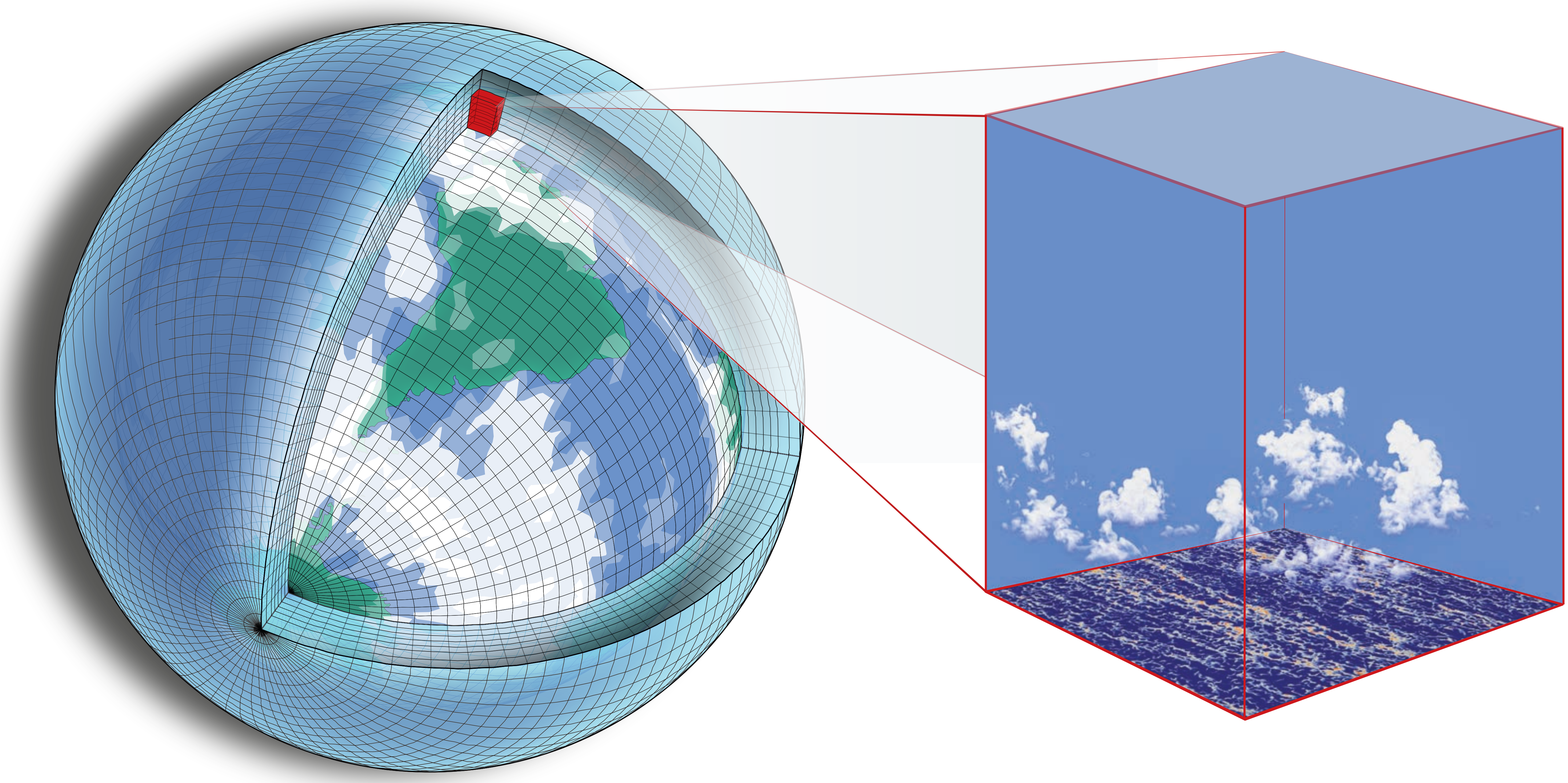
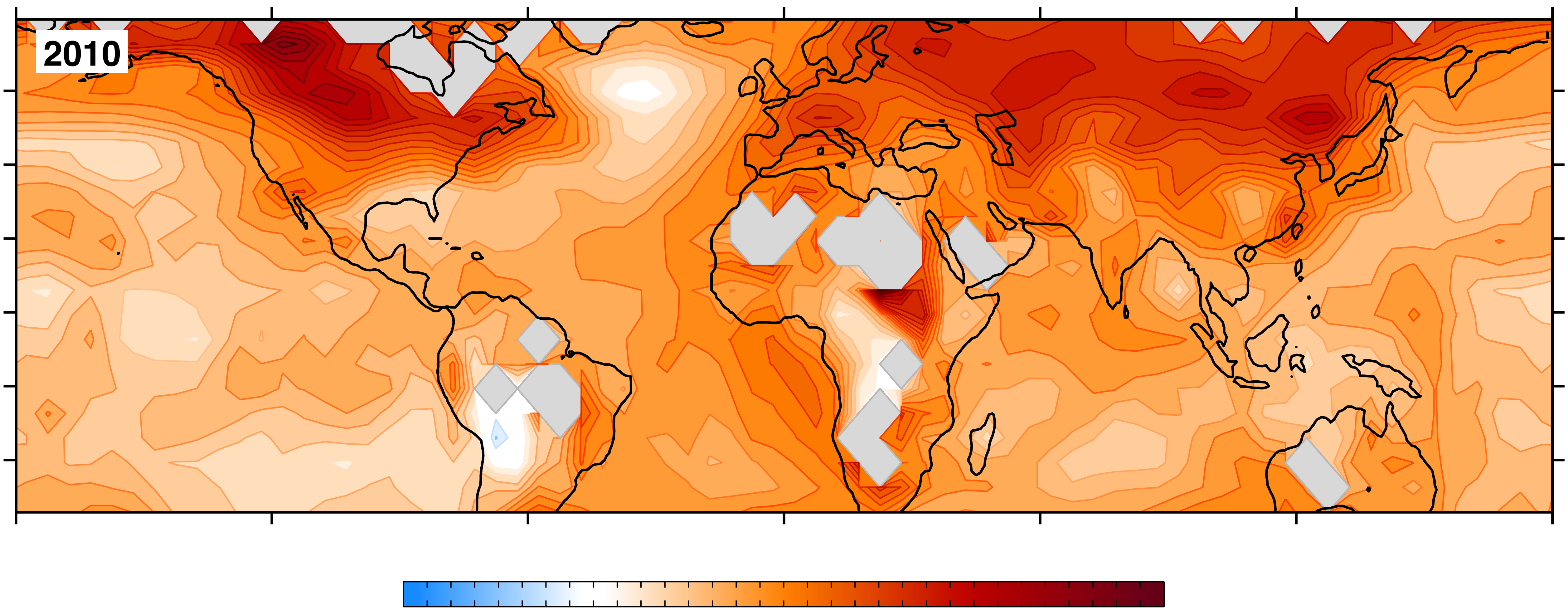


Earth System Modeling 2.0: Toward Data-Informed Climate Models With Quantified Uncertainties

Tapio Schneider, Yair Cohen, Anna Jaruga, Jia He, Ignacio Lopez-Gomez, Emmet Cleary, Alfredo Garbuno, Andrew Stuart

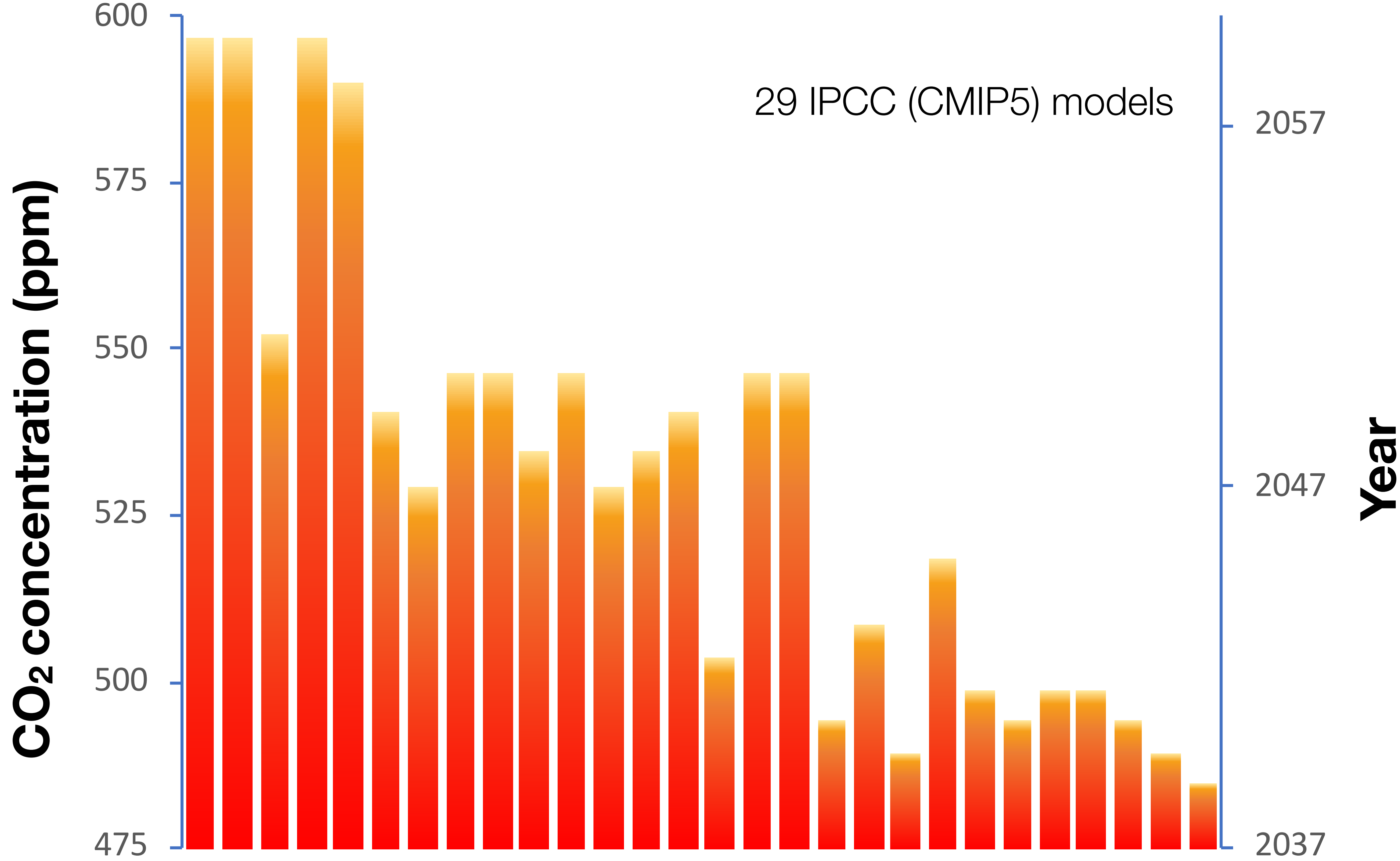
Temperatures have risen over the past 150 years



Temperature change (ºC) from 1850s through 2010s

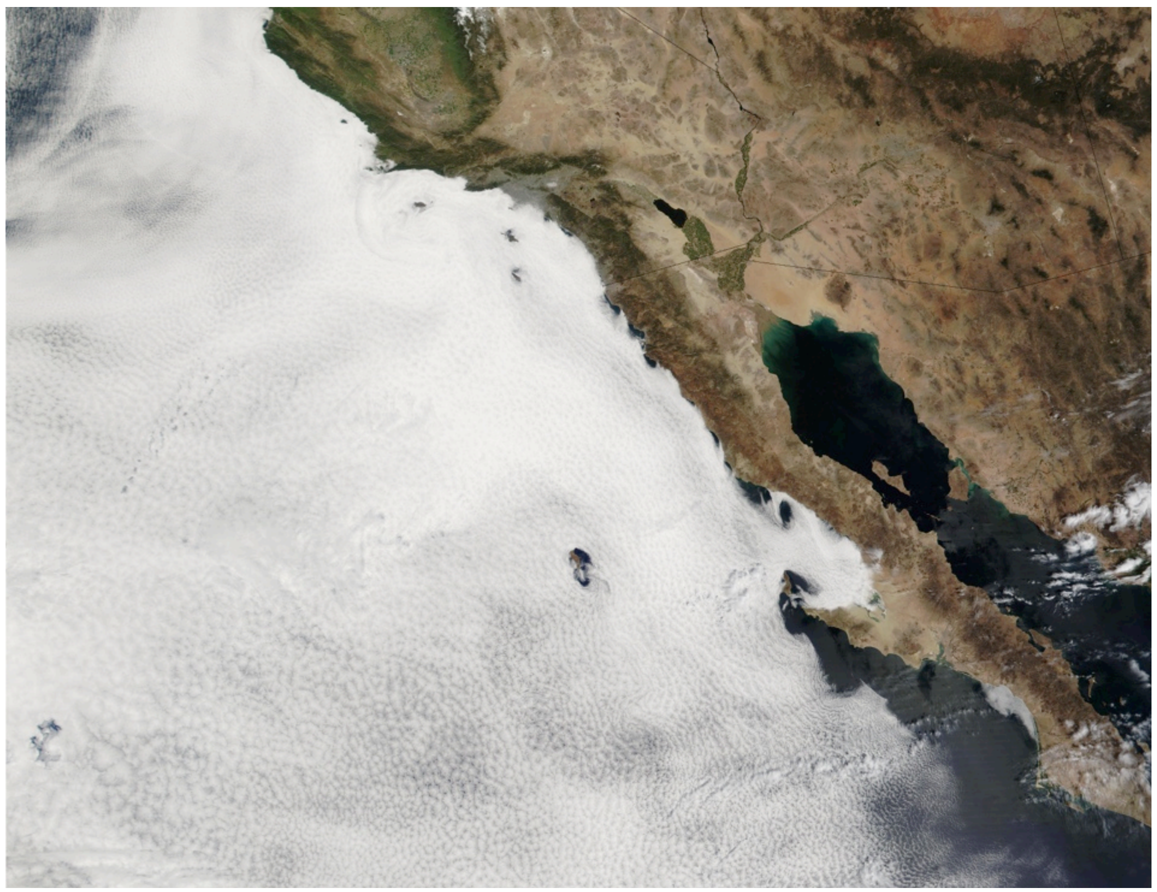
Schneider & Held, J. Climate, 2001; update http://climate-dynamics.org/videos

