CliMA

CLIMATE MODELING ALLIANCE

A New Approach to Climate Modeling

Tapio Schneider
Andrew Stuart
and the CliMA team

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****.*

Our goals

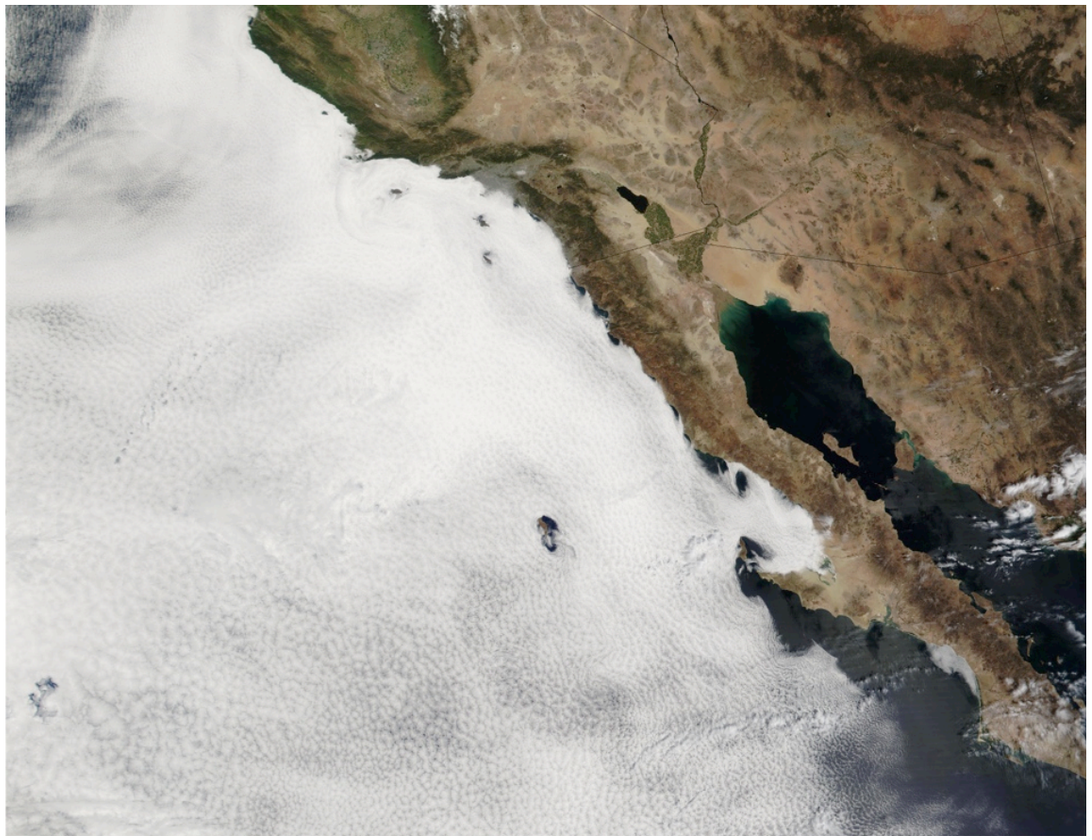
- Harness advances in the data sciences to put Earth system modeling on a data-driven footing
- Aim for a step change in the accuracy and precision of climate simulations and predictions
- Build a modeling platform that learns automatically from
  - Global observations of the climate system
  - Targeted high-resolution simulations (e.g., of ocean turbulence, atmospheric turbulence, clouds, convection)

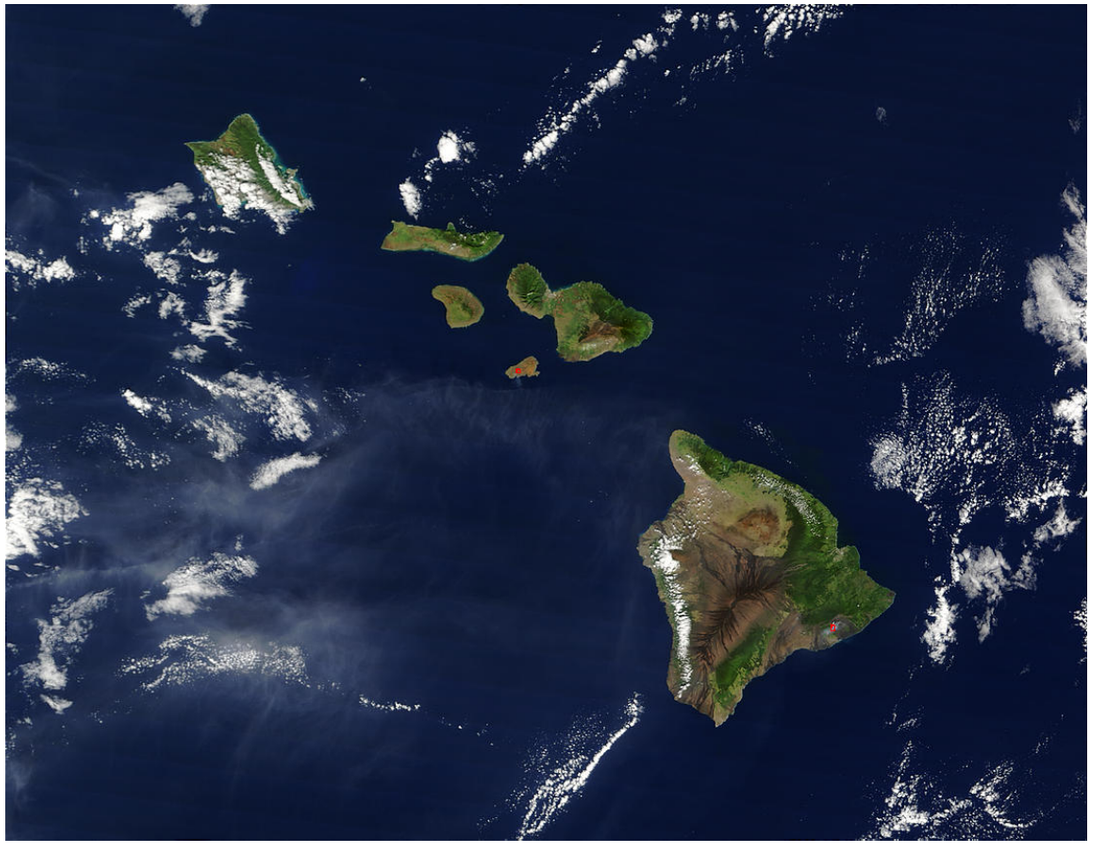
Overview of CliMA Approach and Objectives

