

# Post-processing Data Analysis Tools: From Theory to Application

ChinaFlux Eddy Covariance Training (virtual)

August 26, 2020

Dave Johnson

# Eddy Covariance Data/Workflow Pipeline

Interesting Hypothesis



Raw Data Collection



Data Processing



Post-Processing and Analysis



Answers & Publications



# Eddy Covariance Data/Workflow Pipeline

Interesting Hypothesis



Raw Data Collection



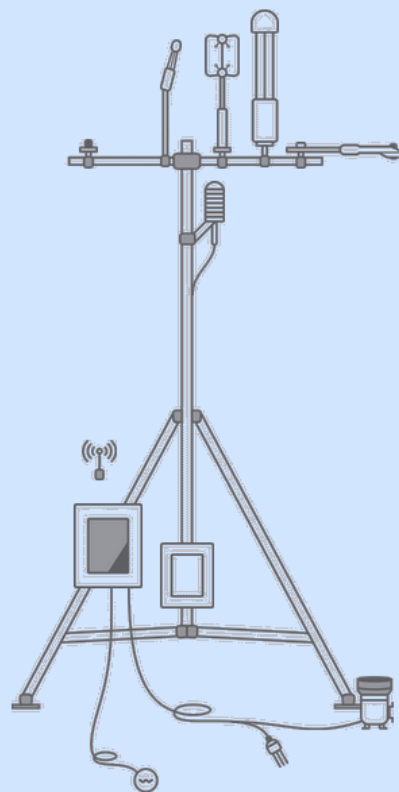
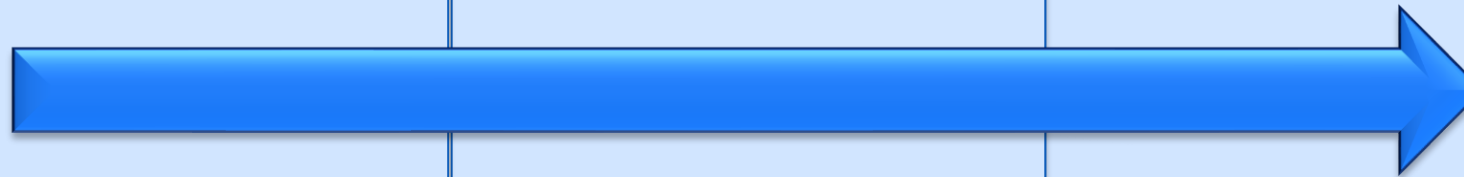
Data Processing



Post-Processing and Analysis



Answers & Publications



# Data Processing – Output

- 17,500 records per year
  - 30 minute files

# Processing versus Post-processing

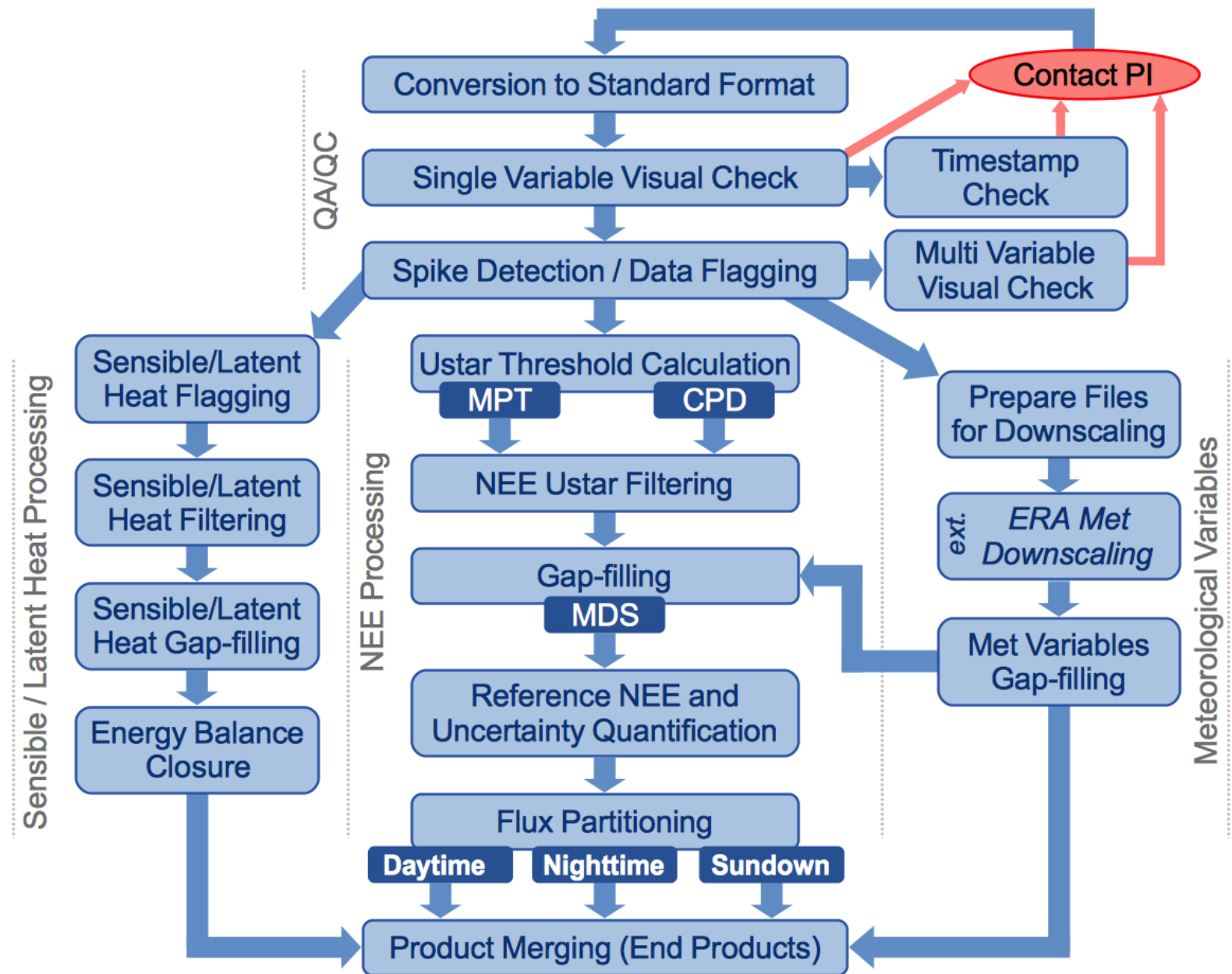
- What is the difference?

# Data Analysis: Common Post-Processing steps

- Quality Control
- Meteorological Gap Filling
- Averaging Replicates and/or Redundant Variables
- Friction Velocity ( $u^*$ ) Threshold Detection
- Flux Gap Filling
- Flux Partitioning
- Energy Balance Closure/Analysis
- Footprint Analysis
- ...

# Typical Post-Processing Pipeline

- Example from the FLUXNET2015 Release



# Primarily four options for Post-Processing...

- Custom
- ONEFlux
- REddyProc
- Tovi



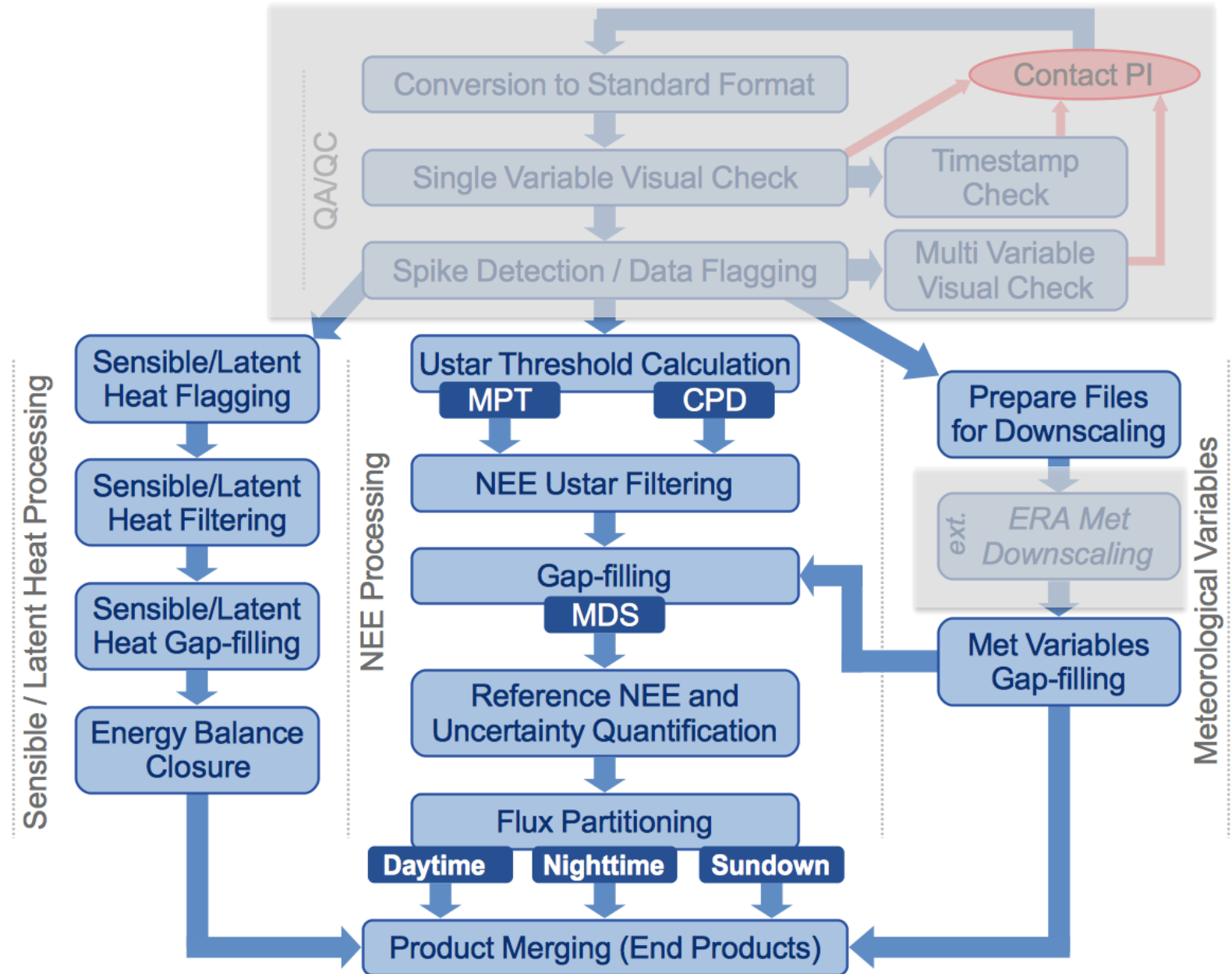
# Custom user-developed

- Using combinations of software programming languages...



# ONEFlux

- Developed in a collaboration between:
  - European Ecosystem Fluxes Database
  - ICOS Ecosystem Thematic Centre (ICOS-ETC)
  - AmeriFlux Management Project (AMP).



# ONEFlux (continued)

```
usage: runoneflux.py [-h] [--perc [PERC [PERC ...]]]
                  [--prod [PROD [PROD ...]]] [-l LOGFILE] [--force-py]
                  [--mcr MCR_DIRECTORY] [--ts TIMESTAMP] [--recint {hh,hr}]
                  COMMAND DATA-DIR SITE-ID SITE-DIR FIRST-YEAR LAST-YEAR
```

## positional arguments:

COMMAND	ONEFlux command to be run [all, partition_nt, partition_dt]
DATA-DIR	Absolute path to general data directory
SITE-ID	Site Flux ID in the form CC-XXX
SITE-DIR	Relative path to site data directory (within data-dir)
FIRST-YEAR	First year of data to be processed
LAST-YEAR	Last year of data to be processed

## optional arguments:

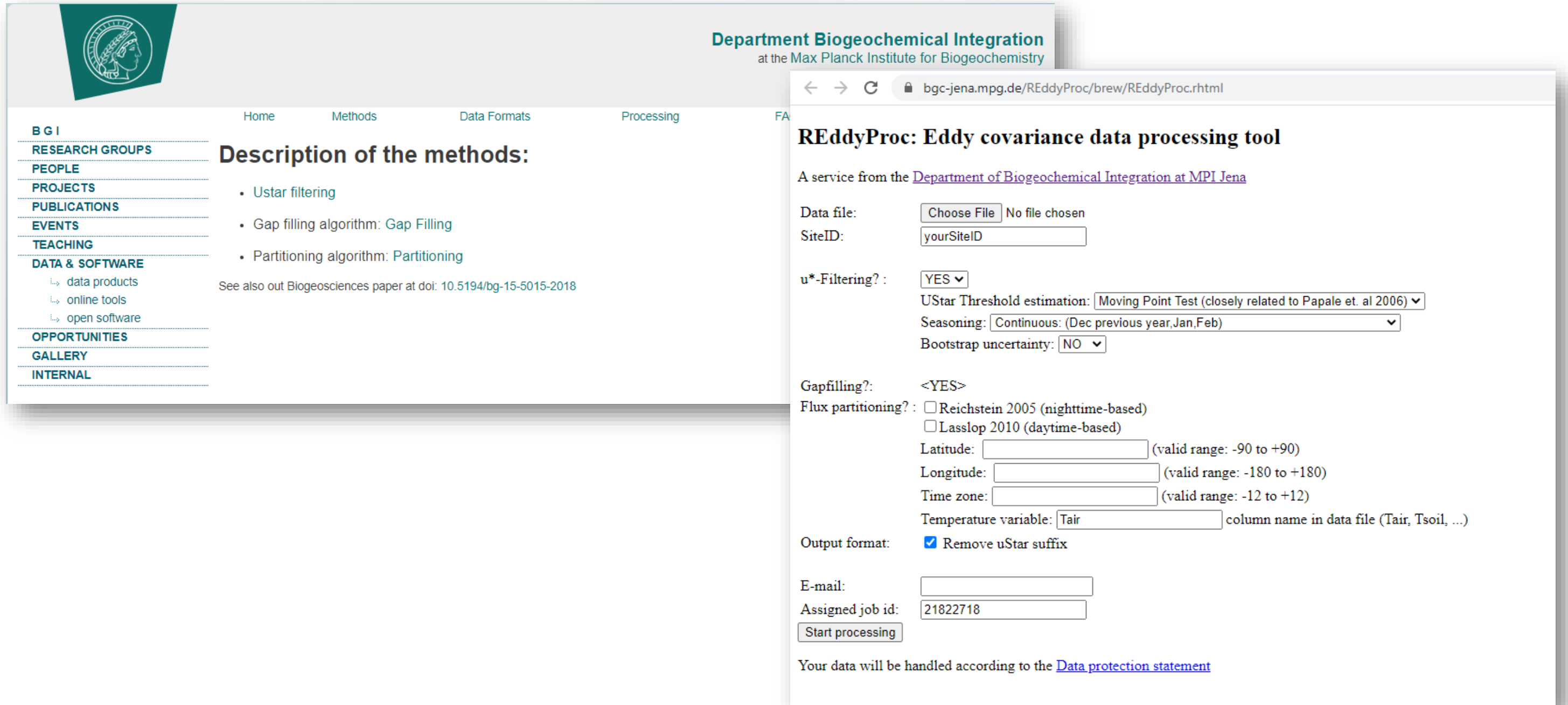
-h, --help	show this help message and exit
--perc [PERC [PERC ...]]	List of percentiles to be processed
--prod [PROD [PROD ...]]	List of products to be processed
-l LOGFILE, --logfile LOGFILE	Logging file path
--force-py	Force execution of PY partitioning (saves original output, generates new)
--mcr MCR_DIRECTORY	Path to MCR directory
--recint {hh,hr}	Record interval for site

COMMAND	ONEFlux command to be run [all, partition_nt, partition_dt]
DATA-DIR	Absolute path to general data directory
SITE-ID	Site Flux ID in the form CC-XXX
SITE-DIR	Relative path to site data directory (within data-dir)
FIRST-YEAR	First year of data to be processed
LAST-YEAR	Last year of data to be processed

## Example:

```
$ python runoneflux.py all \
  "../datadir/" US-ARc "US-ARc_sample_input" \
  2005 2006 -l fluxnet_pipeline_US-ARc.log \
  --mcr ~/bin/matlab/v94/ --recint hh
```

# REddyProc



The image shows a screenshot of the REddyProc web interface. On the left is a navigation menu for the Department of Biogeochemical Integration (BGI) at the Max Planck Institute for Biogeochemistry. The main content area displays the 'Description of the methods' page, which lists three methods: Ustar filtering, Gap filling algorithm: Gap Filling, and Partitioning algorithm: Partitioning. A reference to a Biogeosciences paper is also provided. Overlaid on the right is a browser window showing the REddyProc configuration form. The browser address bar shows 'bgc-jena.mpg.de/REddyProc/brew/REddyProc.rhtml'. The form includes fields for 'Data file' (with a 'Choose File' button), 'SiteID' (text input), 'u\*-Filtering?' (dropdown menu set to 'YES'), 'UStar Threshold estimation' (dropdown menu set to 'Moving Point Test'), 'Seasoning' (dropdown menu set to 'Continuous'), and 'Bootstrap uncertainty' (dropdown menu set to 'NO'). There are also checkboxes for 'Gapfilling?' (set to '<YES>') and 'Flux partitioning?' (with options for Reichstein 2005 and Lasslop 2010). Latitude, Longitude, and Time zone are text input fields with valid ranges. The 'Temperature variable' is a text input field with a hint. The 'Output format' section has a checked checkbox for 'Remove uStar suffix'. At the bottom, there is an 'E-mail' field, an 'Assigned job id' field (containing '21822718'), and a 'Start processing' button. A link to the 'Data protection statement' is also present.

**BGI**  
RESEARCH GROUPS  
PEOPLE  
PROJECTS  
PUBLICATIONS  
EVENTS  
TEACHING  
DATA & SOFTWARE  
↳ data products  
↳ online tools  
↳ open software  
OPPORTUNITIES  
GALLERY  
INTERNAL

Home Methods Data Formats Processing FA

## Description of the methods:

- Ustar filtering
- Gap filling algorithm: Gap Filling
- Partitioning algorithm: Partitioning

See also out Biogeosciences paper at doi: 10.5194/bg-15-5015-2018

## REddyProc: Eddy covariance data processing tool

A service from the [Department of Biogeochemical Integration at MPI Jena](#)

Data file:  No file chosen  
SiteID:

u\*-Filtering? :    
UStar Threshold estimation:    
Seasoning:    
Bootstrap uncertainty:

Gapfilling?:

Flux partitioning? :  Reichstein 2005 (nighttime-based)  
 Lasslop 2010 (daytime-based)  
Latitude:  (valid range: -90 to +90)  
Longitude:  (valid range: -180 to +180)  
Time zone:  (valid range: -12 to +12)  
Temperature variable:  column name in data file (Tair, Tsoil, ...)

Output format:  Remove uStar suffix

E-mail:   
Assigned job id:

Your data will be handled according to the [Data protection statement](#)

# eddy4R and REddyProc

- NEON has a family of eddy-covariance software open-source “eddy4R” packages in the R language for statistical computing. eddy4R intends to provide raw-data processing capabilities that complement existing high-level functionalities of the “REddyProc” R-package.

A screenshot of the RStudio interface. The main editor window shows an R script with the following code:

```
1 #####  
2 #####  
3 ####  
4 #' @title Example workflow for processing  
5   NEON eddy-covariance turbulence data with  
6   eddy4R-Docker 0.2.0  
7  
8 #' @author  
9   David Durden \email{ddurden@battelleecolog  
10  y.org} \cr  
11   Stefan Metzger \email{eddy4R.info@gmail  
12  .com} \cr
```

The console window at the bottom shows the R startup message:

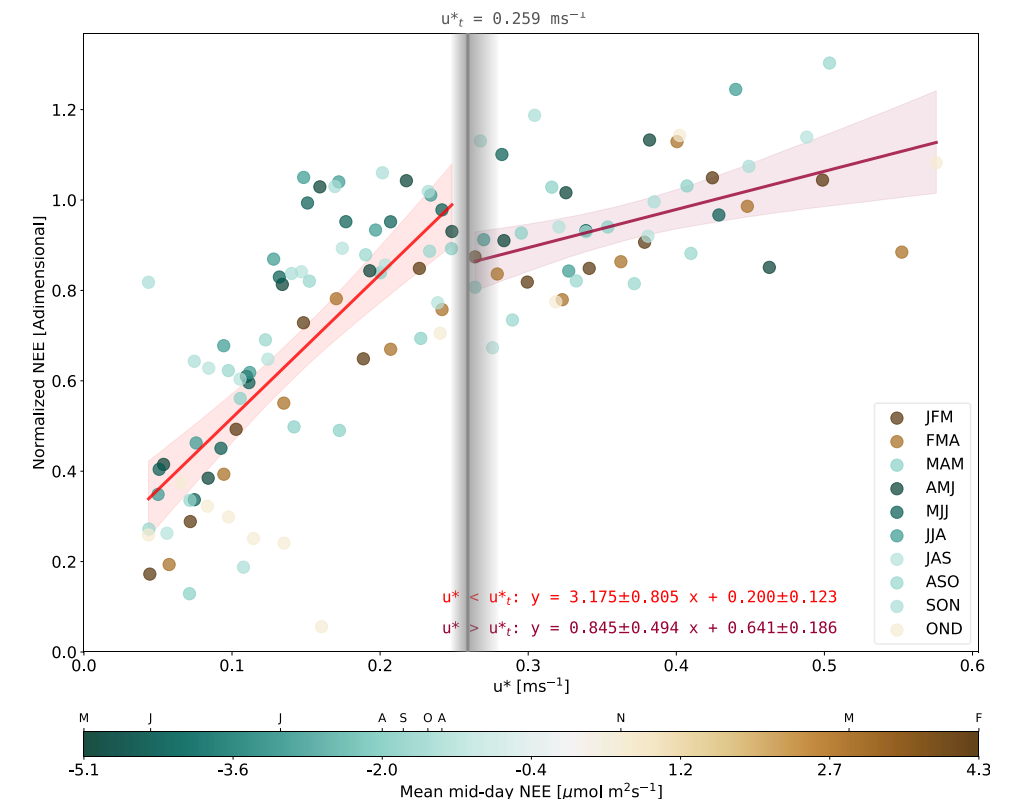
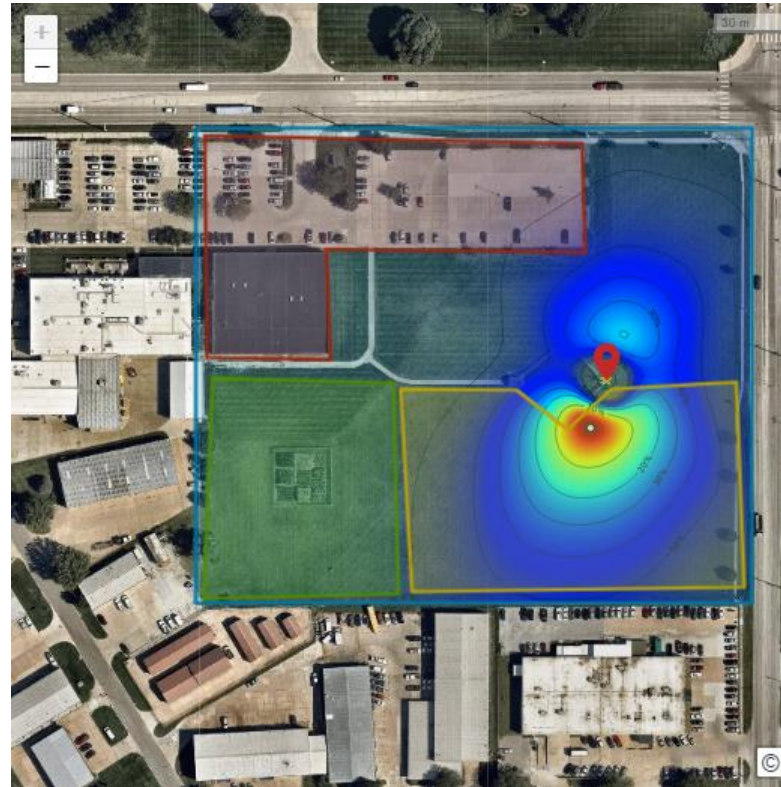
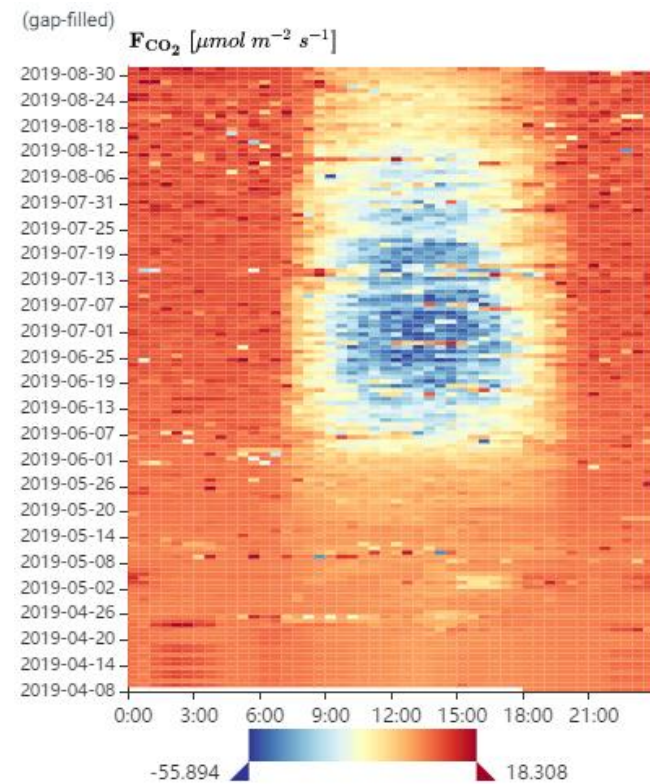
```
R version 3.4.0 (2017-04-21) -- "You Stupid Darkness"  
Copyright (C) 2017 The R Foundation for Statistical  
Computing  
Platform: x86_64-pc-linux-gnu (64-bit)
```

The Environment pane on the right shows an empty global environment. The Packages pane on the right shows a list of installed packages:

Name	Description	Version
<input type="checkbox"/> eddy4R.base	Eddy-covariance calculation for R: Base package	0.2.0
<input type="checkbox"/> eddy4R.qaqc	Eddy-covariance calculation for R: Quality assurance and quality control	0.2.0
<input type="checkbox"/> REddyProc	Data processing and plotting utilities of (half-)hourly eddy-covariance measurements	1.0.0

# Tovi Software

- A suite of tools for analyzing eddy covariance flux and meteorological datasets that have been processed with EddyPro Software.
- Tovi is designed to bring together analysis tools, developed by or with the scientific community, into a unique, interactive and intuitive environment, to streamline typical processes such as quality control, flux gap filling and flux partitioning, as well as to enable novel analysis such as footprint-based flux allocation.



# The first step in using any post-processing package

- Formatting data for importing
  - Files (fluxes, biomet, metadata, etc)
  - Time Steps (30 minute)
  - Variable names (standardized)
  - Units (standardized)
  - Positional Qualifiers (for submission)
  - Defining (for importing in different programming languages)
- Manual versus automated

# Import (and integrate) Ancillary Weather Station Data - Optional

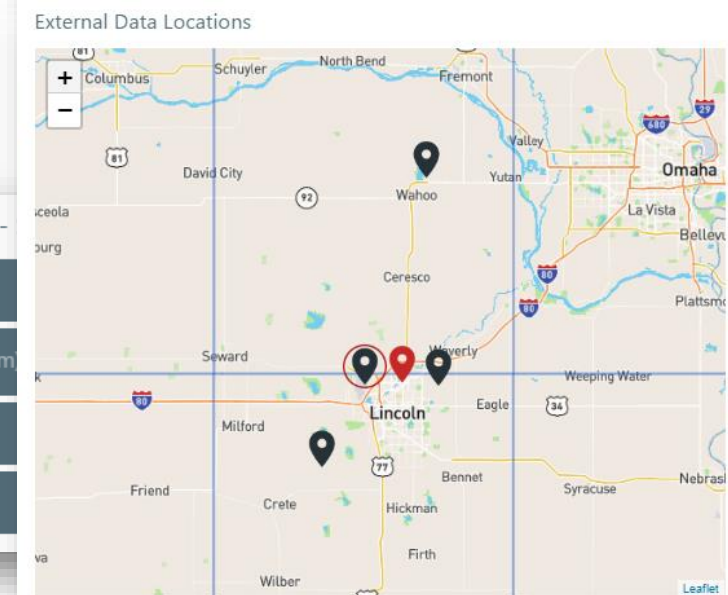
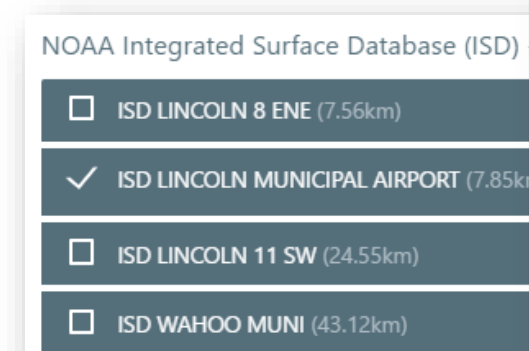
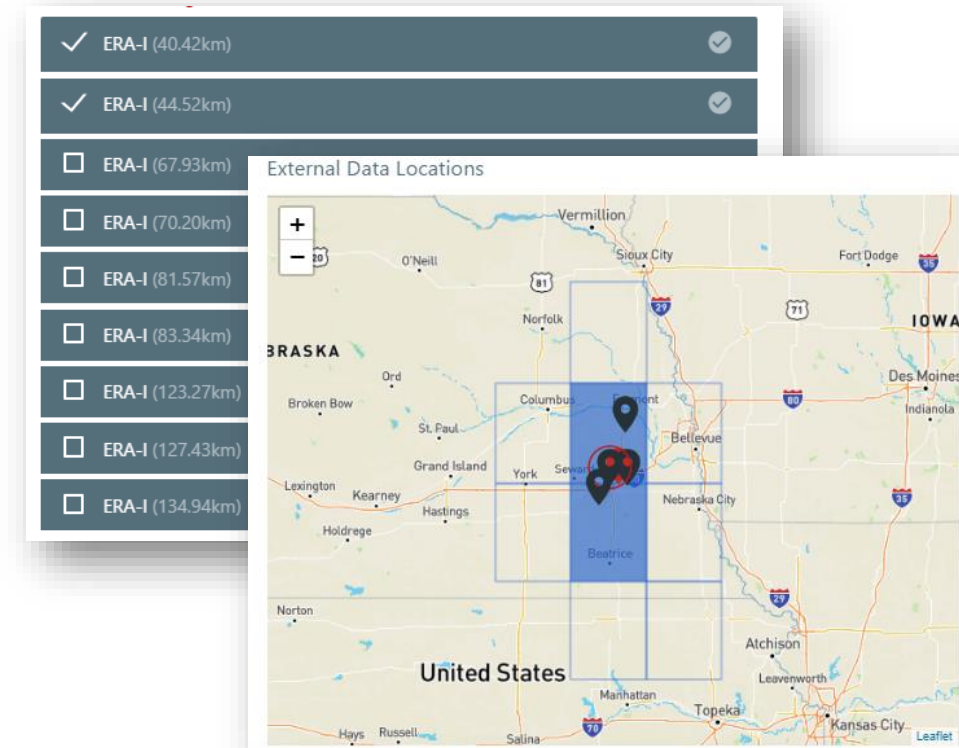
- Biological and Meteorological (Biomet) data are important for:
  - Quality Assurance and Quality Control (QA/QC) - Energy Balance closure.
  - To help 'explain' flux results - The physical environment has profound effects on the biology as well as on the surface-atmosphere exchange.
  - Gap Filling of surface fluxes - A complete record of flux drivers is essential for successful gap filling of surface fluxes
- What if you are missing Biomet data?
  - Before the gaps in the fluxes are filled, gaps are filled in the environmental measurements that are used to drive gap filling of the fluxes.
- Ancillary Weather Station Data
  - Can be used for all the above purposes



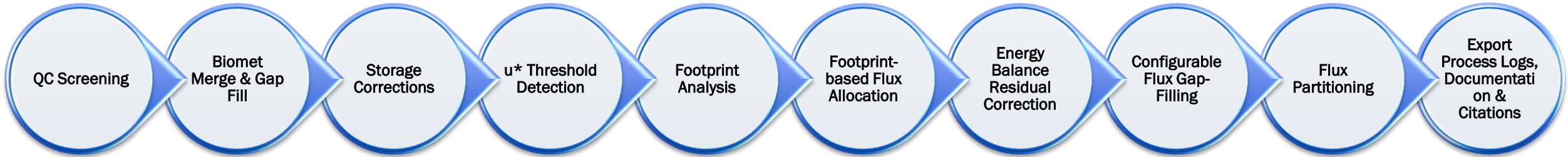
# Example – Automated import of Weather Station

## Data using Tovi

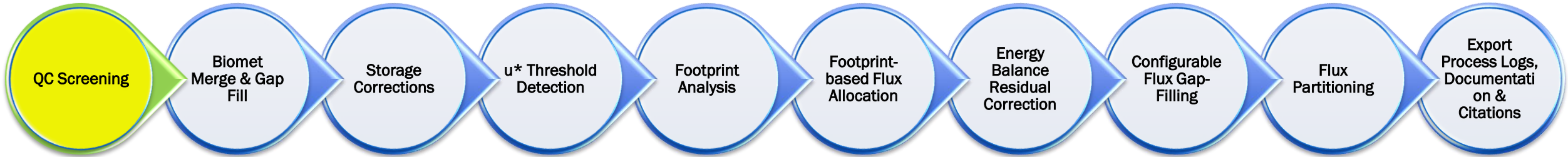
- Tovi locates nearby external datasets that can be imported.
- On the map, select datasets to download by choosing regions on the map (for gridded data) or by selecting the dataset.
- The two external datasets available are:
  - The European Center for Medium Range Weather Forecasting (ECMWF) Re-Analysis Interim (**ERA-I**) data.
    - The ERA-I data set is a gridded, surface meteorological data set that covers the Earth's surface at a horizontal resolution of 80km.
  - The Integrated Surface Dataset (**ISD**, released by NOAA).
    - The ISD data set is a collection of approximately 14,000 active surface meteorological observing stations.
- You can choose both ERA-I and ISD datasets.

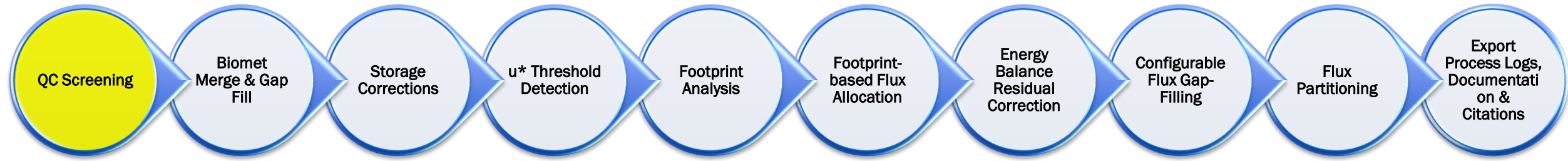


# Typical workflow for Post Processing...

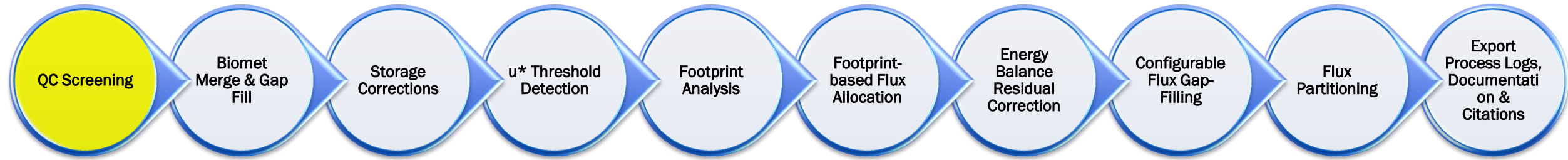


# Typical workflow for Post Processing...

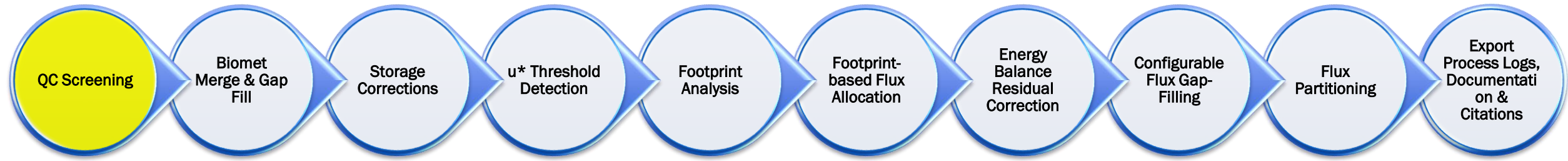




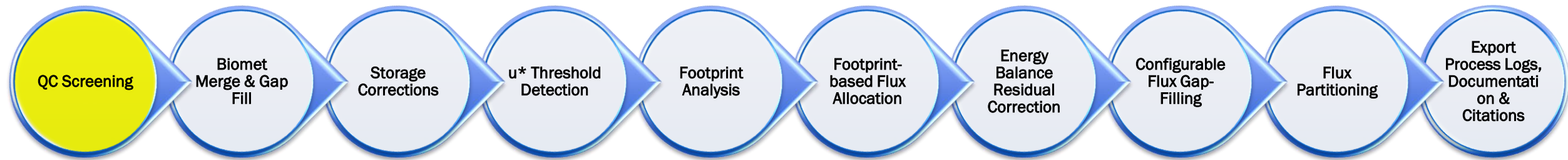
- Quality Check (screening)
  - Performed on the 30-minute fluxes (and averages)
  - Goal is to eliminate errors in processing, outliers, physical limits, flags
  - Will create gaps in the data



- Filtering options for all variables
  - Data Range – set absolute limits based on
    - Ecosystem/Site knowledge
    - Physical Limits
    - Diurnal Patterns
    - Experience
  - Date Exclusions – pick a period or time
  - Dependency – using other variables
  - Smart Y-Range – filter values 10 times the Inter Quantile Range



- Filtering options for fluxes
  - Foken Flags - computed by EddyPro
  - $u^*$  value - after using a  $u^*$  threshold detection tool
  - Gap Filling Flag - after using a gap filling tool
  - Footprint Flag - quality flag based on footprint contribution

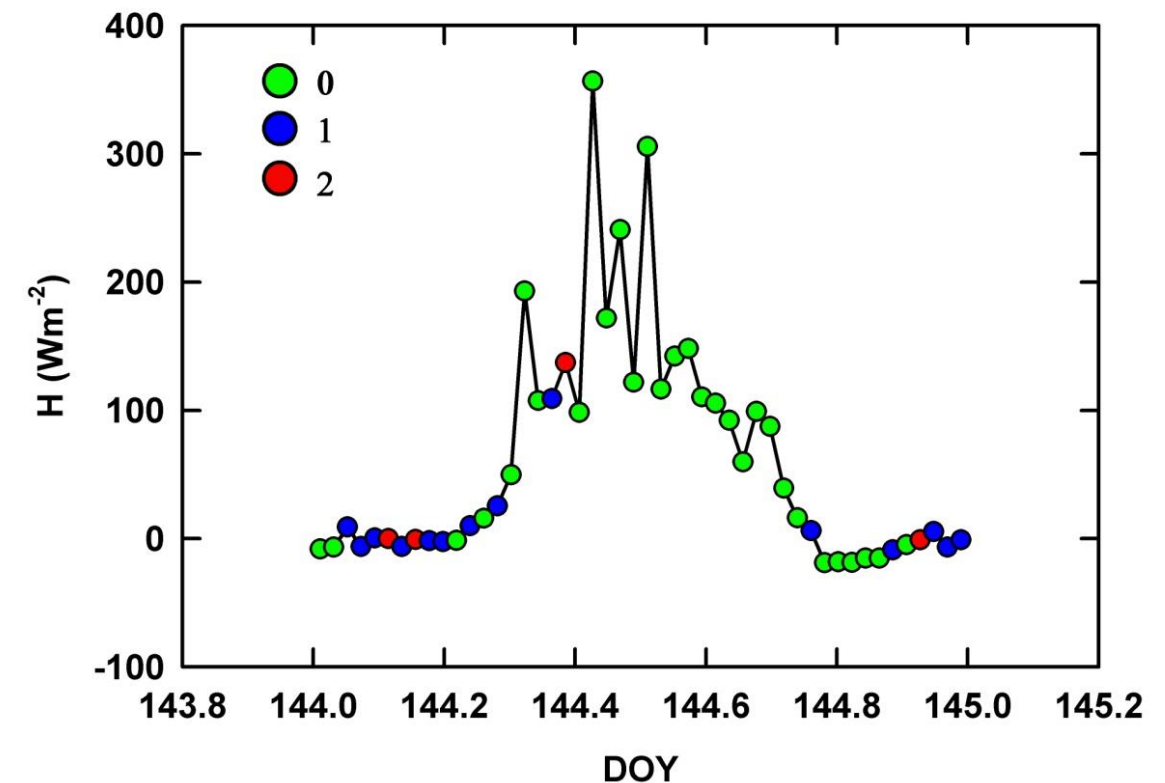


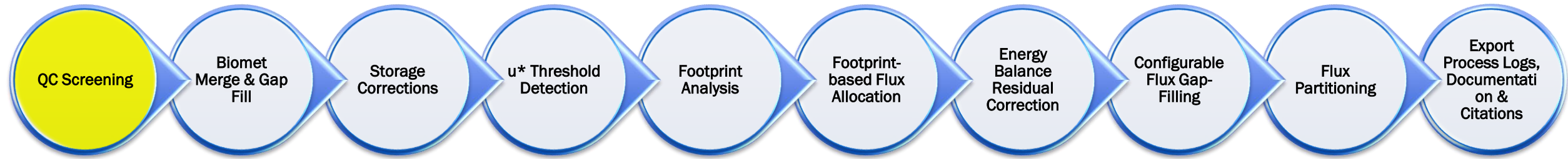
- Foken (Flux Quality) Flags
- Have the theoretical assumptions been met?
  - Do we have steady state conditions?
  - Do we have well developed turbulence?

