy covariance does not work well in rain as the water will obstruct the optical measurements. The signal strength checks are a better judge of this as the rain will impact the signal strength and precipitation can persist on the window post-rainfall. Quality of the precipitation data can be relatively low depending on the type of rain gage used. Default to signal strength checks over precipitation for IRGA-based values. There is this check in the CAF processing stream for completeness.
3. ***Tower maintenance and instrument status:*** 
   1. Site-specific; includes times when towers were off or instruments removed, column headers changed, general upkeep maintenance, etc. May require modification of the code depending on specific needs or could be done separate of the base code. For CAF; half-hours are filtered out if the door\_open\_hist column is non-zero.
4. ***U-star filter:* Not used**
   1. Not needed for AmeriFlux submission – Done in REddyProc for the gap-filling. This is a somewhat contentious filter depending on the site characteristics but is not applied to the CAF flux data apart from the gap-filling processing. Does not contribute to the bit-flag described above. Listed here for completeness sake
5. ***Despike:*** 
   1. Most spikes should be found in the bounds check but still a useful check [See Aubinet et al. (2000), Starkenburg et al. (2016), Vickers & Mahrt (1997)]. Limits and bounds depend on how strict of checks are needed for the data, but the bounds check will also catch the spikes. This is another somewhat contentious but expected step (see references above); multiple methods to do this processing. Only a simple method is used within the code at the moment for the fluxes, none of the meteorology data is despiked.
6. ***Filter for quality control grades:*** 
   1. Allowable values depends on application; each flux dealt with independently [See (Aubinet et al. (2012), Foken & Wichura (1996)]. Typically filter just the highest graded fluxes (QC grade = 2) for the initial processing based on the descriptions of levels of grades. Increased stringency of this filter depends on the application of the data. Grades are specific their individual flux and half-hour and it is possible for the fluxes to have different grade values for the same time period. CAF currently filters data with a QC grade of 2 (or 7-9 depending on system) in the initial processing. These are transformed to 0, 1, and 2 to be consistent with AmeriFlux requirements. For more details see Chapter X.
7. ***Check for completeness of dataset***
   1. This checks that enough samples were recorded for the half-hour to have reliable data coverage. Currently set at 80% data coverage (14400/18000) per half-hour.
8. ***Any other checks deemed appropriate at the site-level***
   1. Includes any bias or calibration corrections, formatting issues, sensor continuity, or other required quality controls deemed appropriate by the site. Depending on the reasoning, this could be folded in with the site maintenance filters. There are a couple of these in the CAF but are hard-coded for specific periods of time as known by the site logbook. (Note: This is still being worked as January, 2021; details to follow as implemented.)

### 1.4.2 Flag Identifiers

Seven binary "bits", each representing a post-processing filter are used to keep track of which filters were triggered at each half-hour. The following format is used:

0000000: No QA/QC was triggered, data passed all checks cleanly

0000001: Insufficient sonic samples flag

0000010: Insufficient IRGA samples flag

0000100: Turbulence threshold flag

0001000: Door open flag

0010000: QA/QC Grading system flag

0100000: Precipitation flag

1000000: Hard Limit flag

As an example, “0110000” is a data point that tripped both the precipitation and QA/QC grade flag and “0000100” is a data point that tripped the turbulence threshold flag. Only one filter is required to remove a data point, but each half-hour can trip multiple flags hence this flagging system. An update to this data is prepared each month or as requested. Each update will be added to the previous months’ data through the crop year (October 1-September 30). A final year-long QA/QC’ed 30-minute flux data file will be done at the end of each crop year (October 1) along with gap-filling for Level 3 data.

### 1.4.3 Column Dropping

Several columns from the EFlux data are dropped during the Level 2 processing to reduce the number of columns for Level 3 processing and to shorten to the primary data outputs to the most useful variables for various workers of the data. A full, unfiltered list is available as needed.

### 1.4.4 AmeriFlux Data Processing

Data processed to the AmeriFlux standard are output and sent to their database at least once per year. AmeriFlux has a specific column and timestamp convention (<http://ameriflux.lbl.gov/data/how-to-uploaddownload-data/>) which will be produced during the full data processing. They require base data **not** be gapfilled or u-star filtered but all other processing steps will be completed from the QA/QC list in Section 1.4.1. Gap-filled data may be uploaded but requires special handling and column names. The column renaming takes place in a separate function from the main QA/QC processing script since it is only used sparingly over the course of the year and the other data is processed more frequently. The max allowable grade can be shifted as needed but the u-star value needs to stay at zero to meet the requirements. Otherwise the same QC checks are done on the flux data. This step is done after all the QA/QC checks to ensure only higher quality data are passed to the AmeriFlux database.

All four towers are registered with AmeriFlux and are assigned the following site IDs:

[Cook East: US-CF1](https://ameriflux.lbl.gov/sites/siteinfo/US-CF1)

[Cook West: US-CF2](https://ameriflux.lbl.gov/sites/siteinfo/US-CF2)

[Boyd North: US-CF3](https://ameriflux.lbl.gov/sites/siteinfo/US-CF3)

[Boyd South: US-CF4](https://ameriflux.lbl.gov/sites/siteinfo/US-CF4)

## Meteorology Data Quality Control and Gap-Filling

The meteorological data quality control (MD-QC) process is based on the Meteorological Assimilation Data Ingest System (MADIS) quality control checks used by NCEP for surface data. Unless and until guidance on MD-QC is produced by the LTAR national network, the local MD-QC will follow the described process. Data must pass through two types of checks to be considered “valid” data:

1. Validity or “hard-limit” check – Values within physically reasonable norms
2. Internal consistency checks – check against unreasonable changes every 30-minutes and if the daily averaged data changes every day (variable dependent)
   1. Wind direction and wind speed are not checked if the daily average changes day to day
   2. Precipitation is not checked if it changes day-to-day since there can be multiple days of zero precipitation in a row. Precipitation has two variable specific checks.
3. Variable specific checks:
   1. RH: Check if over 100%; reduce value to 100% if true.
   2. Precipitation: Check if RH is over 90% (if RH is available) and check if temperature is above freezing (if temperature available).
   3. PAR is not checked if the half-hour difference is zero because overnight the value is zero.

The bounds used are based on the climatological high and low values for the Pullman/Palouse region and on physically possible values (e.g., wind direction between 0-360 degrees). Prior to the validity checks, the measured atmospheric pressure is converted to mean sea level pressure (MSLP) via the hypsometric equation:

|  |  |
| --- | --- |
|  | (1) |

Where *P0* is the MSLP, *Pz* is the measured pressure at altitude *z* (elevation above sea level in kilometers) and *H* is the scale height (in kilometers) defined as:

|  |  |
| --- | --- |
|  | (2) |

Where R is the universal gas constant (8.314 J mol-1 K-1), T is the temperature in Kelvin, *g* is the gravitational constant (9.81 m s-1), and is the molecular weight of air (0.029 kg mol-1).

Each data point is checked against a specified range for that variable:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Upper Limit** | **Lower Limit** |
| Relative humidity (RH) | 100% | 0% |
| MSLP | 110 kPa | 80 kPa |
| Pressure | 100 kPa | 70 kPa |
| Air Temperature | 50 C (323 K) | -40 C (223K) |
| Wind speed | 125 m s-1 | 0 m s-1 |
| Wind direction | 360 degrees | 0 degrees |
| Precipitation | 0 mm | 400 mm |
| PAR | 0 μmol s-1 m-2 | 5000 μmol s-1 m-2 |
| Net Radiation | 1500 W m-2 | -500 W m-2 |

If data is not within the specified range, it is flagged with a Boolean, left in the original data column but filtered in the output data after the remaining data checks. The internal consistency checks calculate the difference between the current and previous timestep to check that the data is consistent in time with itself (i.e., no major, unexplained shifts). Not every variable is run through a consistency check:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Max Allowed Change per 30 minutes** | **Checked for daily change?** |
| Atmospheric Pressure | |3.1 kPa| | Yes |
| MSLP | |3.1 kPa| | Yes |
| Air Temperature | |15 C| | Yes |
| Relative Humidity | |50%| | Yes |
| Wind Speed | |15 ms-1| | No |
| Wind Direction | Any | No |
| PAR | |1500 μmol s-1 m-2| | Yes |
| Net Radiation | |500 W m-2| | Yes |
| Precipitation | Any | No |

Like the validity checks, the data is flagged within a separate Boolean column which is used to filter the suspect points to a final output data set. The output dataset includes all the individual Booleans, False indicating their removal from the data set along with the pre- and post-filtered data. At some point, a knowledgeable person needs to manually check each data column to approve the removal of the suspect data before the final data is archived. This needs to be done at least once a year. No meteorological data is gap-filled at this time (January 2021).

## 1.6 Soil Data QC Checks

The soil data for both the profile and near-surface data are checked to be within physical bounds and if the soil temperature at the respective depth is above freezing. With the change in the headers in the program from the original; a combining of the header names is used in the processing to collapse the columns into one column header. This is a vestige of older data and left in the code. Also, with the constant removal and reinstallation of the soil sensors, a listing of the timings for when the Cook sites had the surface (5cm, 15cm, and 30cm) sensors removed is used to filter out any data recorded during these times. The check against soil temperature occurs prior to the bounds check. Since the same sensor is used at each depth and the measurement is repeated, only one instance of each variable is going to be reported. The currently used bounds are:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Upper** | **Lower** |
| tdr315\_wc\_Avg\_L2 | 75% | 0% |
| tdr315\_tmpr\_Avg\_L2 | 50 C | -40 C |
| tdr315\_E\_Avg\_L2 | 25 | 0 |
| tdr315\_bulkEC\_Avg\_L2 | 500 | 0 |
| tdr315\_poreEC\_Avg\_L2 | 10000 | 0 |

This QC outputs three Boolean values:

1. \*\_tdr315\_tmpr\_Avg(\*)\_gt\_Zero: True is the value is above zero and data passes to L2
2. \*\_tdr315\_tmpr\_Avg(\*)\_L2\_Upper: True if the L2 value is lower than the upper bound
3. \*\_tdr315\_tmpr\_Avg(\*)\_L2\_Lower: True if the L2 value is higher than the lower bound

### 1.7 Flux Data Gap-Filling

Gap-filled data are most used for carbon, water, and other budget applications. It does impart another level of uncertainty into the final values (Soloway et al., 2017) and being derived values, they are not used for some applications since they are not directly measured values. The gap-filling process can also partitions the carbon flux into gross primary production and ecosystem respiration (Desai et al., 2008; Moffat et al., 2007). Some use home-built gap-filling code but there is a slow transition to REddyProc (Wutzler et al., 2018). CAF uses the REddyProc code with the nighttime respiration model. This is a set of R code that takes post-processed flux data and performs some QC checks before gapfilling the provided data. Only eight variables are needed and the format for the input data has to be exact to the specifications or it will not work and/or will error out. Without certain variables, different parts of the code will not be used; see links below for details.

1. [Variables required are:](https://www.bgc-jena.mpg.de/bgi/index.php/Services/REddyProcWebDataFormat)
   1. H, LE, NEE, RG (incoming solar radiation), Tair, Ustar, VPD, rH
   2. Time needs to be in three columns:
      1. 1) YYYY (year), 2) DDD (Day of year), and 3) HH.H (Decimal hour, e.g., 12:00 = 12.0 and 12:30 = 12.5).
   3. The site latitude and longitude coordinates are also needed
2. [Output format](https://www.bgc-jena.mpg.de/bgi/index.php/Services/REddyProcWebFormats) is provided on the REddyProc webpage and not repeated here.
3. Descriptions of the processing methodologies are provided [here](https://www.bgc-jena.mpg.de/bgi/index.php/Services/REddyProcWebMethod) and in Wutzler et al. (2018).

