Lessons learned from an automated parameter calibration effort in the NASA GISS Earth System Model

Gregory Elsaesser[1,2,3] Marcus van Lier-Walqui [1,2,4], Qingyuan (Yang) Yang[1,5] (presenter)

1. Learning the Earth with Artificial Intelligence and Physics (LEAP) NSF Science and Technology Center (STC), NY, USA

2. NASA Goddard Institute for Space Studies, NY, USA

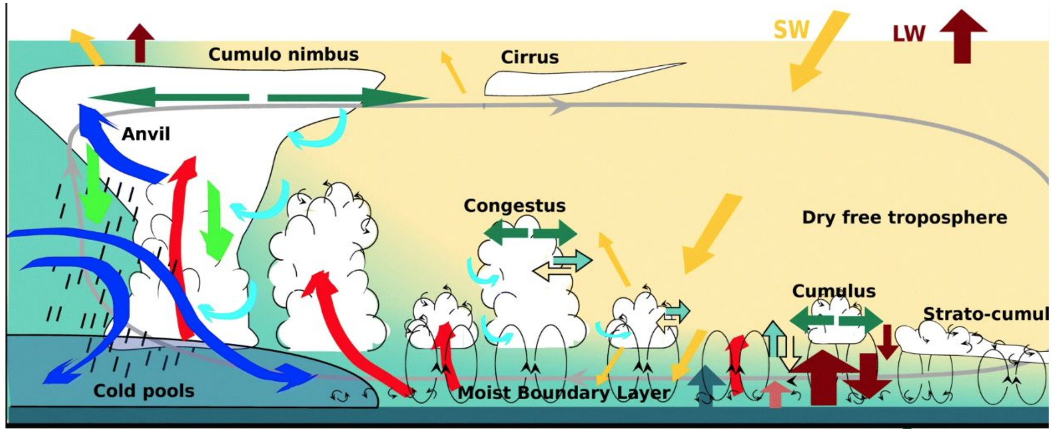
3. Department of Applied Physics and Applied Mathematics, Columbia University, NY, USA

4. Columbia Climate School, Columbia University, NY, USA

5. Department of Earth and Environmental Engineering, Columbia University, NY, USA

MOTIVATION

- ❖ Many processes must be represented in an Earth System Model (ESM) atmosphere. (e.g., figure to right)



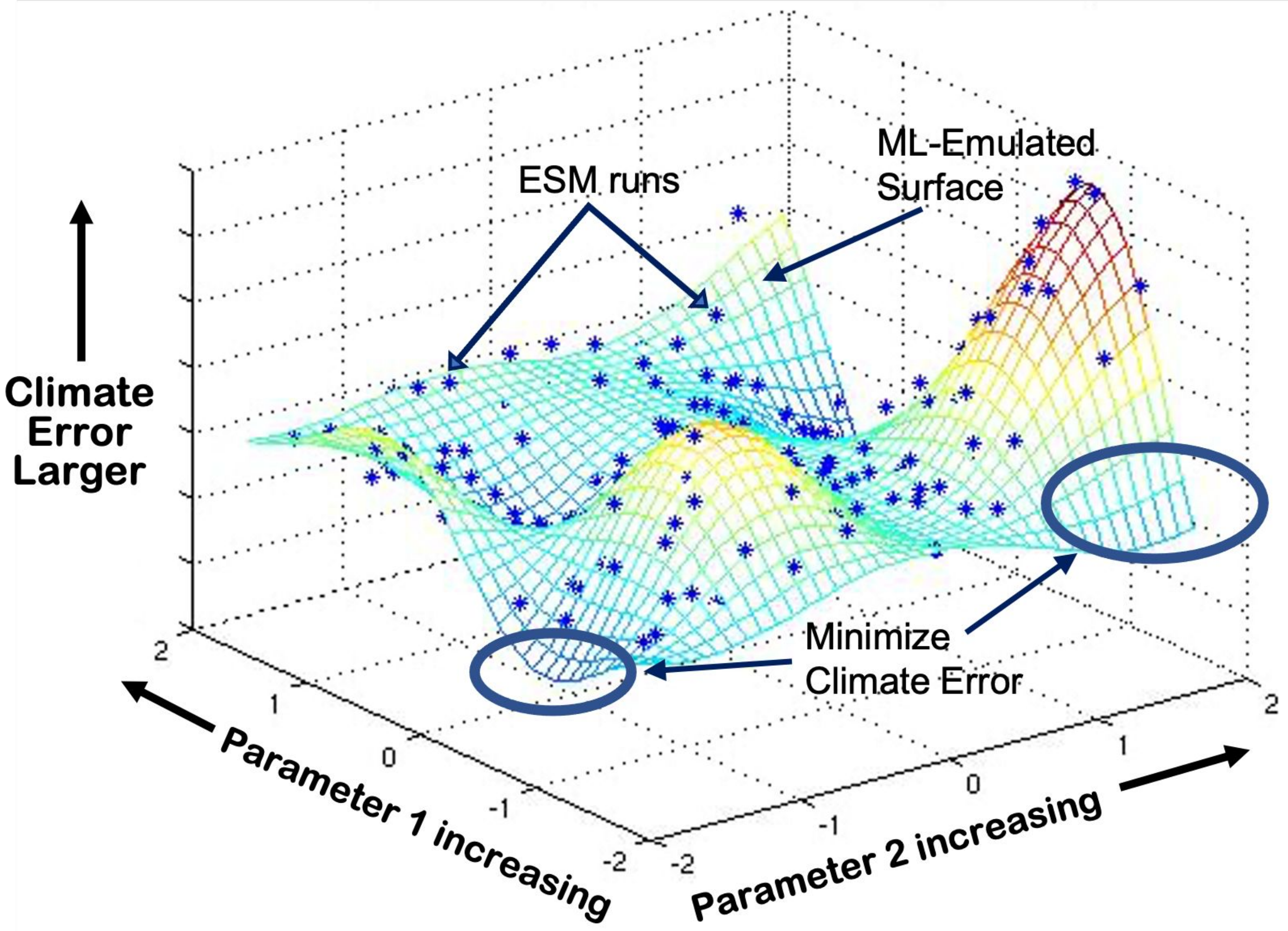
Rio et al. 2019

- ❖ Each process is represented by equations, and each equation has many parameters.

