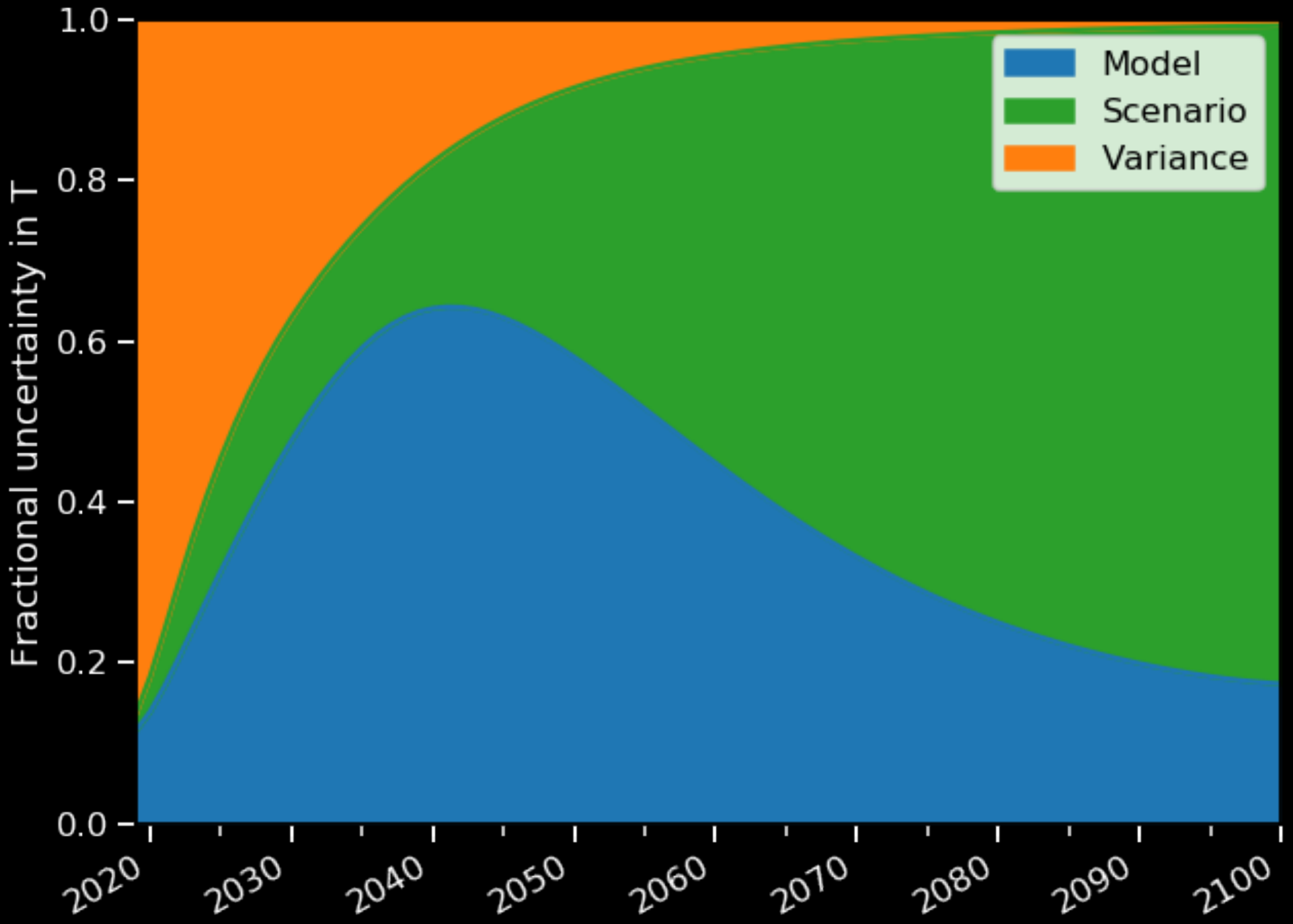
A Community Earth System Emulator for UQ and Calibration

Duncan Watson-Parris, Scripps / HDSI (UC San Diego),
Andrew Williams, Lucia Deaconu and Philip Stier (University of Oxford)

