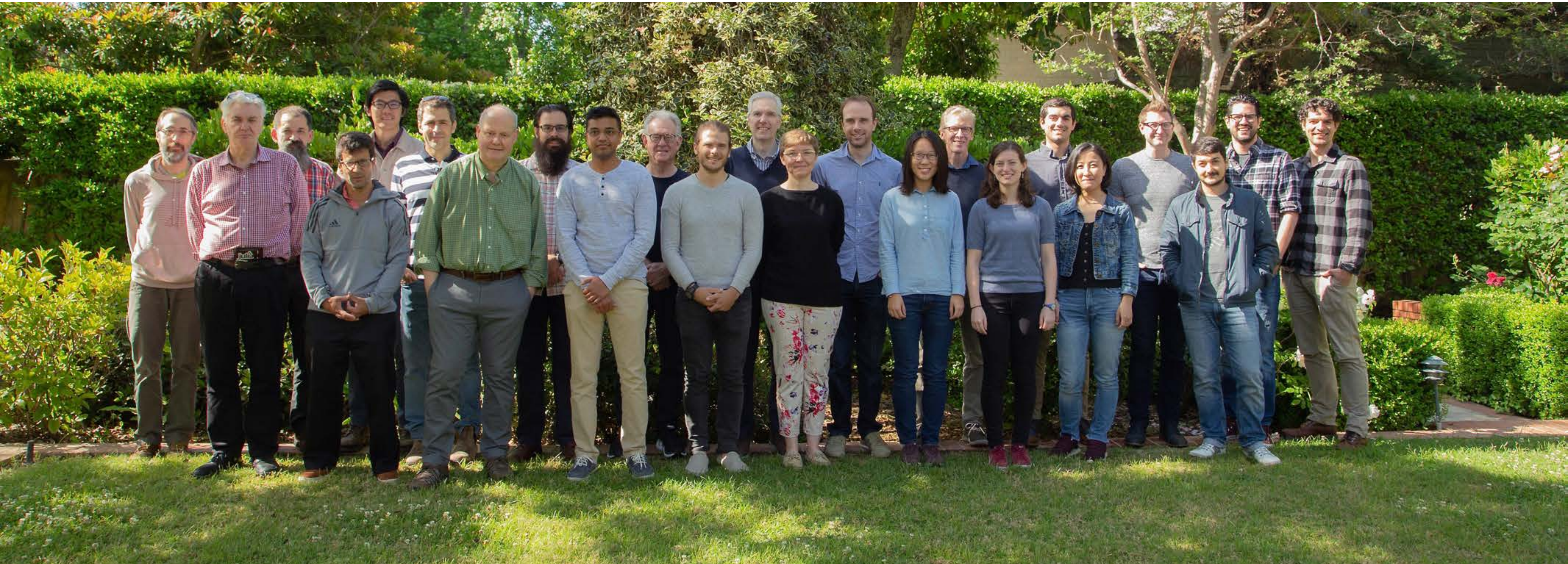
CLIMATE MODELING ALLIANCE

Clouds, Climate, And Data-Informed Earth System Modeling

Tapio Schneider

The Climate Modeling Alliance (CliMA)…

*…is a coalition of scientists, engineers, and applied mathematicians from **Caltech**, **MIT**, the **Naval Postgraduate School**, and the **Jet Propulsion Laboratory.** We are building the first Earth system model that automatically learns from diverse data sources to produce accurate climate predictions with quantified uncertainties.*



