Multi-purpose, modularized, testable, and fully constrained against multiple datasets:

A Vision for the Future of Land Surface Model Development

Natasha MacBean

University of Arizona

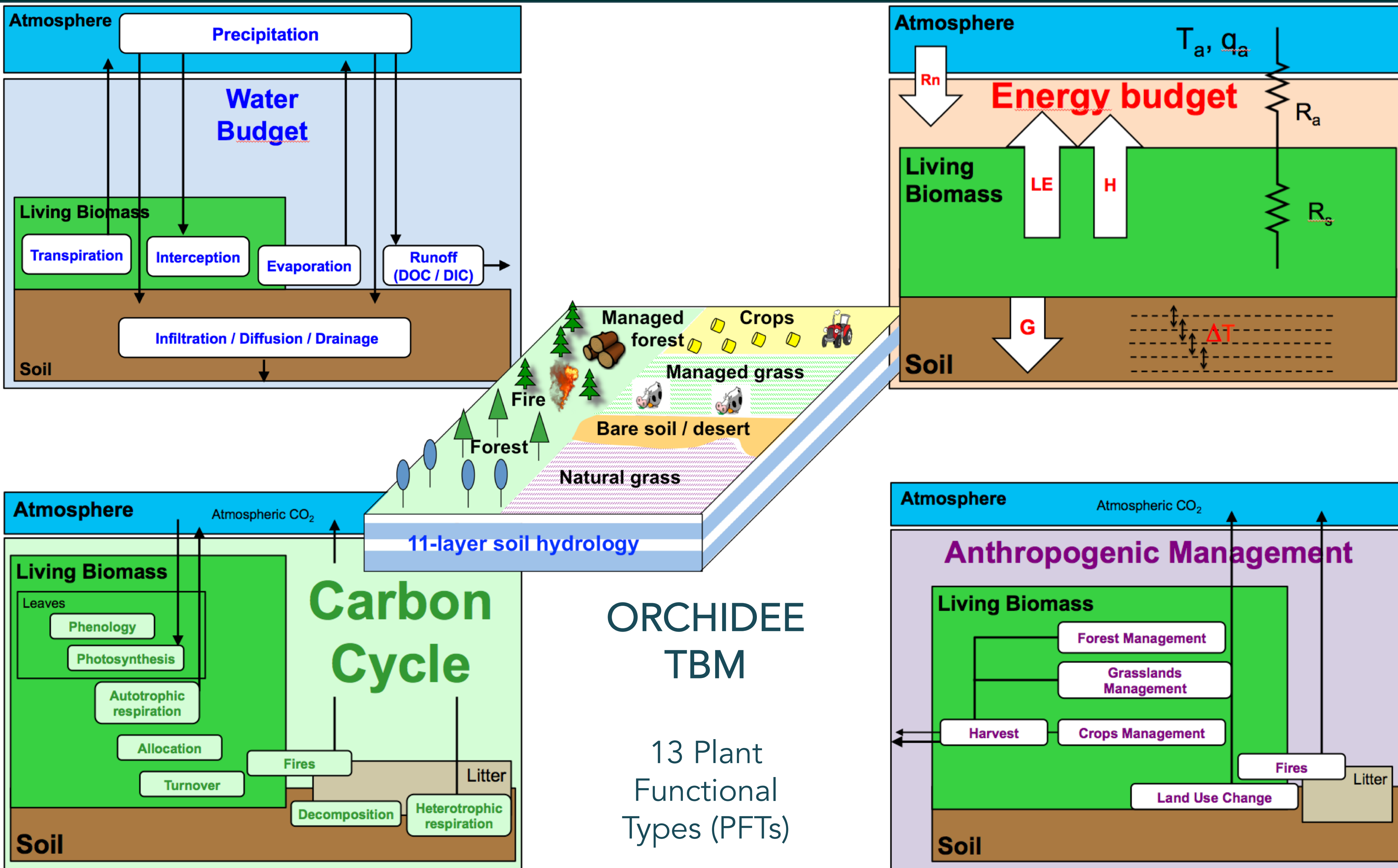
Philippe Peylin, Cédric Bacour, Vladislav Bastrikov, Fabienne Maignan
Andy Fox, Dave Moore

Outline

Data assimilation as a tool for reducing uncertainty in land surface models

- Highlights with the ORCHIDEE LSM
- Challenges we face
- Future perspectives

Global terrestrial biosphere models (TBMs)

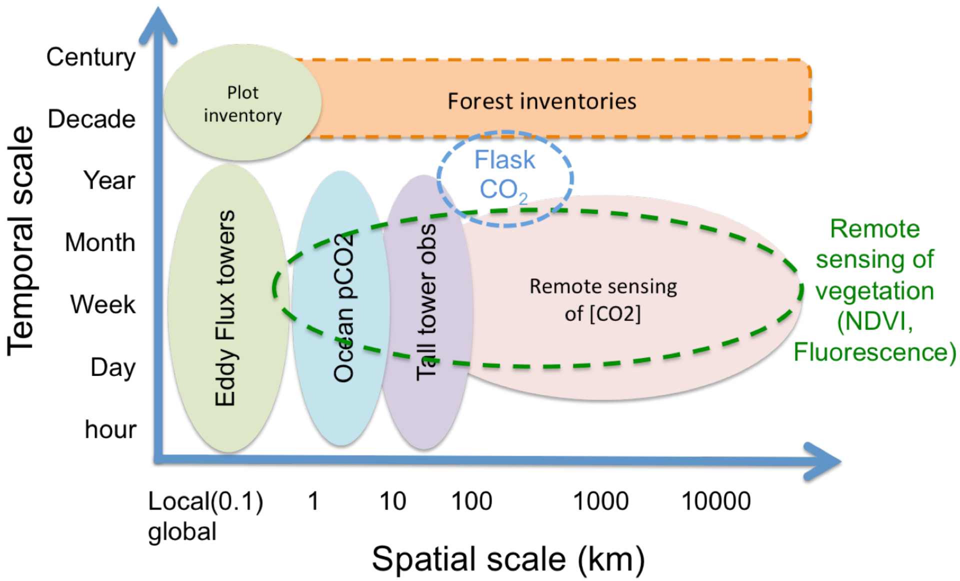


ORCHIDEE TBM

13 Plant Functional Types (PFTs)

Reducing uncertainty: the need for model – data integration

Available C-related data streams



Model

Parameters

Obs

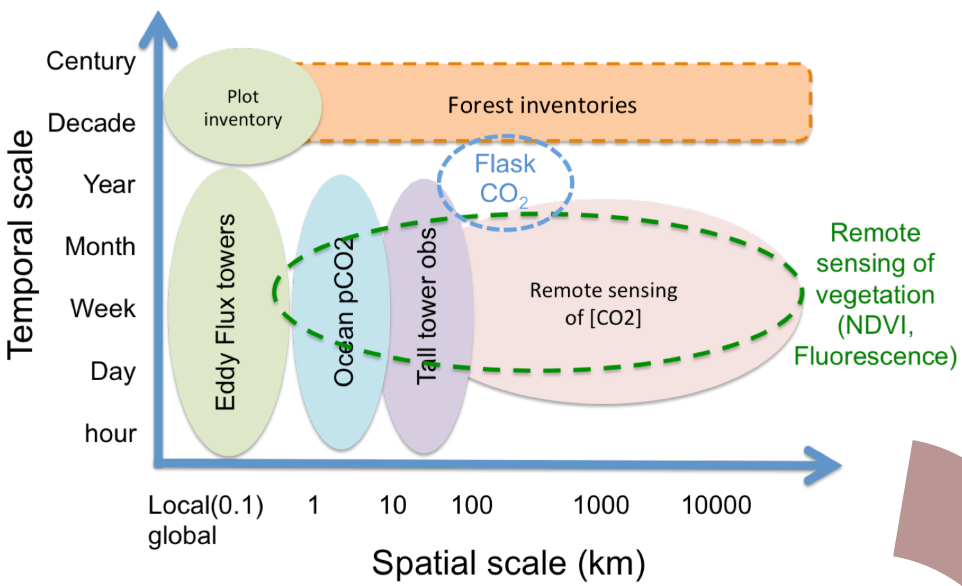
$$J(\mathrm{x}) = \underbrace{\tfrac{1}{2}(\mathbf{M}(\mathrm{x})-\mathrm{y})^{T}\,\mathbf{R}^{-1}\overbrace{(\mathbf{M}(\mathrm{x})-\mathrm{y})}^{\text{Model-data mismatch}}}_{\text{Observation term}}$$

Prior parameters

$$+\ \underbrace{\tfrac{1}{2}(\mathrm{x}-\mathrm{x}_b)^{T}\,\mathbf{B}^{-1}(\mathrm{x}-\mathrm{x}_b)}_{\text{Prior parameter term}}$$

Reducing uncertainty: the need for model – data integration

Available C-related data streams



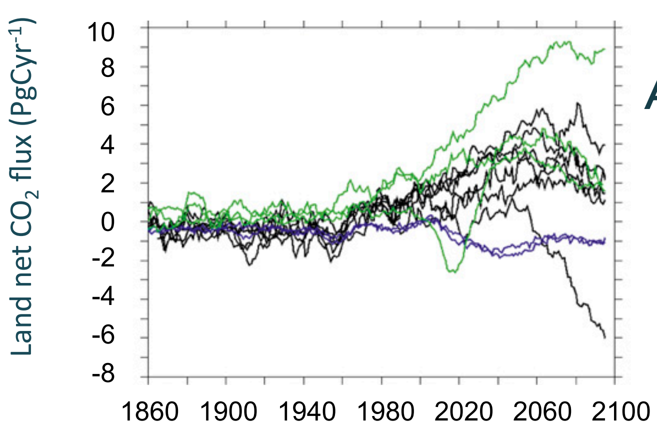
Model Parameters Obs Model-data mismatch

$$J(\mathrm{x}) = \underbrace{\tfrac{1}{2}(\mathbf{M}(\mathrm{x})-\mathrm{y})^{\mathrm{T}}\,\mathbf{R}^{-1}\overbrace{(\mathbf{M}(\mathrm{x})-\mathrm{y})}^{\text{Model-data mismatch}}}_{\text{Observation term}}$$

Prior parameters

$$+\ \underbrace{\tfrac{1}{2}(\mathrm{x}-\mathrm{x}_{\mathrm{b}})^{\mathrm{T}}\,\mathbf{B}^{-1}(\mathrm{x}-\mathrm{x}_{\mathrm{b}})}_{\text{Prior parameter term}}$$

