NEON ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD): EDDY-COVARIANCE DATA PRODUCTS BUNDLE

PREPARED BY	ORGANIZATION	DATE
Stefan Metzger	NEON	04/30/2018
David Durden	NEON	04/24/2018
Ke Xu	University of Wisconsin	04/13/2018
Natchaya Pingintha-Durden	NEON	12/17/2019
Hongyan Luo	NEON	07/26/2021
Christopher Florian	NEON	02/16/2022

APPROVALS	ORGANIZATION	APPROVAL DATE
Kate Thibault	SCI	05/16/2022

RELEASED BY	ORGANIZATION	RELEASE DATE
Tanisha Waters	CM	05/16/2022

See configuration management system for approval history.

The National Ecological Observatory Network is a project solely funded by the National Science Foundation and managed under cooperative agreement by Battelle. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	06/15/2018	ECO-05623	Initial release
B	11/30/2018	ECO-05977	Add MDP wording
C	02/16/2021	ECO-06579	Incorporating sections for QFQM determination, ITC and Stationarity flags, and turbulent CO2 drift correction
D	12/01/2021	ECO-06664	Add LI-850A and LI-7200RS to this doc
E	02/16/2022	ECO-06783	Adding sections for the new L1 CH4 data product
F	05/16/2022	ECO-06818	• Added NEON to document title • Updated NEON logo • Update to reflect change in terminology from relocatable to gradient sites

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

The National Ecological Observatory Network (NEON) seeks to address grand challenges in continental-scale ecology through extensive observational infrastructure across the U.S. One core element are eddy-covariance (EC) flux measurements of ecologically-relevant energy, water, and trace gas fluxes, which are performed at 47 NEON Terrestrial Instrument System (TIS) sites. The underlying NEON observatory design is based on multivariate geographic clustering (11). At each NEON TIS site, tower-based EC-flux measurements are performed in coordination with a wide range of contextual observations (Appendix B). The EC subsystems aim to maximize data coverage and quality through close design integration of tower, a suite of sensors and auxiliary components (Appendix C). The full complement of resulting EC bundled data products (DP) is directly available from the NEON Data Portal. In addition, a set of standard outputs is regularly submitted to AmeriFlux and FLUXNET for cross-network harmonization and access.

1.1 Purpose

This Algorithm Theoretical Basis Document (ATBD) accompanies NEON's EC bundled DPs. It describes the theoretical background and entire algorithmic sequence used for determining the surface-atmosphere exchange (SAE) of momentum, heat, H_2O and CO_2 from sensor readings of the wind vector, temperature, and scalar concentrations. In addition, the isotopic composition of atmospheric H_2O and CO_2 is determined. Additional measurements can be proposed (Sect. 8).

1.2 Scope

The EC system consists of two subsystems, the eddy-covariance turbulent exchange subsystem (ECTE) and the eddy-covariance storage exchange subsystem (ECSE). The command, control, and configuration (C3) documents for ECTE [AD01] and ECSE [AD02] describe the corresponding instruments, set points, control parameters, conditions/constraints, and any necessary error handling for the physical implementation of the subsystems. Some basic information is also summarized in Appendix C. Data product levels relevant for the subsequent processing of sensor readings are:

- dp00: sensor readings in engineering units; e.g. concentration as infrared absorptance.
- dp0p: pre-conditioned data in scientific units; e.g. concentration as mole fraction.
- dp01: descriptive statistics.
- dp02: time-interpolated data.
- dp03: space-interpolated data.
- dp04: flux data.

Prior to the scientific processing described in this ATBD, all dp00 are pre-conditioned to dp0p as detailed in AD[02] and AD[03]. The present document outlines the scientific rationale and process implementation for transitioning dp0p to dp01 – dp04. At the time of writing, direct links to the corresponding workflows and functions with public access to the operational code in the https://github.com/NEONScience/eddy4R GitHub repository are in preparation. This combines a concise overview in this ATBD with the full transparency of each processing step, and facilitates direct community input on the continued development and operation of NEON EC DPs (Sect. 4). In the interim while this GitHub repository is being prepared for public access, the manuals for the eddy4R.base and eddy4R.qaqc R-packages are available in Appendix G.

This ATBD first introduces related documents, acronyms and conventions in Sect. 2. Inputs and outputs are described in Sect. 3, and Sect. 4 introduces the framework for community-driven algorithm development and operation. Throughout Sects. 5–7 the processing steps specific to turbulent exchange, storage exchange and net surface-atmosphere exchange are summarized, respectively. Each section is divided into subsections specific to data analysis, quality assurance and quality control, and uncertainty analysis, and focuses on algorithm theory and its implementation. Future plans and modifications are briefly outlined in Sect. 8.

The algorithms described in this document will be implemented to process data from at all applicable NEON terrestrial sites and up to 5 Mobile Deployment Platform (MDP) sites at a given period. MDP is a PI driven system; its site location varies depending on PI's decision and research needs.

Here and below, "all NEON sites" or "NEON sites" in this document include both terrestrial sites and MDP sites.

2 RELATED DOCUMENTS, ACRONYMS AND VARIABLE NOMENCLATURE

Below applicable and reference documents are available from the NEON Document library. External references are listed in Sect. 11, and acronyms and variable nomenclature are tabulated in Appendix D – Appendix G.

2.1 Applicable Documents

AD[01]	NEON.DOC.000456	Eddy-covariance turbulent exchange subsystem Command, Control and Configuration document
AD[02]	NEON.DOC.000807	NEON Algorithm Theoretical Basis Document (ATBD) – Eddy Covariance Turbulent Exchange Subsystem Level 0 to Level 0 prime
AD[03]	NEON.DOC.004967	NEON Algorithm Theoretical Basis Document (ATBD) – Eddy Covariance Storage Exchange Subsystem Level 0 to Level 0 prime
AD[04]	NEON.DOC.000573	FIU plan for airshed QA/QC development
AD[05]	NEON.DOC.001113	Quality Flags and Quality Metrics for TIS Data Products ATBD
AD[06]	NEON.DOC.011081	ATBD QA/QC plausibility testing
AD[07]	NEON.DOC.001069	Preprocessing for TIS Level 1 Data Products
AD[08]	NEON.DOC.002651	Data Product Naming Convention
AD[09]	NEON.DOC.000465	Eddy-covariance storage exchange subsystem Command, Control and Configuration document

2.2 Reference Documents

RD[01]	NEON.DOC.000008	NEON Acronym List
RD[02]	NEON.DOC.000243	NEON Glossary of Terms

3 DATA PRODUCT DESCRIPTION

3.1 Variables Reported

The eddy-covariance related DPs provided by the algorithms documented in this ATBD are listed in **Table 1**. The DPs are provided in the form of HDF5 files (Sect. 3.2), including a description of all file structure and objects. Each Data Product encompasses several sub-products, and the units of the individual sub-products are provided in the HDF5 file. The data products will be produced and published in three phases, the initial transition, a science reviewed quality transition, and the epoch yearly transition (see Sect. 4 for more information).

Table 1. List of variables reported (below) For column "Location in HDF5 file", SITE refers to the four-letter acronym of a NEON TIS site, and DQU to one of {data, qfqm, ucrt} referring to respective locations of data, quality and uncertainty information. The "*" denotes products that are not in the current NEON HDF5 files, but will be added in future data releases.

Instrument system	Description	Temporal resolution	Data Product Number	Location in HDF5 file
Re-ingested	Triple Aspirated Air Temperature (**tempAirTop**)	1 min, 30 min	NEON.DP1.00003	SITE/dp01/DQU/tempAirTop
	Single Aspirated Air Temperature (**tempAirLvl**)	1 min, 30 min	NEON.DP1.00002	SITE/dp01/DQU/tempAirLvl
	*Soil Temperature (**tempSoil**)	1 min, 30 min	NEON.DP1.00041	SITE/dp01/DQU/tempSoil
	*Soil Heat Flux (**fluxHeatSoil**)	1 min, 30 min	NEON.DP1.00040	SITE/dp01/DQU/fluxHeatSoil
	*Shortwave and Longwave Radiation (**radiNet**)	1 min, 30 min	NEON.DP1.00023	SITE/dp01/DQU/radiNet
	*Soil water content and water salinity (**h2oSoilVol**)	1 min, 30 min	NEON.DP1.00094	SITE/dp01/DQU/h2oSoilVol
	*Barometric pressure (**presBaro**)	1 min, 30 min	NEON.DP1.00004	SITE/dp01/DQU/presBaro
EC turbulent exchange (ECTE)	3D Wind Speed, Direction and Sonic Temperature (**soni**)	1 min, 30 min	NEON.DP1.00007	SITE/dp01/DQU/soni
	3D Wind Attitude and Motion Reference (**amrs**)	1 min, 30 min	NEON.DP1.00010	SITE/dp01/DQU/amrs
	CO_2 Concentration – Turbulent (**co2Turb**)	1 min, 30 min	NEON.DP1.00034	SITE/dp01/DQU/co2Turb
	H_2O Concentration – Turbulent (**h2oTurb**)	1 min, 30 min	NEON.DP1.00035	SITE/dp01/DQU/h2oTurb

EC storage exchange (ECSE)	CO_2 Concentration – Storage (**co2Stor**)	2 min, 30 min	NEON.DP1.00099	SITE/dp01/DQU/co2Stor
	H_2O Concentration – Storage (**h2oStor**)	2 min, 30 min	NEON.DP1.00100	SITE/dp01/DQU/h2oStor
	Atmospheric CO_2 Isotopes (**isoCo2**)	9 min, 30 min	NEON.DP1.00036	SITE/dp01/DQU/isoCo2
	CH4 Concentration (**ch4Conc**)	9 min, 30 min	NEON.DP1.00030	SITE/dp01/DQU/ch4Conc
	Atmospheric H_2O isotopes (**isoH2o**)	9 min, 30 min	NEON.DP1.00037	SITE/dp01/DQU/isoH2o
EC storage exchange (ECSE)	Temperature rate of change (**tempStor**)	30 min	NEON.DP2.00024	SITE/dp02/DQU/tempStor
	CO_2 concentration rate of change (**dp02 co2Stor**)	30 min	NEON.DP2.00008	SITE/dp02/DQU/co2Stor
	H_2O concentration rate of change (**dp02 h2oStor**)	30 min	NEON.DP2.00009	SITE/dp02/DQU/h2oStor
EC storage exchange (ECSE)	Temperature rate of change profile (**dp03 tempStor**)	30 min	NEON.DP3.00008	SITE/dp03/DQU/tempStor
	CO_2 concentration rate of change profile (**dp03 co2Stor**)	30 min	NEON.DP3.00009	SITE/dp03/DQU/co2Stor
	H_2O concentration rate of change profile (**dp03 h2oStor**)	30 min	NEON.DP3.00010	SITE/dp03/DQU/h2oStor
EC profile + turbulence combined (NSAE)	Sensible heat flux (**fluxHeat**)	30 min	NEON.DP4.00002	SITE/dp04/DQU/fluxHeat
	Momentum Flux (**fluxMome**)	30 min	NEON.DP4.00007	SITE/dp04/DQU/fluxMome
	Latent heat flux (**fluxH2o**)	30 min	NEON.DP4.00137	SITE/dp04/DQU/fluxH2o
	Carbon dioxide flux (**fluxCo2**)	30 min	NEON.DP4.00067	SITE/dp04/DQU/fluxCo2
	Footprint characteristics (**foot**)	30 min	NEON.DP4.00201	SITE/dp04/DQU/foot

3.2 HDF5 Representation

The DPs listed in Table 1 can be downloaded as HDF5 data product bundle from the NEON Data Portal: the Hierarchical Data Format (HDF), currently distributed as HDF5, provides a file format with high compressibility, fast efficient reading and writing capabilities, directory-style files, and metadata attachment. The contents of HDF5 files can be explored intuitively e.g. with tools freely available from the HDF Group such as HDFView, and is supported by all major programming languages. The HDF5 file format allows to package various data sets into a single file with built-in structure for managing both data and metadata. In addition, the NEON processing pipeline utilizes HDF5 files for input/output operations.

The specific locations of the individual EC DPs in the HDF5 file is provided in in **Table 1**, column "Location in HDF5 file". The HDF5 file itself contains a description of all terms used for naming objects ("objDesc"), and a readMe with examples and additional information. Notably, missing values in the HDF5 file are expressed as NaN for data and NA for metadata.

The underlying HDF5 file structure was developed following the NEON data product naming convention provided in AD[08], where portions of the naming convention were selected to develop the hierarchical structure of the HDF5 file as described below and illustrated in **Figure 1**.

NEON.DOM.SITE.DPL.PRNUM.REV.TERMS.HOR.VER.TMI, with:

NEON=NEON

DOM=DOMAIN, e.g. D10

SITE=SITE, e.g. STER

DPL=DATA PRODUCT LEVEL, e.g. DP1

PRNUM = PRODUCT NUMBER =>5 digit number. Set in data products catalog. TIS = 00000-09999

REV = REVISION, e.g. 001.

TERMS=From NEON's controlled list of terms. Index is unique across products.

HOR = HORIZONTAL INDEX. Semi-controlled. Examples: Tower=000, HUT=700.

VER = VERTICAL INDEX. Semi-controlled. Examples: Ground level=000, second tower level=020.

TMI=TEMPORAL INDEX. Examples: 001=1 minute, 030=30 minute, 999=irregular intervals.

One departure from the DP naming convention is the use of the data product name (e.g. **irgaTurb**) in place of PRNUM in the Data Product ID group level, this change was made to improve readability. At the top level of the provided HDF5 file a readme and object description (objDesc) are provided to explain the contents of the file. An additional HDF5 group level was added to separate the data (data), quality flags and quality metrics (qfqm), and uncertainty quantification (ucrt). It should also be noted that level 3 and 4 data products (dp03 and dp04) are spatially interpolated and only provided at 30 minute aggregation periods; thus, the HOR_VER_TMI level is not present (see example in **Figure 1**).

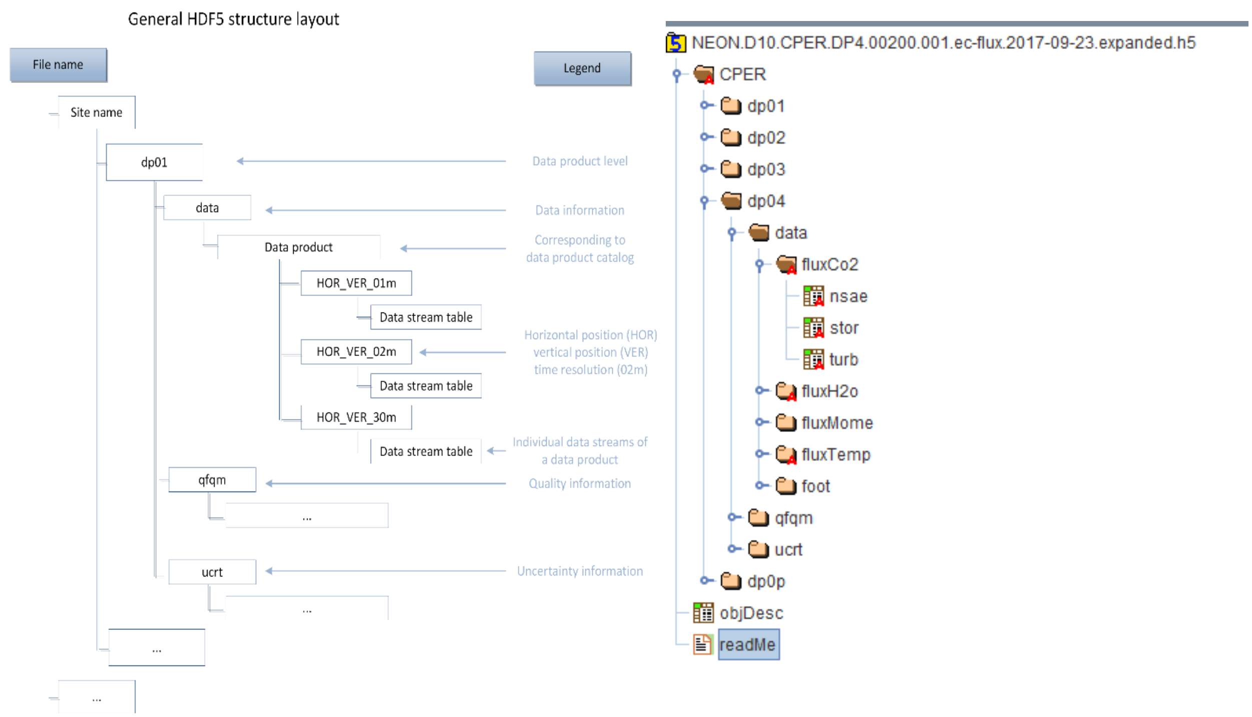

Figure 1. Theoretical diagram depicting the NEON HDF5 file structure following the NEON DP naming convention (left). An actual NEON HDF5 file screenshot depicting the hierarchical layout of the files (right). Note that PRNUM is replaced by the data product name associated with that PRNUM, e.g., **fluxCO2**.

3.3 Input Dependencies

Below **Table 2** - **Table 5** and **Table 12** detail the Eddy-covariance turbulent exchange (ECTE) related dp0p DPs used to produce dp01 DPs in this ATBD.

Table 2. List of eddy-covariance turbulent exchange ultrasonic anemometer/thermometer (**soni**)-related dp0p DPs that are ingested in this ATBD.

dp0p Description	dp0p Term Name	Sample Frequency	Units
Measured along-axis wind speed (u_m)	veloXaxs	20 Hz	ms^{-1}
Measured cross-axis wind speed (v_m)	veloYaxs	20 Hz	ms^{-1}
Measured vertical-axis wind speed (w_m)	veloZaxs	20 Hz	ms^{-1}
Measured speed of sound (c_m)	veloSoni	20 Hz	ms^{-1}
Sonic temperature (T_{SONIC})	tempSoni	20 Hz	K
Sample count	idx	20 Hz	NA
Sensor error flag ($QF_{SONIC,o1}$: Sensor unresponsive)	qfSoniUnrs	20 Hz	NA
Sensor error flag ($QF_{SONIC,o2}$: No data available)	qfSoniData	20 Hz	NA
Sensor error flag ($QF_{SONIC,o3}$: Sensor trigger source lost)	qfSoniTrig	20 Hz	NA
Sensor error flag ($QF_{SONIC,o4}$: SDM communications error)	qfSoniComm	20 Hz	NA
Sensor error flag ($QF_{SONIC,o5}$: Wrong embedded sensor code)	qfSoniCode	20 Hz	NA
Sensor signal flag ($QF_{SONIC,s1}$: Axes T_{SONIC} difference > 4 K)	qfSoniTemp	20 Hz	NA
Sensor signal flag ($QF_{SONIC,s2}$: Poor signal lock)	qfSoniSgnlPoor	20 Hz	NA

dp0p Description	dp0p Term Name	Sample Frequency	Units
Sensor signal flag ($QF_{SONIC,s3}$: High signal amplitude)	qfSoniSgnlHigh	20 Hz	NA
Sensor signal flag ($QF_{SONIC,s4}$: Low signal amplitude)	qfSoniSgnlLow	20 Hz	NA

Table 3. List of eddy-covariance turbulent exchange attitude and motion reference (**amrs**)-related dp0p DPs that are ingested in this ATBD.

dp0p DP	dp0p Term Name	Sample Frequency	Units
Measured along-axis acceleration ($acc_{x,m}$)	accXaxs	40 Hz	m s^{-2}
Measured cross-axis acceleration ($acc_{y,m}$)	accYaxs	40 Hz	m s^{-2}
Measured vertical-axis acceleration ($acc_{z,m}$)	accZaxs	40 Hz	m s^{-2}
Along-axis free acceleration	accXaxsDiff	40 Hz	m s^{-2}, positive forward
Cross-axis free acceleration	accYaxsDiff	40 Hz	m s^{-2}, positive left
Vertical-axis free acceleration	accZaxsDiff	40 Hz	m s^{-2}, positive up
Pitch rate	avelYaxs	40 Hz	rad s^{-1}
Roll rate	avelXaxs	40 Hz	rad s^{-1}
Yaw rate	avelZaxs	40 Hz	rad s^{-1}
Measured pitch angle (θ_{m})	angYaxs	40 Hz	rad
Measured roll angle (ϕ_{m})	angXaxs	40 Hz	rad
Yaw angle (ψ)	angZaxs	40 Hz	rad
Index value	idx	40 Hz	NA
Sensor signal flag: Selftest	qfAmrsVal	40 Hz	NA
Sensor signal flag: Filter Valid	qfAmrsFilt	40 Hz	NA
Sensor signal flag: NoVelocityUpdate status	qfAmrsVelo	40 Hz	NA
Sensor signal flag: Clipping indication	qfAmrsRng	40 Hz	NA

Table 4. List of eddy-covariance turbulent exchange infrared gas analyzer (irgaTurb)-related dp0p DPs that are ingested in this ATBD.

dp0p DP	dp0p Term Name	Sample Frequency	Units
Cell temperature in (at sensor head inlet)	tempIn	20 Hz	K
Cell temperature out (at sensor head inlet)	tempOut	20 Hz	K
Cell temperature (weighted average of head inlet and outlet temperature)	tempMean	20 Hz	K
Block temperature	tempRefe	20 Hz	K
Ambient pressure (LI–7550 box pressure)	presAtm	20 Hz	Pa
Head pressure (differential pressure head-box)	presDiff	20 Hz	Pa
Total pressure (LI–7550 box pressure + head pressure)	presSum	20 Hz	Pa
H_2O sample power	powrH2oSamp	20 Hz	W
H_2O reference power	powrH2oRefe	20 Hz	W
H_2O raw absorptance	asrpH2o	20 Hz	-
H_2O molar density	densMoleH2o	20 Hz	mol m^{-3}
H_2O dry mole fraction	rtioMoleDryH2o	20 Hz	mol mol^{-1}

dp0p DP	dp0p Term Name	Sample Frequency	Units
CO_2 sample power	powrCo2Samp	20 Hz	W
CO_2 reference power	powrCo2Refe	20 Hz	W
CO_2 raw absorptance	asrpCo2	20 Hz	-
CO_2 molar density	densMoleCo2	20 Hz	mol m^{-3}
CO_2 dry mole fraction	rtioMoleDryCo2	20 Hz	mol mol^{-1}
Sequence number	idx	20 Hz	NA
LI-7200 (or LI-7200RS) diagnostic value 2 (sync clocks)	diag02	20 Hz	NA
LI-7200 (or LI-7200RS) cooler voltage	potCool	20 Hz	V
CO_2 signal strength	ssiCo2	20 Hz	-
H_2O signal strength	ssiH2o	20 Hz	-
Sensor flag (f_{L01}: Head detect)	qfIrgaHead	20 Hz	NA
Sensor flag (f_{L02}: Outlet temperature)	qfIrgaTempOut	20 Hz	NA
Sensor flag (f_{L03}: Inlet temperature)	qfIrgaTempIn	20 Hz	NA
Sensor flag (f_{L04}: Aux input)	qfIrgaAux	20 Hz	NA
Sensor flag (f_{L05}: Differential pressure)	qfIrgaPres	20 Hz	NA
Sensor flag (f_{L06}: Chopper)	qfIrgaChop	20 Hz	NA
Sensor flag (f_{L07}: Detector)	qfIrgaDetc	20 Hz	NA
Sensor flag (f_{L08}: PLL)	qfIrgaPll	20 Hz	NA
Sensor flag (f_{L09}: Sync)	qfIrgaSync	20 Hz	NA
Sensor flag (f_{L10}: AGC)	qfIrgaAgc	20 Hz	-

Table 5. List of eddy-covariance turbulent exchange infrared gas analyzer sampling mass flow controller (mfcSampTurb)-related dp0p DPs that are ingested in this ATBD.

dp0p DP	dp0p Term Name	Sample Frequency	Units
Sampling mass flow rate set point	frtSet00	20 Hz	m^3 s^{-1}
Sampling mass flow rate	frt00	20 Hz	m^3 s^{-1}
Sampling volumetric flow rate	frt	20 Hz	m^3 s^{-1}
Sampling gas pressure	presAtm	20 Hz	Pa
Sampling gas temperature	temp	20 Hz	K

Below **Table 6** - **Table 10** detail the eddy-covariance storage exchange (ECSE) related dp0p DPs used to produce dp01 DPs in this ATBD.