- ❖ Some parameters are known (e.g., Coriolis parameter, acceleration due to gravity); **most are not!** (e.g., parameter for conversion of cloud drops to rain drops, relative humidity threshold in a large grid box for determining if a cloud should form).

- ❖ **How one sets all unknown parameters impacts ESM skill & climate projection.**

The Goal: Determine all combinations of free parameters (100s) that yield Earth System Model (ESM) configurations whose mean states look like Earth's.

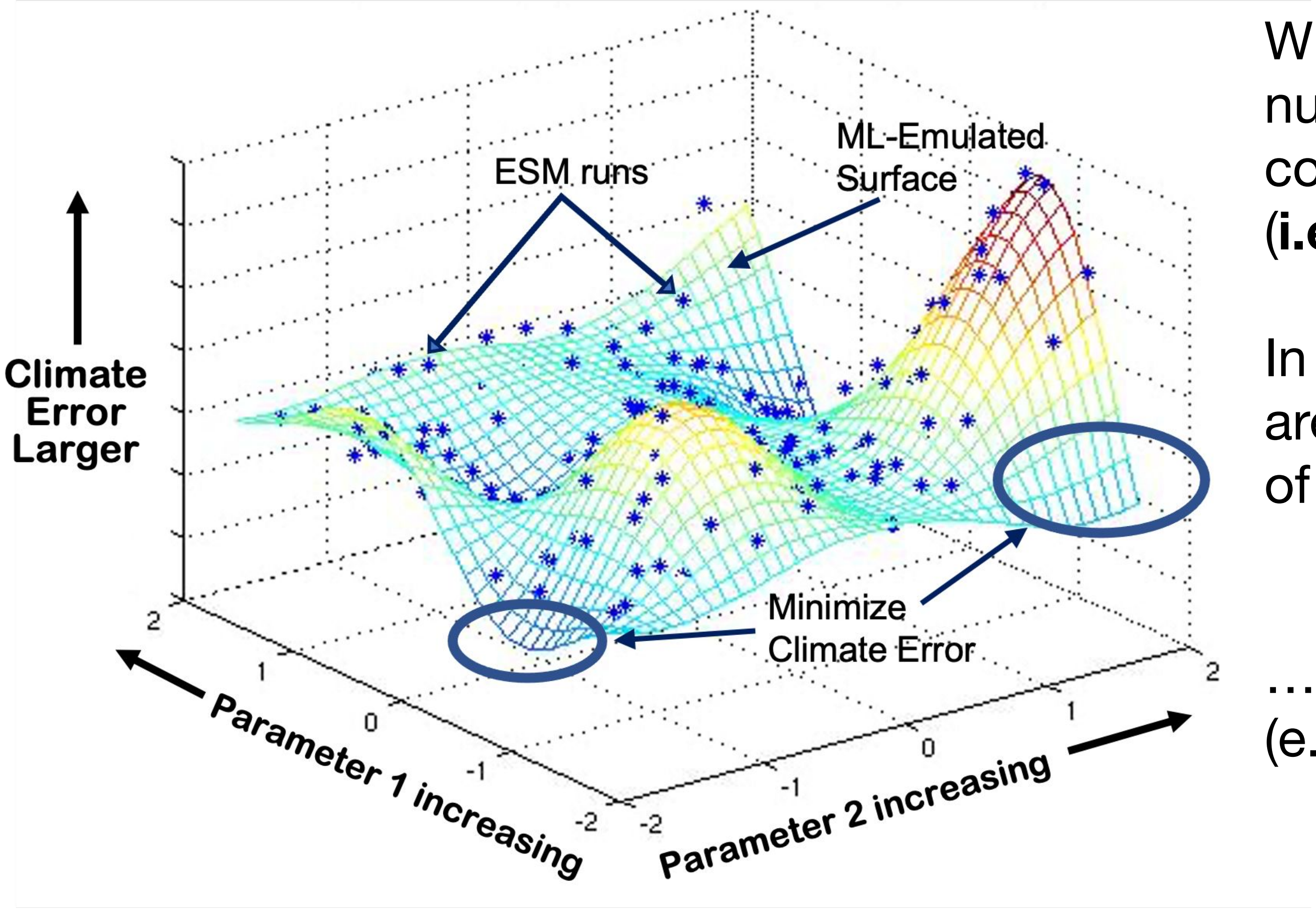


Why? Because then we can run all combinations to see the PDF of climate projections *knowing that individual model mean-states are realistic*.

For a designed model ensemble that is more constrained by the available information, is the spread in projections similar to the CMIP model spread (and can we learn from it)?

In a perturbed parameter ensemble (PPE) that is not constrained by any observations, many members are not Earth-like, which makes it harder to learn realistic spread from the ensemble!

The Goal: Determine all combinations of free parameters (100s) that yield Earth System Model (ESM) configurations whose mean states look like Earth's.



