Data variables and formatting

This document describes variable labels and file formatting for uploading continuously sampled data to AmeriFlux and the European Fluxes databases. The effort to agree on a common and shared system to name and organize the variables collected is an important step toward standardization and improvement of data sharing across networks.

Continuously sampled data are defined as variables that are measured at regular intervals of time, generally daily or more frequent, for a certain period. This means that the time interval between two sequential values is always the same.

The labels for the data variables used here are composed by: a **base name**, indicating the measured or derived physical quantity or quality information; and, **qualifiers** to the base names (e.g., positional information, quality flags, filtering states, gapfilling, processing methods, etc.). Qualifiers are always appended as suffixes to a variable base name.

The rules described in this document apply to all the different steps involved in the measurement life cycle: from the data upload by the tower team to the database, to the centralized processing and QA/QC, to the data distribution to final users. Base names are the same for all the different steps while the suffix qualifiers can be relevant for one or more steps.

Temporal representativeness and timestamps

Two forms of reporting the time associated with a record are used. One using a single timestamp and another using a pair of timestamps. In cases in which the temporal resolution of the period represented matches the temporal resolution of the timestamp being used, there is no ambiguity. For instance: to represent a daily aggregate, a temporal resolution up to the day is sufficient for a timestamps to unambiguously identify the period represented, e.g., 20150728.

However, in situations in which the temporal resolution is different between the period represented and the timestamp, it is necessary to clarify what is being represented by a given timestamp. For instance, using a timestamp with resolution up to the minute -- e.g., 201507281730 -- to identify a single half-hour period can be interpreted in different ways: 5:00pm to 5:30pm, 5:30pm to 6:00pm, or even 5:15pm to 5:45pm.

In the past, a convention for these mismatched cases was used, defining the timestamp as referring to the beginning, middle, or end of the averaging period. Different tower teams, and even different networks, use different conventions. This introduces the problem of keeping track of which convention is being used, which led to many cases of data sets being shifted in time because of confusion on the conventions used.

To address this issue, two variables explicitly referring to start and end of a given period are adopted (TIMESTAMP_START and TIMESTAMP_END), eliminating ambiguity. Data files in **half-hourly**, **hourly**, and **weekly** resolutions use start and end timestamps. Data files using **daily**, **monthly**, and **yearly** resolutions use a single timestamp. Below are examples of resolutions that will use a single TIMESTAMP variable for timekeeping and resolutions requiring the use of both TIMESTAMP_START and TIMESTAMP_END (blank spaces added for legibility).

sample **half-hourly** data file (both timestamps)

```
TIMESTAMP_START, TIMESTAMP_END,  CO2,   ...
201507281700,    20150728173000, 391.1, ...
201507281730,    20150728180000, 391.8, ...
...
```

sample **hourly** data file (both timestamps)

```
TIMESTAMP_START, TIMESTAMP_END,  CO2,   ...
201507281700,  20150728180000, 391.1, ...
201507281800,  20150728190000, 391.8, ...
...
```

sample **daily** data file (single timestamp)

```
TIMESTAMP, CO2,   ...
20150728,  391.1, ...
20150729,  392.8, ...
...
```

sample **weekly** data file (both timestamps)

```
TIMESTAMP_START, TIMESTAMP_END, CO2,   ...
20150714,        20150721,      391.1, ...
20150721,        20150728,      391.8, ...
...
```

sample **monthly** data file (single timestamp)

```
TIMESTAMP, CO2, ...
201507,  391.1, ...
201507,  392.8, ...
...
```

sample **yearly** data file (single timestamp)

```
TIMESTAMP, CO2,   ...
2014,      388.1, ...
2015,      392.8, ...
...
```

Timestamp column ordering (text-based files only)

For text file data representations (i.e., CSV formatted), timestamps must be always in the first column(s) of the file.

Time zone and convention

Time must the reported in local standard time (i.e., without “Daylight Saving Time”). The time zone must be specified using the BADM template for the site.

