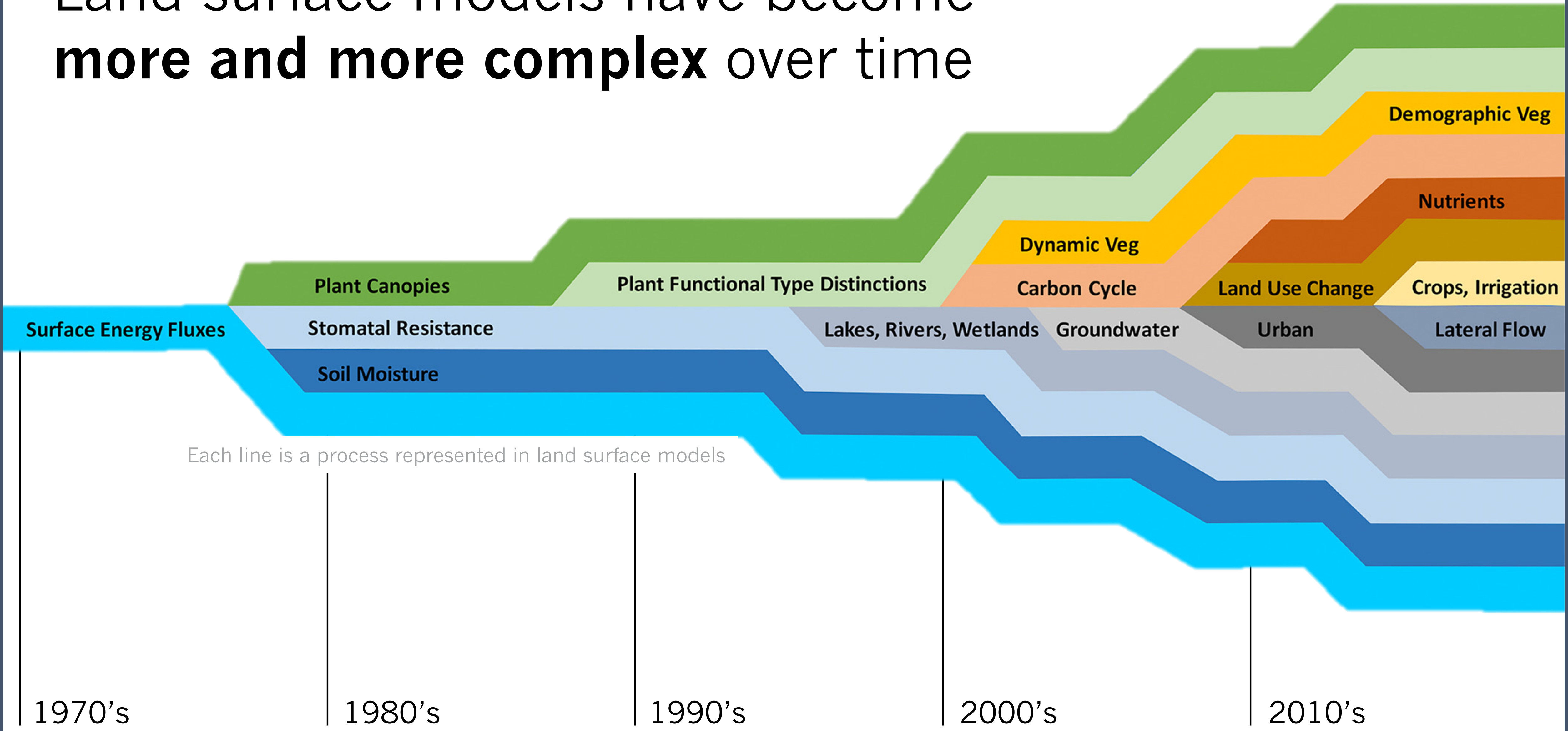
Optimal model complexity for terrestrial carbon cycle prediction using data assimilation

Caroline A. Famiglietti[1], T. Luke Smallman[2,3], Sophie Flack-Prain[2], Rong Ge[2], Paul A. Levine[4], 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

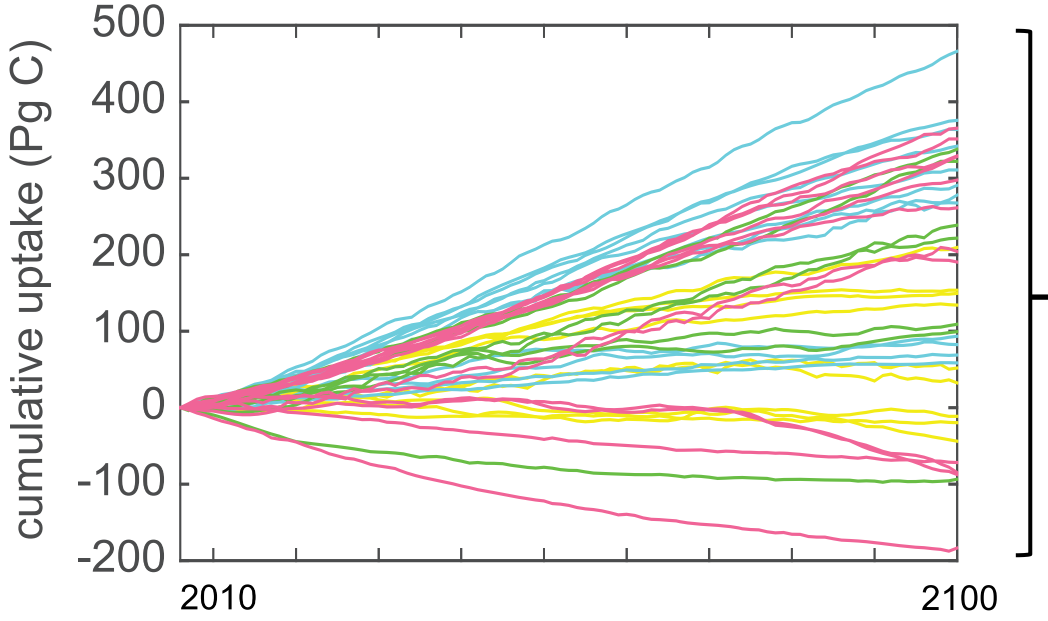
Land surface models have become **more and more complex** over time