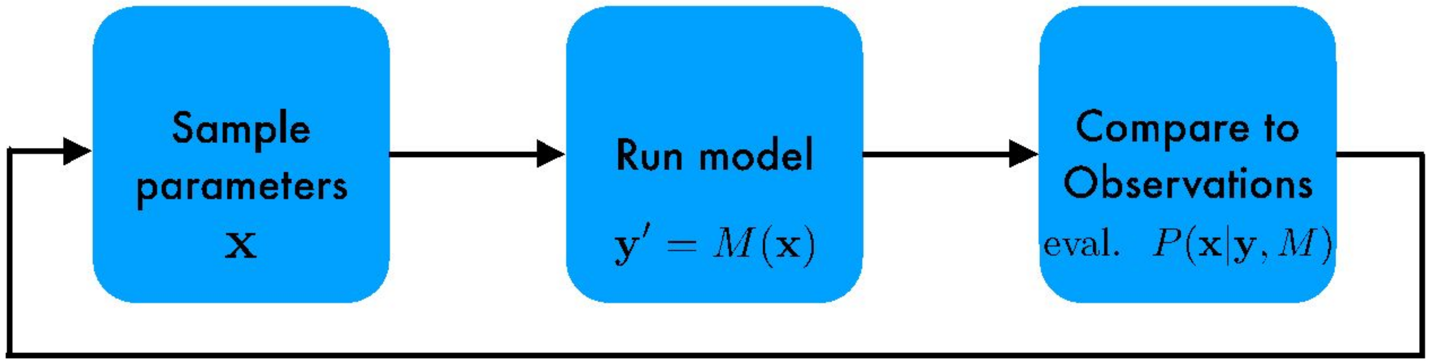
Why might the state space yield a number of diverse parameter combinations giving desired output (**i.e., equifinality**)?

In part, because so many parameters are being explored (i.e., more degrees of freedom).

…and also because of uncertainty (e.g., in observations, in the emulator).

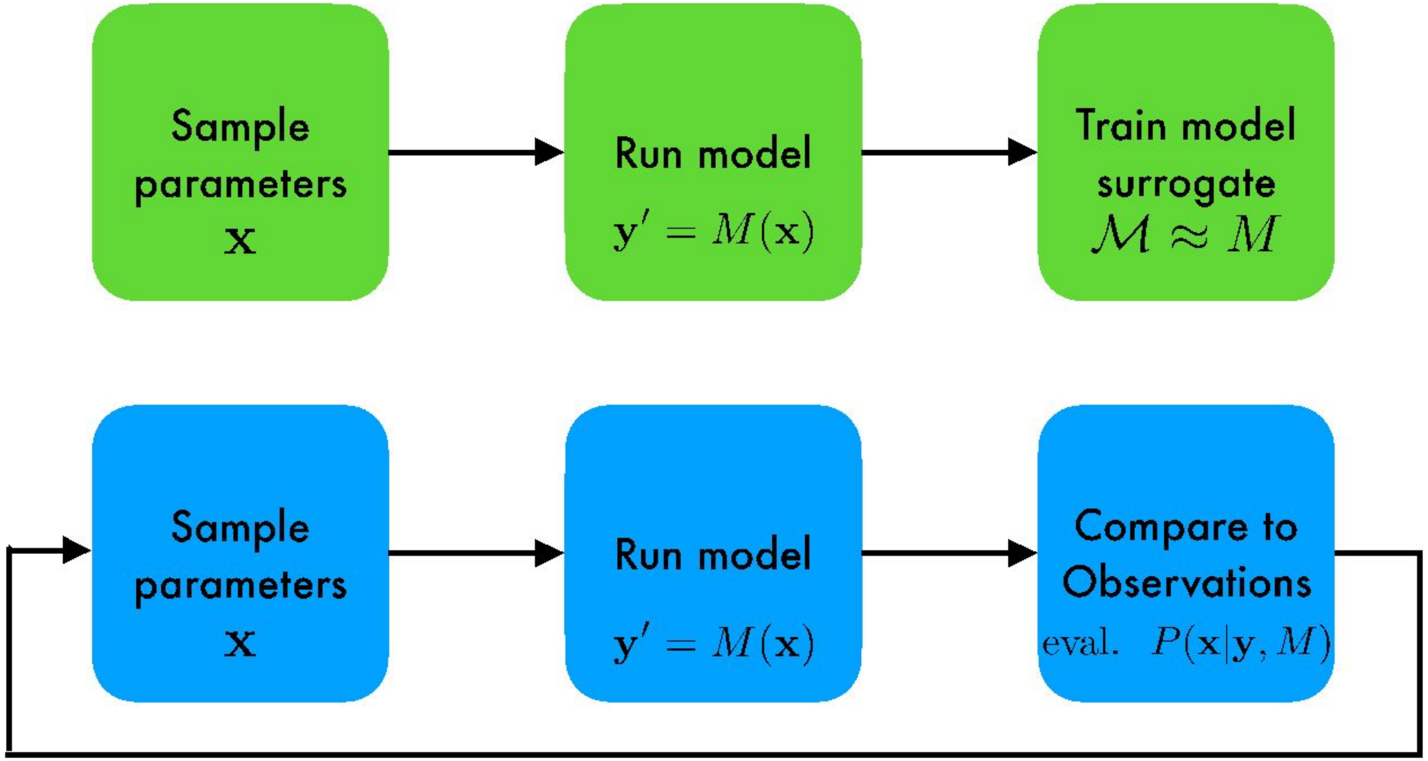
Framework applied to GISS climate model atmosphere

- 45 parameters, 36 observational climatologies to try to match.
- Markov-Chain Monte Carlo (MCMC) can be used to find all parameter combinations, but is prohibitively expensive (requires 10k – 10M sampling iterations)



Framework applied to GISS climate model atmosphere

- We can train a machine learning "emulator" or "model surrogate" to replicate the model's mapping of parameter perturbations to outputs.

