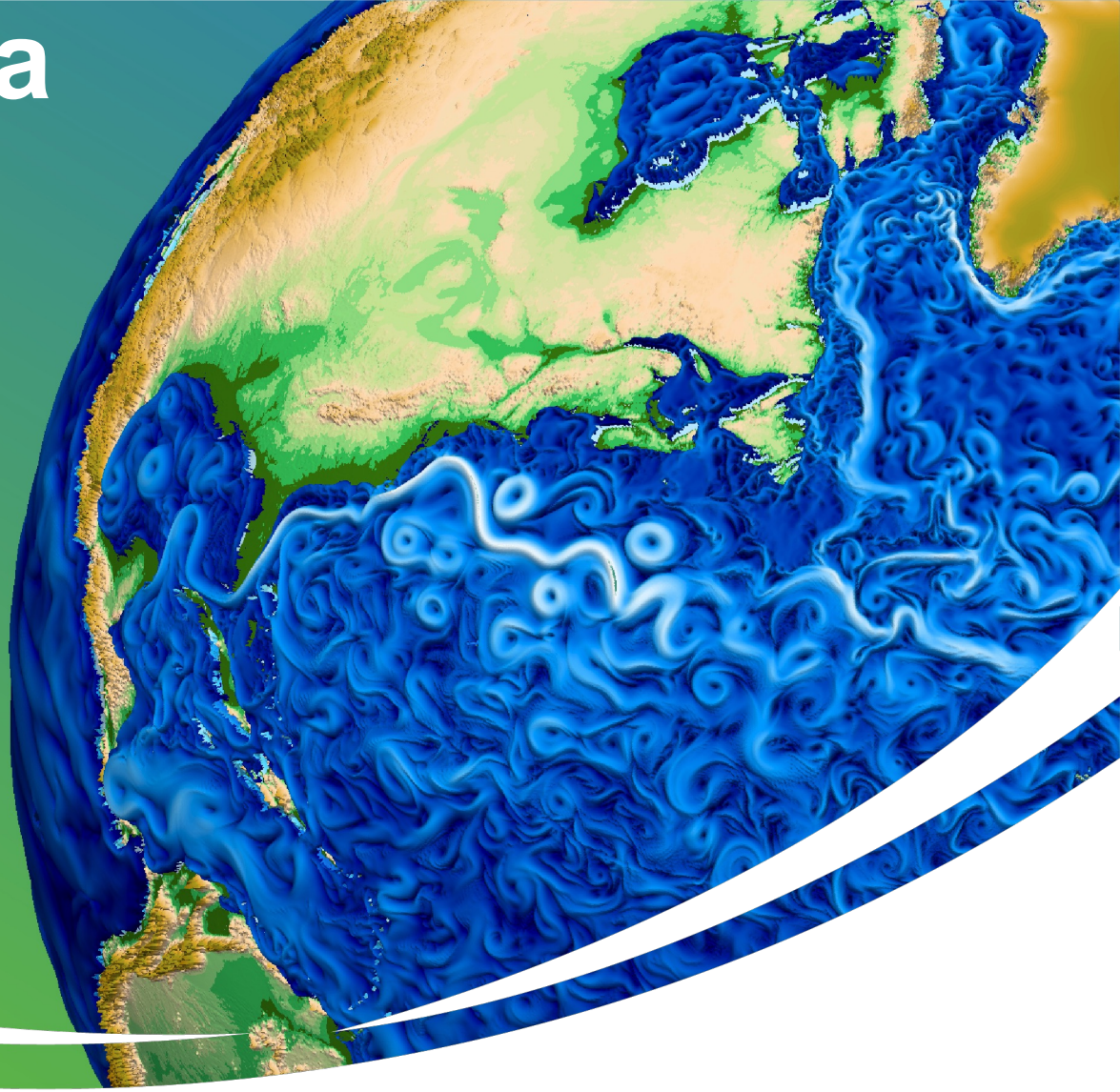
Autotuning E3SM using a surrogate model for climatological spatial fields

Drew Yarger, Benjamin Wagman, Lyndsay Shand, Andy Salinger
Sandia National Labs

Acknowledgments: Kenny Chowdhary, *NVIDIA;* Peter Caldwell, *LLNL*

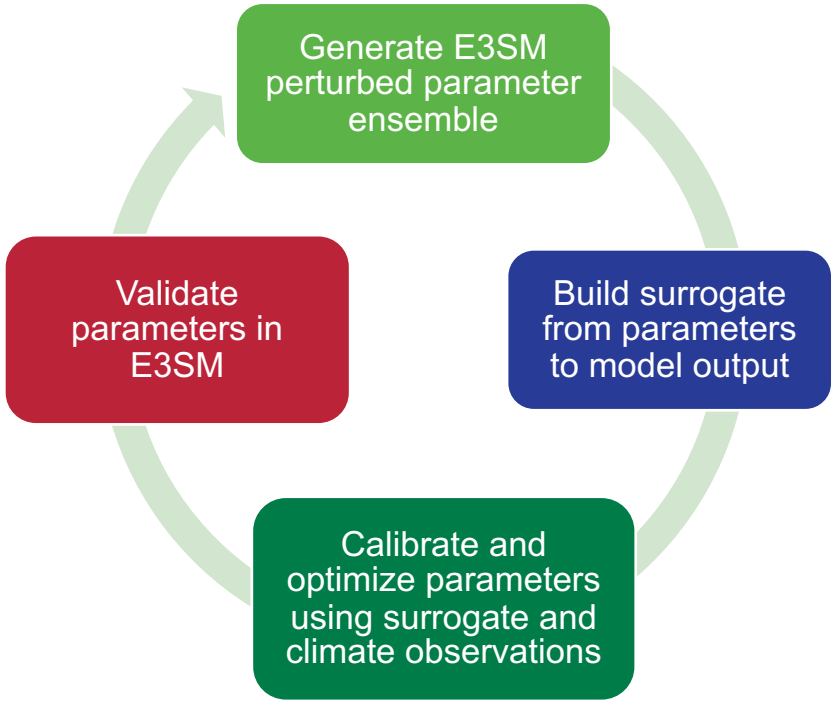


E3SM All-hands seminar
March 2, 2023

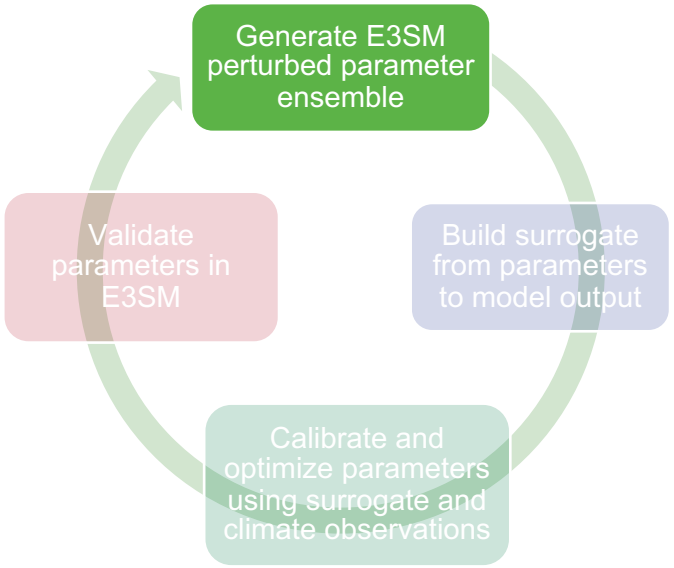
U.S. DEPARTMENT OF ENERGY

E3SM
Energy Exascale Earth System Model

Goal: Use machine learning to expedite and formalize E3SM tuning



E3SM PPE creation using Dakota[1]



Workflow

- Dakota
  - Reads user-defined `config_ensemble.yaml`
  - Does Latin hypercube sampling (LHS)
  - Calls `cime/scripts/create_clone`
  - Propagates LHS parameters to each clone's user_nl_eam
- Additional python scripts
  - Submit simulations in bundles
  - Post-process target fields using zppy

Code

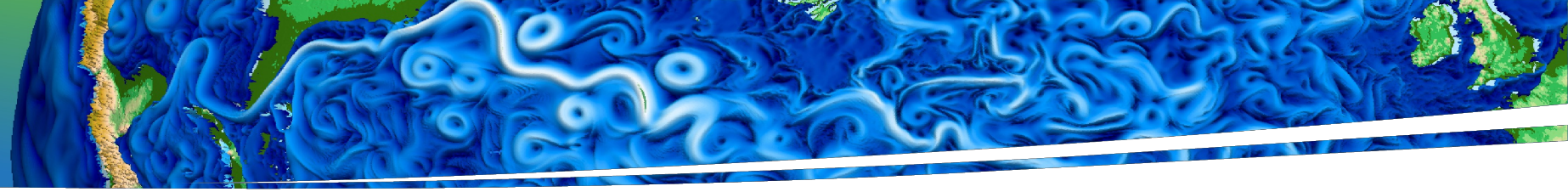
- `https://github.com/E3SM-Project/Autotuning-NGD.git`

[1] Adams et al., 2021

E3SM PPE's

Ensemble	Res.	Config.	N param.	N. simulations	N Yr.	Nodes per bundle	Sims. per bundle	SYPD per bundle	Machine	Complete
V2 ultra-low resolution (ULR)	~7.5°	F2010	5	500	10	100	50	5100	Chrysalis	✅
V2 Low resolution (LR)	~1°	**F2010**	**5**	**250**	**10**	**100**	**10**	**95**	**Chrysalis**	✅

Autotuning PPE's (E3SM confluence)