CESM Parameter Estimation Cross Working Group

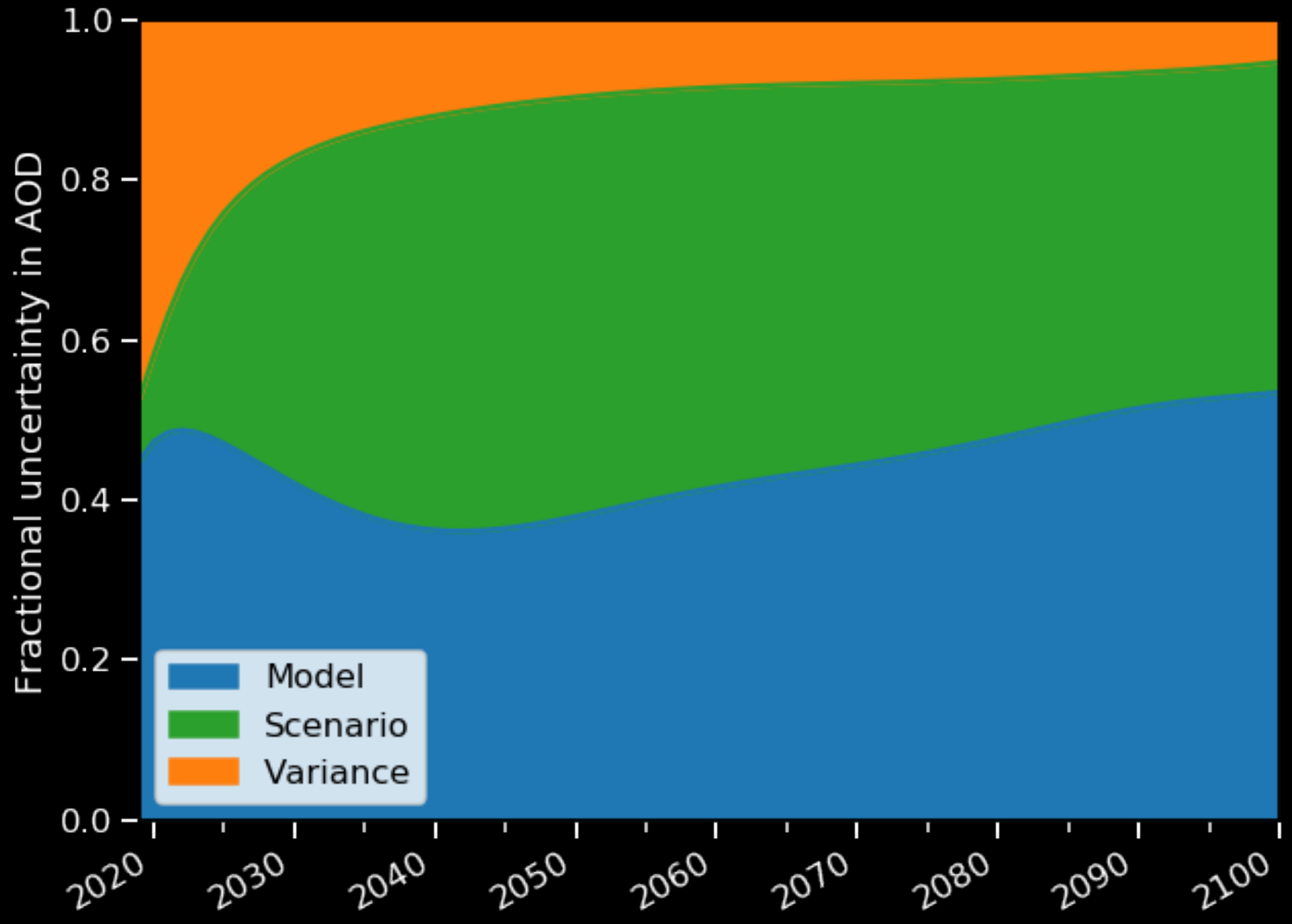
12th June 2023

Sources of uncertainty



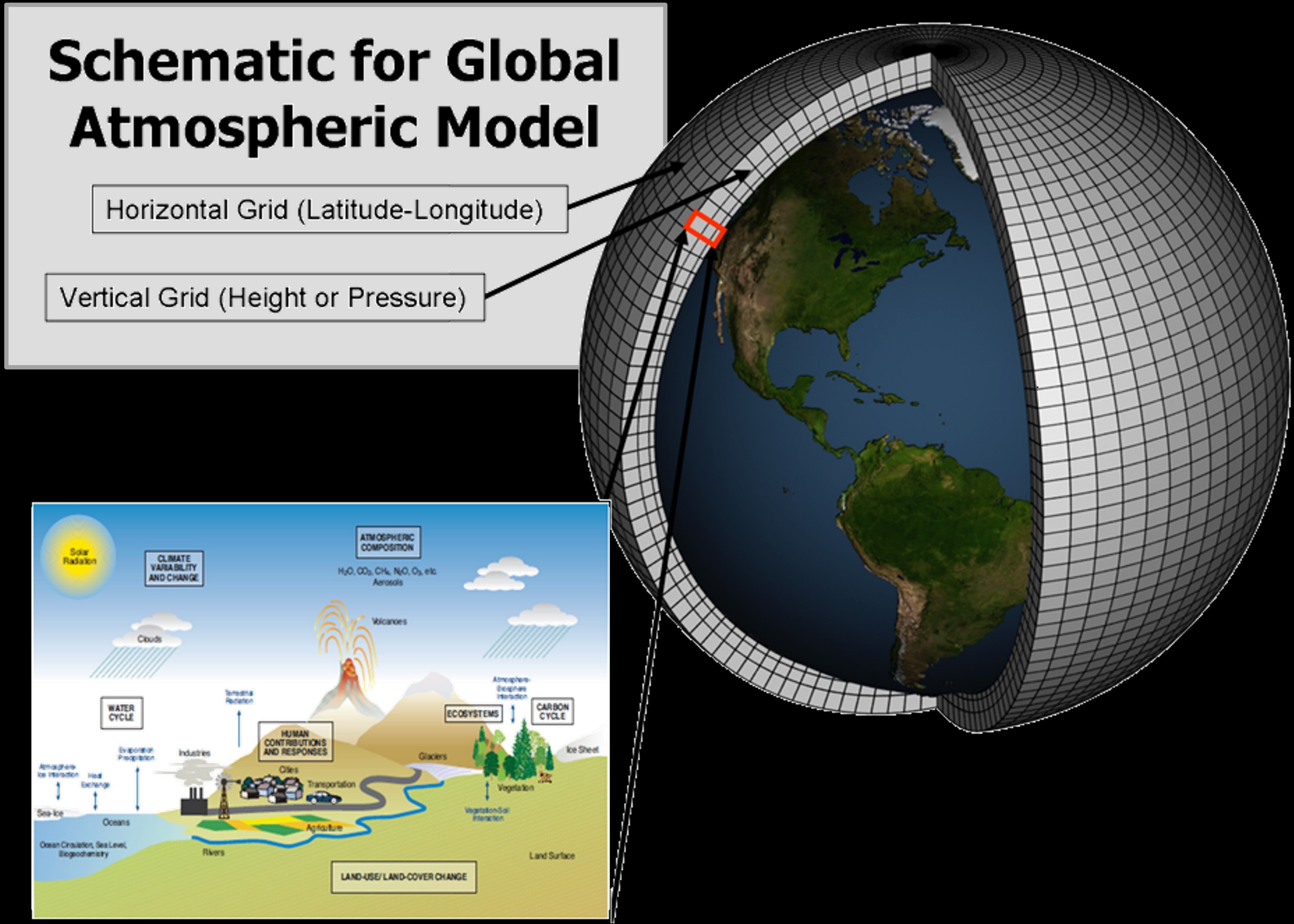
Watson-Parris 2021

Sources of uncertainty

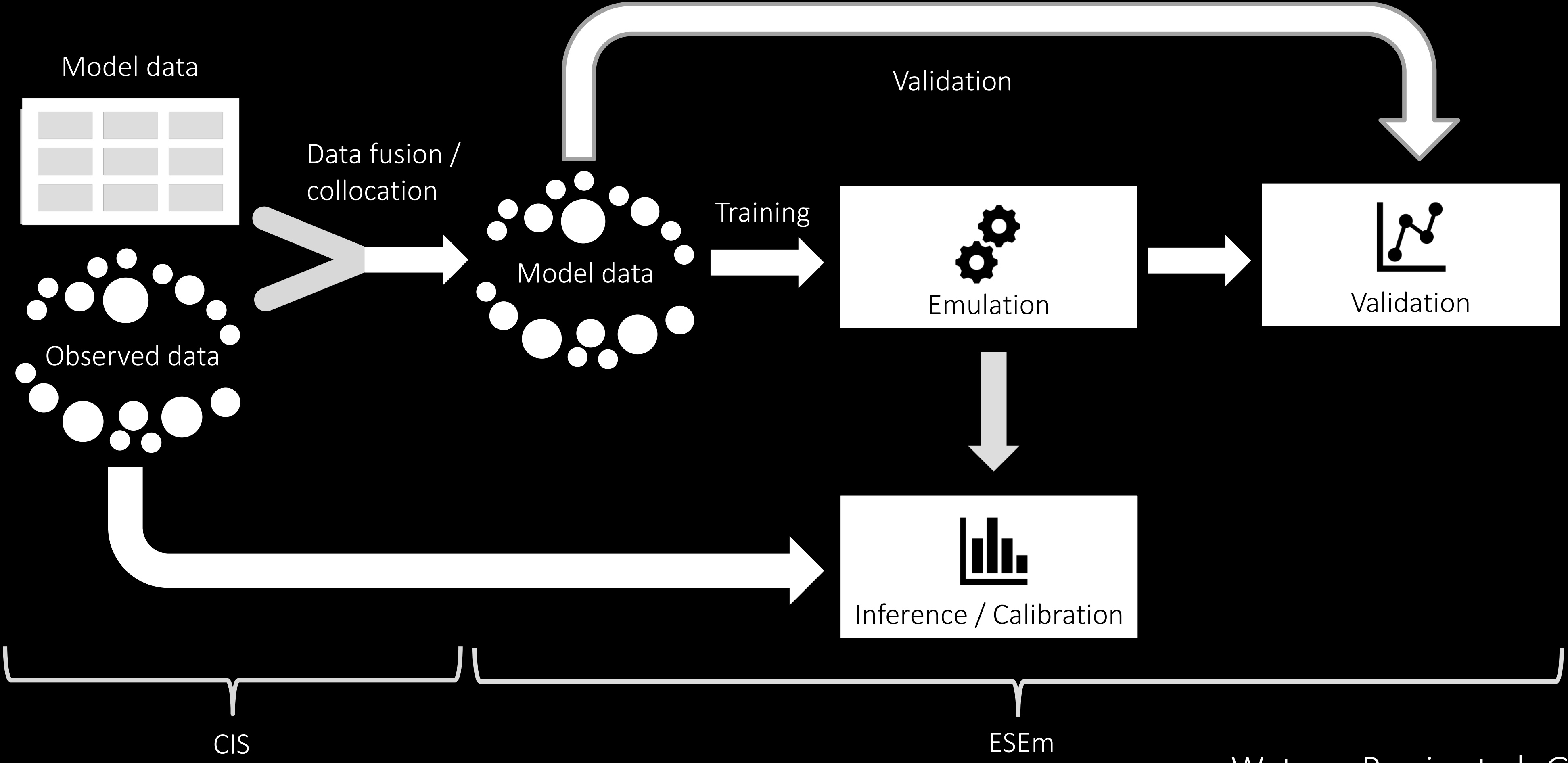


Watson-Parris 2021

Climate models



Earth System Emulator (ESEm)



Watson-Parris et al. *GMD 2016*

Watson-Parris et al. *GMD 2021*

Exploring parametric uncertainty

- A climate model is a function $\mathcal{F}$ of input forcers X and parameters θ and generates outputs Y: $\mathcal{F}(X, \theta) = Y$.
- Given a set of observations of Y, denoted Y^0, we would like to calculate the inverse: $\mathcal{F}^{-1}(Y^0) = \theta$.
- Or, more precisely, we would like to know the posterior probability distribution of the input parameters:

$$p(\theta|Y^0) = \frac{p(Y^0|\theta)p(\theta)}{p(Y^0)}$$

Likelihood-free inference

