An autotuning approach to DOE's earth system model

Lyndsay Shand, Benjamin Wagman, Drew Yarger, Kenny Chowdhary (NVIDIA), Gavin Collins

Sandia National Laboratories, lshand@sandia.gov

August 8, 2023

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Secu DE-NA0003525.

wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. SAND NO. 2023-07335C

Motivation: Global Climate Model Calibration

- The goal of climate model tuning, or calibration, is to find the set of model parameters such that model predictions best match observational records.
- Currently, model tuning is done by ”hand” through a tedious ”plug and check” process. Different sets of parameters can be favored to best predict select variables or represent oscillations (e.g. ENSO, AMOC, NAO, etc.) most accurately. (E.g. Ma et al., 2022)
- We propose a flexible and straightforward framework for the automated tuning of the atmosphere-only model of the Department of Energy’s Energy Exascale Earth System Model (E3SMv2).

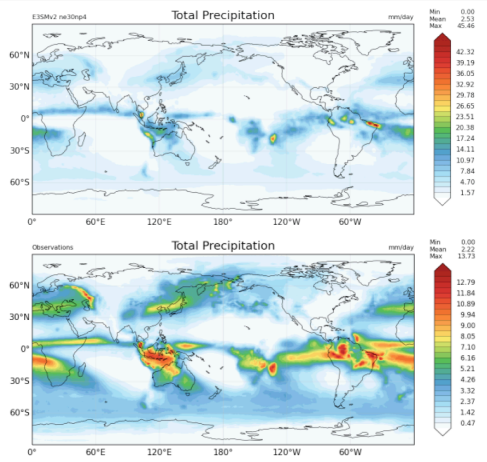
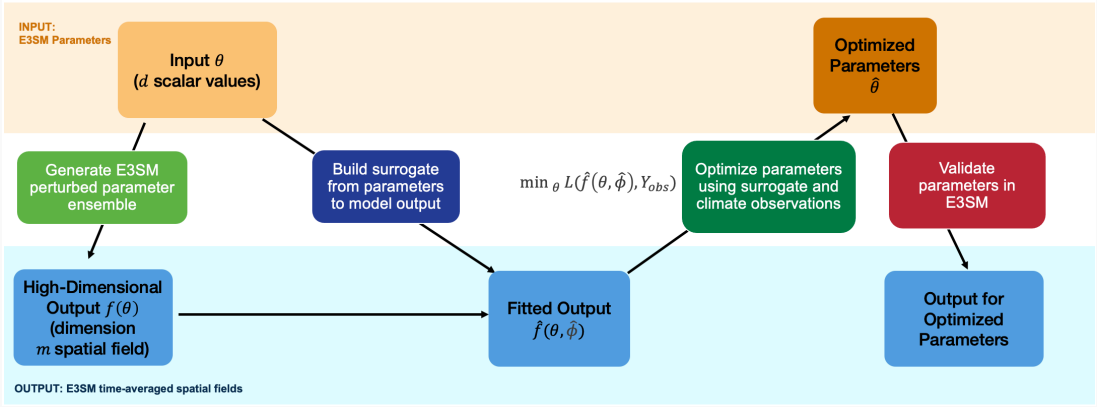


Figure: Changes in 10 year averaged precipitation due to perturbations in five atmospheric E3SM parameters. The goal of calibration is to select these parameters such that the precipitation matches historical observations

Proposed methods in literature

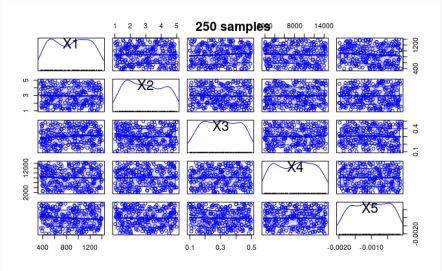
- In the statistical literature, the classical proposed solution is the calibration approach of **Kennedy and O'Hagan (2001)** in which they propose Gaussian Process (GP) emulators within a Bayesian framework to account for all known sources of uncertainty involved in model calibration.
- **Jackson et al. (2008)** propose a multiple very fast simulated annealing scheme (MVFSA) scheme to quantify uncertainty in climate model predictions.
- **Higdon et al. (2012)** demonstrates a multistep ensemble kalman filtering approach which also leverages GP emulators and accounts for model discrepancy on the Community Atmosphere Model (CAM).
- The 'Calibrate, Emulate, Sample" approach of **Cleary et al. (2022)** leverages GPs as emulators within an approximate Bayesian learning framework.
- In the climate modeling literature, the focus has been on using machine learning methods such as convolutional neural networks to build climate model emulators for multiple output variables (e.g. **Fletcher et al., 2022**).

