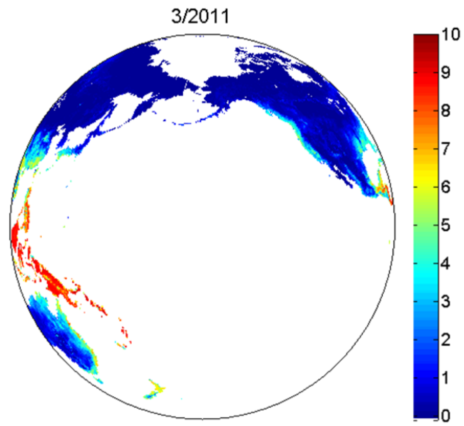


Statistical modelling (upscaling) of Gross Primary Production (GPP)

Martin Jung

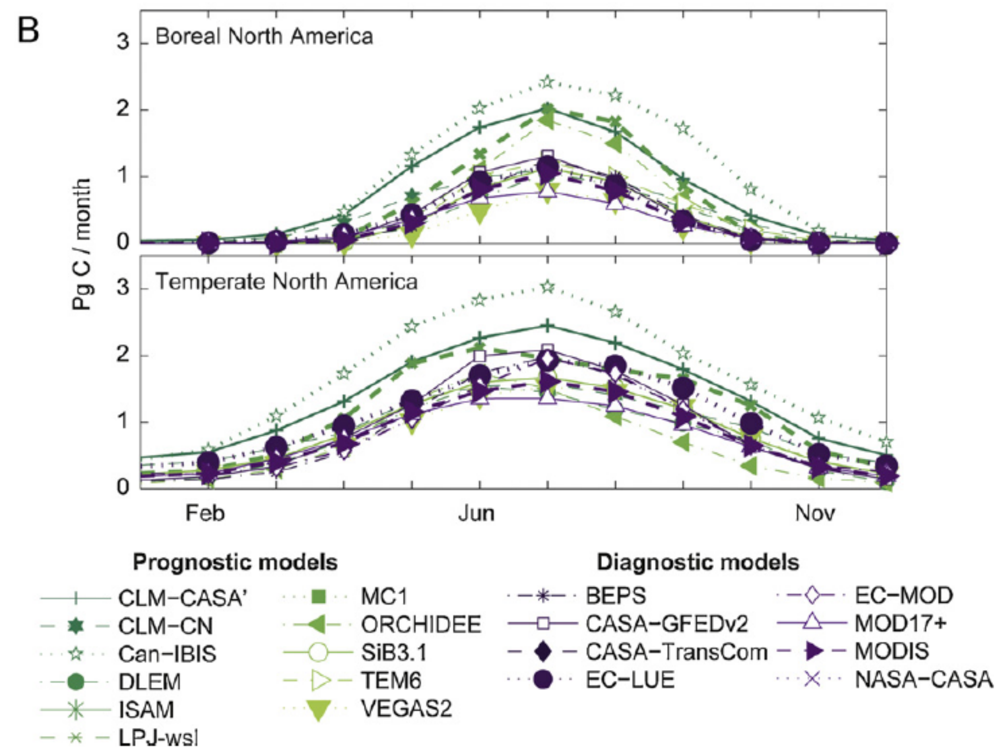
Max-Planck-Institute of Biogeochemistry, Jena, Germany

Department of Biogeochemical Integration

mjung@bgc-jena.mpg.de

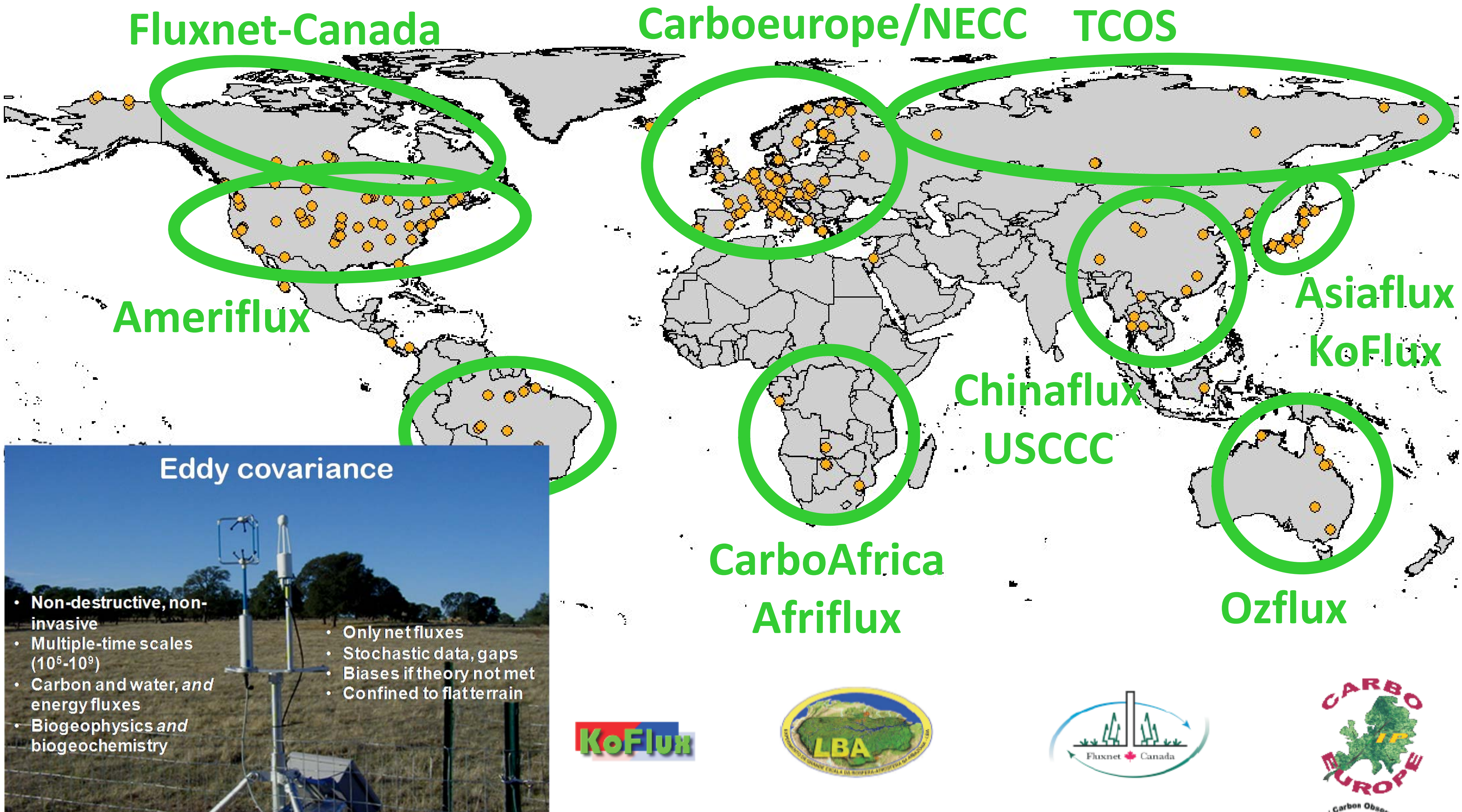
Motivation

- Substantial uncertainty of GPP simulations of Land Surface / Terrestrial Ecosytem Models
- Independent, cor
observation-base
assumptions abo
- Improved diagno
checks, LSM evalu



Huntzinger et al 2012

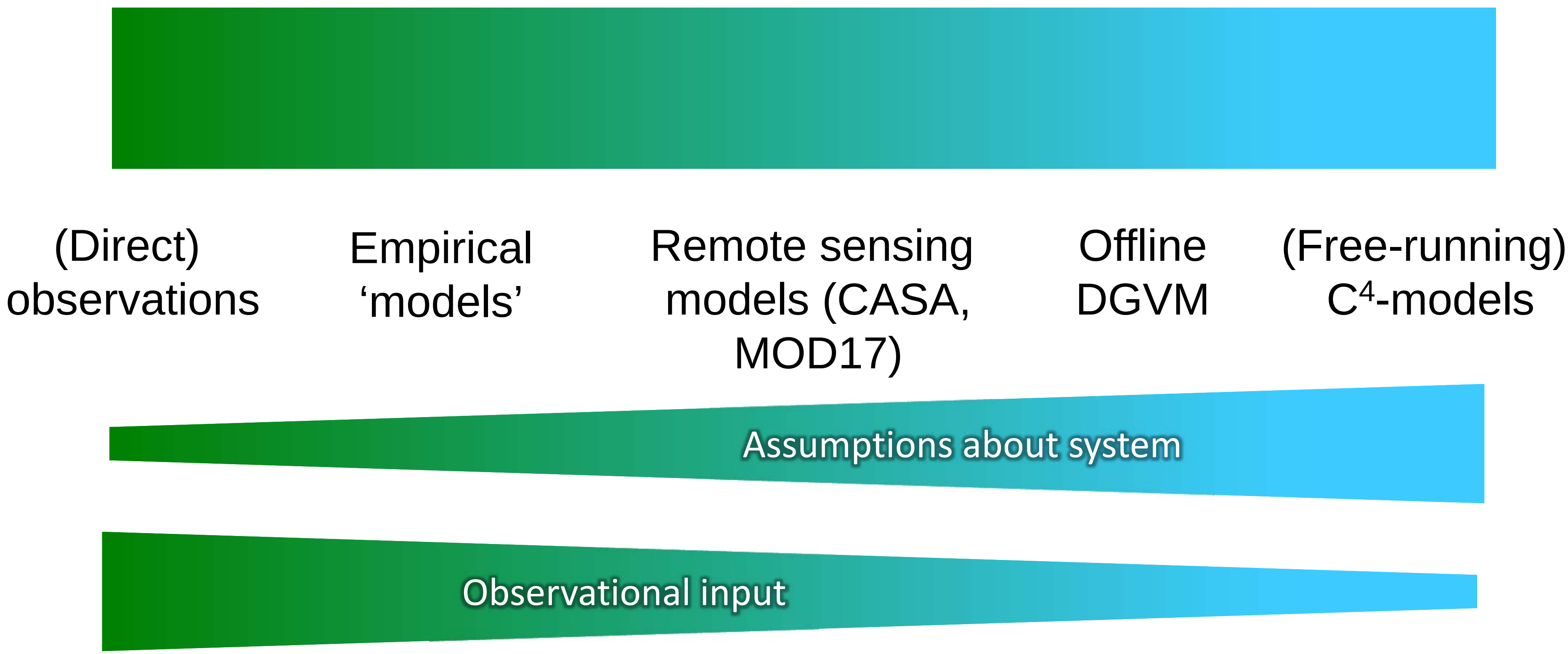
FLUXNET: a network of network of eddy covariance sites



KoFlux LBA Fluxnet Canada CARBO EUROPE IP

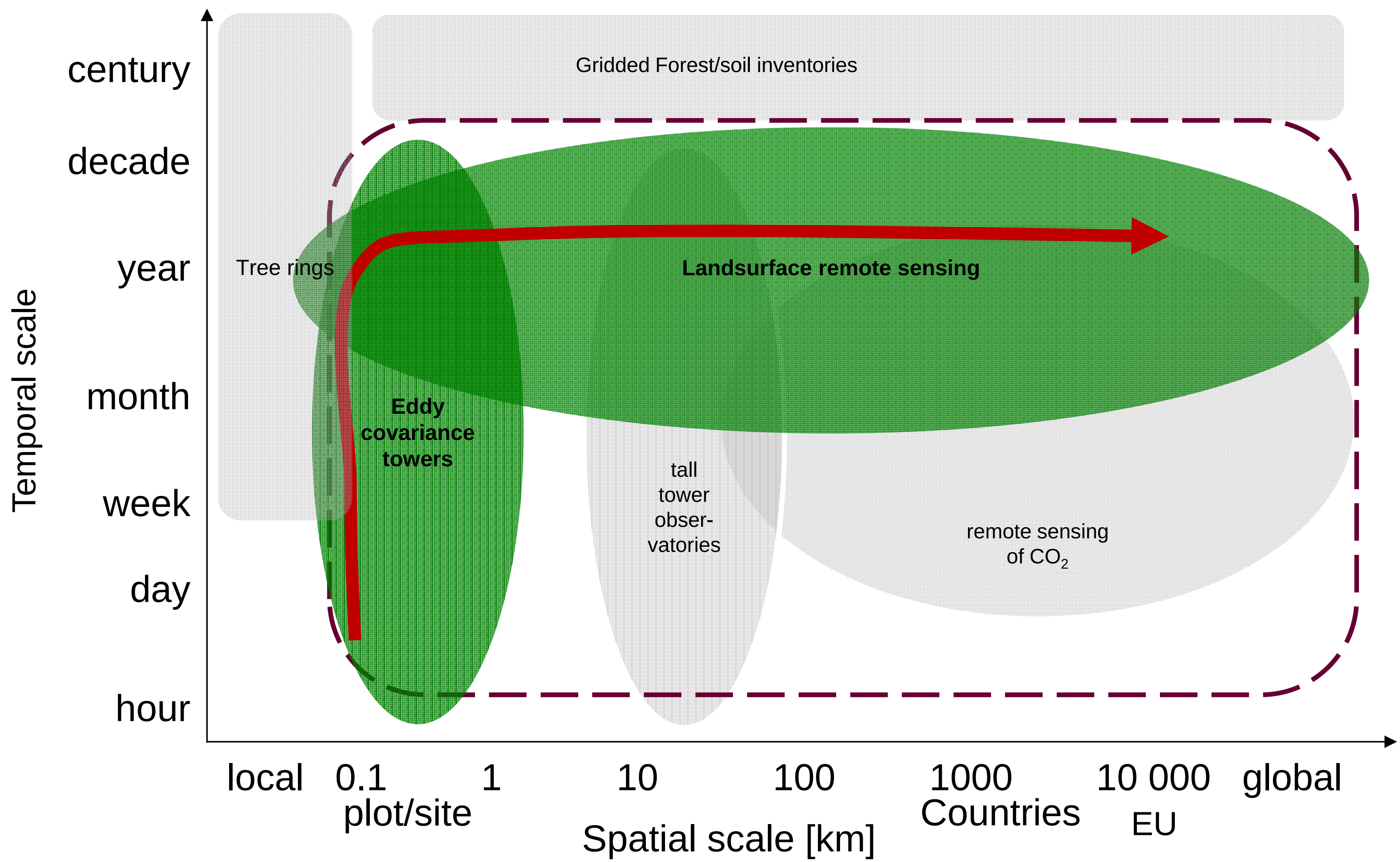
AfriFlux ChinaFLUX OzFlux Australian and New Zealand flux Research and Monitoring USCCC NECC Nordic Centre for Studies of Ecosystem Carbon Exchange and its Interactions with the Climate System TCOS SIBERIA Terrestrial Carbon Observation System

FLUXNET gridded products rationale:

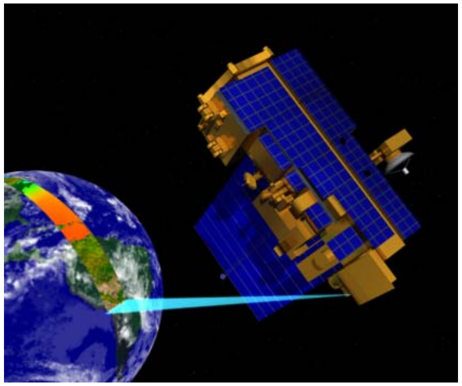


… be as much as possible on observational side!