$RN_{cov}$	$ITC\sigma$	QA/QC FLAG	DATA QUALITY
$\leq 30$	$\leq 30$	0	High
$\leq 100$	$\leq 100$	1	Moderate
$> 100$	$> 100$	2	Low

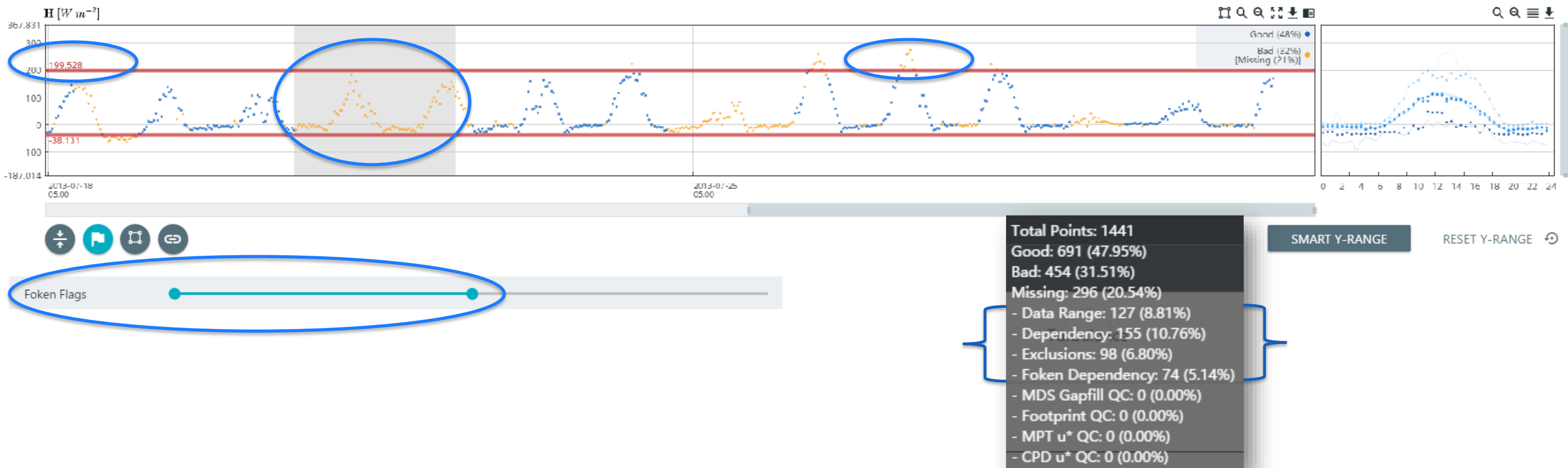
↑  
*Steady State test:*  
 to determine  
 stationarity ( $RN_{cov}$ )

↙  
*Developed Turbulent Conditions*  
*test: Integral Turbulence*  
*Characteristics ( $ITC\sigma$ )*

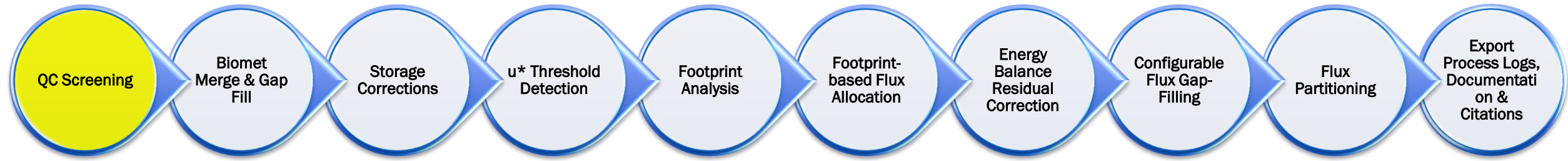




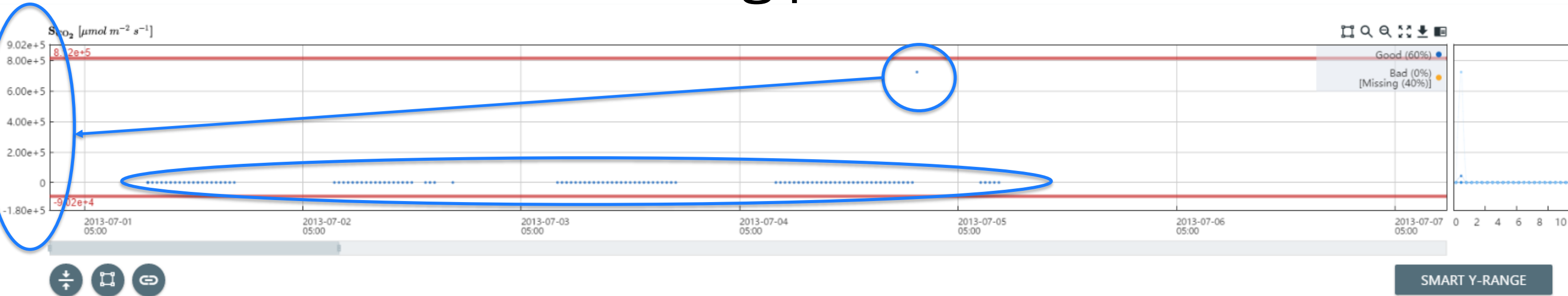
- Example; using min/max thresholds, date exclusions, and Foken flags to clean data in Tovi

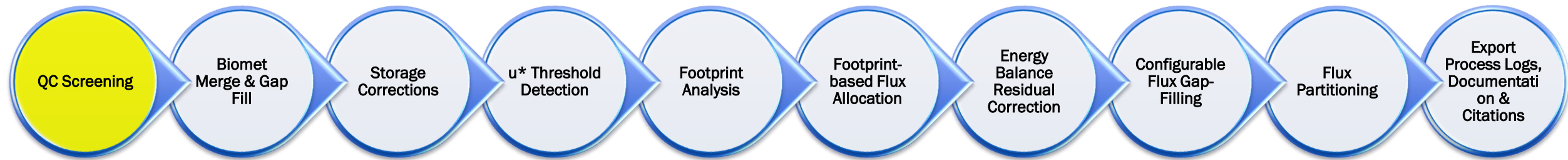




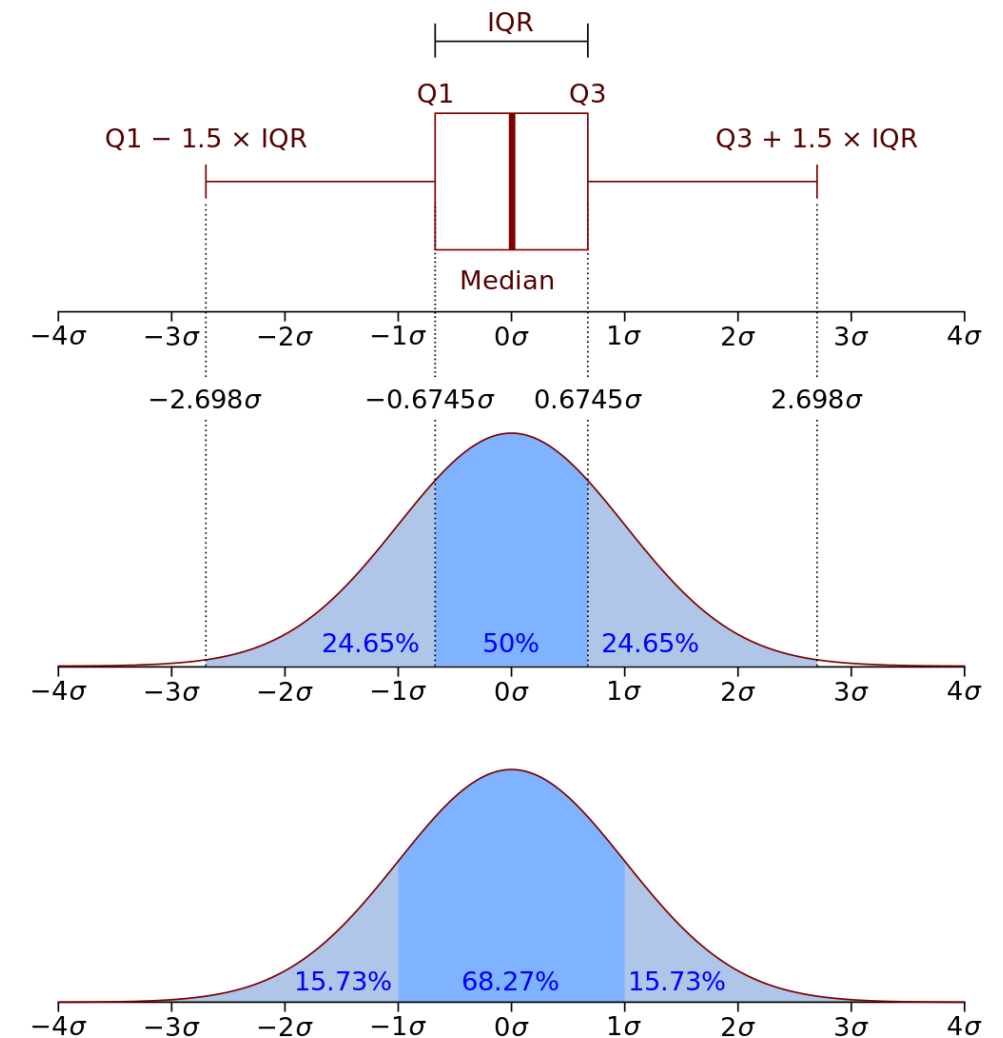


# ■ Outliers create a viewing problem for other data

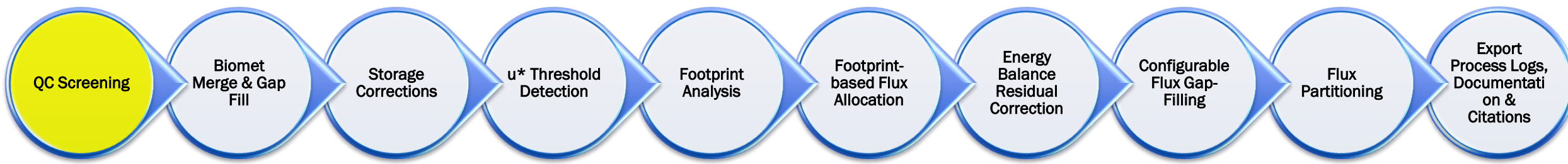




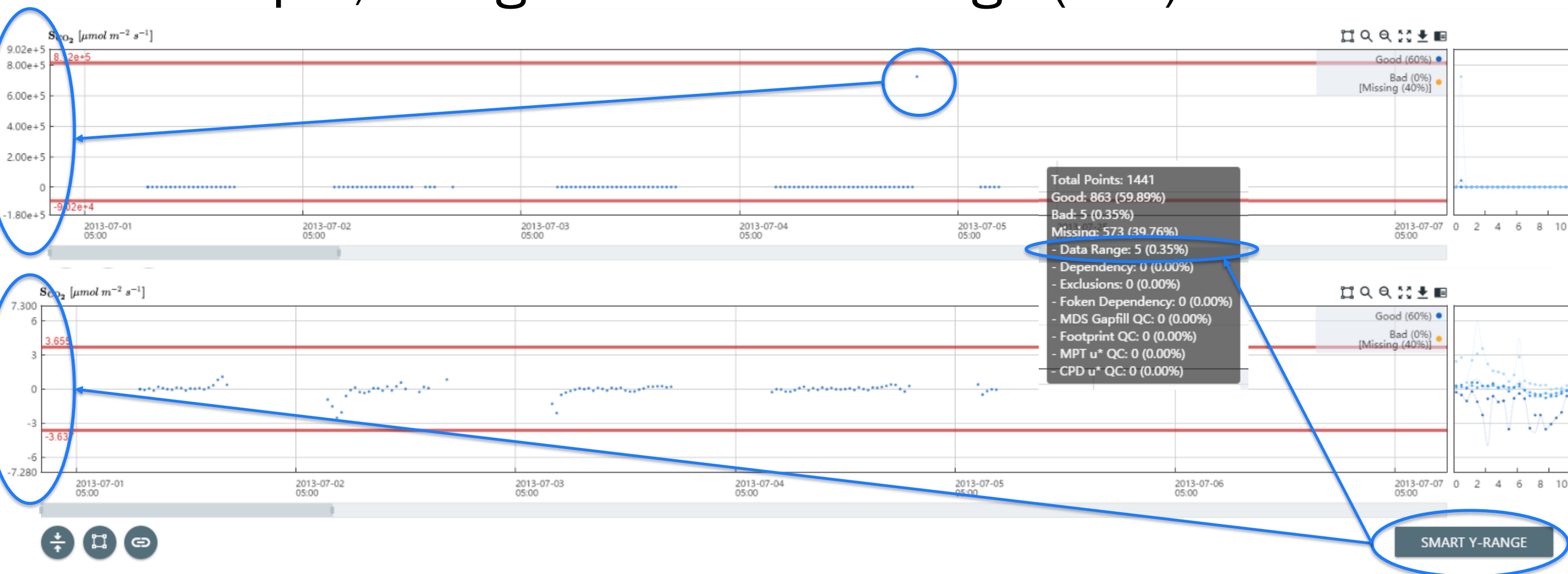
- ‘Smart Y-Range’ – filter values 10 times the Interquartile Range (IQR)
  - The Smart Y-Range is a way to bring both the zoom level and the minimum and maximum lines down to where the bulk of the data is found.
  - The Smart Y-range is computed statistically as the median  $\pm 10$  times the interquartile range (IQR) computed on the entire dataset.
  - Click the [SMART Y-RANGE] button, and then you can adjust minimum and maximum lines.



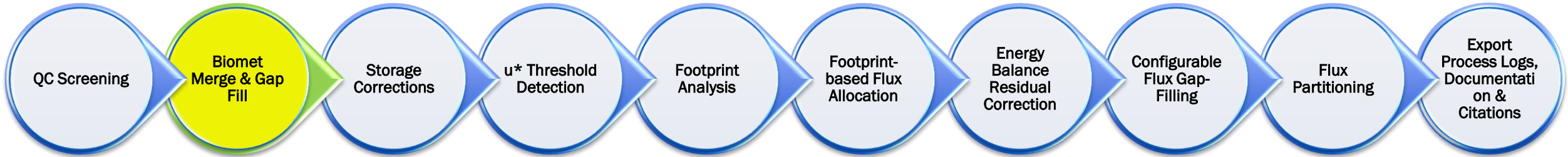
By Jhguch at en.wikipedia, CC BY-SA 2.5,  
<https://commons.wikimedia.org/w/index.php?curid=14524285>

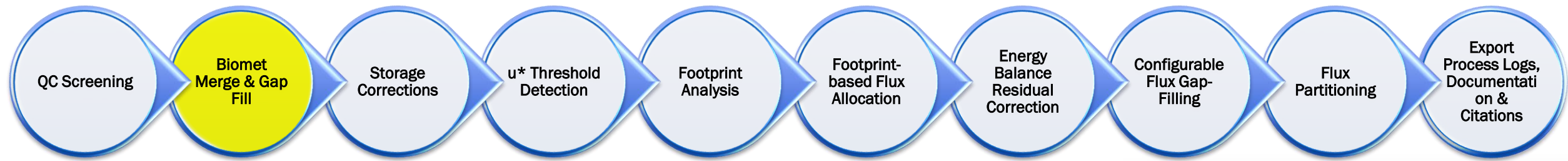


# ■ Example; using the Smart-Y Range (IGR) in Tovi



# Typical workflow for Post Processing...





- Gap-filled time-series are required if site data are to be used by modelers for parameterization and validation of land surface models and to construct annual sums of carbon and water exchange between the land surface and the atmosphere.
- Before the gaps in the fluxes are filled, gaps need to be filled in the environmental measurements that are used to drive gap filling of the fluxes.

Biogeosciences, 14, 2903–2928, 2017  
<https://doi.org/10.5194/bg-14-2903-2017>  
 © Author(s) 2017. This work is distributed under the Creative Commons Attribution 3.0 License.



### OzFlux data: network integration from collection to curation

Peter Isaac<sup>1</sup>, James Cleverly<sup>2</sup>, Ian McHugh<sup>3</sup>, Eva van Gorsel<sup>4</sup>, Cacia Ewenz<sup>5</sup>, and Jason Beringer<sup>6</sup>

<sup>1</sup>OzFlux Central Node, TERN-OzFlux, Melbourne, VIC 3159, Australia

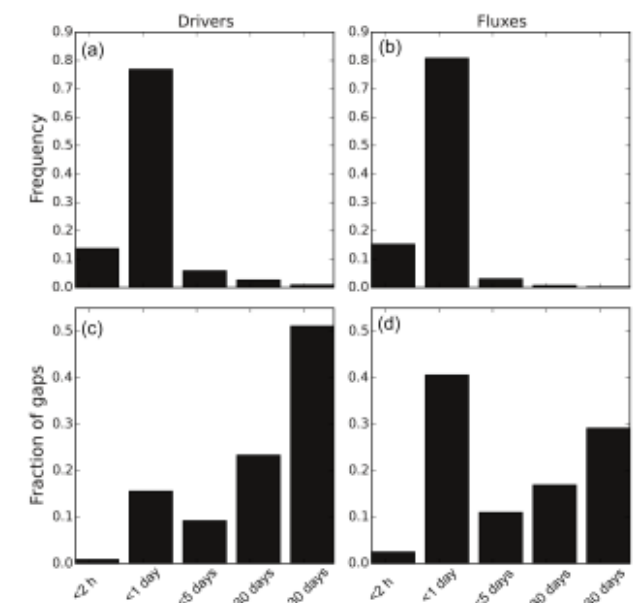
<sup>2</sup>School of Life Sciences, University of Technology, Sydney, NSW 2007, Australia

<sup>3</sup>School of Earth, Atmosphere and Environment, Monash University, Melbourne, VIC 3800, Australia

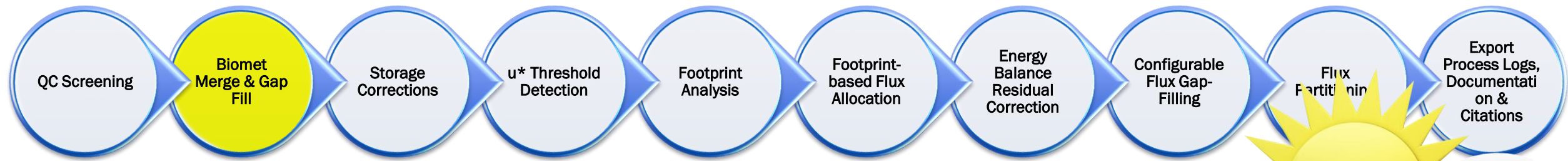
<sup>4</sup>CSIRO Oceans & Atmosphere Flagship, Yarralumla, ACT 2600, Australia

<sup>5</sup>Airborne Research Australia, Adelaide, SA 5106, Australia

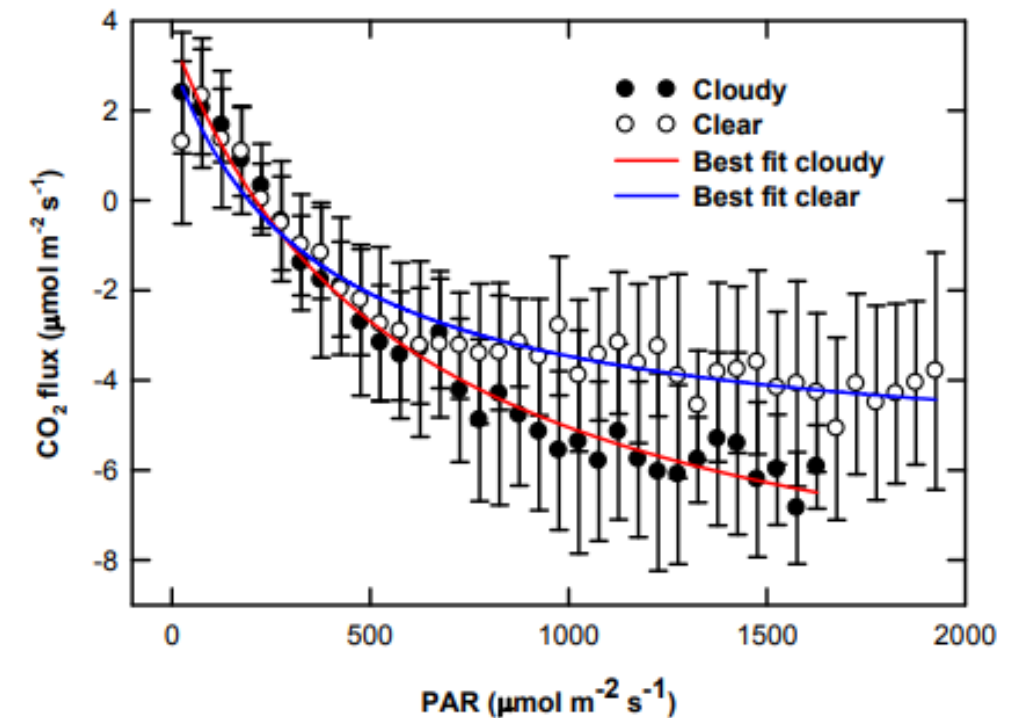
<sup>6</sup>School of Earth and Environment, University of Western Australia, Crawley, WA 6009, Australia

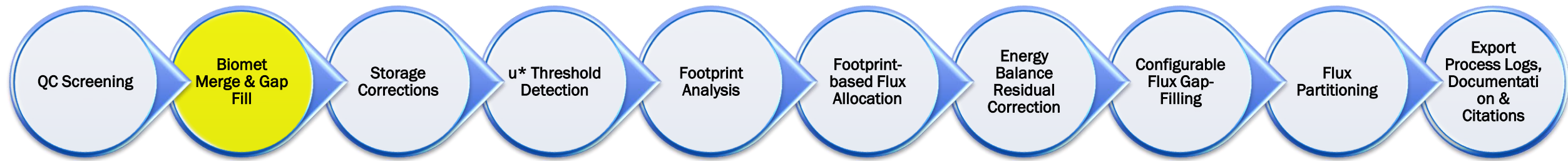


**Figure 3.** Histograms of gap frequency by duration for (a) drivers and (b) fluxes and histograms of the fraction of total gap length for (c) drivers and (d) fluxes.



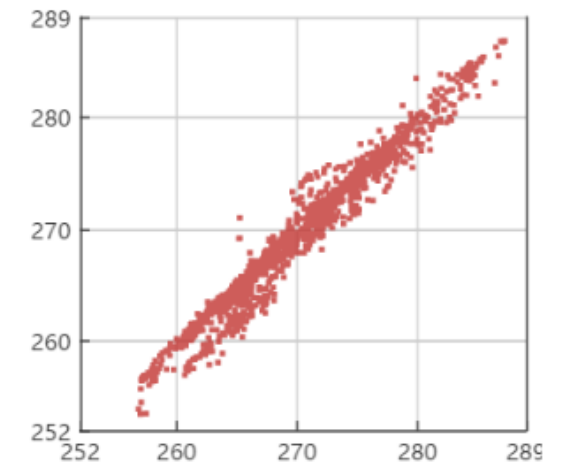
- Biological and Meteorological (Biomet) data are important for:
  - Quality Assurance and Quality Control (QA/QC) - Energy Balance closure.
  - To help ‘explain’ flux results - The physical environment has profound effects on the biology as well as on the surface-atmosphere exchange.
  - Gap Filling of surface fluxes - A complete record of flux drivers is essential for successful gap filling of surface fluxes



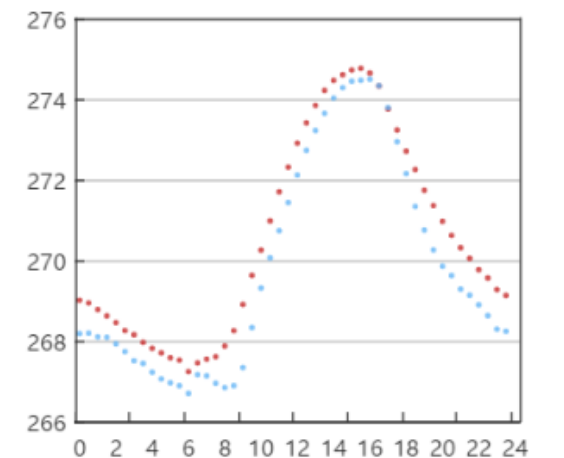


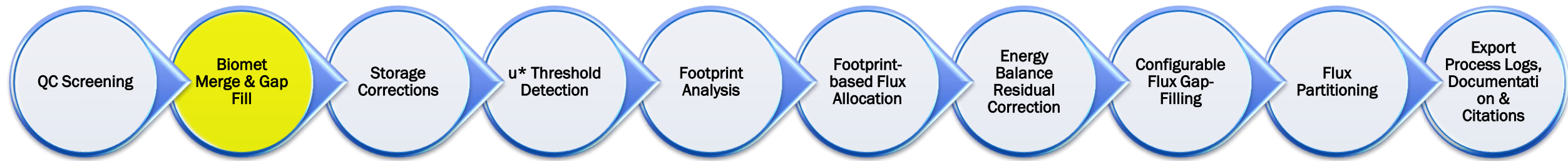
- As a substitute for measured Biomet data, data from three alternative sources can be used to fill time-series of radiation, meteorological and soil data from flux towers:
  - (1) Use observation measurements from redundant sensors at the Site
  - (2) Use data observed from a nearby weather station (example; NOAA-ISD station)
  - (3) Use data modeled from a gridded source (example; ERA-I grid data)
- External Biomet data should overlap with existing Site data
- Gap filling can be done on a Monthly basis or for the entire period
- Gap filling can be done by performing an OLS regression between site data and imported data, or simply replace the gaps with imported data

Correlation

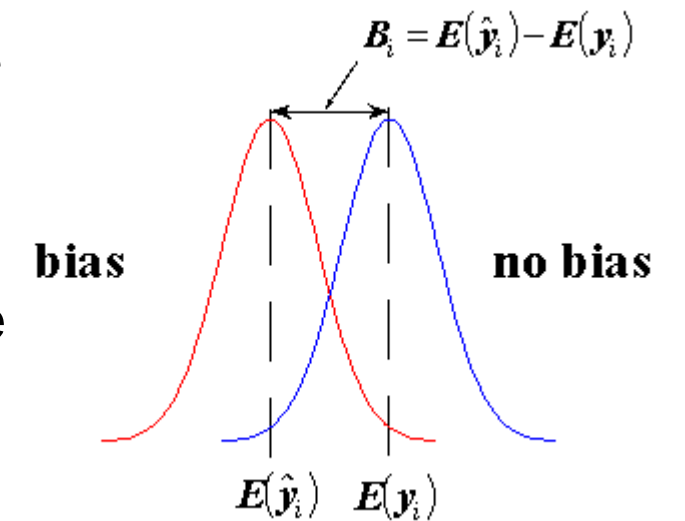


Diurnal Primary Secondary

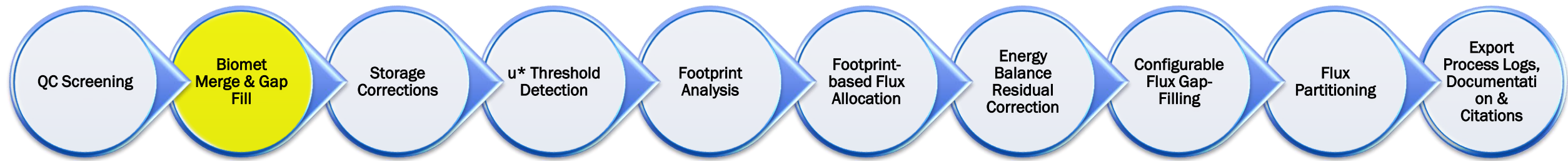




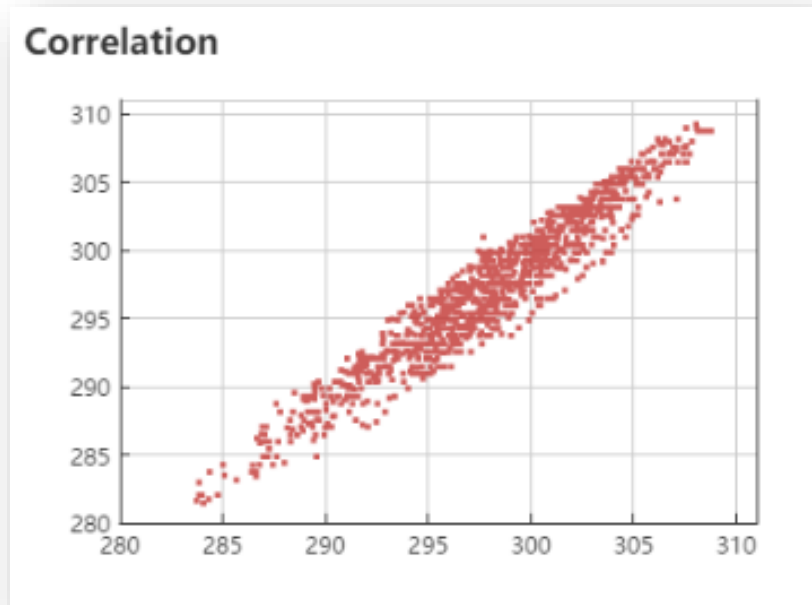
- During gap filling of the drivers, compute and examine the statistics (R, bias, root mean square error (RMSE), ratio of the variances) that describe the fit between data measured at the flux tower and from the alternate sources
  - Bias can be frequently observed between the data from flux towers and the alternative sources and this must be removed before the alternate data can be used for gap filling the flux tower data.
  - Bias correction should be done for each 90-day period using a linear fit calculated for each variable provided based on a minimum percentage of good data available (overlapping) within the window (default 50 %).
- For gaps that do not satisfy the minimum percentage of good data threshold, the window size should be increased by moving the start and end dates of the window period by one day-at-a-time, until the minimum percentage of good data requirement is met.
- This technique ensures that the window length is kept as short as possible while still providing sufficient good data for the linear fit.







- Example; filling gaps in Biomet data using Tovi
  - Choose the data which gives the best regression



Primary Variable: TA | Time Window (months): 1 | Min Overlap Required (%): 50

Coverage: 80.12%  
Filled: 0.00%

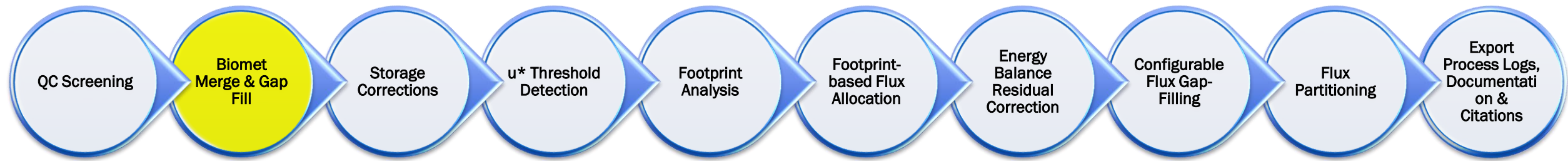
Apply to all Periods  
 Auto Complete

Merged Variables (order of priority)

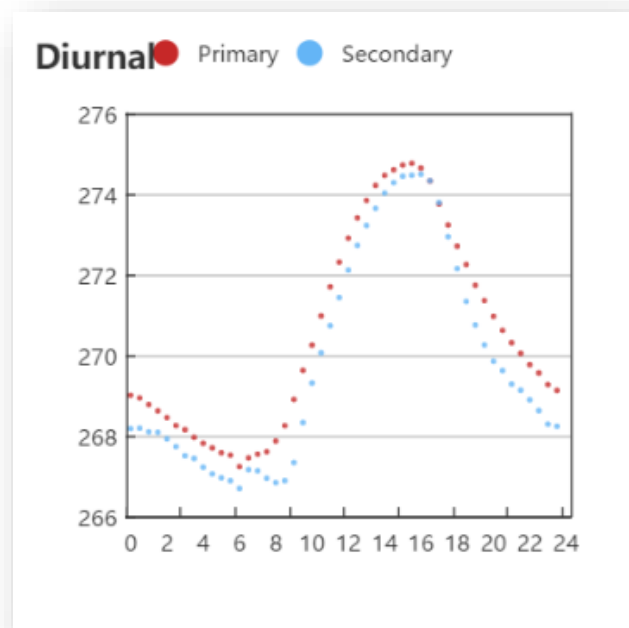
Variable	Type	(Lat, Lon)	Units	r2	Slope	Intercept	Overlap (%)	Fit Type	Lag
No data available									
Unused Variables (available)									
Variable	Type	(Lat, Lon)	Units	r2	Slope	Intercept	Overlap (%)	Fit Type	View (1,2,3)
TA_1_1_1	biomet	40.856, -96.654	K	1.000	1.00	0.00	80.12	OLS	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
TA	ISD	40.851, -96.747	K	0.926	0.87	40.01	80.12	OLS	<input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
TA	ERAI	41.250, -96.750	K	0.767	0.82	52.17	80.12	OLS	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3

Time Window: Jan-2013 to Jan-2014

T<sub>air</sub> [K] vs Time



- Example; filling gaps in Biomet data using Tovi
  - Check the ‘Lag’ box to see if the regression improves



Primary Variable: TA | Time Window (months): 1 | Min Overlap Required (%): 50

← | Jan-2013 | Feb-2013 | Mar-2013 | Apr-2013 | May-2013 | Jun-2013 | Jul-2013 | Aug-2013 | Sep-2013 | Oct-2013 | Nov-2013 | Dec-2013 | Jan-2014 | →

Coverage: 80.12%  
Filled: 19.88%

Apply to all Periods  
 Auto Complete

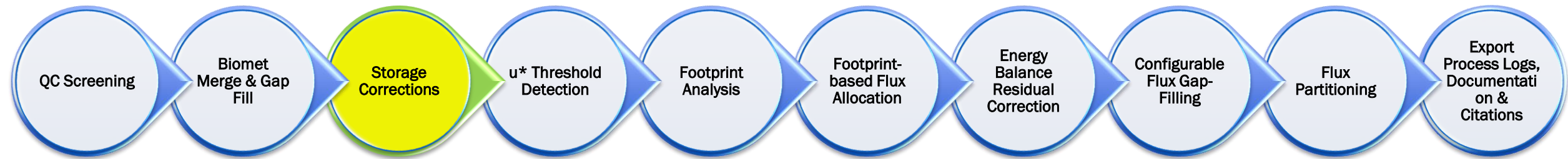
Variable	Type	(Lat, Lon)	Units	r2	Slope	Intercept	Overlap (%)	Fit Type	Lag
TA	ISD	40.851, -96.747	K	0.926	0.87	40.01	80.12	OLS	<input type="checkbox"/>

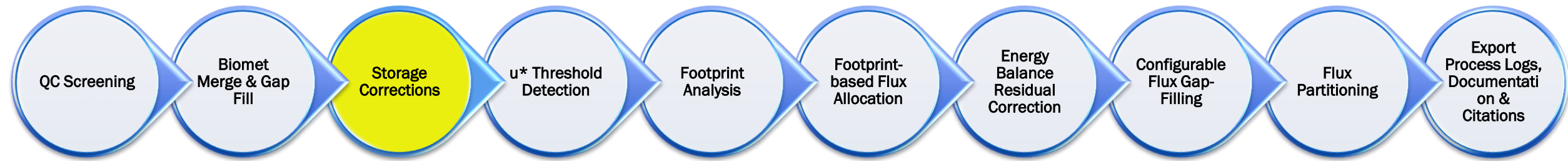
Unused Variables (available)

Variable	Type	(Lat, Lon)	Units	r2	Slope	Intercept	Overlap (%)	Fit Type	View (1,2,3)
TA_1_1_1	biomet	40.856, -96.654	K	1.000	1.00	0.00	80.12	OLS	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
TA	ERA1	41.250, -96.750	K	0.767	0.82	52.17	80.12	OLS	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3

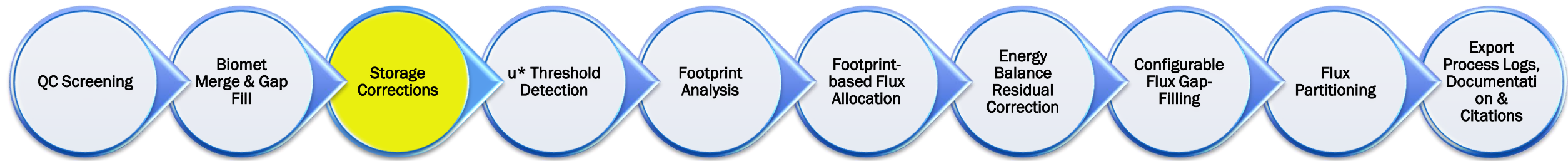
**T<sub>air</sub> [K]**

# Typical workflow for Post Processing...

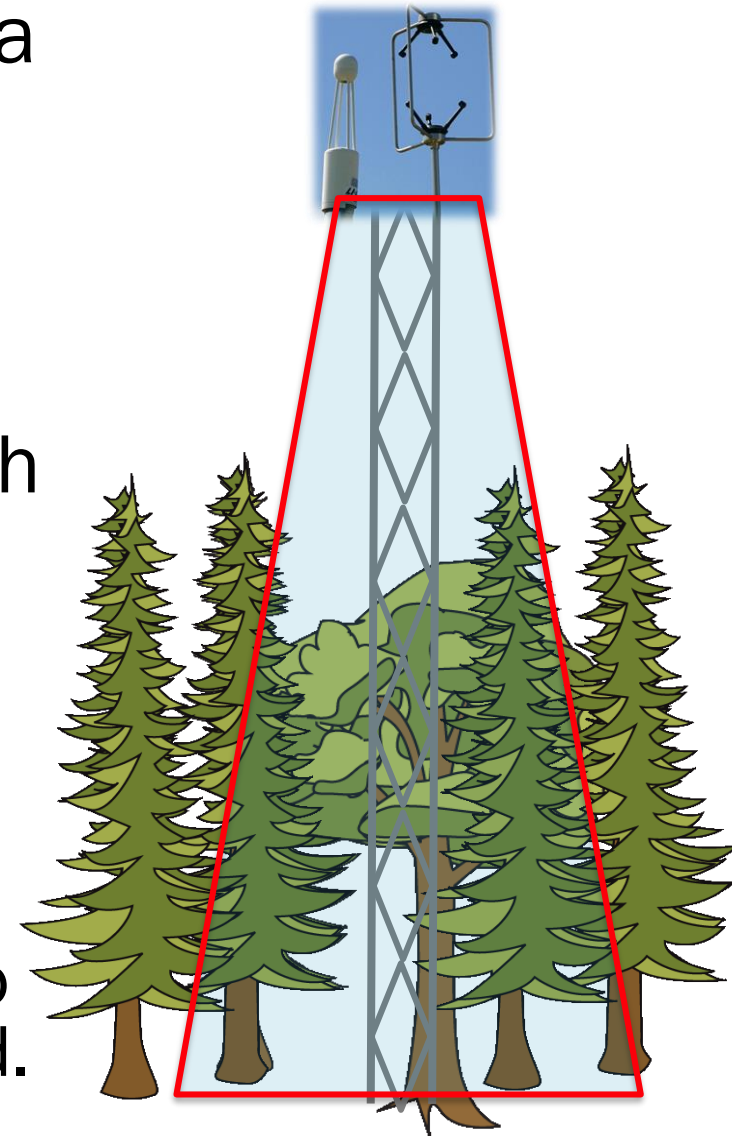


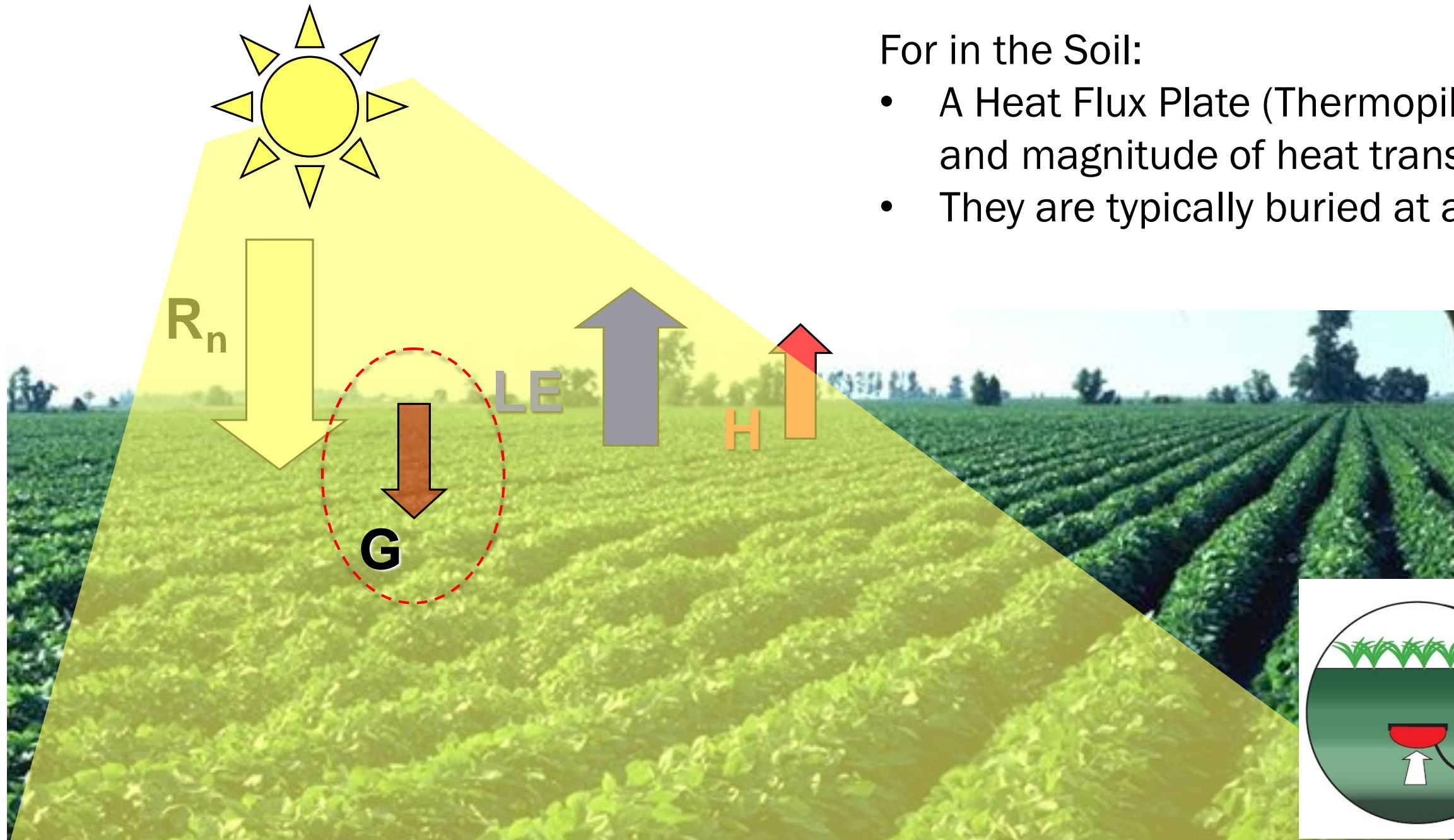
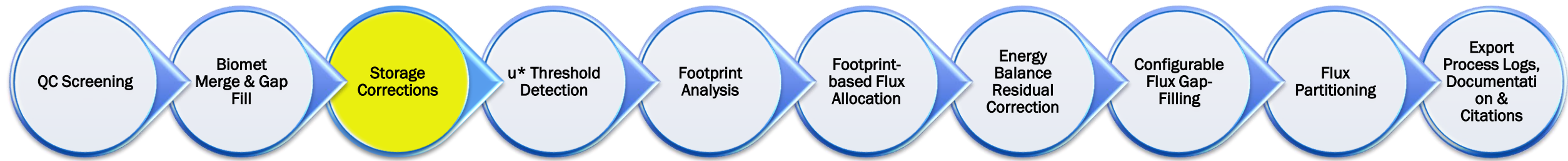


- Storage Corrections are done by computing or measuring storage terms and ‘correcting’ the measured values by adding this storage term.
- Two types of Storage Corrections should be applied for more accurate values for these ecosystem measurements:
  - One is to measure or compute the storage of gases and/or energy between the soil surface and the instrument (eddy covariance system) height. This term is then added to the fluxes to compute Net Ecosystem Exchange (NEE)
  - Two is to measure or compute the storage of energy between the soil surface the instrument (soil heat flux sensor) depth. This term is then added to the Soil Heat Flux measurement to compute the Soil Heat Flux at the soil’s surface.



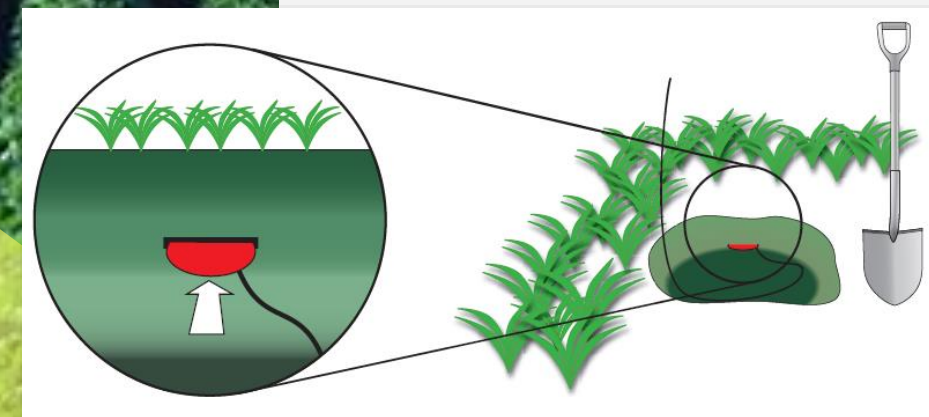
- For in the air: Eddy covariance instruments record flux at a specific measurement height. Below this height, gas can build up or become depleted, especially during calm periods, or within a tall canopy ( $\text{CO}_2$  build up on a calm night or  $\text{CO}_2$  depletion during a calm day, for examples).
- This  $\text{CO}_2$  Storage term ( $\text{SCO}_2$ ) can either be measured with a profile system or computed as part of the raw data processing scheme (for example, in EddyPro)
- This  $\text{SCO}_2$  term can be applied to the  $\text{CO}_2$  Flux ( $\text{FCO}_2$ ) to compute the Net Ecosystem Exchange (NEE).
  - In its simplest form (when profile measurements are not available), the NEE can be computed by simply adding the two components ( $\text{FCO}_2$  and  $\text{SCO}_2$ ) together for the half-hour period.

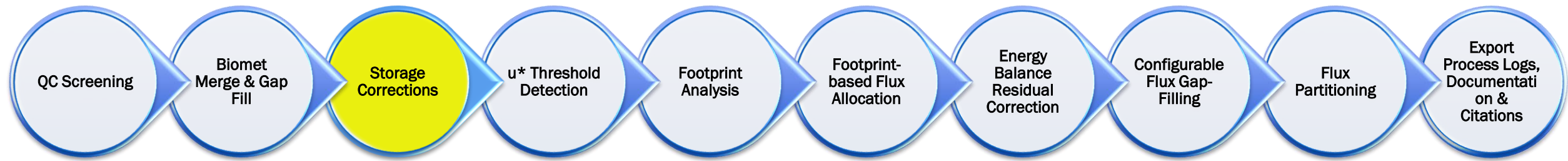




For in the Soil:

- A Heat Flux Plate (Thermopile) measures direction and magnitude of heat transfer ( $G$ ) through soil.
- They are typically buried at a depth of 5 cm.





- The heat flux plate measures  $G$  only at the depth of placement.
- Need to account for the change in heat storage in the soil above the plate (Fuchs & Tanner, 1968).
- Failure to account and correct for heat storage in the soil above the plate can result in large errors.
- Typically the heat storage correction results in significantly greater  $G$  during the daytime and lower  $G$  at night.
- This error would lead to a subsequent underestimate of  $LE + H$  by the same amount.
- For computing Soil Heat Flux Storage, the sources include  $G, T_{SOIL}$ , and  $SWC$ .

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Publications from USDA-ARS / UNL Faculty

U.S. Department of Agriculture: Agricultural Research Service, Lincoln, Nebraska

2005

### Soil Heat Flux

Thomas J. Sauer  
USDA-ARS-NSTL

Robert Horton  
Iowa State University

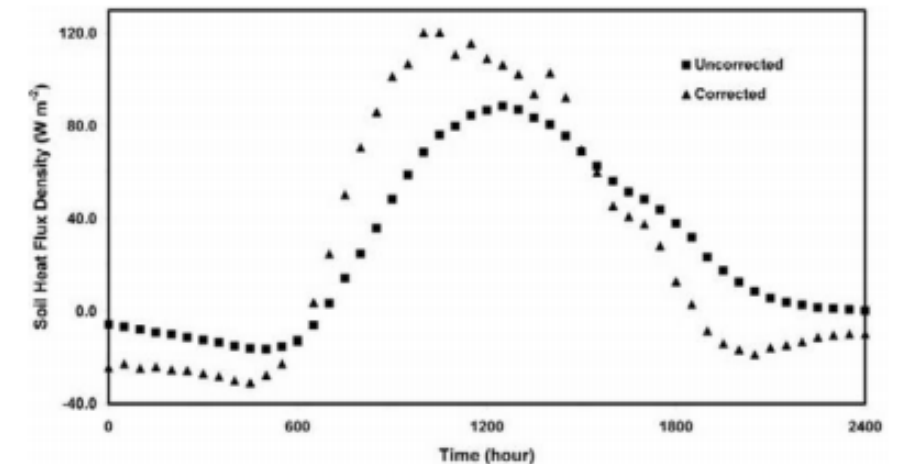


Fig. 7-4. Heat flux density measured at 0.05 m in a silt loam soil in southwestern Minnesota with and without the correction for heat storage in the soil layer above the heat flux plate (T.J. Sauer and N.S. Eash, unpublished data).

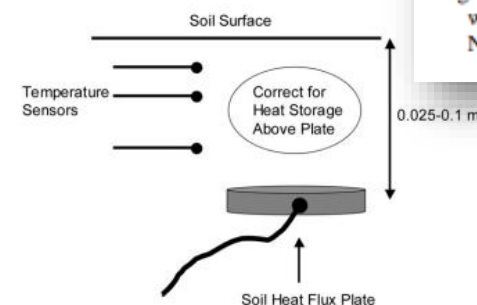
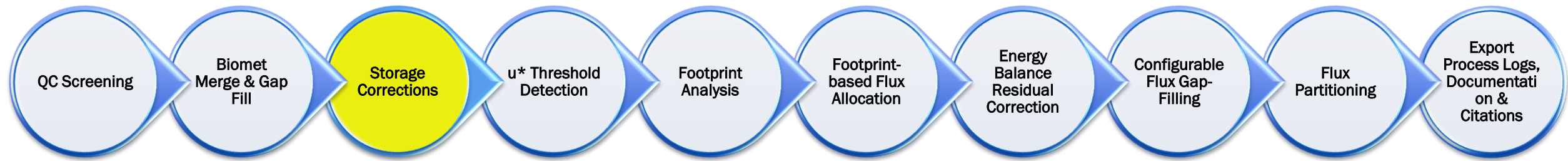


Fig. 7-3. Arrangement of sensors and measurements necessary for application of the heat flux plate method to determine soil heat flux at one location in a homogenous soil.