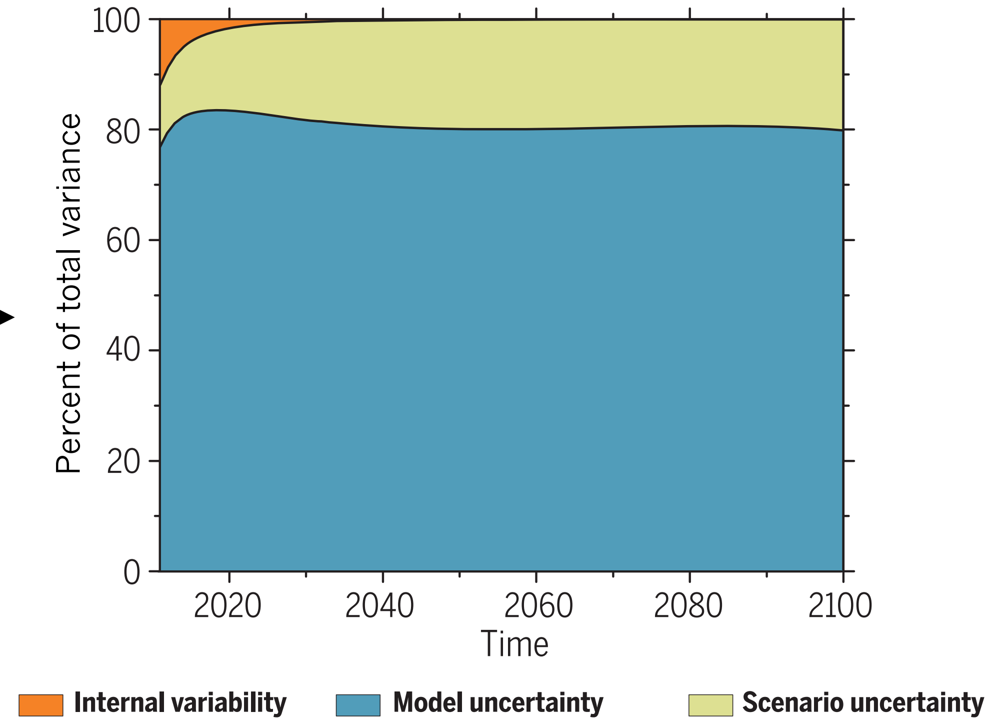


Fisher & Koven, JAMES, 2020

But spread in terrestrial carbon cycle predictions remains and is dominated by **model uncertainty**

Model uncertainty comprises both model **structure** and model **parametrization**.

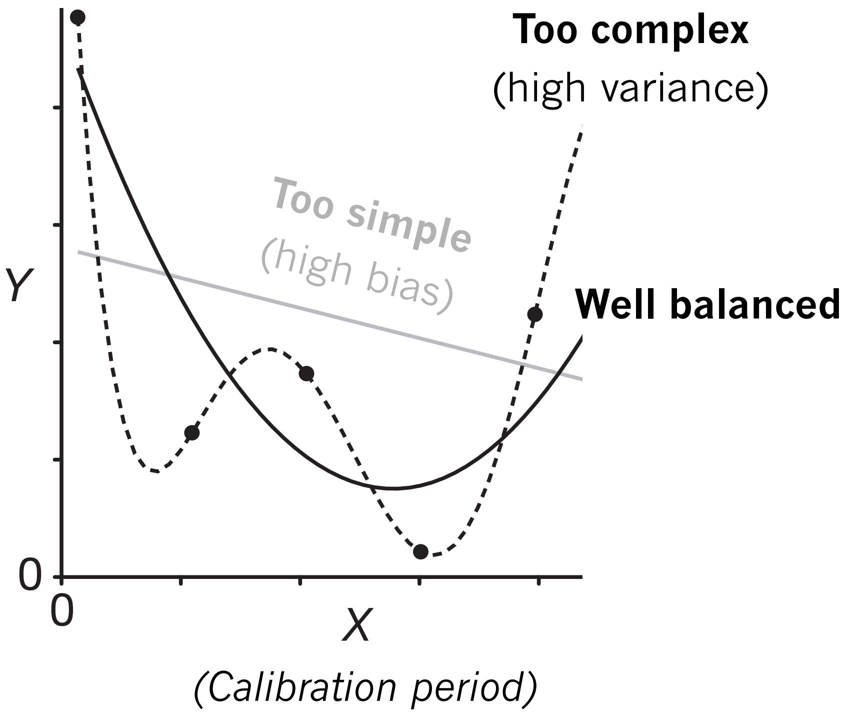




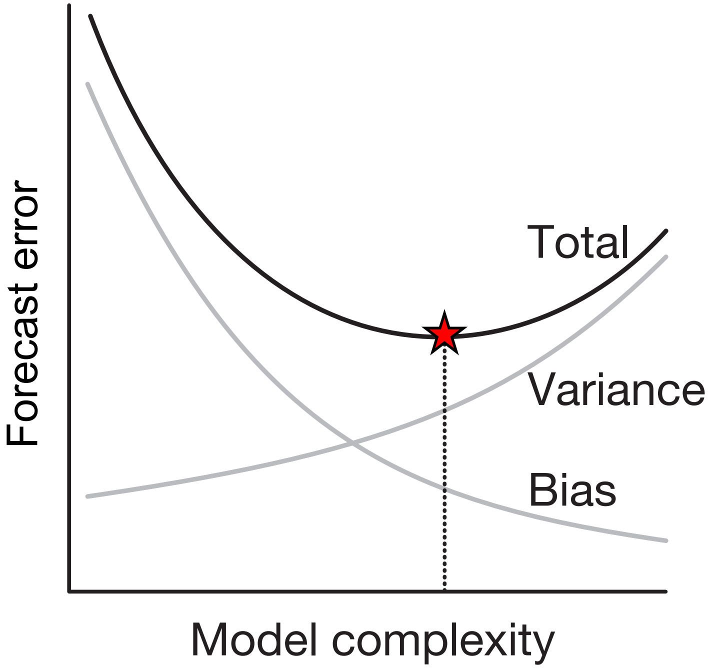
Bonan & Doney, Science, 2018
Lovenduski & Bonan, ERL, 2017