Workflow Overview



Perturbed parameter ensemble for v2

- 250 simulation runs, with varying input parameters chosen by Latin Hypercube Sampling.
- Five input parameters changed: ice_sed_ai, clubb_c1, clubb_gamma_coef, zmconv_tau, zmconv_dmpdz

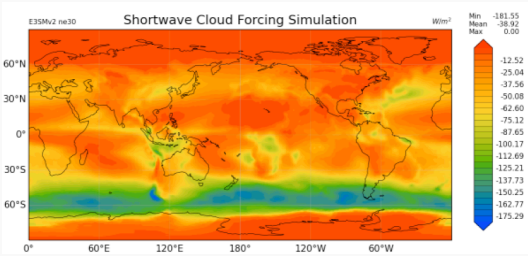


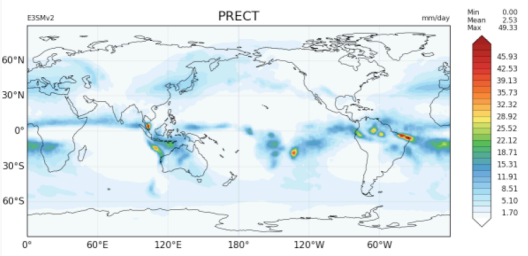
Run info:

- v2 of E3SM
- Prescribed present-day sea surface temperature (SST)
- 10 year runs: then average variables over time for each season DJF, MAM, JJA, SON

High-dimensional output fields

- 8 latitude/longitude fields
  - SWCF, LWCF, PRECT, PSL, Z500, U200, U850, TREFHT
  - consider either 24x48 and 180x360 resolution
  - weight to take into account area represented by each grid point
- 3 latitude/pressure level fields
  - U, RELHUM, T
  - 24x37 resolution
- Consider seasonal target fields: one for each DJF, MAM, JJA, SON
- Also consider globally averaged RESTOM value
- Results in 45 fields





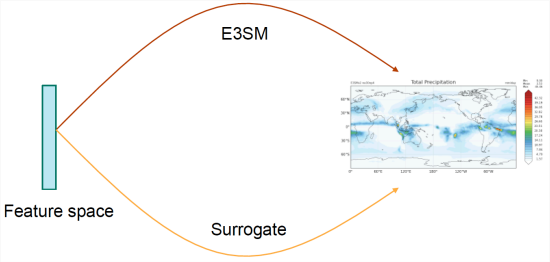
The use of a surrogate model

Advantages

- Surrogates will be able to enable real-time, fast evaluation of E3SM outputs (fractions of a second versus a few days for a single low resolution simulation)
- With a fast surrogate, we can evaluate gradients, perform parameter optimization, Bayesian inference, etc.

Disadvantages

- Generating data for training the surrogates is non-trivial and computationally expensive
- Surrogate is only as good as the training data it is built with



Some Notation

$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$: d-dimensional variable describing the input parameters

$f(\boldsymbol{\theta}) \in \mathbb{R}^m$: $m-$dimensional time-averaged spatial fields (target outputs of the ESM)

$\boldsymbol{Y}_{obs} \in \mathbb{R}^m$: observations of target fields

$\boldsymbol{X} \in \mathbb{R}^{n \times d}$: sampled input values from $\boldsymbol{\theta}$

$\boldsymbol{Y} \in \mathbb{R}^{n \times m}$: predicted target fields for given inputs $\boldsymbol{X}$

Goal

The goal of autotuning is to find values for the vector of E3SM parameters $\boldsymbol{\theta}$ that minimize a loss function $L(\cdot, \cdot)$ between $f(\boldsymbol{\theta})$ and an m-dimensional vector of target observations $\boldsymbol{Y}_{obs}$, i.e. $L(f(\boldsymbol{\theta}), \boldsymbol{Y}_{obs})$.

