

Automating Earth System Model Development with Agentic AI

Kara D. Lamb

ICES Biennial Workshop VIII
AI & Science
October 1, 2026



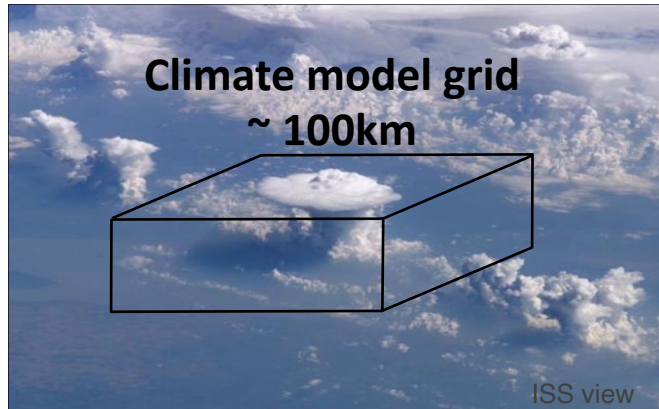
LEAP



COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

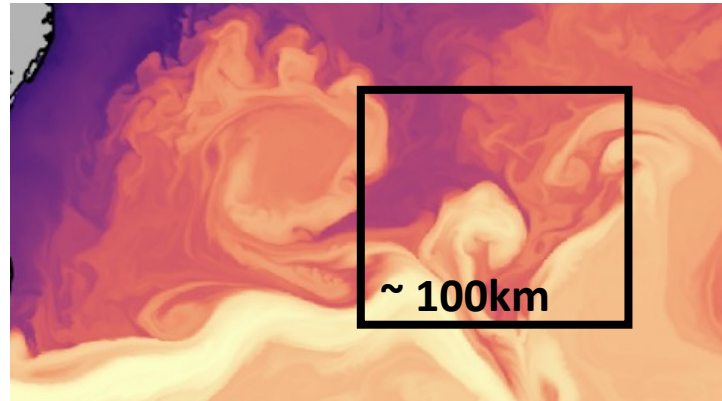
Parameterizations dominate near term uncertainty in climate projections

Atmospheric Processes



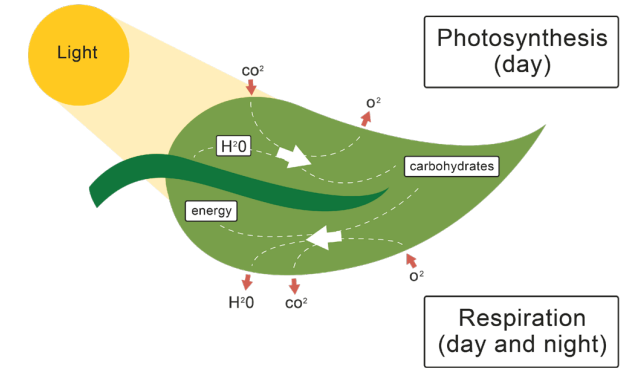
e.g. Gentine et al. 2018, Rasp et al. 2018

Ocean Processes



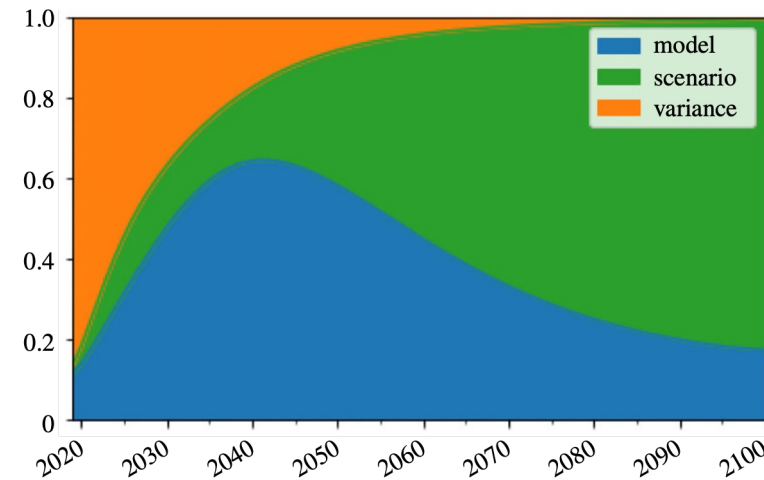
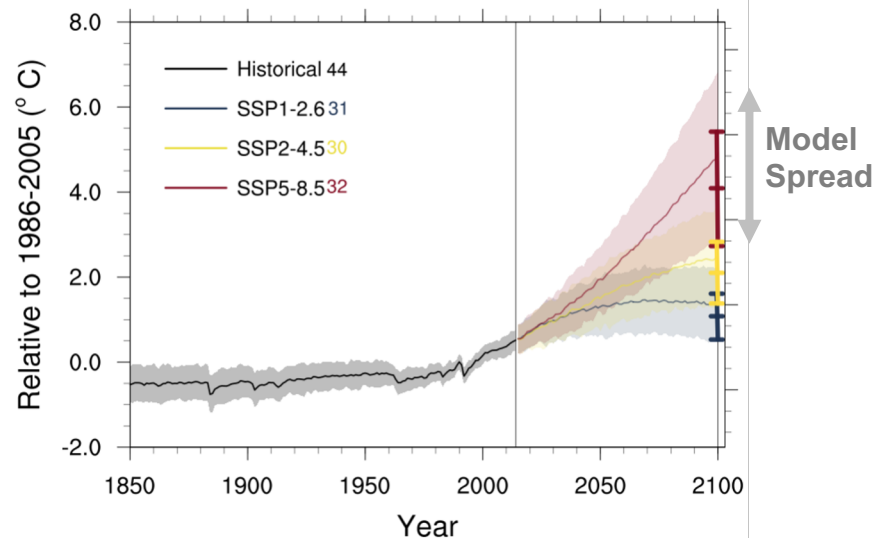
e.g. Bolton & Zanna, 2019; Zanna & Bolton, 2020.