Table 6. List of eddy-covariance storage exchange carbon dioxide (**co2Stor**)-related dp0p DPs that are ingested in this ATBD.

dp0p DP	dp0p Term Name	Sample Frequency	Units
Sampling mass flow rate	frt00	1 Hz	m^3 s^{-1}
Sampling gas pressure	pres	1 Hz	Pa
CO_2 dry mole fraction	rtioMoleDryCo2	1 Hz	mol mol^{-1}
CO_2 wet mole fraction	rtioMoleWetCo2	1 Hz	mol mol^{-1}
Sampling gas temperature	temp	1 Hz	K

Table 7. List of eddy-covariance storage exchange water vapor (**h2oStor**)-related dp0p DPs that are ingested in this ATBD.

dp0p DP	dp0p Term Name	Sample Frequency	Units
Sampling mass flow rate	frt00	1 Hz	$m^3\ s^{-1}$
Sampling gas pressure	pres	1 Hz	Pa
H_2O dry mole fraction	rtioMoleDryH2o	1 Hz	mol mol^{-1}
H_2O wet mole fraction	rtioMoleWetH2o	1 Hz	mol mol^{-1}
Sampling gas temperature	temp	1 Hz	K

Table 8. List of eddy-covariance storage exchange carbon dioxide isotope (**isoCo2**)-related dp0p DPs that are ingested in this ATBD.

dp0p DP	dp0p Term Name	Sample Frequency	Units
Ratio of stable isotopes ^{13}C to ^{12}C in CO_2	dlta13CCo2	1 Hz	‰
Gas spectrum ID	idGas	1 Hz	NA
Pressure	pres	1 Hz	Pa
$^{12}CO_2$ in CO_2 dry mole fraction	rtioMoleDry12CCo2	1 Hz	mol mol^{-1}
$^{13}CO_2$ in CO_2 dry mole fraction	rtioMoleDry13CCo2	1 Hz	mol mol^{-1}
CO_2 dry mole fraction	rtioMoleDryCo2	1 Hz	mol mol^{-1}
H_2O dry mole fraction	rtioMoleDryH2o	1 Hz	mol mol^{-1}
$^{12}CO_2$ in CO_2 wet mole fraction	rtioMoleWet12CCo2	1 Hz	mol mol^{-1}
$^{13}CO_2$ in CO_2 wet mole fraction	rtioMoleWet13CCo2	1 Hz	mol mol^{-1}
CO_2 wet mole fraction	rtioMoleWetCo2	1 Hz	mol mol^{-1}
H_2O wet mole fraction	rtioMoleWetH2o	1 Hz	mol mol^{-1}
CH_4 dry mole fraction	rtioMoleDryCh4	1 Hz	mol mol^{-1}
CH_4 wet mole fraction	rtioMoleWetCh4	1 Hz	mol mol^{-1}
Instrument status	sensStus	1 Hz	NA
Temperature (temp)	temp	1 Hz	K
Temperature (temp) measured at PICARRO warm box	tempWbox	1 Hz	K

Table 9. List of eddy-covariance storage exchange water vapor isotope (**isoH2o**)-related dp0p DPs that are ingested in this ATBD.

dp0p DP	dp0p Term Name	Sample Frequency	Units
Ratio of stable isotopes 18O:16O in H_2O	dlta18OH2o	1 Hz	‰
Ratio of stable isotopes 2H:1H in H_2O	dlta2HH2o	1 Hz	‰
Pressure	pres	1 Hz	Pa
H_2O dry mole fraction	rtioMoleDryH2o	1 Hz	mol mol^{-1}
H_2O (wet mole fraction)	rtioMoleWetH2o	1 Hz	mol mol^{-1}
Instrument Status	sensStus	1 Hz	NA
Signal to indicate if the instrument is processing the data for N2 gas or background air. 0=air mode, 1=N2 mode	stusN2	1 Hz	NA
Temperature (temp)	temp	1 Hz	K

dp0p DP	dp0p Term Name	Sample Frequency	Units
Temperature (temp) measured at PICARRO warm box	tempWbox	1 Hz	K
State of external solenoid valves if attached to PICARRO L2130-i	valvCrdH2o	1 Hz	NA

Table 10. List of eddy-covariance storage exchange sampling mass flow controller (**mfcSampStor**)-related dp0p DPs that are ingested in this ATBD.

dp0p DP	dp0p Term Name	Sample Frequency	Units
Sampling mass flow rate set point	frtSet00	1 Hz	m3 s-1
Sampling mass flow rate	frt00	1 Hz	m3 s-1
Sampling volumetric flow rate	frt	1 Hz	m3 s-1
Sampling gas pressure	presAtm	1 Hz	Pa
Sampling gas temperature	temp	1 Hz	K

In addition, standard NEON TIS sensor plausibility tests are applied at dp00 temporal resolution to the dp0p DPs listed above. The corresponding pass/fail flags per **Table 11** are generated for each test according to AD[06]. (Note. We will not be carrying out the "gap test" or "null test" since the regularization is being applied according to AD[07]).

Table 11. Plausibility quality flags to be applied to all dp0p DPs

Flag	Term modifier	Description
QF_{Cal}	qfCal	Quality flag for the Invalid Calibration test
QF_{Pers}	qfPers	Quality flag for the Persistence test
QF_{Rng}	qfRng	Quality flag for the Range test
QF_{Step}	qfStep	Quality flag for the Step test

The flags are applied to all dp0p DP following a uniform naming convention, whereby the dp0p DP term name is augmented with the plausibility test flag term modifier. For example, the quality flag for the step test for measured along-axis wind speed will be "qfStepVeloXaxs".

Solenoid flags that indicate ECTE validation periods are outlined in **Table 12**. These data do not have quality flag information associated with the measurements.

Table 12. List of ECTE IRGA solenoid for validation gas system in NEMA enclosure (valvWaliNemaTurb) -related dp0p DPs that are ingested in this ATBD

dp0p DP	dp0p Term Name	Sample Frequency	Units
Validation gas 1-5 status NEMA enclosure	qfGas01 – qfGas05	0.2 Hz	NA

3.4 Product Instances

Each NEON Core site with terrestrial infrastructure will produce an instance of the reported variables in **Table 1**. Each NEON gradient site will produce an identical instance of the reported variables in **Table 1** except for Atmospheric H_2O isotopes.

3.5 Temporal Resolution and Extent

The temporal resolution/extent of all reported variables in ECTE instrument system under **Table 1** is 1 min and 30 min. The temporal resolution of most input variables in ECTE instrument system in **Table 2** - **Table 5** is 0.05 s (20 Hz) (design described in AD[01]), with exception of the mean **soniAmrs** variables collected at measurement frequency of 0.025 s (40 Hz). The temporal extent of all input variables is 0.5 h, i.e. a data set of 0.5 h duration shall be considered for each implementation of the presented algorithms.

The temporal resolution/extent of dp01 CO_2 and H_2O concentration data products under **Table 1** is 2 min (or 9 min) and 30 min for ECSE instrument system (dp01, sect. 6.2.1.2). The temporal resolution of all input variables in **Table 6** and **Table 7** is 1 s (1 Hz). The temporal extent of all input variables is 1 day, i.e. a data set of 1 day duration shall be considered for each implementation of the presented algorithms.

3.6 Spatial Resolution and Extent

The input variables for ECTE used in this ATBD are measured at a single position in space at the tower top. Consequently both, input variables and reported variables are not spatially resolved. The 3D boom accelerations (acc_x, acc_y, acc_z) are point measurements. The spatial extent (path length) of all remaining variables is ≈10 cm (AD[01]). The spatial representativeness of the means, variances and covariances reported in this ATBD is a function of several factors such as measurement height $d_{z,m}$, displacement height $d_{z,d}$, wind speed and direction, atmospheric stability and surface roughness. From dispersion modeling (e.g., Schmid, 1994; Vesala et al., 2008) it is found that $\approx 10\ (d_{z,m} - d_{z,d}) < d_{x,FP90} < 100\ (d_{z,m} - d_{z,d})$, where $d_{x,FP90}$ is the cross-wind integrated upwind extent from within which 90% of a measured flux value is sourced. The spatial representativeness for each observation of the reported variables will be quantified during the implementation of AD[04].

ECSE analyzers (CO_2 and H_2O gas analyzer and isotopic CO_2 and H_2O analyzers) are located inside the instrument hut at the bottom of the tower. However, the analyzer's measurements reflect the points in space where the gas sample inlets are located on the tower infrastructure at different vertical levels, which will be site-specific. The array of aspirated air temperature measurements is made at the same vertical heights as above gas sample inlets. The vertical profile measurements of the air temperature, CO_2 and H_2O concentration will be integrated into higher-level derived data products of time rate of change (dp02, sect. 6.2.1.3), vertical resolved time rate of change (dp03, sect. 6.2.1.3) and storage flux (dp04, sect. 0).

4 OVERALL ALGORITHMIC IMPLEMENTATION

NEON utilizes the eddy4R-Docker EC data processing environment (Metzger et al., 2017) to routinely perform the calculations outlined in the following sections. eddy4R-Docker relies on the eddy4R family of open-source packages for EC raw data processing, analyses and modeling in the R Language for Statistical Computing (R Core Team, 2016), wrapped into a Docker filesystem that contains only the minimal context needed to run. The eddy4R-Docker EC data processing environment is publicly available and extensible, and continuously solicits community input through a Development and Systems Operations (DevOps) approach (**Figure 2**). At the time of writing, the repository and a detailed Wiki are being prepared for public access.

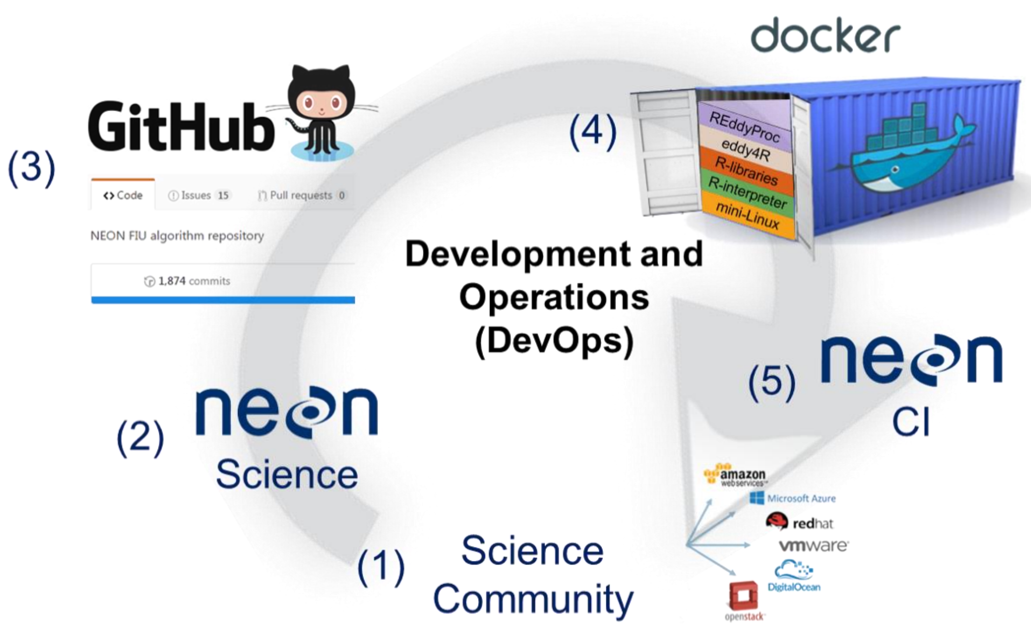

Figure 2. NEON's DevOps framework consists of a periodic sequence: The science community contributes algorithms and best practices (1) which together with NEON Science (2) are compiled into eddy4R packages via the GitHub distributed version control system (3). NEON Science releases an eddy4R version from GitHub, which automatically builds an eddy4R-Docker image on DockerHub as specified in a "Dockerfile" (4). The eddy4R-Docker image is immediately available for deployment by NEON Cyberinfrastructure (CI; 5), the Science Community (1) and NEON Science (2) alike. This DevOps cycle can be repeated for continuous development and integration of requests and future methodological improvements, resulting in the next release.

To then perform a defined series of processing steps, the eddy4R-Docker image is called with an instruction set, resulting in a running instance called Docker container (**Figure 3**). Through this mechanism, an arbitrary number of eddy4R-Docker containers can be run simultaneously performing identical or different services depending on the workflow file. This provides an ideal framework for scaled deployment using e.g. high-throughput compute architectures, cloud-based services etc.

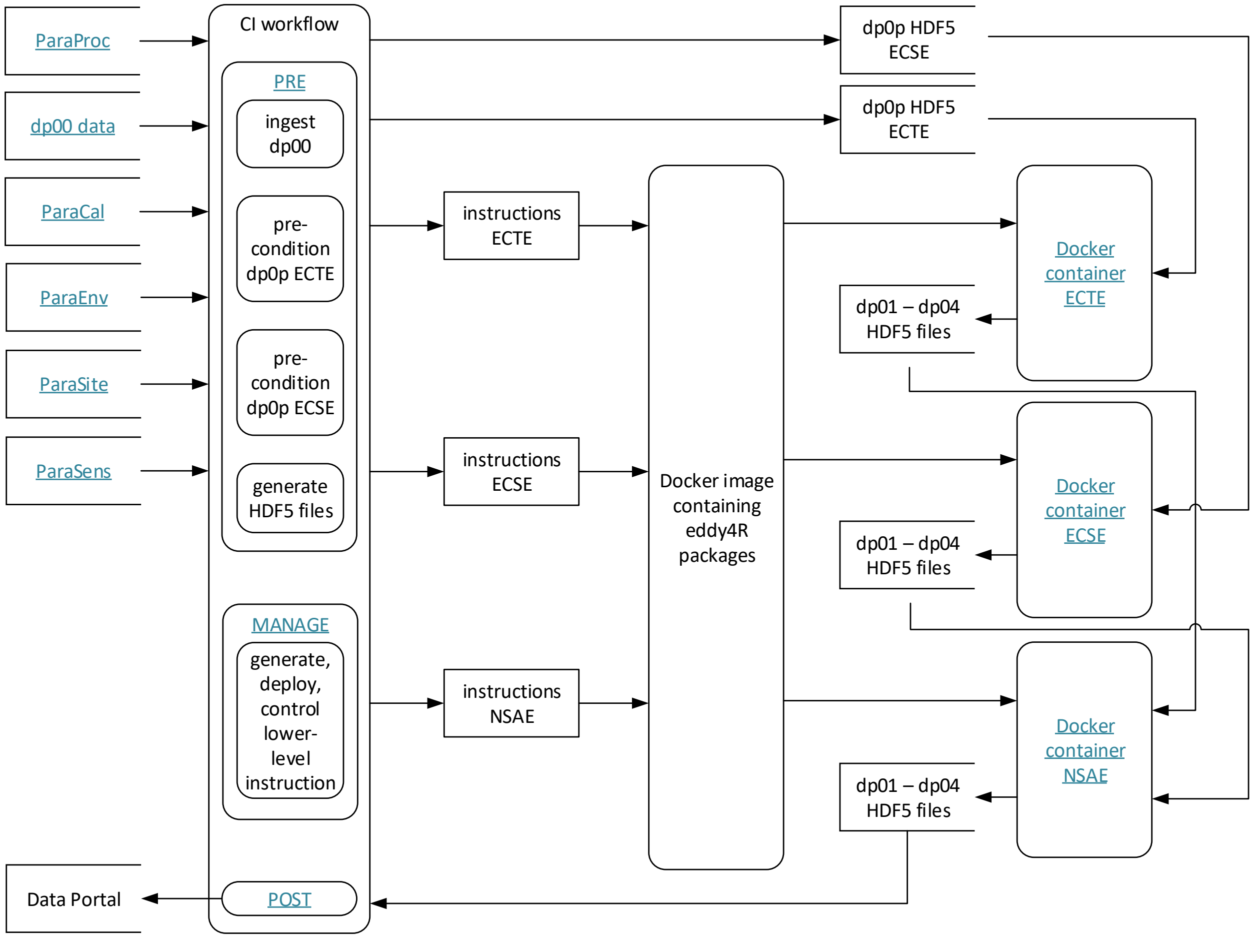

Figure 3. NEON's eddy4R-Docker EC processing framework. Individual components are described in the text.

The overall processing framework begins with ingesting information from various data sources on a site-by-site basis (**Figure 3** top left). This includes EC raw data (Level 0, or dp00 data) alongside contextual information on measurement site (ParaSite), environment (ParaEnv), sensor (ParaSens), calibration (ParaCal), as well as processing parameters (ParaProc). Next, the raw data is preconditioned and all information is hierarchically combined into a compact and easily transferable HDF5 file (**Figure 3** panel "CI workflow"). Each file contains the calibrated raw data (dp0p) and metadata for one site and one day, either for EC turbulent exchange or storage exchange. Together with the corresponding turbulence (ECTE), storage (ECSE) or net surface-atmosphere exchange (NSAE) instruction sets the HDF5 dp0p data file is passed to the eddy4R-Docker image, where a running Docker container is spawned that performs the specified computations (**Figure 3** top right). The resulting higher-level data products (Level 1 – Level 4, or dp01-dp04) are collected and, together with all contextual information, are combined into a daily dp01-dp04 HDF5 data file that is served on the data portal (**Figure 3** bottom left). In addition to the daily output files, monthly concatenated files are also available for download from the NEON data portal.

At the time of writing, the processing described in this ATBD is actively being rolled out across NEON sites, resulting in varying site and temporal coverage for download from the NEON Data Portal. During subsequent nominal operations, we plan to produce and publish the data products in three phases, to accommodate a variety of use cases: the initial near-real-time transition, a science reviewed quality transition, and the epoch yearly transition. The initial near-real-time transition is scheduled to process daily files at a 5-day delay after data collection to accommodate a 9-day centered planar-fit window (see Sect. 5.2.1.2 for details). If the data has not been received from the field it will attempt to process daily for 30 days, and if not all data is available after this window a force execution is performed populating a HDF5 file with metadata and filling data with NaN's. The monthly file will be produced after all daily files are available, no later than 30 days after the last daily file was initially attempted to be processed. An example of this transition schedule is outlined in **Table 13**.

Table 13. Initial data processing transition schedule example.