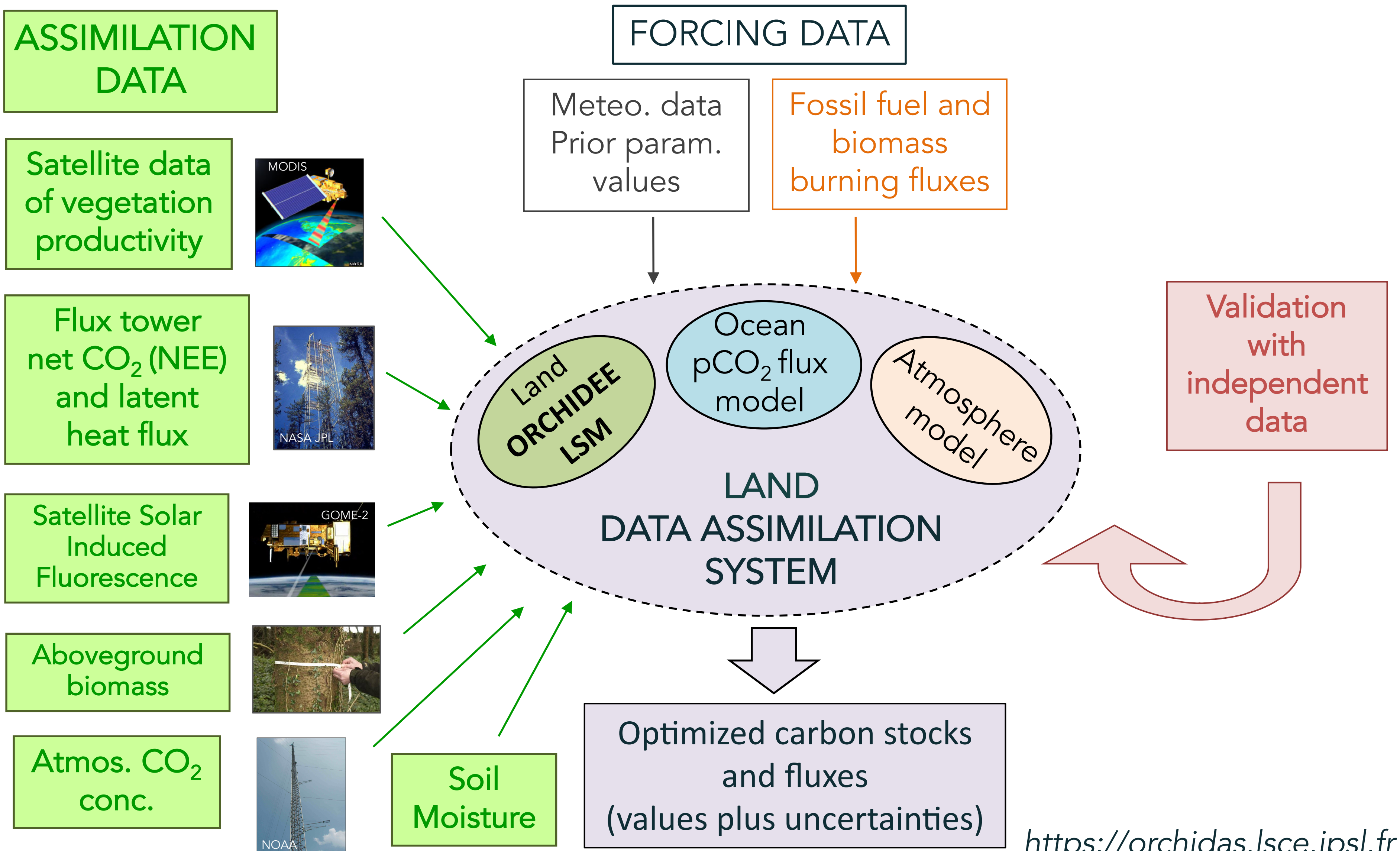
DATA ASSIMILATION



Improve:

- C land budget estimates
- Quantify & reduce uncertainty
- Future climate predictions
- Process understanding

Global Data Assimilation System – ORCHIDEE LSM



Highlights of ORCHIDEE model parameter optimization so far…

https://orchidas.lsce.ipsl.fr

ORCHIDEE Data Assimilation Systems

Institut Pierre Simon Laplace / Laboratoire des Sciences du Climat et de l'Environnement

ORCHIDEE LAND SURFACE MODEL

ORCHIDAS DATA ASSIMILATION SYSTEMS

Overview | Results | Publications | Tutorials | People | Contact

GLOBAL BIOGEOCHEMICAL CYCLES, VOL. 21, GB2013, doi:10.1029/2006GB002834, 2007

Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France

Diego Santaren,[1] Philippe Peylin,[1,2] Nicolas Viovy,[1] and Philippe Ciais[1]

Biogeosciences, 9, 3757–3776, 2012

Constraining a global ecosystem model with multi-site eddy-covariance data

S. Kuppel[1], P. Peylin[1,2], F. Chevallier[1], C. Bacour[3], F. Maignan[1], and A. D. Richardson[4]

Geosci. Model Dev., 7, 2581–2597, 2014

Model–data fusion across ecosystems: from multisite optimizations to global simulations

S. Kuppel[1,2], P. Peylin[1], F. Maignan[1], F. Chevallier[1], G. Kiely[3], L. Montagnani[4], and A. Cescatti[5]

Biogeosciences, 12, 7185–7208, 2015

Using satellite data to improve the leaf phenology of a global terrestrial biosphere model

N. MacBean[1], F. Maignan[1], P. Peylin[1], C. Bacour[2], F.-M. Bréon[1], and P. Ciais[1]

Geosci. Model Dev., 9, 3321–3346, 2016

A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle

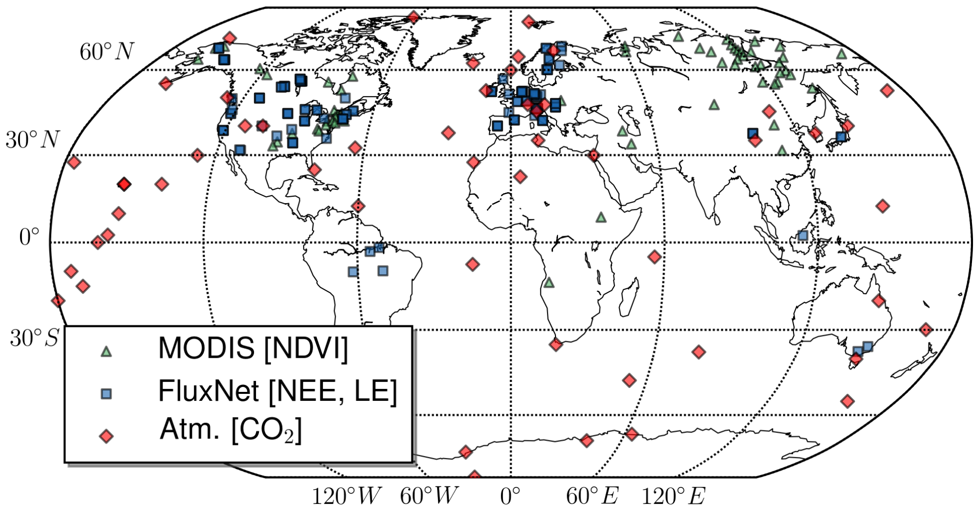
Philippe Peylin[1], Cédric Bacour[2], Natasha MacBean[1], Sébastien Leonard[1], Peter Rayner[1,3], Sylvain Kuppel[1,4], Ernest Koffi[1], Abdou Kane[1], Fabienne Maignan[1], Frédéric Chevallier[1], Philippe Ciais[1], and Pascal Prunet[2]

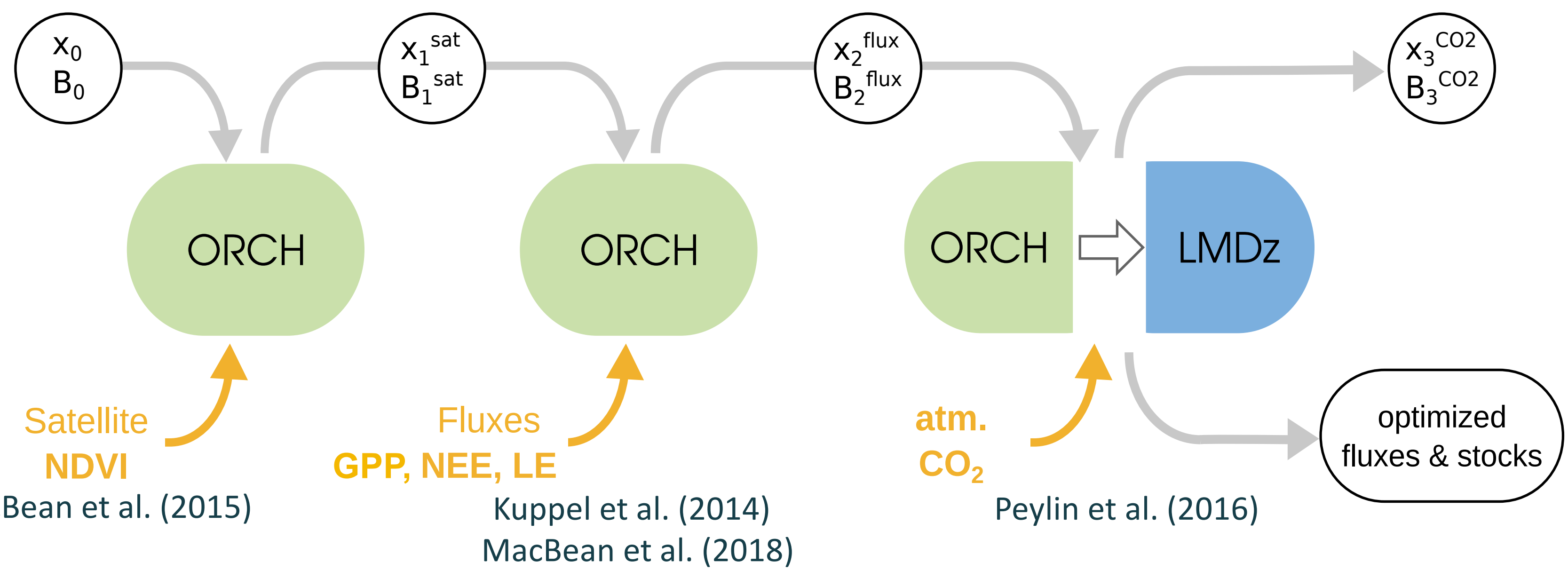
SCIENTIFIC REPORTS | (2018) 8:1973 | DOI:10.1038/s41598-018-20024-w

Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data

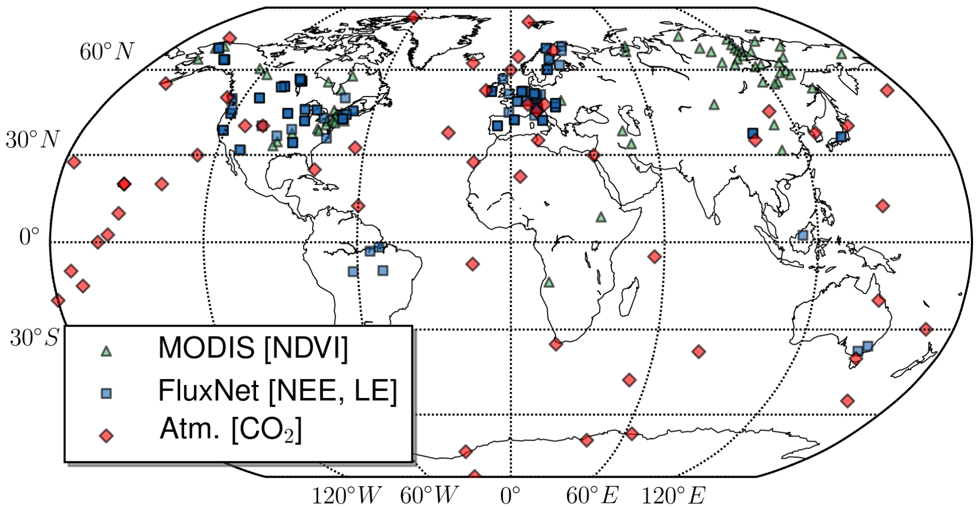
Natasha MacBean[1,2], Fabienne Maignan[1], Cédric Bacour[3], Philip Lewis[4,5], Philippe Peylin[1], Luis Guanter[6], Philipp Köhler[7], Jose Gómez-Dans[4,5] & Mathias Disney[4,5]

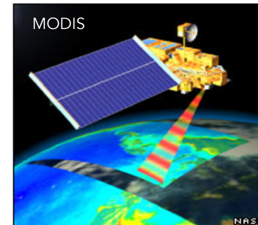
Multiple constraint on global carbon stocks and fluxes

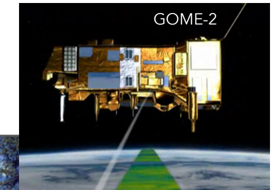


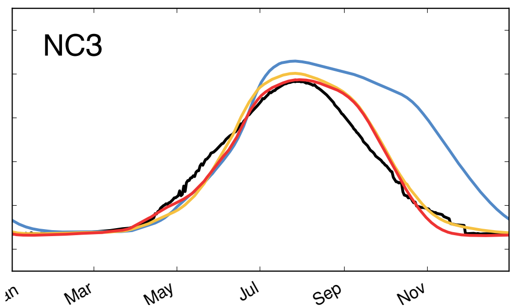


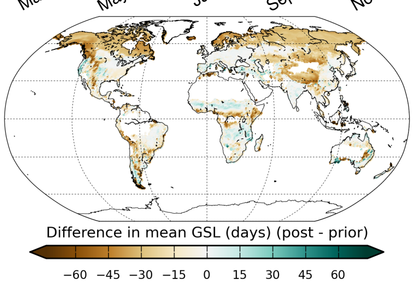
Multiple constraint on global carbon stocks and fluxes