From point to globe via integration with remote sensing



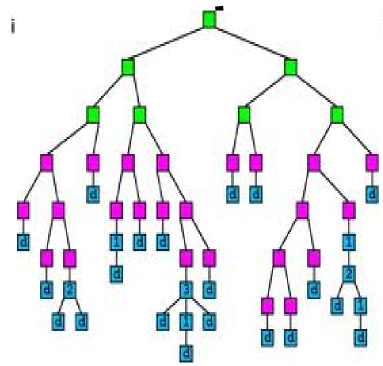
General Principle



Site-level explanatory variables
- **Meteorology**
- **Vegetation type**
- **Remote sensing indices**

Training

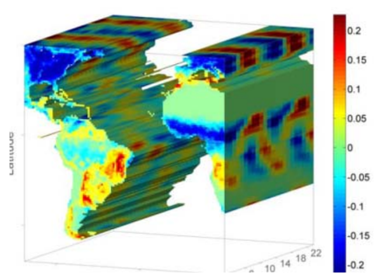
The same gridded explanatory variables



Training

Target variable
ecosystem-atmosphere flux

Gridded target variable



Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate

Christian Beer,[1*] Markus Reichstein,[1] Enrico Tomelleri,[1] Philippe Ciais,[2] Martin Jung,[1] Nuno Carvalhais,[1,3] Christian Rödenbeck,[4] M. Altaf Arain,[5] Dennis Baldocchi,[6] Gordon B. Bonan,[7] Alberte Bondeau,[8] Alessandro Cescatti,[9] Gitta Lasslop,[1] Anders Lindroth,[10] Mark Lomas,[11] Sebastiaan Luyssaert,[12] Hank Margolis,[13] Keith W. Oleson,[7] Olivier Roupsard,[14,15] Elmar Veenendaal,[16] Nicolas Viovy,[2] Christopher Williams,[17] F. Ian Woodward,[11] Dario Papale[18]

Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 110 (2007) 109–122
Remote Sensing of Environment
www.elsevier.com/locate/rse

Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach

…a Yang [a,b,*], Kazuhito Ichii [b,c], Michael A. White [d], Hirofumi Hashimoto [b,e], …ndrew R. Michaelis [b,e], Petr Votava [b,e], A-Xing Zhu [f,a], Alfredo Huete [g], Steven W. Running [h], Ramakrishna R. Nemani [b]

Biogeosciences, 6, 2001–2013, 2009
www.biogeosciences.net/6/2001/2009/

Biogeosciences

Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model

M. Jung[1], M. Reichstein[1], and A. Bondeau[2]

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 116, G00J07, doi:10.1029/2010JG0…

Remote Sensing of Environment 114 (2010) 576–591
Contents lists available at ScienceDirect
Remote Sensing of Environment
journal homepage: www.elsevier.com/locate/rse
ELSEVIER

A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data

Jingfeng Xiao [a,*], Qianlai Zhuang [b], Beverly E. Law [c], Jiquan Chen [d], Dennis D. Baldocchi [e], David R. Cook [f], Ram Oren [g], Andrew D. Richardson [h], Sonia Wharton [i], Siyan Ma [e], Timothy A. Martin [j], Shashi B. Verma [k], Andrew E. Suyker [k], Russell L. Scott [l], Russell K. Monson [m], Marcy Litvak [n], David Y. Hollinger [o], Ge Sun [p], Kenneth J. Davis [q], Paul V. Bolstad [r], Sean P. Burns [m], Peter S. Curtis [s], Bert G. Drake [t], Matthias Falk [i], Marc L. Fischer [u], David R. Foster [v], Lianhong Gu [w], Julian L. Hadley [x], Gabriel G. Katul [g], Roser Matamala [y], Steve McNulty [p], Tilden P. Meyers [z], J. William Munger [aa], Asko Noormets [ab], Walter C. Oechel [ac], Kyaw Tha Paw U [i], Hans Peter Schmid [ad,ae], Gregory Starr [af], Margaret S. Torn [ag], Steven C. Wofsy [ah]

Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations

Martin Jung,[1] Markus Reichstein,[1] Hank A. Margolis,[2] Alessandro Cescatti,[3] Andrew D. Richardson,[4] M. Altaf Arain,[5] Almut Arneth,[6,7] Christian Bernhofer,[8] Damien Bonal,[9] Jiquan Chen,[10] Damiano Gianelle,[11] Nadine Gobron,[12] Gerald Kiely,[13] Werner Kutsch,[14] Gitta Lasslop,[1] Beverly E. Law,[15] Anders Lindroth,[6] Lutz Merbold,[16] Leonardo Montagnani,[17,18] Eddy J. Moors,[19] Dario Papale,[20] Matteo Sottocornola,[11] Francesco Vaccari,[21] and Christopher Williams[22]

'Proof of concept'

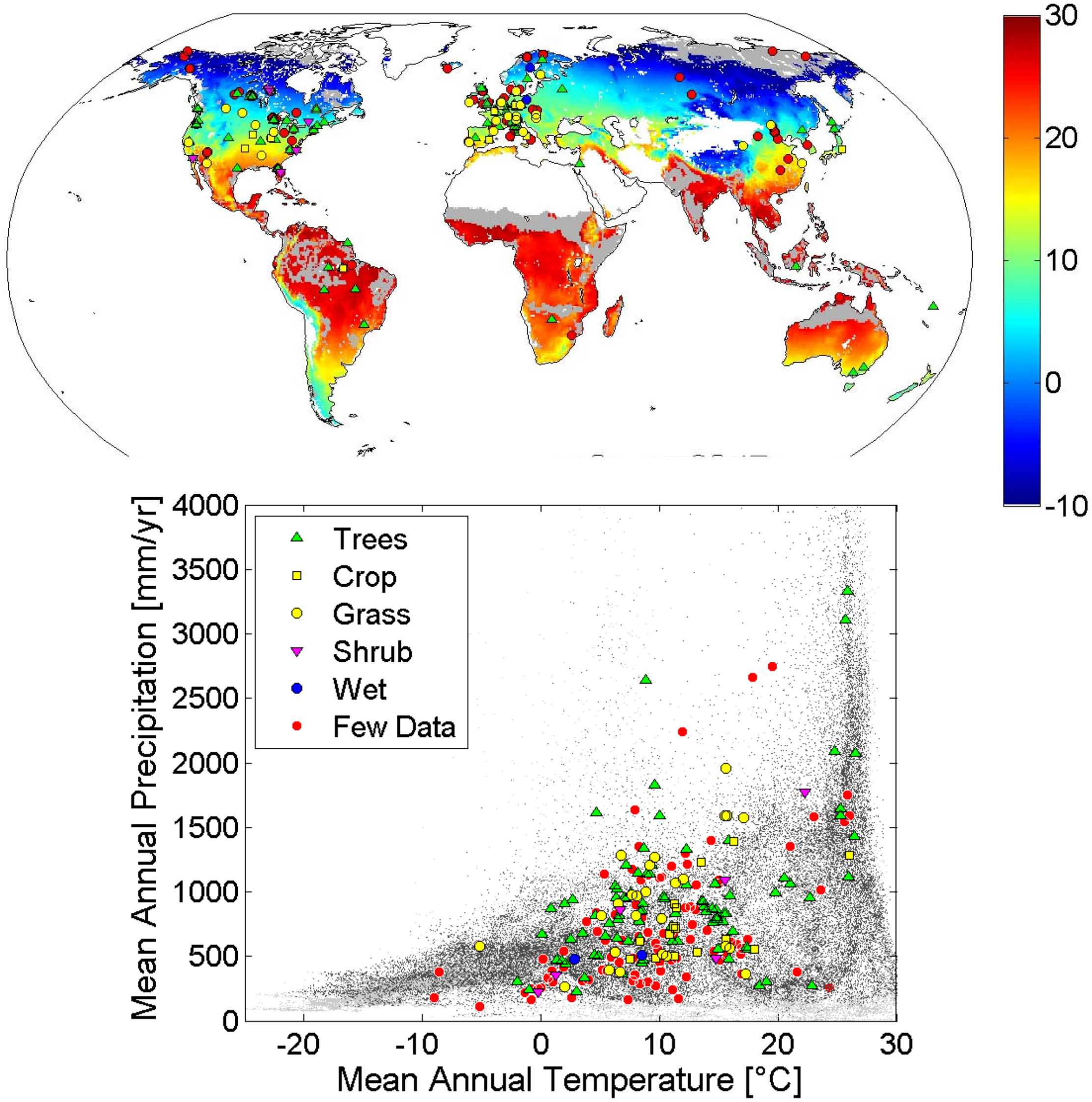
Biogeosciences, 6, 2001–2013, 2009
www.biogeosciences.net/6/2001/2009/

Biogeosciences

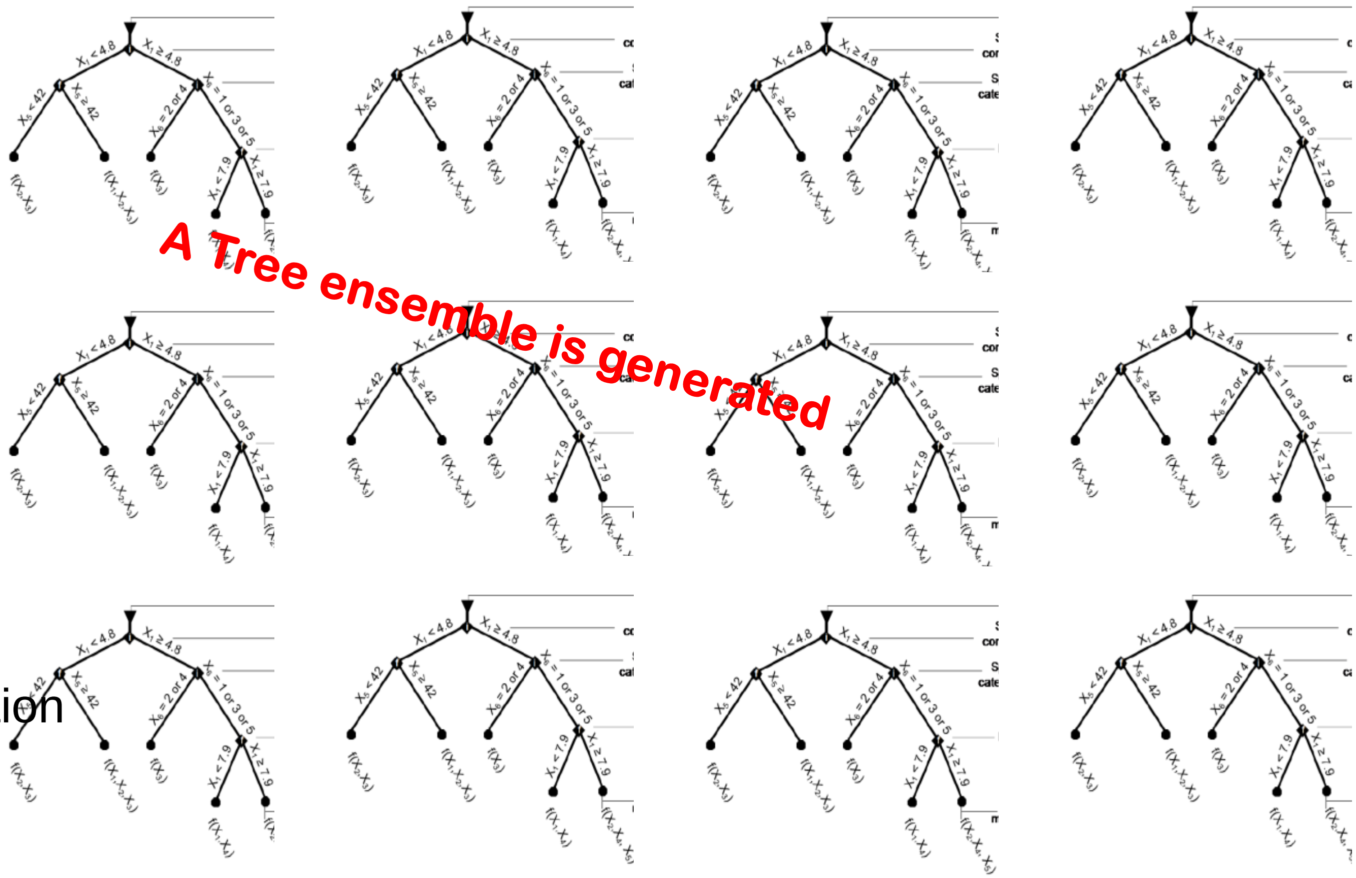
Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model

M. Jung[1], M. Reichstein[1], and A. Bondeau[2]

FLUXNET representativeness



Upscaling strategy using a model tree ensemble (MTE)



Advantages:

- ‚Functional‘ stratification
- Intuitive
- Can cope with switches
- Seemless integration of categorical variables

Jung et al. (2009) BG

Mimicking a process model (LPJmL)

Site level

FLUXNET database

Quality control

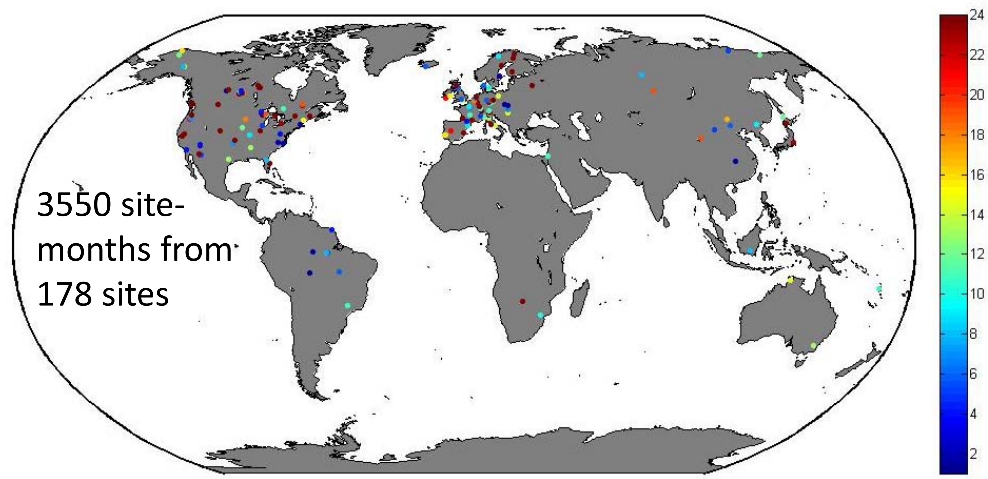
Extracting simulated GPP & model forcing data at locations & times

Train Model Tree Ensemble

Global scale

Apply MTE global

Compare to LPJmL simulation



25 trees, $R^2 = 0.966$

$R^2 = 0.92$

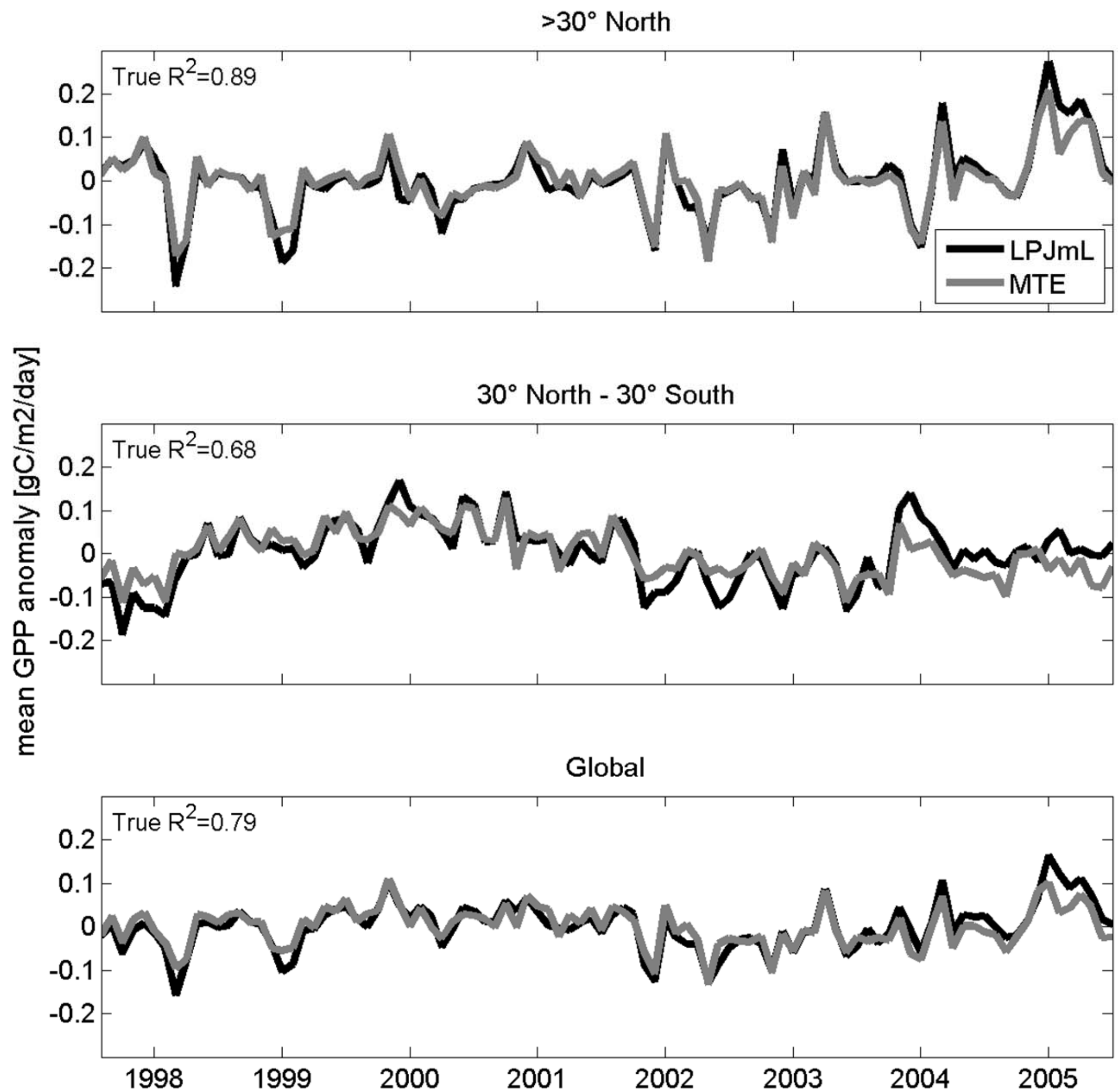
Jung et al. (2009) BG

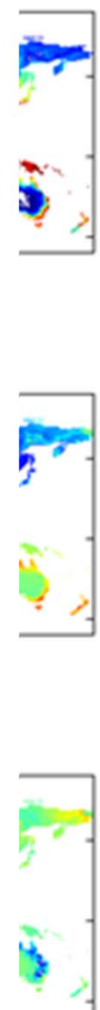
Mimicking a process model (LPJmL)