CliMA is funded by a consortium of private and public foundations

ERIC AND WENDY SCHMIDT

SCHMIDT FUTURES

EARTHRISE ALLIANCE

CHARLES TRIMBLE

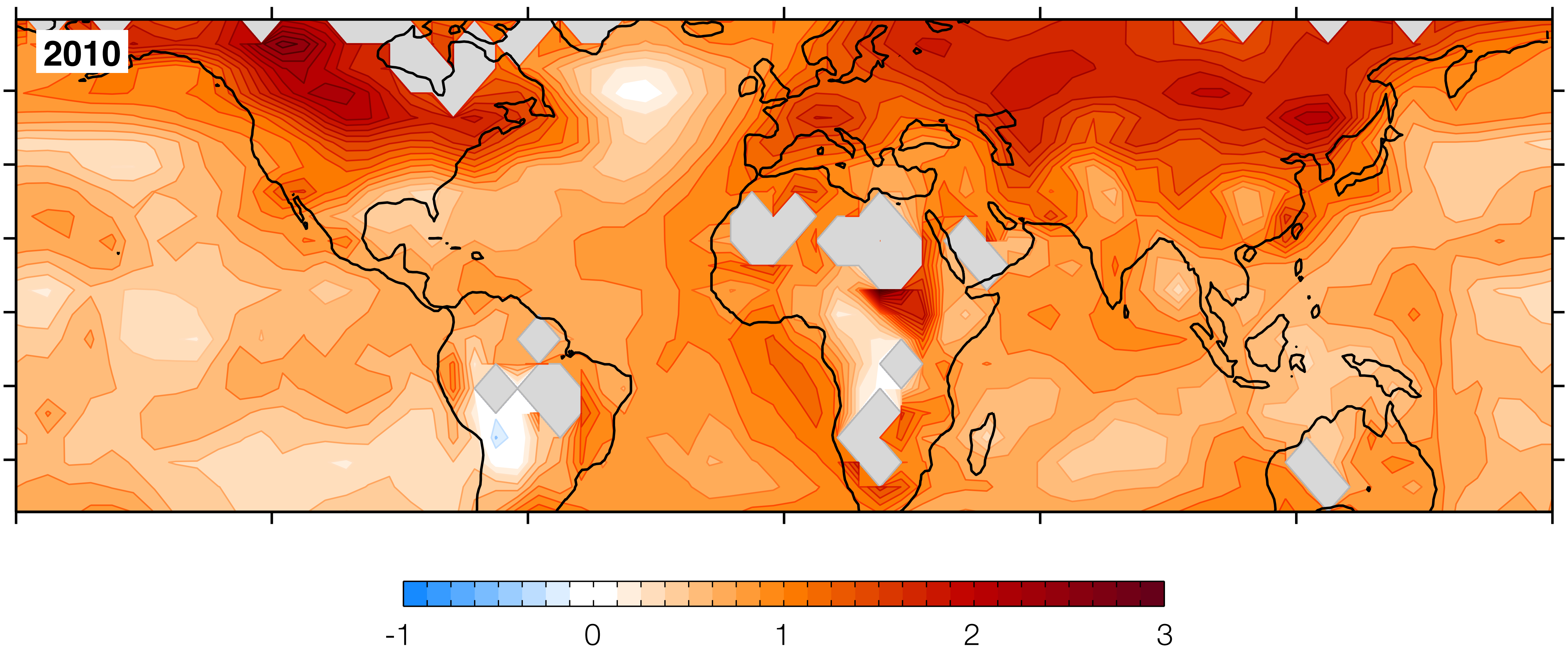
RONALD AND MAXINE LINDE CLIMATE CHALLENGE

PAUL G. ALLEN FAMILY FOUNDATION

NSF

Limitations of Current Models

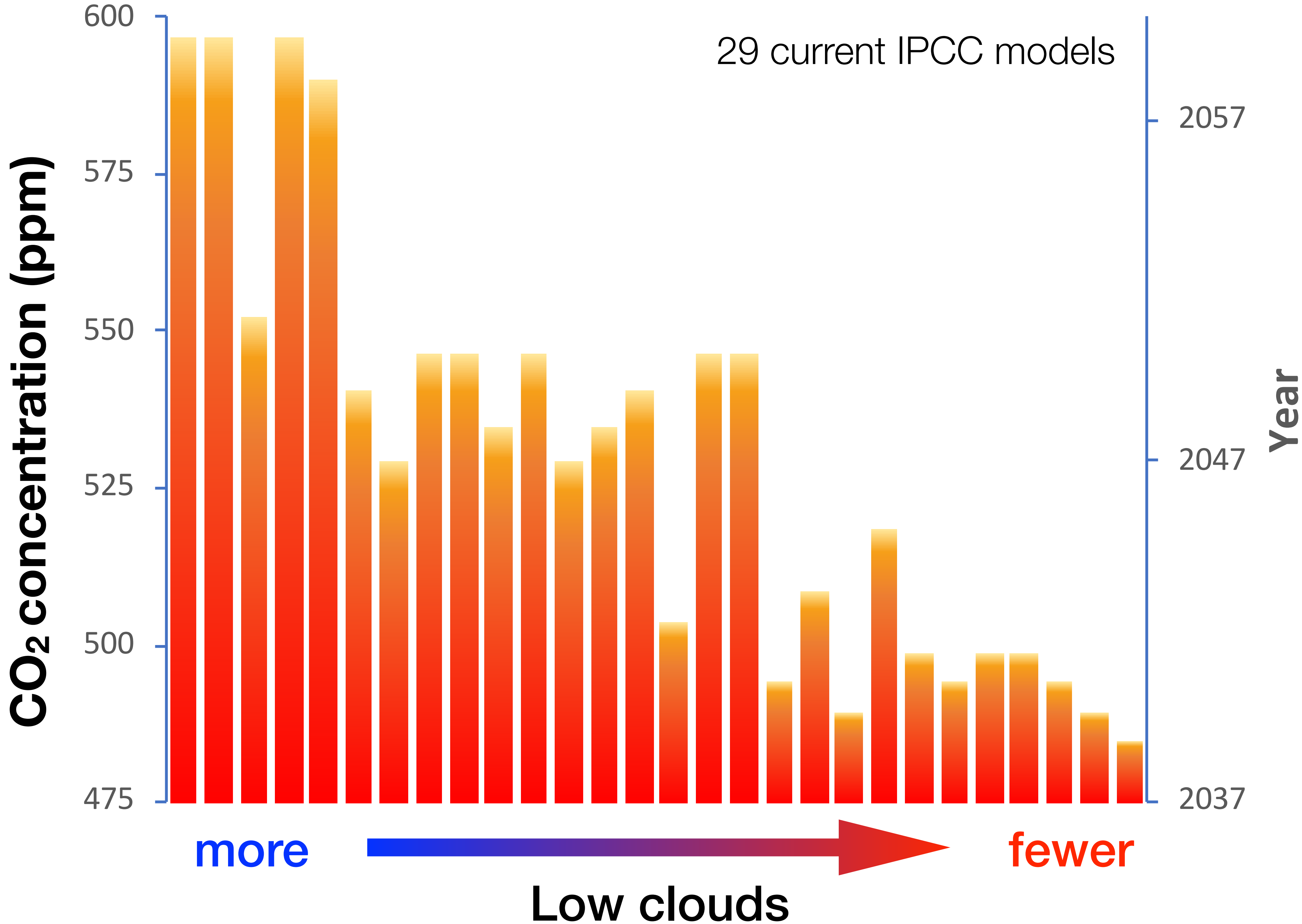
Temperatures have risen over the past 150 years



Temperature change (ºC) from 1850s through 2010s

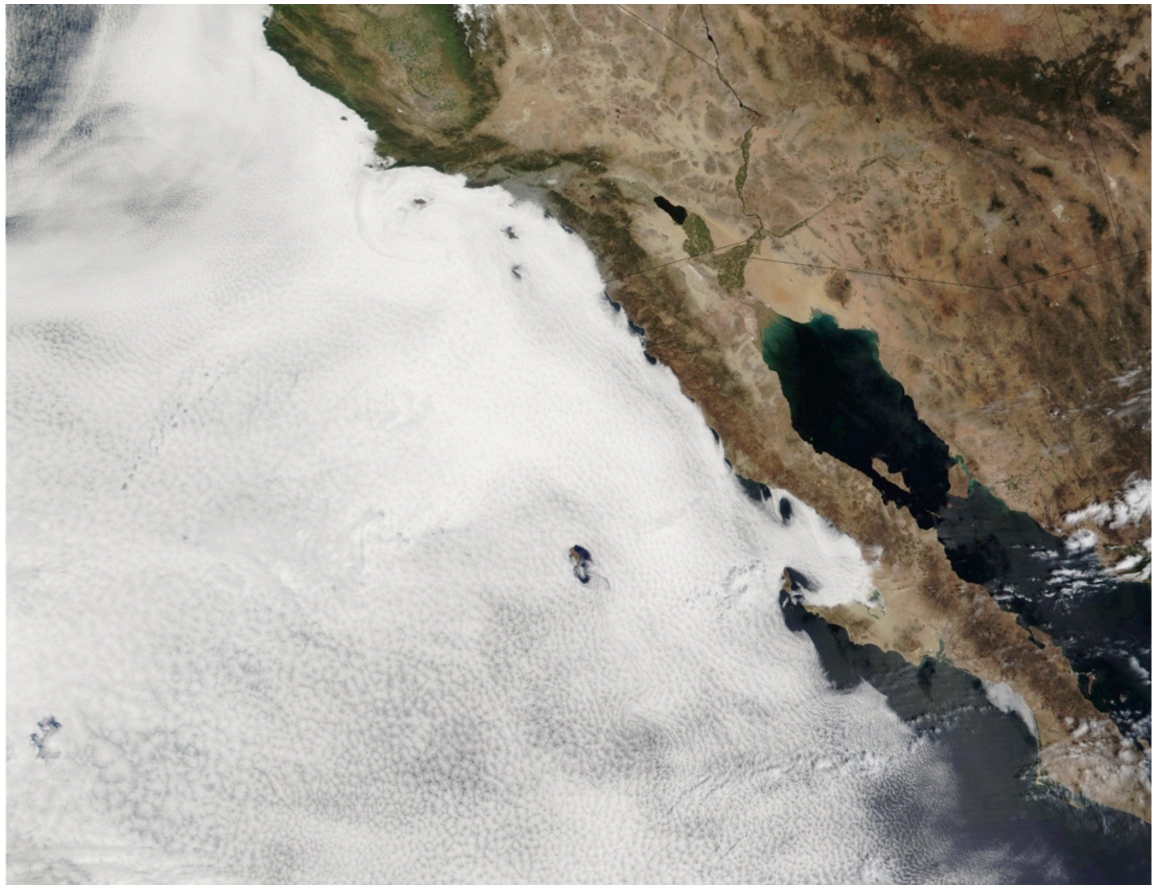
Schneider & Held, J. Climate, 2001; update http://climate-dynamics.org/videos

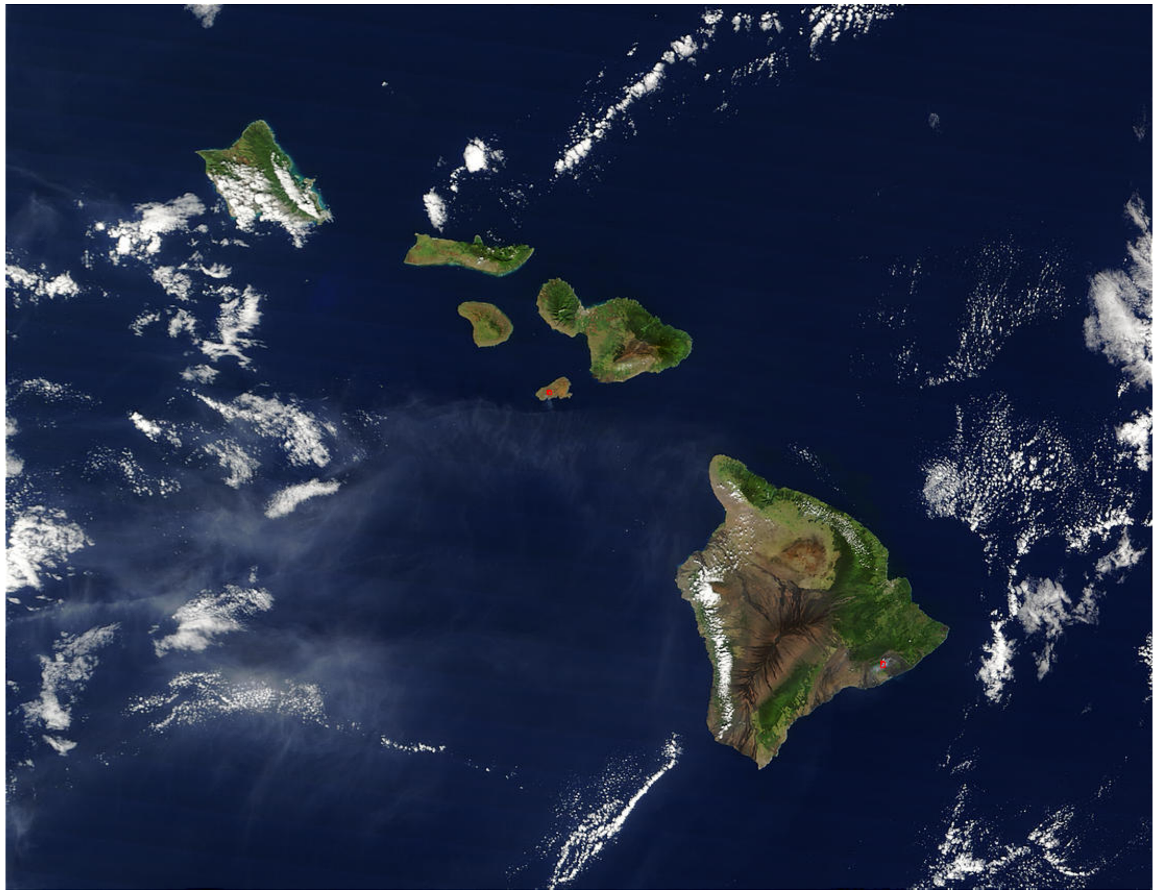
Climate predictions are uncertain: E.g., the CO_2 concentration at which 2°C warming threshold is crossed varies widely across models