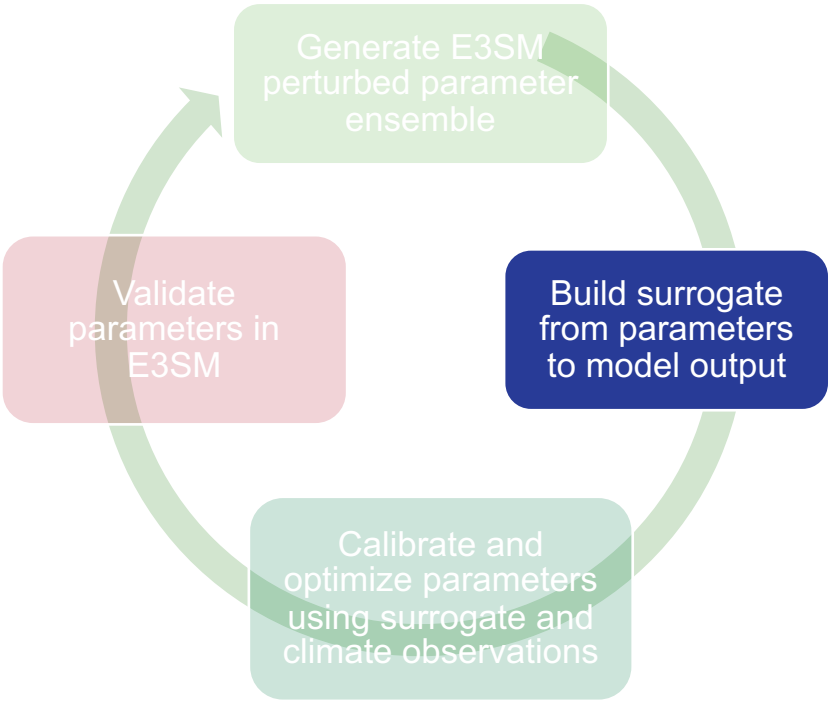
E3SMv2 sampled parameters

Parameter	Description	Low__Default__High [1]
clubb_c1	Constant for dissipation of variance of mean(w'^2)	1.0__**2.4**__5.0
clubb_gamma_coef	Constant of the width of PDF in w coordinate	0.1__**0.12**__0.5
zmconv_tau	Time scale for consumption rate deep CAPE	1800__**3600**__14400
zmconv_dmpdz	Parcel fractional mass entrainment rate	-2.0e-3__**-0.7e-3**__-0.1e-3
micro_mg_ai	Fall speed parameter for cloud ice	350__**500**__1400

[1] Ranges from Qian et al., 2018

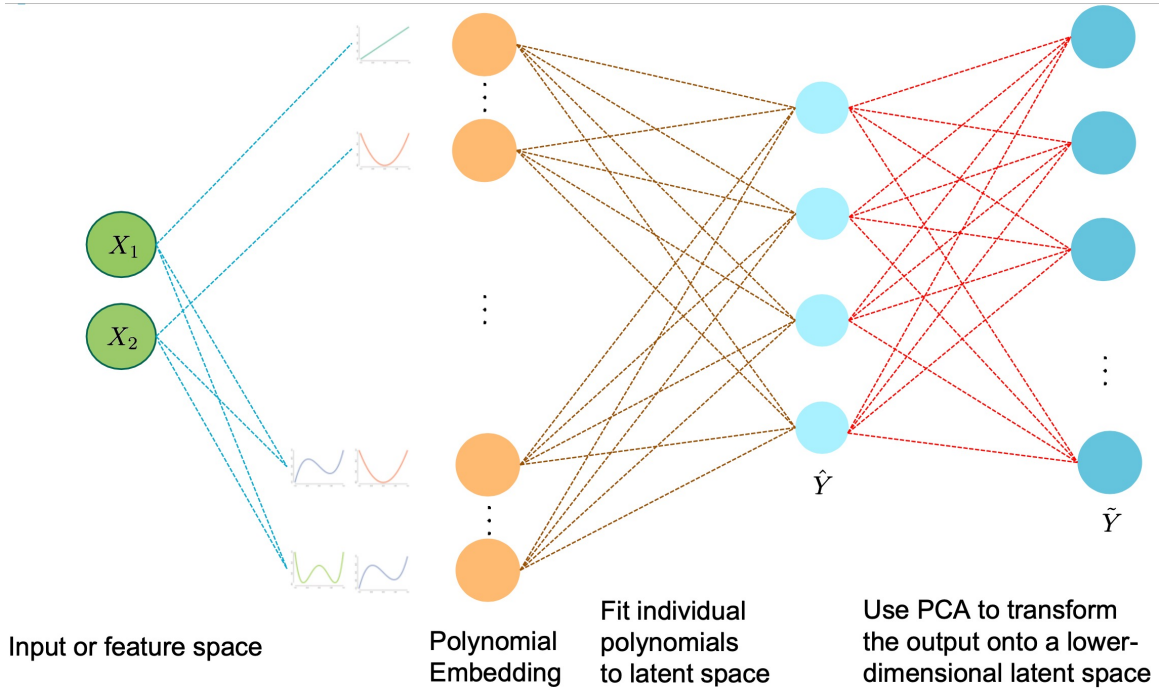
E3SMv2 target seasonal climatologies

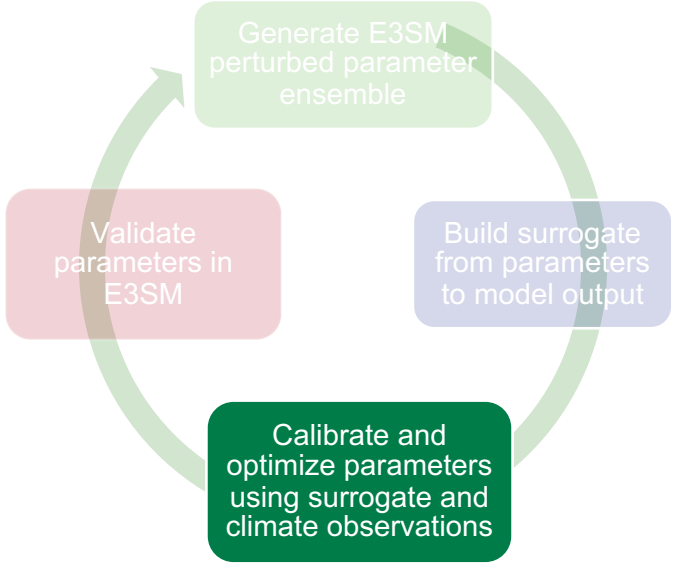
Field	Coordinates	Spatial resolution	Product
TREFHT	lat x lon	24x48, 180x360	ERAI
PRECT	lat x lon	24x48, 180x360	GPCP1DD
SWCF	lat x lon	24x48, 180x360	ceres_ebaf_toa
LWCF	lat x lon	24x48, 180x360	ceres_ebaf_toa
PSL	lat x lon	24x48, 180x360	ERAI
FLNT	lat x lon	24x48, 180x360	N/A
FSNT	lat x lon	24x48, 180x360	N/A
Z500	lat x lon	24x48, 180x360	ERAI
U200	lat x lon	24x48, 180x360	ERAI
U850	lat x lon	24x48, 180x360	ERAI
RELHUM	lat x lev	24x37, 180x37	ERAI
T	lat x lev	24x37, 180x37	ERAI
U	lat x lev	24x37, 180x37	ERAI


Generate E3SM perturbed parameter ensemble
Build surrogate from parameters to model output
Calibrate and optimize parameters using surrogate and climate observations
Validate parameters in E3SM

Surrogate function choice: polynomial chaos expansion

- Uses Python's scikit-learn framework and is based off Kenny Chowdhary's Autotuning 2020-2022 work at Sandia.
  - Python package tesuract: https://github.com/kennychowdhary/tesuract
- PCA and polynomial chaos expansion: similar or better prediction errors compared to
  - neural network,
  - random forest,
  - support vector machines,
  - Gaussian processes.
- Hyperparameters of polynomial chaos expansion chosen automatically by cross-validation.




Generate E3SM perturbed parameter ensemble
Build surrogate from parameters to model output
Calibrate and optimize parameters using surrogate and climate observations
Validate parameters in E3SM

Optimization strategy

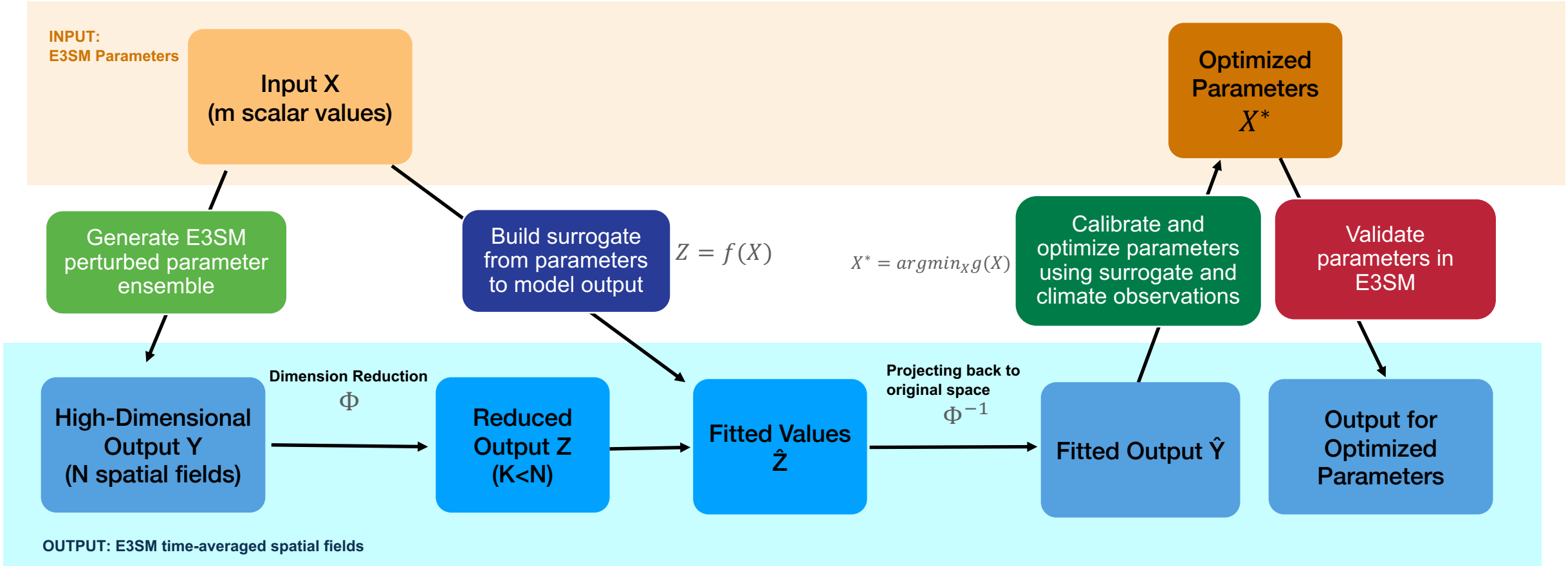
- We minimize the root-mean-squared-error (RMSE) between
  1. the surrogate predicted fields,
  2. the observational data.

