Efficiently constraining parameter uncertainty in a General Circulation Model using targeted data

Oliver Dunbar[1], Tapio Schneider[1], Andrew Stuart[2], **CliMA**

[1]Geological and Planetary Sciences
California Institute of Technology

[2]Computing and Mathematical Sciences
California Institute of Technology

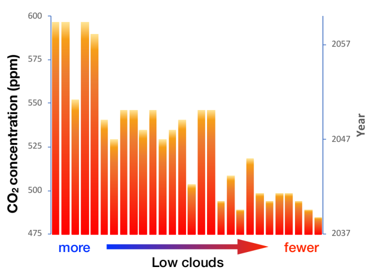
Emerging Technologies for Weather & Climate Modelling, 30th June 2020

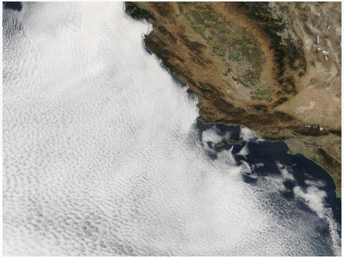
Table of Contents

CliMA CLIMATE MODELING ALLIANCE Caltech

Why cloud parameters? Why uncertainty?

2015 Paris Agreement: $2^\circ C$ threshold temperature increase above preindustrial levels.

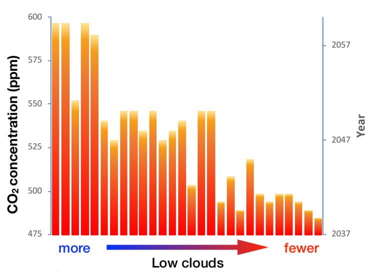


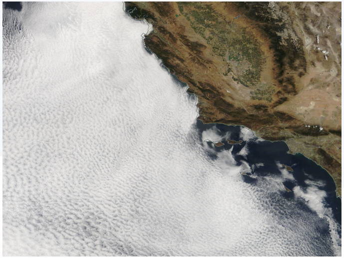


Schneider, Teixeira, et al. 2017 , https://worldview.earthdata.nasa.gov/

Why cloud parameters? Why uncertainty?

2015 Paris Agreement: $2^{\circ}C$ threshold temperature increase above preindustrial levels.





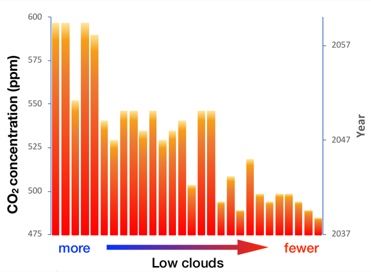
Schneider, Teixeira, et al. 2017 , https://worldview.earthdata.nasa.gov/

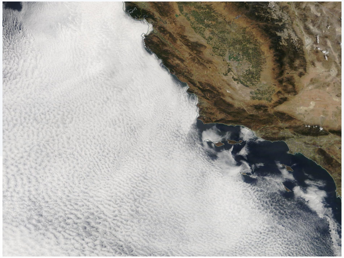
Cloud parameter uncertainty $\implies$ prediction uncertainty

CliMA CLIMATE MODELING ALLIANCE Caltech

Why cloud parameters? Why uncertainty?

2015 Paris Agreement: $2^{\circ}C$ threshold temperature increase above preindustrial levels.





Schneider, Teixeira, et al. 2017 , https://worldview.earthdata.nasa.gov/

Cloud parameter uncertainty $\Longrightarrow$ prediction uncertainty
All of these models give certain predictions.

CliMA CLIMATE MODELING ALLIANCE Caltech

Climate Modelling Alliance (CliMA)

2018 Collaboration to produce a new Earth System Model[1]. clima.caltech.edu

Caltech

Jet Propulsion Laboratory

NPS PRAESTANTIA PER SCIENTIAM 1909

Some aims

- Redesign physical models to better resolve clouds. ('physical over empirical')
- Data assimilation, uncertainty quantification and machine learning framework.
- Include data from high resolution simulation and observations.
- Julia programming language framework unified across all components.

[1]Schneider, Lan, et al. 2017.

Table of Contents

CliMA CLIMATE MODELING ALLIANCE Caltech

The abstract Bayesian setting[2]

Ingredients

- Prior parameters $\theta \sim \mu_0$.
- $\mathcal{G}_T(\theta)$ - **time averaged** data from GCM, (dimension reduction)
- Observational noise $\eta \sim N(0, \Sigma)$.