Heat stored above heat flux plates can be estimated using information on soil properties, temperature, and moisture.

**S** = Storage Term ( $\text{W m}^{-2}$ )

$$S = \frac{\Delta T_s C_s d}{t}$$

$\Delta T_s$  = Change in soil temperature over flux averaging interval  $t$

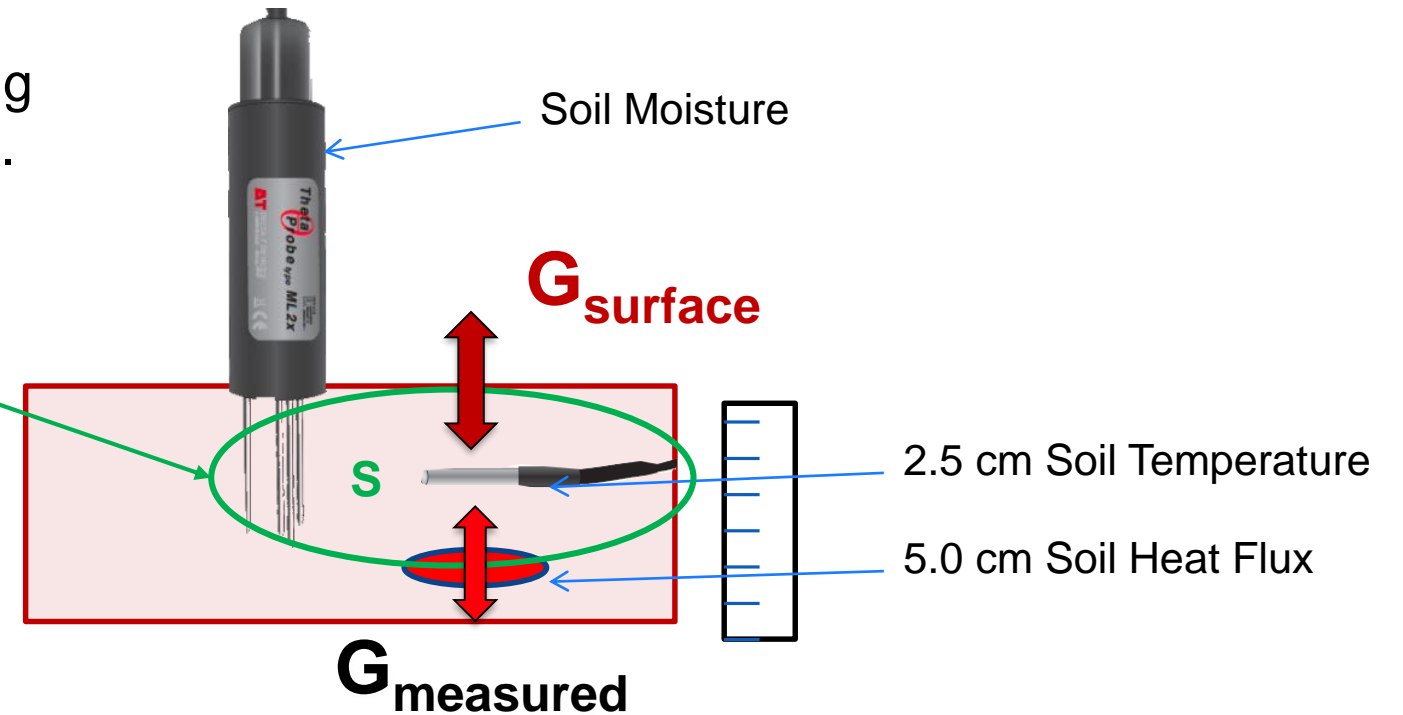
$d$  = depth of the Soil Heat Flux plate

$C_s$  = Heat capacity of moist soil

$$C_s = \rho_b C_d + \theta_v \rho_w C_w$$

$\rho_b$  = soil bulk density ( $\text{Kg/m}^3$ ),  $C_d$  = heat capacity of dry mineral soil ( $\text{J Kg}^{-1} \text{K}^{-1}$ )

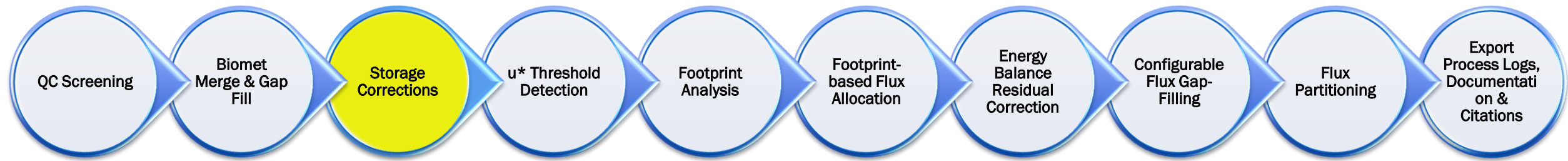
$\theta_v$  = volumetric water content ( $\text{m}^3 \text{m}^{-3}$ ),  $C_w$  = heat capacity of water



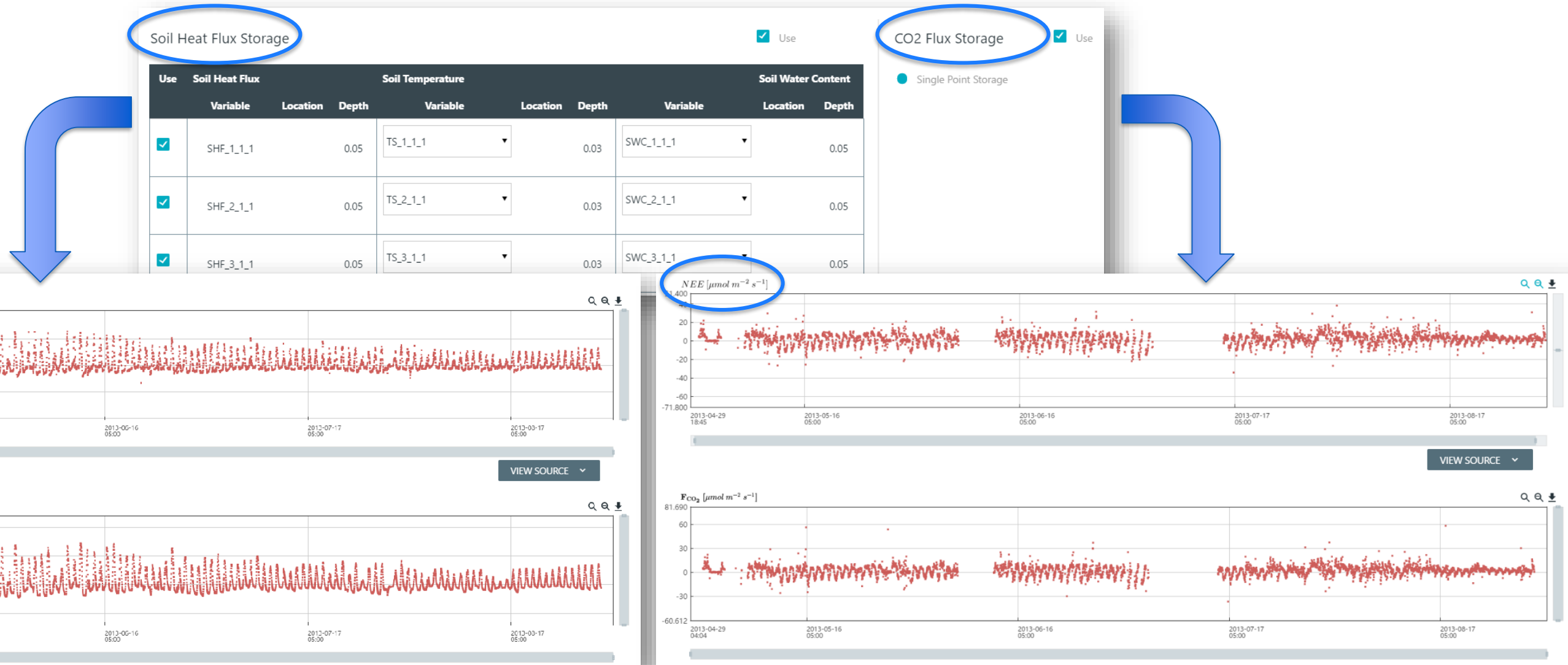
Note: For grasslands, agricultural and chaparral sites, corrected soil heat flux has a larger impact than at the forested sites.

**Heat Flux at the Soil Surface = Measured Soil Heat Flux + Storage Term (S)**

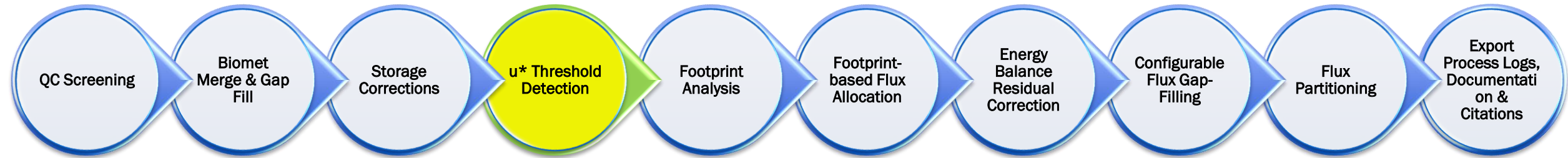


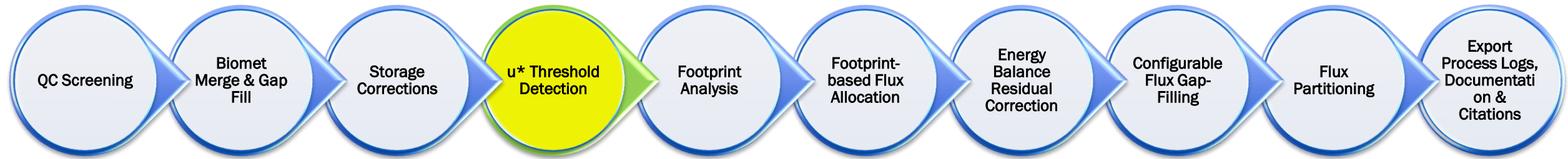


- Example; correcting the Soil Heat Flux and computing NEE using Tovi

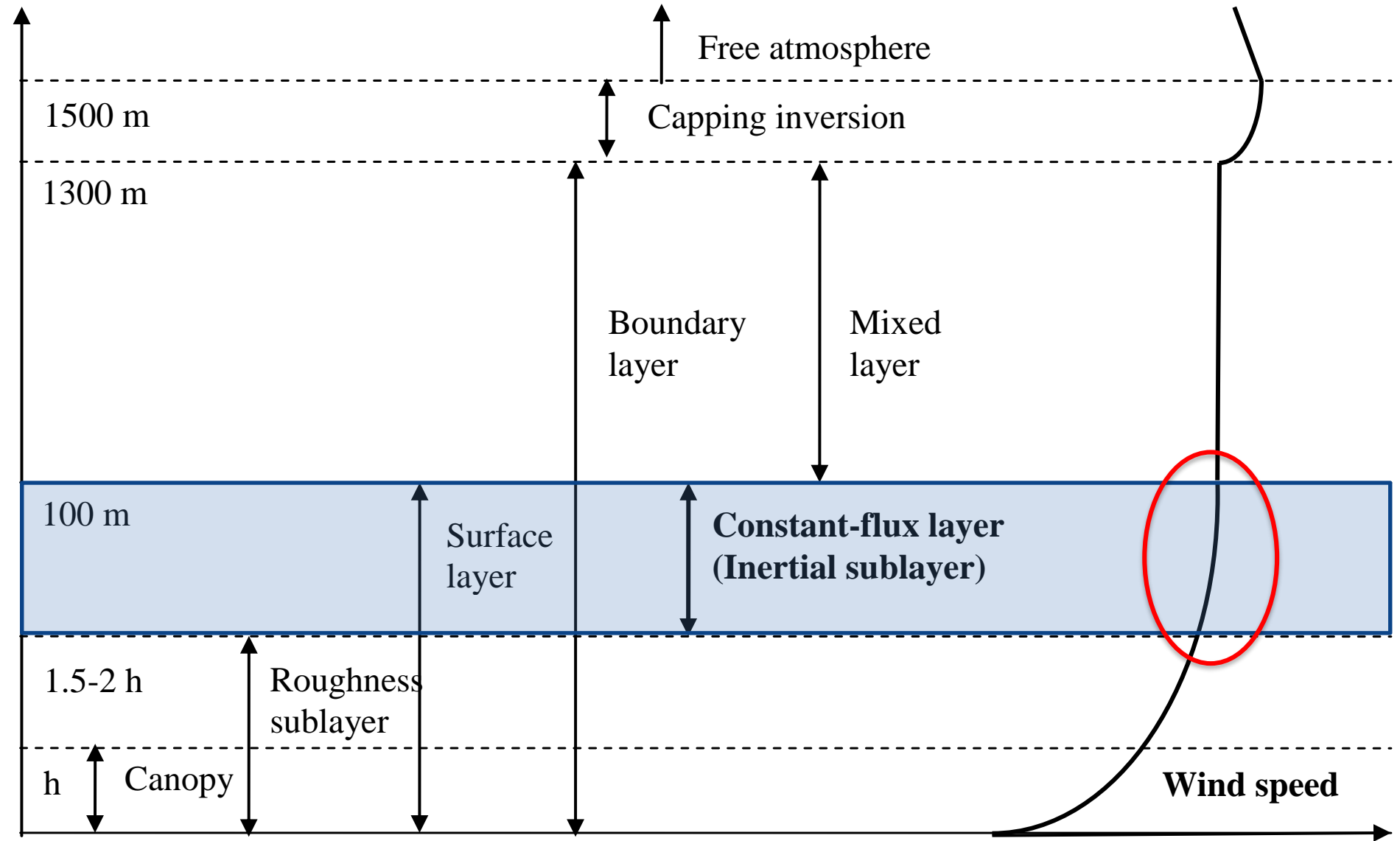


# Typical workflow for Post Processing...



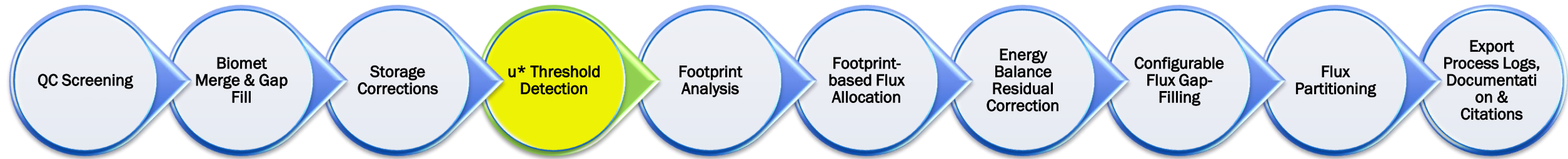


# Atmospheric Boundary Layer



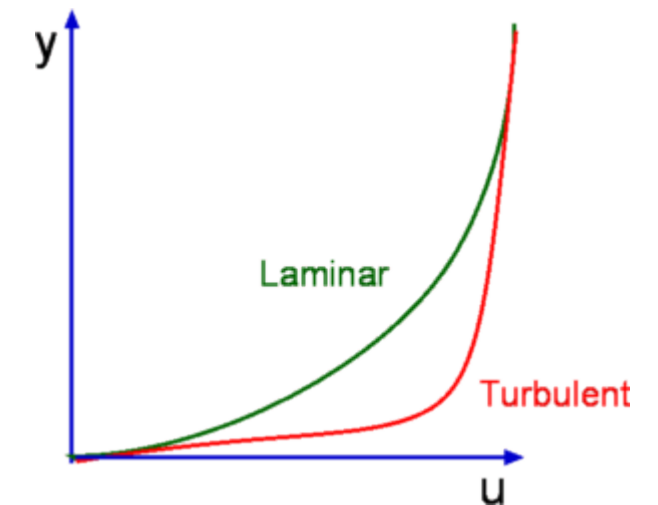
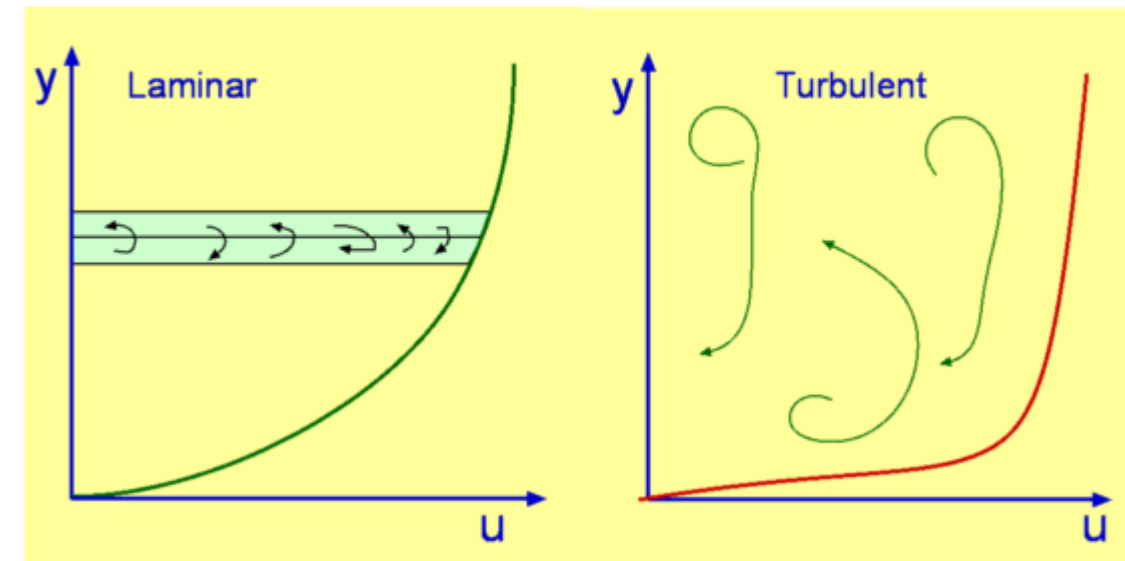
Drawn based on Stull (1988), Denmead *et al.* (1996) and Oke (2007)

Not to scale.

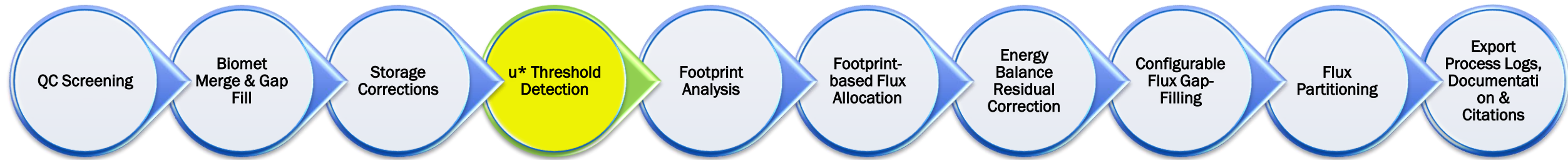


## ■ Friction velocity ( $u^*$ )

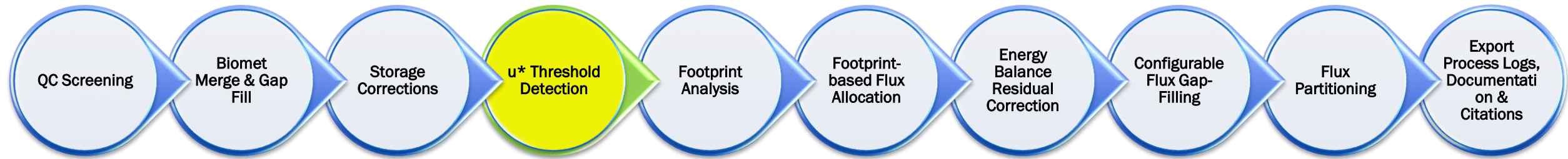
- Friction velocity characterizes the shear at the surface
- The friction velocity is often used as a scaling parameter for the fluctuating component of velocity in turbulent flows.
- For turbulent flow, it is near constant closer to the surface and in the atmosphere



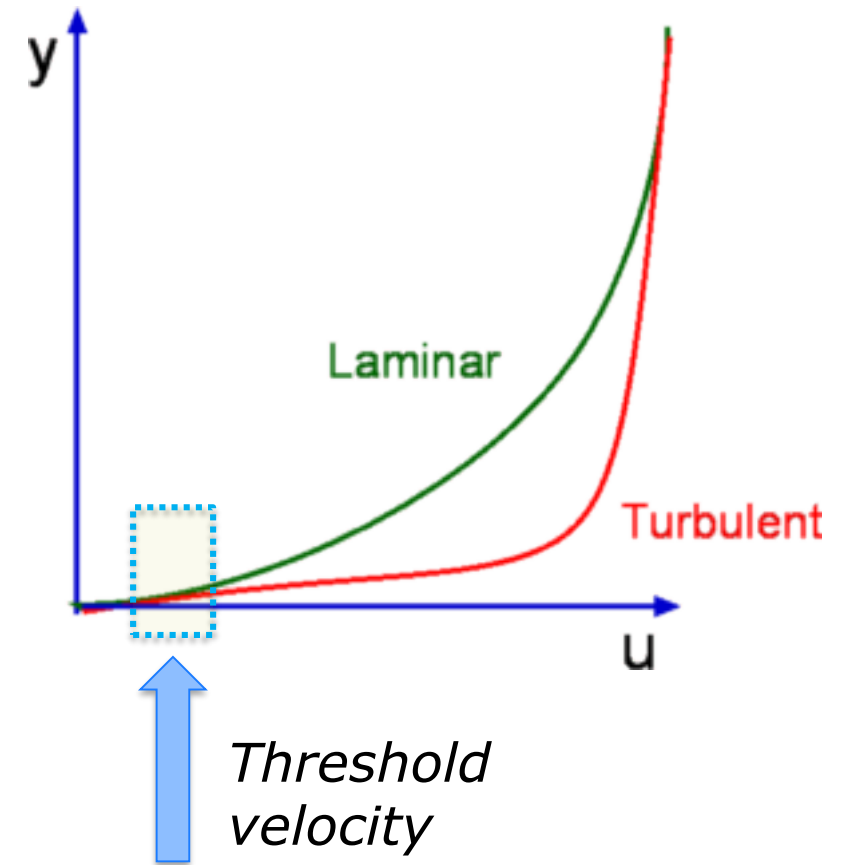
(c) Aerospace, Mechanical & Mechatronic Engg. 2005, University of Sydney

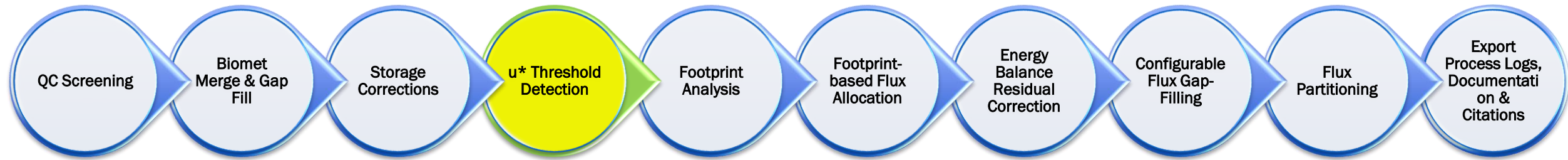


- Can friction velocity ( $u^*$ ) be used to help filter data?
  - EC requirements assume turbulent flow
  - The Foken tests do not capture stable stratification (night-time) well,
  - As a result, during these conditions, these lower-quality measurements are not flagged
  - These incorrect (night-time) measurements bias flux results, leading to an underestimation of fluxes.
  - In addition, advection terms gain importance and are no longer negligible

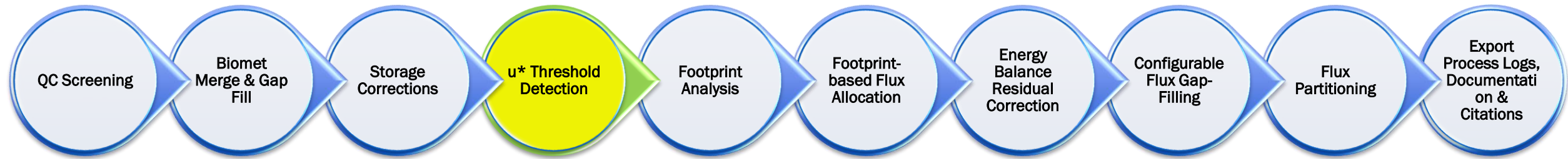


- Is there an alternative, low-turbulence condition test that could be used for night-time flux data (respiration)?
  - Need to compute the ‘**threshold**’ velocity to transition between laminar and turbulent flows (‘bad’ and ‘good’ data).
- Detecting the ‘threshold’ value for the friction velocity ( $u^*$ )
  - Can we measure the sensitivity of the flux measurements, (in particular, the night-time respiration) to the friction velocity?



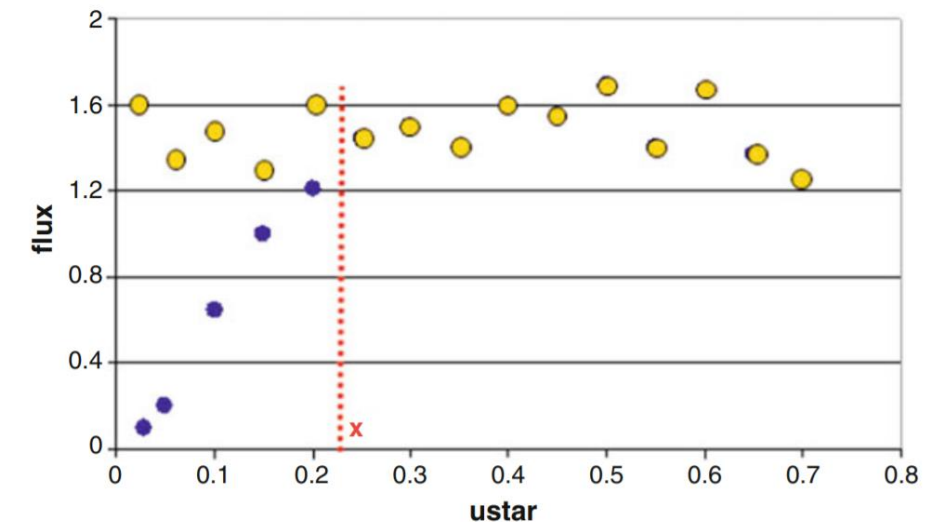


- Determining the  $u^*$  (friction velocity) threshold
- Several factors will effect the value of the  $u^*$  threshold. It depends on;
  - Local average meteorological conditions (frequency of occurrence of nonturbulent periods)
  - Site topography
  - Land cover heterogeneity
  - Soil and plant biology (importance of respiration)
  - Canopy architecture (vegetation height, canopy density).
- One suggestion has been to use a static value of  $0.2 \text{ m s}^{-1}$  and apply this across the data set. This value has been determined as a sort of ‘average’ based on finding at several ecosystems, however, because of the above influences, this is not a good recommendation



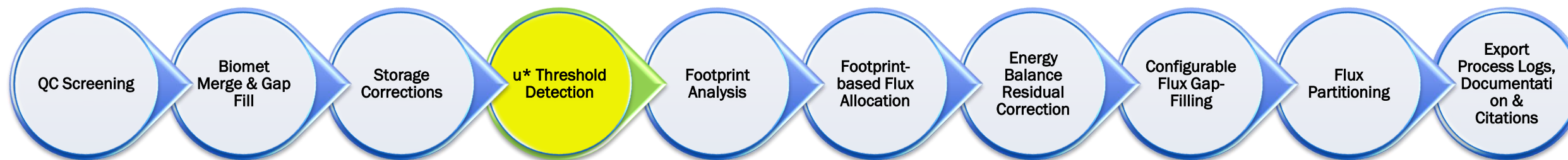
- A better way = detect this from the NEE versus  $u^*$  relationship
- Primarily two methods developed by the community to detect the friction velocity  $u^*$  threshold for your site:
  - Moving Point Test (MPT) methodology - Papale et al. (2006)
  - Change Point Detection (CPD) methodology - Barr et al. (2013)
- Both methods work on data where solar radiation is below a threshold (default:  $R_g < 10 \text{ W m}^{-2}$ )
- They assume that photosynthesis is near zero
- Both ecosystem respiration and  $u^*$  exhibit diurnal and seasonal cycles
- These must be removed before  $u^*$  can be used as a filter criterion.
- Data is subset to similar environmental conditions (adjacent times (seasons) and temperature classes)
- Within each season/temperature subclass the  $u^*$  threshold is estimated at which NEE saturates.
- The  $u^*$  threshold (subclass) estimates are aggregated to a  $u^*$  threshold for each year.

Researchers often find the  $u^*$  threshold by visually examining the scatter plot of nighttime fluxes versus  $u^*$ . Conceptually, it is assumed that the threshold is located where the flux begins to level off as  $u^*$  increases.



**Fig. 5.4**  $u^*$  threshold selection (theoretical and optimal situation): *blue dots* are normalized nighttime fluxes (or nighttime fluxes acquired in a narrow temperature range) storage corrected, *yellow dots* are the theoretical pattern of the fluxes (independent by  $u^*$ ), *red x* is the  $u^*$  threshold





## Moving Point Test (MPT):

- MPT uses an iterative approach to simultaneously determine a valid temperature response function, which is used to normalize nighttime flux measurements, and identify  $u^*$  thresholds based on the normalized fluxes.
  - Night-time data was selected according to a global radiation threshold of  $20 \text{ W m}^{-2}$  and defined as Reco.
  - Data is split into six temperature subclasses
  - Data is then binned into 20  $u^*$  subclasses
- The threshold is defined as the  $u^*$ -class where the night-time flux reaches more than 95% of the average flux at the higher  $u$ -classes.
- The threshold is only accepted if for the temperature class, temperature and  $u^*$  are not or only weakly correlated
- The final threshold is defined as the median of the thresholds of the six temperature classes.
- This procedure is applied to the subsets of four 3-month periods to account for seasonal variation of vegetation structure.
- For each period, the  $u$ -threshold is reported, but the whole data set is filtered according to the highest threshold found (conservative approach).
- In cases where no  $u^*$ -threshold could be found, it is set to  $0.4 \text{ m s}^{-1}$ . A minimum threshold is set to  $0.1 \text{ m s}^{-1}$ .

### Objective threshold determination for nighttime eddy flux filtering

Lianhong Gu<sup>a,\*</sup>, Eva M. Falge<sup>b</sup>, Tom Boden<sup>a</sup>, Dennis D. Baldocchi<sup>c</sup>,  
T.A. Black<sup>d</sup>, Scott R. Saleska<sup>e</sup>, Tanja Suni<sup>f</sup>, Shashi B. Verma<sup>g</sup>,  
Timo Vesala<sup>f</sup>, Steve C. Wofsy<sup>e</sup>, Liukang Xu<sup>c</sup>

<sup>a</sup>Environmental Sciences Division, Building 1509, Mail Stop 6335, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6335, USA

<sup>b</sup>Plant Ecology, Bayreuth University, Bayreuth, Germany

<sup>c</sup>Department of Environmental Sciences, Policy, and Management, University of California, Berkeley, USA

<sup>d</sup>Faculty of Agricultural Sciences, University of British Columbia, Vancouver, Canada

<sup>e</sup>Division of Engineering and Applied Science and Department of Earth and Planetary Sciences,

Harvard University, Cambridge, MA, USA

<sup>f</sup>Department of Physical Sciences, University of Helsinki, Helsinki, Finland

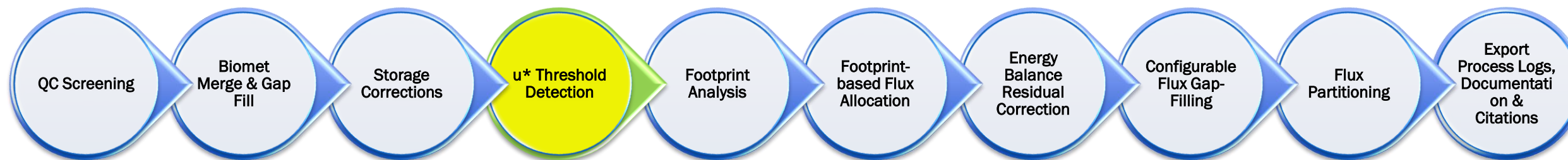
<sup>g</sup>School of Natural Resources, University of Nebraska, Lincoln, USA

Received 15 June 2004; accepted 10 November 2004

Global Change Biology (2005) 11, 1424–1439, doi: 10.1111/j.1365-2486.2005.001002.x

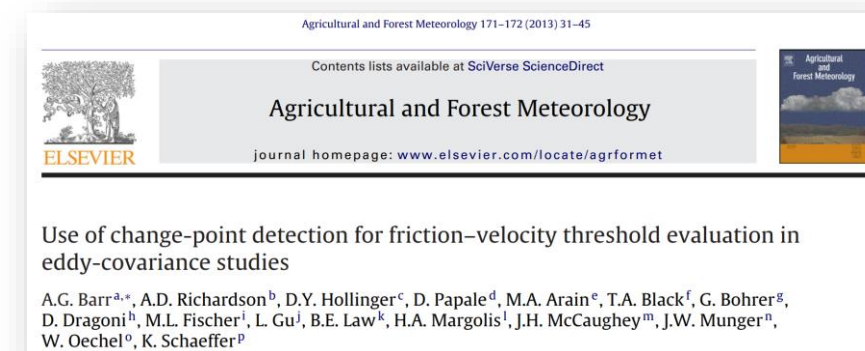
### On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm

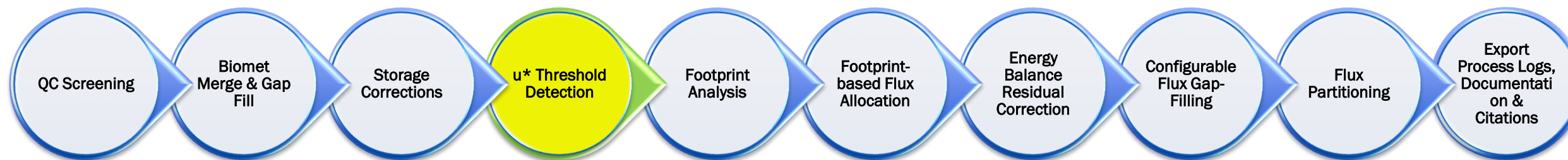
MARKUS REICHSTEIN<sup>††</sup>, EVA FALGE<sup>‡</sup>, DENNIS BALDOCCHI<sup>§</sup>, DARIO PAPALE<sup>\*</sup>,  
MARC AUBINET<sup>¶</sup>, PAUL BERBIGIER<sup>||</sup>, CHRISTIAN BERNHOFER<sup>\*\*</sup>,  
NINA BUCHMANN<sup>†††</sup>, TAGIR GILMANOV<sup>§§</sup>, ANDRÉ GRANIER<sup>¶¶</sup>,  
THOMAS GRÜNWALD<sup>\*\*</sup>, KATKA HAVRÁNKOVÁ<sup>|||</sup>, HANNU ILVESNIEMI<sup>\*\*\*</sup>,  
DALIBOR JANOUS<sup>|||</sup>, ALEXANDER KNOHL<sup>††§</sup>, TUOMAS LAURILA<sup>†††</sup>,  
ANNALEA LOHILA<sup>†††</sup>, DENIS LOUSTAU<sup>†††</sup>, GIORGIO MATTEUCCI<sup>§§§</sup>,  
TILDEN MEYERS<sup>¶¶¶</sup>, FRANCO MIGLIETTA<sup>||||</sup>, JEAN-MARC OURCIVAL<sup>\*\*\*\*</sup>,  
JUKKA PUMPANEN<sup>††††</sup>, SERGE RAMBAL<sup>\*\*\*\*</sup>, EYAL ROTENBERG<sup>††††</sup>, MARIA SANZ<sup>§§§§</sup>,  
JOHN TENHUNEN<sup>¶¶¶¶</sup>, GÜNTHER SEUFERT<sup>§§§§</sup>, FRANCESCO VACCARI<sup>|||||</sup>,  
TIMO VESALA<sup>|||||</sup>, DAN YAKIR<sup>†††††</sup> and RICCARDO VALENTINI<sup>\*</sup>



## Change Point Detection (CPD)

- Provides an objective, robust procedure to determine if the relationship between  $u^*$  and NEE changes at some value of  $u^*$  and includes a test of statistical significance.
- The time series of nocturnal NEE measurements is divided into windows, or “seasons”, of 1000 points with successive windows overlapping by 500 points.
- Data for each window are divided into four equally populated temperature classes and within each temperature class the nocturnal NEE is averaged over 50  $u^*$  bins.
- Bins with fewer than five points (30-min averaging period) are removed from further analysis.
- The  $u^*$  threshold for each temperature class in each 1000-point “season” is then found by successive fitting of a two-stage linear regression to the binned NEE data while varying the center point of the linear fit from the lowest  $u^*$  bin to the highest.
- The  $u^*$  thresholds for all “seasons” in a year are then averaged to get annual threshold values after rejecting those “seasons” for which the sign of the regression slopes differed from the majority (dominant mode).
- The fraction of “seasons” rejected for this reason is typically less than 5%.





- Bootstrapping
- MPT - This procedure is repeated 100 times within a bootstrapping technique to assess the uncertainty of the  $u^*$  threshold detection
  - The 5% and 95% percentiles of the 100 bootstrapped threshold estimates are taken as confidence interval boundaries
  - The gap filling and partitioning is then applied using those different thresholds to propagate the uncertainty of  $u^*$  threshold estimation to NEE, GPP and Reco.
- CPD - Uncertainty in the  $u^*$  threshold value is estimated by repeating the CPD method 1000 times and randomizing the time order of the input data prior to each run (bootstrapping).
  - Bootstrapping yields a mean threshold value and 95% confidence limits that are used to estimate uncertainty in NEE due to uncertainty in the  $u^*$  threshold.
  - In practice, mean and confidence limits of the  $u^*$  threshold are insensitive to the number of bootstrapping iterations between 100 and 1000 runs

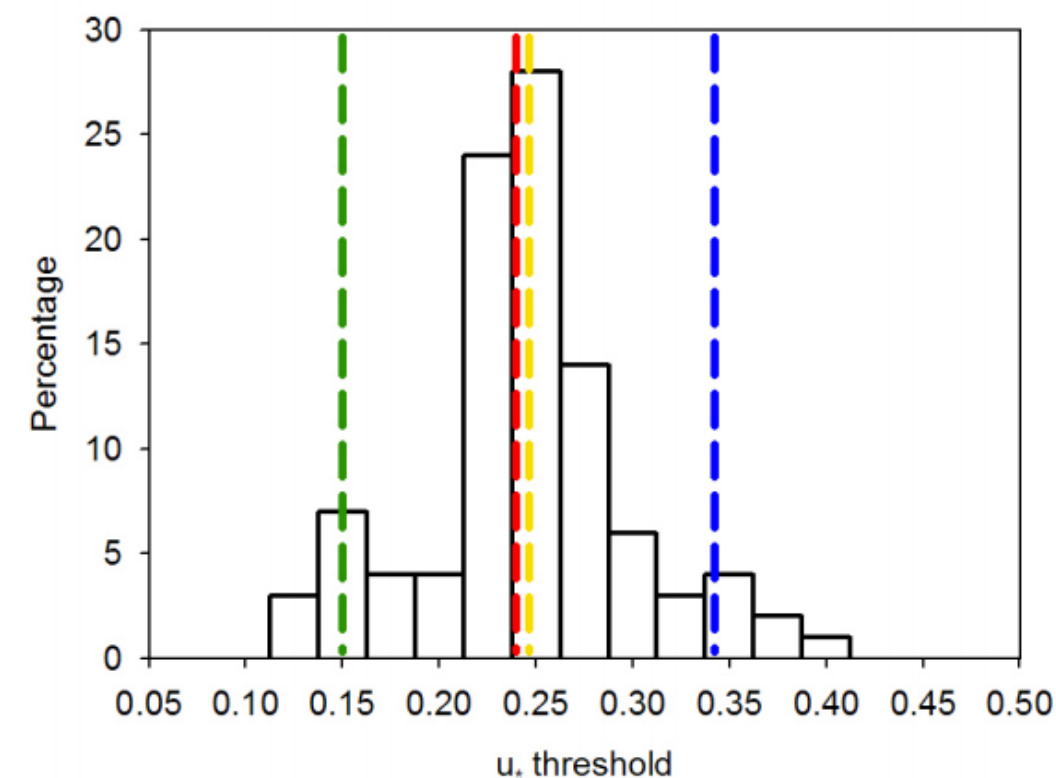
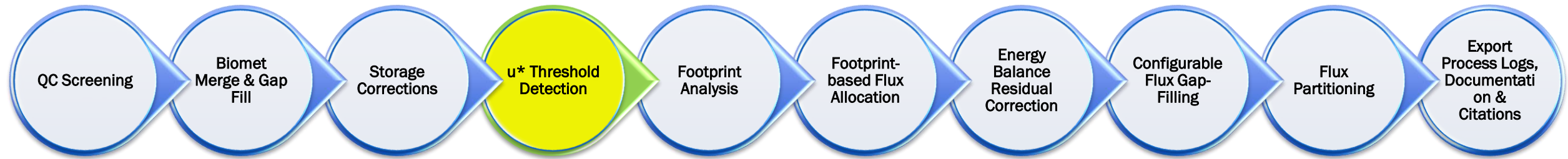
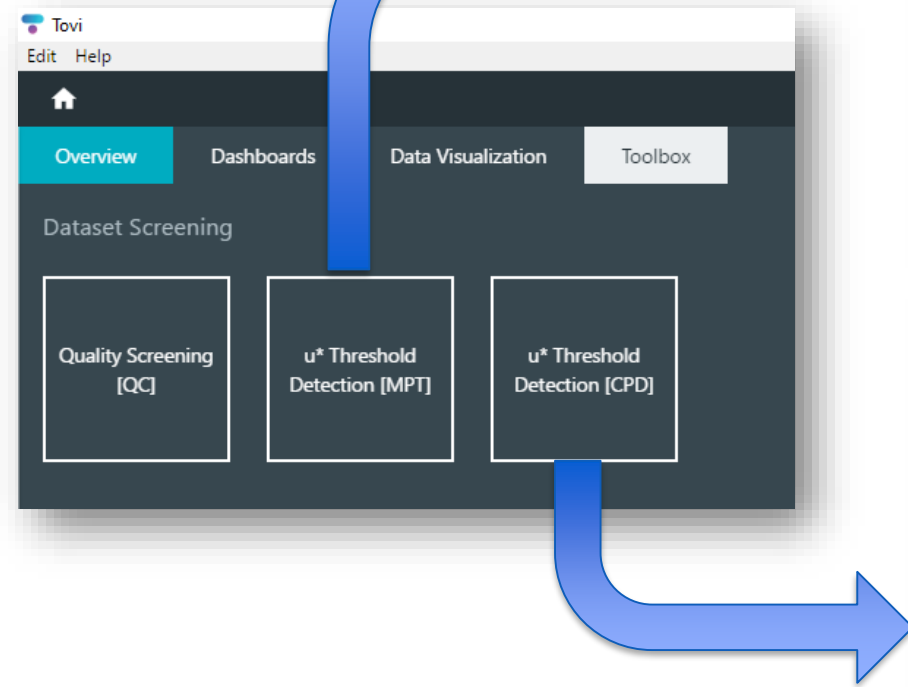


Figure 1: example of the variability in the  $u^*$  threshold found with the bootstrapping. Lines indicate mean (yellow), median (red), 5% (green) and 95% (blue) values.