Mean Annual

PCA1 of Seasonal cycle

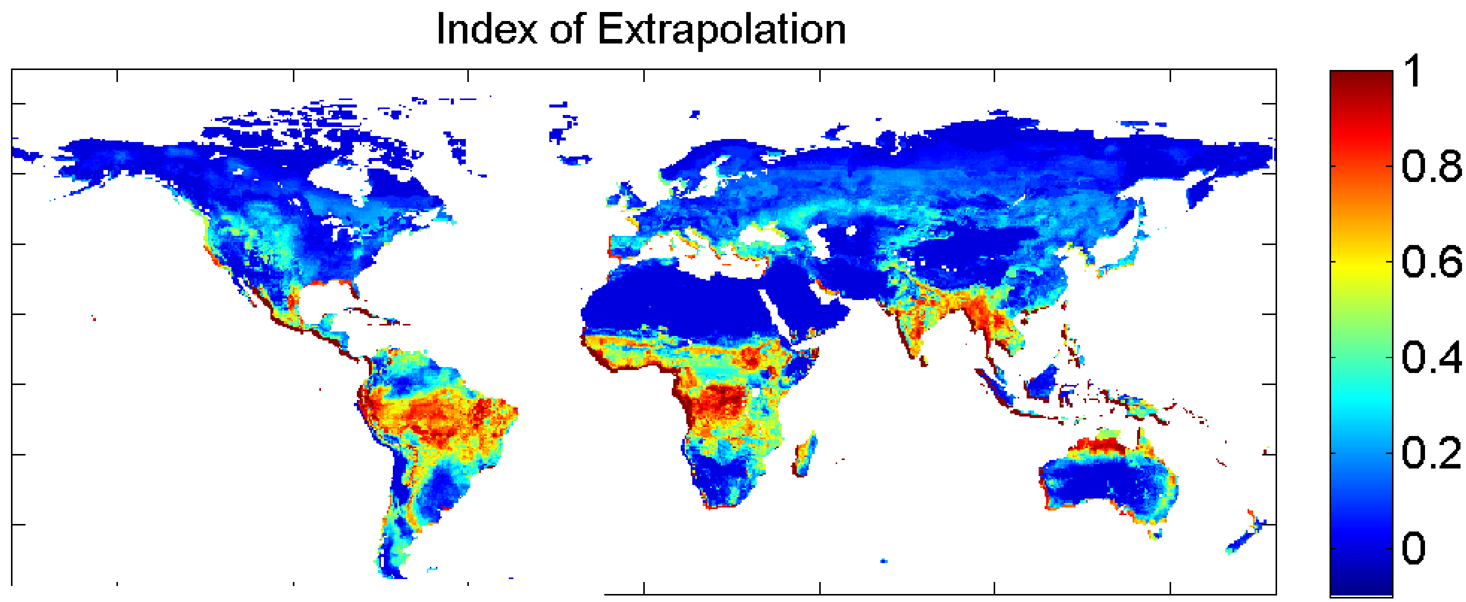
PCA1 of IAV

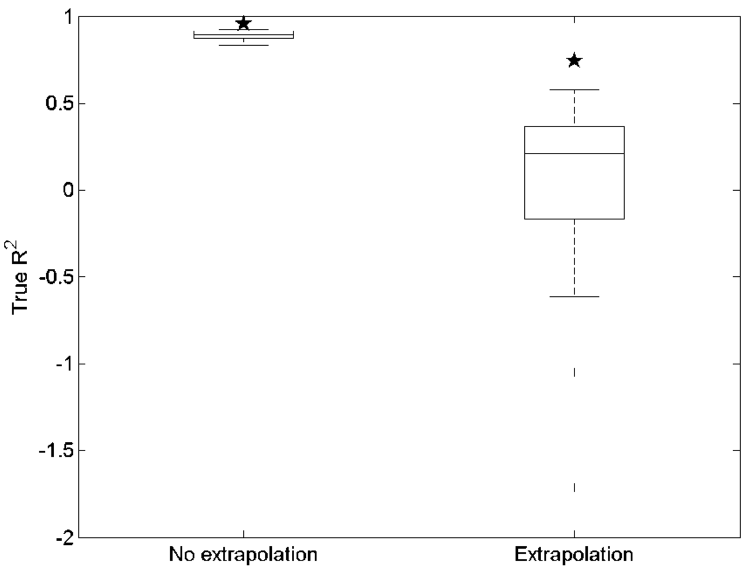


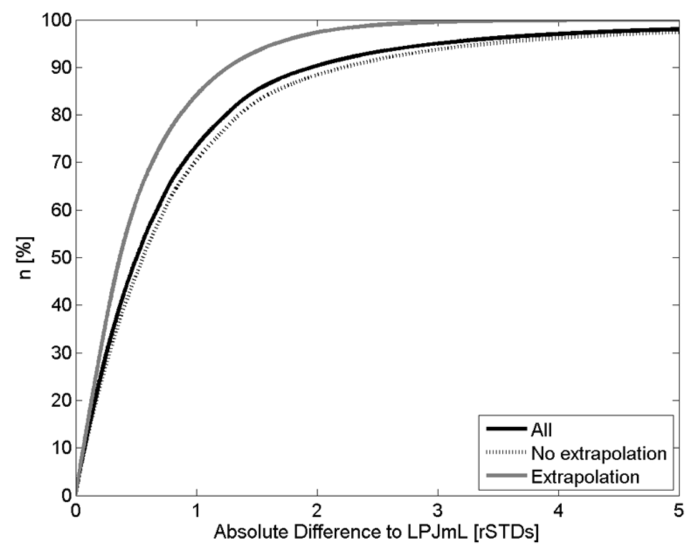


Jung et al. (2009) BG

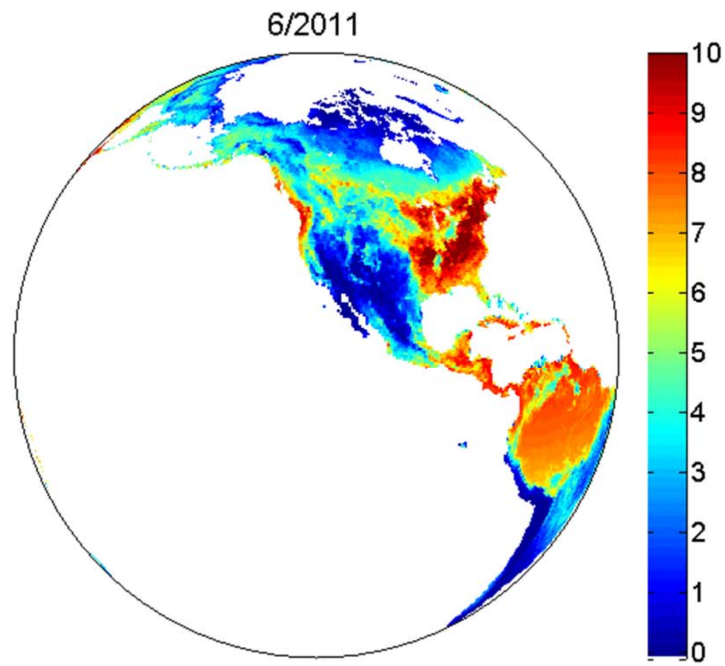
The value of 'ensemble magic'



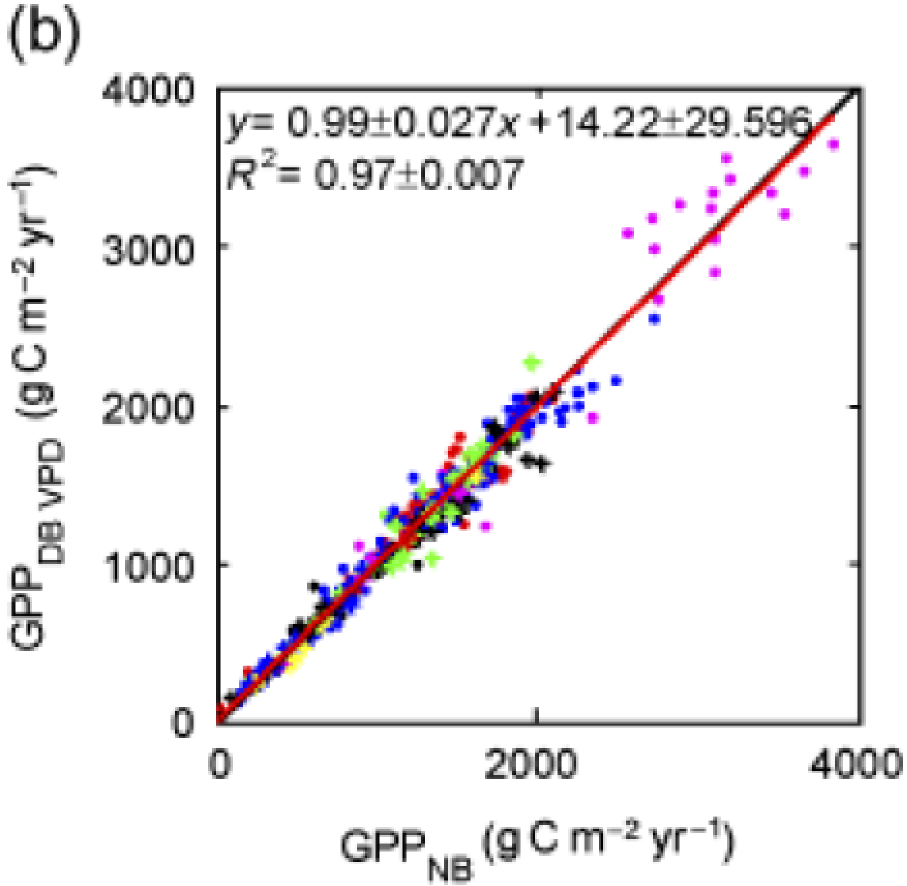




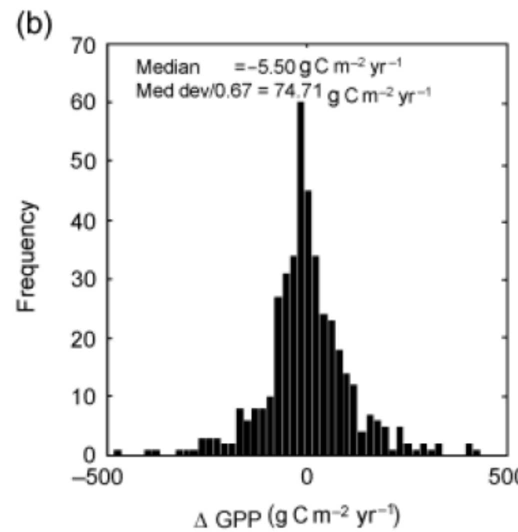
Moving to upscaling with real world observations

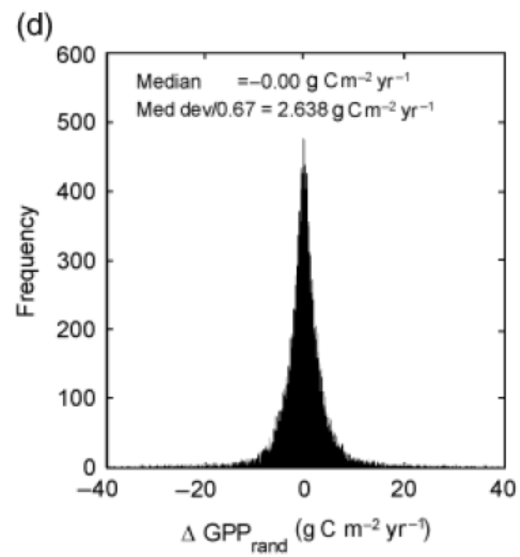


Comparison of two flux partitioning methods

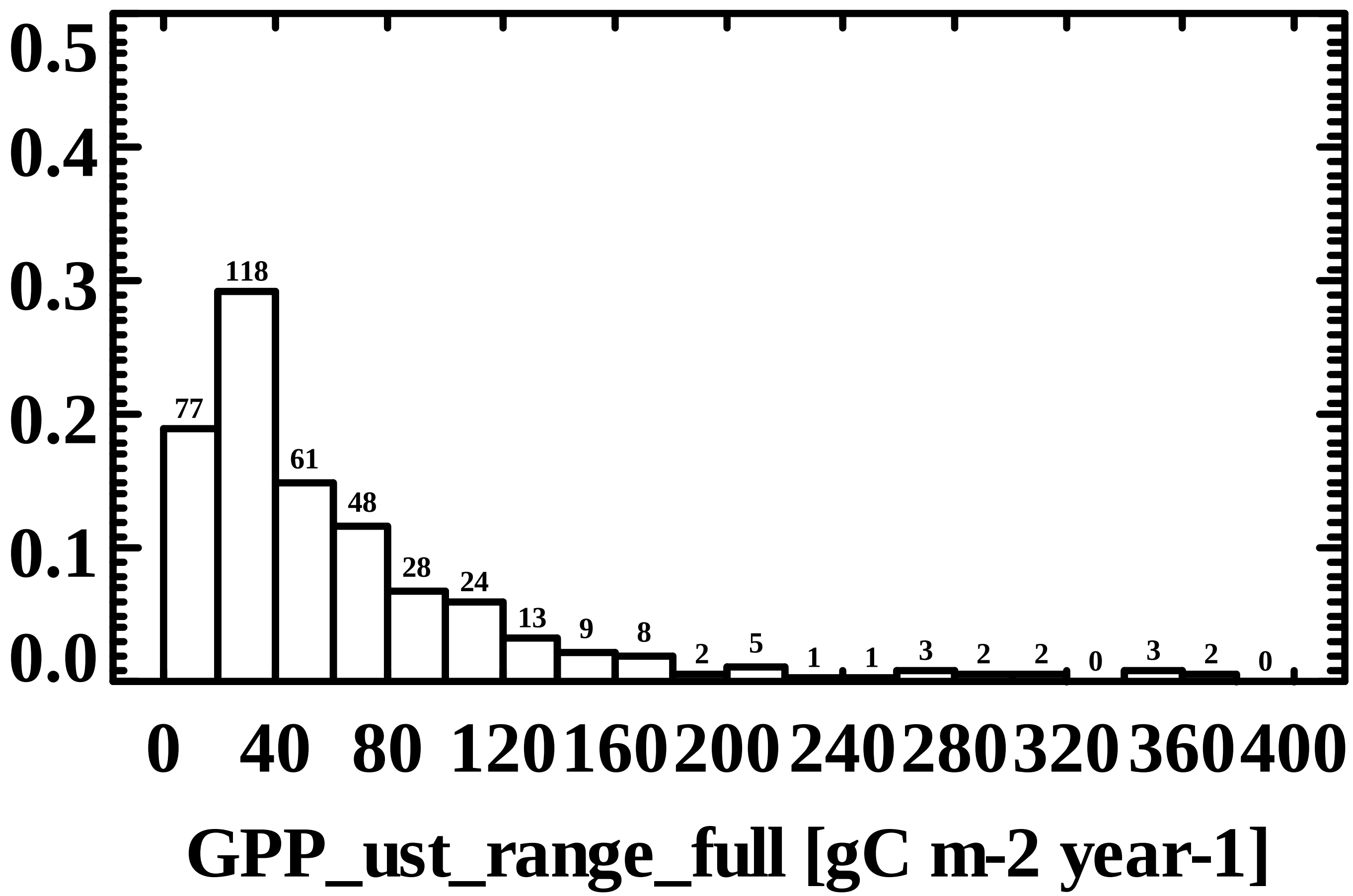


Lasslop et al 2010





Uncertainty due to u* filtering



Reichstein et al in prep

Quality control

- Remove data points with more than 20% of gapfilling
- Remove outliers from the comparison of Reichstein vs Lasslop flux partitioning methods (monthly data)
- Remove 5% of data that show largest uncertainty due to u* filtering (monthly data)

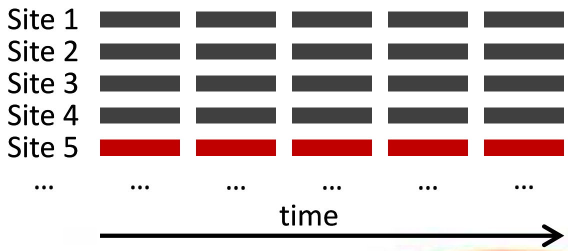
Inputs

Table 1. List of Explanatory Variables Used for the Training of MTEs[a]