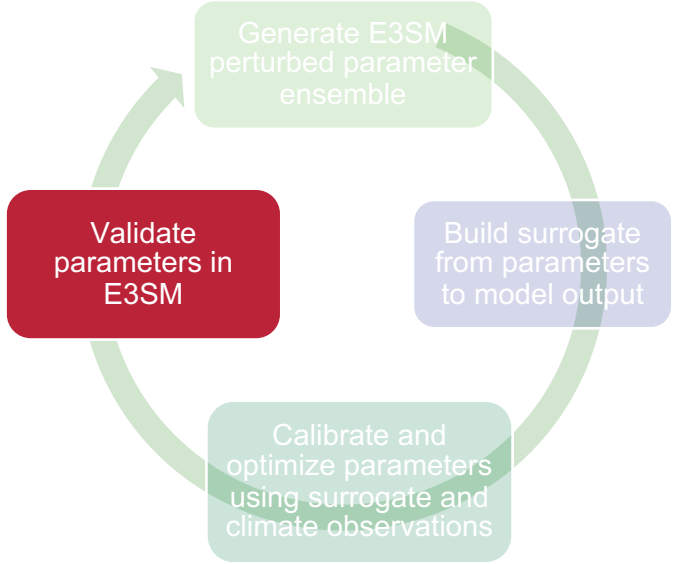
- Regularization terms that normalize the errors from different output fields
  - Any one output field does not dominate the optimization.
  - Allows Bayesian formulation of the normalization.

- Use L-BFGS-B[1] optimization algorithm on the minimization problem.
  - Can also use Markov Chain Monte Carlo (MCMC) sampling to visualize uncertainty in optimized parameters.

[1] Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, **16**, 1190–1208. doi:10.1137/0916069.

Current Workflow



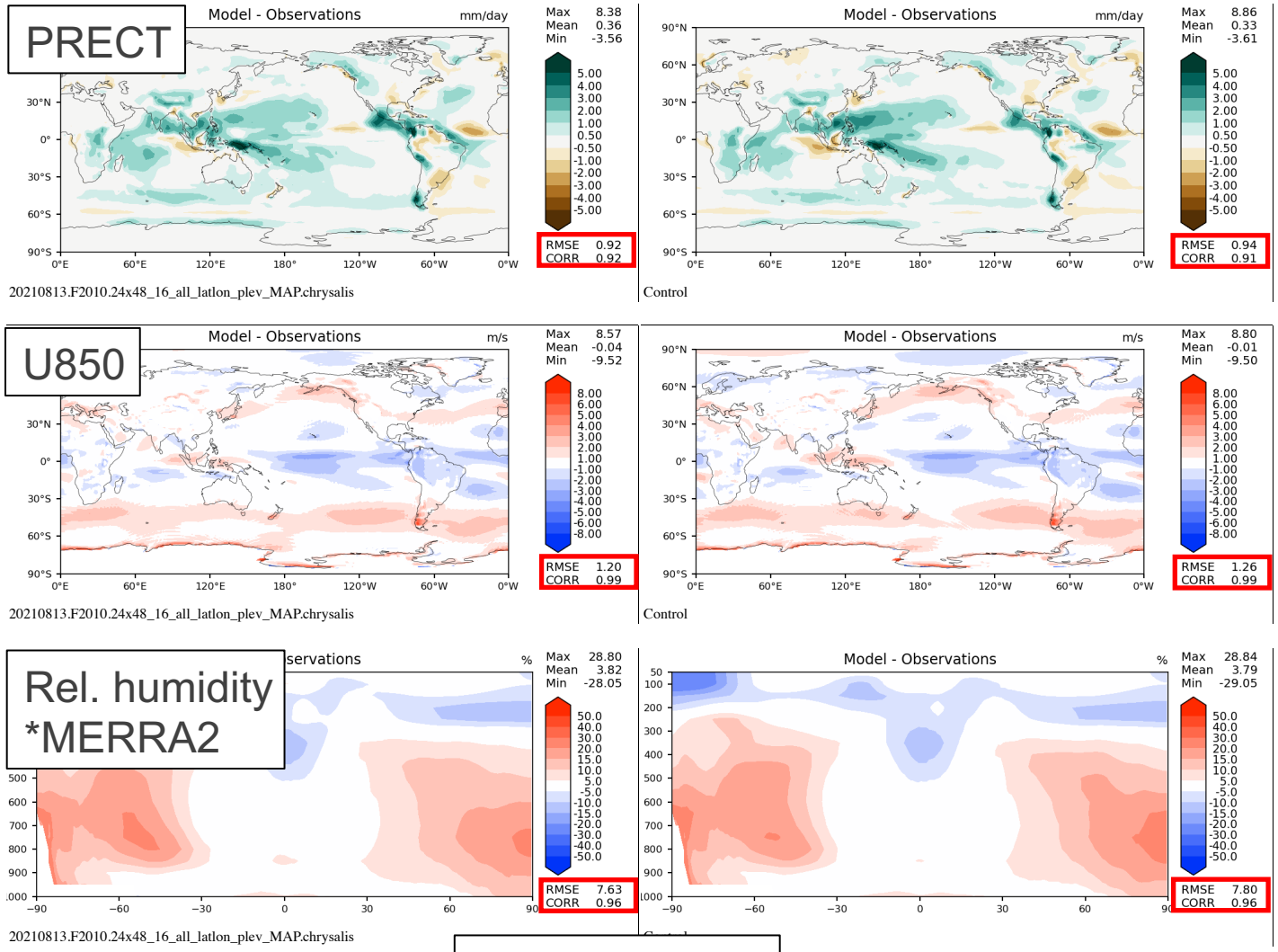

Generate E3SM perturbed parameter ensemble
Build surrogate from parameters to model output
Calibrate and optimize parameters using surrogate and climate observations
Validate parameters in E3SM

Autotuned E3SMv2 vs. v2 control: % Change in RMSE

	DJF	MAM	JJA	SON	Average
LWCF	4.5	-2.4	-1.9	2.9	0.8
PRECT	1.8	-2.8	-4.5	3.9	-0.4
PSL	-9.1	-7.1	-1.3	-8.5	-6.5
RELHUM	-0.0	-0.3	1.4	-0.5	0.1
SWCF	2.7	0.1	-3.5	-1.3	-0.5
T	-1.6	-2.8	2.4	-5.4	-1.9
TREFHT	-8.1	-13.8	-4.4	-6.0	-8.1
U	-7.9	-6.2	-0.4	-1.1	-3.9
U200	4.6	-10.7	-17.6	-4.0	-6.9
U850	1.2	-12.6	-14.8	-3.7	-7.5
Z500	-0.9	-15.1	-8.9	-6.8	-7.9
Average	-1.2	-6.7	-4.9	-2.8	-3.9

- Area-weighted root-mean-squared-error (RMSE) evaluation on
  - 180x360 (lat x lon) grid
  - 180x37 (lat x pressure) grid
- **Average of 3.9% decrease in RMSE across variables and seasons**

