Author response to interactive comments on "Optimal model complexity for terrestrial carbon cycle prediction"

Reviewer: Famiglietti and colleagues explored relationship between model complexity and forecast skill either with or without assimilated data using a data assimilation system. The authors found that without assimilated data, a complex model has a poorer forecast skill than a simple model; with assimilated data, the opposite is true. The findings make sense and highlight the importance of using data to inform model before forecasting. The manuscript is very interesting and well written. I have only a few minor concerns about the manuscript below.

We thank the reviewer for their feedback. We have responded to each minor concern below (author response shown in blue).

Reviewer: L230-232: Will there be any difference in key results and conclusion obtained between using the histogram interaction and using the more familiar metrics?

Our key result—the decline in performance attributable to the most complex models under extreme parametric uncertainty scenarios—is preserved across metrics, as shown on the next page (Figure 1) for the normalized root-mean-square error (RMSE; note that larger values correspond to poorer performance) and coefficient of determination (R^2).

We selected the histogram intersection for use in the manuscript because it accounts for prediction accuracy along with prediction and observational uncertainties. Note that because the RMSE and R^2 metrics account for only the first, they are less sensitive to the effects of the different factorial combinations (Table 3, main text) on model parameterization across the effective complexity axis and are therefore less interpretable. We chose not to report these results in the manuscript because they only provide an assessment of individual model skill, and do not provide an integrated assessment of both prediction accuracy and uncertainty.

For clarity, however, we have added the following text to section 3.2: "The decline in performance attributable to the most extreme effective complexity scenarios is also preserved across RMSE and R^2 metrics (not shown; further comparison between different metrics is beyond the scope of this paper)."