Land Processes



e.g. Dagan et al. 2020

TAS, global, CMIP6.

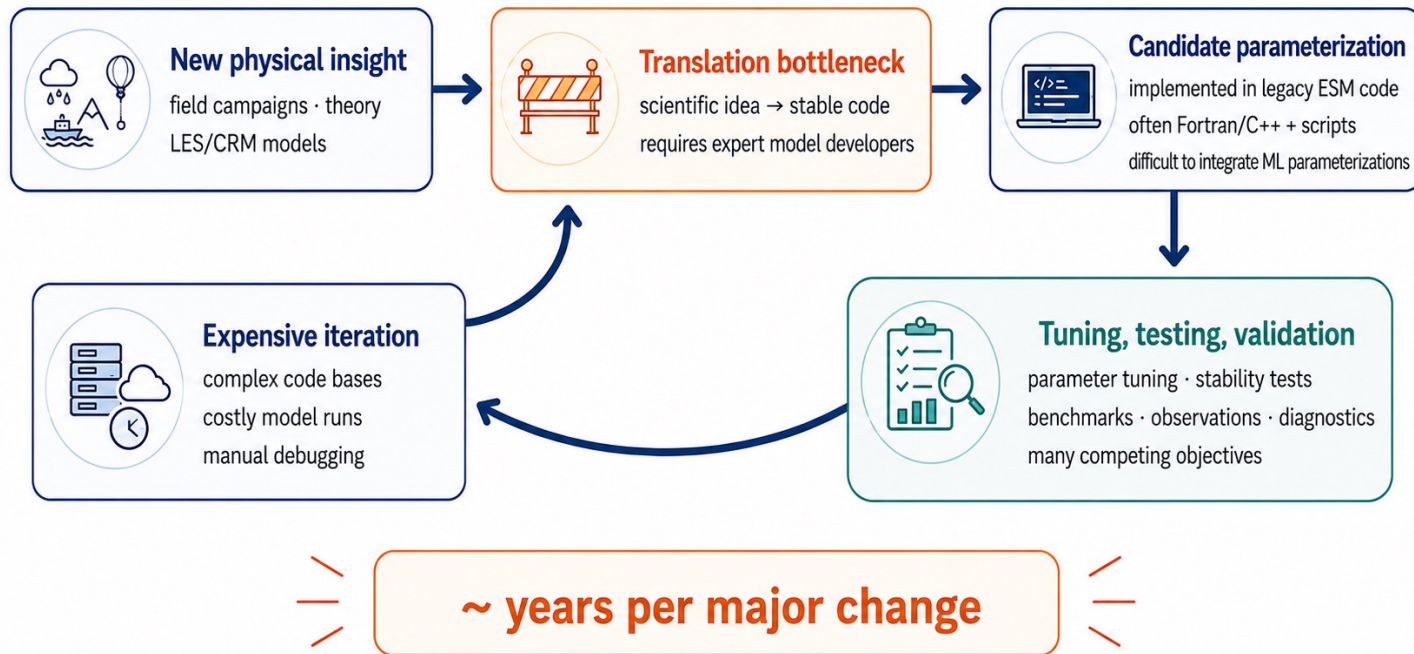


D. Watson-Parris, *Philosophical Transactions A*, 2025.



The current Earth System Model (ESM) development process is slow!