Recipe for observation y:

$$y = \mathcal{G}_T(\theta) + \eta \tag{1}$$

[2] Stuart 2010; Cleary et al. 2019.

CliMA CLIMATE MODELING ALLIANCE Caltech

Our GCM example

Idealized moist GCM: Aquaplanet[3]. Moist convection scheme in quasi-equilibrium (Betts Miller type).

$$\text{flux}(x) \to \frac{x - x_{\text{ref}}(\alpha)}{\tau}, \qquad \theta = (\alpha, \tau)$$

Physical Priors

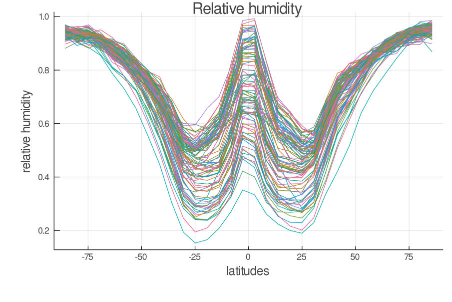
$$0 < \alpha < 1 \text{ relative humidity. Prior: } \text{logit}(\alpha) \sim \text{Normal}$$
$$0 < \tau \text{ relaxation time. Prior: } \log(\tau) \sim \text{Normal}$$

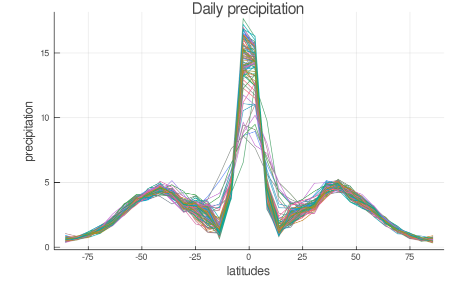
[3] Frierson 2007; O'Gorman and Schneider 2008.

CliMA CLIMATE MODELING ALLIANCE Caltech

Our GCM example

Forward map $\mathcal{G}_T$: 2 time averaged (20 day) quantities: relative humidity at 5km, and daily precipitation.





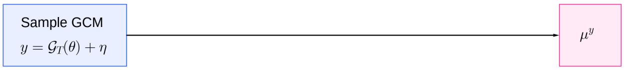
100 samples of $\mathcal{G}_T(\theta)$ at $\theta \sim \mu_0$.

Observed data point: $y = \mathcal{G}_T((0.7, 7200s))$

CliMA CLIMATE MODELING ALLIANCE Caltech

The Bayesian inverse problem setting[5]

Learning objective: Sample $\theta | y \sim \mu^y$ posterior, constrained by data y.

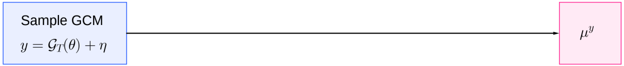


[4] Cleary et al. 2019.

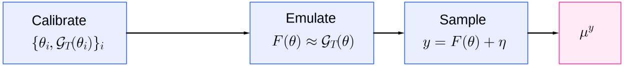
[5] Stuart 2010; Cleary et al. 2019.

The Bayesian inverse problem setting[5]

Learning objective: Sample $\theta | y \sim \mu^y$ posterior, constrained by data y.



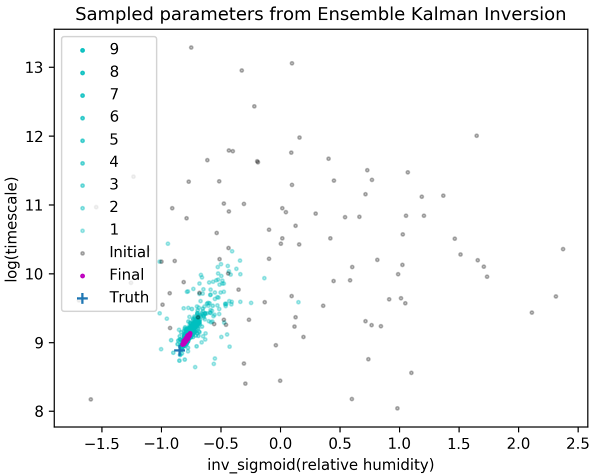
Calibrate **E**mulate **S**ample framework (CES)[4]



[4]Cleary et al. 2019.
[5]Stuart 2010; Cleary et al. 2019.