Schneider et al., *Nature Climate Change* 2017

Low clouds dominate uncertainty in projections





http://eoimages.gsfc.nasa.gov

Stratocumulus: colder

Cumulus: warmer

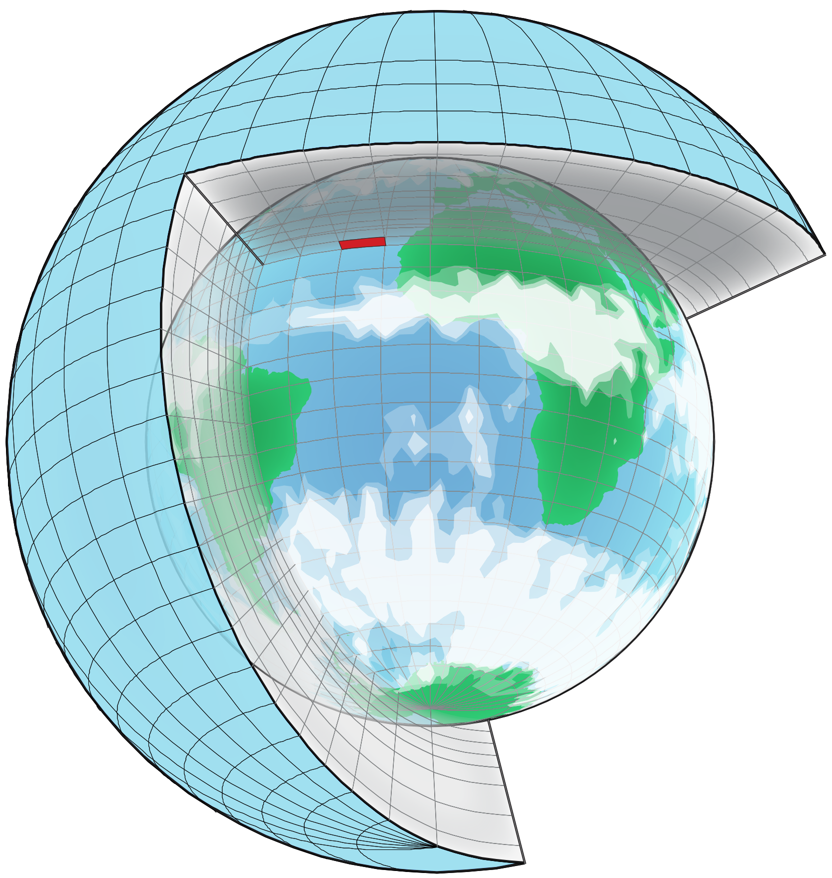
We don't know if we will get more low clouds (damped global warming), or fewer low clouds (amplified warming)

More accurate climate projections with quantified uncertainties would enable…

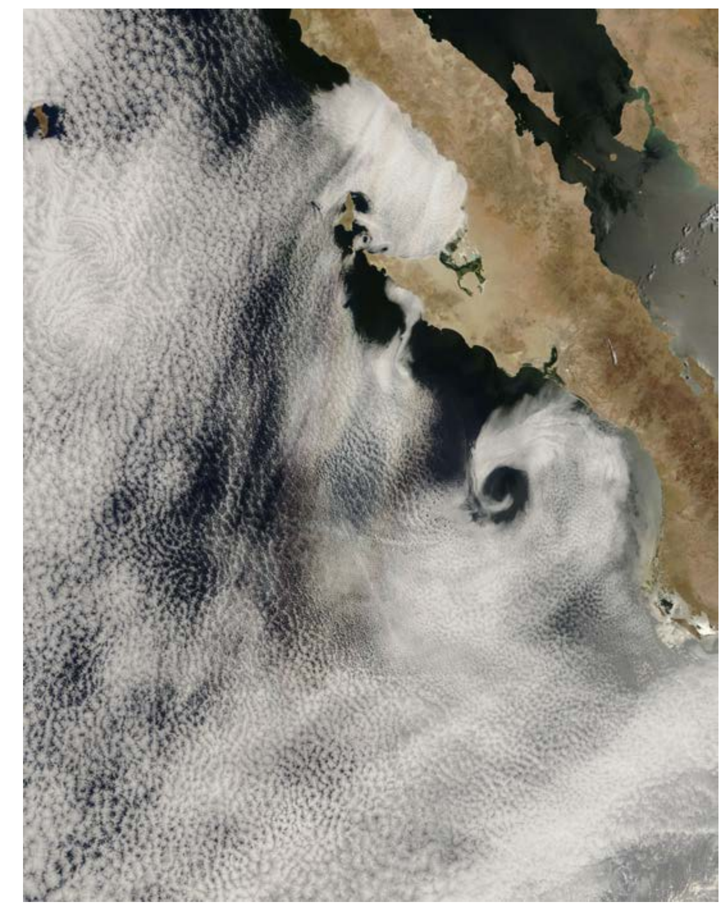
- Data-driven decisions about infrastructure planning, e.g.,
  - *How high a sea wall should New York City build to protect itself against storm surges in 2050?*
  - *What water management infrastructure is needed to ensure food and water security in sub-Saharan Africa?*
- Rational resource allocation for climate change adaptation: costs estimated to reach >$200B annually by 2050 (UNEP 2016)

Cumulative socioeconomic value of more accurate predictions estimated to lie in the trillions of USD (Hope 2015)

Small-scale processes (e.g., clouds) are the primary sources of uncertainty in climate projections