Substituting $f(\boldsymbol{\theta})$ with a surrogate model, $\hat{f}(\boldsymbol{\theta}, \boldsymbol{\phi})$, we can write the solution as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} L(\hat{f}(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}), \boldsymbol{Y}_{obs}). \tag{1}$$

$\boldsymbol{\theta}$ refers to the input parameters

$\boldsymbol{\phi}$ represents parameters of the surrogate model $\hat{f}$ itself

Surrogate Model

We construct our surrogate on a reduced space through PCA,

$$Y \approx \sum_{j=1}^{k} \boldsymbol{\eta}_j \boldsymbol{\psi}_j^\top,$$

and fit a surrogate model, $\hat{f}_j(\boldsymbol{\theta}, \boldsymbol{\phi}_j),\ j = 1, \ldots, k$, to to each of the k projection coefficients, $\boldsymbol{\eta}_j$ using regularized polynomial-based regression models such that our surrogate can be written as

$$\hat{f}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_{j=1}^{k} \hat{f}_j(\boldsymbol{\theta}, \boldsymbol{\phi}_j) \boldsymbol{\psi}_j. \tag{2}$$

Surrogate Fitting

$$\hat{\phi}_j = \min_{\boldsymbol{\phi}_j} \sum_{i=1}^{n} ((\boldsymbol{\eta}_j)_i - \hat{f}_j(\boldsymbol{X}_{i\cdot}, \boldsymbol{\phi}_j))^2 + \lambda \mathrm{Pen}(\boldsymbol{\phi}_j)$$

Through cross-validation, we automate the learning of the following

- surrogate parameters ϕ_j
- Polynomial order (up to order 12)
- Penalty Form: linear (λPen(ϕ_j)=0), lasso or elastic net
- regularization parameter λ for lasso or elastic net

Optimization

Let $Y_{p,\ell}$ denote the entry of $\boldsymbol{Y}_{obs}$ corresponding the the p-th target field and the ℓ grid point of that target field. The normalized likelihood for this output can then be written as

$$\frac{Y_{p,\ell}}{\sigma_p} \overset{ind}{\sim} \mathcal{N}\left(\frac{\hat{f}_{p,\ell}(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}})}{\sigma_p}, \frac{s_p^2}{w_{p,\ell}}\right) \tag{3}$$

where $\{w_{p,\ell}\}_{\ell=1}^{m_p}$ are area weights on grid points for the p-th output field, and $\sigma_p^2 = \mathrm{Var}_\ell(Y_{p,\ell})$ is the variance of the observational spatial field for the p-th output field.

Optimization

We can construct our Gaussian likelihood-based loss function through the joint log-likelihood over all target fields:

$$\mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{s}^2, \hat{\boldsymbol{\phi}}, \boldsymbol{Y}_{obs}\right) \propto \sum_p \left(-\frac{e_p(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}, \boldsymbol{Y}_{obs})}{s_p^2} + m_p \log\left(s_p^2\right) \right) \tag{4}$$

where $\propto$ means "proportional to" and $e_p(\cdot)$ is the weighted mean-squared error (MSE) for variable p,

$$e_p(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}, \boldsymbol{Y}_{obs}) = \sum_{\ell=1}^{m_p} w_{p,\ell} \frac{\left(\hat{f}_{p,\ell}(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}) - Y_{p,\ell}\right)^2}{\sigma_p^2}.$$

