Quantifying parameter uncertainty within a climate model

Oliver Dunbar[1], Alfredo Garbuno-Iñigo[2], Tapio Schneider[1], Andrew Stuart[1],
CliMA

[1]California Institute of Technology
[2]Instituto Tecnológico Autónomo de México (ITAM)

Eric and Wendy Schmidt, Mountain Philanthropies, NSF, ONR, Paul G. Allen Family Foundation

NYSDS, 20th October 2020

CliMA
CLIMATE MODELING ALLIANCE

Caltech

The Climate Modelling Alliance (CliMA)

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

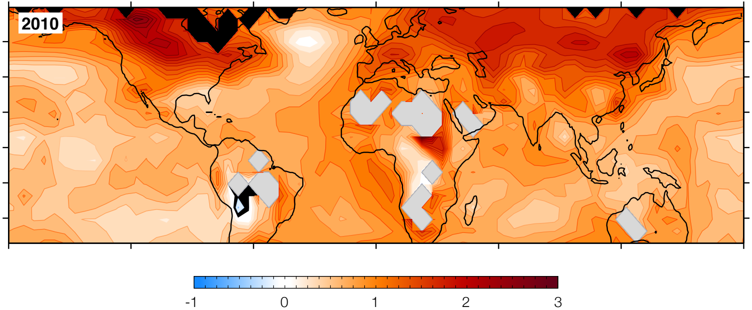
Caltech MIT JPL Jet Propulsion Laboratory NPS PRAESTANTIA PER SCIENTIAM 1909

[1]Schneider et al. 2017.

Table of Contents

Table of Contents

Global temperatures are increasing

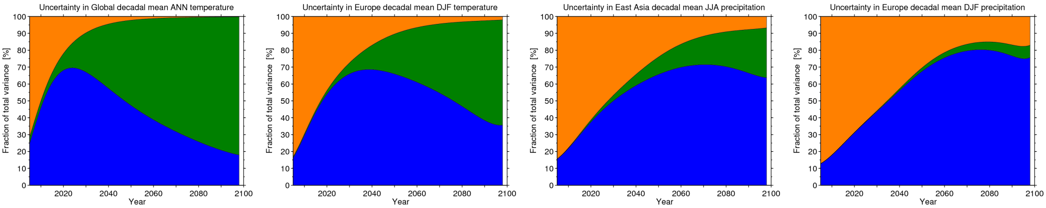


Temperature change (°C) from 1850s through 2010s

Schneider & Held, J. Climate, 2001; update http://climate-dynamics.org/videos

But climate model predictions are uncertain

( ) Internal Variability, ( ) Scientific (model) Uncertainty, ( ) Scenario Uncertainty.

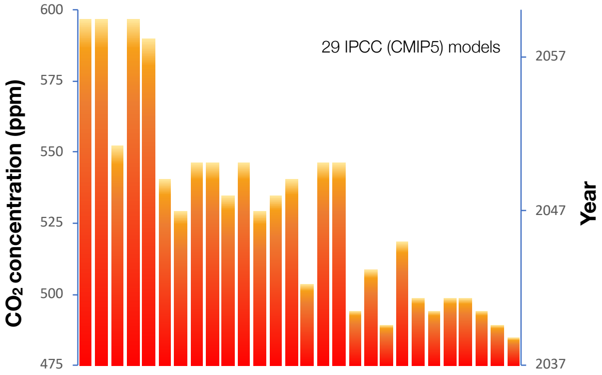


[IPCC 2013, Climate Change 2013: The Physical Science Basis, Working Group 1 Contribution to the fith Assessment Report (AR5)]

CliMA
CLIMATE MODELING ALLIANCE

Caltech

CO_2 concentration that breaches 2°C warming?