- We cannot compute the full likelihood but can approximate it using Approximate Bayesian computation (ABC; or rejection sampling):
    1. Draw θ from $p(\theta)$
    2. Calculate $\mathcal{F}(\theta)$ to determine Y
    3. Accept θ if $\rho(Y^0, Y) \leq \epsilon$
- As $\epsilon \rightarrow \infty$, we get observations from the prior, $p(\theta)$
- As $\epsilon \rightarrow 0$, we generate observations from $p(\theta|Y^0)$
- In practice we need many, many samples and so a surrogate model, or emulator, is used: $\mathcal{E}(\theta) \sim \mathcal{F}(\theta)$

Implausibility

- In practice, we need to define the distance, $\rho(Y^0, Y)$
- We use the concept of an implausibility metric:

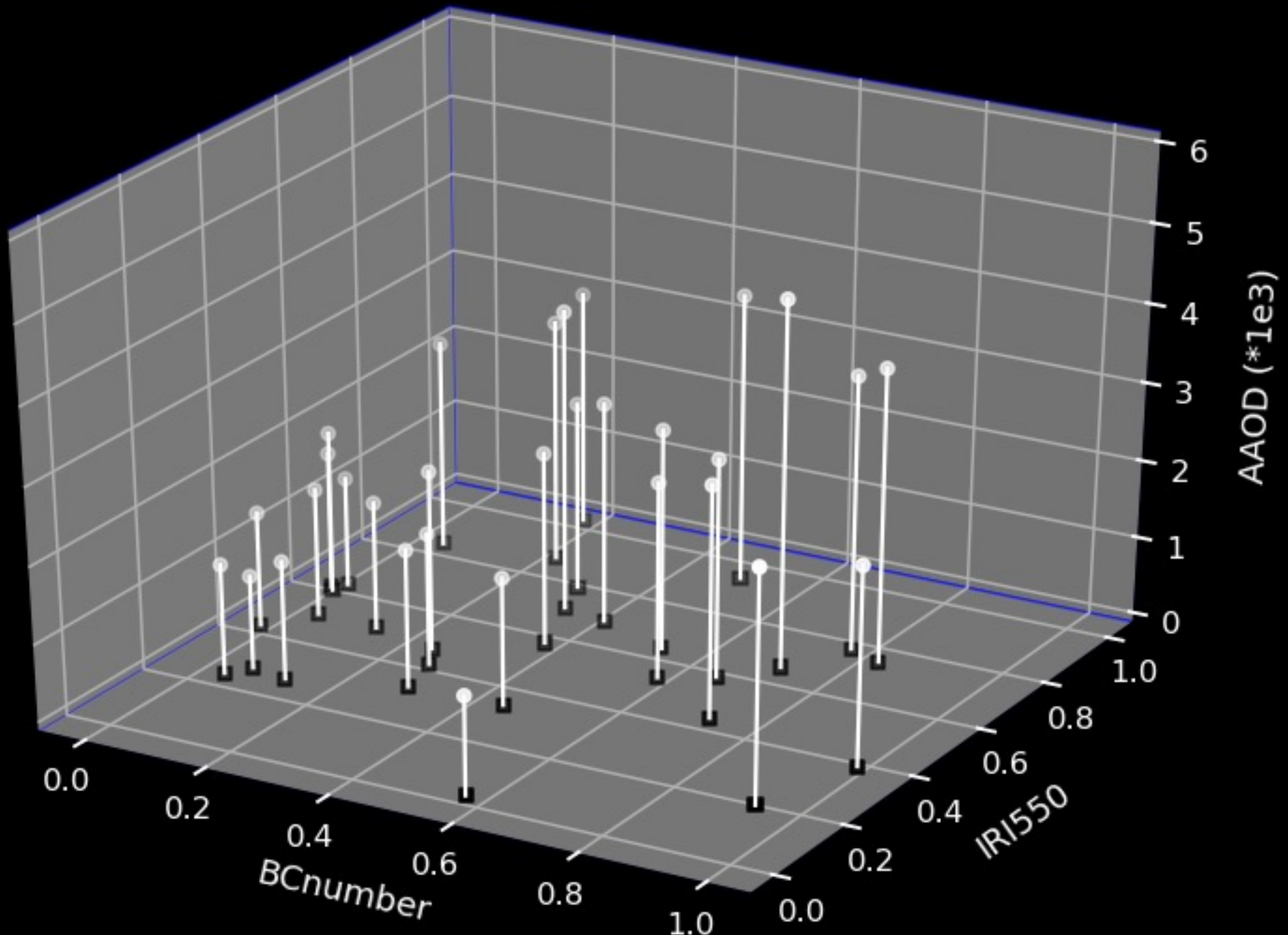
$$I(\theta) = \frac{|Y^0 - E[\eta(\boldsymbol{\theta})]\,|}{\sqrt{[Var(\phi(\boldsymbol{\theta})) + Var(\epsilon) + Var(\delta)]}}$$