Why new parameterizations take years to enter ESMs



Current workflow is fragmented, slow, manual, & relies on expert intuition

THE ART AND SCIENCE OF CLIMATE MODEL TUNING

FRÉDÉRIC HOURDIN, THORSTEN MAURITSEN, ANDREW GETTELMAN, JEAN-CHRISTOPHE GOLAZ, VENKATRAMANI BALAJI, QINGYUN DUAN, DORIS FOLINI, DUOYING JI, DANIEL KLOCKE, YUN QIAN, FLORIAN RAUSER, CATHERINE RIO, LORENZO TOMASSINI, MASAHIRO WATANABE, AND DANIEL WILLIAMSON

We survey the rationale and diversity of approaches for tuning, a fundamental aspect of climate modeling, which should be more systematically documented and taken into account in multimodel analysis.

How to systematically integrate new physical understanding into ESMs?



1) Agentic AI

Code translation, literature review, parameterization development and integration

→ *Enabling complex scientific workflows*

- *Automated laboratories*
- *AI Co-Scientists*
- *Automating model selection for Scientific AI*

2) Differentiable Programming

Differentiable programming (JAX) for fast, gradient-enabled models

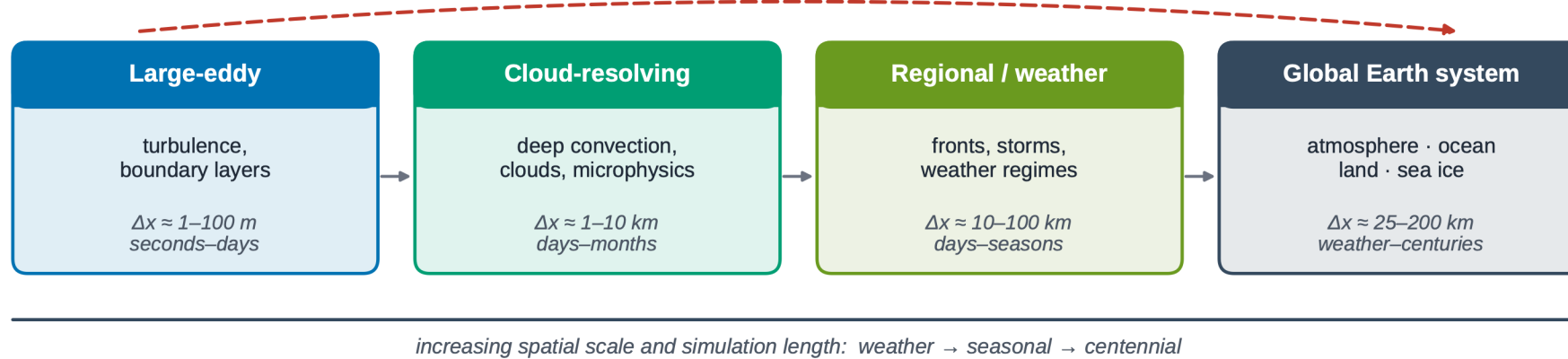
→ *Enables hybrid-physics ML models*

- *Direct access to model adjoint*
- *Easily integrate with ML parameterizations*
- *Simultaneous tuning of physics & ML model parameters*
- *Native on GPUs & TPUs for fast parallel computation*

Agentic AI to develop an intermediate complexity differentiable ESM

a One differentiable code base across scales

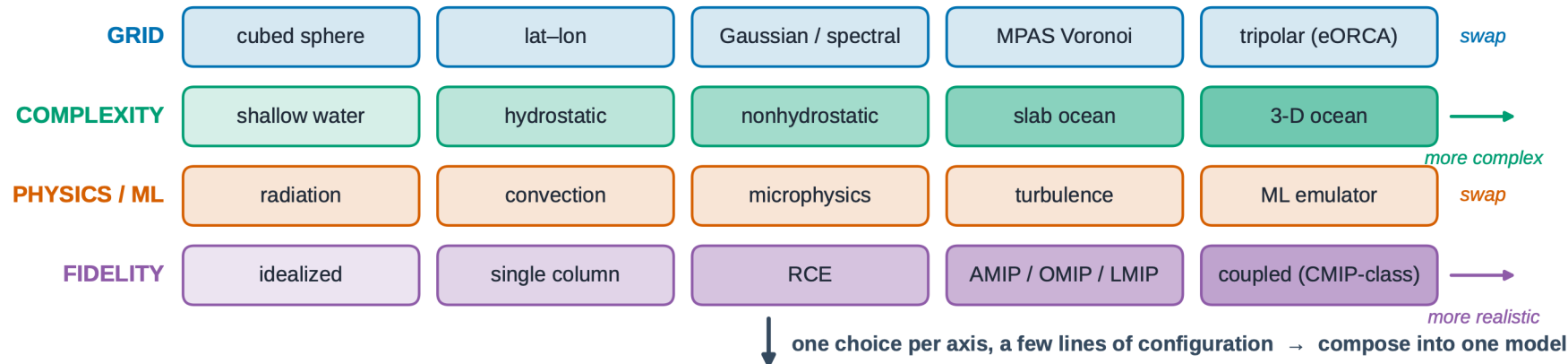
high-resolution simulations supply training targets for ML components, which then run in the global model