- Structural deficiencies in current models
- Our approach to data-driven Earth system modeling
- Optimization and uncertainty quantification for ESMs
- Objectives for 2021

Structural Deficiencies in Current Models

Low clouds dominate spread in climate predictions



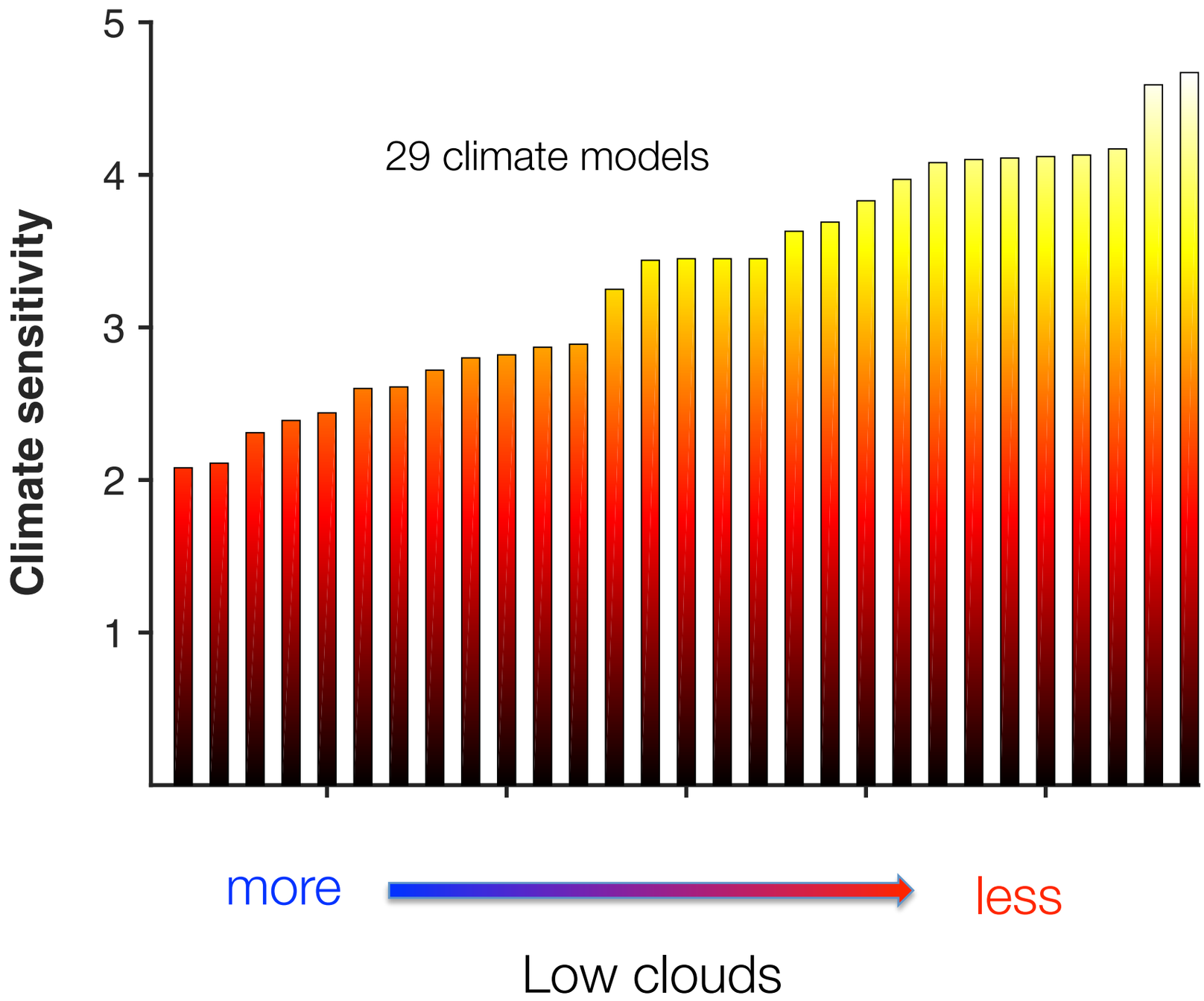


http://eoimages.gsfc.nasa.gov

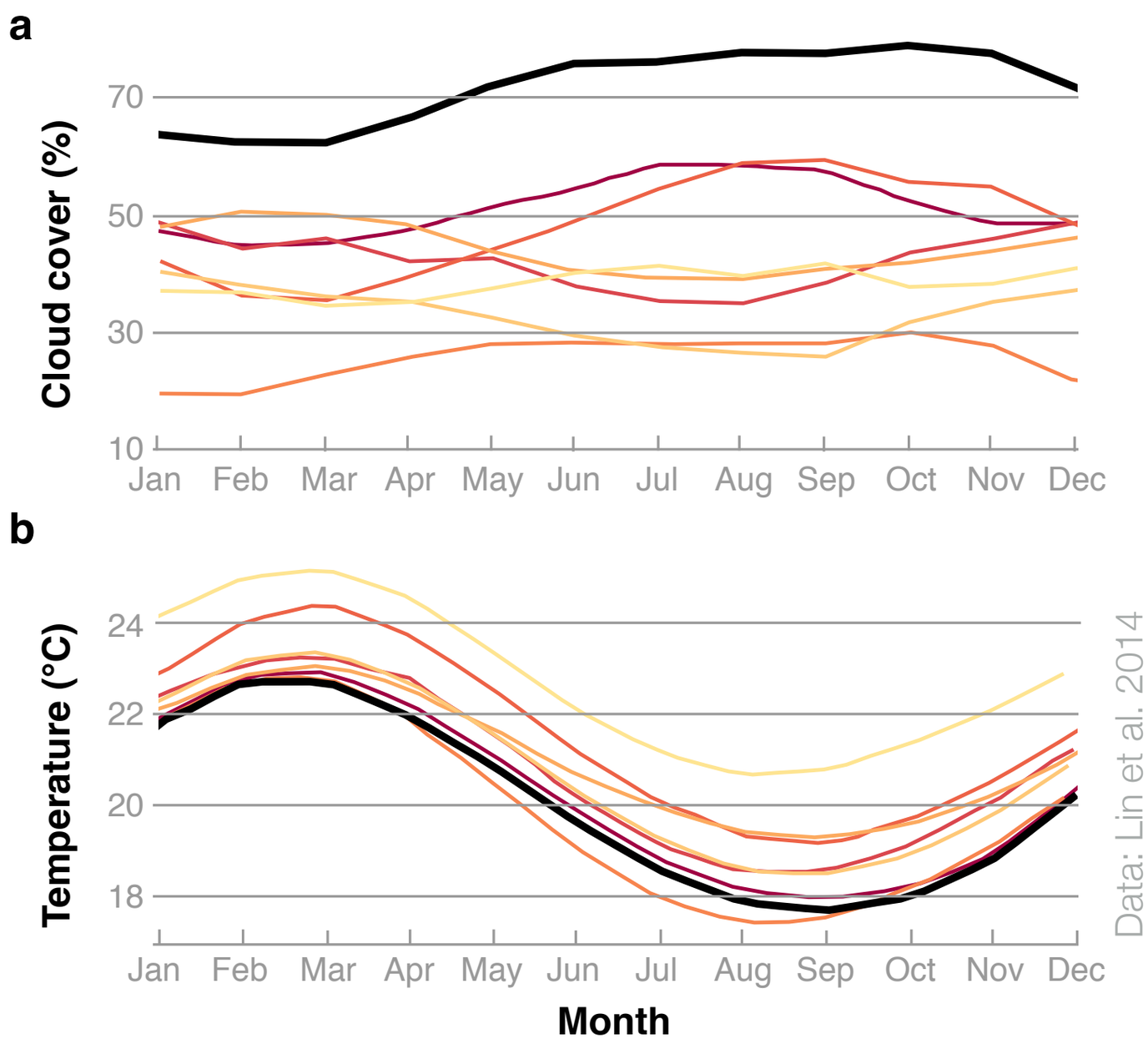
Stratocumulus: colder

Cumulus: warmer

In some models, low clouds dampen warming; in some, they amplify warming



But no climate model simulates stratocumulus well (here, off west coast of South America)