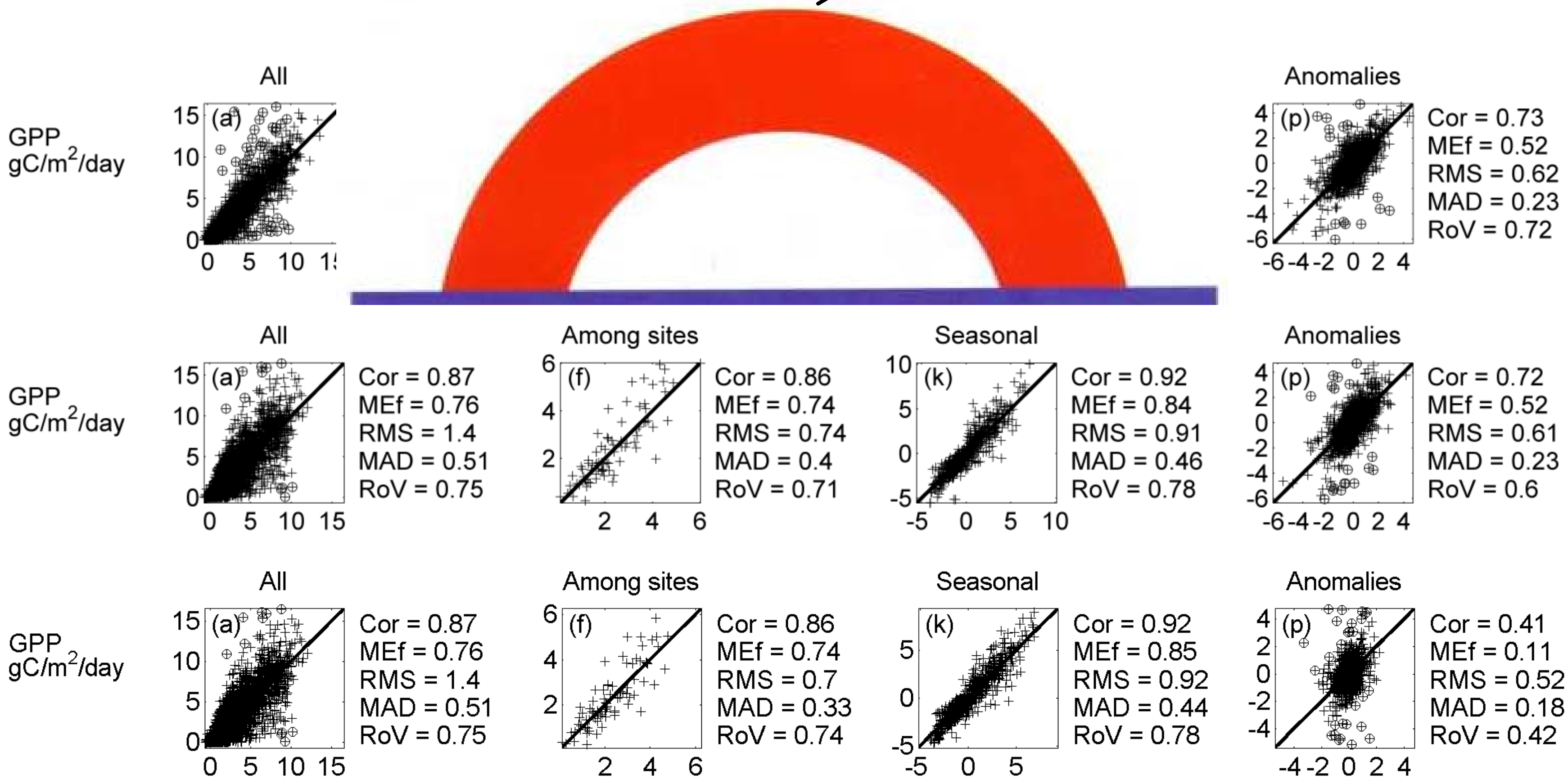
Variable	Type	Type of Variability
Climate (for Data Stratification)		
Mean annual temperature	Split	static
Mean Annual precipitation sum	Split	static
Mean annual climatic water balance	Split	static
Mean annual Potential evaporation	Split	static
Mean annual sunshine hours	Split	static
Mean annual number of wet days	Split	static
Mean annual relative humidity	Split	static
Mean monthly temperature	Split	Monthly but static over years
Mean monthly precipitation sum	Split	Monthly but static over years
Mean monthly climatic water balance	Split	Monthly but static over years
Mean monthly potential evaporation	Split	Monthly but static over years
Mean monthly sunshine hours	Split	Monthly but static over years
Mean monthly number of wet days	Split	Monthly but static over years
Mean monthly relative humidity	Split	Monthly but static over years
Vegetation Structure		
Maximum fAPAR of year	Split	yearly
Minimum fAPAR of year	Split	yearly
Maximum–minimum fAPAR	Split	yearly
Mean annual fAPAR	Split	yearly
Sum of fAPAR over the growing season	Split	yearly
Mean fAPAR of the growing season	Split	yearly
Growing season length derived from fAPAR	Split	yearly
Sum of fAPAR × potential radiation of year	Split	yearly
Maximum of fAPAR × potential radiation of year	Split and regression	yearly
IGBP vegetation type	Split	static
Meteorology		
Temperature	Split and regression	monthly
Precipitation	Split and regression	monthly
Potential radiation	Split and regression	Monthly but static over years
Vegetation Status		
fAPAR	Split and regression	monthly
fAPAR × potential radiation	Split and regression	monthly

[a]Variables that are only split variables are only used for data stratification and do not enter regressions. Please note that not all variables are automatically selected by the model trees. The "type of variability" refers to if and when the values of the respective variable change for a given pixel. "Static" variables never change and can be used by MTE to stratify into spatial domains (e.g., according to long-term mean annual temperature). "Monthly but static over years" refers to mean seasonal cycles, i.e., the values change monthly, but the same monthly values are repeatedly used for all years. "Yearly" variables have the same value within a year, but this value is updated for each year, which is primarily used for variables describing vegetation structure to capture possible effects of land cover change. "Monthly" variables exhibit different values for each month and year; that is, they are continuously updated for each month.

Cross-validation results



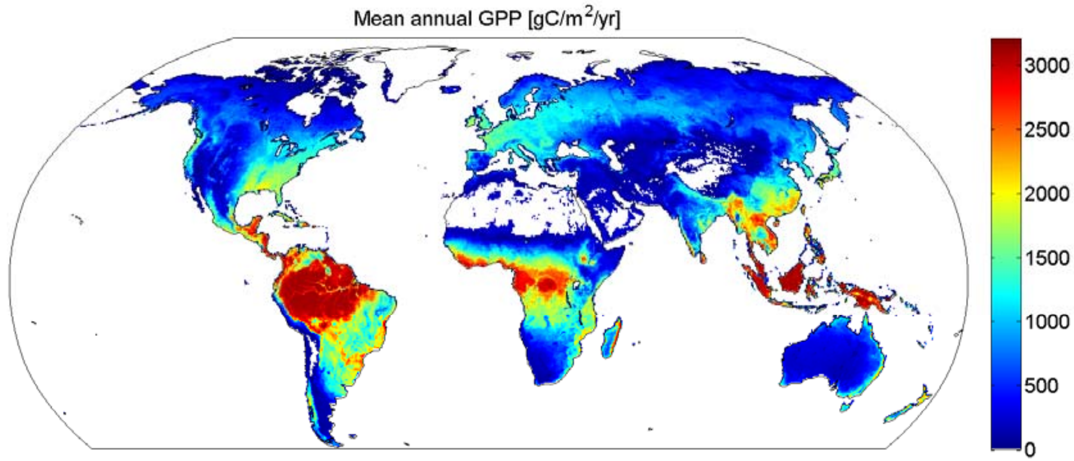
d data from entire sites were left out

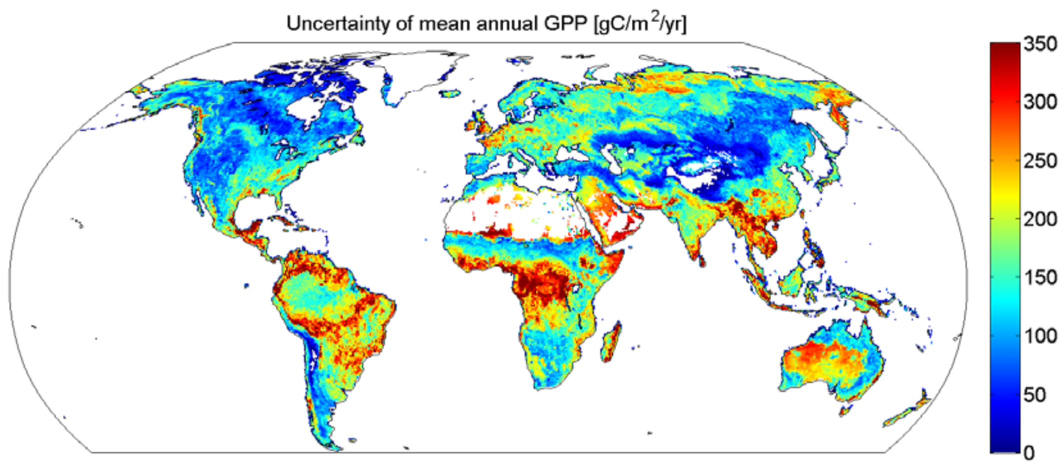


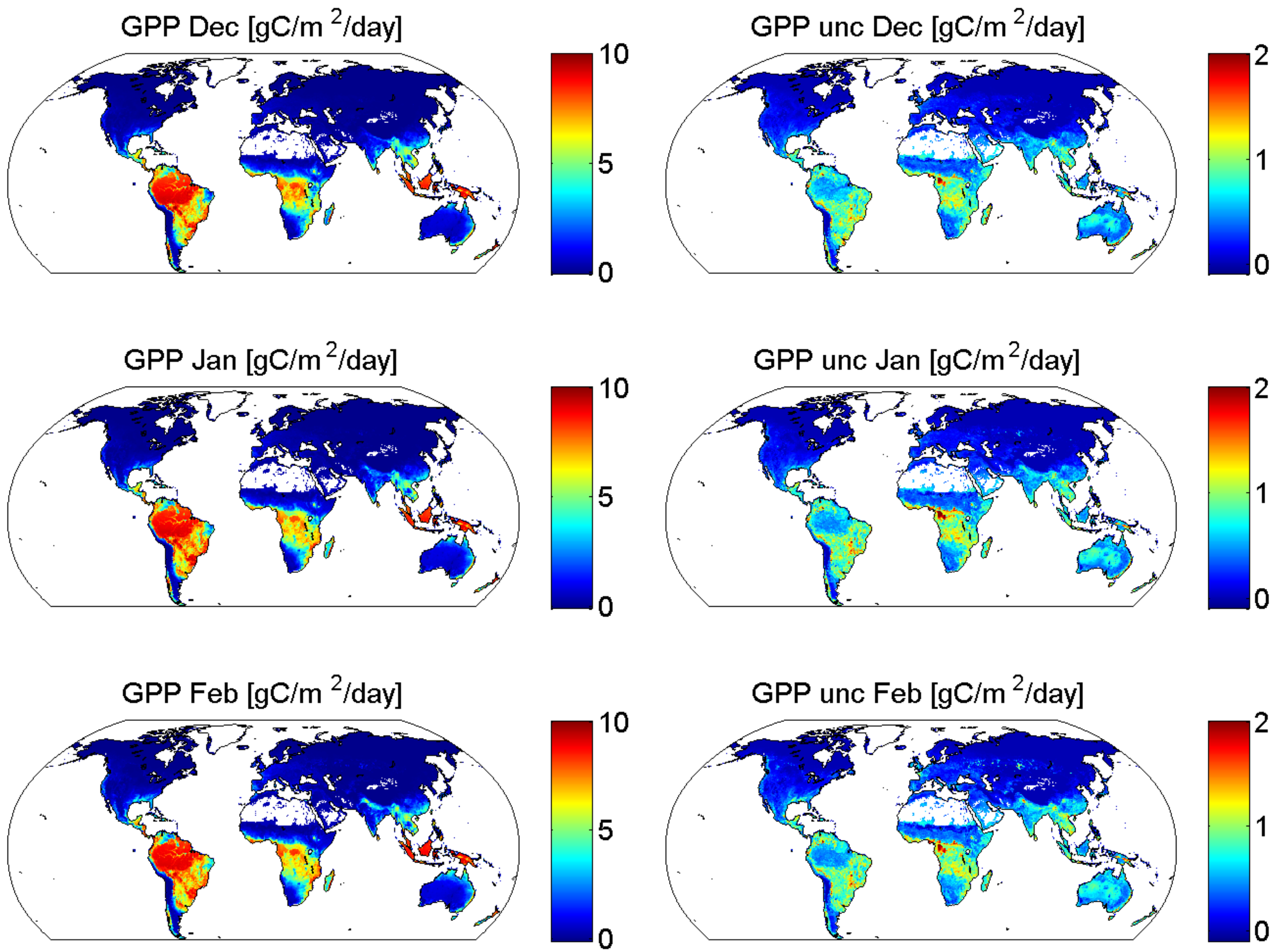
Possible reasons why anomalies are (apparently) predicted poorly

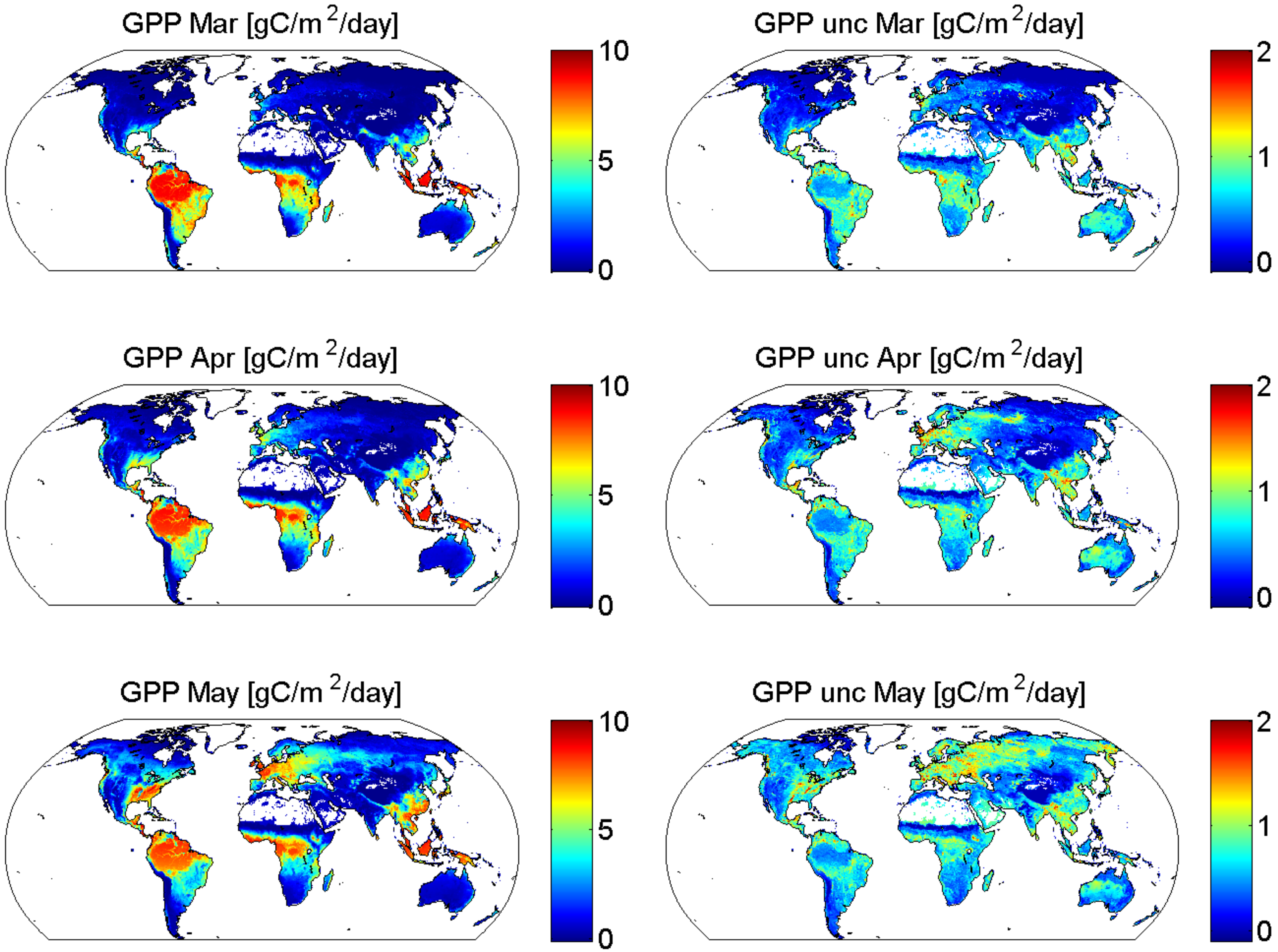
- low signal to noise ratio of monthly anomalies of eddy covariance data
- Large noise in site-level remote sensing data
- Substantial ‘non-climatically’ caused anomalies in FLUXNET data (e.g. management, disturbances, manipulations)
- Anomalies due to ecophysiological variations that change the sensitivity to meteo and biophysical properties

