Flux Data Characteristics: What are the Data?

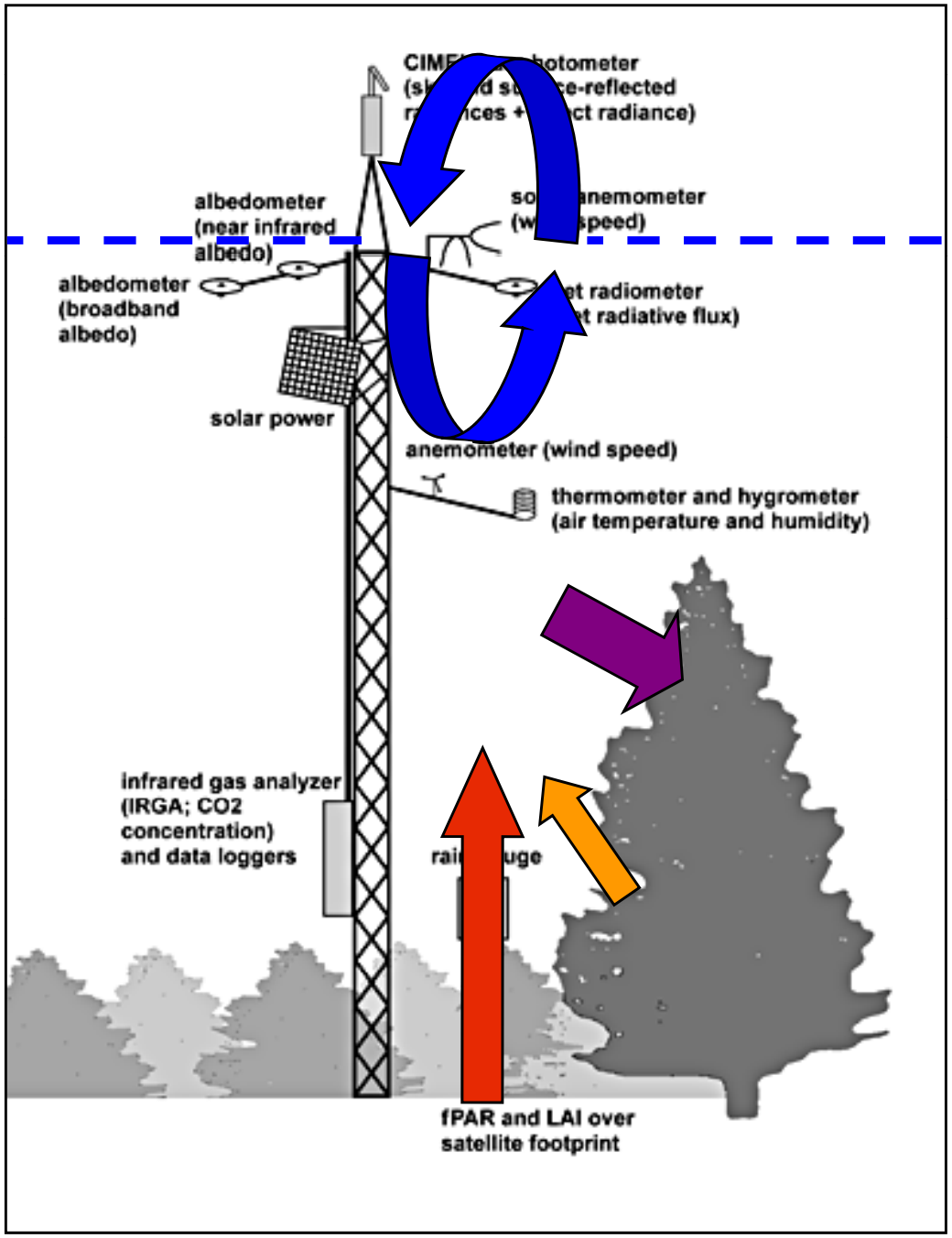
(how are they processed, uncertainties, ancillary data)

Dario Papale

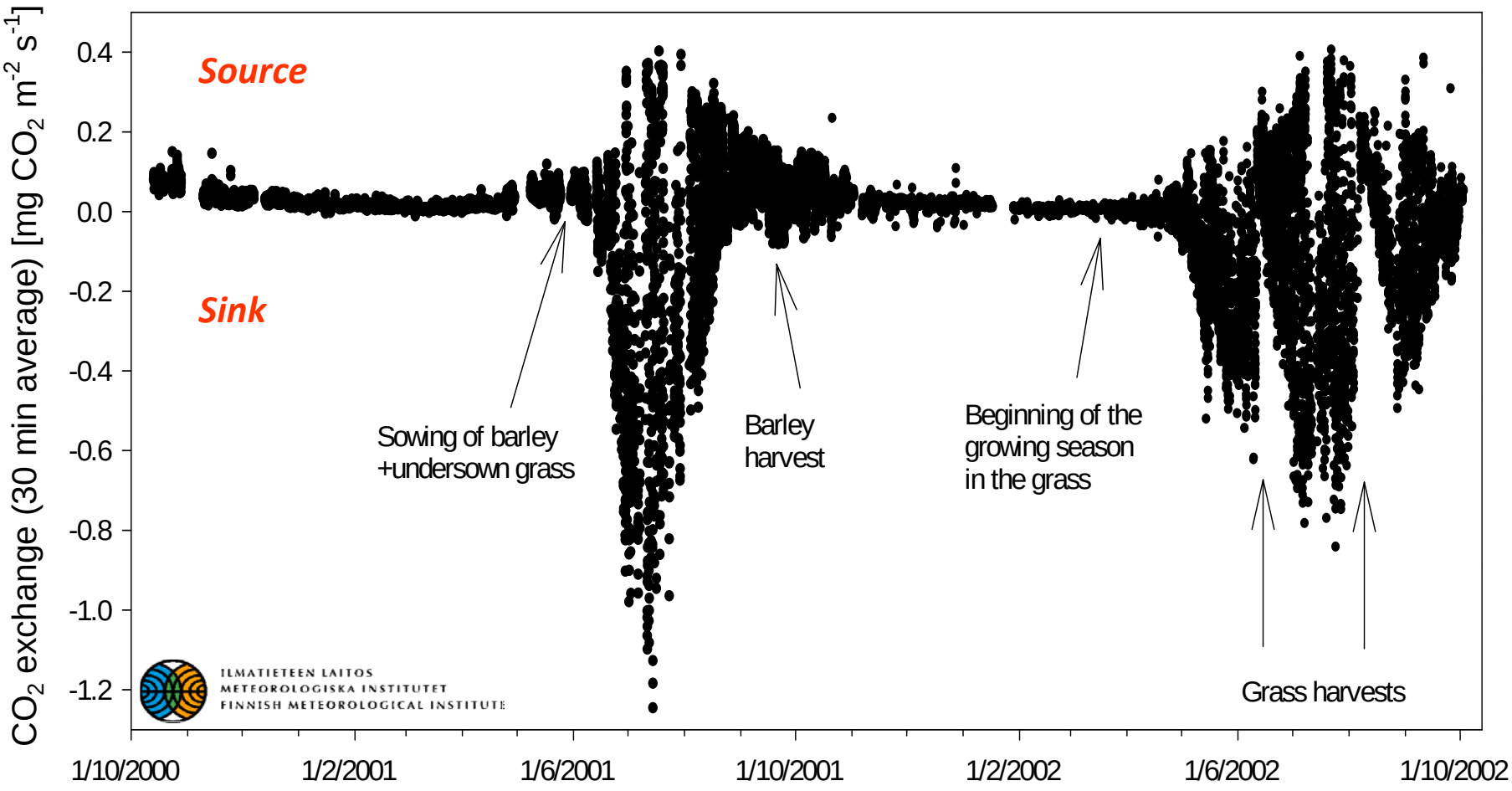
Eddy covariance technique

+ Measures whole net ecosystem exchange of CO_2, H_2O and energy
+ Non-destructive & continuous
+ Time-scale hourly to interannual

Does not deliver compartment fluxes, but:
NEP = GPP - Reco

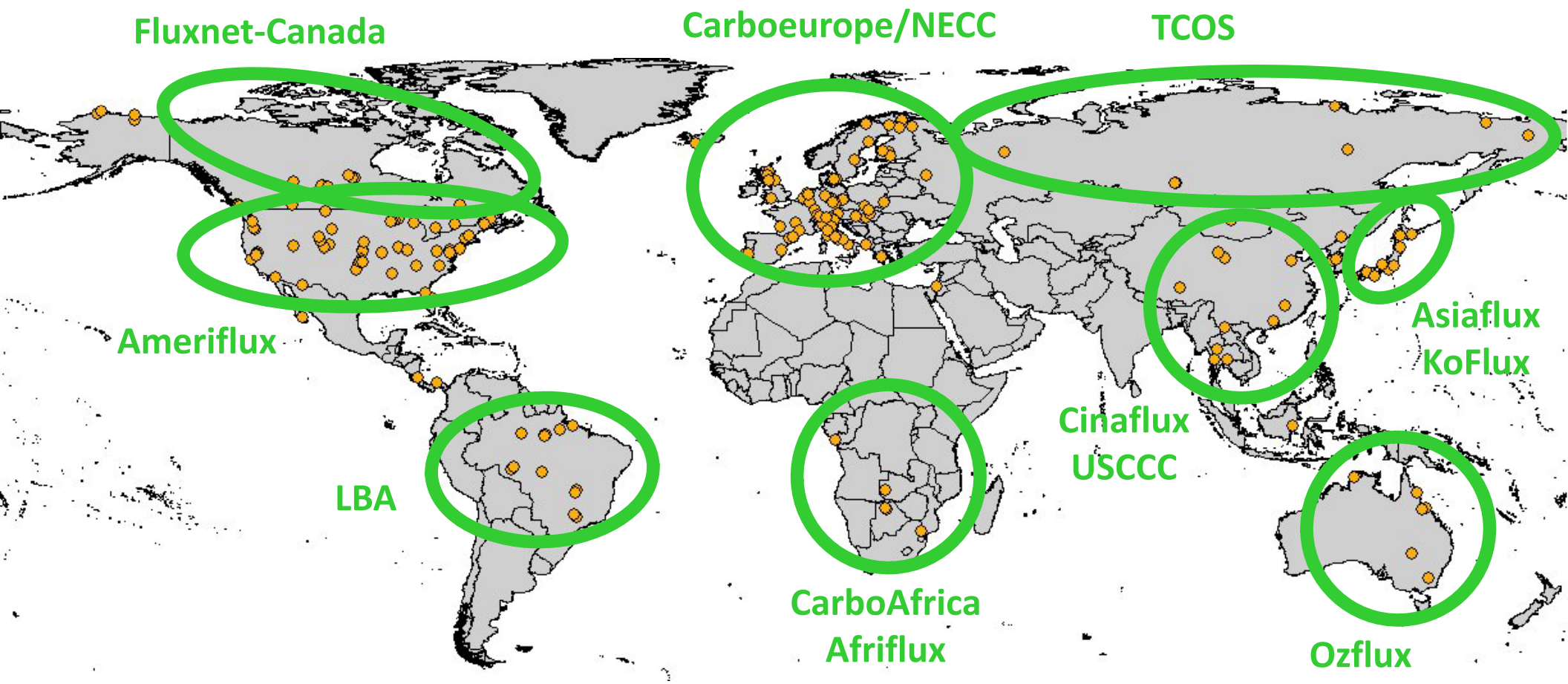


Two years of NEE measurements in the peat soil



T. Laurila et al.

The FLUXNET network



AmeriFlux AsiaFlux KoFlux LBA Fluxnet Canada CARBO EUROPE IP

AfriFlux ChinaFLUX OzFlux Australian and New Zealand flux Research and Monitoring USCCC NECC Nordic Centre for Studies of Ecosystem Carbon Exchange and its Interactions with the Climate System Terrestrial Carbon Observation System SiBERIA

But where are the data available in the databases coming from?

The long road of the eddy covariance measurement…

EDDY COVARIANCE SITE

Where? (the footprint)

What? (the set-up)

Raw data (20 Hz)

Raw data correction and processing, fluxes calculation

Random errors

Half hourly data

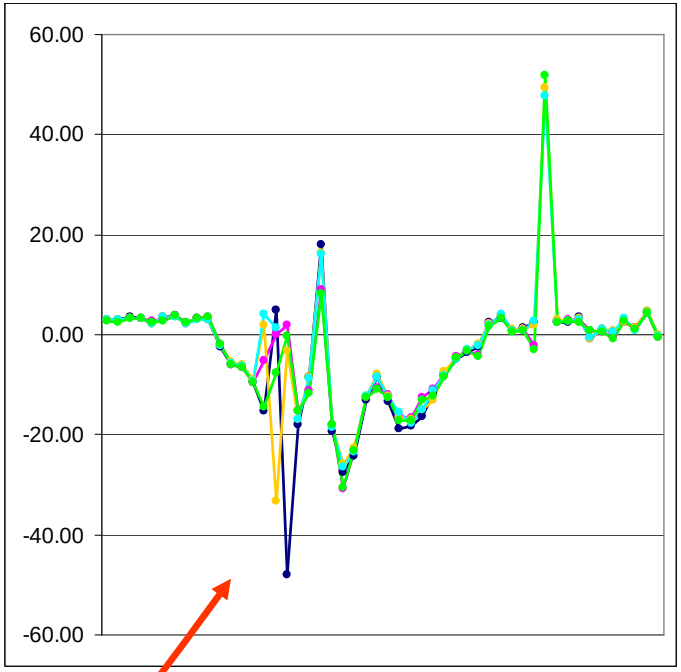
HH data quality check, filtering and correction