Global model:
~10-50 km resolution

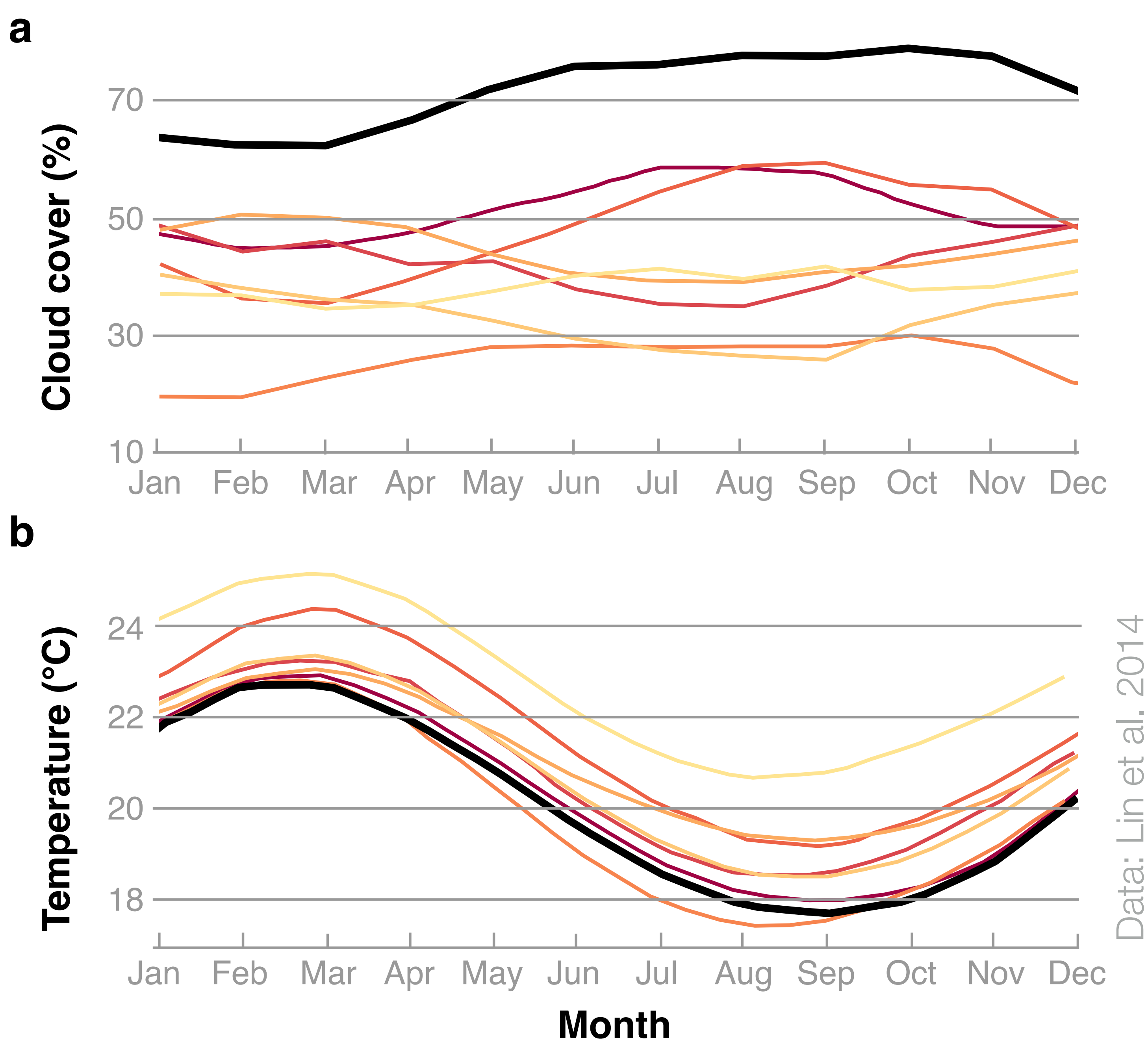


NASA MODIS

Cloud scales: ~10-100 m

Subgrid-scale processes (e.g., clouds and turbulence) are represented semi-empirically

The models' inability to predict low clouds is also manifest in failure to simulate present climate: E.g., no model simulates stratocumulus well



Cloud cover (low bias)

SST (warm bias)

"Too few, too bright bias" leads to large rainfall biases

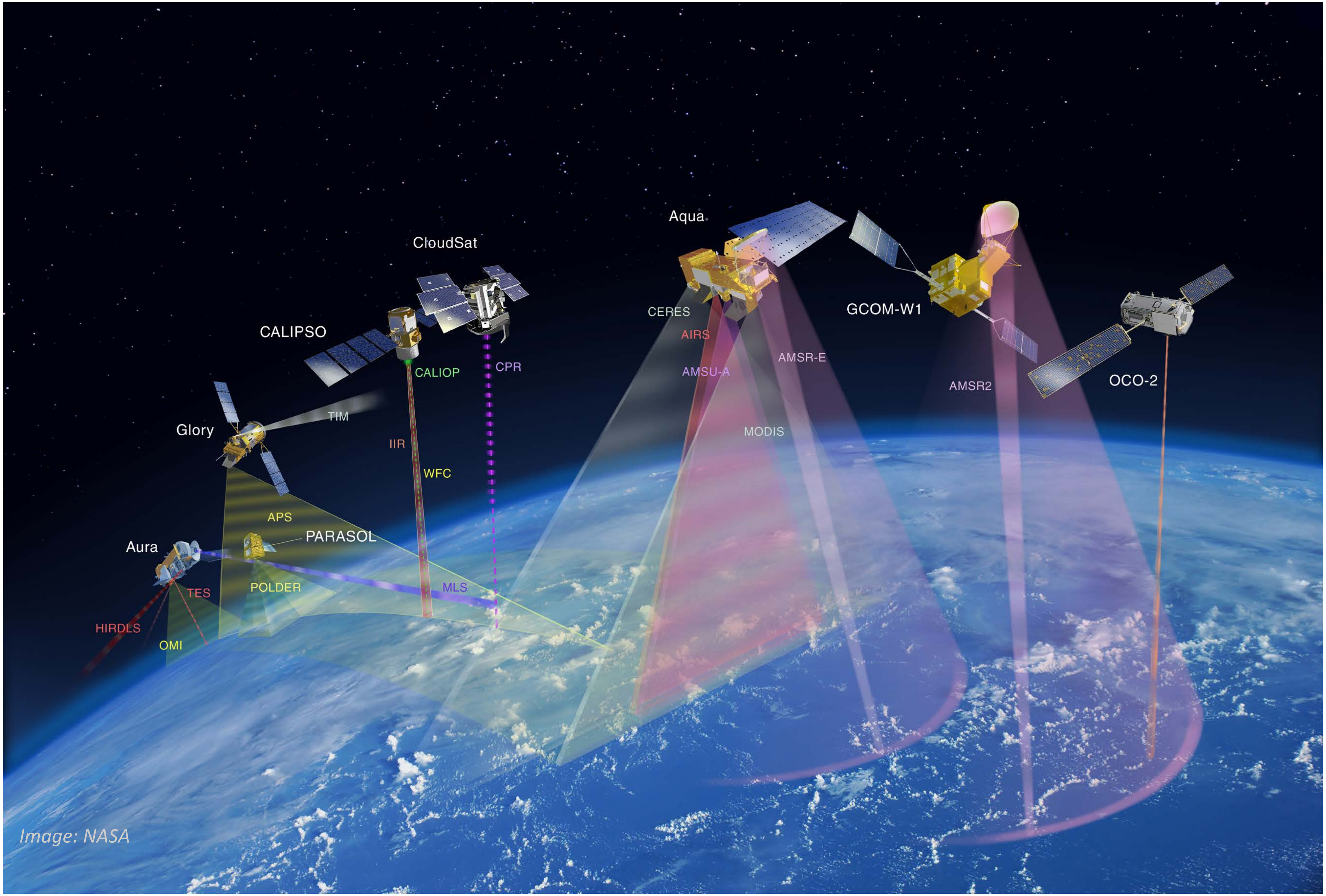
A new approach: climate models that learn from diverse data sources

We are building an Earth system model (ESM) that learns automatically from two data sources