Optimization

The maximum a posteriori (MAP) estimate is then defined as

$$\hat{\boldsymbol{\theta}}_{MAP}, \hat{\mathbf{s}}^2_{MAP} = \underset{\boldsymbol{\theta}, \mathbf{s}^2}{\arg\max}\, g\left(\boldsymbol{\theta}, \mathbf{s}^2 \middle| \hat{\boldsymbol{\phi}}, \mathbf{Y}_{obs}\right).$$

$$g\left(\boldsymbol{\theta}, \mathbf{s}^2 \middle| \boldsymbol{\phi}, \mathbf{Y}_{obs}\right) = \mathcal{L}\left(\boldsymbol{\theta}, \mathbf{s}^2, \boldsymbol{\phi}, \mathbf{Y}_{obs}\right) + \sum_p \log\left(\mathcal{P}\left(s_p\right)\right).$$

We assign inverse gamma priors $\mathcal{P}$ to variance terms s_p^2 and uniform priors to $\boldsymbol{\theta}$.

Results: optimized parameters

Input Parameter	v2 control	Autotuned	Minimum	Maximum
ice_sed_ai	500.00	1400.00	350.00	1400.00
clubb_c1	2.40	1.00	1.00	5.00
clubb_gamma_coef	0.120	0.312	0.100	0.500
zmconv_tau	3600.00	4787.46	1800.00	14400.00
zmconv_dmpdz	-0.00070	-0.00042	-0.00200	-0.00010

Minimum and maximum specify the range considered for the Autotuned parameters.

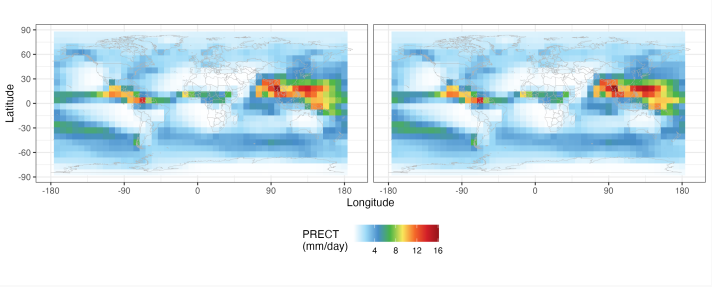
Results: RMSE Improvement

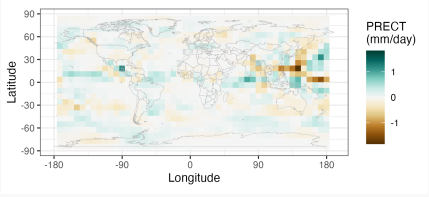
Comparison of autotuned and v2 default parameters: percentage change in root-mean-squared-error (RMSE) between time-averaged E3SMv2 output and observation. Green represents improvements when using the autotuning parameters.

Variable	DJF	MAM	JJA	SON	Avg.
LWCF	9.7	−1.3	0.4	10.0	4.7
PRECT	9.5	4.1	−0.3	11.8	6.3
PSL	4.3	−6.9	−5.3	−18.0	−8.6
RELHUM	−1.7	0.3	1.9	0.4	0.2
SWCF	5.1	−0.3	−6.2	2.0	0.1
T	−0.3	−3.3	1.9	−4.0	−1.4
TREFHT	−7.2	−10.0	−2.5	−10.3	−7.5
U	1.4	−10.6	−6.7	−10.8	−6.7
U200	7.4	−12.8	−18.0	−7.3	−4.0
U850	5.7	−11.8	−16.1	0.7	−5.4
Z500	4.0	−9.8	−7.1	−15.0	−2.7
Average	2.7	−5.7	−5.3	−2.4	−2.7

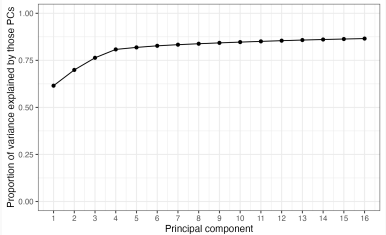
Results

E3SM and surrogate-predicted output for time-averaged PRECT during JJA, plotted on a 24x48 grid. (Top Left) E3SM output, (Top Right) surrogate prediction, (Bottom) the difference between the E3SM output and the surrogate prediction.