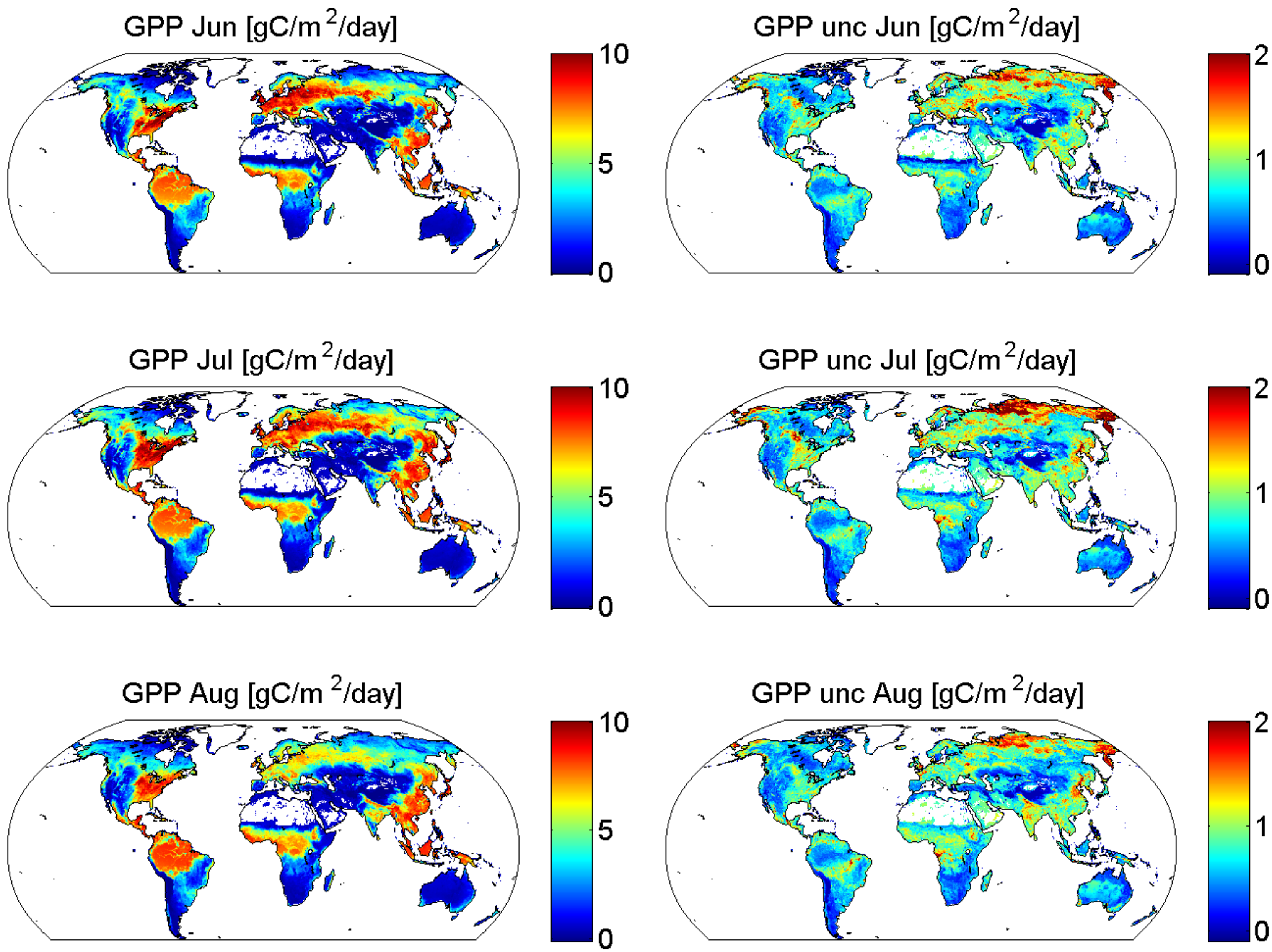
Global Results

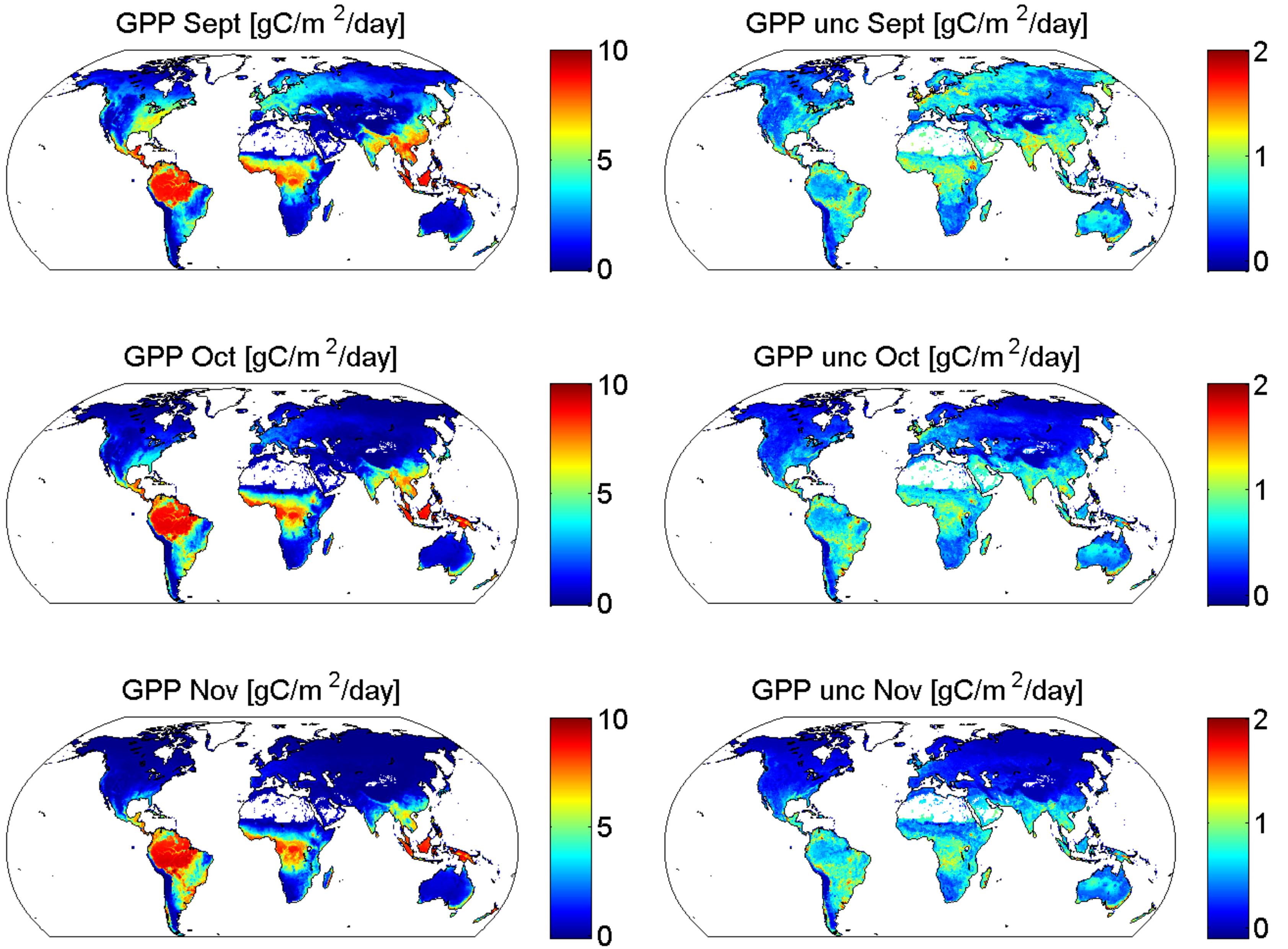





GPP Dec [gC/m 2/day]
10
5
0
GPP unc Dec [gC/m 2/day]
2
1
0
GPP Jan [gC/m 2/day]
10
5
0
GPP unc Jan [gC/m 2/day]
2
1
0
GPP Feb [gC/m 2/day]
10
5
0
GPP unc Feb [gC/m 2/day]
2
1
0

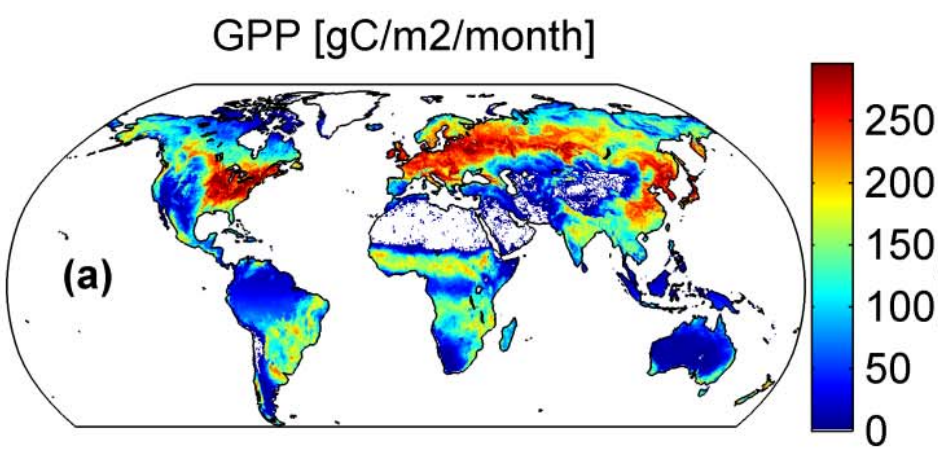

GPP Mar [gC/m 2/day]
10
5
0
GPP unc Mar [gC/m 2/day]
2
1
0
GPP Apr [gC/m 2/day]
10
5
0
GPP unc Apr [gC/m 2/day]
2
1
0
GPP May [gC/m 2/day]
10
5
0
GPP unc May [gC/m 2/day]
2
1
0


GPP Jun [gC/m 2/day]
10
5
0
GPP unc Jun [gC/m 2/day]
2
1
0
GPP Jul [gC/m 2/day]
10
5
0
GPP unc Jul [gC/m 2/day]
2
1
0
GPP Aug [gC/m 2/day]
10
5
0
GPP unc Aug [gC/m 2/day]
2
1
0

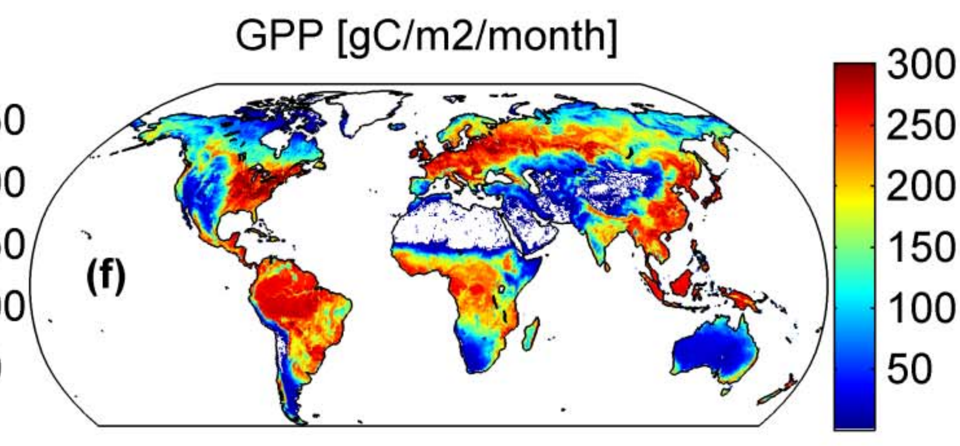

GPP Sept [gC/m 2/day]
10
5
0
GPP unc Sept [gC/m 2/day]
2
1
0
GPP Oct [gC/m 2/day]
10
5
0
GPP unc Oct [gC/m 2/day]
2
1
0
GPP Nov [gC/m 2/day]
10
5
0
GPP unc Nov [gC/m 2/day]
2
1
0