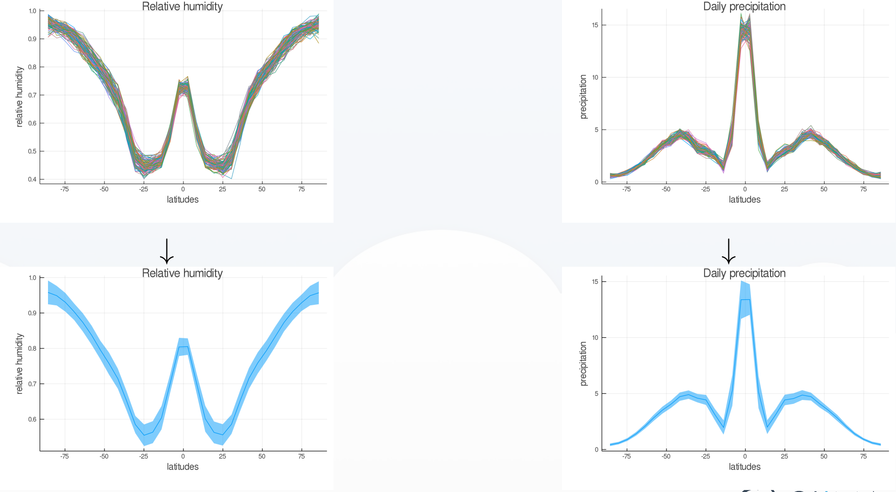
CliMA CLIMATE MODELING ALLIANCE Caltech

Calibrate: Ensemble Kalman Inversion of the GCM parameters.[6]



[6] Iglesias, Law, and Stuart 2013.

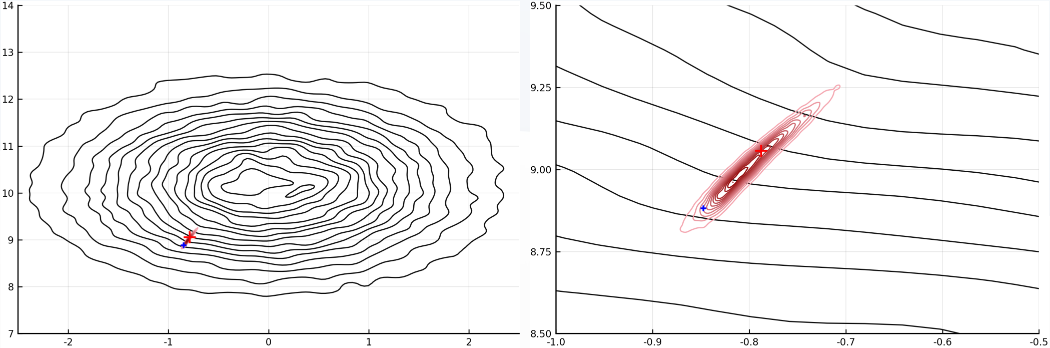
Emulate: Gaussian Process emulator.[7]



[7] Kennedy and O'Hagan 2001; Santner et al. 2018.

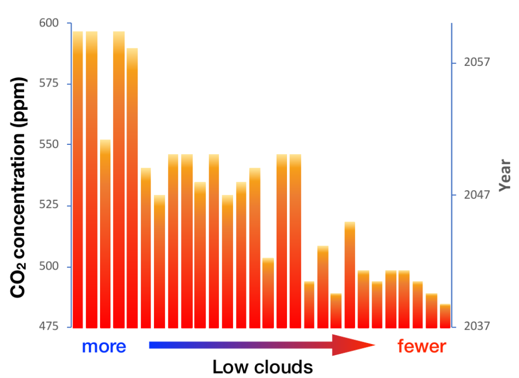
Sample: Random Walk Metropolis

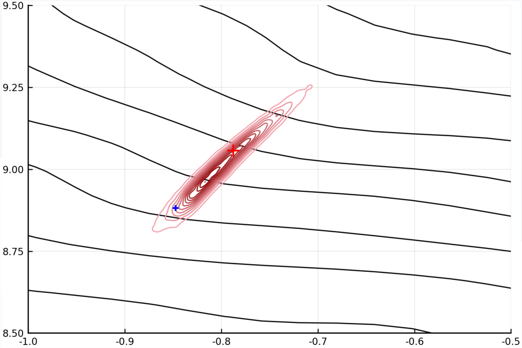
Blue + Truth, Red + Optimal,
Black – Prior, Red – Posterior