### 1.7.3 Formatting Level 3 Data

### *1.7.4 Level 3 Data Naming Convention*

*Level 3 data will have the naming convention as:*

*CAF\_LTAR\_L3\_sitename\_­yyyy\_mm\_dd\_to\_yyyy\_mm\_dd*

*Where the first ­yyyy\_mm\_dd is the start date of the gap-filling and second ­yyyy\_mm\_dd is the end date. Full crop year L3 data has the convention as:*

*CAF\_LTAR\_L3\_sitename\_CY­yyyy*

*Preliminary data will contain the suffix “Preliminary” to differentiate from more final data. This data has been processed but not inspected by an end user, does not represent a final, complete data set, or is for testing purposes only, not publication/research.*

## *Transpiration and Evaporation Partitioning*

*The evaporation was estimated as the difference between ET and T (E =ET-T). The water values are all converted to mm per 30 minutes prior to output. ET is converted directly from LE via:*

|  |  |
| --- | --- |
|  | *(6)* |

*Where Lv = 2501000-2370\*Tair, is the latent heat of vaporization based on temperature (Tair). ET, T, and E all have units of mm per 30mins so taking a daily sum will generate mm day-1.*

## *Soil heat flux and soil storage*

*The soil heat flux being located 5 cm below the soil surface requires the calculation of the soil heat storage in the layer between it and the surface. This is done by determining the heat transfer through the soil or the heat capacity of the layer of soil each time step (t) using measurements of temperature (T) and soil moisture (ϴ) in this layer (z). Typically, the heat capacity () determination is done through a calorimetry process, using the change in temperature and moisture over time, the methodology used here follows that used in Gao et al. (2017) and Russell et al. (2015). Briefly; the equation for the surface soil heat flux () combines the heat stored in the upper layer () and the measured (Gz(t)):*

|  |  |
| --- | --- |
|  | *(7)* |

*Where:*

|  |  |
| --- | --- |
|  | *(8)* |

*Where is the heat capacity of the soil at time t, is the soil mineral fraction (ratio of bulk density to rho\_m), is the density of the soil minerals (2650 g m-3), is the volumetric heat capacity of the soil minerals, is the layer averaged volumetric water content as a percent, is the density of water (1000 kgm-3), and is the volumetric heat capacity of water (4180). Both the Gsfc and are output from the calculation.*

## Contact Information

LTAR CAF Site PI: Dave Huggins: [David.Huggins@ars.usda.gov](mailto:David.Huggins@ars.usda.gov)

Eric Russell eric.s.russell@wsu.edu

Data Management: Bryan Carlson: bryan.carlson@ars.usda.gov

Tower/Field Management: Patrick O’Keeffe: pokeeffe@wsu.edu

Ian Leslie: Ian.Leslie@ars.usda.gov

Caven Halsey

# Chapter 2:

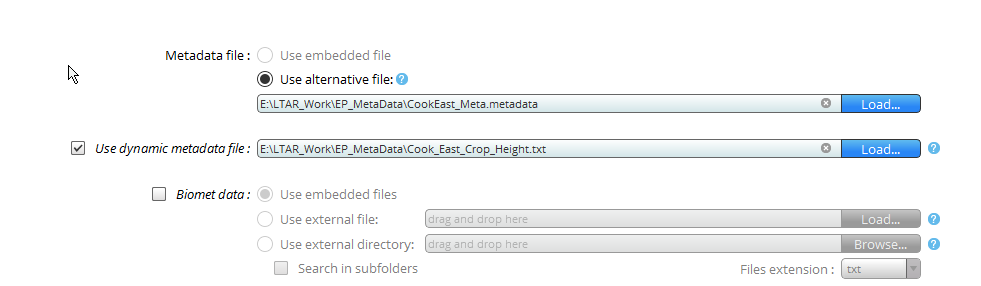
## 2.1 Introduction

*Data is collected from the towers directly on the compact flash cards and brought back to the office. This is done approximately every 4 weeks at all sites. The compressed raw data (time series, flux, and Met files) are all uploaded to the LAR ftp server and anywhere else as required under the specified site name. Each card collection is placed in a new folder that is named as ‘yyyymmdd’ based on the day of collection.*

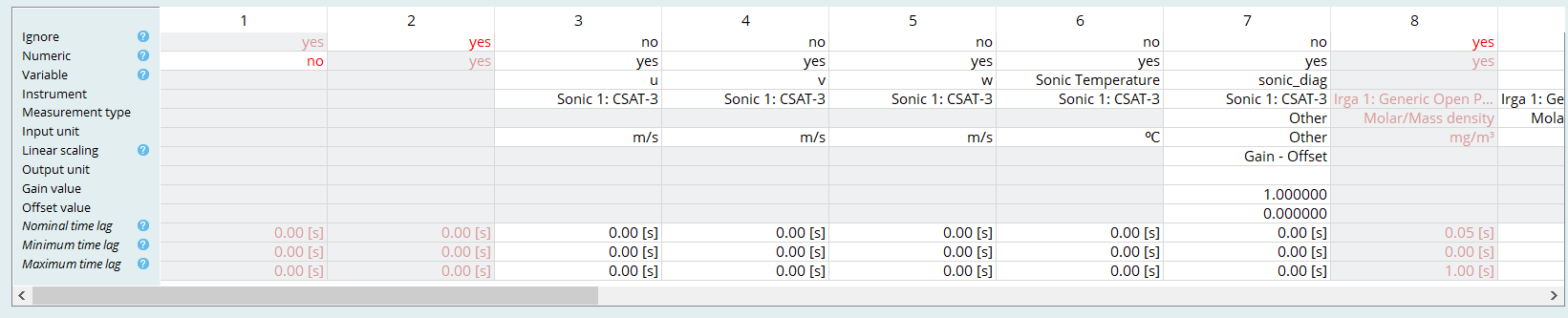
All the data needs to be converted from the compressed version to the flat file to run through EddyPro and the processing scripts. These are collected in whatever directory the user has for the data. EddyPro is used to run the high frequency data for the high-frequency CO2 data until the logger code is updated to handle high-frequency data. The other fluxes have been shown to be near-equivalent between the two processing programs so these can be used interchangeably though consistency is useful (Chapter 4). A metadata file exists for each site for EddyPro that has the specific geographic information for each LTAR site. These files also contain the file structure for the time series and are set-up for the specific corrections.

## 2.2 Step-by-step processing for EddyPro

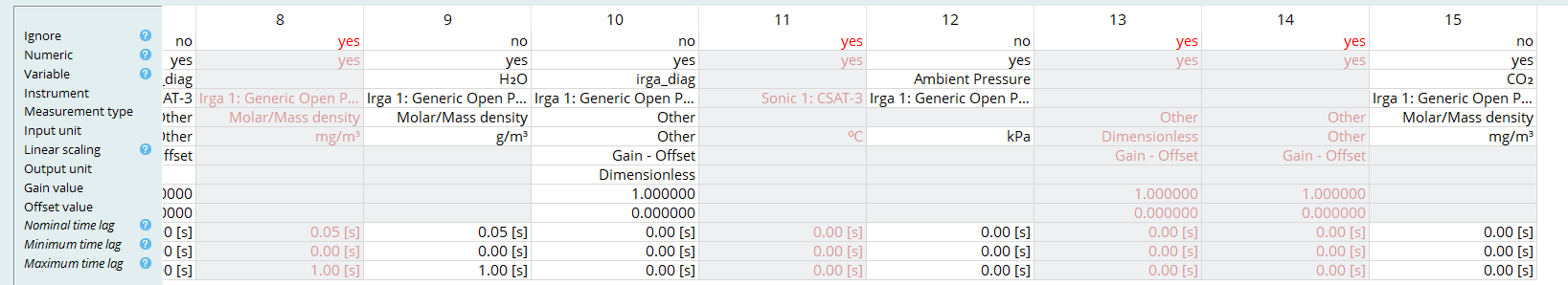
1. Update timeseries directory that houses the files with the new data.
2. Update the crop-height data file if the crop height has changed since the last run of EddyPro.
3. Update the crop-height metadata file (Project Creation) if crop height has changed over time:



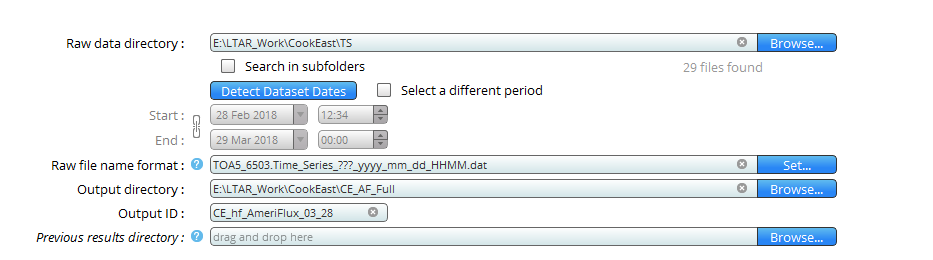
1. Load appropriate metadata file into EPro
   1. Double check that the columns are correctly assigned and labeled, and the CO2, H2O, and ambient pressure are correctly selected:



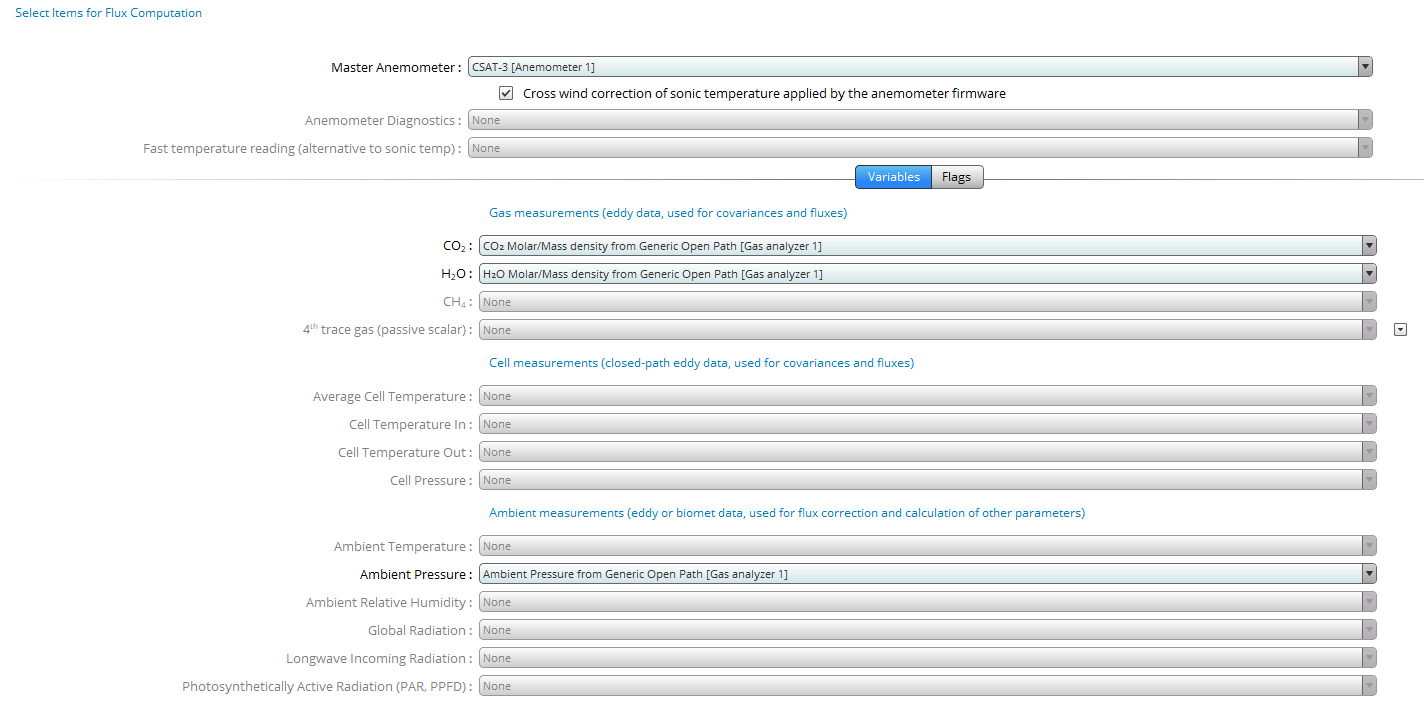
And



1. Update the input directories (“Raw data directory”) to include the latest data and the raw file format to match the newest timeseries data. Beware that the logger puts a counting number into the timeseries file.
   1. Change output directory to appropriate directory
   2. Update the filename format and raw data directory
   3. Set output ID as see fit.



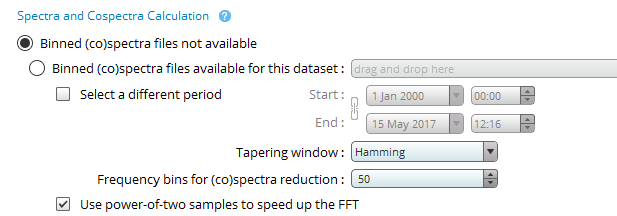
* 1. Make sure all the appropriate instruments are selected for the flux calculation:



1. Under advanced settings double check the following; all other values should be left as their default value:
   1. Processing Options:
      1. Flagging policy should be Foken (2003) (1 to 9 system)
      2. Footprint method: Kljun et al., 2004



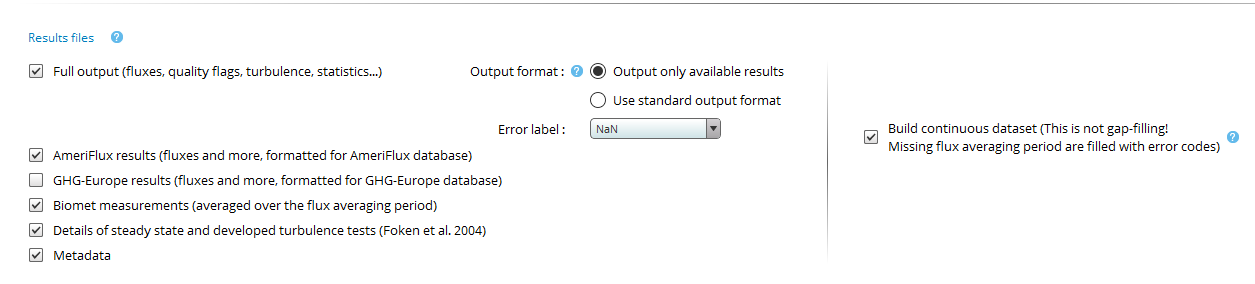
* 1. Statistical Analysis: Leave alone
  2. Spectral Analysis and Corrections:
     1. Tapering window: Hamming



* + 1. Correction of low-pass filtering effects: Massman (2000, 2001) – Fully analytic



1. Under Output Files check “Full Output”, use the error label ‘NaN’, and check “Build continuous dataset”



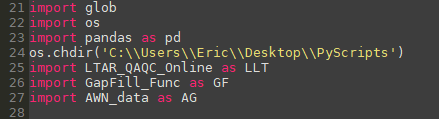
1. Click “Advanced Mode” and wait for data to process. Time depends on how much data is being processed and computer specs.