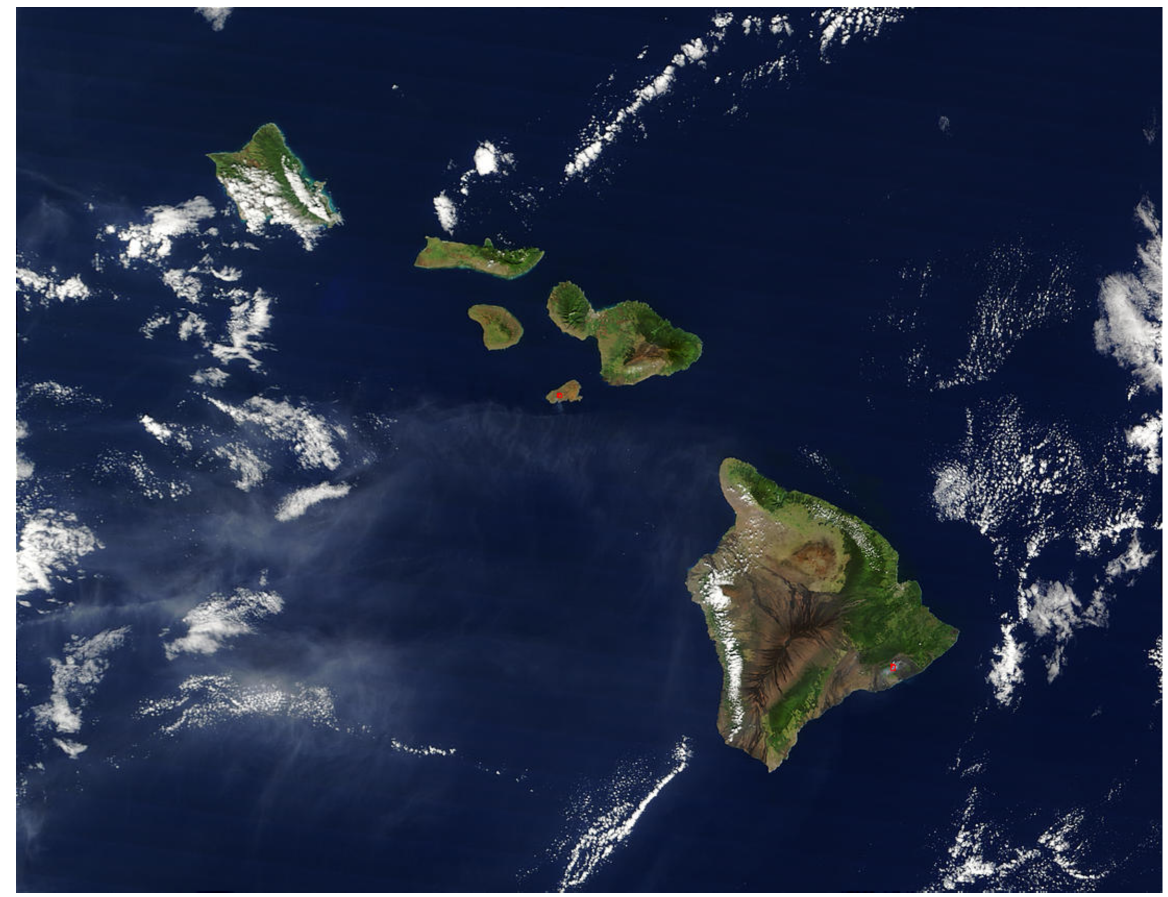
But climate predictions remain uncertain: E.g., the CO_2 concentration at which 2°C warming threshold is crossed varies widely across models



Schneider et al., *Nature Climate Change* 2017

The primary (but not only) source of uncertainties in climate predictions is the representation of low clouds in models





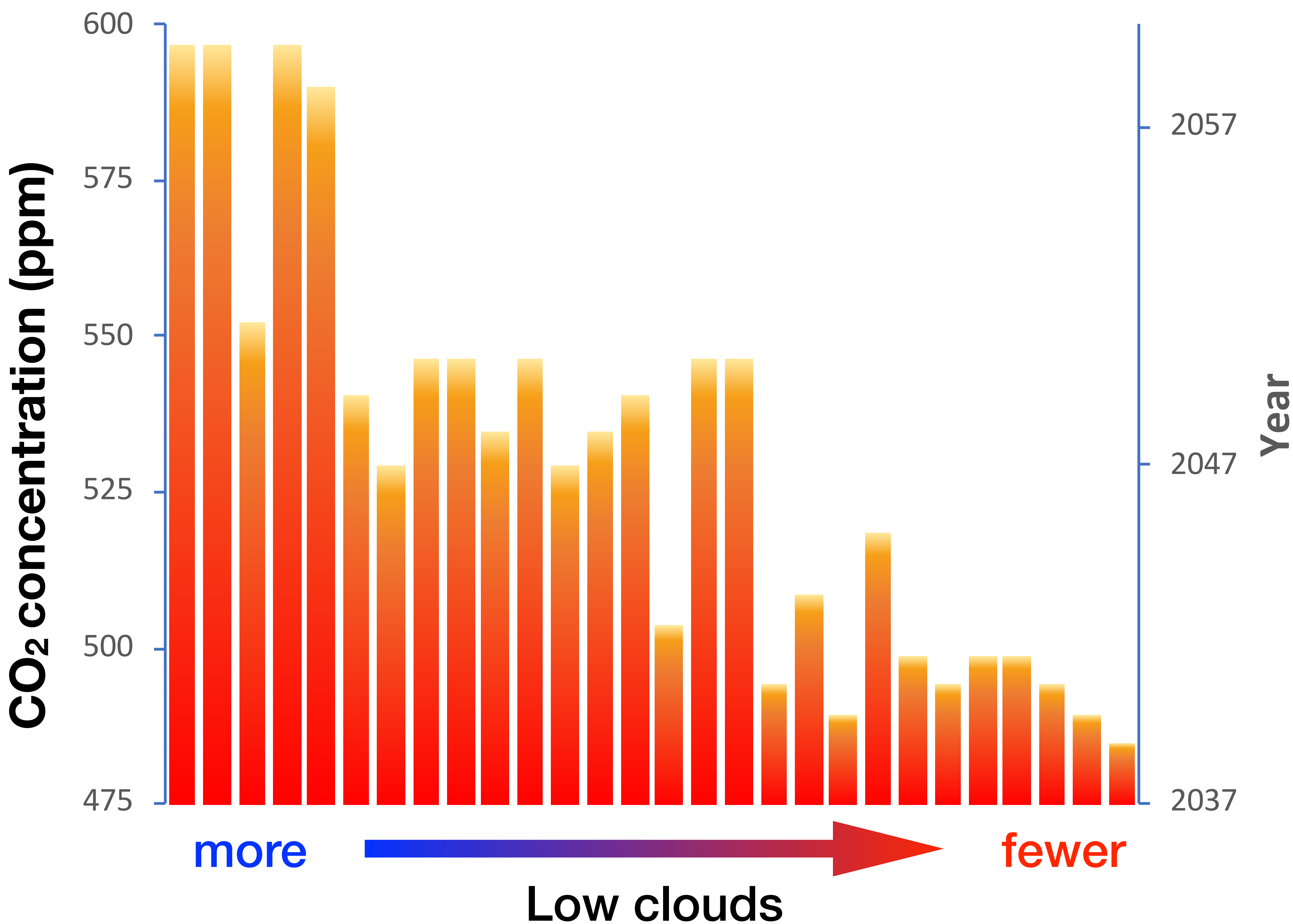
http://eoimages.gsfc.nasa.gov

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

Spread in predictions for next ~30-50 years is dominated by uncertainties in low clouds; uncertainties are poorly quantified



Schneider et al., *Nature Climate Change* 2017

More accurate climate projections with quantified uncertainties would enable…

- Data-driven decisions about infrastructure planning, e.g.,
    - How high a sea wall should New York City build to protect itself against storm surges in 2050?
    - What water management infrastructure is needed to ensure food and water security in sub-Saharan Africa?
- Rational resource allocation for climate change adaptation: costs estimated to reach >$200B annually by 2050 (UNEP 2016)
- Realization of the socioeconomic value of more accurate predictions, which is estimated to lie in the trillions of USD (Hope 2015; CDP 2019)

"The climate information needs of Federal, State, Local, and Private Sector decision makers are not being fully met." U.S. GAO (2015)

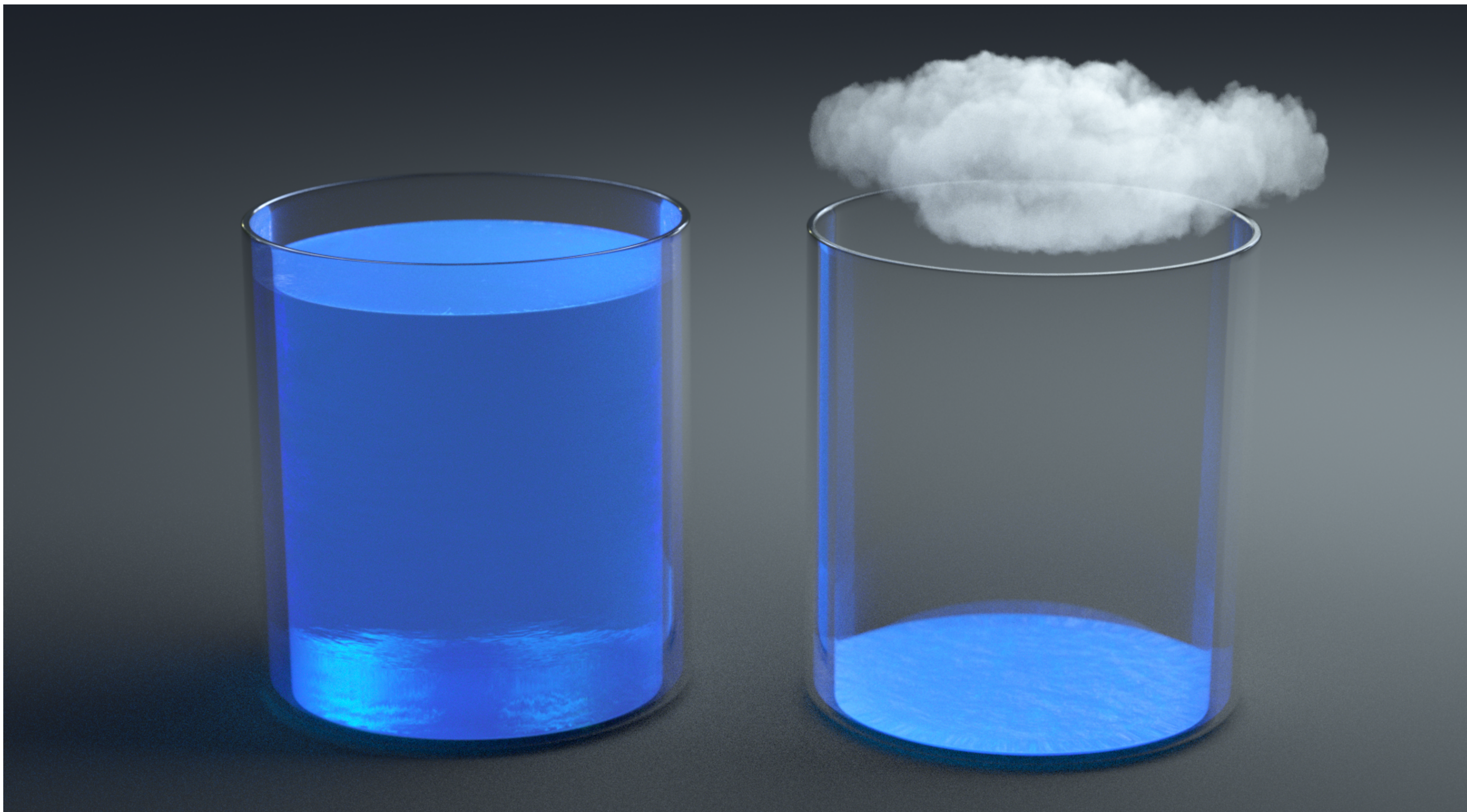
Clouds in climate predictions:

Why are they difficult but important?

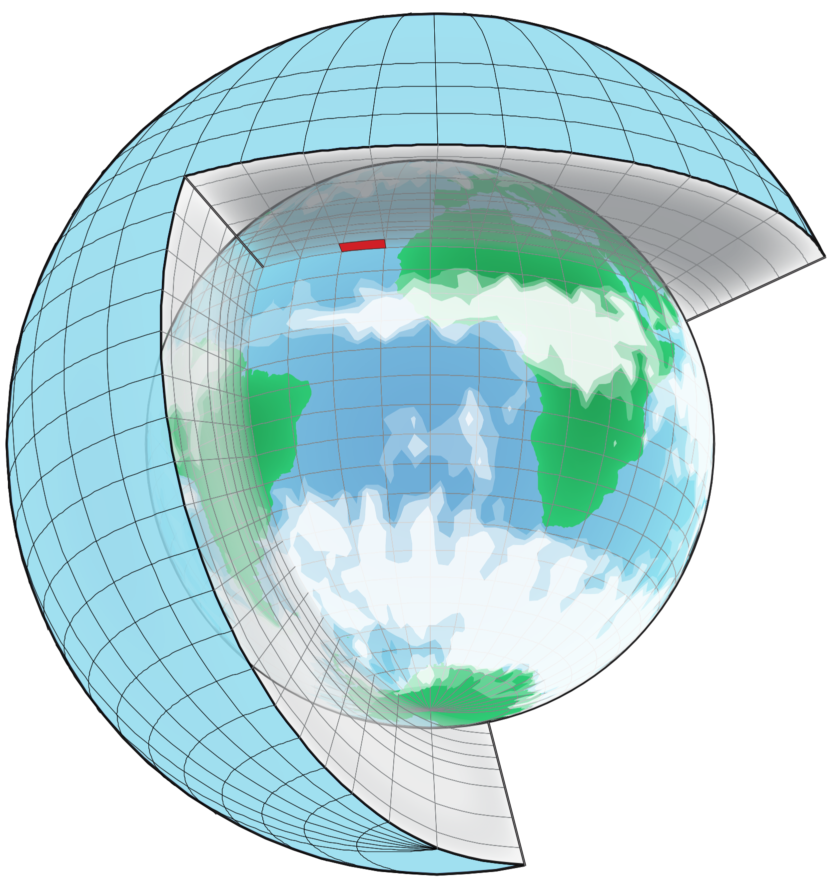
Clouds are difficult to simulate because they contain very little water

Water vapor:
25 mm

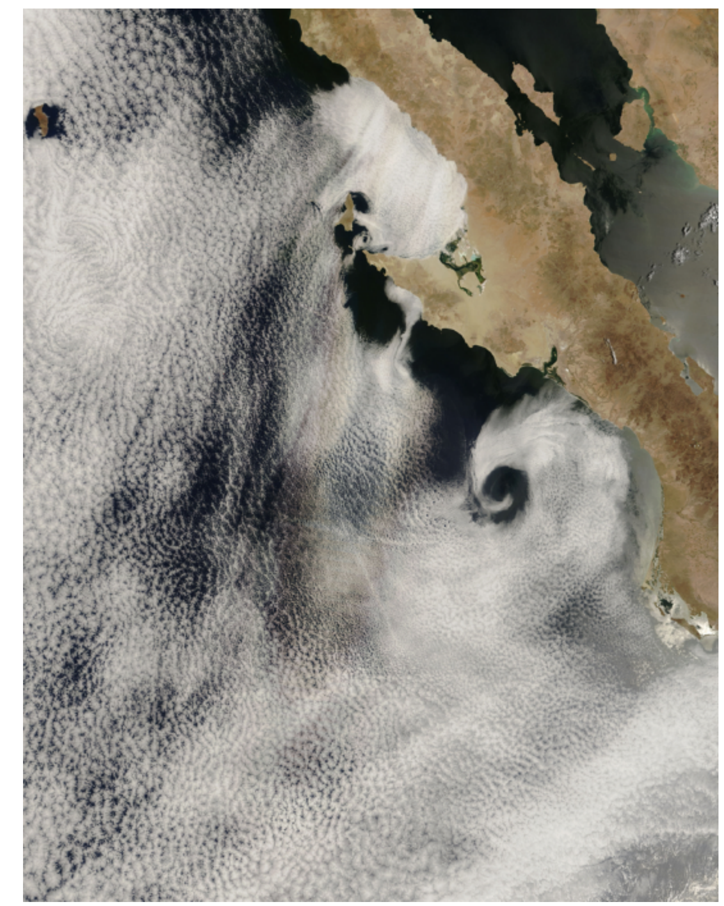
Cloud droplets:
0.1 mm



The small-scale cloud-controlling processes cannot be computed globally in climate models