CliMA
CLIMATE MODELING ALLIANCE
Caltech

We can now make predictions with uncertainty





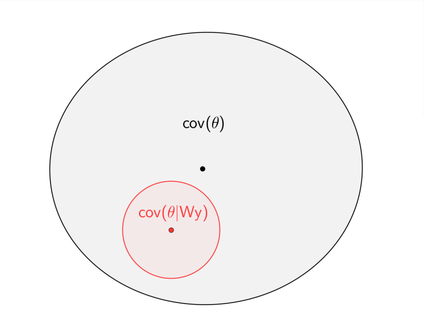
CliMA
CLIMATE MODELING ALLIANCE
Caltech

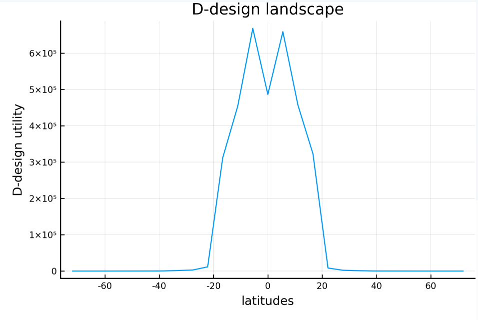
Table of Contents

Design objectives

Design objectives:

1. Find design W so that $\theta | Wy$ is **maximally** informative.





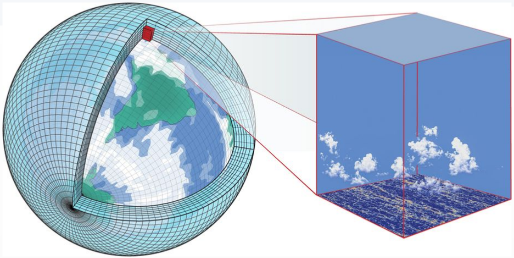
$\Longrightarrow$ Equatorial region is highly informative.

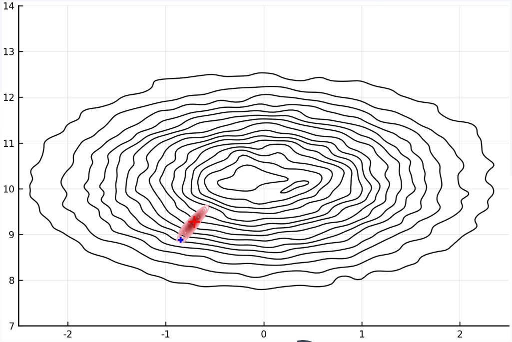
CliMA CLIMATE MODELING ALLIANCE Caltech

Design objectives

Design objectives:

1. Find design W so that $\theta | Wy$ is **maximally** informative.
2. **Target computation:** Uncertainty quantification **only** at optimal location:





Looking ahead

1. Using imperfect data in the design stage. (model error)
2. Targeted simulation independent of parameterization. (LES/DNS)
3. More complex GCM, with $\mathcal{O}(100)$ parameters (Gaussian Process?)
4. Online (sequential) design (Multiple design locations)