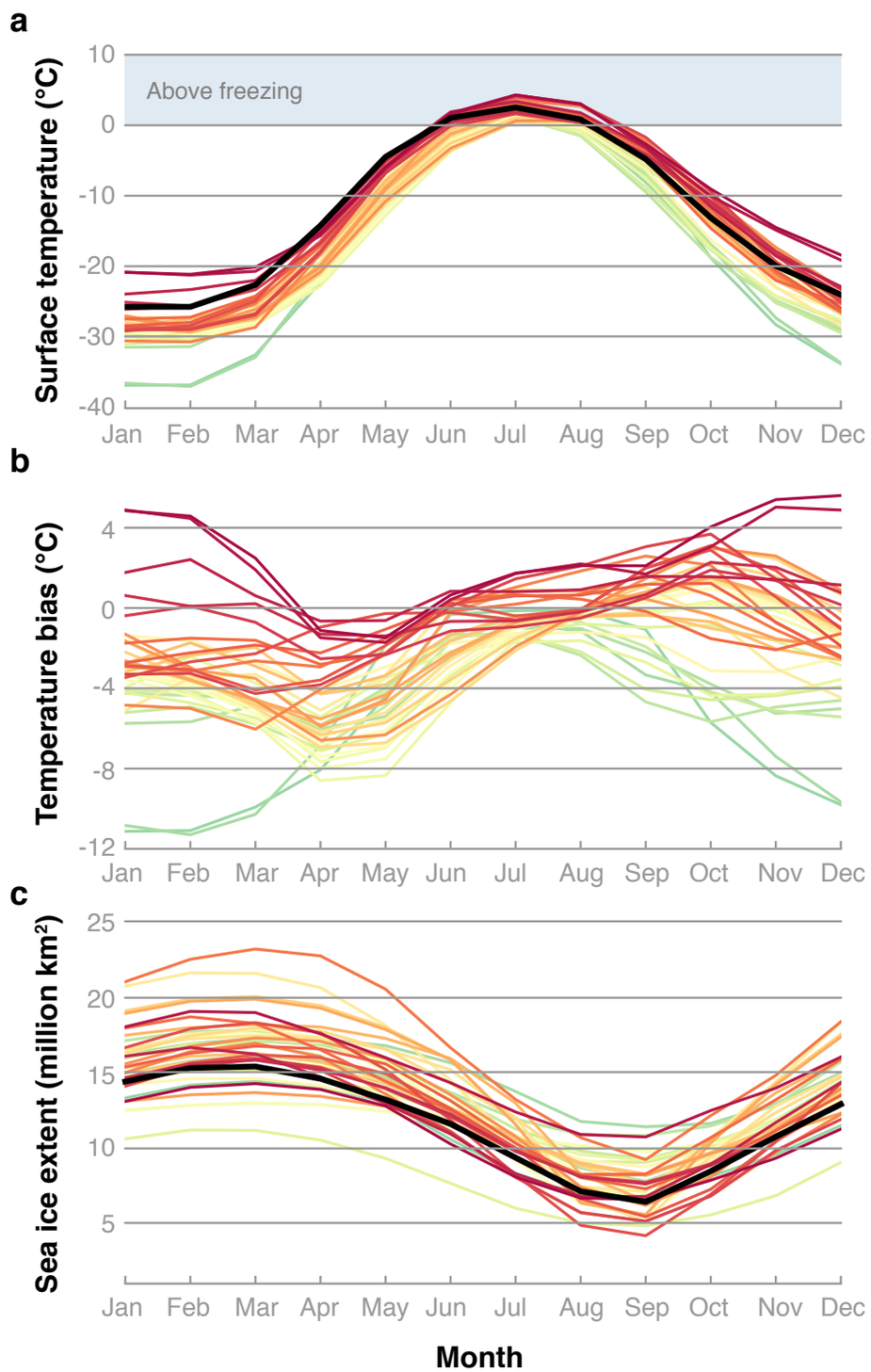


Cloud cover
(low bias)

SST
(warm bias)

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

There are also large biases, e.g., in polar regions



Arctic temperature in current models

Arctic temperature bias

Sea ice extent

These biases represent an opportunity

- We have observations revealing the biases, and we can conduct reliable high-resolution simulations of some of the processes that exhibit biases in current models