legoESM — differentiable JAX ESM



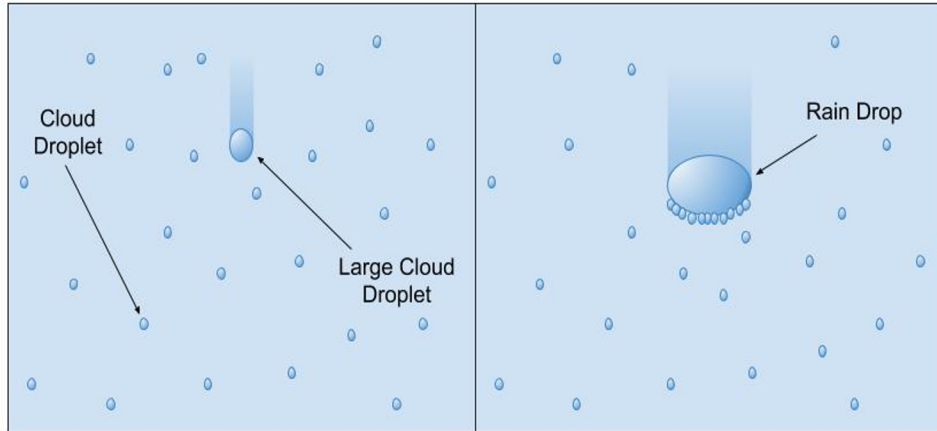
b Composable along four independent axes



~ 10^5 lines of code
< 6 mo. development time

P. Gentine, D. Balwada, A. Paçal, L. Hawkins, A. Adcroft, J. Fang, A. Lahlou, K. Lamb, D. Lawrence, J. Le Sommer, J. Mouallem, D. Watson-Parris, V. Eyring.
“legoESM: a modular, differentiable, multiscale, AI-ready Earth system model built with AI agents”, submitted.