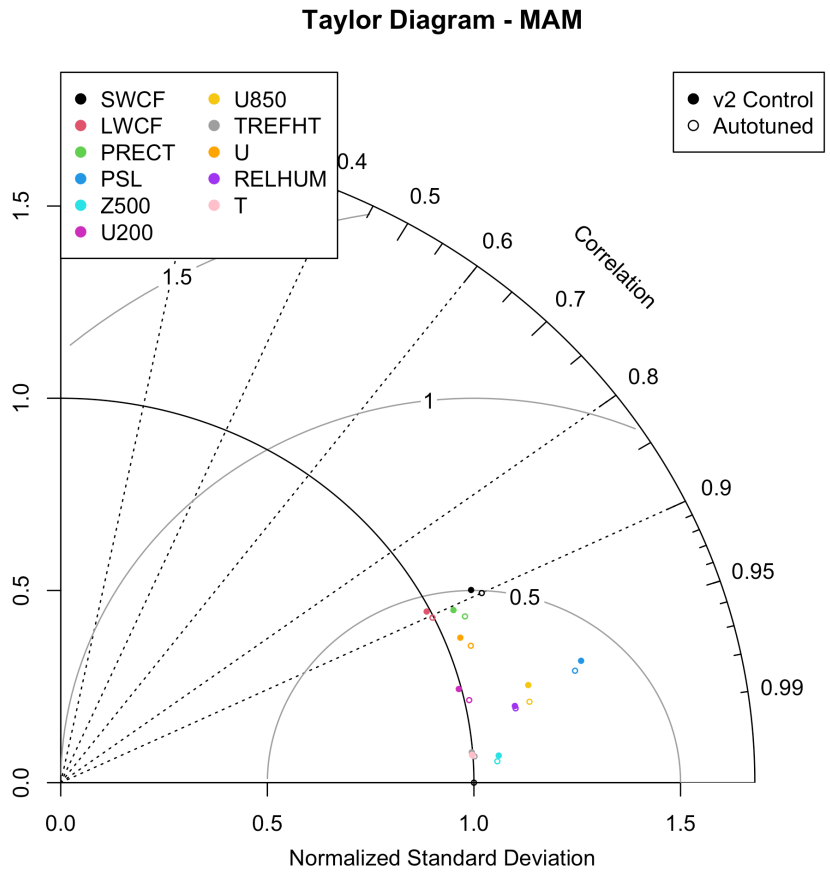
Autotuned Control

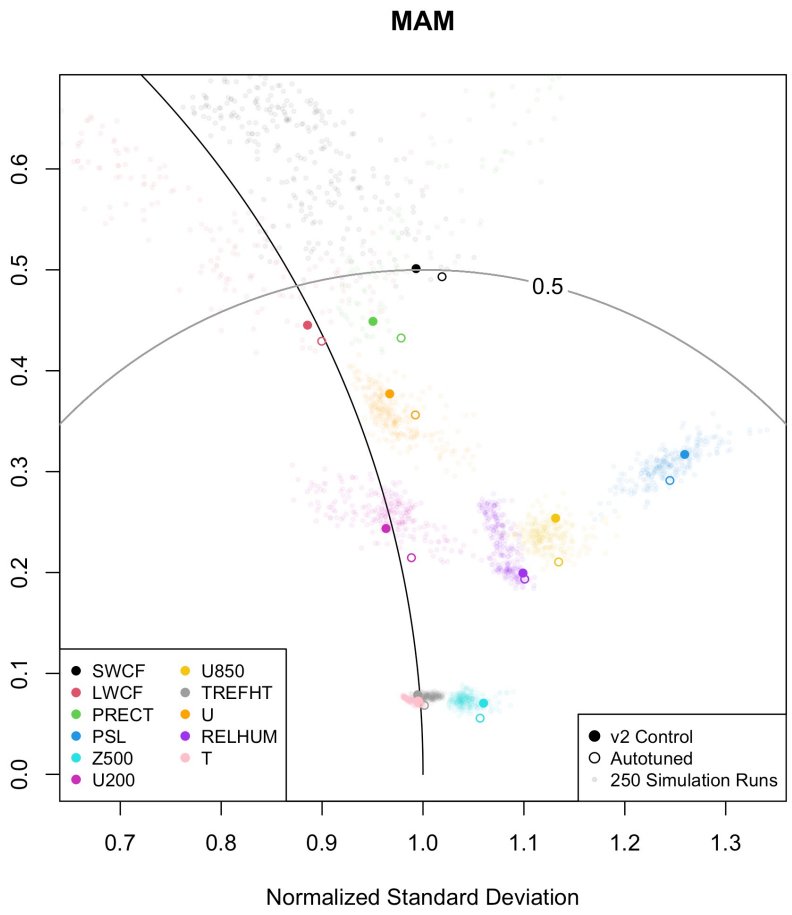


Annual means

- Autotuned improvements are evident in E3SM-diags.
  - IICE: Autotuned vs. Control.
- Improvements are generally not product dependent.
  - E.g. autotuning to RH from ERAI improves the simulation even when evaluated against MERRA2.
- Biases have the same spatial pattern in autotuned and control simulations.
  - Some biases are structural;
  - Other biases are sensitive to parameters that were not sampled.

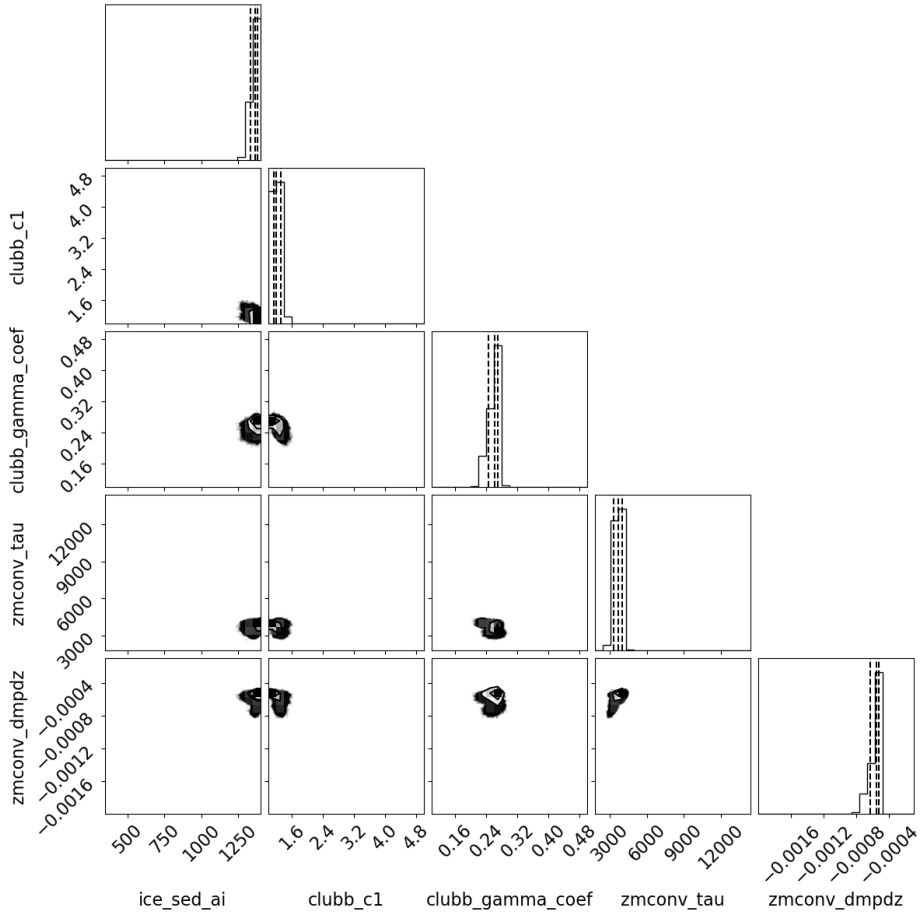
Taylor diagram

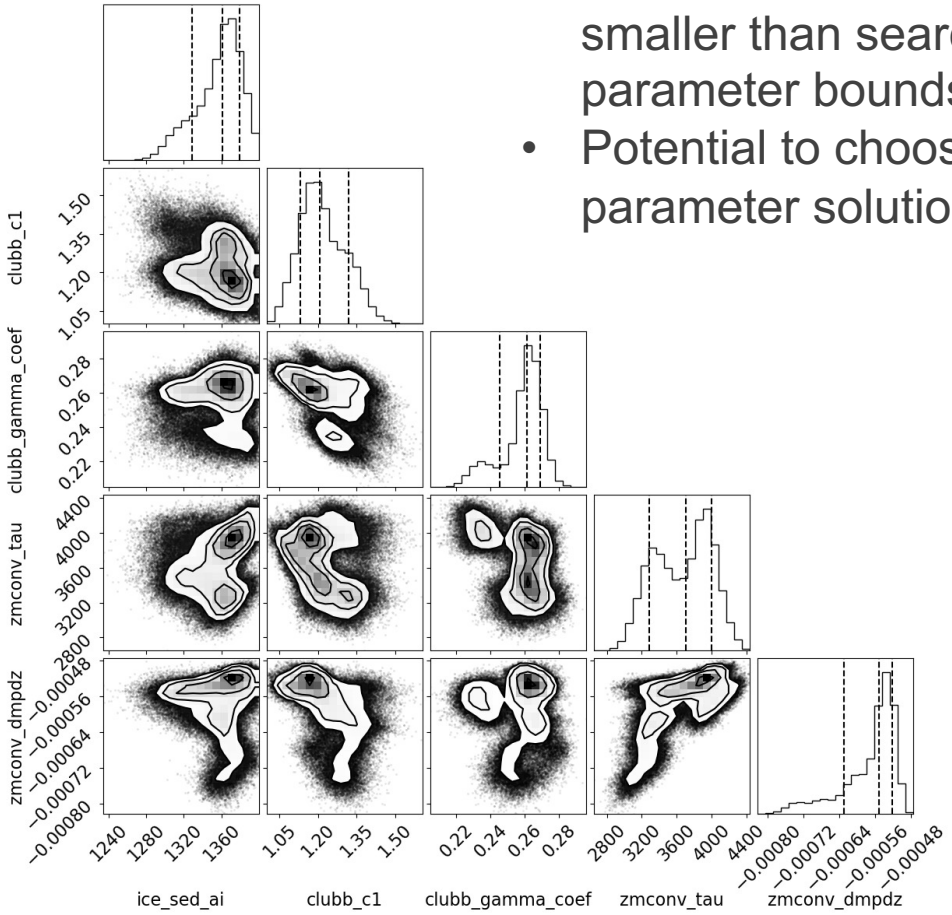




Calibration: Results

(Left) Searched parameter bounds (Right) Zoomed in





Preliminary MCMC results:

- Posterior distribution currently much smaller than searched parameter bounds
- Potential to choose multiple parameter solutions

Summary

- Created a flexible workflow for automated tuning and demonstrated on the E3SMv2 atmosphere model
  - Leverages a surrogate model that predicts high dimensional E3SM climatological spatial fields
  - Surrogate is independent of choice of loss function

- Autotuned "optimized" E3SMv2 reduces RMSE for 11 spatial fields and 4 seasons by an average of 3.9%.

- A Bayesian implementation provides a distribution of optimized tuning parameters.

