Challenges with multi-data stream parameter optimization

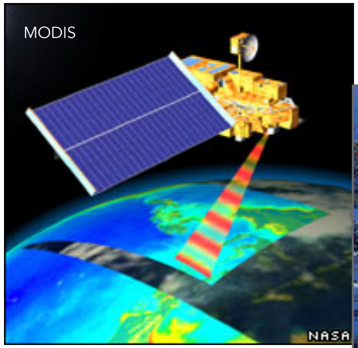
Natasha MacBean, Philippe Peylin, Cédric Bacour, Vladislav Bastrikov
(& ORCHIDAS Team)

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]

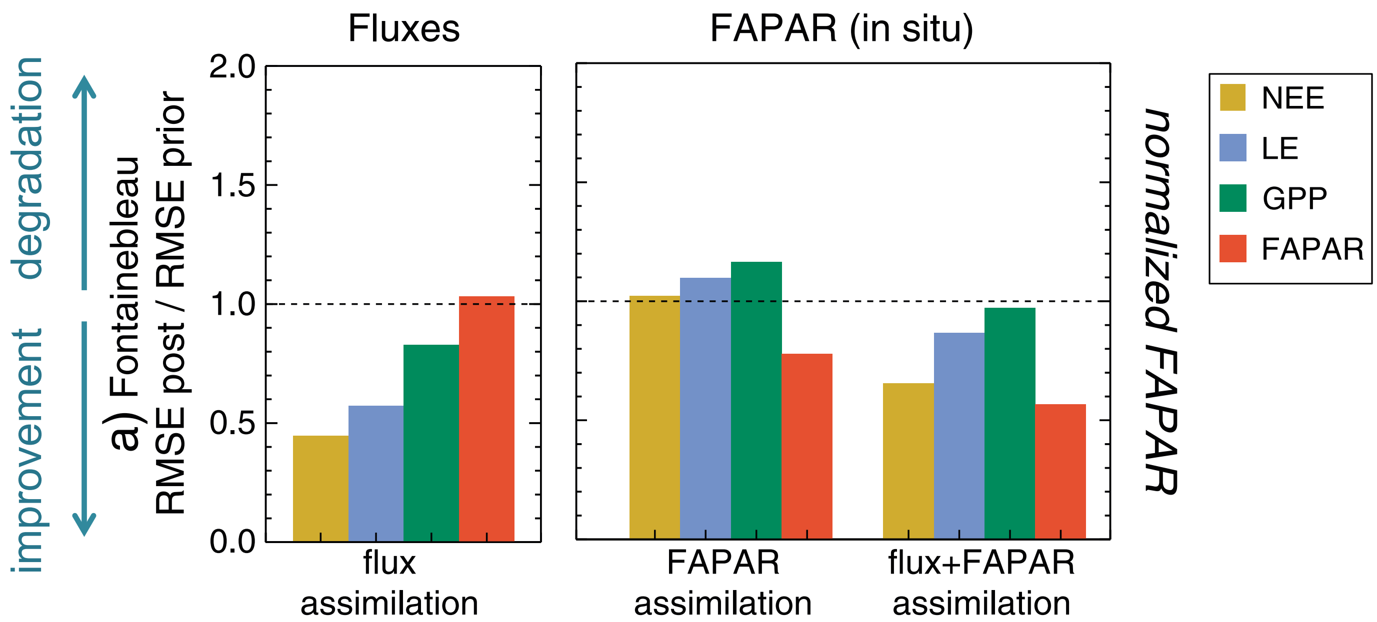


Journal of Geophysical Research: Biogeosciences

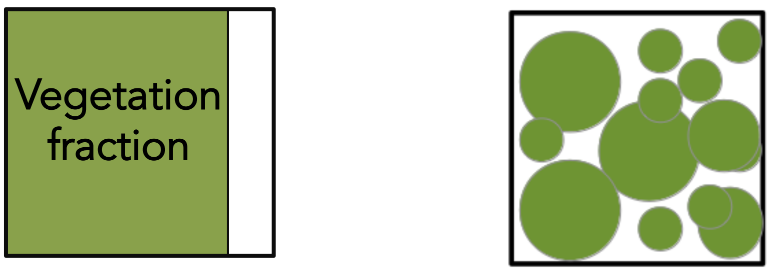
Joint assimilation of eddy covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model

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



- Inconsistencies between model and data → aliased onto parameters that then result in degradation of fit to other variables

