A new climate model PPE emulator that enables new insights on host model behavior

One of the insights

Original title: detecting and identifying the **impact of parameter interaction** on climate model outputs based on two Perturbed Parameter Ensembles (PPEs)

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

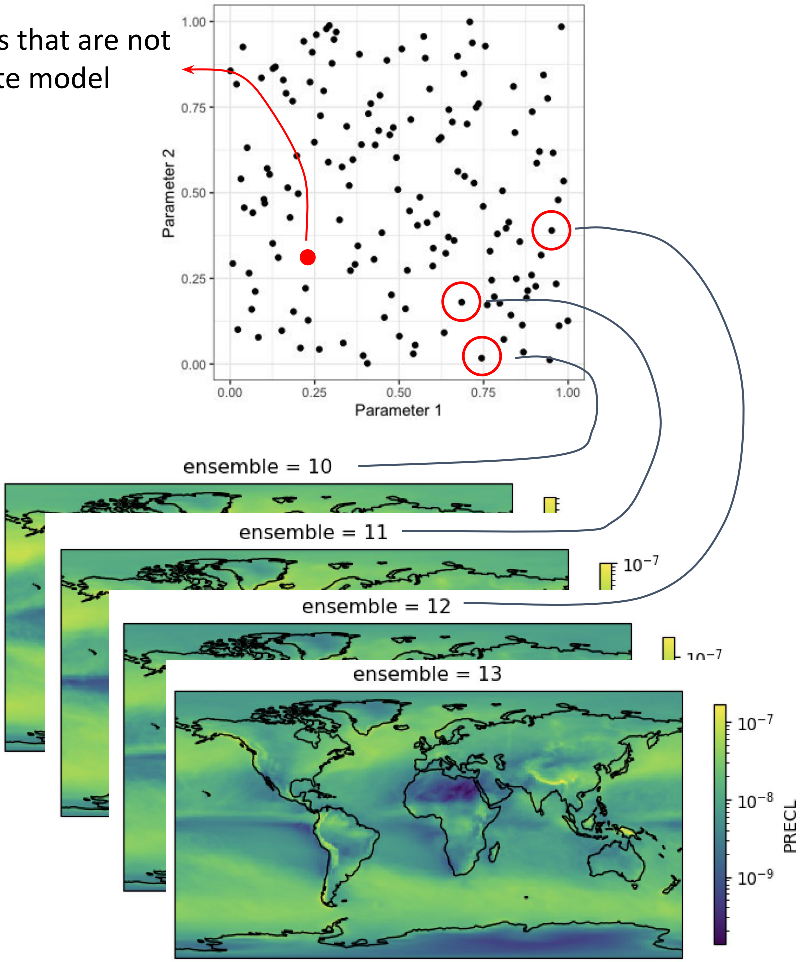
1. Learning the Earth with Artificial Intelligence and Physics (LEAP) National Science Foundation (NSF) Science and Technology Center, Columbia University; 2. Department of Earth and Environmental Engineering, Columbia University; 3. NASA Goddard Institute for Space Studies; 4. Department of Applied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

National Aeronautics and Space Administration
Goddard Institute for Space Studies

Background

- Perturbed Parameter Ensembles (PPEs)
  - A collection of climate model simulations
  - Each simulation corresponds to different parameter sets
  - Ensemble has a few hundred members, typically
  - Key feature: sparsity
- Input: climate model parameters (e.g., size threshold for cloud ice-snow autoconversion)
- Output: processed climate model output (e.g., global average longwave cloud forcing)
- Uses:
  - Studying how parameters affect the model output
  - Training emulators for parameter tuning
  - Uncertainty connected to parameter settings

Parameter values that are not run by the climate model



Considerations for emulator development

Parameter – model output relationships	How non-linear	Linear	Very non-linear
	Parameter interaction	No interaction	All parameters interacting
Optimal ways to train the emulators?	One emulator for one or all variables?	One emulator for one variable	One emulator for all variables
	How many ensemble members?	100, 200, … , 600... how many is enough for skillful emulation?	
	How to set target variables	Output (e.g., global climatologies) or scores (e.g., differences from observations)	
	Parameter sensitivity	How insensitive is truly insensitive?	

Propose a new method: *sage* (Simplified Additive Gaussian processes Emulator)

A new additive emulator

- Simplified from additive Gaussian Processes (GPs)
- Focuses on global, zonal, or local climatologies
- Works with one variable at a time
- A target variable is decomposed as the sum of means of GPs
  - Each GP corresponds to a parameter or parameter group:

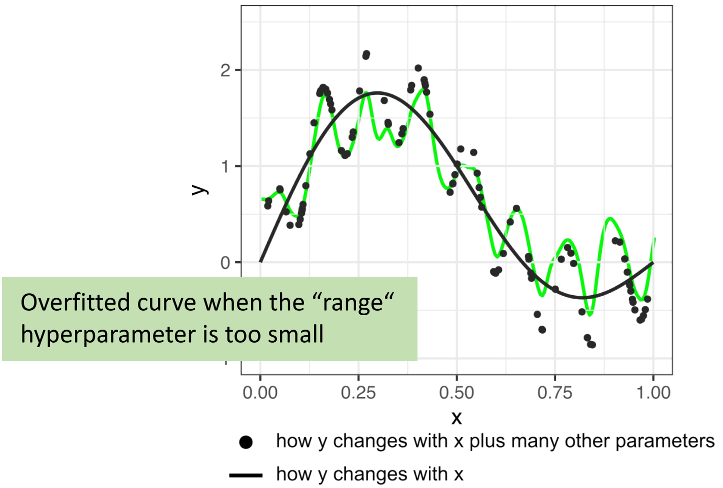
$$\text{Prediction} = \Sigma f_i(\theta_i) + \Sigma f_j(\boldsymbol{\theta}_j) + \Sigma f_k(\boldsymbol{\theta}_k)$$