Patterns of Seasonality

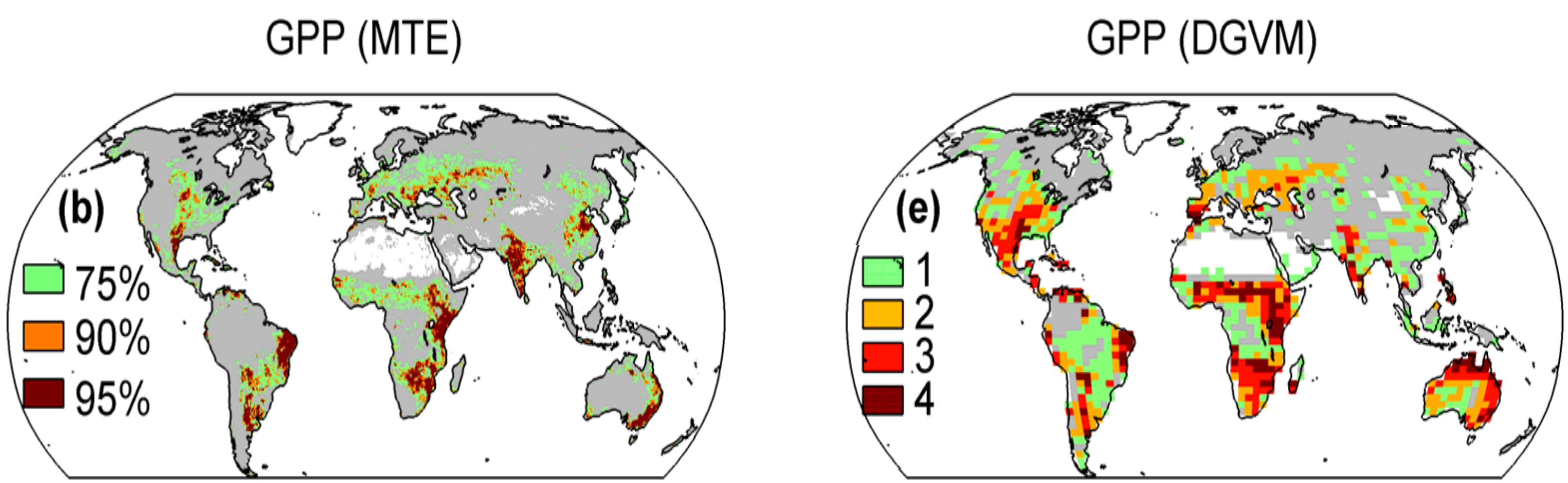
Amplitude of mean seasonal cycle

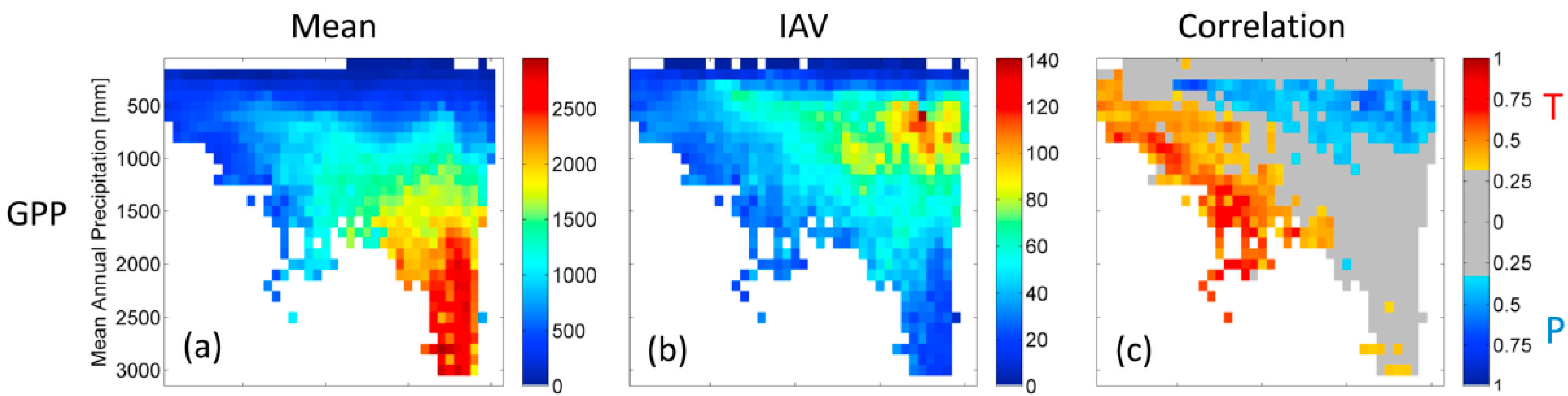


Max of mean seasonal cycle

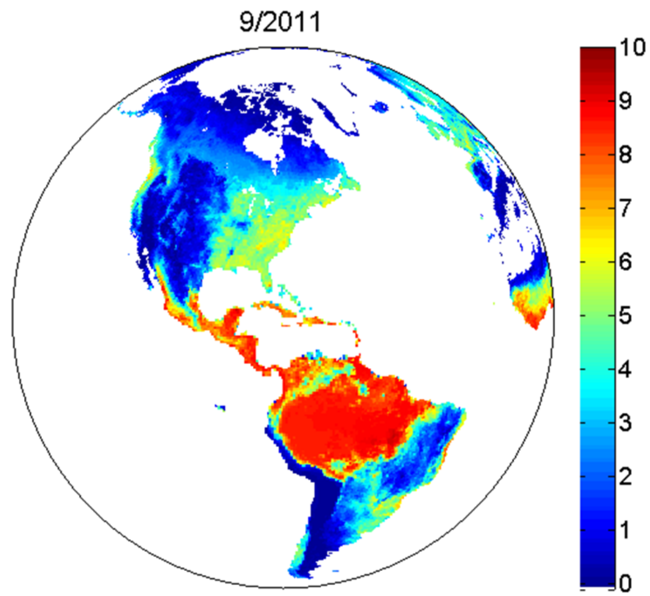


Hot spots of interannual variability

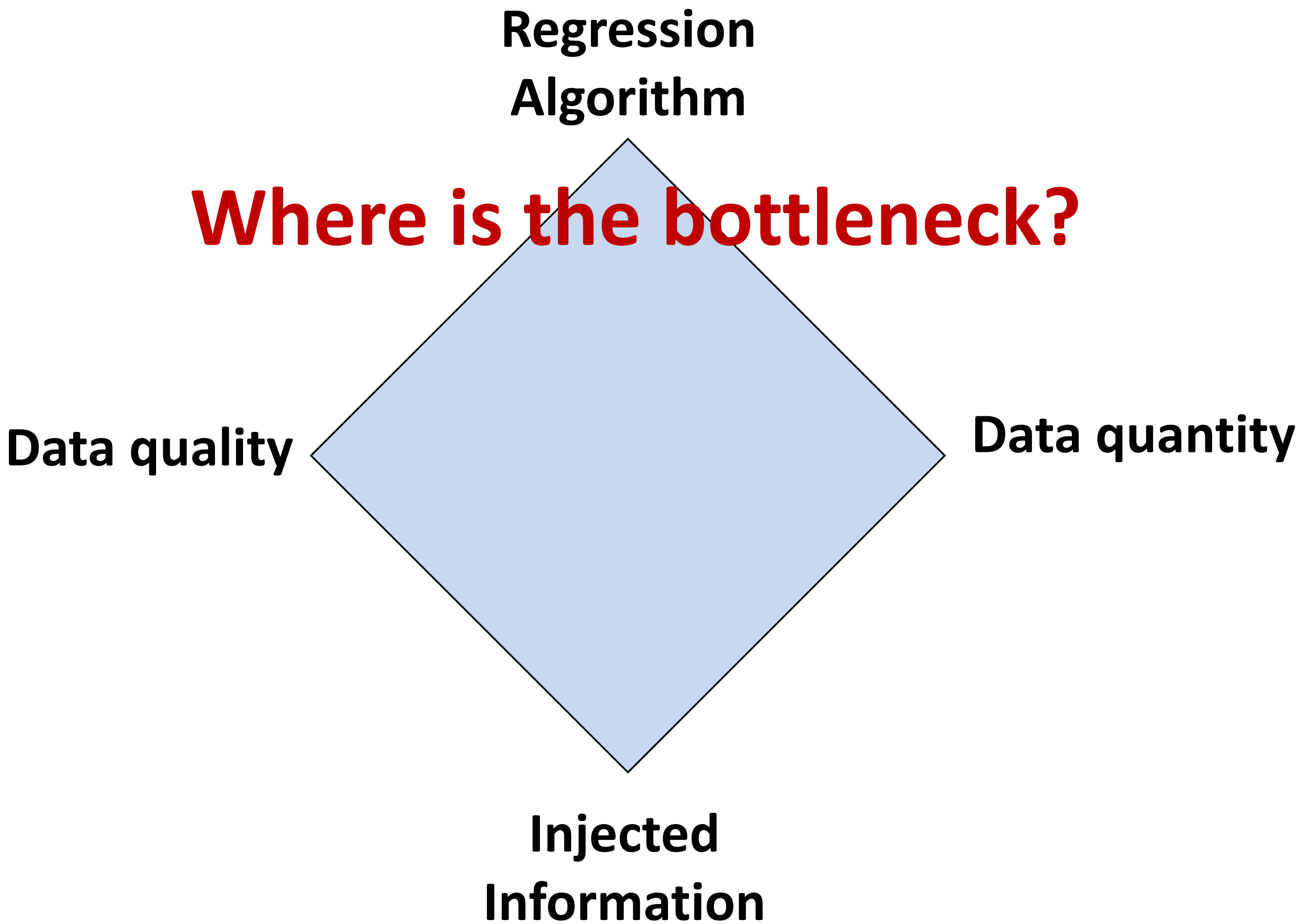




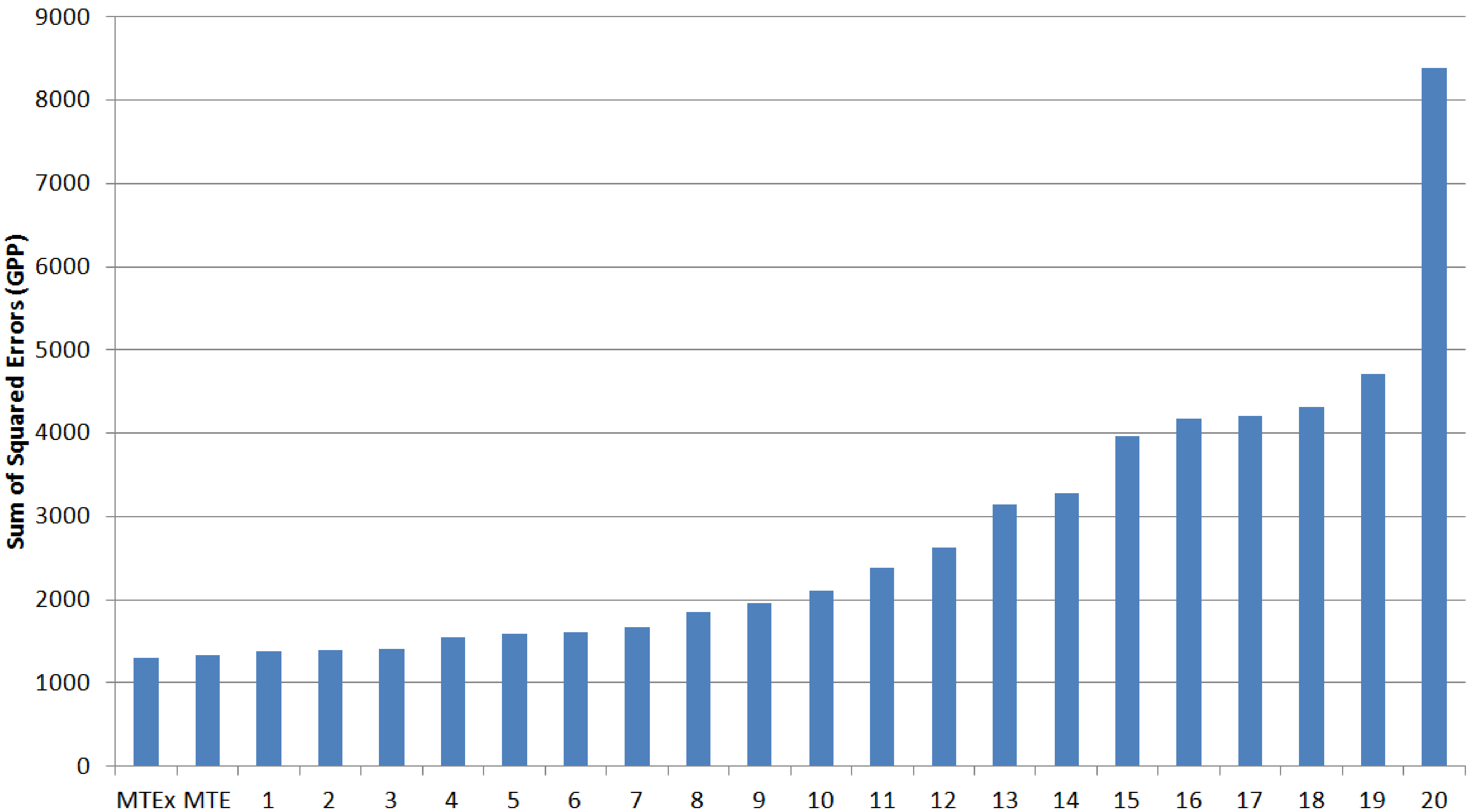
Towards improving …



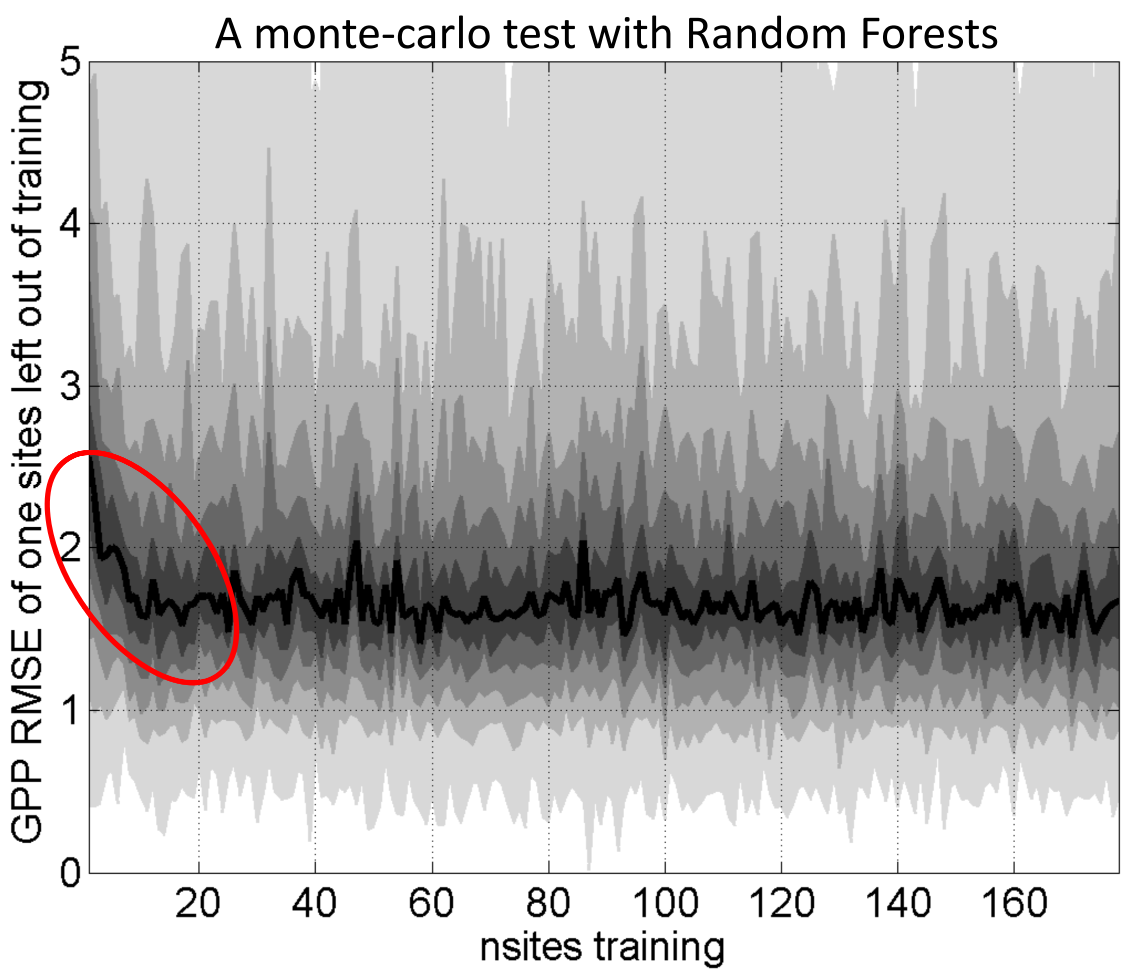
The 4 principal dimensions



Sandbox analysis I: Do we use a suitable algorithm?



Sandbox analysis II: Are we limited by the number of sites?

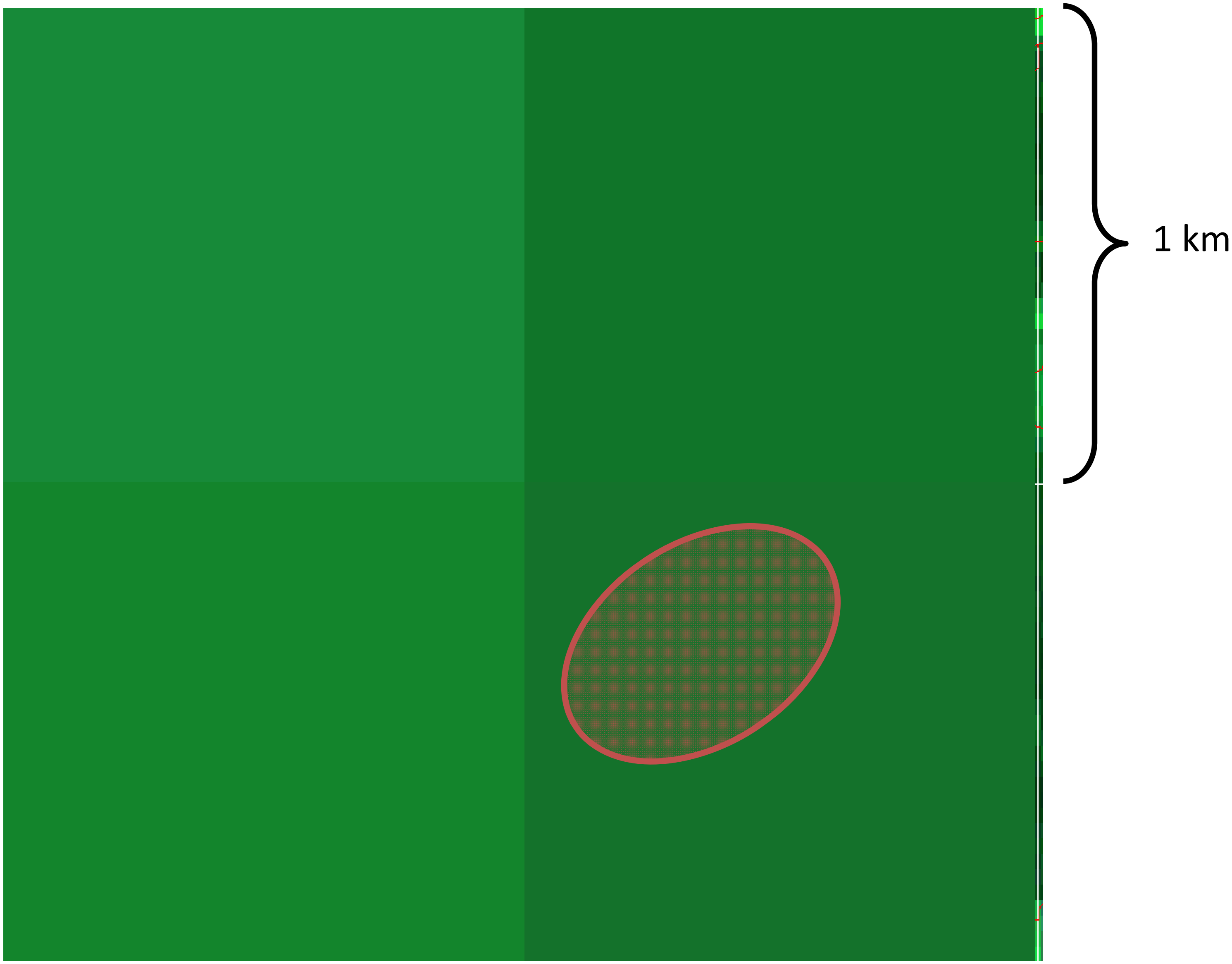


Remote sensing data quality issues

Aerial photo

Landsat

MODIS



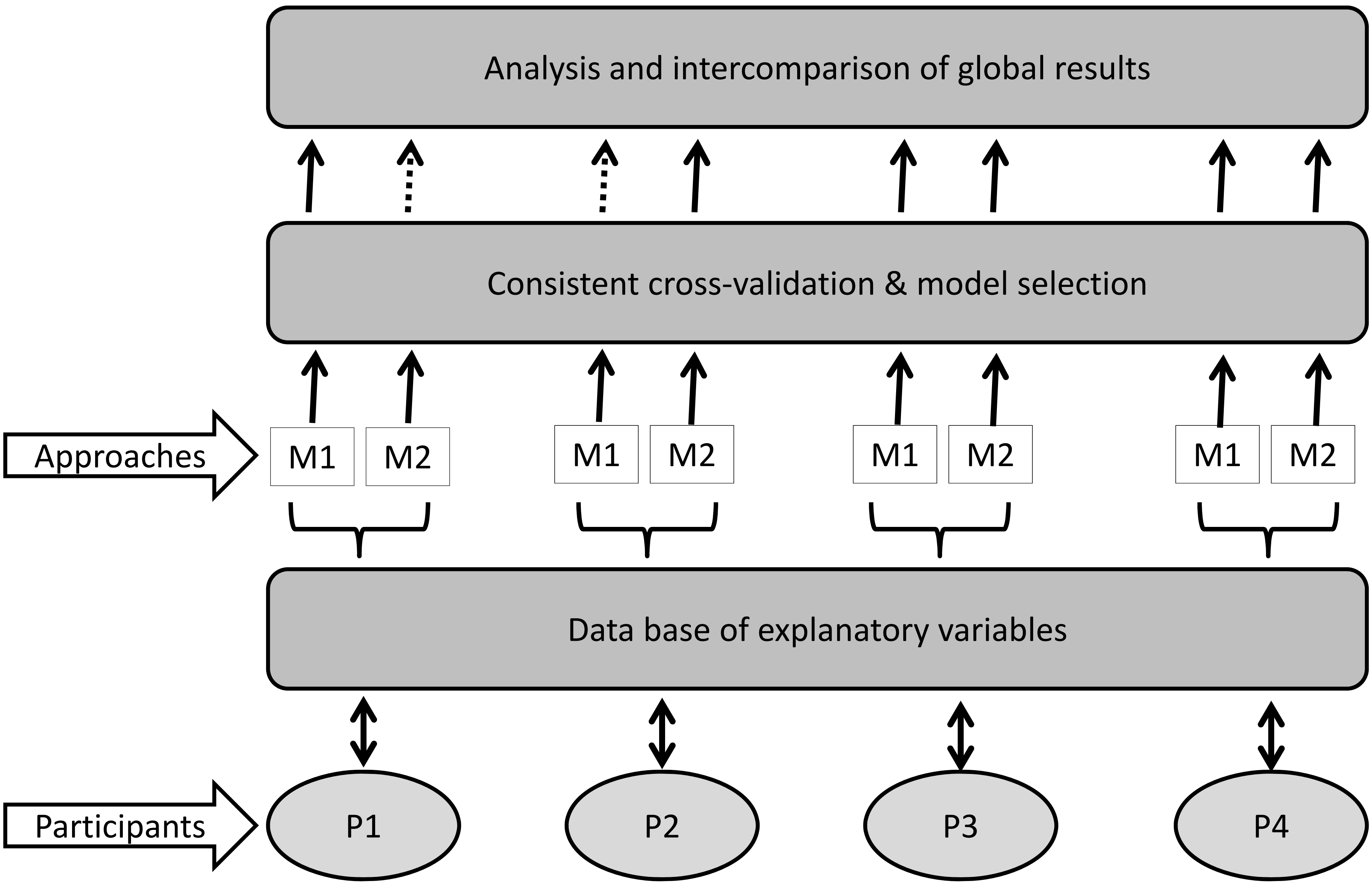
Towards a coordinated set of experiments and intercomparisons:

FluxCom

Concrete FLUXCOM goals

- To deliver a **best estimate ensemble** product of carbon and energy fluxes from an ensemble of diverse data-oriented and FLUXNET based approaches.
- To **study** and if possible **quantify** sources of **uncertainty** which hopefully leads to improved strategies with reduced uncertainty along the way of the FLUXCOM activity.
- To come up with a **practice guidance** regarding regionalization of FLUXNET data.