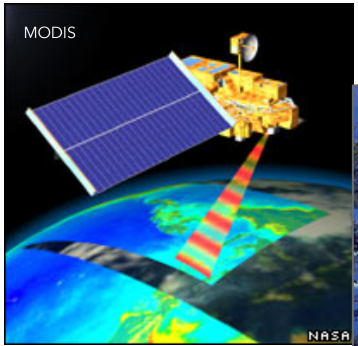


Progress in using multiple datasets to constrain models

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
www.geosci-model-dev.net/9/3569/2016/
doi:10.5194/gmd-9-3569-2016

Geoscientific Model Development
Open Acce

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
→ Toy model examples

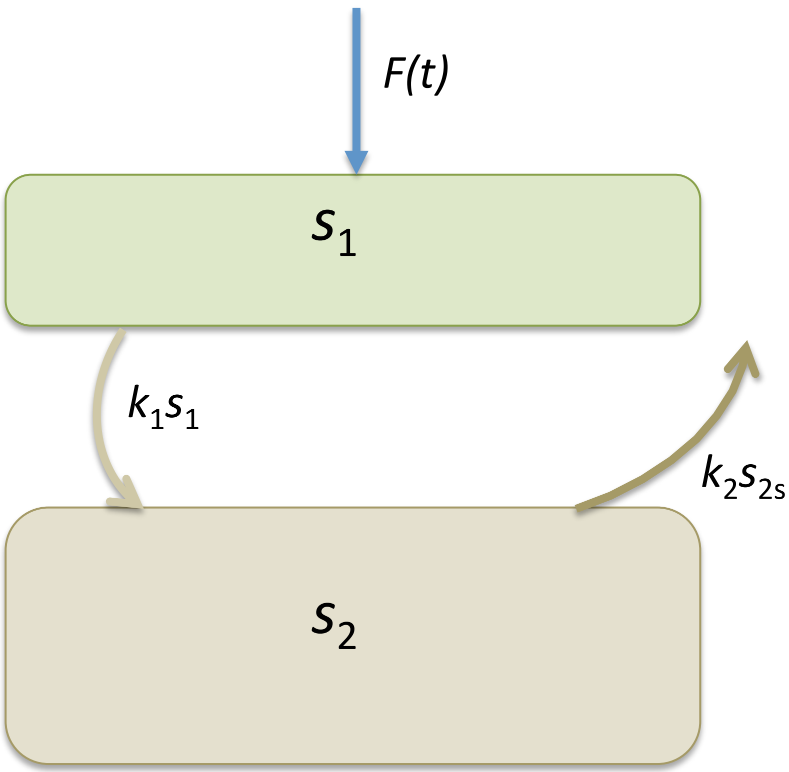
Two toy models:

1) Simple C cycle model (2 pools)

- Synthetic DA experiment with "pseudo" observations

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

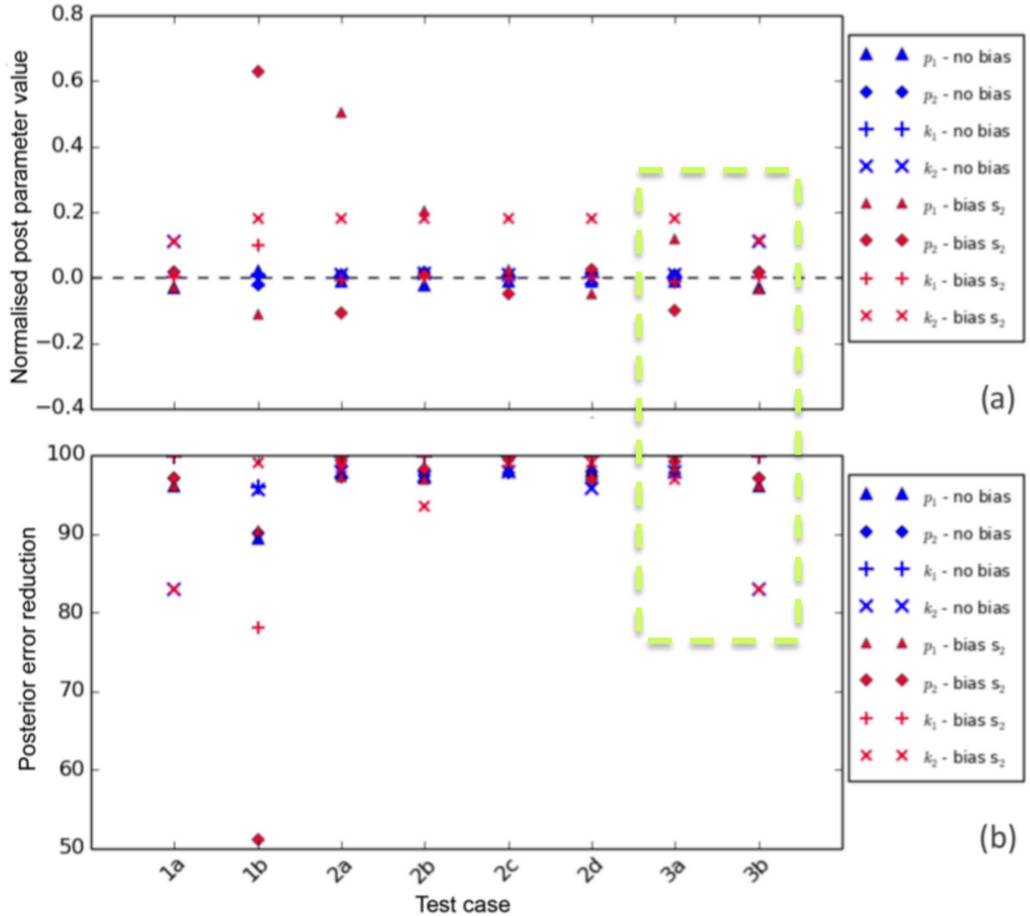
$$\frac{\mathrm{d}s_2}{\mathrm{d}t} = k_1 s_1 - k_2 s_2.$$



