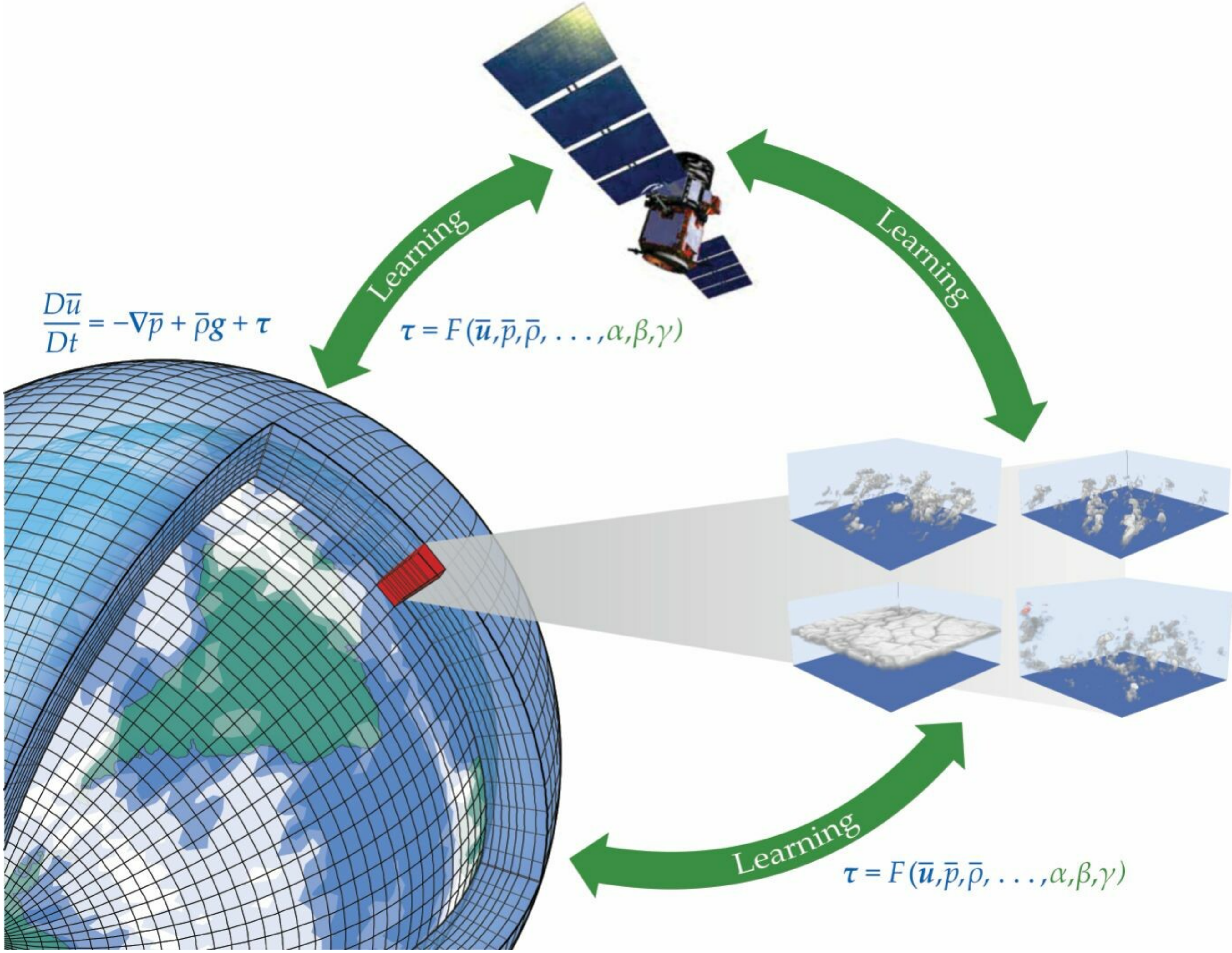


AI-Aided Hybrid Modeling

Tapio Schneider and the CliMA Team (clima.caltech.edu)