Global model:
~10-50 km resolution

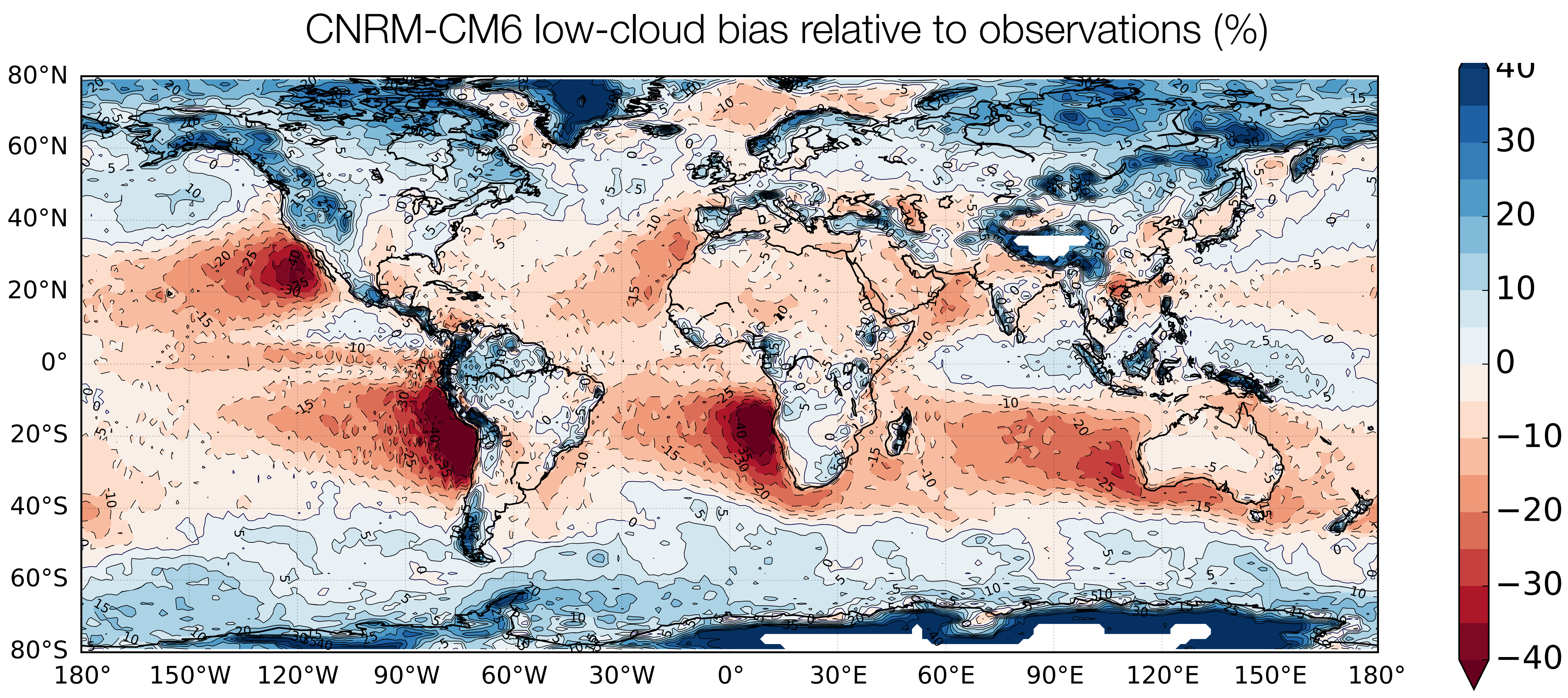


NASA MODIS

Cloud scales: ~10-100 m

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

No climate model simulates low clouds well, leading to large energy flux biases (~50 W m^{-2})

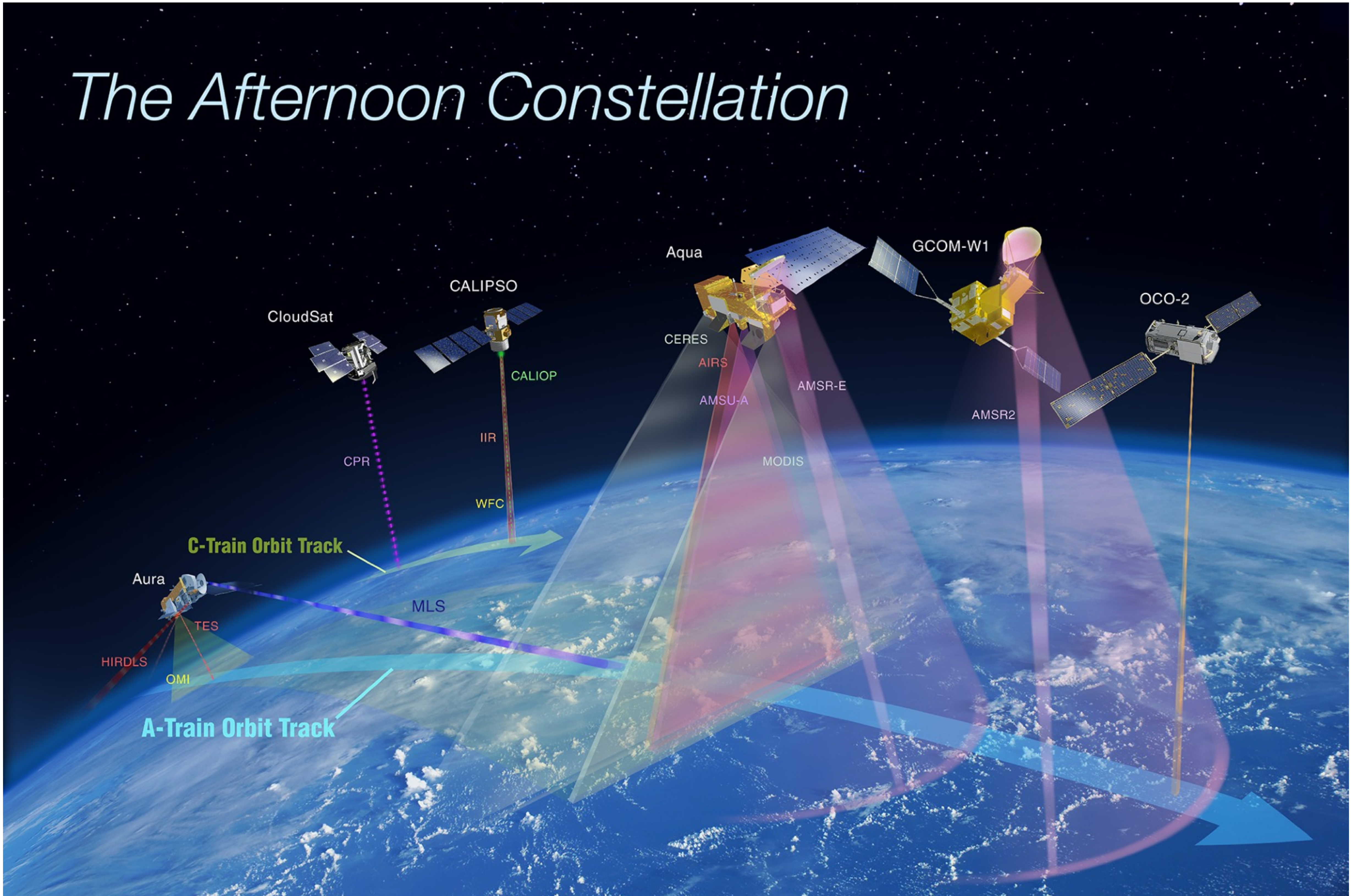


Brient et al. *JAMES*, 2019

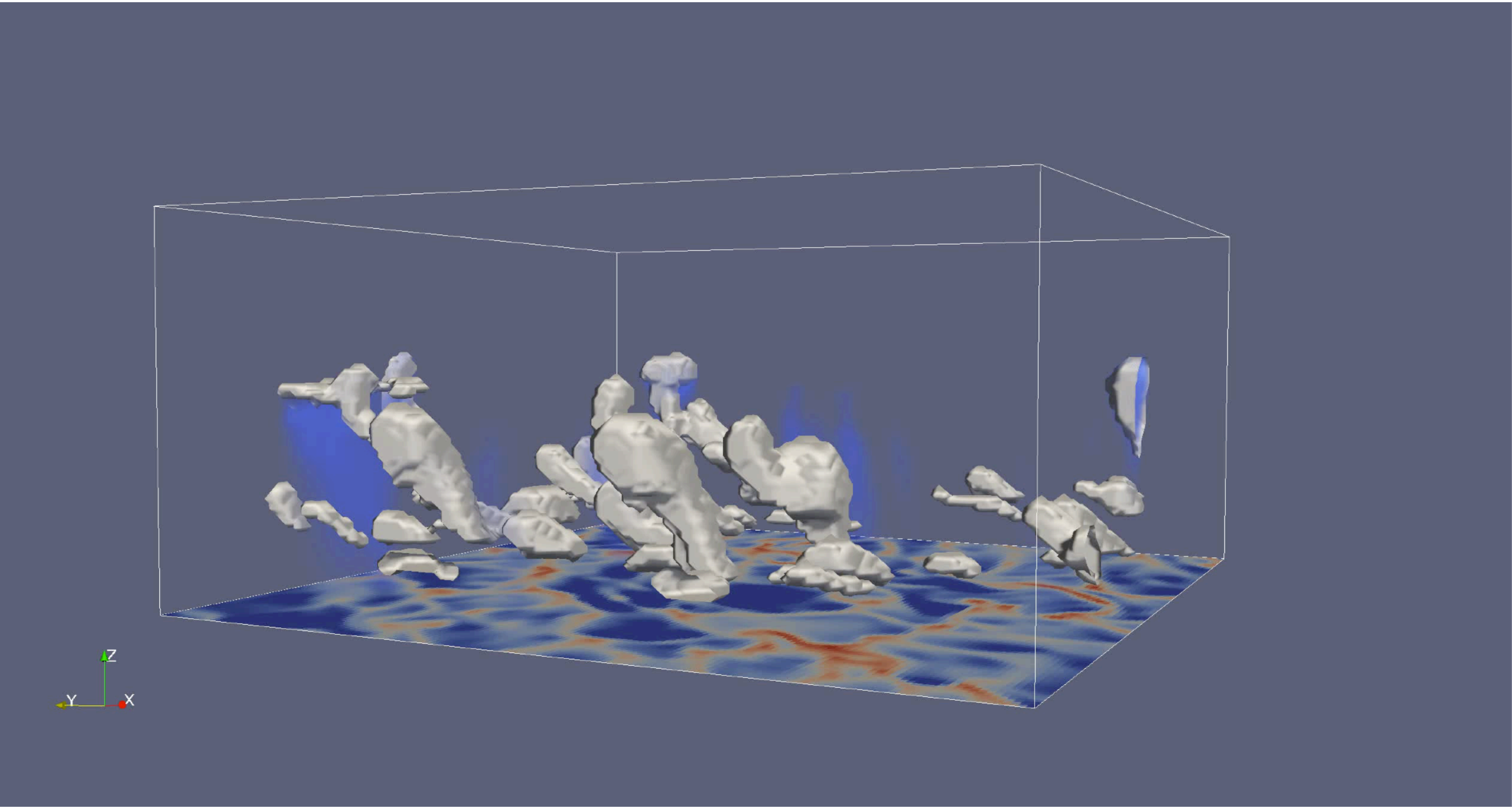
Improving predictions is urgent.

How can we make progress?

We have a wealth of global climate data, whose potential to improve models has not been tapped



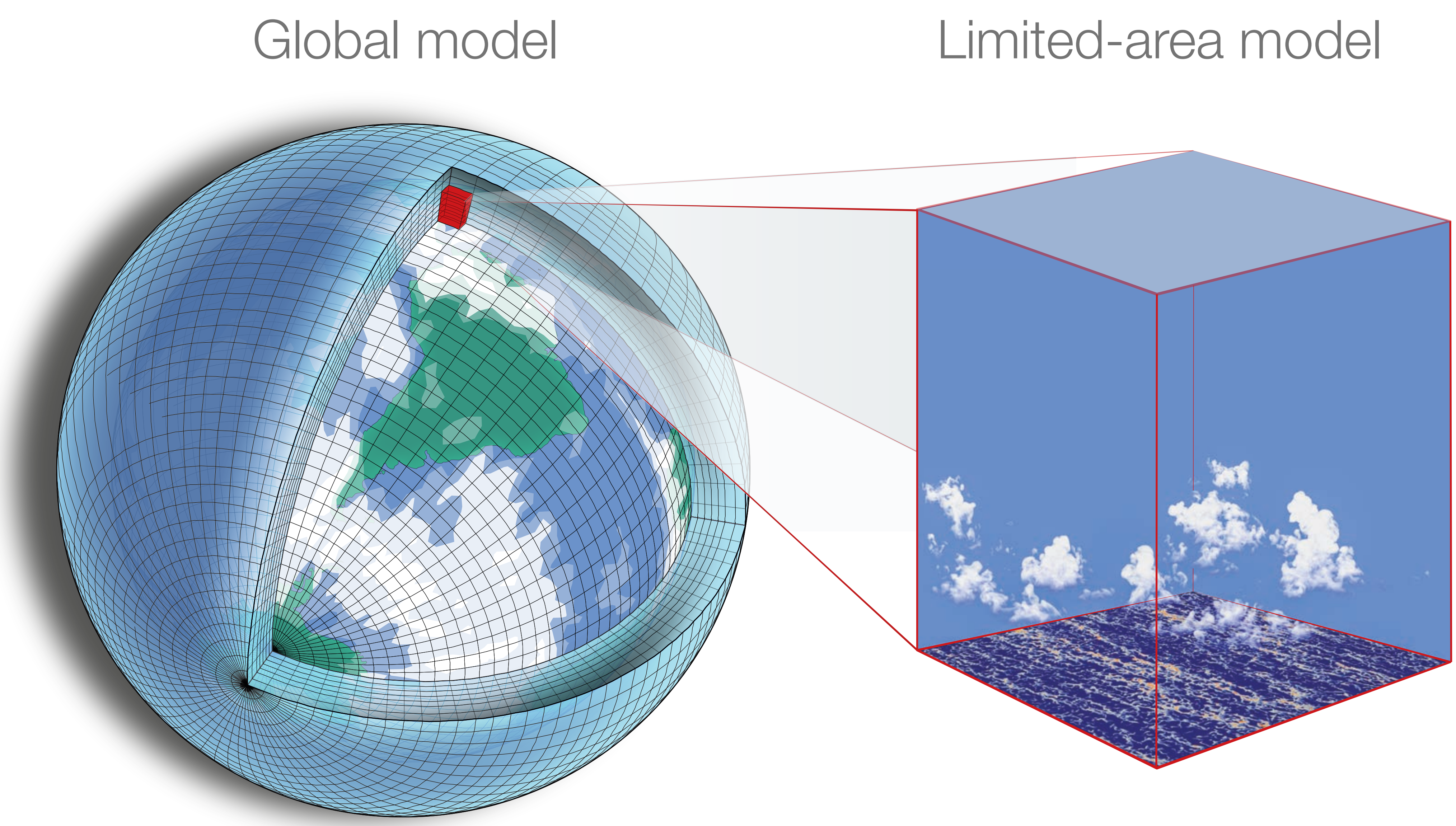
We can also simulate some small-scale processes (e.g., clouds) faithfully, albeit only in limited areas