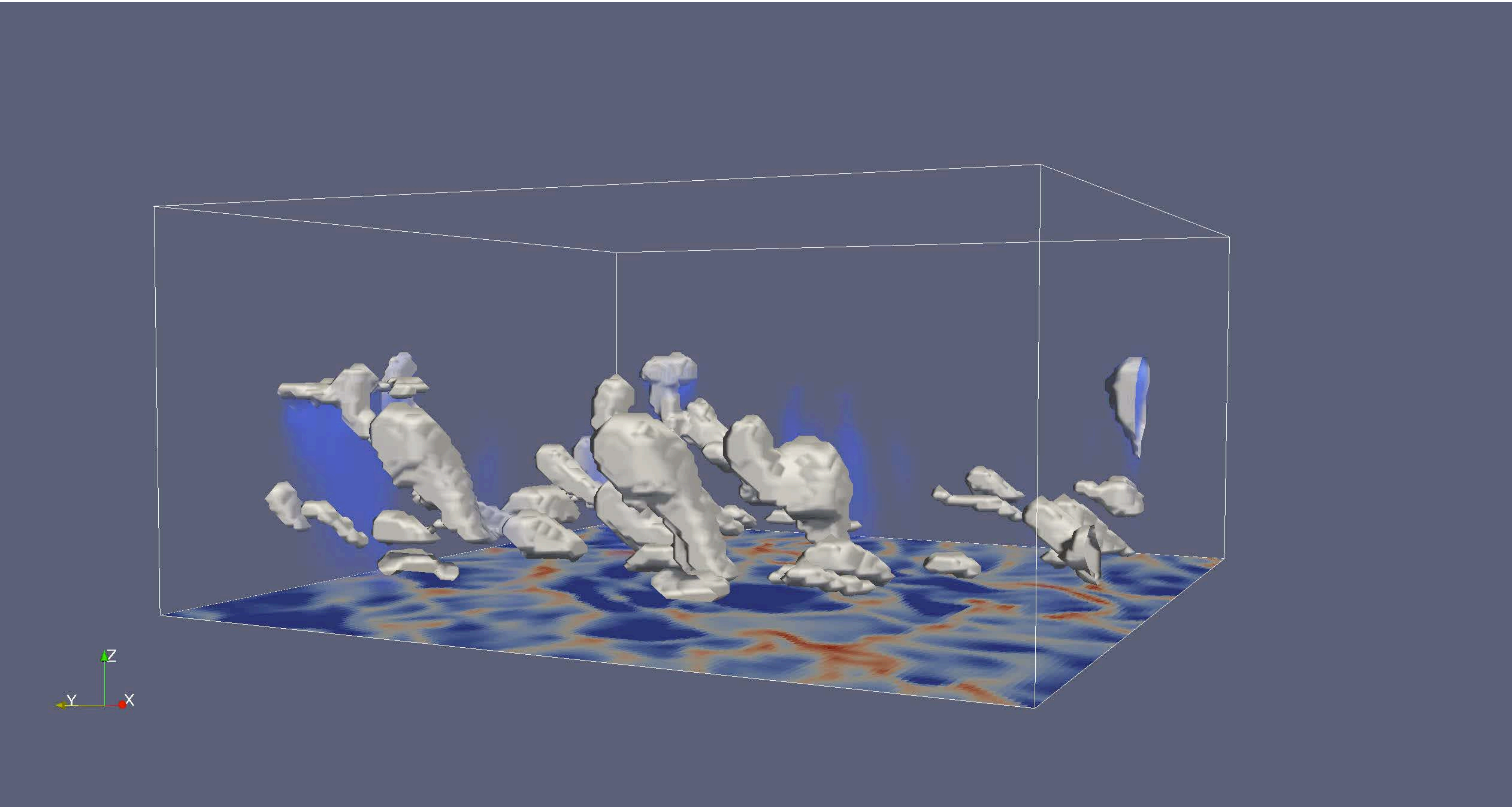
1. ***Global observations***: Our ESM will learn from space-based measurements of temperature, humidity, clouds, ocean surface currents, and sea ice cover

2. ***Local high-resolution simulations***: Our ESM will learn from targeted high-resolution simulations of computable processes such as ocean turbulence, clouds, and convection

A wealth of Earth observations is available, whose potential to improve models has not been tapped



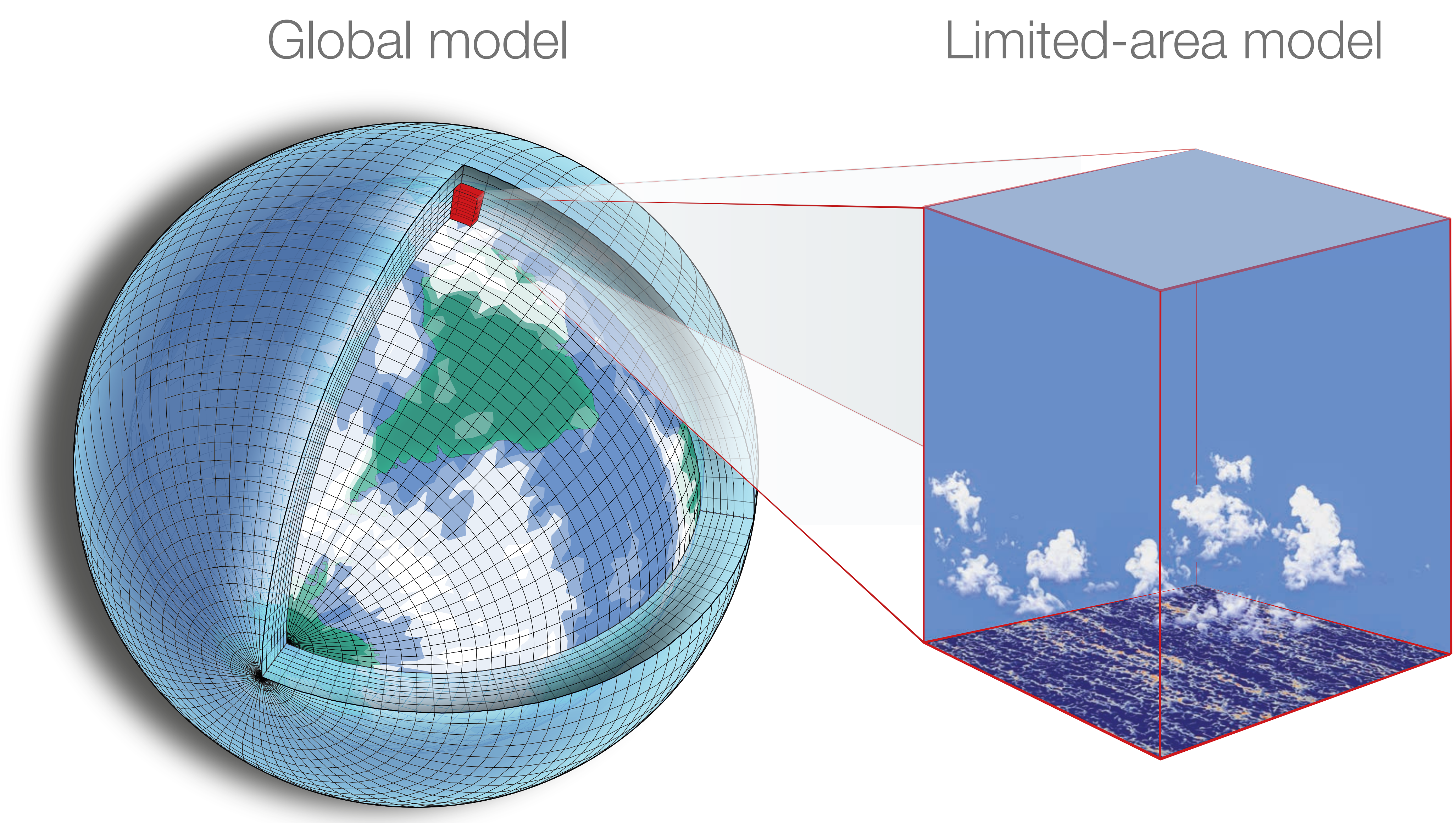
We can also simulate some processes (e.g., clouds) faithfully, albeit only in limited areas



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 or tens of thousands of high-resolution simulations can be embedded in a global model, and the global model can learn from them

Our ESM will learn from observations and targeted high-resolution simulations by optimizing over climate statistics