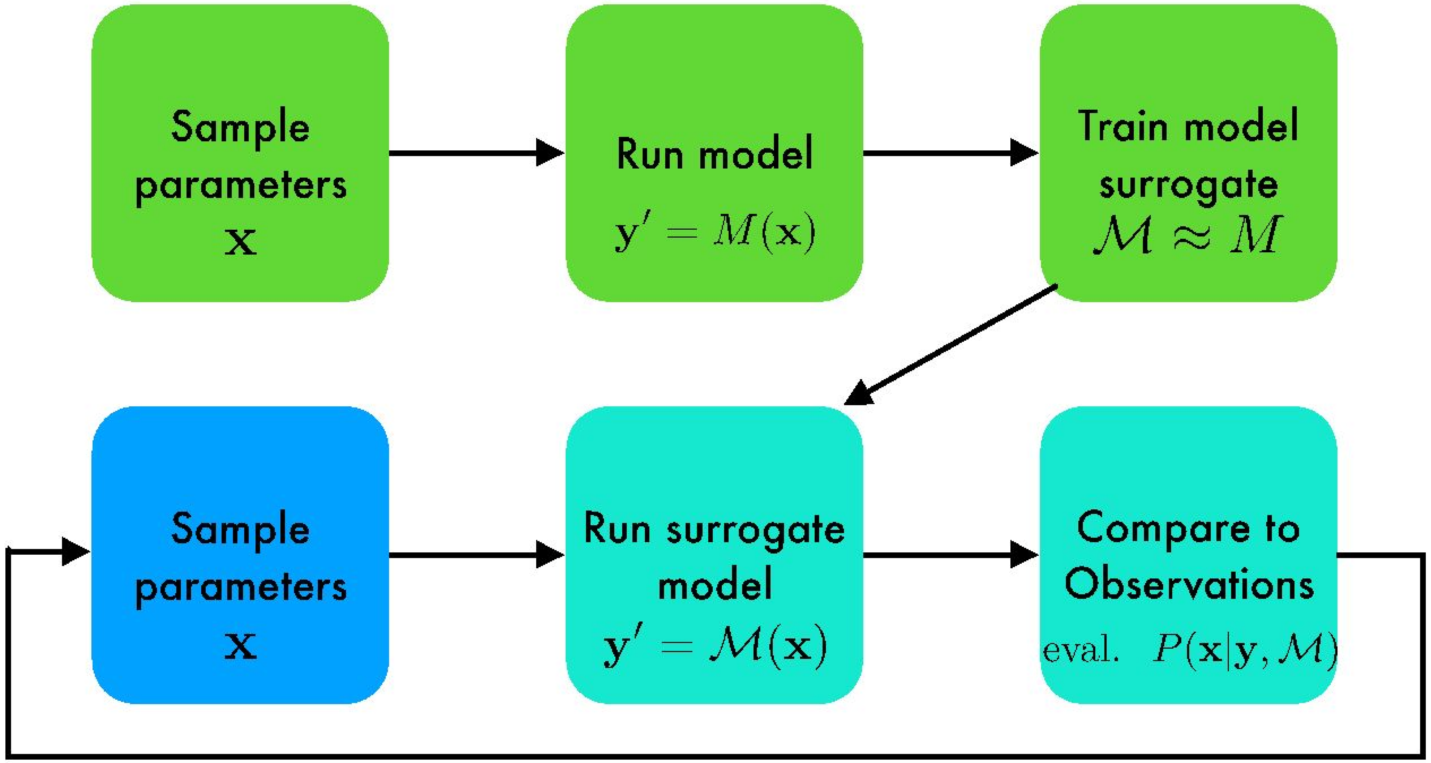


COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

LEAP

Framework applied to GISS climate model atmosphere

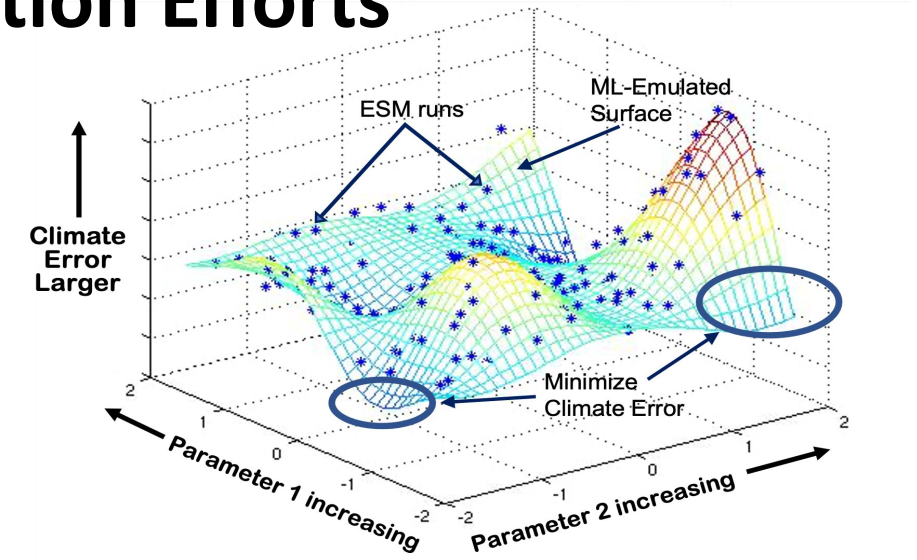
- By replacing model with the machine learning surrogate, we can do MCMC with low computational cost