- Minimizing the biases systematically represents an opportunity to improve models
- Minimizing biases will likely improve predictive capabilities because many present-day biases (e.g., in sea ice, cloud cover, precipitation) correlate with the climate response of models

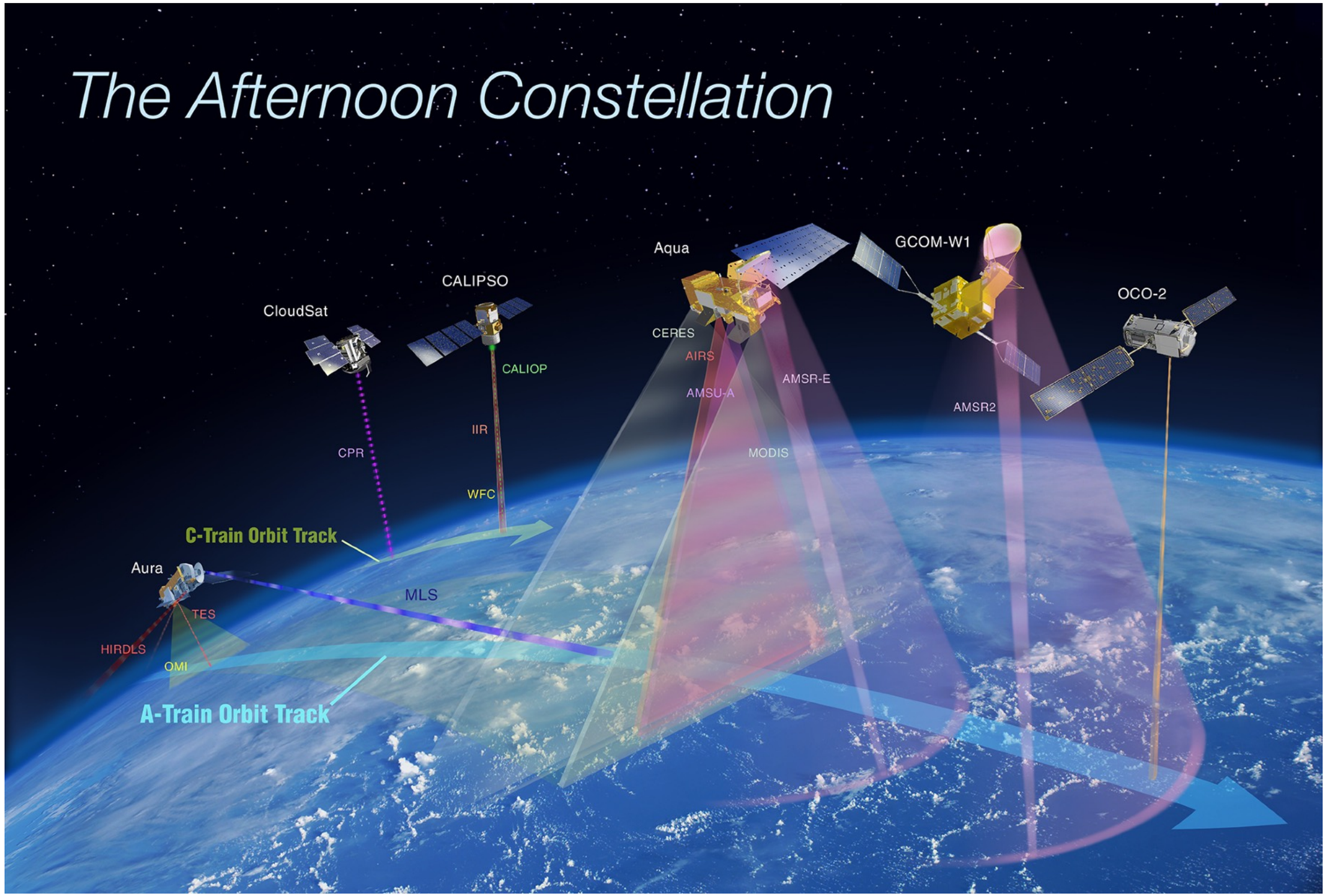
Our Approach to Data-driven Earth System Modeling

