# Gaining insights into climate model behavior from Perturbed Parameter Ensembles (PPEs) using an additive emulator

Qingyuan Yang[1,2,*], Gregory S Elsaesser[1,3,4], Marcus Van Lier-Walqui[3,5], Trude Eidhammer[6], Linnia Hawkins[1,2]

1. Learning the Earth with Artificial Intelligence and Physics (LEAP) National Science Foundation (NSF) Science and Technology Center, Columbia University; 2. Department of Earth and Environmental Engineering, Columbia University; 3. NASA Goddard Institute for Space Studies; 4. Department of Applied Physics and Applied Mathematics, Columbia University; 5. Columbia Climate School; 6. National Center for Atmospheric Research
* qy2288@columiba.edu

# Outline

Climate model Perturbed Parameter Ensembles (PPEs)

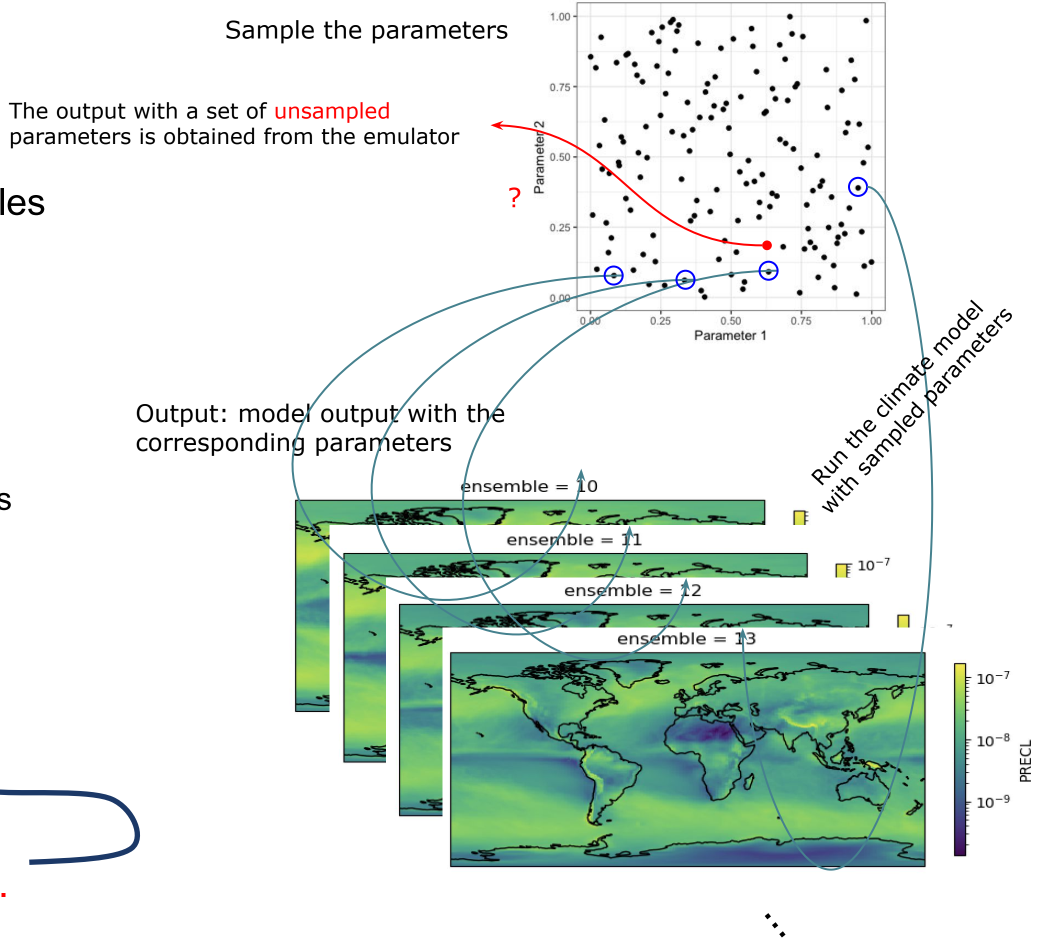
- Generation;
- Importance;
- Sparsity:
    1. A few hundreds of ensemble members;
    2. Tens of parameters perturbed.
- Draw the coastline of a country given three points

Method introduction

Performance evaluation

Significance:

1. Insights from working with two PPEs;
2. Analyzing a PPE ≠ analyzing emulator-generated data trained based on the PPE.



# Method

A simple emulator for climate model PPEs

Simplified from additive GP (Gaussian Process)

Additive;

Accounts for parameter interaction;

One variable at a time;

Does not quantify uncertainty.