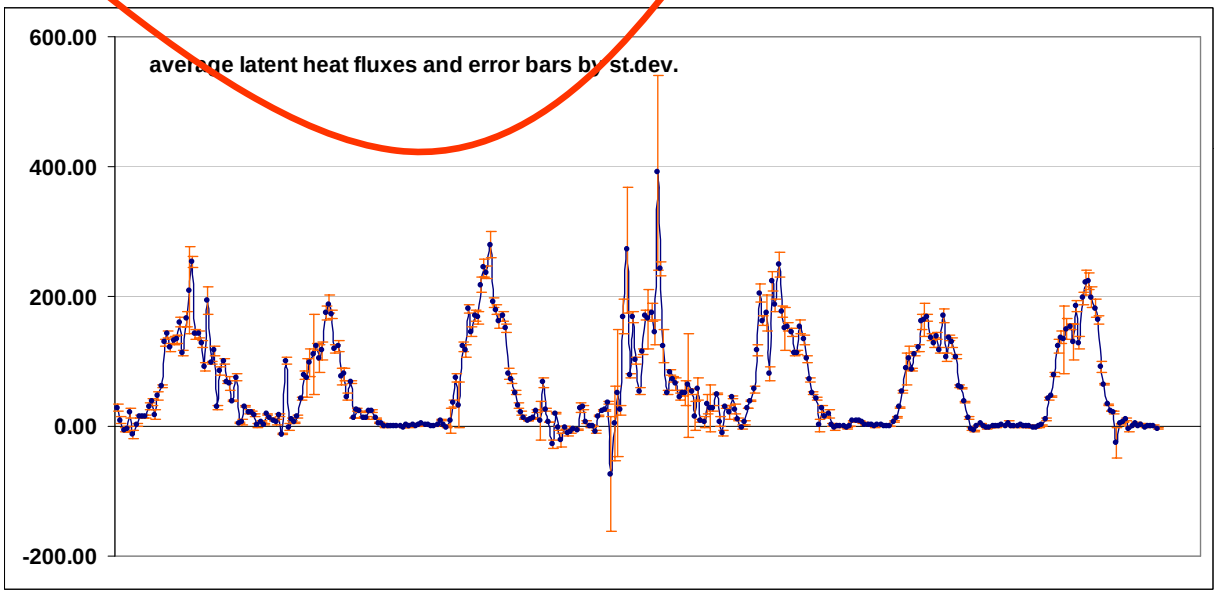
Gap filling

Partitioning

Half hourly, daily, monthly, annual data

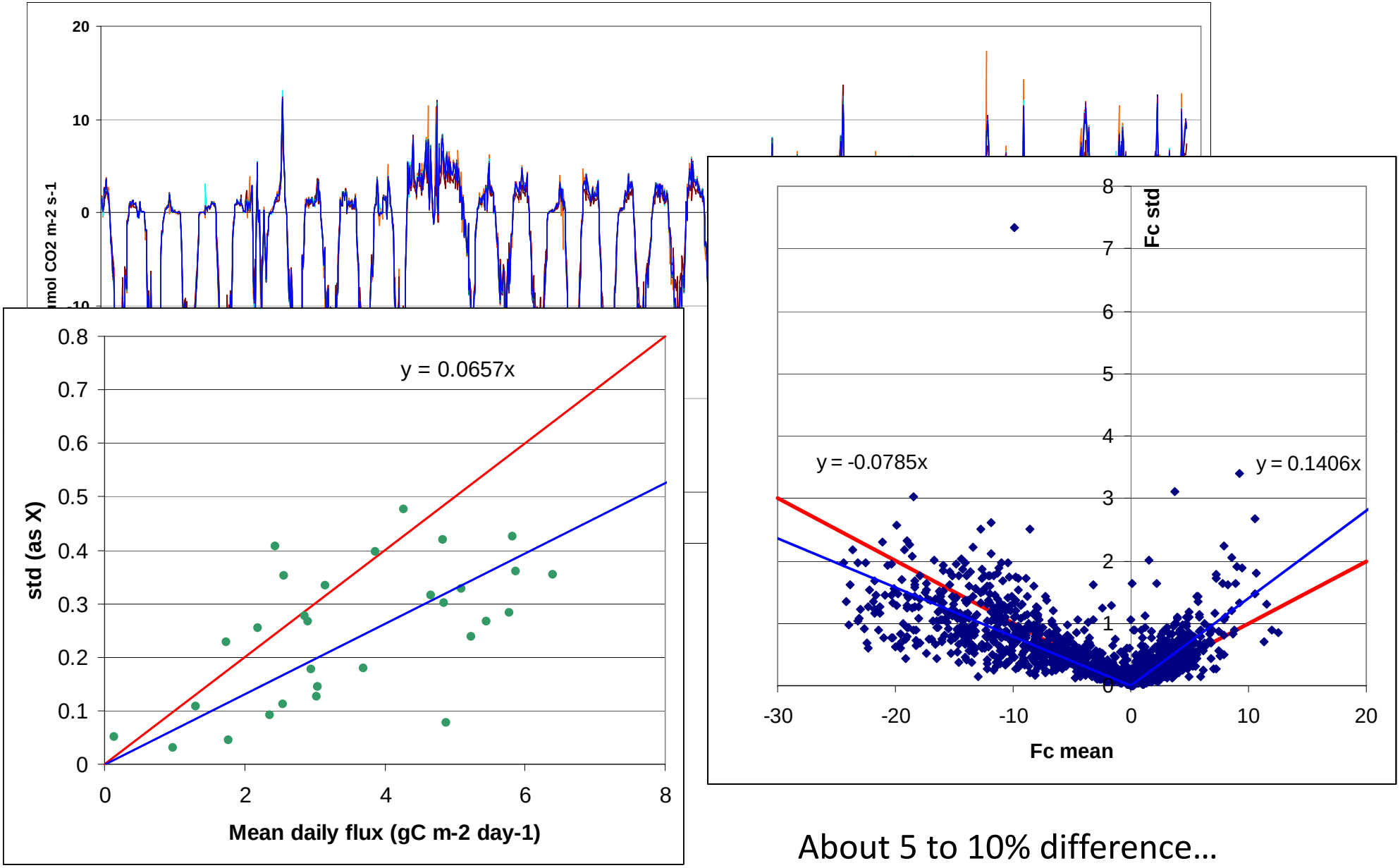
Raw data processing: 5 software comparison…





Fratini et al. in prep.

Raw data processing: same software, different settings

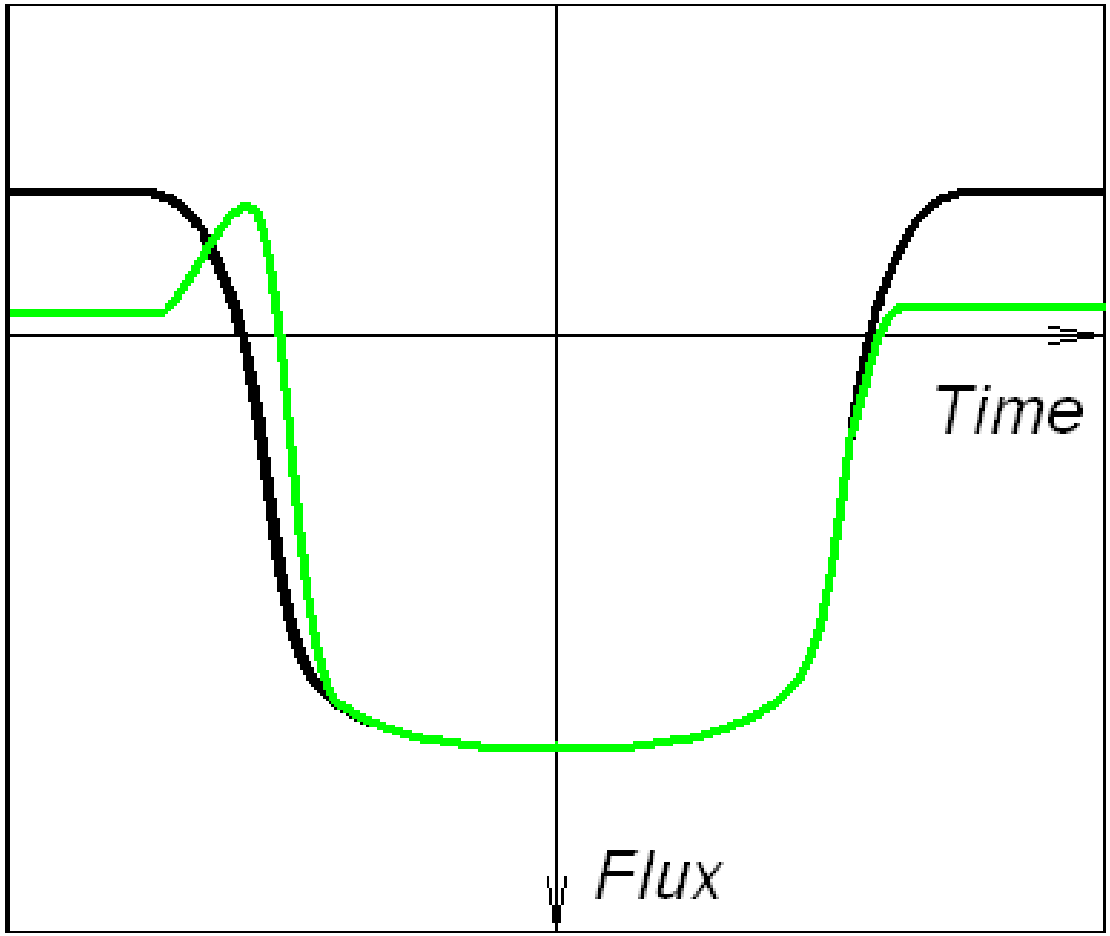


About 5 to 10% difference…
Often the processing scheme used is not documented!!