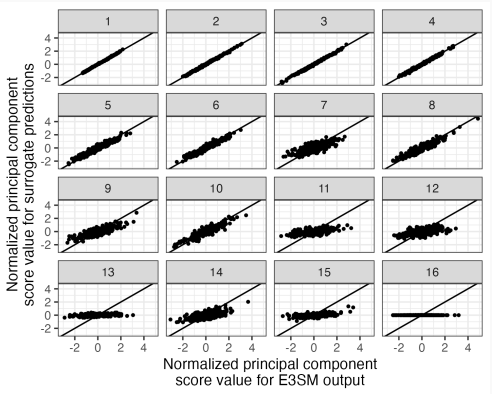




Results



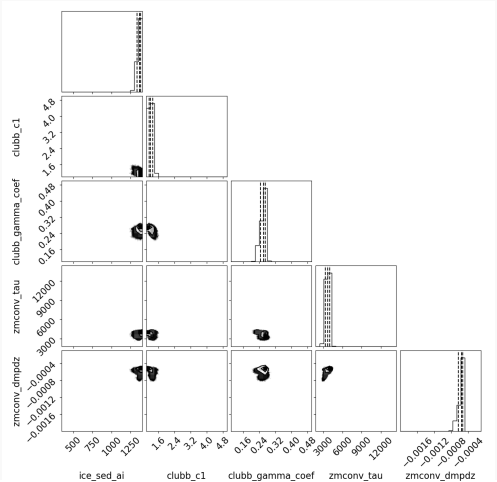
The cumulative proportion of variance in the data explained by the first k principal components, for $k = 1, 2, \ldots, 16$.

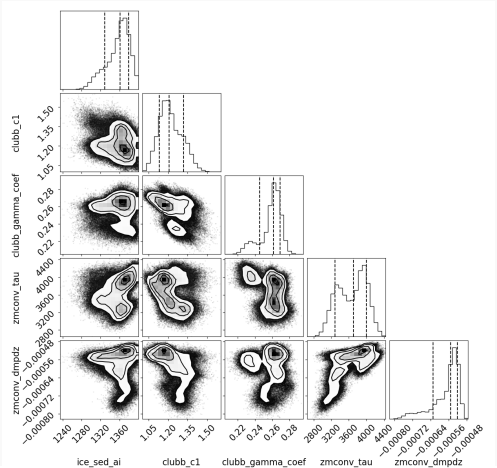


Principal component (PC) score/coordinate values for each principal component, comparing the simulation output and the surrogate predicted fields. The line y=x is plotted in each plot.

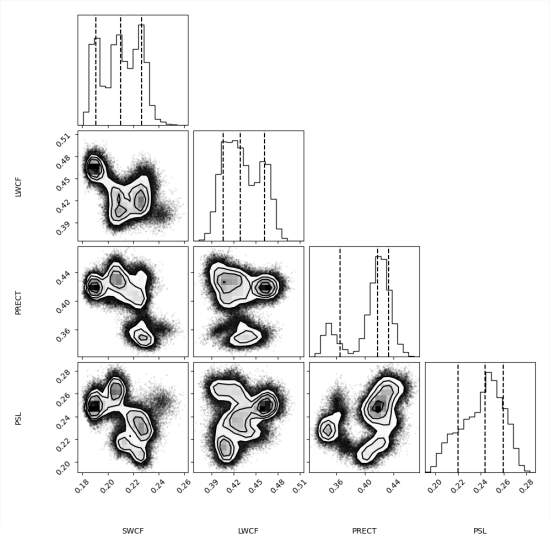
Exploration of posterior distribution of $\boldsymbol{\theta}$