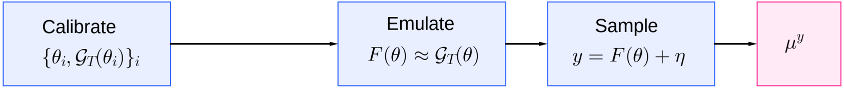
References

- Alexanderian, Alen, Philip J Gloor, and Omar Ghattas (2016). "On Bayesian A-and D-optimal experimental designs in infinite dimensions". In: *Bayesian Analysis* 11.3, pp. 671–695.
- Chaloner, Kathryn and Isabella Verdinelli (1995). "Bayesian Experimental Design: A Review". In: *Statistical Science* 10.3, pp. 273–304.
- Cleary, Emmet et al. (2019). "Calibrate, Emulate, Sample". In: *arXiv preprint arXiv:1912.*
- Frierson, Dargan MW (2007). "The dynamics of idealized convection schemes and their effect on the zonally averaged tropical circulation". In: *Journal of the Atmospheric Sciences* 64.6, pp. 1959–1976.
- Huan, Xun and Youssef M. Marzouk (Jan. 2013). "Simulation-based optimal Bayesian experimental design for nonlinear systems". In: *Journal of Computational Physics* 232.1.
- Iglesias, Marco A, Kody JH Law, and Andrew M Stuart (2013). "Ensemble Kalman methods for inverse problems". In: *Inverse Problems* 29.4, p. 045001.
- Kennedy, Marc C and Anthony O'Hagan (2001). "Bayesian calibration of computer models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3, pp. 425–464.
- O'Gorman, Paul A. and Tapio Schneider (2008). "The Hydrological Cycle over a Wide Range of Climates Simulated with an Idealized GCM". In: *Journal of Climate* 21.15, pp. 3815–3832. DOI: 10.1175/2007JCLI2065.1.
- Santner, Thomas J et al. (2018). *The Design and Analysis of Computer Experiments*. 2nd. Springer Series in Statistics. New York, NY: Springer.
- Schneider, Tapio, Shiwei Lan, et al. (2017). "Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations". In: *Geophysical Research Letters* 44.24, pp. 12, 396–12, 417. DOI: 10.1002/2017GL076101.
- Schneider, Tapio, João Teixeira, et al. (2017). "Climate goals and computing the future of clouds". In: *Nature Climate Change* 7.1, p. 3.
- Stuart, Andrew M (2010). "Inverse problems: a Bayesian perspective". In: *Acta Numerica* 19, pp. 451–559.

CliMA CLIMATE MODELING ALLIANCE Caltech

Calibrate Emulate Sample methods (CES)[10]

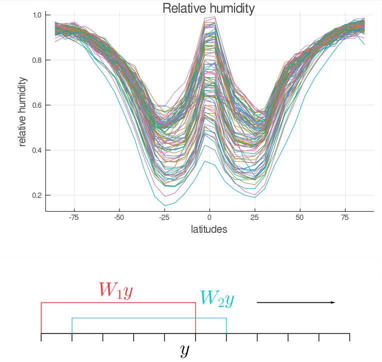
- (Calibrate) Ensemble Kalman Inversion of the GCM parameters.[8]. Derivative free. Optimal parameters in 100s of samples.
- (Emulate) Gaussian Process emulator.[9]
- (Sample) Random Walk Metropolis.

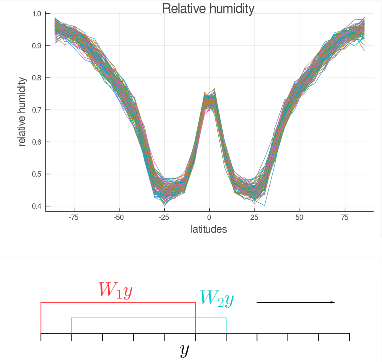


[8] Iglesias, Law, and Stuart 2013.
[9] Kennedy and O'Hagan 2001; Santner et al. 2018.
[10] Cleary et al. 2019.

Finding the optimal design

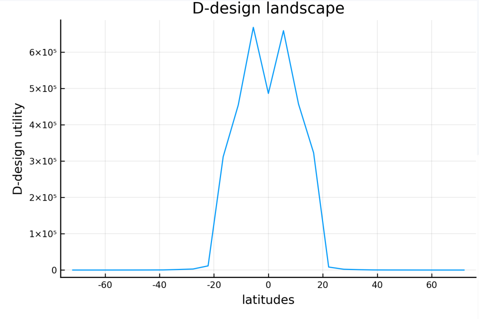
Regional (in latitude) data $W_i y$ give different posterior distributions $\mu^{W_i y}$

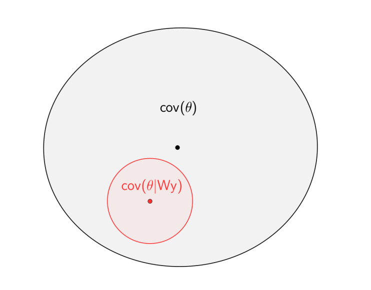




Maximal information

W^* that maximizes an information entropy[11] gives the most concentrated posterior distribution





E.g $W* = \arg\max(U(W))$ where $U(W) = \det(\mathrm{cov}(\theta|Wy))^{-1}$

[11] Chaloner and Verdinelli 1995; Huan and Marzouk 2013; Alexanderian, Gloor, and Ghattas 2016.