Missing data

Missing data must be reported using -9999 as replacing value.[1]

Data Variables

1. Data Variable Labels: Base names

Base names indicate fundamental quantities that are either measured or calculated/derived. They can also indicate quantified quality information.

TABLE 1: Base names for data variable labels[2]

Variable	Units	Description
TIMEKEEPING		
TIMESTAMP	YYYYMMDDHHMM	ISO timestamp - short format
TIMESTAMP_START	YYYYMMDDHHMM	ISO timestamp start of averaging period - short format
TIMESTAMP_END	YYYYMMDDHHMM	ISO timestamp end of averaging period - short format
GASES		
CO2	µmolCO2 mol-1	Carbon Dioxide (CO2) mole fraction
H2O	mmolH2O mol-1	Water (H2O) vapor mole fraction
CH4	nmolCH4 mol-1	Methane (CH4) mole fraction
NO	nmolNO mol-1	Nitric oxide (NO) mole fraction
NO2	nmolNO2 mol-1	Nitrogen dioxide (NO2) mole fraction
N2O	nmolN2O mol-1	Nitrous Oxide (N2O) mole fraction

[1] Other values such as -6999 are not acceptable as indication of a missing value for any reason
[2] Please see Appendix A for timekeeping base names used for transitional and compatibility purposes.

O3	nmolO3 mol-1	Ozone (O3) mole fraction
FC	µmolCO2 m-2 s-1	Carbon Dioxide (CO2) flux
FCH4	nmolCH4 m-2 s-1	Methane (CH4) flux
FNO	nmolNO m-2 s-1	Nitric oxide (NO) flux
FNO2	nmolNO2 m-2 s-1	Nitrogen dioxide (NO2) flux
FN2O	nmolN2O m-2 s-1	Nitrous oxide (N2O) flux
FO3	nmolO3 m-2 s-1	Ozone (O3) flux
SC	µmolCO2 m-2 s-1	Carbon Dioxide (CO2) storage flux
SCH4	nmolCH4 mol-1	Methane (CH4) storage flux
SNO	nmolNO mol-1	Nitric oxide (NO) storage flux
SNO2	nmolNO2 mol-1	Nitrogen dioxide (NO2) storage flux
SN2O	nmolN2O mol-1	Nitrous oxide (N2O) storage flux
SO3	nmolO3 mol-1	Ozone (O3) storage flux
FOOTPRINT		
FETCH_MAX	m	Distance at which footprint contribution is maximum
FETCH_90	m	Distance at which footprint cumulative probability is 90%
FETCH_80	m	Distance at which footprint cumulative probability is 80%
FETCH_70	m	Distance at which footprint cumulative probability is 70%
FETCH_FILTER	adimensional	Footprint quality flag: 0 identifies data measured when wind coming from direction that should be discarded
FC_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for FC according to Foken et al 2004
FCH4_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for FCH4 according to Foken et al 2004
FNO_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for FNO according to Foken et al 2004
FNO2_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for FNO2 according to Foken et al 2004
FN2O_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for FN2O according to Foken et al 2004
FO3_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for FO3 according to Foken et al 2004
HEAT		
G	W m-2	Soil heat flux
H	W m-2	Sensible heat flux