Prediction = $\Sigma f_i(\theta_i) + \Sigma f_j(\boldsymbol{\theta_j}) + \Sigma f_k(\boldsymbol{\theta_k})$

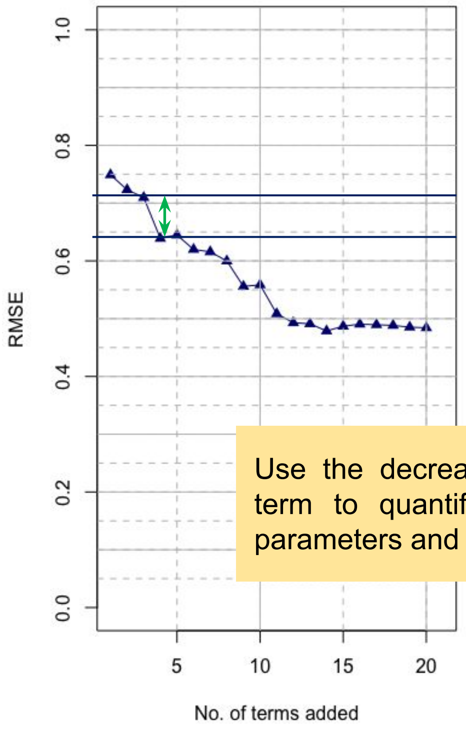
$f_i$, $f_j$, $f_k$: means of GPs with fixed hyperparameters (tested later)

$\theta_i$: a single parameter

$\boldsymbol{\theta_j}$: a parameter pair

$\boldsymbol{\theta_k}$: a parameter group of three

The RMSE decreases as more added terms are summed




Contribution from a single parameter or parameter group

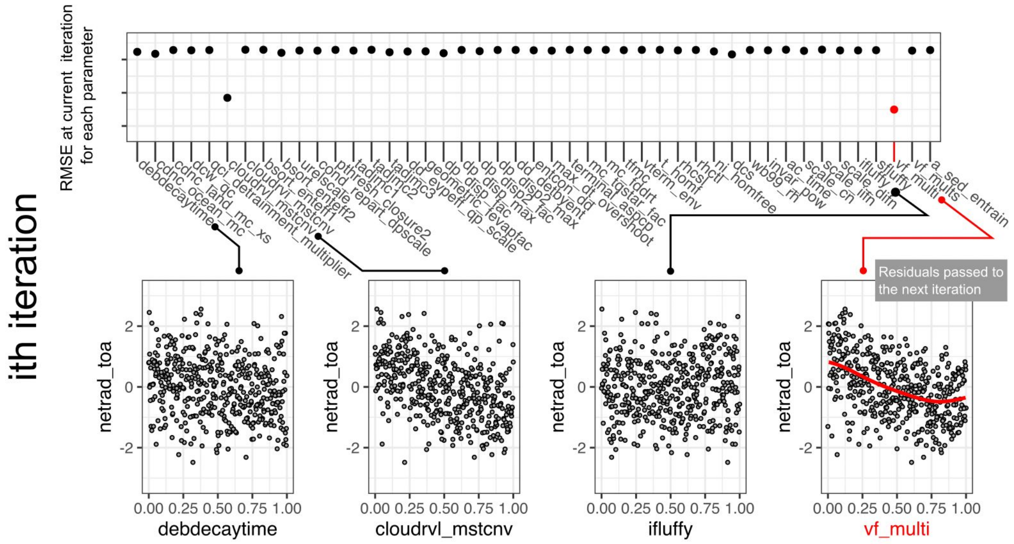
Use the decrease in RMSE of each term to quantify the contribution of parameters and parameter groups.

# Method implementation

Works through iterations (example for single parameters shown on the right)
In each iteration:

1. Fit a GP for each parameter;
2. Select the parameter with the lowest RMSE (red point);
3. Use the GP mean as the term for this iteration (red line)
4. Pass the residuals to the next iteration.