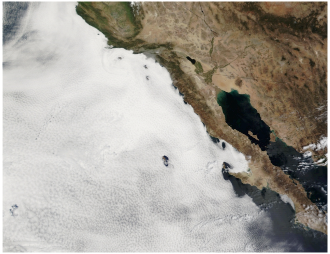


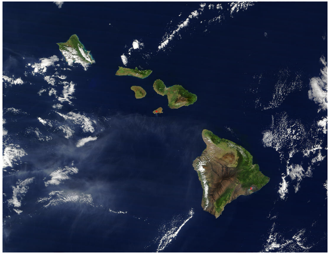
Schneider et al., *Nature Climate Change* 2017

CliMA
CLIMATE MODELING ALLIANCE

Caltech

Ability to predict cloud cover is key





http://eoimages.gsfc.nasa.gov

Stratocumulus: colder

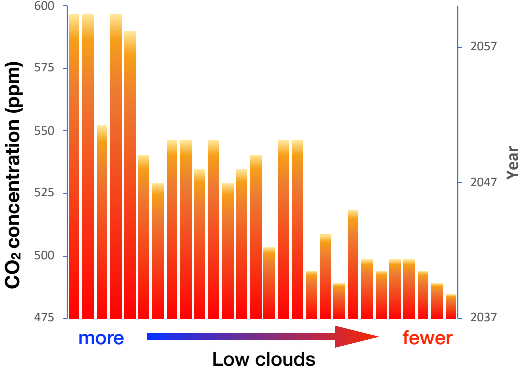
Cumulus: warmer

We don't know if we will get more low clouds (damped global warming), or fewer low clouds (amplified warming) with rising CO_2 levels

CliMA CLIMATE MODELING ALLIANCE

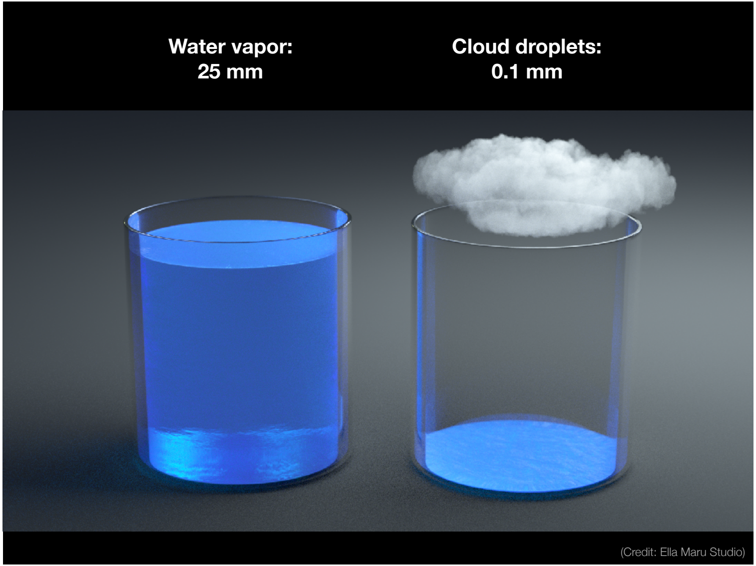
Caltech

Cloud cover correlates with CO_2 concentration predictions



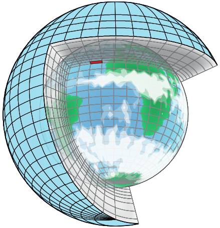
Schneider et al., *Nature Climate Change* 2017

Range of scales hampers ability to predict cloud cover

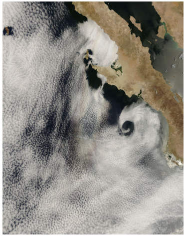


(Credit: Ella Maru Studio)

Range of scales hampers ability to predict cloud cover



Global model:
~10-50 km resolution

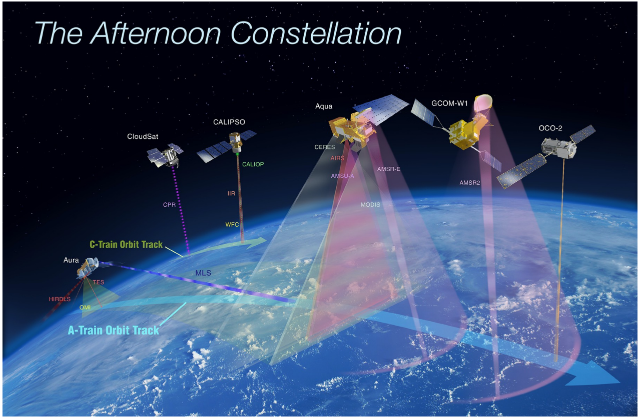


NASA MODIS

Cloud scales: ~10-100 m

Subgrid-scale processes (e.g., clouds and turbulence) are represented in ad-hoc fashion (not data-driven)