It is not clear if increased carbon cycle model complexity implies **increased predictive skill**

Excessive complexity can lead to **overfitting**



Theoretically, a model that **balances** the tradeoffs of under- and over-fitting can minimize forecast error

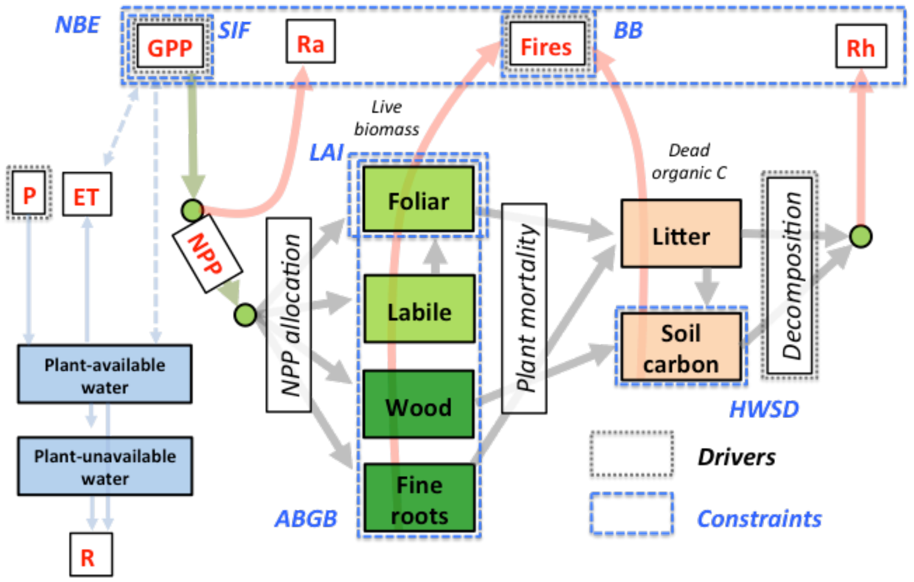


Adapted from Lever et al., Nature Methods, 2016

We seek to understand:

How is carbon cycle **model complexity** related to **forecast skill**?

We used **the CARbon DAta MOdel fraMework (CARDAMOM)**, a Bayesian data assimilation system that allows for flexibility in defining the underlying model structure (variants of the **DALEC model**; *see right*), to explore the relationship between model complexity and predictive skill.



Bloom & Williams, Biogeosciences, 2015
Bloom et al., PNAS, 2016

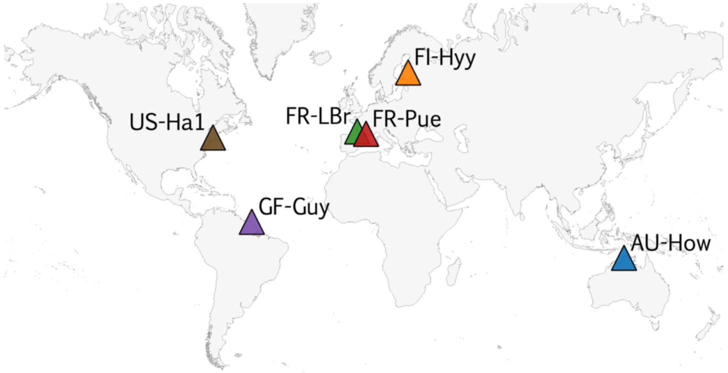
We tested **4032** combinations of **14** models, **6** sites, and **48** data scenarios.

We compared the skill of 14 different CARDAMOM/DALEC **model versions of varying complexity** at forecasting net ecosystem exchange (NEE).

We tested **4032** combinations of **14** models, **6** sites, and **48** data scenarios.

We compared the skill of 14 different CARDAMOM/DALEC **model versions of varying complexity** at forecasting net ecosystem exchange (NEE).

We ran the suite of models at 6 globally distributed FLUXNET sites **across biomes**, each with at least 10 years of data.

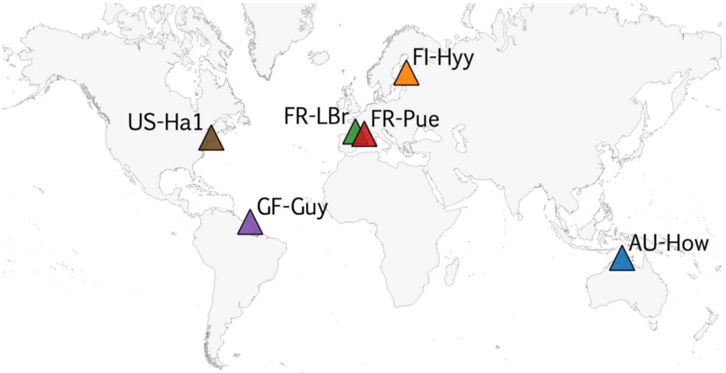


We tested **4032** combinations of **14** models, **6** sites, and **48** data scenarios.

We compared the skill of 14 different CARDAMOM/DALEC **model versions of varying complexity** at forecasting net ecosystem exchange (NEE).