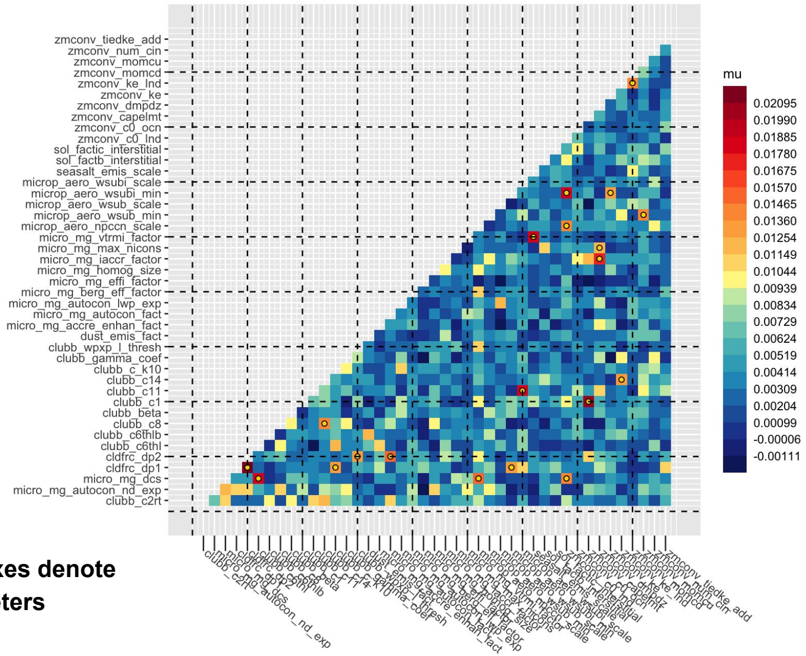
Similar procedure for parameter groups




Number of additive terms determined based on training and validation of 80% and 20% of the training data.

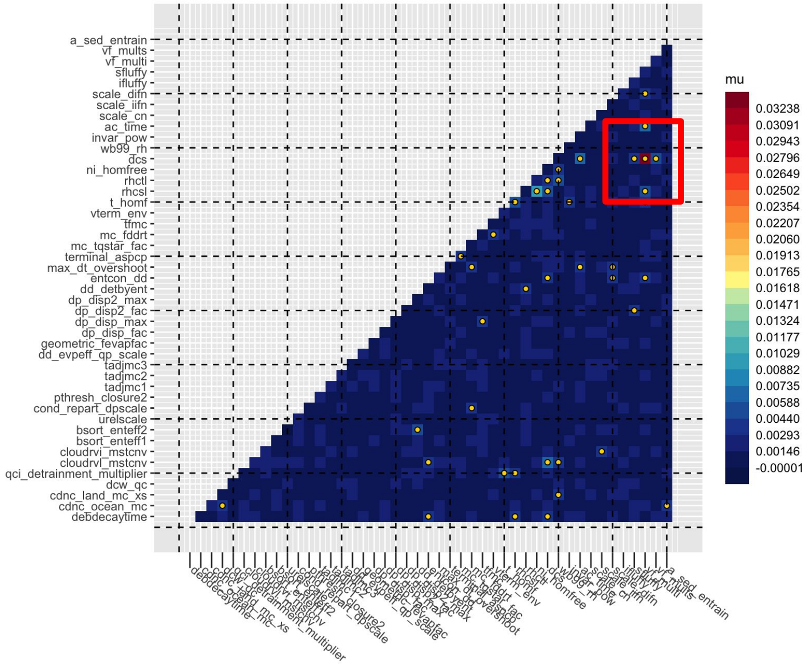
# The only difference in working with parameter groups: the measure to select the parameter groups

Assumption: if a parameter group is important, its importance should not be sensitive to the GP hyperparameters.
The measure we propose reflects the benefit of emulating a variable using two parameters **jointly** compared to using them **independently**.




**Both axes denote parameters**

A variable without a parameter pair contributing significantly to explaining its variability




A variable with a parameter pair contributing significantly to explaining its variability

# Method evaluation (R-sq based)

Compared with Neural Network with (80% training; 20% testing):

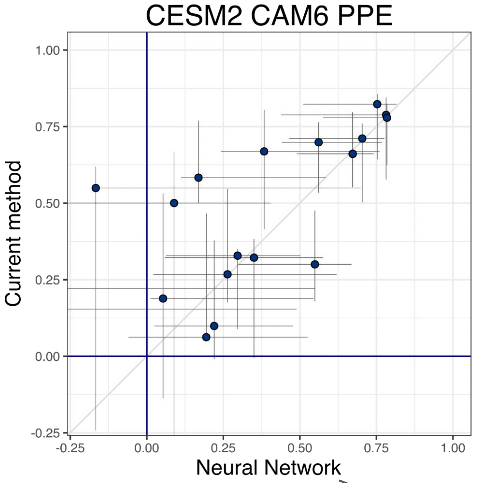
CESM2 CAM6 PPE (262 ensemble members)