- Example; both MPT and CPD methods for detecting  $u^*$  threshold are available in

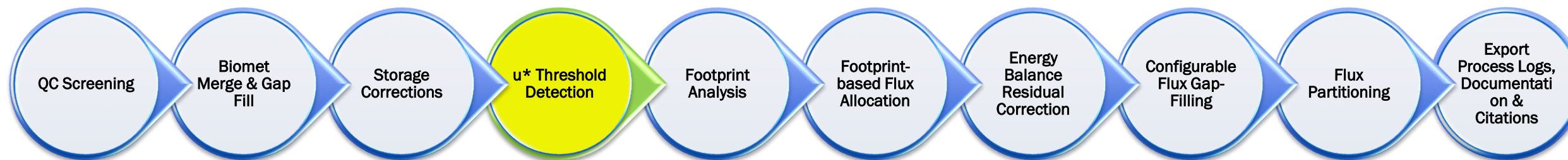


Toolbox **u\* Threshold Detection [MPT]** START

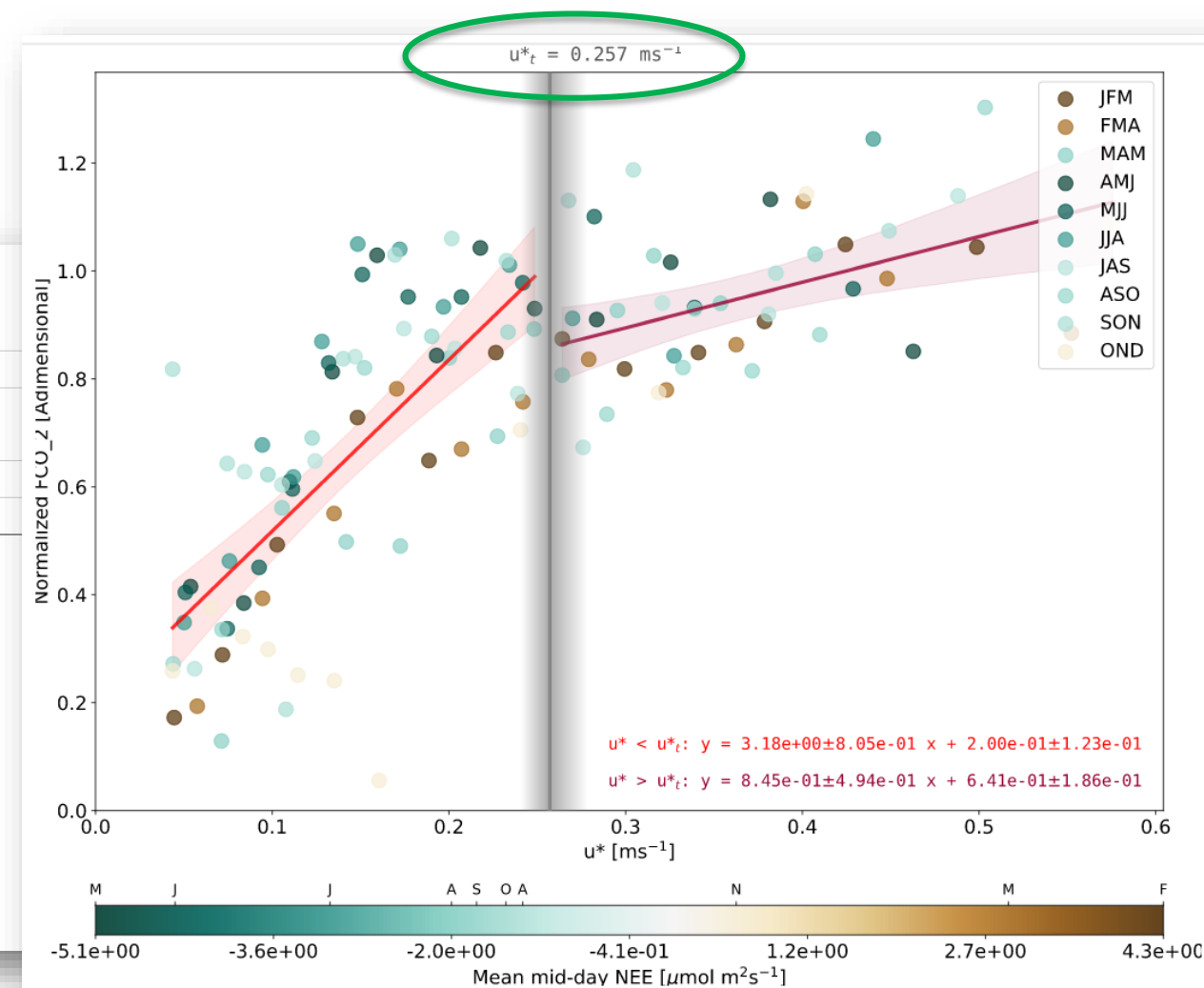
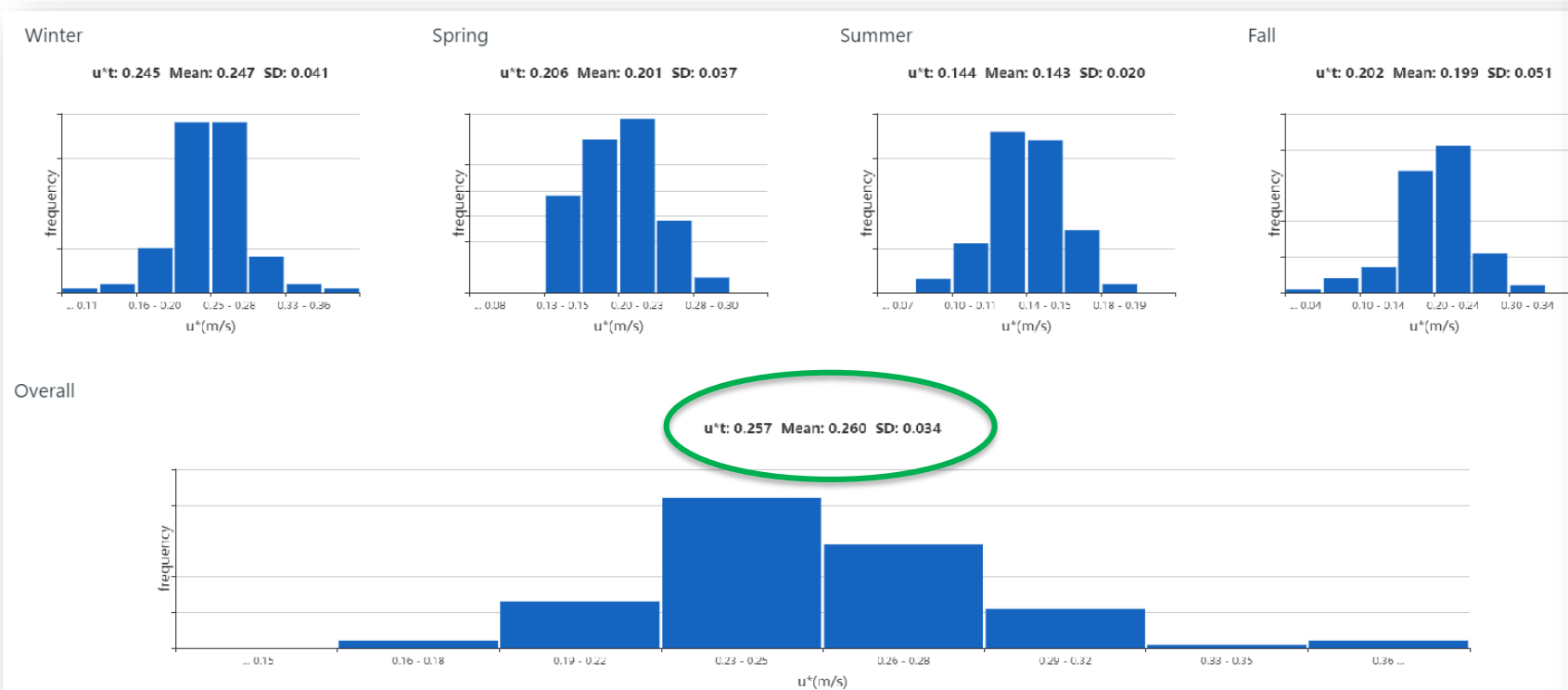
NEE (CO <sub>2</sub> Flux)	FCO2	Bootstraps	100
Friction Velocity	USTAR	Temperature Classes	7
Air Temperature	TA	u* Classes	20
Radiation	SW_IN_1_1_1		

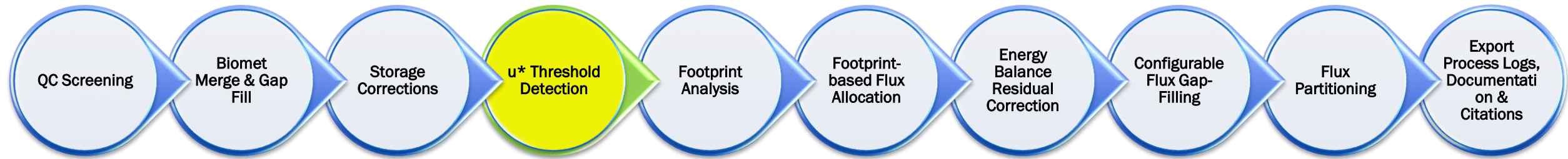
Toolbox **u\* Threshold Detection [CPD]** START

NEE (CO <sub>2</sub> Flux)	FCO2	Bootstraps	10
Friction Velocity	USTAR		
Air Temperature	TA		
Radiation	SW_IN_1_1_1		



■ Example; results of  $u^*$  detection from the MPT method available in Tovi

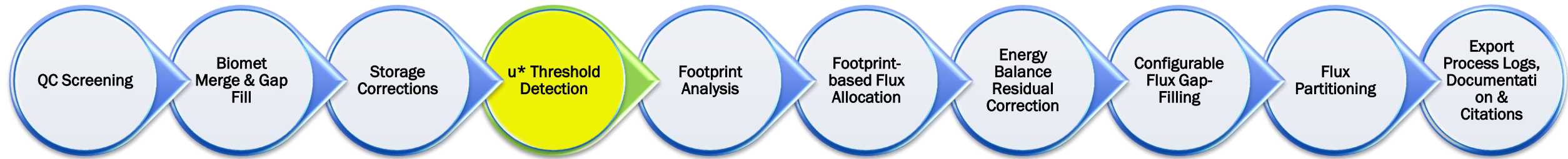




# Example; filtering data based on $u^*$ threshold (Tovi)



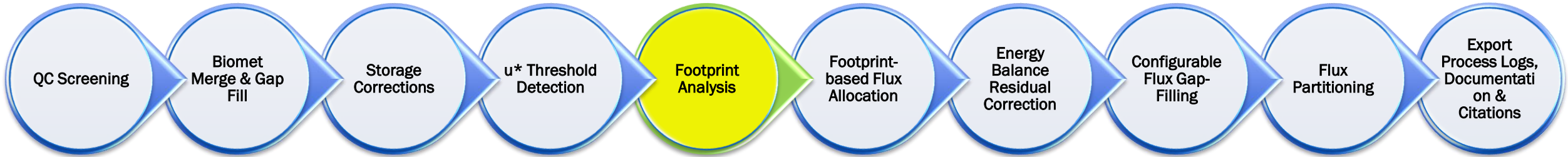
In this example, the CPD technique was used to determine  $u^*$  threshold.  $FCO_2$  was filtered for when  $u^*$  values were below the detected threshold. This technique eliminated 34.7% of the data. (The MPT technique eliminated 38.4%)



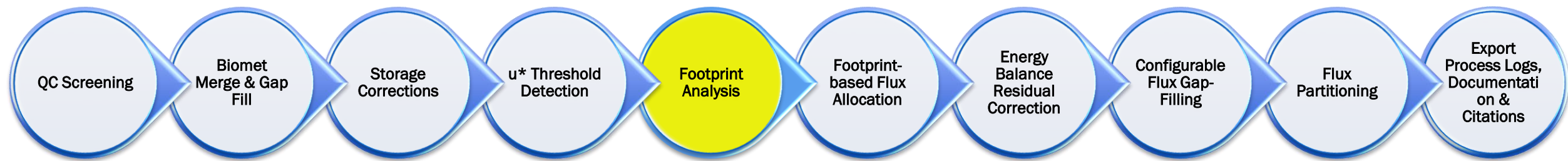
In summary ( $u^*$  threshold detection)...

- It is possible that you will exclude a lot of data
  - Filters out mostly night-time data (primarily respiration) so to avoid a bias in your budgets or cumulative values, you need to appropriately gap fill (more on this later)
- Run both models to see which might work best for your site
- Try using different drivers, different class sizes, or changing the amount of bootstraps to get the best results
- The lower the  $u^*$  value, the less amount of data will be filtered out
- Choose to use this filter or not...
  - Some researchers have sometimes found this is not an ideal, final solution to the nighttime flux issue.
  - Continuous theoretical and experimental research is still needed to overcome the challenges in measuring nighttime fluxes accurately.

# Typical workflow for Post Processing...

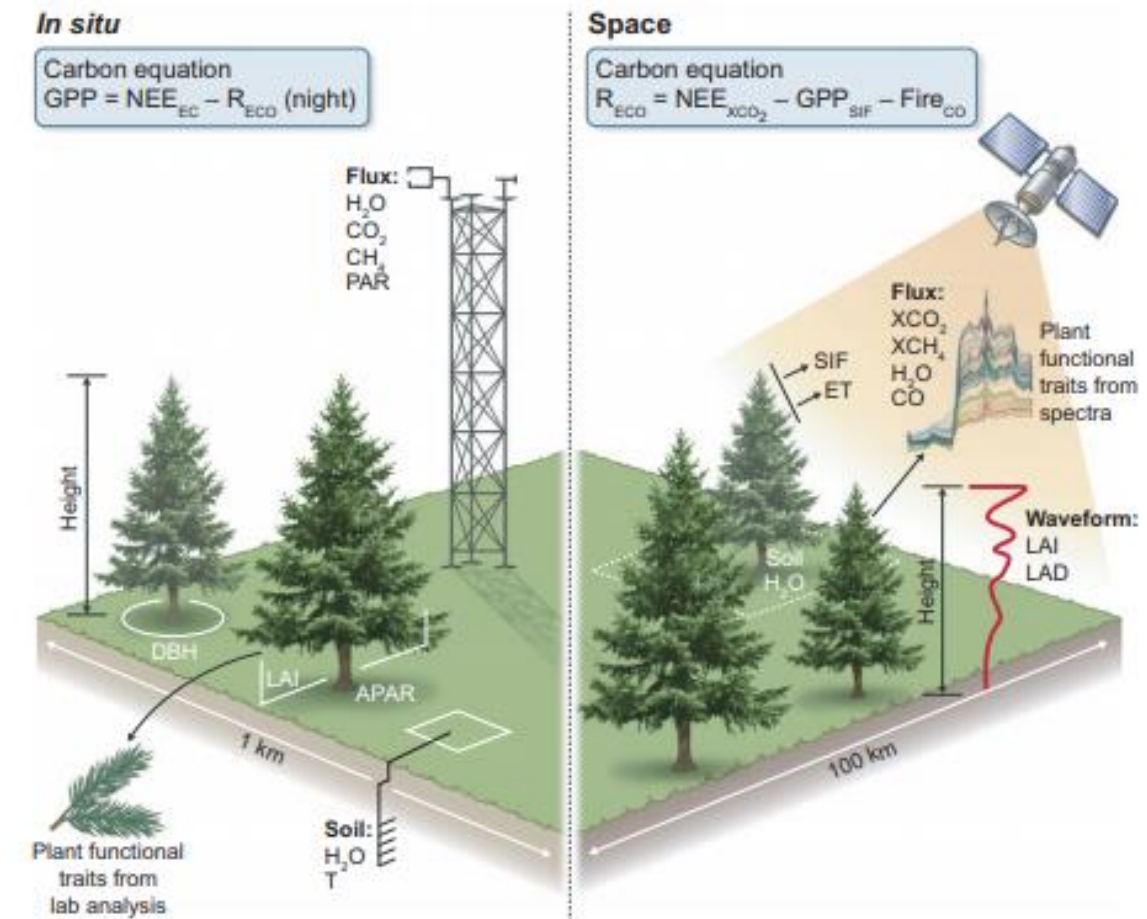




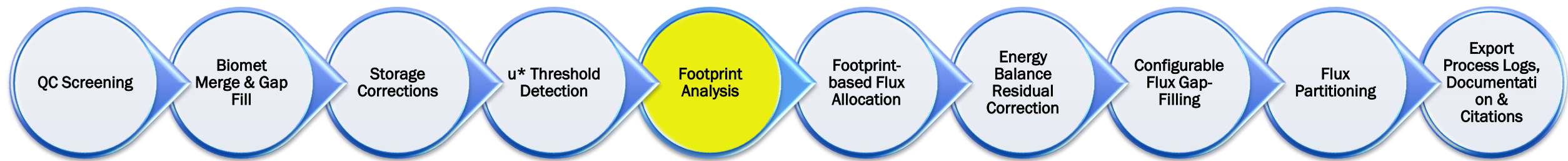


## Flux Footprint models can help in

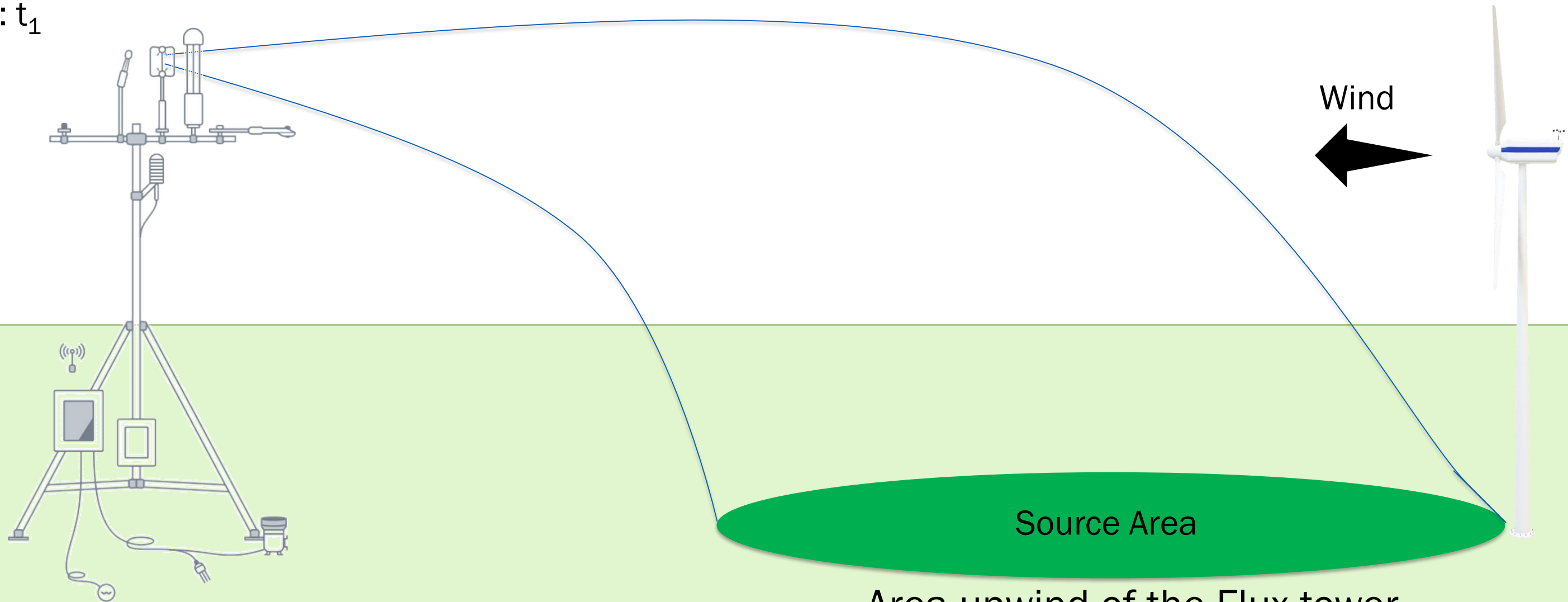
- Estimate position, size and relative contribution of surface source areas
- Upscaling single point measurements to local or regional scale
- Ground truthing satellites estimates of GPP/NPP with pixel matching using footprints
- Important tool for determining ideal location for flux tower
- Flux footprint models are often used for interpretation of flux tower measurements, to estimate position and size of surface source areas, and the relative contribution of passive scalar sources to measured fluxes.



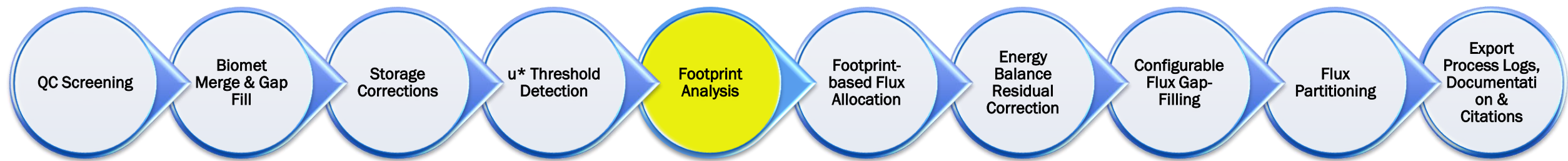
From: Schimel et al., 2019 New Phytologist



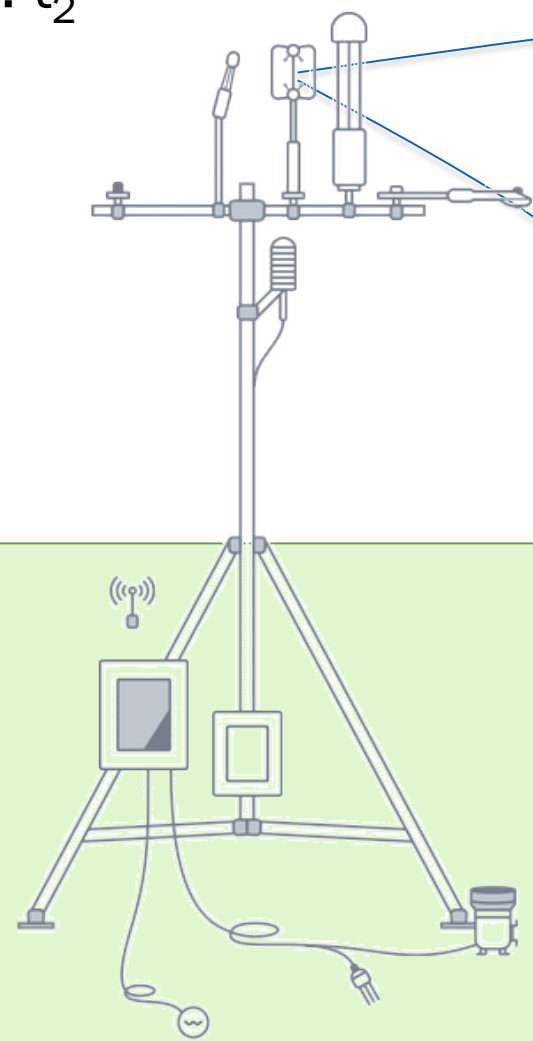
Time:  $t_1$



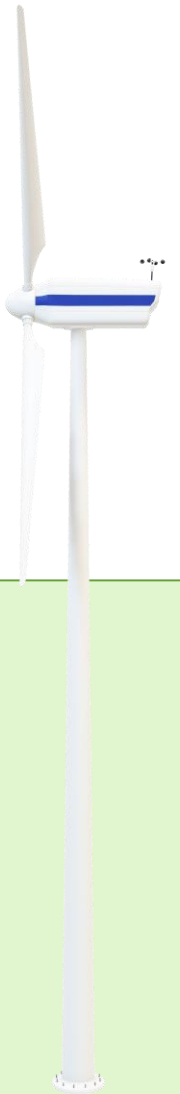
Area upwind of the Flux tower which is being “sensed”



Time:  $t_2$

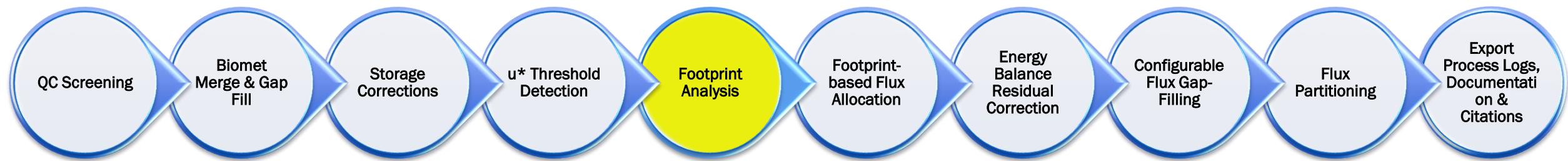


Wind

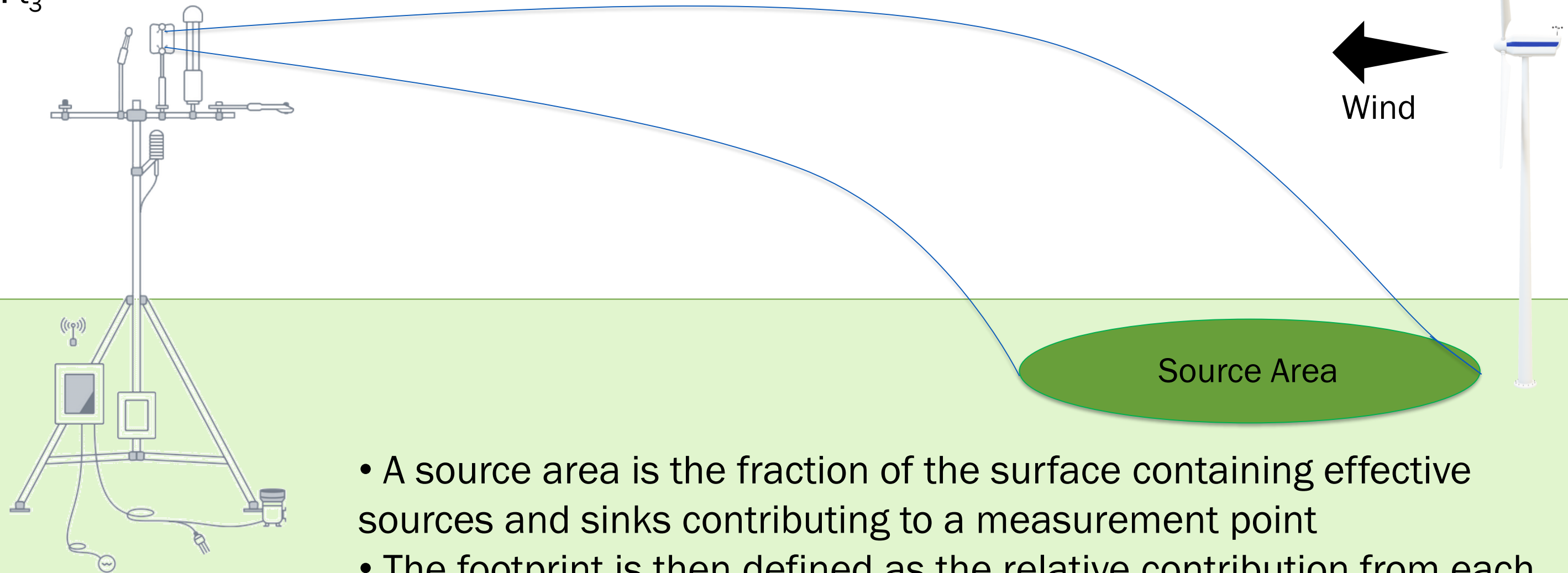


The 'source area' depends on measurement height, wind speed and wind direction, atmospheric stability, and surface characteristics.

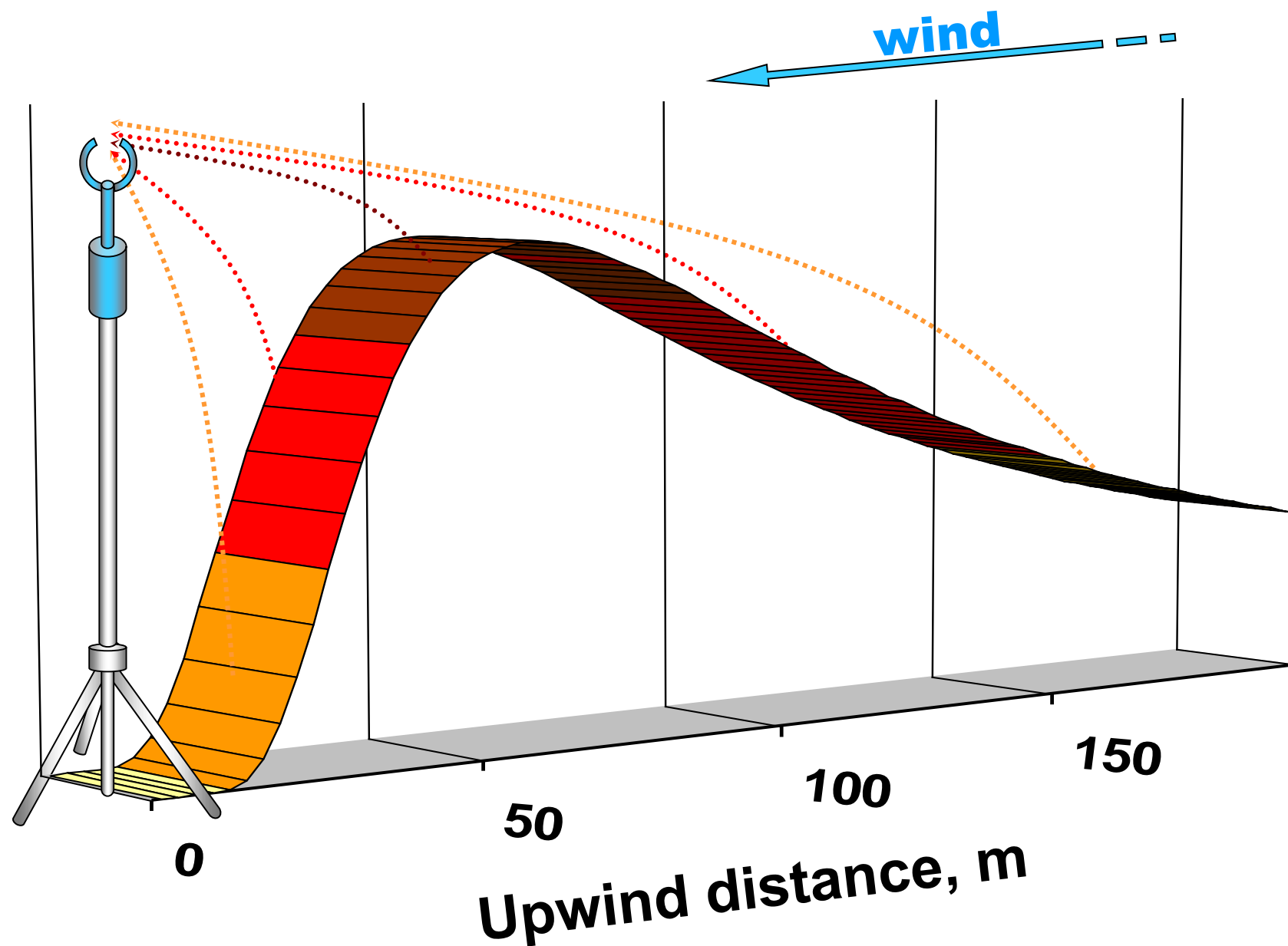
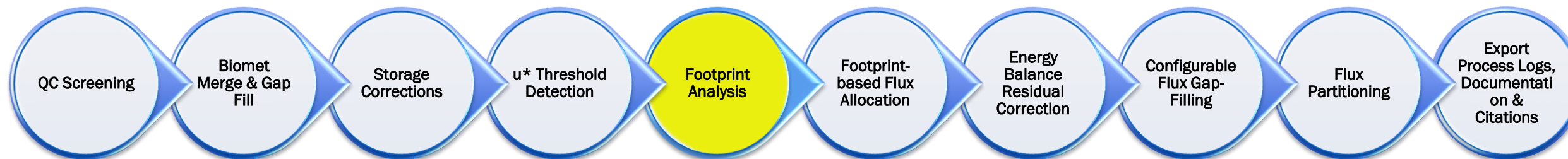
Source Area



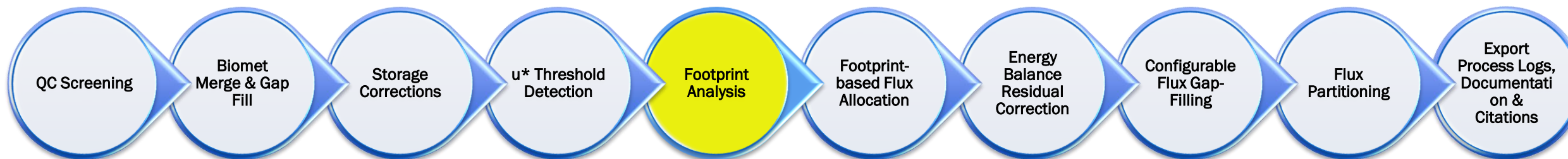
Time:  $t_3$



- A source area is the fraction of the surface containing effective sources and sinks contributing to a measurement point
- The footprint is then defined as the relative contribution from each element of the surface area source/sink to the measured vertical flux

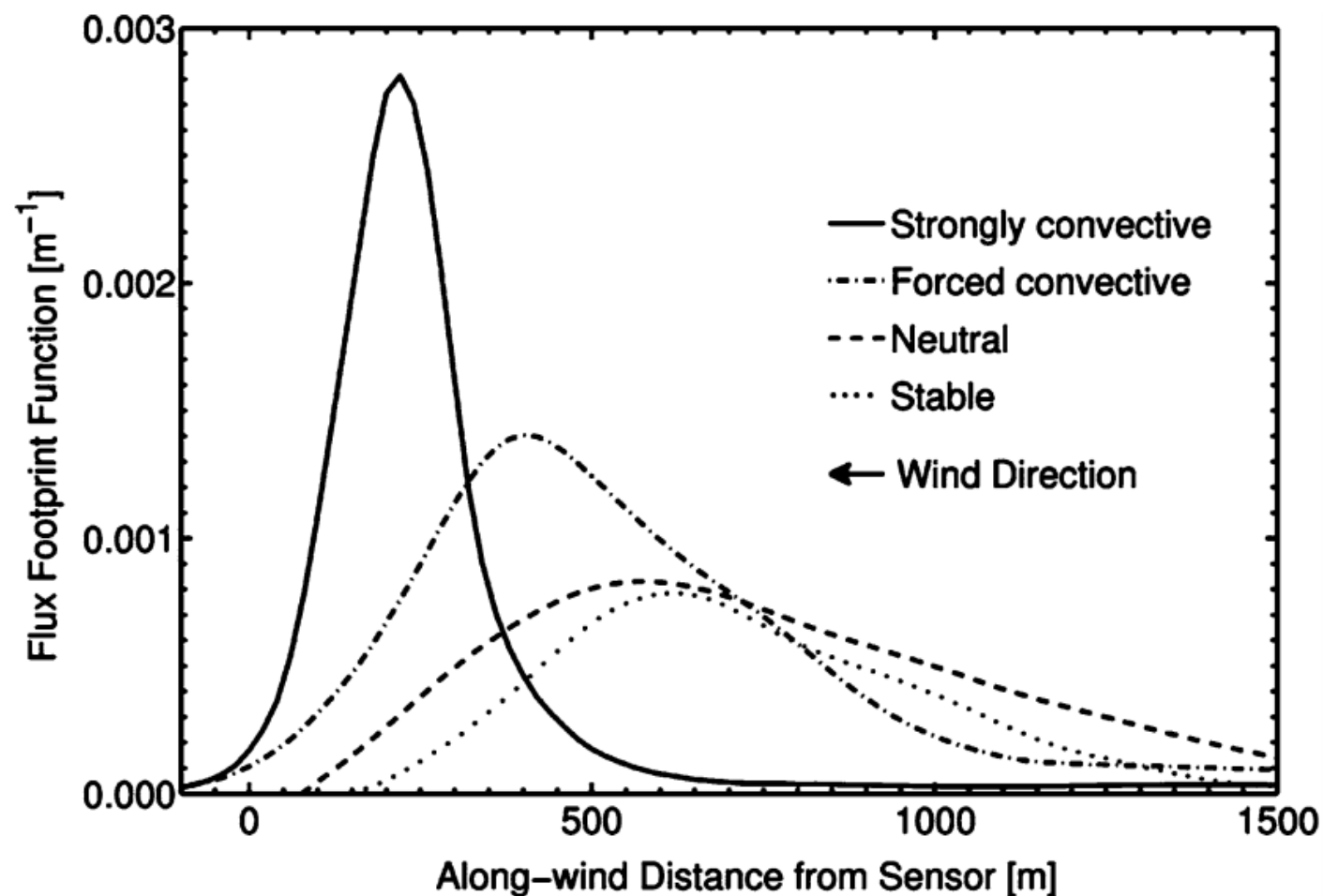


- Flux footprint models provide a probability distribution/density function for “sensing fluxes” as a function of distance from the tower
- The highest probability is indicated by the darker areas and it is usually an area away from the bottom of the tower
- The peak contributing area is not a fixed point as it changes with the wind direction and stability of the atmosphere for a given tower height
- Mathematically, the surface area of influence on the entire flux goes to infinity



## FLUX FOOTPRINT DEPENDS ON:

*Measurement height*  
*Surface roughness*  
*Thermal stability*

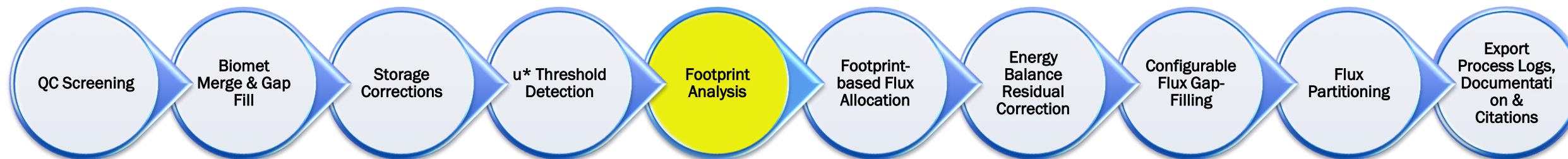


## SIZE OF FOOTPRINT INCREASES WITH:

- ↑ Increased measurement height
- ↓ Decreased surface roughness
- Change in stability from unstable to stable

## AREA NEAR TOWER MAY CONTRIBUTE A LOT IF:

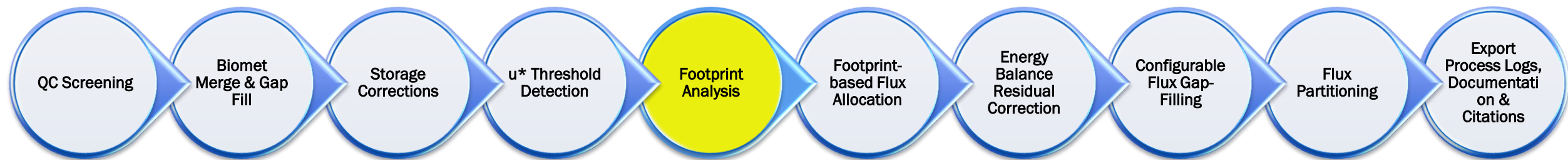
- ↓ Measurement height is low
- ↑ Surface roughness is high
- ~~ Conditions are very unstable



- Several atmospheric scientists and modelers worked on a variety of model types over a variety of ecosystems:
  - Complex footprint models based on large-eddy simulation (LES; e.g. Luhar and Rao, 1994; Leclerc et al., 1997; Steinfeld et al., 2008; Wang and Davis, 2008)
  - Eulerian models of higher-order turbulence closure (e.g. Sogachev and Lloyd, 2004)
  - Lagrangian stochastic particle dispersion (LPD; e.g. Leclerc and Thurtell, 1990; Horst and Weil, 1992; Flesch, 1996; Baldocchi, 1997; Rannik et al., 2000; Kljun et al., 2002; Hsieh et al., 2003)
  - Combination of LES and LPD (e.g. Markkanen et al., 2009; Hellsten et al., 2015) can, to a certain degree, offer the ability to resolve complex flow structures (e.g. flow over a forest edge or a street canyon) and surface heterogeneity.
- However, many models are laborious to run, still highly CPU intensive, and hence can in practice be applied only for case studies over selected hours or days, and most are constrained to a narrow range of atmospheric conditions.

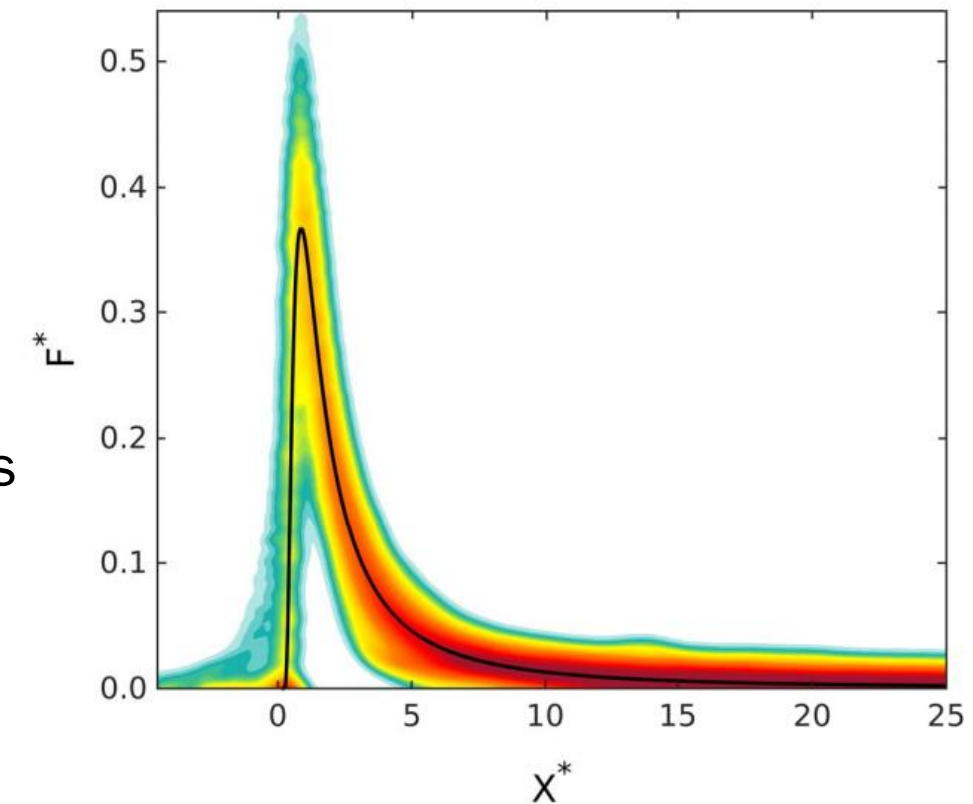
Author	Remarks
Pasquill (1972)	First model description, concept of effective fetch
Gash (1986)	Neutral stratification, concept of cumulative fetch
Schuepp et al. (1990)	Use of source areas, but neutral stratification and averaged wind velocity
Leclerc and Thurtell (1990)	Lagrangian footprint model
Horst and Weil (1992)	One-dimensional footprint model
Schmid (1994, 1997)	Separation of footprints for scalars and fluxes
Leclerc et al. (1997)	LES model for footprints
Baldocchi (1997)	Footprint model within forests
Rannik et al. (2000, 2003)	Lagrangian model for forests
Kormann and Meixner (2001)	Analytical model with exponential wind profile
Kljun et al. (2002)	Three-dimensional Lagrangian model for various turbulence stratifications with backward trajectories
Sogachev and Lloyd (2004)	Boundary-layer model with 1.5 order closure
Sogachev et al. (2004)	Footprint estimates for a non-flat topography
Strong et al. (2004)	Footprint model with reactive chemical compounds
Cai and Leclerc (2007)	Footprints from backward and forward in-time particle simulations driven with LES data
Klaassen and Sogachev (2006)	Footprint estimates for a forest edge
Vesala et al. (2008a)	Footprint estimates for a complex urban surface
Steinfeld et al. (2008)	Footprint model with LES-embedded particles

Adopted from Foken (2008) and Vesala et al. (2010)

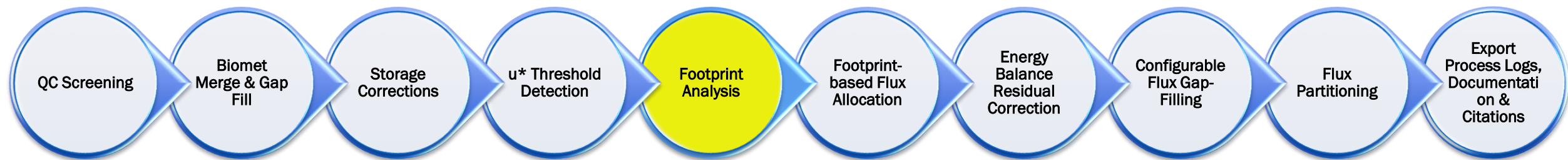


- Some footprint models offer simple parameterizations (Horst and Weil (1992, 1994), Weil and Horst (1992), Schmid (1994) or Hsieh et al. (2000)).
- The primary drawback of these is their limitation to a particular turbulence scaling domain or to a limited range of stratifications.
- Kljun et al. (2004b) introduced a footprint parameterization based on a fit to scaled footprint estimates derived from the Lagrangian stochastic particle dispersion model, LPDM-B (Kljun et al., 2002).
  - However, the Kljun et al. (2004b) model comprises only the crosswind-integrated footprint, i.e. it describes the footprint function's upwind extent but not its width (one-dimension only)
- The three most used one-dimensional models are from Hsieh et al. (2000) with crosswind extension of Detto et al. (2006), the model of Kormann and Meixner (2001), and the footprint parameterization of Kljun et al. (2004b),

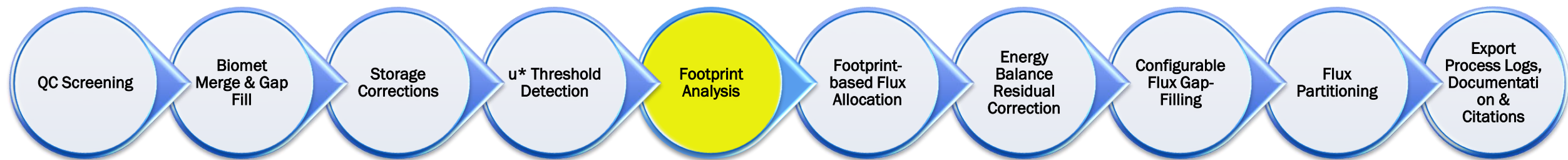
One-dimensional footprint output







- As a result, Kljun et al. developed a **two-dimensional** parameterization for Flux Footprint Prediction (FFP)
- Based on a novel scaling approach for the crosswind distribution of the flux footprint and on an improved version of the footprint parameterization of Kljun et al. (2004b).
- Compared to the original model, it now also provides the width and shape of footprint estimates, and explicit consideration of the effects of the surface roughness length.
- Like the Kljun et al., 2002 model, the parameterization is valid for a broad range of boundary layer conditions and measurement heights over the entire planetary boundary layer. Thus, it can provide footprint estimates for a wide range of real-case applications.



- The two-dimensional footprint function is calculated by applying the crosswind dispersion to the crosswind integrated footprint.
- The new footprint model is computationally inexpensive and can be run easily for several years of data in, for example, half-hourly time steps.

Geosci. Model Dev., 8, 3695–3713, 2015  
 www.geosci-model-dev.net/8/3695/2015/  
 doi:10.5194/gmd-8-3695-2015  
 © Author(s) 2015. CC Attribution 3.0 License.