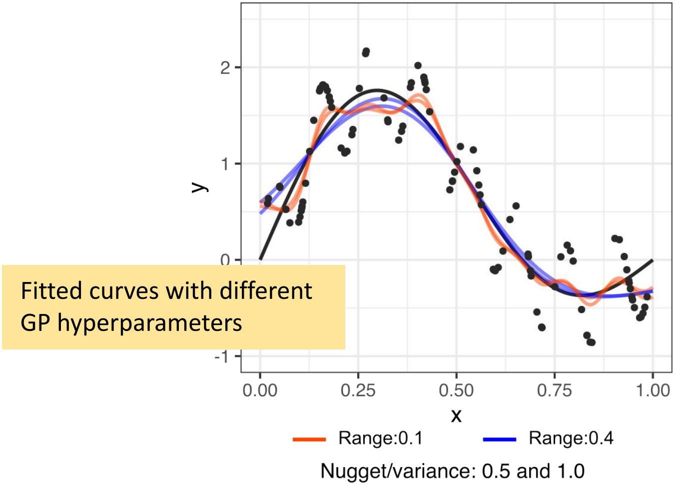
f_i, f_j, f_k: means of GPs with fixed hyperparameters (tested later)

θ_i: 1 parameter

$\boldsymbol{\theta}_j$: **2** parameters

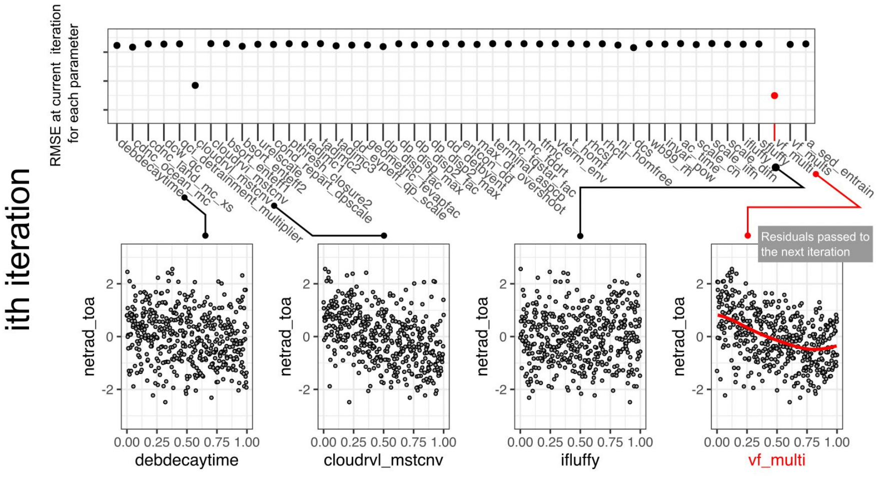
$\boldsymbol{\theta}_k$: **3** parameters





A new additive emulator

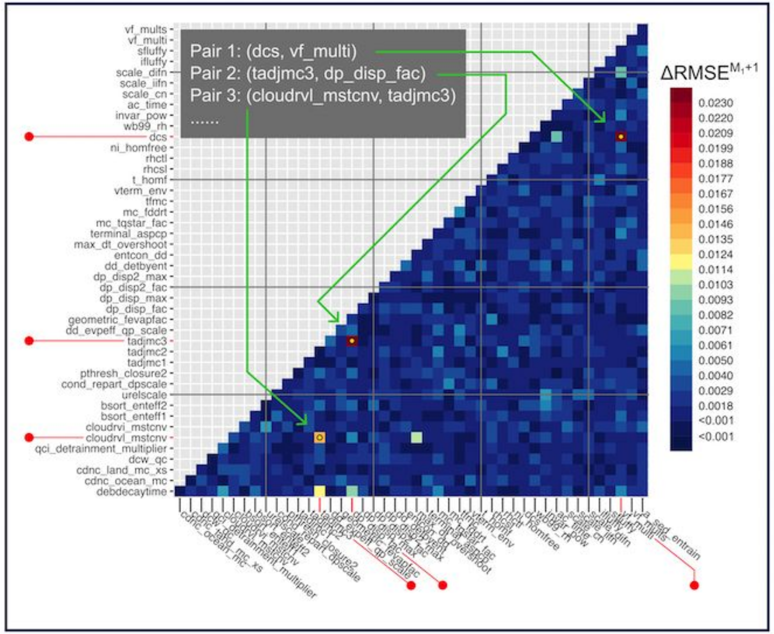
- Iteration for single parameter:
  - Select parameter based on RMSE
  - Fit to the target variable or residual from the previous iteration
  - Pass the residuals to the next iteration
- Similar procedure for parameter pairs and groups of three



x-axis: parameter; y-axis: RMSE for the corresponding parameter

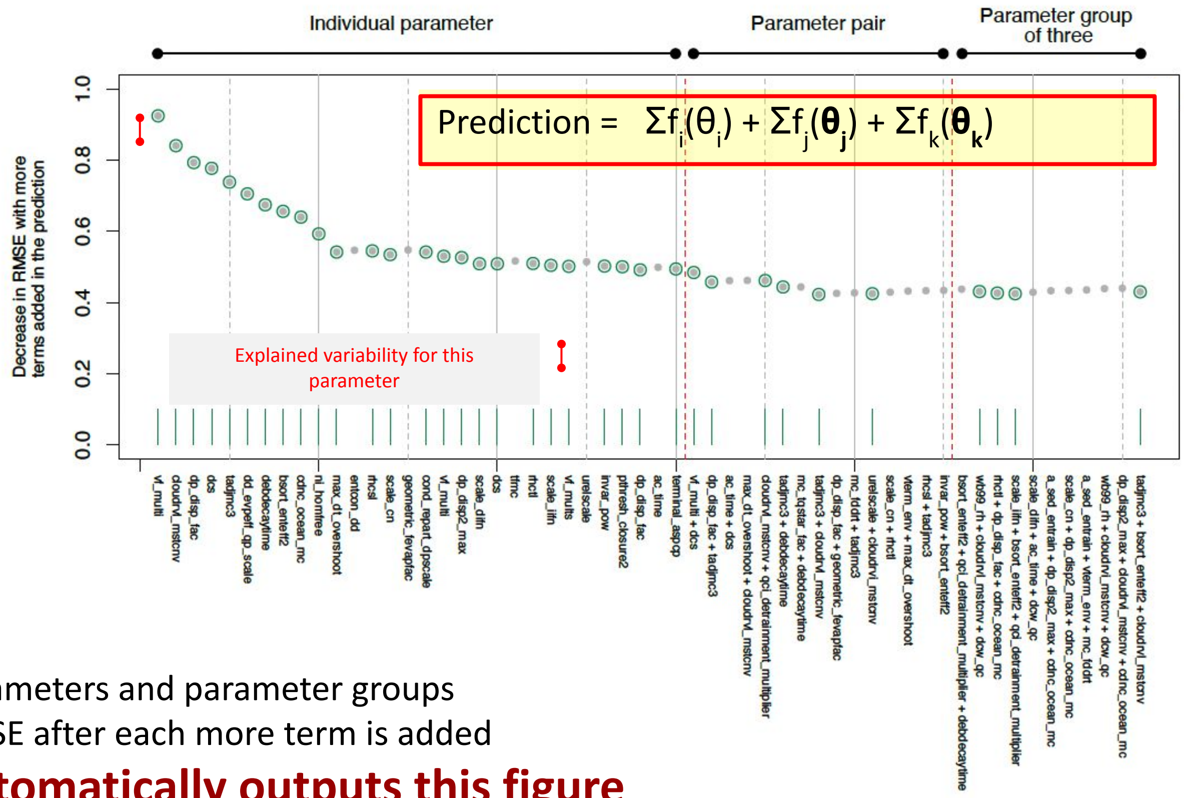
Selecting parameter pairs:

- Calculate $\Delta RMSE^{M1+1}$ for **all parameter pairs**
- $\Delta RMSE^{M1+1}$: the **additional** benefit of emulating two parameter impacts **jointly**



x- and y-axes: parameter

Interpretability: explained variability



x-axis: parameters and parameter groups

y-axis: RMSE after each more term is added

sage automatically outputs this figure

Method evaluation (R-sq based)

Comparison with fully connected Neural Network (**which emulates all variables all at once**); partitioning of data: 80% training; 20% testing.

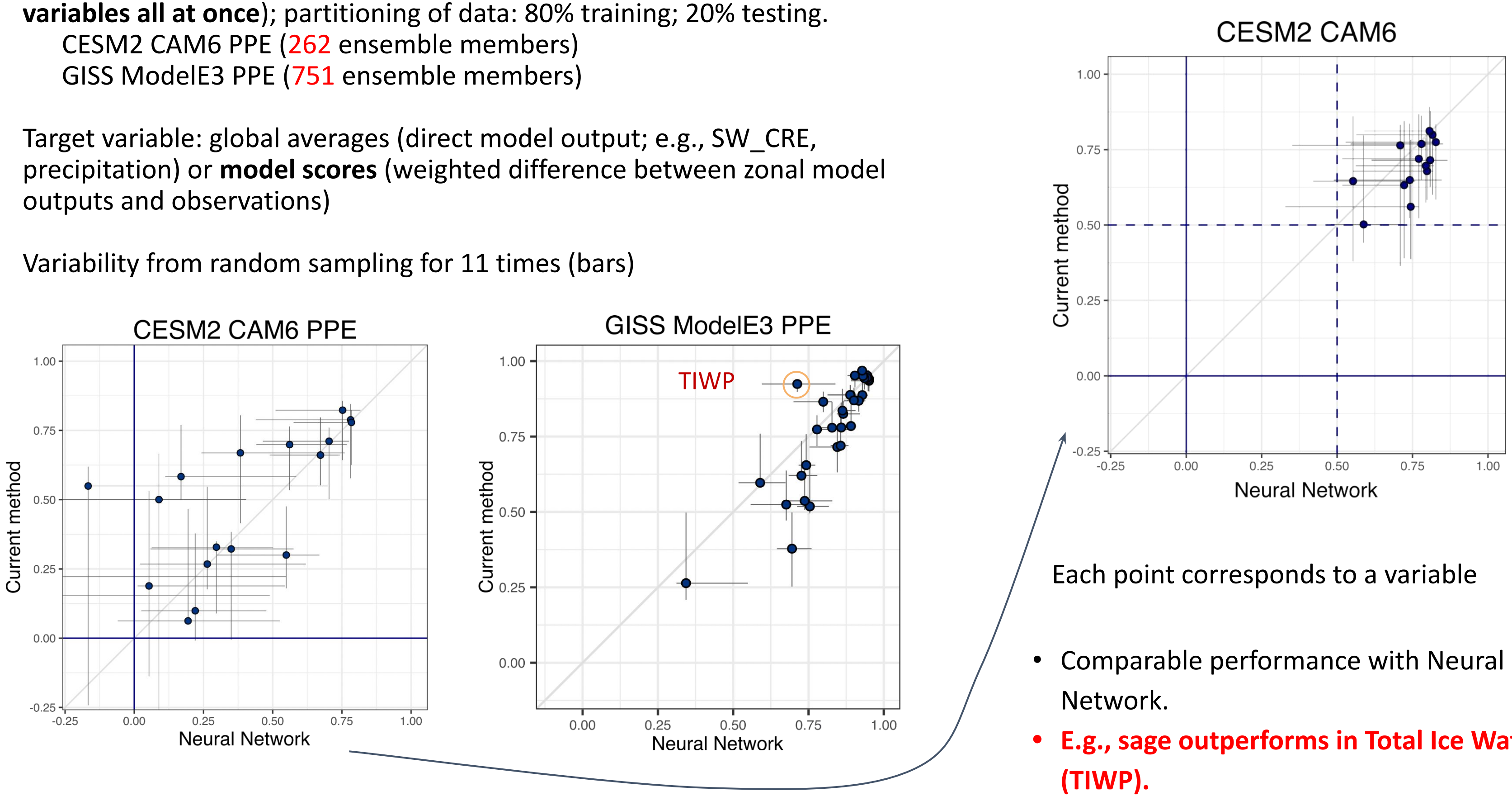
CESM2 CAM6 PPE (262 ensemble members)

GISS ModelE3 PPE (751 ensemble members)