## ~~2.3 Python L3 Processing and Gap-Filling Script~~

~~The gap-filling code requires several input file names and other variables to be defined. The functions called do work as standalone functions and do not require their structure to be changed unless a fundamental change is needed. None of the functions create an output file, this is left to the user to define as they see fit.~~

### ~~2.3.1 Import Libraries~~

~~Multiple libraries need to be imported into the main driver; all the functions contain the required library imports. The Driver libraries are all already listed at the top of the script and~~ ***~~do not~~*** ~~need to be changed or the script will error out for lack of functionality or improper calls. The directory that contains the non-Python native scripts (LTAR\_QCQA\_Online, GapFill\_Func, and AWN\_data) does need to be updated to the appropriate directory.~~

~~~~

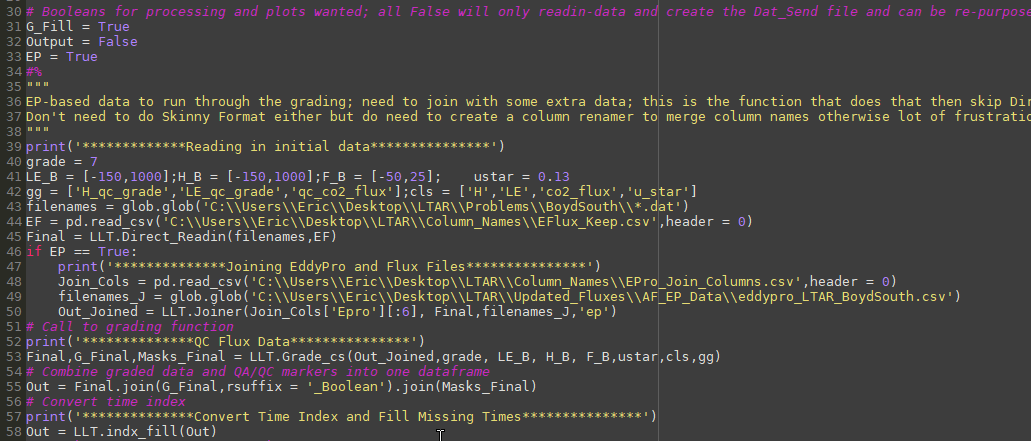
### ~~2.3.2 Variable Definitions~~

#### ~~2.3.2.1 Boolean Values~~

1. ~~Three Boolean values at the top of the script need to be set to either TRUE or FALSE depending on what is needed:~~
   1. ~~G\_Fill: If TRUE, will call the gap-filling script to run~~
   2. ~~Output: If TRUE will call the output functions and output datasets~~
   3. ~~EP: If TRUE, will call the EddyPro loop to combine with the EFlux dataset.~~

#### ~~2.3.2.2 Inputs~~

1. ~~In the driver script (LTAR\_Level\_3\_Driver) the following input directories need to be updated. Not all need to change with each run for the different sites. The~~ **~~bolded~~** ~~and~~ **~~underlined~~** ~~variables do need to change per site-run as they are site-specific (See figure)~~
   1. **~~filenames~~**~~: Directory that contains the flux filenames from the logger-produced data (EFlux)~~
   2. ~~EF: Columns to keep from EFlux dataset, read in as a \*.csv file~~
   3. ~~Join\_Cols: If needed, columns from a separate dataset to join to the main dataframe, currently a set of data from the EPro output~~
   4. **~~filenames\_J~~**~~: Directory that contains the files with the columns that are to be joined with the main data frame~~
   5. ~~Tracker\_cols: List of columns that are to be used to track the happenings of the Met and Gap-filling processing, \*.csv file~~
   6. **~~g\_files~~**~~: List of directories that contain the files for the meteorology gap-filling if using a nearby EC tower site. Only needs to change if the input files need to change.~~
   7. ~~AN: Directory that contains AgWeatherNet datafile if needed for gap-filling of meteorology data.~~

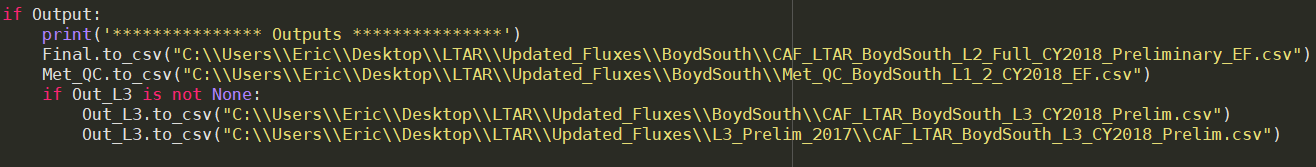
~~~~

#### ~~2.3.2.3 Limit and Processing values~~

1. ~~Several limit values also need to be updated or checked depending on the user (See figure above). These do not need to be changed each pass through the script:~~
   1. ~~grade: Maximum allowed QC grade based on the grading system~~
   2. ~~LE\_B: Lower (first) and upper (second) bound for a hard-limit check on the latent heat flux.~~
   3. ~~H\_B: Lower (first) and upper (second) bound for a hard-limit check on the sensible heat flux.~~
   4. ~~F\_B: Lower (first) and upper (second) bound for a hard-limit check on the CO2 flux~~
   5. ~~ustar: Minimum value for u-star as per the u-start threshold. If set to zero, then no values will be filtered via u-star~~
   6. ~~gg: Names of the columns that house the QC grades from the end combined flux data~~
   7. ~~cls: Name of the columns that contain the raw flux data from the combined flux data~~
   8. **~~Site~~**~~: Names of the sites matching the g\_files used for the meteorology gap-filling for tracking purposes.~~

#### ~~2.3.2.4 Output variables~~

1. ~~The variables that can be output are (See figure below). These need to change with each site that is run through the script:~~
   1. ~~Out\_L3: Level 3 gap-filled 30-minute data with the tracker values for the QC processing~~
   2. ~~Met\_QC: Quality controlled meteorology data at input frequency with metadata columns; no flux data contained in this output.~~
   3. ~~Final: Level 2 flux data with appropriate tracker columns.~~

~~~~

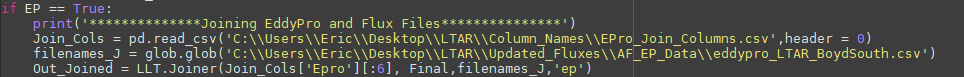
1. ~~Some minor checks are needed in the gap-filling function with respect to the user’s choice of values to be used.~~
   1. ~~Distance and multiplier for the despike function~~
   2. ~~Acceptable number of data points for the Rref and E0 derivation~~
      1. ~~Addition of any upper and/or lower limit to these values~~
   3. ~~Upper and lower limits to the GPP parameters, currently based off the REACCH values~~
   4. ~~Number of days on either side of missing value to use for the MDV gap-filling algorithm, currently set at 10 days~~

### ~~2.3.3 Order of Function Calls and Definitions~~

1. ~~Order of function calls and inputs required for each:~~
   1. *~~Final~~* ~~= LLT.Direct\_Readin(~~*~~filenames, EF~~*~~)~~ 
      1. *~~Final~~* ~~Combined dataframe from all files read-in and columns kept~~
      2. *~~filenames~~*~~: Glob list of filenames that contain the flux data to be processed~~
      3. *~~EF~~*~~: List of columns to keep within the output file; any column in the input data that does not match~~ ***~~exactly~~*** ~~with this list is dropped~~

~~~~

* 1. *~~Out\_Joined~~* ~~= LLT.Joiner(~~*~~Join\_Cols,Final, filenames\_J,’ep’~~*~~)~~
     1. *~~Out\_Joined~~*~~: Dataframe output that combined the extra columns (Join\_Cols) with the base data set (Final)~~
     2. *~~Join\_Cols~~*~~: List of columns to join to the base data set (Final) from a second data set (filenames\_J)~~
     3. *~~Final~~*~~: Same as Final from previous function~~
     4. *~~filenames\_J~~*~~: Glob list of files to be read-in and contain the Join\_Cols to be combined with the base dataset (Final)~~
     5. ~~‘~~*~~ep’~~*~~: denotes input file is an EddyPro file; can be ‘cs’ for a Campbell Sci file from EFlux.~~

~~~~

* 1. *~~Final~~*~~,~~ *~~G\_Final, Masks~~* ~~= LLT.Grade\_cs(~~*~~Out\_Joined, grade, LE\_B, H\_B, F\_B, ustar, cls, gg~~*~~):~~
     1. *~~Final~~*~~: Graded data output, overwrite the Final variable from above~~
     2. *~~G\_Final~~*~~: Boolean of the total grade for each flux value~~
     3. *~~Masks\_Final~~*~~: Individual Booleans for each QC check in the grade function~~
     4. *~~Out\_Joined~~*~~: Joined dataframe from the previous step~~
     5. *~~grade~~*~~: Maximum allowed QC grade for each flux~~
     6. *~~LE\_B~~*~~: Upper and lower bound for latent heat flu~~
     7. *~~H\_B~~*~~: Same as LE\_B but for sensible heat~~
     8. *~~F\_B~~*~~: Same as LE\_B but for CO2 flu~~
     9. *~~ustar~~*~~: Threshold value for the friction velocity (ms-1)~~
     10. *~~cls~~*~~: Column names for the flux and ustar values, see above~~
     11. *~~gg~~*~~: Names of the columns that contain the QC grades for the flues~~

~~~~

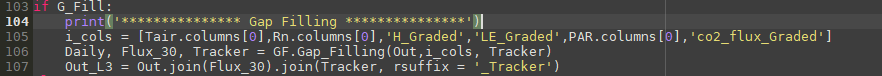
* 1. *~~Met\_QC, qn3~~* ~~= LLT.Met\_QAQC(~~*~~RH~~*~~,~~ *~~P~~*~~,~~ *~~Tair~~*~~,~~ *~~WS~~*~~,~~ *~~WD~~*~~,~~ *~~Precip~~*~~,~~ *~~PAR~~*~~,~~ *~~Rn~~*~~,~~ *~~z~~*~~)~~
     1. *~~Met\_QC~~*~~: Output dataframe for the QCed meteorology dataframe~~
     2. *~~qn3~~*~~: Placeholder for the 3-hour average of the meteorology data~~
     3. *~~RH~~*~~: relatively humidity data column in percent (0-100)~~
     4. *~~P~~*~~: Atm. pressure data column in kiloPascals (kPa)~~
     5. *~~Tair~~*~~: Ambient air temperature data column (degrees Celsius)~~
     6. *~~WS~~*~~: Wind speed data column (m/s)~~
     7. *~~WD~~*~~: Wind direction data column (degrees)~~
     8. *~~PAR~~*~~: PAR data column (μmol m-2 s-1)~~
     9. *~~Rn~~*~~: Net radiation data column (W m-2)~~
     10. *~~z~~*~~: Elevation site is above sea level (kilometers)~~

~~~~

* 1. *~~Tracker~~*~~,~~ *~~Var\_Name~~* ~~= LLT.Met\_Track(~~*~~Data\_Name~~*~~, AWN,~~ *~~Var\_Name.columns[0], col\_name, site,cl, interpolate~~*~~,~~ *~~first~~*~~)~~
     1. ~~Tracker:~~
     2. *~~Var\_Name:~~* ~~Name of dataframe to assign gap-filled meteorology data to~~
     3. *~~Data\_Name:~~* ~~Dataframe of data to be gap-filled, typically same as~~ *~~Var\_Name~~*
     4. ~~AWN: Dataframe to use for the gap-filling process~~
     5. *~~Data\_Name.columns[0]:~~* ~~Name of data column for~~ *~~Data\_Name~~*
     6. *~~col\_name:~~* ~~Column name of variable to be used for gap-fill~~ *~~Data\_Name~~*
     7. *~~site:~~* ~~Undefined string for where the gap-filling data came from~~
     8. *~~cl~~*~~: List of column names for the Tracker variable to assign to~~
     9. *~~interpolate~~*~~: If 1 then will interpolate between any missing data-points not gap-filled from AWN data~~
     10. *~~First~~*~~: Boolean to indicate if first pass through the data gap-filling routine if multiple sites are used. If only one site is used, set to True.~~

~~~~

* 1. *~~Daily~~*~~,~~ *~~Flux\_30, Tracker~~* ~~= GF.Gap\_Filling(~~*~~Out,i\_cols,Tracker~~*~~)~~
     1. *~~Daily~~*~~: Output at daily time frequency of the gap-filled flux data~~
     2. *~~Flux\_30~~*~~: Output of 30-minute gap-filled data~~
     3. *~~Out~~*~~: Dataframe that contains the needed data for the gap-filling process~~
     4. *~~i\_cols~~*~~: Name of the columns needed for the gap-filling. Columns required (in order) are:~~
        1. *~~Tair~~*
        2. *~~Rn~~*
        3. *~~QCed H~~*
        4. *~~QCed LE~~*
        5. *~~PAR~~*
        6. *~~QCed~~* ~~Fc~~
     5. ~~Tracker: Tracker for the QC trace, likely to go away in future versions~~

~~~~

## ~~2.4 Step-by-step Processing:~~

~~1. Update Booleans values; all TRUE for typical processing (See 3.2.1)~~

~~~~

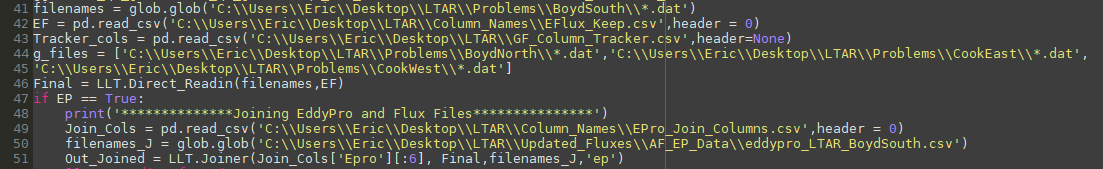
~~2. Update grade, flux hard-limit, and u-star threshold values (See 3.2.3)~~

~~~~

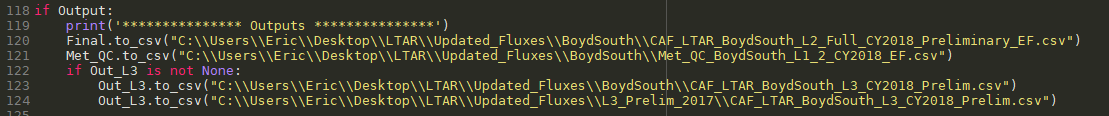
~~3. Update flux and grade column names (See 3.2.3)~~

~~~~

~~4. Update input directories and filenames (See 3.2.2)~~

~~~~

~~5. Update output filenames and values (See 3.2.4)~~

~~~~

~~6. Double check all the function inputs are correct and match with the data column names and expected units. (See 3.3)~~

~~7. Click run on the driver and wait for processing to happen or fix errors as needed.~~

## 2.5. Author comments on Processing Procedure:

I’ve tried to eliminate as much ‘guess-work’ in the set-up as possible between this procedure and the overall SOP document, hence all the detail and screenshots of the different snippets of code and processing steps. I’m still (slowly) working on generalizing everything so there are less hard-coded values in some of the functions and more ability for the user to change things as they see fit. This is a slow process because changing one hard-code value in the function means I must trace it back to the function call and figure out the best way to pass it through (multiple individual values or array of multiple values in a specific order). I’ve put some exit conditions and errors within functions as I thought needed but haven’t finished this and Python is likely to error out as well if files don’t track through.