Observation
Model output
[Prediction from emulator]
Emulator prediction uncertainty
Observational uncertainty
Structural uncertainty

- Assuming this distribution is unimodal, we can then use the 3σ to rule out parameter combinations at 95% confidence

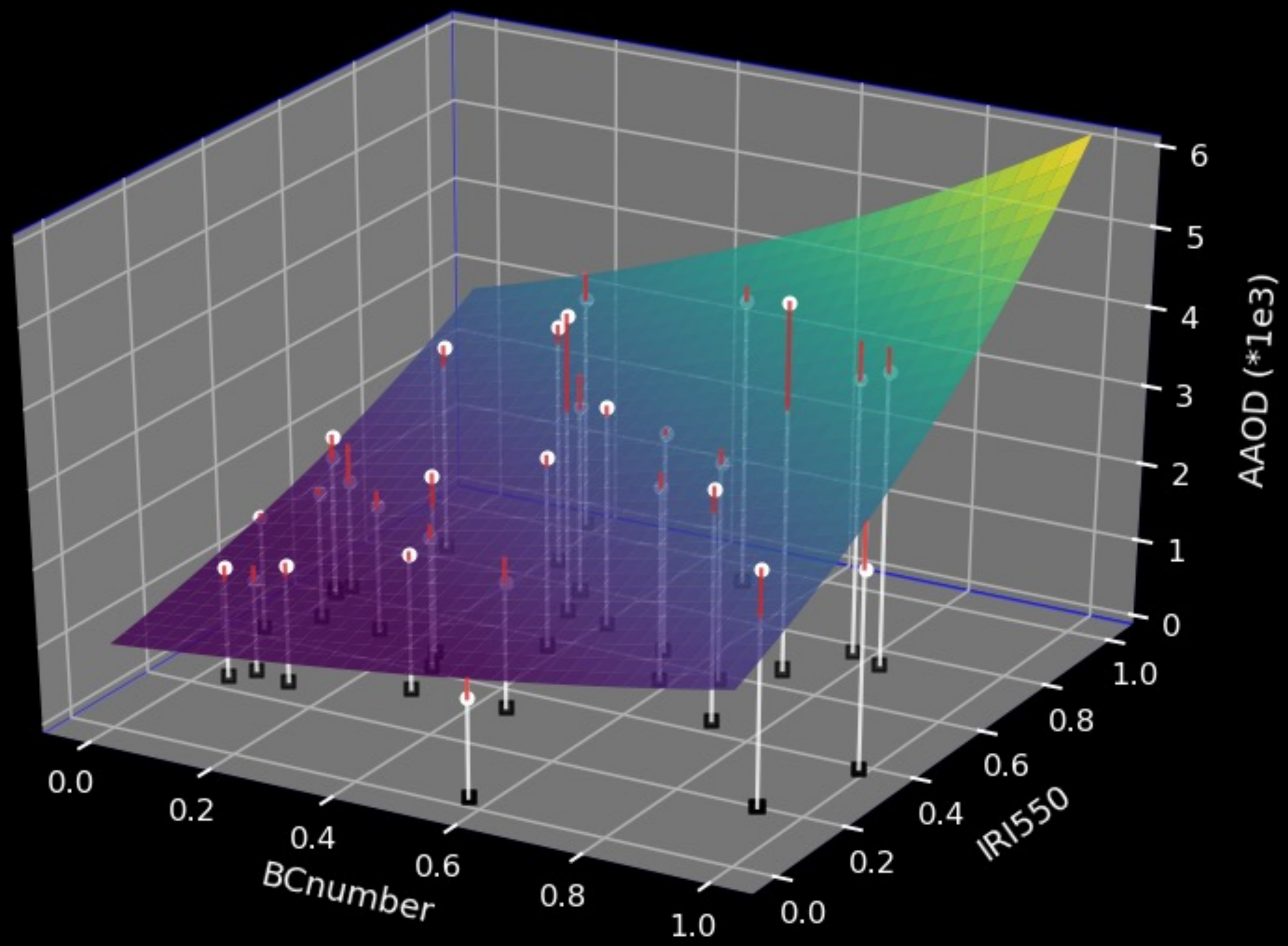
Vernon et al. *2010*

Emulation



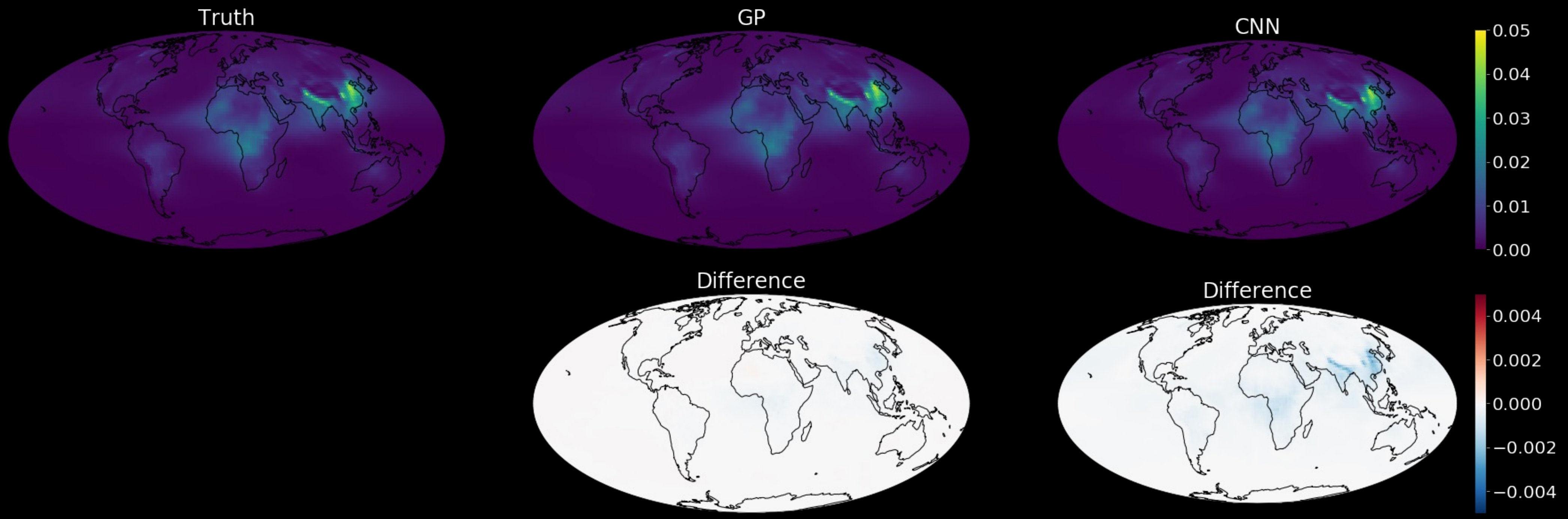
AAOD (*1e3)
0
1
2
3
4
5
6
BCnumber
0.0
0.2
0.4
0.6
0.8
1.0
IRI550
0.0
0.2
0.4
0.6
0.8
1.0