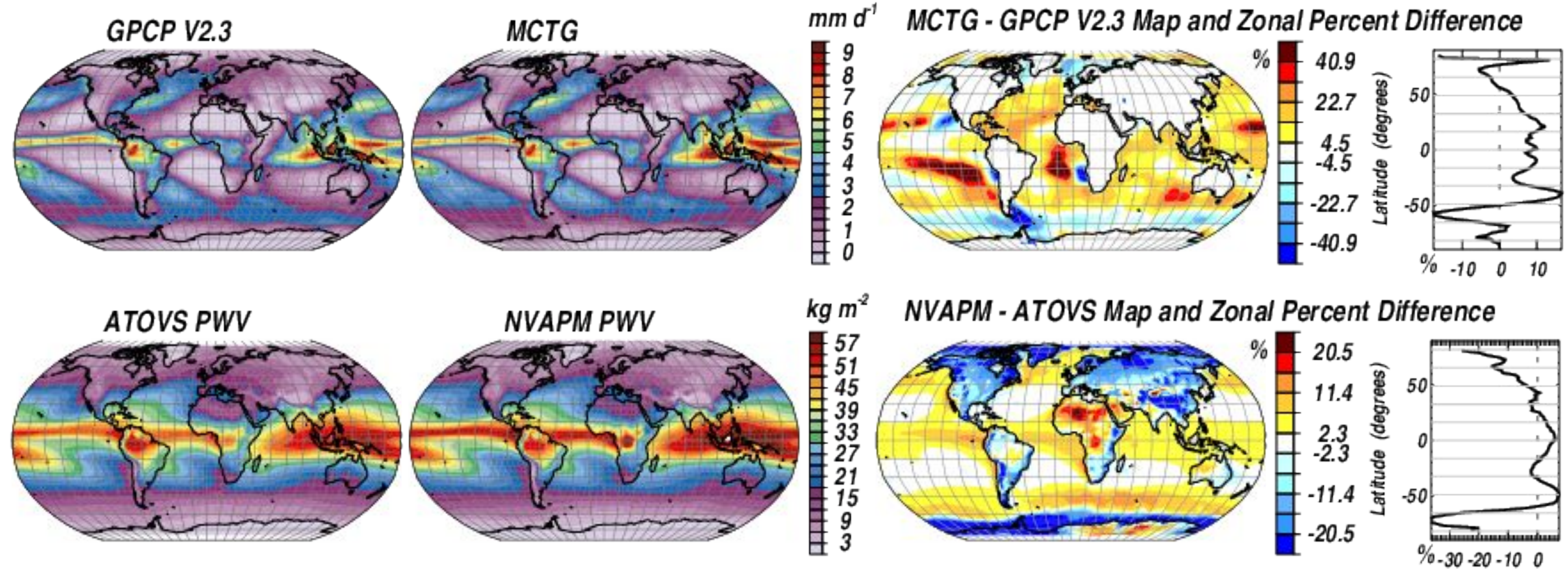


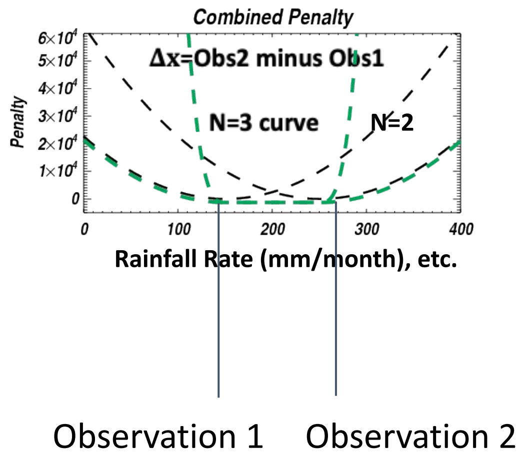
Earth System Model (ESM) Parameter Calibration Efforts

Example: accounting for observational uncertainty in the emulation of climate model error.

Examples below using two surface rainfall and column integrated water vapor products. Note discrepancies, which are accounted for quantitatively in the model penalty (i.e., climate error) functions.