Large-eddy simulation of tropical cumulus

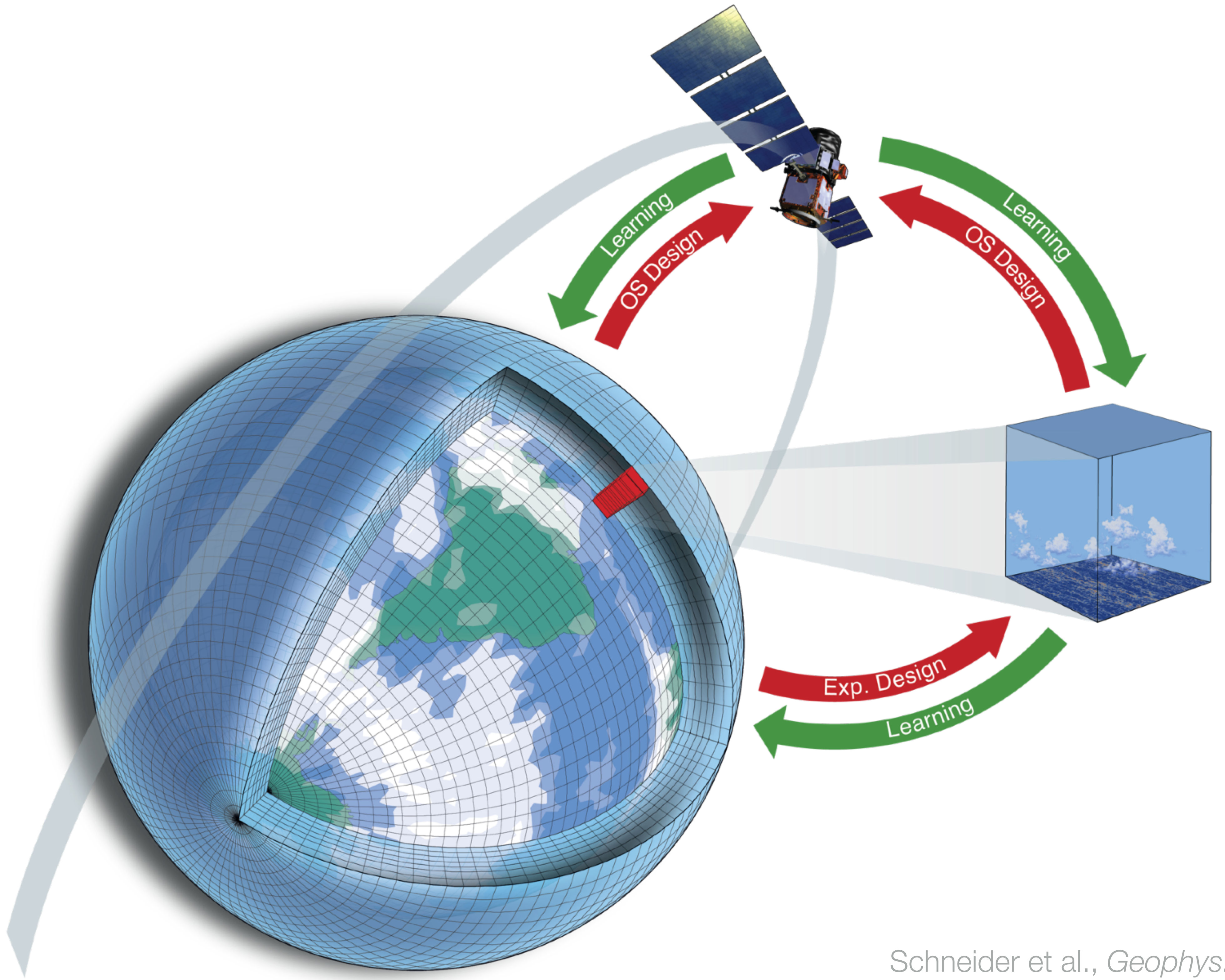
Simulation with PyCLES (Pressel et al. 2015)

Such limited area models can be nested in a global model and can, in turn, inform the global model



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

Vision: build a model that learns automatically both from observations and targeted high-resolution simulations



Schneider et al., *Geophys. Res. Lett.* 2017

Out of these ideas was born CliMA (fall 2018)

CliMA
CLIMATE MODELING ALLIANCE

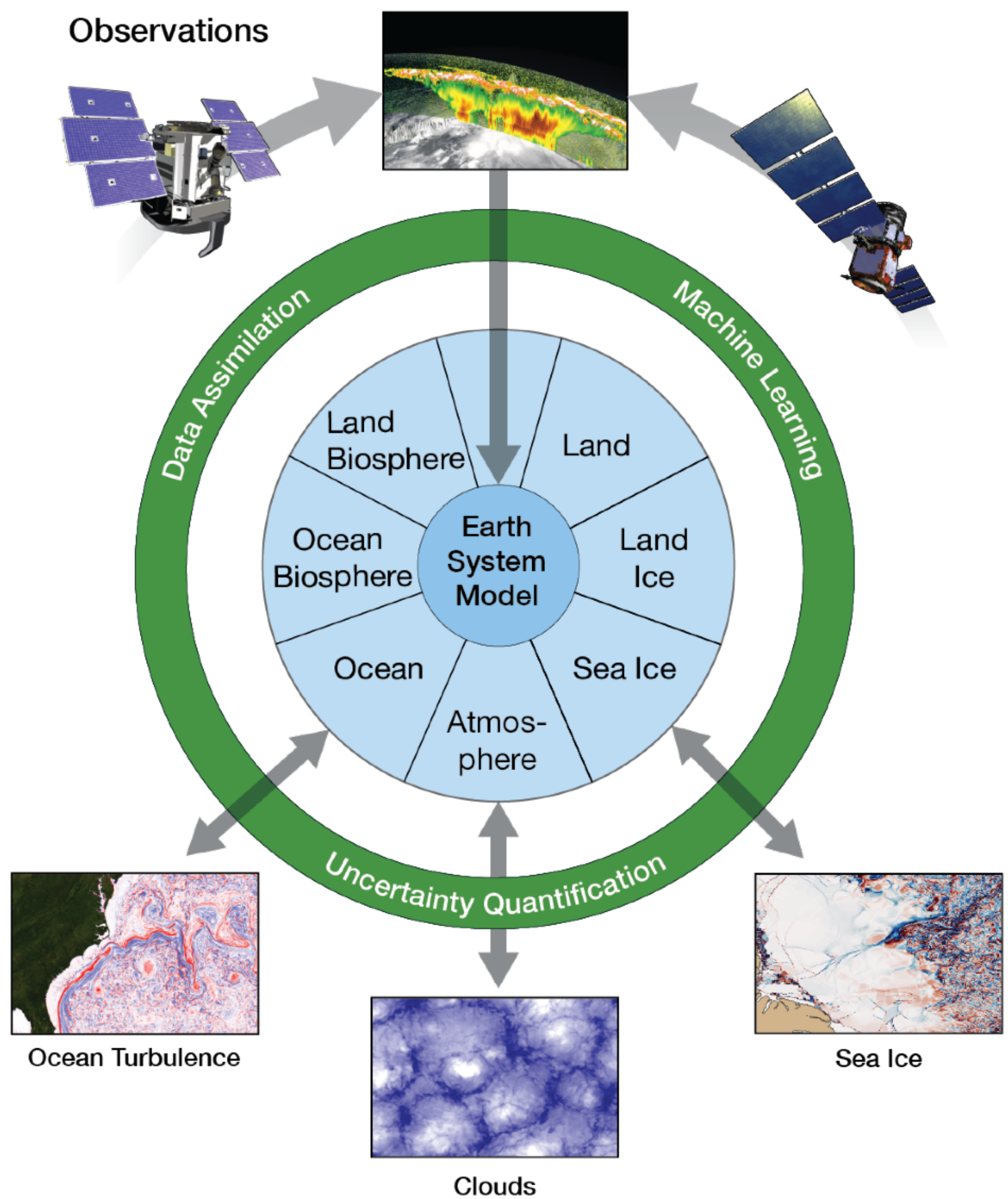
About 50 Earth scientists, engineers, and applied mathematicians at 4 institutions:

Caltech

JPL
Jet Propulsion Laboratory
California Institute of Technology

NPS
PRAESTANTIA PER SCIENTIAM

CliMA is building an Earth system model that wraps a *joint* data assimilation/machine learning layer around all component models



Targeted High-Resolution Simulations

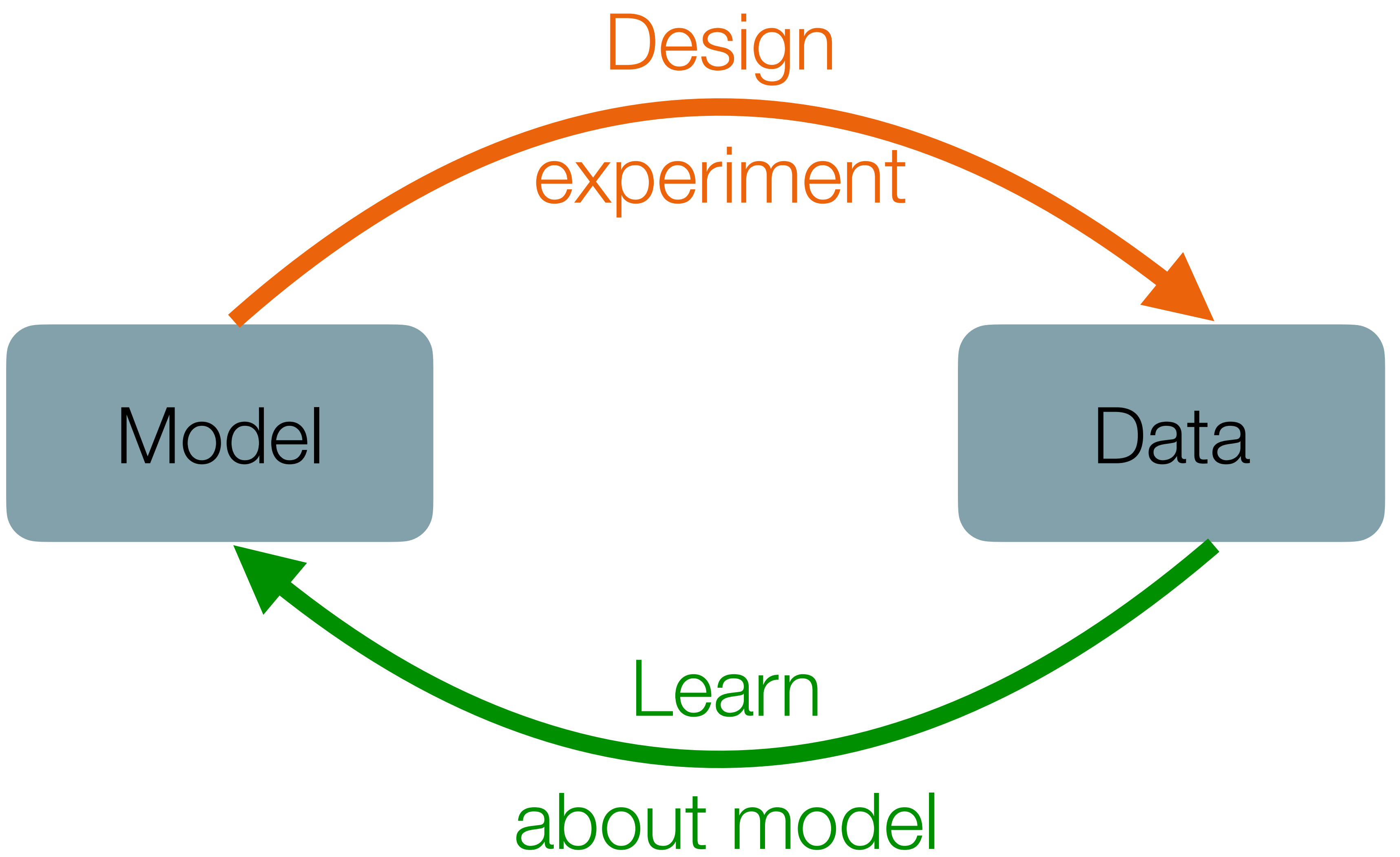
How does this actually work?