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



Each point corresponds to a variable

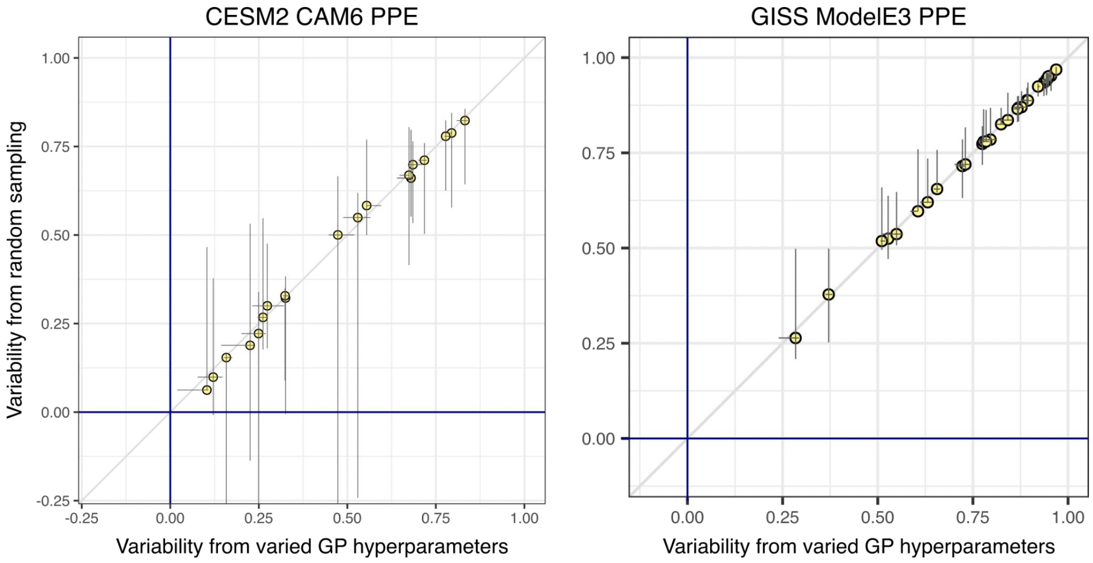
- Comparable performance with Neural Network.
- **E.g., sage outperforms in Total Ice Water Path (TIWP).**

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**



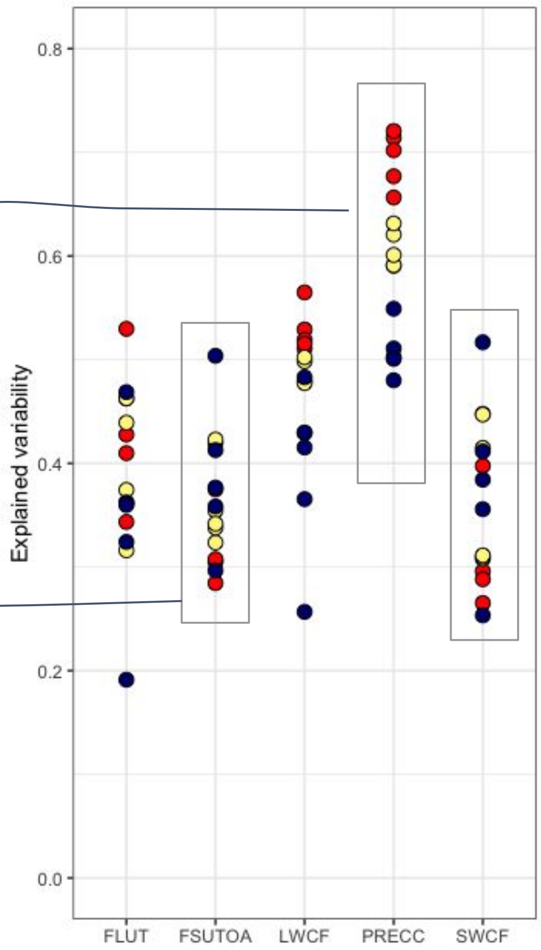
Set name	Range for 1-D GP	Range for 2-D GP	Range for 3-D GP	Nugget to variance Ratio
Default Test	0.60	$\sqrt{0.5^2+0.5^2}$	$\sqrt{0.4^2+0.4^2+0.4^2}$	2.00
Set 1	0.60	$\sqrt{0.5^2+0.5^2}$	$\sqrt{0.4^2+0.4^2+0.4^2}$	4.00
Set 2	0.60	$\sqrt{0.5^2+0.5^2}$	$\sqrt{0.4^2+0.4^2+0.4^2}$	1.00
Set 3	0.80	$\sqrt{0.6^2+0.6^2}$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

Comparison with linear regression (using CAM6 as example)

Heat maps (or correlation maps) showing parameter-output relationships are often based on linear regression, but these can be misleading in quantifying relationship strength

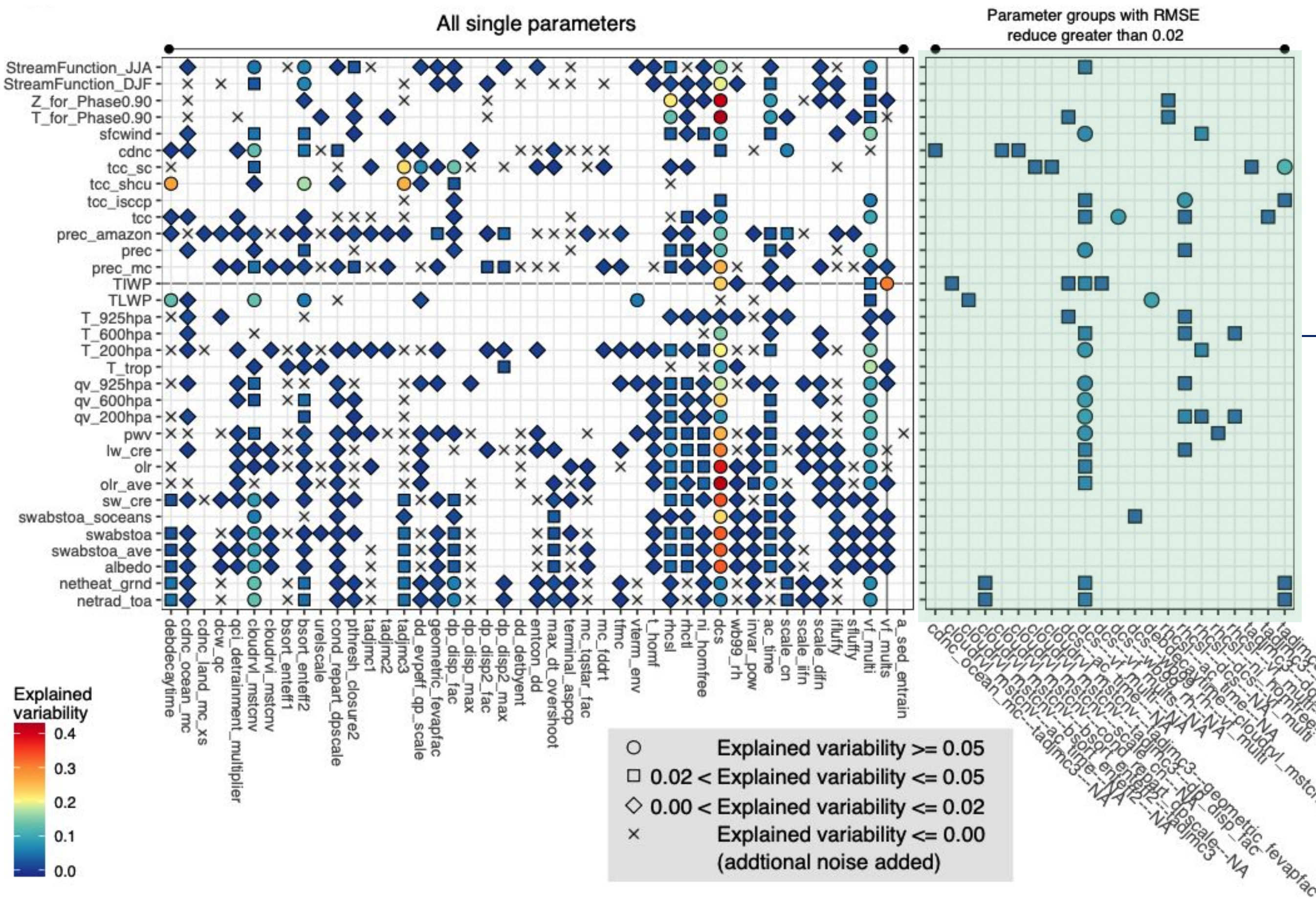
The relationship between parameters and targets is more complicated than a simple linear relationship, at least for some variables