Next: Auto-tuning E3SMv3

E3SMv3 PPE (March-April 2023)

- 8 tuning parameters
- ~200-300 10-yr F2010 simulations and 1-yr SST+4K simulations
- Machine: Chrysalis or Compy

Surrogate and optimization

- Loss function: user can override automatic field weighting with pre-specified weights

Deliverables

- Optimal set of tuning parameters for F2010 to assist E3SMv3 expert tuning
- Distribution of tuning parameters from MCMC sampling
- Alternative optimized parameters using different field weightings
- Alternative tunings with a range of climate sensitivities
- Potentially: adapt tools to tune other E3SMv3 configurations

Our PPE's

Confluence: Autotuning PPE's

Our software

E3SM Autotuning github: E3SM-Project/Autotuning-NGD

Tesuract (surrogate construction): https://github.com/kennychowdhary/tesuract

Other software

Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Hooper, R.W., Hough, P.D., Hu, K.T., Jakeman, J.D., Khalil, M., Maupin, K.A., Monschke, J.A., Ridgway, E.M., Rushdi, A.A., Seidl, D.T., Stephens, J.A., Swiler, L.P., and Winokur, J.G., "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.15 User's Manual," Sandia Technical Report SAND2020-12495, November 2021.

Additional slides

E3SM PPE's

Ensemble	Res.	Config.	N param.	N. simulations	N Yr.	Nodes per bundle	Sims. per bundle	SYPD per bundle	Machine	Complete
V2 ultra-low resolution (ULR)	~7.5°	F2010	5	500	10	100	50	5100	Chrysalis	✅
V2 Low resolution (LR)	~1°	**F2010**	**5**	**250**	**10**	**100**	**10**	**95**	**Chrysalis**	✅
V3 candidate LR	~1°	F2010	8	200-300	10	170?	10	?	Chrysalis or Compy	
V3 candidate LR	~1°	F2010 + 4K	8	200-300	1	170?	10	?	Chrysalis or Compy	

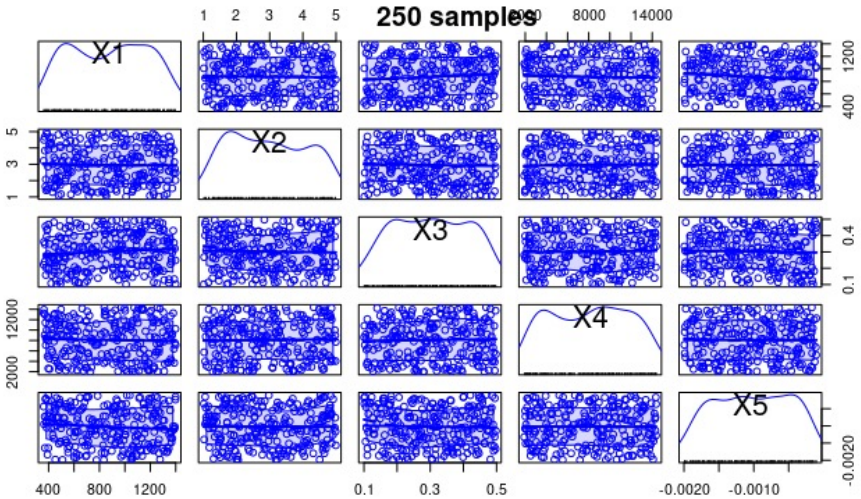
Autotuning PPE's (E3SM confluence)

Input parameters for E3SMv3 PPE

Scheme	Parameter	Description	Low-Def-High
Clubb	clubb_c1	Constant for dissipation of variance of mean(w'^2)	1.0__1.335__5.0
Clubb	clubb_gamma_coef	Constant of the width of PDF in w coordinate	0.1__0.32__0.5
ZM conv.	zmconv_tau	Time scale for consumption rate deep CAPE	1800__3600__14400
ZM conv.	zmconv_dmpdz	Parcel fractional mass entrainment rate	-2.0e-3__-0.7e-3__-0.1e-3
ZM micro	zmconv_micro_dcs	Autoconversion size threshold for cloud ice (to snow)	100.e-6__150.e-6__400e.-6
P3 Micro	nucleate_ice_subgrid	Subgrid parameter for ice nucleation	1.0__1.35__1.4 (Range from original developers) -consider 1.2__1.35__1.4 based on Wuyin's experience.
P3 Micro	p3_embryonic_rain_size	p3_embryonic_rain_size (micrometers)	15e-6__25e-6__40e-6
P3 Micro	p3_qc_accret_expon	Accretion qc/qr exponent. Accretion: growth of a precipitation particle by the collision of a frozen particle with a supercooled liquid water droplet which freezes upon impact.	1.0__1.15__2.0

Optimization: Results

Parameter	Description	Low-Default-High (autotuned)
clubb_c1	Constant for dissipation of variance of mean(w'^2)	1.0__**1.335**__5.0 (1.13)
clubb_gamma_coef	Constant of the width of PDF in w coordinate	0.1__**0.32**__0.5 (0.265)
zmconv_tau	Time scale for consumption rate deep CAPE	1800__**3600**__14400 (4049.74)
zmconv_dmpdz	Parcel fractional mass entrainment rate	-2.0e-3__**-0.7e-3**__-0.1e-3 (-0.49e-3)
micro_mg_ai	Fall speed parameter for cloud ice	350__**500**__1400 (360.25)



V2 Control parameters are likely influenced by output fields we have not considered here (climate sensitivity, El Niño, AMOC, coupling).

Calibration strategy (SWAP in from pdf or eliminate)

- Bayesian formulation for how different output fields are weighted in the optimization problem.

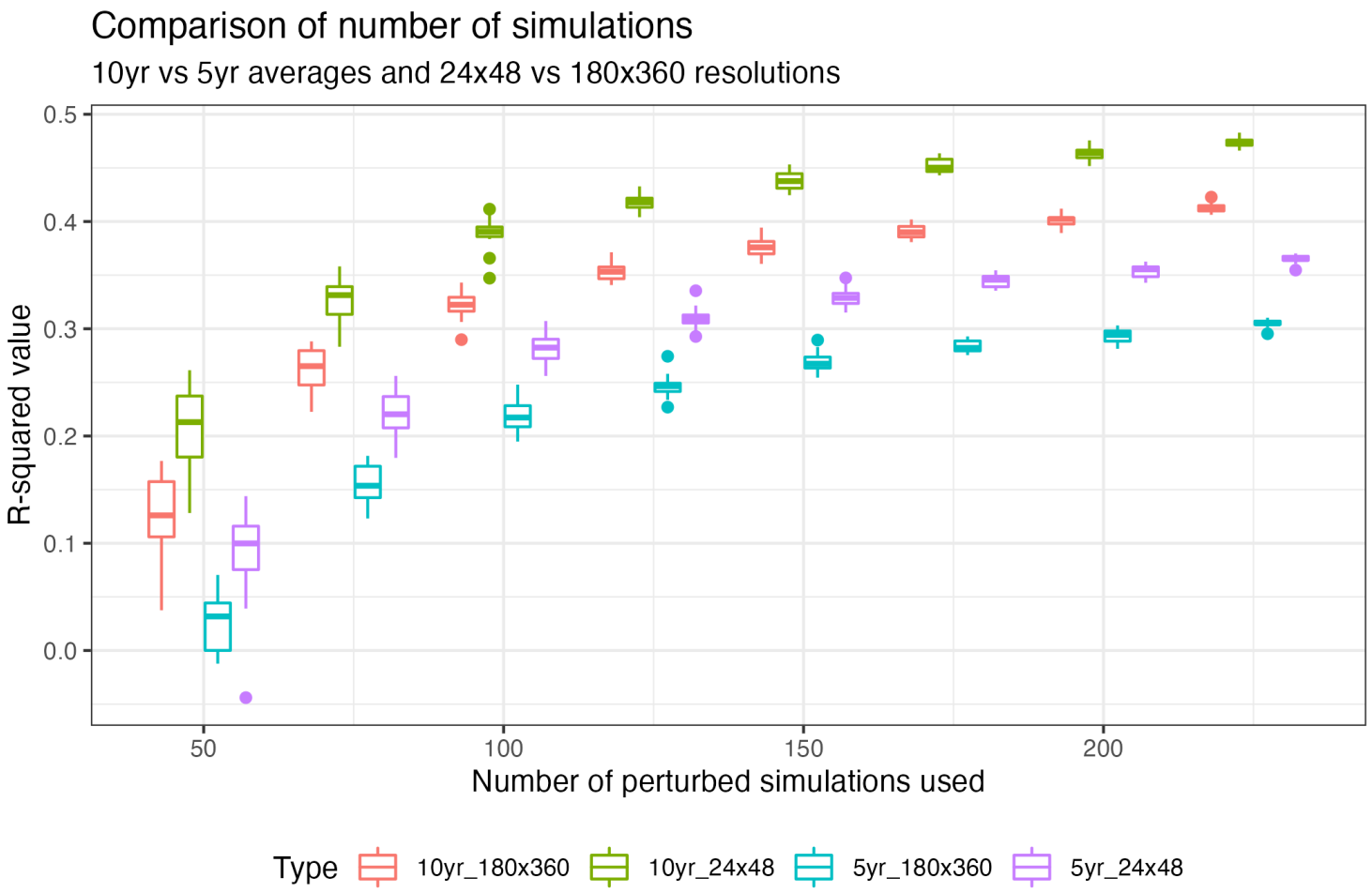
- Markov Chain Monte Carlo (MCMC) sampling

How many simulations do we need?

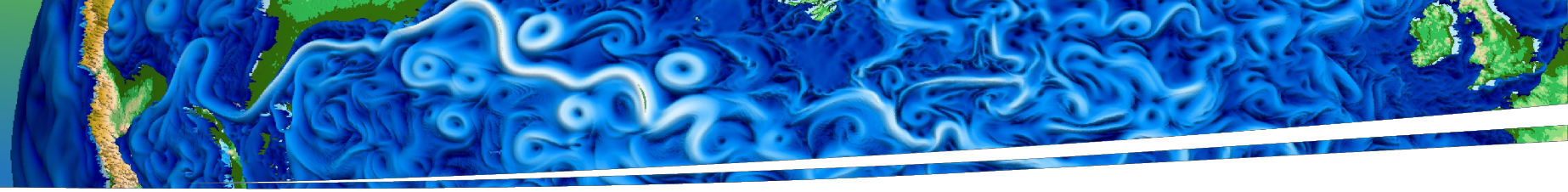
Randomly sample k = 50, 75, 100, …, 225 of the 250 perturbed parameter ensemble runs.