“Soft AI”

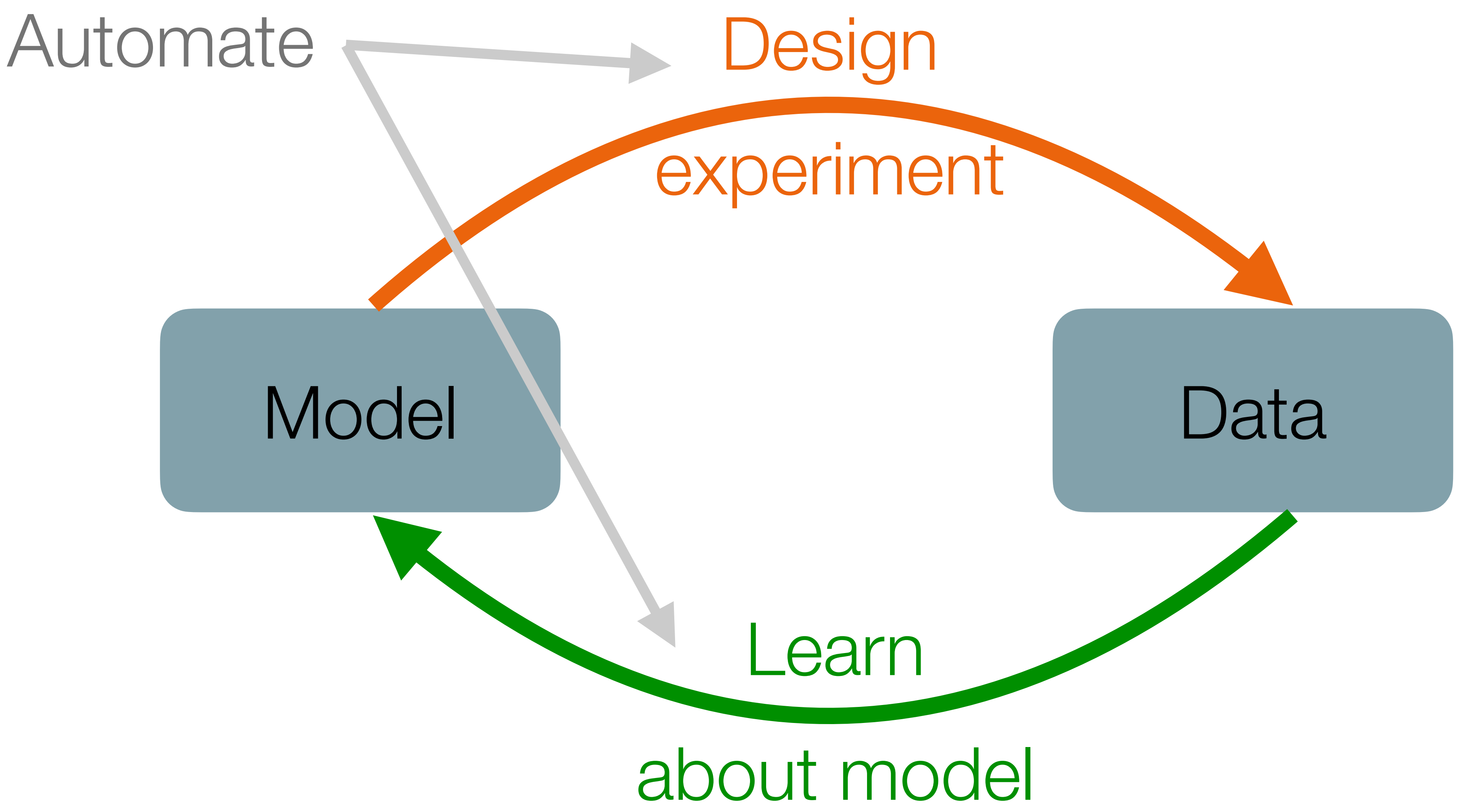
We want to use observations, yet need out-of-sample predictive capabilities and computational feasibility

- We need out-of-sample predictive capabilities (predict a climate we have not seen), yet want to use present-day observations
  - Use known equations of motion to the extent possible to **minimize number of adjustable parameters** and **avoid overfitting**
- Climate data often do not have high temporal resolution but do provide informative time aggregate statistics
  - Learn from **climate statistics** (in contrast to weather states in NWP)
- Running climate models is computationally extremely expensive
  - Need **fast algorithms** for learning about models from data (with judicious use of ML tools)

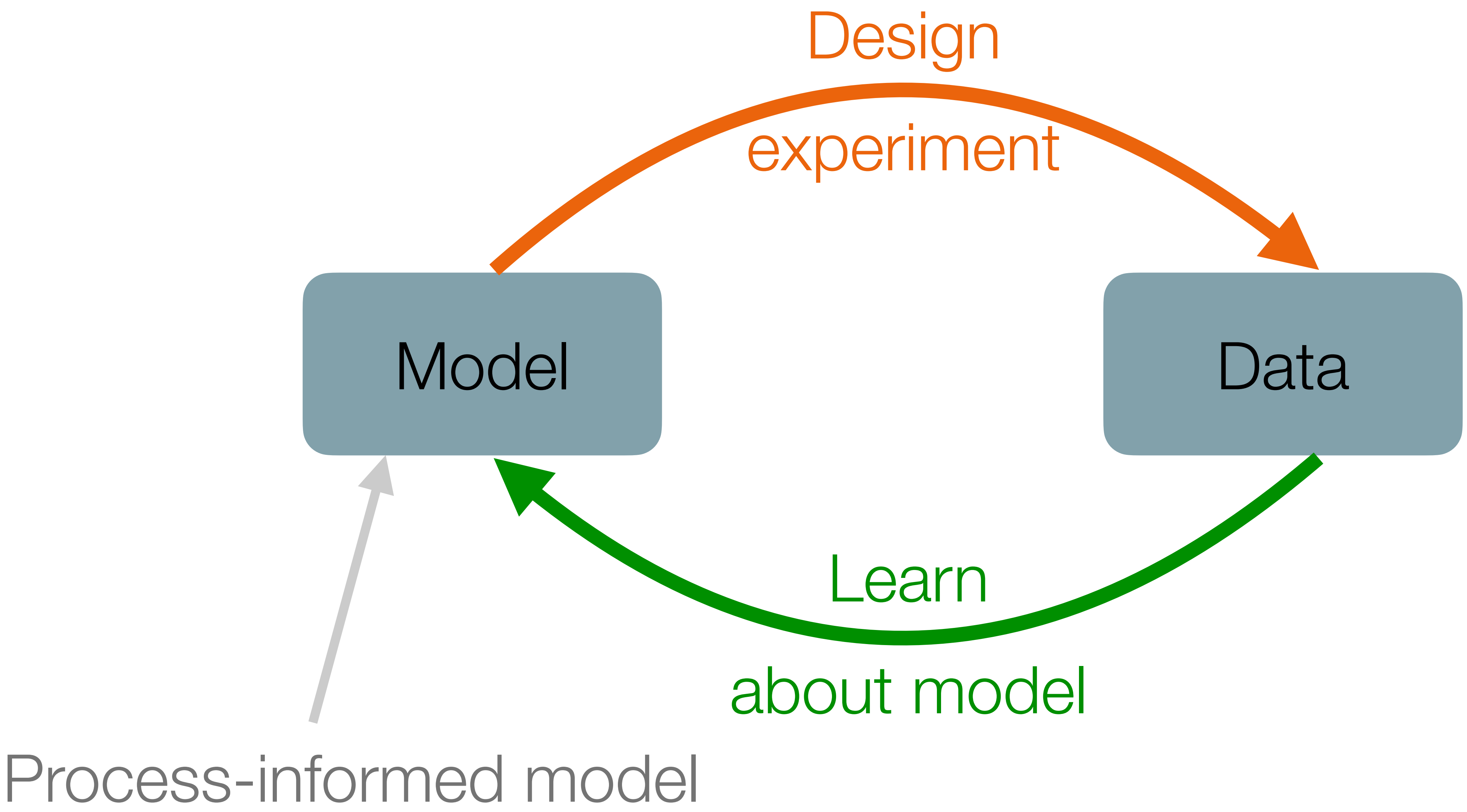
Our strategy: Close, automate, and accelerate the scientific discovery loop



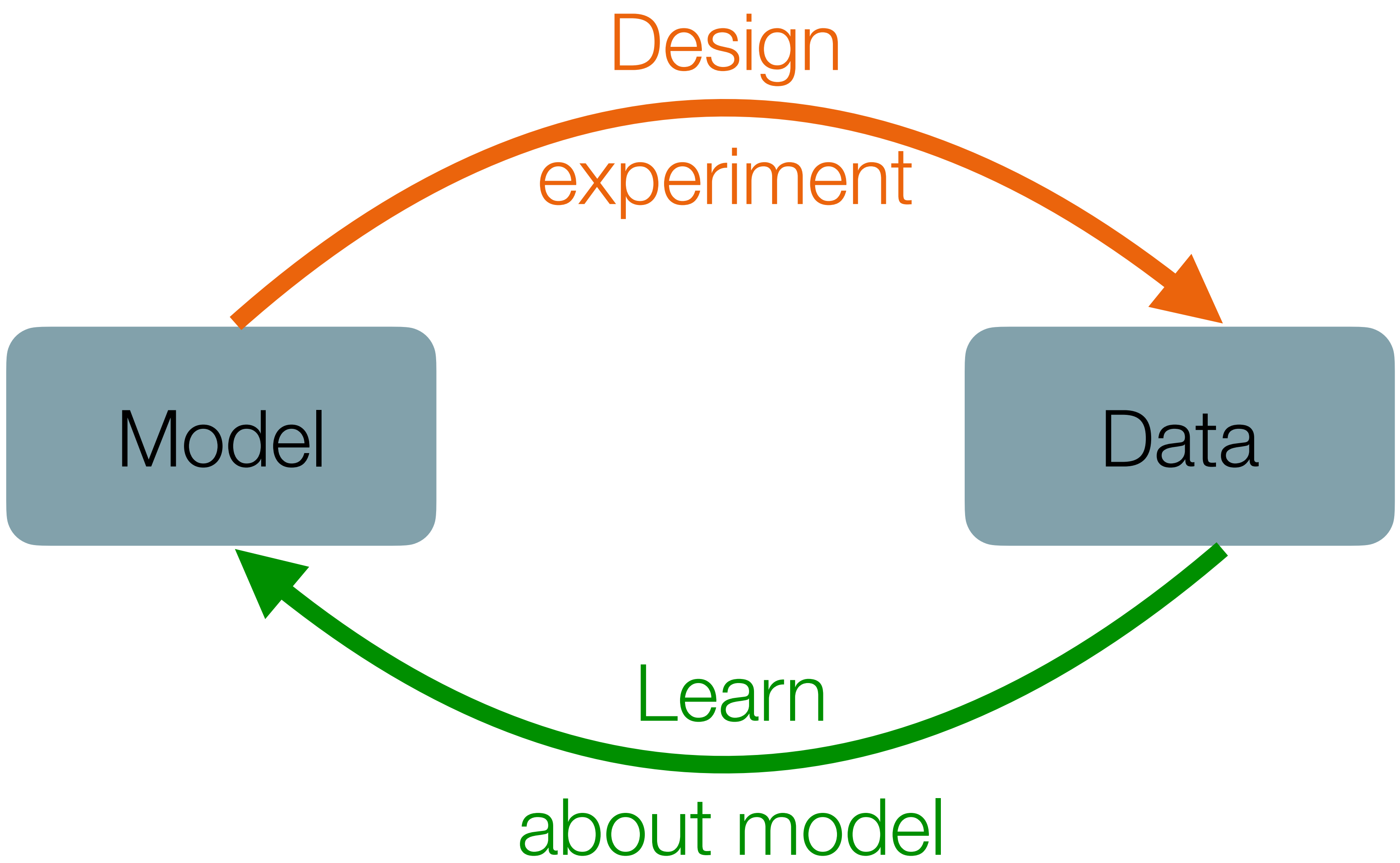
Our strategy: Close, automate, and accelerate the scientific discovery loop



Our strategy: Close, automate, and accelerate the scientific discovery loop



Our strategy: Close, automate, and accelerate the scientific discovery loop



Qualitative progress from doing 10^4 times more computational experiments and using $>10^6$ times more observational degrees of freedom than before

An example: Reduced-order models for turbulence, convection, and clouds

Yair Cohen

Jia He

Anna Jaruga

Ignacio Lopez Gomez

Cloud/boundary layer turbulence schemes in current GCMs have unphysical discontinuities and many correlated parameters

- **Deep convection**: Often mass flux schemes (e.g., Arakawa & Schubert 1974, Tiedtke 1989; Arakawa & Wu 2013)
- **Shallow convection**: Often also mass flux schemes, but with discontinuously different parameters (e.g., entrainment rates)
- **Boundary layer turbulence**: Often diffusive; difficult to match with cloud layer (e.g.,Troen & Mahrt 1986)

Parametric and structural discontinuities for processes with common (e.g., dry) limits; plethora of parameters

We use a unified, physics-based model, derived by coarse graining of equations of motion and adaptable in complexity to data availability

Decomposes domain into environment (i=0) and coherent plumes (i=1, …, N):

- Continuity:

$$\frac{\partial(\rho a_i)}{\partial t}+\frac{\partial(\rho a_i\overline{w}_i)}{\partial z}+\nabla_h\cdot(\rho a_i\langle\mathbf{u}_h\rangle)=\boxed{\underbrace{\rho a_i\overline{w}_i\left(\sum_j\epsilon_{ij}-\delta_i\right)}_{\text{Mass entrainment/detrainment}}}$$

- Scalar mean:

$$\frac{\partial(\rho a_i\overline{\phi}_i)}{\partial t}+\frac{\partial(\rho a_i\overline{w}_i\overline{\phi}_i)}{\partial z}+\nabla_h\cdot(\rho a_i\langle\mathbf{u}_h\rangle\overline{\phi}_i)=\boxed{\underbrace{-\frac{\partial(\rho a_i\overline{w_i'\phi_i'})}{\partial z}}_{\text{Turbulent transport}}}+\boxed{\underbrace{\rho a_i\overline{w}_i\left(\sum_j\epsilon_{ij}\overline{\phi}_j-\delta_i\overline{\phi}_i\right)}_{\text{Entrainment/detrainment}}}+\underbrace{\rho a_i\overline{S}_{\phi,i}}_{\text{Sources/sinks}},$$

- Scalar covariance

$$\frac{\partial(\rho a_i\overline{\phi_i'\psi_i'})}{\partial t}+\frac{\partial(\rho a_i\overline{w}_i\overline{\phi_i'\psi_i'})}{\partial z}+\nabla_h\cdot(\rho a_i\langle\mathbf{u}_h\rangle\overline{\phi_i'\psi_i'})=\underbrace{-\rho a_i\overline{w_i'\psi_i'}\frac{\partial\overline{\phi}_i}{\partial z}-\rho a_i\overline{w_i'\phi_i'}\frac{\partial\overline{\psi}_i}{\partial z}}_{\text{Generation/destruction by cross-gradient flux}}$$

$$+\underbrace{\rho a_i\overline{w}_i\left[\sum_j\epsilon_{ij}\left(\overline{\phi_j'\psi_j'}+(\overline{\phi}_j-\overline{\phi}_i)(\overline{\psi}_j-\overline{\psi}_i)\right)-\delta_i\overline{\phi_i'\psi_i'}\right]}_{\text{Covariance entrainment/detrainment}}-\underbrace{\frac{\partial(\rho a_i\overline{w_i'\phi_i'\psi_i'})}{\partial z}}_{\text{Turbulent transport}}+\underbrace{\rho a_i(\overline{S'_{\phi,i}\psi_i'}+\overline{S'_{\psi,i}\phi_i'})}_{\text{Sources/sinks}}.$$

(Tan et al., JAMES 2018)

We use a unified, physics-based model, derived by coarse graining of equations of motion and adaptable in complexity to data availability

Decomposes domain into environment (i=0) and coherent plumes (i=1, …, N):

- Continuity:

$$\frac{\partial(\rho a_i)}{\partial t}+\frac{\partial(\rho a_i\overline{w}_i)}{\partial z}+\nabla_h\cdot(\rho a_i\langle\mathbf{u}_h\rangle)=\boxed{\underbrace{\rho a_i\overline{w}_i\left(\sum_j\epsilon_{ij}-\delta_i\right)}_{\text{Mass entrainment/detrainment}}}$$