Date (data recorded at site)	**2017-09-01**	**2017-09-30**	**comments**
Date (first daily processing attempt)	2017-09-06	2017-10-05	Processing needs to accommodate 5 days delay (for 9 day planar-fit window)
Date (last daily processing attempt)	2017-10-06	2017-11-04	Try for up to 30 days
Date (daily processing force execution)	2017-10-07	2017-11-05	After last daily processing attempt (this example: 1 day after), process eddy4R with dp0p file that is populated with metadata + data (NaNs)
Date (first monthly processing attempt)	-	2017-10-06	First opportunity for monthly processing to succeed (provided all daily file processing attempts succeeded)
Date (last monthly processing attempt)	-	2017-11-05	Try for up to 30 days
Date (monthly processing force execution)	-	2017-11-06	After last monthly processing attempt (this example: 1 day after)

After the initial transition, the NEON science team has a one month window to manually flag data that were identified as suspect through field-based problem tracking and resolution tickets or through additional manual data quality analysis. Then, the science-reviewed transition will occur, and the data will be republished to the data portal. The last transition type is part of the yearly epoch versioning, which provides a fully quality assured and quality controlled version of the data using the latest full release of the processing code. This transition is scheduled to occur 18 months after the initial data collection, which is to provide sufficient time for all sensors to be re-calibrated in the calibration and validation laboratory (CALVAL) to determine and apply drift corrections.

5 TURBULENT EXCHANGE

The calculation of eddy-covariance momentum, heat, water vapor and carbon dioxide fluxes provides higher-level DPs with ecological relevance. These DPs have many applications within ecology and atmospheric science, and play a crucial role in constraining, calibrating and validating process-based models (e.g., Rastetter et al., 2010). This shall enable the detection of continental scale ecological change and the forecasting of its impacts.

5.1 Theory of Measurement

The exchange of momentum, heat, water vapor, CO_2 and other scalars between the earth's surface and the atmosphere is mainly governed by turbulent transport. Buoyancy as well as shear stress result in a turbulent wind field for most of the day (e.g., Stull, 1988). The eddy-covariance (EC) technique measures the properties of the turbulent wind field directly. This makes it the least invasive method currently available for direct and continuous observations of the surface-air exchange. The technique is based on the concept of mass conservation and makes use of the Reynolds decomposition (isolation of mean and fluctuating part) of relevant terms in the Navier-Stokes equation (e.g., Foken, 2008; Stull, 1988). With several restrictions (AD[04]) the net flux *F* into or out of an ecosystem can be expressed as (e.g., Loescher et al., 2006);

$$F = \underbrace{\int_0^{d_{z,m}} \frac{\partial \bar{X}}{\partial t} dz}_{\text{I}} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \overline{u'X'}}{\partial x} dz}_{\text{II}} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \overline{v'X'}}{\partial y} dz}_{\text{III}} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \overline{w'X'}}{\partial z} dz}_{\text{IV}} \tag{1}$$

$$+ \underbrace{\int_0^{d_{z,m}} \frac{\partial \bar{u}\bar{X}}{\partial x} dz}_{\text{V}} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \bar{v}\bar{X}}{\partial y} dz}_{\text{VI}} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \bar{w}\bar{X}}{\partial z} dz}_{\text{VII}},$$

with overbars denoting means, and primes denoting deviations from the mean. Here, *X* is a scalar quantity such as H_2O or CO_2 mixing ratios; *u*, *v* and *w* are along-, cross-, and vertical wind speeds with respect to the Cartesian coordinates *x*, *y*, and *z*; *t* is time, and $d_{z,m}$ is the measurement height. Term I in Eq. (1) represents the positive or negative rate of change of X in the vertical column below the sensor, equivalent to storage. Terms II–IV represent the turbulent flux divergence, and terms V–VII represent advection through the layer between the surface and sensor. If the conditions at the measurement site fulfill several assumptions (details provided in AD[04]), terms I–III and V–VII cancel from Eq. (1), and term IV can be further simplified to;

$$F = \overline{w'X'}. \tag{2}$$

That is, in this case the net flux into or out of an ecosystem can be expressed as the covariance between the vertical wind and the scalar, which can be computed from ECTE measurements alone. Whether or not this reduction of Eq. (1) is valid is assessed in a series of tests during the complete implementation of AD[04]. Wherever possible, auxiliary measurements will be used to re-substitute non-negligible terms in Eq. (2), e.g. the storage term I (Sect. 6).

5.2 Data Analysis

5.2.1 Theory of Algorithm

The subject of this ATBD is the mathematical derivation of statistical quantities in Eq. (1). These quantities are used to (i) express the net flux according to Eq. (2), and (ii) quantify the fulfillment of assumptions on the site conditions during the implementation of AD[05].

5.2.1.1 De-spiking

The time series signal despiking algorithm by (Brock, 1986) is used, including the additional threshold by (Starkenburg et al., 2016). This study concluded that the median filter approach resulted in robust despiking results with little to no misclassification of spikes.

5.2.1.2 Planar-fit Coordinate Rotation

After careful consideration of several options, the planar-fit coordinate rotation method is used to align the vector basis of the mass conservation Eq. (1) with the average streamlines over a synoptic timescale (default: 9 days). **Table 14** provides an overview of the advantages and disadvantages of three principal coordinate rotation methods. The main advantages of the double rotation method (Kaimal and Finnigan, 1994; McMillen, 1988; Tanner and Thurtell, 1969) are its applicability for online flux computation, and its robustness against alignment changes of the sonic anemometer. However, the method also suffers from several disadvantages which are overcome by the planar-fit (Kondo and Sate, 1982; Lee et al., 2004; Mahrt et al., 1996; Wilczak et al., 2001) and surface-fit (Baldocchi et al., 2000; Finnigan, 1999; Lee, 1998; Paw U et al., 2000) methods. In particular, instrument offsets, low wind periods or transient mean vertical flows $\overline{w} \neq 0$ can result in over-rotation. E.g., 0.05 m s^{-1} mean vertical flow at 2 m s^{-1} mean horizontal flow results in 1.5° over-rotation. Errors >10% per 1° over-rotation and ≤5% per 2° over-rotation have been reported for measurements of shear stress (Wilczak et al., 2001) and scalar flux (Lee et al., 2004), respectively. These errors are not distributed randomly, but a function of the 3-D flow pattern at the measurement site. Over complex terrain with diurnal flow patterns, resulting biases of the daily flux integrals in the order of 5% have been observed (Turnipseed et al., 2003).

In contrast, the planar-fit and surface-fit methods eliminate over-rotation by identifying and distinguishing (i) an ensemble mean regression offset, and (ii) the transient mean vertical flows during each averaging period. For flows over uniformly tilted slopes the planar-fit method is applicable, while over more complex surfaces only the surface fit method is capable of this differentiation, because it not only considers wind direction (e.g., sectorial planar fit), but also wind magnitude. The transient mean vertical flows can contribute up to 25% to the total surface-atmosphere exchange over complex topography (Finnigan et al., 2003), and are only quantifiable with latter regression methods. Moreover, these methods (i) avoid high-pass filtering and cross-axis folding, (ii) provide a consistent frame of reference to assess the quality of the flux measurements over multiple days, (iii) enable tracking of the instrument alignment, and (iv) enable quantification of the uncertainty related to coordinate rotations.

In an application of the sectorial planar-fit method, Yuan et al. (2011) find that two hundred 30-min data sets (i.e., little over four days of data) are sufficient to derive stable planar fit coefficients. Based on the time-frequency work by Xu et al. (2017), we extended the initial planar fit period to 9 days. This intends to be long enough to capture a full cycle of synoptic-scale atmospheric motions, and to be short enough to resolve changing ecosystem phenology, both of which physical determinants of a stable (and meaningful) aerodynamic reference plane. As with all processing parameters, 9 days represents an observatory-wide initial value that is intended to be adjusted on site-specific (and potentially season-specific) basis during nominal operations of the NEON.

Table 14. Properties of three coordinate rotation methods for aligning the vector basis of the mass conservation equation with the mean streamlines. Advantages of individual methods are highlighted with underline.

Property	**Double rotation**	**Planar fit**	**Surface fit**
References	Kaimal and Finnigan (1994); McMillen (1988); Tanner and Thurtell (1969)	Kondo and Sate (1982); Lee et al. (2004); Mahrt et al. (1996); Wilczak et al. (2001)	Baldocchi et al. (2000); Finnigan (1999); Lee (1998); Paw U et al. (2000)
Vector basis	Average streamline	Aerodynamic plane	Aerodynamic surface
Data basis	Individual averaging period	Ensemble of averaging periods	Ensemble of averaging periods
Computation	<u>Real-time</u>	Delayed	Delayed
Change in anemometer alignment	<u>Automatic adaptation</u>	New set of rotation angles required	New set of rotation angles required
Over-rotation	Problematic	<u>Eliminated for simple slopes</u>	<u>Eliminated for complex terrain</u>
Information on vertical advection	No	<u>For simple slopes</u>	<u>For complex terrain</u>
High-pass filtering and cross-axis folding	Yes	<u>No</u>	<u>No</u>
Consistent vector basis for flux QA/QC	No	<u>Yes</u>	<u>Yes</u>
Tracking instrument tilt	No	<u>Yes</u>	<u>Yes</u>
Uncertainty propagation	No	<u>Yes</u>	<u>Yes</u>

5.2.1.3 Infrared gas analyzer validation correction

The readings of environmentally-exposed infrared gas analyzers (IRGAs) can be affected by several types of drift. Diurnal variations of incoming radiation and air temperature can cause small short-term temperature drifts on the order of hours. Cell contamination can lead to longer-term drifts between cell cleaning events, on the order of weeks to months. Aging of electronic components in uncalibrated instruments can lead to long-term drifts on the order of years. Left unmitigated, these drifts can lead to hourly flux errors on the order of several percent, which can aggregate to significant biases in estimating the terrestrial carbon sink, e.g. in the form of annual budgets. This section describes a spatio-temporal hierarchy of validation measurements and a corresponding algorithmic process to consistently and continuously correct drifts in IRGA measurements.

Specifically, every 23.5 hours traceable gas standards of known concentrations are applied to the IRGA. This allows correction of any potential consistent drifts on the order of days to years, which can be achieved by measurement of a zero gas and three traceable known CO2 concentrations (hereafter, validation gases) supplied to the IRGA through a solenoid valve at a flow rate of ~1.5 L min^{-1}. We determined that with the system volume of approximately 0.1 L, a flushing time of 3.5 min at a flow rate of ~1.5 L min^{-1} is sufficient to flush the sample cell and ensure no residual sample from the previous stream is left (AD[01]). Therefore, each validation gas is sampled 5 min and the data from the last 1.5 min will be used for the actual validation.

For each validation cycle, the slope and offset will be determined by the linear relationship between the validation gases and the IRGA's measured values (Eq. (3) and **Figure 4**(a)). Linear function relationships will be estimated by the maximum likelihood (Ripley and Thompson, 1987) method which considers the error in the measured values and the references (validation gas).

$$y = \left(c_{1,\mathrm{CO2,MLFR}} \pm SE\right) \cdot x + \left(c_{2,\mathrm{CO2,MLFR}} \pm SE\right) \tag{3}$$

Where x is the measured CO2 dry mole fraction, y is the known CO2 dry mole fraction, $c_{1,\mathrm{CO2,MLFR}}$ is the resulted slope from MLFR for each validation cycle, $c_{2,\mathrm{CO2,MLFR}}$ is the resulted zero offset from MLFR for each validation cycle, and SE is the resulted standard error from MLFR.

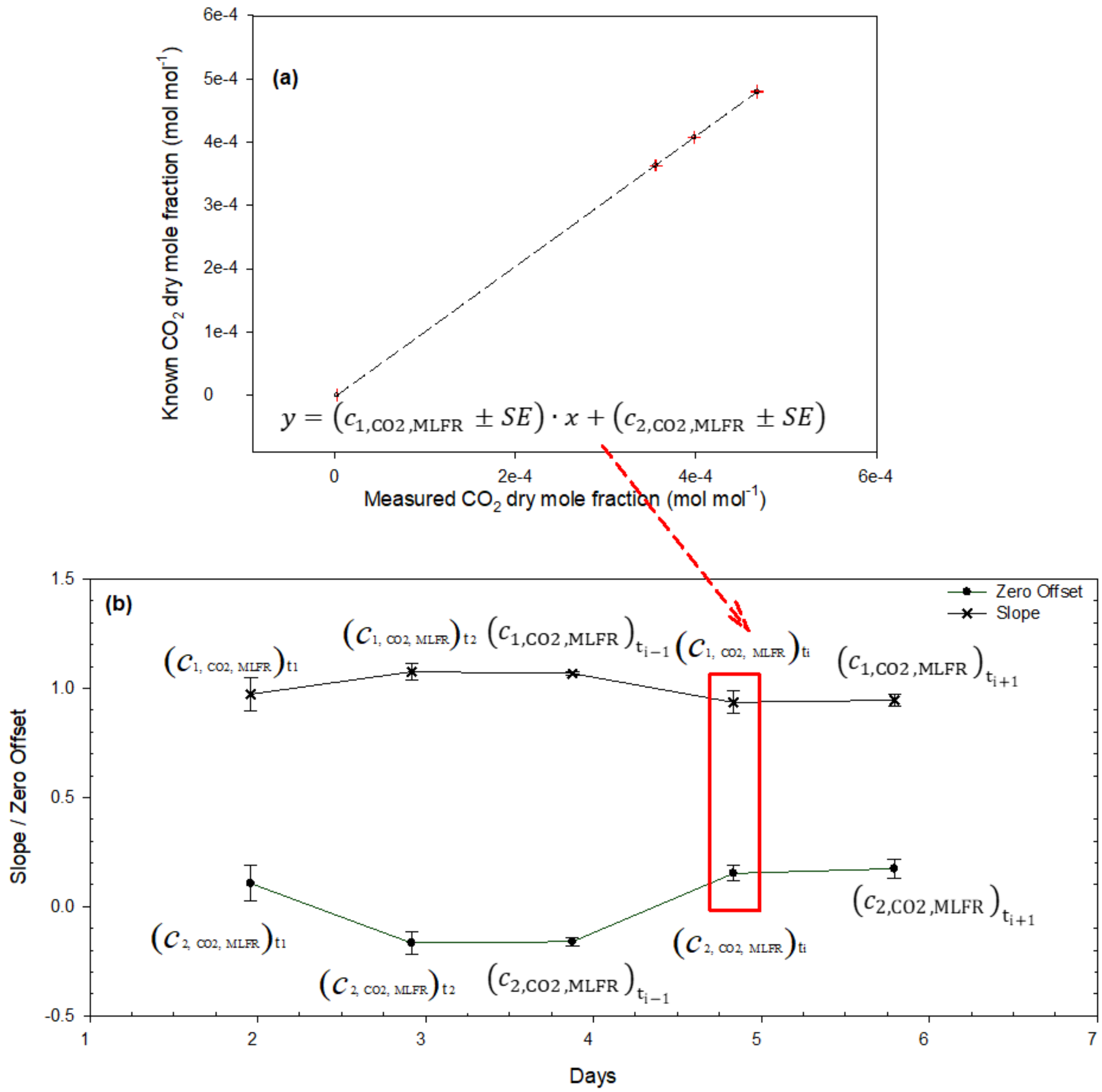

Figure 4. (a) The linear relationship between the known CO2 of the validation gas and measured CO2 from the IRGA using the maximum-likelihood fitting (b) slope and zero offset with their standard errors for each validation cycle, determined by (a). Dashed line in (a) is the maximum-likelihood fitting of a functional relationship (MLFR) line. For each data point in (a), bars indicate standard error of mean CO2 for the known validation gas and IRGA measurements.

By assuming that the slope and zero offset change linearly between each pair of adjacent validation cycles (Welles and McDermitt, 2005), the time-series of slope and zero offset can be interpolated as:

$$\left(c_{1,\mathrm{CO2,int}}\right)_{\mathrm{t}} = \left(c_{1,\mathrm{CO2,MLFR}}\right)_{\mathrm{t_i}} + \left[\left(c_{1,\mathrm{CO2,MLFR}}\right)_{\mathrm{t_{i+1}}} - \left(c_{1,\mathrm{CO2,MLFR}}\right)_{\mathrm{t_i}}\right]\frac{(\mathrm{t}-\mathrm{t_i})}{(\mathrm{t_{i+1}}-\mathrm{t_i})} \tag{4}$$

$$\left(c_{2,\mathrm{CO2,int}}\right)_{\mathrm{t}} = \left(c_{2,\mathrm{CO2,MLFR}}\right)_{\mathrm{t_i}} + \left[\left(c_{2,\mathrm{CO2,MLFR}}\right)_{\mathrm{t_{i+1}}} - \left(c_{2,\mathrm{CO2,MLFR}}\right)_{\mathrm{t_i}}\right]\frac{(\mathrm{t}-\mathrm{t_i})}{(\mathrm{t_{i+1}}-\mathrm{t_i})} \tag{5}$$

where $\mathrm{t_i} < \mathrm{t} < \mathrm{t_{i+1}}$, t_i is the previous validation cycle, and t_{i+1} is the next validation cycle (**Figure 4**(b)). The corrected CO2 dry mole fraction at time t can then be calculated as:

$$(rtioMoleDryCo2Cor)_{\mathrm{t}} = \frac{(rtioMoleDryCo2)_{\mathrm{t}} - \left(c_{2,\mathrm{CO2,int}}\right)_{\mathrm{t}}}{\left(c_{1,\mathrm{CO2,int}}\right)_{\mathrm{t}}} \tag{6}$$

5.2.1.4 Lag-correction

Application of Eq. (1) requires that the instantaneous vertical wind *w* and scalar *X* are measured at the same place and at the same time, which is not presently possible, primarily due flow distortion issues with the sonic anemometer. Consequently, before applying Eq. (1), the recorded time series must be adjusted by a certain time lag to ensure spatiotemporal coincidence. The delay between the two time series is mainly caused by differences in electronic signal treatment, spatial separation between wind and scalar sensors, and air travel through the tubes in closed-path gas analyzers. Assuming joint stationarity, the lag time *l* can be estimated for each averaging interval by performing a cross correlation analysis between the quantities of interest;

$$\mathrm{abs}\left(\frac{\overline{w\prime(t)\cdot X\prime(t+l)}}{\overline{w\prime}\cdot\overline{X\prime}}\right) \rightarrow \mathrm{max}, \tag{7}$$

for samples collected at times *t* and *t+l*. This is equivalent to comparing the correlations between the quantities lagged by different delays (**Figure 5**). The time lag that results in the highest correlation is selected. However, when correlations are small this procedure can result in ambiguous lag times. Hence, high-pass filtering and pre-defining the maximum size of the cross-correlation search window aids in constraining the lag times to physically feasible values. The maximum size of the search window is found on the basis of known electronic delays, sensor separation and typical wind speeds, as well as mass flow and tube dimensions of closed-path gas analyzers. In cases where these limits are exceeded, Rebmann et al. (2012) recommend to use the value of the preceding averaging interval.

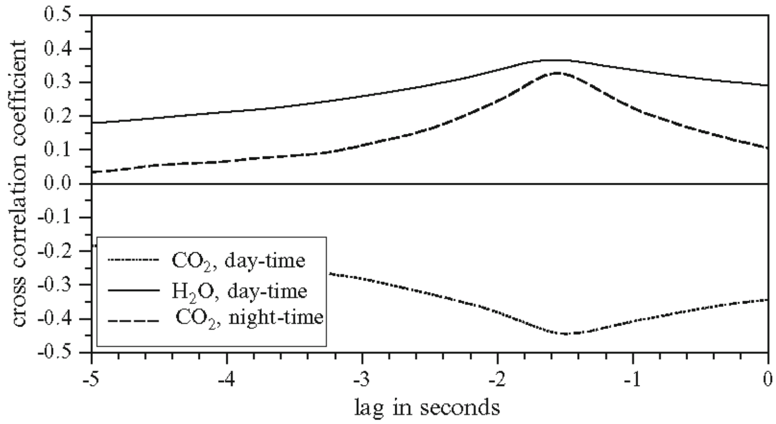

Figure 5. From Rebmann et al. (2012): Cross-correlation between the vertical wind component and CO_2 and H_2O for different lag times.

5.2.1.5 Sonic temperature conversions

A cross-wind correction (Campbell Scientific, 2011; Liu et al., 2001) is not necessary for the sonic anemometer operated at NEON sites (Campbell Scientific Inc., model CSAT-3 firmware: 3.0f; Logan, Utah, USA; Appendix C). However, the speed of sound in air is not only a function of air temperature, but also of humidity. Hence the temperature measurement by an ultrasonic anemometer/thermometer (SONIC) T_{SONIC} does not equal the air temperature, but includes a cross-dependence on humidity. A conversion is required to cancel this humidity dependence and to yield means, variances and covariances of air temperature T_{air}, respectively (Schotanus et al., 1983);

$$\overline{T_{\mathrm{air}}} = \frac{\overline{T_{\mathrm{SONIC}}}}{1+0.51\,\overline{FW_{\mathrm{mass,H2O}}}}, \tag{8}$$

$$\overline{T_{\mathrm{air}}'^2} = \overline{T_{\mathrm{SONIC}}'^2} - 1.02\,\overline{T_{\mathrm{air}}}\,\overline{T_{\mathrm{air}}' FW_{\mathrm{mass,H2O}}'} - (0.51\,\overline{T_{\mathrm{air}}})^2\,\overline{FW_{\mathrm{mass,H2O}}'^2}, \tag{9}$$

$$\overline{w' T_{\mathrm{air}}'} = \overline{w' T_{\mathrm{SONIC}}'} - 0.51\,\overline{T_{\mathrm{air}}}\;\overline{w' FW_{\mathrm{mass,H2O}}'}, \tag{10}$$

with wet mass fraction (specific humidity) $FW_{\mathrm{mass,H2O}}$. Eqs. (8)–(10) are linear approximations and ignore higher-order terms in their exact definitions. The magnitude of these conversions is in the order of 1–2%, and the accuracy of the approximation for temperature is ≤0.03 K for $0<FW_{\mathrm{mass,H2O}}<40$ g kg^{-1} H_2O, i.e. better than the accuracy of a sonic thermometer. It can be seen that the conversion of variance and covariance (Eqs. (9)–(10)) are subject to cross-dependence on ambient temperature T_{air} in terms $\overline{T_{\mathrm{air}}}$ and $\overline{T_{\mathrm{air}}' FW_{\mathrm{mass,H2O}}'}$ on the right-hand side. Hence Eqs. (9)–(10) must be solved iteratively, by first substituting T_{air} in respective terms on the right hand side with T_{SONIC}, and subsequently updating T_{air} with the outcomes after each cycle until the results for Eqs. (9)–(10) change by no more than 0.01% between iterations (e.g., Mauder and Foken, 2011). Moreover, T_{SONIC} closely resembles the virtual temperature T_{v}, with a difference in the humidity-related conversion in the order of 0.1%;

$$\overline{T_{\mathrm{air}}} = \frac{\overline{T_\mathrm{v}}}{1+0.61\,\overline{FW_{\mathrm{mass,H2O}}}}. \quad (11)$$

Consequently, $\overline{w'T'_{\mathrm{SONIC}}}$ is often used as surrogate for the buoyancy flux, e.g. in the computation of the Monin-Obukhov length (Rebmann et al., 2012).