The FLUXCOMP vision



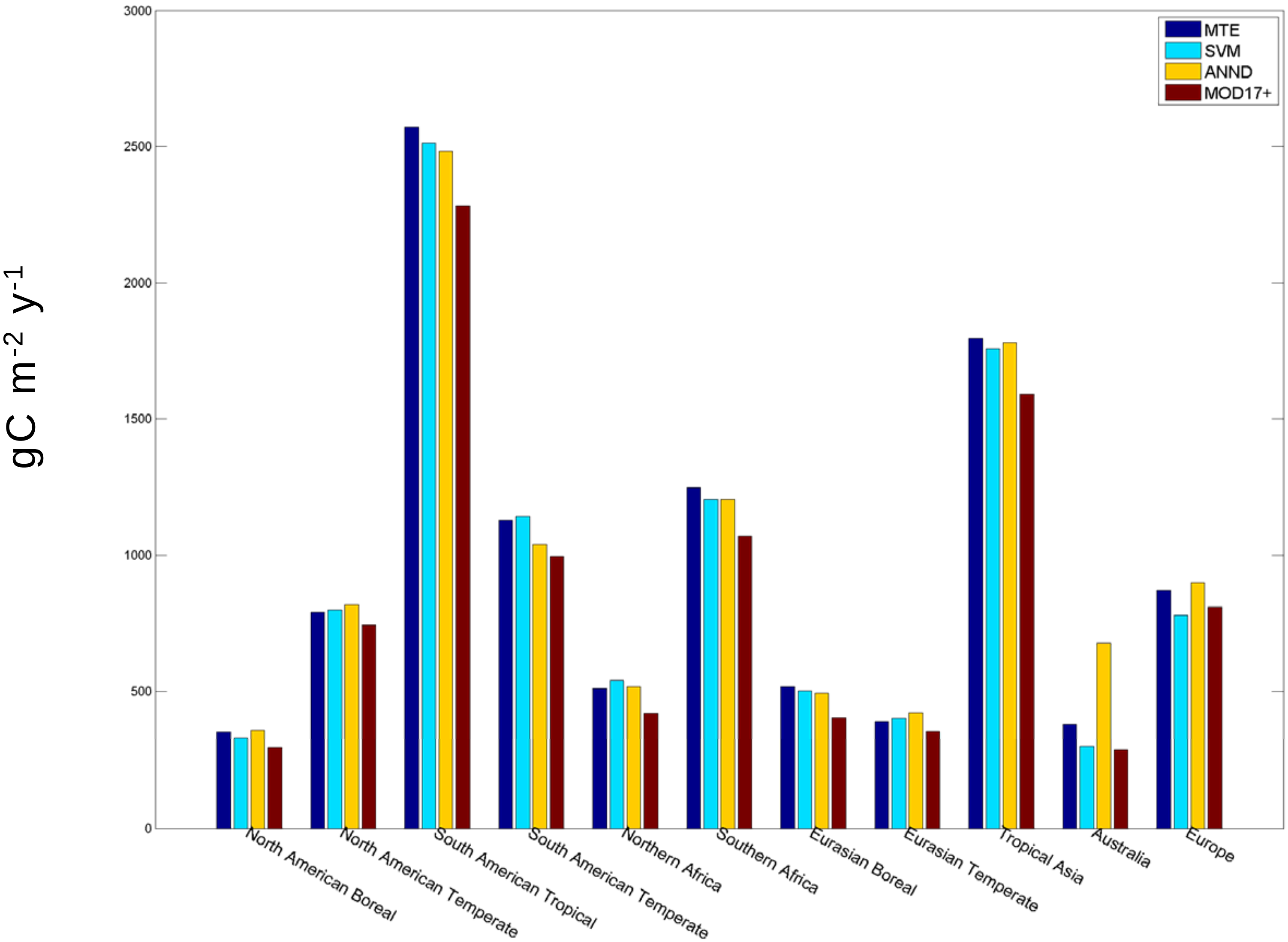
Current Participants

Name	Method/Model	Approach
Kazuhito Ichii	SVM	Machine Learning
Dario Papale	ANN	Machine Learning
Antje Moffat	ANN	Machine Learning
Gustavo Camps-Valls	ANN	Machine Learning
Christopher Schwalm	ANN	Machine Learning
Martin Jung	MTE, RF	Machine Learning
Enrico Tomelleri	MOD17+	Model-Data-Fusion
Nuno Carvalhais	CASA	Model-Data-Fusion
Timothy Hilton	VPRM	Model-Data-Fusion
Anthony Bloom	DALEC	Model-Data-Fusion

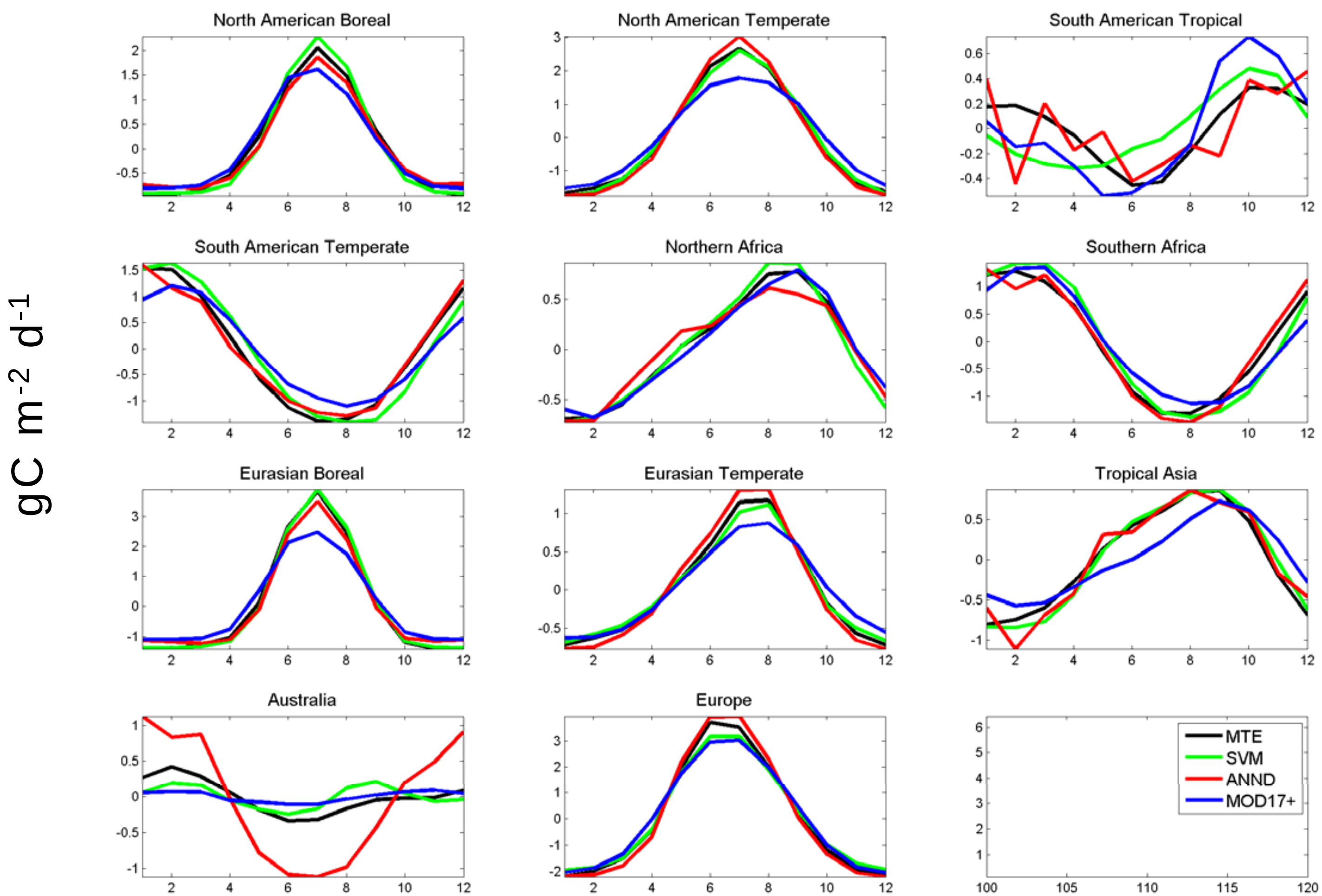
Some quick pre-FLUXCOM comparison (fully unharmonized)

- SVM (Kazuhito Ichii)
- ANN ensembles (Dario Papale)
- MTE (Martin Jung)
- MOD17+ (Enrico Tomelleri)

GPP Mean annual fluxes



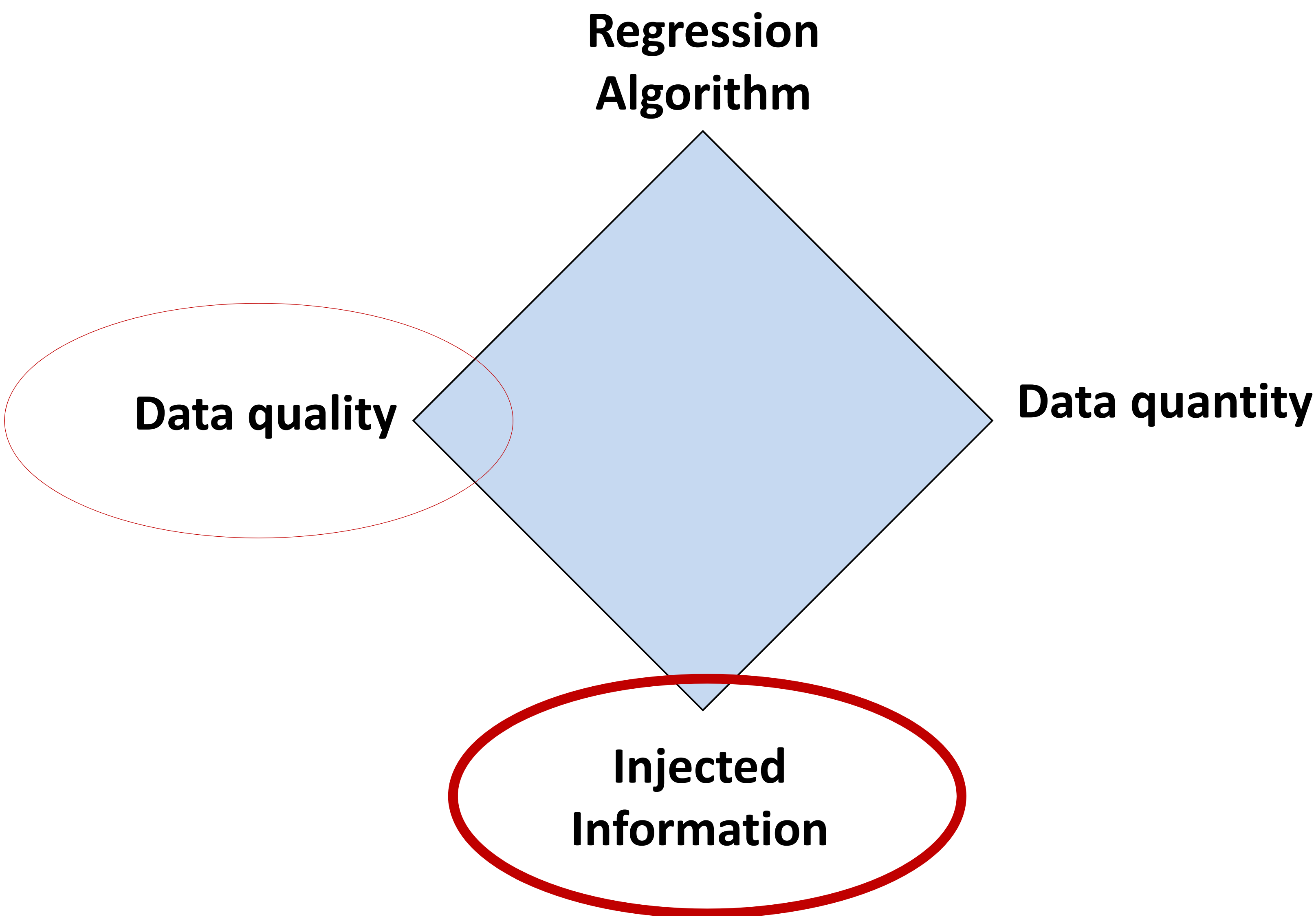
Mean Seasonal cycles



FLUXCOM products will capitalize on:

- An ensemble (of ensembles ☺) of estimates
- Higher temporal and spatial resolution
- More injected information (variables)
- ~twice the amount of La Thuille FLUXNET data

The 4 principal dimensions



Which and how many variables are needed?

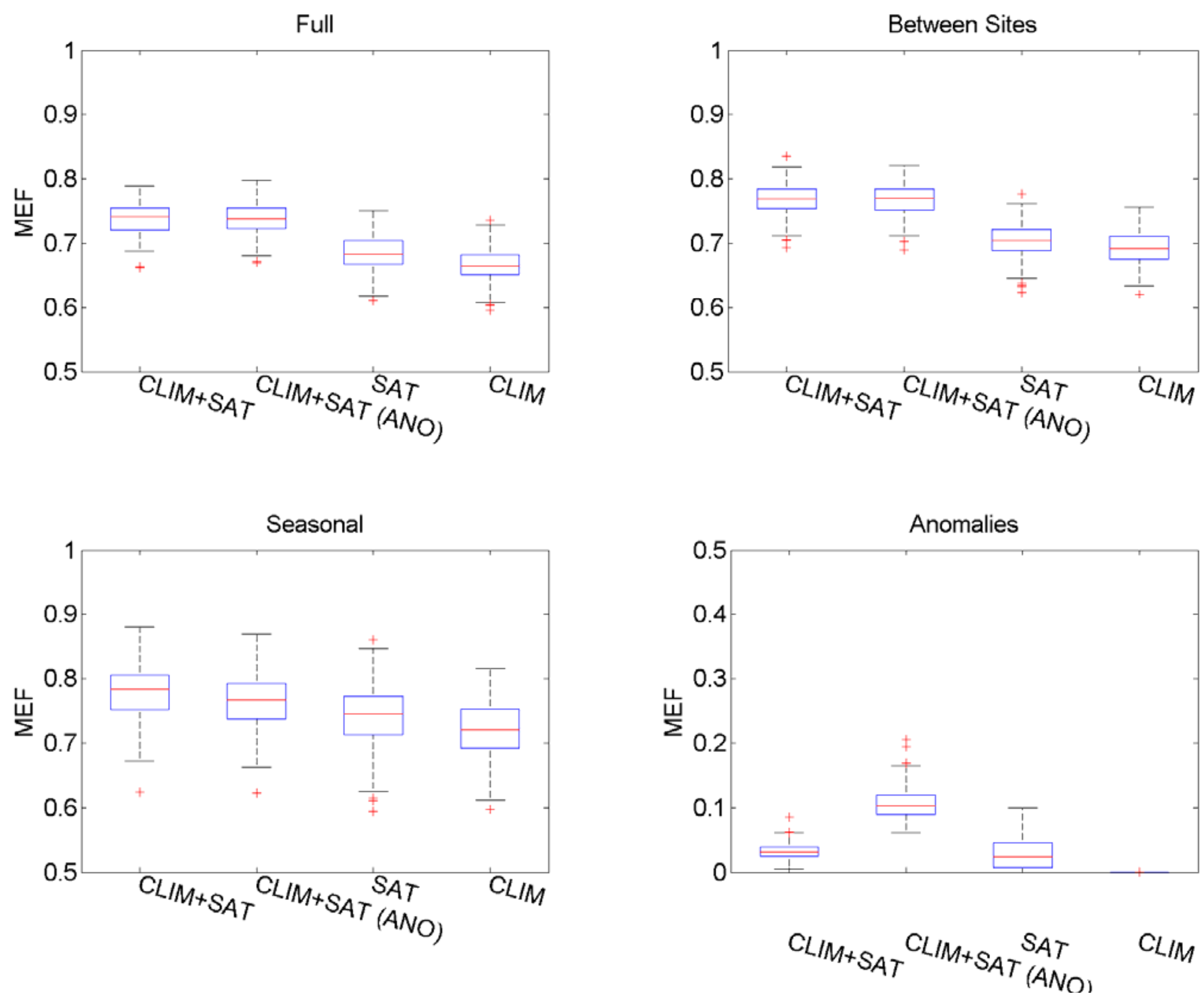
- Injected Information = f(availability, imagination)
- → potentially large set → feature selection problem

	☺	☹
Climate variables	Precise in-situ measurements	Need coarse resolution climate data for global upscaling Uncertainty and biases of global climate fields
Satellite variables	No product switch necessary from site to globe High res global output possible	Very noisy at site-level Tower satellite footprint mismatch gaps

Some experiments with a feature selection algorithm and Random Forests?