Closure functions

- Scalar mean:

$$\frac{\partial(\rho a_i\overline{\phi}_i)}{\partial t}+\frac{\partial(\rho a_i\overline{w}_i\overline{\phi}_i)}{\partial z}+\nabla_h\cdot(\rho a_i\langle\mathbf{u}_h\rangle\overline{\phi}_i)=\boxed{\underbrace{-\frac{\partial(\rho a_i\overline{w_i'\phi_i'})}{\partial z}}_{\text{Turbulent transport}}}+\boxed{\underbrace{\rho a_i\overline{w}_i\left(\sum_j\epsilon_{ij}\overline{\phi}_j-\delta_i\overline{\phi}_i\right)}_{\text{Entrainment/detrainment}}}+\underbrace{\rho a_i\overline{S}_{\phi,i}}_{\text{Sources/sinks}}$$

- Scalar covariance

$$\frac{\partial(\rho a_i\overline{\phi_i'\psi_i'})}{\partial t}+\frac{\partial(\rho a_i\overline{w}_i\overline{\phi_i'\psi_i'})}{\partial z}+\nabla_h\cdot(\rho a_i\langle\mathbf{u}_h\rangle\overline{\phi_i'\psi_i'})=\underbrace{-\rho a_i\overline{w_i'\psi_i'}\frac{\partial\overline{\phi}_i}{\partial z}-\rho a_i\overline{w_i'\phi_i'}\frac{\partial\overline{\psi}_i}{\partial z}}_{\text{Generation/destruction by cross-gradient flux}}$$

$$+\underbrace{\rho a_i\overline{w}_i\left[\sum_j\epsilon_{ij}\left(\overline{\phi_j'\psi_j'}+(\overline{\phi}_j-\overline{\phi}_i)(\overline{\psi}_j-\overline{\psi}_i)\right)-\delta_i\overline{\phi_i'\psi_i'}\right]}_{\text{Covariance entrainment/detrainment}}-\underbrace{\frac{\partial(\rho a_i\overline{w_i'\phi_i'\psi_i'})}{\partial z}}_{\text{Turbulent transport}}+\underbrace{\rho a_i(\overline{S'_{\phi,i}\psi_i'}+\overline{S'_{\psi,i}\phi_i'})}_{\text{Sources/sinks}}.$$

(Tan et al., JAMES 2018)

Parametric functions requiring closure appear in the coarse-grained equations; can be refined with data

- Entrainment and detrainment (exchange between subdomains):

  Represented by a physical entrainment length ($|b|/w^2$) and an adjustable function of nondimensional parameters

$$\varepsilon, \delta = c_\varepsilon \frac{1}{L} f\left(RH \ldots\right)$$

- Nonhydrostatic pressure gradients

  Represented by a combination of buoyancy reduction (virtual mass) and pressure drag

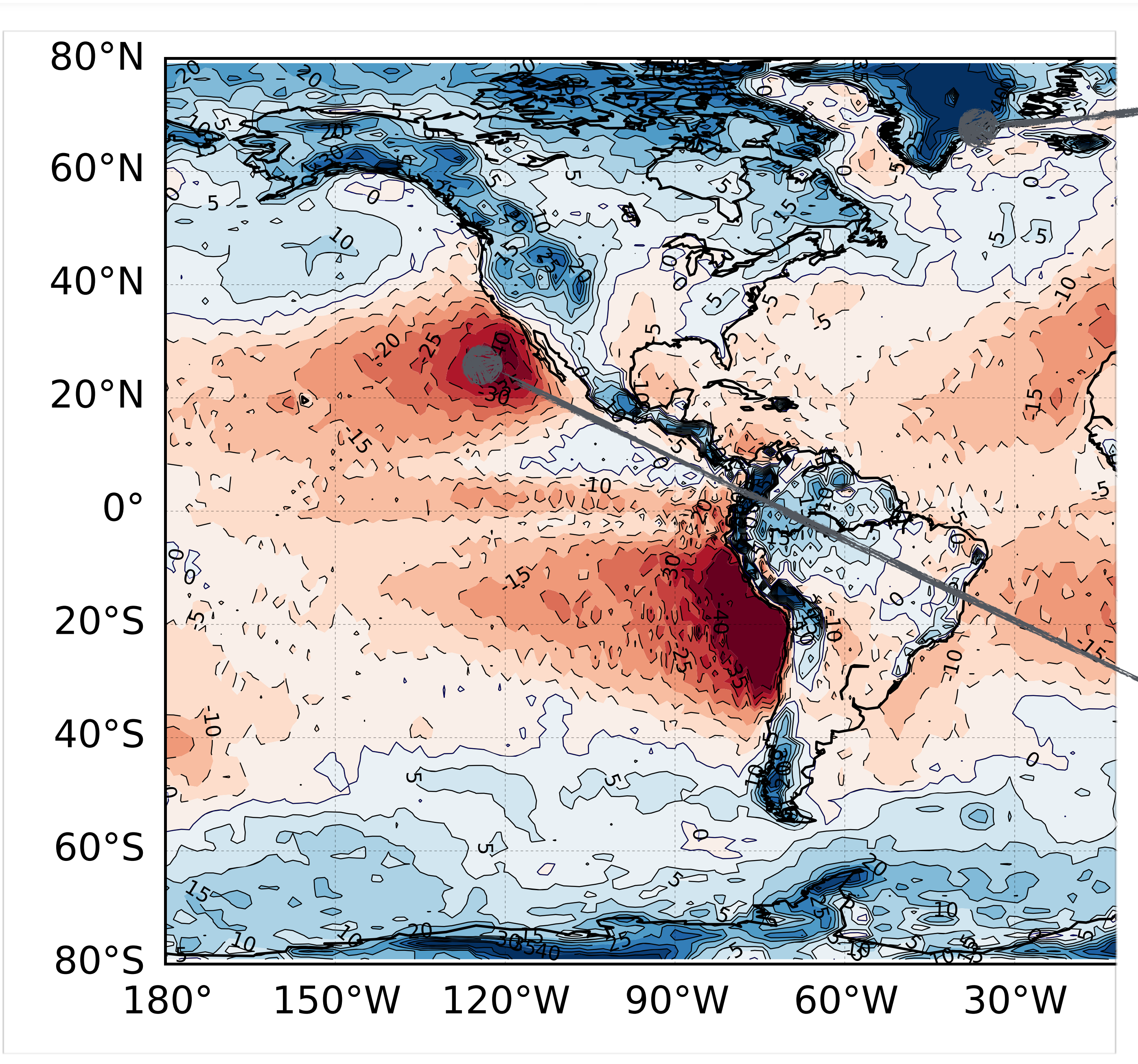
$$-\frac{\partial p_{nh}}{\partial z} = -\rho a \left(\alpha_b \bar{b} + \alpha_d \frac{\left(\bar{w}^{up} - \bar{w}^{env}\right)\left|\bar{w}^{up} - \bar{w}^{env}\right|}{Ha^{1/2}} \right)$$

- Eddy diffusion/mixing length

  Mixing length as soft minimum of all possible balances between production and dissipation of TKE

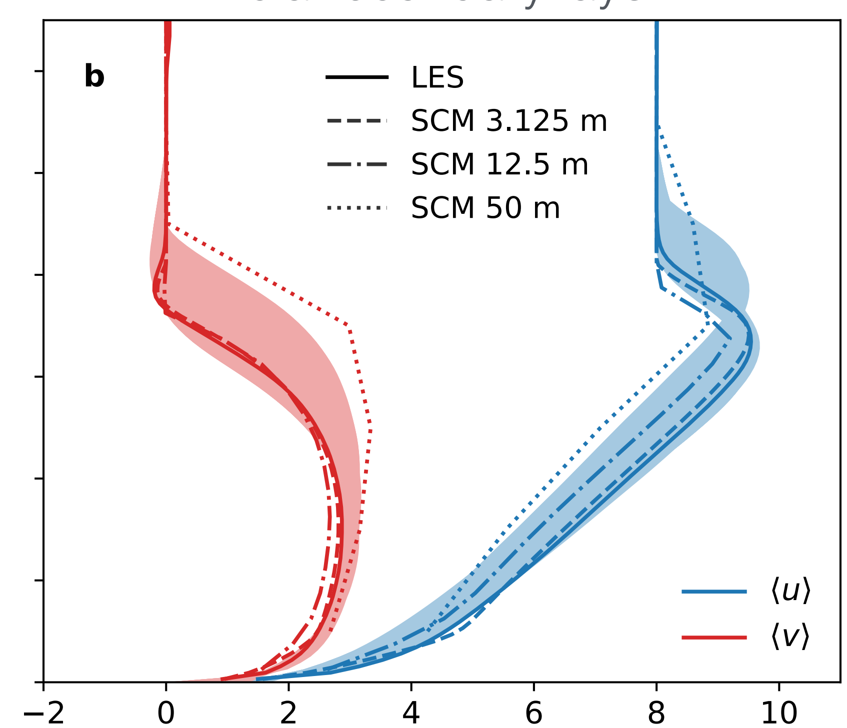
$$K = c_k l \sqrt{TKE}$$

Reduced-order model captures polar and subtropical boundary layer and clouds (which have vexed climate models for decades)

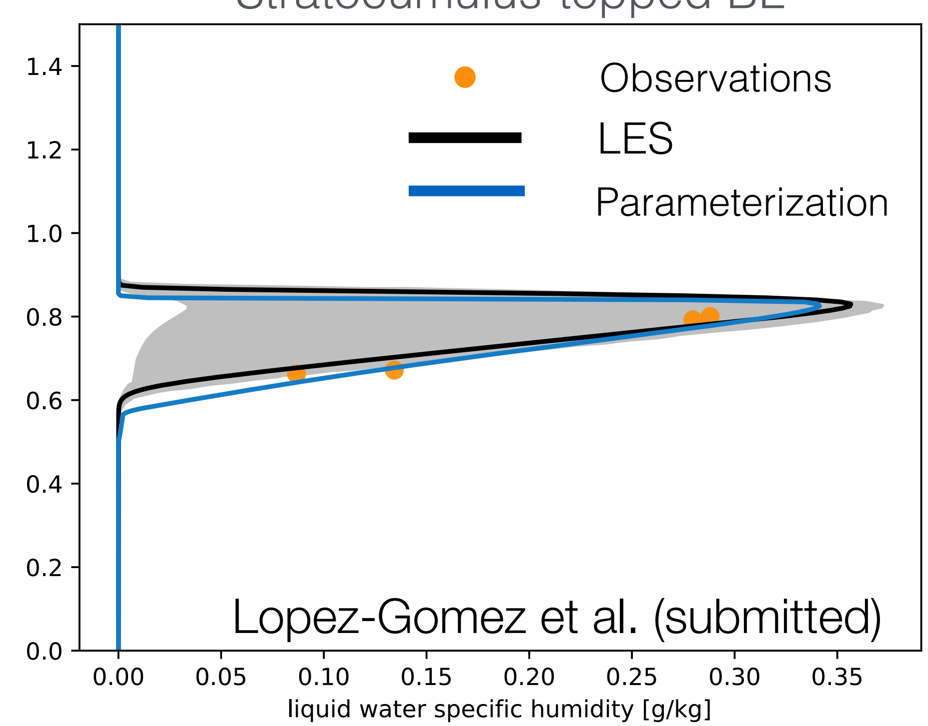


Low-cloud bias (from earlier)