(Left) Searched parameter bounds (Right) Zoomed in

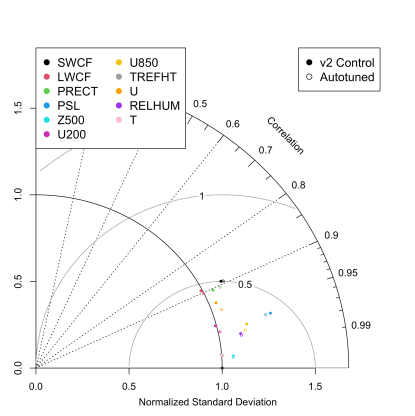


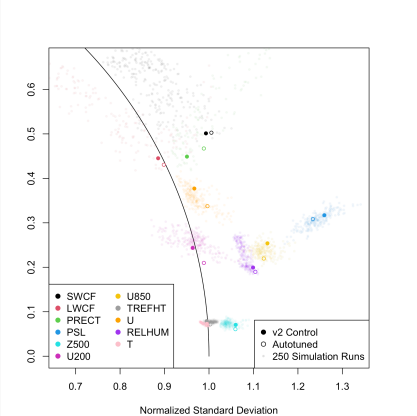


Exploration of posterior distribution of s_j^2 for DJF



Taylor Diagram



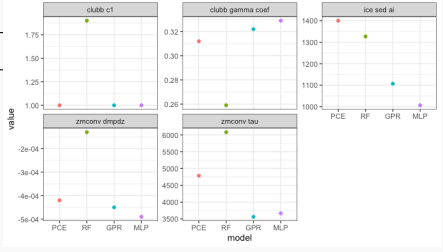


The Taylor diagram visualizes the components of the centered mean-squared error between the model run and observations in one plot with polar coordinates.

Surrogate model choice

Comparison of surrogate models using cross-validated scores: polynomial chaos expansion (PCE), random forest (RF), Gaussian process regression (GPR), and multilayer perceptron (MLP). For each surrogate model, we run on a single node in Chrysalis.

Name	R-squared	RMSE	MAE	Time (mm:ss)
PCE	0.478	7.36	5.04	07:42
RF	0.444	7.60	5.19	45:01
GPR	0.468	7.45	5.08	06:52
MLP	0.466	7.46	5.10	42:45



Summary

- We present an intuitive approach to autotune GCMs and demonstrate on E3SM.
- This approach produces an improved set of tuning parameters over the tuning parameters chosen by the expert while considering a large number of output fields.
- Approach allows us to flexibly weight some output fields more than others if needed to obtain alternative solutions.
- Can easily interchange surrogate choices, loss functions, etc.

Further Research

- Choice of Loss Function
- Iterative approach to sampling ensemble members - ongoing research in this area at Sandia
- surrogate model choice - have only scratched the surface here

  - PCE not easily implemented within a fully Bayesian framework

  - currently exploring promising options: Bayesian Adaptive Smoothing Splines (BASS) and Bayesian Adaptive Regression Trees (BART)
- Bayesian Model Averaging (BMA) approaches to avoid model selection all together
- E3SM is ultimately a multi-step procedure for the coupled model. Can this method easily extend to the coupled model?
- Are there clear metrics that could ensure we are accurately capturing climate oscillations? e.g. AMOC, QBO (ongoing SciDAC project at Sandia)

References

Chowdhary, K. K., & USDOE. (2021, 10). tesuract v.1.0, version 1.0. Retrieved from https://www.osti.gov/biblio/1861275 doi: 10.11578/dc.20220405.

Fletcher, C. G., McNally, W., Virgin, J. G., & King, F. (2022). Toward efficient calibration of higher-resolution Earth system models. Journal of Advances in Modeling Earth Systems, 14, e2021MS002836.

Higdon, D., Gattiker, J., Lawrence, E., Jackson, C., Tobis, M., Pratola, M., Habib, S., Heitmann, K. & Price, S. (2013) Computer Model Calibration Using the Ensemble Kalman Filter, Technometrics, 55:4, 488-500

Jackson, C. S., Sen, M. K., Huerta, G., Deng, Y., & Bowman, K. P. (2008). Error Reduction and Convergence in Climate Prediction, Journal of Climate, 21(24), 6698-6709

Kennedy, M.C. and O'Hagan, A. (2001), Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63: 425-464.

Ma, P.-L. et al. (2022). Better calibration of cloud parameterizations and subgrid effects increases the fidelity of the E3SM Atmosphere Model version 1, Geosci. Model Dev., 15, 2881–2916.