LE	W m-2	Latent heat flux
SG	W m-2	Heat storage in the soil above the soil heat fluxes measurement
SH	W m-2	Heat storage in the air
SLE	W m-2	Latent heat storage flux
SB	W m-2	Heat storage in biomass
H_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for H according to Foken et al 2004
LE_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for LE according to Foken et al 2004
MET_WIND		
WD	Decimal degrees	Wind direction
WS	m s-1	Wind speed
WS_MAX	m s-1	maximum WS in the averaging period
USTAR	m s-1	Friction velocity
ZL	adimensional	Stability parameter
TAU	Kg m-1 s-2	Momentum flux
MO_LENGTH	m	Monin-Obukhov length
U_SIGMA	m s-1	Standard deviation of velocity fluctuations (towards main-wind direction after coordinates rotation)
V_SIGMA	m s-1	Standard deviation of lateral velocity fluctuations (cross main-wind direction after coordinates rotation)
W_SIGMA	m s-1	Standard deviation of vertical velocity fluctuations (after coordinate rotation)
TAU_SSITC_TEST	adimensional	Results of the Steady State and Integral Turbulence Characteristics for TAU according to Foken et al 2004
MET_ATM		
PA	kPa	Atmospheric pressure
RH	%	Relative humidity, range 0-100
TA	deg C	Air temperature
VPD	hPa	Vapor Pressure Deficit
T_DP	deg C	Dew point temperature
T_SONIC	deg C	Sonic temperature
T_SONIC_SIGMA	deg C	Standard deviation of sonic temperature
PBLH	m	Planetary boundary layer height

MET_SOIL		
SWC	%	Soil water content (volumetric), range 0-100
TS	deg C	Soil temperature
WATER_TABLE_DEPTH	m	Water table depth
MET_RAD		
ALB	%	Albedo, range 0-100
APAR	µmol m-2 s-1	Absorbed PAR
FAPAR	%	Fraction of absorbed PAR, range 0-100
FIPAR	%	Fraction of intercepted PAR, range 0-100
NETRAD	W m-2	Net radiation
PPFD_IN	µmolPhoton m-2 s-1	Photosynthetic photon flux density, incoming
PPFD_OUT	µmolPhoton m-2 s-1	Photosynthetic photon flux density, outgoing
PPFD_BC_IN	µmolPhoton m-2 s-1	Photosynthetic photon flux density, below canopy incoming
PPFD_BC_OUT	µmolPhoton m-2 s-1	Photosynthetic photon flux density, below canopy outgoing
PPFD_DIF	µmolPhoton m-2 s-1	Photosynthetic photon flux density, diffuse incoming
PPFD_DIR	µmolPhoton m-2 s-1	Photosynthetic photon flux density, direct incoming
SW_IN	W m-2	Shortwave radiation, incoming
SW_OUT	W m-2	Shortwave radiation, outgoing
SW_BC_IN	W m-2	shortwave radiation, below canopy incoming
SW_BC_OUT	W m-2	shortwave radiation, below canopy outgoing
SW_DIF	W m-2	Shortwave radiation, diffuse incoming
SW_DIR	W m-2	Shortwave radiation, direct incoming
LW_IN	W m-2	Longwave radiation, incoming
LW_OUT	W m-2	Longwave radiation, outgoing
LW_BC_IN	W m-2	Longwave radiation, below canopy incoming
LW_BC_OUT	W m-2	Longwave radiation, below canopy outgoing
SPEC_RED_IN	µmolPhoton m-2 s-1	Radiation (red band), incoming
SPEC_RED_OUT	µmolPhoton m-2 s-1	Radiation (red band), outgoing
SPEC_RED_REFL	adimensional	Reflectance (red band)
SPEC_NIR_IN	µmolPhoton m-2 s-1	Radiation (near infra-red band), incoming
SPEC_NIR_OUT	µmolPhoton m-2 s-1	Radiation (near infra-red band), outgoing
SPEC_NIR_REFL	adimensional	Reflectance (near infra-red band)
SPEC_PRI_TGT_IN	µmolPhoton m-2 s-1	Radiation for PRI target band (e.g., 531 nm), incoming