	# candidate variables	# selected
CLIM+SAT	166	10
CLIM+SAT (ANO)	166	16
SAT	165	9
CLIM	61	7

Results based on 10-fold cross-validation and bootsrapping



On the evaluation of the products

Site-level cross-validation is important and valuable to compare different approaches and set-ups but…

Is a not sufficient evaluation of global fields because:

- Biased sampling of biosphere by tower sites
- Noise and site-specific peculiarities at site-level
- Doesn’t capture uncertainties of global drivers

Need for an independent, observational based approach

Fluorescence

If photosynthesis is so complicated why can we get the big picture of gpp relatively easily from sparse data?

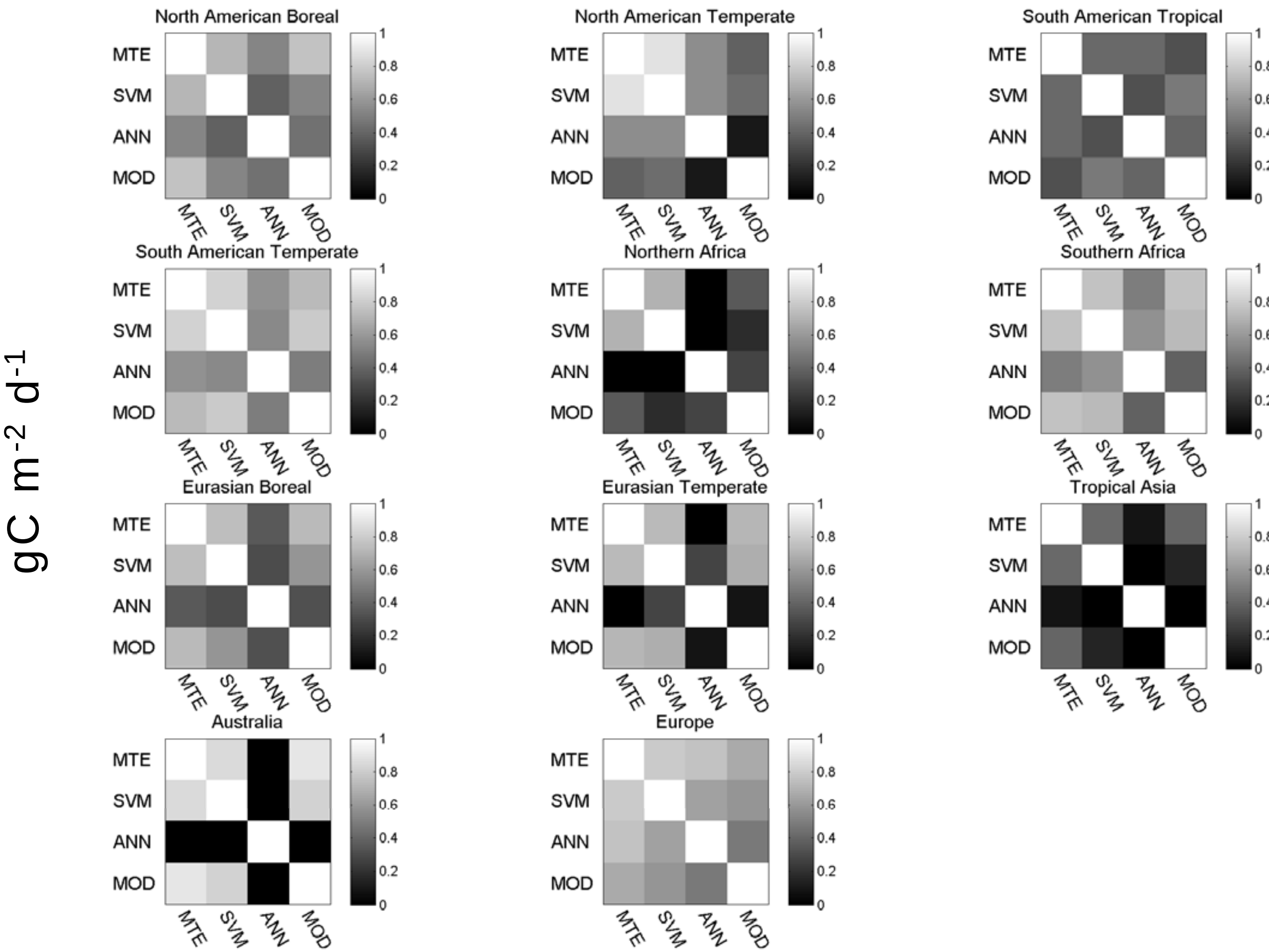
- Adaptation to the environment
- At ecosystem scale everything scales with LAI/FPAR

What is the added value of fluoresence?

Does it track physiological regulations that control interannual variability?

Thank you very much
for your attention!

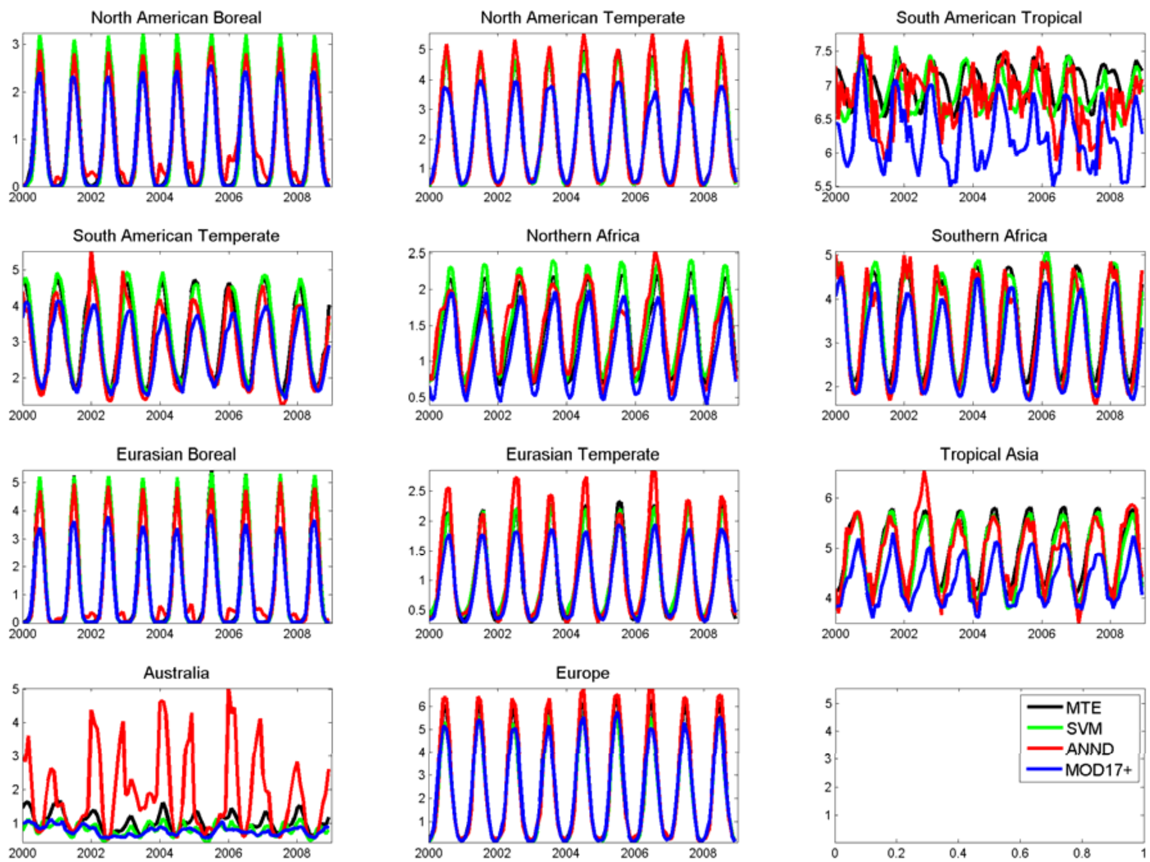
Interannual variability



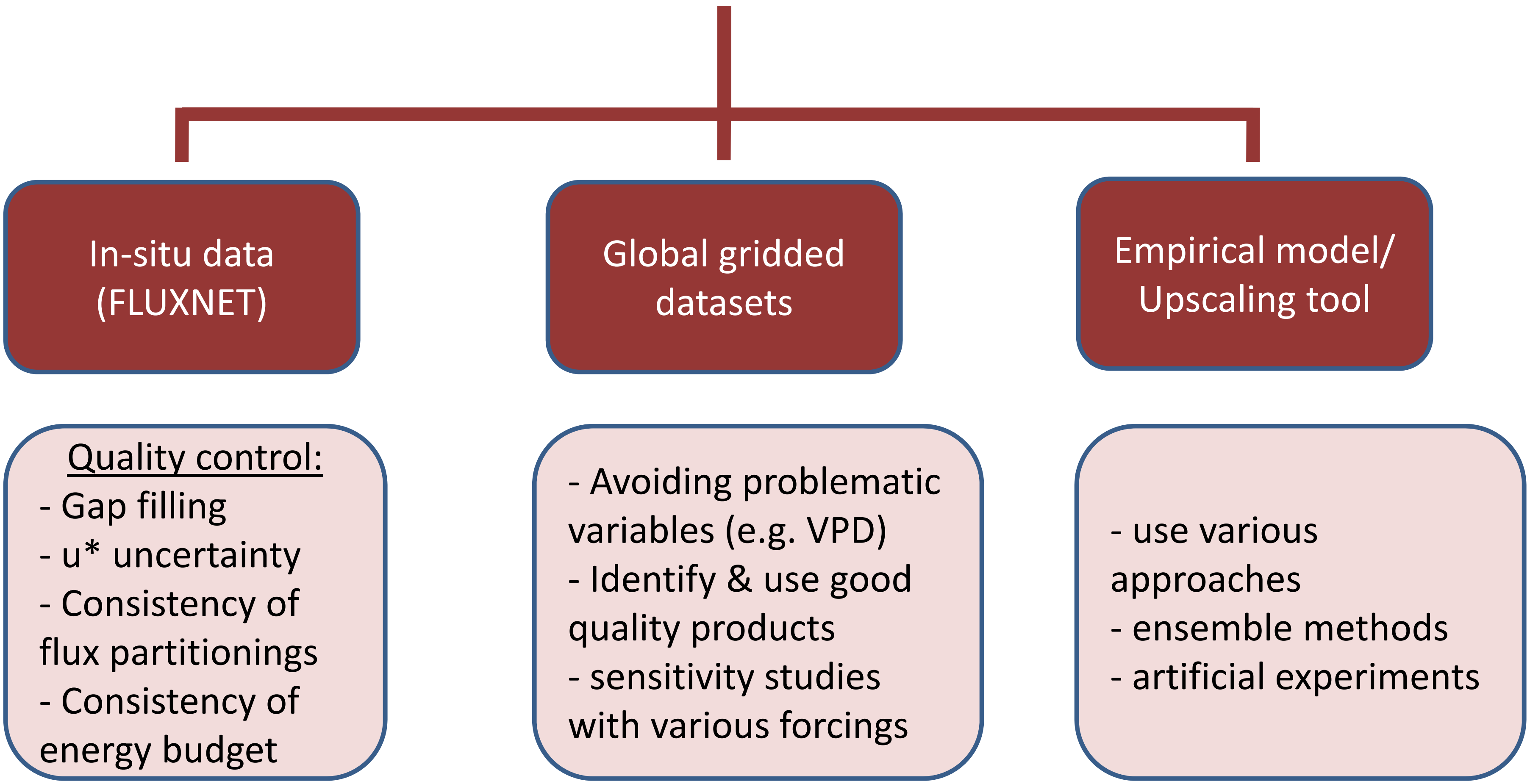
Conclusions

- Global upscaling from FLUXNET is feasible
- A new observation driven data stream
- Known issues:
  - Spatial distribution of mean NEE
  - Underestimation of interannual variability
- Largest potential for future improvements: adding informative variables
- Plan of a coordinated intercomparison (FLUXCOMP)
- Global products are available on request

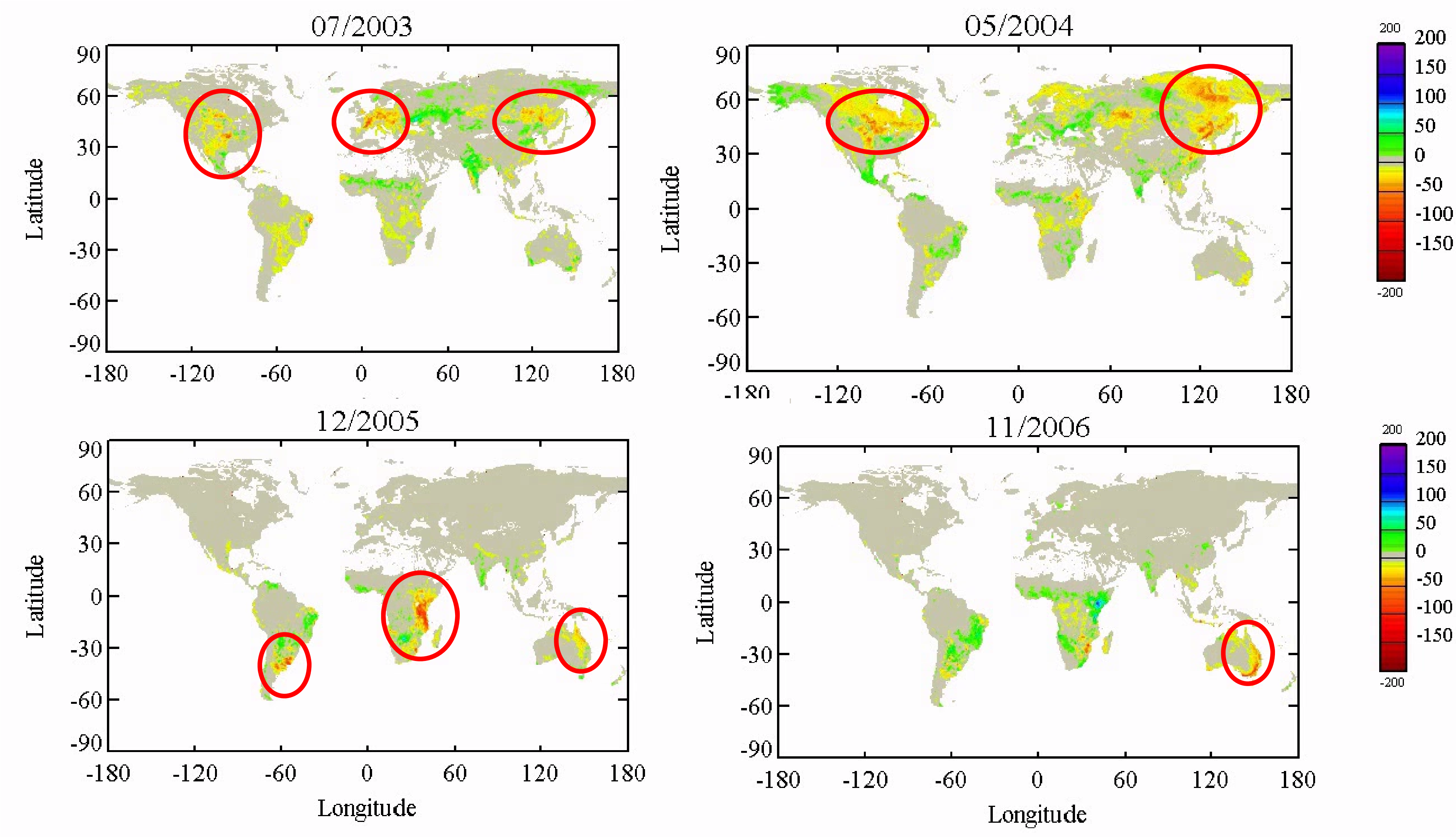
GPP


North American Boreal
North American Temperate
South American Tropical
South American Temperate
Northern Africa
Southern Africa
Eurasian Boreal
Eurasian Temperate
Tropical Asia
Australia
Europe
MTE
SVM
ANND
MOD17+

Uncertainties of empirical upscaling



GPP anomaly snapshots (gC m^{-2} mo^{-1})



Jung et al. in prep