We are designing an ESM that learns automatically from two data sources

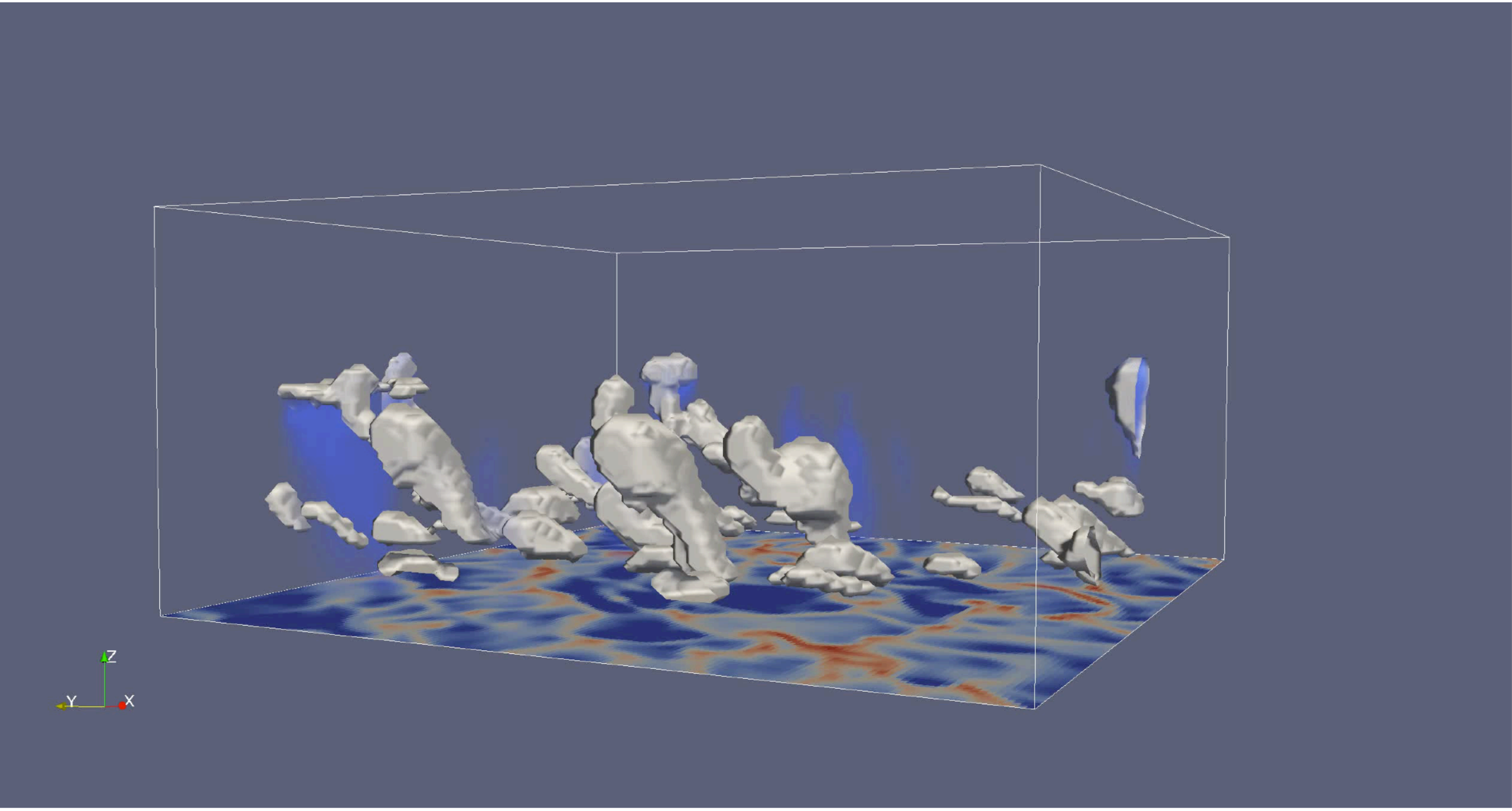
1. ***Global observations***: Our ESM will learn, e.g., from space-based measurements of temperature, humidity, clouds, ocean surface currents, and sea ice cover

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

We will harness the potential of coordinated global climate measurements to improve models



We also have faithful simulations of, e.g., clouds in limited areas, which we use to systematically inform global models



Large-eddy simulation of tropical cumulus

Simulation with PyCLES (Pressel et al. 2015)

We will optimize over aggregate climate statistics

We will use **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 (incl. high-res simulations), 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, or ecosystem carbon uptake/surface temperature covariances, which are known to correlate with the climate response of models. This amounts to exploiting “emergent constraints” in model development. Higher-order statistics relevant for predictions (e.g., precipitation extremes) can also be included in objective function.