SPEC_PRI_TGT_OUT	µmolPhoton m-2 s-1	Radiation for PRI target band (e.g., 531 nm), outgoing
SPEC_PRI_TGT_REFL	adimensional	Reflectance for PRI target band (e.g., 531 nm)
SPEC_PRI_REF_IN	µmolPhoton m-2 s-1	Radiation for PRI reference band (e.g., 570 nm), incoming
SPEC_PRI_REF_OUT	µmolPhoton m-2 s-1	Radiation for PRI reference band (e.g., 570 nm), outgoing
SPEC_PRI_REF_REFL	adimensional	Reflectance for PRI reference band (e.g., 570 nm)
NDVI	adimensional	Normalized Difference Vegetation Index
PRI	adimensional	Photochemical Reflectance Index
R_UVA	W m-2	UVA radiation, incoming
R_UVB	W m-2	UVB radiation, incoming
MET_PRECIP		
P	mm	Precipitation
P_RAIN	mm	Rainfall
P_SNOW	mm	Snowfall
D_SNOW	cm	Snow depth
RUNOFF	mm	Run off
BIOLOGICAL		
DBH	cm	Diameter of tree measured at breast height (1.3m) with continuous dendrometers
LEAF_WET	%	Leaf wetness, range 0-100
SAP_DT	deg C	Difference of probes temperature for sapflow measurements
SAP_FLOW	mmolH2O m-2 s-1	Sap flow measurement
STEMFLOW	mm	Stemflow
THROUGHFALL	mm	Excess water from wet leaves reaching the ground
T_BOLE	deg C	Bole temperature
T_CANOPY	deg C	Temperature of the canopy
PRODUCTS		
NEE	µmolCO2 m-2 s-1	Net Ecosystem Exchange
RECO	µmolCO2 m-2 s-1	Ecosystem Respiration
GPP	µmolCO2 m-2 s-1	Gross Primary Productivity

2. Data Variable Labels: Qualifiers

Qualifiers are suffixes adding information about the variable. Multiple qualifiers can be added to a variable base name and they must follow the order in which they are presented here.

Qualifiers are classified into types: PRESENT and CHOICE. A qualifier is of PRESENT type if it indicates the occurrence of the qualifier (e.g., a data variable is gapfilled). A qualifier is of CHOICE type if it indicates one of many possible choices for its occurrence (e.g., which method was used for gapfilling a variable).

In general, qualifiers are reserved for use at the network level (network teams only) and should not be used for data uploads by tower teams. Exceptions are noted in the use documentation for individual qualifiers.

2.1. Qualifiers: General

General qualifiers indicate additional information about a variable.

2.1.1. _PI (Provided by PI/tower team)

- Type: PRESENT
- Use: network team only
- Details: It is the variable version after filtering, gapfilling or any other specific processing by the tower team, independent from the version created at the network level (database team). Must be always associated to metadata describing processing applied to variable in versions distributed to the users. This flag can only be combined with the _F and _QC flags to indicate gapfilling of variable (see below) or quality flags (see below), with the condition that the method is described in the BADM Instrument template; it cannot be combined with method qualifiers, for instance.

2.1.2. _QC (Quality control flag)

- Type: PRESENT
- Use: network team only
- Details: Used only by the network team to report quality check resulting from standard and centralized quality control of the data.

2.1.3. _F (Gapfilled variable)

- Type: PRESENT
- Use: tower team and network team
- Details: Indicates that the variable has been gapfilled.

2.1.4. _IU (Instrument units)

- Type: PRESENT
- Use: tower team or network team
- Details: It indicates that the variable is using instrument units (e.g., counts, mV, absorbance) instead of standard units (e.g., mm, degC, µmol mol-1). This qualifier is in general used only in the data uploads to the network teams and only for specific variables.

2.2. Qualifiers: Theme, Methods, and Uncertainty

Placeholder for theme, methods, and uncertainty related qualifiers.
This will be their position in the order of suffixes to variable labels.
These qualifiers are currently being defined along with the post-processing results.

2.2.1 Meteo

2.2.2 Energy

2.2.3 NEE

2.2.4 Partitioning

2.3. Qualifiers: Positional (_H_V_R)

Positional qualifiers indicate relative positions of sensors originating variable time series. Variables submitted to the database should be results of single sensor measurements. There are variables that are measured in different points (e.g. along a vertical profile or in different positions in the horizontal plane) or monitored in the same position but using two or more sensors. The sensor position information is recorded in the BADM (Instrument template)[3]. A given data variable is mapped to a particular sensor also using the BADM Instruments template. The identification of the variable is done via the variable code plus the positional qualifier.