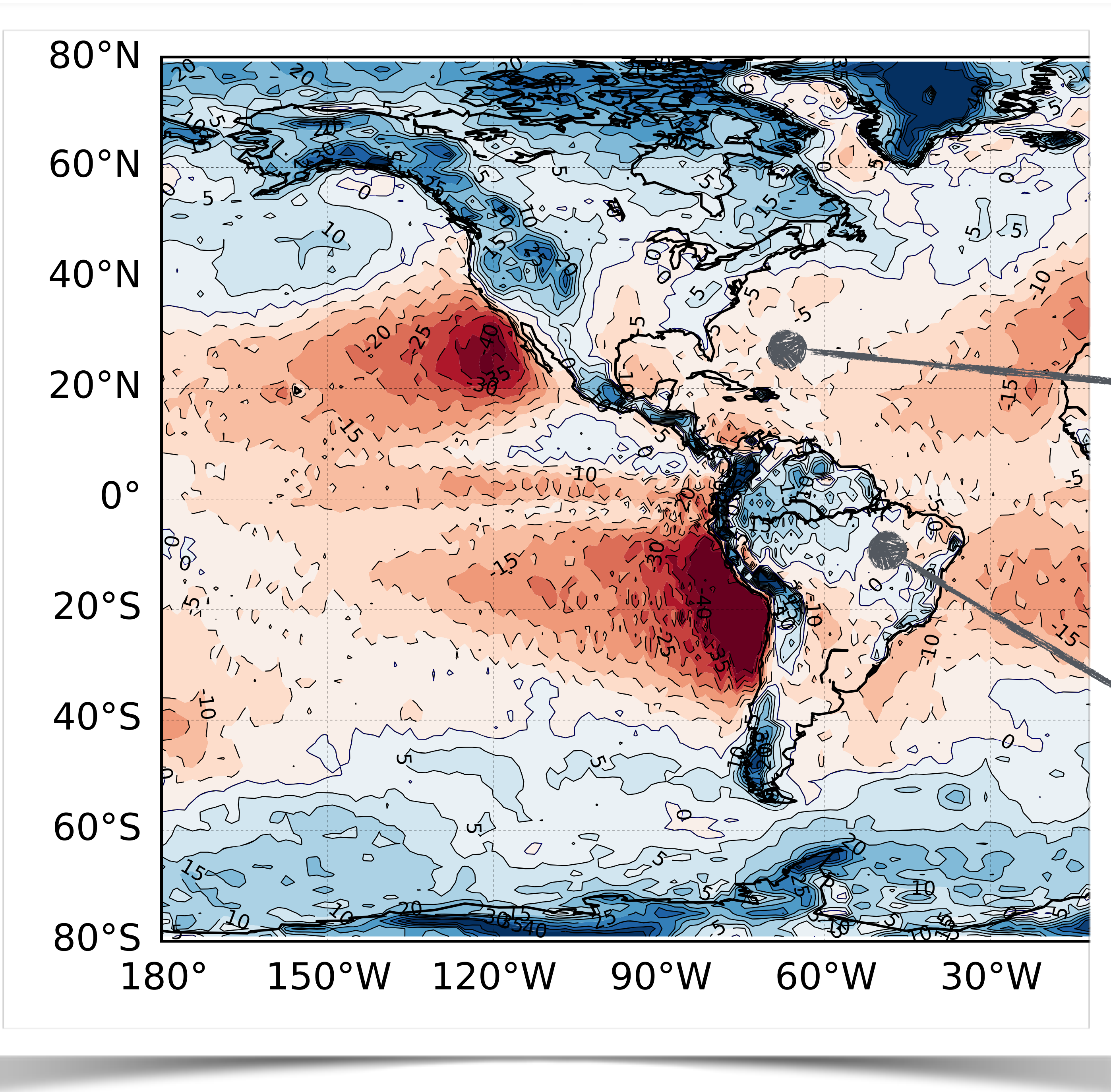
Polar boundary layer



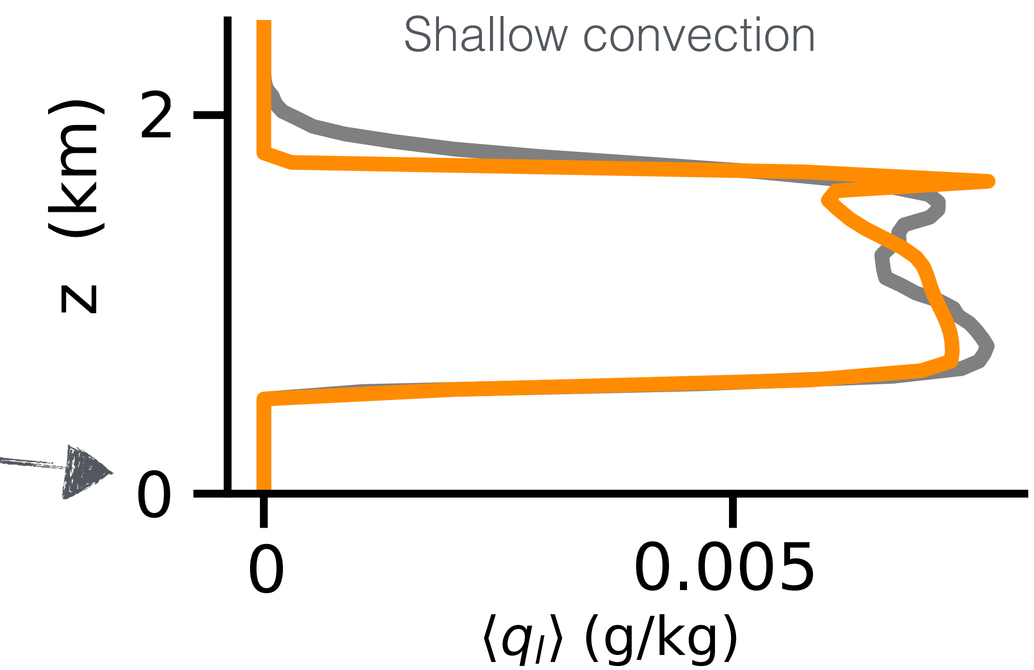
Stratocumulus-topped BL

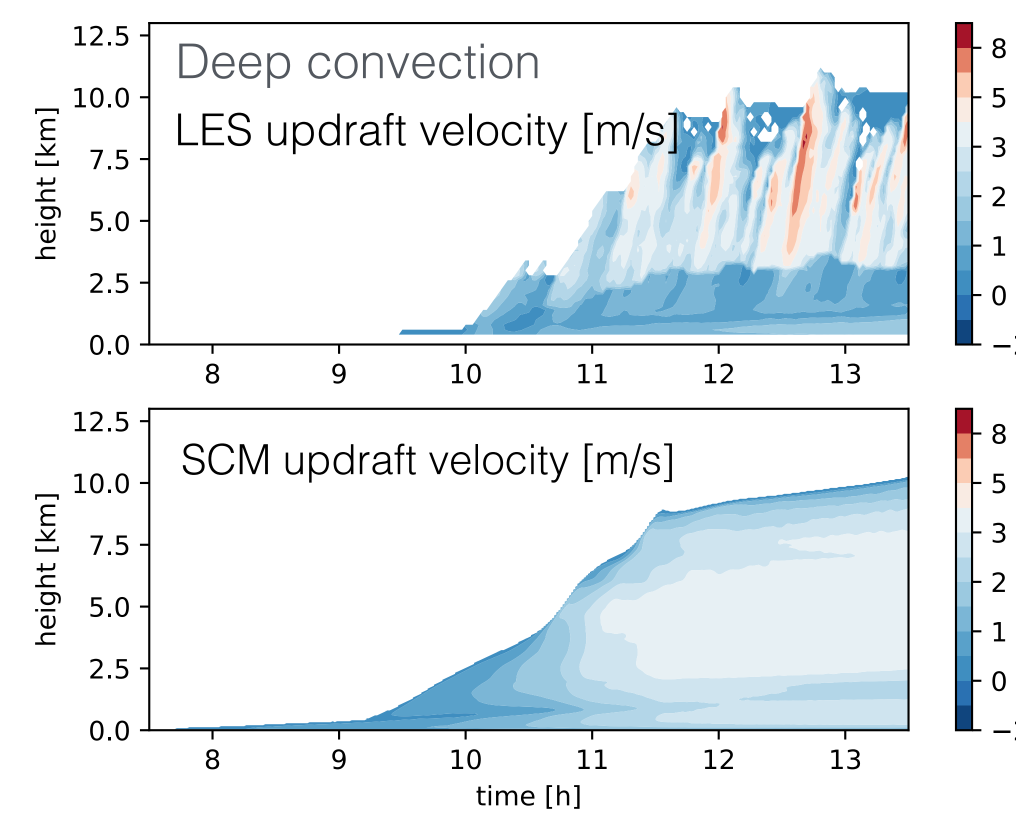


It also captures shallow and deep cumulus convection



Low-cloud bias (from earlier)





The new unified turbulence and convection scheme…

- is prognostic (essential at high host model resolution)
- captures dynamical regimes from boundary layer turbulence to deep convection
- reduces number of adjustable parameters relative to the plethora of parameters in traditional schemes

Next step is implementation in climate model, calibration and UQ with ~10,000 LES driven by climate model (first dozen running on Google Cloud Platform right now)

Calibrating a climate model and quantifying its uncertainties

Andrew Stuart

Emmet Cleary

Alfredo Garbuno

We want to improve climate models in a similar way that weather forecasts have improved, though data assimilation approaches

We are using **statistics accumulated in time** (e.g., over seasons) to calibrate model components jointly by:

1. ***Minimizing model biases,*** *especially biases that are known to correlate with the climate response of models. That is, we will minimize mismatches between time averages of ESM-simulated quantities and data, directly targeting quantities relevant for climate predictions.*

2. ***Minimizing model-data mismatches in higher-order Earth system statistics,*** e.g., covariances such as cloud-cover/surface temperature covariances, which are known to correlate with the climate response of models. Higher-order statistics relevant for predictions (e.g., precipitation extremes) are also included in objective function.

Learning from climate statistics presents challenges and opportunities

- Matching statistics results in **smoother** objective functions than matching trajectories (as is done in weather prediction)

- **Climate-relevant statistics** such as covariances between cloud cover and temperature (*emergent constraints*) and precipitation extremes **can be included in objective function**

- But objective function evaluation (accumulation of averages) is **extremely expensive**

Our setting for learning about parameters (or parametric or nonparametric functions)

Find Parameter θ From Data y

Let $\mathsf{G} : \Theta \mapsto \mathcal{Y}$, and η be noise. Then data and parameter are related by

$$y = \mathsf{G}(\theta) + \eta, \quad \eta \sim \mathsf{N}(0, \gamma^2 I).$$

Our Setting

- Calibration and UQ for θ are both important.
- G is expensive to evaluate.
- G is only approximately available.
- Derivatives of G are not available.

Optimization approach

Formulation

$$\theta^\star = \operatorname{argmin}_{\theta\in\Theta} \Phi(\theta; y),$$
$$\Phi_0(\theta; y) = \frac{1}{2\gamma^2}|y - G(\theta)|^2,$$
$$\Phi(\theta; y) = \frac{1}{2\gamma^2}|y - G(\theta)|^2 + \frac{1}{2}\langle\theta, \Sigma^{-1}\theta\rangle.$$

Algorithms: parameter θ calibration
(e.g., derivative-free ensemble methods, $O(10^2)$ evaluations of *G;*
scale well to high-dimensional data and parameter spaces)

Bayesian approach

Formulation

$$\mathbb{P}(\theta|y) \propto \mathbb{P}(y|\theta) \times \mathbb{P}(\theta),$$

$$\begin{aligned}\mathbb{P}(\theta|y) &\propto \exp\Big(-\Phi_0(\theta; y)\Big) \times \exp\Big(-\frac{1}{2}\langle\theta, \Sigma^{-1}\theta\rangle\Big)\\ &\propto \exp\Big(-\Phi(\theta; y)\Big)\end{aligned}$$

Algorithms: parameter θ sampling
(e.g., MCMC, $O(10^5)$ evaluations of *G;* not feasible for climate models)

We combine calibration and Bayesian approaches in a three step process for fast Bayesian learning

Calibrate: $y = \mathcal{G}(\theta) + \eta$ → **Emulate**: $\mathcal{G}^{(M)}(\theta) \approx \mathcal{G}(\theta)$ → **Sample**: $y = \mathcal{G}^{(M)}(\theta) + \eta(\theta)$

Gradient-free ensemble methods (EKS/EKI)

Gaussian processes/NNs

MCMC

- Experimental design (where to place high-resolution simulations) can be incorporated into CES pipeline

- Gives approximate Bayesian posterior (i.e., quantified uncertainties, including covariance structure of error etc.)

Cleary, Garbuno-Inigo, Lan, Schneider, Stuart, https://arxiv.org/abs/2001.03689, 2020

Proof-of-concept in idealized general circulation model (GCM)

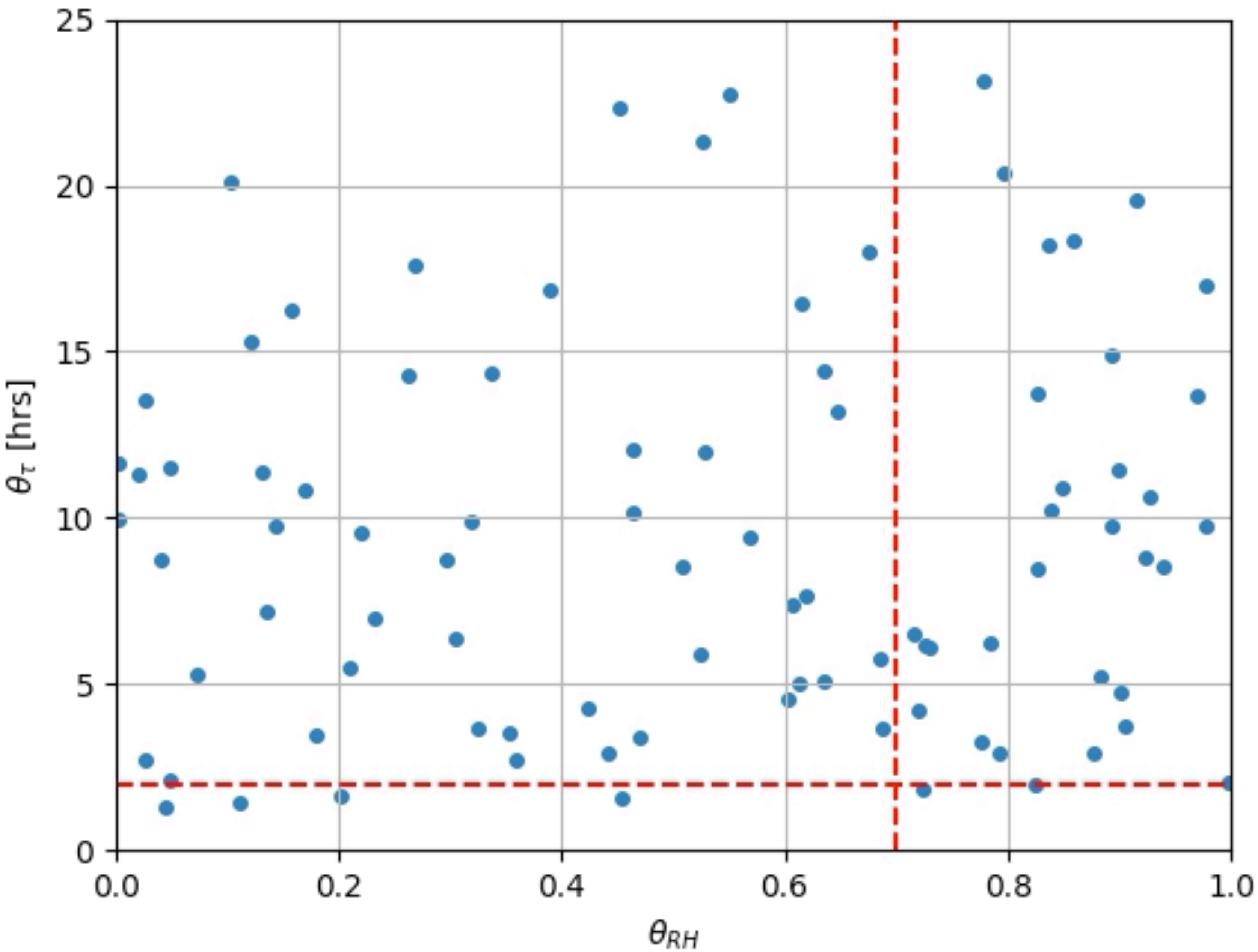
- GCM is an idealized aquaplanet model
- It has a simple convection scheme that relaxes temperature and specific humidities to reference profiles

$$\partial_t T + v \cdot \nabla T + \cdots = -\frac{T - T_{\mathrm{ref}}}{\tau}$$

$$\partial_t q + v \cdot \nabla q + \cdots = -\frac{q - \mathrm{RH}_{\mathrm{ref}} q^*(T_{\mathrm{ref}})}{\tau}$$

- Two closure parameters: **timescale** τ and **reference relative humidity** $\mathbf{RH_{ref}}$