# Chapter 3. Summary of Flux Corrections for CAF LTAR at Washington State University

## 3.1. Introduction

This chapter provides a summary of the major corrections used in the EasyFlux CR3000 data logger program, (Campbell Scientific), EddyPro data processing (LiCor Biogenosciences), TK3 (Mauder and Foken, 2011), NEON’s flux processing R-package, edd4R (Metzger et al., 2017), and listed in a number of flux network methodologies and books [e.g., *Aubinet et al.*, 2000, 2012; *Foken*, 2008; *Mauder et al.*, 2008]. Procedures outlined below are the current used “best practices” in the eddy covariance/eddy flux community based on the procedures described in the above programs’ documentation. More recent iterations of these corrections and proposed changes in methods that are still in their infancy and being vetted are not included. The attempt here is provide background information for users less familiar with eddy covariance fluxes and their processing to use the data appropriately and understand (at least conceptually) the processes that produces fluxes. Further information, discussion, and details of each method can be found in the reference list at the end of this document. *This description is not meant to be a flow chart for flux processing or define the full output of data produced, but to describe the major processing steps to calculate latent, sensible, and carbon dioxide fluxes from the raw time series data.* The order of the list is the order in which these corrections are traditionally applied. Sensor-specific corrections (e.g. cross-wind for different anemometers, transducer shadowing, cross-sensitivity for krypton hygrometers) are not described here.

## 3.2. Data filtering and Pre-Processing

The first step in most data processing is to filter “bad” data based on hard limits of physically possible data points. Possible reasons for bad data include: sufficiently blocked transducers on the gas analyzer and sonic from dust, dirt, ice, or other debris and electronic noise. There are three ways to accomplish this, each approaching the problem in a different manner: despike (Vickers and Mahrt, 1997), instrument diagnostic flags, and physical hard limits. Specifics are often left to the user as each flux site and instrument has its own set of acceptable limits.

### 3.2.1 Despiking

A despike algorithm checks the data against a running mean and standard deviation filter to identify and remove electronic noise or otherwise bad data (Starkenburg et al., 2016; Vickers and Mahrt, 1997). Despike methods traditionally work by taking the mean (μ) and standard deviation (σ) for a set data window then checking each point within that window against μ±xσ, where x is a multiplier, commonly between 3 and 4.5. If multiple iterations are completed, the multiplier may be lowered incrementally as more spikes are removed. If a certain number of consecutive spikes are detected (varies based on program, typically more than 3 or 4), then the spikes will pass-through without being flagged or eliminated. For single spikes, some methods will replace the data with the mean of the two neighboring values; this is not done with more than 1 spike.

### 3.2.2 Diagnostic Flags

The diagnostic flags are based off the diagnostic codes produced by the individual sensors. The codes’ structures, respective flags, and triggers are specific to each sensor and manufacturer. One example is the signal strength value produced by infrared gas analyzers. A minimum good value is allowed but below a certain threshold, the window can be considered sufficiently blocked that the data (even if reasonable), needs to be filtered out. From the LI7500A instruction manual (Li-Cor Inc): “There is no absolute guideline for good or bad signal strength, but 100% is very good and 0% is very bad.” The sonic anemometer has its own set of diagnostic codes. To decipher the meaning of these codes, one should consult the relevant instruction manual.

### 3.2.3 Hard-limit filter

Filtering due to hard limits is often not reported in the literature and are lumped under “despiking” the data set. All data should go through a “sanity test” to make sure the data falls within acceptable and physical limits specific to the field site and measurement limits of the instruments. The data filtered or removed via all these methods is typically counted or reported as a percent for each flux averaging period or data block. For the time series data; often this is done based on the sensors’ maximum/minimum measurement thresholds as reported in their instruction manuals.

## 3.3. Coordinate Rotation

Coordinate rotation is accomplished via either double rotation or planar fit method (Tanner and Thurtell, 1969; Wilczak et al., 2001). A third method, triple rotation, is not used. The effect of the coordinate rotation is to put the mean wind into a single coordinate (the x-direction, or *u*) and zero the mean vertical velocity (the z-direction, or *w*) and y-wind (*v*, a.k.a., the cross-wind component) direction. The variability in the wind speeds remain similar between pre- and post-rotated values. The goal of a coordinate rotation is to level the sonic to the underlying terrain (rotate wind into natural streamlines) and correct for deficiencies in the leveling of the sonic during installation (even a relatively minor misalignment can cause a bias in the scalar fluxes, tolerance is to 0.1 degree from level) (Kaimal and Haugen, 1969; Lee et al., 2005; Wilczak et al., 2001). Leveling is done relative to gravity, not to the underlying slope (Sun, 2007) Coordinate rotation also removes cross-contamination of the 3-D winds and creates three orthogonal wind vectors. The rotated wind data is used in the calculation of the turbulent fluxes.

### 3.3.1. Double Rotation

The double rotation method (DR) has been used since the 1960’s (Kaimal and Haugen, 1969) and works best under more ideal sites and conditions. At more complex sites (e.g. hilly terrain like CAF), and under non-ideal conditions (stable, unstable), its usefulness is reduced. Even so, it is still widely utilized in flux work, in part for its relative simplicity. The advantage of DR is that it does not require any *a priori* information on rotation angles meaning it can be run in real time on a data logger without any user-based inputs. DR operates on each data block independently.

DR is completed through two rotations; the first as:

|  |  |
| --- | --- |
|  | (A.1a) |
|  | (A.1b) |
|  | (A.1c) |

Where *u, v,* and *w* are the measured wind components at 10Hz, subscript *‘t’* are the transitional wind components and , where the overbars represent the 30-minute mean of the *u* and *v* wind components. The second rotation is done as:

|  |  |
| --- | --- |
|  | (A.2a) |
|  | (A.2b) |
|  | (A.2c) |

Where , , and are the final rotated wind coordinates and .

### 3.3.2. Planar Fit Method

The planar fit method (PFM) was put forth to address the deficiencies in the double rotation scheme, namely that both the local terrain and sonic are not level to each other or to gravity [*Sun*, 2007]. The PFM places the wind coordinates into a 3-D orthogonal coordinate system based on the mean streamline coordinate system. To do this, the x-axis is rotated to be parallel with the mean local wind; the z-axis is then placed in an axis vertically orthogonal to the x-axis, and the y-axis is placed in a plane such that a right-handed coordinate systems is defined (Wilczak et al., 2001). While the mean *w* and *v* winds are zero over the averaging period used, the variation within these directions remains the same as before the rotation.

The downside for this method is the rotation angles need to be calculated independent of the actual rotation calculations because they require multiple weeks of data to settle. If the sonic anemometer is moved for any reason then new rotation angles need to be calculated for the new positioning, again requiring multiple weeks of data to generate. Once generated, they can be applied as needed to new data.

The PFM can be broken into a wind-direction based system or used across all wind direction (global rotation). There is no strong consensus on which rotation method (DR, global PFM, or sector-wise PFM) is best over different terrains and land covers. (Oldroyd et al., 2015; Shimizu, 2015; Stiperski and Rotach, 2015; Yuan et al., 2011, 2007). The drawback of sector-wise PFM is it requires rotation coefficients in each the defined wind-direction sectors to be empirically derived. That is, unlike DR, PFM requires a sufficient length of time for the rotation coefficients to converge (approximately 1 month) and assumes the sonic has not been moved in the interim. As such, calculating these values in real time will need sufficient time to reach consistent values.

## 3.4. Spatial Sensor Separation

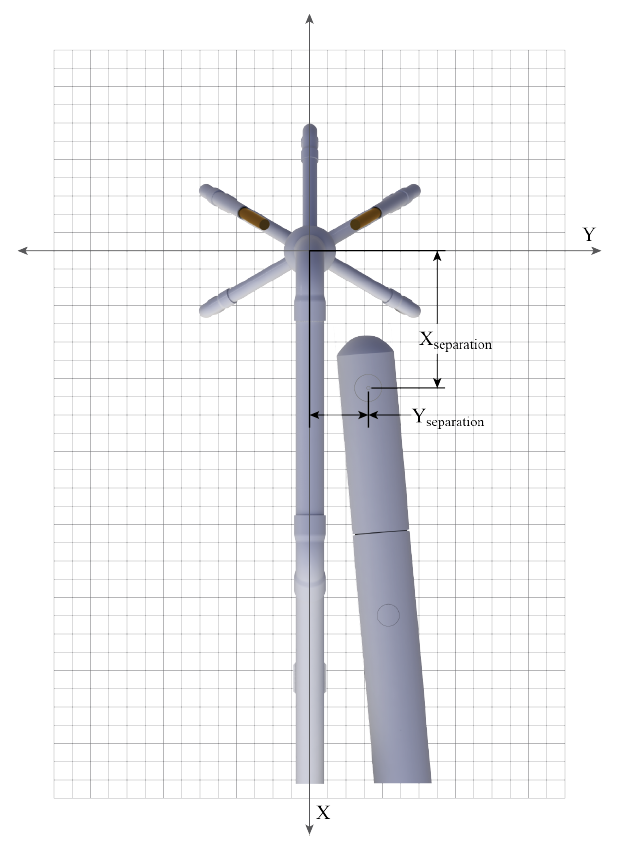


Figure D-2 from EasyFlux Manual, Campbell Scientific

The distance between the sonic anemometer and IRGA combined with the frequency of the measurement introduces a “lag” between the measurement of the same volume of air. The distance between the center of the path-length for the two sensors are parted into the x and y directions or defined as along (x) and across (y)-wind directions (See Figure). The end effect of the lag-time is a decrease in the magnitude of the covariances between the two sensors. Lag effects in the x-direction is removeable via maximizing the covariance through a cross-covariance maximization process (Horst and Lenschow, 2009). Lag effects in the across-wind direction are resolved through a frequency transfer function as part of the total transfer function for spectral corrections (Aubinet et al., 2012) (Section 5). Integrated sensors (e.g., IRGASON from Campbell Scientific) do not require a lag correction because they measure the same volume of air. Trace gas measurements with inlet tubes require a more rigorous definition and treatment of the spatial and temporal separation.

## 3.5. Frequency corrections

Frequency corrections are required because the overall measurement system acts as a filter for both high and low frequency information leading to an underestimation of the flux values.Experimental and theoretical-based frequency corrections are used with each being better suited to open (theory-based) or closed (experimental-based) path sensors*.* Using the “opposite” correction leads to a continued underestimation of the flux as well.

### 3.5.1 High Frequency Loss

High frequency information is filtered through the response times of the sensors and data logger, line-averaging over the sonic and IRGA’s path-length, sensor separation (Section 4), and passage through closed-path tubing systems(Aubinet et al., 2012; Moncrieff et al., 2005, 1997; Moore, 1986; van Dijk, 2002)*.* High frequency loss (low-pass filter) are noted from a faster roll-off of the scalar flux co-spectra (See figure at right) at high frequencies (, where x is the scalar quantity). These issues are corrected by using a transfer function based off the w-T co-spectrum and supporting information. Both corrections have their drawbacks; the theoretical approach relies on modeled co-spectra which are not representative of real co-spectra in all terrain types (i.e. forests) and all the physical processes are not fully described, particularly in closed-path systems. The experimental approach assumes that the at least one co-spectra (generally, ) can be measured with negligible errors and requires sufficient periods (>3 hours) of stationary conditions to be most effective. As the transfer function is based on the ratio of the normalized co-spectra, the experimental approach assumes similarity in the turbulent transport of scalars.