2.3.1. **_H_V_R** (Three-index positional qualifier)

- Type: PRESENT
- Use: tower team and network team
- Details: The three components of the qualifier are integer numbers that represent:

  H: horizontal position index

[3] Note that the indices might be reassigned from the upload time to the publication time at network level. Any such change will be based on BADM reported positions and feedback from tower teams.

V: vertical position index
R: replicate index

- Note: The numbers indicate positional indices in their respective planes, and not measurements of distances. H, V, and R above are to be replaced with numerical indices.

Indices:

Horizontal position (H): same value identifies the same position in the horizontal plane. For example all the variables associated to sensors in a vertical profile would have the same H qualifier.

Vertical position (V): indexes must be in order, starting from the highest (for example V=1 for the highest temperature sensor of a profile or for the higher, i.e. more superficial, soil temperature sensor in a profile). The indexes are assigned on the basis of the relative position for each vertical profile separately.

Replicates (R): index identifying a variable measured in the same position (H and V) but with different sensors. Two collocated sensors should be considered “replicates” if the differences in the values measured are mainly due to differences in the instruments/technique and not to the difference in the position. This is clearly different for different variables. For example two radiometers for incoming radiation at 1 meter of distance could be considered replicates while two soil water content sensors at the same distance could be treated as different positions (different H values).

Example:

Two profiles of soil temperature in two different horizontal positions: First profile has 4 sensors at -2, -5, -10 and -30 cm, second profile has 3 sensors, one at -5 and two at -30 cm (e.g. different models). The codes will be:

Sensor	Code
Profile 1, -2 cm	TS_1_1_1
Profile 1, -5 cm	TS_1_2_1
Profile 1, -10 cm	TS_1_3_1
Profile 1, -30 cm	TS_1_4_1
Profile 2, -5 cm	TS_2_1_1
Profile 2, -30 cm, Sensor A	TS_2_2_1
Profile 2, -30 cm, Sensor B	TS_2_2_2

Adding sensors:

- when a new sensor is added in the horizontal space, a new value of the H qualifier is added
- when a new level is added in an existing vertical profile the whole profile should be renamed but it is enough to use a different code (even if not in the correct order) and metadata about the position using the BADM. The whole profile will be renamed centrally in the database, including also years when the level was not measured and where the values will be filled with -9999.

Following the example above, if two new sensors are added, one in a new position at -30 cm and the other along profile number 2 at -20 cm the codes will become:

Sensor	Code
Profile 1, -2 cm	TS_1_1_1
Profile 1, -5 cm	TS_1_2_1
Profile 1, -10 cm	TS_1_3_1
Profile 1, -30 cm	TS_1_4_1
Profile 2, -5 cm	TS_2_1_1
Profile 2, -20 cm	TS_2_2_2
Profile 2, -30 cm, Sensor A	TS_2_3_1
Profile 2, -30 cm, Sensor B	TS_2_3_2
Profile 3, -30 cm	TS_3_1_1

Positional (and aggregation) qualifiers are the last qualifiers in a variable label.

2.4. Qualifiers: Aggregation

The sensor level data identified by the _H_V_R qualifier are aggregated in the database based on the base variable code, position qualifiers, metadata and discussion with the tower team.

2.4.1. _H_V_*A* (Aggregation of replicates)

- Type: PRESENT
- Use: network team only
- Details: If replicates can be aggregated (e.g. because the sensors are with similar quality level) they are averaged and the result has as third qualifier in the _H_V_R the letter "A". For example still in the case presented above, if the TS_2_3_1 and TS_2_3_2 can be averaged, the result will be named TS_2_3_A
- Note: H and V above are to be replaced with numerical indices, while the character A is to be used as is.