To accelerate climate modeling and climate science in the service of society, we need models that…

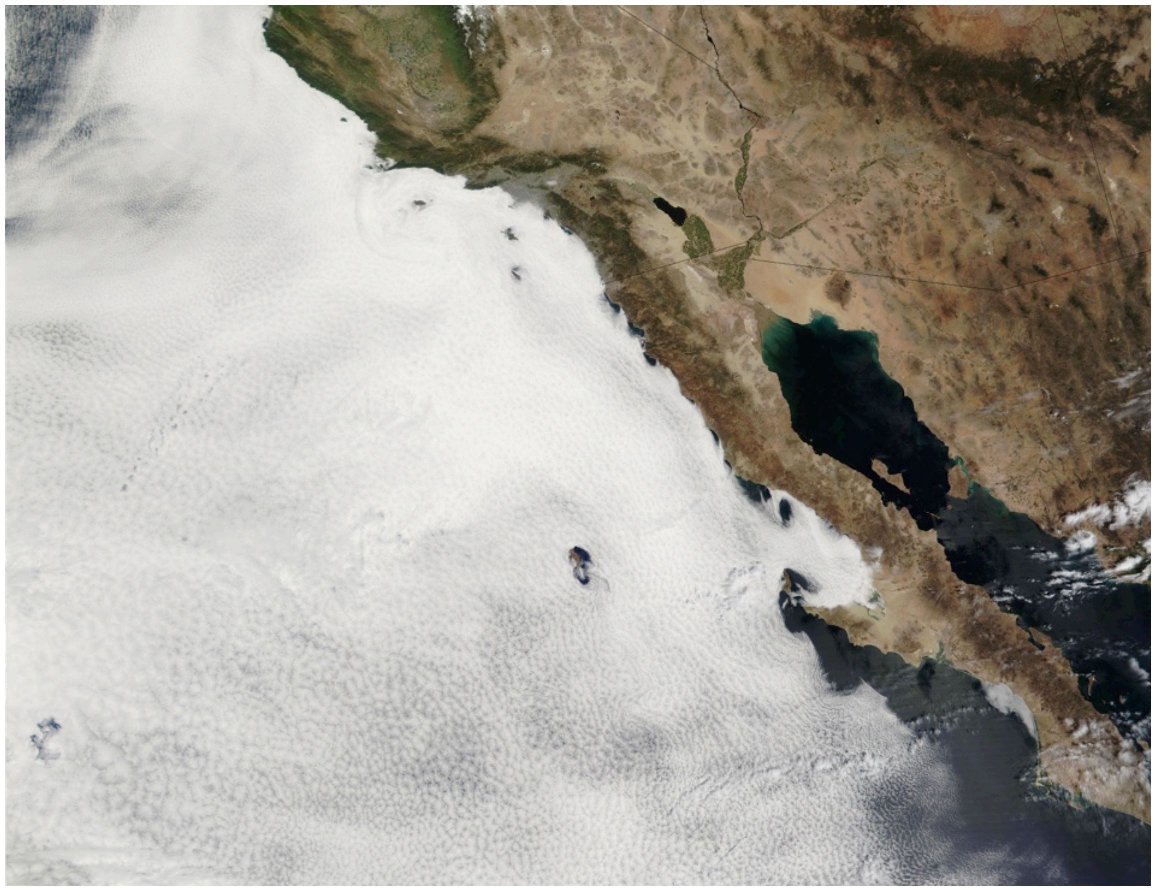
- Allow rapid iteration (e.g., run ensemble integrations, study emergent physics in hierarchies of complexity etc.)