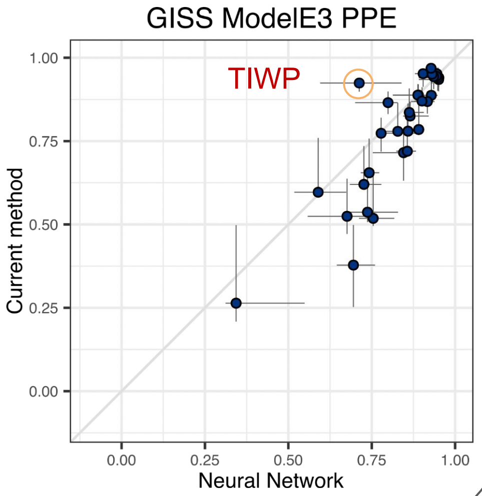
GISS ModelE3 PPE (751 ensemble members)

Compared with fully connected Neural Network.

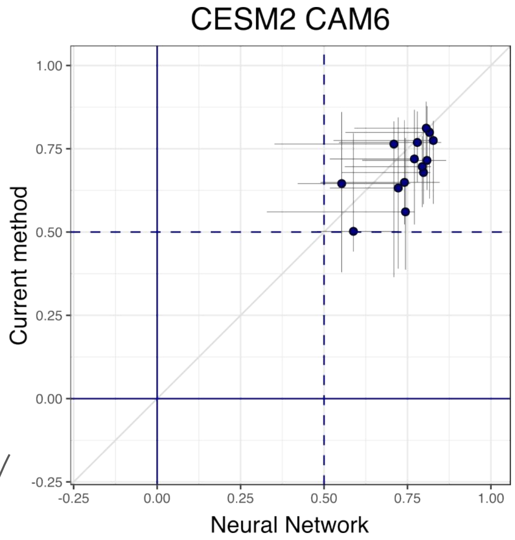
Target variable: global averages (direct model output; e.g., SW_CRE, precipitation) or **model scores** (weighted difference between zonal model outputs and observations)

Variability from random sampling for 11 times (bars)




Performance of both emulators when focusing only on global averages for the CAM6 PPE



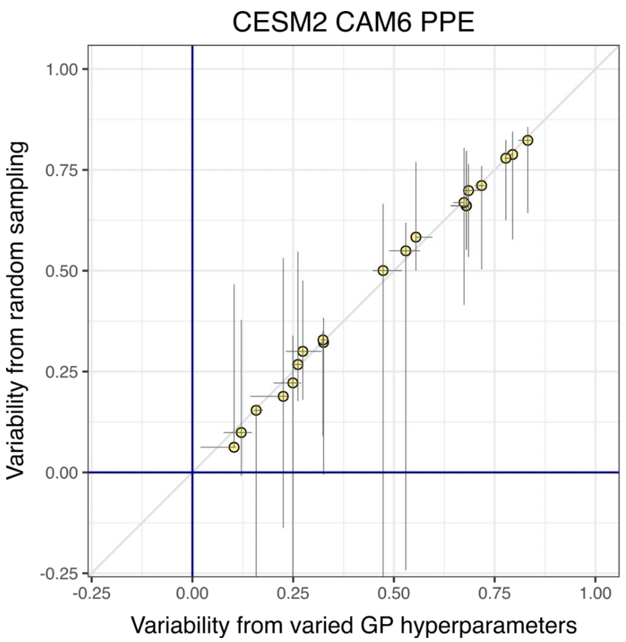

- Comparable performance between the current method and Neural Network.
- **The "learnable" relationship between parameters and output variables is not that complicated, given the sparsity of the PPEs.**

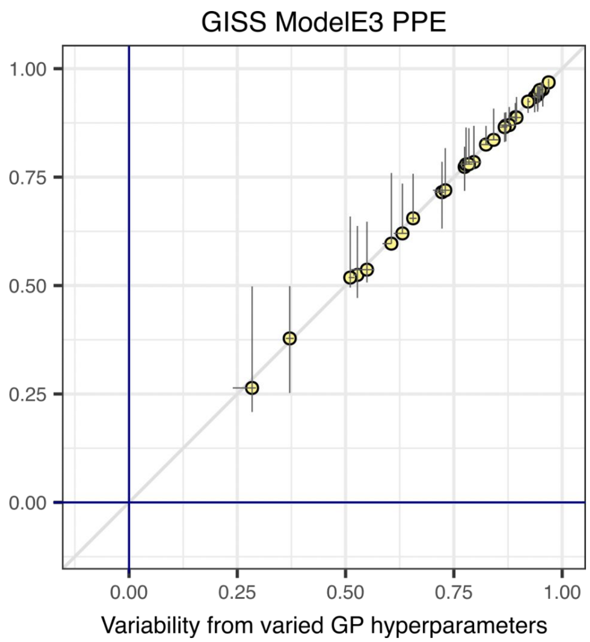
# Variability from random sampling vs from varied hyperparameters