5.2.1.6 Calculation of means, variance and standard error

The arithmetic mean of a quantity *X* (such as wind components *u*, *v*, *w*) with sample size *N* is calculated as;

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_\mathrm{i}. \quad (12)$$

From here, the sample variance (*N*−1) and standard deviation of *X* are calculated;

$$\overline{X'^2} = \frac{1}{N-1}\sum_{i=1}^{N}(X_\mathrm{i} - \bar{X})^2, \quad (13)$$

$$\frac{\mathrm{std_{err}}(X)=\sqrt{\overline{X'^2}}}{\sqrt{N}}. \quad (14)$$

5.2.1.7 High-frequency spectral correction

EC measurement systems, like all instruments, act as filters, removing both high- and low-frequency components of a signal. High-frequency losses are mainly due to inadequate sensor frequency response, line averaging, sensor separation and, in closed-path infrared gas analyzer (IRGA) systems, air transport through a filter and a tube (Foken et al., 2012). Figure 6 schematically illustrates the impact of high frequency loss in the measurement of an atmospheric scalar *X*, such as CO_2 or H_2O dry mole fraction, on spectral density. The frequency range of attenuation depends on the instrumental setup and especially the length and conditioning of the sample tube. It is often confined to frequencies beyond the spectral peak, which is referred to as the inertial subrange (ISR) of atmospheric turbulence, and for the NEON ECTE system design beyond 1 Hz under most conditions (Metzger et al., 2016). As can be seen from Eq. (2), high frequency losses in *X* (and, to a lesser degree, *w*) propagate into the ECTE flux measurement, and the corresponding cospectrum CO(*w*,*X*). Low-frequency losses result from the finite sampling duration, with the averaging period not always being sufficiently long to include all relevant low frequencies. The subject of this section is the correction of high-frequency losses, in order to avoid underestimating the variances and covariances of outputs from ECTE sensors. Low-frequency losses are planned to be addressed as part of future developments (Sect. 8).

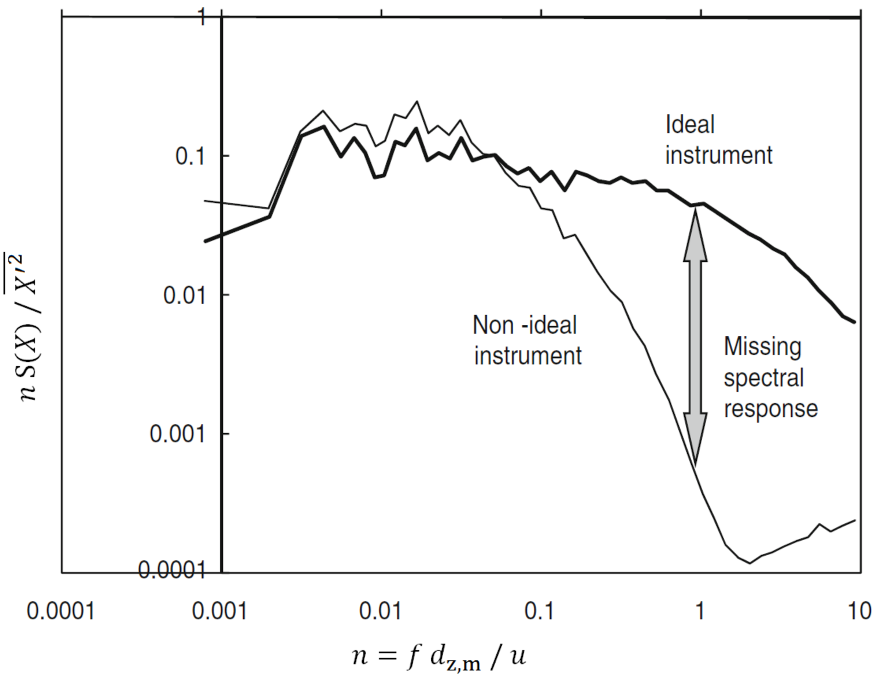

Figure 6. Normalized power spectrum for an ideal instrument which measures the unaffected spectrum of turbulence, and for a non-ideal instrument (Modified after Foken et al. (2012)).

The missing energy between both response curves must be corrected (normalized frequency *n*; measurement frequency *f*, measurement height $d_{z,m}$, along-wind speed *u*, power spectrum and variance of atmospheric scalar *X*, S(*X*) and $\overline{X'^2}$, respectively). The calculation of power spectra are detailed in AD[07].

Nordbo and Katul (2012, in the following referred to as NK12) have presented a Wavelet-based approach to high-frequency spectral corrections which (i) directly corrects the high-frequency data instead of the cospectrum, (ii) corrects each individual averaging period, and is thus able to take into account variations in environmental conditions (e.g., flow rate, relative humidity), (iii) does not assume cospectral similarity with heat, and (iv) does not rely on a theoretical shape for the velocity–scalar cospectrum, thereby making it advantageous to employ in non-ideal conditions. Furthermore, the method is not gas-specific, and can be used with very little input information at various sites. The method's largest insufficiency is its inability to correct attenuation starting already near the peak of power spectra, which however is explicitly taken into account in the design of NEON's ECTE. Consequently this is the method of choice for NEON, as it overcomes the drawbacks of the conventional "theoretical" and "empirical" approaches (Foken, 2017), and is fully automatable. Cospectral attenuation through sensor separation is not considered by the NK12 method. Instead it is explicitly addressed in Sect. 5.2.1.3 through maximization of the cross-correlation and by using an exponential decay model.

NEON currently uses a simplified version of the NK12 method, which is described in Sect. 5.2.2.

5.2.1.8 Footprint modeling

A footprint model is used to determine where on the ground surface emissions measured by the ECTE system originated from. This allows interpretation of observed emission rates against hour-to-hour variations in flux footprint over surface properties such as land cover, soil moisture etc. e.g. from gridded remote-sensing data products (**Figure 7**). An in-depth review into footprint models and their continued development can be found in Leclerc and Foken (2014). Here, we initially use the footprint model described by Metzger et al. (2012). This builds upon the cross-wind integrated footprint model of Kljun et al. (2004), which quantifies the flux contribution relative to the distance away from the measurement position, into the prevailing wind direction. Metzger et al. (2012) coupled the model with a cross-wind distribution function, permitting to spatially resolve also flux contributions perpendicular to the wind direction. We intend to add outputs for the Kljun et al. (2015) footprint model as part of future developments (Sect. 8).

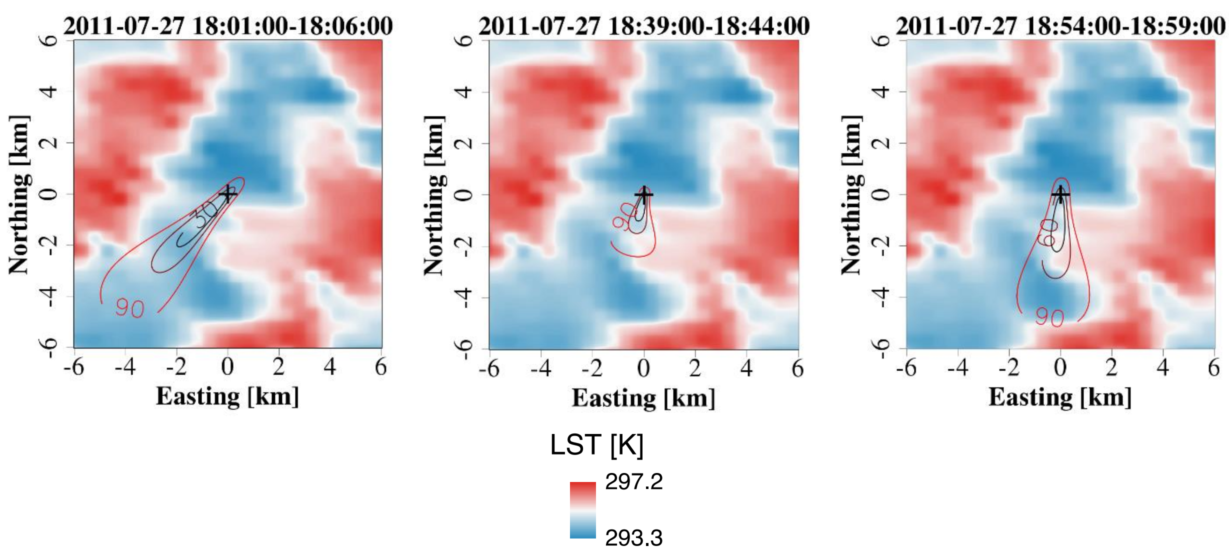

Figure 7. From Xu et al. (2017): example flux footprints (30%, 60% and 90%, contour lines) over MODIS-land surface temperature (LST).

5.2.2 Algorithmic Implementation

The EC turbulent exchange data analysis is implemented as part of the eddy4R-Docker EC processing framework (Sect. 4). The corresponding R workflow flow.turb.tow.neon.dp04.r (link to public GitHub repo in preparation) and its algorithmic sequence are summarized in **Figure 8**.

The calculations described in Sects. 5.2.1.1 – 5.2.1.5 are applied to yield the ECTE dp01 and dp04 data products in **Table 1**. The process is coded and documented in detail in the R-packages eddy4R.base and eddy4R.qaqc, consisting of the following sequence (incl. function references):

- Calculation is performed for datasets of 30 min time resolution (plus 1 min in case of dp01).

- De-spiking is performed at dp0p temporal resolution using the eddy4R.qaqc::def.dspk.br86() function.
- Derived variables at dp0p temporal resolution are calculated using the eddy4R.base::wrap.derv.prd.day() function. For example, specific humidity is calculated from dp0p inputs, so it can later be used in Eqs. (8)–(10).
- Regression of the planar-fit coefficients is performed using the eddy4R.turb::PFIT_det() function over a moving, centered window of 9 days of 20 Hz dp0p data. Data points corresponding to bad sensor diagnostics and spikes are omitted from the regression. The eddy4R.turb::PFIT_apply function is then used to apply the regression coefficients and perform the planar-fit coordinate rotation for the central day of the moving window (day 5).
- The slope and zero offset for a central day, as well as the previous and subsequent days, are determined using the eddy4R.base::wrap.irga.vali() function. The eddy4R.base::def.irga.vali.cor() function is then used to apply the validation correction to the central day of the moving window. The validation results are available in the daily expanded HDF5 file at SITE/dp01/data/co2Turb/000_ver_01m/rtioMoleDryCo2Vali.
- Lag-correction is performed at dp0p temporal resolution using the eddy4R.base::def.lag() function.
- Sonic temperature is converted to air temperature.
- Descriptive statistics are calculated for averaging periods of 1 min and 30 min using the eddy4R.base::wrap.neon.dp01() function, and are available in the HDF5 file at:
    - SITE/dp01/data/amrs
    - SITE/dp01/data/co2Turb (Note that the rtioMoleDryCo2 with validation correction applied are available in the basic package under rtioMoleDryCo2 table and the expanded package under rtioMoleDryCo2 and rtioMoleDryCo2Cor tables. The raw rtioMoleDryCo2 data without the drift correction applied are also available in the expanded package under rtioMoleDryCo2Raw table.)
    - SITE/dp01/data/h2oTurb
    - SITE/dp01/data/soni
- Turbulent fluxes are calculated for averaging periods of 30 min using the eddy4R.turb::REYNflux_FD_mole_dry() function, and are available in the HDF5 file at:
    - SITE/dp04/data/fluxCo2 (Note that the CO_2 fluxes with drift correction applied are available in the basic package under turb/flux and the expanded package under turb/flux and turb/fluxCor. The CO_2 fluxes calculated from DP01 data without the drift correction applied are also available in the expanded package under turb/fluxRaw.)
    - SITE/dp04/data/fluxH2o
    - SITE/dp04/data/fluxMome
    - SITE/dp04/data/fluxTemp
- Wavelet-based high-frequency spectral correction is performed on 30-min basis through the following sequence automated in the wrapper function eddy4R.turb::wrap.wave().

- Periods with missing values >10% are being omitted. For all other periods, < 10% missing values are linearly interpolated.
- The Waves::cwt() function then uses a Morlet mother Wavelet to perform the continuous Wavelet transform of the 3-D wind components, air temperature, as well as H_2O and CO_2 concentration.
- In the eddy4R.turb::def.vari.wave() function the cross-scalograms with the vertical wind are calculated, and the absolute spectral power is scale-wise integrated to co-spectra. Then the power-law coefficient in the ISR of the unweighted co-spectrum is regressed in the frequency range 0.1 … 0.5 Hz. In case the coefficient exceeds the range of −1.8 … −1.3, the standard −5/3 power law decay is used. The reference spectral coefficients following the power slope are calculated, and the transfer function against the observed co-spectra is determined in the frequency range >0.5 Hz. The transfer function is then applied directly to the corresponding cross-scalogram. The ratio of the global Wavelet covariance after and before application of the transfer function provides the flux-specific correction factor, which is applied to the classical EC flux Eq. (2).

- Footprint calculation is performed on 30-min basis though the following sequence.
    - footprint model inputs incl. turbulence statistics are prepared and constrained to within the valid range of the Kljun et al. (2004) parameterization
        - relative measurement height above displacement (distZaxsMeasDisp)
        - wind direction (angZaxsErth)
        - standard deviation of the cross-wind (veloYaxsHorSd) and vertical wind (veloZaxsHorSd)
        - friction velocity (veloFric)
        - roughness length (distZaxsRgh; calculated via call to eddy4R.turb::def.dist.rgh())
        - boundary layer height (distZaxsAbl) is set to 1000 m by default
        - footprint matrix cell size (distReso) is set equal to relative measurement height above displacement, and rounded to 10 m
    - the square footprint weight matrix with 301 x 301 cells and the tower at its center is calculated through calling eddy4R.turb::footK04()
    - footprint statistics are calculated
        - along-wind distance of the 90 percent crosswind-integrated cumulative footprint (distXaxs90)
        - along-wind distance of contribution peak (distXaxsMax)
        - one-sided cross-wind distance of the 90 percent along-wind integrated cumulative footprint (distYaxs90)
    - location of results in HDF5 file
        - model inputs and footprint statistics are included in the basic and expanded HDF5 files: SITE/dp04/DQU/foot/stat
        - half-hourly footprint weight matrices are only included in the expanded HDF5 files: SITE/dp04/DQU/foot/grid/turb

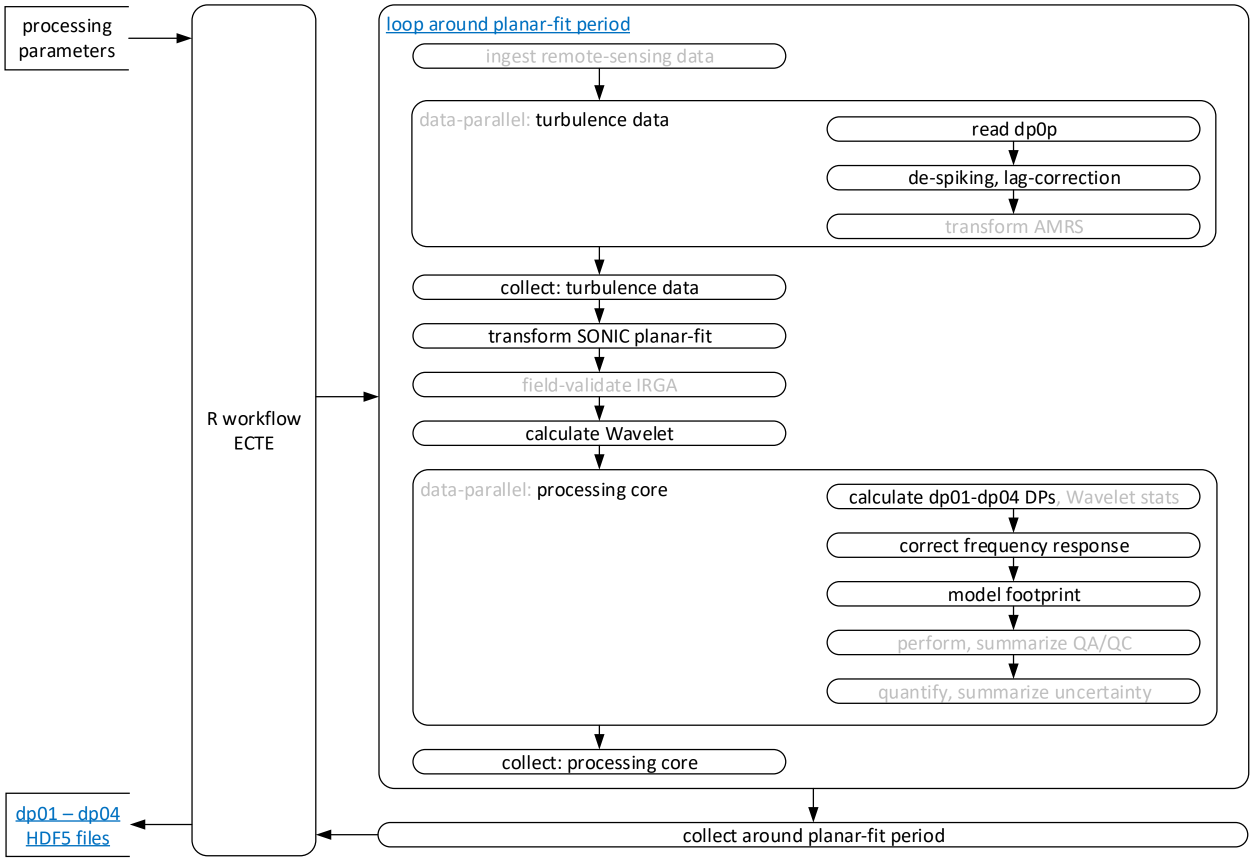

Figure 8. The EC turbulent exchange workflow within the eddy4R-Docker EC processing framework (Sect. 4).

5.3 Quality Assurance and Quality Control analysis

In general, the quality flags (*QF*s) are generated for each test and each *QF* can be set to one of three states as shown in Eq. (15) (AD[06]).

$$QF = \begin{cases} 1 \text{ if the quality test failed} \\ 0 \text{ if the quality test passed} \\ -1 \text{ if NA i.e. not able to be run due to a lack of ancillary data} \end{cases} \qquad (15)$$

In extension, specifically for EC data products, combinations of data quality and data availability signifiers are used to express a number of conditions in the HDF5 file (**Table 15**):

- Condition A: Data are available and good
  - Data are expected and available [Data ≠ NaN]
  - Data pass a critical number of quality tests [QF = 0]
- Condition B: Data are available but bad
  - Data are expected and available [Data ≠ NaN]
  - However, data fail a critical number of quality tests [QF = 1]
- Condition C: Data are available but user discretion advised

- Data are expected and available [Data ≠ NaN]
- However, not all quality tests can be evaluated due to missing dependency data [QF = −1]

- Condition D: Data are not available but expected
  - Data are expected from a particular sensor or measurement level, but are not available [Data = NaN]
  - Data quality cannot be assessed due to missing data or dependency [QF = −1]
  - For example: a quality test requires variables from auxiliary sensors such as the mass flow controller: in the case that mass flow controller data are not available the test cannot be executed and the test result of QF=-1 is assigned.
- Condition E: Data are not available and not expected
  - Data are not expected from a particular sensor or measurement level [Data = NaN]
  - Data quality is not assessed as data is not expected [QF = NA]
  - For example: Profile system with a single analyzer cycles through measurement levels, thus data availability at individual levels is discontinuous. After regularization to create a continuous time series, these data points are not expected to be measured by the analyzer and are represented by NaN, thus the QF = NA.

Table 15. Lookup table of joint data availability and data quality conditions which apply to data products in this ATBD.

	QF = 0	**QF = 1**	**QF= −1**	**QF= NA**
Data ≠ NaN	A	B	C	Not applicable
Data = NaN	Not applicable	Not applicable	D	E

Sensor and statistical QA/QC tests are performed on and reported for the dp0p data (e.g. 20 Hz), while flux QA/QC tests are reported on time-integrated data per flux averaging period (e.g. 30 min). Here, we utilize the NEON data quality framework as described in AD[06] and Smith et al. (2014) to summarize the results from sensors test and QA/QC tests in a way that is transparent and easily interpretable. In the following, these sensor health and statistical QA/QC tests are first aggregated to the flux averaging period, and then combined with the results for flux QA/QC tests to determine the QF_{FINAL}.