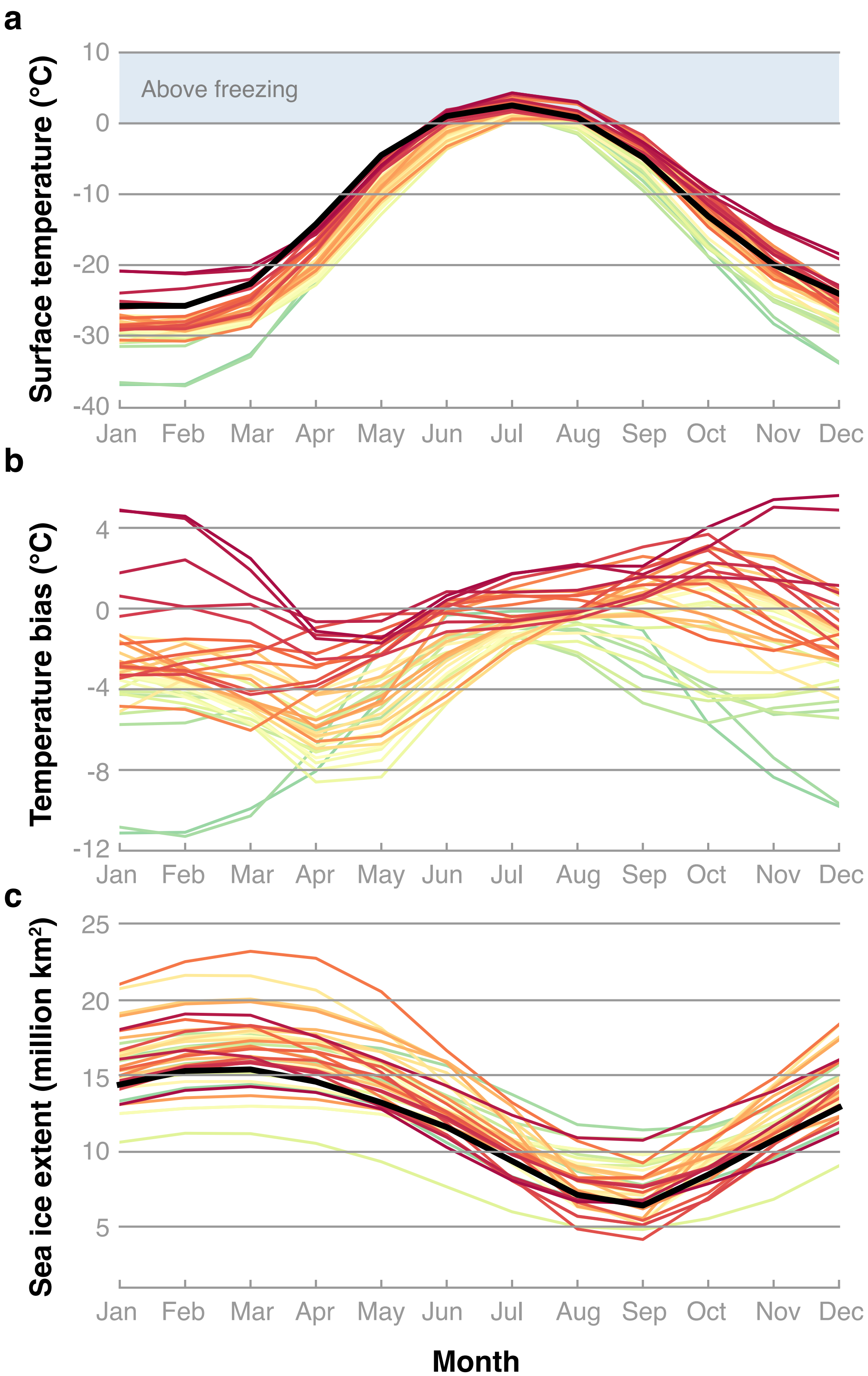
---

We are using **statistics accumulated in time** (e.g., over seasons) to

1. ***Minimize 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. ***Minimize 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.

Example of large biases in climate models: temperature and sea ice in Arctic



- Arctic temperatures and sea ice cover in current climate models have large biases
- This has enormous implications, e.g., for cryosphere, ecosystems, and hydrological impacts (CA drought)
- Reducing biases represents opportunity to improve models, including predictive capabilities

Keys to predictive success and computational feasibility

- We need out-of-sample predictive capabilities (predict a climate we have not seen)
  - Use known equations of motion to the extent possible to minimize number of adjustable parameters and avoid overfitting
- Running ESMs is computationally extremely expensive, hence computational efficiency is essential
  - For optimization, use ensemble methods (Kalman inversion and variants) that easily parallelize
  - For uncertainty quantification, use ML tools (e.g., Gaussian process emulation) to create surrogate models

One example of new SGS scheme: turbulence/ convection scheme for all forms of SGS turbulence

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

- Continuity:

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

- Scalar mean:

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

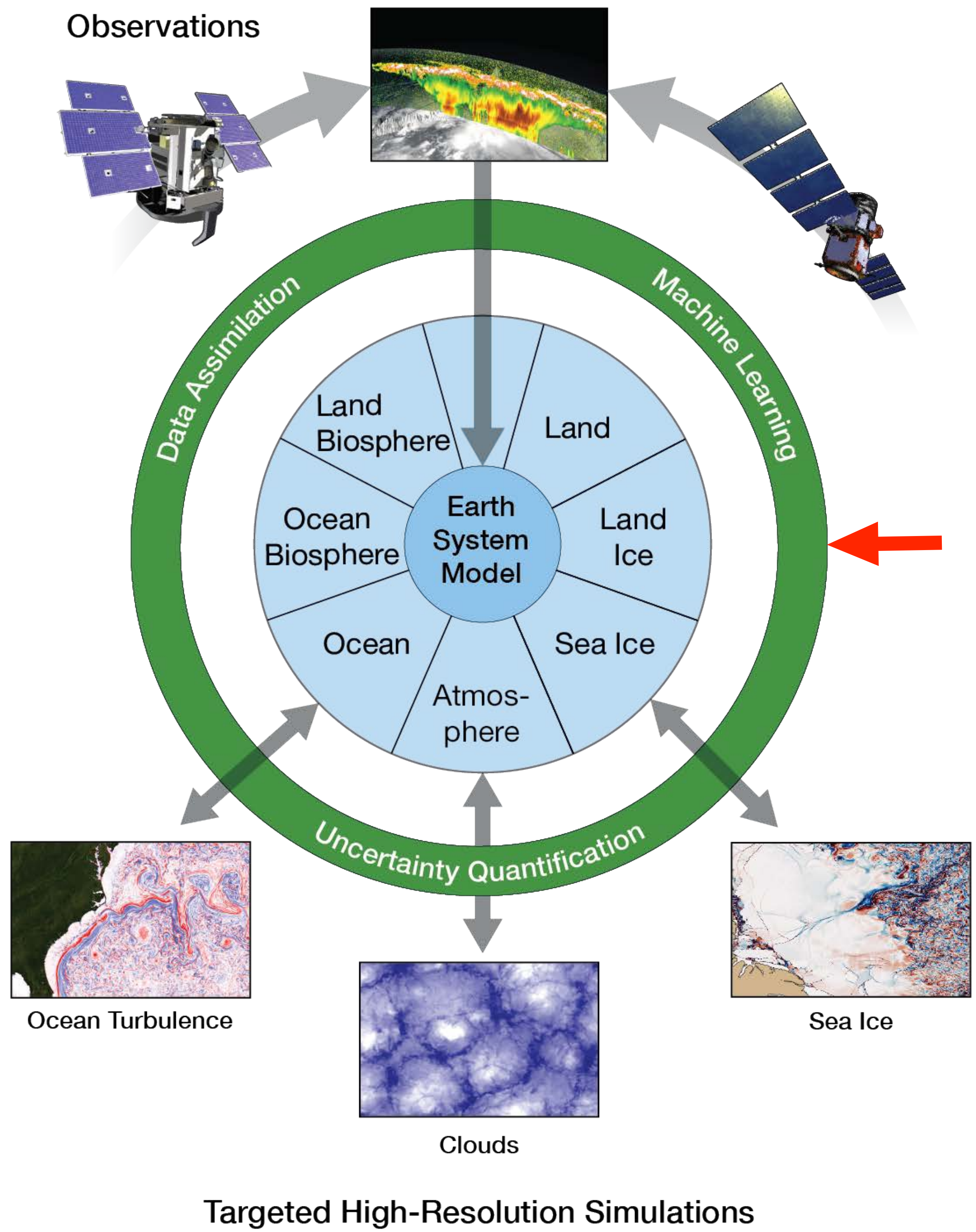
- Scalar covariance

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

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

(Tan et al., JAMES 2018)

We are currently building a modeling platform with a fresh architecture that integrates all of these elements



Approximate Bayesian calibration and uncertainty quantification for climate models: Calibrate, emulate, sample

Andrew Stuart

Emmet Cleary

Alfredo Garbuno

Learning about parameters in convection scheme of an idealized climate model as proof-of-concept