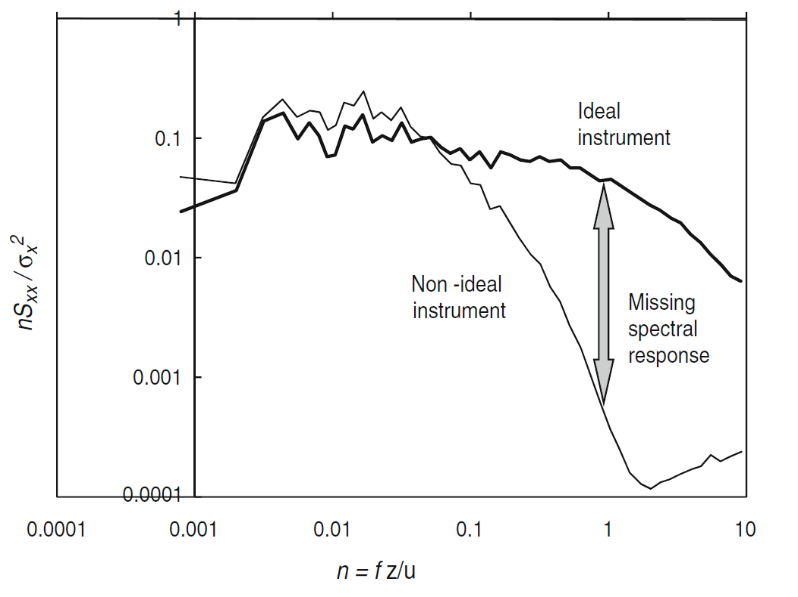
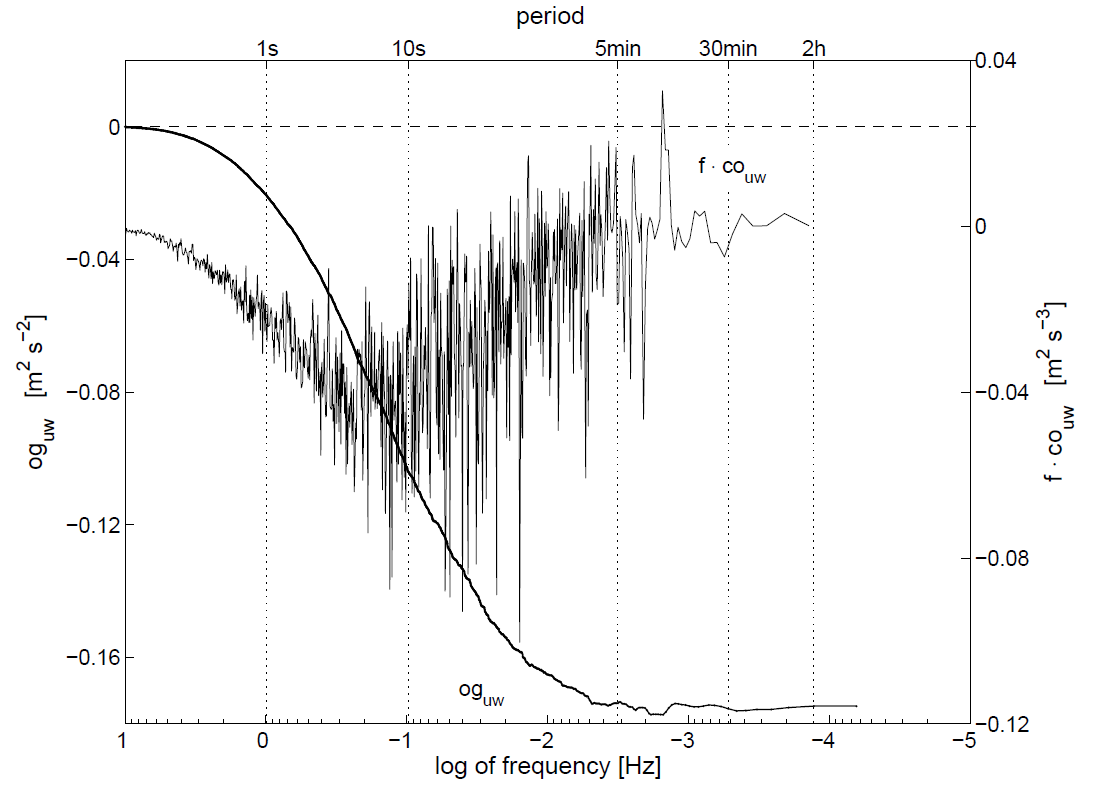


Figure 4.1 from *Aubinet et al.* [2012]

### 3.5.2 Low Frequency Loss

Loss of low frequency information is the result of insufficient averaging time and sample duration with respect to the flux processes. The “standard” averaging time for flux measurements is 30 minutes starting from the 10 Hz data. This period may not capture all the contributions to the low frequency fluxes, again leading to an underestimation of the flux. To check, an ogive test can be used to determine if 30 minutes is an appropriate averaging time (Aubinet et al., 2012; Foken et al., 2006)*.* The ogive is the cumulative integral of the co-spectra starting at the highest frequency, working toward the lowest frequencies (see figure). Mathematically, the Ogive is defined as:



Trace of the ogive

Figure 2a from *Foken et al.* [2006]

|  |  |
| --- | --- |
|  | (A.3) |

Where is the co-spectrum of the vertical velocity (*w*) and some scalar value (*x*) over a range of frequencies (*f*).When plotted, the ogive will (theoretically) level out at a constant value; the frequency where this near-constant value is reached represents the appropriate flux averaging time where the averaged flux will be maximized (see Figure).

## 3.6. SND Buoyancy Correction

Named after the authors of the original paper (Schotanus et al., 1983), the SND correction is used to convert the sonic virtual temperature to ambient air temperature. The sonic anemometer measures temperature through using a speed of sound measurement but the speed of sound changes with the density of the medium and in air, this is affected by the change in the water vapor density. Per *Schotanus et al.* [1983]: “the sonic anemometer only measures the real temperature fluctuation, T’, when the humidity and normal velocity fluctuations are zero”. This correction is used prior to the sensible heat flux calculation and removes the buoyancy (water vapor) effects on the measured high-frequency temperature (Aubinet et al., 2012; Van Dijk, 2001). The sonic temperature prior to the SND correction is akin to the potential virtual temperature (θv). As such, the SND correction is:

|  |  |
| --- | --- |
|  | (A.4) |

Where is the mean sonic temperature, is the ‘corrected’ or air temperature, and is the specific humidity. To correct the sensible heat flux to account for this same buoyancy, the fluctuating component is derived and combined with the sensible heat flux term as:

|  |  |
| --- | --- |
|  | (A.5) |

## 3.7. WPL Density Fluctuation Correction

Also a density correction named after the authors of the original paper [*Webb et al.*, 1980]. The WPL correction is a density correction used for the water vapor and other scalar trace gas fluxes (Leuning, 2007). Temperature and humidity fluctuations cause fluctuations in the gas being measured that are not a result of turbulence. This correction is specific for use with open-path systems (such as the EC150 system used here); the frequency corrections (Section 5) areapplied before the WPL correction.

|  |  |
| --- | --- |
|  | (A.6) |

Where *c* is the scalar value of interest (typically water vapor or CO2), which is the ratio of the molar mass of dry air to water vapor, and as the ratio of the density of wet and dry air.

Closed-path systems do not require the WPL correction but do require more attention in the lag-time correction for tube length and flow rate. Even with the added high-frequency correction, the closed-path system can report lower flux values than an open-path system (Aubinet et al., 2012). At this stage, the 30-minute turbulent energy and scalar fluxes have been calculated using the filtered and corrected data. All that remains is to check the quality of the resultant data.

## 3.8. Data quality grades

Data quality grades are assigned based on three criteria. The flagging system ranges from 1-9 (highest to lowest quality) based on a steady-state test, integral turbulence characteristics and wind direction. Grades 1-3 are considered “research grade” for use in fundamental research, grades 4-6 are still acceptable data but only for more general usage (overall system health, broad-scale research). Grades 7-8 are considered relatively poor quality and only for orientation usage. Grade 9 data should not be used for any application. Each of the three component receives its own grade (1-9 for steady state and turbulence tests, and 1-3 for wind direction) (Foken and Wichura, 1996). A look-up table provides the overall grade for each data point based on the combination of the individual scores (Table F3, recreated from EasyFlux documentation, pp F4-F5):

|  |  |  |  |
| --- | --- | --- | --- |
| **Overall quality**  **Grade (reported)** | ***RNcov***  **Relative non-stationarity** | **ITCsw**  **Relative integral turbulence characteristic** | **wnd\_dir\_sonic**  **Wind Direction** |
| 1 (best) | 1 | 1-2 | 1 |
| 2 | 2 | 1 – 2 | 1 |
| 3 | 1 – 2 | 3 – 4 | 1 |
| 4 | 3 – 4 | 1 – 2 | 1 |
| 5 | 1-4 | 3-5 | 1 |
| 6 | 5 | 5 | 2 |
| 7 | 6 | 6 | 2 |
| 8 | 7-8 | 7-8 | 2 |
| 9 (worst) | 9 | 9 | 3 |

Note there is not a direct comparison between the overall grade and individual grades or with the grading system in (Aubinet et al., 2012). Table below lists the ranges for each of the three data quality that combine to make up the grades (Recreated from Table 4.4; (Aubinet et al., 2012), sonic anemometer column uses directions and grades based on Table F-1 EasyFlux-dl documentation). Some variation may exist depending on the exact application of the grading system depending on the ranges used for each grade.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Steady State Test  (Section 8.1) | | Integral turbulence characteristics (Section 8.2) | | Sonic anemometer horizontal orientation (Section 8.3) | |
| Class | Range | Class | Range | Class | Range | |
| 1 | 0–15% | 1 | 0–15% | 1 | 0-150º, 210-360º | |
| 2 | 16–30% | 2 | 16–30% | 2 | 150-170º, 190-210º | |
| 3 | 31–50% | 3 | 31–50% | 3 | 170-190º | |
| 4 | 51–75% | 4 | 51–75% |  |  | |
| 5 | 76–100% | 5 | 76–100% |  |  | |
| 6 | 101–250% | 6 | 101–250% |  |  | |
| 7 | 251–500% | 7 | 251–500% |  |  | |
| 8 | 501–1,000% | 8 | 501–1,000% |  |  | |
| 9 | >1,000% | 9 | >1,000% |  |  | |

### 3.8.1 Steady State

The steady state test checks the level of stationarity within the covariances (Aubinet et al., 2012; Cava et al., 2014; Foken and Wichura, 1996). EC work assumes a high degree of stationarity within the flux periods. The test compares the average of a series of short interval (SI) covariances (5 minutes) to the whole interval (WI) average (30 minutes) The comparison is done by a percent difference of the average of the short interval covariances compared to the whole interval covariance. A larger difference between the SI and WI values indicates less stationary conditions and produces a higher-grade value.

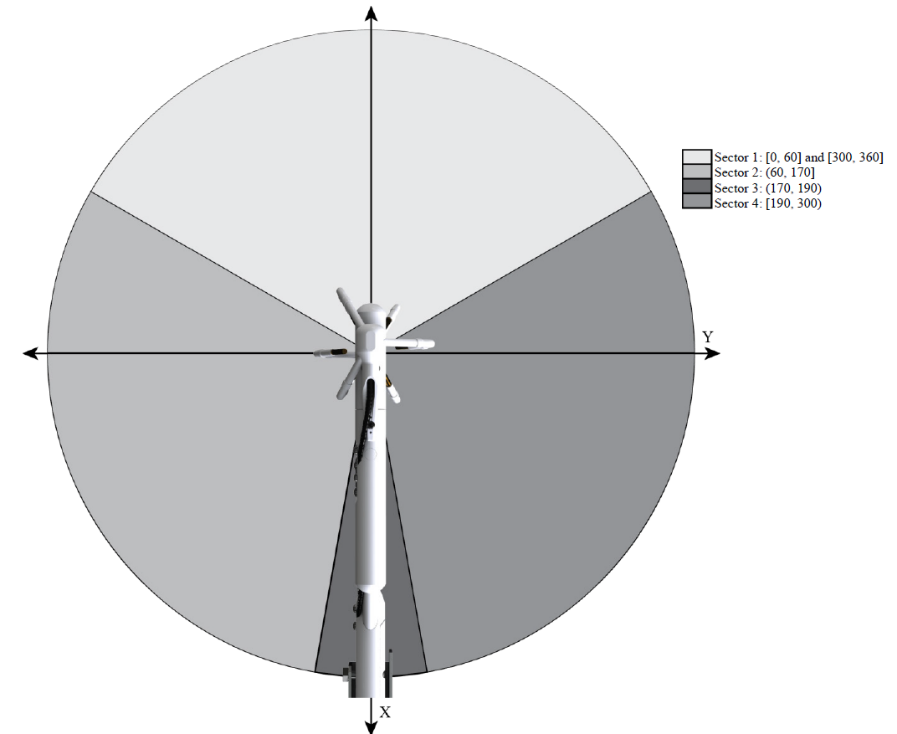
### 3.8.2 Surface layer turbulence characteristics

This data check is based on a modeled surface layer compared to the measured values of the same result. The modeled values utilize two constants from a look-up table that are based on the velocity or temperature parameter and stability via z/L (where L is the Obukhov length). Like the steady-state test, the determining factor is the percent of deviation from the modeled developed turbulence relationship. The more similar the modeled and measured values, the higher data quality grade the data period receives.