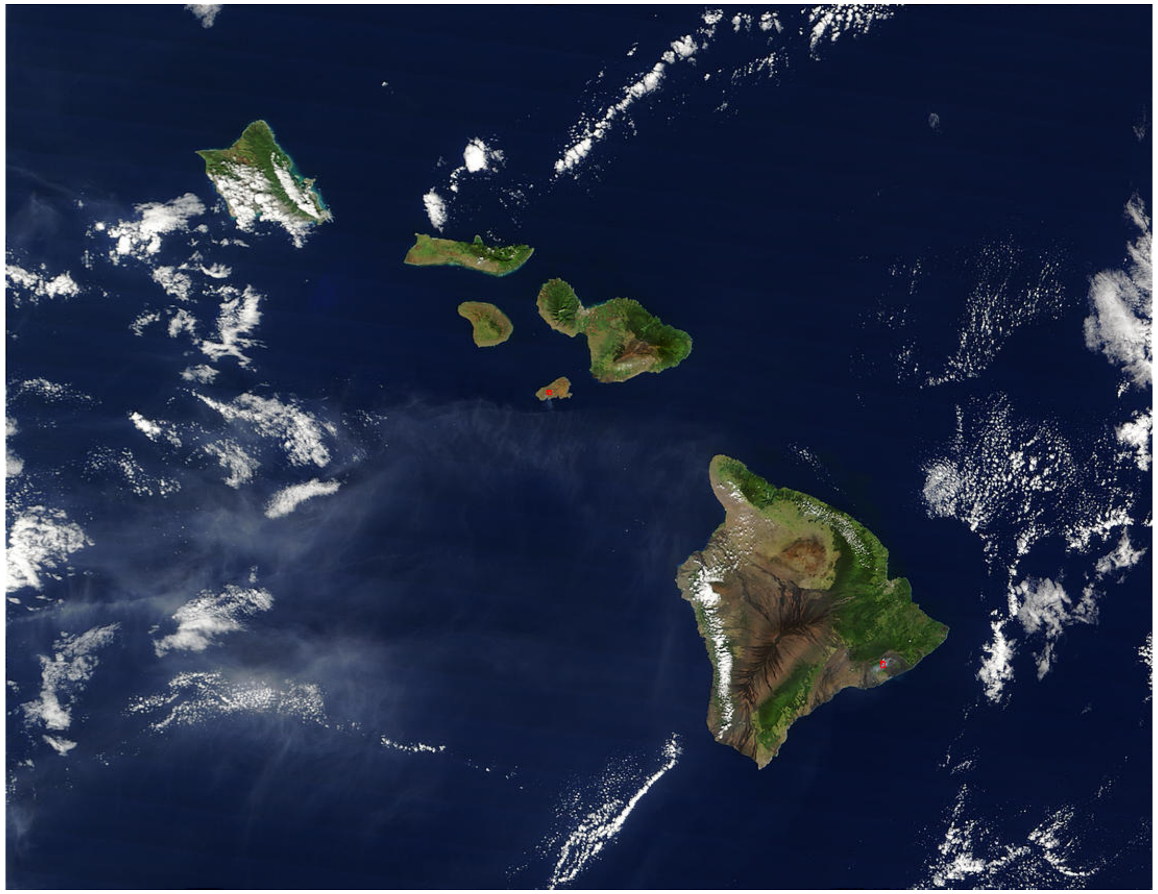
- Are more accurate than existing ones (e.g., better precipitation simulations, including extremes)

- Have quantified uncertainties (e.g., to quantify tail risks)

*To get there, we need to go beyond model (and data) comparison to **learning** from data*

We also need kilometer-scale models, but resolution alone will not break through the primary uncertainties, e.g., from low clouds and microphysics





http://eoimages.gsfc.nasa.gov

Stratocumulus: colder

Cumulus: warmer

They will remain globally unresolvable for decades to come

Schneider et al., *Nature Climate Change* 2017

We can get generalizable, interpretable models with UQ by *combining* the best of reductionist science with data science approaches

- **Deep learning**’s success rests on *overparameterization*:
    - Leads to expressive models and data-hungry methods
    - Makes generalizability, interpretability, and UQ challenging
- **Reductionist science**’s success rests on *parametric sparsity:*
    - Generalizable and interpretable (e.g., Newton’s Law of Universal Gravitation)
    - Reaches limits in complex systems such as the Earth system