Half hourly data QAQC: storage and u* filtering

Actual flux

Measured flux

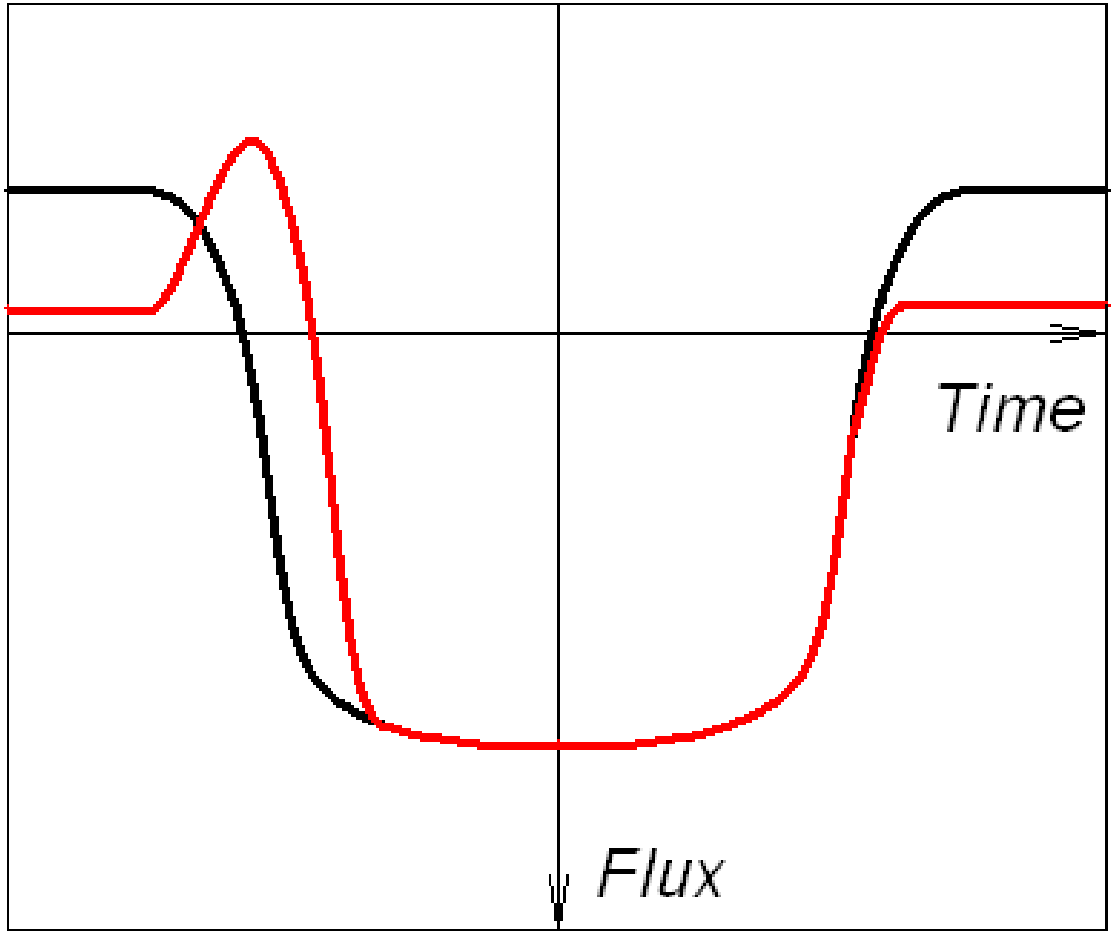


Marc Aubinet

Half hourly data QAQC: storage and u* filtering

Actual flux

Turbulent flux if storage

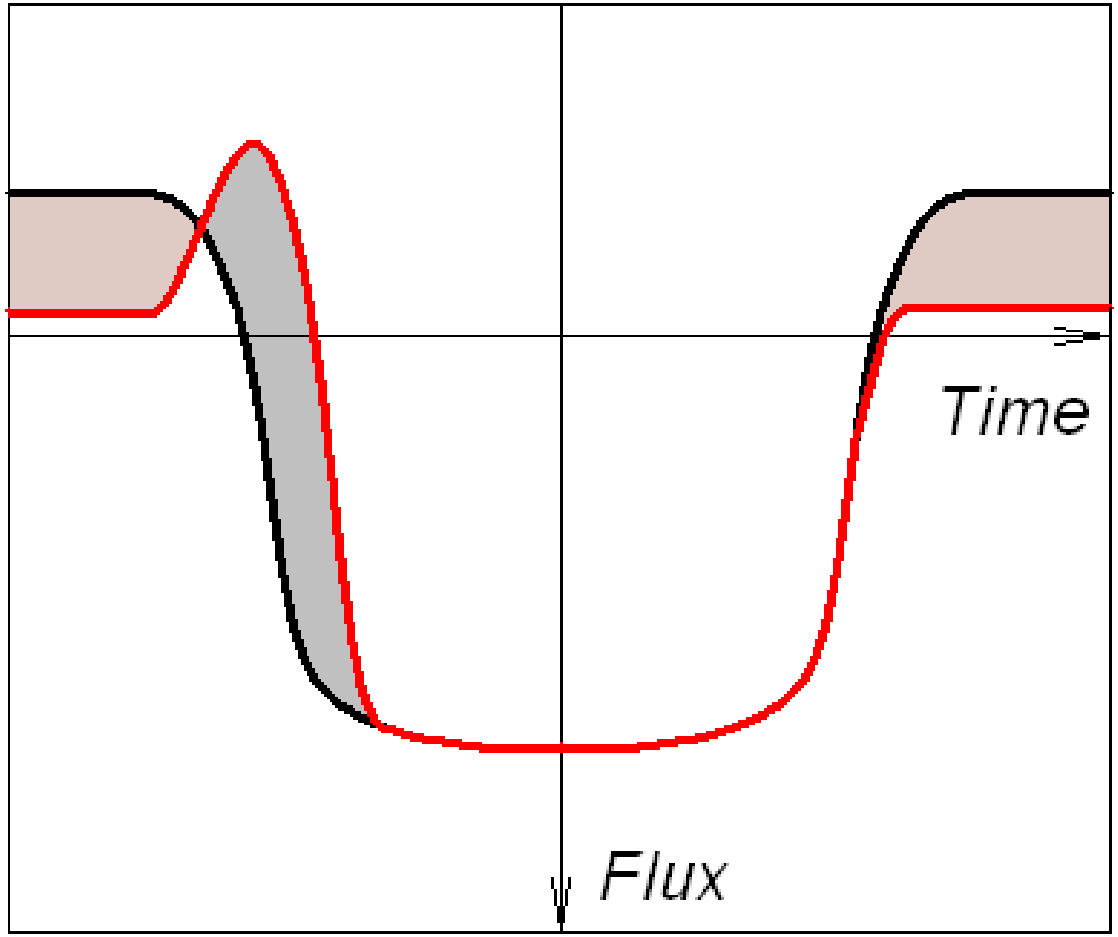


Marc Aubinet

Half hourly data QAQC: storage and u* filtering

Real flux

Turbulent flux if storage

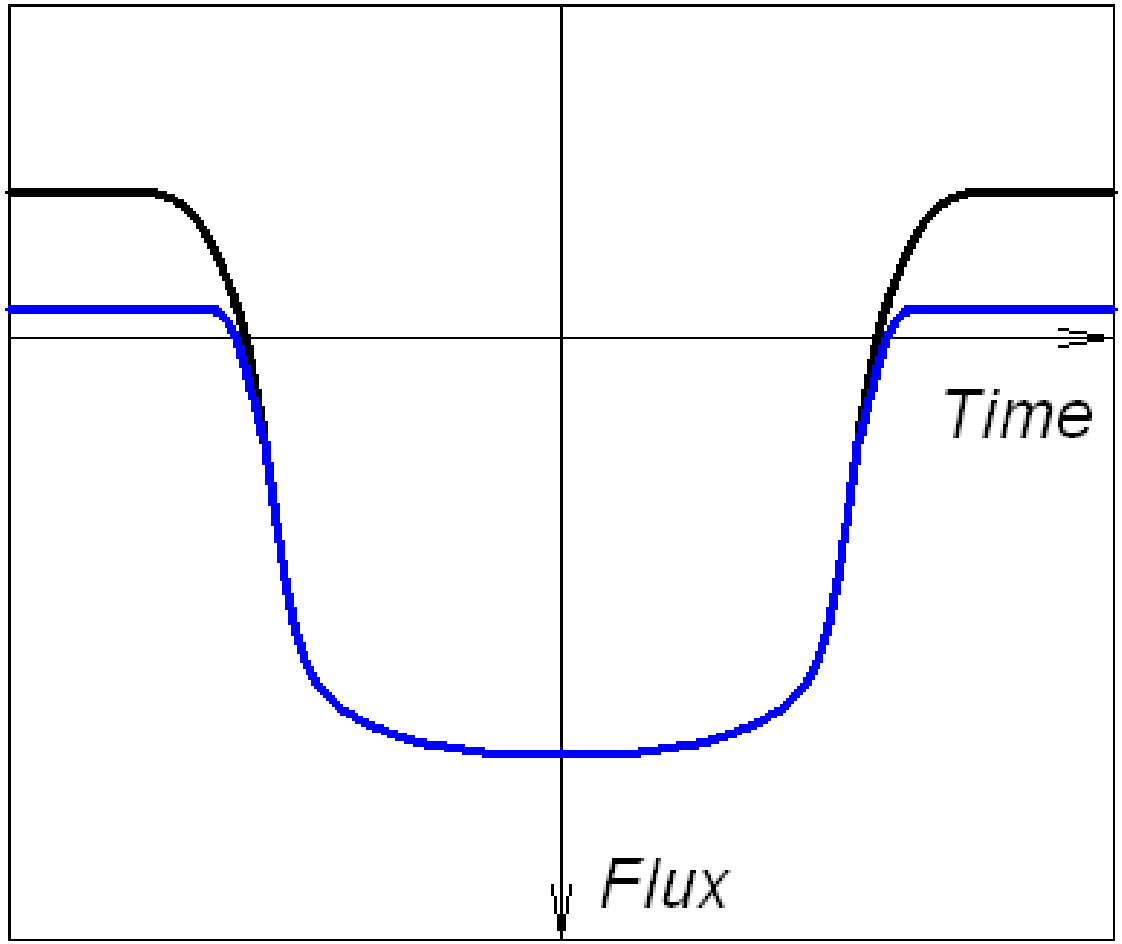


Marc Aubinet

Half hourly data QAQC: storage and u* filtering

Actual flux

Turbulent flux if advection

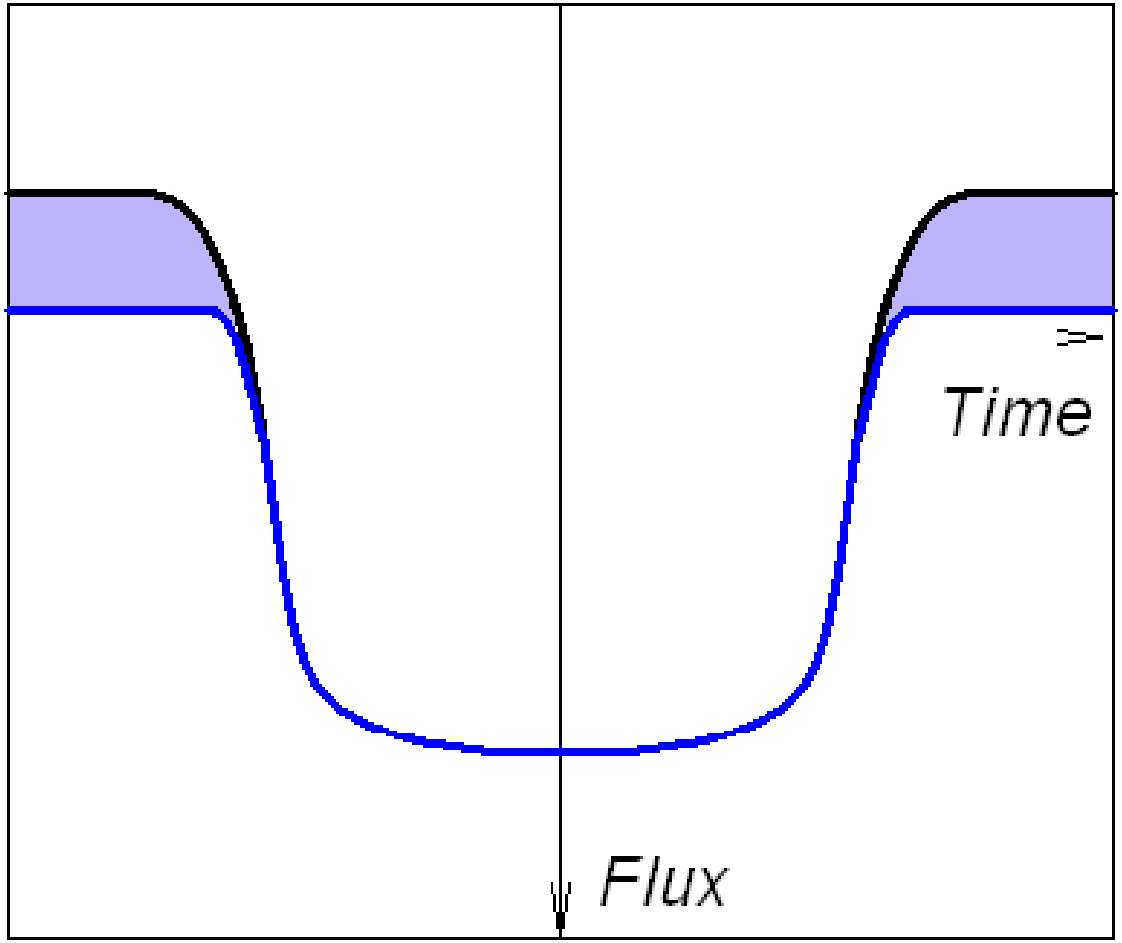


Marc Aubinet

Half hourly data QAQC: storage and u* filtering

Actual flux

Turbulent flux if advection

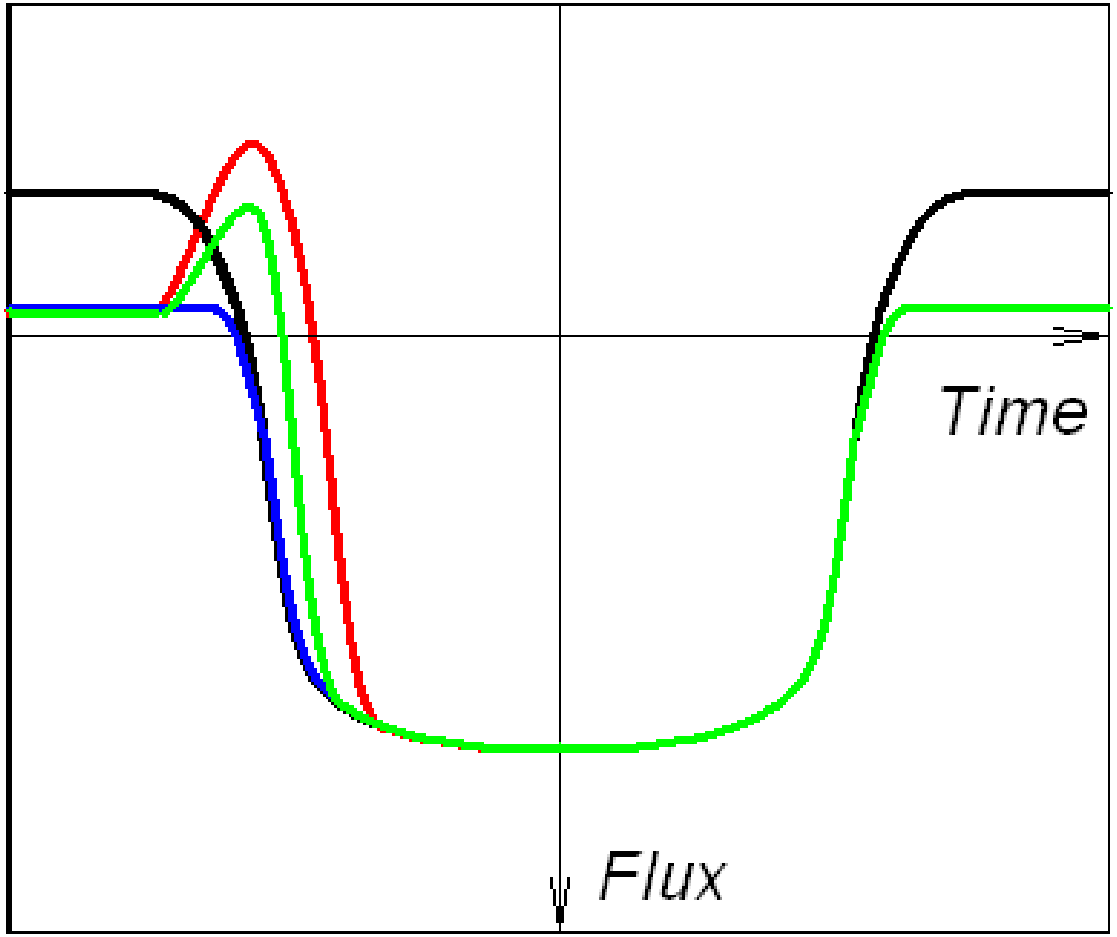


Marc Aubinet

Half hourly data QAQC: storage and u* filtering

Actual flux

Most probable situation

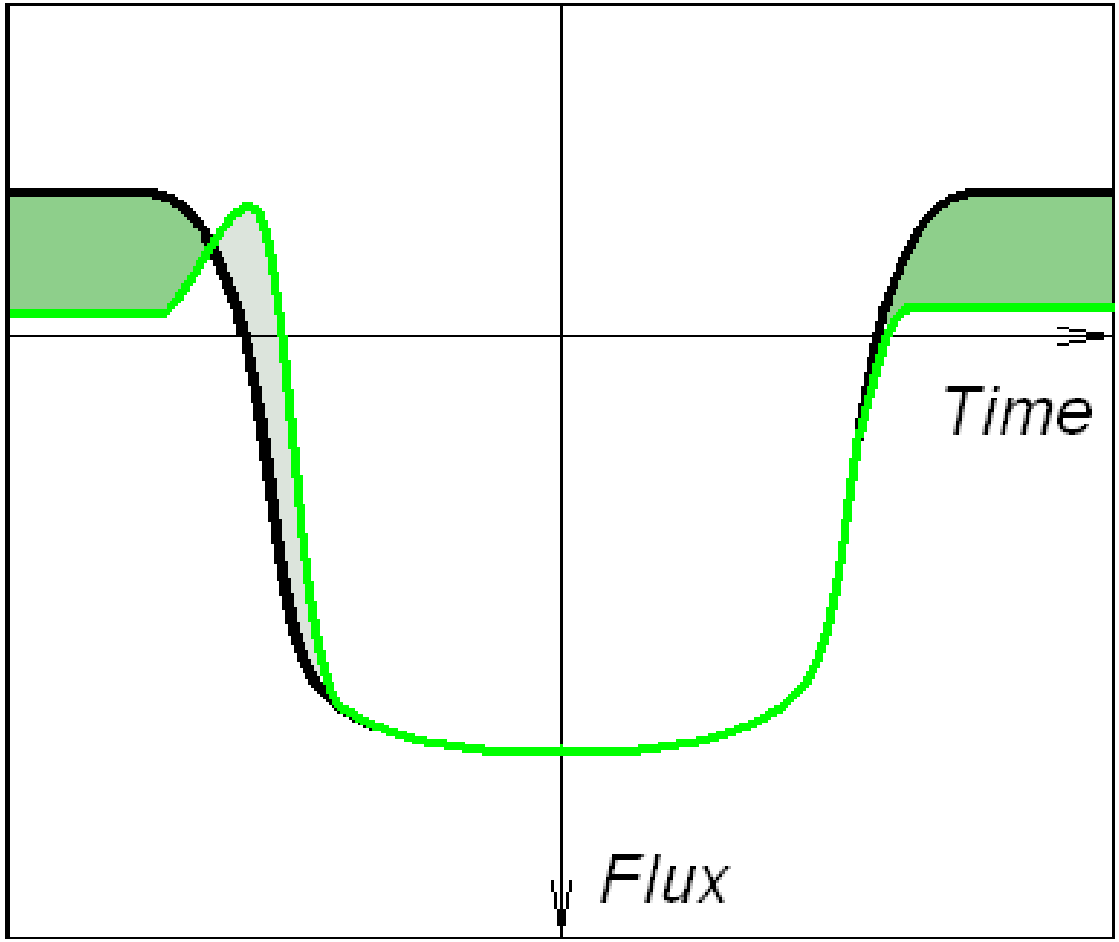


Marc Aubinet

Half hourly data QAQC: storage and u* filtering

Actual flux

Most probable situation

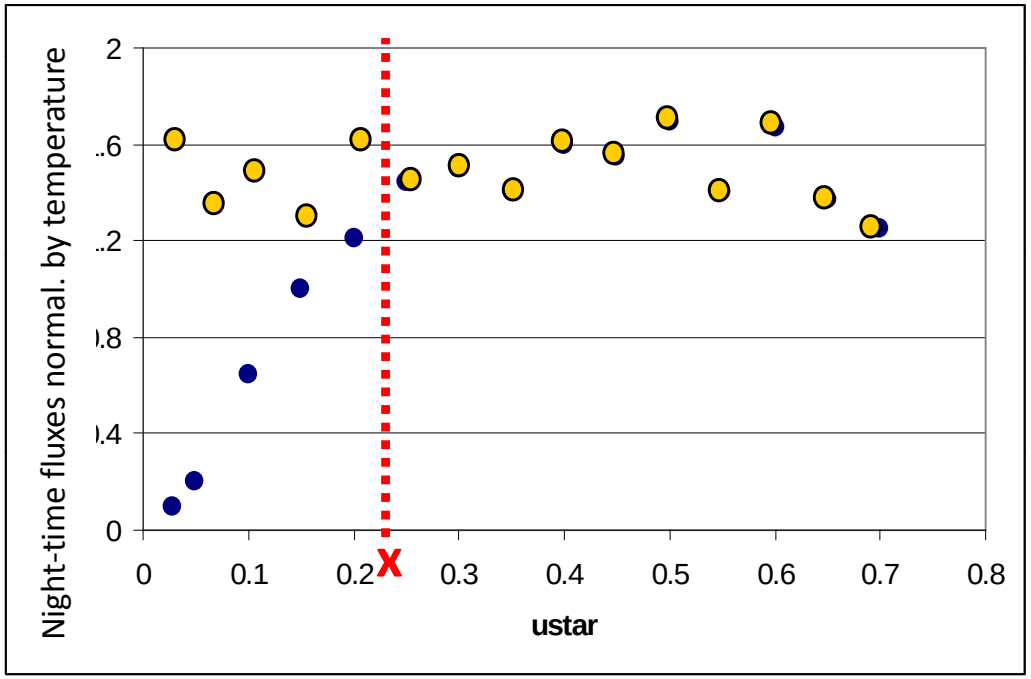


Marc Aubinet

Half hourly data QAQC: storage and u* filtering

How to correct for this problem?