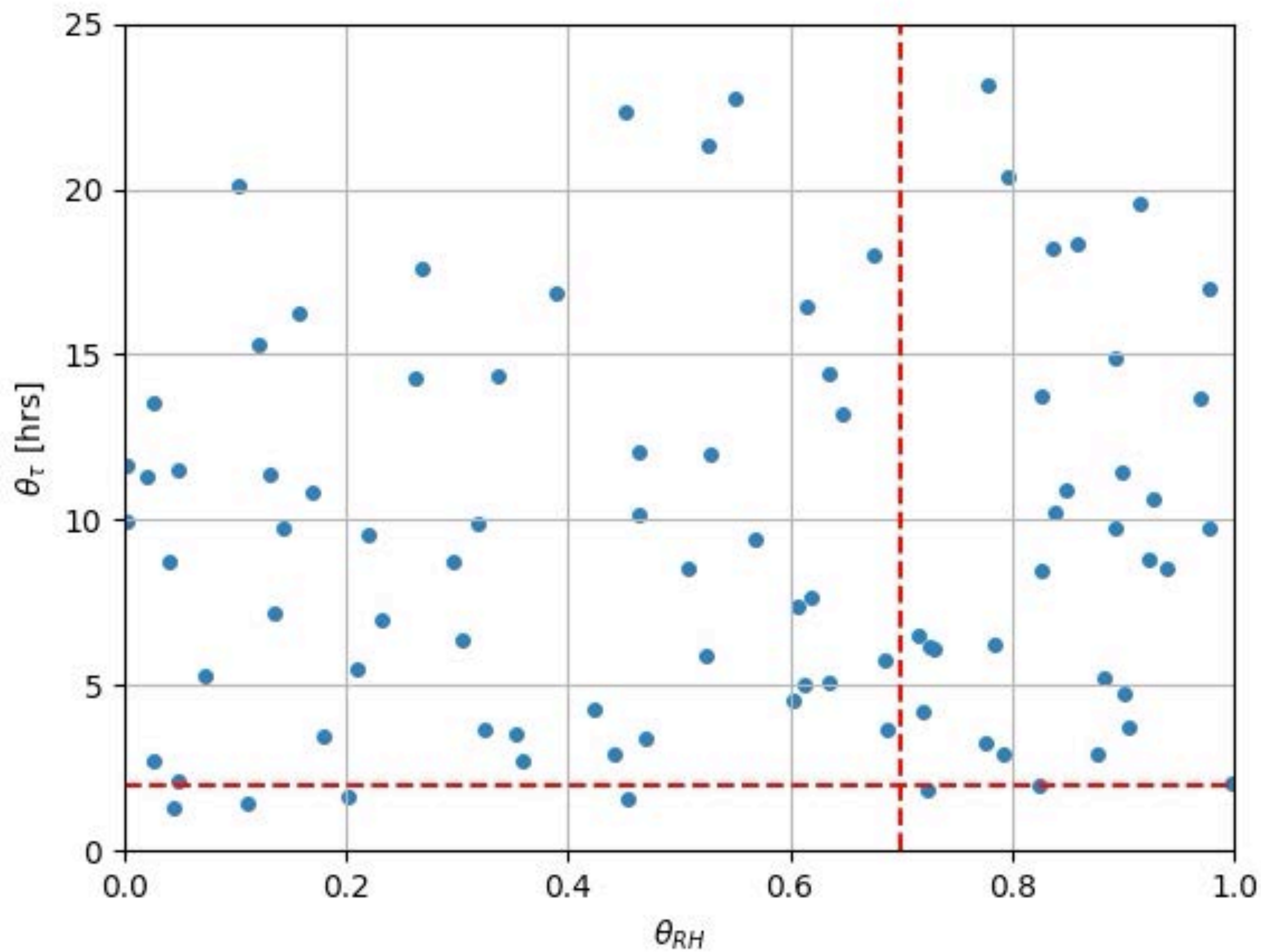
- GCM is an idealized aquaplanet model
- It has a convection scheme that relaxes temperature T and specific humidities q to reference profiles

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

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

- Two closure parameters: timescale τ and reference relative humidity $\mathrm{RH}_{\mathrm{ref}}$
- Objective function contains 96 terms (tropospheric relative humidity, precipitation, and precipitation extremes in 32 latitude bands)

Three steps: (1) Calibration by ensemble Kalman inversion (converges quickly, but ensemble collapses)



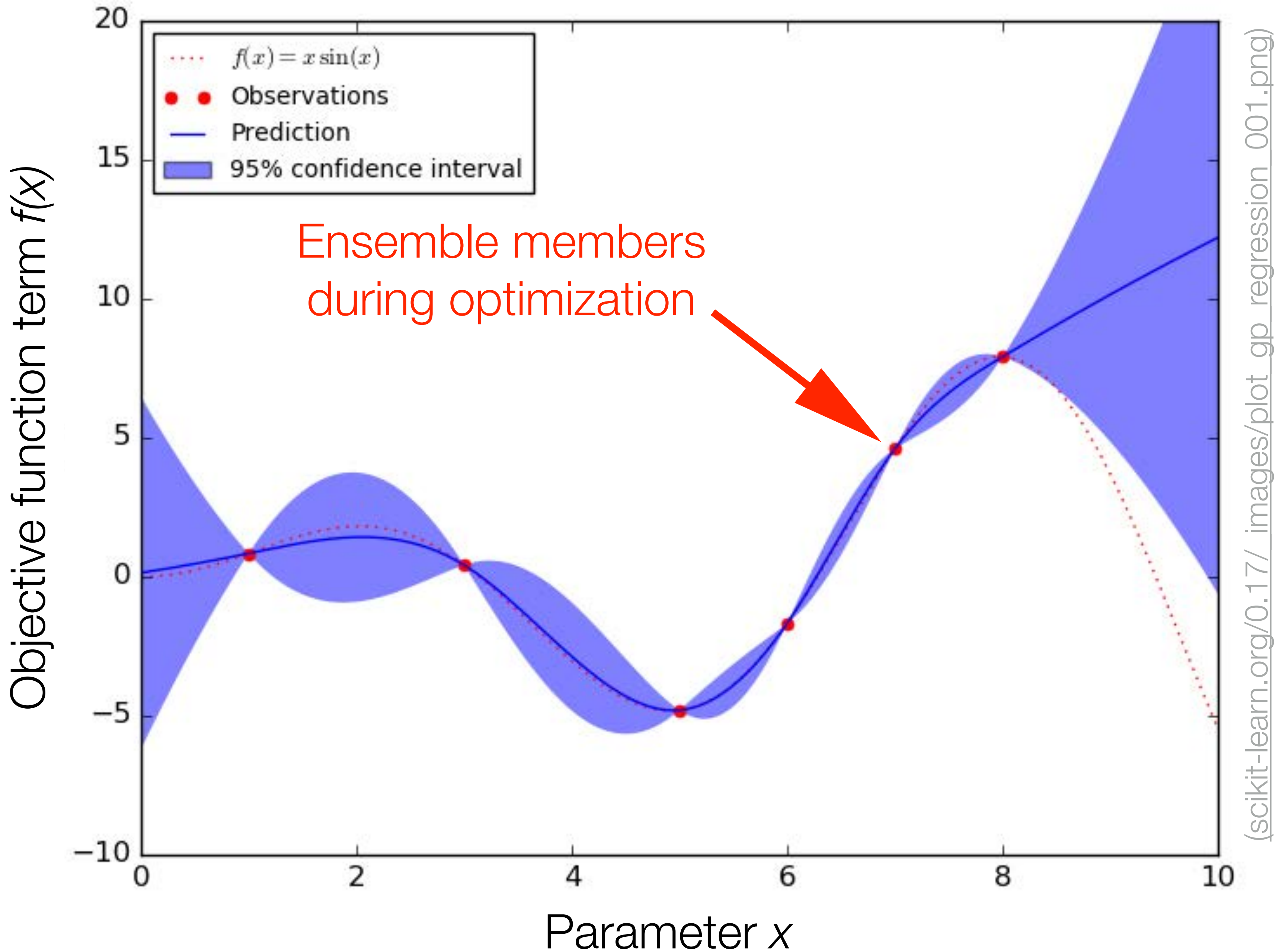
Optimization of parameters in convection scheme in an idealized GCM: ensemble of size 100 converges in ~5 iterations