Hybrid models combine both, traditional reductionist science with AI where reductionism reaches its limits

Schneider, Jeevanjee, Socolow, *Accelerating Progress in Climate Science*, *Physics Today* 6/2021

E.g., to model turbulence, convection, and clouds, we use a unified model, derived by conditional averaging of equations of motion

Coarse-graining fluids equations by conditionally averaging over coherent plumes (*i*=1, …, *N*) and environment (*I*=0), leading to exact conservation laws:

- 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:

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

Building on work by Siebesma, Teixeira et al. (ECMWF); Tan et al., *JAMES* 2018, Cohen et al. *JAMES* 2020, Lopez-Gomez et al. *JAMES* 2020

E.g., to model turbulence, convection, and clouds, we use a unified model, derived by conditional averaging of equations of motion

Coarse-graining fluids equations by conditionally averaging over coherent plumes (*i*=1, …, *N*) and environment (*I*=0), leading to exact conservation laws:

- 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:

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

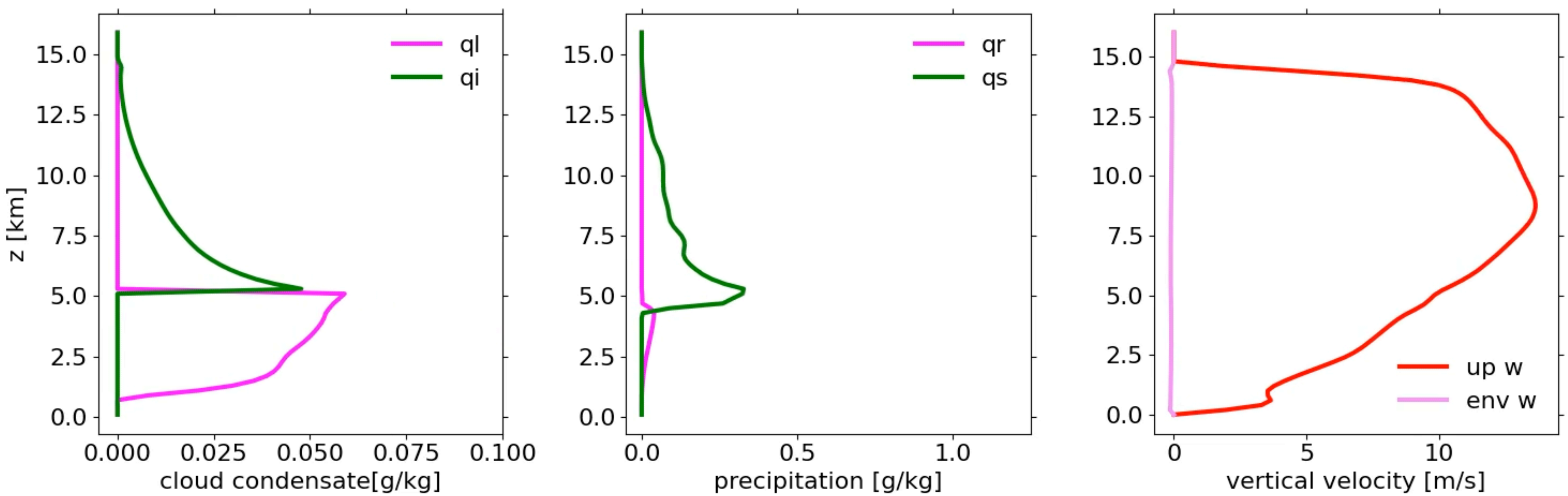
Closure functions are excellent targets for (explainable!) ML approaches; they can be stochastic and should include structural error

Building on work by Siebesma, Teixeira et al. (ECMWF); Tan et al., *JAMES* 2018, Cohen et al. *JAMES* 2020, Lopez-Gomez et al. *JAMES* 2020