The impact of hyperparameters is small.

x-axis: R-square variability from **varied hyperparameters**

y-axis: R-square variability from **random sampling**



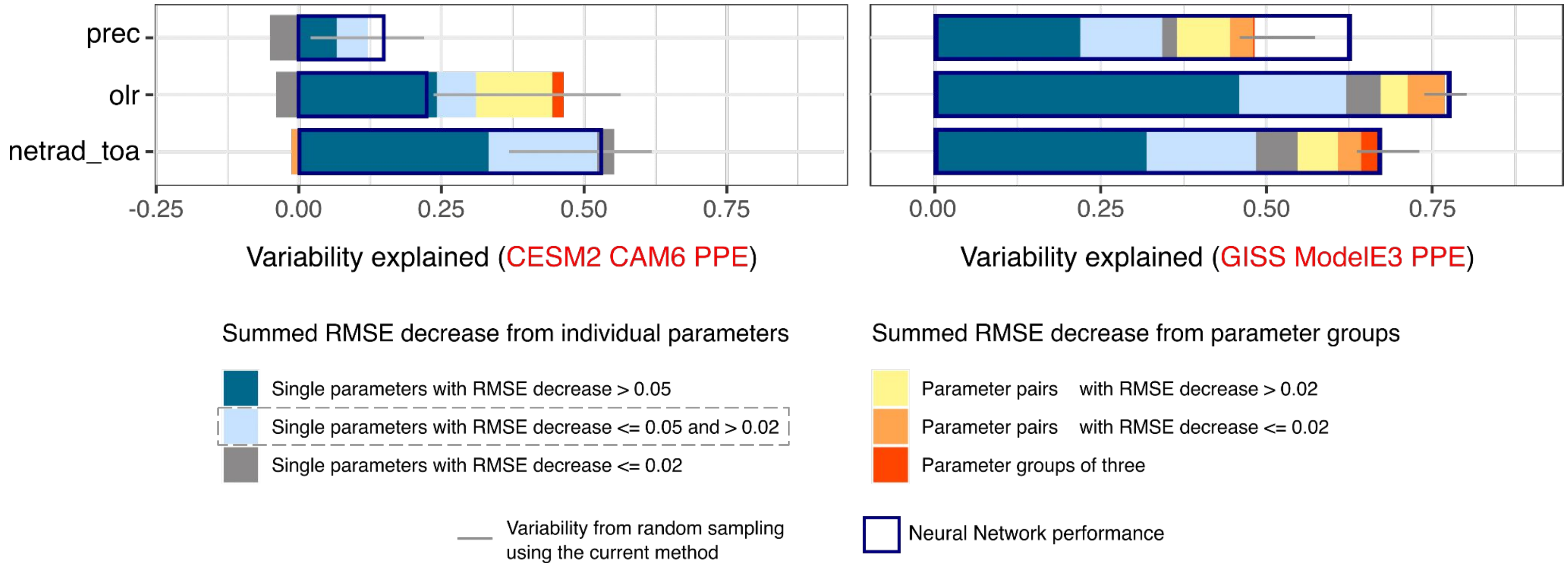




| Set name | Range for 1-D GP | Range for 2-D GP | Range for 3-D GP | Nugget to variance Ratio |
|---|---|---|---|---|
| Default Test | 0.60 | $\sqrt{0.5^2+0.5^2}$ | $\sqrt{0.4^2+0.4^2+0.4^2}$ | 2.00 |
| Set 1 | 0.60 | $\sqrt{0.5^2+0.5^2}$ | $\sqrt{0.4^2+0.4^2+0.4^2}$ | 4.00 |
| Set 2 | 0.60 | $\sqrt{0.5^2+0.5^2}$ | $\sqrt{0.4^2+0.4^2+0.4^2}$ | 1.00 |
| Set 3 | 0.80 | $\sqrt{0.6^2+0.6^2}$ | $\sqrt{0.4^2+0.4^2+0.4^2}$ | 2.00 |
| Set 4 | 1.00 | $\sqrt{0.8^2+0.8^2}$ | $\sqrt{0.6^2+0.6^2+0.6^2}$ | 2.00 |
| Set 5 | 0.50 | $\sqrt{0.4^2+0.4^2}$ | $\sqrt{0.3^2+0.3^2+0.3^2}$ | 2.00 |