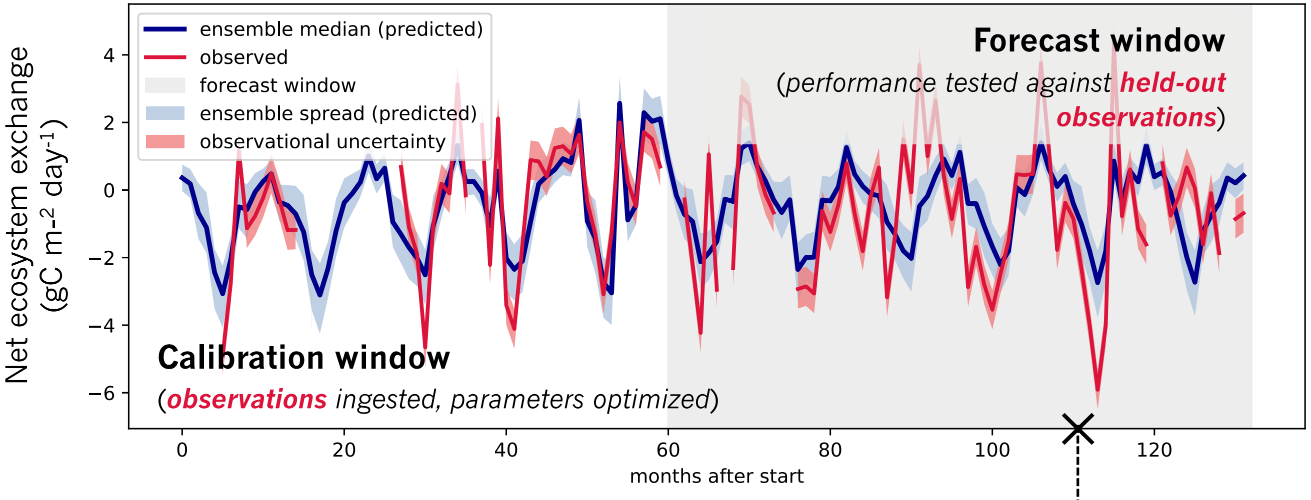
During calibration, we assimilated different combinations of **NEE** *(eddy covariance)*, leaf area index (**LAI**) *(Copernicus)*, and **biomass** *(in situ surveys),* along with additional functional constraints. We also tested different magnitudes of observational uncertainty.

We ran the suite of models at 6 globally distributed FLUXNET sites **across biomes**, each with at least 10 years of data.



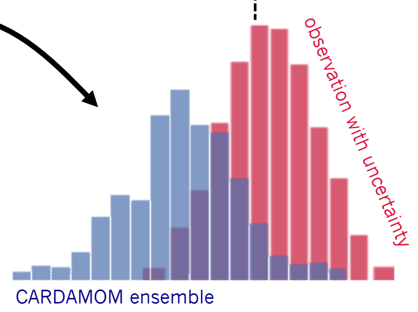
How was forecast skill determined?

We computed the **histogram overlap** between CARDAMOM ensembles and a Gaussian distribution centered at each observation in the forecast window. Unlike an R^2 or RMSE, **this metric explicitly accounts for uncertainty in observations.**



1. Each CARDAMOM run outputs **an ensemble of net ecosystem exchange predictions** based on posterior parameter distributions.

2. We computed the histogram overlap between the **CARDAMOM ensemble** and the **observation with uncertainty** at each timestep.



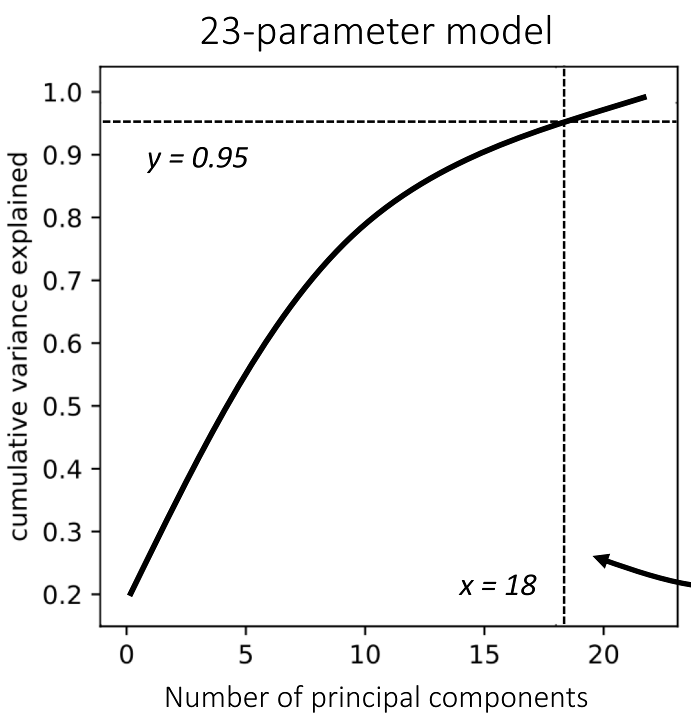
3. The **forecast skill** is the average of the overlaps at every timestep in the forecast window.

How was model complexity determined?

Model structure and assimilated data (**not just number of model parameters**) impact effective model complexity.

Thus, to quantify the complexity of a model-site-experiment combination (run), we performed a **principal component analysis (PCA)** on the parameter space.

We defined the **"inherent dimensionality"** of a given run by the number of principal components at which 95% variance in the parameter set is explained.

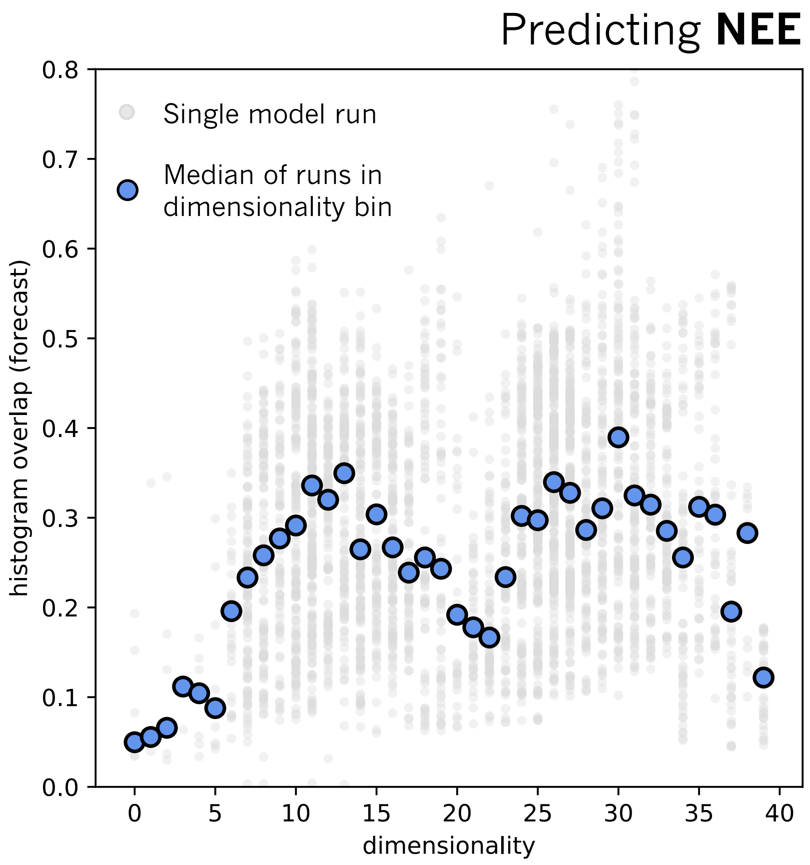


Example: This run (using a 23-parameter model) has an inherent dimensionality of 18.