Emulation



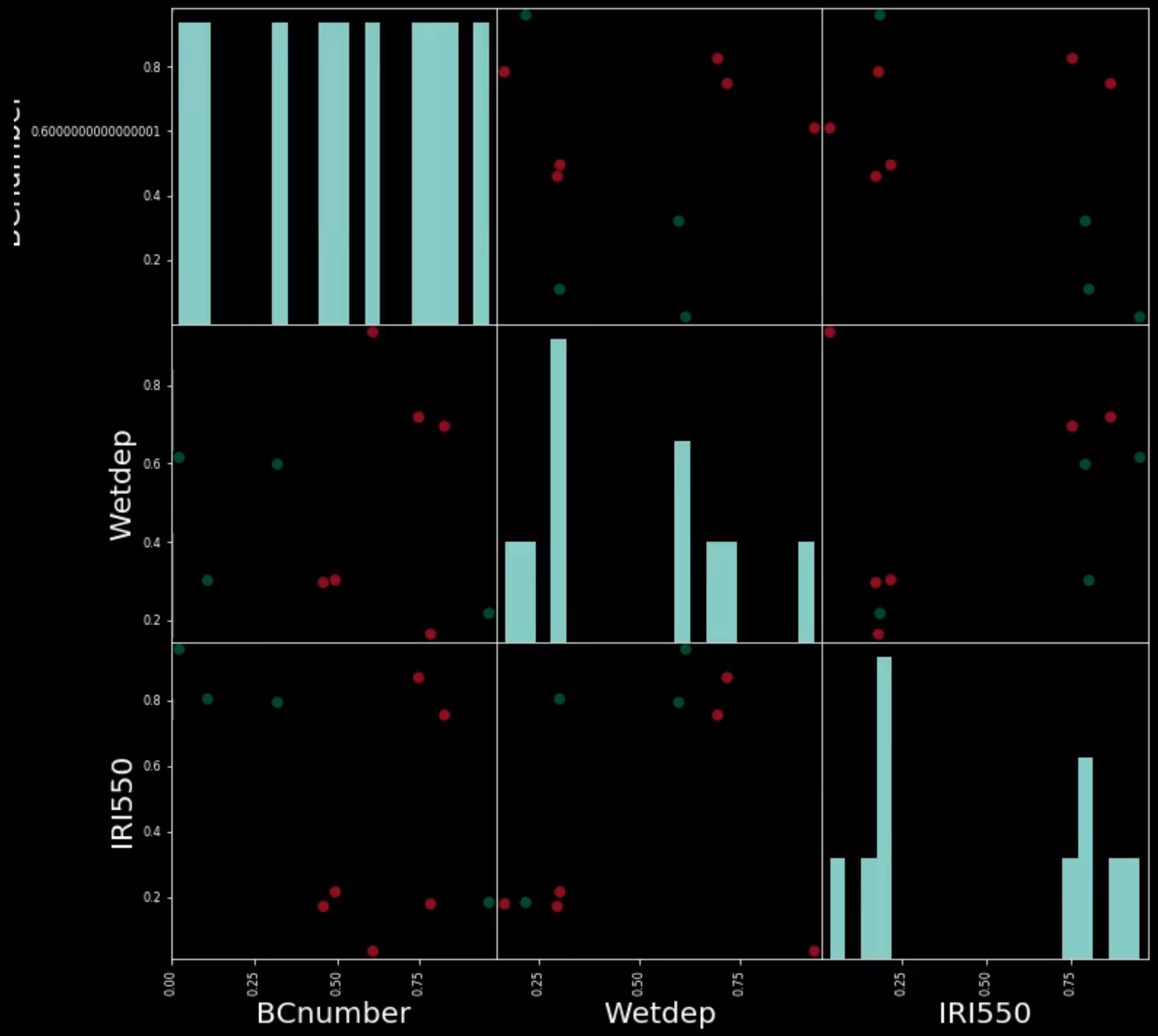
AAOD (*1e3)
0
1
2
3
4
5
6
BCnumber
0.0
0.2
0.4
0.6
0.8
1.0
IRI550
0.0
0.2
0.4
0.6
0.8
1.0

Emulation



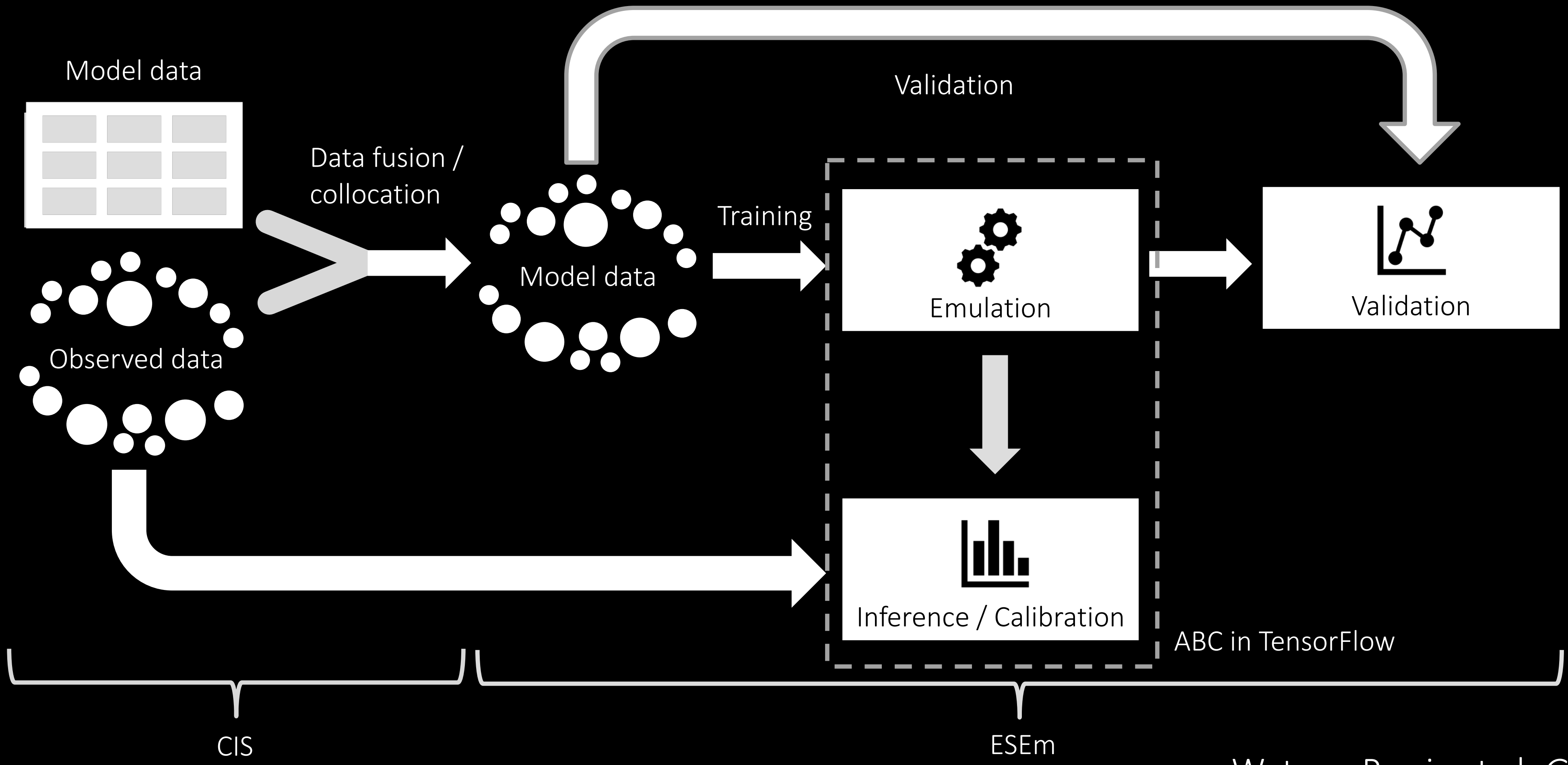
Kasim et al. 2020
Watson-Parris et al. GMD 2021