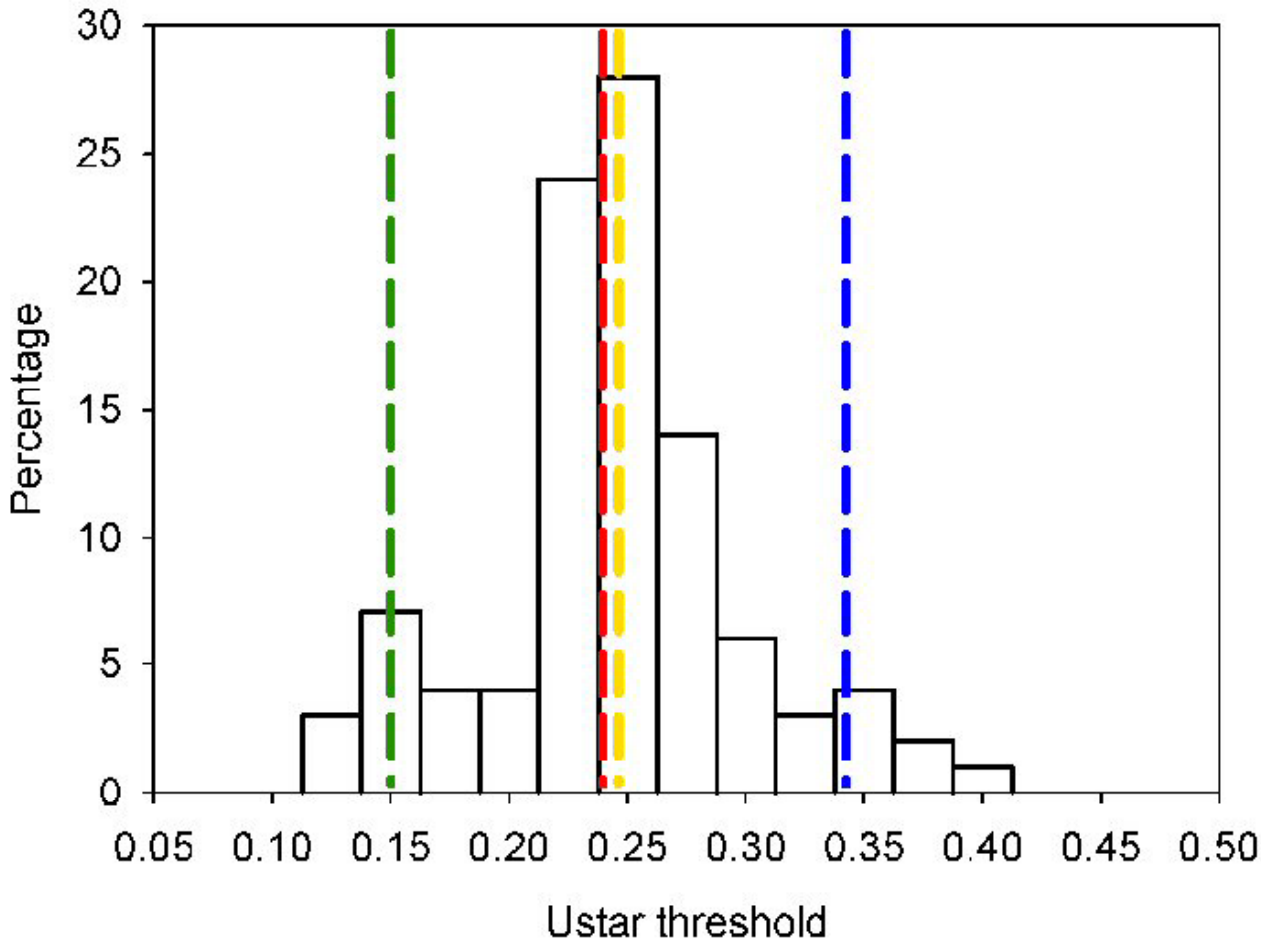
- Storage: measuring the CO_2 concentration dynamic below the EC system and adding this term
- Advection: the most commonly applied and accepted method is the u* filtering



There are different methods proposed to find the u* threshold value (Reichstein et al., Papale et al, Gu et al., Barr et al…) plus other alternative methods (Hollinger et al., Thomas et al., …). All the different approaches give different results => uncertainty.

Half hourly data QAQC: the u* filtering

Bootstrapping technique is used to assess the uncertainty in the ustar threshold selection

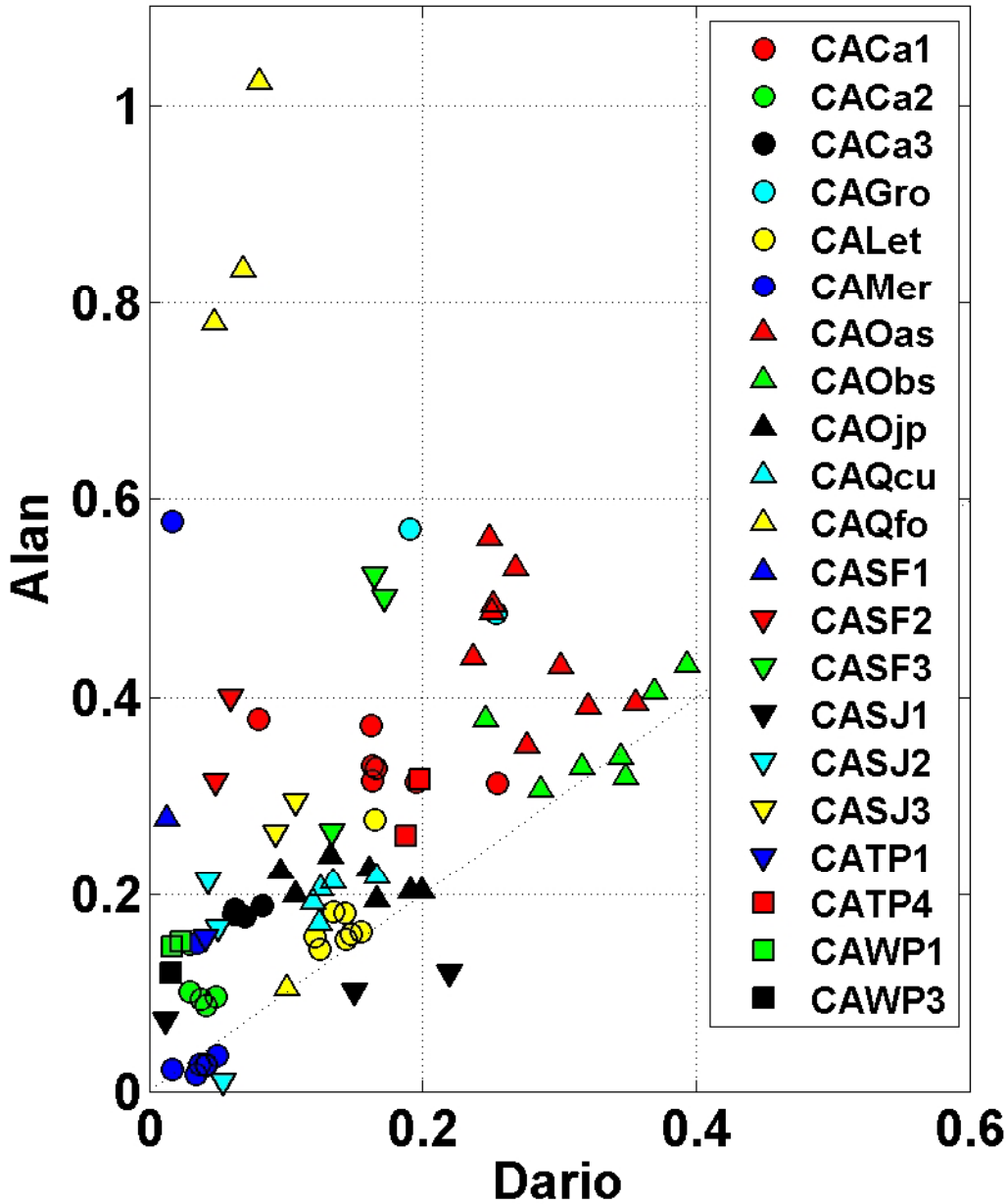


BE-Vie 2001

100 bootstrap estimates

Papale et al. 2006

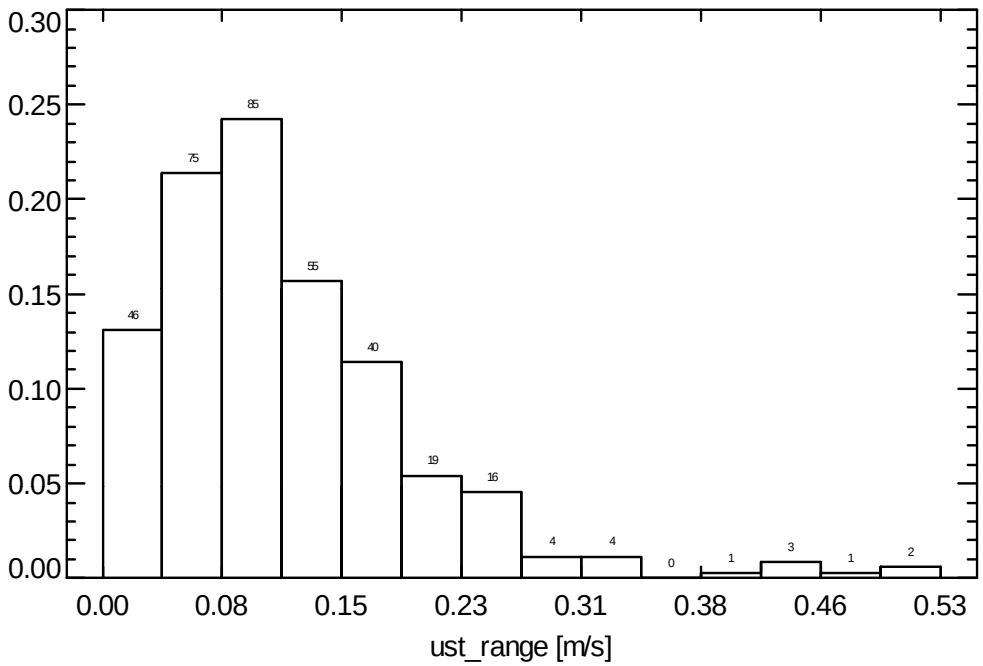
Half hourly data QAQC: differences between u* calcul. methods