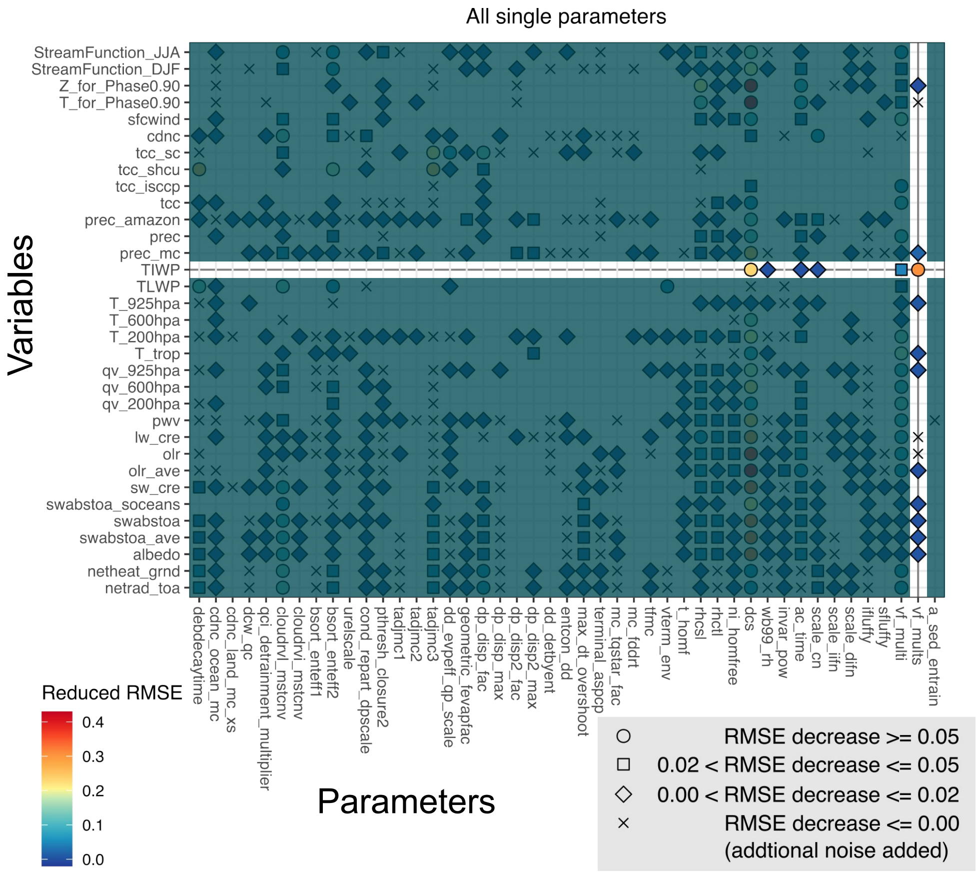
# Insights from working with the method

**1. Cumulative contribution from individually less important parameters and parameter groups (more pronounced in the ModelE3 PPE) should not be neglected;**

Output variables normalized to have standard deviation of one.



# Insights from working with the method (example from GISS ModelE3 PPE)




The method outperforms Neural Network in emulating TIWP (unshaded)

The current method emulates one variable at a time;

Neural Network emulates all variables all at once;

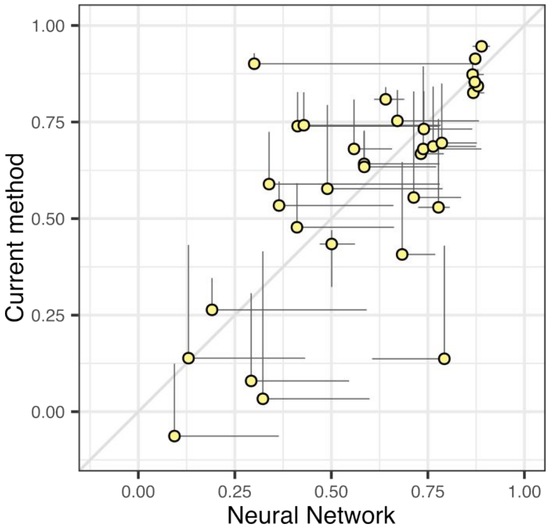
The simple, **stand-alone** relationship between *TIWP* and *vf_mults* is too “weak” and hence overlooked by Neural Network.