Collision-coalescence as a test case for agentic parameterization development

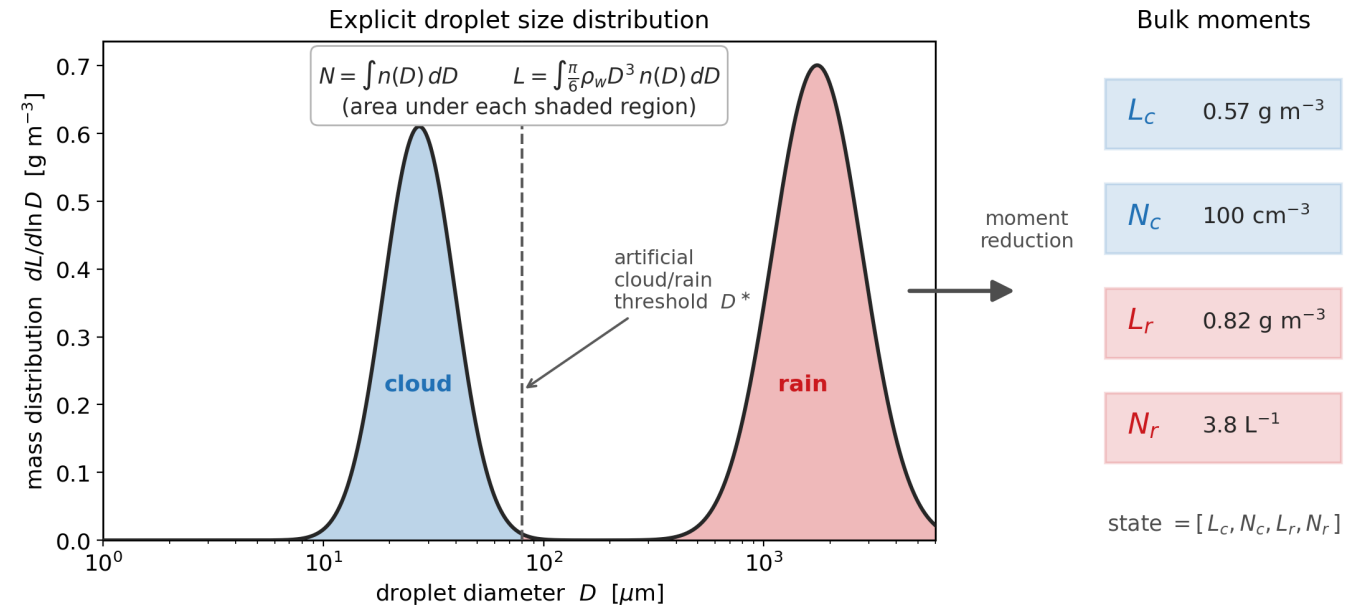


- 50+ year lineage of published schemes of increasing complexity
- High fidelity ground truth (superdroplet/bin) can be used to evaluate simplified bulk parameterizations used in ESMs
- Known limitations in current parameterizations (cloud/rain threshold)

Typical bulk process rates: coupled ODE's that are power laws of cloud and rain moments [L_c , N_c , L_r , N_r]

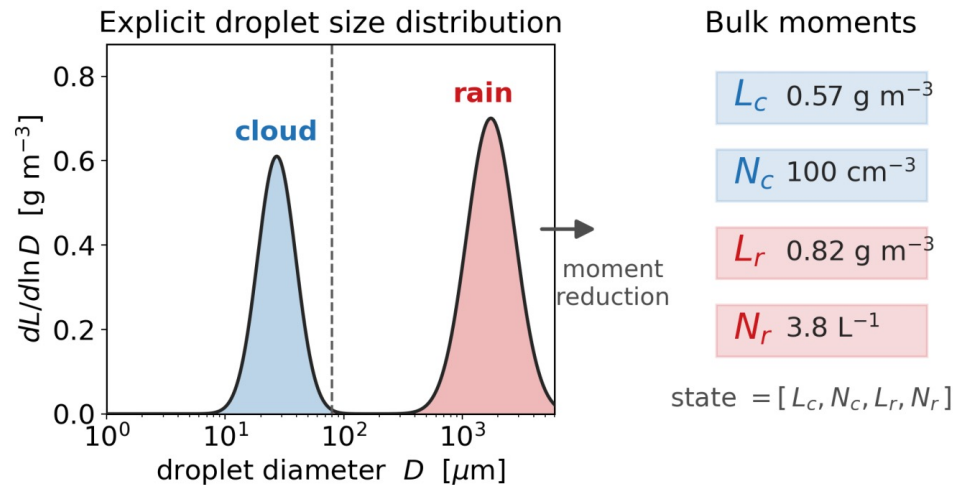
Explore:

- Human-expert-in-the-loop development
- Single shot and iterative API calls to an LLM
- Fully automated agentic workflow