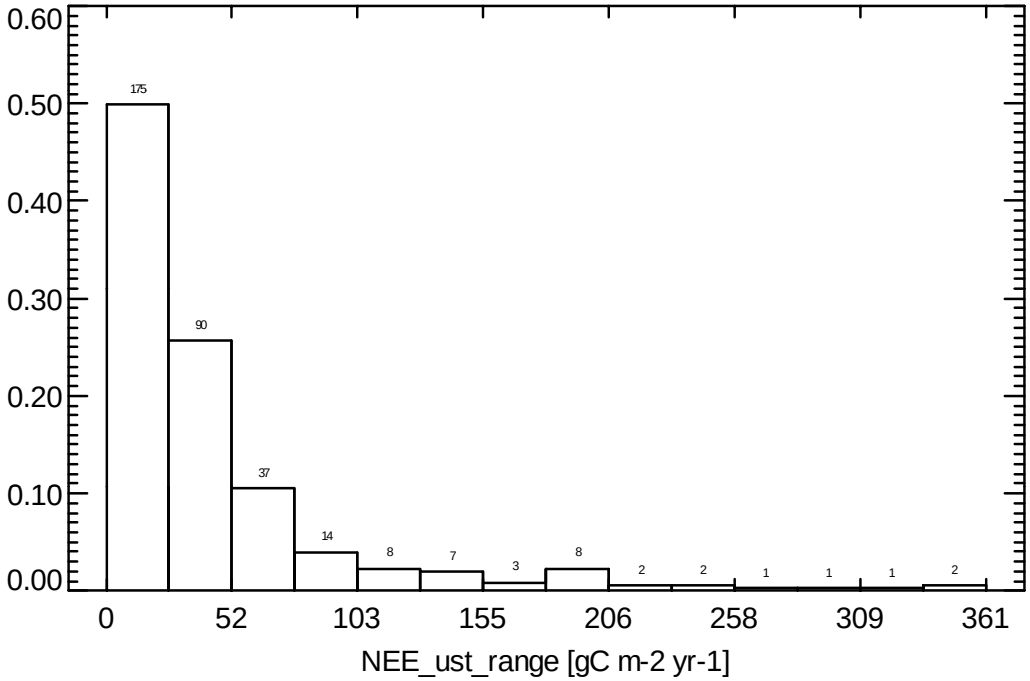


A. Barr

Half hourly data QAQC: Effect of different u* threshold on NEE

Distribution of u* and derived NEE uncertainty





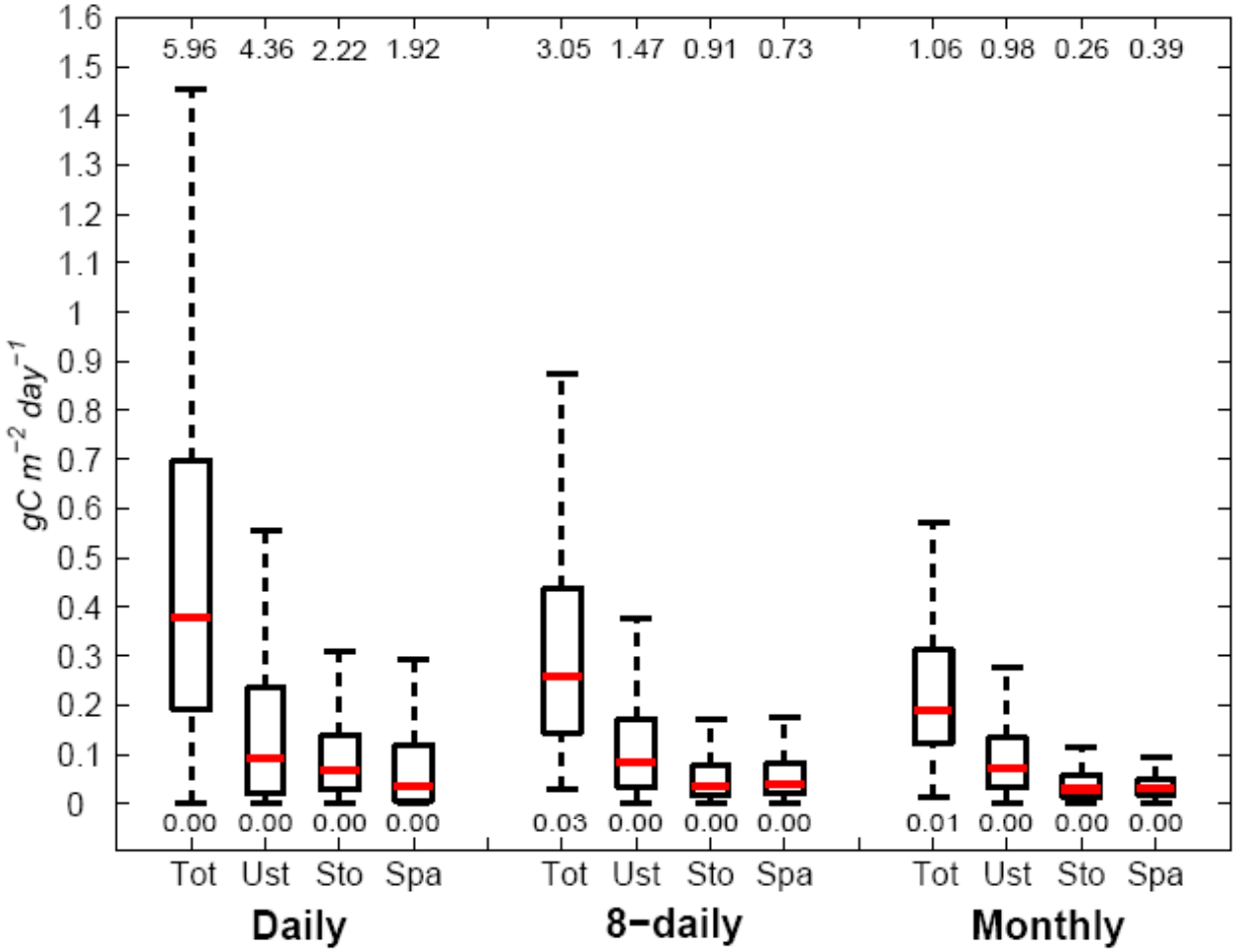
Range = percentile 97.5 – percentile 2.5 ~ 95% conf. interval

M. Reichstein

Half hourly data QAQC: Effect of u*, storage and spike detection

Uncertainty due to quality check and filtering

Difference between minimum and maximum value obtained at different time resolution using different correction setting (u* thresholds, spike thresholds, storage measurement)



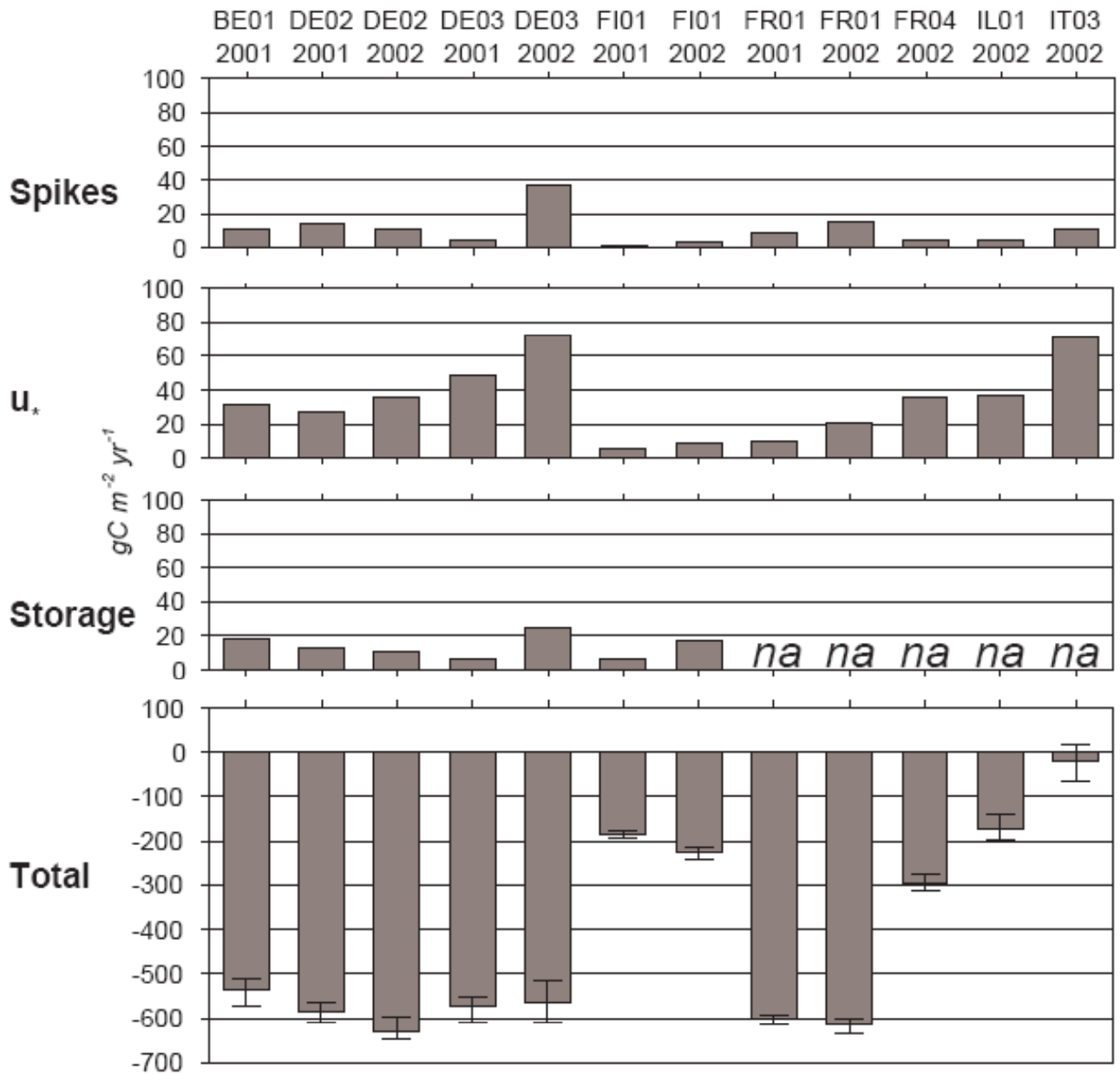
Papale et al. 2006

Half hourly data QAQC: Effect of u*, storage and spike detection

NEE

Uncertainty due to different corrections settings (u* thresholds, spike thresholds, storage measurement) at annual scale.

< 100 gC m-2 (50 gC m-2)



Papale et al. 2006

Gapfilling: do we need to fill gaps?

Gapfilling is not always needed, but it is necessary when we have to integrate to daily-annual scale

Where are the gaps coming from?

- Power problems
- Instrument problems
- Calibrations
- Quality tests and filtering

NOT RANDOM DISTRIBUTED

For this reason we can not just calculate the average of the integration period

We need other methods

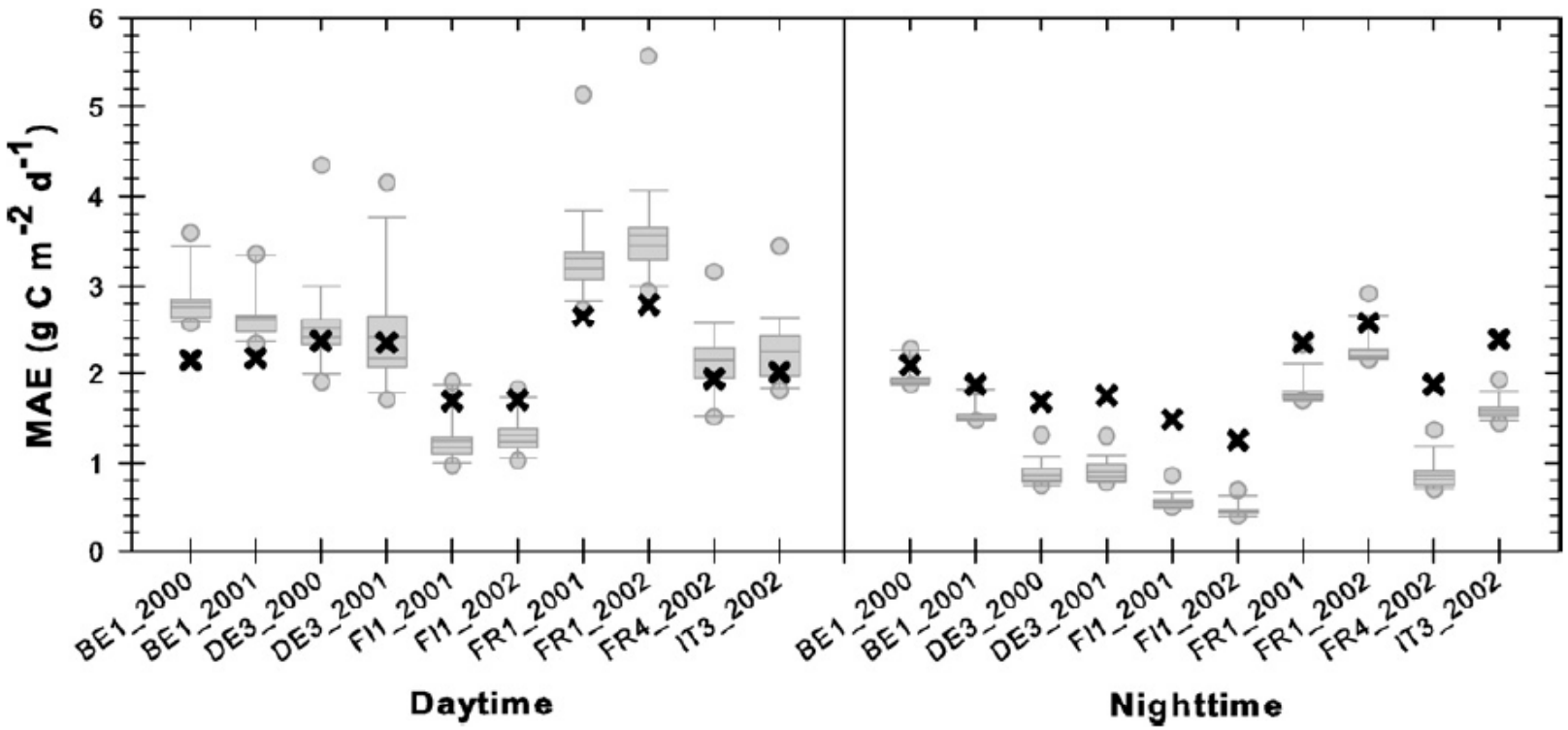
Gapfilling: which methods are available?

Not possible if also meteo is missing

Technique	Meteo Input	Process Based	Auto-Correlations	Conservation of Noise
NLR	x	x		
UKF	x	x	x	x
ANN	x		(x)	
LUT	x		x	
MDS	(x)		x	(x)
SPM	x		x	(x)
MDV			x	(x)
MIM	(x)			x
BETHY	x	x		

Not (or problematic) for model validation and parameterization

Gapfilling: gapfilling methods comparison



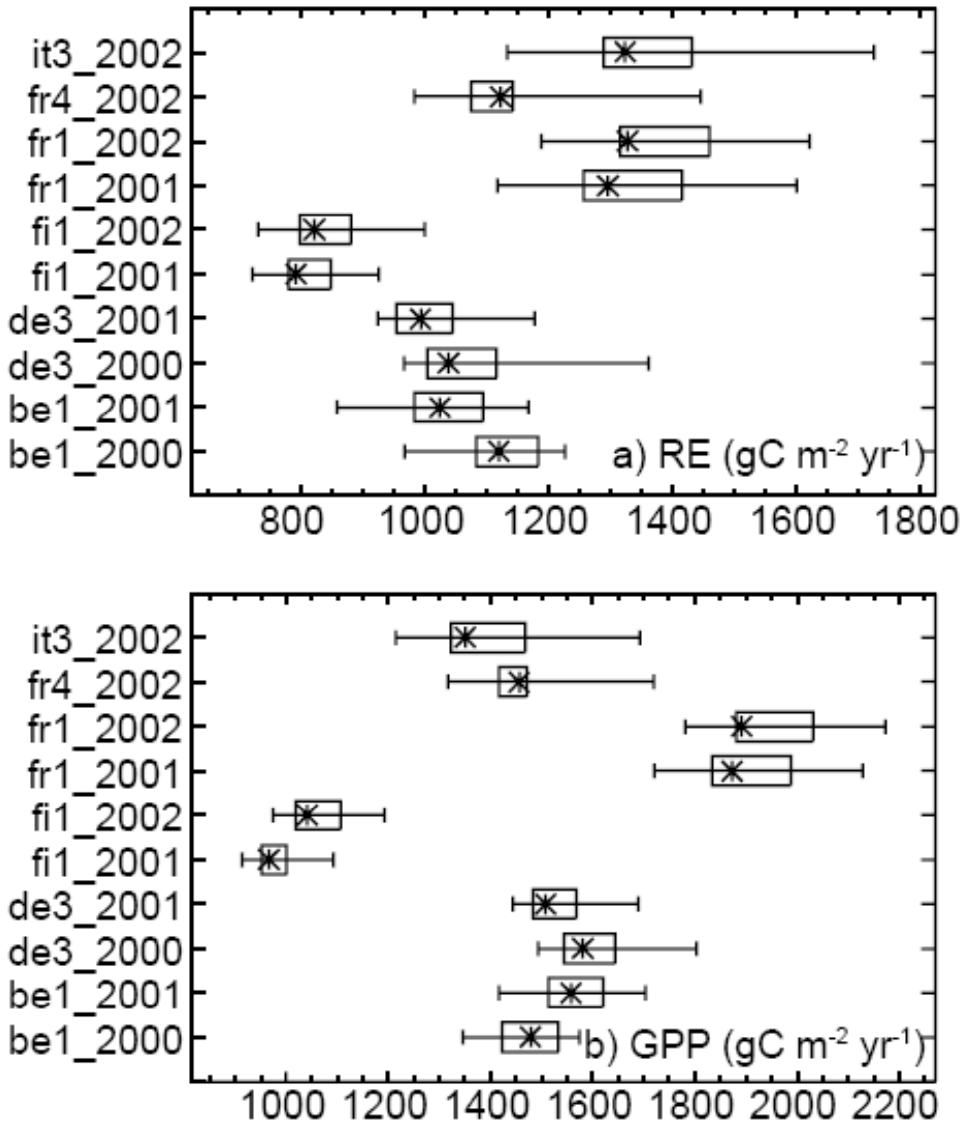
MAE boxplot of the different techniques and random uncertainty estimation using Richardson et al. method