Key to learning from diverse data sources: Treat learning problem as **inverse problem** (rather than supervised learning) and learn from ***time-averaged climate statistics***

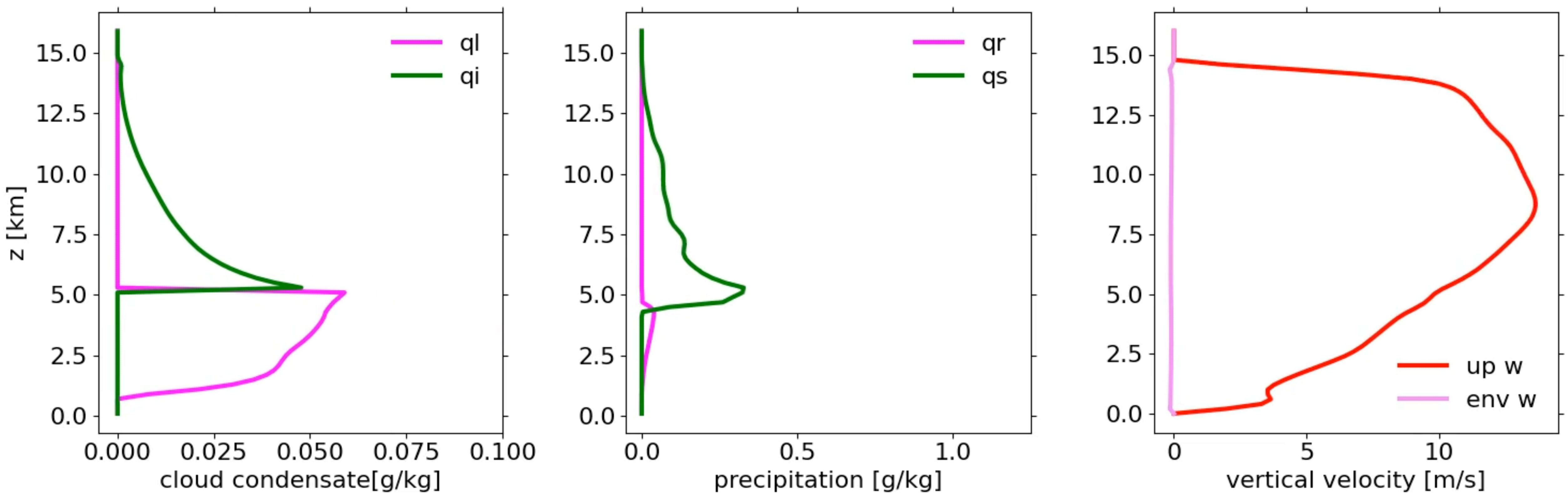
- Spatial smoothness of statistics **overcomes observation/simulation resolution mismatch**
- **Climate-relevant statistics** can include, e.g., emergent constraints and precipitation extremes
- Most **ML methods** (e.g., neural networks, neural operators, random feature models) embedded in host models (e.g., for entrainment) **can be trained in this way** (Kovachki & Stuart 2019; Lopez-Gomez et al. 2022)

One example: continuous transition from BL turbulence, to shallow convection, to deep convection in one unified parameterization