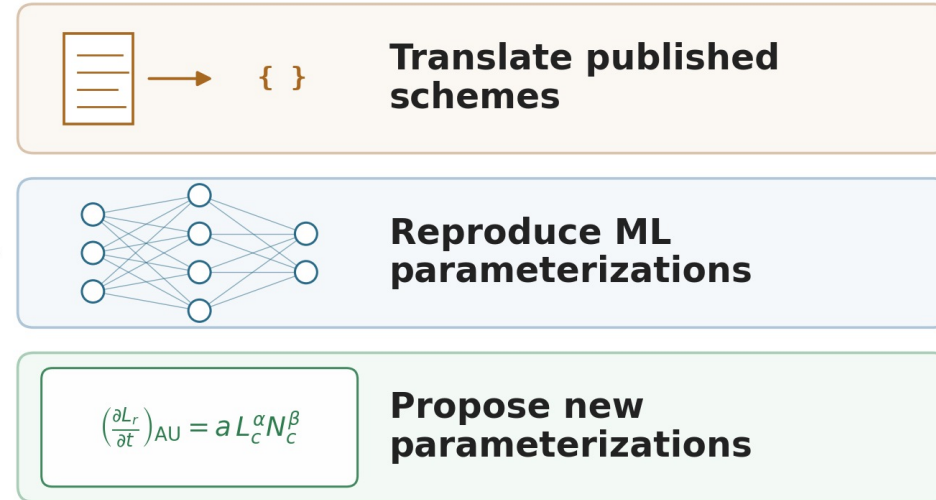


Common evaluation framework and reference high fidelity simulations

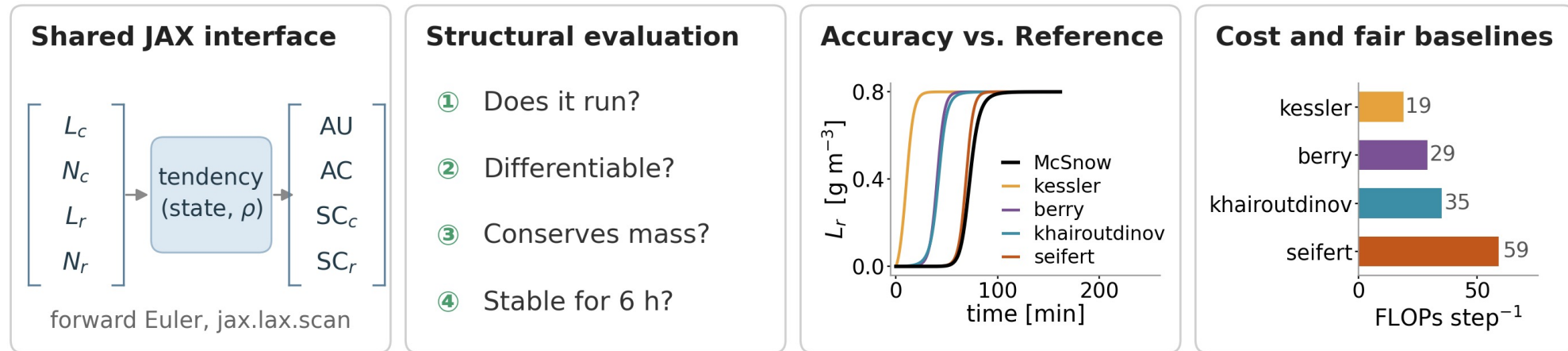
a) Reference and bulk state



b) Experiments



c) Common evaluation framework



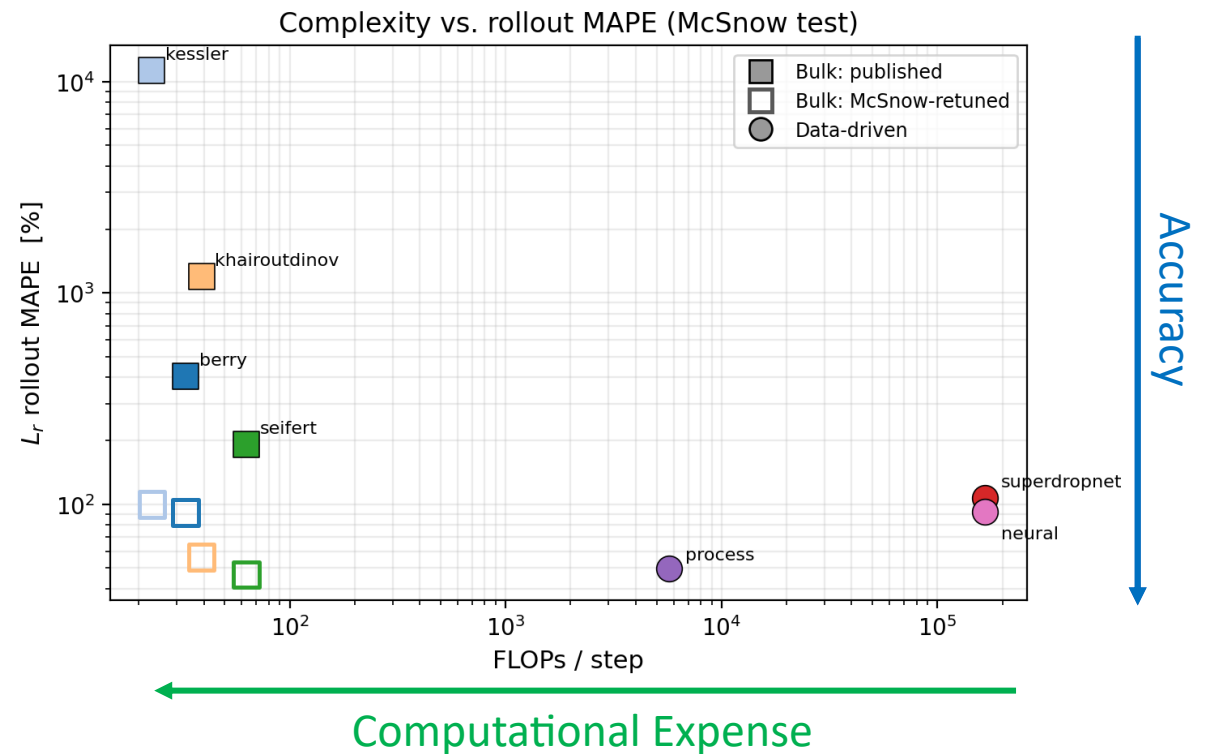
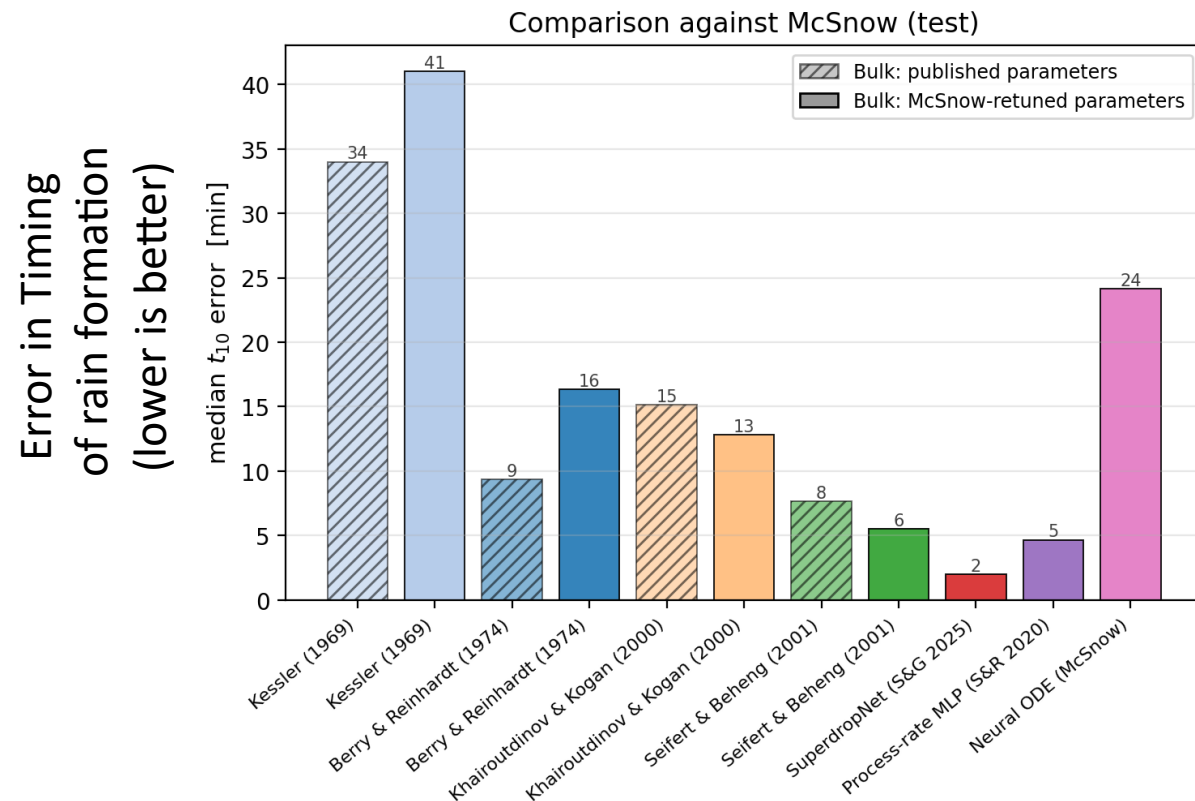
K. Lamb, "Can LLM Agents Improve Imperfect Parameterizations in ESMs?", Accepted, *NeurIPS Sim2Sci Workshop*, 2026

Expert-in-the-loop JAX based development of parameterizations

✓ JAX-based implementation
tests of historic parameterizations

✓ Reconstruction of data-driven
parameterizations trained on high-fidelity
box model simulations, based on existing
ML baselines

⚠ Develop novel scientific machine
learning approaches – applying neural
ODEs for stable, physically consistent
online process rates