Across **all site-run combinations**, an **intermediate-complexity model** had the highest forecast skill

Better skill

(greater overlap between predictions and observations)



More complex

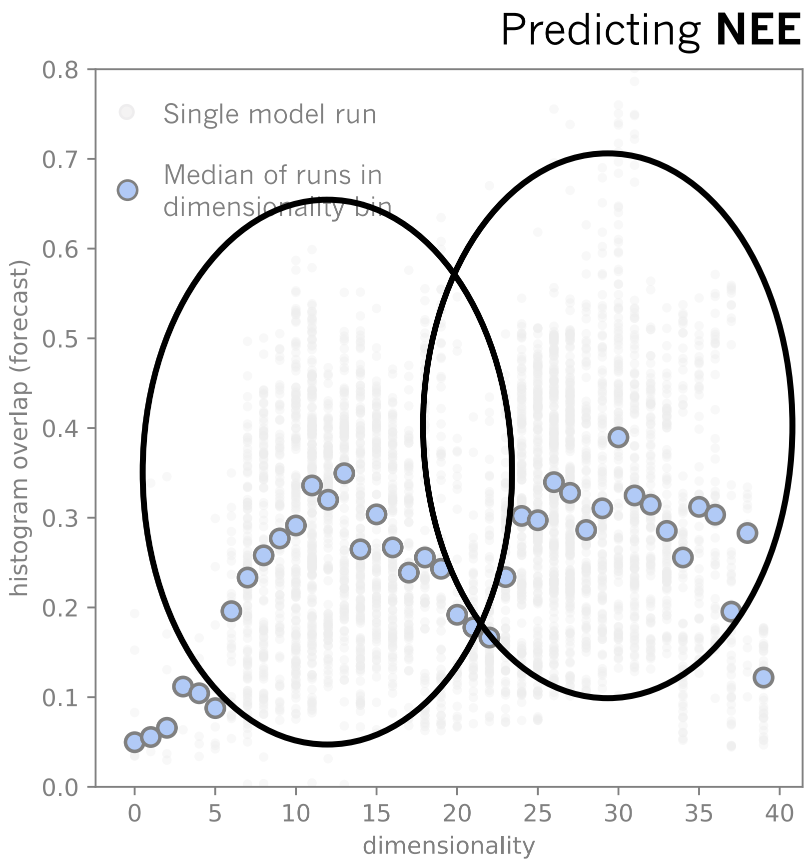
(higher inherent dimensionality)

Results

Across **all site-run combinations**, an **intermediate-complexity model** had the highest forecast skill

Better skill

(greater overlap between predictions and observations)



More complex

(higher inherent dimensionality)

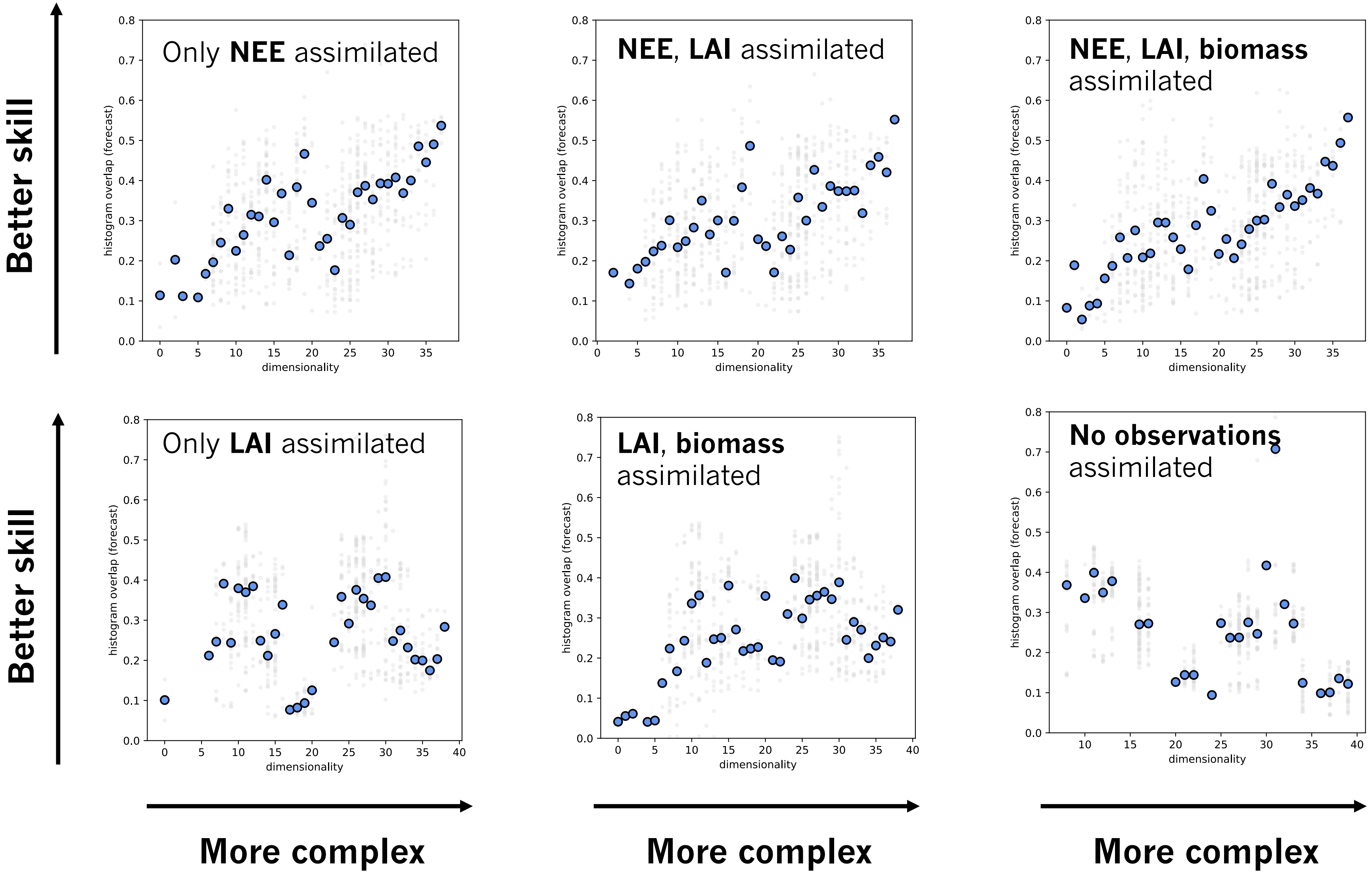
Our models fall into **two populations**.