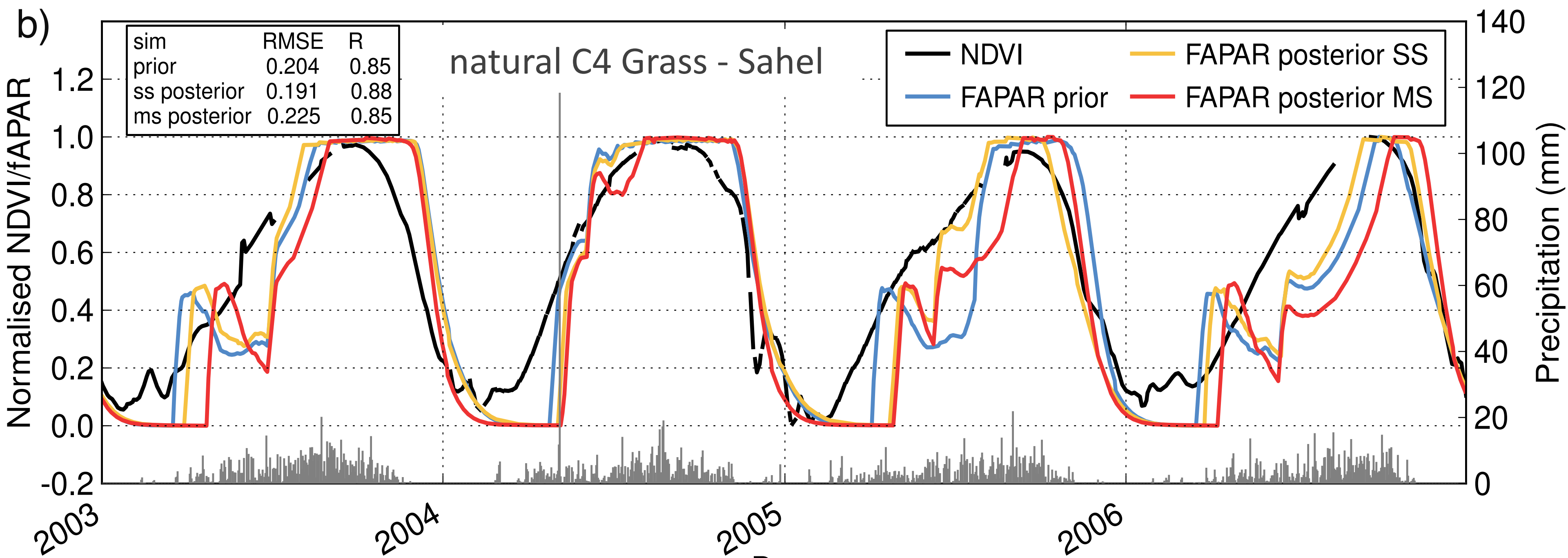






MacBean et al. (2015)

→ Insight into vegetation dynamics in semi-arid/dryland ecosystems…

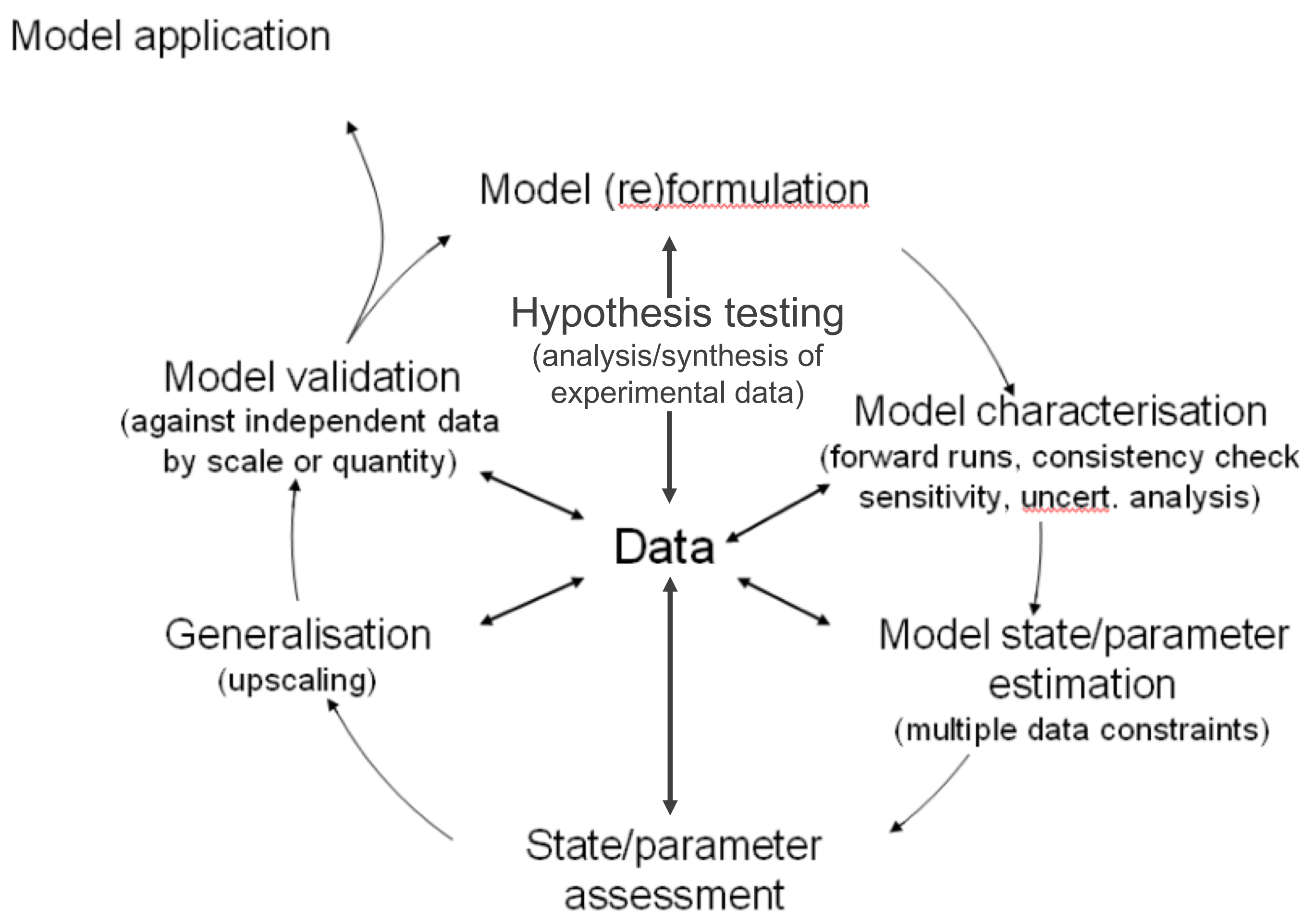


➢ Leaf onset/senescence controlled by moisture availability in these ecosystems (time since moisture minimum)

➢ *How does moisture availability control leaf dynamics?*

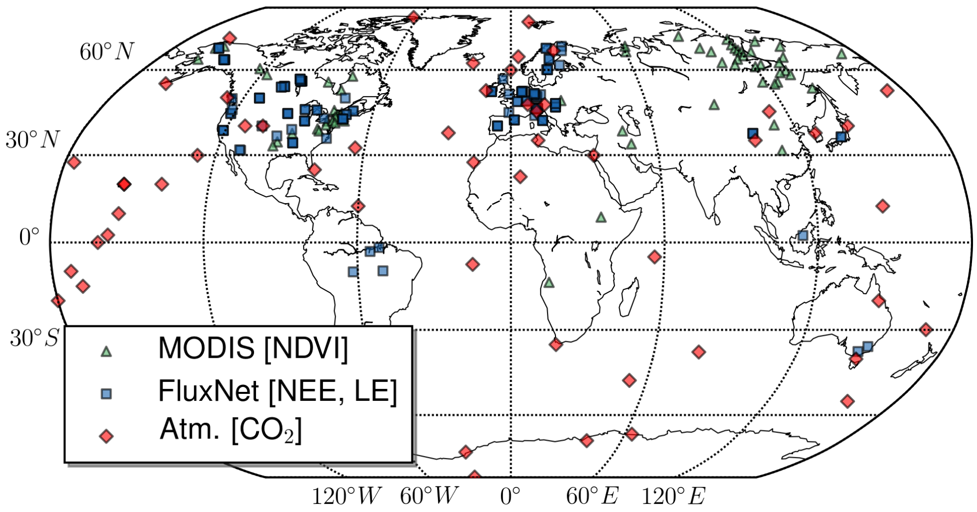
MacBean et al. (2015) Using satellite data to improve the leaf phenology of a global Terrestrial Biosphere Model, *Biogeosciences,* 12, 7185-7208

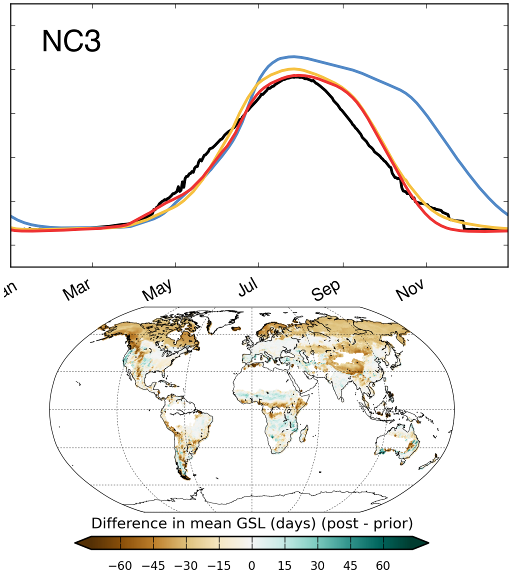
Improving models | Reducing uncertainty → the model development cycle



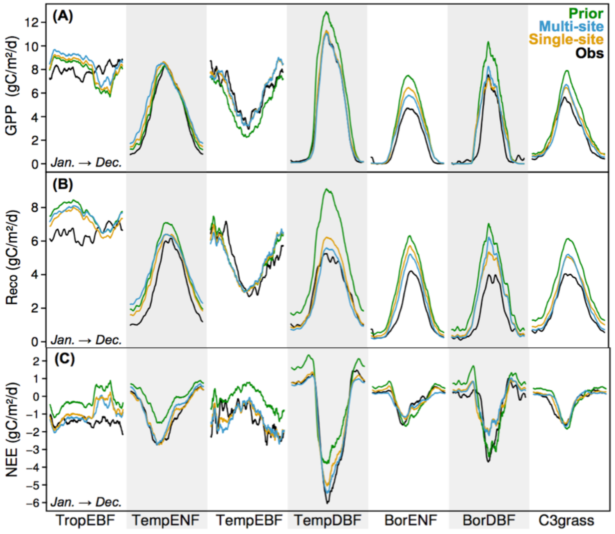
Adapted from Williams et al. (2009)

Multiple constraint on global carbon stocks and fluxes



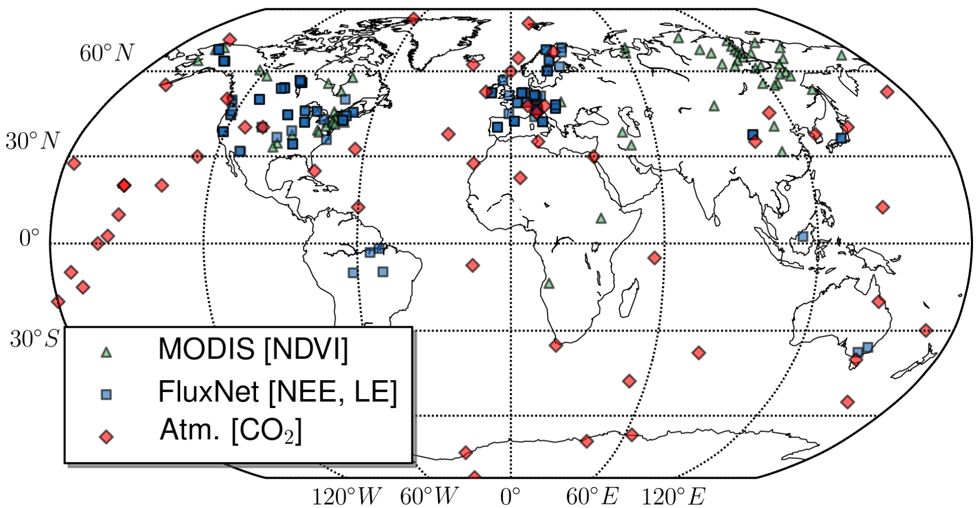


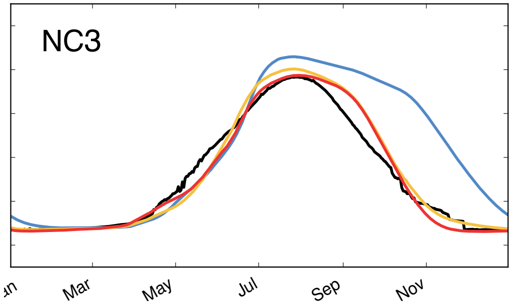
MacBean et al. (2015)