More precisely, optimizing over aggregate climate statistics means…

- **Accumulate statistics** over timescales >10 days (so atmospheric initial condition is forgotten):

$$\langle \phi \rangle_T = \frac{1}{T} \int_{t_0}^{t_0+T} \phi(t)\, dt.$$

- Objective function should contain terms penalizing, e.g., **mean deviations** (bias) and **covariance mismatch** ("emergent constraints"):

$$J_o(\boldsymbol{\theta}) = \frac{1}{2} \| \langle \boldsymbol{f}(\boldsymbol{y}) \rangle_T - \langle \boldsymbol{f}(\tilde{\boldsymbol{y}}) \rangle_T \|^2_{\Sigma_y}$$

with moment function

$$\boldsymbol{f}(\boldsymbol{y}) = \begin{pmatrix} \boldsymbol{y} \\ y_i' y_j' \end{pmatrix}$$

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

Learning by moment matching represents challenge and opportunity

- An objective function of the form

$$J_o(\boldsymbol{\theta}) = \frac{1}{2}\|\langle \boldsymbol{f}(\boldsymbol{y})\rangle_T - \langle \boldsymbol{f}(\tilde{\boldsymbol{y}})\rangle_T\|^2_{\Sigma_y}$$

  creates computational challenges because objective function evaluation (accumulation of averages) is costly

- But matching statistics results in smoother objective functions than matching trajectories (as in NWP)

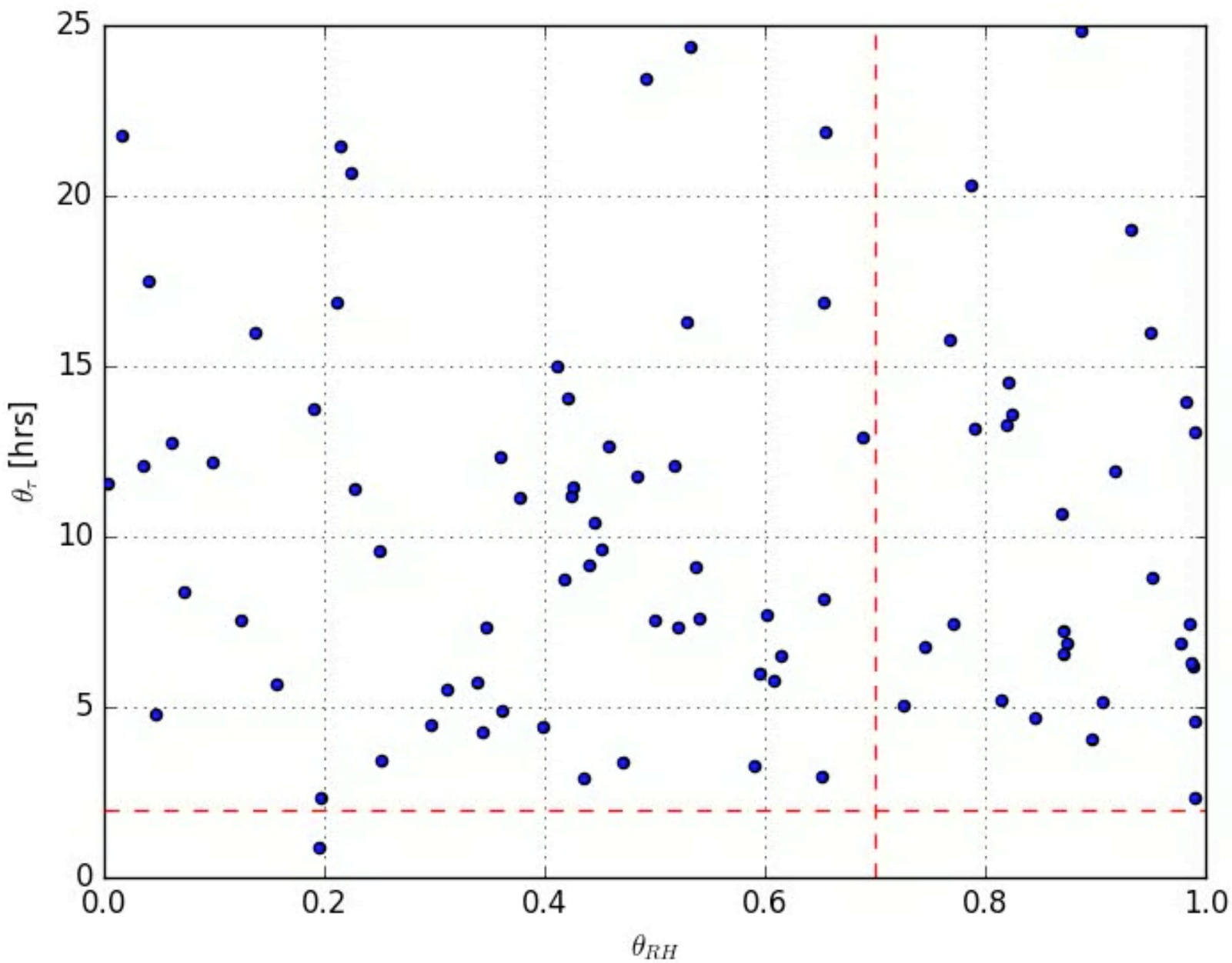
- It creates *the* opportunity to radically improve ESMs, similarly to how data assimilation has improved NWP

- Minimizing such objective functions and quantifying uncertainties is just feasible

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

Optimization and uncertainty quantification for ESMs

Key to computational feasibility is separation of optimization and UQ: (1) Ensemble flow methods for optimization