**2. Analyzing a PPE ≠ analyzing emulator generated data trained based on the PPE.**

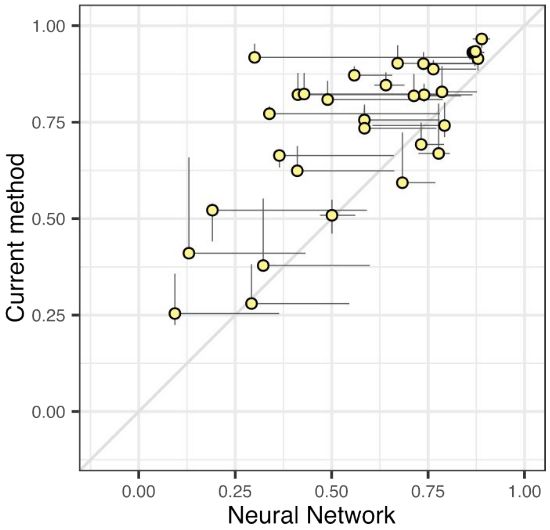
# Insights from working with the method (example from GISS ModelE3 PPE)

**3. The number of ensemble members matters.**

GISS ModelE3 trained based on
250 ensemble members




GISS ModelE3 trained based on
250 ensemble members
+ using selected parameters
and paramter groups trained
based on 611 ensemble memebrs
using the current method



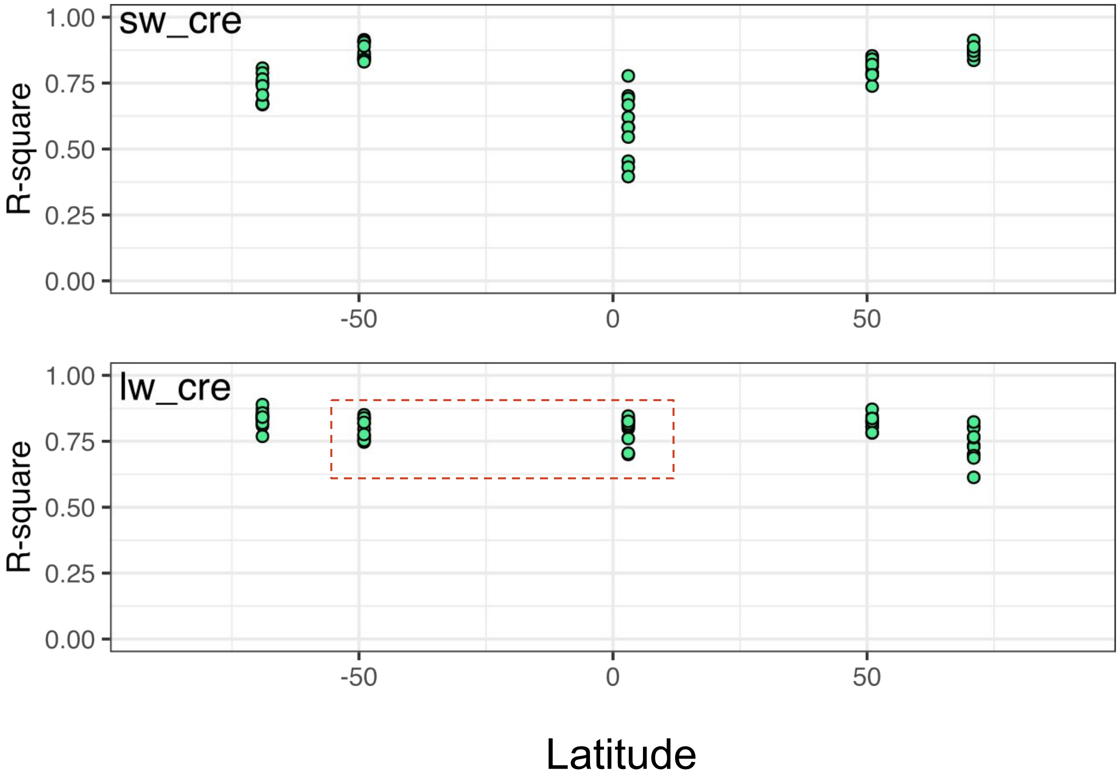

The x-axis is the same in the two figures.