### 3.8.3 Wind directions sectors

Figure C-1 from EasyFlux documentation,

Campbell Scientific

Wind direction checks are aimed at determining if the wind is coming through the tower or the backside of the sonic anemometer body. The grading system used is based on (Aubinet et al., 2012). This assigns a grade of 1-3 depending on the general wind sector based on the sonic orientation, *not* the actual wind direction. The original system presented 9 grades of wind direction but there was overlap for wind direction grades and in the final data quality grade. The sonic-based wind sectors are (Figure:). Sector 1 is labeled as grade 1, Sectors 2 and 3 are grade 2 wind direction, and Sector 4 is grade 3 wind direction. Wind from Sector 4 automatically places the overall data quality grade at 9 due to the distortion of the flow through the main sonic body. Flow distortion through the tower equipment and shadowing from the transducers (Horst et al., 2016, 2015) are not calculated within current QA/QC procedures. Interruption of the flow due to other field equipment and tower structure do need to be considered and their sonic-relative wind directions do need to be flagged as potential lower quality data.

## Final Comments

The above represents the current best-practices with respect to flux corrections and calculations. The reader is advised to read the original papers and references (listed below) for more thorough and nuanced discussions of these procedures. The was written to provide a conceptual “roadmap” for those using eddy covariance flux data without much experience. It is not out of the ordinary to lose 20-40% of the flux data from QA/QC flagging (Section 8) and data filtering (Section 2). These data loses are filled through various gap-filling procedures (e.g. (Chi et al., 2016)) to estimate seasonal to yearly scale carbon and water balances or are left empty for more specific research questions. Other post-processing flags can be devised per the users’ interests (e.g., u-star filtering, footprint overlaps, etc).

## 3.10. CAF-Specific Post-Processing

Post-processing of the CAF LTAR is done in two stages: 1) Processing of flux data following the above procedures either online (EasyFlux) or offline (EddyPro) and 2) filtering of processed data based on the QA/QC and site-specific values (offline). Currently, there are seven post-processing filters that are used to filter flux data (Section 3). So long as each of these criteria are met, the data are passed through and considered of good quality. This does not include any gap-filling for flux data, merely the initial processing. Data after this point are considered “post-processed” Level 2 data and sufficiently filtered for research work and gap-filling.

# Chapter 4: EddyPro – EasyFlux Comparison

***Comparison of Offline EddyPro and Online EasyFlux flux processing programs***

Eric S. Russell, Victoria Dziekan, Patrick O’Keeffe, Shelley Pressley, Brian Lamb

Department of Civil and Environmental Engineering, Washington State University, Pullman, WA, USA

**1 Introduction**

Flux calculations for eddy covariance data is an intensive process with multiple processing steps required to generate high-quality 30-minute flux data from the raw, high-frequency measurements (Aubinet et al., 2000, 2012; Finnigan, 2008; Foken, 2008; Fratini & Mauder, 2014; Mauder et al., 2007, 2008; Twine et al., 2000). As eddy covariance measurements and ecosystem flux data permeates the broader environmental and agronomic research fields (Baldocchi et al., 2001; Finnigan, 2008), an increasing variety of publicly available (e.g. TK2/TK3, EddyPro, ECPack) and in-house flux correction programs are used to process flux data. Research reporting flux data report the corrections used and relevant reference, but do not always report which program and version was used to accomplish the processing.

Differences in the latent (LE), sensible (H), and CO2 (Fc) fluxes between sites and studies can result from different implementations of the same correction processes via differences in physical constants and constraints used, and the exact implementation of each correction (Mauder et al., 2007). Mauder et al., (2007) reported up to a 15% difference in H and LE using different correction programs with the sensor separation step being of high importance. Mauder et al., (2008) reported differences of 5-10% between 30-minute fluxes across multiple flux correction programs with discrepancies in the CO2 fluxes attributed to the data preparation (despiking and lag/sensor separation correction), coordinate rotation type, and correction of high-frequency spectral loses. Fratini and Mauder, (2014) noted that a difference in the spectral correction used in EddyPro (utilizing Moncrieff et al., 1997) and TK3 (utilizing Moore, 1986) caused the observed differences in the turbulent fluxes. After forcing the spectral correction to match they achieved similar flux values with both programs.

The correction procedures, equations, and constraints currently in use have been discussed in the literature over the past few decades with a standard set of steps being established (Aubinet et al., 2012; Finnigan, 2008; Foken, 2008; Lee et al., 2004)*.* Discussions about the appropriate flux averaging time (Charuchittipan et al., 2014), despiking process (Starkenburg et al., 2016; Vickers and Mahrt, 1997), which coordinate rotation is best in heterogeneous terrains (Shimizu, 2015; Stiperski and Rotach, 2015; Sun, 2007; Yuan et al., 2007), how to best “grade” flux data (Cava et al., 2014; Foken & Wichura, 1996), and usage of spectral corrections(Cheng et al., 2017; Fratini & Mauder, 2014; Fratini et al., 2012; Mamadou et al., 2016) are still ongoing. The need for these steps are undisputed, but their best practice implementation under different atmospheric conditions, topography, and land cover is evolving with the increased variety of field sites.

Campbell Scientific (Logan, UT, USA) recently released a correction program to calculate fully-corrected fluxes in real-time directly on CR3000 and CR6 data loggers in lieu of off-line flux calculations called EasyFlux-dl (https://www.campbellsci.com/easyflux-dl). As new flux correction programs become available, comparing against the current generation and broadly accepted processing programs is essential to confirm similarities for compatibility across sites and studies and quality of the data (Fratini & Mauder, 2014). With several repositories of flux data and measurement networks reporting flux data; consistency of flux corrections and calculations for every correction program is imperative so synthesis from data across sites and time are possible regardless of the processing program used. The goal for this work is to compare fluxes calculated using EddyPro and EasyFlux to determine the sameness of the fluxes produced via each*.*

**2 Data and Methods**

***2.1 Data sites***

Data used here were collected at three sites at the Long-Term Agroecosystem Research site near Washington State University, Pullman, WA (Table 1). One of the sites (Cook East) was located over spring wheat and the two sites (Boyd North and South) were located over winter wheat. All three tower sites were installed in the summer of 2017 with data records of varying length (Table 1). Each tower contained identical instrumentation: CSAT3A ultrasonic anemometer (Campbell Scientific Inc., Logan, UT, USA), EC150 infrared gas analyzer (Campbell Scientific Inc., Logan, UT, USA), HMP115A (Vaisala Inc., San Jose, CA, USA), NRlite2 net radiometer (Kipp&Zoen B.V., Delft, The Netherlands), LI190SB quantum photosynthetically active radiation sensor (LI-COR Biosciences, Lincoln, NE, US) operated through a CR3000 data logger (Campbell Scientific Inc., Logan, UT, USA). The EC data were collected at 10Hz and stored locally onto compact flash cards. Flux processing via EasyFlux was accomplished in real time on the sites’ data loggers and saved onto the compact flash cards, collected every few weeks along with the 10Hz data. Offline flux processing was accomplished using EddyPro V6.2.0. Fluxes were calculated as 30-minute averages for both processing programs following the standard convention.

***2.2 Data Processing and Flux Calculation***

***2.2.1 EasyFlux-DL***

EasyFlux-dl (hereafter “EFlux”) is a flux correction program written in CRBasic for CR3000 and CR6 data loggers (https://www.campbellsci.com/easyflux-dl). EFlux runs directly on the data logger and process fluxes in real-time, attempting to eliminate the need for offline flux correction. User-provided data is required for site characteristics, instrument type and constants as with any flux correction scheme (e.g., vegetation and measurement height, sonic azimuth, GPS location, etc.). Like EddyPro, the fluxes produced need to post-processed via the end-user to ensure their final quality (Section 2.2.3). References for the processing steps implemented are presented in Table 2. No modifications of the flux correction portion of the code were made.

***2.2.2 EddyPro***

EddyPro (hereafter “EPro”) is anoffline flux calculation program produced and maintained by LiCor Biosciences, Inc., to calculate eddy covariance flux data from the raw data. Version 6.2.0 was used for this work with the metadata set-up to reflect each site’s characteristics (Table 1) and to be consistent with the set-up values used in EFlux (i.e., sensor separation distances, crop and measurement height, sonic azimuth, etc.). EPro was run using both the Massman, (2000, 2001) – Fully analytic spectral correction (hereafter “MM”), which is based off the Moore (1986) correction (used in EFlux), and the Moncrieff et al., (1997) – Fully analytic (hereafter “MC”) spectral correction. The MC correction is the “effective default” in EPro since it is the initially selected procedure unless changed by the user. Other than the spectral correction, the other processing methods were held consistent throughout the analysis (Table 2).

***2.2.3 Post-Processing***

Post-processing was accomplished using a Python (v3.5) script with identical criteria for each data set so no differences were introduced after the flux calculations. No gap-filling was done so only directly measured data were compared. Most gaps occurred overnight when turbulence was low and did not meet the friction velocity threshold ( ≥ 0.13 m s-1). Data were also filtered for times with recorded precipitation at each site, if there were less than 80% of 10 Hz data coverage for the half-hour, low signal strength from the IRGA or sonic (<80%), if the QA/QC grade was ≥ 8 (Foken & Wichura, 1996), or during times the site was being serviced. Data retained for each site, discounting data gaps from site malfunctions, after the entire processing and QA/QC procedure ranged from 60% to 70%. The processed data were compared using an ordinary least squares (OLS) linear regression assuming no dependent or independent variables with the intercept forced through zero. Also reported are the Pearson correlation coefficient () and fractional bias (FB = ).

**3 Results**

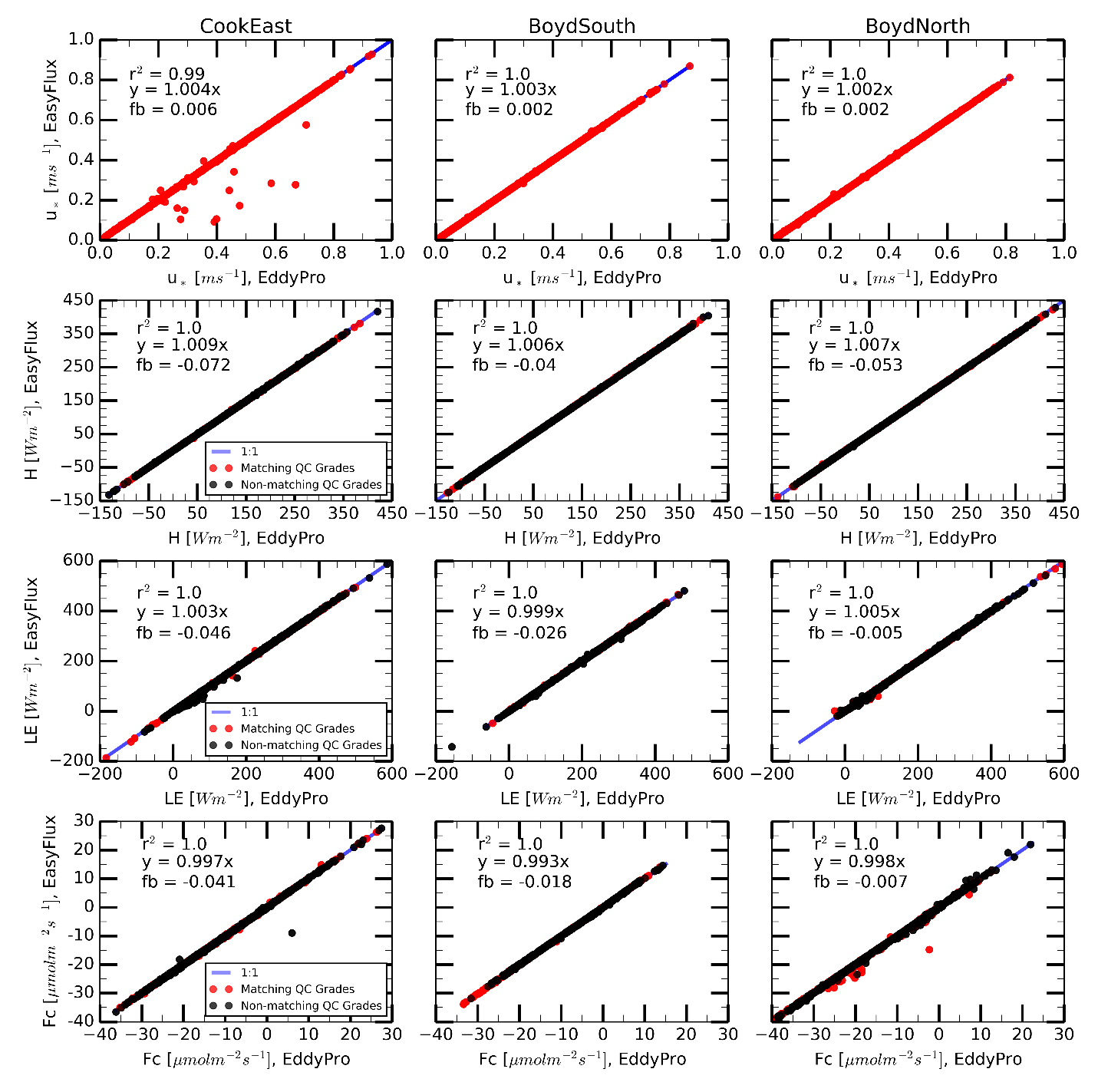
Using the MM option in EPro, the fluxes from EPro and EFlux compare within 1% of each other with R2 values of 1 across all fluxes and sites (Figure 1). Differences in the implementation and other aspects of the correction programs likely contributed to the accumulation of the minor differences observed in the final flux values. Using the MC spectral correction for EPro increased the difference between the corrections to 1-8% depending on flux and site (Figure 2). EFlux produced lower LE and Fc values compared to EPro (Figure 2) at each site with H and comparing favorably. The differences observed here did not have any relationship with the QA/QC grades because of the high R2 values so differences were due to the difference in the spectral correction used. These values are akin to what was observed by Fratini & Mauder (2014) in their comparison of EPro and TK3 before forcing the spectral corrections to match. The R2 values were still near 1 but the slopes of the regression and FB were relatively high. The change in the slope from MM to MC for LE and Fc imply that using different accepted spectral corrections, with every other aspect the same, can contribute to the overall uncertainty of the flux estimate (Mamadou et al., 2016).