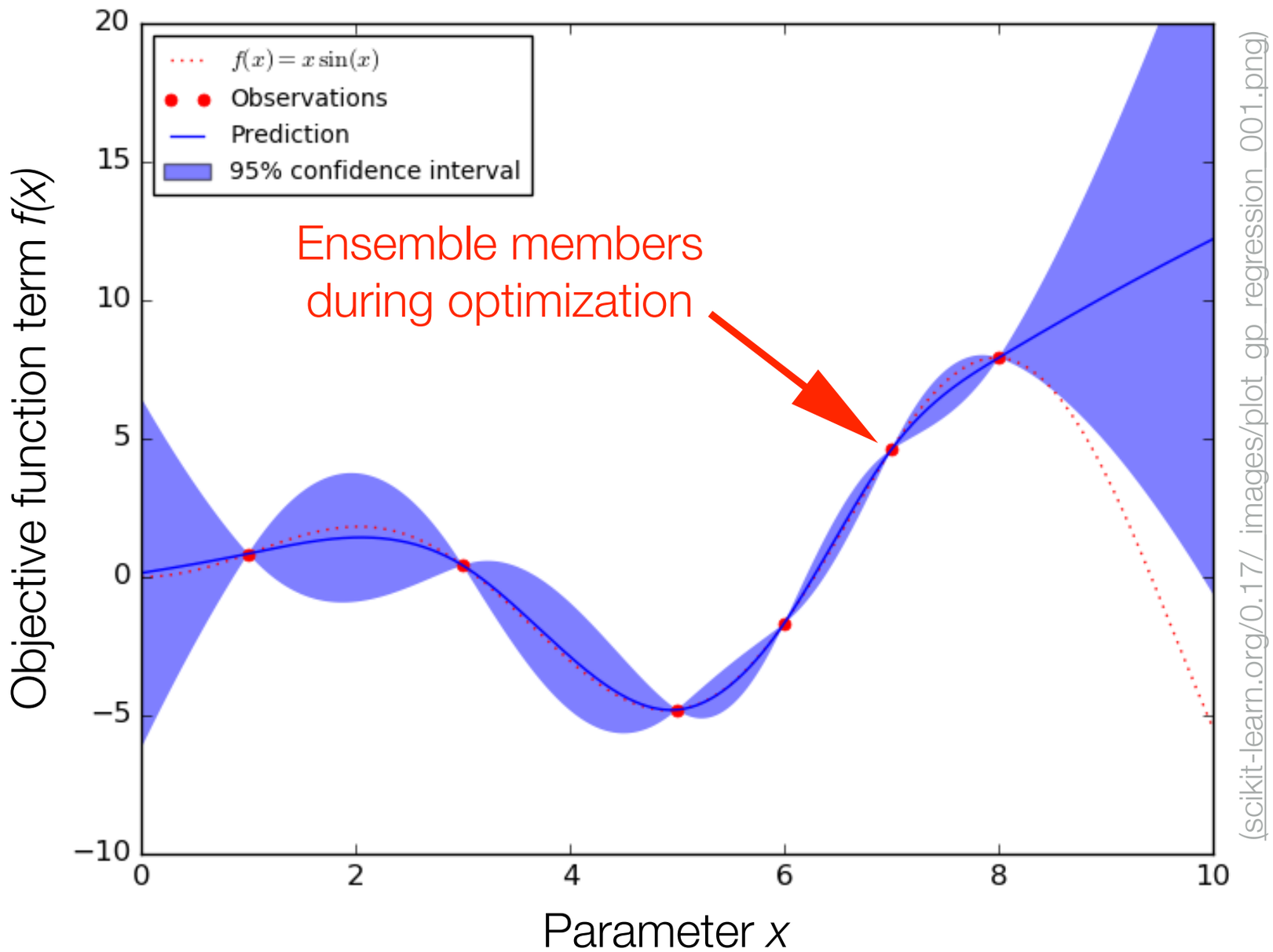
Optimization of parameters in convection scheme in an idealized GCM: ensemble of size 100 converges in ~5 iterations

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

Courtesy Emmet Cleary

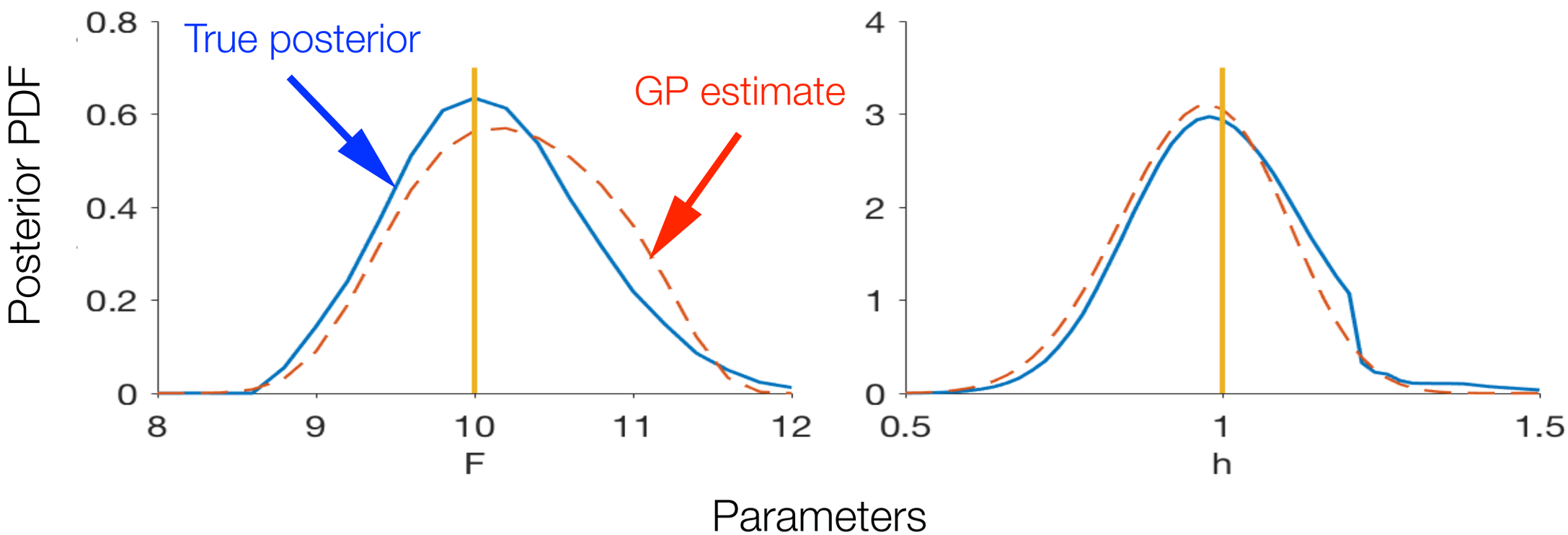
(2) Model emulation to recover the uncertainty lost in optimization