5.3.1 Theory of Algorithm

A wide range of qualitative and quantitative algorithmic processing routines are applied to EC data products including:

1. Tests related to sensor diagnostics (AD[02]);
2. Statistical plausibility tests, e.g. range, persistence, step (AD[02] and AD[06]);
3. EC-specific tests based on the degree of fulfillment of one or several methodological assumptions, e.g. detection limit, homogeneity and stationarity, development of turbulence tests.

Sensor diagnostics and statistical plausibility tests are performed on and reported for high frequency data (e.g. 20 Hz), while EC-specific tests are reported on time-integrated data per flux averaging period (e.g. 30 min).

5.3.1.1 Sensor quality flags

Most of quality flags due to the sensor health and statistical plausibility tests are generated as part of the dp0p report variables (AD[02]). In addition, the IRGA validation flag (qfIrgaVali) and IRGA automatic gain control quality flag (qfIrgaAgc) were also generated in this ATBD, which are defined below.

1. **IRGA Validation flag** (qfIrgaVali) – is generated to indicate when the sensor is operated under validation period (1 = validation period, 0 = normal operating condition, −1 = NA). The IRGA validation flag is determined from the IRGA sampling mass flow controller flow rate set point as follow:

$$\mathrm{qfIrgaVali} = \begin{cases} 1 \text{ if } \mathrm{frtSet00} = 0 \\ 0 \text{ if } 0.0001333 \leq \mathrm{frtSet00} \leq 0.00025 \\ -1 \; otherwise. \end{cases} \quad (16)$$

where frtSet00 is the flow rate set point from IRGA sampling mass flow controller (irgaMfcSamp) in the unit of $m^3 s^{-1}$.

2. ***IRGA automatic gain control quality flag*** *(qfIrgaAgc)* is indicating when the sensor is operating with low signal strength using 50 percent as the default threshold (1 = when qfIrgaAgc <= 0.50, 0 = when qfIrgaAgc >= 0.50, -1 = NA).

5.3.1.2 EC-Specific tests

1. Homogeneity and stationarity

Homogeneity and stationarity of the flow field is one of the flux QFQM algorithms used to test fulfillment of the theoretical requirements for EC measurements. Stationarity in the first moment ($qiStnaTrnd$) is tested by comparing the covariances of the original time series with the covariance of the detrended time series (Vickers and Mahrt, 1997). Stationarity in the second moment ($qiStnaSubSamp$) is tested by comparing the covariances determined for the entire 0.5h averaging period with the mean covariance derived from six of 5 min subintervals within this averaging period (Foken and Wichura, 1996). Then the greater (worse) of both test results will be used to indicate the stationarity quality flag. If the quality indicator is found less than or equal to 1, a stationarity quality flag is flagged as zero ($qfStna$ = 0) (Foken et al. 2004); otherwise, the test is failed and the data are flagged ($qfStna$ = 1).

2. Testing the development of turbulence

In the derivation of the eddy-covariance (Eq. 2), it is assumed that the flow field in the layer between the surface and sensor is homogenous and well mixed. This allows for the cancellation of the storage term I in Eq. (1) for low canopies, and the substitution of the integral over the vertical flux divergence with a point measurement of the eddy-covariance. Under these conditions, the ratio between the standard deviation and the scaling variable of an atmospheric quantity X ($X_* = u_*$ for X = *u*, *v*, or *w* and $X_* = T_* = -\overline{w'T'}/u_*$ for X= *T*);

$$XMeasItc = \frac{\sigma(X)}{X_*} \tag{17}$$

is a predictable function of stability of stratification (Wyngaard et al., 1971). These similarity characteristics are called integral turbulence characteristics (ITCs), because they characterize the state of turbulence integral over all frequencies (Tillman, 1972). Based on the flux-variance similarity, ITCs are a suitable measure for describing atmospheric turbulent conditions in the surface layer which can be used to identify whether or not the turbulence is well developed. To use ITC as a measure of flux data quality, the measured $XMeasItc$ are compared against theoretically-derived ITC values $XModlItc$

$$qiItcX = \left|\frac{XMeasItc - XModlItc}{XModlItc}\right|. \tag{18}$$

Parameterizations of $XModlItc$ follow Foken et al., 2004 and Thomas and Foken, 2002.
ITC quality is passed and flagged as zero ($qfItc$ = 0) if the quality indicator ($qiItc$) are less than or equal to 1 (Foken et al. 2004); otherwise, the test is failed and the data are flagged ($qfItc$ = 1).

5.3.1.3 Quality budget (QFQM)

The theory of algorithm, the definition of quality flag (*QF*), quality metric (*QM*), alpha (α) and beta (β) *QF*s and *QM*s are detailed in AD[05]. Each of EC DP will have QF_{FINAL}, QM_α, and QM_β associated with it. Aside from QF_{FINAL}, QM_α, and QM_β, each EC DP will also be accompanied by *QM* results for individual tests, representing the fractional occurrence of each state that a quality flag can take.

In order to determine the QF_{FINAL} (Eq.(19) individually for each DP, the sensor health and statistical plausibility tests are first used to calculate QM_α, and QM_β over the averaging period. However, to be able to calculate QM_α and QM_β, the QF_α and QF_β for each observation need to first be determined. The calculation of QF_α and QF_β are similar, the difference is that α determines whether or not at least one quality flag was set to 1, while β determines whether or not at least one quality flag was set to -1. Then, QM_α, and QM_β over the averaging period can be calculated. Once, QM_α and QM_β were calculated, the QF_{FINAL} can be determined using Eq. (19):

$$\begin{cases} 1 \quad if\ (a \cdot QM_\beta) + (b \cdot QM_\alpha) \geq 20\%\ \text{or} \end{cases}$$

$$QF_{\mathrm{FINAL}} = \quad QF_{\mathrm{sciRevw}} = 1 \qquad (19)$$

$$0 \quad otherwise$$

where a and b are the ratio of QM_{α} to QM_{β} with maximums of 10% for QM_{α} and 20% for QM_{β} (more details can be found in AD[05] and (Smith and Metzger, 2013). Therefore, by default a and b are set to 1 and 2, respectively. If the scientific review flag (QF_{sciRevw}) is set high during science operation management (SOM) review then QF_{FINAL} will be set high.

For the higher level data products (e.g. dp04 turbulent flux), QF_{FINAL} of dp01 is propagated to determine the QF_{FINAL} of each dp04 (see **Table 16**). Then, the results of EC specific (QF_{spec}) tests (i.e., detection limit, homogeneity and stationarity, development of turbulence tests) are taken into account to determine whether the data product is flagged as valid (QF_{FINAL} = 0) or invalid (QF_{FINAL} = 1).

Table 16. Indicates which dp01 final quality flags are propagated to determine the final quality flag of each dp04 turbulent flux.

Turbulent flux (dp04)	dp01
fluxCo2	qfqm/soni/qfFinl/veloZaxsErth and qfqm/co2Turb/qfFinl/rtioMoleDryCo2
fluxH2o	qfqm/soni/qfFinl/veloZaxsErth and qfqm/co2Turb/qfFinl/rtioMoleDryH2o
fluxMome	qfqm/soni/qfFinl/veloZaxsErth and qfqm/soni/qfFinl/veloXaxsErth
fluxTemp	qfqm/soni/qfFinl/veloZaxsErth and qfqm/soni/qfFinl/tempAir
foot	qfqm/soni/qfFinl/veloZaxsErth and qfqm/soni/qfFinl/veloXaxsErth

5.3.2 Algorithmic Implementation

The EC turbulent exchange data quality analysis is implemented as part of the eddy4R-Docker EC processing framework (Sect. 4), and the algorithmic sequence is summarized in **Figure 8**.

The calculations described in Sects. 5.3.1.1 – 5.3.1.3 are applied to all data products of product level "dp01 statistics" in **Table 1**. The process is coded and documented in detail in the R-package eddy4R.qaqc, consisting of the following sequence (incl. function references):

- Calculation is performed individually for datasets of 1 min and 30 min duration.
- Derived quality flags at dp0p temporal resolution for IRGA validation period and AGC are calculated using the eddy4R.qaqc::def.qf.irga.vali() and eddy4R.qaqc::def.qf.irga.agc() functions, respectively.
- Using eddy4R.qaqc::def.dp01.qf.grp() function to indicate which quality flags are used as the input variables to determine alpha and beta quality metrics, and final quality flag for each reported dp01.

- Quality flags are combined into quality metrics using the eddy4R.qaqc::wrap.neon.dp01.qfqm() function.
    - The final quality flag for each reported dp01 are included in the basic and expanded hdf5 files:
        - SITE/dp01/qfqm/amrs
        - SITE/dp01/qfqm/co2Turb
        - SITE/dp01/qfqm/h2oTurb
        - SITE/dp01/qfqm/soni
    - The quality metrics, alpha and beta quality metrics are only included in expanded HDF5 files:
        - SITE/dp01/qfqm/amrs
        - SITE/dp01/qfqm/co2Turb
        - SITE/dp01/qfqm/h2oTurb
        - SITE/dp01/qfqm/soni

        Note that a full set of quality metrics (including pass, fail, and Na) for quality flags that are used in the determination of final quality flag for each reported dp01 are available in 30 min averaging folder.

- The homogeneity and stationarity ($qfStna$) and development of turbulence tests ($qfItc$)) quality flags for each reported dp04 turbulent fluxes are calculated using for averaging periods of 30 min using eddy4R.turb::def.stna() and eddy4R.turb::def.itc(), respectively.
- QF_{FINAL} of dp01 is propagated to determine the QF_{FINAL} of each dp04 (**Table 16**). Then, the results of $qfStna$ and $qfItc$ are taken into account to determine whether the data product is flagged as valid (QF_{FINAL} = 0) or invalid (QF_{FINAL} = 1).
    - The final quality flag for each reported dp04 are included in the basic and expanded hdf5 files:
        - SITE/dp04/qfqm/fluxCo2/turb
        - SITE/dp04/qfqm/fluxH2o/turb
        - SITE/dp04/qfqm/fluxMome/turb
        - SITE/dp04/qfqm/fluxTemp/turb
        - SITE/dp04/qfqm/foot/turb
    - The EC specific test results (quality indicators and quality flags) and dp01 propagation flags are only included in expanded HDF5 files:
        - SITE/dp04/qfqm/fluxCo2/turb
        - SITE/dp04/qfqm/fluxH2o/turb
        - SITE/dp04/qfqm/fluxMome/turb
        - SITE/dp04/qfqm/fluxTemp/turb

5.4 Uncertainty Analysis

5.4.1 Theory of Algorithm

Random errors are defined as the errors due to time averaging over an insufficient period for the time mean to converge to the ensemble mean by the ergodic hypothesis (Lenschow and Stankov, 1986; Lenschow et al., 1994; Lumley and Panofsky, 1964; Mann and Lenschow, 1994).

Here, the random sampling error is estimated using the method of Salesky et al. (2012). In comparison to other available approaches (e.g., Finkelstein and Sims, 2001; Hollinger and Richardson, 2005; Lenschow et al., 1994), the Salesky et al. (2012) method does not require an estimate of the integral time scale or replicate tower measurements. It is also equally applicable to statistical moments of any order, i.e. means, variances and covariance alike. Principally, the method consists of three parts, (i) a local time-series decomposition, (ii) the fitting of a power-law, and (iii) the inter- or extrapolation of the power law.

(i) The dp0p time-series is low-pass filtered using a running mean filter. This is performed for several filter window sizes $time_{\mathrm{filt}}$ in the range 10 $time_{\mathrm{scal}}$ < $time_{\mathrm{filt}}$ < $time_{\mathrm{agr}}$ / 10. Here, $time_{\mathrm{scal}}$ is the integral time scale of the process (assumed to be ~1 s), and $time_{\mathrm{agr}}$ the duration of the dataset available for aggregation (1,800 s). For each low-pass filtered time-series the standard deviation is calculated as representation of the random error associated with averaging over $time_{\mathrm{filt}}$. The random error decreases with increasing window-size of the low-pass filter (**Figure 9**).

(ii) Next, a power-law in the form of $\sigma = coef_{01}\ time_{\mathrm{filt}}{}^{coef_{02}}$ is regressed to the results (**Figure 9**). Here, $coef_{01}$ and $coef_{02}$ define the slope and convexity of the uncertainty reduction with increasing window-size of the low-pass filter, respectively. Salesky et al. (2012) relate a value of $coef_{02} = -1/2$ to the power law decay of random error as derived e.g. by Lenschow et al. (1994) for Gaussian and stationary turbulence. Salesky et al. (2012) restrict their analysis to stationary data and thus permit regression only of $coef_{01}$. In order to also accommodate non-stationary data we additionally permit regression of $-1/2 < coef_{02} < 0$. It should be noted that for $coef_{02} \rightarrow 0$ the power law becomes less convex, resulting in less uncertainty reduction with increasing window-size. The resulting algorithm such provides a conservative random error estimate for non-stationary data.

(iii) Lastly, $time_{\mathrm{filt}}$ in the resulting power law is substituted with the target averaging periods, yielding the corresponding random error.

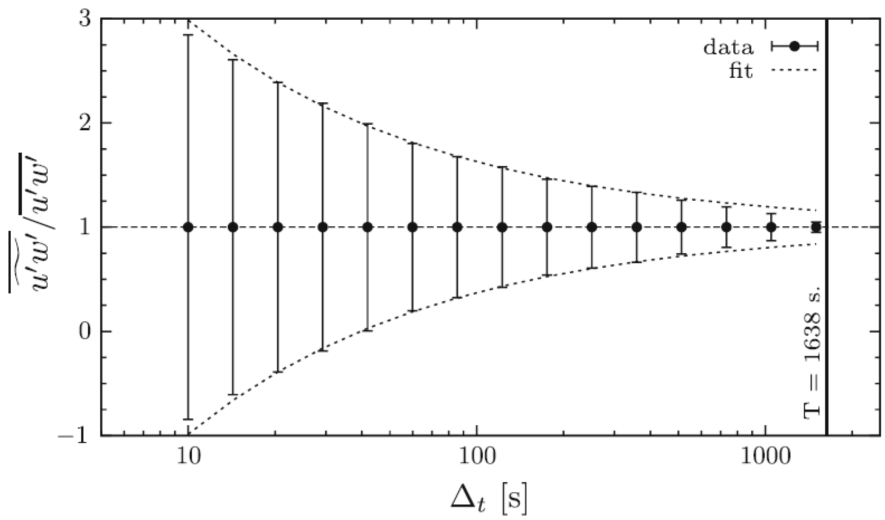

Figure 9. Reduction of standard deviation with increasing window size of the low-pass filter, from Salesky et al. (2012). The error bars denote the standard deviation, and the dashed line denotes a power-law fit.

5.4.2 Algorithmic Implementation

The EC turbulent exchange data uncertainty analysis is implemented as part of the eddy4R-Docker EC processing framework (Sect. 4), and the algorithmic sequence is summarized in **Figure 8**.

The random error calculation described in Sect. 5.4.1 is applied to all data products in **Table 1**. It is coded and documented in detail in the R-function eddy4R.ucrt:: def.ucrt.samp.filt(). In short:

- Calculation is performed individually for datasets of 30 min duration.
- Calculation is only performed if there are less than 10% missing values in the dataset. If less than 10% missing values, those are filled using linear interpolation.
- The signal is de-trended and tapered.
- Filtering is performed using Fast Fourier transform for 10 exponentially spaced filter widths in the range 10 s < $time_{filt}$ < 180 s.
- Nonlinear least squares regression is used to fit the power law.
- The random sampling error is calculated for averaging periods of 1 min and 30 min, and available in the HDF5 file at:
    - SITE/dp01/ucrt/amrs
    - SITE/dp01/ucrt/co2Turb
    - SITE/dp01/ucrt/h2oTurb
    - SITE/dp01/ucrt/soni
- As part of future developments (Sect. 8) we plan to implement end-to-end uncertainty quantification and propagation to dp04 data products (fluxes).

6 STORAGE EXCHANGE

The Eddy Covariance Storage Exchange Assembly (or EC profile assembly, hereafter referred to as the ECSE) consists of a suite of sensors such as temperature, CO_2 and H_2O gas analyzer and isotopic CO_2 and H_2O analyzers. The EC profile assembly is served to provide the measurements of temperature, CO_2 and H_2O concentration, the stable isotope of δ^{13}C in CO_2, δ^{18}O, and δ^{2}H in water vapor in the atmosphere at each tower measurement level. The vertical profile measurements of temperature, CO_2 and H_2O concentration will be used to calculate the storage fluxes, which will be incorporated into the calculation of the net ecosystem exchange of temperature, CO_2 and H_2O.

6.1 Theory of Measurement

In Eq. (2), NSAE of the control volume is expressed by the turbulent flux alone, based on several assumptions. Strict stationary is one of those assumptions, which implies that storage flux is negligible, i.e. the abundance of the scalar in the control volume remains constant. Because in tall and dense canopies this assumption is frequently violated, NEON measures and calculates the storage flux explicitly. With storage flux data products, NSAE is then calculated as the sum of storage flux and turbulent flux, which is described in more detail in Sect. 7.

An IRGA that is housed in the instrument hut switches between different measurement levels (usually between 4 and 8 levels at terrestrial sites and 2 to 4 at MDP sites) of each flux tower. At each measurement level, the IRGA measures CO_2 and H_2O for about 2 minutes. The storage of heat is calculated from the temperature profile measurements. Here we calculate the storage flux of the control volume based on the assumption that the temperature profile and the switched IRGA measurements at several levels can represent the vertical integral of time rate of change of the scalar over the entire control volume.

6.2 Data Analysis

6.2.1 Theory of Algorithm

The subject of this ATBD section is to describe the theory of algorithms to process the stable isotope of δ^{13}C in CO_2, δ^{18}O and δ^{2}H in water vapor in the atmosphere, and describe the mathematical derivation of the storage term, $\int_0^{d_{z,m}} \frac{\partial \bar{X}}{\partial t} dz$, in Eq. (1) in Sect. 5.1 and expressed as $fluxStor$ in Eq. (21)- below in Sect. 0. Only dp01 stable isotope data will be computed in this ATBD, while dp01 to dp04 data products will be computed for CO_2 ,H_2O , and temperature. The computation for dp01-dp04 data products follows the steps described below.

6.2.1.1 De-spiking

This part is the same as Sect. 5.2.1.1. This is applied to all time series signals collected for ECSE assembly.

6.2.1.2 Calculation of means, variance and standard error

This part is the same as Sect. 5.2.1.6, but X in the equations refers to temperature, CO_2 and H_2O concentration, stable isotope of δ^{13}C in CO_2, δ^{18}O and δ^{2}H in water vapor in the atmosphere in ECSE dp01, instead of turbulent fluxes in ECTE dp01.

6.2.1.3 Calculation of time rate of change

The time rate of change in ECSE dp02 is calculated from the time average of the four-minute measurement at the end of a half hour minus the time average of four minute measurements at the beginning of the same half hour (Eq. (20)). For example, assuming that storage flux estimates are to be computed for timestamps 07:30:00, then by convention the 07:30:00 timestamp represents the flux corresponding to observations between 07:30:00 and 08:00:00. dX, where X stands for temperature, CO_2 concentration, or H_2O concentration, will then be computed from the time average of measurements from 07:58:00 to 08:02:00 minus the time average of measurements from 07:28:00 to 07:32:00.

$$\frac{\mathrm{dX}}{dt} = \frac{\overline{\mathrm{X}}_{t\geq t_e-120s\ and\ t<t_e+120s} - \overline{\mathrm{X}}_{t\geq t_b-120s\ and\ t<t_b+120s}}{30\ min} \tag{20}$$

Where t_b is the beginning time of the first minute in the 30 minute block, and t_e is the last minute of the 30 minute block.

To calculate ECSE dp02, ECSE dp01 are firstly linearly interpolated into 1 minute resolution. Both the interpolation method and resolved temporal resolution can be adjusted for different locations and in different applications.

The averaging time is chosen as four minutes at the beginning and end of each half hour window because following Finnigan (2006), storage flux estimates influenced by single eddies penetrating inside the canopy should be avoided. The time period for the storage flux computation should be long enough to capture an adequate number of these eddies biasing the profiles or single observational points. Here, the period of storage flux computation is chosen based on 10 times of the time variable τ , the integral time scale of the turbulent time series between these eddies. Given the wide range in turbulence characteristics existing at the NEON ecosystem sites, and initially missing experimental evidence for all site conditions, we consider an adequate time of integration a period between 180 s and 300 s, with the shorter integration time to be reserved to turbulent conditions as observed in short canopies, 300 s to be reserved for dense canopies of 30 m or above, and the median value 240 s as default value. The temporal resolution of ECSE dp02 is set to be half hour for combination with ECTE data products.

ECSE dp03 is the vertically resolved time rate of change based on ECSE dp02. The interpolation method is linear interpolation. The resolved vertical resolution is prescribed as 0.1 m. Both the interpolation method and vertical resolution can be adjusted for different locations and different applications.

6.2.1.4 Calculation of storage flux

The storage flux in ECSE dp04 is integrated from vertical profiles of time rate of change for each date product using the equation:

$$fluxStor(X) = \frac{d\overline{X}}{dt} \cdot (\max(DistZaxsLvlMeasTow)) \tag{21}$$

Where *fluxStor* is storage flux, *X* is a scalar quantity such as H_2O or CO_2 mixing ratios, $DistZaxsLvlMeasTow$ is profile measurement heights.

