EGU2020-10722
https://doi.org/10.5194/egusphere-egu2020-10722
EGU General Assembly 2020

Optimal model complexity for terrestrial carbon cycle prediction using data assimilation

Caroline A. Famiglietti[1], T. Luke Smallman[2,3], Sophie Flack-Prain[2], Rong Ge[2], Victoria Meyer[5], Nicholas C. Parazoo[4], Gregory R. Quetin[1], Andrew Revill[2], Stephanie G. Stettz[4], Yan Yang[4], Yuan Zhao[2], Penghui Zhu[2], A. Anthony Bloom[4], Mathew Williams[2,3], and Alexandra G. Konings[1]

[1]Department of Earth System Science, Stanford University, Stanford, USA
[2]School of GeoSciences, University of Edinburgh, Edinburgh, UK
[3]National Centre for Earth Observation, University of Edinburgh, Edinburgh, UK
[4]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, USA
[5]School of the Art Institute of Chicago, Chicago, USA

The future role of the terrestrial biosphere in the global carbon cycle is highly uncertain. Modeling and predicting the terrestrial net carbon balance is difficult due to the numerous processes driving variability of gross fluxes. Many approaches to reducing this model uncertainty have focused on model structure, namely by adding additional processes (*e.g.,* nutrient dynamics or vegetation demography) and thus increasing complexity. While these developments seek to achieve greater structural realism by mirroring the complexity of the natural world, they often rely, by necessity, on poorly-determined or over-generalized parameters. Furthermore, increased structural complexity may increase the risk that parameters with compensating errors are found during model development, thereby reducing model accuracy in prediction. It is not clear whether or to what extent carbon cycle predictability scales with structural complexity, or whether an intermediate, optimum level of complexity exists that may balance the costs of a low (more biased) or high (more variant) complexity model. Here, we explore and define the relationship between carbon cycle model complexity and prediction accuracy. To do so, we leverage the CARbon Data MOdel fraMework (CARDAMOM), a Bayesian data assimilation system that retrieves terrestrial carbon cycle variables (including pools, fluxes, and static parameters) by combining multiple observations with a relatively simple ecosystem carbon balance model. CARDAMOM includes several ecological and dynamical constraints that can prevent ecologically unrealistic parameter combinations and reduce compensating errors between parameters (also known as equifinality). Furthermore, it is a flexible framework to which process representations, parameters, and constraints can easily be added and removed. We used CARDAMOM to develop a suite of model versions spanning a broad range of structural complexity, including the number of carbon pools and the allocation of carbon to the canopy. We assessed a model's complexity based on its inherent dimensionality, determined via a principal component analysis that reduces the parameter space to its principal components. We tested and compared the training and forecast accuracies of net ecosystem exchange predictions using 14 increasingly complex versions of CARDAMOM, each with 48 different experimental designs (*i.e.,* combinations of data constraints

and error assumptions) at 5 globally-distributed eddy covariance sites representing a range of biomes and vegetation types across a total of 70 site-years. We also compared the model performance values against a range of machine learning approaches, which are assumed to represent the limit of infinite model complexity due to their large number of underlying parameters. In this presentation, we use this population to demonstrate and explain patterns in the mapping of model complexity and other assimilation choices to prediction accuracy, offering theoretical and empirical insights into the optimal structure of a carbon cycle model.