Note:
The result on the right cannot be used to indicate better performance of the current method, as additional information is used.

**The 250 points have information that could help with emulation, but such information is difficult to capture by both the current method and Neural Network.**

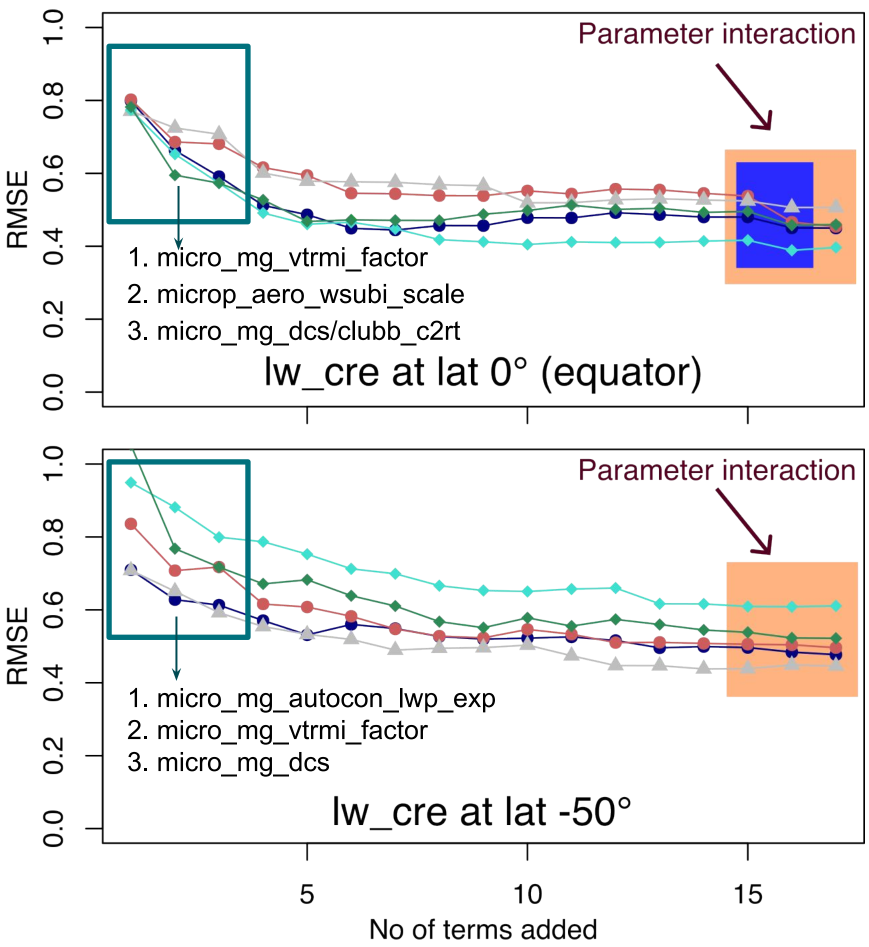
# Insights from working with the method (example from CESM2 CAM6 PPE)

**4. The difficulty in emulating a variable varies by latitude/location (emulating zonal average)**




Point spread from random sampling

# LW_CRE at -50° and 0° (CAM6 PPE)




The most important single parameters are different at the two latitudes.

The decrease in blue rectangle:
small but robust decrease in RMSE from parameter pair (micro_mg_dcs, micro_mg_vtrmi_factor)

No RMSE decrease even if we force the method to consider parameter pair interaction

Different lines corresponds to random sampling tests to ensure robustness.

# Conclusions

- We present a new method that implements **analysis and** emulation for climate model PPEs;
- The method is applied to two PPEs with performance comparable to fully connected Neural Network;
- Insights from working with this method, e.g.,
  - The “learnable” relationship between parameters and output variables is not that complicated, given the sparsity of the PPEs;
  - A group of individually less sensitive parameters could have a non-negligible and cumulative impact on the overall emulator performance;
  - Emulating the variables all at once is not always the perfect emulator design;
  - A PPE with small ensemble size could have useful information that would benefit the emulator performance, but such information is difficult to identify by an emulator;
  - The difficulty in emulating the same variable could vary by region.
- Analyzing a PPE ≠ analyzing emulator-generated data trained based on the PPE;

# Thank you!

National Aeronautics and Space Administration
Goddard Institute for Space Studies
Columbia Climate School
NCAR
Operated by UCAR