Geoscientific  
 Model Development  
 Open Access  
 EGU

### A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP)

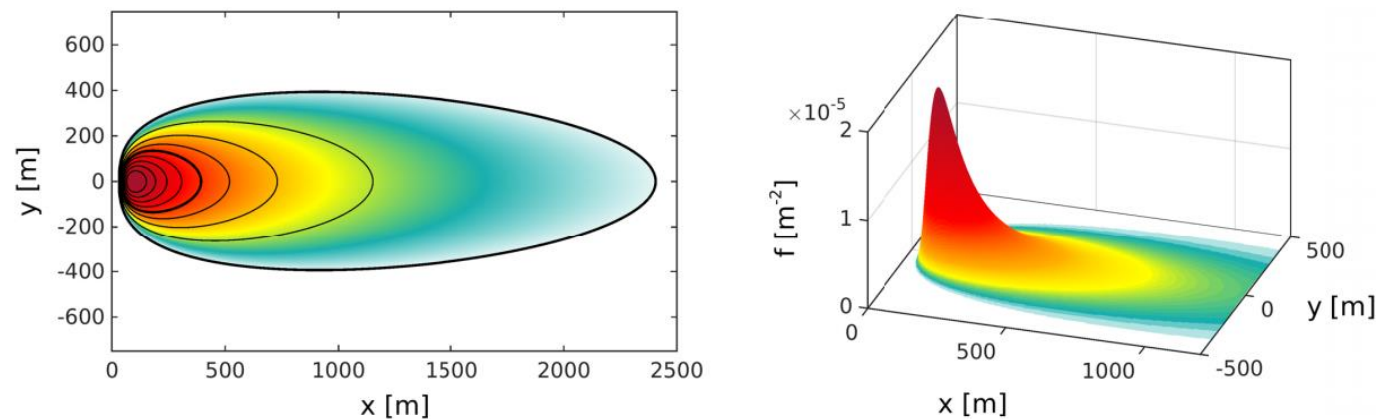
N. Kljun<sup>1</sup>, P. Calanca<sup>2</sup>, M. W. Rotach<sup>3</sup>, and H. P. Schmid<sup>4</sup>

<sup>1</sup>Department of Geography, Swansea University, Swansea, UK

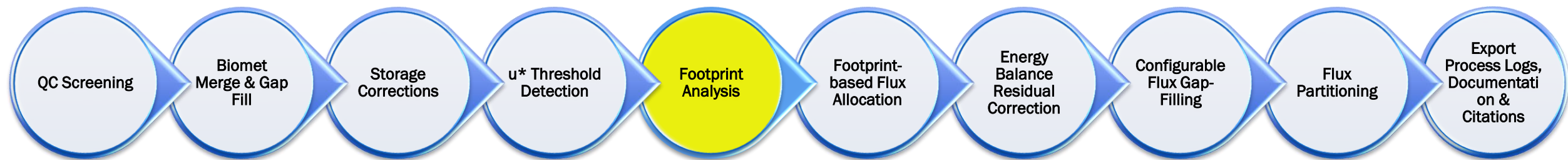
<sup>2</sup>Agroscope, Institute for Sustainability Sciences, Zurich, Switzerland

<sup>3</sup>Institute of Atmospheric and Cryospheric Sciences, Innsbruck University, Innsbruck, Austria

<sup>4</sup>KIT, Institute of Meteorology and Climate Research, Garmisch-Partenkirchen, Germany

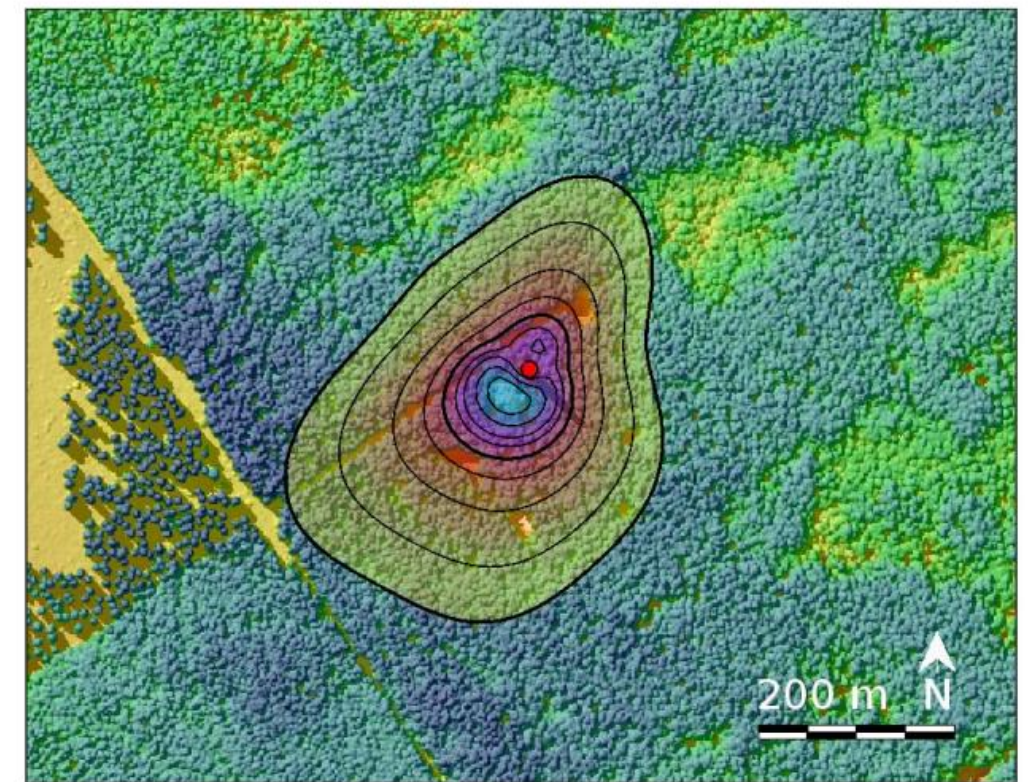


**Figure 4.** Example footprint estimate for the convective Scenario 1 of Table 1, a measurement height of 20 m, and a roughness length of 0.01 m. The receptor is located at (0/0) m and the  $x$  axis points towards the main wind direction. Footprint contour lines (left panel) are shown in steps of 10 % from 10 to 90 %.

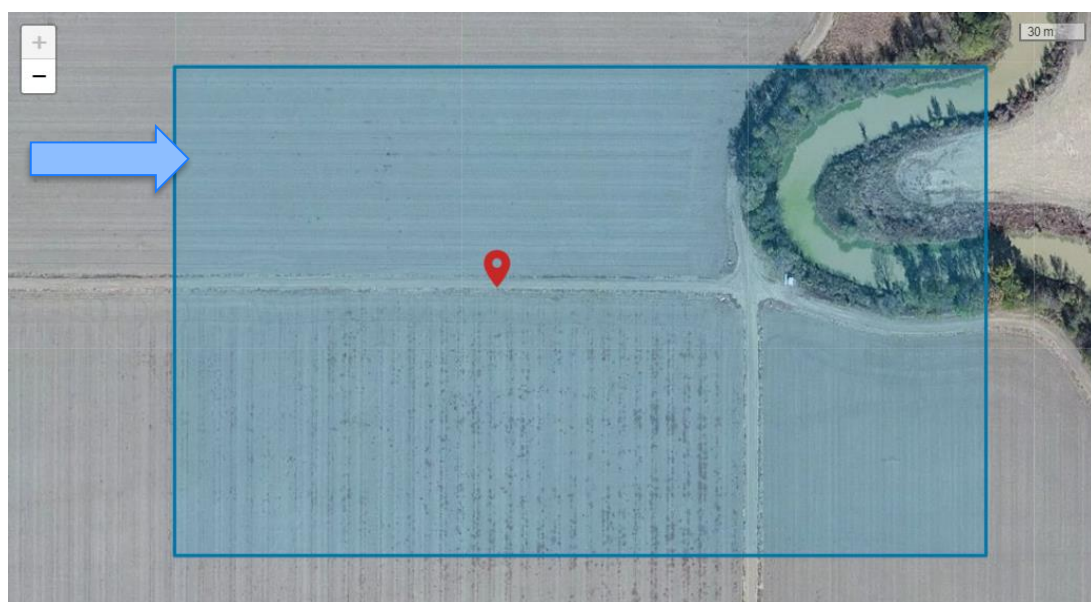
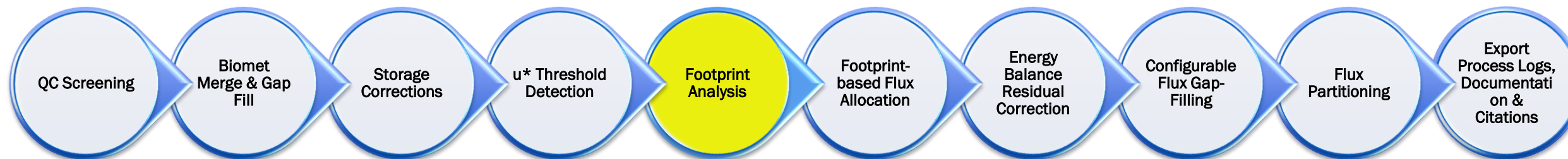


- Each single data point can be associated with its source area by converting the footprint coordinate system to geographical coordinates and positioning the spatial array onto a map or aerial image surrounding the measurement site.
- In many cases, an aggregated footprint (footprint climatology) is of more interest than a series of footprint estimates.

One month (May) of footprint aggregates

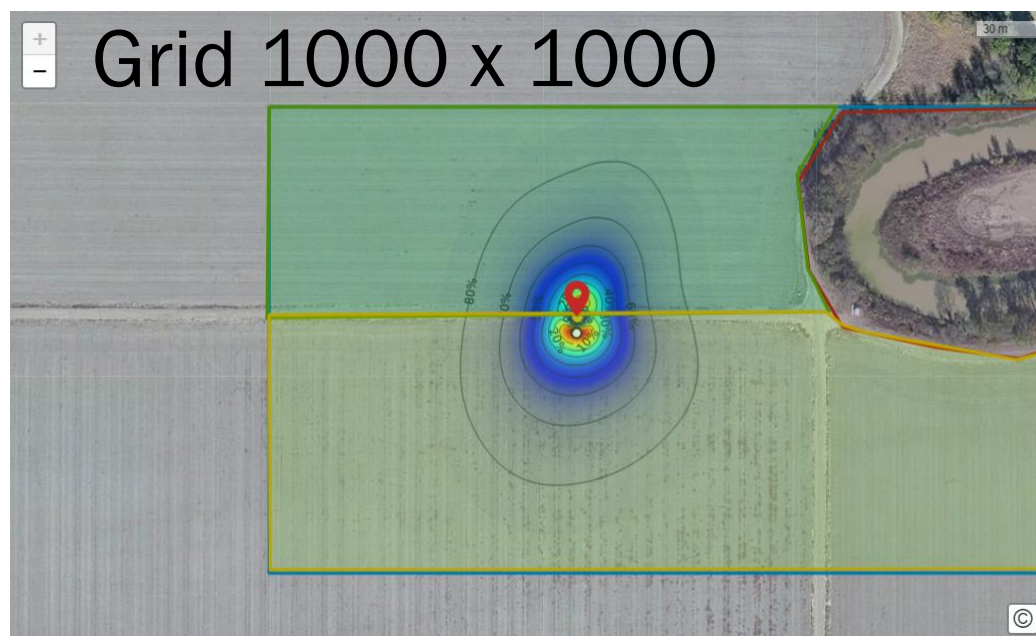
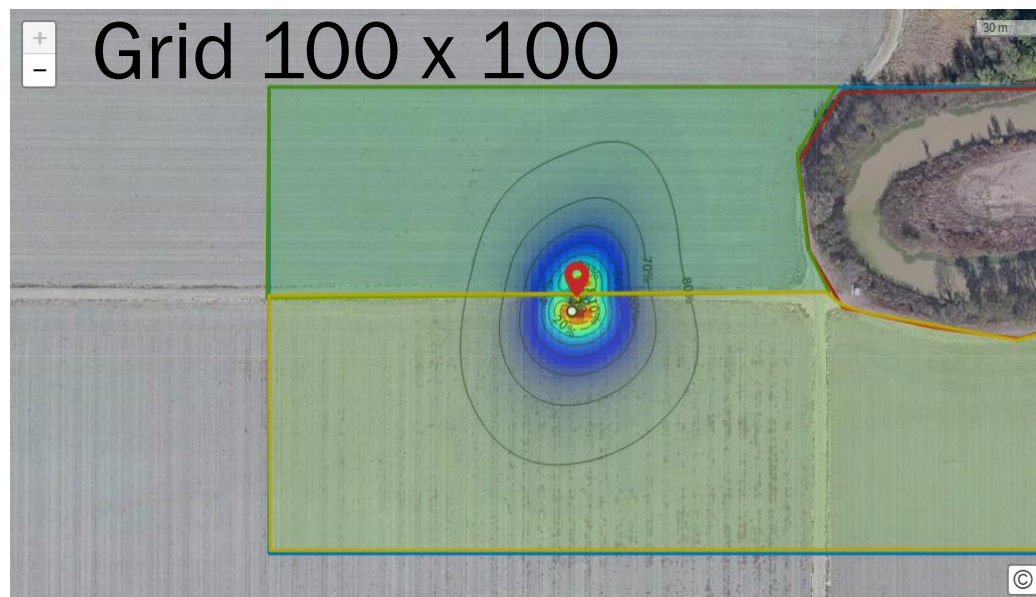
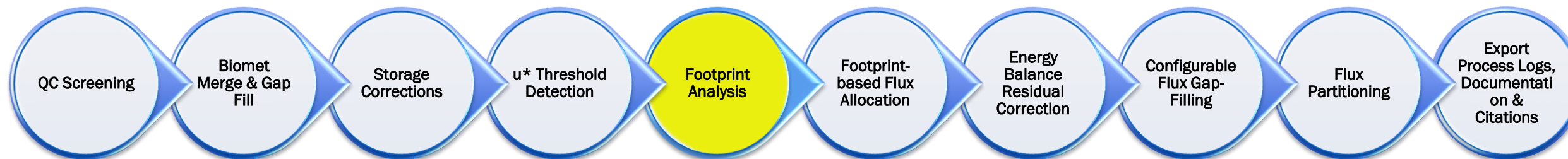


**Figure 5.** Example footprint climatology for the ICOS flux tower Norunda, Sweden, for 1–31 May 2011. The red dot depicts the tower location with a receptor mounted at  $z_m = z_{\text{receptor}} - z_d = 12$  m. Footprint contour lines are shown in steps of 10% from 10 to 90%. The background map is tree height derived from an airborne lidar survey.



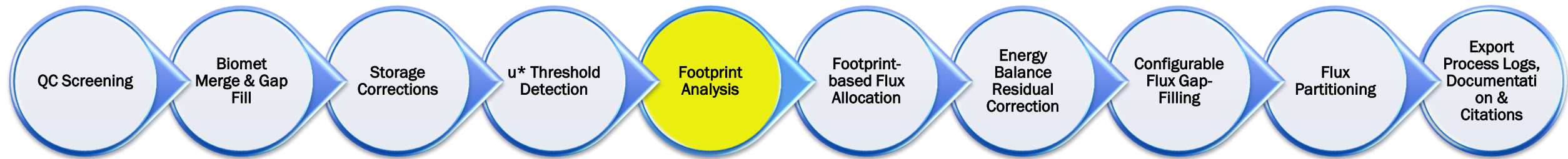
## Computation Grid

- The area or domain in which the footprint model runs.
- Size of the domain should be selected so that the footprint result details are easily distinguished.
  - Often, the interest lies in the extent and location of the area contributing to, for example, 80 % of the measured flux.
- Good starting point is the 100:1 fetch-to-measurement height (rule of thumb)

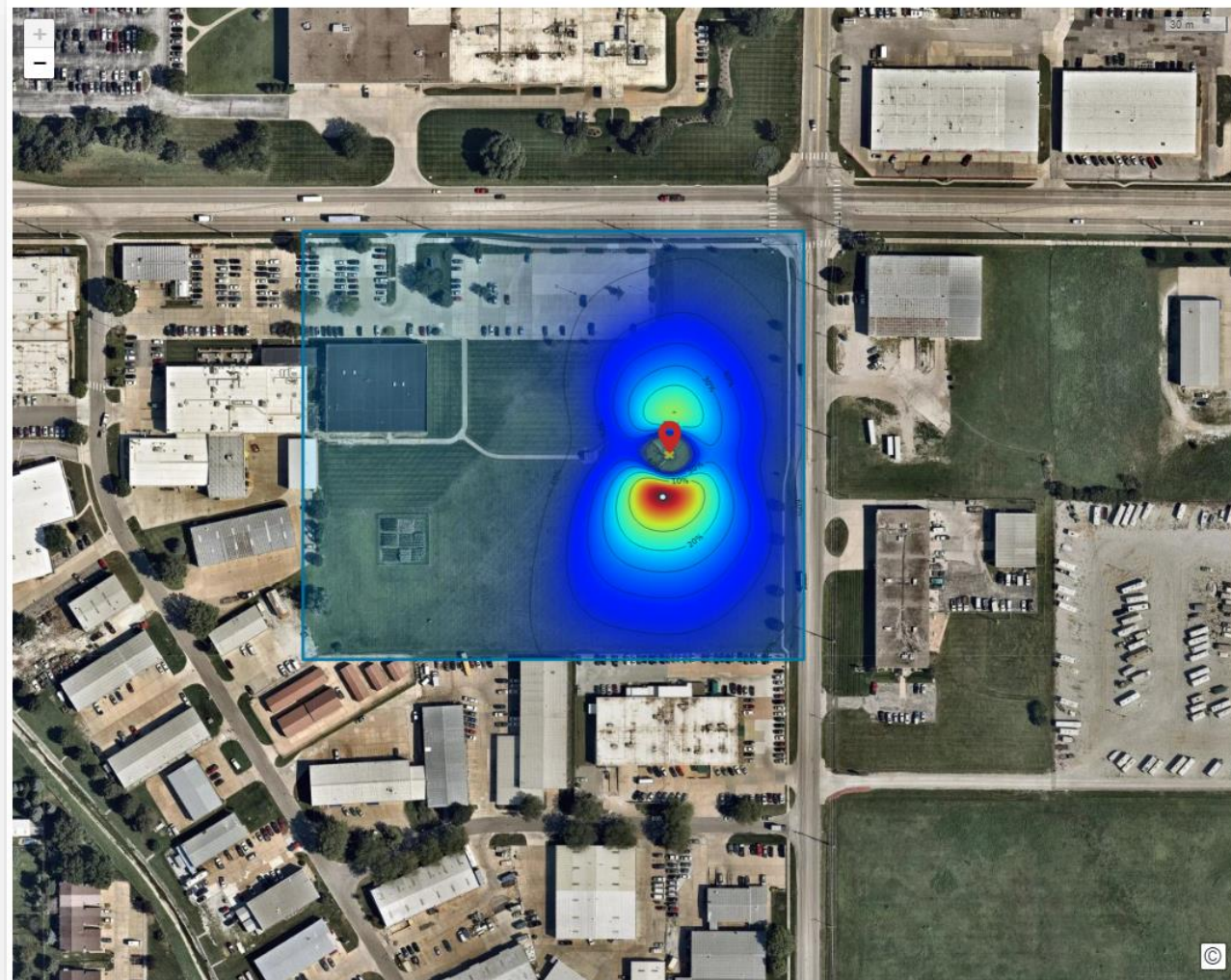


## Grid size

- Increasing the grid resolution can produce figures with higher clarity
- However, increased resolution increases the time required to run a Footprint model



# Example; Footprint results from Tovi



Save Map Image    Export All Data

**Climatology**

- Contours
- Heat Map

**Drawing**

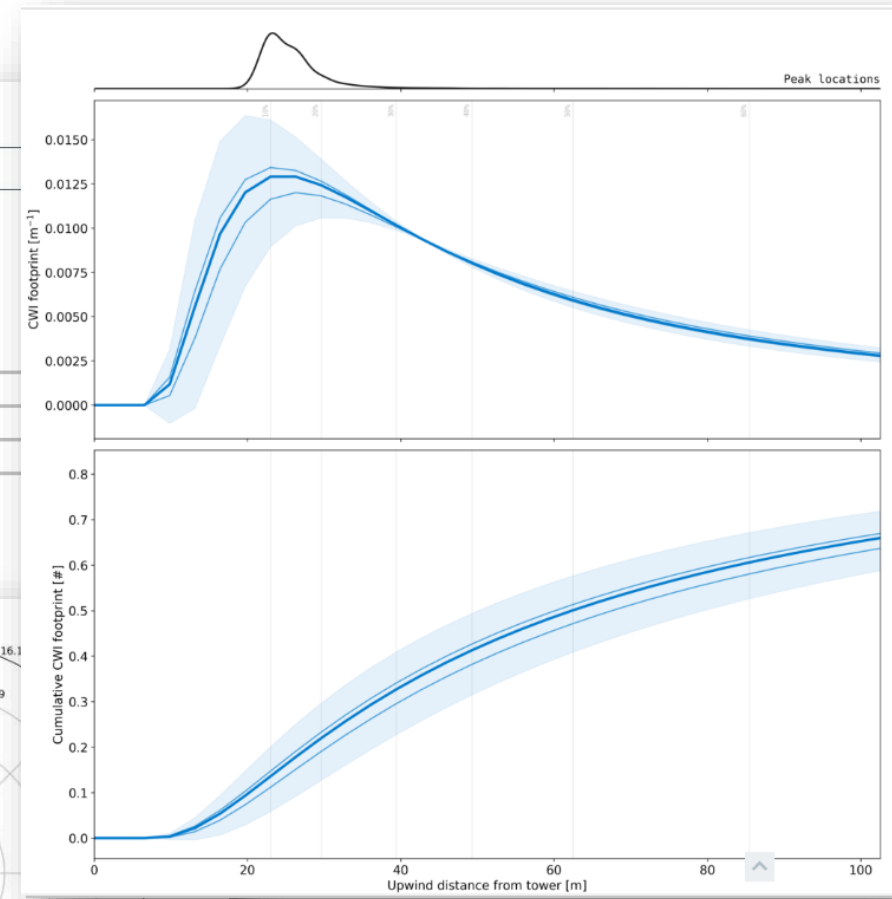
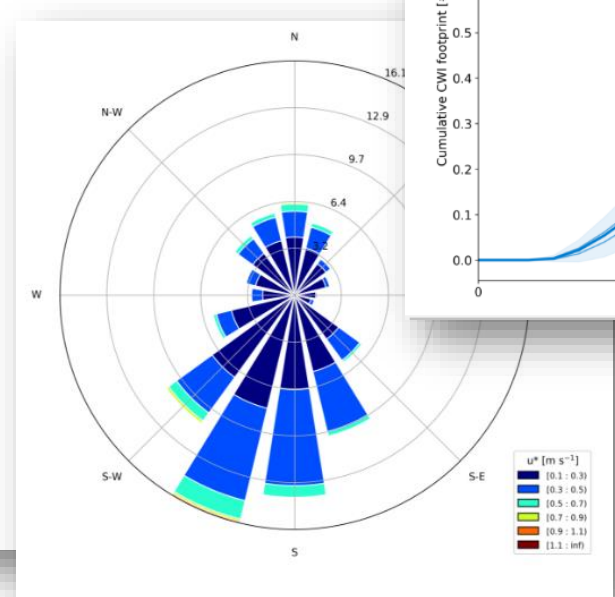
- Domain
- Flagging
- Class One
- Class Two
- Tiles

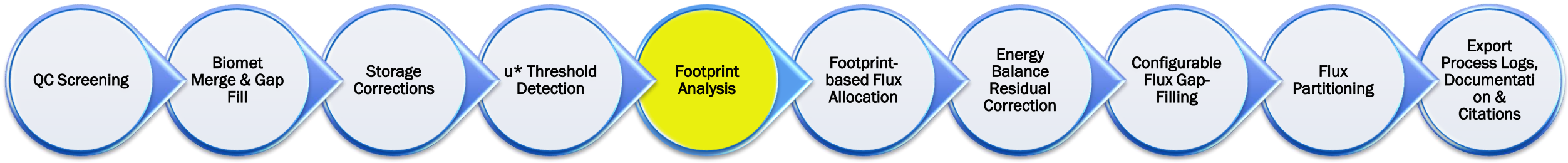
**Tools**

- Marker
- Site
- Zoom
- Scale

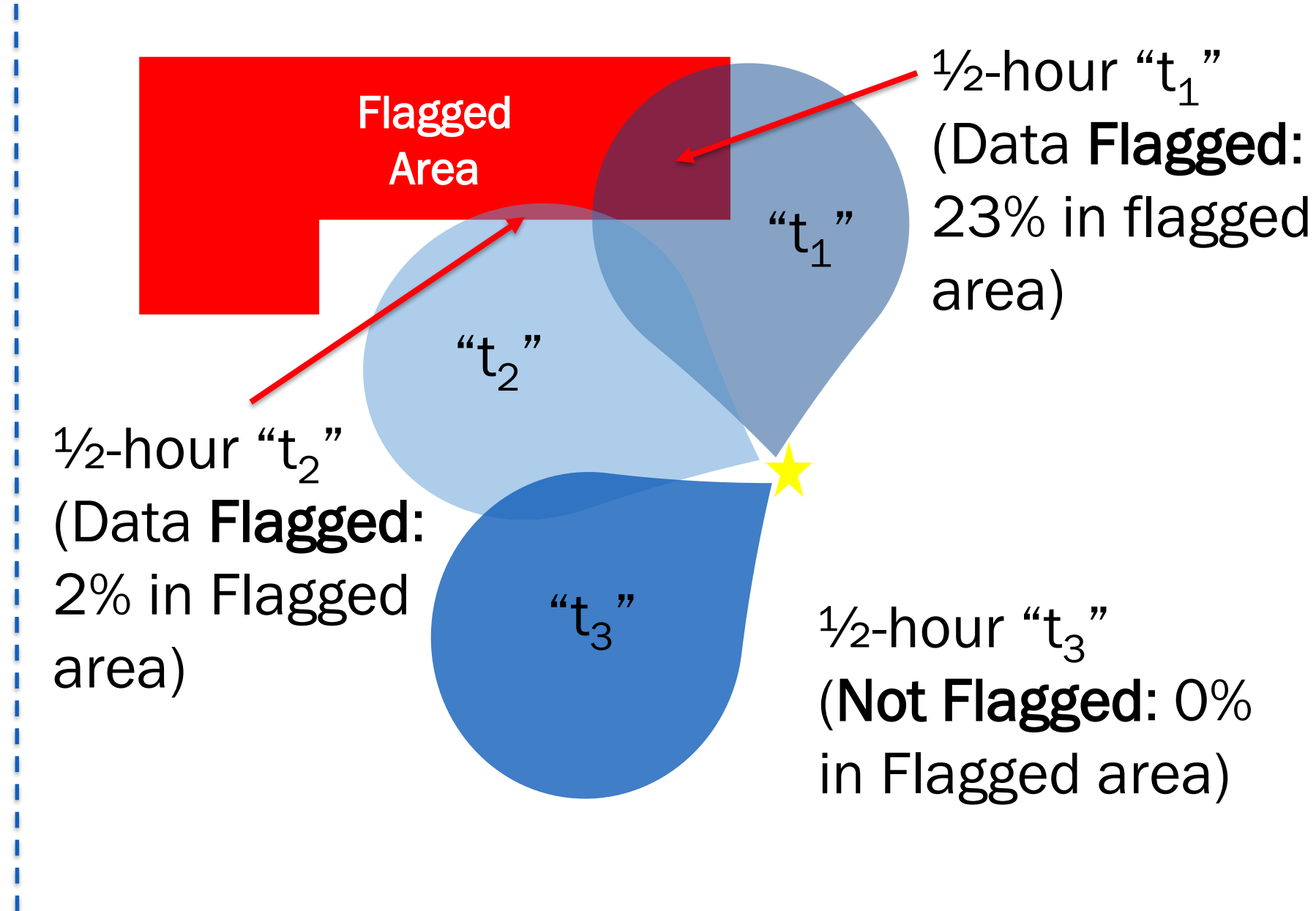
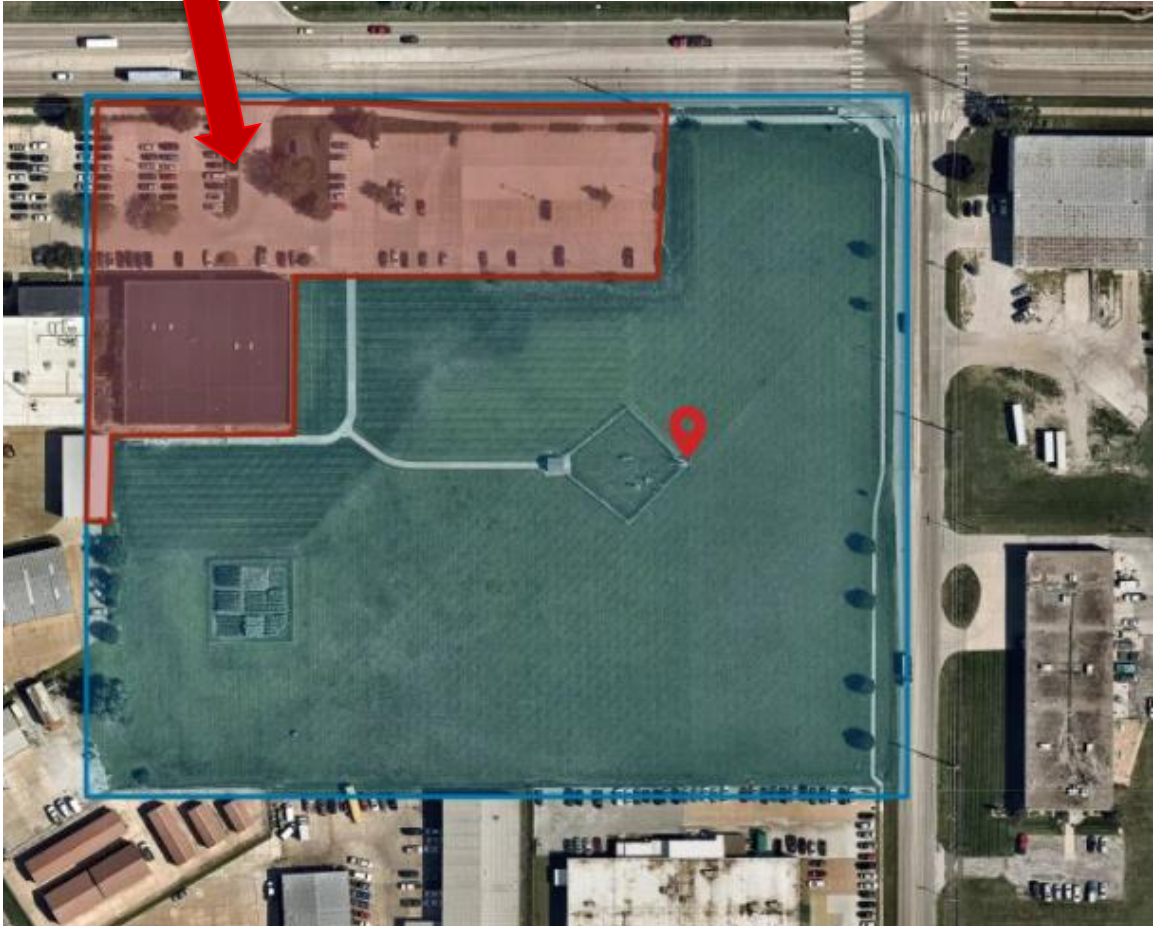
**Map Settings**

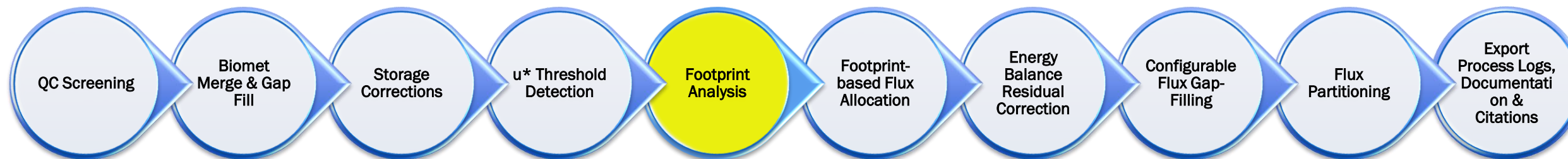
- Border
- Shading
- Shading
- Shading
- Shading
- Shading
- Map
- Terrain
- Geo Lines
- Geo Ticks
- Compass
- Attribution





Draw 'Quality Flag' Regions in Tovi (ex; a walking path, parking lot or pond)





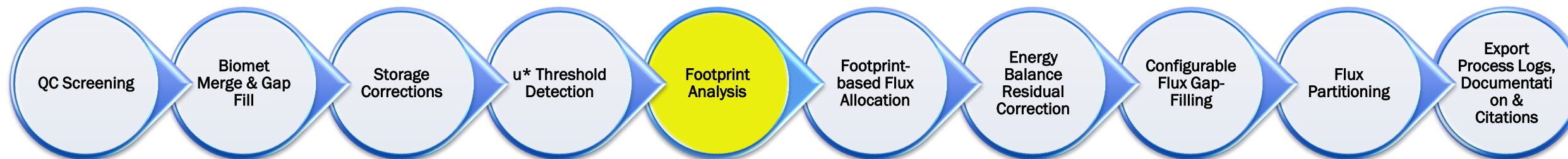
# ■ Example; Filter out 'Footprint Flagged' flux results





Tovi gives an option in QA/QC tool to exclude fluxes when flux footprint probability is within certain limits (fraction) from the quality flagging region

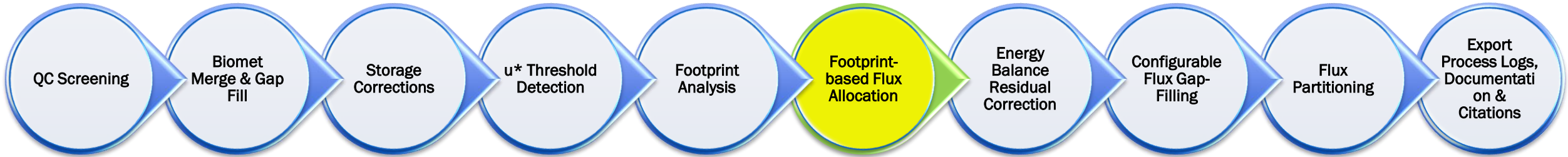


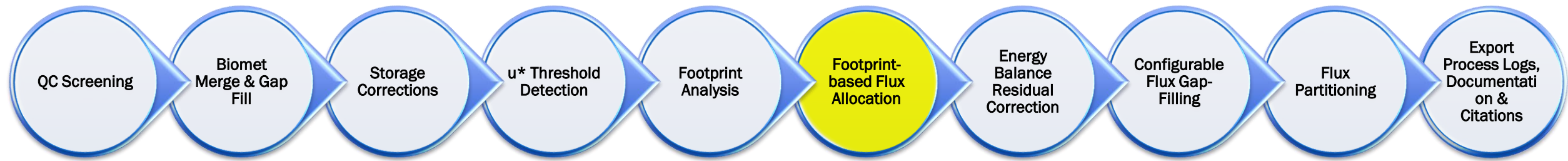




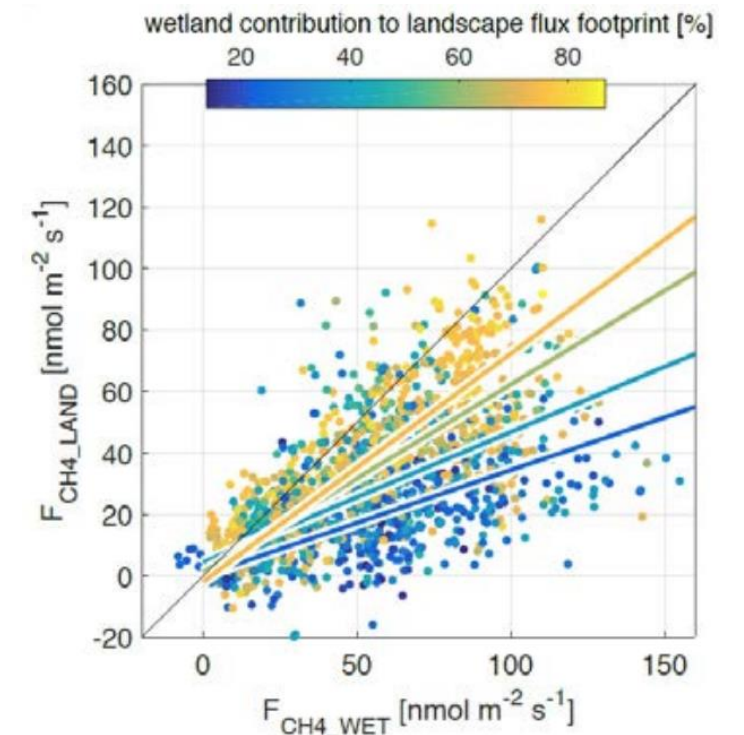
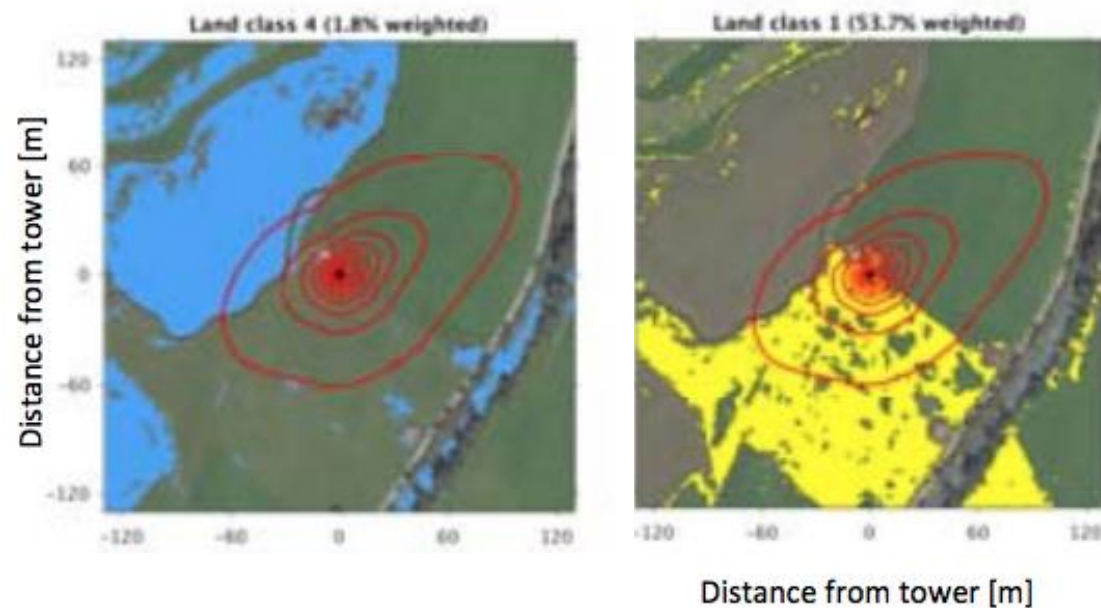
EddyPro 	Tovi 
Three footprint model choices a) Kljun <i>et al.</i> (2004) b) Kormann and Meixner (2001) c) Hsieh <i>et al.</i> (2000)	Kljun <i>et al.</i> (2015) Tovi runs this footprint model internally with inputs from EddyPro output file
All models provided are one dimensional	Two dimensional model
Not configurable – using information from metadata	Footprint domain can be configured, with areas of interest and QA/QC regions.
Outputs include distances from the tower for various percentage contributions	Entire output is provided in graphs and downloadable maps and gridded data.
No visualization	Map based visualization and plots
No climatological outputs based on stability, seasons or time of day	Climatological aggregation of data is available

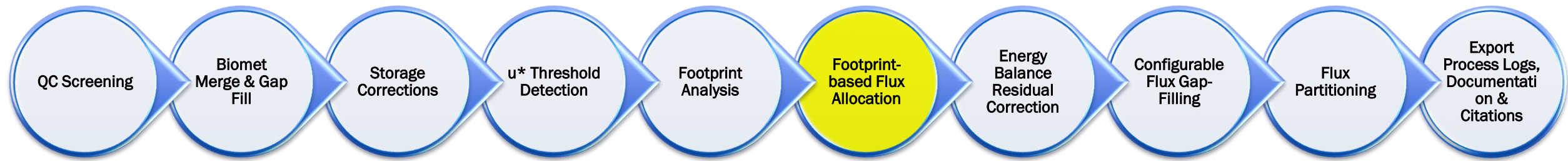
# Typical workflow for Post Processing...





- Footprint fractions from a given land cover type can be used to interpret the data.
- Fluxes can be allocated to a given land cover type, based on footprint fraction.

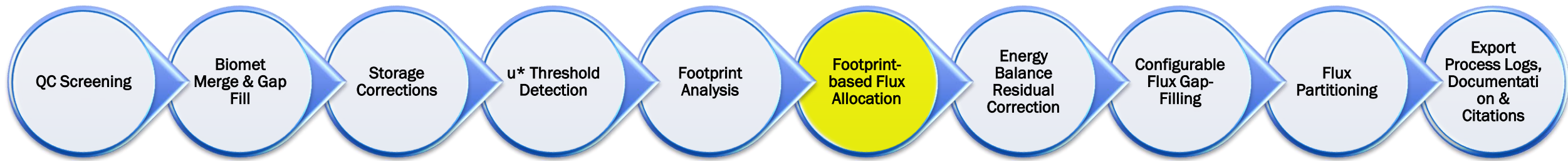




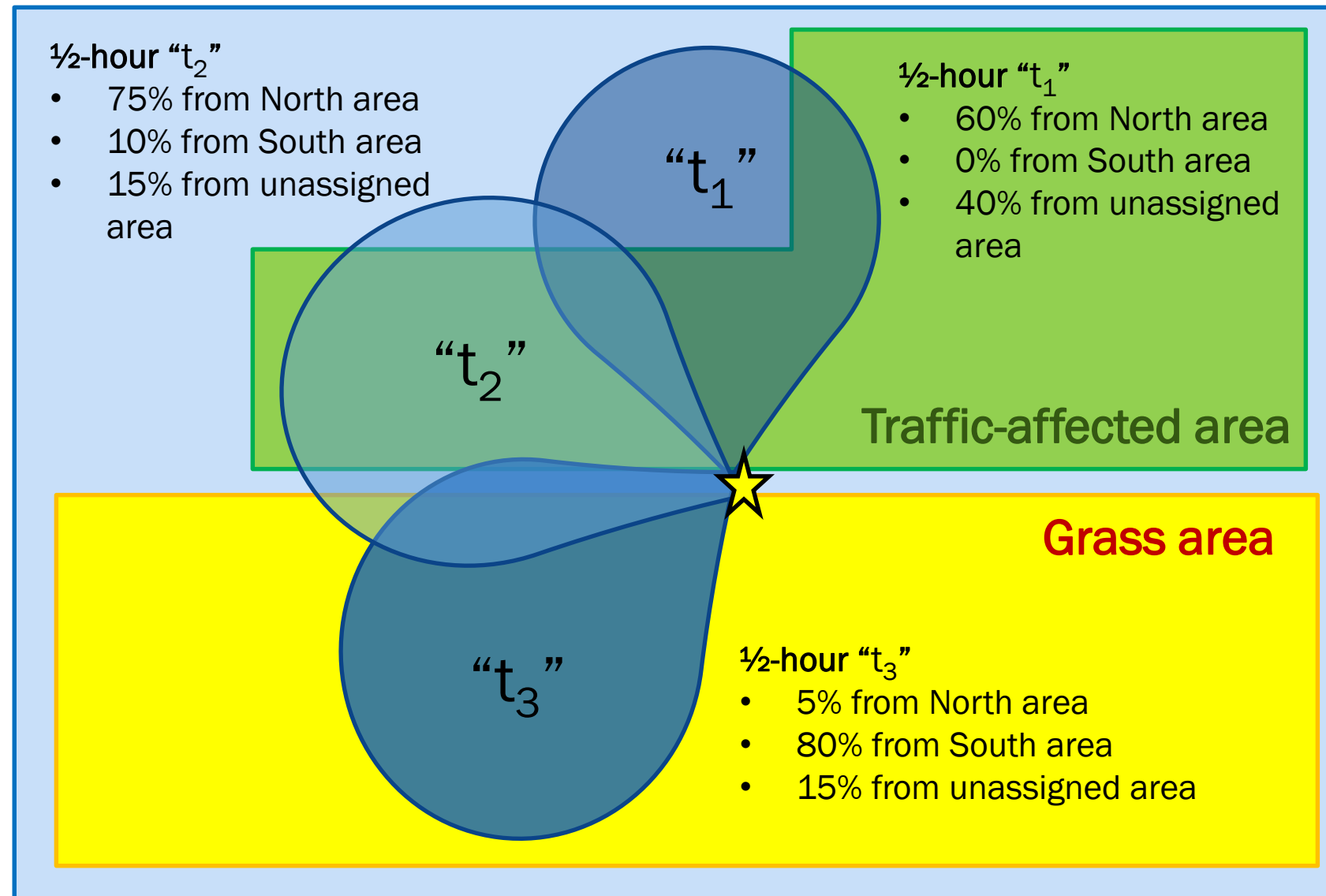
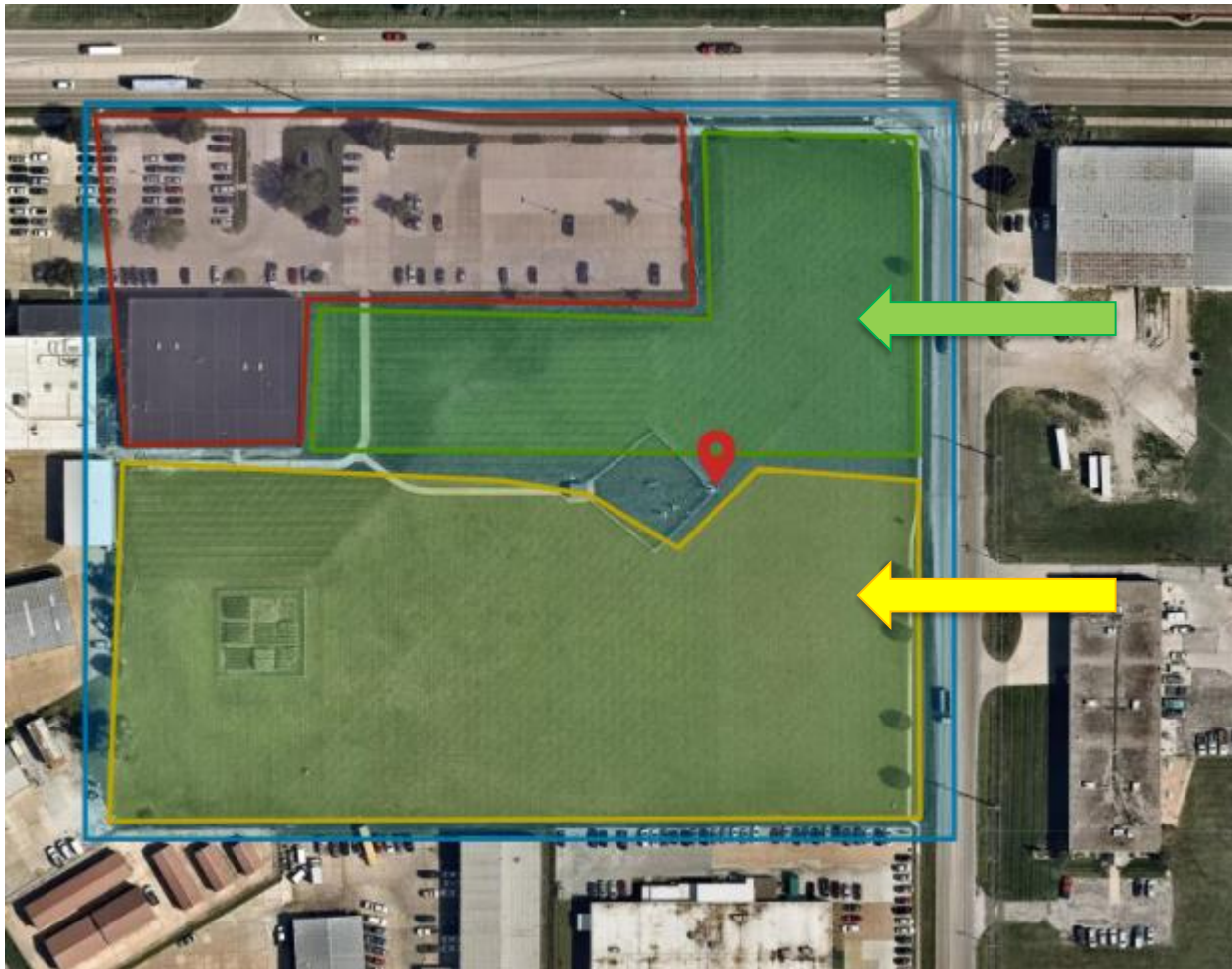
## Area of Interest

