Supplementary Material for:

Optimal model complexity for terrestrial carbon cycle prediction

Caroline A. Famiglietti¹,*, T. Luke Smallman², Paul A. Levine³, Sophie Flack-Prain², Gregory R. Quetin¹, Victoria Meyer⁴, Nicholas C. Parazoo³, Stephanie G. Stettz³, Yan Yang³, Damien Bonal⁵, A. Anthony Bloom³, Mathew Williams², and Alexandra G. Konings¹

¹Department of Earth System Science, Stanford University, Stanford, USA
²School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, Edinburgh, UK
³Jet Propulsion Laboratory, California Institute of Technology, Pasadena, USA
⁴School of the Art Institute of Chicago, Chicago, USA
⁵Université de Lorraine, AgroParisTech, INRAE, UMR Silva, 54000 Nancy, France

Correspondence to: Caroline A. Famiglietti (cfamigli@stanford.edu)

Contents of this file:

Figures S1-S6
**Fig. S1:** Effect of observational error scalar on effective complexity. Models are ordered from fewest (S1) to greatest (G4) number of parameters.
Fig. S2: Effects of EDCs on effective complexity. Models are ordered from fewest (S1) to greatest (G4) number of parameters.
**Fig. S3:** Complexity–skill relationship for LAI predictions. (a) All model runs included in the experiment; (b) all model runs for which data was assimilated. Dark gray shading spans the 25th to 75th percentile of runs; light gray shading spans 5th to 95th percentile; blue points are medians of complexity bins. Average forecast skill is computed using the histogram intersection metric.
**Fig. S4:** Complexity–skill relationship for NEE predictions, where a bootstrapping procedure has been performed to equalize number of runs within each complexity bin. Error bars represent one standard deviation of skill across the different bootstrap combinations.
**Fig. S5:** Complexity–skill relationship for NEE predictions, split by observational error scalar (title of each subplot). Dark gray shading spans the 25th to 75th percentile of runs; light gray shading spans 5th to 95th percentile; blue points are medians of complexity bins. Average forecast skill is computed using the histogram intersection metric.
Fig. S6: Complexity–skill relationship for LAI predictions, split by assimilated data subset (title of each subplot). Dark gray shading spans the 25th to 75th percentile of runs; light gray shading spans 5th to 95th percentile; blue points are medians of complexity bins. Average forecast skill is computed using the histogram intersection metric.