This storage flux is part of the carbon dioxide flux listed as dp04 data product in **Table 1**.

6.2.2 Algorithmic implementation

The ECSE data analysis is implemented as part of the eddy4R-Docker EC processing framework (Sect. 4). The corresponding R workflow flow.stor.towr.neon.R (link to public GitHub repo in preparation) and its algorithmic sequence are summarized in **Figure 10**. The calculations described in Sects. 6.2.1.1–6.2.1.4 are applied to generate the ECSE dp01–dp04 in **Table 1**. The process is coded and documented in detail in the R-packages eddy4R.base, eddy4R.stor, and eddy4R.qaqc.

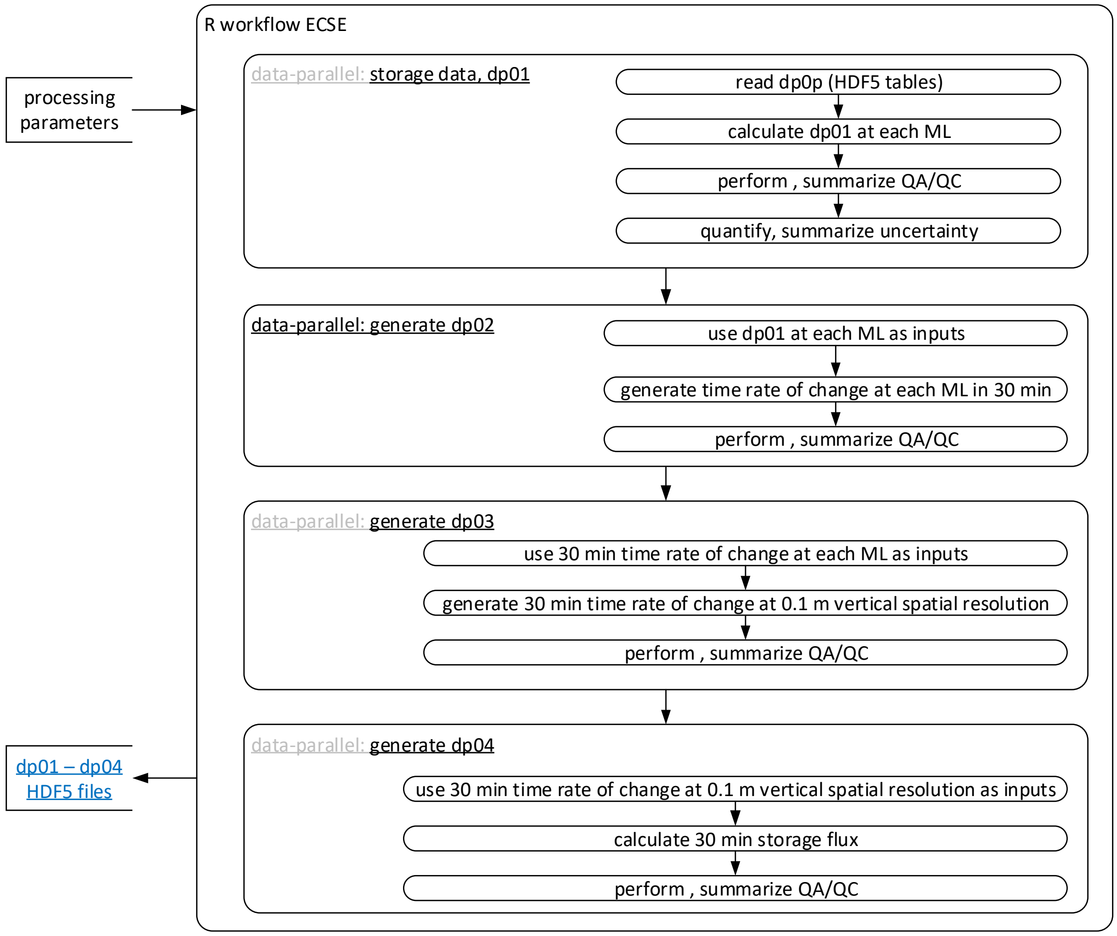

Figure 10. R workflow for eddy-covariance storage exchange (ECSE). Note: ML stands for measurement level.

6.2.2.1 ECSE dp01

IRGA CO_2 concentration (co2Stor) and IRGA H_2O concentration (h2oStor)

ECSE dp01 data (appears as SITE/dp01/data/co2Stor and dp01/data/h2oStor in HDF5 files) includes the descriptive statistics, mean, minimum, maximum, variance, standard error, number of samples, as well as begin time and end time, of **co2Stor** and **h2oStor** sub-data products at 2 min and 30 min resolution. The current NEON processing design utilizes the eddy4R package within a Docker framework to read in ECSE dp0p HDF5 files, do de-spiking, calculate descriptive statistics, and output HDF5 dp01.

Data flow for signal processing of dp01 IRGA CO_2 concentration (**co2Stor**) and IRGA H_2O concentration (**h2oStor**) will be treated in the following order.

- Calculation is performed individually for datasets of 2 min and 30 min duration.

- For each measurement of sampling data, e.g. the data under 000_0n0 folder, the middle two minute data after the first one minute critical time are selected for further calculation, while the critical data are set to be NaN.
- For each measurement of validation data, e.g. the data under co2XXX, the middle two minute before the last 20 s are selected for further calculation, while the last 20 s are set to be NaN.
- De-spiking is performed at dp0p temporal resolution using the eddy4R.qaqc::def.dspk.br86() function.
- Descriptive statistics are calculated using the eddy4R.stor::wrap.dp01.ecse() function.
- 2 and 30 min averages of description statistics are included in the basic and expanded HDF5 files: SITE/dp01/data/co2Stor and SITE/dp01/data/h2oStor.

Stable isotope of δ^{13}C in CO_2 (isoCo2)

ECSE dp01 data (appears as dp01/data/isoCo2 in HDF5 files) includes the descriptive statistics, mean, minimum, maximum, variance, standard error, number of samples, as well as begin time and end time, of **isoCo2** sub-data products at 9 and 30 min resolution for sampling and validation periods. The current NEON processing design utilizes the eddy4R package within a Docker framework to read in ECSE dp0p HDF5 files, do de-spiking, calculate descriptive statistics, and output HDF5 dp01.

Before the descriptive statistics are calculated, the data in the prescribed temporal interval are selected. During the sampling and validation period, each measurement level or each validation gas type is sampled for 10 min. Each time the first 1 min existing data are discarded in order to make sure the gas from previous sample has been cleaned from the flow. Therefore, the descriptive statistics of **isoCo2** sub-data products are calculated only using the next 9 min measurements.

Data flow for signal processing of dp01 is as follows:

- De-spiking is performed at dp0p temporal resolution using the eddy4R.qaqc::def.dspk.br86() function.
- Determine which data in the timestamp will be used in descriptive statistics calculation for each temporal interval using the eddy4R.base::def.idx.agr() function.
- Descriptive statistics are calculated using the eddy4R.stor::wrap.dp01.ecse() function

9 and 30 min averages of description statistics are included in the basic and expanded HDF5 files: SITE/dp01/data/isoCo2

CH4 concentration (ch4Conc)

ECSE dp01 data (appears as dp01/data/ch4Conc in HDF5 files) includes the descriptive statistics, mean, minimum, maximum, variance, standard error, number of samples, as well as begin time and end time, of **ch4Conc** sub-data products at 9 and 30 min resolution for sampling and validation periods. The current NEON processing design utilizes the eddy4R package within a Docker framework to read in ECSE dp0p HDF5 files, do de-spiking, calculate descriptive statistics, and output HDF5 dp01.

Before the descriptive statistics are calculated, the data in the prescribed temporal interval are selected. During the sampling and validation period, each measurement level or each validation gas type is sampled for 10 min. Each time the first 1 min existing data are discarded in order to make sure the gas from previous sample has been cleaned from the flow. Therefore, the descriptive statistics of **ch4Conc** sub-data products are calculated only using the next 9 min measurements.

Data flow for signal processing of dp01 is as follows:

- De-spiking is performed at dp0p temporal resolution using the eddy4R.qaqc::def.dspk.br86() function.
- Determine which data in the timestamp will be used in descriptive statistics calculation for each temporal interval using the eddy4R.base::def.idx.agr() function.
- Descriptive statistics are calculated using the eddy4R.stor::wrap.dp01.ecse() function
    - 9 and 30 min averages of description statistics are included in the basic and expanded HDF5 files: SITE/dp01/data/ch4Conc

Stable isotopes of δ^{18}O, and δ^{2}H in water vapor (isoH2o)

ECSE dp01 data (appears as dp01/data/isoH2o in HDF5 files) includes the descriptive statistics, mean, minimum, maximum, variance, standard error, number of samples, as well as begin time and end time, of **isoH2o** sub-data products at 9 and 30 min resolution for sampling period and 3 and 30 min for validation period. The current NEON processing design utilizes the eddy4R package within a Docker framework to read in ECSE dp0p HDF5 files, do de-spiking, calculate descriptive statistics, and output HDF5 dp01.

Before the descriptive statistics are calculated, the data in the prescribed temporal interval are selected. During the sampling period, each measurement level is sampled for 10 min. Each time the first 1 min existing data are discarded in order to make sure the gas from previous sample has been cleaned from the flow. Therefore, the descriptive statistics of **isoH2o** sub-data products are calculated only using the next 9 min measurements.

During the routine field validation of the CRD H_2O, the analyzer will cease to measure the atmospheric vapor from the tower profiles and measure water standards by using the zero air as a carrier gas. Water standards are injected through the vaporizer using the autosampler and a syringe. Field validation is performed for 3 standards (NEON Tertiary Low, Mid, and High standard) and each standard is injected 6 times. The procedure typically takes around 9 minutes per injection, the descriptive statistics are calculated using the data when the water concentration is stabilized which defined as the 3 min measurements right before the last 15 s in each injection time.

Data flow for signal processing of dp01 **isoH2o** is as follows:

- De-spiking is performed at dp0p temporal resolution using the eddy4R.qaqc::def.dspk.br86() function.

- Determine which data in the timestamp will be used in descriptive statistics calculation for each temporal interval using the eddy4R.base::def.idx.agr() function.
- Descriptive statistics are calculated using the using the eddy4R.stor::wrap.dp01.ecse() function.
  - 9 and 30 min averages of description statistics during sampling period are included in the basic and expanded HDF5 files: SITE/dp01/data/isoH2o
  - 3 and 30 min averages of description statistics during validation period are included in the basic and expanded HDF5 files: SITE/dp01/data/isoH2o

6.2.2.2 ECSE dp02

Data flow for signal processing of ECSE dp02 is as follows:

- Linear interpolation is performed for ECSE dp01 at temporal resolution into 1 min resolution with the maximum gap of 40 min.
- Time rate of change of temperature, CO_2 concentration and H_2O concentration for each measurement level was calculated using Eq. (20) in Sect. 6.2.1.3.
  - half-hourly time rate changes of temperature are included in the basic and expanded HDF5 files: SITE/dp02/data/tempStor
  - half-hourly time rate changes of CO_2 concentration are included in the basic and expanded HDF5 files: SITE/dp02/data/co2Stor
  - half-hourly time rate changes of H_2O concentration are included in the basic and expanded HDF5 files: SITE/dp02/data/h2oStor

6.2.2.3 ECSE dp03

Data flow for signal processing of ECSE dp03 is as follows:

- Linear interpolation is performed for all ECSE dp02 spatially into 0.1 m spatial resolution by default. If only one measurement level is available at a time, it is assigned to all vertical levels at the time in ECSE dp03. If no measurement level is available at a time, NaN is assigned to all vertical levels at the time.
  - half-hourly time rate changes at 0.1 m vertical spatial resolution of temperature are included in the basic and expanded HDF5 files: SITE/dp03/data/tempStor
  - half-hourly time rate changes at 0.1 m vertical spatial resolution of CO_2 concentration are included in the basic and expanded HDF5 files: SITE/dp03/data/co2Stor
  - half-hourly time rate changes at 0.1 m vertical spatial resolution of H_2O concentration are included in the basic and expanded HDF5 files: SITE/dp03/data/h2oStor

6.2.2.4 ECSE dp04

Data flow for signal processing of ECSE dp04 is as follows:

- Storage flux is calculated based on all ECSE dp03 according to Eq. (21)

- half-hourly storage fluxes of temperature are included in the basic and expanded HDF5 files: SITE/dp04/data/fluxTemp/stor
- half-hourly storage fluxes of CO_2 are included in the basic and expanded HDF5 files: SITE/dp04/data/fluxCo2/stor
- half-hourly storage fluxes of H_2O are included in the basic and expanded HDF5 files: SITE/dp04/data/fluxH2o/stor

6.3 Quality Assurance and Quality Control analysis

The basic quality assurance and quality control analysis can be found in Sect. 5.3. However, sensor and statistical QA/QC tests are performed on and reported for the 1 Hz data for ECSE.

6.3.1 Theory of Algorithm

Details can be found in Sect. 5.3.1.

6.3.1.1 Sensor Quality Flags

Most of quality flags due to the sensor health and statistical plausibility tests are generated as part of the dp0p report variables (AD[03]). In addition, the water validation quality flag (qfValiH2o) and mass flow meter flow rate flag (qfFrt00) are also generated in this ATBD as defined below.

Water Validation quality flag (qfValiH2o) – is indicating when the validation of crdH2o sensor is good (0) or bad (1). Field validation of crdH2o is performed using 3 standards (NEON Tertiary Low, Mid, and High standard) and each standard is injected 6 times. The qfValiH2o of the first 3 injections of each standard (injection number 1, 2, 3, 7, 8, 9, 13, 14, and 15) will be set to 1. For the rest (injection number 4, 5, 6, 10, 11, 12, 16, 17, and 18), the qfValiH2o is set to 1 if the measurements values of dlta18OH2o and dlta2HH2o are greater or less than the thresholds. As default, thresholds can be can calculated as the reference value ± 30% of reference value.

Mass flow meter flow rate flag (qfFrt00) – is generated for each measurement level to indicate when the sensor is operated under low flow or pump failure condition (1 = low flow or pump failure condition, 0 = normal operating condition, −1 = NA). The flag is determined from the mass flow meter flow rate as follows:

$$\mathrm{qfIFrt00} = \begin{cases} 1 \text{ if } \mathrm{frt00} \leq 0.000051 \\ 0 \text{ if } \mathrm{frt00} > 0.000051 \\ -1 \; otherwise. \end{cases} \tag{22}$$

where frt00 is the flow rate from mass flow meter (mfm) in the unit of $m^3\ s^{-1}$.

6.3.1.2 Quality Budget (QFQM)

Details of the basic quality budget that applied to ECSE dp01 will be identical to ECTE, which can be found in Sect. 5.3.1.2. The QF_{FINAL} of ECSE dp02 and dp03 is estimated from surrounding values. For example, if one or more QF_{FINAL} of ECSE dp01 which used to determine ECSE dp02 is equal to 1, assign QF_{FINAL} of that ECSE dp02 to 1. Similarly, QF_{FINAL} of that ECSE dp03 is assigned to 1 if one or more QF_{FINAL} of ECSE dp02 which used to determine ECSE dp03 is equal to 1.

6.3.2 Algorithmic Implementation

The calculations described in Sects. 6.3.2.1 - 6.3.2.4 are applied to ECSE dp01–dp04 in Table 1. The process is coded and documented in detail in the R-packages eddy4R.base and eddy4R.qaqc, consisting of the following sequence (incl. function references).

6.3.2.1 ECSE dp01

Data flow for QA/QC processing of ECSE "dp01 statistics" (dp01) will be treated in the following order.

- Using eddy4R.stor::wrap.prd.day.ecse() to calculate qfFrt00 at dp0p temporal resolution for all mass flow meters and replace flagged data by NaN.
- Calculation is performed individually for datasets of each duration.Using the eddy4R.base::def.idx.agr() function to determine the datasets of 2 min and 30 min duration for **co2Stor** and **h2oStor**, 9 min and 30 min duration for **isoCo2**, 9 min and 30 min duration for **isoH2o** during sampling period and 3 min and 30 min duration for **isoH2o** during validation period.
- Derive qfValiH2o at dp0p temporal resolution for **isoH2o**.
- Using eddy4R.qaqc::def.dp01.qf.grp() function to indicate which quality flags are used as the input variables to determine alpha and beta quality metrics, and final quality flag for each reported dp01.
- Calculate quality metrics, alpha and beta quality metrics, and final quality flag for each reported dp01 using the eddy4R.qaqc::wrap.dp01.qfqm.ecse() function.
    - The final quality flag for each reported dp01 are included in the basic and expanded HDF5 files:
        - SITE/dp01/qfqm/co2Stor
        - SITE/dp01/qfqm/h2oStor
        - SITE/dp01/qfqm/isoCo2
        - SITE/dp01/qfqm/ch4Conc
        - SITE/dp01/qfqm/isoH2o
    - The quality metrics, alpha and beta quality metrics are only included in expanded HDF5 files:
        - SITE/dp01/qfqm/co2Stor
        - SITE/dp01/qfqm/h2oStor
        - SITE/dp01/qfqm/isoCo2

- SITE/dp01/qfqm/ch4Conc
- SITE/dp01/qfqm/isoH2o

Note that detailed information of quality metrics (including pass, fail, and Na) for quality flags used in determination of final quality flag for each reported dp01 are available in 30 min averaging folder.

6.3.2.2 ECSE dp02

Data flow for QA/QC processing of ECSE dp02 is as follows:

- Interpolated QF_{FINAL} of ECSE dp01 at temporal resolution into 1 min resolution by assigning the QF_{FINAL} to 1 to ECSE dp01 that fell in between two adjacent available data and if one or more QF_{FINAL} of two adjacent available data is equal to 1. Otherwise, interpolated QF_{FINAL} values are equal to 0.
- For each half-hourly, calculated QF_{FINAL} of ECSE dp02 by assigning the QF_{FINAL} of that half-hourly to 1 if one or more QF_{FINAL} of ECSE dp01 which used to determine that ECSE dp02 is equal to 1. Otherwise, QF_{FINAL} of ECSE dp02 at that half-hourly is equal to 0.
    - half-hourly summarized of the final quality flag of time rate changes of temperature are included in the basic and expanded HDF5 files: SITE/dp02/qfqm/tempStor
    - half-hourly summarized of the final quality flag of time rate changes of CO_2 are included in the basic and expanded HDF5 files: SITE/dp02/qfqm/co2Stor
    - half-hourly summarized of the final quality flag of time rate changes of H_2O are included in the basic and expanded HDF5 files: SITE/dp02/qfqm/h2oStor

6.3.2.3 ECSE dp03

Data flow for QA/QC processing of ECSE dp03 is as follows:

- For each half-hourly, interpolated QF_{FINAL} of ECSE dp03 into 0.1 m spatial resolution from QF_{FINAL} of ECSE dp02. Assign the QF_{FINAL} to 1 to ECSE dp03 that fell in between two adjacent measurement levels and if one or more QF_{FINAL} of two adjacent measurement levels is equal to 1. Otherwise, spatial interpolated QF_{FINAL} values are equal to 0.
    - half-hourly summarized of the final quality flag of time rate changes at 0.1 m vertical spatial resolution of temperature are included in the basic and expanded HDF5 files: SITE/dp03/qfqm/tempStor
    - half-hourly summarized of the final quality flag of time rate changes at 0.1 m vertical spatial resolution of CO_2 concentration are included in the basic and expanded HDF5 files: SITE/dp03/qfqm/co2Stor
    - half-hourly summarized of the final quality flag of time rate changes at 0.1 m vertical spatial resolution of H_2O concentration are included in the basic and expanded HDF5 files: SITE/dp03/qfqm/h2oStor

6.3.2.4 ECSE dp04

Data flow for QA/QC processing of ECSE dp04 is as follows:

- For each half-hourly, calculated QF_{FINAL} of ECSE dp04 by assigning the QF_{FINAL} of that half-hourly to 1 if one or more QF_{FINAL} of ECSE dp03 which used to determine that ECSE dp04 is equal to 1. Otherwise, QF_{FINAL} of ECSE dp04 at that half-hourly is equal to 0.
    - half-hourly summarized of the final quality flag of storage fluxes of temperature are included in the basic and expanded HDF5 files: SITE/dp04/qfqm/fluxTemp/stor
    - half-hourly summarized of the final quality flag of storage fluxes of CO_2 are included in the basic and expanded HDF5 files: SITE/dp04/qfqm/fluxCo2/stor
    - half-hourly summarized of the final quality flag of storage fluxes of H_2O are included in the basic and expanded HDF5 files: SITE/dp04/qfqm/fluxH2o/stor

6.4 Uncertainty Analysis

6.4.1 Theory of Algorithm

Similar to ECTE (Sect. 5.4.1), the random sampling error is estimated using the method of Salesky et al. (2012). However, the minimum and maximum time filter width are adjusted as recommended by (Salesky et al. (2012). Therefore, the filtering is performed using Fast Fourier transform for 10 exponentially spaced filter widths in the range 2 $time_{scal} < time_{filt} < time_{agr} / 4$.

6.4.2 Algorithmic Implementation

The calculations described in this section are applied to ECSE dp01–dp04 in **Table 1**. The process is coded and documented in detail in the R-packages eddy4R.base and eddy4R.ucrt, consisting of the following sequence (incl. function references).