- This can be areas such as water bodies, different crop types, or simply separating the north and south side of a location.
- These are collection of areas where the computed footprint contribution is summed to give one value which represents the entire region of interest
- Enables classification of fluxes based on footprint and allows creation of mean diurnal flux computations for each area of interest
- Multiple area of interests can be defined

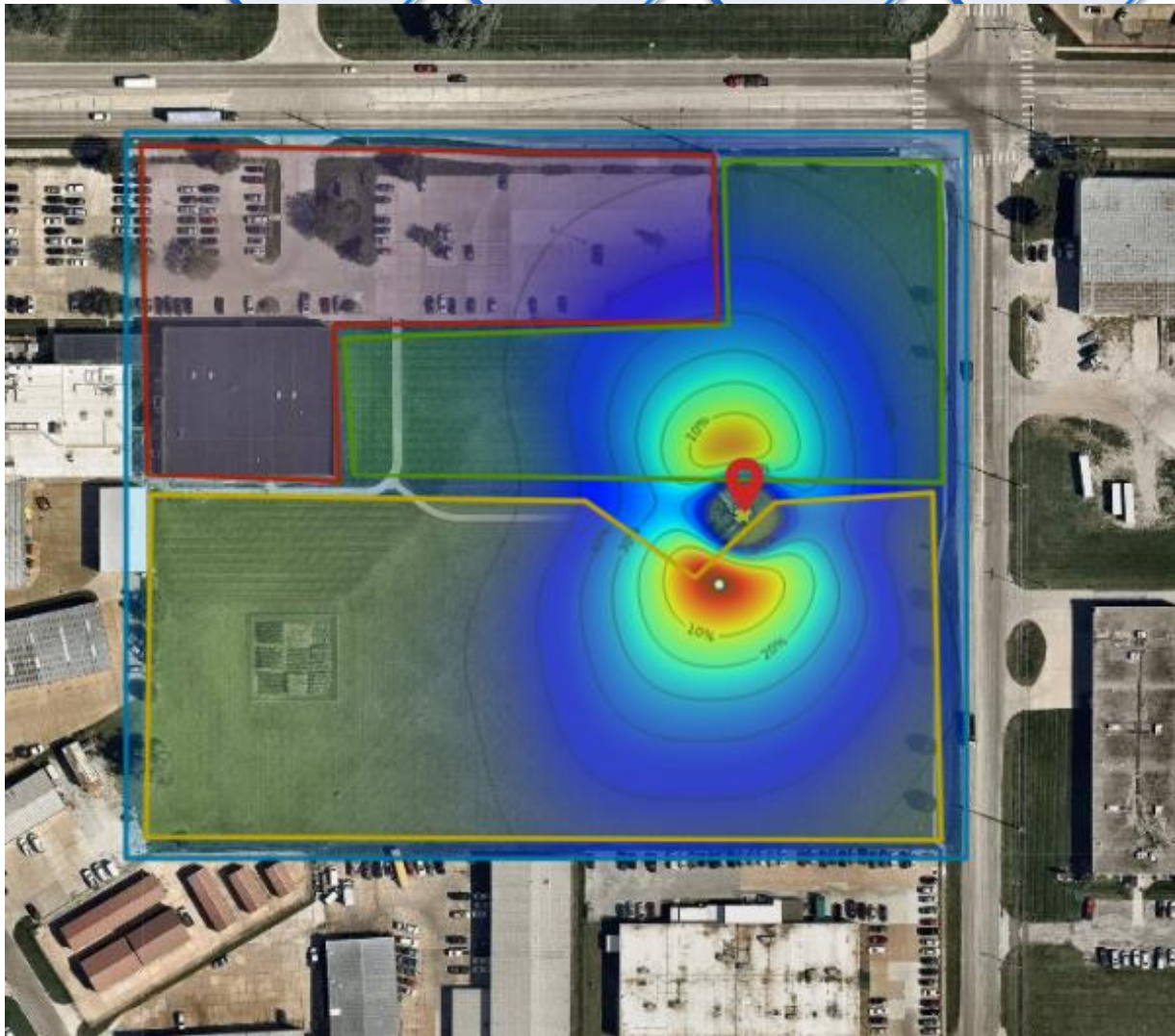
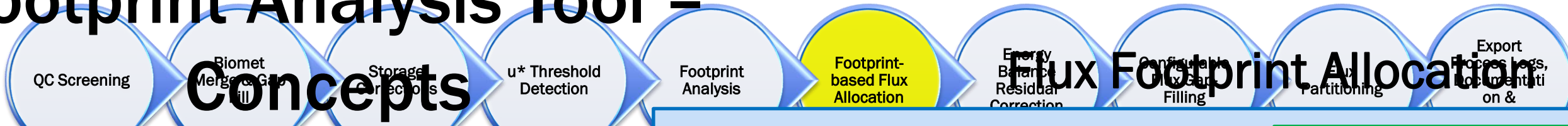




## Draw 'Areas of Interest' in Tovi (ex; different crops, ecosystems, etc.)



# Footprint Analysis Tool -



### 1/2-hour "Y"

- 75% from North area
- 10% from South area
- 15% from unassigned area

### 1/2-hour "X"

- 60% from North area
- 0% from South area
- 40% from unassigned area

NORTH

H

**SOUTH**

H

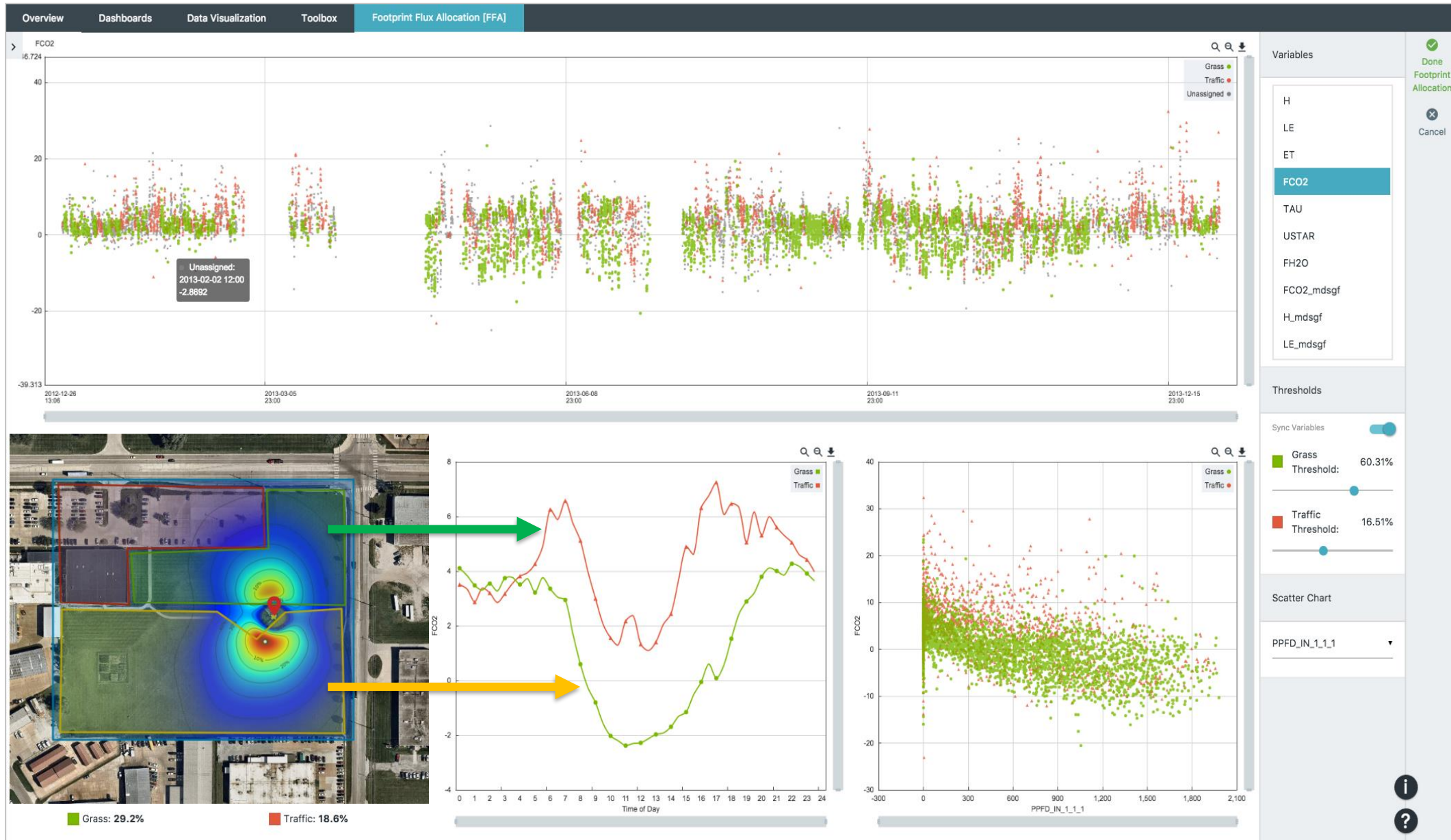
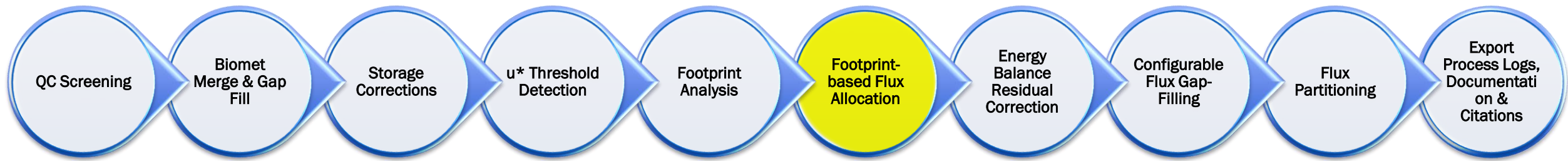
### 1/2-hour "Z"

- 5% from North area
- 80% from South area
- 15% from unassigned area

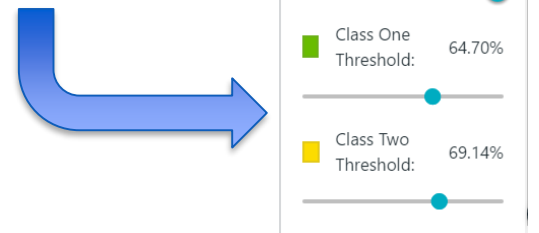
Sync Variables

Class One Threshold: 64.70%

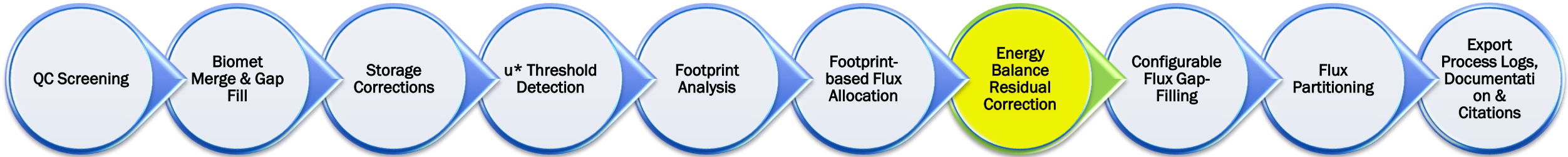
Class Two Threshold: 69.14%



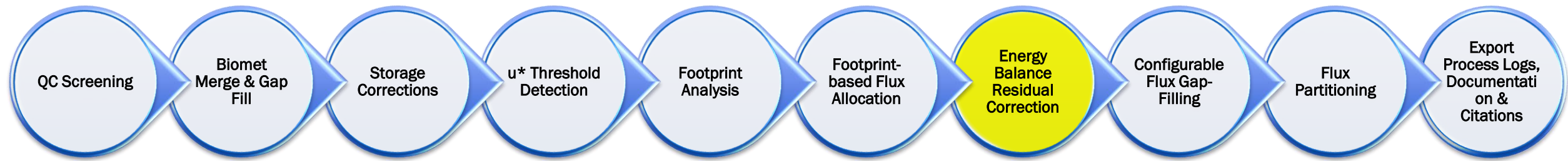
- Ability to classify each individual 30-minute flux based on Flux Footprint
- Visualize average diurnal flux patterns from different regions of interest
- Develop relationship between fluxes and environmental drivers for different regions of interest
- Customize your footprint thresholds



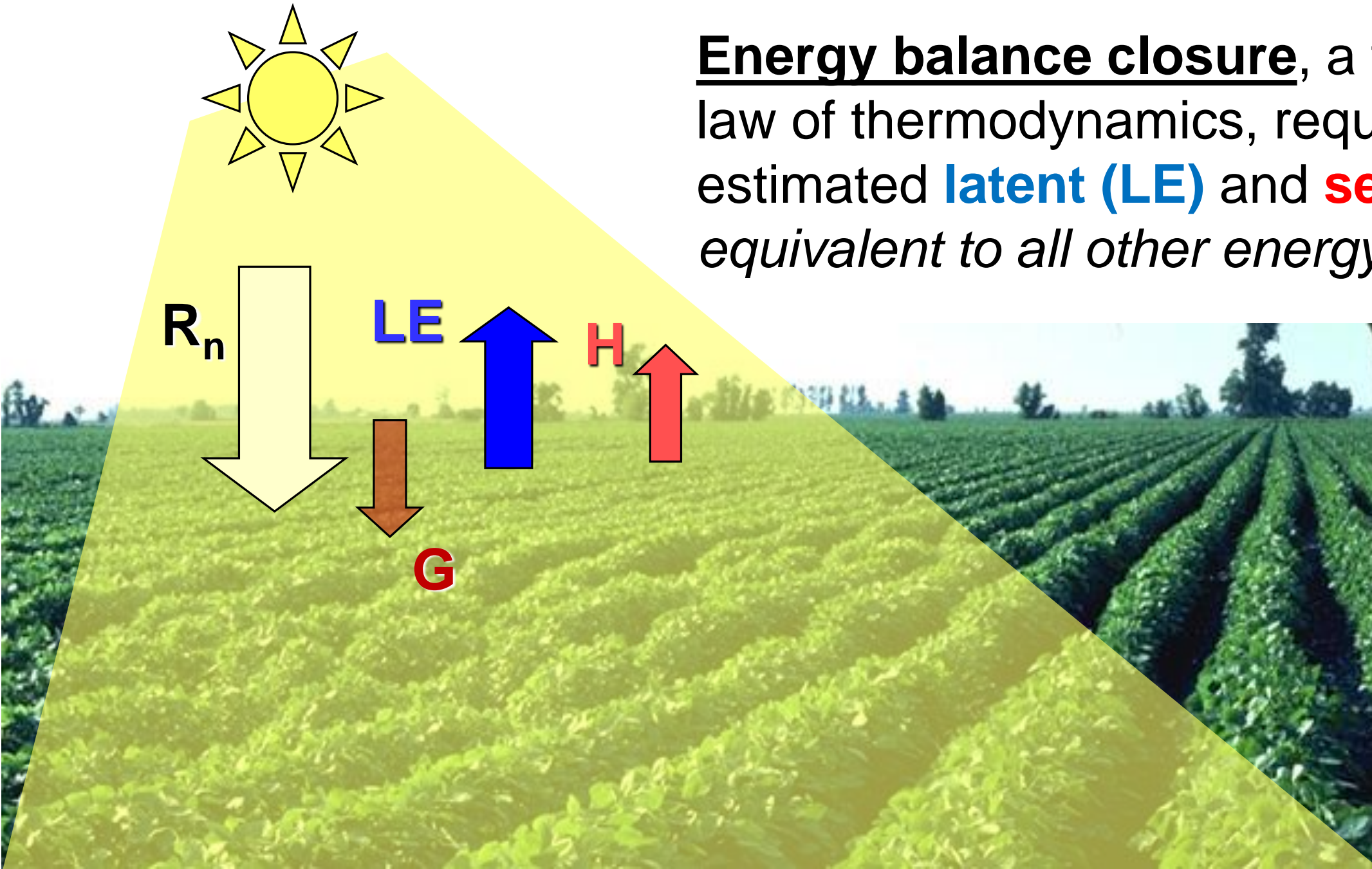
# Typical workflow for Post Processing...







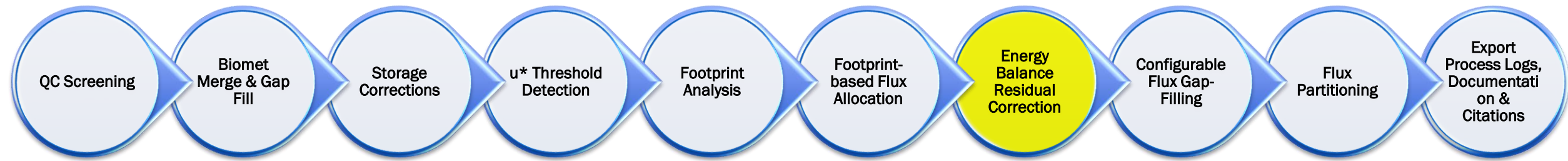
**Energy balance closure**, a formulation of the first law of thermodynamics, requires that the sum of the estimated **latent (LE)** and **sensible (H)** heat flux be *equivalent to all other energy sinks and sources*



$$R_n = H + LE + G + S + Q$$

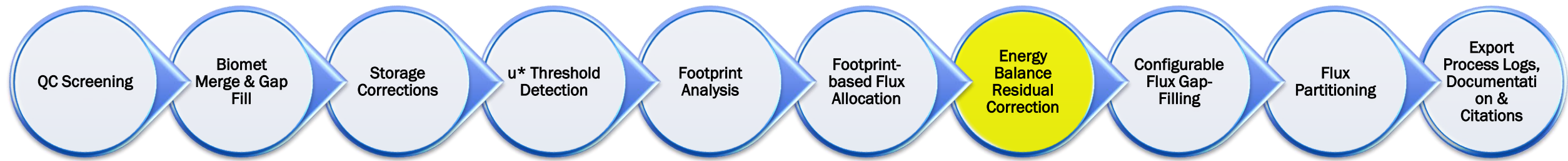


$$H + LE \approx R_n - G$$



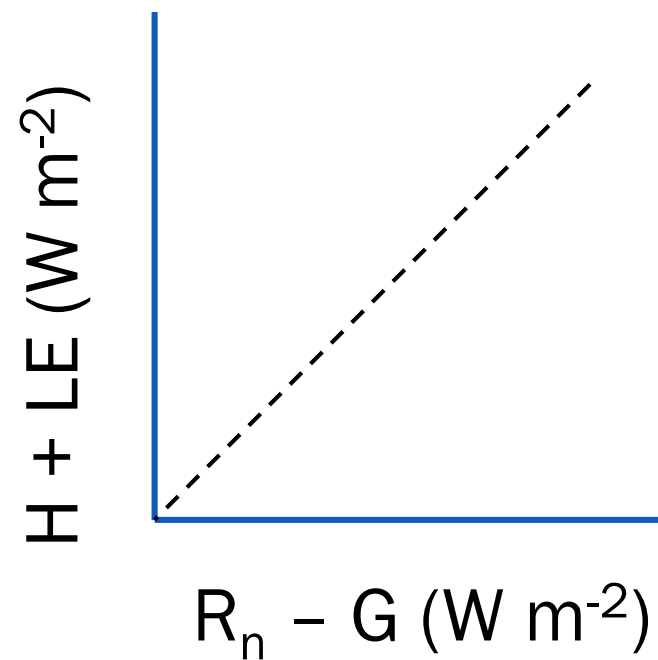
- Energy Balance can act as a tool for verifying eddy covariance instrumentation ( $\text{CO}_2/\text{H}_2\text{O}/\text{CH}_4$  analyzers and sonic anemometers) are working accurately and are installed properly.
  - In turn, this helps to verify that the final computed flux values are correct and accurate.
- Investigate relationships between half-hourly estimates of dependent flux variables ( $LE + H$ ) against independently derived available energy ( $R_n - G$ ).

**Remember ...  $H + LE \approx R_n - G$**



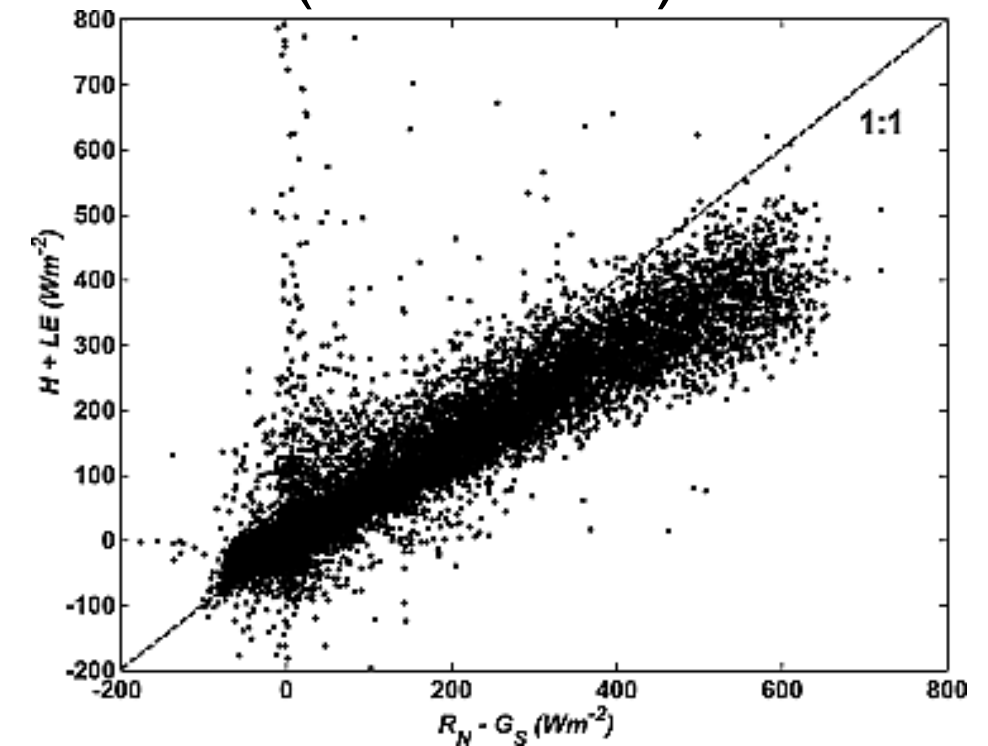
**Ideal closure** is represented by a slope of 1 and an intercept of 0.

Measured by the **EC System**



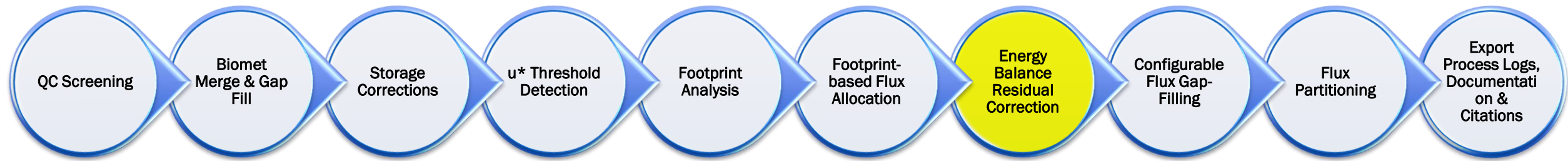
Measured by the **Biomet System**

**Realistic (measured) closure**



**If not ideal:**

- Sampling errors?
- Systematic biases?
- Neglected energy sinks?
- Other?



- Remember the reasons the Energy Sources are NOT balanced?

Cause of imbalance

Examples

Rn, G, H, LE

Sampling

Source areas differ

Rn, G

Instrument bias

Net radiometer biased

Neglected energy sinks

Storage above soil heat plates

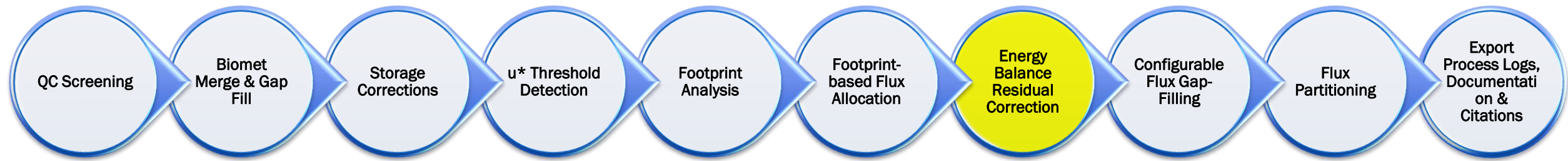
H, LE

High/low frequency loss

Sensor separation/large eddies

Advection

Regional circulation



- Let's assume the Biomet Sensors and measurements are 'good', and look at the measurements of H & LE...

Cause of imbalance

Examples

Sampling

Source areas differ

Instrument bias

Net radiometer biased

Neglected energy sinks

Storage above soil heat plates

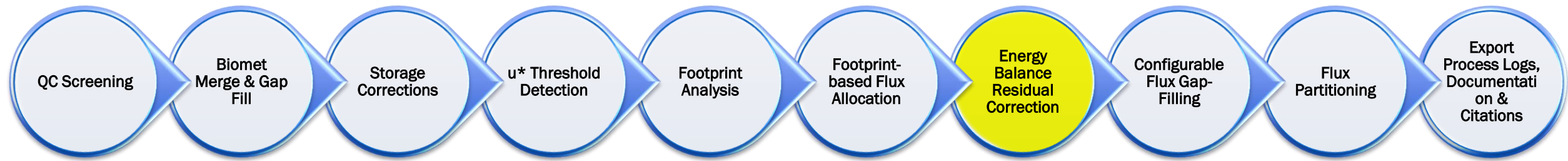
H, LE

High/low frequency loss

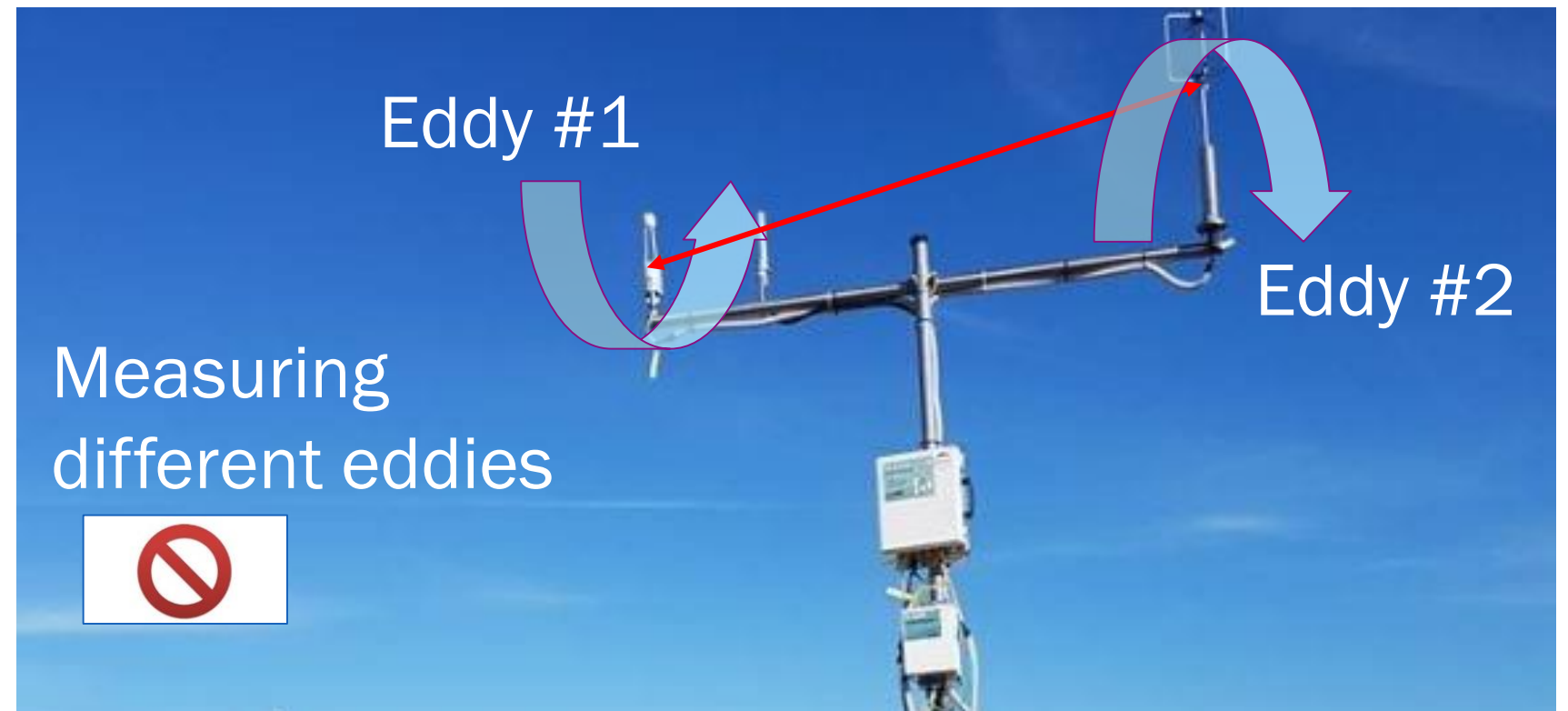
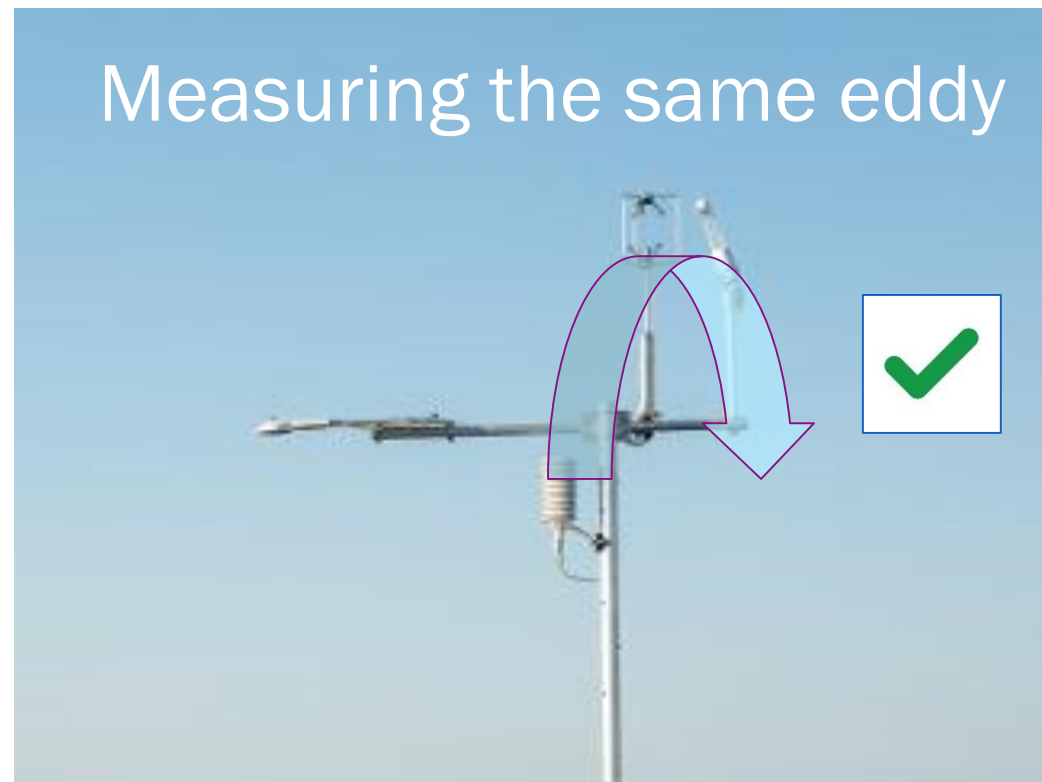
Sensor separation/large eddies

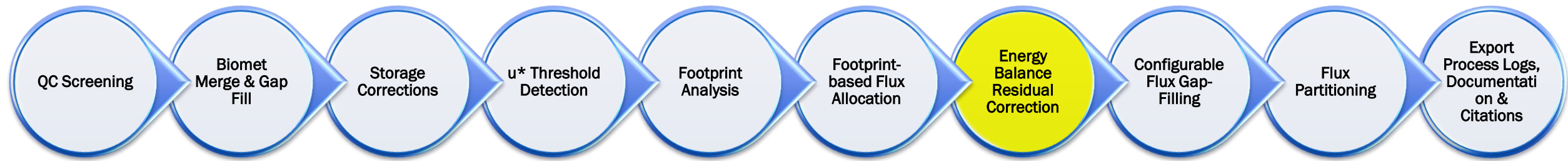
Advection

Regional circulation

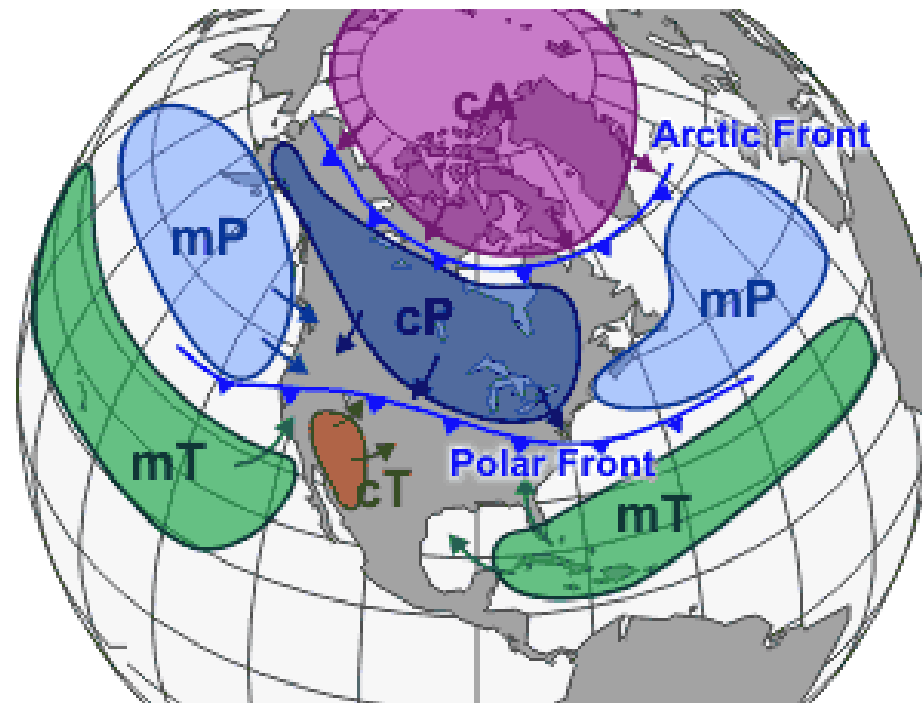


- Sensor Separation (High-Frequency Loss): If the instrument **separation** is too big to measure the same eddy, covariance is lost.
- H (anemometer) and LE (analyzer) are not from the same eddy, it will potentially create an imbalance in the energy closure

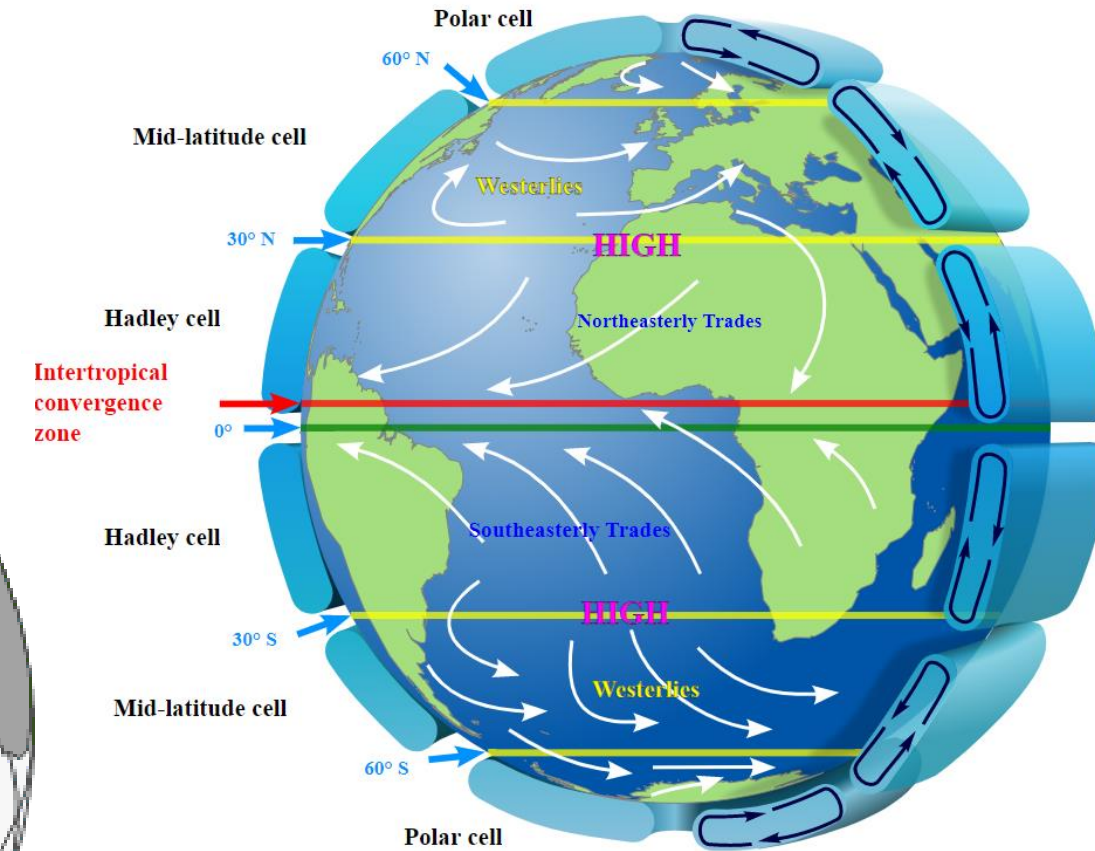




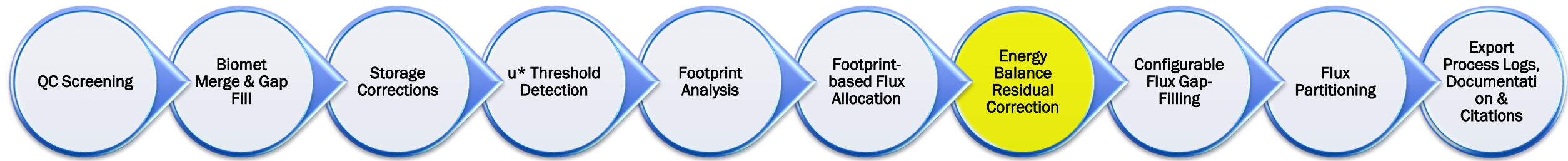
- Low-Frequency Loss: Large (synoptic) Scale, slow-moving structures transport portions of H and LE on scales longer than 30 minutes, creating residual energy terms



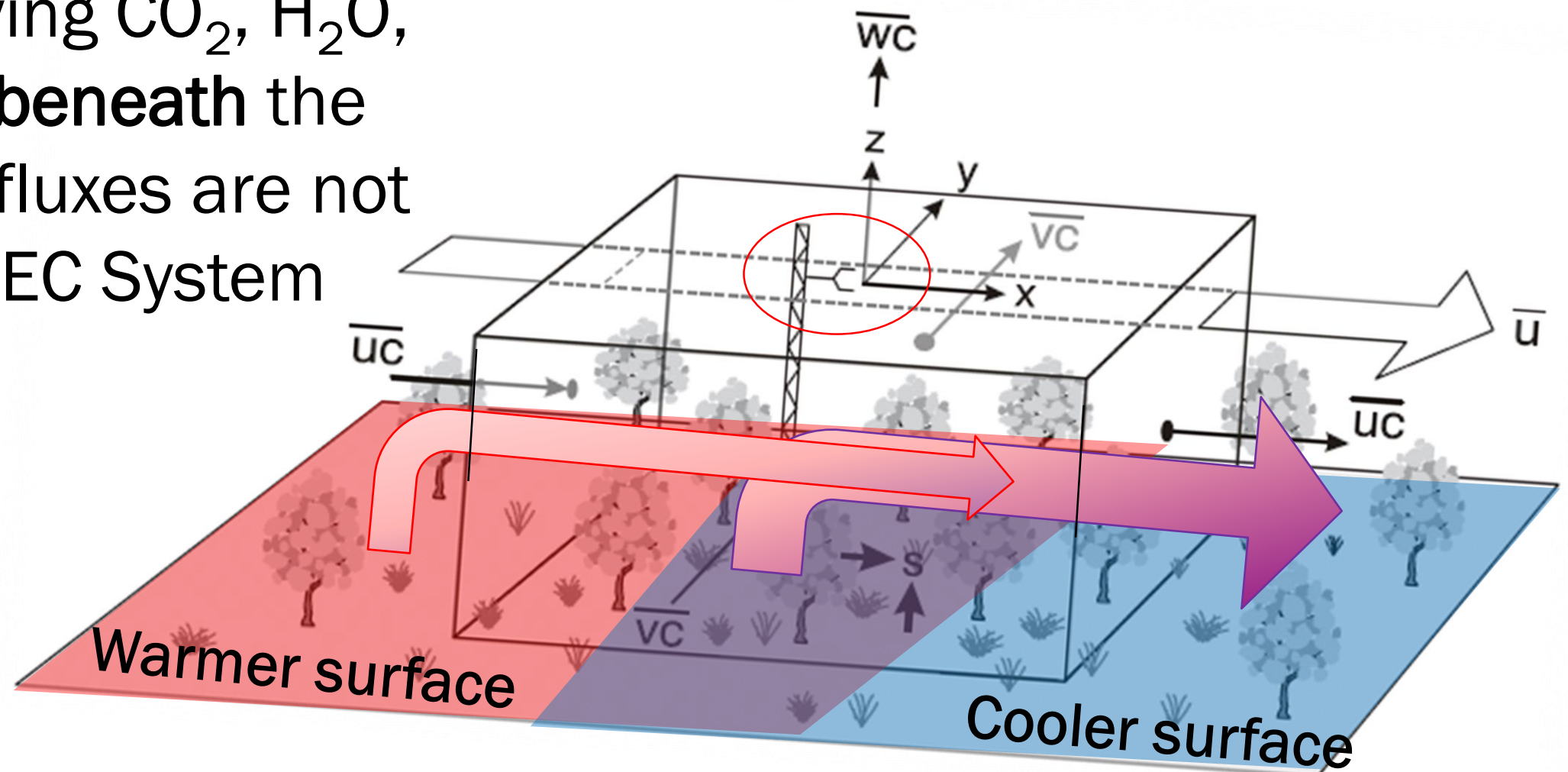
By The original uploader was Gopher backer at English Wikipedia. - Transferred from en.wikipedia to Commons by Undead\_warrior., Public Domain, <https://commons.wikimedia.org/w/index.php?curid=34671644>



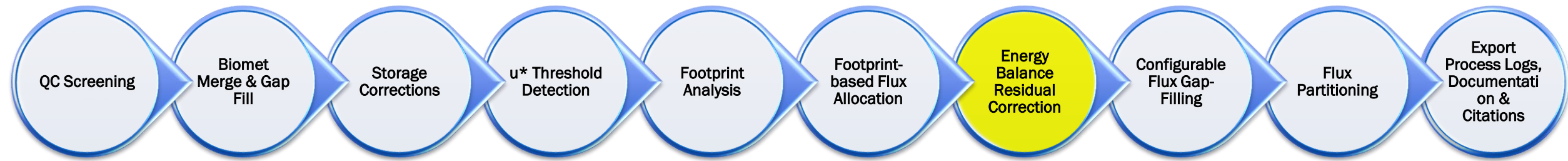
By Kaidor - Own work based on File:NASA depiction of earth global atmospheric circulation.jpg, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=23902538>



- Advection: Air mass slowly moves horizontally, carrying  $\text{CO}_2$ ,  $\text{H}_2\text{O}$ , and Heat Energy **beneath** the tower, and these fluxes are not measured by the EC System



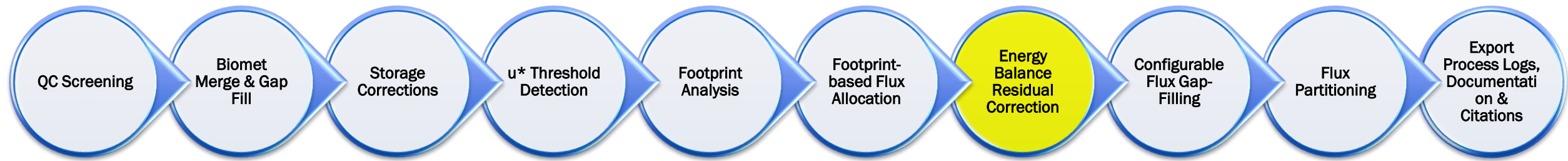




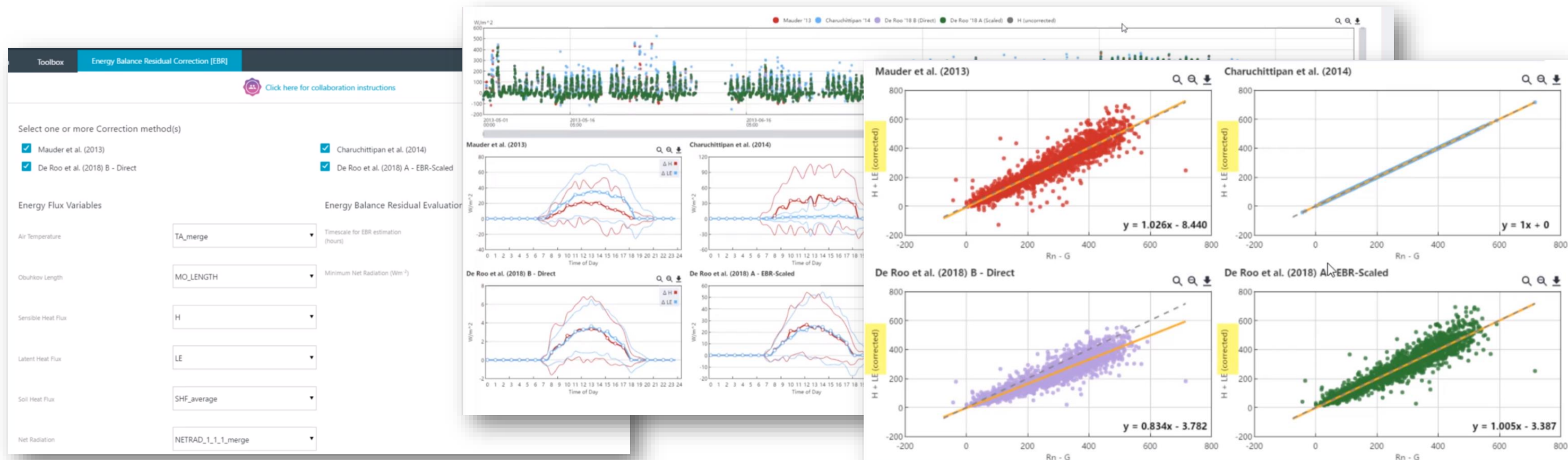
- Non-closure of the surface energy balance is a frequently observed phenomenon of hydrometeorological field measurements, when using the eddy covariance method, which can be ascribed to an underestimation of the turbulent fluxes (Mauder et al., 2018).
- The available four methods under consideration are:
- **Mauder et al., (2013)** forces the EBC on a daily time scale and partitions residual energy in such a way as to preserve the Bowen Ratio ( $H/LE$ ) before and after the correction on a 30-minute basis.
- **Charuchittipan et al., (2014)** forces the EBC on a 30-minute basis and attributes a larger portion of the residual energy to  $H$ , and hence significantly increases the Bowen Ratio for most periods. The rationale is that the large-scale coherent structures not captured by the eddy covariance method are driven by buoyancy and hence entail primarily additional sensible heat fluxes.
- **De Roo et al. (2018)** proposes two methods. With both methods, informed by a Large Eddy Simulation (LES) study, the partition between  $H$  and  $LE$  varies with the measurement height and attributes a larger portion of the energy to  $H$  with respect to  $LE$ . The two methods differ for the fact that one is purely derived from the LES results without any consideration of the observed non-closure, while the second is constrained by the non-closure evaluated on a daily basis.