There is available satellite data (JPL)

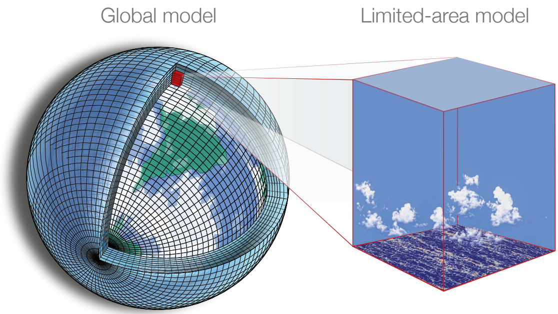


[https://atrain.nasa.gov/]

CliMA
CLIMATE MODELING ALLIANCE

Caltech

There are affordable limited-area simulations



Thousands of high-resolution simulations can be embedded in global model in a distributed computing environment (cloud), and the global model can learn from them

The Climate Modelling Alliance (CliMA)[2]

Our Aims

- Redesign physical models to better resolve clouds.
- Data assimilation, uncertainty quantification and machine learning framework to learn about parameter uncertainty within physical models.
- Include data from both high resolution simulation and observations.
- Julia language framework unified across all components. (v0.1.0 released!)

[2] Schneider et al. 2017.

CliMA
CLIMATE MODELING ALLIANCE

Caltech

Table of Contents

Going from parameters to data[3]

Ingredients

- Parameter (prior) distribution $\theta \sim \mu_0$, parameter space $\theta \in \Theta$.
- Data space $y \in \mathcal{Y}$.
- Forward map $\mathcal{G}\colon \Theta \to \mathcal{Y}$,
- Noise covariance Σ.

Recipe for data y:

$$y = \mathcal{G}(\theta) + N(0, \Sigma) \tag{1}$$

[3] Stuart 2010; Garbuno-Inigo et al. 2019.

The perfect climate model setting[4]

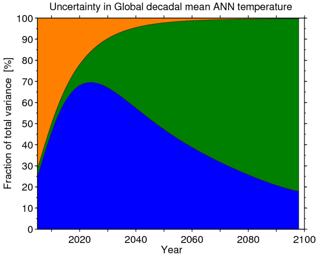
Recipe for data y:

$$y = \mathcal{G}_T(\theta; z_0) = \mathcal{G}_\infty(\theta) + N(0, \Sigma) \quad (2)$$

- Forward map $\mathcal{G}_T$, T time averaged forward run of length T.
- No observational noise η ('perfect setting'), **but** $\mathcal{G}_T$ is noisy $N(0, \Sigma)$.

Perfect doesn't mean easy!

- $\mathcal{G}_T$ is noisy, no access to $G_\infty(\theta)$.
- $\mathcal{G}_T$ expensive, especially for large T.
- $\mathcal{G}_T$ is non-differentiable.



[4] Dunbar et al. 2020.

An idealized GCM[5]

Idealized moist GCM: Aquaplanet. T21 Spectral discretization (32 discrete latitudes).

Moist convection in quasi-equilibrium (Betts Miller type). Features sources in temperature and humidity equations

$$\text{Source}(x) = \frac{x - x_{\text{ref}}(\alpha)}{\tau}$$

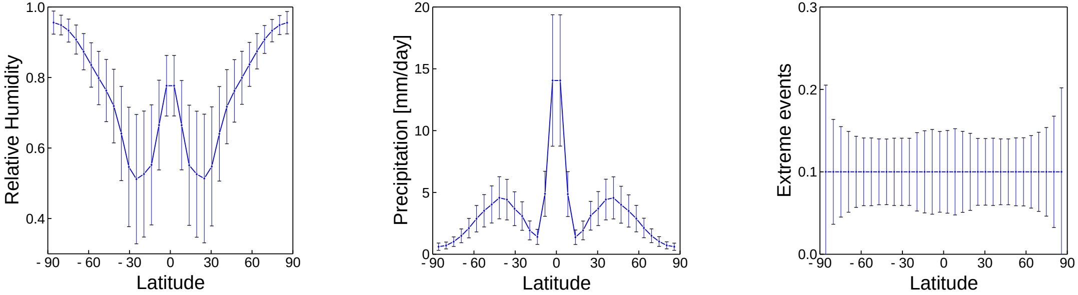
$0 < \alpha < 1$ relative humidity. $0 < \tau$ relaxation time. Priors enforce constraints.