Each population has an intermediate-complexity optimum.

Results

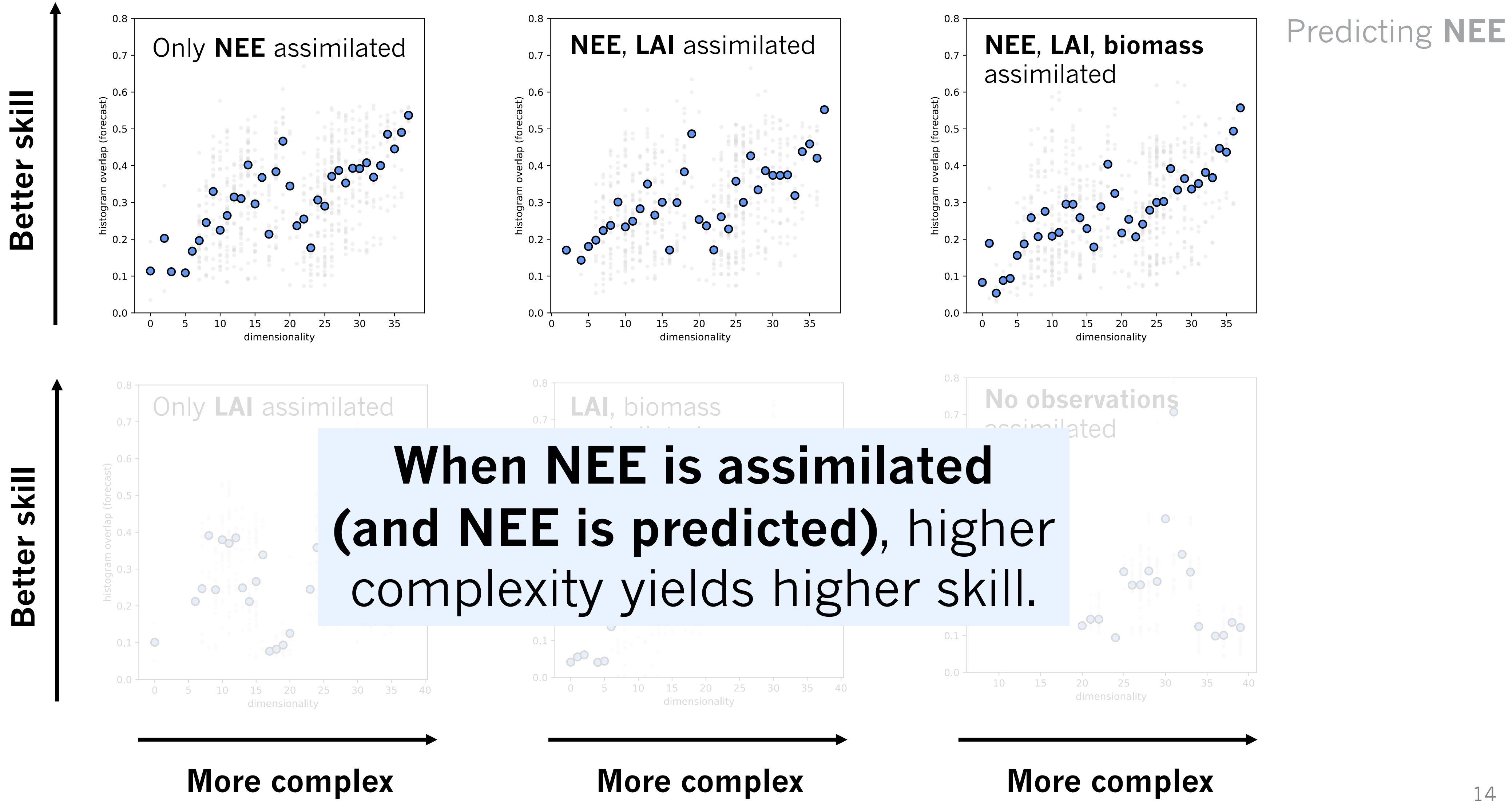
However, overall, optimal complexity is a **function of which type of data are assimilated** *(different subplots below)*