Related to the presence of outliers being included or excluded in the training/validation



- *sage*
- *sage* excluding parameter interaction
- Linear regression

Additional analysis of the ModelE3 PPE



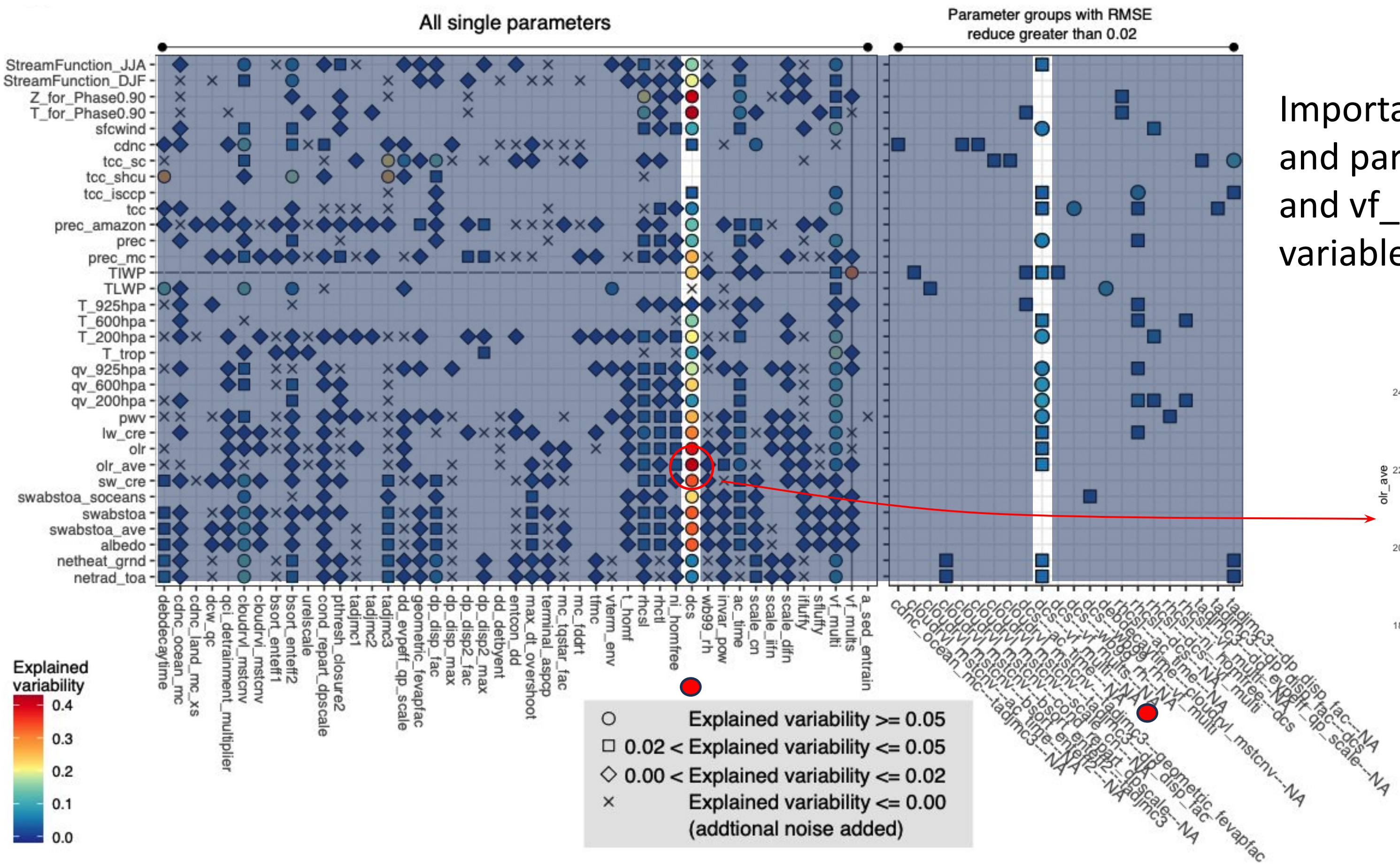
y-axis: target variables

x-axis: parameters and parameter groups

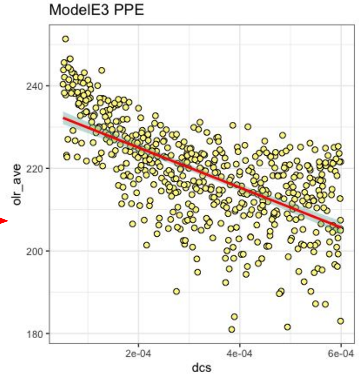
Color: explained variability

Variability explained by parameter groups (and beyond linear relationship)