6.4.2.1 ECSE dp01

The random error calculation described in Sect. 5.4.1 and 6.4.1 is applied to all ECSE dp01 data products in **Table 1**. It is coded and documented in detail in the R-function eddy4R.ucrt::def.ucrt.samp.filt() and eddy4R.ucrt::wrap.neon.dp01.ucrt.ecse(). Data flow is as follows:

- Calculation is performed individually for datasets of each duration.
- Using the eddy4R.base::def.idx.agr function to determine the input datasets of 2 min and 30 min duration for **co2Stor** and **H2oStor**, 9 min and 30 min duration for **isoCo2**, 9 min and 30 min duration for **isoH2o** during sampling period and 3 min and 30 min duration for **isoH2o** during validation period.
- As default, the calculation is only performed if there are less than 10% missing values in the dataset. However, the missing values are adjusted to

- 50% for dlta13CCo2, rtioMoleDry12CCo2, rtioMoleDry13CCo2, rtioMoleDryCo2, rtioMoleWet12CCo2, rtioMoleWet13CCo2, and rtioMoleWet13CCo2 measured in **isoCo2**
- 85% for rtioMoleDryH2o and rtioMoleWetH2o measured in **isoCo2**, and rtioMoleDryCh4 and rtioMoleWetCh4 in **ch4Conc**

- If the missing data are less than the values as mentioned above, those are filled using linear interpolation.
- Filtering is performed using Fast Fourier transform for 10 exponentially spaced filter widths in the range of:
  - 2 s < $time_{filt}$ < 30 s for **co2Stor** and **h2oStor** during both sampling and validation period
  - 2 s < $time_{filt}$ < 135 s for **isoCo2** during both sampling and validation period
  - 2 s < $time_{filt}$ < 135 s for **isoH2o** during sampling period and 2 s < $time_{filt}$ < 45 s for **isoH2o** during sampling period
- Nonlinear least squares regression is used to fit the power law.
- The random sampling error is calculated for each target averaging periods of:
  - 2 min for **co2Stor** and **h2oStor** during both sampling and validation period
  - 9 min for **isoCo2** during both sampling and validation period
  - 9 min for **ch4Conc** during both sampling and validation period
  - 9 min for **isoH2o** during sampling period and 3 min for **isoH2o** during sampling period
- The random sampling error for 30 min averaging period can be determine by:
  - First, determined how many small averaging periods falling into each 30 min window.
  - Then, calculated the random sampling error for each of small averaging period that falling into each 30 min window.
  - Lastly, calculated the median out of the random sampling error from the previous step and used that results to represent the random sampling error of that 30 min averaging period
- The uncertainty results for each reported dp01 are included in the basic and expanded HDF5 files:

  - SITE/dp01/ucrt/co2Stor
  - SITE/dp01/ucrt/h2oStor
  - SITE/dp01/ucrt/isoCo2
  - SITE/dp01/ucrt/ch4Conc
  - SITE/dp01/ucrt/isoH2o

7 NET SURFACE-ATMOSPHERE EXCHANGE

7.1 Theory of Measurement

7.2 Data Analysis

7.2.1 Theory of Algorithm

Here, net surface-atmosphere exchange (NSAE) is defined as the sum of storage flux and turbulent flux, on a 30 min basis, Eq. (23). The constituent terms I and II are derived, respectively, in Sect 5 and Sect. 6.

$$NSAE = \underbrace{\int_0^{d_{z,m}} \frac{\partial \overline{X}}{\partial t} dz}_{\text{I}} + \underbrace{\overline{w'X'}}_{\text{II}} \qquad (23)$$

7.2.2 Algorithmic Implementation

The NSAE data analysis is implemented as part of the eddy4R-Docker EC processing framework (Sect. 4). The corresponding R workflow flow.nsae.R (link to public GitHub repo in preparation) and its algorithmic sequence are summarized in **Figure 11**. The calculations are applied to all data products of product level "dp04" in **Table 1**. The process is coded and documented in detail in the R-packages eddy4R.base, consisting of the following sequence:

- calculation is performed for datasets of 30 min time resolution.
- total (wet) air density, dry air density, and latent heat of vaporization are calculated.
- ECTE and ECSE heat and water vapor fluxes are converted from kinematic units to units of energy [W m-2].
- ECTE and ECSE CO_2 fluxes are converted to units [$\mu molCO_2\ m^{-2}\ s^{-1}$].
- ECTE and ECSE fluxes are combined per Eq. (23) to yield NSAE fluxes.
- location of results in HDF5 file: SITE/dp04/data/FLUX/nsae, where FLUX is one of {**fluxCo2**, **fluxH2o**, **fluxTemp**}.

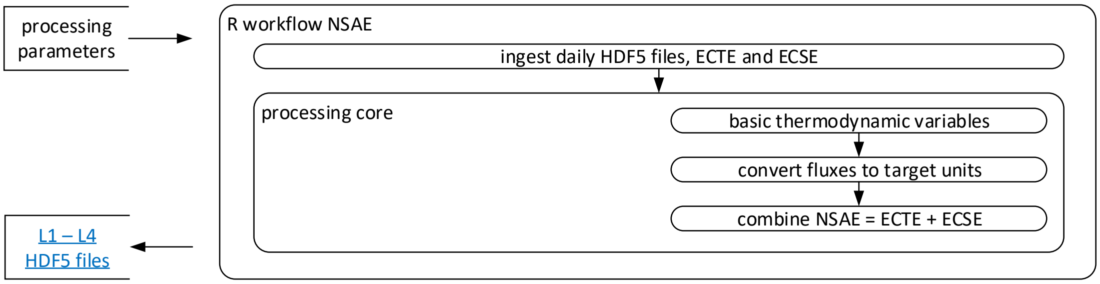

Figure 11. The EC net surface-atmosphere exchange workflow within the eddy4R-Docker EC processing framework (Sect. 4).

7.3 Quality Assurance and Quality Control Analysis

7.3.1 Theory of Algorithm

The final quality flag of NSAE is determined using the information of storage flux final quality flag (qfFinlStor) and turbulent flux final quality flag (qfFinlTurb) as:

$$\mathrm{qfFinl} = \begin{cases} 0 \ if \ \mathrm{qfFinlStor} = 0 \text{ and } \mathrm{qfFinlTurb} = 0 \\ 1 \ \ otherwise. \end{cases} \quad (24)$$

7.3.2 Algorithmic Implementation

The NSAE quality assurance and quality control analysis is implemented as part of R workflow flow.nsae.R. The calculations are applied to fluxCo2, fluxH2o, and fluxTemp data products of product level "dp04" in **Table 1**. The process consists of:

- The final quality flag of NSAE is determined per Eq. (24).
  - The NSAE final quality flag for each reported dp04 are included in the basic and expanded hdf5 files and the information of qfFinlStor and qfFinlTurb are only included in expanded HDF5 files:
    - SITE/dp04/qfqm/fluxCo2/nsae
    - SITE/dp04/qfqm/fluxH2o/nsae
    - SITE/dp04/qfqm/fluxTemp/nsae

7.4 Uncertainty Analysis

Scheduled for implementation as part of future plans and modifications (Sect. 8).

7.4.1 Theory of Algorithm

N/A

7.4.2 Algorithmic Implementation

N/A

8 FUTURE PLANS AND MODIFICATIONS

Additional measurements can be proposed through NEON's Assignable Asset program. At the time of writing, implementation of NEON's EC processing focusses on completing data products (**Table 1**) and making the DevOps framework (Sect. 4) publicly accessible for continued development. This DevOps framework intends to incentivize justified and reasonable community requests and contributions, and thus to continuously tailor NEON EC DPs and the publicly available eddy4R-Docker software to user needs. Requests and contributions are kept and moderated in a central backlog. The Surface Atmosphere Exchange Technical Working Group will be consulted for regular prioritization of scientific return-on-investment, and activation of capability development following requests. Prominent capability requests include end-to-end quality and uncertainty budgets, adding more footprint parameterizations, mapping naming conventions to other networks, and many more.

9 ACKNOWLEDGEMENTS

We thank the members of the Surface Atmosphere Exchange Technical Working Group for many years of constructive advise in creating NEON's eddy-covariance infrastructure. Special thanks for reviewing and editing this version of the Algorithm Theoretical Basis Document go to George Burba, Ankur Desai, Tarek El-Madany, Tobias Gerken, Jordan Goodrich, Eugenie Paul-Limoges, and Sebastian Wolf.

10 CITATION

Metzger, S., Durden, D., Florian, C., Luo, H., Pingintha-Durden, N., and Xu, K.: Algorithm theoretical basis document: eddy-covariance data products bundle, National Ecological Observatory Network, NEON.DOC.004571, Revision A (2018-04-30), http://data.neonscience.org/documents, Boulder, U.S.A., 55 pp., 2018.

11 BIBLIOGRAPHY

Baldocchi, D., Finnigan, J., Wilson, K., Paw U, K. T., and Falge, E.: On measuring net ecosystem carbon exchange over tall vegetation on complex terrain, Boundary Layer Meteorol., 96, 257-291, doi:10.1023/a:1002497616547, 2000.

Brock, F. V.: A nonlinear filter to remove impulse noise from meteorological data, J. Atmos. Oceanic Technol., 3, 51-58, doi:10.1175/1520-0426(1986)003<0051:anftri>2.0.co;2, 1986.

Campbell Scientific: CSAT3 three dimensional sonic anemometer instruction manual, Campbell Scientific, Logan, USA, 72, 2011.

Dyer, A. J., and Hicks, B. B.: Flux-gradient relationships in the constant flux layer, Q. J. R. Meteorolog. Soc., 96, 715-721, 10.1002/qj.49709641012, 1970.

Finkelstein, P. L., and Sims, P. F.: Sampling error in eddy correlation flux measurements, J. Geophys. Res. Atmos., 106, 3503-3509, doi:10.1029/2000JD900731, 2001.

Finnigan, J.: A comment on the paper by Lee (1998): "On micrometeorological observations of surface-air exchange over tall vegetation", Agric. For. Meteorol., 97, 55-64, doi:10.1016/s0168-1923(99)00049-0, 1999.

Finnigan, J.: The storage term in eddy flux calculations, Agric. For. Meteorol., 136, 108-113, doi:10.1016/j.agrformet.2004.12.010, 2006.

Finnigan, J. J., Clement, R., Malhi, Y., Leuning, R., and Cleugh, H. A.: A re-evaluation of long-term flux measurement techniques. Part 1: Averaging and coordinate rotation, Boundary Layer Meteorol., 107, 1-48, doi:10.1023/A:1021554900225, 2003.

Foken, T.: Micrometeorology, Springer, Berlin, Heidelberg, 306 pp., 2008.

Foken, T., Leuning, R., Oncley, S. P., Mauder, M., and Aubinet, M.: Corrections and data quality control, in: Eddy covariance: A practical guide to measurement and data analysis, edited by: Aubinet, M., Vesala, T., and Papale, D., Springer, Dordrecht, Heidelberg, London, New York, 85-131, 2012.

Foken, T.: Micrometeorology, 2 ed., Springer, Berlin, Heidelberg, 362 pp., 2017.

Hargrove, W. W., and Hoffman, F. M.: Using multivariate clustering to characterize ccoregion borders, Computing in Science and Engineering, 1, 18-25, doi:10.1109/5992.774837, 1999.

Hargrove, W. W., and Hoffman, F. M.: Potential of multivariate quantitative methods for delineation and visualization of ecoregions, Environmental Management, 34, S39-S60, doi:10.1007/s00267-003-1084-0, 2004.

Hicks, B. B.: Wind profile relationships from the 'wangara' experiment, Q. J. R. Meteorolog. Soc., 102, 535-551, doi:10.1002/qj.49710243304, 1976.

Hollinger, D. Y., and Richardson, A. D.: Uncertainty in eddy covariance measurements and its application to physiological models, Tree Physiol., 25, 873-885, doi:10.1093/treephys/25.7.873, 2005.

Kaimal, J. C., and Finnigan, J. J.: Atmospheric boundary layer flows: Their structure and measurement, Oxford University Press, New York, USA, 289 pp., 1994.

Kljun, N., Calanca, P., Rotach, M. W., and Schmid, H. P.: A simple parameterisation for flux footprint predictions, Boundary Layer Meteorol., 112, 503-523, doi:10.1023/B:BOUN.0000030653.71031.96, 2004.

Kljun, N., Calanca, P., Rotach, M. W., and Schmid, H. P.: A simple two-dimensional parameterisation for flux footprint prediction (FFP), Geosci. Model Dev., 8, 3695-3713, doi:10.5194/gmd-8-3695-2015, 2015.

Kondo, J., and Sate, T.: The Determination of the von Kármán Constant, Journal of the Meteorological Society of Japan, 60, 461-471, 1982.

Kormann, R., and Meixner, F. X.: An analytical footprint model for non-neutral stratification, Boundary Layer Meteorol., 99, 207-224, doi:10.1023/A:1018991015119, 2001.

Leclerc, M. Y., and Foken, T.: Footprints in micrometeorology and ecology, 1st ed., Springer, Berlin, Heidelberg, Germany, 239 pp., 2014.

Lee, X.: On micrometeorological observations of surface-air exchange over tall vegetation, Agric. For. Meteorol., 91, 39-49, doi:10.1016/s0168-1923(98)00071-9, 1998.

Lee, X., Finnigan, J., and Paw U, K. T.: Coordinate systems and flux bias error, in: Handbook of micrometeorology: A guide for surface flux measurement and analysis, 1 ed., edited by: Lee, X., Law, B., and Massman, W., Springer, Dordrecht, 33-66, 2004.

Lemon, E. R.: Photosynthesis under field conditions. II. An aerodynamic method for determining the turbulent carbon dioxide exchange between the atmosphere and a corn field, Agron. J., 52, 697-703, doi:10.2134/agronj1960.00021962005200120009x, 1960.

Lenschow, D. H., and Stankov, B. B.: Length scales in the convective boundary layer, Journal Of The Atmospheric Sciences, 43, 1198-1209, doi:10.1175/1520-0469(1986)043<1198:LSITCB>2.0.CO;2, 1986.

Lenschow, D. H., Mann, J., and Kristensen, L.: How long is long enough when measuring fluxes and other turbulence statistics?, J. Atmos. Oceanic Technol., 11, 661-673, doi:10.1175/1520-0426(1994)011<0661:HLILEW>2.0.CO;2, 1994.

Liu, H. P., Peters, G., and Foken, T.: New equations for sonic temperature variance and buoyancy heat flux with an omnidirectional sonic anemometer, Boundary Layer Meteorol., 100, 459-468, doi:10.1023/A:1019207031397, 2001.

Loescher, H. W., Law, B. E., Mahrt, L., Hollinger, D. Y., Campbell, J., and Wofsy, S. C.: Uncertainties in, and interpretation of, carbon flux estimates using the eddy covariance technique, J. Geophys. Res. Atmos., 111, D21S90-, 2006.

Lumley, J. L., and Panofsky, H. A.: The structure of atmospheric turbulence, Interscience Publishers, New York, 1964.

Mahrt, L., Vickers, D., Howell, J., Højstrup, J., Wilczak, J. M., Edson, J., and Hare, J.: Sea surface drag coefficients in the Risø Air Sea Experiment, J. Geophys. Res., 101, 14327-14335, doi:10.1029/96jc00748, 1996.

Mann, J., and Lenschow, D. H.: Errors in airborne flux measurements, J. Geophys. Res. Atmos., 99, 14519-14526, 1994.

Mauder, M., and Foken, T.: Documentation and instruction manual of the eddy-covariance software package TK3, Universität Bayreuth, Arbeitsergenisse Abteilung Mikrometeorologie, 46, Bayreuth, Germany, 60 pp., ISSN 1614-8924, 2011.

McMillen, R. T.: An eddy correlation technique with extended applicability to non-simple terrain, Boundary Layer Meteorol., 43, 231-245, doi:10.1007/bf00128405, 1988.

Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H. P., and Foken, T.: Eddy-covariance flux measurements with a weight-shift microlight aircraft, Atmos. Meas. Tech., 5, 1699-1717, doi:10.5194/amt-5-1699-2012, 2012.

Metzger, S., Burba, G., Burns, S. P., Blanken, P. D., Li, J., Luo, H., and Zulueta, R. C.: Optimization of an enclosed gas analyzer sampling system for measuring eddy covariance fluxes of H_2O and CO_2, Atmos. Meas. Tech., 9, 1341-1359, doi:10.5194/amt-9-1341-2016, 2016.

Metzger, S., Durden, D., Sturtevant, C., Luo, H., Pingintha-Durden, N., Sachs, T., Serafimovich, A., Hartmann, J., Li, J., Xu, K., and Desai, A. R.: eddy4R 0.2.0: a DevOps model for community-extensible processing and analysis of eddy-covariance data based on R, Git, Docker, and HDF5, Geosci. Model Dev., 10, 3189-3206, doi:10.5194/gmd-10-3189-2017, 2017.

Monin, A. S., and Obukhov, A. M.: Basic laws of turbulent mixing in the surface layer of the atmosphere, Tr. Akad. Nauk SSSR Geofiz. Inst., 24, 163-187; English translation by John Miller, 1959, 1954.

Monteith, J. L., and Unsworth, M. H.: Principles of environmental physics, 3 ed., Elsevier, Amsterdam, Boston, 418 pp., 2008.

Nordbo, A., and Katul, G.: A wavelet-based correction method for eddy-covariance high-frequency losses in scalar concentration measurements, Boundary Layer Meteorol., 146, 81-102, doi:10.1007/s10546-012-9759-9, 2012.

Paw U, K. T., Baldocchi, D. D., Meyers, T. P., and Wilson, K. B.: Correction of eddy-covariance measurements incorporating both advective effects and density fluxes, Boundary Layer Meteorol., 97, 487-511, doi:10.1023/a:1002786702909, 2000.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2016.

Rastetter, E. B., Williams, M., Griffin, K. L., Kwiatkowski, B. L., Tomasky, G., Potosnak, M. J., Stoy, P. C., Shaver, G. R., Stieglitz, M., Hobbie, J. E., and Kling, G. W.: Processing arctic eddy-flux data using a simple carbon-exchange model embedded in the ensemble Kalman filter, Ecological Applications, 20, 1285-1301, doi:10.1890/09-0876.1, 2010.

Rebmann, C., Kolle, O., Heinesch, B., Queck, R., Ibrom, A., and Aubinet, M.: Data acquisition and flux calculations, in: Eddy covariance: A practical guide to measurement and data analysis, edited by: Aubinet, M., Vesala, T., and Papale, D., Springer, Dordrecht, Heidelberg, London, New York, 59-83, 2012.

Ripley, B.D. and Thompson, M.: Regression Technique for the detection of analytical bias, Analyst, 112, 377-383, 1987.

Salesky, S., Chamecki, M., and Dias, N.: Estimating the random error in eddy-covariance based fluxes and other turbulence statistics: The filtering method, Boundary Layer Meteorol., 144, 113-135, doi:10.1007/s10546-012-9710-0, 2012.Schmid, H. P.: Source areas for scalars and scalar fluxes, Boundary Layer Meteorol., 67, 293-318, doi:10.1007/bf00713146, 1994.

Schotanus, P., Nieuwstadt, F. T. M., and Bruin, H. A. R.: Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes, Boundary Layer Meteorol., 26, 81-93, doi:10.1007/BF00164332, 1983.

Smith, D., and Metzger, S.: Algorithm theoretical basis document: Time series automatic despiking for TIS level 1 data products, National Ecological Observatory Network, NEON.DOC.000783, Boulder, U.S.A., 15 pp., 2013.

Starkenburg, D., Metzger, S., Fochesatto, G. J., Alfieri, J. G., Gens, R., Prakash, A., and Cristóbal, J.: Assessment of de-spiking methods for turbulence data in micrometeorology, J. Atmos. Oceanic Technol., 33, 2001 - 2013, doi:10.1175/jtech-d-15-0154.1, 2016.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Kluwer Academic Publishers, Dordrecht, The Netherlands, 670 pp., 1988.

Tanner, C. B., and Thurtell, G. W.: Anemoclinometer Measurements of Reynolds Stress and Heat Transport in the Atmospheric Surface Layer, University of Wisconsin, Madison, 200, 1969.

Turnipseed, A. A., Anderson, D. E., Blanken, P. D., Baugh, W. M., and Monson, R. K.: Airflows and turbulent flux measurements in mountainous terrain: Part 1. Canopy and local effects, Agric. For. Meteorol., 119, 1-21, doi:10.1016/s0168-1923(03)00136-9, 2003.

Vesala, T., Kljun, N., Rannik, U., Rinne, J., Sogachev, A., Markkanen, T., Sabelfeld, K., Foken, T., and Leclerc, M. Y.: Flux and concentration footprint modelling: State of the art, Environ. Pollut., 152, 653-666, doi:10.1016/j.envpol.2007.06.070, 2008.

Welles, J.M. and McDermitt, D.K.: Measuring carbon dioxide in the atmosphere, in: Micrometeorology in agricultural systems, edited by: Hatfield, J.L. and Baker, J.M., Agronomy Monograph, Madison, WI, 47, 287-320, 2005.

Wilczak, J. M., Oncley, S. P., and Stage, S. A.: Sonic anemometer tilt correction algorithms, Boundary Layer Meteorol., 99, 127-150, doi:10.1023/A:1018966204465, 2001.

Xu, K., Metzger, S., and Desai, A. R.: Upscaling tower-observed turbulent exchange at fine spatio-temporal resolution using environmental response functions, Agric. For. Meteorol., 232, 10-22, doi:10.1016/j.agrformet.2016.07.019, 2017.