Sampling



0.8
0.6000000000000001
0.4
0.2
Wetdep
IRI550
0.00
0.25
0.50
0.75
BCnumber
Wetdep
IRI550

Earth System Emulator (ESEm)



Watson-Parris et al. *GMD 2016*

Watson-Parris et al. *GMD 2021*

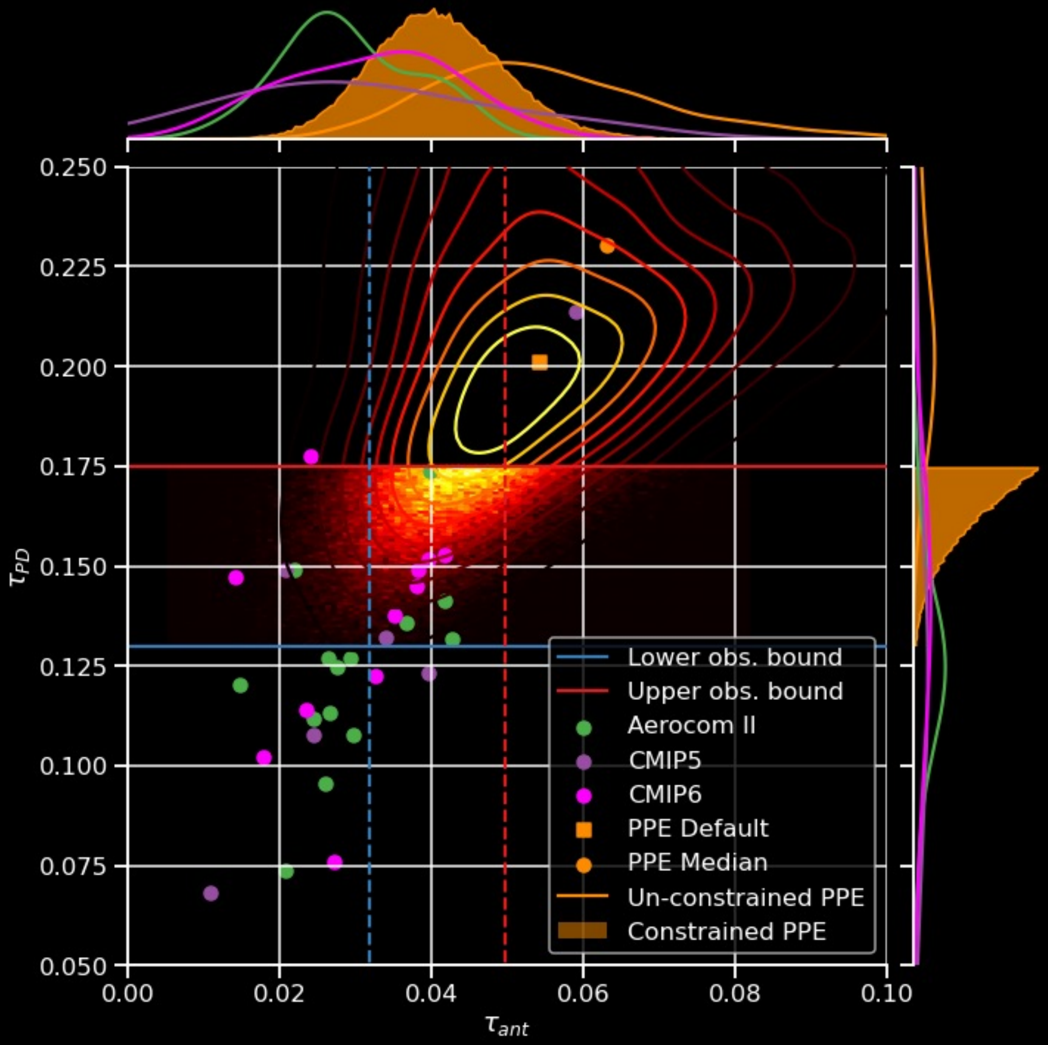
Parametric uncertainty

Table 1. The uncertain parameters and emissions factors

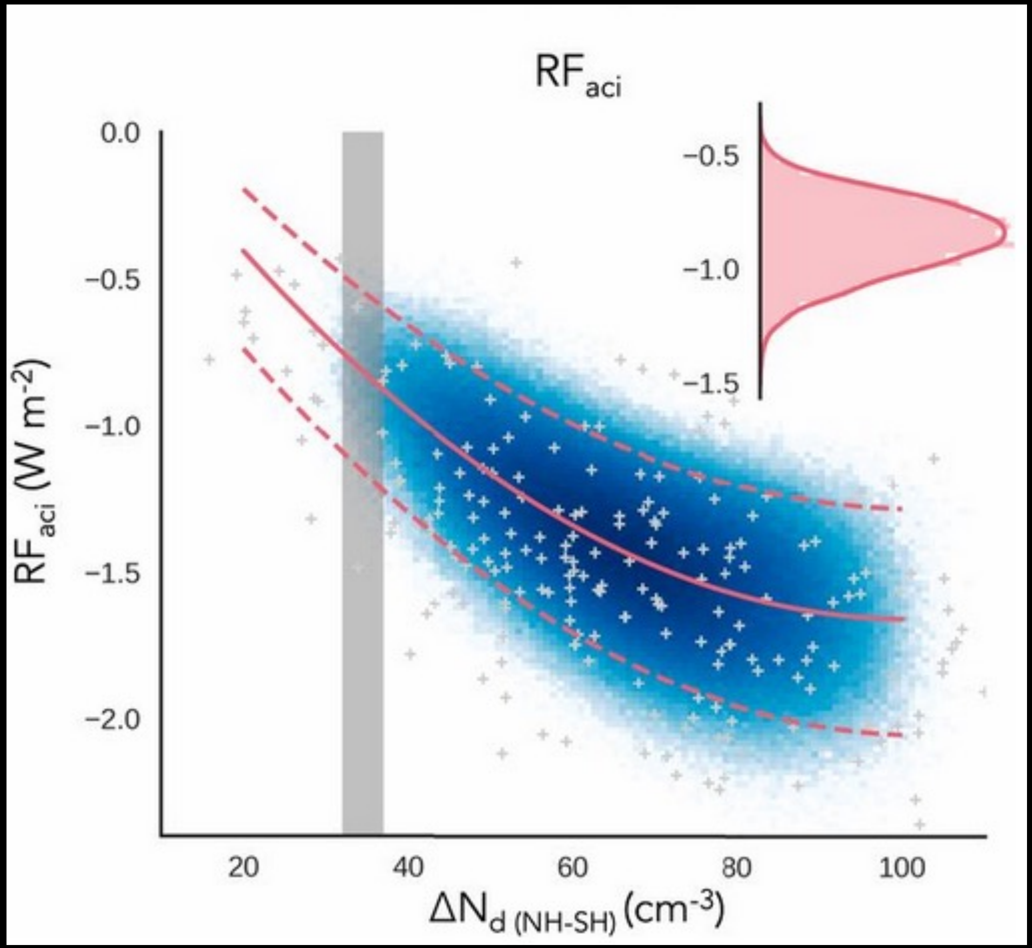
Parameter Key	Parameter name	Description of parameter	Uncertainty range	Effect
P1	BL_NUC	Boundary layer nucleation rate coeff (A)	$3.2e^{-7}$–$2e^{-4}$ s^{-1}	Absolute
P2	FT_NUC	Free troposphere nucleation rate	0.01–10	Scaled
P3	AGEING	Ageing "rate" from insoluble to soluble	0.3–5 monolayer	Absolute
P4	ACT_DIAM	Cloud drop activation dry diameter	50–100 nm	Absolute
P5	SO2O3_CLEAN	pH of cloud drops (controls $SO_2 + O_3$)	pH 4–6.5	Absolute
P6	SO2O3_POLL	pH of cloud drops ($SO_2 + O_3$)	pH 3.5–5	Absolute
P7	NUC_SCAV_DIAM	Nucleation scavenging diameter offset dry diameter	0–50 nm	Absolute
P8	NUC_SCAV_ICE	Nucleation scavenging fraction (accumulation mode) in mixed and ice clouds ($T < -15\,^{\circ}C$)	0–1	Scaled
P9	DRYDEP_AER_AIT	Dry deposition velocity of Aitken mode aerosol	0.5–2	Scaled
P10	DRYDEP_AER_ACC	Dry deposition velocity of accumulation mode aerosol	0.1–10	Scaled
P11	ACC_WIDTH	Modal width (accumulation soluble/insoluble)	1.2–1.8	Absolute
P12	AIT_WIDTH	Modal width (Aitken soluble/insoluble)	1.2–1.8	Absolute
P13	NUCAIT_WIDTH	Mode separation diameter (nucleation/Aitken)	9–18 nm	Absolute
P14	AITACC_WIDTH	Mode separation diameter (Aitken/accumulation)	0.9–2 × ACT_DIAM	Scaled
P15	FF_EMS	BC/OC mass emission rate (fossil fuel)	0.5–2	Scaled
P16	BB_EMS	BC/OC mass emission rate (biomass burning)	0.25–4	Scaled
P17	BF_EMS	BC/OC mass emission rate (biofuel)	0.25–4	Scaled
P18	FF_DIAM	BC/OC emitted mode diameter (fossil fuel)	30–80 nm	Absolute
P19	BB_DIAM	BC/OC emitted mode diameter (biomass burning)	50–200 nm	Absolute
P20	BF_DIAM	BC/OC emitted mode diameter (biofuel)	50–200 nm	Absolute
P21	PRIM_SO4_FRAC	Mass fraction of SO_2 converted to new SO_4^{2-} particles in sub-grid power plant plumes	0–1 %	Absolute
P22	PRIM_S04_DIAM	Mode diameter of new sub-grid SO_4^{2-} particles	20–100 nm	Absolute