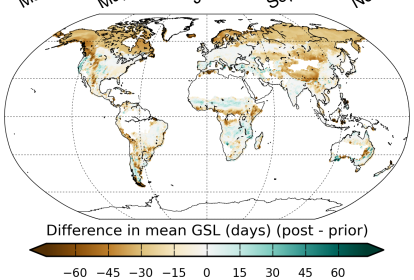


Kuppel et al. (2014)

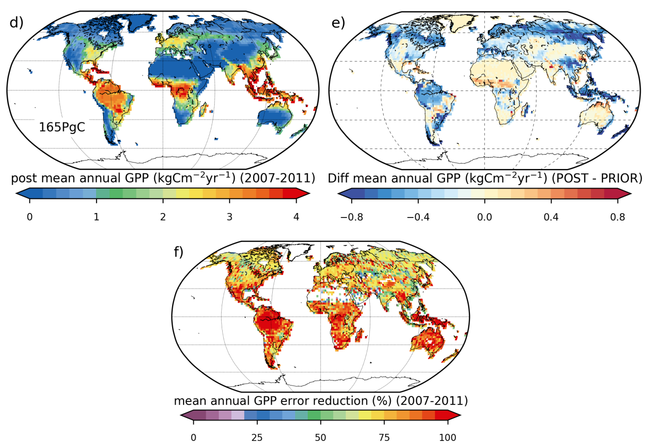
Multiple constraint on global carbon stocks and fluxes





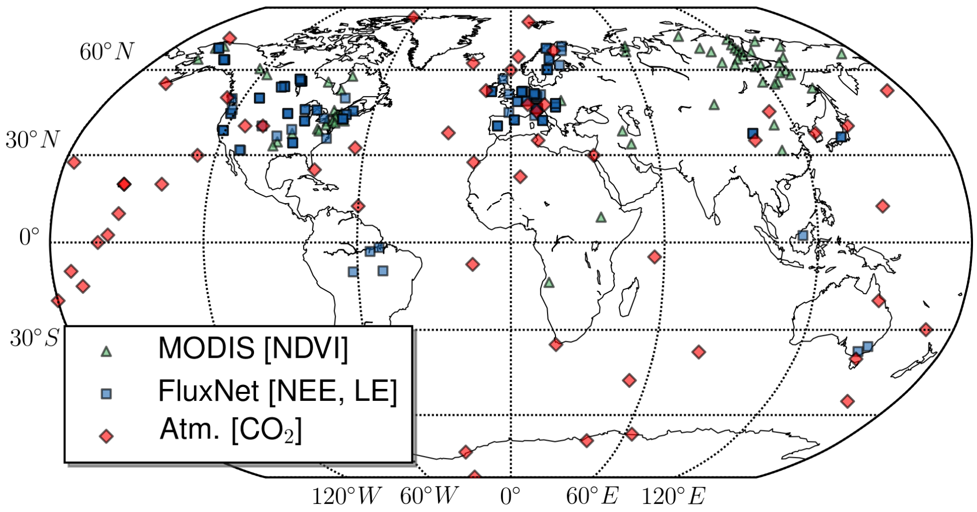


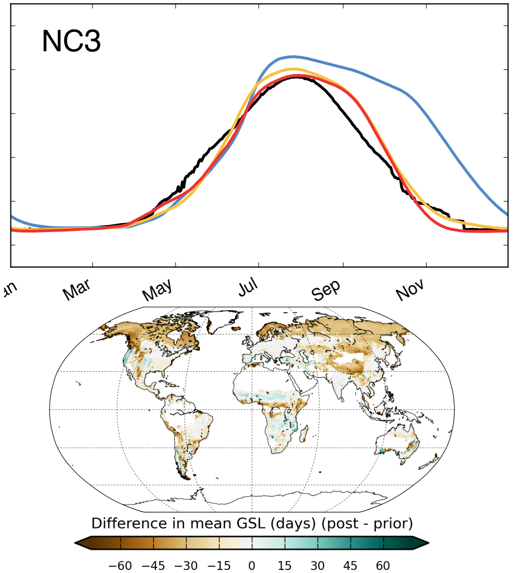
MacBean et al. (2015)



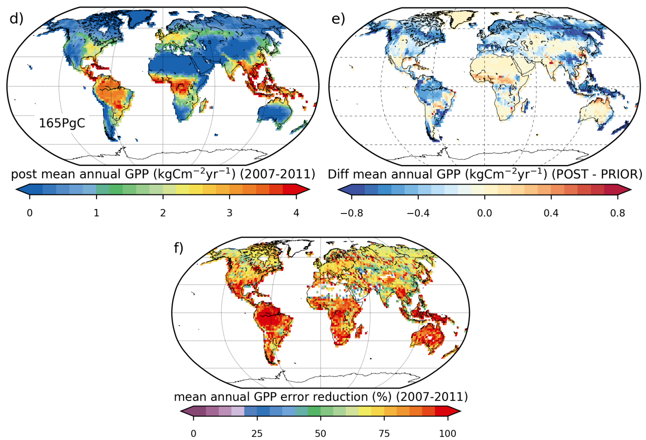
MacBean et al. (2018)

Multiple constraint on global carbon stocks and fluxes

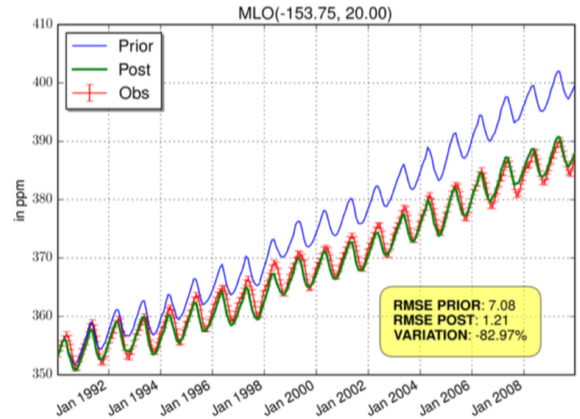




MacBean et al. (2015)

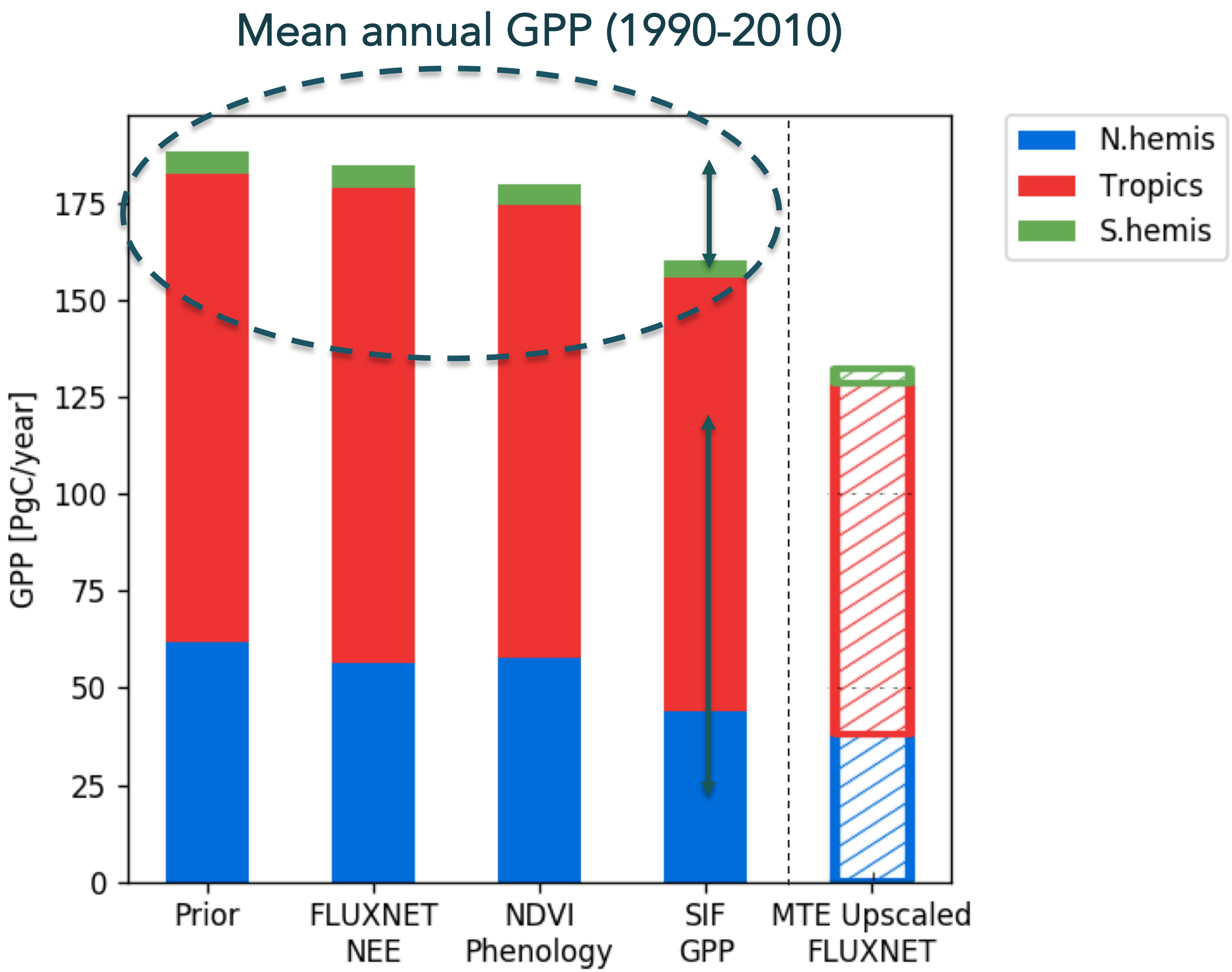


MacBean et al. (2018)



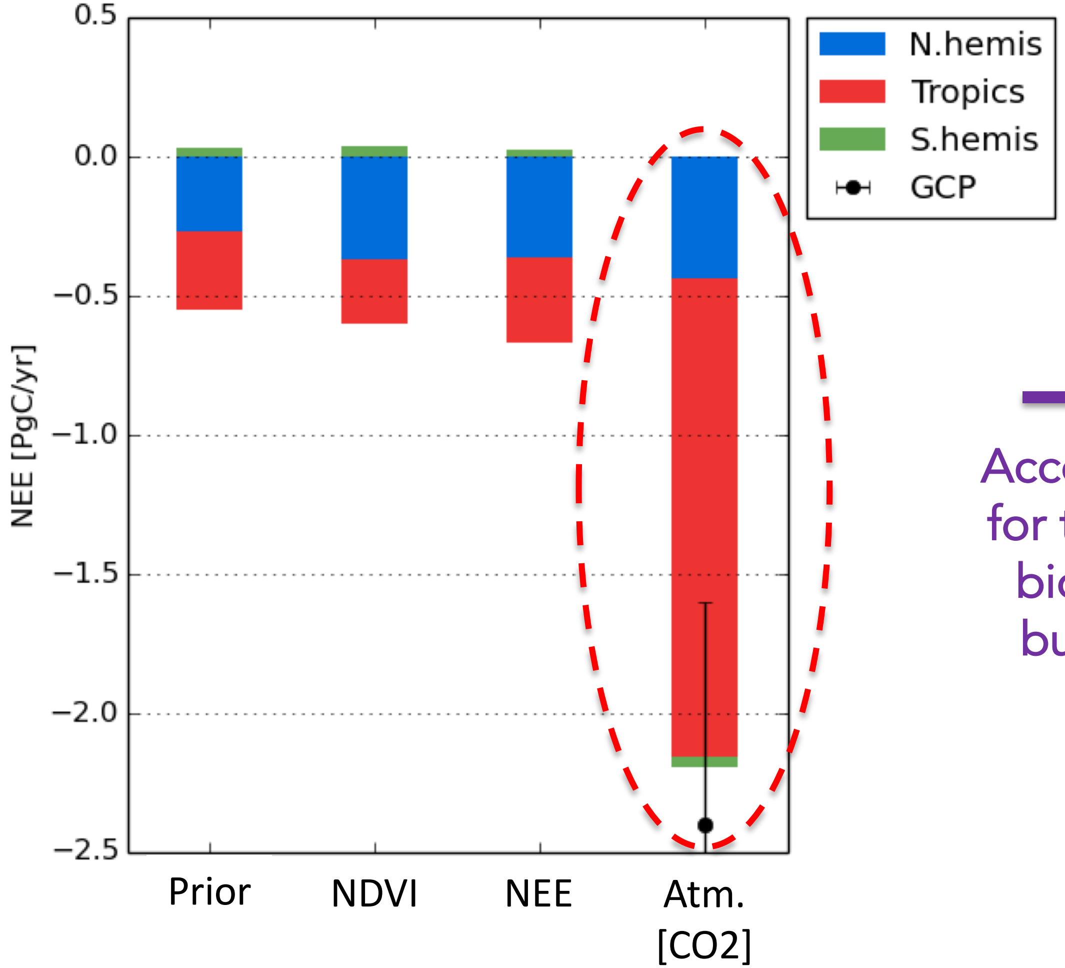
Peylin et al. (2016)