Properties: stationary statistics, zonally symmetric, $\theta = (\alpha, \tau)$

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

The data y

3 time averaged ($T = 30$ day) quantities



$$y^\dagger = \mathcal{G}_T(\theta^\dagger; z_0), \text{ where } \theta^\dagger = (\alpha^\dagger, \tau^\dagger) = (0.7,\ 2\text{hours})$$

Table of Contents

(Traditional) Optimization approach: find optimal parameter set

Calibration objective: Find optimal $\theta^* \in \Theta$ that best fits with prior μ_0 and data $y^\dagger$.

Ensemble Kalman Inversion[6] (EKI) to find θ^*:

- $(+)$ Cheap! $\sim$ 500 evaluations of $\mathcal{G}_T$,
- $(+)$ Doesn't require differentiation! $\mathcal{G}_T$ non-differentiable,
- $(+)$ Works with noisy model evaluations! noisy objective function,

But...

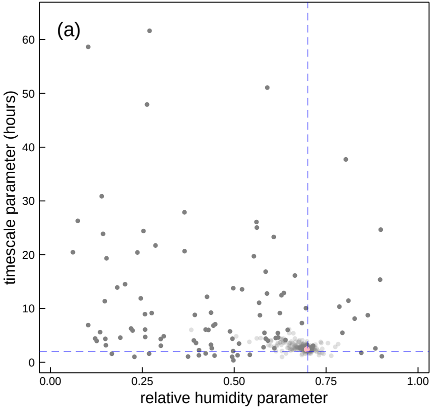
- $(-)$ **Uncertainty greatly underpredicted.** (Ensemble Collapse)

CliMA CLIMATE MODELING ALLIANCE

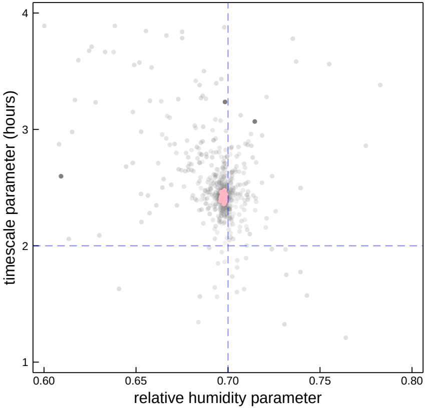
[6] Iglesias, Law, and Stuart 2013.

Caltech

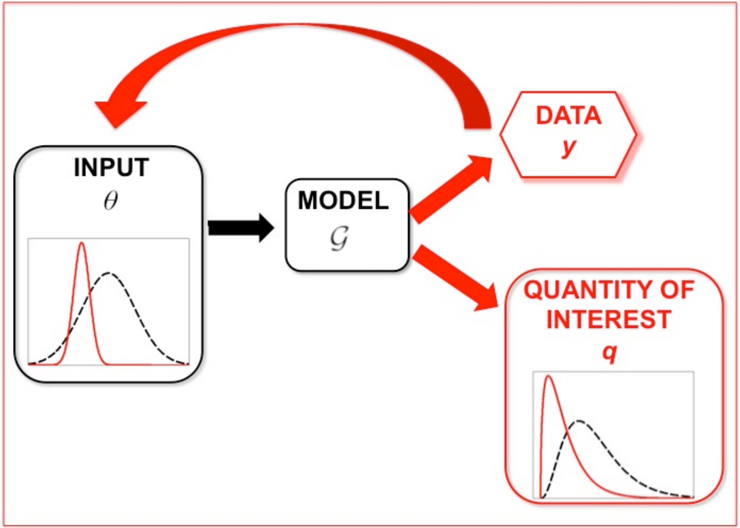
EKI optimization



Zooming in...