P23	SS_ACC	Sea spray mass flux (coarse/accumulation)	0.2–5	Scaled
P24	ANTH_S02	SO_2 emission flux (anthropogenic)	0.6–1.5	Scaled
P25	VOLC_SO2	SO_2 emission flux (volcanic)	0.5–2	Scaled
P26	DMS_FLUX	DMS emission flux	0.5–2	Scaled
P27	BIO_SOA	Biogenic monoterpene production of SOA	5–360 $Tg\,a^{-1}$	Absolute
P28	ANTH_SOA	Anthropogenic VOC production of SOA	2–112 $Tg\,a^{-1}$	Absolute

Lee et al. *2013.*

Constraining parametric uncertainty



Watson-Parris et al. *GRL* 2020



McCoy et al. 2020

Calibrating CESM2 PPE on Casper: Setup

```
[ ]: import xarray as xr
     import numpy as np
     import pandas as pd

     from esem import gp_model
     from esem.utils import validation_plot, get_param_mask
```

```
[ ]: def global_mean(ds):
         weights = np.cos(np.deg2rad(ds.lat))
         return ds.weighted(weights).mean(['lat', 'lon'])

     def get_ensemble_member(ds):
         fname = ds.encoding['source']
         member = int(fname.split('.')[-4])
         return ds.assign_coords(member=member).expand_dims('member')
```

https://gist.github.com/duncanwp/89175a17b7221e4d3639765621c7f7f9

Calibrating CESM2 PPE on Casper: Data Reading

```
[ ]: data_path = '/glade/campaign/cgd/projects/ppe/cam_ppe/rerun_PPE_250/'
     params=xr.open_dataset(data_path+"parameter_262_w_control.nc").to_pandas().drop(columns=['Sample_nmb'])
     ds = xr.open_mfdataset(data_path+'PD/PD_timeseries/*/atm/hist/cc_PPE_250_ensemble_PD.*.h0.SWCF.nc', preprocess=get_ensemble_member)
     SWCF = global_mean(ds['SWCF']).mean('time').compute()
```

```
[ ]: # Some of the PPE members are missing data so just select the params we actually have
     sub_params = params.iloc[SWCF.member.values]
     # Unit normalise all the parameters
     ppe_params = (sub_params - sub_params.min()) / (sub_params.max() - sub_params.min())
```

```
[ ]: # We can use an information criterion to choose the best parameters automatically:
     best_params = ppe_params[ppe_params.columns[get_param_mask(ppe_params, SWCF)]]
     best_params.columns
```

```
[ ]: n_test = 25

     X_test, X_train = best_params[:n_test], best_params[n_test:]
     Y_test, Y_train = SWCF[:n_test], SWCF[n_test:]
```