Additional analysis of the ModelE3 PPE

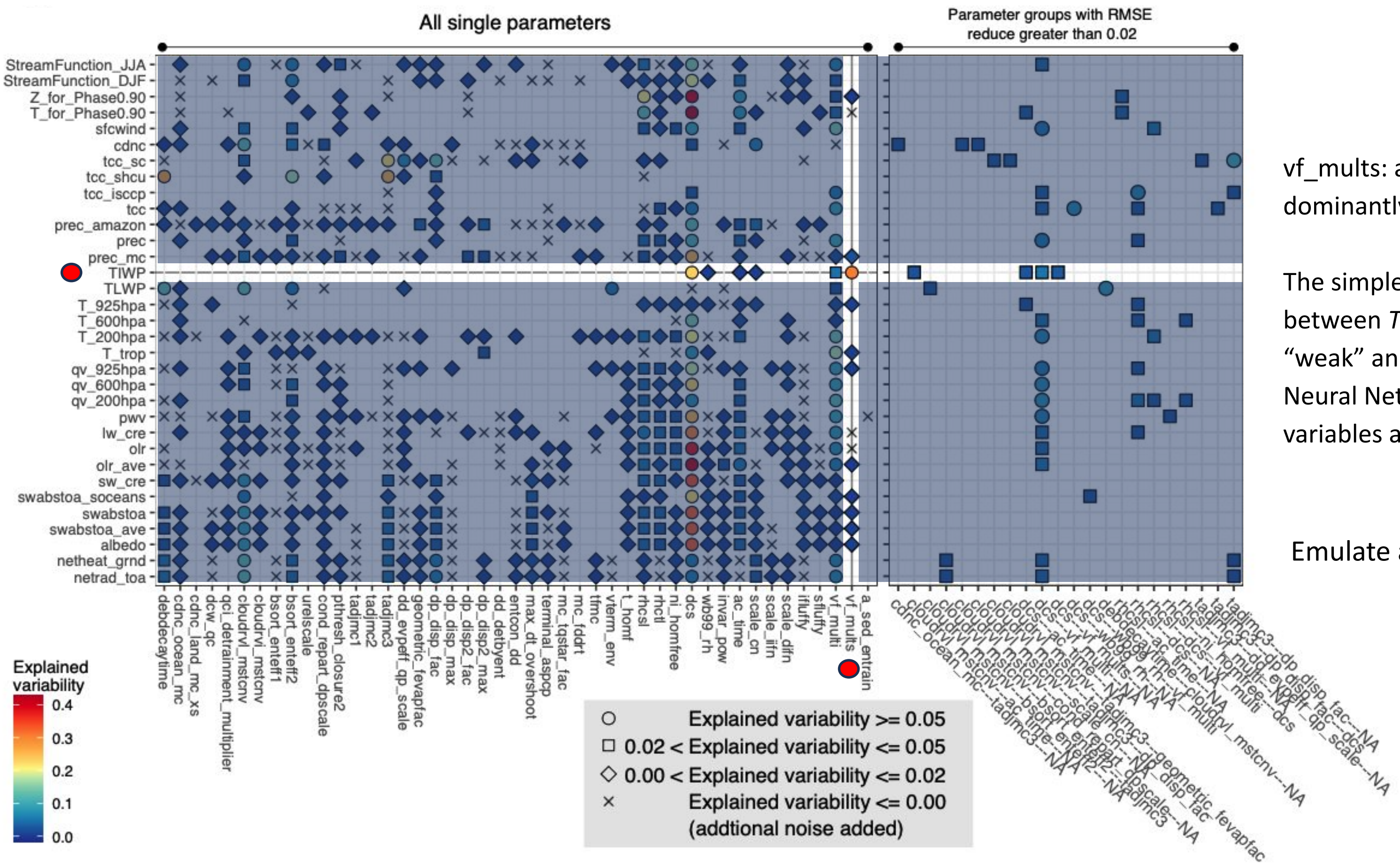


Important parameter (dcs) and parameter pair (dcs and vf_multi) for most variables



Linear relationship is not enough

Additional analysis of the ModelE3 PPE

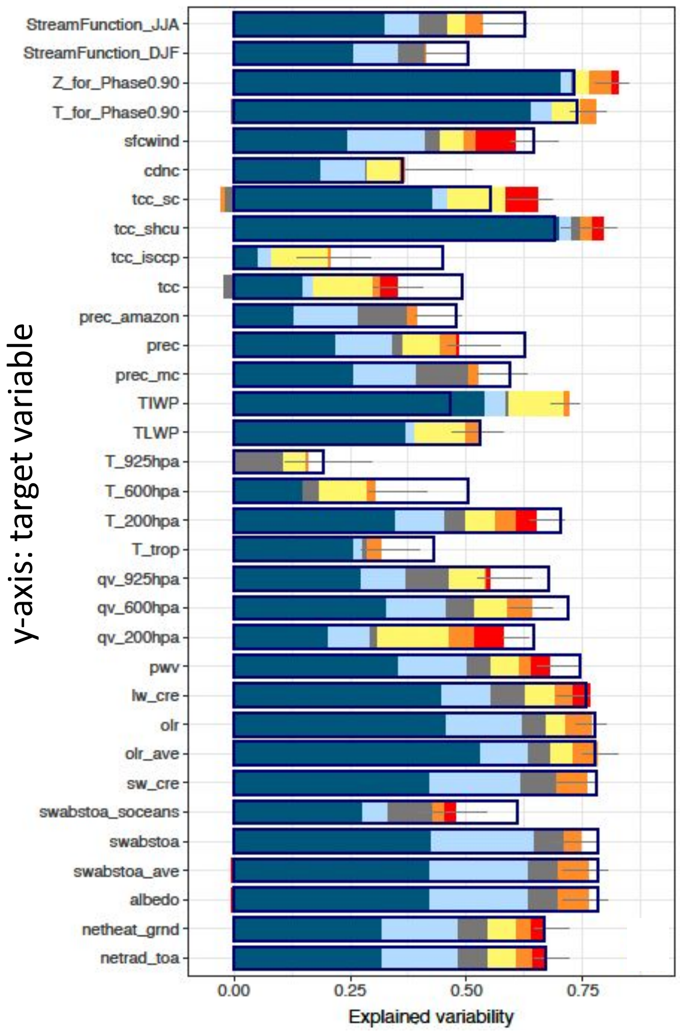


vf_mults: a “lonely” parameter that dominantly controls one target variable

The simple, **stand-alone** relationship between *TIWP* and *vf_mults* is too “weak” and hence overlooked by Neural Network which emulates all variables all at once