What is uncertainty quantification?



Bayesian approach: find the data-informed distribution μ^y

Sampling objective: Sample the distribution $\theta \mid y^\dagger$, given a prior μ_0.

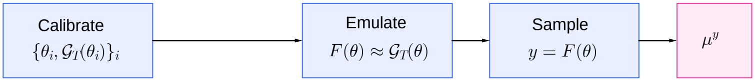
Markov Chain Monte Carlo (MCMC):

- $(+)$ **Uncertainty is quantified!**

But...

- $(-)$ Expensive. $\sim 500,000$ evaluations of $\mathcal{G}_T$,
- $(-)$ Gets stuck in local minima. noisy objective function.

Calibrate-Emulate-Sample[7]



Calibrate-Emulate-Sample (CES):

- (+) **Uncertainty is quantified!**
- (+) Cheap! $\sim$ 500 evaluations of $\mathcal{G}_T$,
- (+) Doesn't require differentiation! $\mathcal{G}_T$ non-differentiable,
- (+) Works with noisy model evaluations! Emulator smoothes the objective.

CliMA CLIMATE MODELING ALLIANCE

[7] Cleary et al. 2019; Dunbar et al. 2020.

Caltech

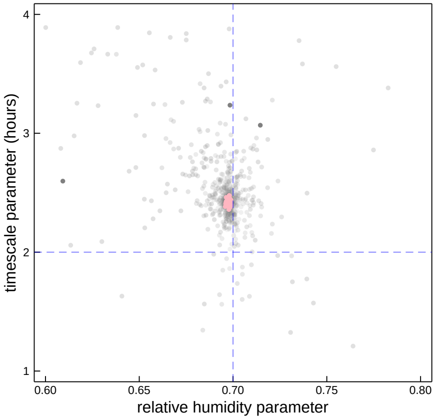
Calibrate-Emulate-Sample[7]

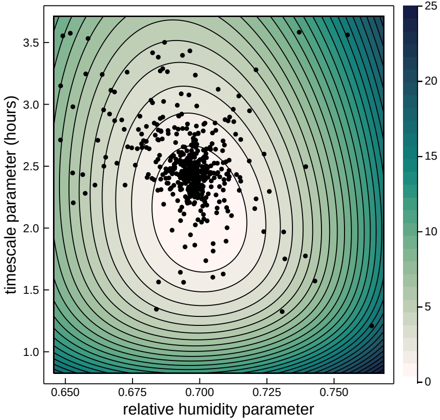
Calibrate $\{\theta_i, \mathcal{G}_T(\theta_i)\}_i$ → Emulate $F(\theta) \approx \mathcal{G}_T(\theta)$ → Sample $y = F(\theta)$ → μ^y

Calibrate **EKI** → Emulate **GP** → Sample **MCMC** → μ^y

[7] Cleary et al. 2019; Dunbar et al. 2020.