https://gist.github.com/duncanwp/89175a17b7221e4d3639765621c7f7f9

Calibrating CESM2 PPE on Casper: Emulation

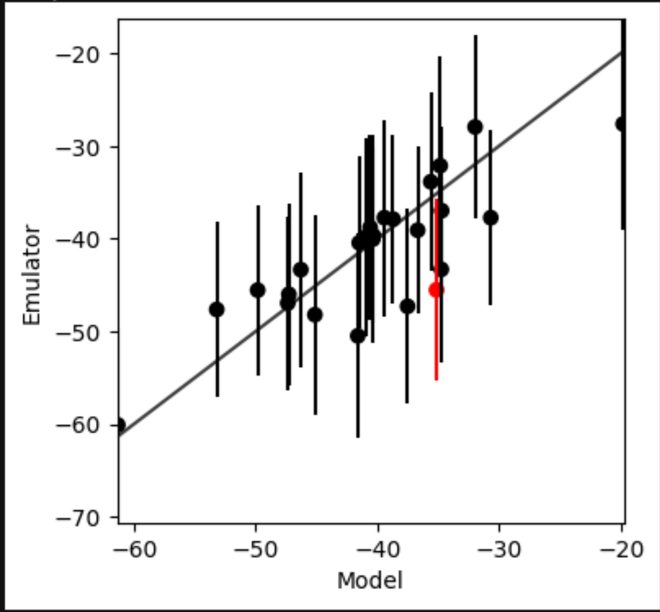
```
[ ]: # Can try different kernels here
     gp = gp_model(X_train, Y_train, kernel=['Linear', 'RBF'])

[ ]: gp.train()

[ ]: m, v = gp.predict(X_test)

[21]: validation_plot(Y_test.data, m, v, figsize=(4,4))
```

Proportion of 'Bad' estimates : 4.00%



https://gist.github.com/duncanwp/89175a17b7221e4d3639765621c7f7f9

Calibrating CESM2 PPE on Casper: Calibration

```
[ ]: from esem.utils import get_random_params
     from esem.abc_sampler import ABCSampler, constrain

[ ]: # Setup sampler with 1 million points
     sample_points = pd.DataFrame(data=get_random_params(23, int(1e6)), columns=X_train.columns)
     sampler = ABCSampler(gp, np.asarray([-40.5]), obs_uncertainty=0.5)

[ ]: valid_samples = sampler.batch_constrain(sample_points, batch_size=10000)

[28]: print("Remaining points: {}".format(valid_samples.sum()))

      Remaining points: 100
```

https://gist.github.com/duncanwp/89175a17b7221e4d3639765621c7f7f9

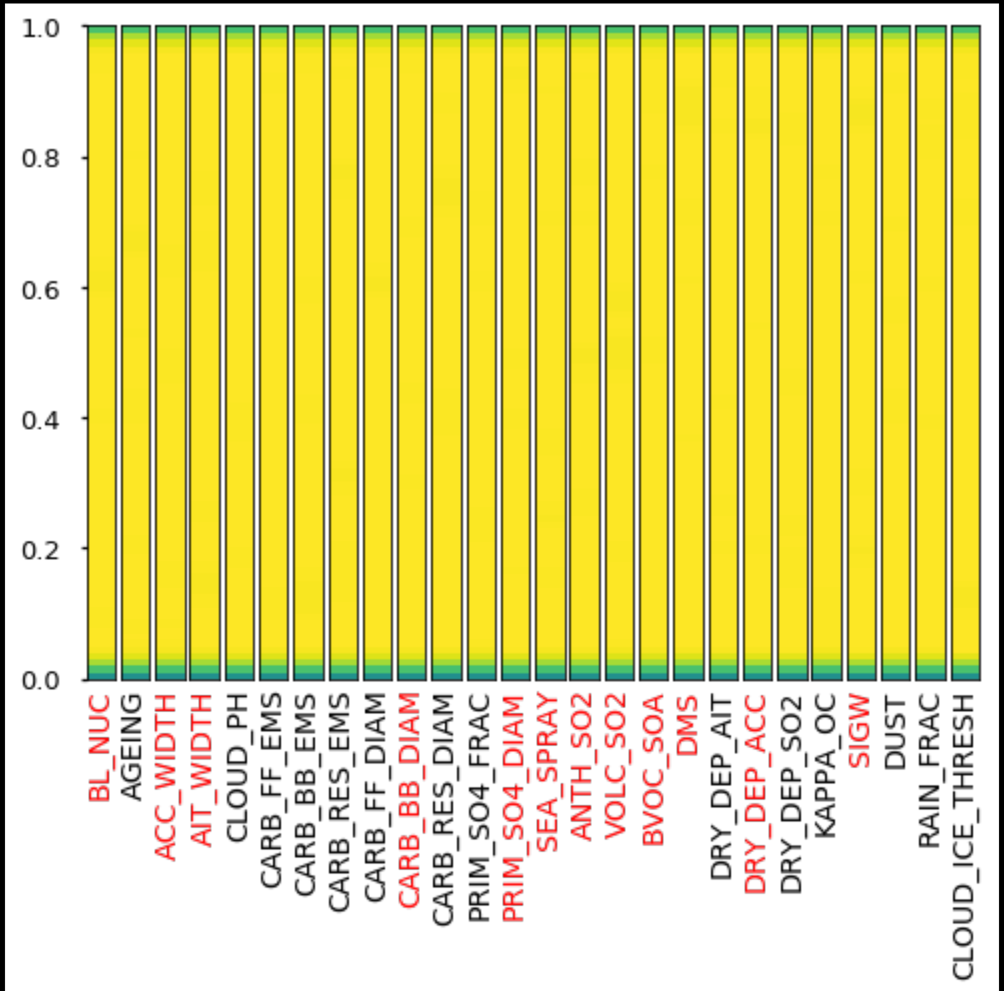
Conclusions and outlook

- Climate model emulation is an important tool for constraining parameter uncertainty, and has been used for a while
- A fast, open tool leveraging modern ML libraries is introduced which makes the whole workflow straightforward and reproducible
- We also demonstrate how such tools can be used for exploring uncertainty in future emissions scenarios
- ESEm is designed to be easily extended to include other ML models
- We welcome feedback and suggestions, especially with pull requests!

https://github.com/duncanwp/ESEm

Spare slides

Getting your priors straight



Getting your priors straight