Predicting **NEE**



Results

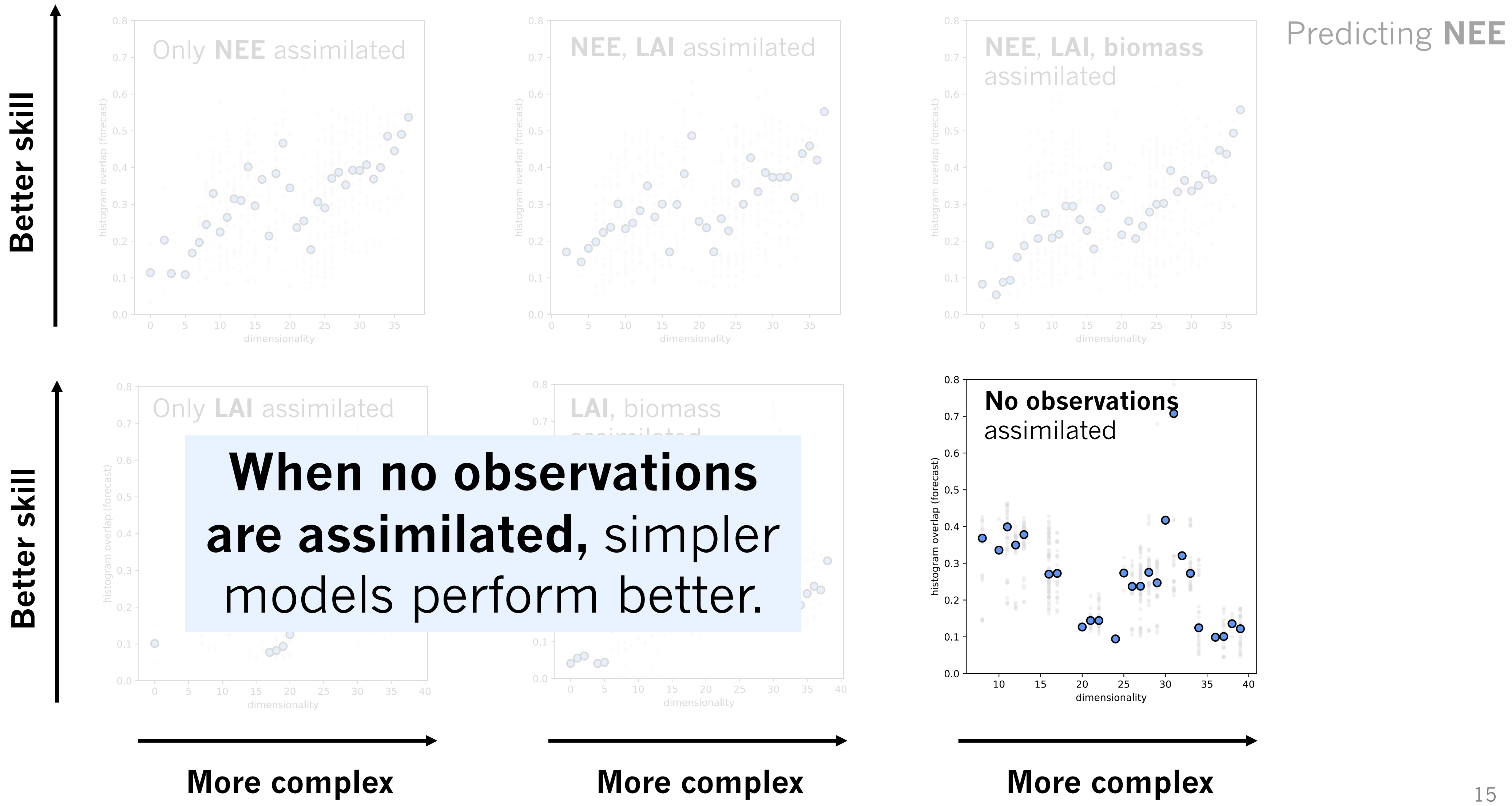
However, overall, optimal complexity is a **function of which type of data are assimilated** *(different subplots below)*



When NEE is assimilated (and NEE is predicted), higher complexity yields higher skill.

Results

However, overall, optimal complexity is a **function of which type of data are assimilated** *(different subplots below)*