Difference in calculated Fc have implications for the overall carbon budget given the difference in the final flux values. Differences in carbon accumulation ranged from 0-3 gC-day-1 between EPro and EFlux for both spectral correction comparisons (Figure 3). Over time this caused the total accumulated carbon at each site to diverge, leading to a 3-8% difference in total carbon between EFlux and EPro using the MM correction. This difference increased to 5%-10% when using MC implying differences in annual budget calculations can be injected into the data from different correction procedures; not just from differences between measurement sites. The differences between the correction procedures were largest for the highest Fc values which overlap with the growing season and were reduced as Fc reduced during senescence and prior to crop emergence. The magnitude of the difference from the different correction process was similar to the magnitude of the uncertainty introduced in data at similar sites through gap-filling of the Fc values (Chi et al., 2016; Waldo et al., 2016).

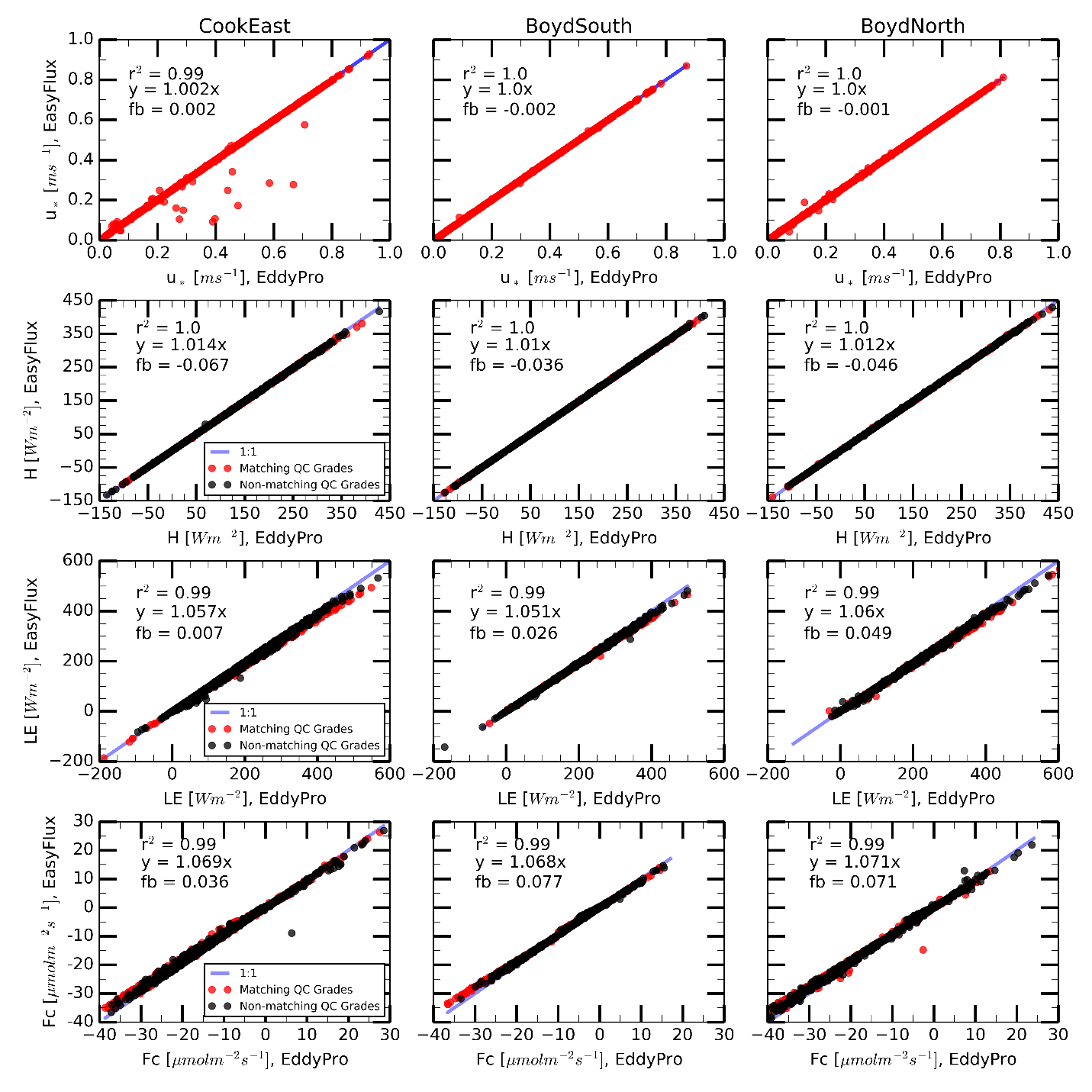
**4 Summary and Conclusions**

The turbulent fluxes from EasyFlux compared favorably with EddyPro. The instantaneous fluxes matched within 1% using the MM-based spectral correction in EddyPro. The minor differences in the 30-minute fluxes values lead to a 3-8% difference in the accumulated carbon over the measurement period. The differences between the correction programs and settings were not entirely random so they can be accounted for between studies and sites so long as the relationship between the correction procedures is known. Systematic uncertainties can be introduced by using different processing methods implying uncertainties and differences between sites can be introduced by the processing methodology which need to be accounted for in synthesis studies. When comparing data between sites or time, any differences in the flux calculations need to be noted so sources of the differences in the fluxes are correctly assigned. Methodology sections need to make clear both the correction methods and the program and version used. As with any correction and post-processing procedure, care needs to be taken to make sure it is implemented correctly, is done following accepted methods, and all steps are well-documented from start to finish.

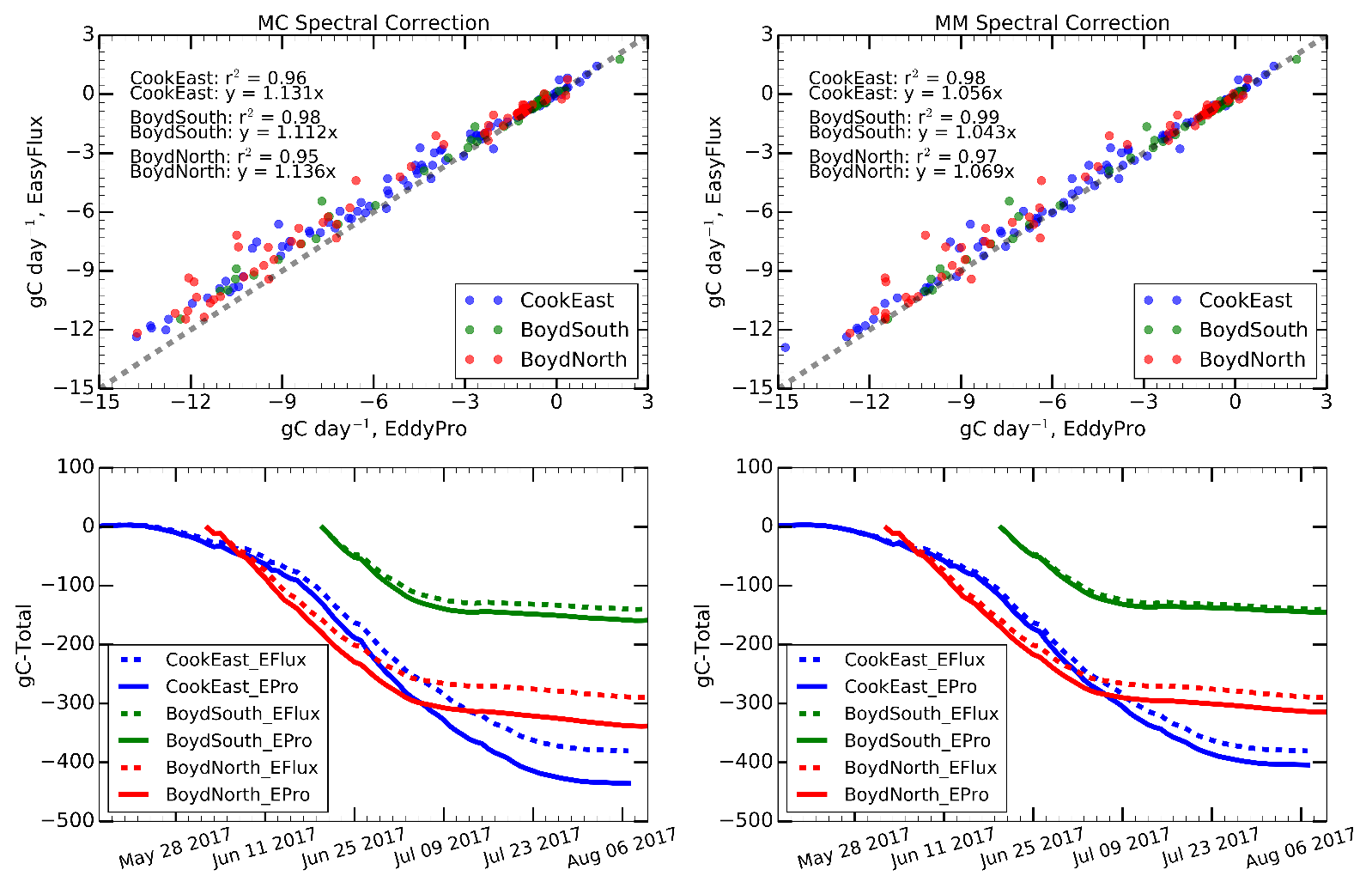
**Figures**



**Figure 1.** Comparison between u-star (top row), sensible heat flux (2nd row), latent heat flux (3rd row), and CO2 flux (bottom row) for EddyPro (x-axis) and EasyFlux (y-axis) at the three sites (columns). Each figure is annotated with the Pearson’s R2, slope of the OLS regression, and fractional bias (fb) where the x-value for each test was the EddyPro data.



**Figure 2.** Same as Figure 1 but using the MC spectral correction for EPro.



**Figure 3.** Integrated daily carbon flux and cumulative sum of accumulated carbon for the data period for each site. The left column is EPro with the MC spectral correction and the right column EPro was run using the MM spectral correction. The top row is annotated with the OLS regression slope and correlation coefficient.

**Tables**

**Table 1.** Location, data period, planting and harvest date, and data retention from the QA/QC procedure for the sites used here.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CookEast** | **BoydSouth** | **BoydNorth** |
| **Crop** | Spring Wheat | Winter wheat | Winter wheat |
| **Data Record Used** | May 12 –  August 6, 2017 | June 20 –  August 10, 2017 | June 02 –  August 10, 2017 |
| **Location** | 46.78152 N  117.08205 W | 46.751809 N  117.128502 W | 46.755104 N  117.126052 W |
| **%Data Retained** | 71% | 59% | 60% |

**Table 2.** Comparison of flux corrections used for each processing program based on references cited in the respective manuals and program menu selections.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **EddyPro** | **EasyFlux** | **Notes** |
| **Despike** | Vickers & Mahrt (1997) | Filtered from sensor diagnostics, signal strength, and measurement range thresholds | EasyFlux did not cite a reference for the despike process. |
| **Density Corrections** | Schotanus et al., (1983) and Webb et al., (1980) | Schotanus et al., (1983), Van Dijk, (2001), and Webb et al., (1980) | EasyFlux followed van Dijk, (2001) implementation of Schotanus el al. (1983) |
| **Spectral Correction** | Massman (2000) and (Moncrieff et al. (1997) | Based on Moore, (1986) | Massman, (2000) is based on Moore, (1986) |
| **Lag/Sensor separation** | Covariance maximization (Horst and Lenschow, 2009) | Covariance maximization (Aubinet et al., 2012; Horst and Lenschow, 2009) | EasyFlux manual indicates additional constraints for physical possible lag times |
| **Grading System** | 0-9; Aubinet et al. (2012) | 0-9; Aubinet et al. (2012) | Differences in grading with respect to wind direction grading system |
| **Coordinate Rotation** | Double Rotation; Wilczak et al. (2001) | Double Rotation; Tanner & Thurtell (1969) | Planar fit was not tested. |

# References

Aubinet, M., Grelle, A., Ibrom, A., Rannik, U., Moncrieff, J., Foken, T., Kowalski, A.S., Martin, P.H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grunwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., Vesala, T., 2000. Exchange of Forests: The EUROFLUX Methodology. Adv. Ecol. Res. 30, 114–175.

Aubinet, M., Vesala, T., Papale, D. (Eds.), 2012. Eddy Covariance: A Practical Guide to Measurement and Data Analysis. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-2351-1

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., others, 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc. 82, 2415–2434.