- Perform surrogate construction and optimization using only the sampled runs.
- Repeat with 20 different random samples for each k.
- Evaluate:
  - 5 year and 10 year time-averaged fields
  - 24x48 and 180x360 latitude/longitude resolution
- Gives very rough idea of sensitivity to the number of simulations used and the replicability of autotuned parameter sets.

How many simulations? Surrogate fit in terms of R^2



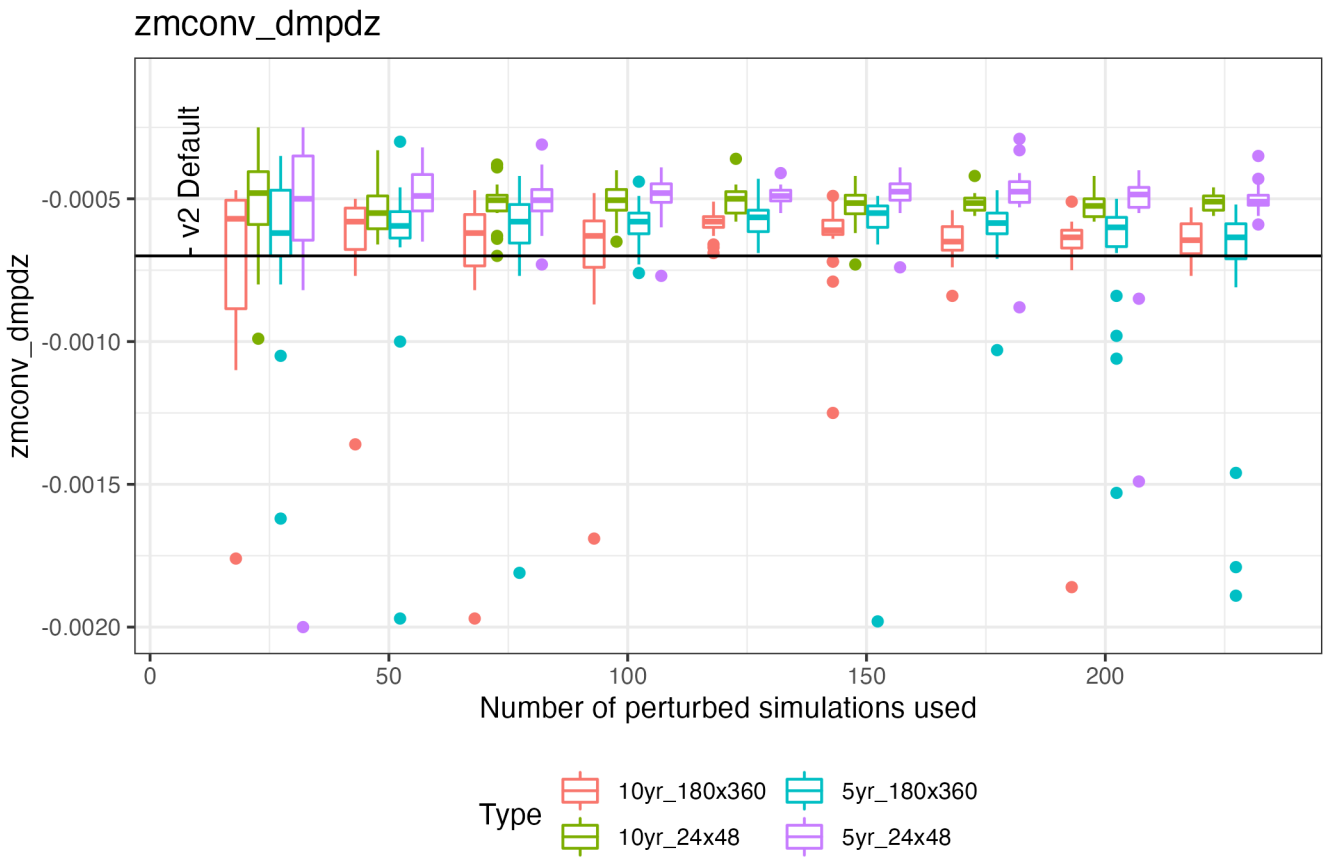
Based on this, we use 10 yrs, 24x48



Abstract

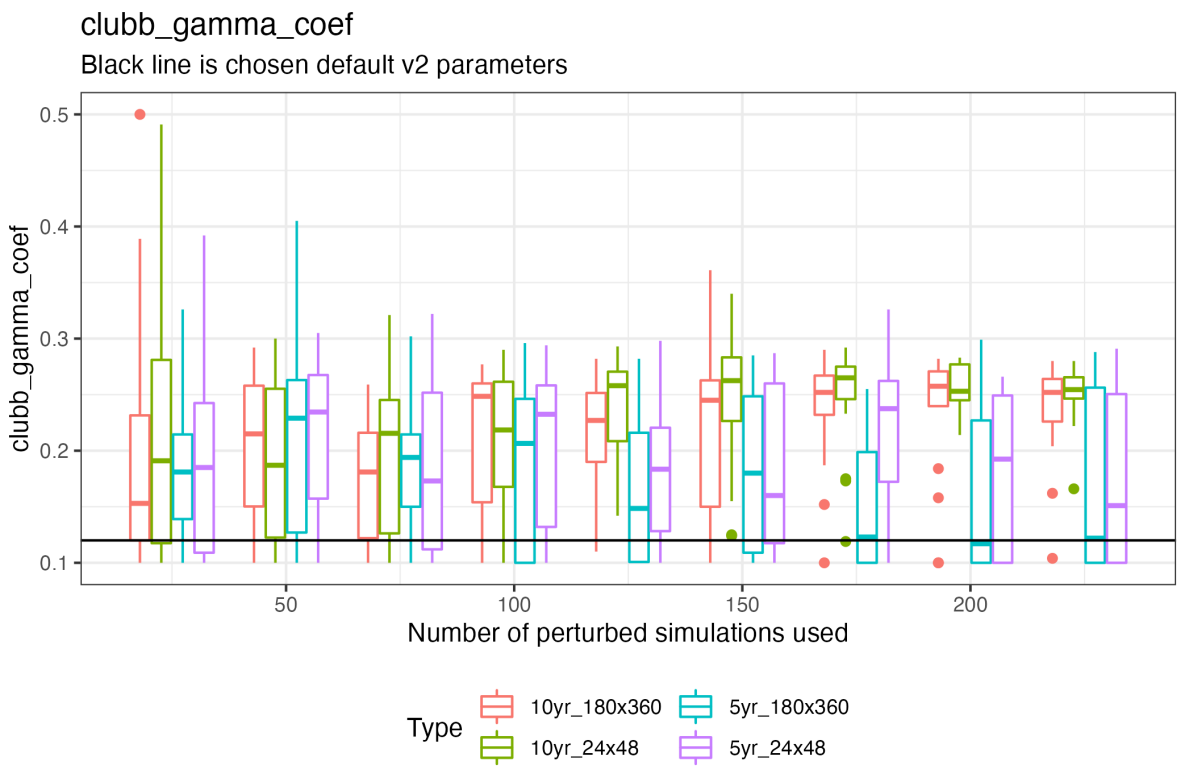
Tuning an Earth system model is a time-consuming effort traditionally led by subject matter experts. We are developing an automated tuning method that uses a perturbed parameter ensemble (PPE) and machine learning (ML)/uncertainty quantification (UQ)/optimization tools to select parameter values. We present the workflow and the results of our optimization for E3SMv2. The automated tuning method reduces the RMSE relative to the E3SMv2 release by 3.9% on average for a set of eleven spatial targets over four seasons. We also study the effects of PPE sample size, PPE climatology length (number of simulated years), and surrogate spatial resolution, to determine the most efficient methods for our next effort, the automated tuning of E3SMv3.

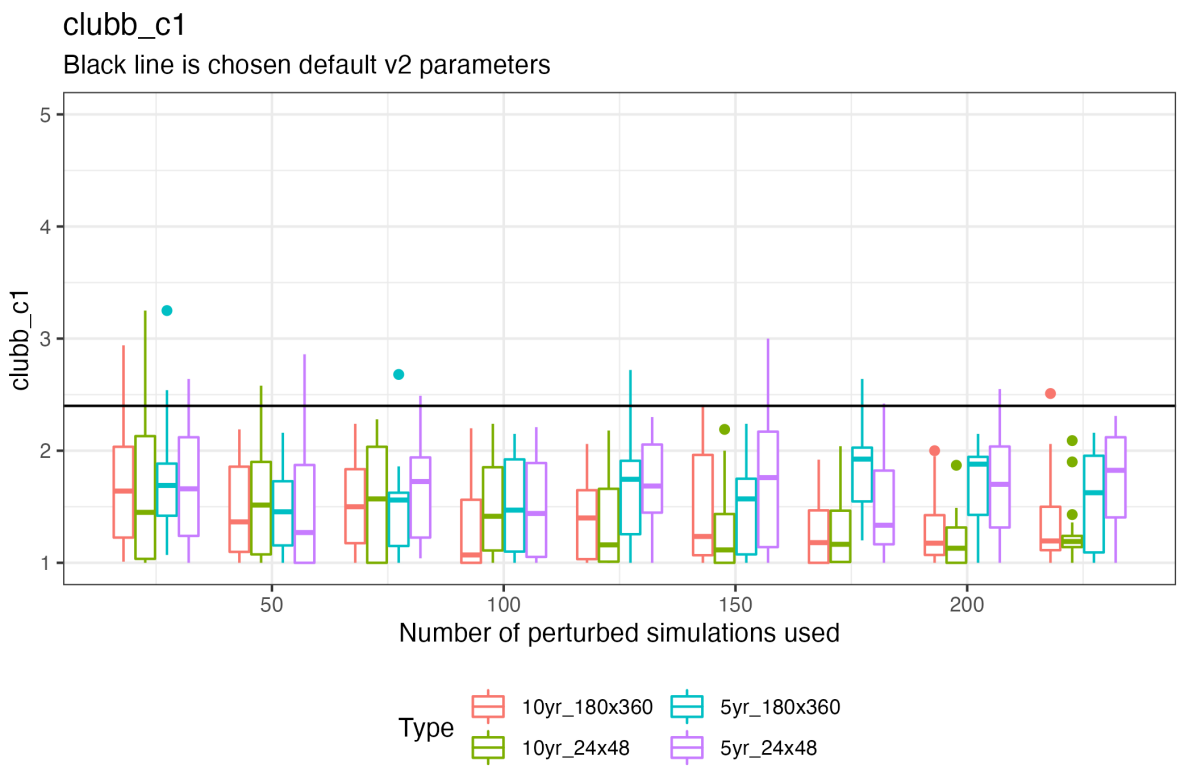
What resolution/sim length should we use for the surrogate?