Emulate all variables all at once? X

Additional analysis of the ModelE3 PPE



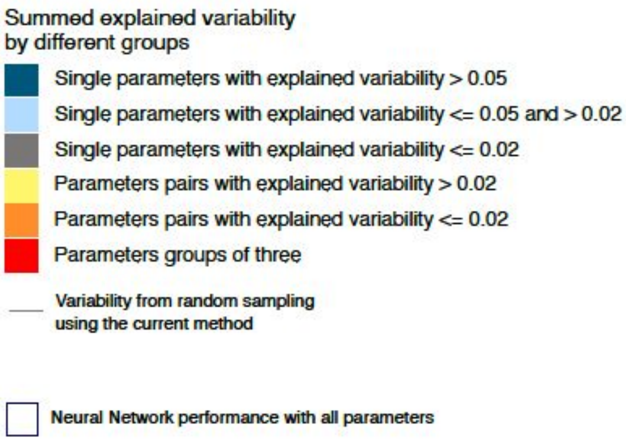
Insights based on ModelE3 PPE:

1. The cumulative impact of less sensitive parameters is not negligible

Intuition: y = $0.700x_0 + 0.018x_1 + 0.013x_2 + 0.010x_3 + \ldots + 0.005x_{45}$

2. Parameter pairs (yellow and orange): small but not negligible

3. Parameter groups of three (red): small



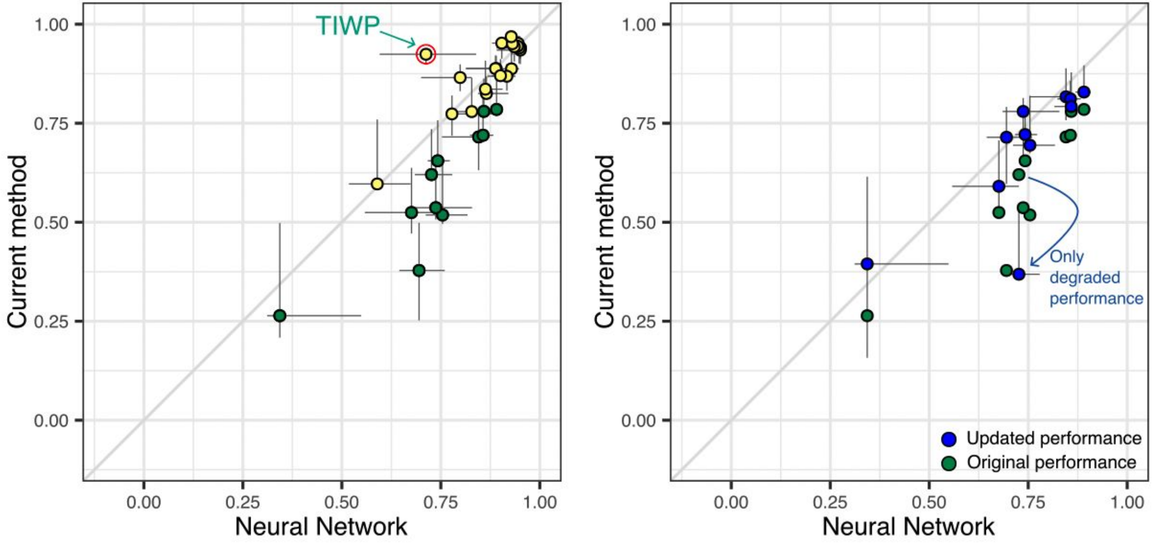
Individually less important parameters

Additional analysis of the ModelE3 PPE: including the relationship between target variables

Left: $y_1 = f_1(x_1, x_2, x_3)$ Easy to emulate (yellow points with high R-square)

$y_2 = f_2(x_1, x_2, x_3)$ Difficult to emulate (green points)

Right: $y_2 = f_2(x_1, x_2, x_3, f_1(x_1, x_2, x_3))$



R-square

Performance more comparable to NN (blue points)

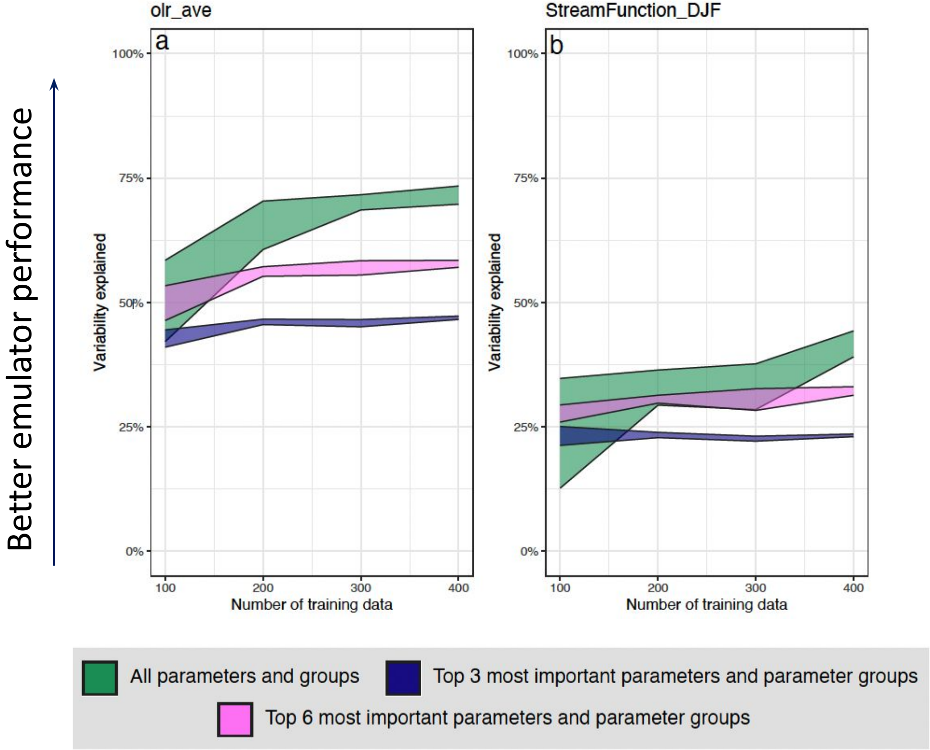
Emulating the variables one-at-a-time? X