Cava, D., Donateo, A., Contini, D., 2014. Combined stationarity index for the estimation of turbulent fluxes of scalars and particles in the atmospheric surface layer. Agric. For. Meteorol. 194, 88–103. https://doi.org/10.1016/j.agrformet.2014.03.021

Charuchittipan, D., Babel, W., Mauder, M., Leps, J.-P., Foken, T., 2014. Extension of the Averaging Time in Eddy-Covariance Measurements and Its Effect on the Energy Balance Closure. Bound.-Layer Meteorol. 152, 303–327. https://doi.org/10.1007/s10546-014-9922-6

Cheng, Y., Sayde, C., Li, Q., Basara, J., Selker, J., Tanner, E., Gentine, P., 2017. Failure of Taylor’s hypothesis in the atmospheric surface layer and its correction for eddy-covariance measurements. Geophys. Res. Lett. 2017GL073499. https://doi.org/10.1002/2017GL073499

Chi, J., Waldo, S., Pressley, S., O’Keeffe, P., Huggins, D., Stöckle, C., Pan, W.L., Brooks, E., Lamb, B., 2016. Assessing carbon and water dynamics of no-till and conventional tillage cropping systems in the inland Pacific Northwest US using the eddy covariance method. Agric. For. Meteorol. 218–219, 37–49. https://doi.org/10.1016/j.agrformet.2015.11.019

Desai, A.R., Richardson, A.D., Moffat, A.M., Kattge, J., Hollinger, D.Y., Barr, A., Falge, E., Noormets, A., Papale, D., Reichstein, M., Stauch, V.J., 2008. Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. Agric. For. Meteorol. 148, 821–838. https://doi.org/10.1016/j.agrformet.2007.11.012

Finnigan, J., 2008. An Introduction to Flux Measurements in Difficult Conditions. Ecol. Appl. 18, 1340–1350. https://doi.org/10.1890/07-2105.1

Foken, T., 2008. Micrometeorology. Springer-Verlag Berlin Heidelberg, Berlin.

Foken, T., Wimmer, F., Mauder, M., Thomas, C., Liebethal, C., 2006. Some aspects of the energy balance closure problem. Atmospheric Chem. Phys. 6, 4395–4402.

Foken, Th., Wichura, B., 1996. Tools for quality assessment of surface-based flux measurements. Agric. For. Meteorol. 78, 83–105. https://doi.org/10.1016/0168-1923(95)02248-1

Fratini, G., Ibrom, A., Arriga, N., Burba, G., Papale, D., 2012. Relative humidity effects on water vapour fluxes measured with closed-path eddy-covariance systems with short sampling lines. Agric. For. Meteorol. 165, 53–63. https://doi.org/10.1016/j.agrformet.2012.05.018

Fratini, G., Mauder, M., 2014. Towards a consistent eddy-covariance processing: an intercomparison of EddyPro and TK3. Atmospheric Meas. Tech. 7, 2273–2281. https://doi.org/10.5194/amt-7-2273-2014

Gao, Z., Russell, E.S., Missik, J.E.C., Huang, M., Chen, X., Strickland, C.E., Clayton, R., Arntzen, E., Ma, Y., Liu, H., 2017. A novel approach to evaluate soil heat flux calculation: An analytical review of nine methods. J. Geophys. Res. Atmospheres 122, 6934–6949. https://doi.org/10.1002/2017JD027160

Helbig, M., Wischnewski, K., Gosselin, G.H., Biraud, S.C., Bogoev, I., Chan, W.S., Euskirchen, E.S., Glenn, A.J., Marsh, P.M., Quinton, W.L., Sonnentag, O., 2016. Addressing a systematic bias in carbon dioxide flux measurements with the EC150 and the IRGASON open-path gas analyzers. Agric. For. Meteorol. 228–229, 349–359. https://doi.org/10.1016/j.agrformet.2016.07.018

Horst, T.W., Lenschow, D.H., 2009. Attenuation of Scalar Fluxes Measured with Spatially-displaced Sensors. Bound.-Layer Meteorol. 130, 275–300. https://doi.org/10.1007/s10546-008-9348-0

Horst, T.W., Semmer, S.R., Maclean, G., 2015. Correction of a Non-orthogonal, Three-Component Sonic Anemometer for Flow Distortion by Transducer Shadowing. Bound.-Layer Meteorol. 155, 371–395. https://doi.org/10.1007/s10546-015-0010-3

Horst, T.W., Vogt, R., Oncley, S.P., 2016. Measurements of Flow Distortion within the IRGASON Integrated Sonic Anemometer and CO\_2/H\_2O Gas Analyzer. Bound.-Layer Meteorol. 160, 1–15. https://doi.org/10.1007/s10546-015-0123-8

Kaimal, J.C., Haugen, D.A., 1969. Some Errors in the Measurement of Reynolds Stress. J. Appl. Meteorol. 8, 460–462.

Lee, X., Finnigan, J., Paw U, K., 2005. Coordinate systems and flux bias error. Handb. Micrometeorology 33–66.

Lee, X., Massman, W., Law, B. (Eds.), 2004. Handbook of micrometeorology: a guide for surface flux measurements. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Leuning, R., 2007. The correct form of the Webb, Pearman and Leuning equation for eddy fluxes of trace gases in steady and non-steady state, horizontally homogeneous flows. Bound.-Layer Meteorol. 123, 263–267. https://doi.org/10.1007/s10546-006-9138-5

Mamadou, O., Gourlez de la Motte, L., De Ligne, A., Heinesch, B., Aubinet, M., 2016. Sensitivity of the annual net ecosystem exchange to the cospectral model used for high frequency loss corrections at a grazed grassland site. Agric. For. Meteorol. 228, 360–369. https://doi.org/10.1016/j.agrformet.2016.06.008

Massman, W.J., 2000. A simple method for estimating frequency response corrections for eddy covariance systems. Agric. For. Meteorol. 104, 185–198. https://doi.org/10.1016/S0168-1923(00)00164-7

Mauder, M., Foken, T., 2011. Documentation and instruction manual of the eddy-covariance software package TK3.

Mauder, M., Foken, T., Clement, R., Elbers, J.A., Eugster, W., Grünwald, T., Heusinkveld, B., Kolle, O., 2008. Quality control of CarboEurope flux data? Part 2: Inter-comparison of eddy-covariance software. Biogeosciences 5, 451–462.

Mauder, M., Oncley, S.P., Vogt, R., Weidinger, T., Ribeiro, L., Bernhofer, C., Foken, T., Kohsiek, W., De Bruin, H.A.R., Liu, H., 2007. The energy balance experiment EBEX-2000. Part II: Intercomparison of eddy-covariance sensors and post-field data processing methods. Bound.-Layer Meteorol. 123, 29–54. https://doi.org/10.1007/s10546-006-9139-4

Metzger, S., Durden, D., Sturtevant, C., Luo, H., Pingintha-Durden, N., Sachs, T., Serafimovich, A., Hartmann, J., Li, J., Xu, K., Desai, A.R., 2017. eddy4R: A community-extensible processing, analysis and modeling framework for eddy-covariance data based on R, Git, Docker and HDF5. Geosci. Model Dev. Discuss. 1–26. https://doi.org/10.5194/gmd-2016-318

Moffat, A.M., Papale, D., Reichstein, M., Hollinger, D.Y., Richardson, A.D., Barr, A.G., Beckstein, C., Braswell, B.H., Churkina, G., Desai, A.R., Falge, E., Gove, J.H., Heimann, M., Hui, D., Jarvis, A.J., Kattge, J., Noormets, A., Stauch, V.J., 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agric. For. Meteorol. 147, 209–232. https://doi.org/10.1016/j.agrformet.2007.08.011

Moncrieff, J., Clement, R., Finnigan, J., Meyers, T., 2005. Averaging, detrending, and filtering of eddy covariance time series. Handb. Micrometeorology 7–31.

Moncrieff, J.B., Massheder, J.M., de Bruin, H., Elbers, J., Friborg, T., Heusinkveld, B., Kabat, P., Scott, S., Soegaard, H., Verhoef, A., 1997. A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. J. Hydrol., HAPEX-Sahel 188, 589–611. https://doi.org/10.1016/S0022-1694(96)03194-0

Moore, C.J., 1986. Frequency response corrections for eddy correlation systems. Bound.-Layer Meteorol. 37, 17–35.

Oldroyd, H.J., Pardyjak, E.R., Huwald, H., Parlange, M.B., 2015. Adapting Tilt Corrections and the Governing Flow Equations for Steep, Fully Three-Dimensional, Mountainous Terrain. Bound.-Layer Meteorol. 159, 539–565. https://doi.org/10.1007/s10546-015-0066-0

Russell, E.S., Dziekan, V., Chi, J., Waldo, S., Pressley, S.N., O’Keeffe, P., Lamb, B.K., 2019. Adjustment of CO2 flux measurements due to the bias in the EC150 infrared gas analyzer. Agric. For. Meteorol. 276–277, 107593. https://doi.org/10.1016/j.agrformet.2019.05.024

Russell, E.S., Liu, H., Gao, Z., Finn, D., Lamb, B., 2015. Impacts of soil heat flux calculation methods on the surface energy balance closure. Agric. For. Meteorol. 214–215, 189–200. https://doi.org/10.1016/j.agrformet.2015.08.255

Schotanus, P., Nieuwstadt, Ftm., De Bruin, H.A.R., 1983. Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes. Bound.-Layer Meteorol. 26, 81–93.

Shimizu, T., 2015. Effect of Coordinate Rotation Systems on Calculated Fluxes over a Forest in Complex Terrain: A Comprehensive Comparison. Bound.-Layer Meteorol. 156, 277–301. https://doi.org/10.1007/s10546-015-0027-7

Soloway, A.D., Amiro, B.D., Dunn, A.L., Wofsy, S.C., 2017. Carbon neutral or a sink? Uncertainty caused by gap-filling long-term flux measurements for an old-growth boreal black spruce forest. Agric. For. Meteorol. 233, 110–121. https://doi.org/10.1016/j.agrformet.2016.11.005

Starkenburg, D., Metzger, S., Fochesatto, G.J., Alfieri, J.G., Gens, R., Prakash, A., Cristóbal, J., 2016. Assessment of Despiking Methods for Turbulence Data in Micrometeorology. J. Atmospheric Ocean. Technol. 33, 2001–2013. https://doi.org/10.1175/JTECH-D-15-0154.1

Stiperski, I., Rotach, M.W., 2015. On the Measurement of Turbulence Over Complex Mountainous Terrain. Bound.-Layer Meteorol. https://doi.org/10.1007/s10546-015-0103-z

Sun, J., 2007. Tilt corrections over complex terrain and their implication for CO2 transport. Bound.-Layer Meteorol. 124, 143–159. https://doi.org/10.1007/s10546-007-9186-5

Tanner, C.B., Thurtell, G.W., 1969. Anemoclinometer measurements of Reynolds stress and heat transport in the atmospheric surface layer. WISCONSIN UNIV-MADISON DEPT OF SOIL SCIENCE.

Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, Pr., Meyers, T.P., Prueger, J.H., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agric. For. Meteorol. 103, 279–300.

van Dijk, A., 2002. Extension to 3D of “The Effect of Line Averaging on Scalar Flux Measurements with a Sonic Anemometer near the Surface” by Kristensen and Fitzjarrald. J. Atmospheric Ocean. Technol. 19, 80–82. https://doi.org/10.1175/1520-0426(2002)019<0080:ETOTEO>2.0.CO;2

Van Dijk, A., 2001. The principles of surface flux physics. internal report, Research Group of the Royal Netherlands Meteorological Institute and Department of Meteorology and Air Quality with Agricultural University Wageningen.

Vickers, D., Mahrt, L., 1997. Quality Control and Flux Sampling Problems for Tower and Aircraft Data. J. Atmospheric Ocean. Technol. 14, 512–526.

Waldo, S., Chi, J., Pressley, S.N., O’Keeffe, P., Pan, W.L., Brooks, E.S., Huggins, D.R., Stöckle, C.O., Lamb, B.K., 2016. Assessing carbon dynamics at high and low rainfall agricultural sites in the inland Pacific Northwest US using the eddy covariance method. Agric. For. Meteorol. 218–219, 25–36. https://doi.org/10.1016/j.agrformet.2015.11.018

Webb, E. K., Pearman, G. I., Leuning, R, 1980. Correction of flux measurments for density effects due to heat and water vapour transfer. Q. J. R. Meteorol. Soc. 106, 85–100.

Wilczak, J.M., Oncley, S.P., Stage, S.A., 2001. Sonic anemometer tilt correction algorithms. Bound.-Layer Meteorol. 99, 127–150.

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, 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

Yuan, R., Kang, M., Park, S., Hong, J., Lee, D., Kim, J., 2007. The effect of coordinate rotation on the eddy covariance flux estimation in a hilly Koflux forest site. Agric. For. Meteorol. 9, 100–108.

Yuan, R., Kang, M., Park, S.-B., Hong, J., Lee, D., Kim, J., 2011. Expansion of the planar-fit method to estimate flux over complex terrain. Meteorol. Atmospheric Phys. 110, 123–133. https://doi.org/10.1007/s00703-010-0113-9

# Appendix A: Added columns from EddyPro Output to EasyFlux data

These columns are added to the logger output flux data until the logger code is updated to the high frequency CO2 values. Once that happens, this addition of extra columns will be obsolete and will be noted as such. All the other fluxes and turbulent components have been determined to be equivalent, so the other data will not be different regardless of processing program used.

|  |  |
| --- | --- |
| **Variable Name** | **Description and Comments** |
| co2\_flux | CO2 flux from EddyPro processing as high frequency CO2 values |
| qc\_co2\_flux | Grading for CO2 flux from EddyPro |
| co2\_mole\_fraction | CO2 mole fraction using high frequency CO2 from EddyPro |
| co2\_mixing\_ratio | CO2 mixing ratio using high frequency CO2 from EddyPro |
| co2\_var | Variance of CO2 using high frequency value from EddyPro |
| w/co2\_cov | Covariance of w and c using high frequency CO2 from EddyPro |

# Appendix B: General workflow for CAF EC data processing.