Satellites help us constrain GPP magnitude and global spatial distribution

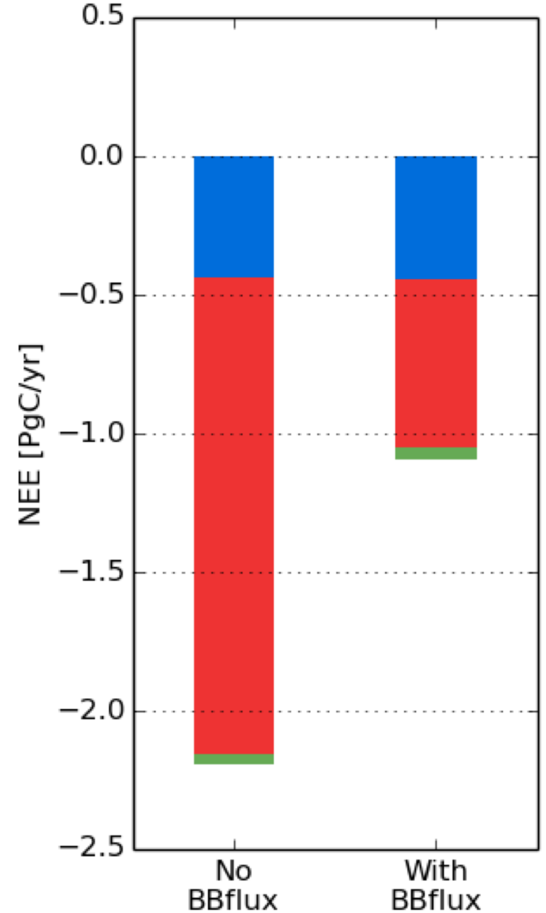


Atmospheric [CO_2] data help us constrain spatial distribution of net CO_2 fluxes

Posterior land net CO_2 flux



Accounting for tropical biomass burning

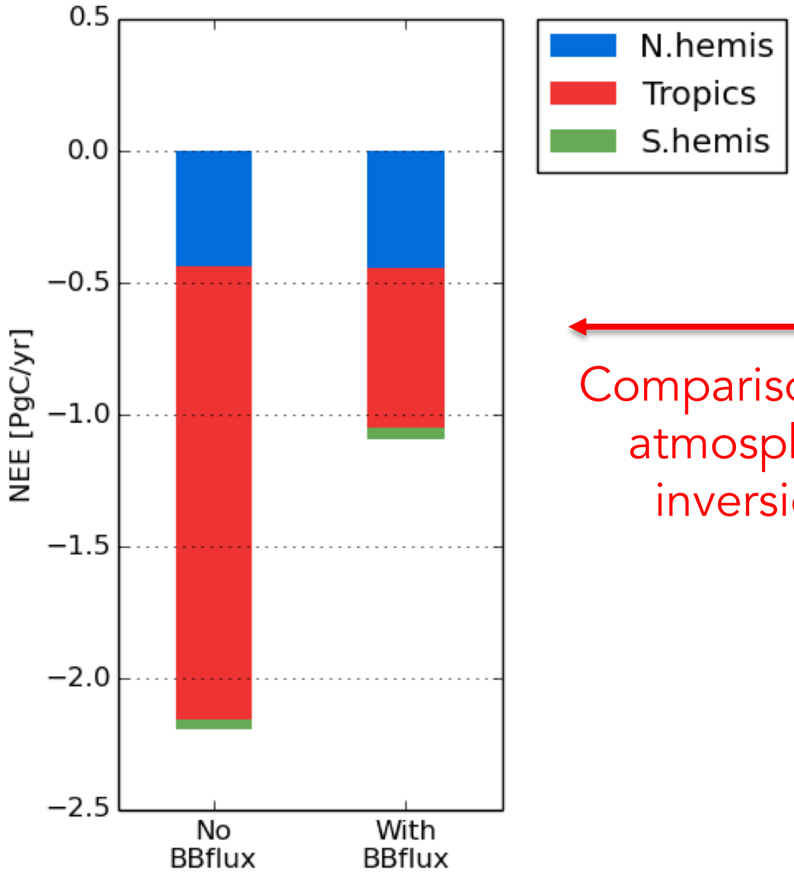


P. Peylin et al. (2016)

P. Peylin et al. (2013)

Toward a comparison with atmospheric inversions…

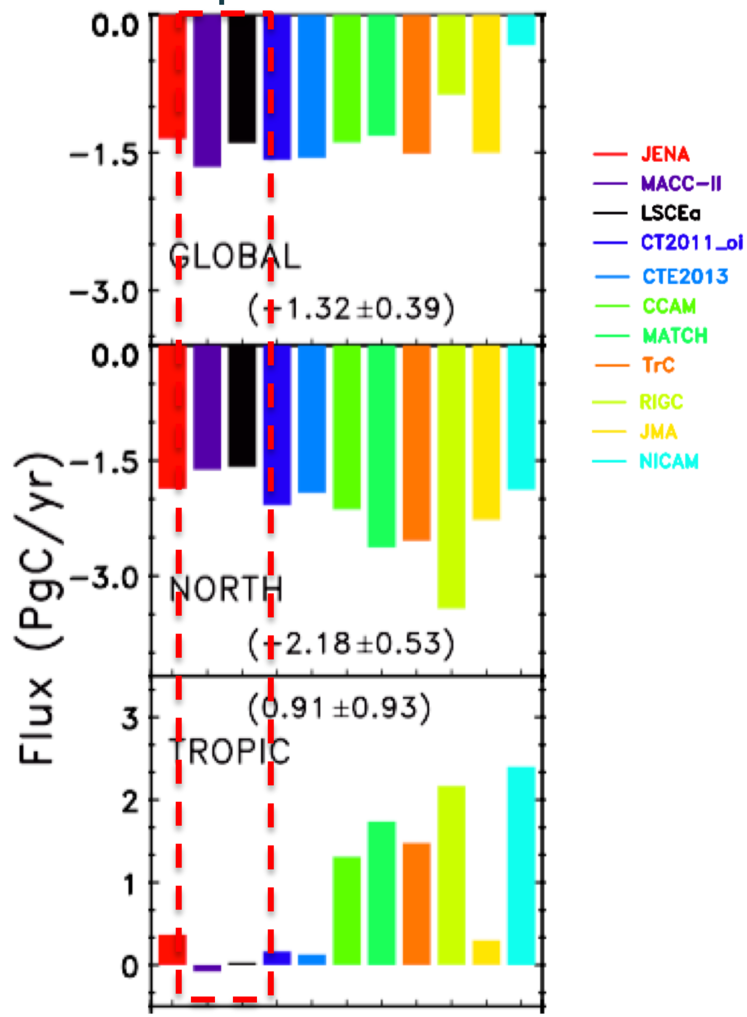
Posterior land net CO_2 flux



P. Peylin et al. (2016)

Comparison with atmospheric inversions

Land net CO_2 flux from suite of atmospheric inversions



P. Peylin et al. (2013)

Outline

Data assimilation as a tool for reducing uncertainty in land surface models

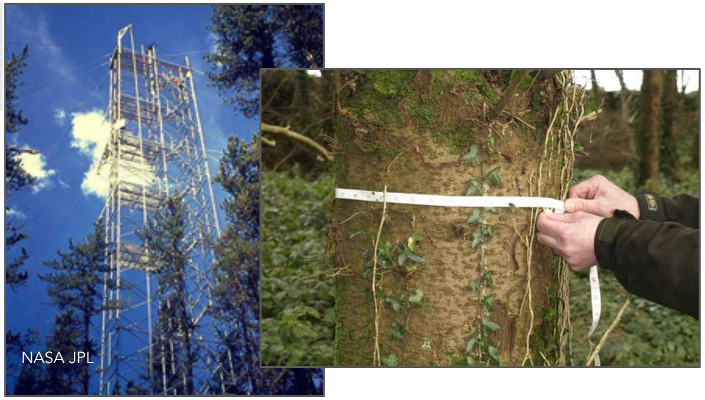
- Highlights with the ORCHIDEE LSM
- *But we face challenges…*
- Future perspectives

Progress in using multiple datasets to constrain models

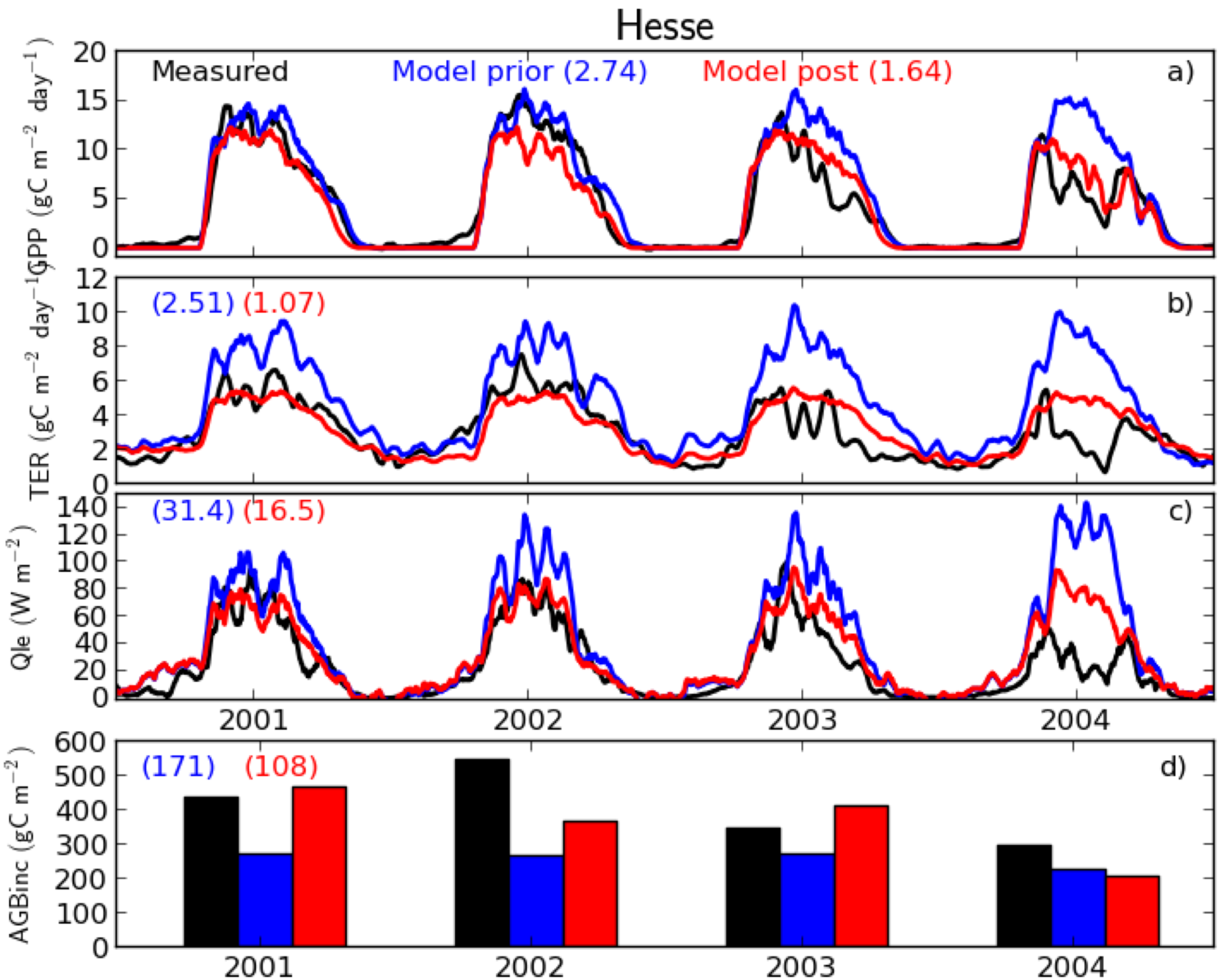
Agricultural and Forest Meteorology

The potential benefit of using forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites

T. Thum[a,*], N. MacBean[b], P. Peylin[b], C. Bacour[c], D. Santaren[b], B. Longdoz[d], D. Loustau[e], P. Ciais[b]



Challenges of multiple data stream assimilation → fluxes + aboveground biomass increment



➔ Assimilate obs:

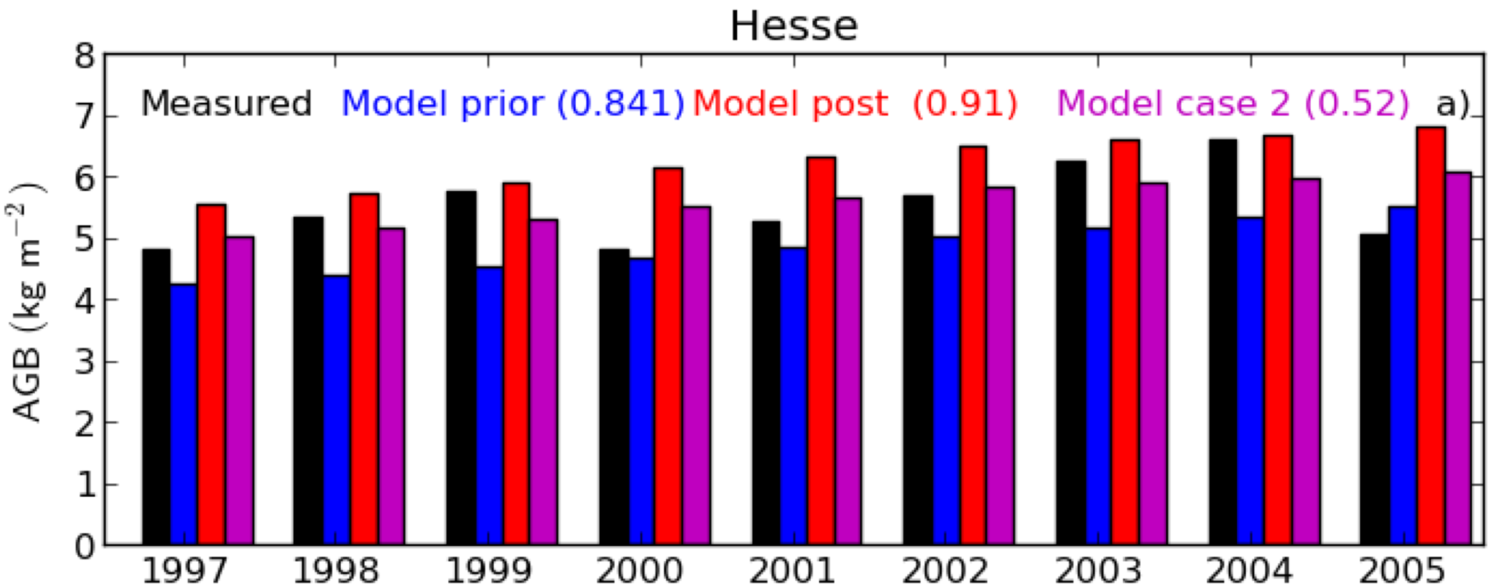
- GPP
- Reco (TER)
- Latent Heat (Qle)
- ***AGB increment***

➔ Optimise params:

- photosynthesis
- respiration
- energy balance
- soil water avail.
- phenology
- ***allocation***

Thum, T., et al. (2017), The potential of forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites, *Agric. For. Meteorol*, 234, 48-65

Challenges of multiple data stream assimilation → fluxes + aboveground biomass increment



- → Assimilate AGB increment degrades fit to *total* AGB
- → Assimilate total AGB
- → Optimise turnover rate
- → *BUT missing model processes?*

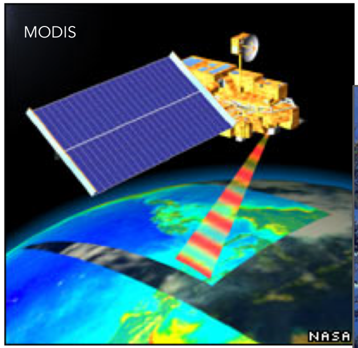
Thum, T., et al. (2017), The potential of forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites, *Agric. For. Meteorol*, 234, 48-65.

Progress in using multiple datasets to constrain models

Agricultural and Forest Meteorology

The potential benefit of using forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites

T. Thum [a,*], N. MacBean [b], P. Peylin [b], C. Bacour [c], D. Santaren [b], B. Longdoz [d], D. Loustau [e], P. Ciais [b]

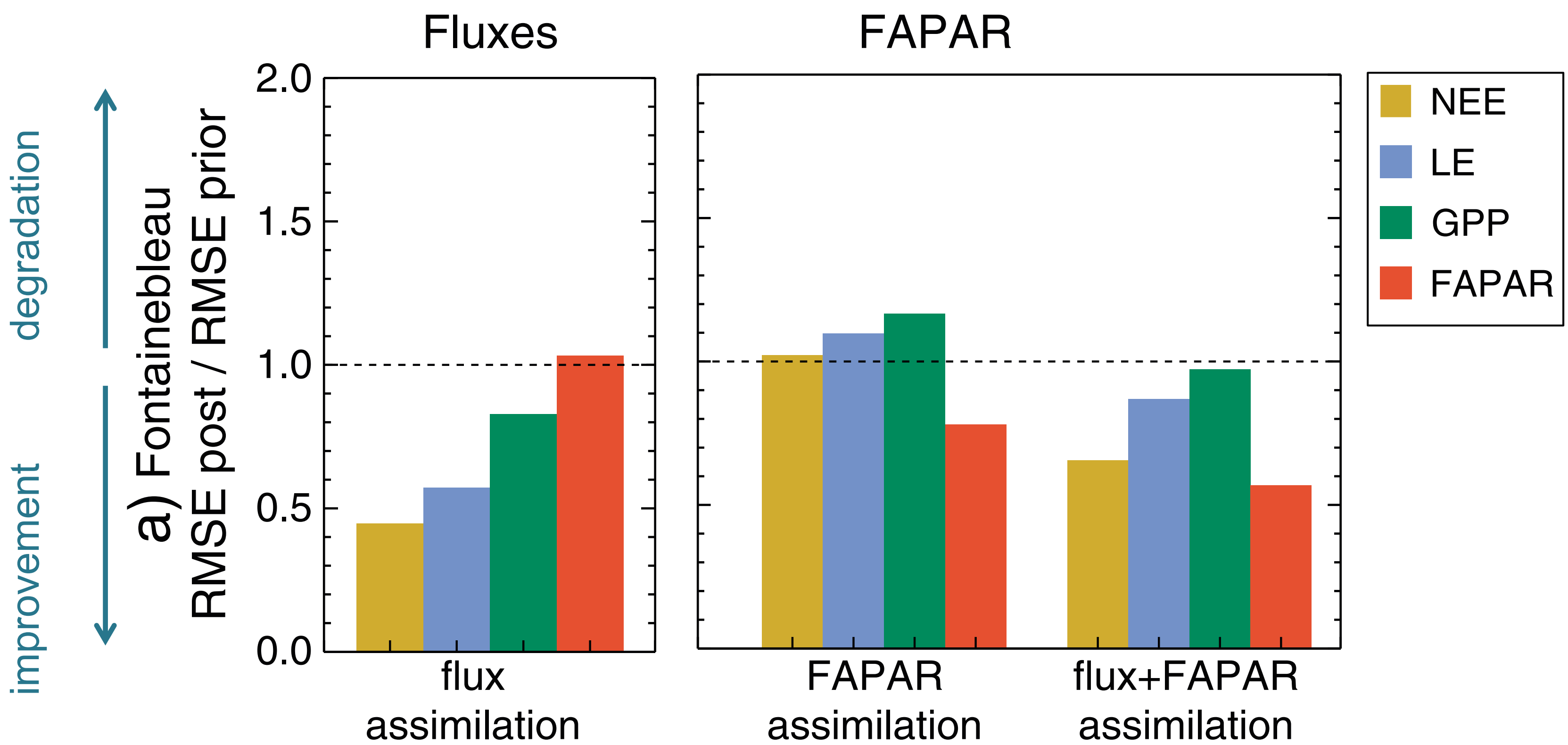


Journal of Geophysical Research: Biogeosciences

Joint assimilation of eddy covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model

C. Bacour[1,2], P. Peylin[2], N. MacBean[2], P. J. Rayner[2,3], F. Delage[2,4], F. Chevallier[2], M. Weiss[5], J. Demarty[5,6], D. Santaren[7,8], F. Baret[5], D. Berveiller[9], E. Dufrêne[9], and P. Prunet[1]

Challenges of multiple data stream assimilation
→ fluxes + satellite FAPAR

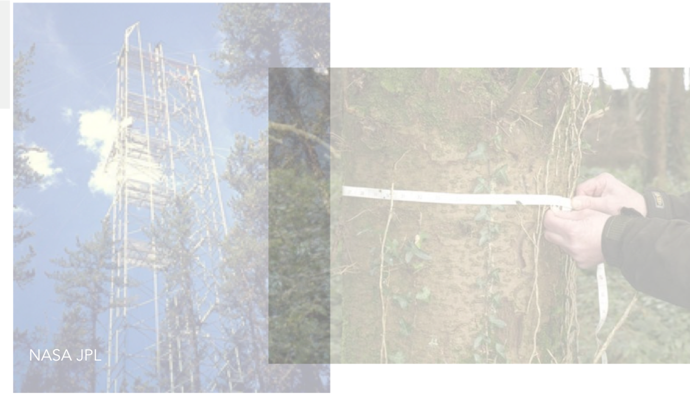


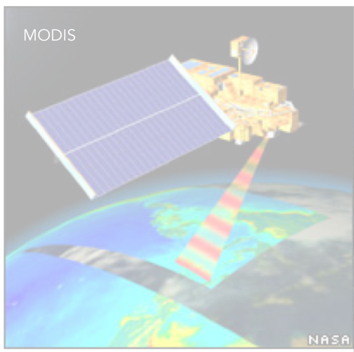
Challenges and progress in using multiple datasets to constrain models

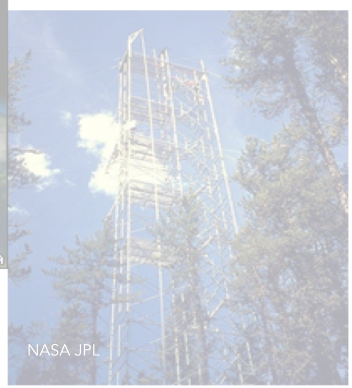
Agricultural and Forest Meteorology

The potential benefit of using forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites

T. Thum[a,*], N. MacBean[b], P. Peylin[b], C. Bacour[c], D. Santaren[b], B. Longdoz[d], D. Loustau[e], P. Ciais[b]







Journal of Geophysical Research: Biogeosciences

Joint assimilation of eddy covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model

C. Bacour[1,2], P. Peylin[2], N. MacBean[2], P. J. Rayner[2,3], F. Delage[2,4], F. Chevallier[2], M. Weiss[5], J. Demarty[5,6], D. Santaren[7,8], F. Baret[5], D. Berveiller[9], E. Dufrêne[9], and P. Prunet[1]