Moffat et al. 2007

Partitioning in GPP and TER: methods comparison

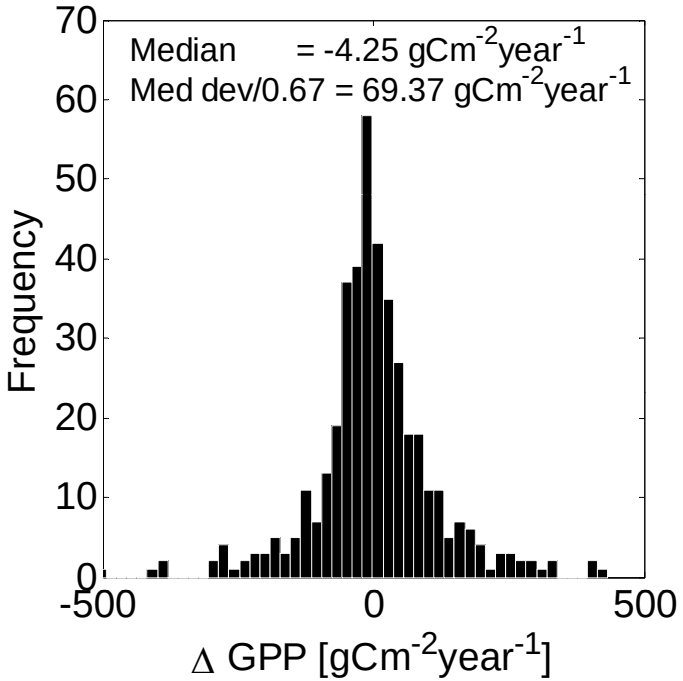
Uncertainty due to partitioning (23 methods, 10 artificial gaps scenario)

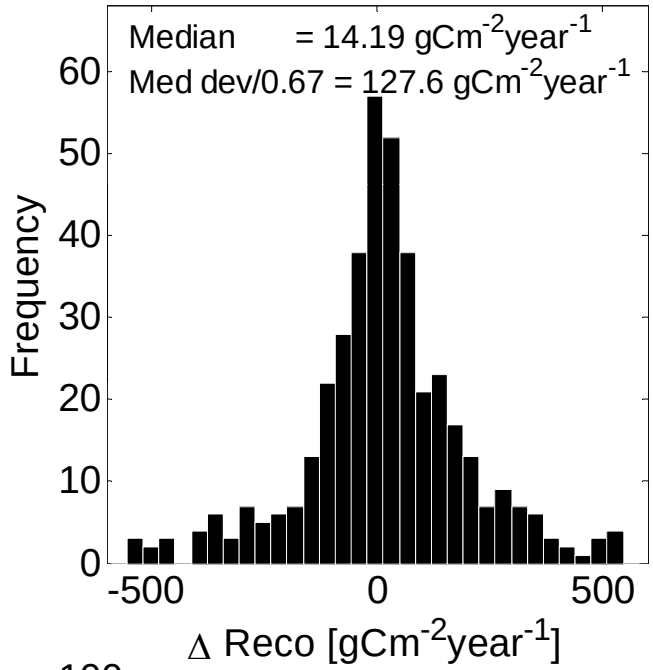
GPP and RE boxplot for each dataset using all the methods. Large part of the methods are in about 100 gC m 2 yr-1



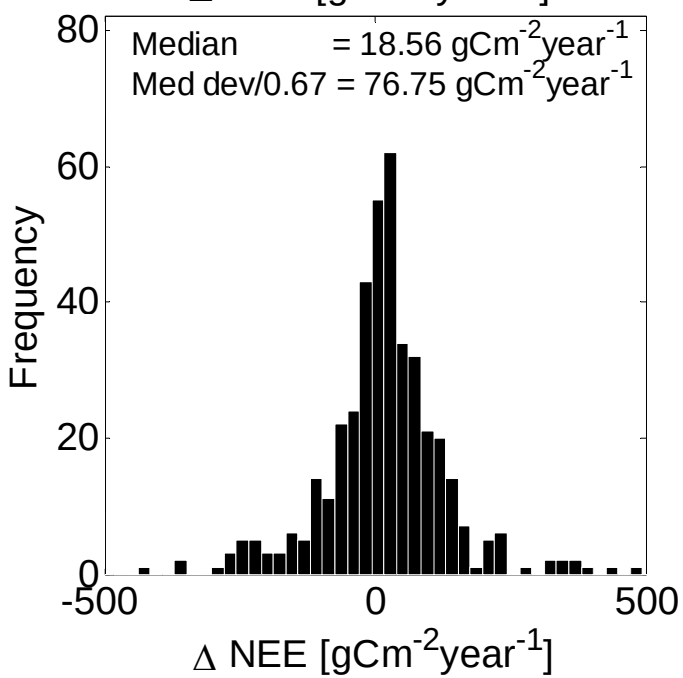
Desai et al. 2007

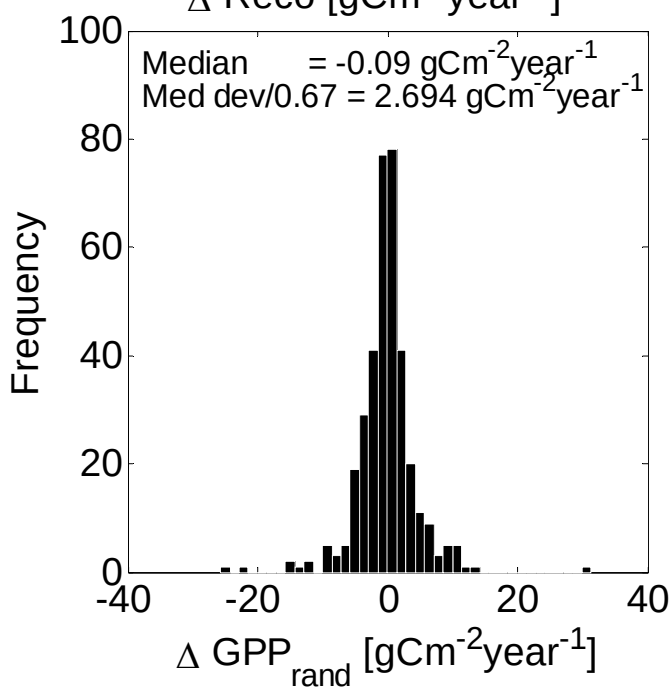
Partitioning in GPP and TER: daytime and nighttime based methods