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.

Challenges of multiple data stream assimilation
→ bias in obs not accounted for in error covariance matrix

Simple C cycle (2 pools) → bias in s_2 variable



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 → inversions assumptions (e.g. linear vs non linear models)

- You will get similar issues if you do not use adhere to the assumptions of the inversion algorithm…

  → e.g. if your inversion algorithm is meant for linear models (which many of the computationally efficient ones are) then you may get incorrect optimized parameters that ****look**** like they are well constrained from the uncertainty reduction.

- Read more here: ☺

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.

Geosci. Model Dev., 9, 3569–3588, 2016
www.geosci-model-dev.net/9/3569/2016/
doi:10.5194/gmd-9-3569-2016

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]

Parameter sensitivity study

- Motivation: Need to figure out parameters to which variables (e.g. NEE, soilC) are most sensitive:

  i) to know which parameters to optimize (reduce computational load) for which variables

  ii) to know relationships between parameters and generally how model is working

- Propose: global scale SA (or at least many grid points over many different PFTs/biomes) – e.g. Morris method

- Will get help from Indiana University software engineers as part of the Research Technologies "Deep Learning" division (who have a lot of expertise in this type of algorithm and in optimizing "big" codes).

- Scripts will be available for future SA studies

- Would like collaborators from the ORCHIDEE Project Group (and associated postdocs/PhD students etc) to help with:

  - Define which parameters to include and their bounds (max, min value and PDF if appropriate → this is a crucial part of the study (and hopefully useful for the group as a whole)

  - Design of study

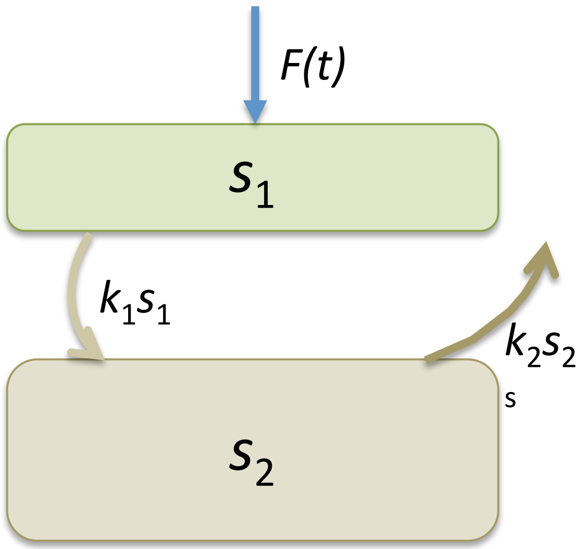
  - How to analyze the results of the SA…

Challenges of multiple data stream assimilation → Toy model examples

Two toy models:

Simple C cycle model (2 pools)

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



Test case	Step 1	Step 2	Parameter error covariance terms propagated in step 2?
Separate			
1a	s_1	–	–
1b	s_2	–	–
Step-wise			
2a	s_1	s_2	yes
2b	s_1	s_2	no
2c	s_2	s_1	yes
2d	s_2	s_1	no
Simultaneous			
3a	s_1 ands_2	–	–
3b	s_1 and only 1 obs for 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.