(1) **Calibrate** with ensemble Kalman inversion

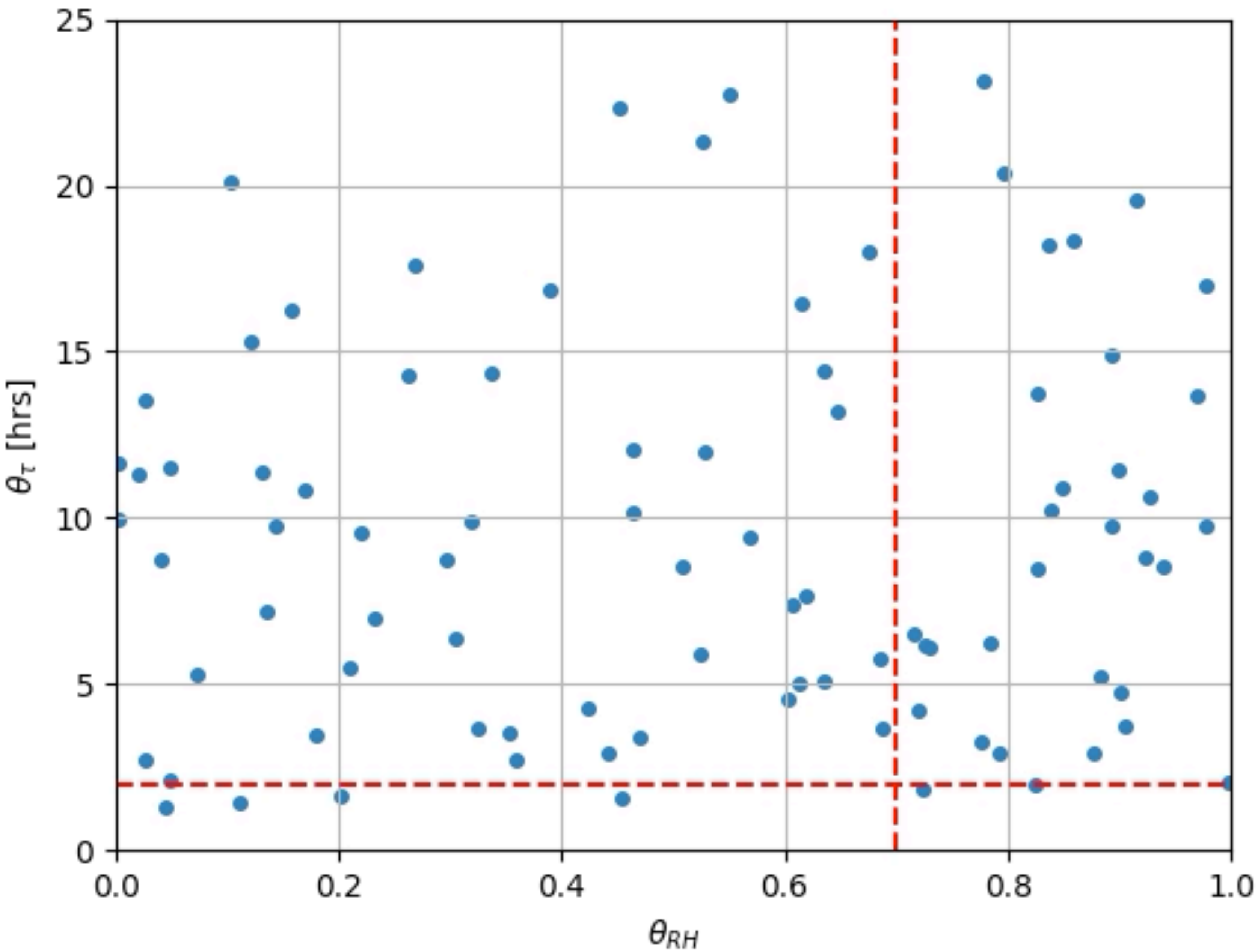


Objective function has ***relative humidity***, ***mean precipitation***, *and* ***precipitation extremes***

Ensemble Kalman inversion for parameters in convection scheme: ensemble of size 100 converges in ~5 iterations

Courtesy Emmet Cleary

(1) **Calibrate** with ensemble Kalman inversion

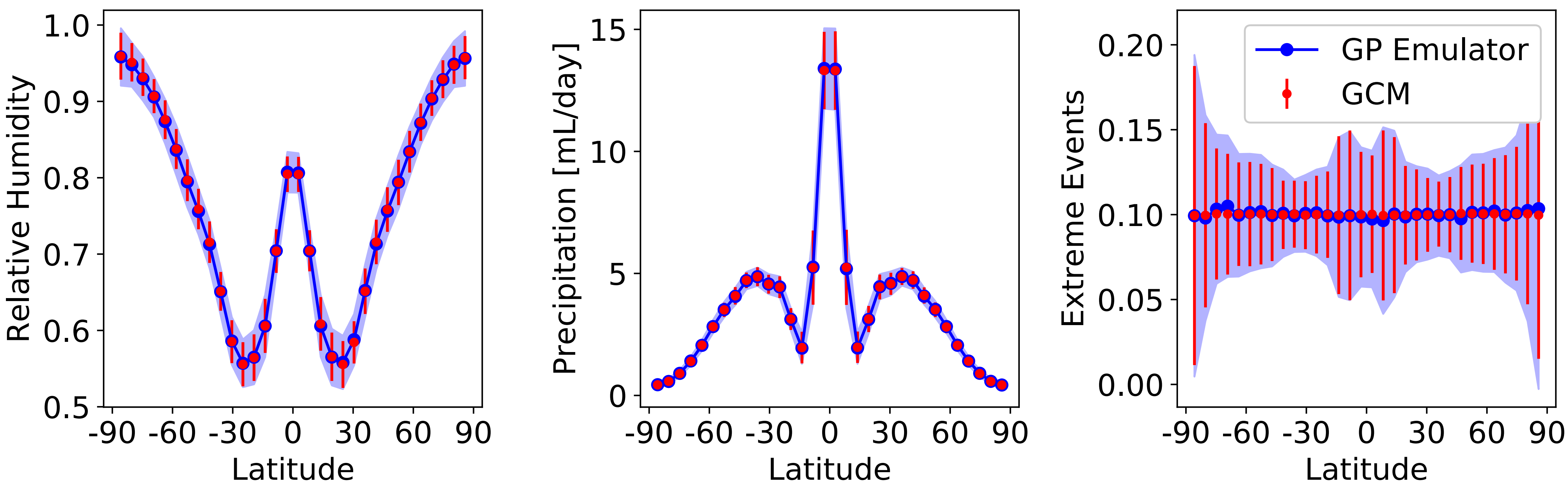


Objective function has ***relative humidity***, ***mean precipitation***, *and* ***precipitation extremes***

Ensemble Kalman inversion for parameters in convection scheme: ensemble of size 100 converges in ~5 iterations

Courtesy Emmet Cleary

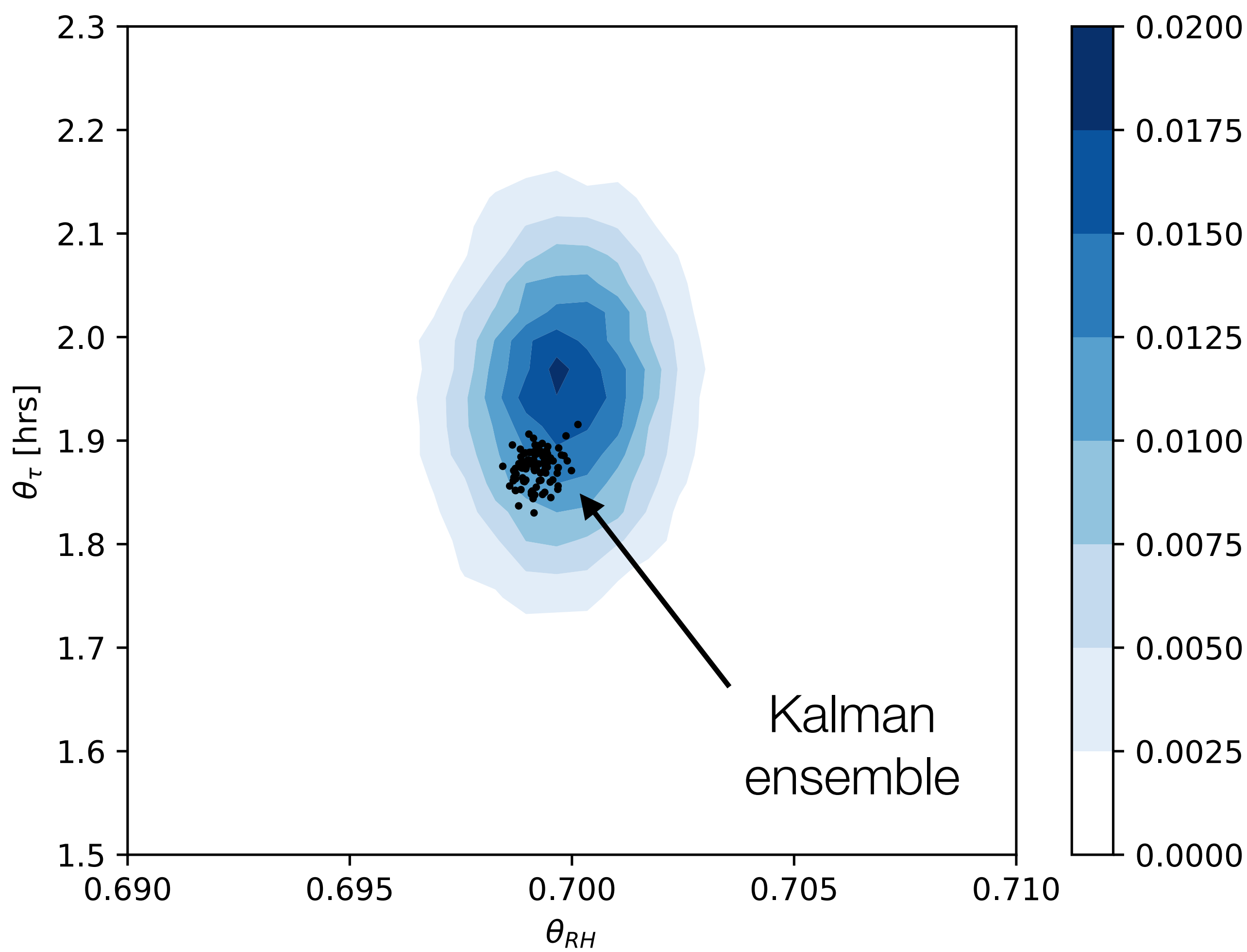
(2) **Emulate** parameters-to-statistics map during calibration step with Gaussian processes



Effective emulation of model statistics at vanishing marginal cost; additional important advantage: smoothing of objective function (can be replace by NNs for better scaling)

(3) **Sample** emulator to obtain posterior PDF for uncertainty quantification

MCMC (500,000 iterations) on GP trained on ensemble gives good estimate of posterior PDF

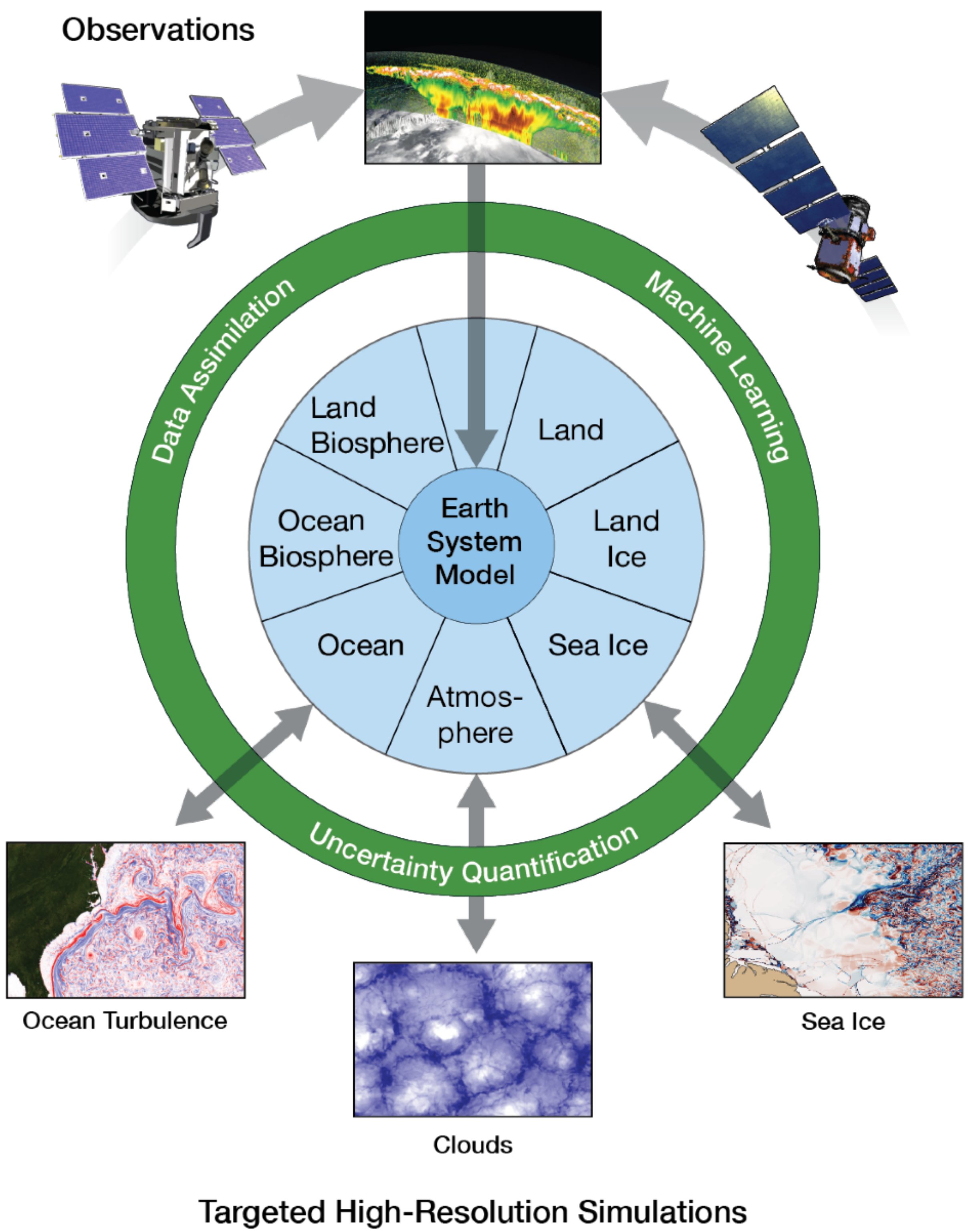


	σ_{RH}	σ_τ (hrs)
MCMC (EKI-GP)	$3.32\text{x}10^{-3}$	0.168
MCMC (Gold Standard)	$2.82\text{x}10^{-3}$	0.169

Approximate Bayesian inversion at 1/1000th the cost of standard methods
First calibrate-emulate-sample paper: https://arxiv.org/abs/2001.03689

Courtesy Emmet Cleary

We are pursuing the same approach for all components of the new Earth system model



5-year goals

- Build a model that learns automatically from observations and high-resolution simulations
- Achieve at least factor 2 reduction in rms error of climate simulations and impacts (e.g., in rainfall extremes)
- Serve as anchor of ecosystem of downstream apps, e.g., for infrastructure planning or projections of wildfire and flood risks.

Core design principles for CliMA's model

- Require performance-portability and scalability across different hardware architectures with accelerators (facilitated by Julia programming paradigms and collaboration with MIT Julia Lab)

- Atmosphere, ocean, land, and (eventually) sea ice share computational kernels, maximizing code re-use and facilitating coupling and optimization

- Use consistent thermodynamics, microphysics etc. across the entire model

- Develop unified parameterizations through hierarchical approximations that can be refined as more data become available

- Couple parameterized processes consistently with their underlying distributional assumption (e.g., subsample microphysics from subgrid-scale distributions of dynamical quantities)

Conclusions

- **Reducing and quantifying uncertainties** in climate models is urgent but within reach
- To reduce and quantify uncertainties, we **combine *process-informed* models with *data-driven* approaches using *climate statistics***
- **Physics-based subgrid-scale models** can capture turbulence and cloud regimes that have vexed climate models for decades
- Our subgrid-scale models will **learn both from observations and (where possible) from high-resolution simulations spun off on the fly**
- **Calibrate-emulate-sample** forms the core of the data assimilation/machine learning layer and achieves up to 1,000x speed-up relative to traditional Bayesian learning methods

Much interesting work (SGS models, more effective filtering strategies, optimal targeting of high-res simulations…) remains to be done!

With thanks to CliMA's funders

ERIC AND WENDY SCHMIDT

SCHMIDT **FUTURES**

MOUNTAIN PHILANTHROPIES

CHARLES TRIMBLE

RONALD AND MAXINE LINDE CLIMATE CHALLENGE

PAUL G. ALLEN FAMILY FOUNDATION

NSF