Geosci. Model Dev., 9, 3569–3588, 2016

Consistent assimilation of multiple data streams in a carbon cycle data assimilation system

Natasha MacBean[1], Philippe Peylin[1], Frédéric Chevallier[1], Marko Scholze[2], and Gregor Schürmann[3]

Challenges of multiple data stream assimilation → perspectives for land surface modelers

- Biases in observations or inconsistencies between observations and the model

- Proper characterization of the parameter and observation error covariance matrices, particularly issues of #obs and covariance between obs

- Step-wise vs simultaneous approaches in the case of computational constraints

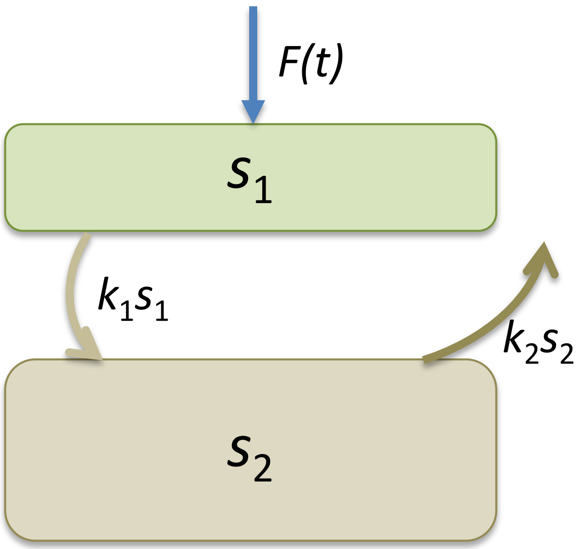
- Assumptions of the inversion algorithm (derivative vs global random search)

MacBean, N., P. Peylin, F. Chevallier, M. Scholze and G. Schürmann (2016), Consistent assimilation of multiple data streams in a carbon cycle data assimilation system, *Geosci. Model Dev.*, 9, 3569-3588.

Challenges of multiple data stream assimilation → Toy model examples

E.g. Simple C cycle model (2 pools)

$$\frac{ds_1}{dt} = F(t)\left(\frac{s_1}{p_1+s_1}\right)\left(\frac{s_2}{p_2+s_2}\right) - k_1 s_1 + s_0$$
$$\frac{ds_2}{dt} = k_1 s_1 - k_2 s_2.$$



- Observations of s_1 and s_2 pools
- 'k' and 'p' parameters for each pool
- Synthetic "pseudo obs" tests
- Bias in s_2 pool → not taken into account in cost function

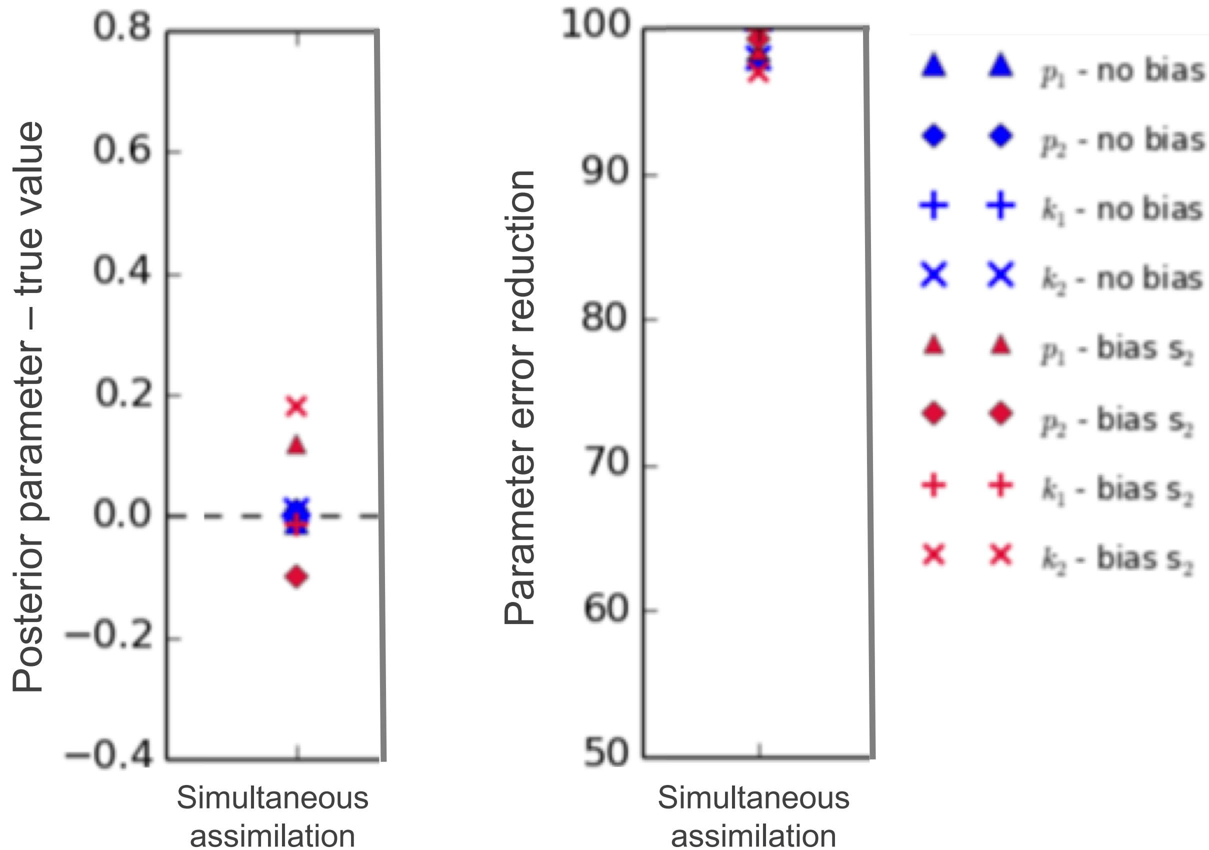
MacBean, N., P. Peylin, F. Chevallier, M. Scholze and G. Schürmann (2016), Consistent assimilation of multiple data streams in a carbon cycle data assimilation system, *Geosci. Model Dev.*, 9, 3569-3588.

Challenges of multiple data stream assimilation → e.g. bias in obs not accounted for cost function

→ Comparison with and without bias in s_2 observation

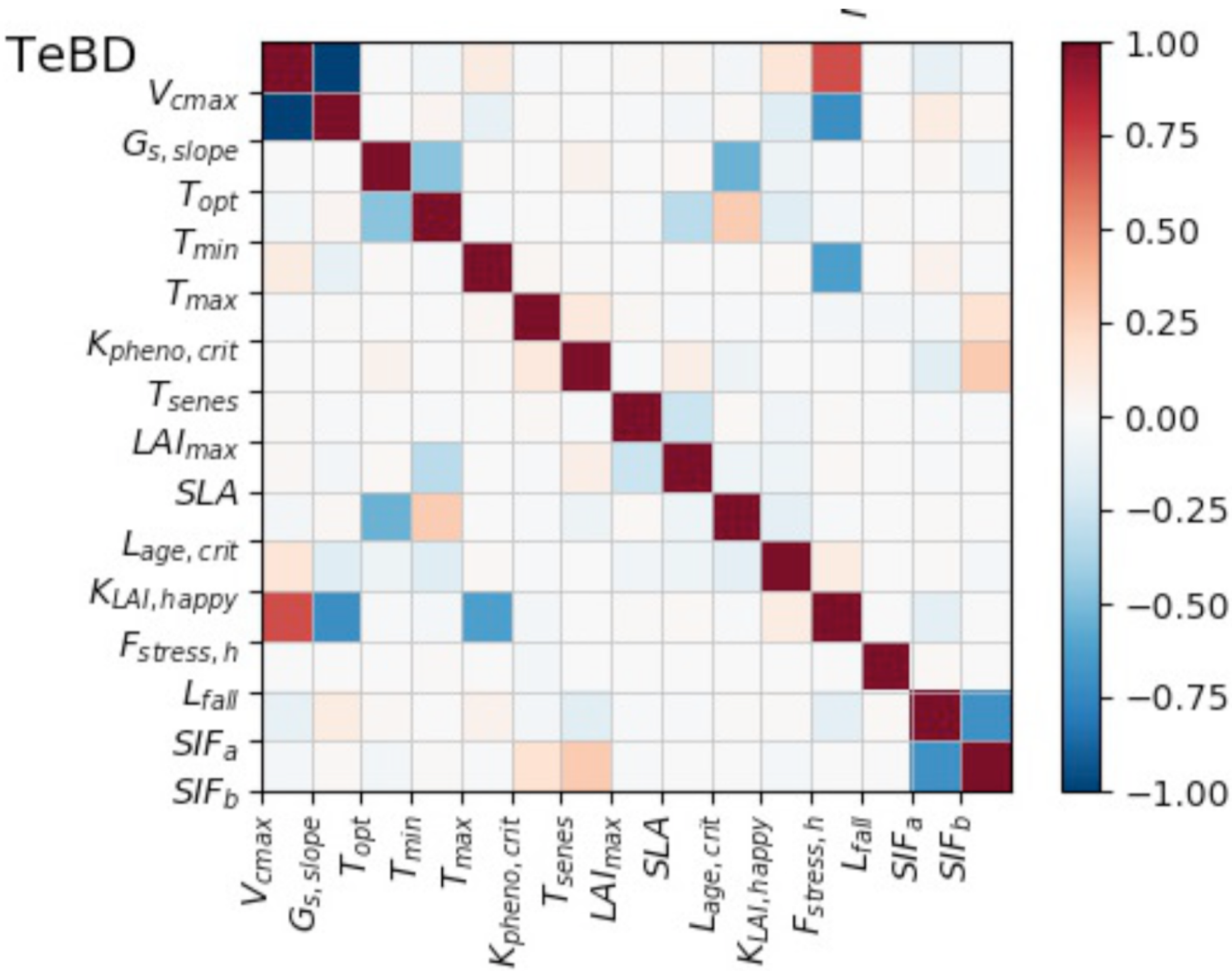
$$\frac{ds_1}{dt} = F(t)\left(\frac{s_1}{p_1+s_1}\right)\left(\frac{s_2}{p_2+s_2}\right) - k_1 s_1 + s_0$$

$$\frac{ds_2}{dt} = k_1 s_1 - k_2 s_2.$$



MacBean, N., P. Peylin, F. Chevallier, M. Scholze and G. Schürmann (2016), Consistent assimilation of multiple data streams in a carbon cycle data assimilation system, *Geosci. Model Dev.*, 9, 3569-3588.