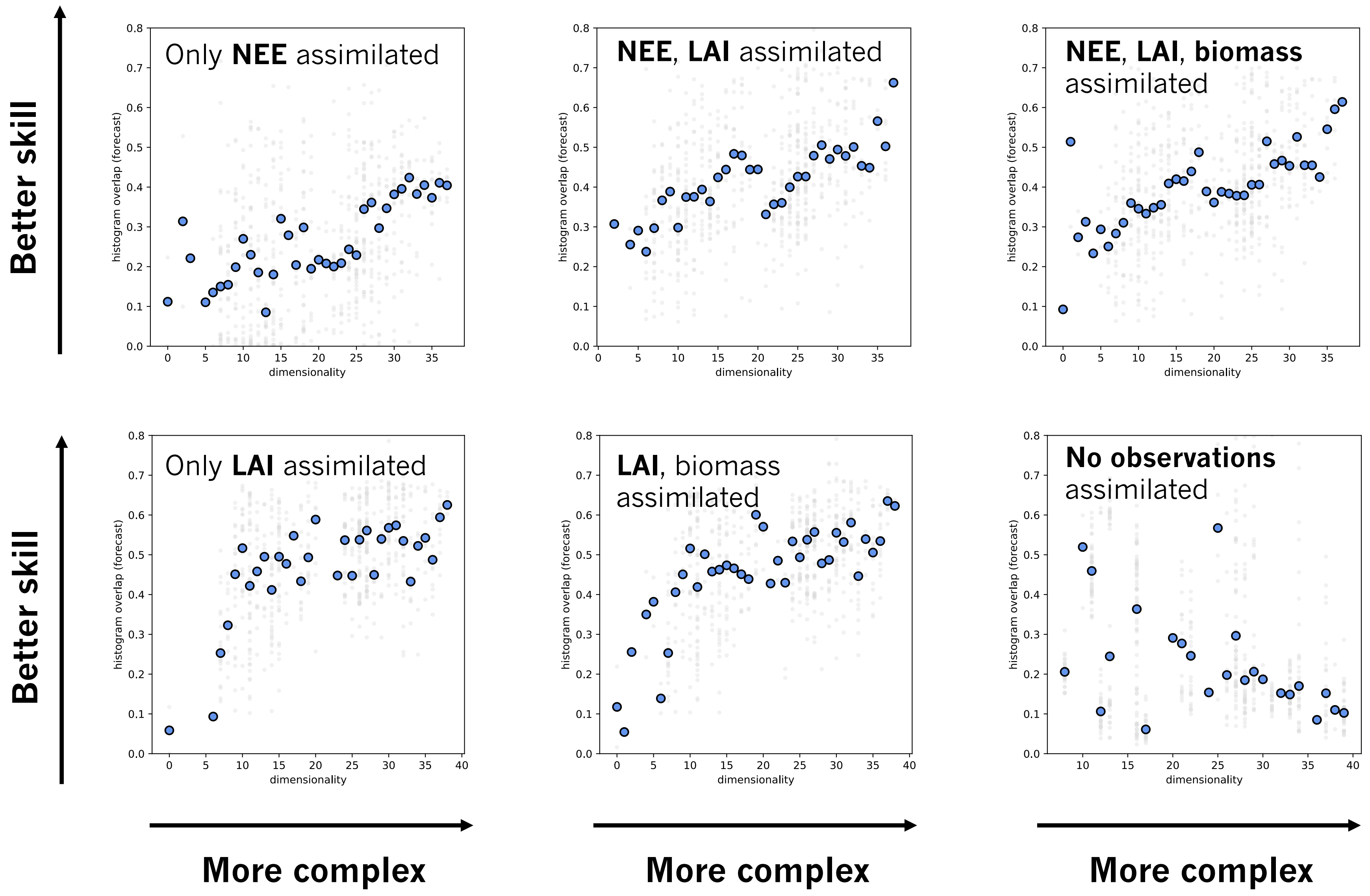


When no observations are assimilated, simpler models perform better.

Results

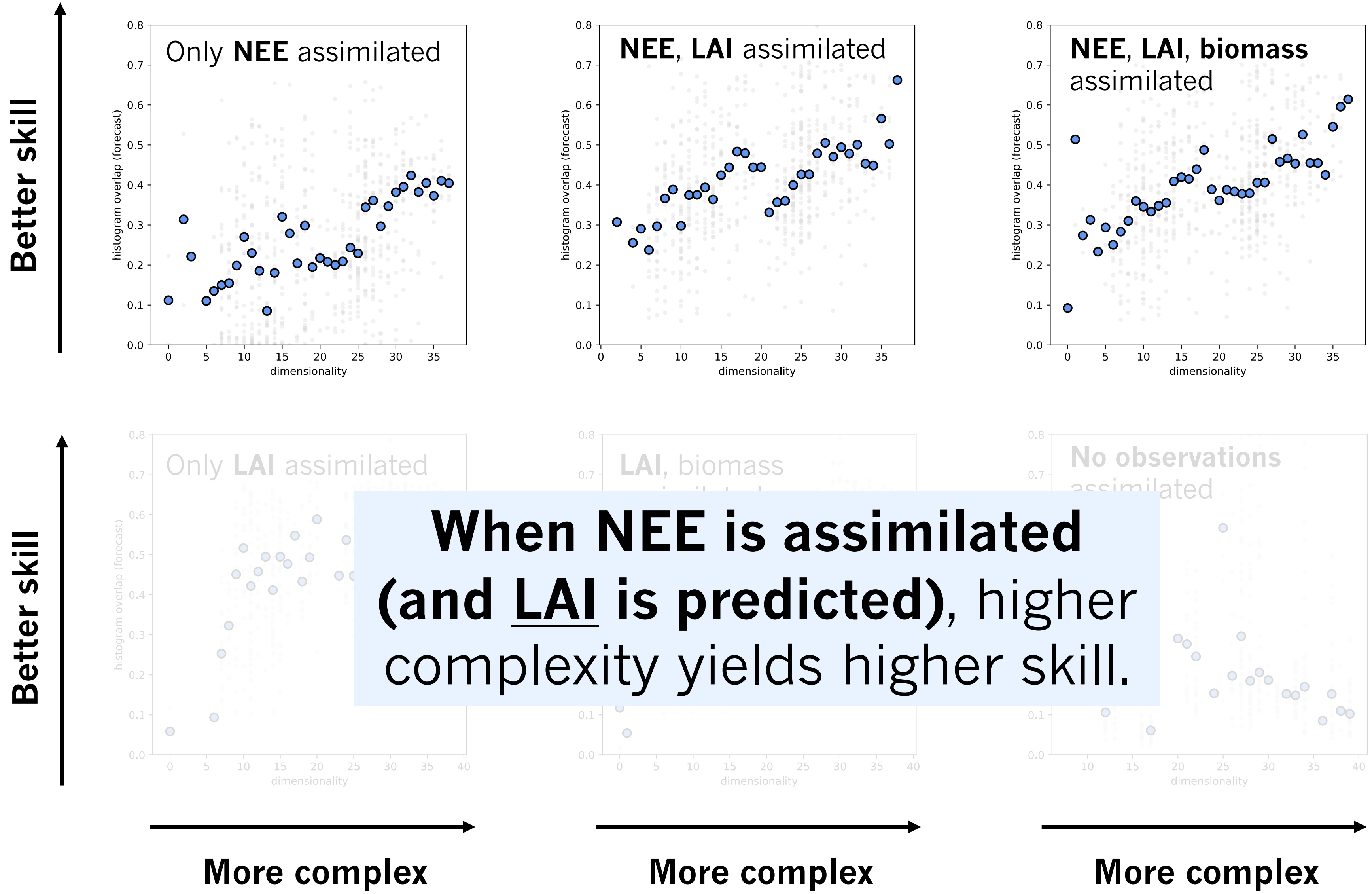
These results remain robust when **predicting LAI instead of NEE**

Predicting **LAI**



These results remain robust when **predicting LAI instead of NEE**

Predicting **LAI**



When NEE is assimilated (and LAI is predicted), higher complexity yields higher skill.

Results

Summary and implications

- **Model complexity matters** for understanding forecast skill.
- When there is not enough information (*e.g.*, data volume and quality) to adequately constrain parameters, **increased complexity can degrade skill.** *(slide 15)*
- However, under specific conditions (*e.g.*, when NEE is assimilated), increased complexity can yield increased forecast skill. *(slide 14)*
- This highlights the importance of **robust model parametrization** for land surface modeling.

Future work

- We will evaluate the relationship between uncertainty metrics (histogram overlap), **precision metrics (R^2, RMSE),** and model complexity.
- We will compare the skill of machine learning models (highly complex) to our suite of process-based carbon cycle models.