Learning: calibrate for training points

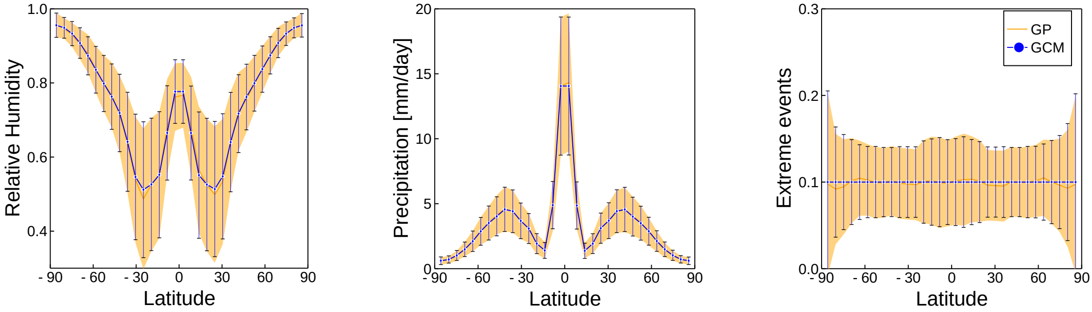




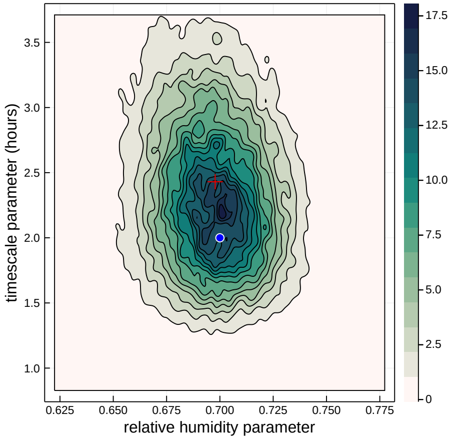
CliMA
CLIMATE MODELING ALLIANCE

Caltech

Learning: emulate with Gaussian process



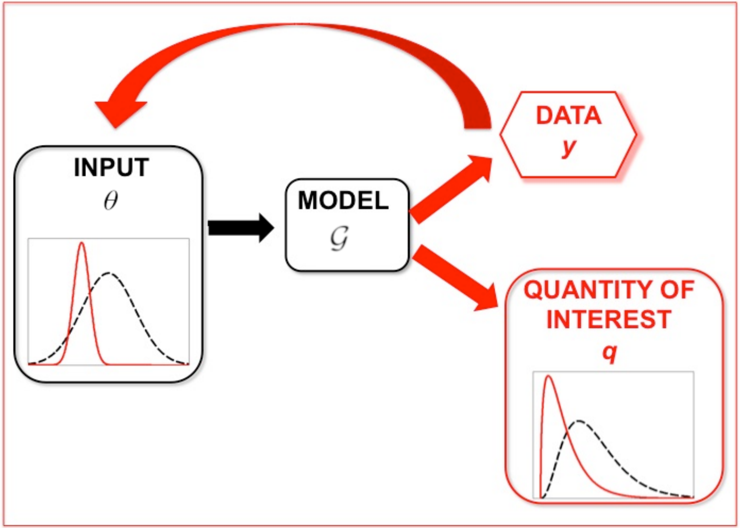
Learning: Sample with MCMC



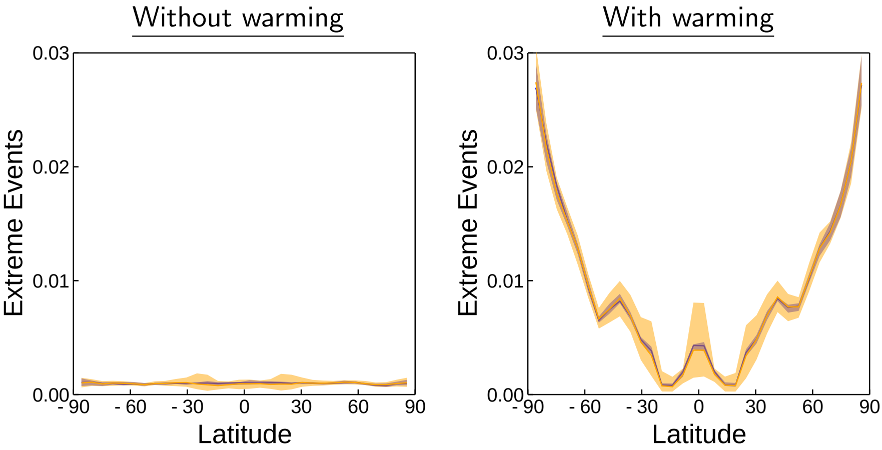
(+) EKI Optimal θ^*,

(●) Truth $\theta^\dagger$.

Now we can make predictions of QoIs!



A warming experiment



(■) without parameter uncertainty, (■) with parameter uncertainty.