A note on parameter correlation and model equifinality
→ problem in complex *and* simple models…



Outline

Data assimilation as a tool for reducing uncertainty in land surface models

- Highlights with the ORCHIDEE LSM
- Challenges we face
- *Where to next?*

Global Data Assimilation System – global CCDAS

GLOBAL BIOGEOCHEMICAL CYCLES, VOL. 19, GB2026, doi:10.1029/2004GB002254, 2005

Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS)

P. J. Rayner,[1,2] M. Scholze,[3,4] W. Knorr,[4,5] T. Kaminski,[6] R. Giering,[6] and H. Widmann[5]

JOURNAL OF GEOPHYSICAL RESEARCH: BIOGEOSCIENCES, VOL. 118, 1–13, doi:10.1002/jgrg.20118, 2013

The BETHY/JSBACH Carbon Cycle Data Assimilation System: experiences and challenges

T. Kaminski,[1] W. Knorr,[2] G. Schürmann,[3] M. Scholze,[2] P. J. Rayner,[4] S. Zaehle,[3] S. Blessing,[1] W. Dorigo,[5] V. Gayler,[6] R. Giering,[1] N. Gobron,[7] J. P. Grant,[2]

Geosci. Model Dev., 9, 2999–3026, 2016

Constraining a land-surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System V1.0

Gregor J. Schürmann[1], Thomas Kaminski[2,a], Christoph Köstler[1], Nuno Carvalhais[1], Michael Voßbeck[2,a], Jens Kattge[1], Ralf Giering[3], Christian Rödenbeck[1], Martin Heimann[1], and Sönke Zaehle[1,4]

Biogeosciences, 11, 7025–7050, 2014

Identifying environmental controls on vegetation greenness phenology through model–data integration

M. Forkel[1], N. Carvalhais[1,2], S. Schaphoff[3], W. v. Bloh[3], M. Migliavacca[1], M. Thurner[1,4], and K. Thonicke[3]

Geosci. Model Dev., 9, 2833–2852, 2016

Land-surface parameter optimisation using data assimilation techniques: the adJULES system V1.0

Nina M. Raoult, Tim E. Jupp, Peter M. Cox, and Catherine M. Luke

Biogeosciences, 10, 2011–2040, 2013

Multiple observation types reduce uncertainty in Australia's terrestrial carbon and water cycles

V. Haverd[1], M. R. Raupach[1], P. R. Briggs[1], J. G. Canadell[1], P. Isaac[1], C. Pickett-Heaps[1], S. H. Roxburgh[2], E. van Gorsel[1], R. A. Viscarra Rossel[3], and Z. Wang[1,4]

PNAS

The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times

A. Anthony Bloom[a,b,c,1], Jean-François Exbrayat[b,c], Ivar R. van der Velde[d], Liang Feng[b,c], and Mathew Williams[b,c]

Geosci. Model Dev., 9, 3321–3346, 2016

A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle

Philippe Peylin[1], Cédric Bacour[2], Natasha MacBean[1], Sébastien Leonard[1], Peter Rayner[1,3], Sylvain Kuppel[1,4], Ernest Koffi[1], Abdou Kane[1], Fabienne Maignan[1], Frédéric Chevallier[1], Philippe Ciais[1], and Pascal Prunet[2]

Biogeosciences, 10, 583–606, 2013

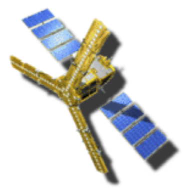
The climate dependence of the terrestrial carbon cycle, including parameter and structural uncertainties

M. J. Smith, D. W. Purves, M. C. Vanderwel, V. Lyutsarev, and S. Emmott

Model data assimilation in the "Big Data" era → short-term processes

➢ New ISS and satellite products, FLUXNET, SAPFLUXNET, ISMN

International soil moisture network (ISMN)



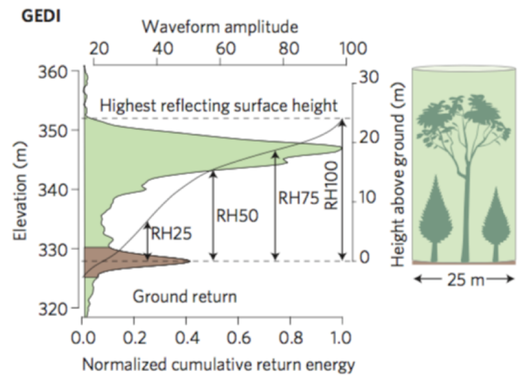
SMOS satellite

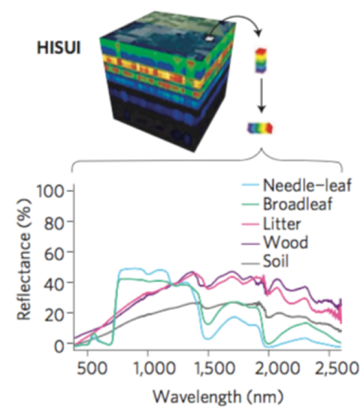
SMAP satellite

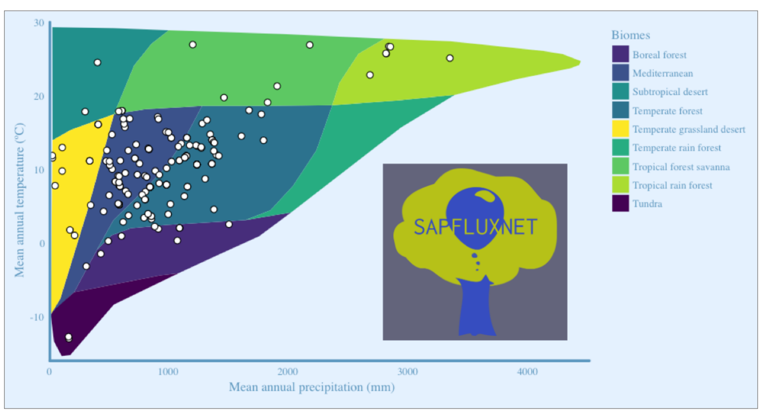
FLUXNET A Global Network

AMERIFLUX

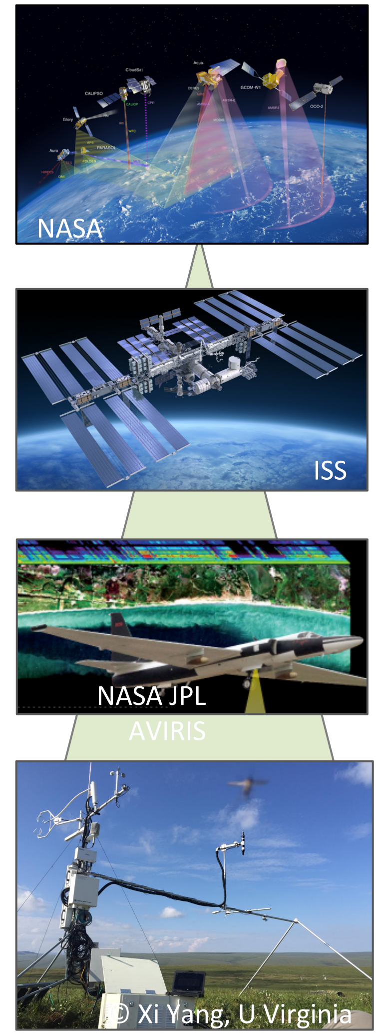
ICOS integrated carbon observation system







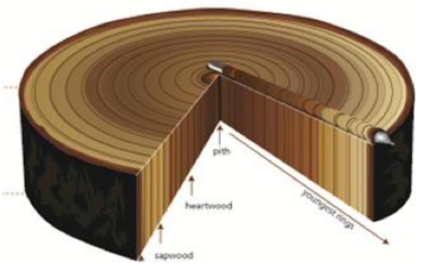
ISCN International Soil Carbon Network

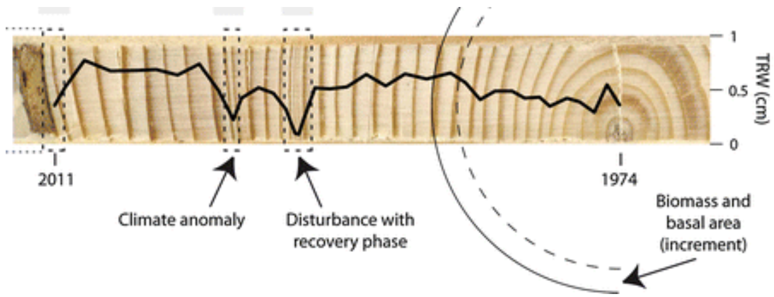


Relationships across scales

Model data assimilation in the "Big Data" era
→ long-term processes and CO_2/climate sensitivities

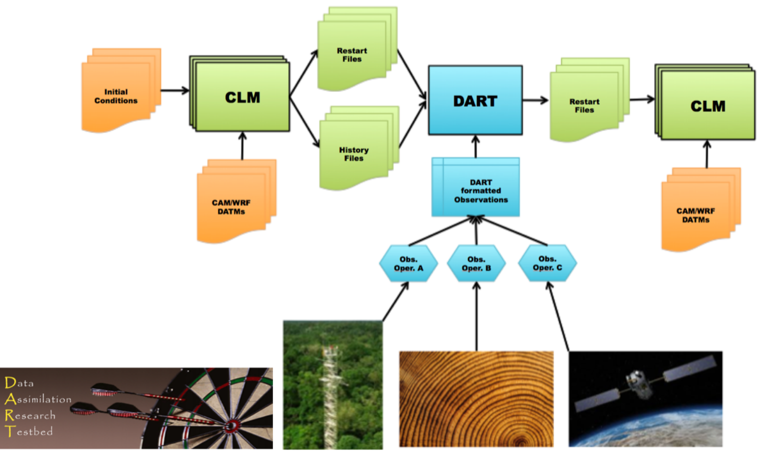
➢ FACE sites, manipulation expts (DroughtNet), Tree rings (ITRDB)



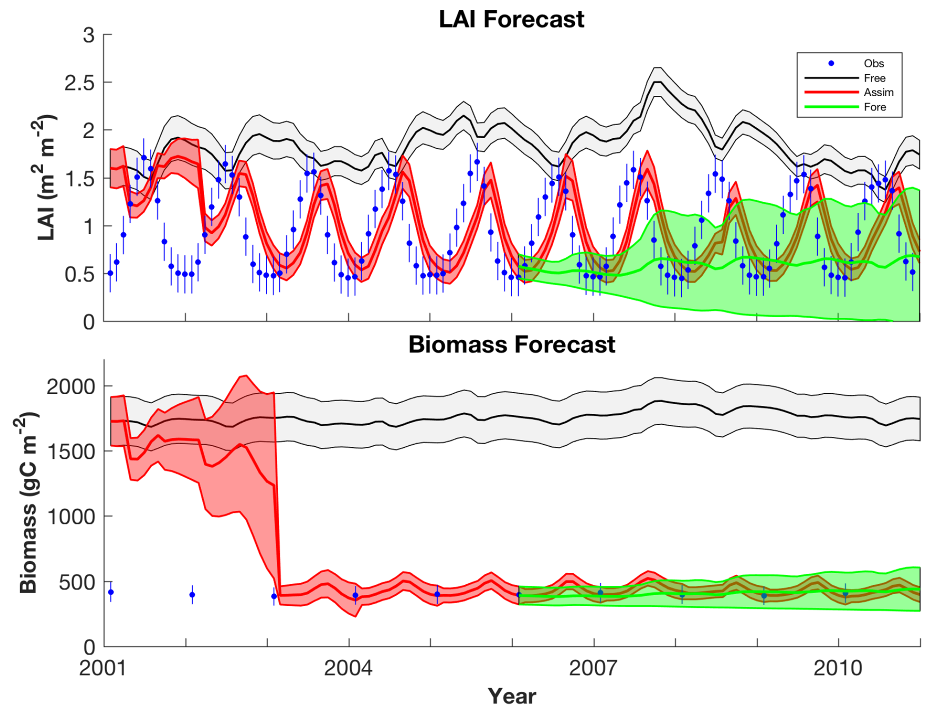


Improving carbon cycle prediction using *state* data assimilation with CLM-DART (led by Andy Fox + Tim Hoar & Jeff Anderson)

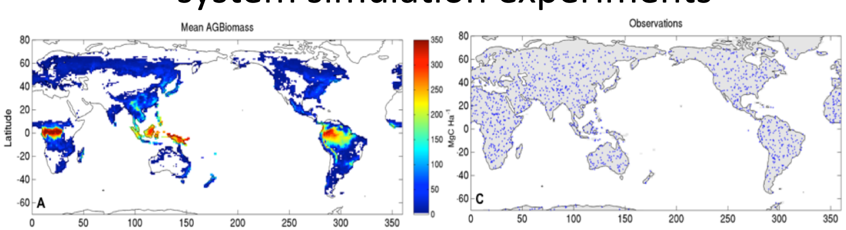
1. Added carbon cycle relevant observations to CLM-DART workflow



2. Assimilating leaf area and biomass obs. improves seasonality and prediction



3. Currently testing approach globally with observing system simulation experiments



Data assimilation wish list

- Better quantification of observation errors and their correlations
- Methods for quantification of model structural error
- Sensitivity analyses
- Resolve inconsistencies between model representation and observations → keep data in mind when developing models
- Modularized code → can optimize more easily & test different processes/structure/complexity
- Think about the terms in the cost function (relationships between variables, sensitivities, emerging constraints)
- Updated inversion algorithms
- Operational DA with newer versions of models

Wish list for land surface model development

- There is a clear need to better quantify and reduce uncertainty of C budget estimates/projections
- Need more time, people, funding, and recognition for technical testing, experiments, and developments

- *Move to a more diverse set of uses for these models?*

  → *Hypothesis testing*

  → *Stakeholder adaptation/mitigation planning and policy decision making? E.g. short-term agricultural productivity and drought monitoring?*

Thank you for listening! Any questions?

nlmacbean@gmail.com
www.nmacbean.wordpress.com
@NatashaMacBean