Emulate all variables all at once? X

Group target variables and emulate them separately! ✓

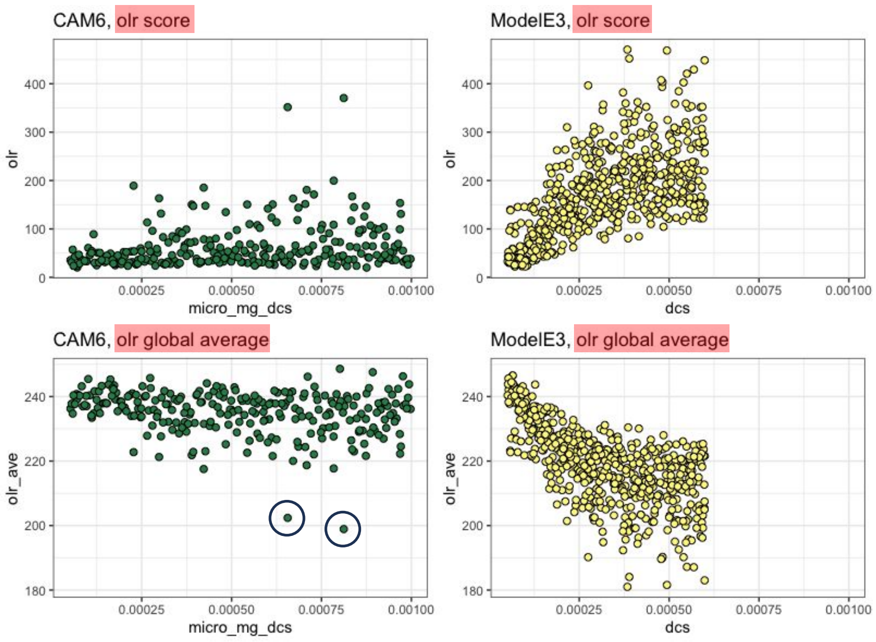
Additional analysis of the ModelE3 PPE

Method performance changes with training dataset size (2 variables as examples)?



- Train with 100, 200, 300, 400 ensemble members, predict another 100
  - Variability from random sampling
- Method performance does not change linearly with the # of training data samples
  - Depending on the variables, we may not need that many samples of parameters!

Why CAM6 PPE is more difficult to emulate?



1. Different models (an obvious conclusion)
2. PPE parameter ranges

Therefore:

Emulator performance and interpretation (e.g., parameter importance) are conditioned on the analyzed PPE

Ongoing work: uncertainty quantification in the emulator prediction and parameter-output relationships

- Focus:
- 1. The occurrence of outliers
  - Including or excluding the outliers in the training/validation greatly affects the emulator performance metric
  - The outliers are not errors!
- 2. Emulating regional climatologies for new parameter-output relationships

Link to

Manuscript Github repo

Conclusions

- A new additive emulator method (**sage**) is proposed for climate model PPEs
    - Comparable performance with a Neural Network
    - But, with new insights on climate model PPEs, and emulator design.

How non-linear?	Not very non-linear, but more complicated than linear relationship
Parameter interaction	Individual parameters and parameter pair interactions dominate
One emulator for one or all variables?	Emulate grouped target variables separately
How many ensemble members?	Examine how the emulator performance changes with the # of training data and decide
How to set target variables	Global climatologies: easier to emulate Model score: observational uncertainty: built into score, or during parameter estimation? (see talk by G Elsaesser at 8:55 AM tomorrow)
Parameter sensitivity	The cumulative impact of individually less sensitive parameters is not negligible

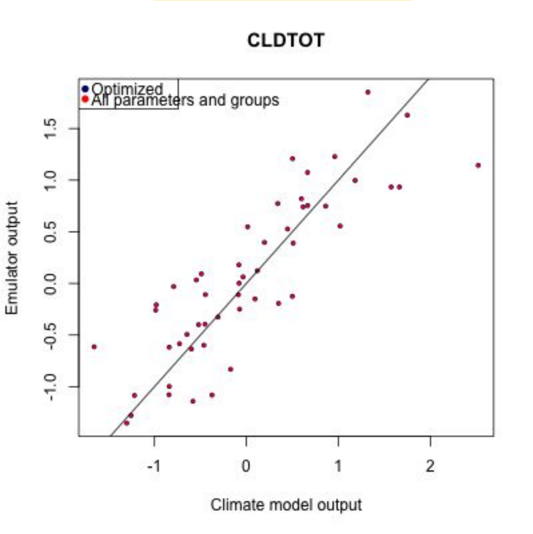
- Ongoing work:
    - Emulating regional climatologies
    - The occurrence of outliers in PPEs

Manuscript Github repo

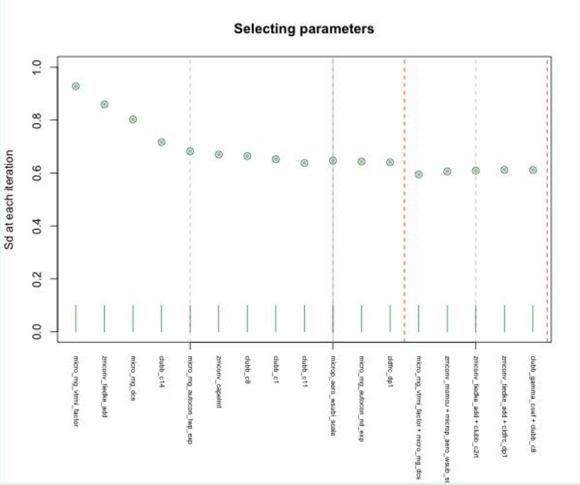
Thank you!

sage automated output

Validation



How explained variability decreases with more terms



Explained variability for parameters and parameter groups

Para1	Para2	Explained variability
micro_mg_vtrmi_factor	NA	0.144
zmconv_tiedke_add	NA	0.014
micro_mg_dcs	NA	0.043
clubb_c14	NA	0.063
micro_mg_autocon_lwp_exp	NA	0.038
zmconv_capelmt	NA	0.030
clubb_c8	NA	0.060
clubb_c1	NA	-0.009
clubb_c11	NA	0.023
microp_aero_wsubi_scale	NA	-0.012
micro_mg_autocon_nd_exp	NA	-0.002
cldfrc_dp1	NA	0.014
micro_mg_vtrmi_factor	micro_mg_dcs	0.039
zmconv_momcu	microp_aero_wsub_scale	0.007
zmconv_tiedke_add	clubb_c2rt	0.012
zmconv_tiedke_add	cldfrc_dp1	0.002
clubb_gamma_coef	clubb_c8	0.013