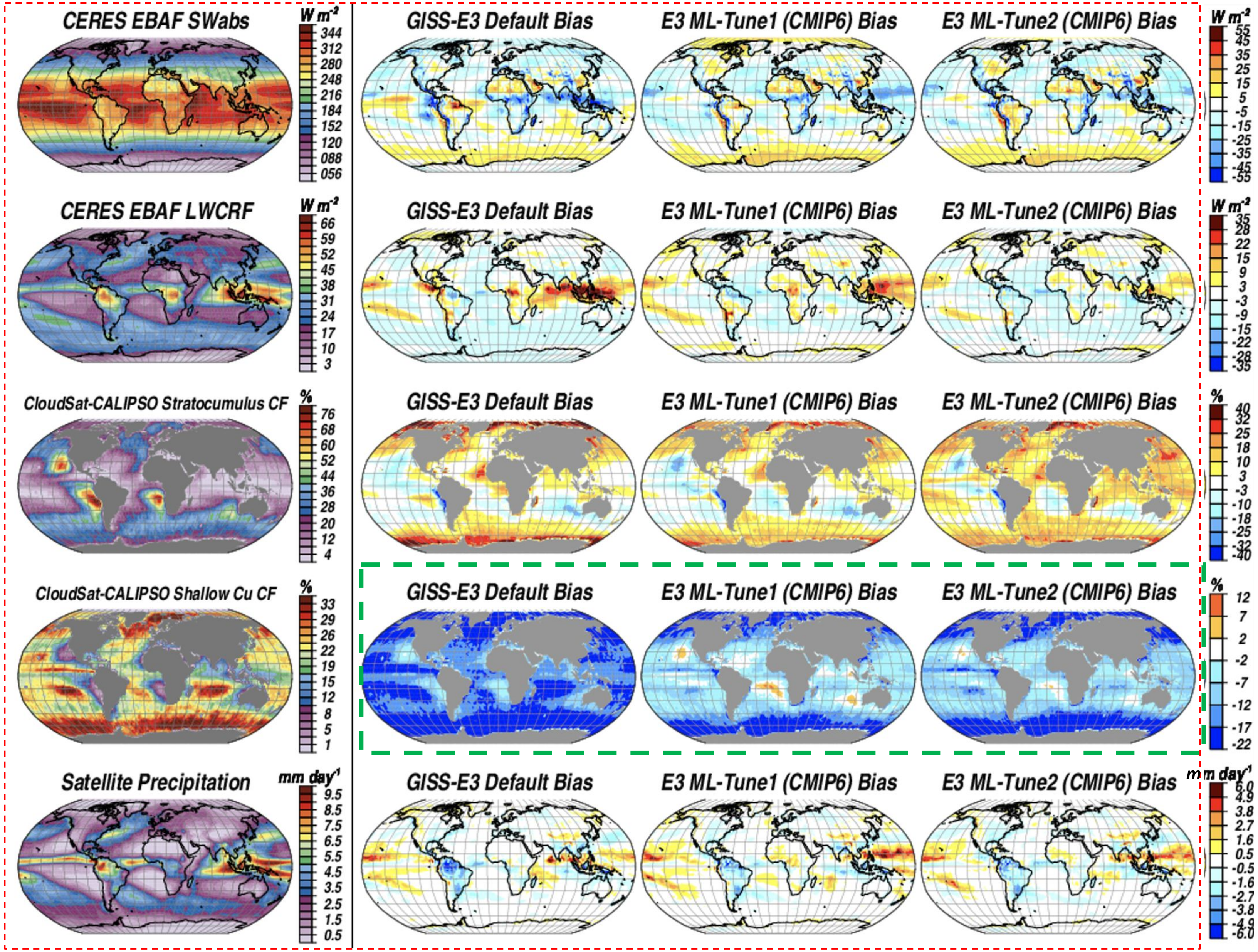






Elsaesser, van-Lier Walqui, Q. Yang, & GISS

Lessons Learned from GISS ESM Atmosphere Parameter Estimation Efforts



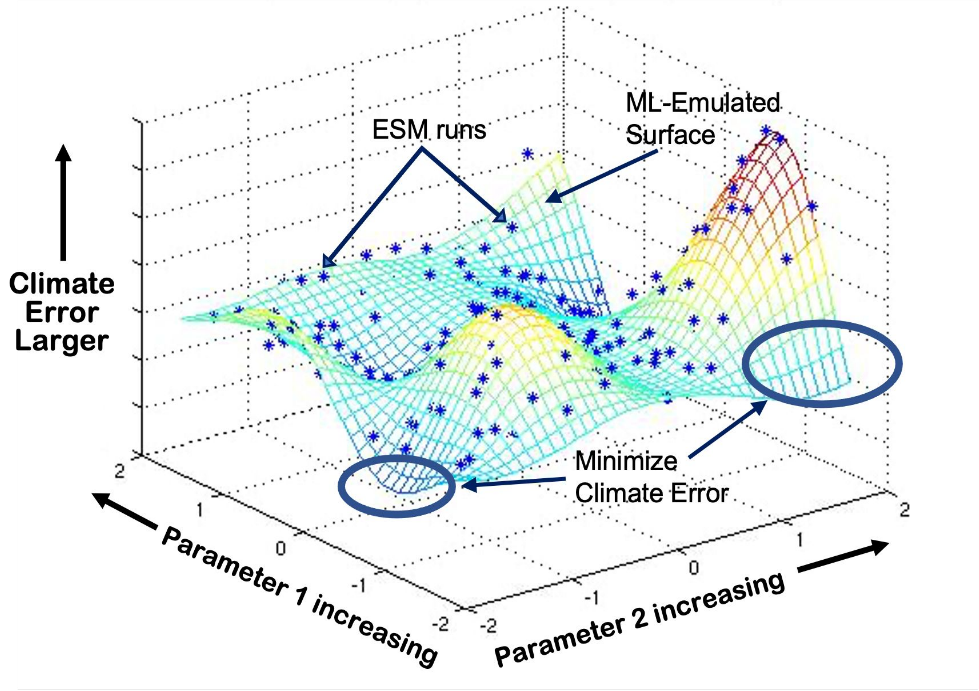
What did ML tuning do relative to improved default physics? First, it provides alternate parameter settings yielding similar mean states.

Finds diversity in parameter combinations that retain the default improvement in clouds (e.g., stratocumulus clouds).

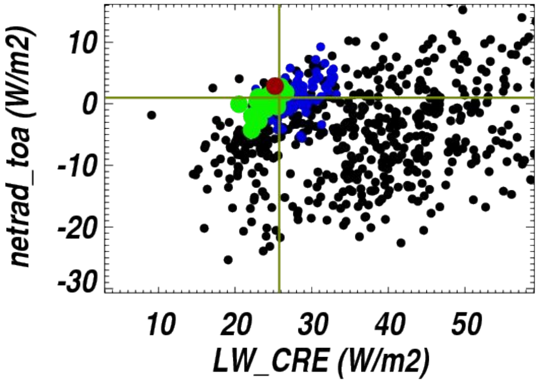
Sometimes ML parameter estimation can fix a problem that Latin Hypercube (or human) sampling of parameters could not (e.g., shallow cumulus, Amazon precip bias reduced by 50%)

It does all this while minimizing SW and LW radiation biases.