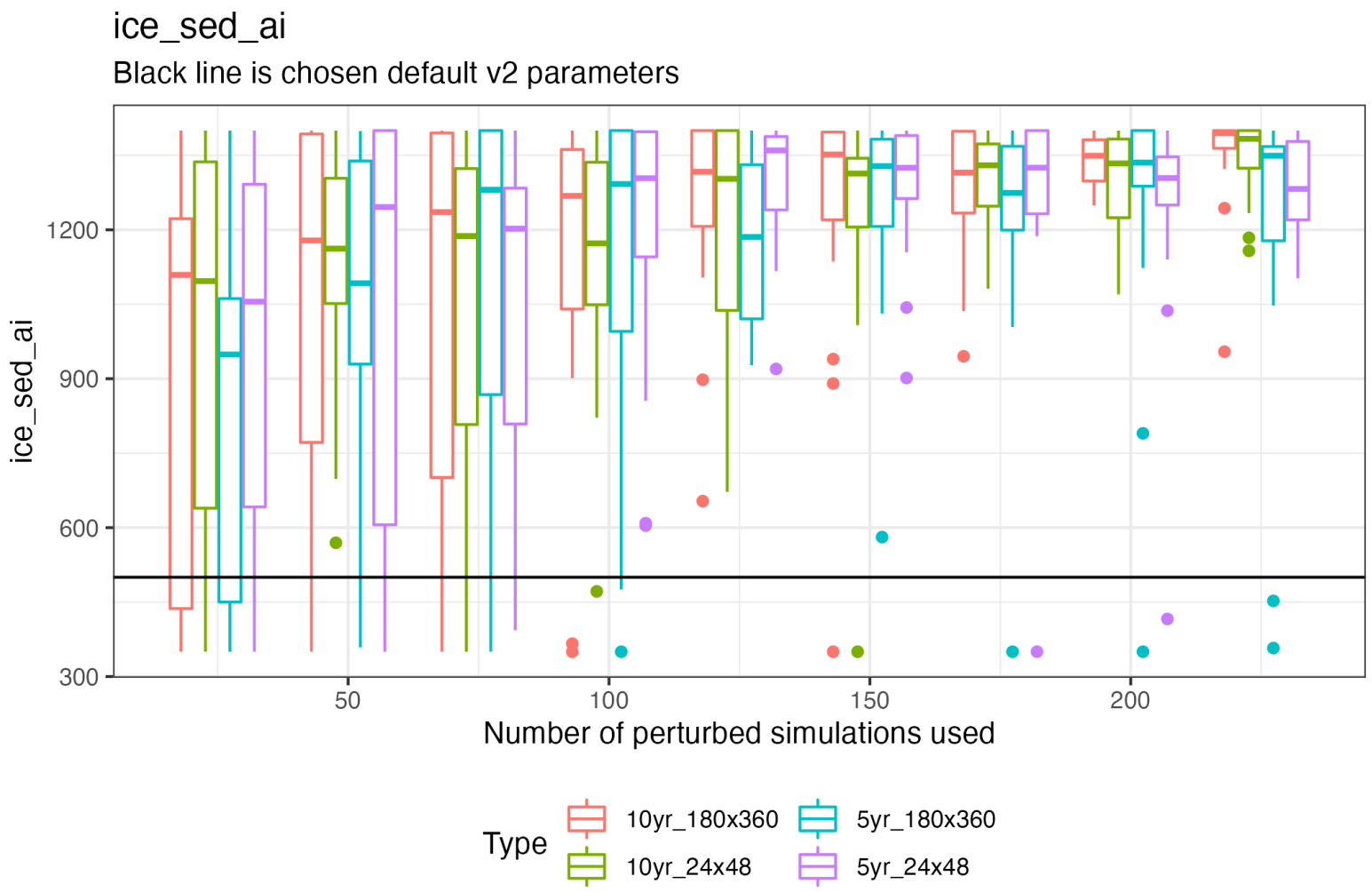
Based on this, we use 10 yrs, 24x48 to reliably beat the default

How many simulations? Optimized parameters





How many simulations? Optimized parameters



1. Generate Ensemble

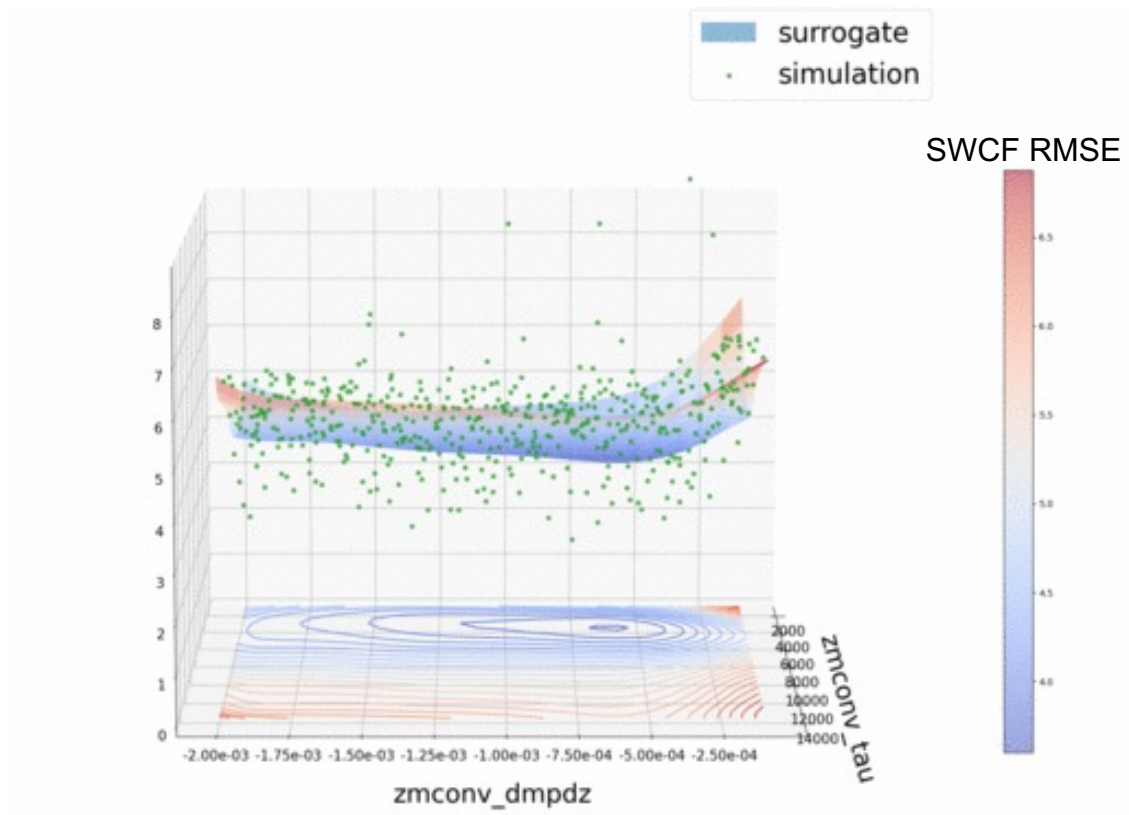


Figure: SWCF RMSE [Wm^{-2}] for autotuning ensemble (green dots) as a function of two deep convection parameters. Error computed with respect to default model tuning.

2. Build Surrogate Model

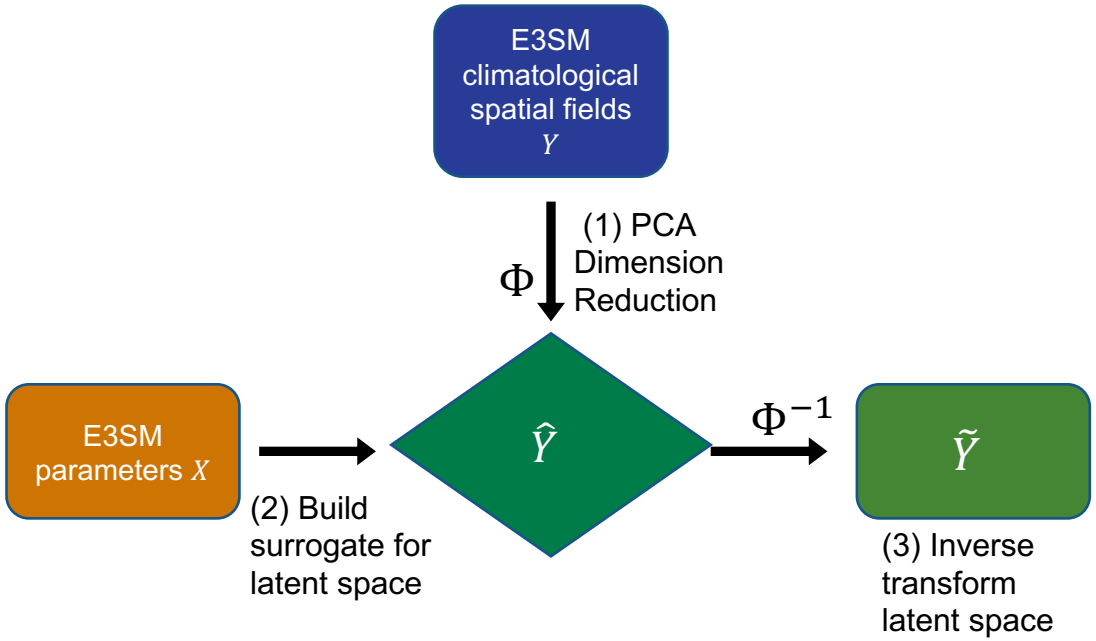


Figure: Surrogate model construction steps for spatial fields of E3SM output $\tilde{Y}$ as a function of E3SM uncertain parameters X.