417 sites years

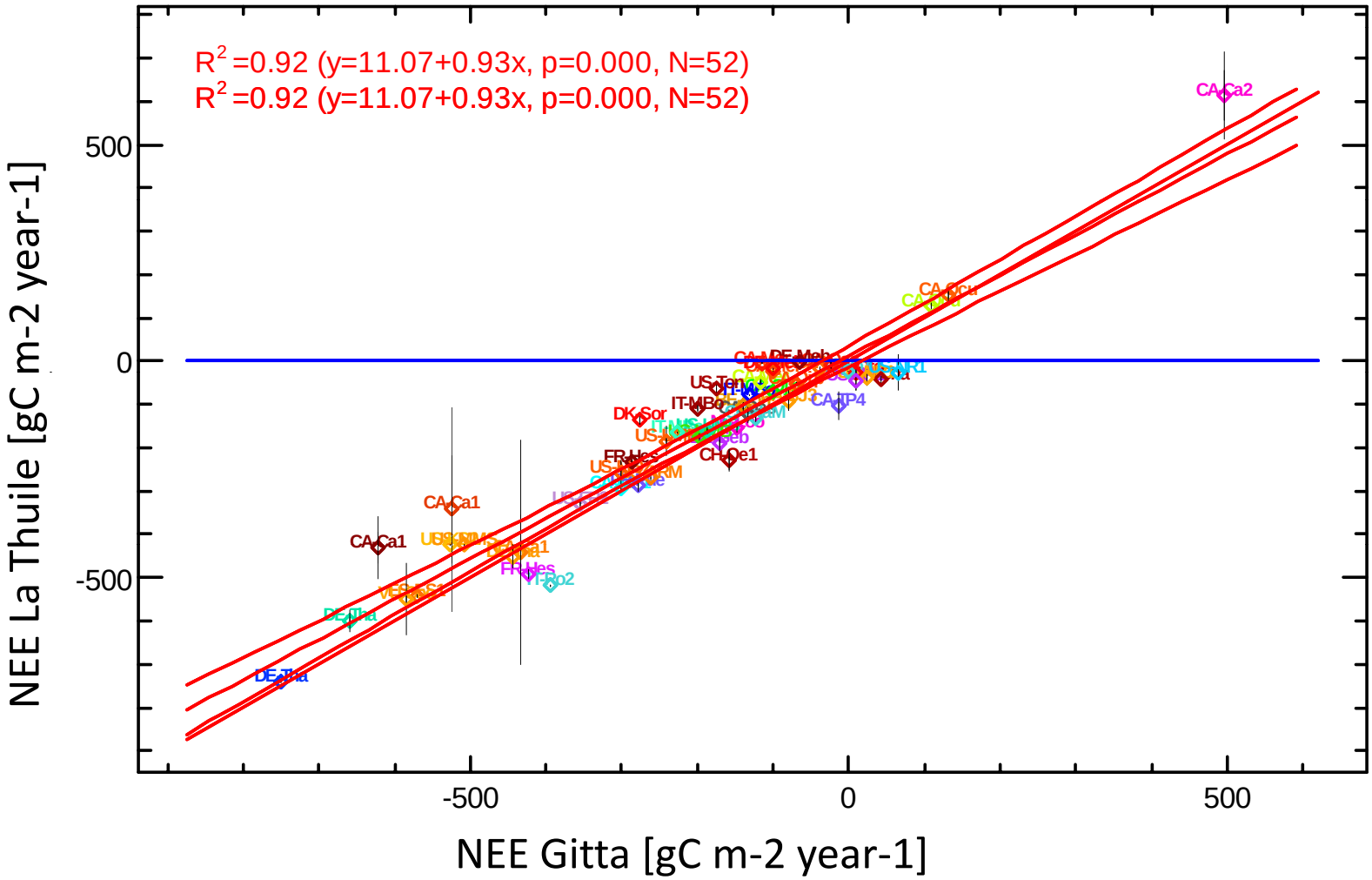




Lasslop et al. submitted

Partitioning in GPP and TER: daytime and nighttime based methods

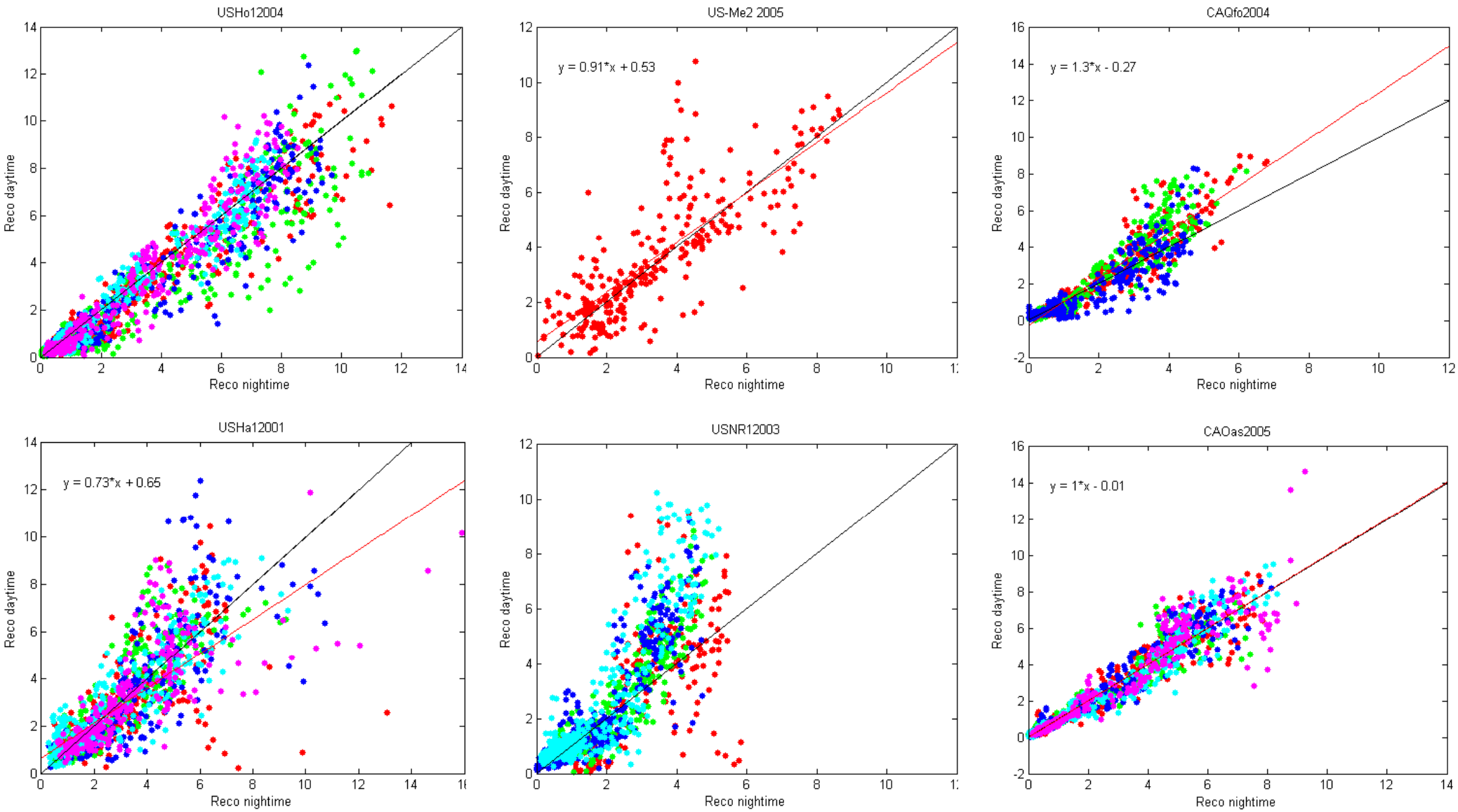
Multi-annual sums



M. Reichstein

Partitioning in GPP and TER: daytime and nighttime based methods

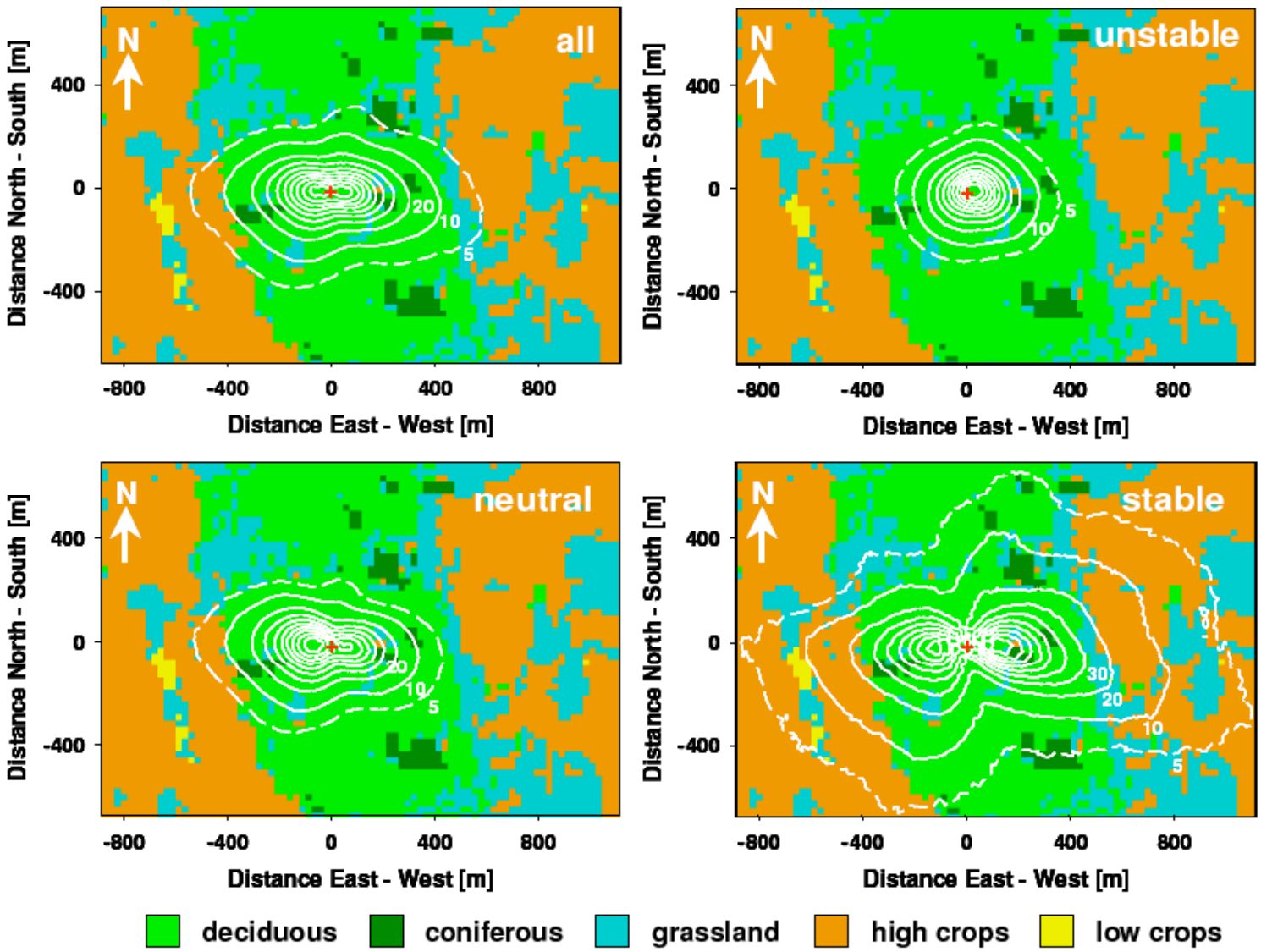
Daily analysis site level…



The tower footprint: what are we measuring?

Berkeley Water Center – Microsoft Research

The tower footprint: modeling approach – DK-Sor example



Lagrangian Stochastic footprint algorithm – lines are percentage, dashed is 5%

Göckede et al. 2008

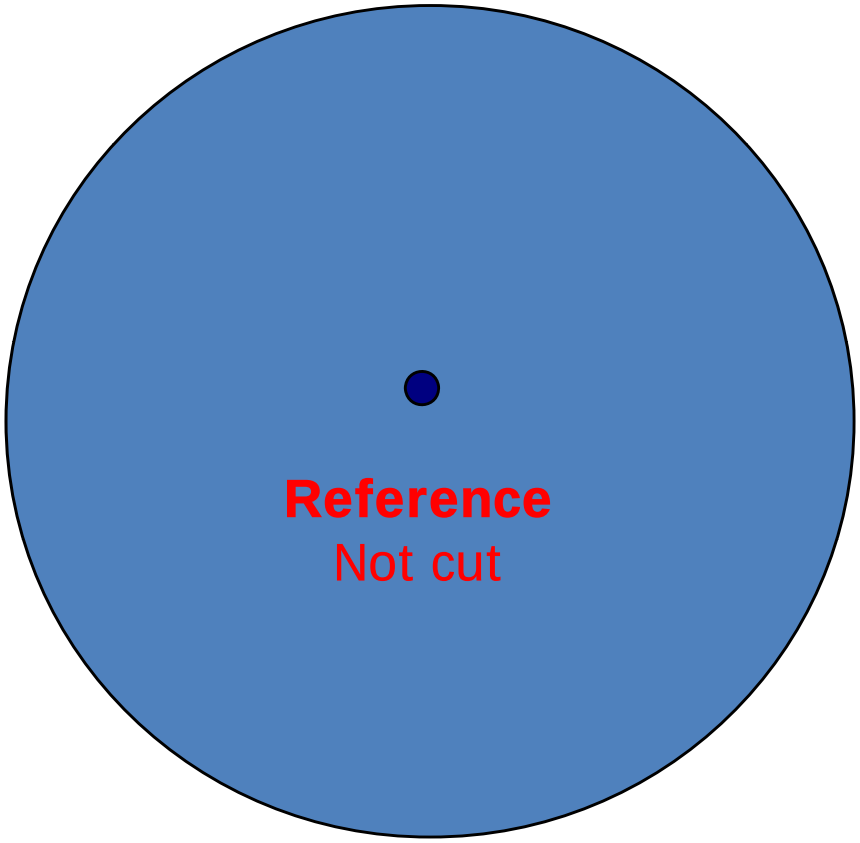
The tower footprint: footprint manipulation experiment