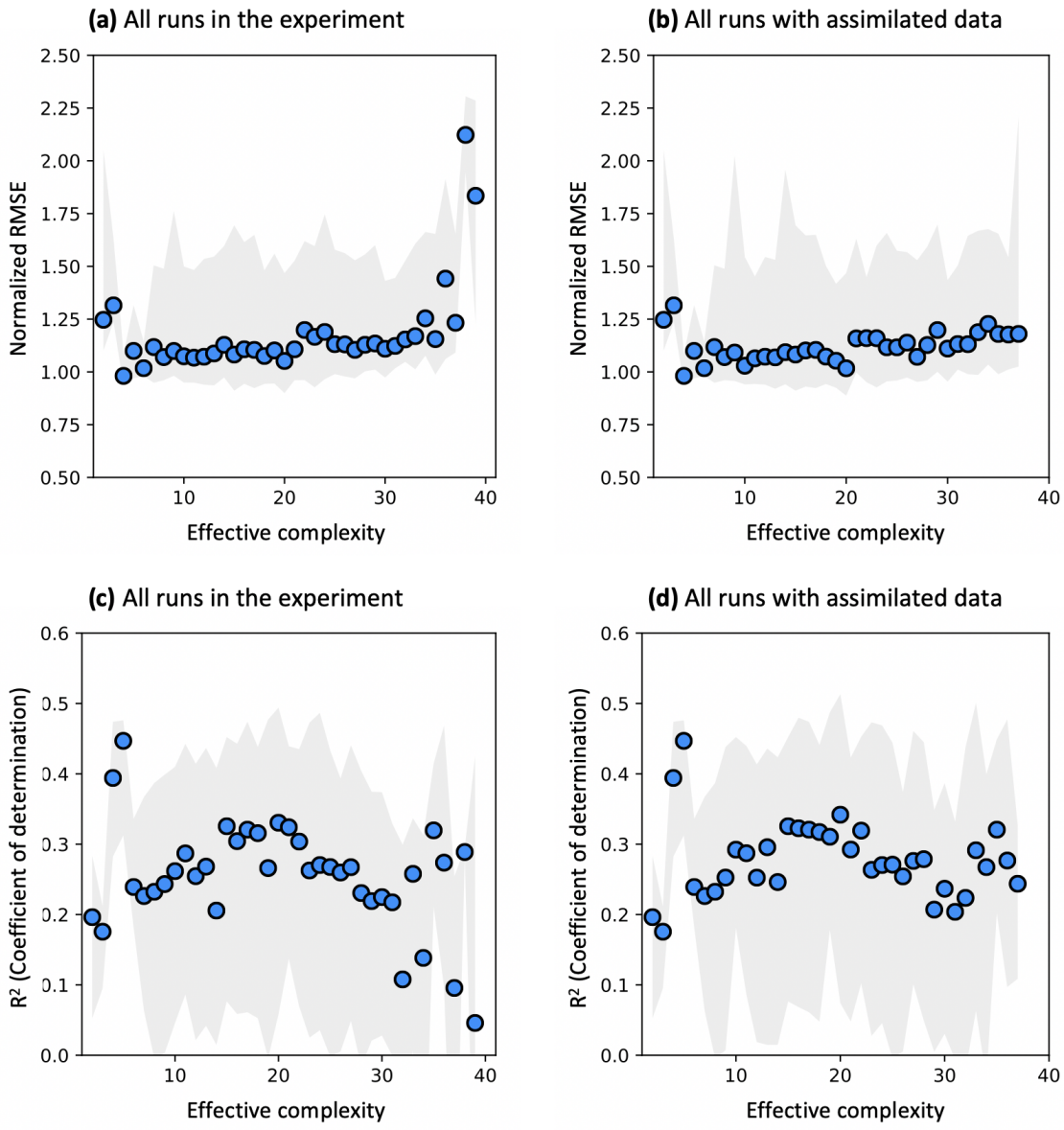

Figure 1: Comparison of model performance across effective complexity axis for RMSE (top row) and R^2 (bottom row) metrics. Left column (a, c) shows all runs included in the experiment; right column (b, d) shows only the subset of runs for which data were assimilated.

Reviewer: L233-239: n value (number of bins) used is?

We have specified the number of bins (n = 50) in the indicated lines.

Reviewer: L405: "assimilate diverse data types" operates blindly. Some datasets are more useful to constrain a specific variable than other datasets. We can do better than just "diverse".

We agree with the reviewer and have amended the indicated line as follows: "assimilate *well-characterized, repeat-observation* datasets".

Reviewer: Table 1. Explain the meanings of the IDs (e.g., C groups, S groups). Why the sub-models ordered in the current way in the Table?

Models are ordered in Table 1 alphabetically by model ID. Models are grouped according to common characteristics, as follows: C models all share the Combined Deciduous Evergreen Analytical (CDEA or CDEA+) phenology sub-model; G models use the Growing Season Index (GSI) phenology sub-model; E models use the evergreen (constant allocation) phenology sub-model; and S models are simple, reduced-complexity variants of other models.

We have added this description to the caption of Table 1.

Reviewer: Fig. 5a: "(a) All runs", do you mean all runs without assimilating data? You may make it clearer.

"All runs" refers to all runs included in the experiment (that is, both with and without assimilating data). We have amended the subplot's title to read "All runs in the experiment" and the figure caption to read "(a) all model runs in the experiment and (b) the subset of runs in panel (a) for which data were assimilated."

Reviewer: Fig. 6: arrangement of the panels are not in a good logic to me. Probably as (a) None ... (f) NEE, LAI, biomass.

The reviewer brings up a worthwhile point. We have rearranged the subplots in Figure 6 so that they are ordered from strongest (NEE, LAI, biomass) to weakest (None) assimilated data constraint, which aligns with our interpretation in the text. The figure caption has also been amended to include the line "Ordering of subplots reflects strongest (a) to weakest (f) data constraint." For consistency, we have made the same changes to Figure S6.

Reviewer: Description of skill metric and complexity metric as well as the model structure are clear, but deposit the code to produce this manuscript could be more helpful to others to use the approaches here.

We agree with the reviewer and have made our analysis code publicly available via Github. We have included this statement and URL (github.com/cfamigli/COMPLEX) in the acknowledgments section of the revised manuscript.