3. Optimize and calibrate surrogate

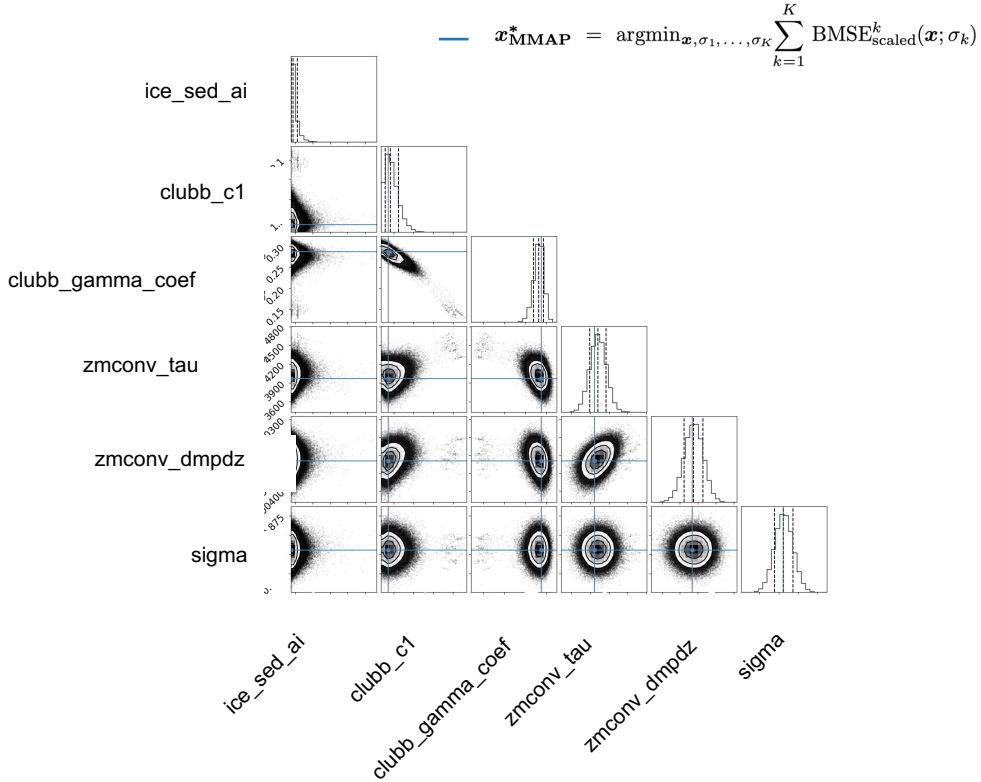


Figure: Corner plot of E3SMv2 ne30 parameter calibration for and PRECT against observations. More details on E3SM con

Season: DJF-only
Fields: PRECT, SWCF, LWCF
: MAP
Sampling: MCMC
Chains: 200

4. Validate in E3SM

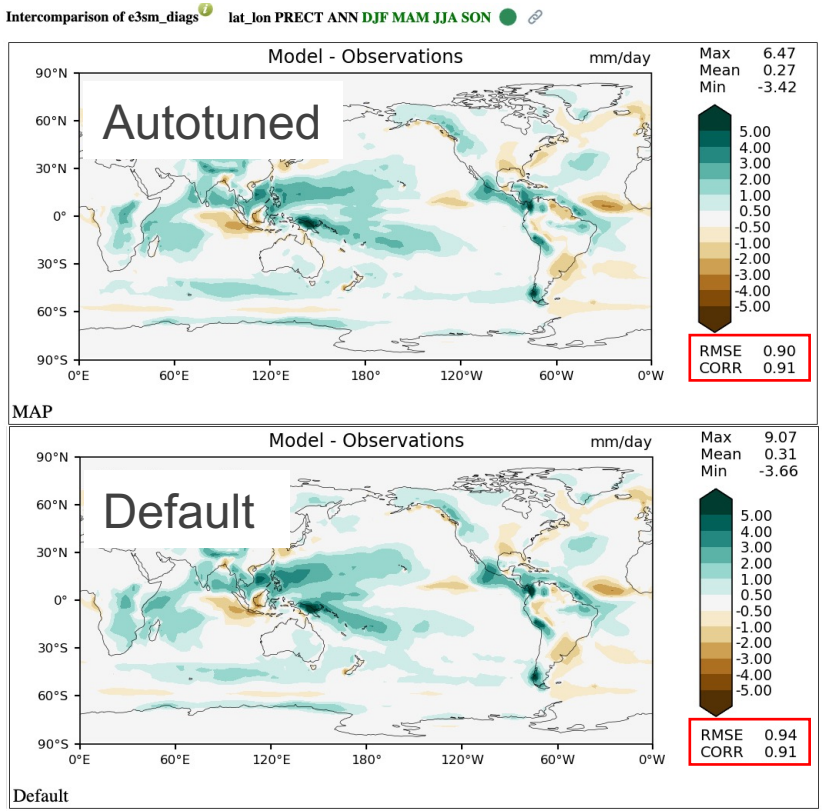
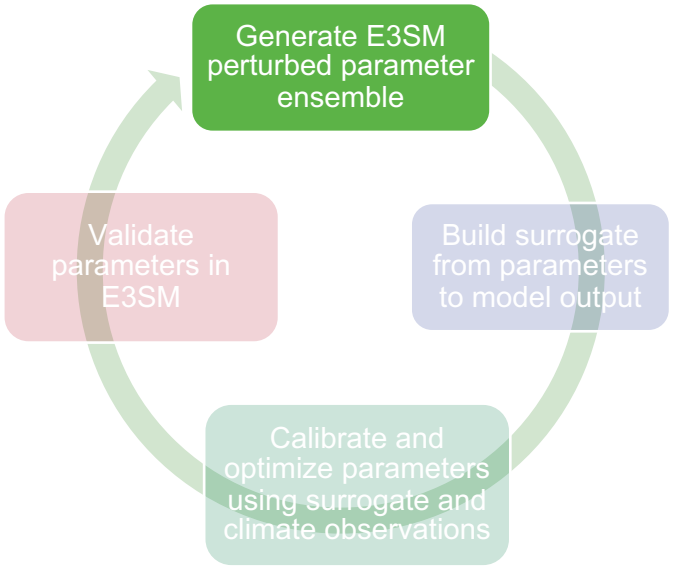


Figure: Intercomparison of E3SM diagnostics of precipitation climatology vs. observations for autotuned E3SMv2 (top) and default E3SMv2 (bottom). Intercomparison of e3sm diags

E3SM PPE creation using Dakota[1]



Screen recording example

[1] Adams et al., 2021

Current Workflow

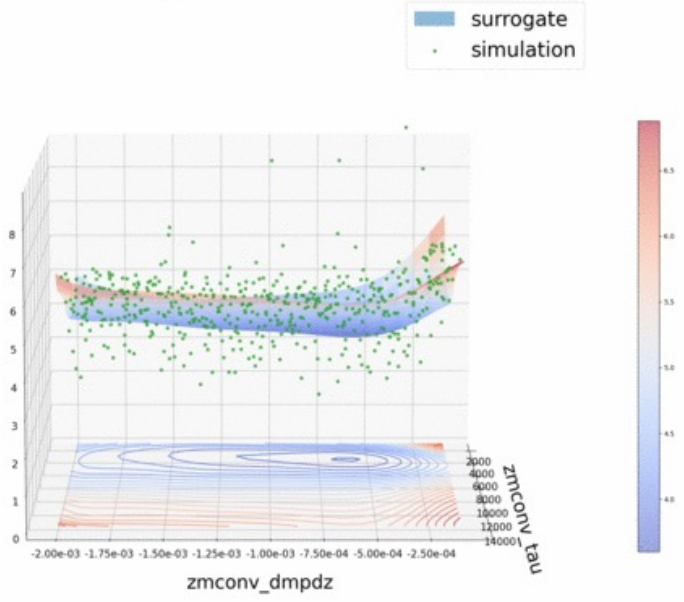


Figure: SWCF RMSE [Wm^{-2}] for autotuning ensemble (green dots) as a function of two deep convection parameters. Error computed with respect to default model tuning.

INPUT:
E3SM Parameters

High-Dimensional Input X
(m scalar values)

Optimal Parameter X^*

Perturbed Parameter Ensemble Generation

Surrogate Construction
$Z = f(X)$

Optimization over specified loss function
$X^* = argmin_X g(X)$

OUTPUT

High-Dimensional Output Y
(N spatial fields)

Dimension Reduction
Φ

Reduced Output Z
(K<N)

Fitted Values
$\hat{Z}$

Projecting back to original space
Φ^{-1}

Fitted Output $\hat{Y}$