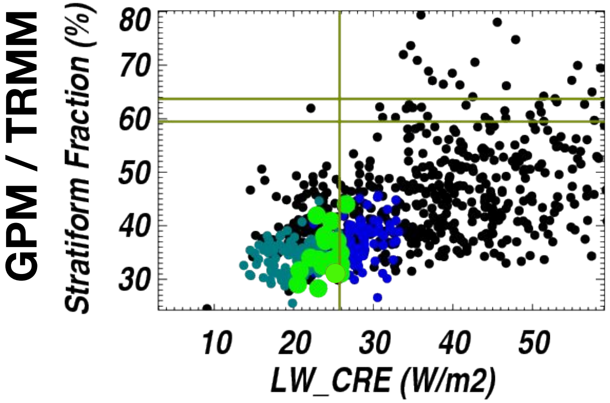
Lessons Learned from GISS ESM Atmosphere Parameter Estimation Efforts



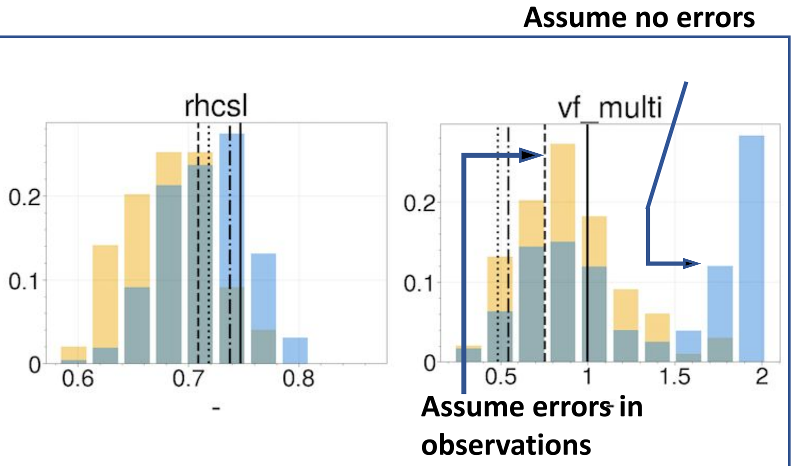
Results1: Process matches obs. target cross-hairs well (posterior **green dots**), much better than random Latin Hypercube search of parameters (**black dots**). Example: LW cloud radiative forcing and net radiation at TOA.



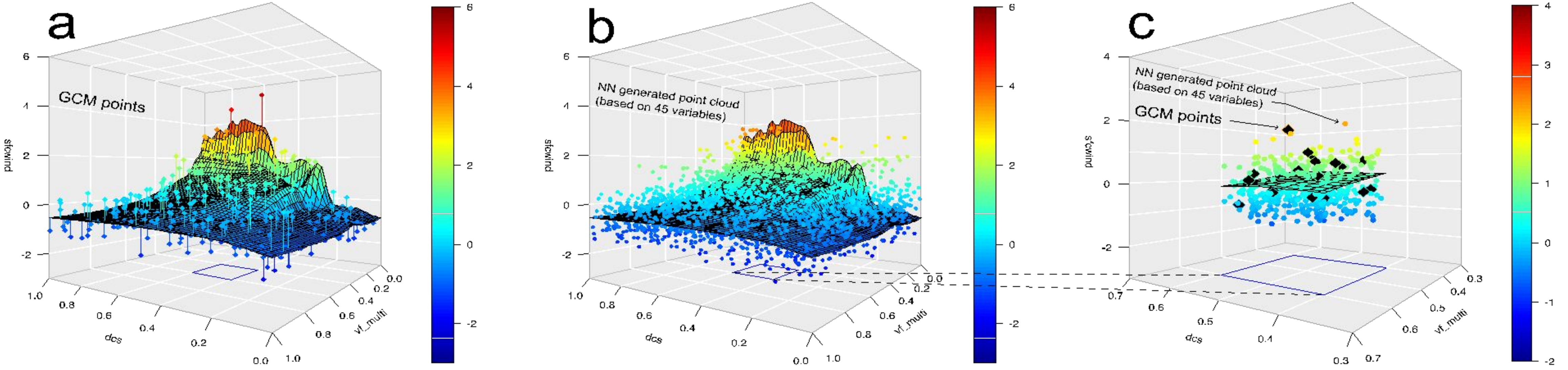
Results2: Sometimes, we can never match targets. Example: for stratiform rainfall fraction, or the amount of large-scale (light) precipitation that falls relative to total precip. **Structural problem!**



Tackling Error in Obs.: If error in observations are not accounted for, histograms of "optimal" parameter combinations are sometimes different (e.g., RH for cloud formation, or ice fall speed). So, what we assume for error is important!



Improving emulator for next round ESM atmosphere parameter estimation

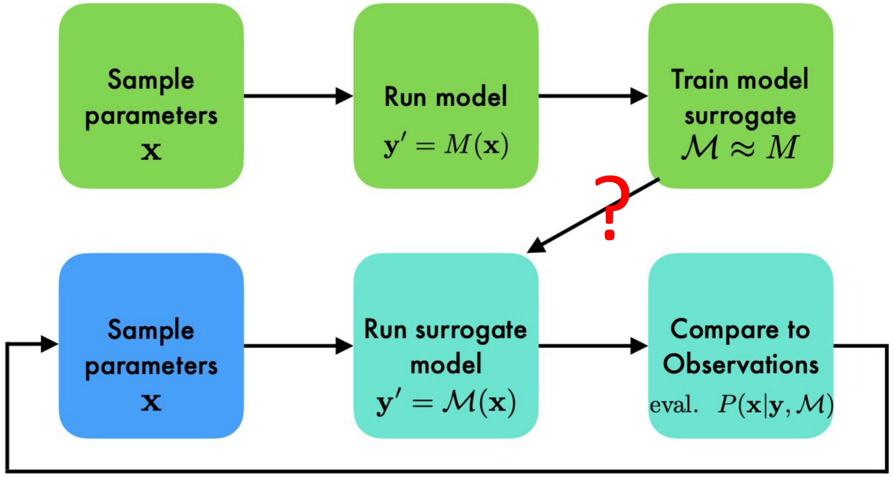


e.g., model surface wind as a function of two microphysics parameters (**points in a**); GISS model surrogate model (a neural net; NN) filling the empty space also shown (**points in b**). The **NN** "cloud of points" in (**b**) is flatter than in (**a**), denoting **NN** error. Ongoing work is to improve the **NN** so it can re-produce any model field!