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



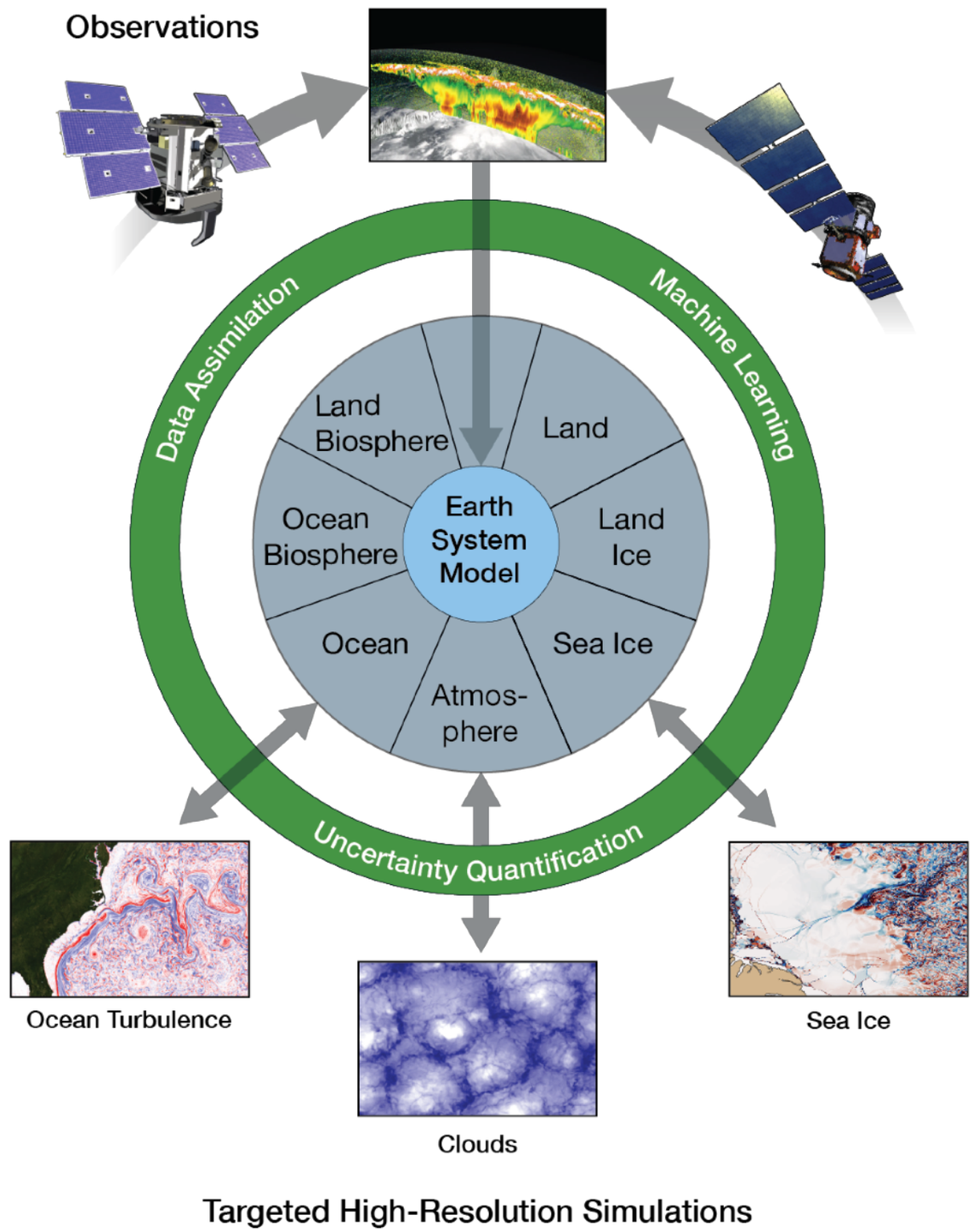
(scikit-learn.org/0.17/_images/plot_gp_regression_001.png)

GP emulation successfully recovers uncertainty in toy problems (here, Lorenz-96)



Demonstration in GCM is ongoing
(Ground-truthing UQ in idealized GCM with MCMC is expensive but feasible)

We are building a model with a fresh architecture that integrates component models within overarching DA/ML environment



Objectives for 2021

- Co-develop new atmosphere and ocean dynamical cores that exploit emerging computing architectures (CPUs, GPUs, etc.) and can run high-resolution simulations on the fly
- Develop atmosphere, ocean, and land process models suitable for DA/ML approaches
- Demonstrate fast and efficient DA/ML algorithms
- Demonstrate learning from observational data and nested high-resolution simulations in atmosphere and ocean models
- Proof-of-concept of uncertainty quantification in process models and climate predictions

Our funders

We are supported by a consortium of private foundations and federal agencies, led by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program, and including Mission Control for Earth, Paul G. Allen Philanthropies, Charles Trimble, and NSF

SCHMIDT FUTURES

MISSION CONTROL EARTH

PAUL G. ALLEN PHILANTHROPIES

NSF

CHARLES TRIMBLE

HEISING-SIMONS FOUNDATION

RONALD AND MAXINE LINDE CLIMATE CHALLENGE