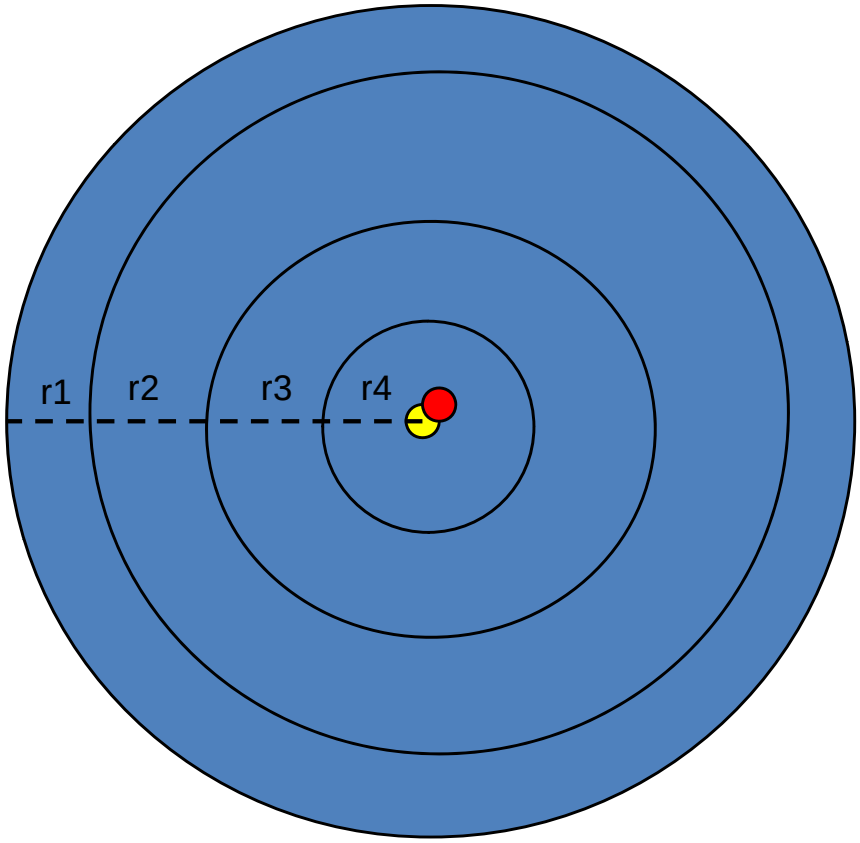
Vegetation height = 1-1.2 m

E.C. reference h_m = 2.3 m

E.C. cut h_m = 2.3 m

E.C. cut h_m = 1.4 m



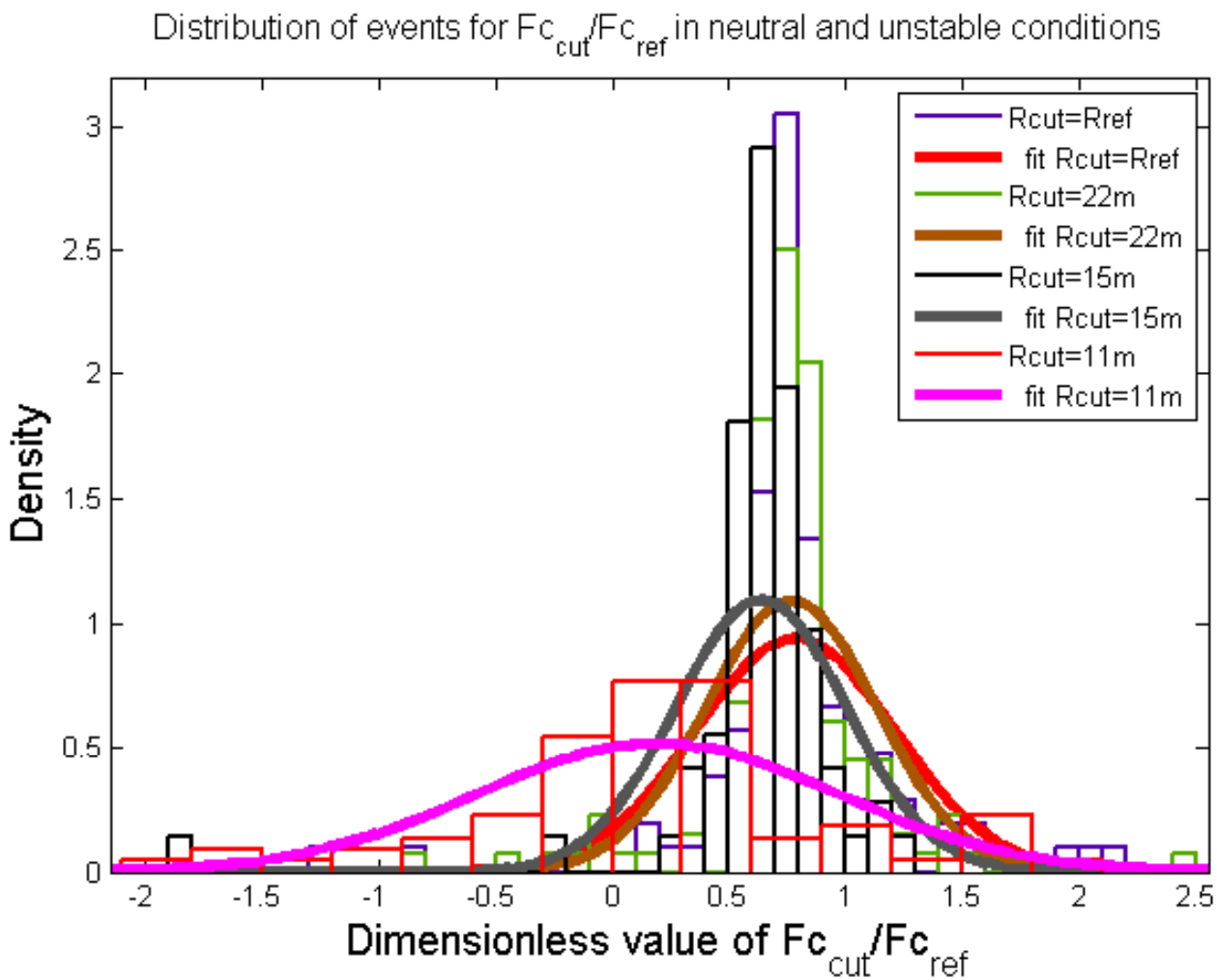


E.C.: Gill R3, Licor 7500

r1=29m, r2=22m, r3=15m, r4=11m

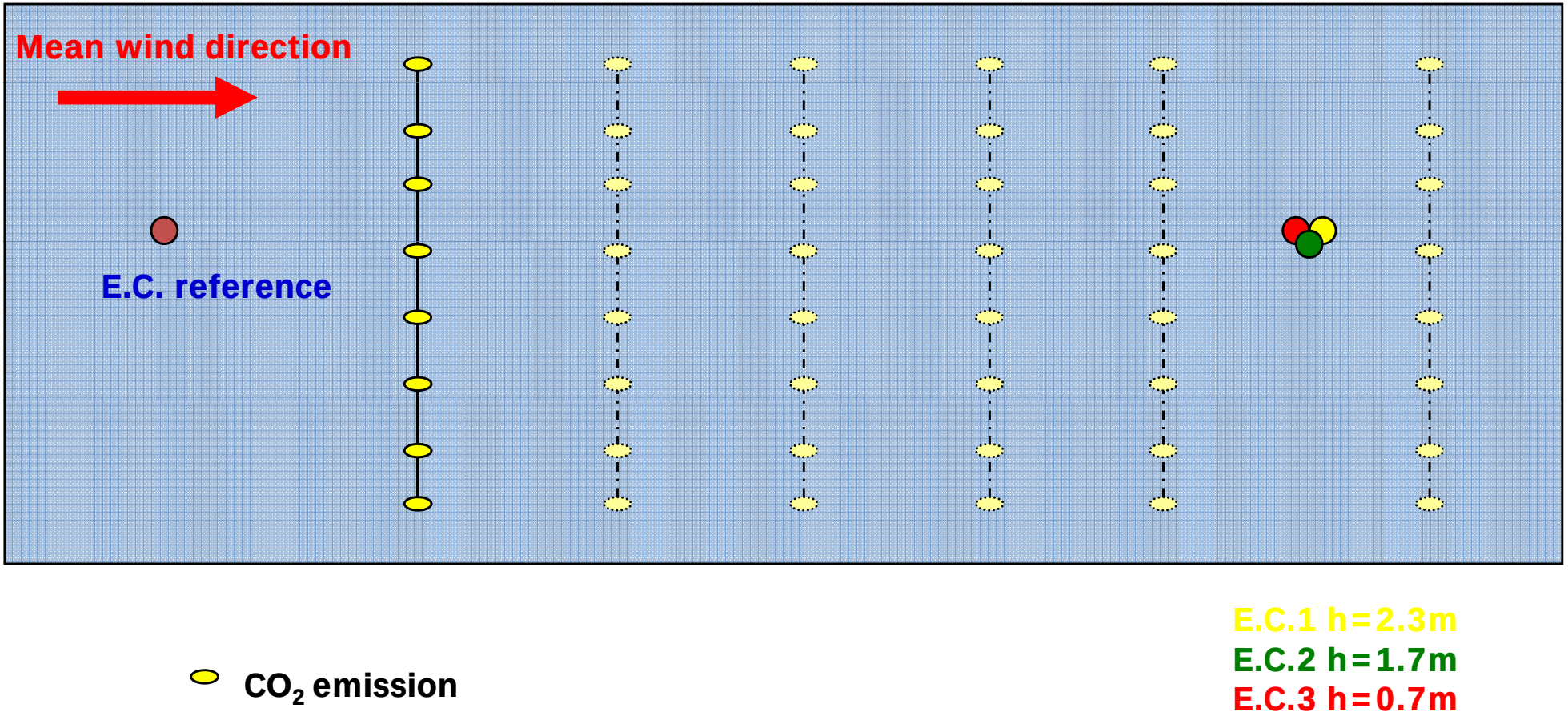
The tower footprint: footprint manipulation experiment

PDF $Fc_{cut}/Fc_{reference}$ with different cut area extension

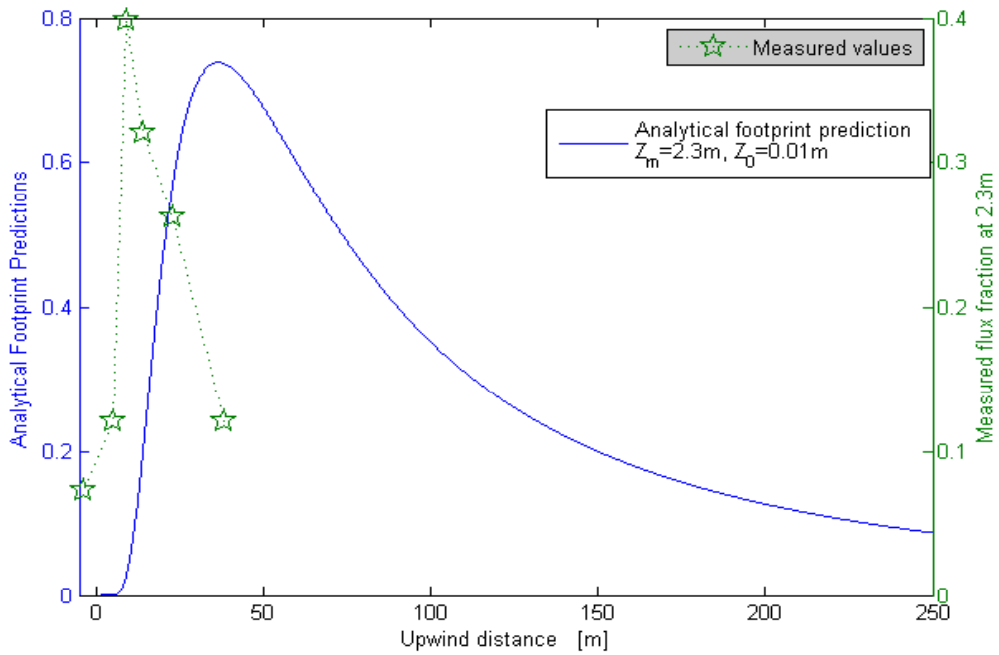


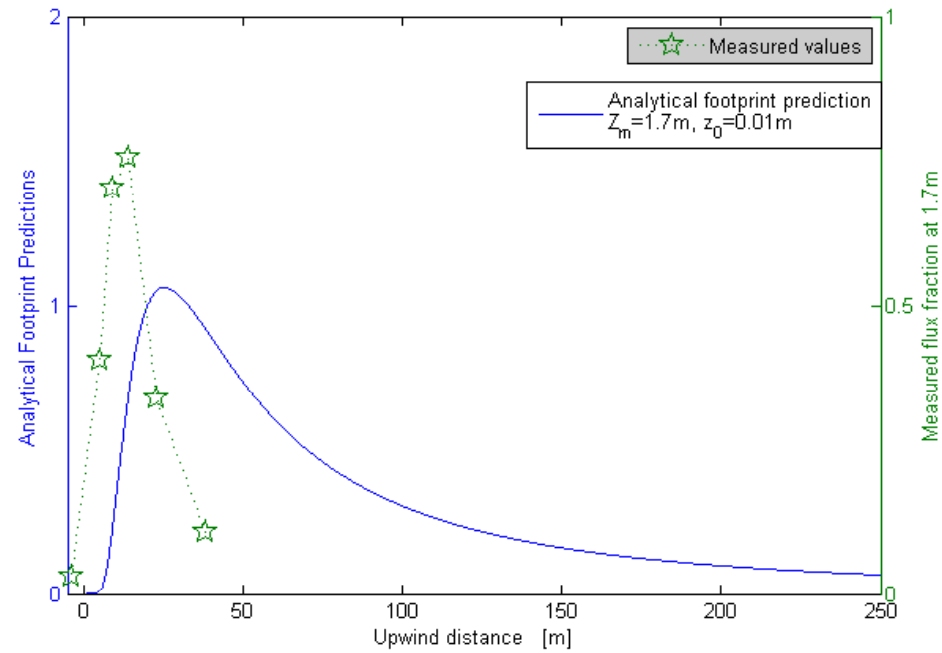
The tower footprint: artificial footprint manipulation

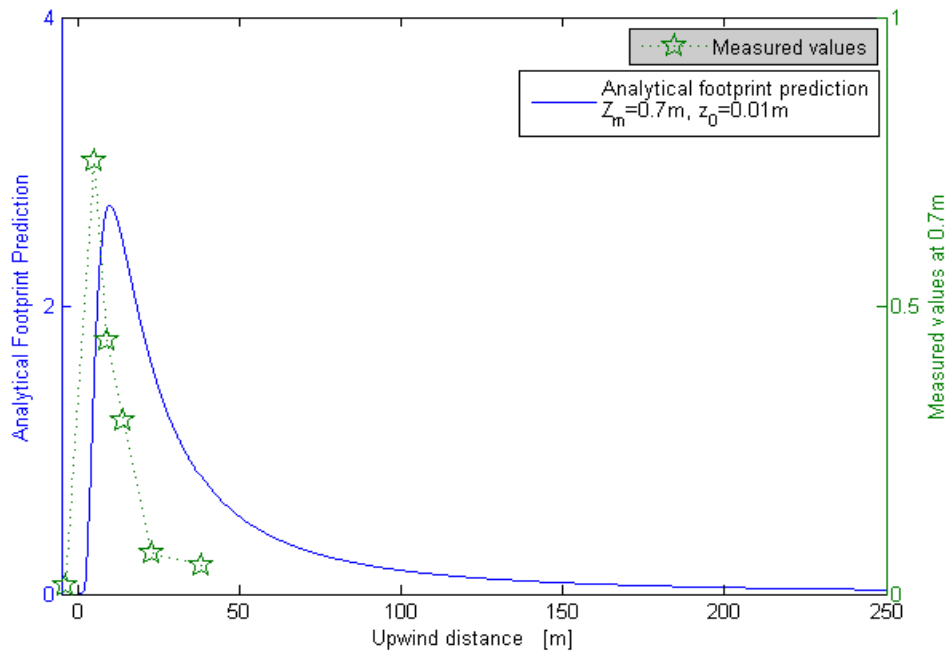
Footprint analysis with artificial CO2 sources



The tower footprint: artificial footprint manipulation







The ancillary data: essential information

Biomass, LAI, soil characteristics, management, disturbances, canopy parameters, chamber measurements, but also system set-up, instruments used, raw data processing scheme…

All these data have been not systematically collected and standardized in the past. Now there are ongoing efforts in the different networks to collect these data.

Consistency cross-check of eddy covariance flux based and biometric estimates of forest ecosystem carbon balance

Data needed: independent estimates of the different component
NEP, GPP and Re from eddy
NPP from biomass inventory
Soil, heterotrophic and autotrophic respirations from chambers

CO2 balance closure equations

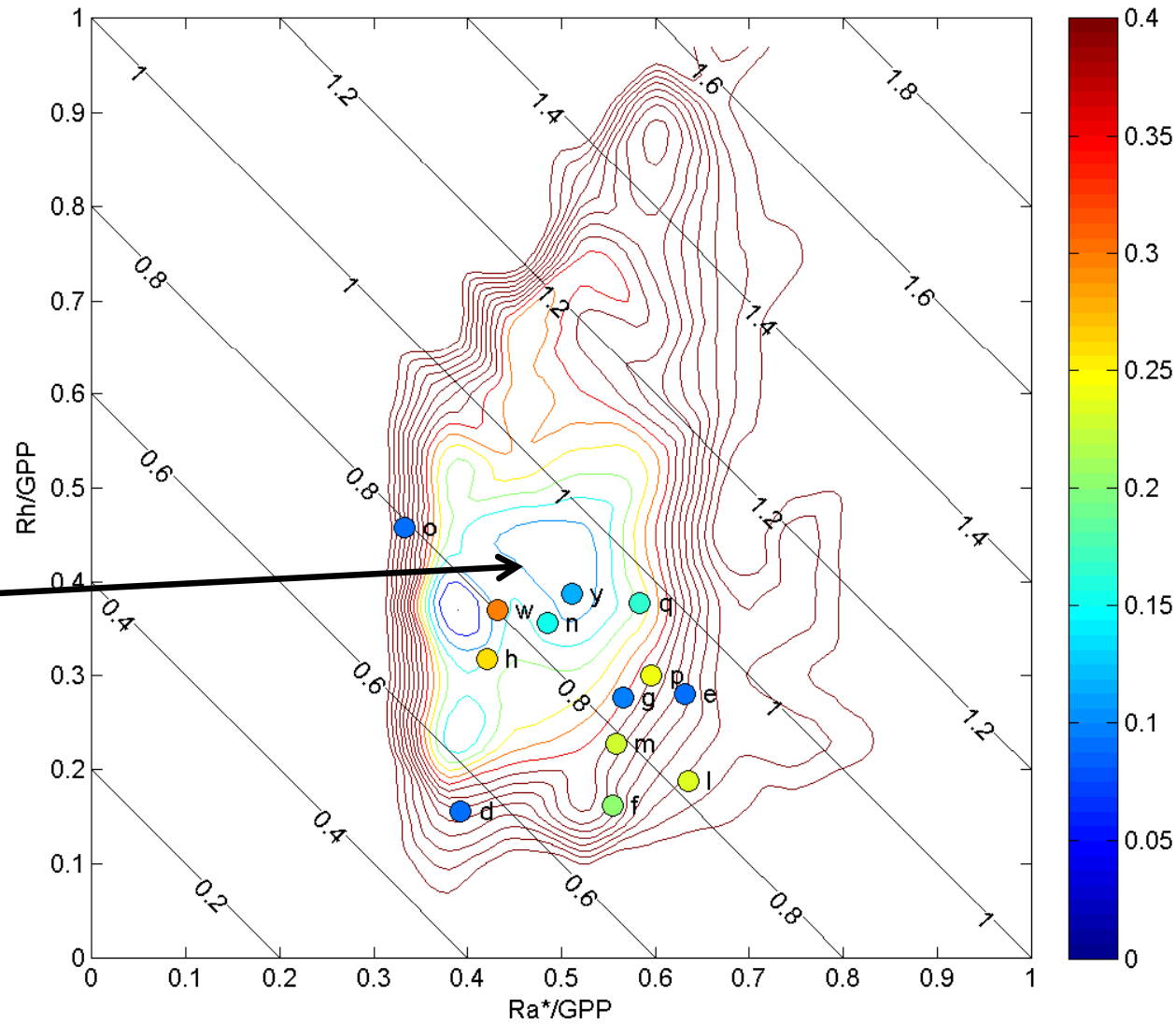
$$GPP + \delta GPP = NPP + \delta NPP + Ra + \delta Ra$$
$$NPP + \delta NPP = NEP + \delta NEP + Rh + \delta Rh$$
$$Re + \delta Re = Ra + \delta Ra + Rh + \delta Rh$$

Thresholds for the 6 closure terms δ imposed

Luyssaert et al, submitted

Uncertainty and biases: cross consistency checks

Experts panel
Rh:GPP and Ra:GPP
estimates



Luyssaert et al, submitted

Conclusions

- There are different uncertainty sources in the eddy covariance data. Advection and its correction using u* filtering is the largest source of uncertainty, the footprint problem seems to be slightly overestimated, gapfilling is not more a big issue

- It is important to take into consideration the uncertainty using the data, but more work is still needed (and ongoing) to better quantify it.

- There is the risk to have biases in some sites due to insufficient correction, in these cases cross-consistency checks using additional data can help

- Standard processing applied to these large databases may be not optimal for some sites, however the users have to decide what they prefer between "*site optimum with heterogeneity between sites*" and "*standard processing but not optimum for some site*"

Conclusions

- Water fluxes: the processing scheme applied to the carbon fluxes is valid also for the water fluxes. Specific uncertainty estimation and standardization methods studies are still missing but gapfilling methods used with carbon work well also with water. In addition, during the night there are no/very small fluxes so the u* problem is not a big issue.

- All the results from the QAQC tests, the corrections applied, the gapfilling quality, are indicated in the dataset using quality flags.

- It is very important to have strong collaboration between “eddy community” and “modelers community” not only for data policy issues (that we will also discuss) but because putting together experiences and knowledge can only improve the quality of the work and ensure the correct use of both data and model results.

…. and this is why we are here ….