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

Appendix A NEON Observatory Design

The NEON observatory design is based on multivariate geographic clustering (Hargrove and Hoffman, 1999, 2004). Using national data sets for eco-climatic variables, the continental US, including Hawaii, Alaska, and Puerto were partitioned into 20 eco-climatic domains (**Figure 12**). These domains capture the full range of US ecological and climatic diversity as well as distinct regions of vegetation, landforms, and ecosystem, dynamics. In each domain, a core (30-year) site that represents the predominant "wildlands" ecosystem is accompanied by additional research sites designed to address specific scientific questions (e.g. land use, management, disturbance, or recovery). A detailed, interactive map is available from the NEON website.

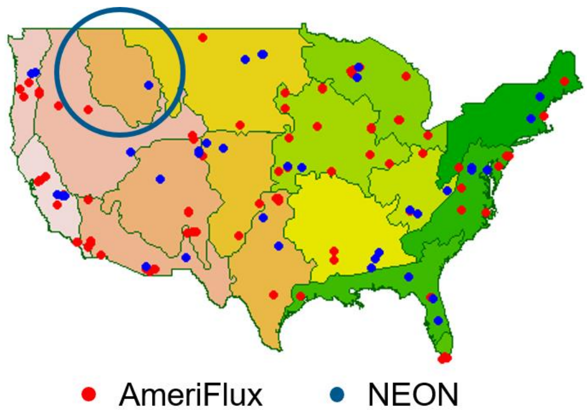

Figure 12. Eco-climatic zones across the contiguous United States after Hargrove and Hoffman (1999, 2004). Superimposed are AmeriFlux sites (prior to NEON site registration) and the NEON terrestrial instrumented site network. The NEON design adds previously underrepresented eco-climatic zones to the joint site distribution (blue circle).

Appendix B NEON Site Design

At each NEON TIS site, tower-based EC-flux measurements are performed in coordination with a wide range of contextual observations. These include meteorological, atmospheric composition, and soil measurements, alongside airborne remote sensing and characterization of soils, plants, insects, birds, mammals and phenology, as well as lakes and streams (**Figure 13**). Each NEON flux tower is placed and oriented with the design goal to represent a target ecosystem during 90% of the time, based on wind statistics from temporary deployments and source area modeling (Kormann and Meixner, 2001). The detailed spatial configuration of each site is available from the NEON website.

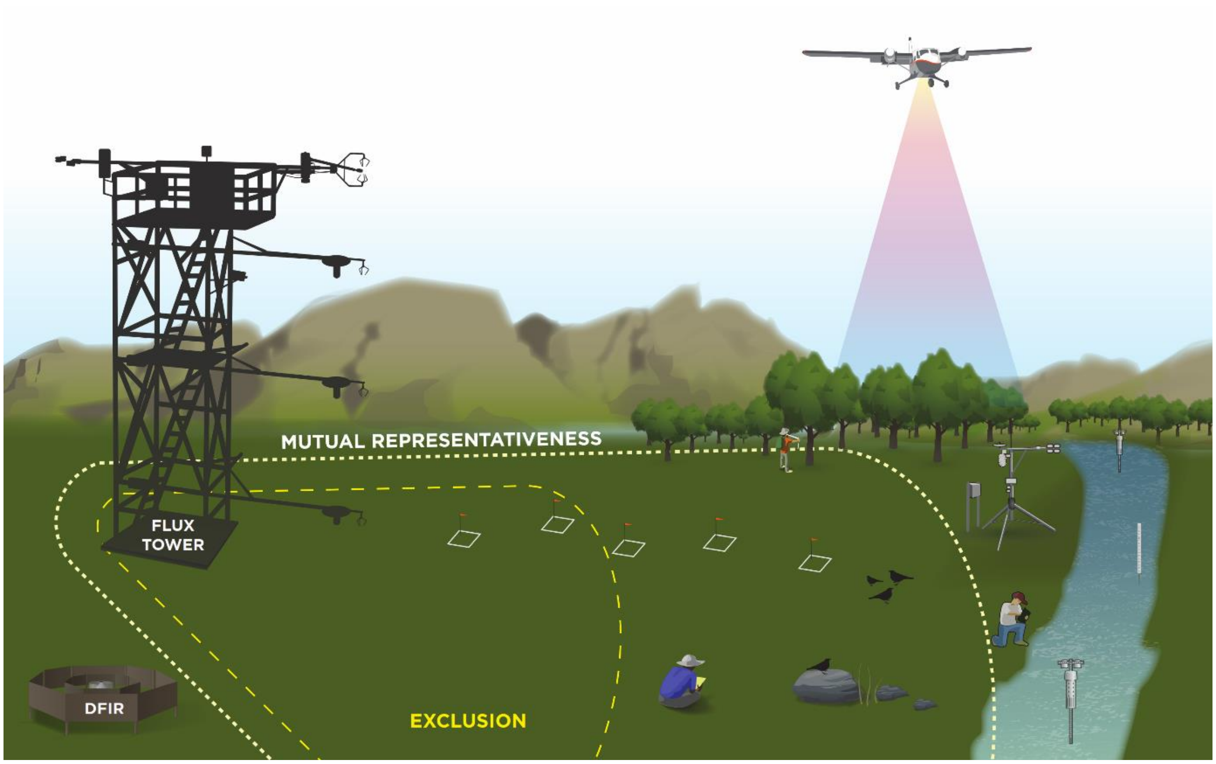

Figure 13. NEON TIS site design with instrument and observation systems covering a wide range of scales.

Appendix C NEON Flux Tower Design

The projected base area of a NEON flux tower is 2 m x 2 m to mimic existing natural ecosystem structures and openings for most forest ecosystem found across NEON sites. Two criteria are applied to determine the height of the tower to ensure that the tower-top extends beyond the roughness sublayer: (i) A fixed tower-measurement height (h_m) of 8 m is used above all short stature ecosystems (e.g., grasslands, shrublands, or agricultural crops) when the mean canopy height is below 3 m. (ii) Over forested or more structurally complex ecosystems, the tower height is determined as $h_m \approx d + 4(h_c - d)$, where h_c is the mean canopy height and d is the zero plane displacement height (Dyer and Hicks, 1970; Hicks, 1976; Lemon, 1960; Monin and Obukhov, 1954; Monteith and Unsworth, 2008). The number of vertical measurement levels on a tower is a function of the ecosystem structure at a specific site. The number of levels varies from four to eight across NEON sites in order to capture ecological meaningful observations across vertical strata. All EC instruments are deployed on booms that extend 4 m from the tower, which is two times the face-width of the tower to reduce the impact of radiation load and flow distortion caused by the tower on the measurements (**Figure 14**).

The command, control and configuration of the eddy-covariance turbulent exchange (ECTE) subsystem is described in detail in AD[01]; in short: it consists of a suite of sensors that record wind speed, temperature, CO_2 and H_2O concentration on the tower top (**Figure 14**, location T08), which are used to calculate turbulent fluxes (Eq. (1) term IV). Wind components are measured in three dimensions by a sonic anemometer (Campbell Scientific Inc., model CSAT-3 firmware: 3.0f; Logan, Utah, USA) operating at 20 Hz. An attitude and heading reference system (Xsens North America Inc., model MTI-300-2A5G4; Culver City, California, USA) is attached to the sonic anemometer collecting data from a gyroscope, accelerometer, and magnetometer at 40 Hz to quantify and correct boom motions and allow rotated sensor deployment. H_2O and CO_2 concentration data are measured at 20 Hz by an enclosed-path infrared gas analyzer (IRGA; Li-Cor Inc., model LI-7200 or LI-7200RS, firmware: 7.3.1; Lincoln, Nebraska, USA). Lastly, a validation system supplies reference gas concentrations to the IRGA for periodic validation enabling thorough uncertainty quantification.

The command, control and configuration of the eddy-covariance storage exchange (ECSE) subsystem is described in detail in AD[01]; in short: it consists of a suite of sensors that record vertical atmospheric profiles of temperature, CO_2 and H_2O concentration (**Figure 14**, locations T02, T04, T06, T07), which are used to calculate storage fluxes (Eq. (1) term I). The air temperature profile is measured at 1 Hz with aspirated temperature sensors (MetOne Instruments, Inc., model 076B-7388; Grant Pass, Oregon, USA). CO_2 and H_2O concentrations are measured at 1 Hz with a closed-path IRGA (Li-Cor, Inc., model LI-840A or LI-850A; Lincoln, Nebraska, USA). The analyzer is located in the instrument hut, and is programmed to operate in two modes, sampling and field validation. During sampling mode, the analyzer will measure air samples from different measurement levels on the tower. During field validation, the analyzer will cease measuring the air samples from the tower levels, and measure known CO_2 gas transfer standards instead. Using a similar strategy, gaseous phase stable carbon and water isotopes are measured at 1 Hz along the tower profile with cavity ring-down spectrometers (CRDS; Picarro Inc., model G2131-I and model L2130-i, firmware 1.5.0-N; Santa Clara, California, USA).

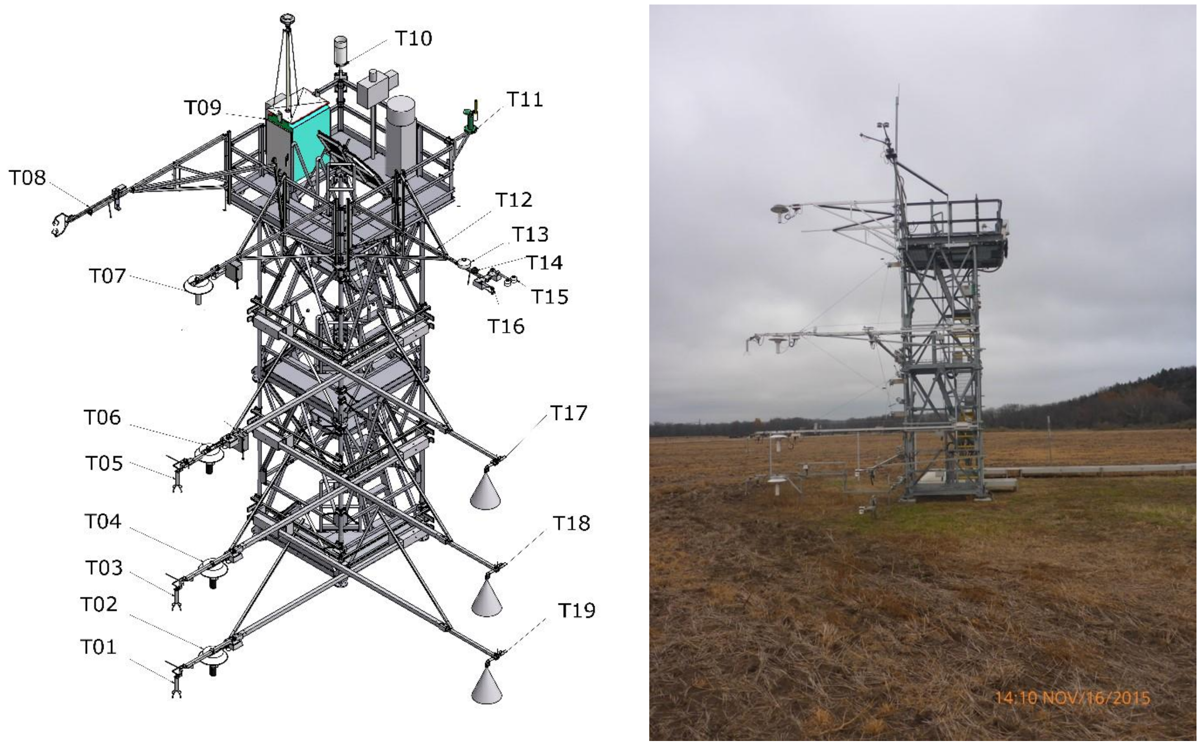

Figure 14. The NEON tower design. Left panel: conceptual design and location of individual instrument assemblies for a 4-level tower: 2D sonic anemometer (T01, T03, T05), air temperature sensor (T02, T04, T06, T07), eddy covariance boom (T08; including 3-D sonic anemometer, infrared gas analyzer, attitude and motion reference sensor), environmental enclosure (T09), secondary precipitation gauge (T10), spectral photometer (T11), radiation boom (T12), pyranometer (T13), sunshine pyranometer (T14), net radiometer (T15), up-facing and down-facing PAR sensors (T16), mid-level radiation boom (T17, T18, T19; including an up-facing PAR sensor and an infrared temperature sensor). Right panel: example of a 4-level tower at NEON CPER site.

Appendix D Acronyms

Acronym	Description
AD	Applicable Documents
ATBD	Algorithm Theoretical Basis Document
C3	Command control and configuration
CI	NEON Cyberinfrastructure project team
CRDS	Cavity ring-down spectrometer
DOM	DOMAIN, e.g. D10
DP	Data product
dp00	sensor readings in engineering units; e.g. concentration as infrared absorptance
dp01	descriptive statistics
dp02	time-interpolated data
dp03	space-interpolated data
dp04	flux data
dp0p	pre-conditioned data in scientific units; e.g. concentration as mole fraction
DPL	DATA PRODUCT LEVEL, e.g. DP1
DQU	Data/qfqm/uncertainty
EC	Eddy covariance
ECSE	Eddy covariance storage exchange
ECTE	Eddy covariance turbulent exchange
HDF	Hierarchical data format
HOR	HORIZONTAL INDEX. Semi-controlled. Examples: Tower=000, HUT=700.
IRGA	Infrared gas analyzer
ISR	Inertial subrange
LST	Land surface temperature
max	Maximum
mfc	Mass flow controller
mfm	Mass flow meter
ML	Measurement level
NA	Not available/not applicable
NaN	Not a number
NEON	National Ecological Observatory Network
NK12	Nordbo and Katul (2012)
NSAE	Net surface-atmosphere exchange
PAR	Photosynthetically active radiation
PRNUM	PRODUCT NUMBER =>5 digit number. Set in data products catalog. TIS = 00000-09999
QA/QC	Quality Assurance/Quality Control
QF	Quality flag
QM	Quality metric
REV	REVISION, e.g. 001
SAE	surface atmosphere exchange
SITE	SITE, e.g. STER
SOM	Science operation management
SONIC	Ultrasonic anemometer/thermometer

Acronym	Description
TERMS	From NEON's controlled list of terms. Index is unique across products
TIS	Terrestrial Instrument System
TMI	TEMPORAL INDEX. Examples: 001=1 minute, 030=30 minute, 999=irregular intervals
VER	VERTICAL INDEX. Semi-controlled. Examples: Ground level=000, second tower level=020

Appendix E Functions

Function	Description
$\sum$	Sum operator
$\int$	Integral operator
∂	Partial differential operator
σ	Standard deviation
abs()	Absolute value
CO	Cospectrum
S	Power spectrum
$\bar{X}$	Short-term (e.g., 30 min) arithmetic mean of atmospheric quantity *X*
$\hat{X}$	Longer-term (e.g., 1 week) arithmetic mean of atmospheric quantity *X*
X'	Immediate deviation from the arithmetic mean of atmospheric quantity *X*
$\overline{X'X'}$, $\overline{X'^2}$	Short-term (e.g., 30 min) sample variance of atmospheric quantity *X*
$\widehat{X'X'}$, $\widehat{X'^2}$	Longer-term (e.g., 1 week) sample variance of atmospheric quantity *X*
$\overline{X'Y'}$	Short-term (e.g., 30 min) sample covariance of atmospheric quantities *X* and *Y*
$\widehat{X'Y'}$	Longer-term (e.g., 1 week) sample covariance of atmospheric quantities *X* and *Y*

Appendix F Parameters, Variables and Subscripts

Parameter subscript or variable	Description	Unit (if applicable)
0	Potential quantity, unless otherwise specified (i.e., under NIST (National Institute of Standards and Technology) standard conditions T_0 = 293.15 K, p_0 = 101.325 kPa	
1…N	Numeric identifier	
1 Hz	1 s temporal resolution	
20 Hz	0.05 s temporal resolution	
40 Hz	0.025 s temporal resolution	
d	Distance/length/height	m
$d_{z,m}$	Measurement height	m
f	Measurement frequency	Hz
F	Flux into or out of an ecosystem	Depending on unit of scalar
FW_{mass}	Wet mass fraction	kg kg^{-1}
i	Running index	
l	Lag time	s
n	Normalized frequency	Dimensionless
N	Sample size	Dimensionless (count)
QF_{Cal} (qfCal)	Quality flag for the Invalid Calibration test	1 = quality test failed 0 = quality test passed -1 = NA
QF_{FINAL} (qfFinal)	Final quality flag	1 = quality test failed 0 = quality test passed -1 = NA
qfIrgaVali	IRGA validation flag, generated to indicate when the sensor is operating under a validation period	1 = validation period 0 = normal operating conditions -1 = NA
QF_{Pers} (qfPers)	Quality flag for the Persistence test	1 = quality test failed 0 = quality test passed -1 = NA
QF_{Rng} (qfRng)	Quality flag for the Range test	1 = quality test failed 0 = quality test passed -1 = NA
$QF_{sciRevw}$ (qfsciRevw)	Flag set during science operation management review	1 = quality test failed 0 = quality test passed

Parameter subscript or variable	Description	Unit (if applicable)
		-1 = NA
QF_{spec} (qfSpec)	EC specific tests (i.e. detection limit, homogeneity and stationarity, development of turbulence tests)	1 = quality test failed 0 = quality test passed -1 = NA
QF_{Step} (qfStep)	Quality flag for the Step test	1 = quality test failed 0 = quality test passed -1 = NA
QM_{α}	Alpha quality metric	%
QM_{β}	Beta quality metric	%
t	Time/duration/period	s
T	Absolute temperature	K
T_{air}	Air temperature	K
t_b	beginning time of the first minute in the 30 minute block when calculating time rate of change	minute
t_e	the last minute of the 30 minute block when calculating time rate of change	minute
T_{SONIC}	SONIC temperature measurement	K
T_v	Virtual temperature	K
u, v, w	Along-, cross- and vertical wind speed	m s^{-1}
x, y, z	Along-, cross- and vertical axes of a Cartesian coordinate system	Dimensionless
X, Y	Placeholder for atmospheric quantities	Depending on unit of atmospheric quantity

Appendix G eddy4R Functions

eddy4R Function	Description
eddy4R.qaqc::def.dspk.br86()	Median filter de-spiking after Brock (1986), Starkenburg et al. (2014)
eddy4R.turb::PFIT_det()	Regression of the planar-fit coefficients over a moving, centered window
eddy4R.base::wrap.derv.prd.day()	Reads the list inpList in the format provided by function eddy4R.base::wrap.neon.read.hdf5.eddy(). For the list entries in inpList the following derived quantities are calculated, each through the call to a separate definition function: inpList$data$time: fractional UTC time, fractional day of year, local standard time; inpList$data$irgaTurb: average signal strength, delta signal strength, total pressure, average temperature, water vapor partial pressure, water vapor saturation pressure, relative humidity, molar density of air (dry air and water vapor), molar density of dry air, wet mass fraction (specific humidity); inpList$data$soni: sonic temperature
eddy4R.base::def.lag()	Lag two datasets, so as to maximize their cross-correlation
eddy4R.base::wrap.neon.dp01()	Compute NEON Level 1 data product descriptive statistics (mean, minimum, maximum, variance, number of non-NA points) across list elements.
eddy4R.turb::REYNflux_FD_mole_dry()	Calculate turbulent vertical flux and auxiliary variables
Waves::cwt()	Morlet mother Wavelet to perform the continuous Wavelet transform of the 3-D wind components, air temperature, as well as H_2O and CO_2 concentration
eddy4R.turb::wrap.wave()	Calculate Wavelet spectrum/cospectrum using the Waves package. The frequency response correction using Wavelet techniques described in Norbo and Katul, 2012 (NK12)
eddy4R.qaqc::def.qf.irga.vali	Definition function to generate the validation flags for IRGA from the IRGA sampling mass flow controller flow rate set point or from the IRGA validation solenoid valves.
eddy4R.qaqc::def.qf.irga.agc	Definition function to generate the signal strength flags for the IRGA from the diagnostic output quality metric
eddy4R.qaqc::wrap.neon.dp01.qfqm	Pre-processing and calculating the random sampling error for the NEON eddy-covariance storage exchange (ECSE) Level 1 data products
eddy4R.ucrt:: def.ucrt.samp.filt()	Calculates the random sampling error via the Salesky et al. (2012) method. Can be used for turbulent moments of any order. If the provided value of \code{NumFilt} is too small and would fail with an error, \code{NumFilt} is incrementally increased in steps of one order of magnitude.
eddy4R.turb::def.vari.wave()	funtion to determine the temporally resolved variance/covariance from continuous wavelet transform
eddy4R.turb::footK04()	Flux footprint after Kljun et a. (2004), Metzger et al. (2012)

eddy4R Function	**Description**
eddy4R.turb::def.dist.rgh()	Aerodynamic roughness length
eddy4R.qaqc::def.qf.irga.vali	Definition function to generate the validation flags for IRGA from the IRGA sampling mass flow controller flow rate set point or from the IRGA validation soleniod valves
eddy4R.qaqc::def.qf.irga.agc	Definition function to generate the signal strength flags for the IRGA from the diagnostic output quality metric
eddy4R.ucrt:: def.ucrt.samp.filt()	Calculates the random sampling error via the Salesky et al. (2012) method
eddy4R.base::wrap.neon.dp01()	Compute NEON Level 1 data product descriptive statistics (mean, minimum, maximum, variance, number of non-NA points) across list elements
eddy4R.base::def.idx.agr()	Definition function to produce a dataframe of indices and corresponding times for aggregation periods
eddy4R.qaqc::def.neon.dp01.qf.grp	Grouping the quality flags of each NEON ECTE and ECSE L1 data product into a single dataframe for further use in the calculation of Alpha, Beta, and Final flag