2.4.2. _*#* (Aggregation per layer)

- Type: PRESENT
- Use: tower team or network team
- Details: Variables measured along one or more vertical profiles are renamed/aggregated per layer in order to provide a reduced number of variables. This is done with respect to the single sensor type and gives the best possible representation of the footprint.
- Note: # above is to be replaced by a numerical index.
- Note: variables that are representative of the footprint of the tower, either through aggregation or spatial resolution might not need the positional qualifiers (with a few exceptions like soil temperature where the qualifiers always persist indicating the vertical layer).

2.4.3. _SD (Standard deviation - spatial variability)

- Type: PRESENT
- Use: network team only
- Details: Standard deviation of the per layer aggregation.

2.4.4. _N (Number of samples - spatial variability)

- Type: PRESENT
- Use: network team only
- Details: Number of samples of the per layer aggregation.

Example:
When the variable is measured by sensors in different positions in the horizontal plane but at "similar" height/depth, they are averaged. The decision to aggregate variables from two or more sensors is based on the metadata (for the position) and discussion with the tower team. In case of sensors with replicates, the values used in this aggregation is the _X_Y_A (already aggregated across replicates)

The results of the renaming/aggregation is labeled with a qualifier indicating the horizontal layer _#. Following the example above the renamed/aggregated variables could be:

TS_1 = TS_1_1_1 (-2 cm)
TS_2 = TS_1_2_1 + TS_2_1_1 (-5 cm)
TS_3 = TS_1_3_1 (-10 cm)
TS_4 = TS_2_2_2 (-20 cm)
TS_5 = TS_1_4_1+TS_2_3_A+TS_3_1_1 (-30 cm)

When for a specific layer (_#) two or more sensors exists additional variables are also created such the Standard Deviation between sensors, identified with _SD and the number of sensors in the layer, identified with _N. In the case above this would happen for TS_2 and TS_5, producing TS_2_SD, TS_2_N, TS_5_SD and TS_5_N

Note:
If a variable is not measured along a vertical profile, the _# qualifier is not used. For example if there is only one radiation sensors measuring SW_IN, SW_IN_1 is not created. Similarly if there are different PPFD sensors below canopy measuring PPFD_BC_IN, they are averaged and standard deviation calculated but the _# is not used (the variables are named directly PPFD_BC_IN and PPFD_BC_IN_SD).

APPENDIX A. Transitional Timekeeping Support

Alternate timekeeping formats are supported for transitional purposes. However, use of the official format proposed in the main table is strongly encouraged. Existing data sets using the timekeeping conventions below can be supported **ONLY IF PREVIOUSLY AGREED WITH THE NETWORK TEAM**. The alternate versions are listed in the preferred order. Also note that **only one timekeeping format should be used** (i.e., only the preferred standard format OR only **one** of the alternates below). Any of the alternate versions of timestamps must report the end of the averaging period. For example, if the timestamps are 12:30; 13:00; 13:30 etc., the values associated to the 13:30 timestamp are representative of the measurements done between 13:00 and 13:30. This means midnight must be reported as 00:00 of the following day and the last value of the year has a timestamp 00:00 of January 1st of the next year.

TIME KEEPING (ALTERNATE VERSIONS AMERIFLUX)		
ALTERNATE TIMEKEEPING 1		
YEAR	YYYY	Four digit year
DOY	DDD	Day of year
HRMIN	HHMM	Hour and Minute of the day (indicating end of averaging period)
ALTERNATE TIMEKEEPING 2		
YEAR	YYYY	Four digit year
DOY	DDD	Day of year
HOUR_DEC	HH.DECMIN	Hour of the day and decimal minutes (indicating end of averaging period)
ALTERNATE TIMEKEEPING 3		
YEAR	YYYY	Four digit year
DTIME	DDD.DECTOD	Day of year and decimal time of the day (indicating end of averaging period)

TIME KEEPING (ALTERNATE VERSION EUROPEAN DB)		
ALTERNATE TIMEKEEPING 4		
DATE	DD/MM/YYYY	Date
TIME	HH:MM	Hour and Minute of the day from 00:00 to 23:30 (indicating end of averaging period)