Table of Contents

CliMA
CLIMATE MODELING ALLIANCE

Caltech

Summary

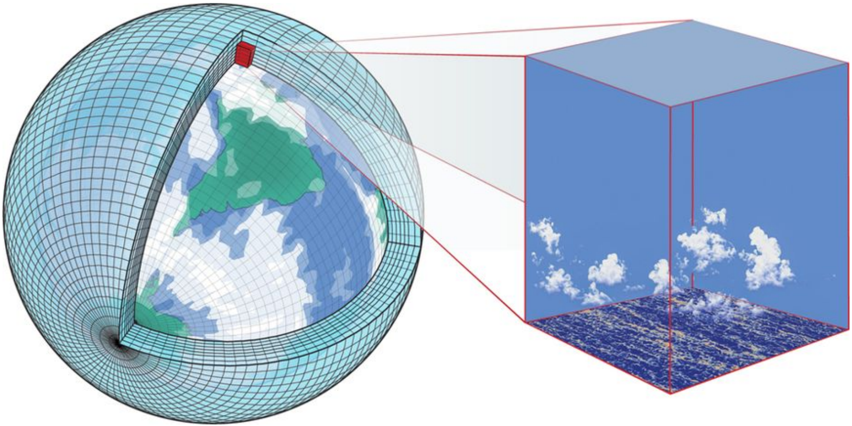
- Uncertainty is important, particularly parameter uncertainty.
- To make trustworthy predictions we must quantify it.
- CES pipeline automates and accelerates uncertainty quantification for expensive and noisy models in a black-box fashion.[8] (Julia package forthcoming on Github)
- We benchmarked this for moist convection in an idealized aquaplanet.[9]

Looking ahead

- Structural model error,
- Higher dimensional parameter learning (and non-parametric function learning),
- Online learning,
- **Automated optimal design...**

[8] Cleary et al. 2019.
[9] Dunbar et al. 2020.

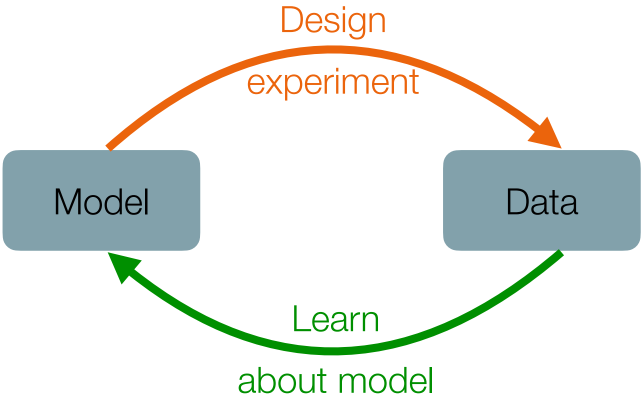
Optimal placement of limited-area simulations



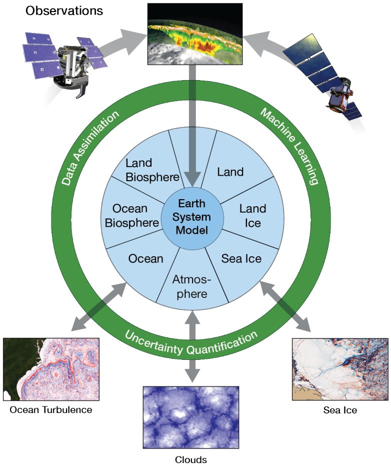
CliMA
CLIMATE MODELING ALLIANCE

Caltech

CES automates and accelerates the scientific loop!



The bigger picture



Targeted High-Resolution Simulations

References

- Cleary, Emmet et al. (2019). "Calibrate, Emulate, Sample". In: *arXiv preprint arXiv:1912.*
- Dunbar, Oliver R. A. et al. (2020). "Calibration and Uncertainty Quantification of Convective Parameters in an Idealized GCM". In: *arXiv preprint - to appear within 1 month.*
- 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.
- Garbuno-Inigo, Alfredo et al. (2019). "Interacting Langevin Diffusions: Gradient Structure And Ensemble Kalman Sampler". In: arXiv: 1903.08866 [math.DS].
- Iglesias, Marco A, Kody JH Law, and Andrew M Stuart (2013). "Ensemble Kalman methods for inverse problems". In: *Inverse Problems* 29.4, p. 045001.
- 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.
- Schneider, Tapio 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.
- Stuart, Andrew M (2010). "Inverse problems: a Bayesian perspective". In: *Acta Numerica* 19, pp. 451–559.