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

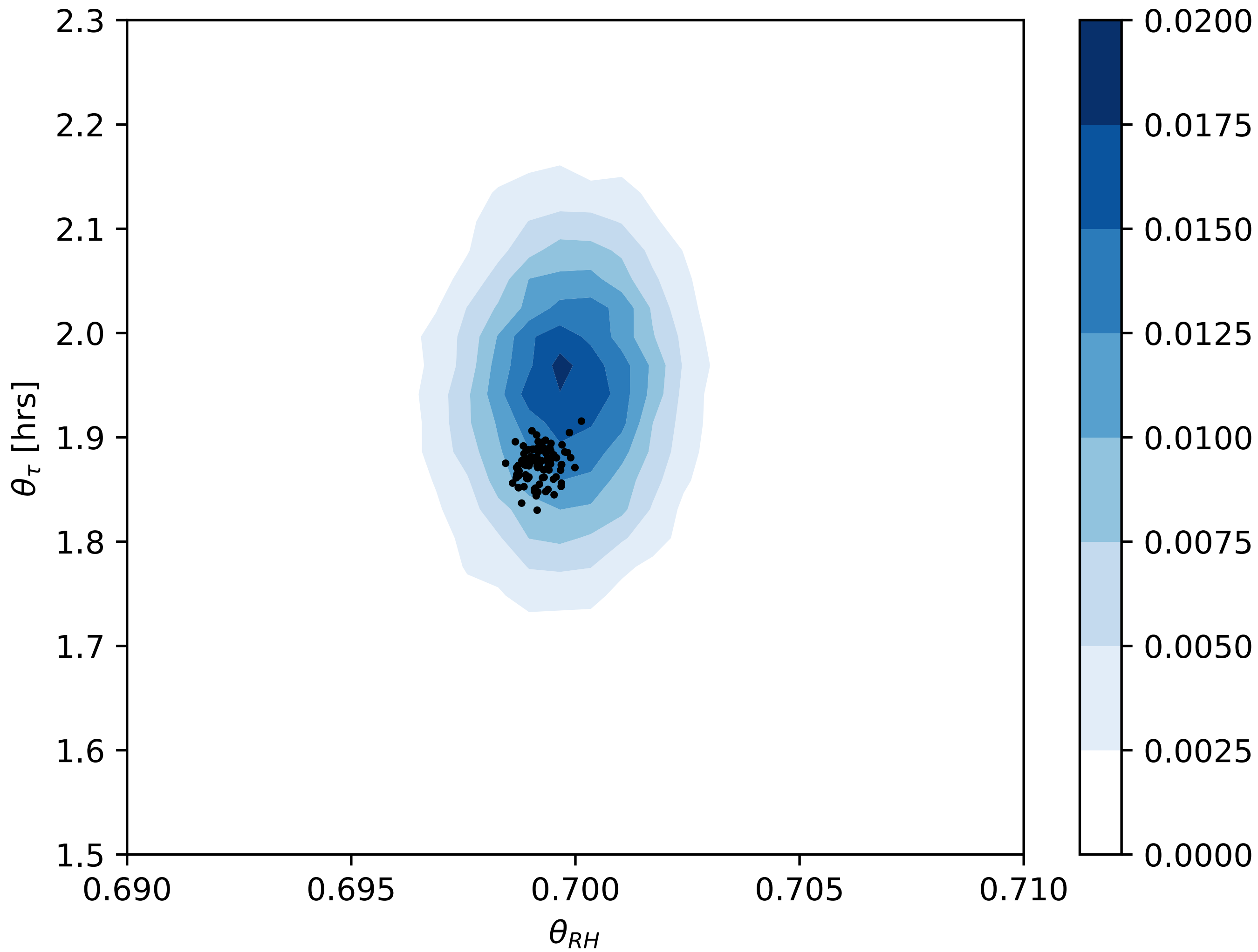
Courtesy Emmet Cleary

(2) Model emulation to recover the posterior distribution lost in optimization

- Train a Gaussian process model during the ensemble optimization, at minimal marginal computational cost



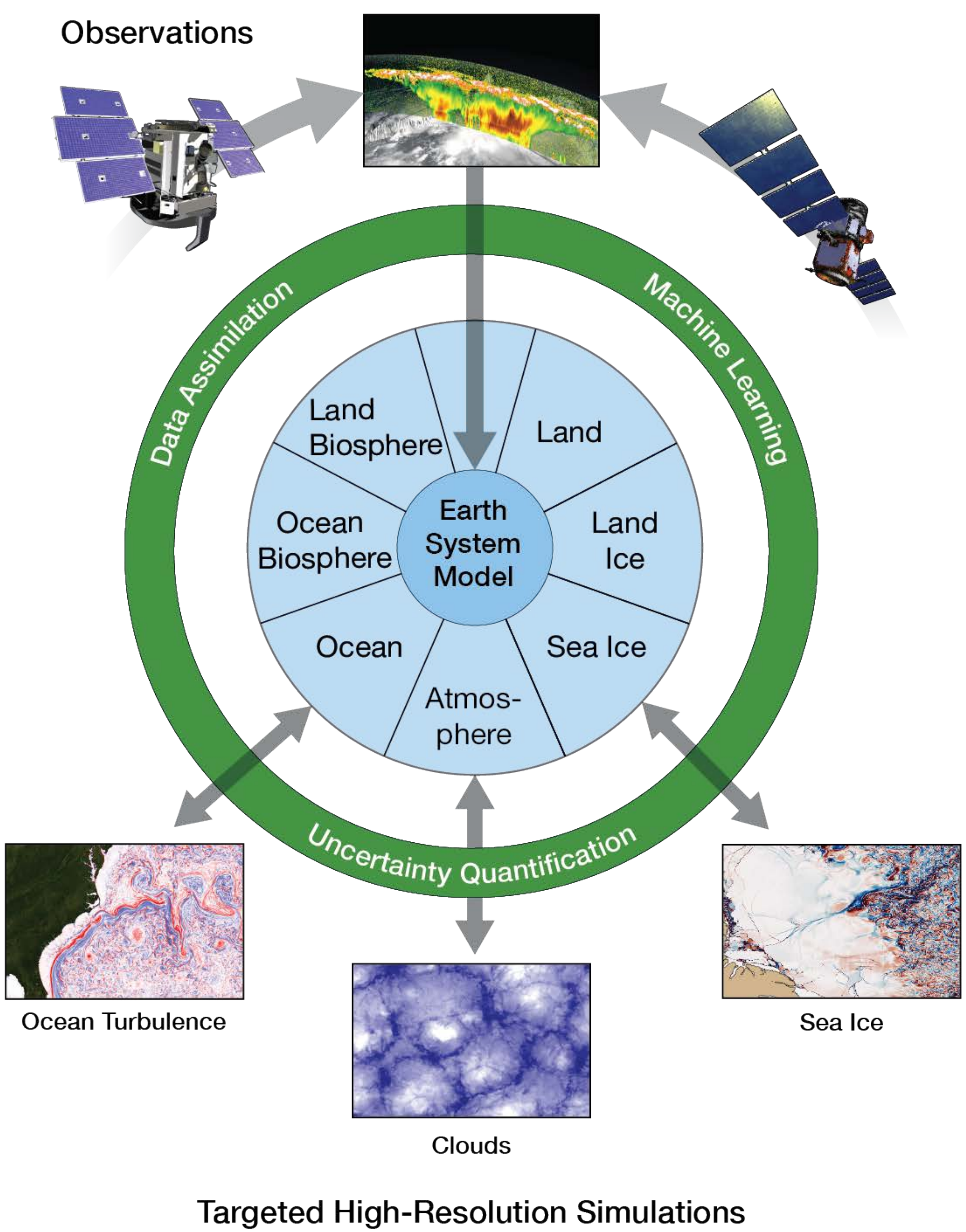
(3) Sample from emulated posterior for uncertainty quantification



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

Courtesy Emmet Cleary

We will use the same approach for calibration *all* components of the ESM *jointly*



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

Within 5 years, we will build an ESM platform that…

- Integrates data and nested high-resolution simulations from the outset in a learning environment
- Implements data assimilation and machine learning algorithms that are efficient enough for ESMs
- Has quantified uncertainties
- Will form basis for an ecosystem of applications (infrastructure planning, flood risk assessment, disaster planning etc.)

To ensure a sustainable educational pipeline in quantitative Earth science, we are establishing cross-links between graduate programs in computational and applied mathematics and in environmental science and engineering