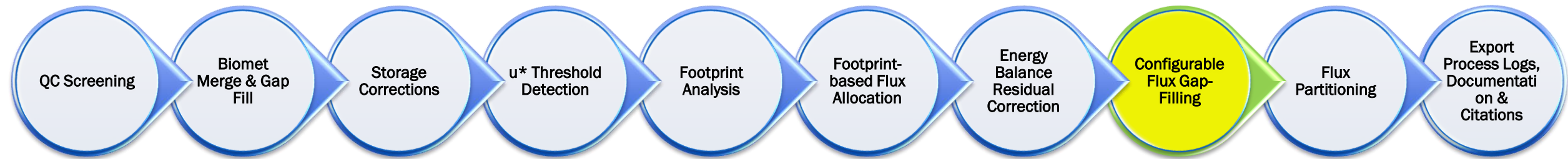
Dr. Matthias Mauder  
 Karlsruhe Institute of Technology (KIT),  
 Institute of Meteorology and Climate Research  
 Atmospheric Environmental Research  
 (during the collaboration)

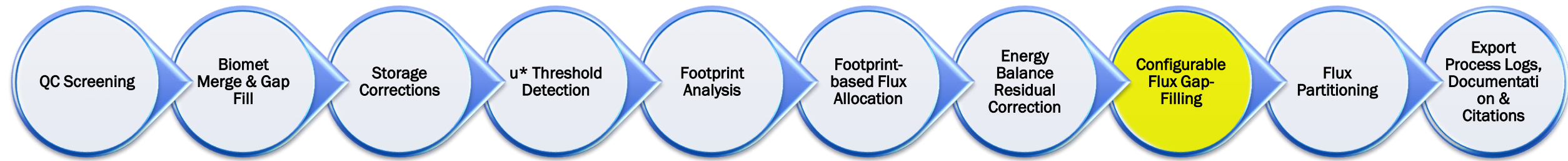


- Example, investigating the four methods for correcting the energy balance (H and LE) in Tovi

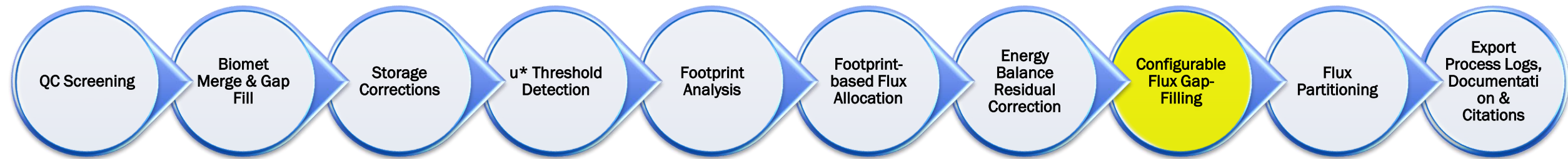


# Typical workflow for Post Processing...



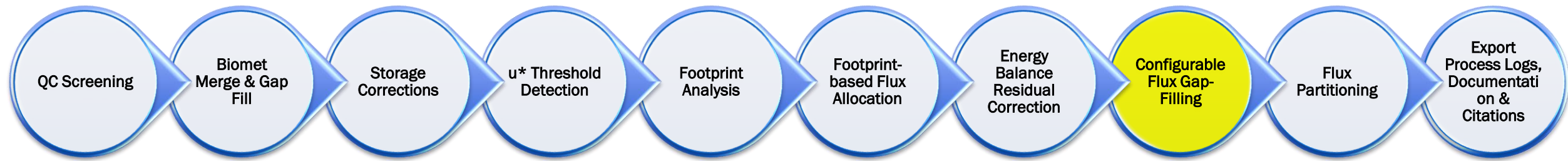


- Gap-filled time series are required if site data are to be used by modelers for parameterization and validation of land surface models and to construct annual sums of carbon and water exchange between the land surface and the atmosphere.
- The eddy covariance method attempts to deliver continuous data sets. However, gaps are created due to unfavorable micro-meteorological conditions and due to instrument failure.
- A standardized filling of those gaps is necessary (gap-filling), e.g. to obtain daily, monthly or annually integrated balances.
- To calculate aggregated values, for example, sums to estimate annual budgets or daily averages needed in model evaluations, the completeness of the data set is required.
- If missing and rejected values in the half-hourly data set would be perfectly random distributed, the calculation of an integrated value could be easily performed by taking the average of all available data.
- Due to its high temporal resolution, the EC technique provides a large amount of data that are often acquired under similar situation in terms of vegetation status and meteorological conditions. This “redundancy” of data is fundamental in the gap-filling methods.
- Unfortunately, data gaps do not occur randomly. For example,  $u^*$  filtering removes mainly nighttime data or power failures occur principally in winter and night when the solar panels are used. This non-randomness of the gaps in the data set leads to the need to apply more sophisticated gap-filling methods to reconstruct the missing periods.
- All of them (except Mean Diurnal Variation) however require as input meteorological variables that for this reason should be available as continuous and gap-free data set.



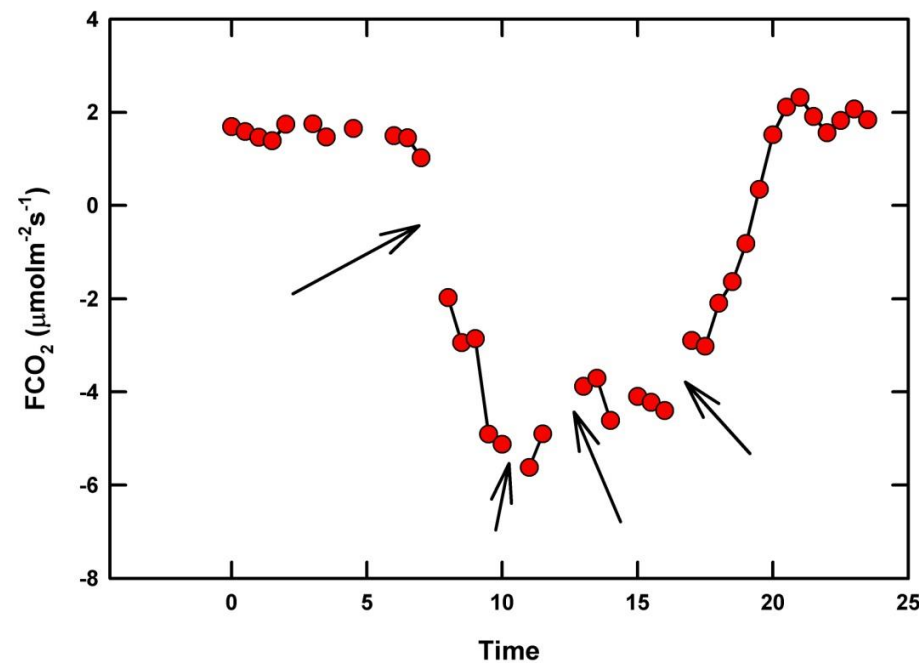
### Five typical methods for Gap Filling fluxes

- **Mean Diurnal Variation method (MDV)** - an interpolation technique based on the temporal auto-correlation of the fluxes (Falge et al. 2001). Missing observations are replaced by the mean of valid values measured on adjacent days at the same time. The averaging period (window) can vary, but a window < 2 weeks is recommended. The MDV method does not require drivers. The implementation is easy, but the accuracy and performances are lower with respect to the other methods
- **Look-up table (LUT)** - an empirical method, easy to implement, where the missing values are replaced with the average of valid measurements occurring under similar meteorological conditions. A multidimensional table is created where the missing value can be “looked up” based on the values of the meteorological drivers. Tables are created for the different seasons. Missing data points are replaced with the NEE value in the table with same combination of drivers. In general, 3–4 variables selected among incoming and diffuse radiation, air and soil temperature, soil water content, and vapor pressure deficit are sufficient.
- **Marginal Distribution Sampling (MDS)** - proposed by Reichstein et al. (2005), they consider both the covariation of fluxes with meteorological variables and the temporal auto-correlation of the fluxes. Similar meteorological conditions are sampled in the temporal vicinity of the gap in a window as small as possible. Incoming radiation, air temperature, and vapor pressure deficit are first considered; then the next step could be to increase the size of the window or reduce the numbers of drivers considered.
- **Artificial Neural Networks (ANNs)** are purely empirical nonlinear regression models with a medium level of implementation difficulties. The ANN consists in a set of nodes, often organized in layers and connected by weights (Bishop 1995; Rojas 1996). The first step to use an ANN is the network parameterization process called “training.” The ANN is trained by presenting it with sets of input data (drivers) and associated output data (valid fluxes). Once the ANN is trained, driver variables are weighted. The ANNs require a gap-free driver dataset and for this reason it is needed to first gap-fill the meteorological variables
- **Nonlinear regressions method** - based on parameterized non-linear equations which express semi-empirical relationships between the flux and environmental variables, often temperature and light for CO<sub>2</sub> fluxes. In general two different equations are used, one for nighttime data (as function of temperature) and one for daytime data (using a light response function). Both the equations have temperature as a driver and can use either air or soil temperatures.

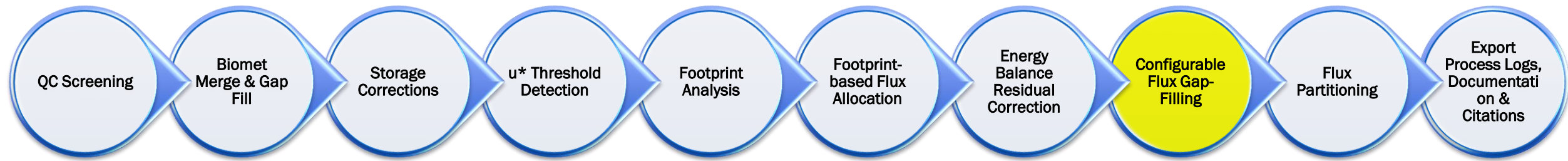


## Five typical methods for Gap Filling fluxes

- (1) **Mean Diurnal Variation method (MDV)** - an interpolation technique based on the temporal auto-correlation of the fluxes (Falge et al. 2001). Missing observations are replaced by the mean of valid values measured on adjacent days at the same time. The averaging period (window) can vary, but a window < 2 weeks is recommended. The MDV method does not require drivers. The implementation is easy, but the accuracy and performances are lower with respect to the other methods

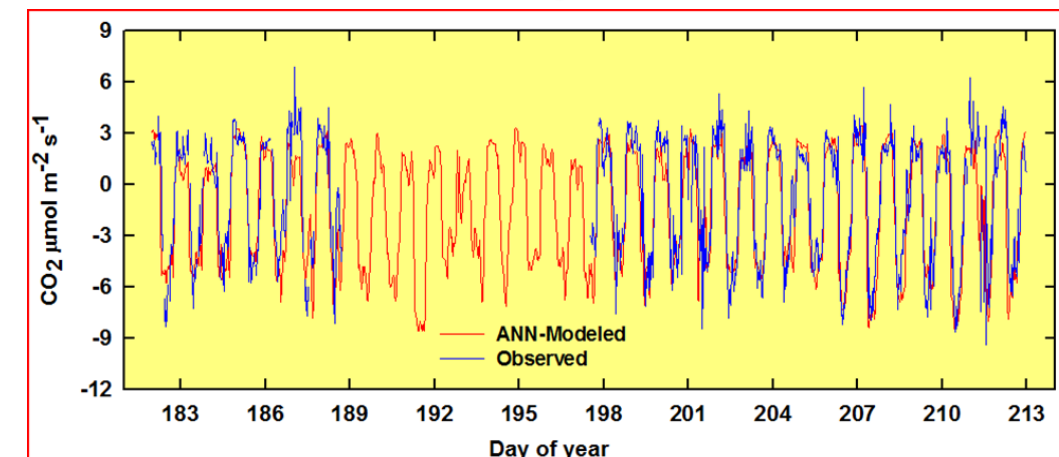
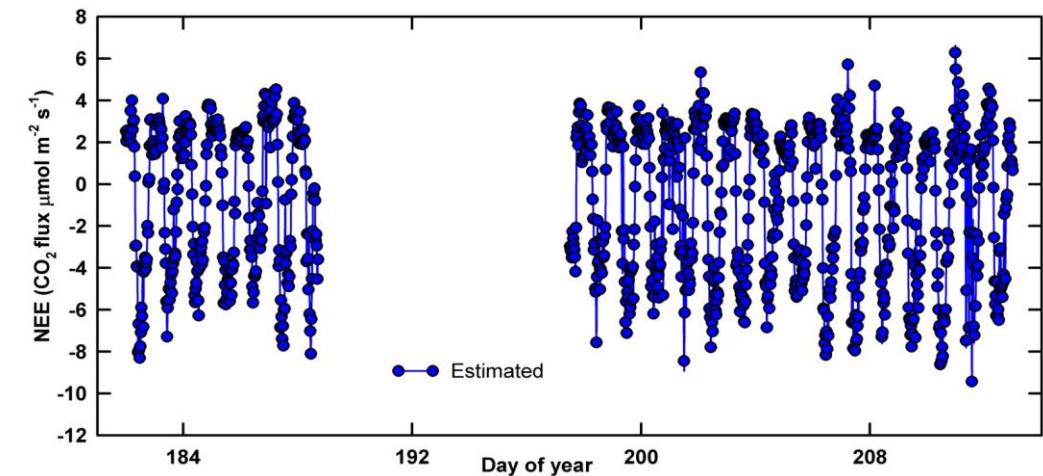


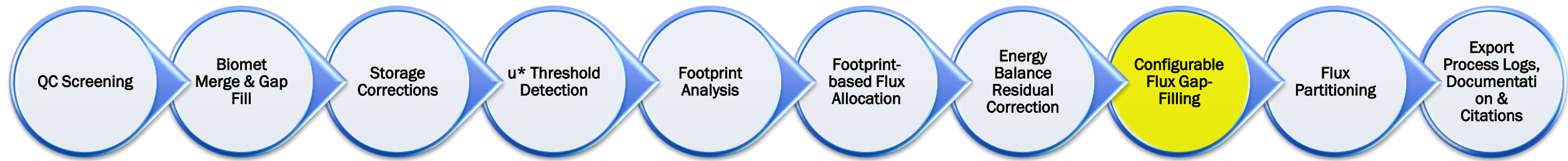
- Gaps occurs due to sensor failure
- Power supply issues
- Data flagged for bad quality
- Spikes in data due to rain events
- Data flagged for low  $U^*$



Five typical methods for Gap Filling fluxes (continued):

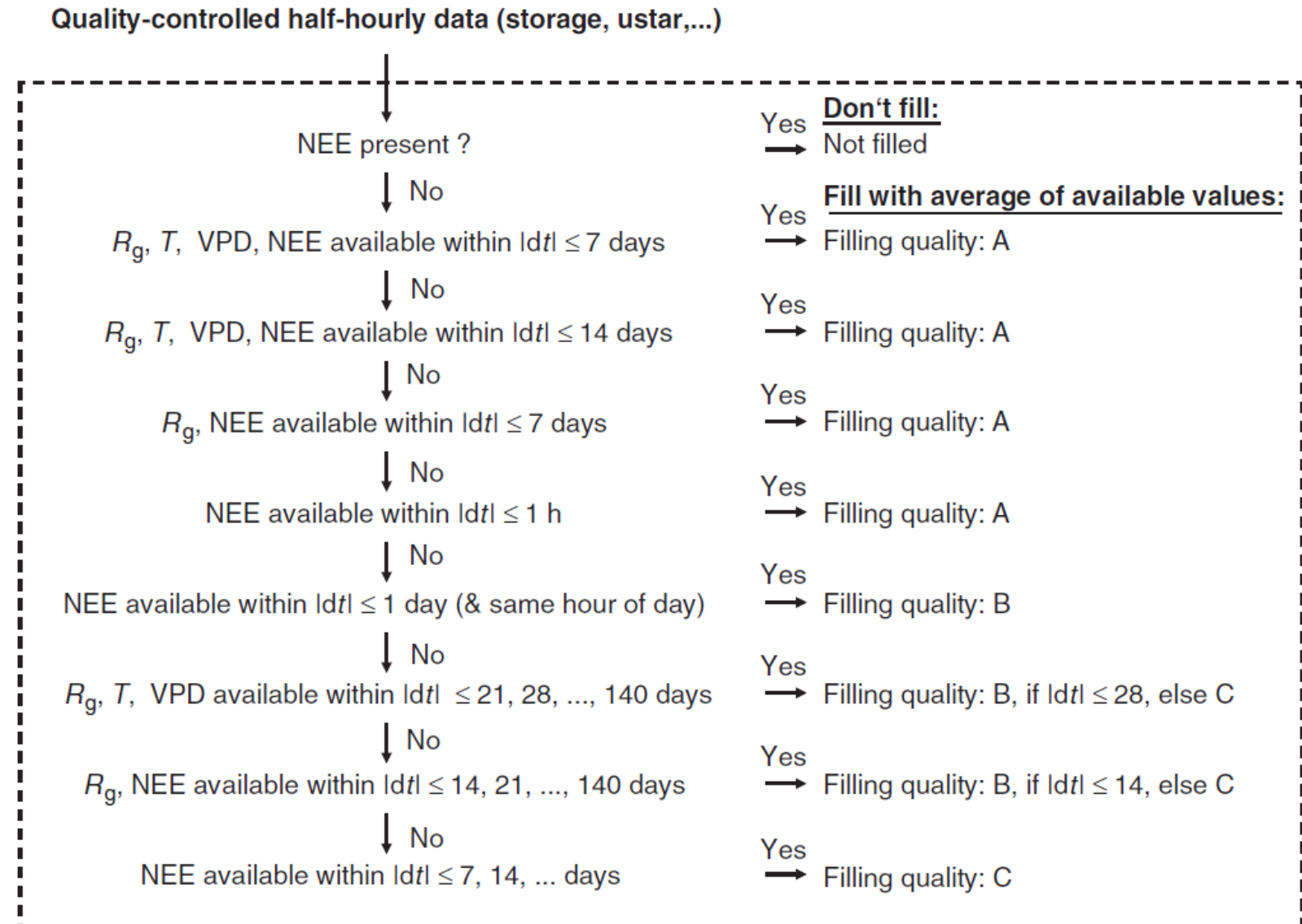
- **(2) Look-up table (LUT)** - an empirical method, easy to implement, where the missing values are replaced with the average of valid measurements occurring under similar meteorological conditions. A multidimensional table is created where the missing value can be “looked up” based on the values of the meteorological drivers. Tables are created for the different seasons. Missing data points are replaced with the NEE value in the table with same combination of drivers. In general, 3-4 variables selected among incoming and diffuse radiation, air and soil temperature, soil water content, and vapor pressure deficit are sufficient.
- **(3) Artificial Neural Networks (ANNs)** are purely empirical nonlinear regression models with a medium level of implementation difficulties. The ANN consists in a set of nodes, often organized in layers and connected by weights (Bishop 1995; Rojas 1996). The first step to use an ANN is the network parameterization process called “training.” The ANN is trained by presenting it with sets of input data (drivers) and associated output data (valid fluxes). Once the ANN is trained, driver variables are weighted. The ANNs require a gap-free driver dataset and for this reason it is needed to first gap-fill the meteorological variables



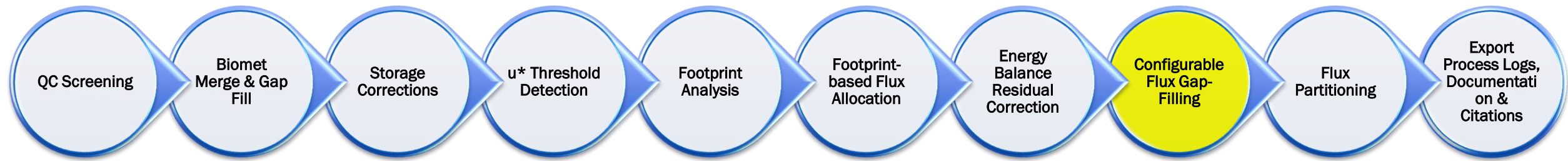


Five typical methods for Gap Filling fluxes (continued):

- (4) Marginal Distribution Sampling (MDS)**
  - proposed by Reichstein et al. (2005), they consider both the covariation of fluxes with meteorological variables and the temporal auto-correlation of the fluxes. Similar meteorological conditions are sampled in the temporal vicinity of the gap in a window as small as possible. Incoming radiation, air temperature, and vapor pressure deficit are first considered; then the next step could be to increase the size of the window or reduce the numbers of drivers considered.



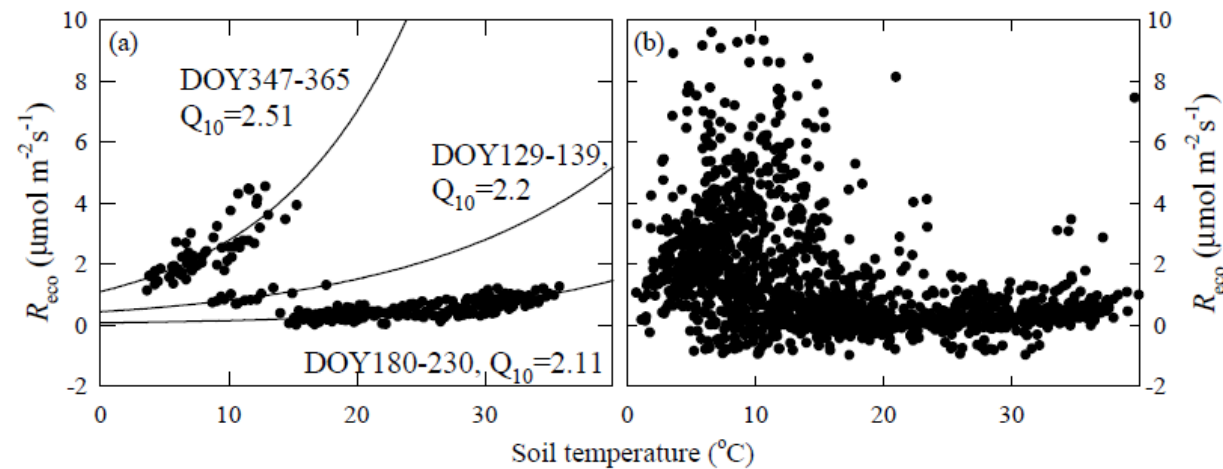




Five typical methods for Gap Filling fluxes (continued):

- (5) Nonlinear regressions method** - based on parameterized non-linear equations which express semi-empirical relationships between the flux and environmental variables, often temperature and light for CO2 fluxes. In general two different equations are used, one for nighttime data (as function of temperature) and one for daytime data (using a light response function). Both the equations have temperature as a driver and can use either air or soil temperatures.

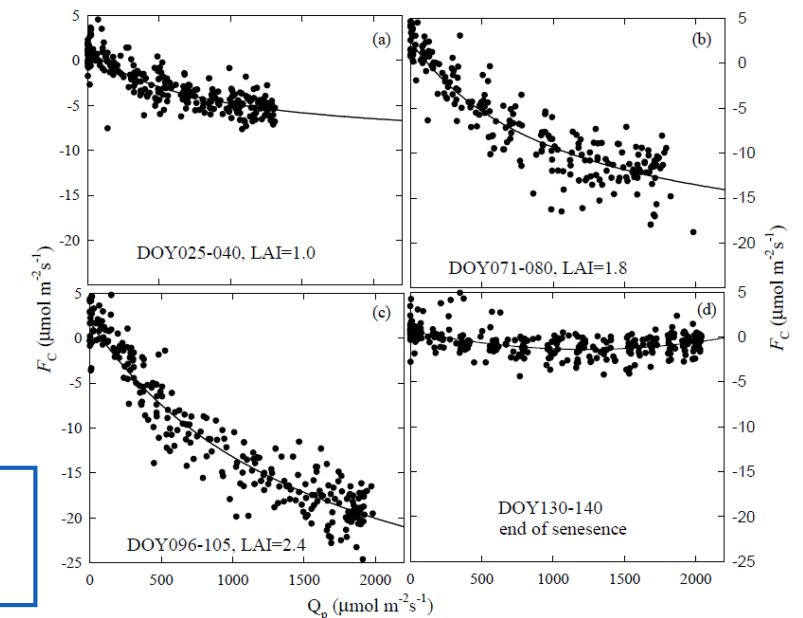
*L. Xu, D.D. Baldocchi/Agricultural and Forest Meteorology 1232 (2004) 79–96*



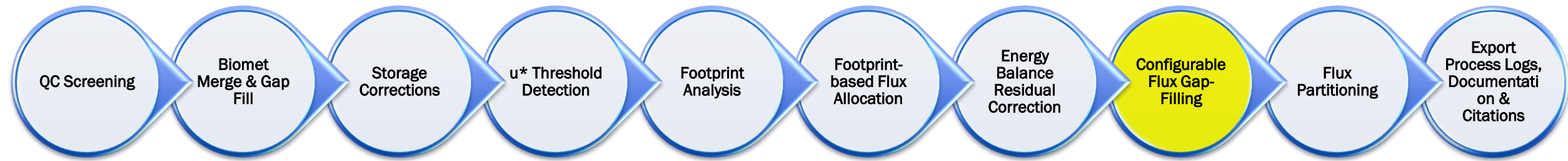
$$F_c = b_0 \exp(bT_{\text{soil}})$$

$$Q_{10} = \exp(10b)$$

*L. Xu, D.D. Baldocchi/Agricultural and Forest Meteorology 1232 (2004) 79–96*

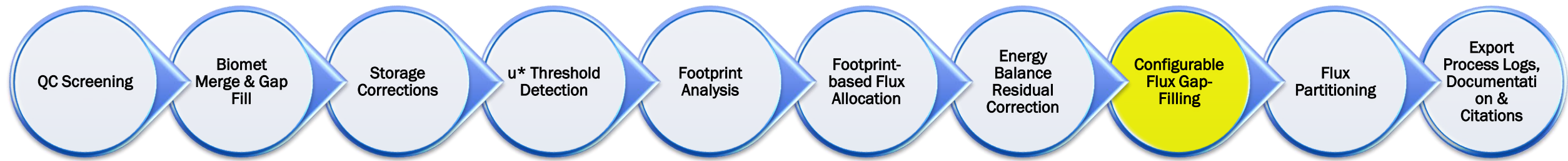


$$F_c = \frac{F_{\text{max}} \alpha \text{PAR}}{\alpha \text{PAR} + F_{\text{max}}} + R_{\text{eco}}$$

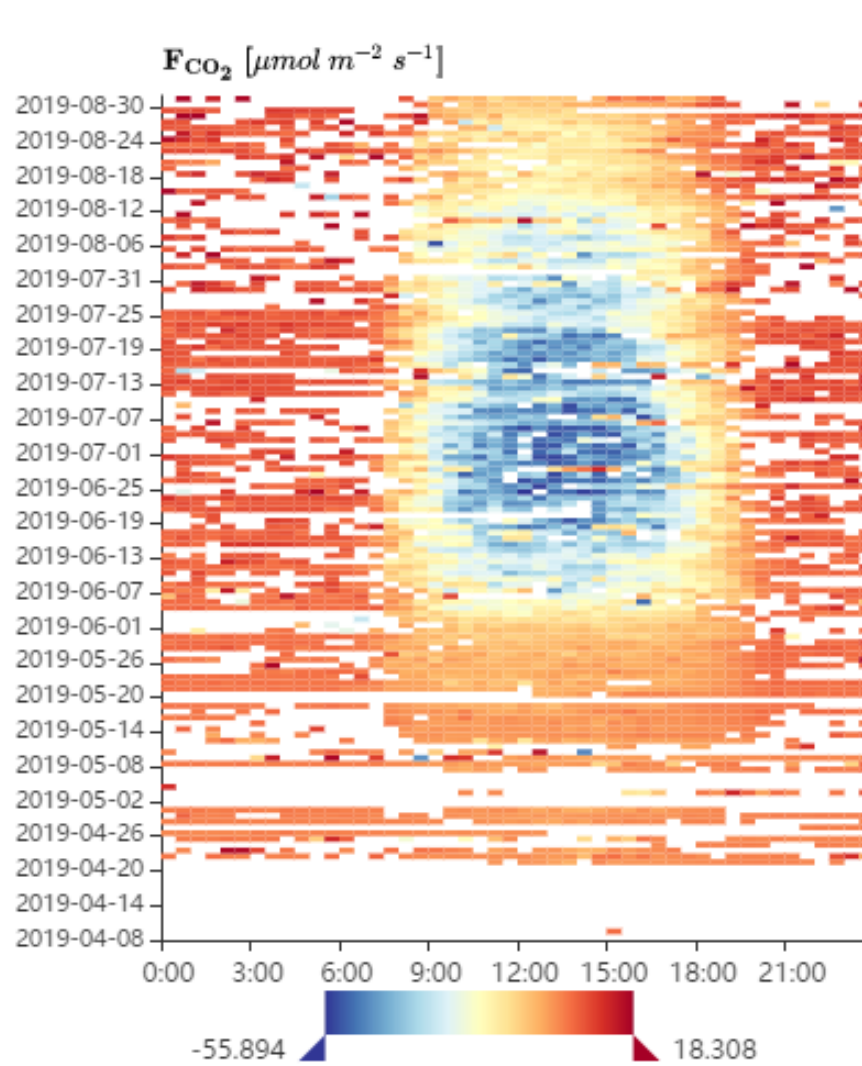
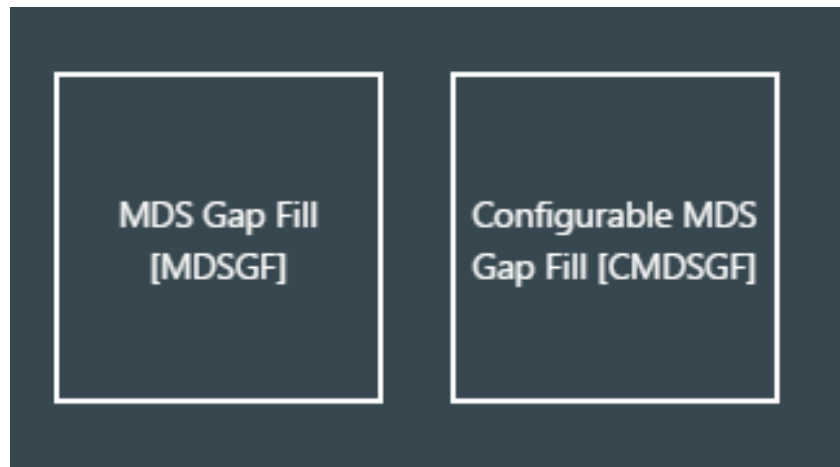


## Uncertainty and Quality Flags

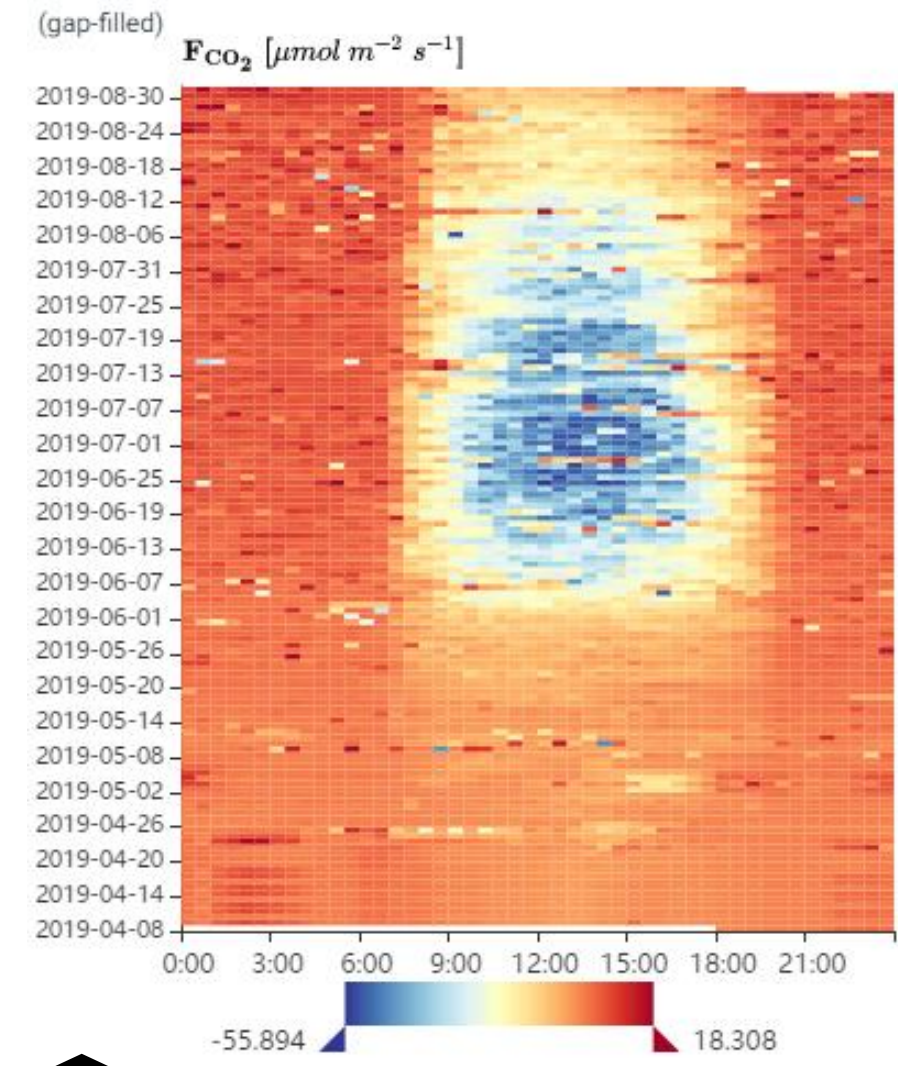
- Uncertainty estimation is an important information that should be always included when data are gap filled. There are two main different uncertainty sources in the gapfilled values:
  - The diverse estimations that different gap-filling methods give for the same missing data point
  - Uncertainty in the selected gap-filling model parameterization.
- The uncertainty related to the selection of the gap-filling method is relatively small when one of the high-performance methods is used if the gap length is not too long and if the data set available to set the gap-filling model parameters is sufficiently large and with of good quality.
- Long gaps, during which the general ecosystem conditions can change (e.g., growing season phase, ground water table, nutrients availability), are more difficult to fill and the uncertainty associated with the gap-filled values will be higher with respect to short gaps.
- It is fundamental to carefully register all the ancillary data about the site and to install a back-up, independent meteorological station close to the EC tower



- Example; two gap filling algorithms inside Tovi

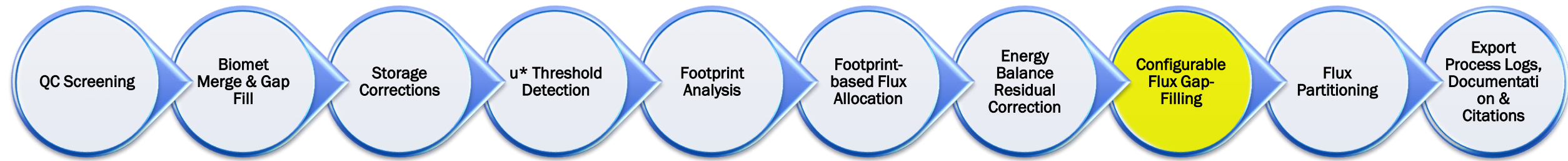


With gaps



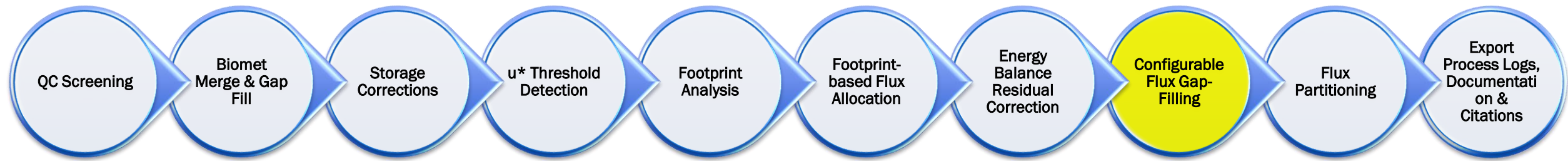
Gap Filled





## Marginal Distribution Sampling (MDS) Gap Filling in Tovi

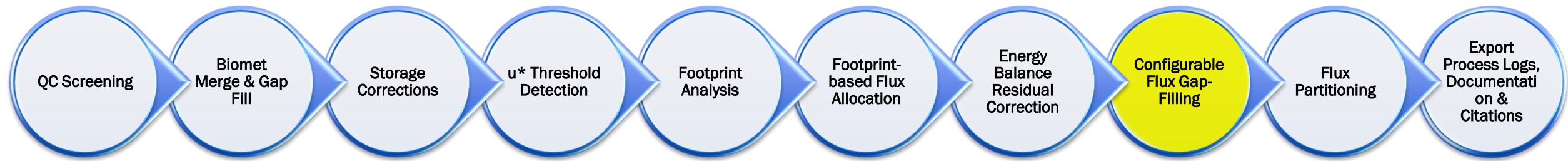
- Based on Reichstein *et al.* 2005
- Drivers: Shortwave incoming, VPD, and Air Temperature
- Combination of two methods
  - Look-Up Table
  - Mean Diurnal Variation/Course (MDV/MDC)
- Dynamic averaging window size - temporal auto-correlation of fluxes. (Performs gap filling operations on different windows around the gap.)



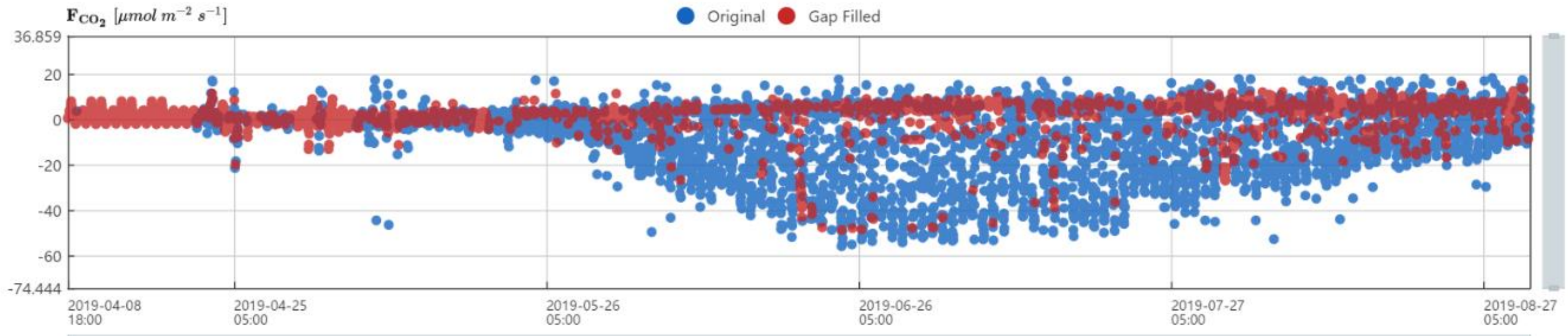
- Example, MDS Gap Filling in Tovi
- Similar Meteorological conditions exists when
  - SWIN is within  $20 \text{ W m}^{-2}$  in low light, or  $50 \text{ W m}^{-2}$  in high light
  - VPD does not deviate by 5 hPa
  - TA does not deviate by  $2.5 \text{ }^\circ\text{C}$

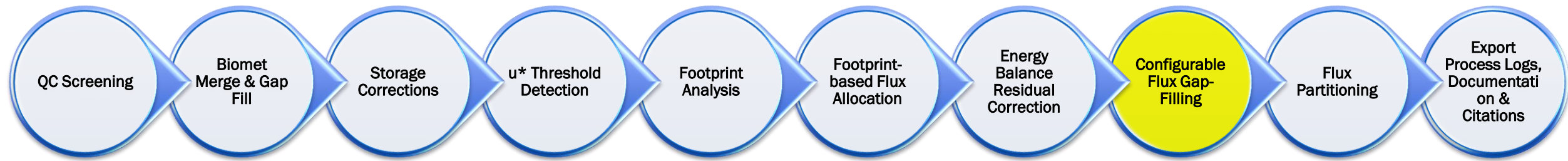
Drivers

Variable for Gap-Filling	Value	Tolerance
Solar Radiation ( $\text{W/m}^2$ )	SW_IN_1_1_1	Low: 20, High: 50
Water (Pa)	VPD_merge	5
Temperature ( $^\circ\text{C}$ )	TA_average_merge	2.5



- Example, **Results** of MDS Gap Filling in Tovi
- Tolerances are used to classify fluxes from similar meteorological conditions

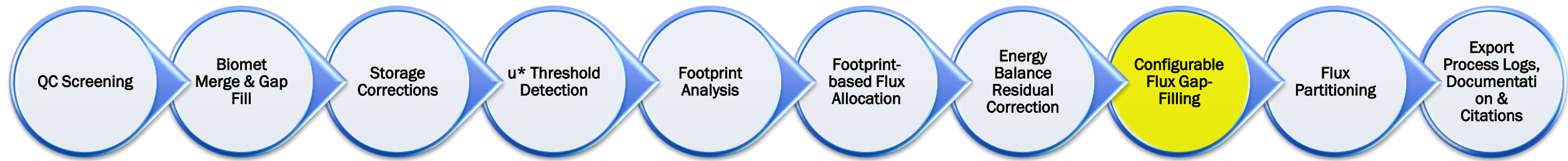




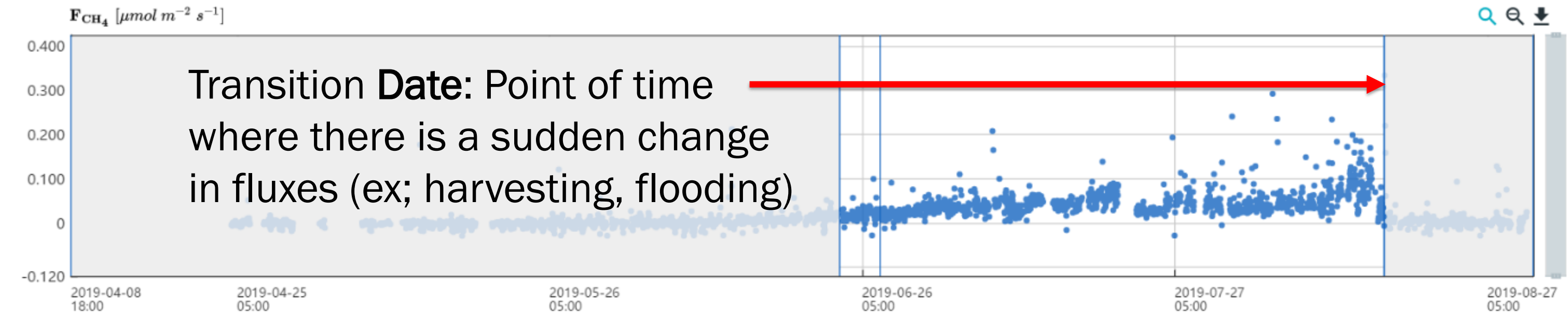
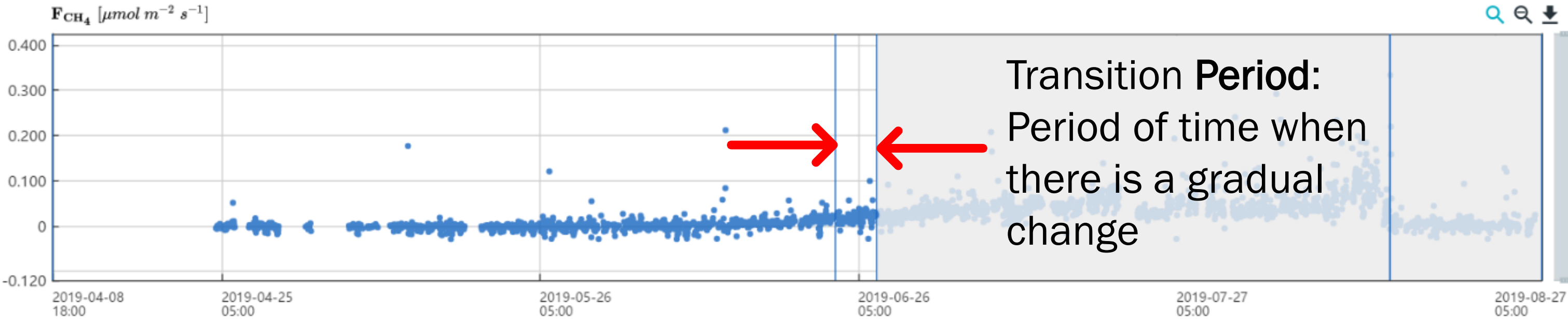
## NEW! Configurable MDS Gap Filling in Tovi only