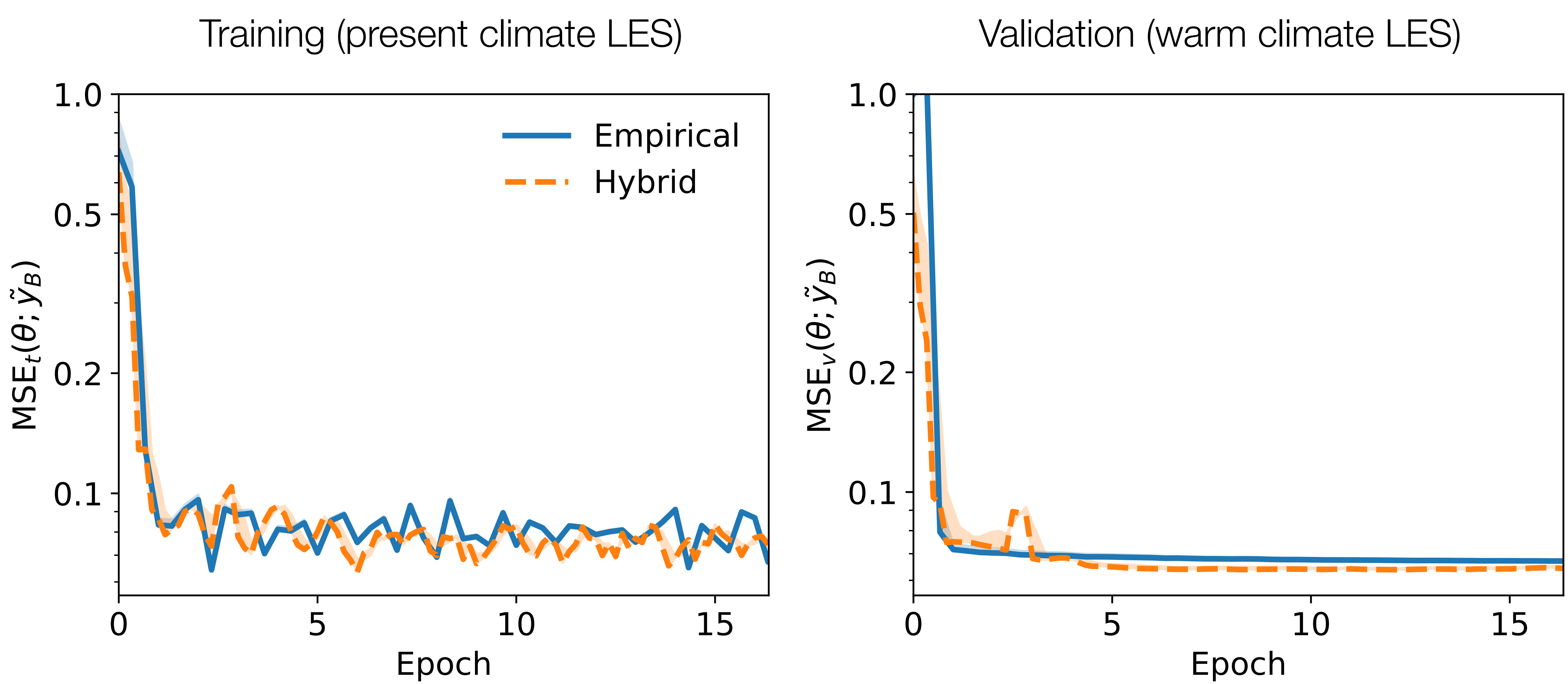
(Anna Jaruga, in prep.)

One example: continuous transition from BL turbulence, to shallow convection, to deep convection in one unified parameterization



(Anna Jaruga, in prep.)

Replacing empirical entrainment/detrainment rates by shallow NN and calibrating with LES library improves parameterization and generalizes well



Training epoch (Kalman iteration with mini-batching)

(Ignacio Lopez Gomez, in prep.)

Main Messages

- Let's move beyond MIPs and *comparing* with observations to **learning from diverse data**, be they observations or computationally generated data
- Retain decades of hard-won domain knowledge: **Augment process models with AI approaches;** do not, by default, replace them.
- **Learning from climate statistics** (rather than, e.g., states or tendencies) **circumvents many issues that have limited supervised learning approaches so far** (e.g., unavailability of sufficient labeled data)
- **Algorithms** for solving these learning and UQ tasks (borrowing on ensemble Kalman methods from NWP) are **now available** (Cleary et al 2021; Dunbar et al. 2021; Howland et al. 2022; Lopez-Gomez et al., in prep.; see clima.caltech.edu/publications)