The uncertainty in the posterior distribution also contains the uncertainty from the emulator;

What is an ideally good emulator?

- Avoid overfitting
- Be able to estimate the uncertainty
- The estimated uncertainty should vary based on the provided parameter values

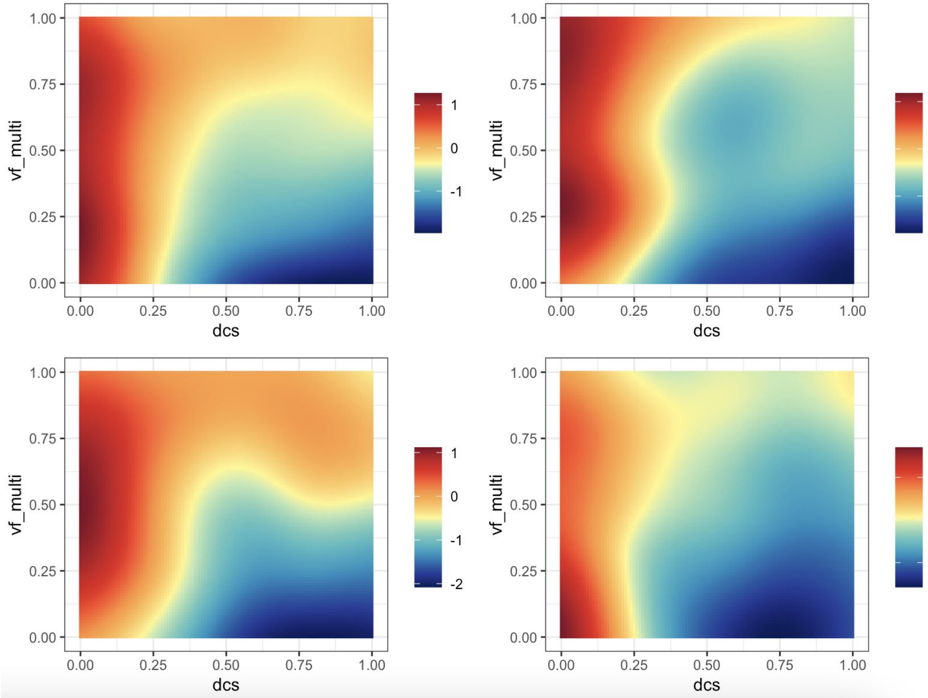


Improve the performance of the emulator

- Explore different kinds of structures in the data
- 1. Features that can be easily identified on a biplot (below)
- 2. The interaction of these variables (3D image below)
- 3. Local variabilities (right), e.g.,
  - When some other variables are within certain range, the pattern of StreamFunction_DJF with respect to dcs and vf_multi starts to look different from the overall trend
  - Subset A (144 pts): 0.2 < scale_cn < 0.5
  - Subset B (153 pts): 0.5 < ni_homfree < 0.8
  - Subset C: All other points

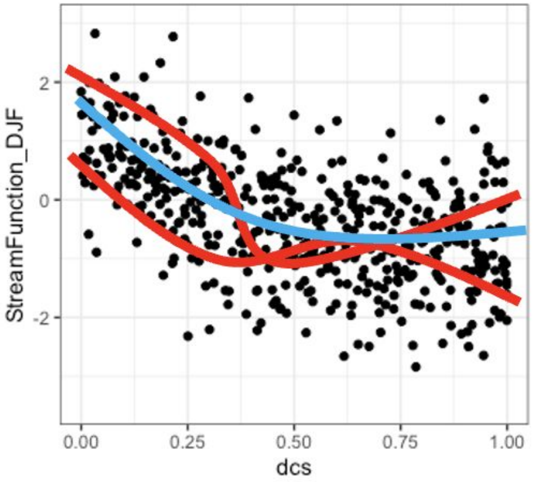
Surface based on complete dataset (0.70; shown before)

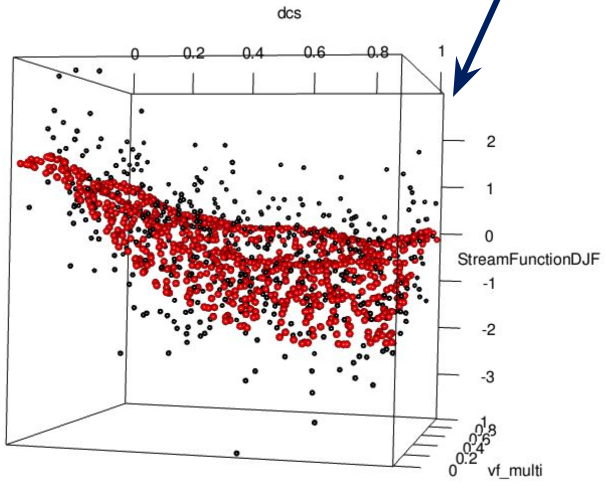
Surface based on subset A (0.58)



Surface based on subset B (0.57)

Surface based on subset C (0.74)





What's next?

Find ways to model such characteristics while avoid overfitting

Estimate the uncertainty

Conclusions

- The current framework could help find multiple possible configurations (different combinations of parameter values) to run ESMs with similar outputs. For different ESM parameters but similar ESM climatologies, how does this impact projections?
- Some observations can be well fitted by the ESMs based on the configurations derived from this workflow, but some are not (i.e., structural error exists!).
- Observational error in the observation is critical and may greatly affect the posterior distribution of the parameters.
- We are working on improving the performance of the emulator by exploring local structures (i.e., input and output relationship) of the training data.
- We hope that these lessons could be informative to the ESM parameter estimation community, and look forward to collaborate with CAM.