- Gap fill any fluxes using MDS methodology
- Select your own drivers of choice
- Select transition periods and dates
- Choose different drivers for different transition periods
- Copy settings and drivers between fluxes
- Five options for filling fluxes via MDS during transition periods
  - Left window, Right Window, Weighted Mean by proximity, Mean, or Weighted mean by QC Flagging

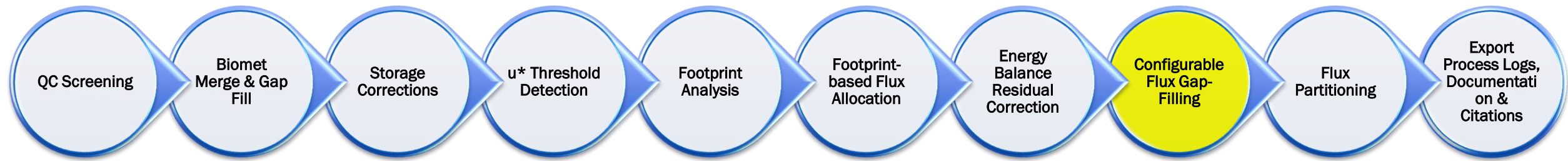




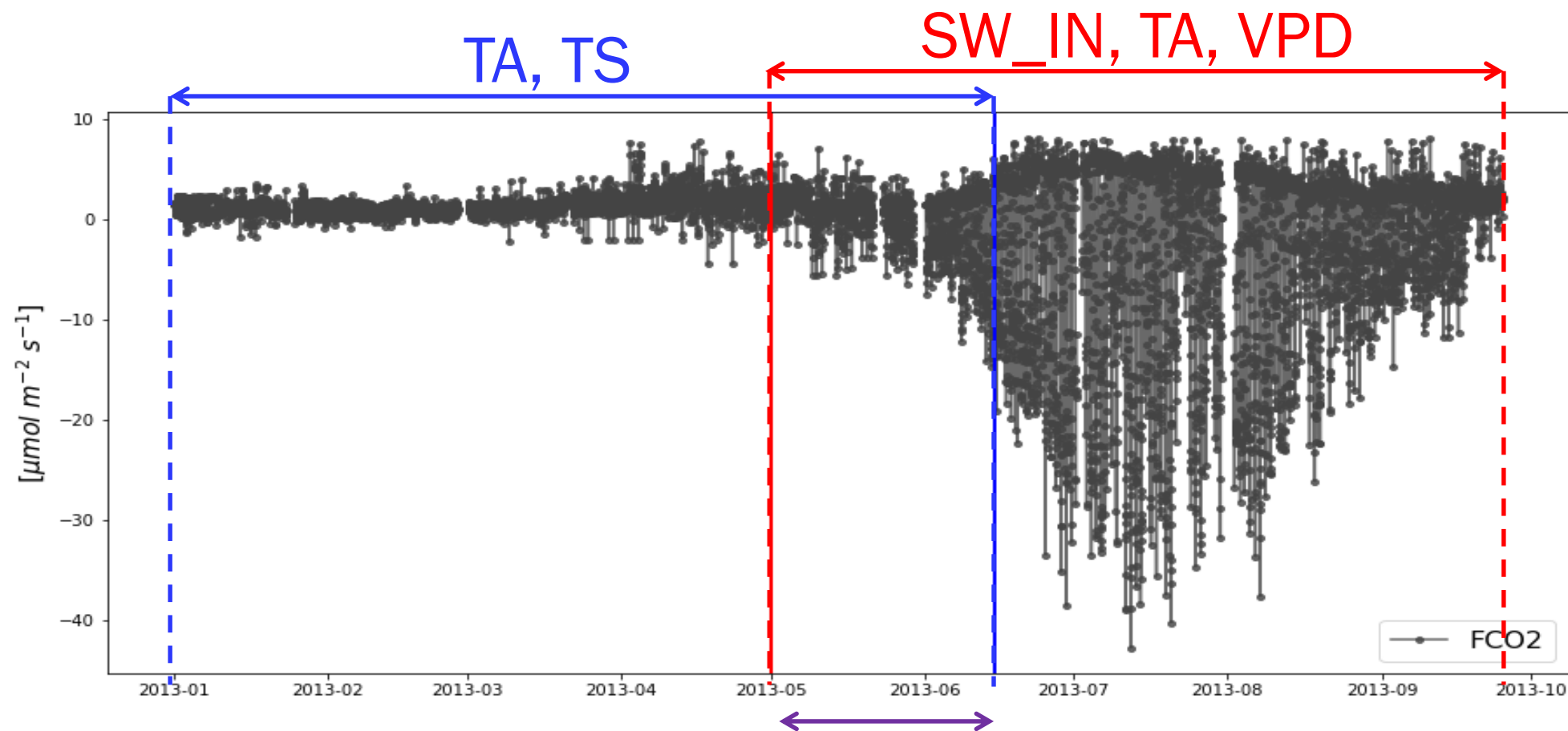
■ Example, using **Transitions in Tovi** for Gap filling







- Example, selecting **Drivers** in Tovi for Gap filling
  - Different drivers can be selected for different windows
  - Different drivers can be set up for daytime and nighttime



Drivers Settings for FCH4 - Time Window 2019/04/08 - 2019/08/31

Setup separately for daytime and nighttime

Primary Driver

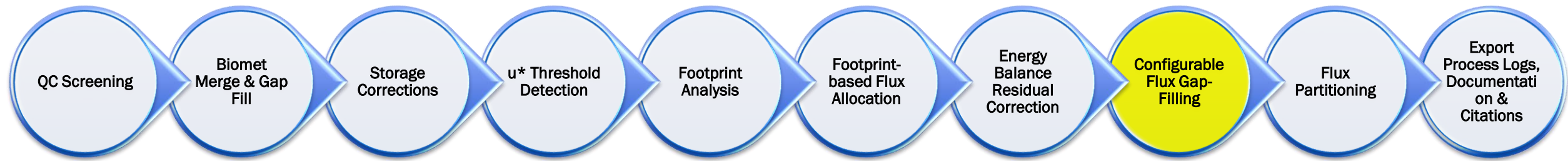
Variable:  Tolerance (K):  Single  Low/High

Additional Driver (Optional)

Variable:  Tolerance (K):  Single  Low/High

Additional Driver (Optional)

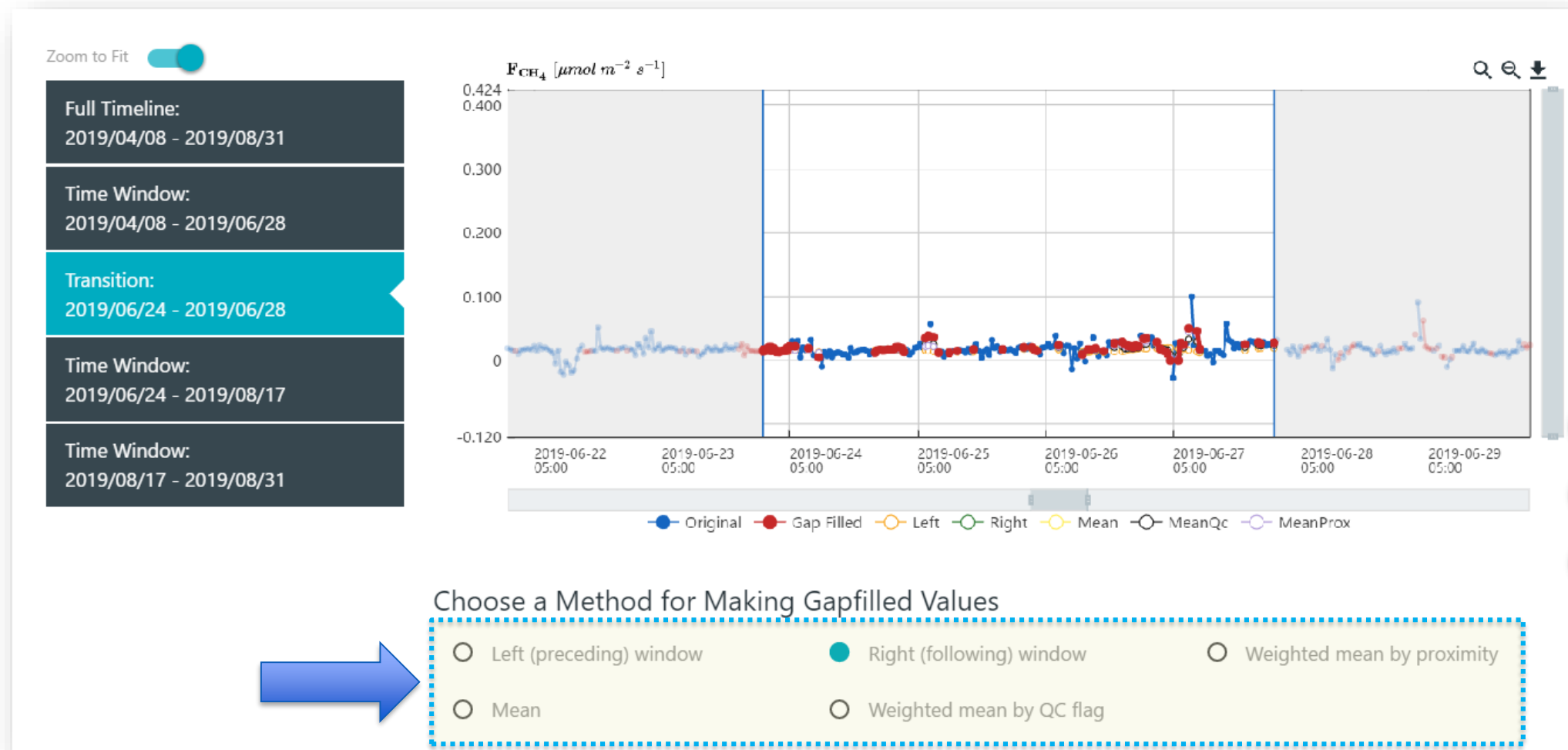
Variable:  Tolerance (m<sup>3</sup>/m<sup>3</sup>):  Single  Low/High



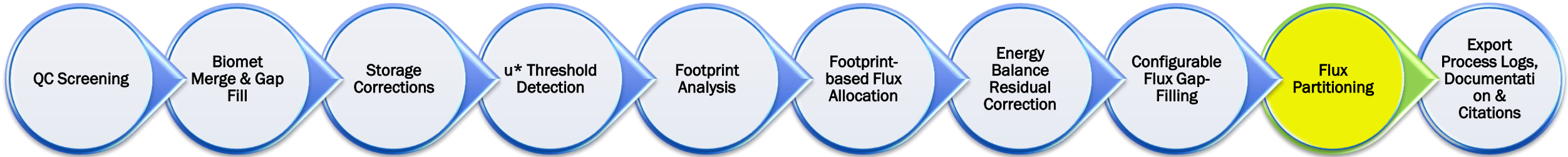
■ Example, selecting **best fits** in Tovi for Gap filling

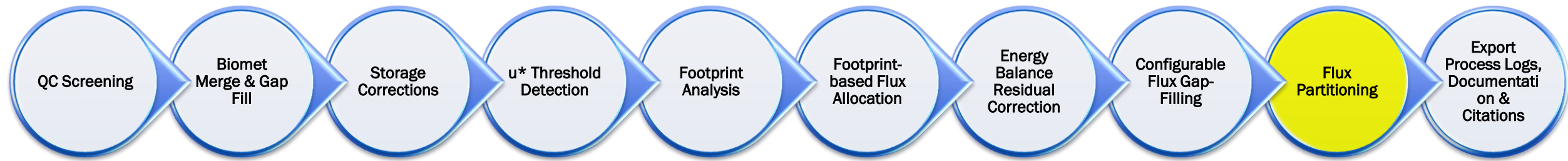
Transitions periods can be gap filled by different methods (using average fluxes from similar periods or average fluxes with similar drivers)

Choose the best option to gap fill by visual examination from different windows



# Typical workflow for Post Processing...

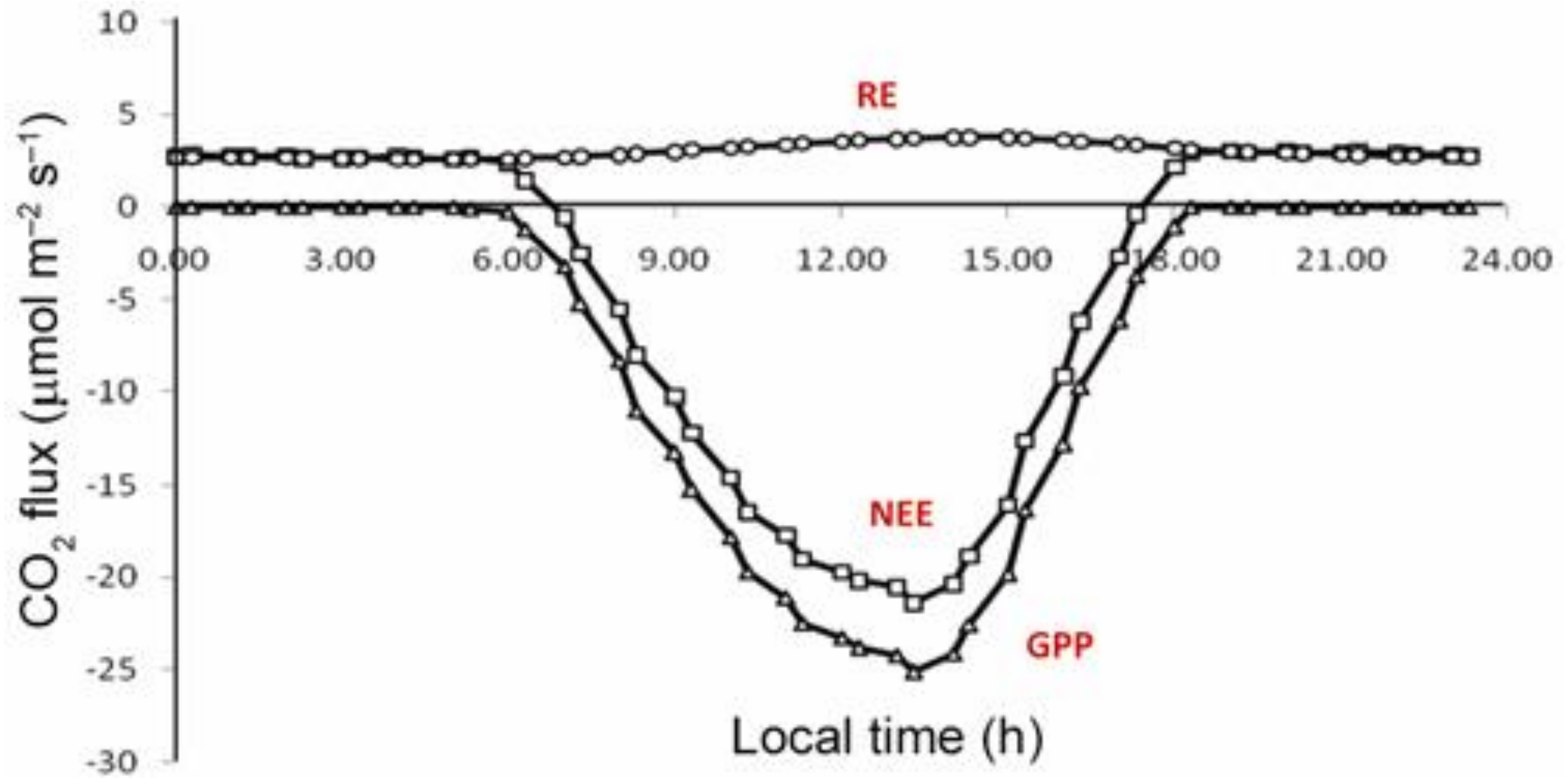
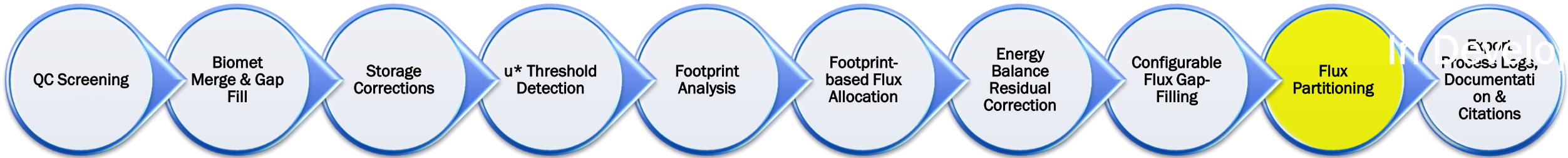


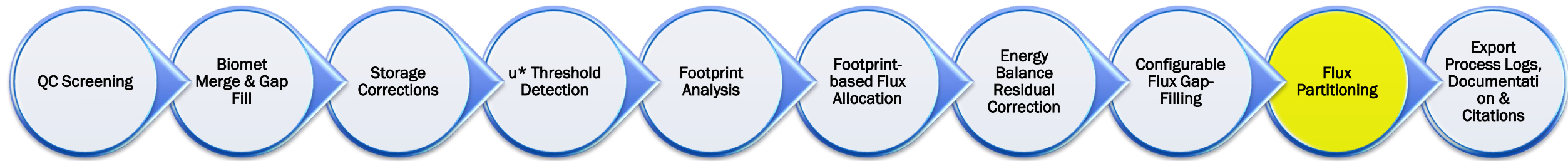


## Partitioning Fluxes

- NEE gives a valuable measure of ecosystem carbon sequestration, but by itself does not describe the processes responsible for carbon flux.
- To obtain better insights into the process-level controls over Net Ecosystem Exchange (NEE), there is considerable interest in partitioning the measured net flux of CO<sub>2</sub> to two component fluxes, gross ecosystem productivity (GEP) and total ecosystem respiration (Reco)
  - GEP is CO<sub>2</sub> uptake by photosynthesis (negative or zero at nighttime)
  - Reco is CO<sub>2</sub> release from ecosystem respiration (always positive)
- At night, the partitioning is simple, as Reco = NEE. During the day, the partitioning is dependent on the model used.
- Most strategies are based on only Reco occurring at night, while GEP is virtually zero. The challenge comes in extrapolating nighttime Reco measurements to daytime conditions to estimate GEP.
- These difficulties are compounded because nighttime flux measurements are often compromised by stable atmospheric conditions (insufficient turbulence) to satisfy the assumptions of the eddy covariance measurement system.

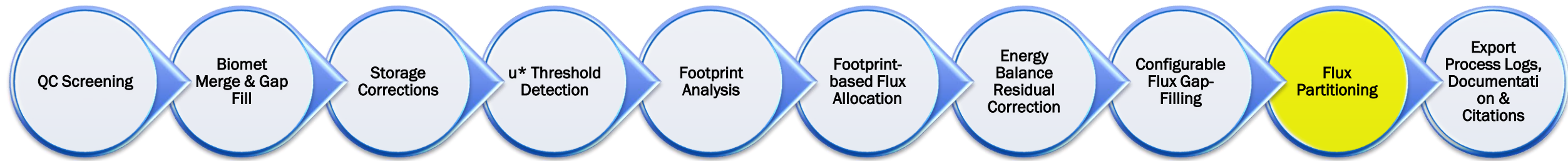
$$NEE = R_{eco} + GEP$$





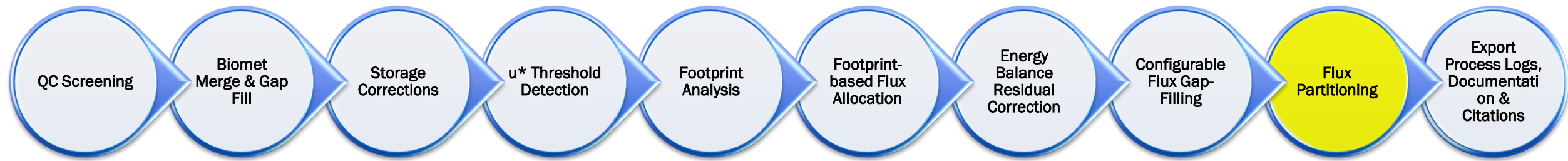
## Partitioning of fluxes – **Nighttime** Data-Based Method (Reichstein et al. 2005a)

- Flux partitioning techniques that rely on nighttime data must first ensure that the quality of these data is reliable.
- The flux partitioning algorithms follows three steps:
  - First, the temperature sensitivity  $E_0$  is estimated for the whole dataset.
  - Second, the reference temperature  $R_{ref}$  throughout the year is estimated with a seven-day sliding window in steps of four days. Periods where no  $R_{ref}$  could be found are linearly interpolated.
  - Third, using the estimates of  $E_0$  and  $R_{ref}$ , the net ecosystem fluxes  $NEE_f$  are partitioned into the gross primary production  $GPP_f$  and ecosystem respiration  $R_{eco}$ .



## Partitioning of fluxes – **Daytime** Data-Based Method (Lasslop 2010)

- Nighttime data for Reco modeling could involve bad quality input data. The alternative is to fit daytime NEE that accounts for the effects of radiation and vapor pressure deficit (VPD)
- Respiration is estimated from fitting the Light-Response Curve ( $NEE \sim \text{Radiation}$ ). NEE is modelled using the common rectangular hyperbolic light-response curve (Falge et al., 2001). The steps of the partitioning procedure are:
  - Reference temperature sensitivity  $E_0$  is estimated for a moving window from night-time data.
  - The  $E_0$  estimates smoothed across successive windows.
  - Parameters of the light-response curve are fitted using only day-time data for a moving window using the previously determined temperature sensitivity ( $E_0$ ).
  - For each  $NEE_f$  record GPP and  $R_{eco}$  are estimated with the parameter set of the previous and next window and results are interpolated linearly by the time difference to the window centers.

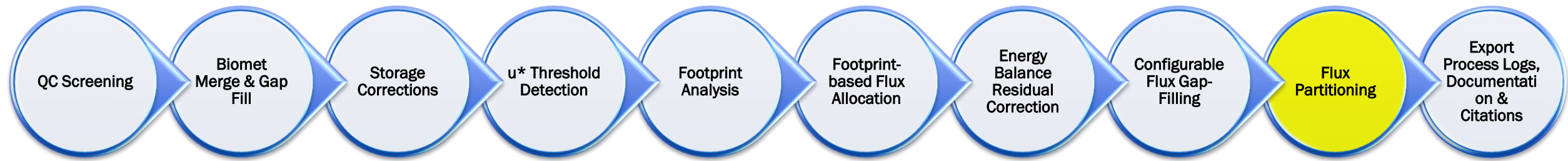


# Advantages and Disadvantages of each model

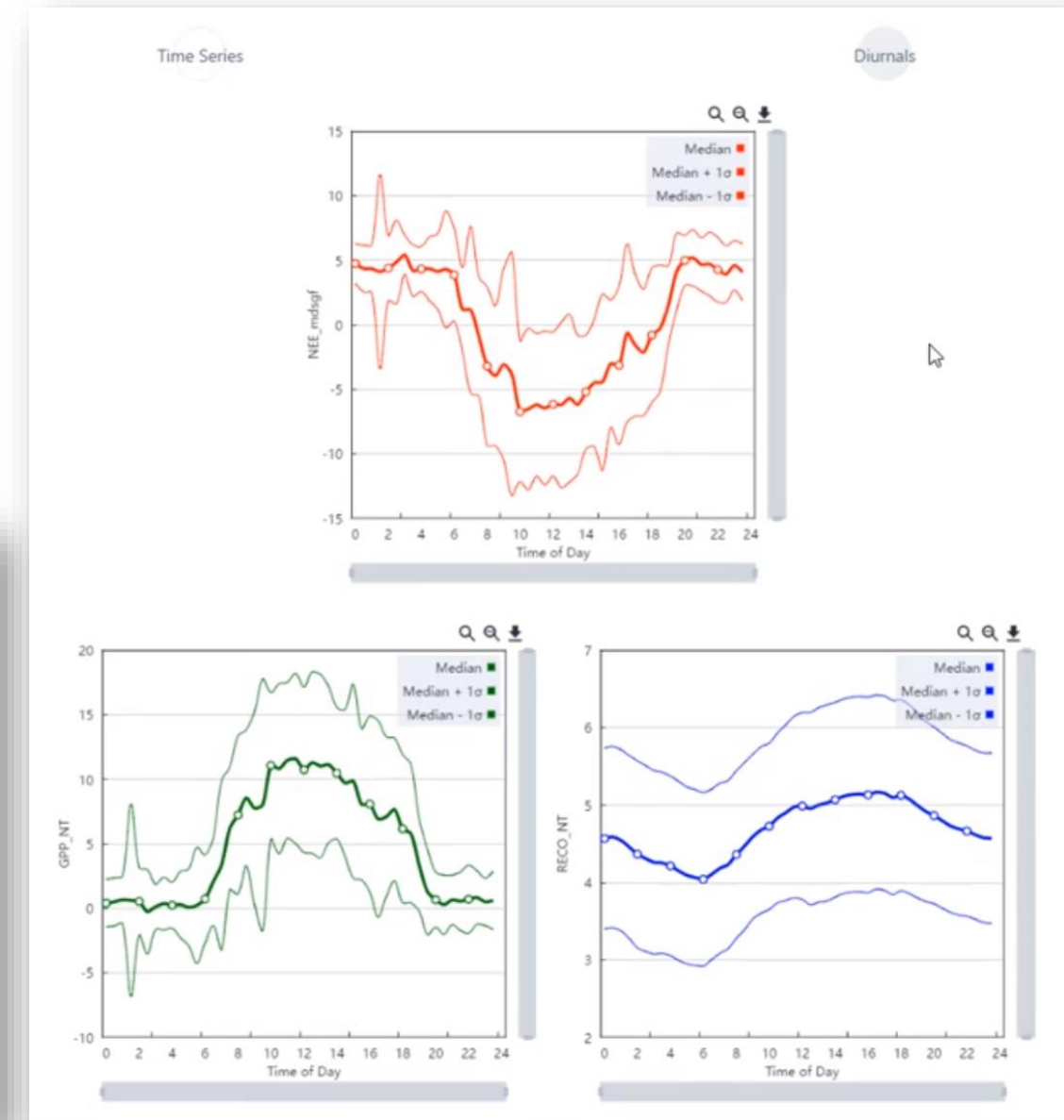
Approach	Advantages	Disadvantages
A: Nighttime data	Flux data represents $R_{eco}$	Extrapolation to daytime periods is necessary, nighttime eddy covariance data quality concerns
1. $R_{eco}$ is modeled as a time-invariant function of temperature Hollinger et al. (1994)	Simplicity, global applicability	Only applicable where no other factors than temperature influence $R_{eco}$ significantly, not generic
2. $R_{eco}$ is modeled as a time-invariant function of temperature and other environmental drivers (Rambal et al. 2003; Reichstein et al. 2002)	Simplicity, accounts for additional drivers of $R_{eco}$ , allows for seasonally varying temperature sensitivity	Results in selection of site specific factors that determine $R_{eco}$ , ancillary variables (e.g., SWC) may not be measured, or not measured uniformly, across ecosystems
3. $R_{eco}$ is modeled using temporally varying base respiration (Rref) parameters and one single temperature sensitivity derived from annual data set (Falge et al. 2002; Law et al. 2002)	Accounts for temporally varying respiration rates at reference temperature resulting from varying substrate availability	Long-term temperature sensitivity from annual data may not reflect short-term responses to environmental variability, introduction of systematic error when extrapolating to daytime
4. $R_{eco}$ is modeled using temporally varying functions of temperature (both Rref and temperature sensitivity varying) (this study)	Accounts for temporally varying respiration rates at reference temperature, caused by any factor, seasonally varying temperature sensitivity is accounted for (Davidson and Janssens 2006)	Often noisiness of eddy covariance data does not allow derivation of temperature sensitivity for large periods of the year, i.e., limited practical applicability

Approach	Advantages	Disadvantages
B: $R_{eco}$ derived from daytime NEE observations	Increases size of data set used to make inference; reduces reliance on night-time data.	Depends on specific light-response curve model; light-response curve can be confounded by other factors (e.g., vapor pressure deficit), sometimes yields unstable parameter estimates (high standard errors); $R_{eco}$ estimate susceptible to storage flux problems, since those occur in the morning and evening during low-light conditions
1. $R_{eco}$ as y-intercept from light-response curve of GEP (Falge et al. 2002)	Day-to-day variation of $R_{eco}$ reflected	Only daily $R_{eco}$ can be derived
2. $R_{eco}$ (temperature driven) and GEP (radiation driven) are simultaneously modeled as parts of one fixed model equation (Gilmanov et al. 2003)	Uses all data (night- and daytime)	Resulting GEP is from a model and thus constrained by model assumptions (disallows comparison with other models), temperature sensitivity maybe confounded by response of GEP to environmental factors, that are hard to separate (e.g., is afternoon drop in NEE caused by $R_{eco}$ as $f(T)$ or by high VPD, or even by plant-internal hydraulic constraints)
3. $R_{eco}$ and GEP are simultaneously modeled as parts of one model equation with state dependent parameters (data-based mechanistic modeling approach)	Uses all data (night- and daytime); very flexible approach; parameters can evolve with time and state	Statistical assumptions (e.g., noncorrelated residuals) and robustness against violations may be problematic; maybe affected by confounding factors similar to B2.
4. $R_{eco}$ (temperature driven) and GEP (radiation and VPD driven) are simultaneously modeled as parts of one model equation with state dependent parameters (Lasslop et al. 2010)	Uses nighttime data to derive the t-sensitivity and daytime data for GEP parameters and flux magnitude of $R_{eco}$	Equifinality and/or local minima in the cost function. Drought limitation of respiration. $-R_{eco}$ is strongly determined by morning and evening measurements where storage peaks can occur.
5. $R_{eco}$ and GEP are derived via a posteriori analysis of an artificial neural network conditioned with all data. (cf. Papale and Valentini 2002) for gap-filling; flux-partitioning not explored	Uses all data (night- and daytime); very flexible approach; influence of different input data can be evaluated for best description of the data set	Extrapolation problem, since zero radiation has to be assumed estimating $R_{eco}$ from neural network during the day; potentially confounded by other factors similar to B2.





- Example, Partitioning fluxes (NEE) in Tovi – setting and results



Toolbox **Night-time Partition [NTP]**

**START**

Variable for Night-time Partition

- FCO2
- NEE
- NEE\_mdsgf**

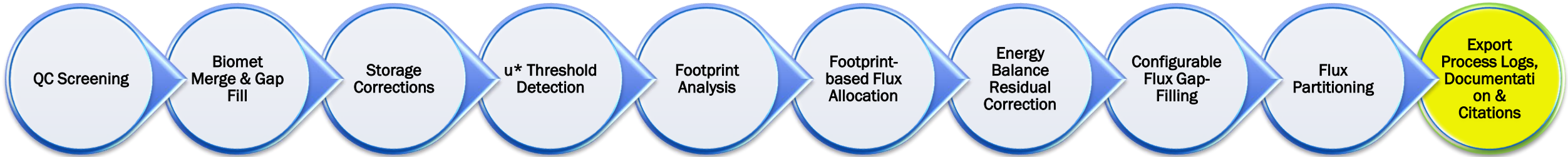
Night-time Partition Inputs

Solar Radiation W/m<sup>2</sup>

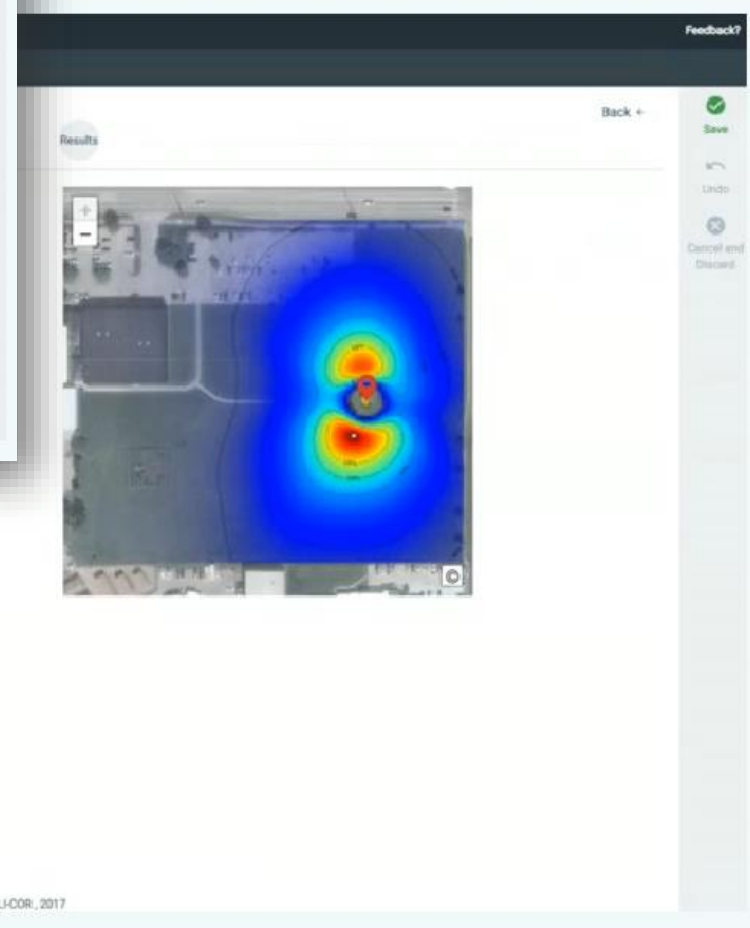
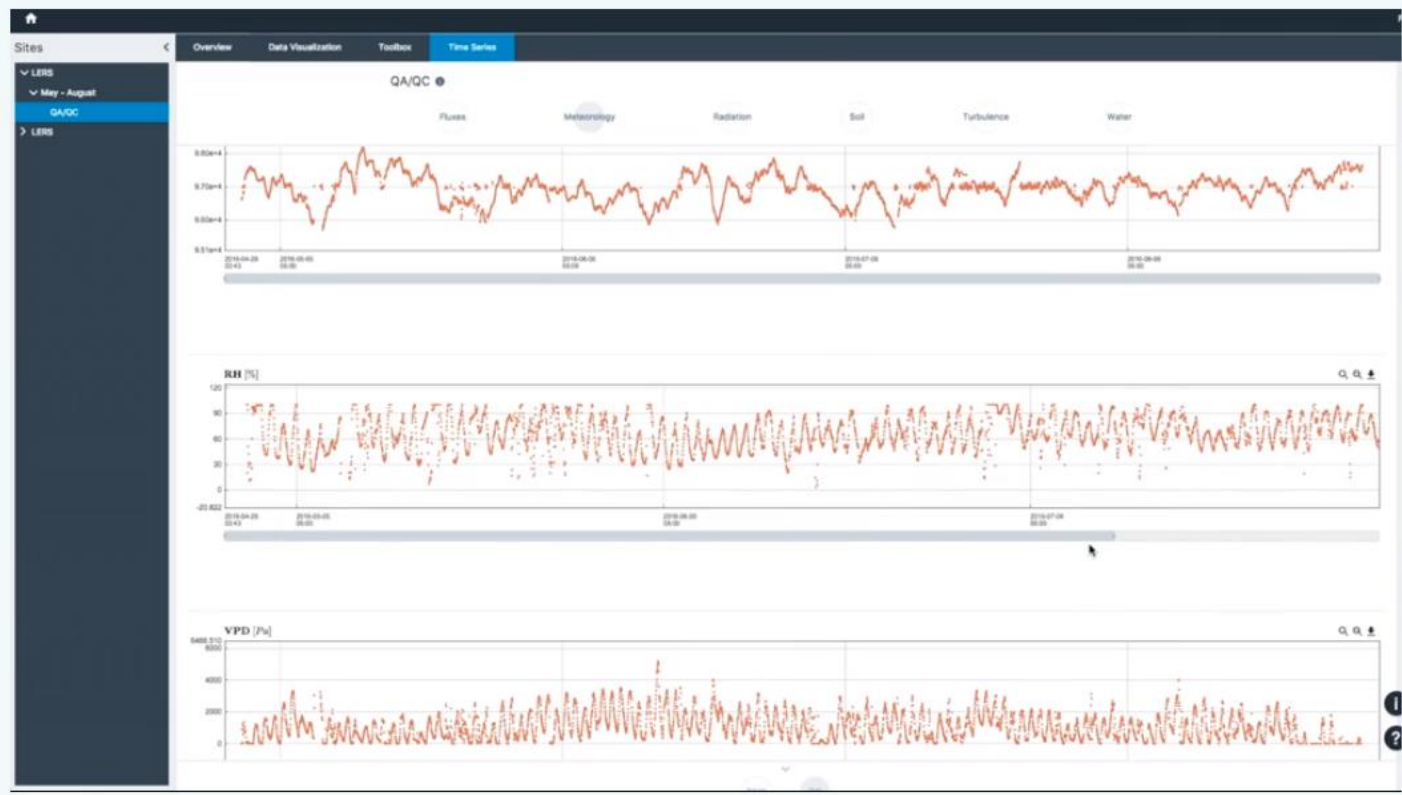
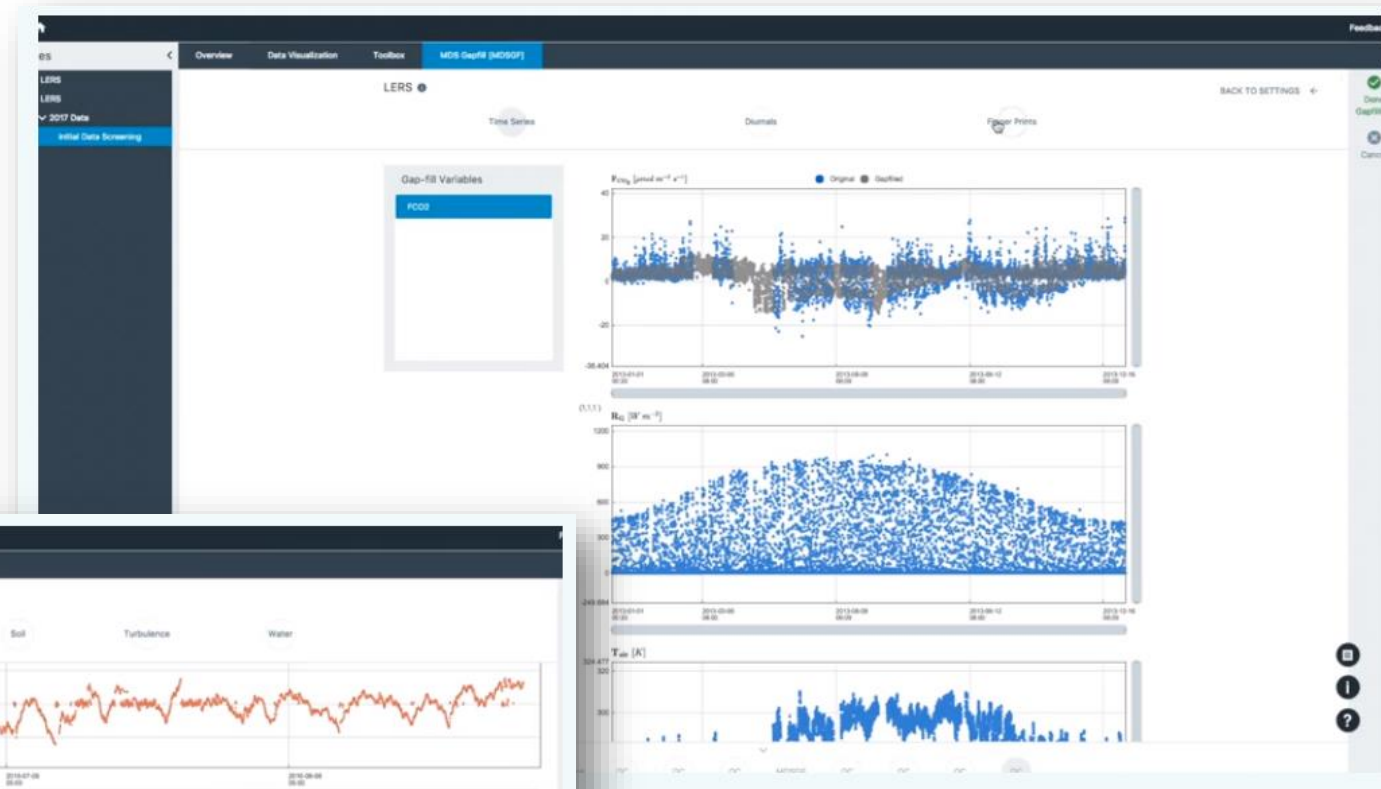
Temperature C

VPD Pa

# Typical workflow for Post Processing...



# Tovi Data Analysis Software



# Data Analysis: Community Collaboration

- LI-COR's post-processing software goals...in line with EddyPro strategy and our mission statement:
  - Collaborate with the community.
  - Assist in the furthering of scientific discovery.
  - Help scientists expand the reach of their new developments to a larger community (wider usage) and citing their work

# Data Analysis: Community Collaboration

- LI-COR's post-processing software goals...in line with EddyPro strategy and our mission statement:
  - Pull everything together in a single platform.
  - Design an easy-to-use UI, a highly-visual experience.
  - Build a robust software.
  - Provide support through scientific staff and documentation.
  - Reproducibility, Transparency (open-source), Citable Documentation.
  - Continual additions, improvements, innovations...

# Tovi Software Collaboration Efforts

- Began design and development three years ago.
- Collaboration with **OzFlux**
  - QC Screening, Meteorological Gap Filling, Averaging
    - *Isaac, P. et al., 2017*
- Collaboration with **ICOS, AmeriFlux, Euroflux**
  - $u^*$  Threshold Detection, Flux Gap Filling, Flux Partitioning
    - *Reichstein, M. et al., 2005*
- Collaboration with world-class individual scientists
  - Footprint Analysis
    - *Kljun, N. et al., 2015.*
  - Energy Balance Correction
    - *Mauder, M et al., ongoing*



# Getting Started: Importing data into Tovi

- Tovi needs the following four files for full operation of all the Tools...

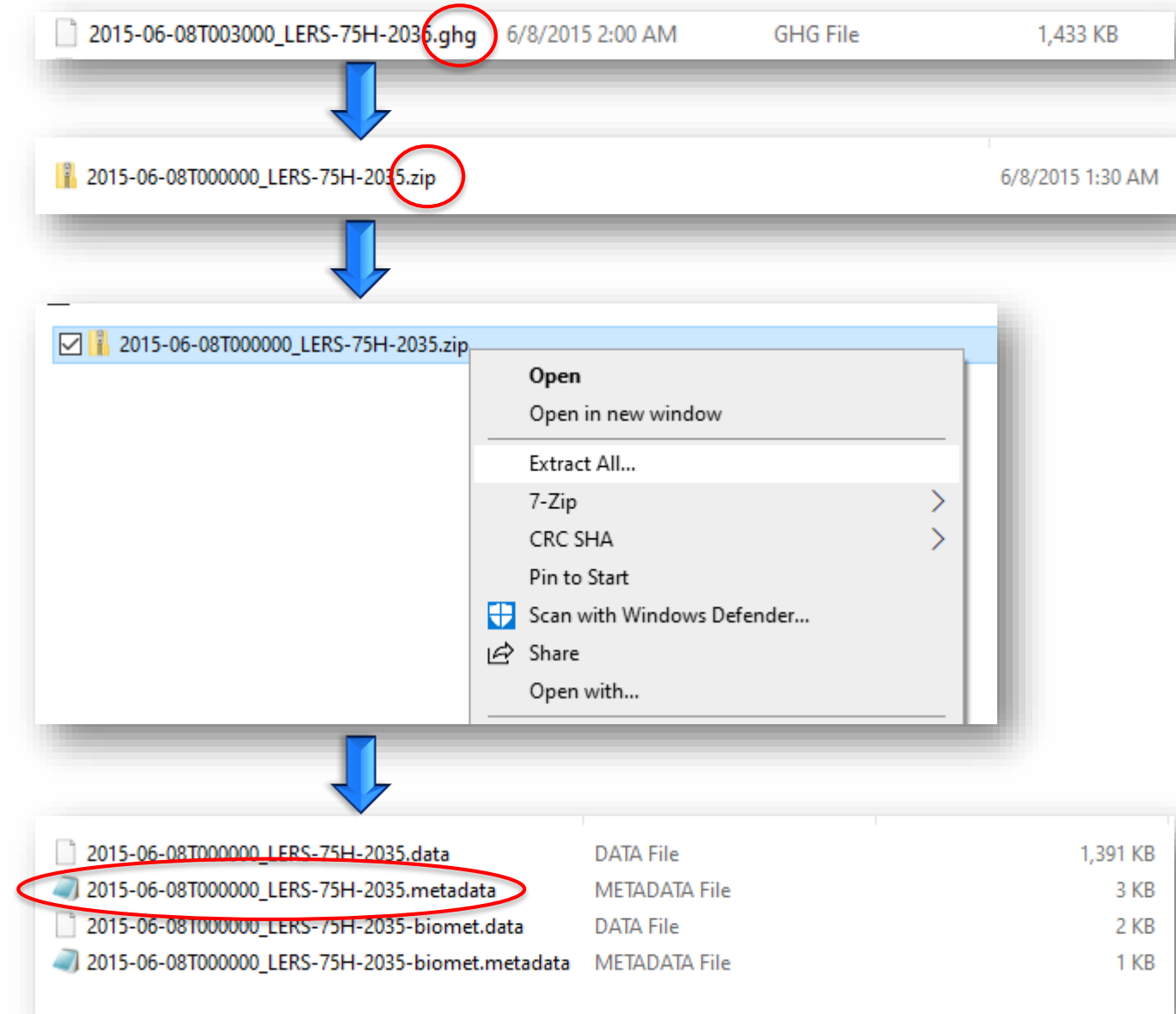
These three are provided  
in the EddyPro outputs

- eddypro\_\*\_full\_output.csv
- eddypro\_\*\_biomet.csv
- \*.eddypro
- \*.metadata

- Depending on your data collection scheme and your data processing pathway, you may need to modify your files for importing to Tovi...

# Getting Started: Importing data into Tovi

- Getting the (\*.metadata) file
  - Select any one of the (\*.ghg) files and copy it to a temporary directory
  - Rename the file by changing the extension from (\*.ghg) to (\*.zip)
  - Select the file and then right-click to 'Extract All...'
  - Open folder and select the (\*.metadata) file. Copy/paste this with your EddyPro outputs, where you can find and import it later.





# Data Analysis – Files for Importing to Tovi