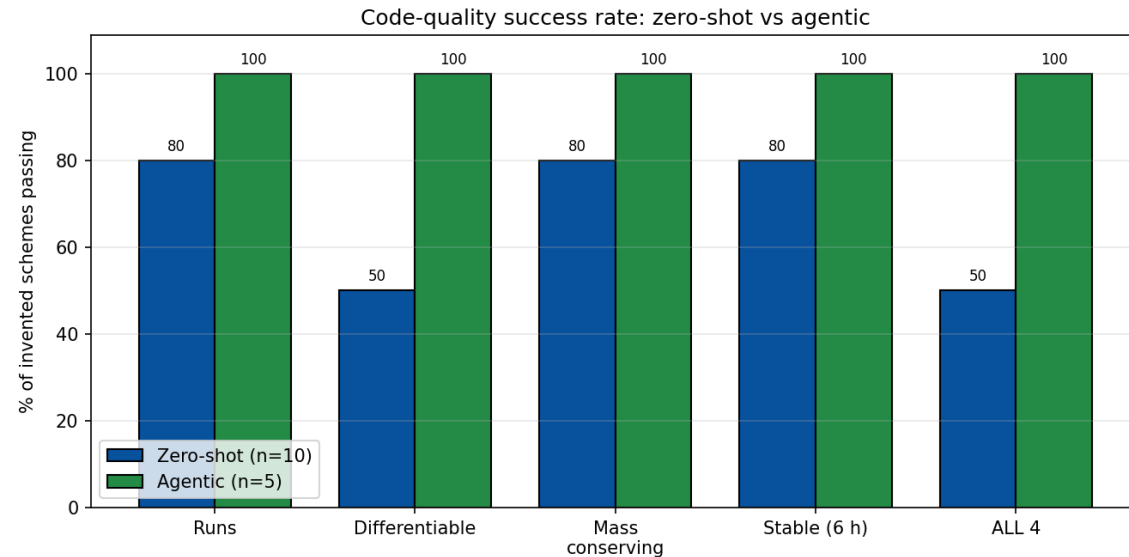


K. Lamb, "Can LLM Agents Improve Imperfect Parameterizations in ESMs?", Accepted, *NeurIPS Sim2Sci Workshop*, 2026

Fully automated development of parameterizations

Construct new parameterizations given high fidelity reference simulations and iterative feedback

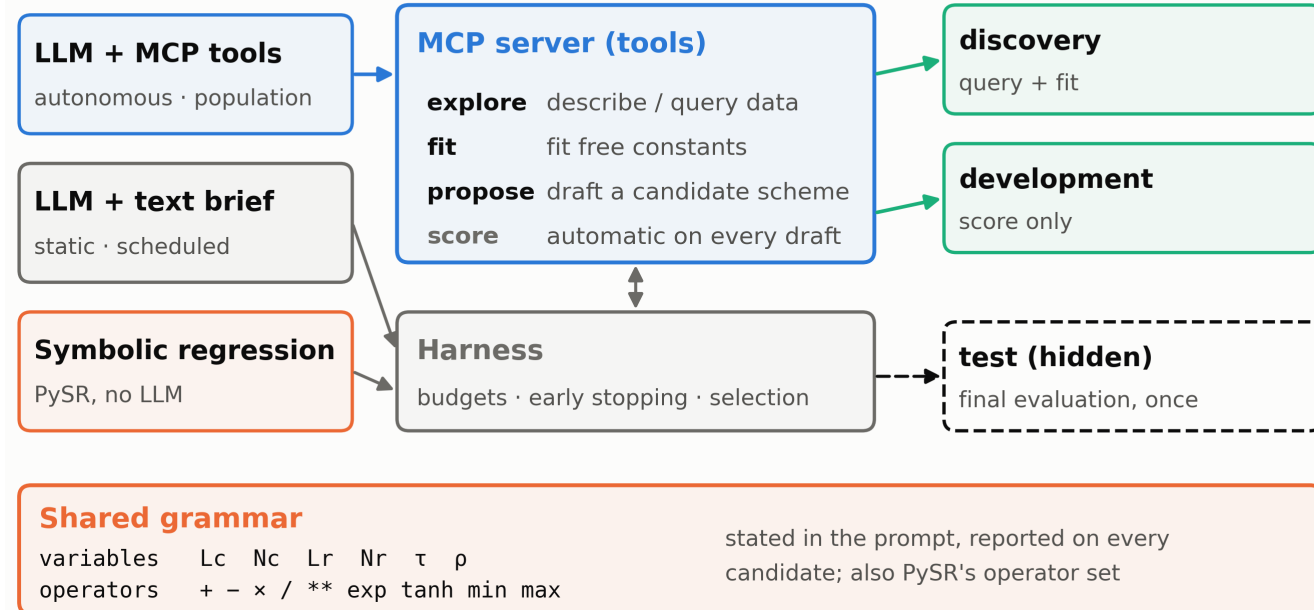
- Evaluate typical failure modes for LLM development



MCP (Model Context Protocol) server where agents can interact with tools, fit data, run simulations & systematically improve

LLM-driven parameterization search via MCP

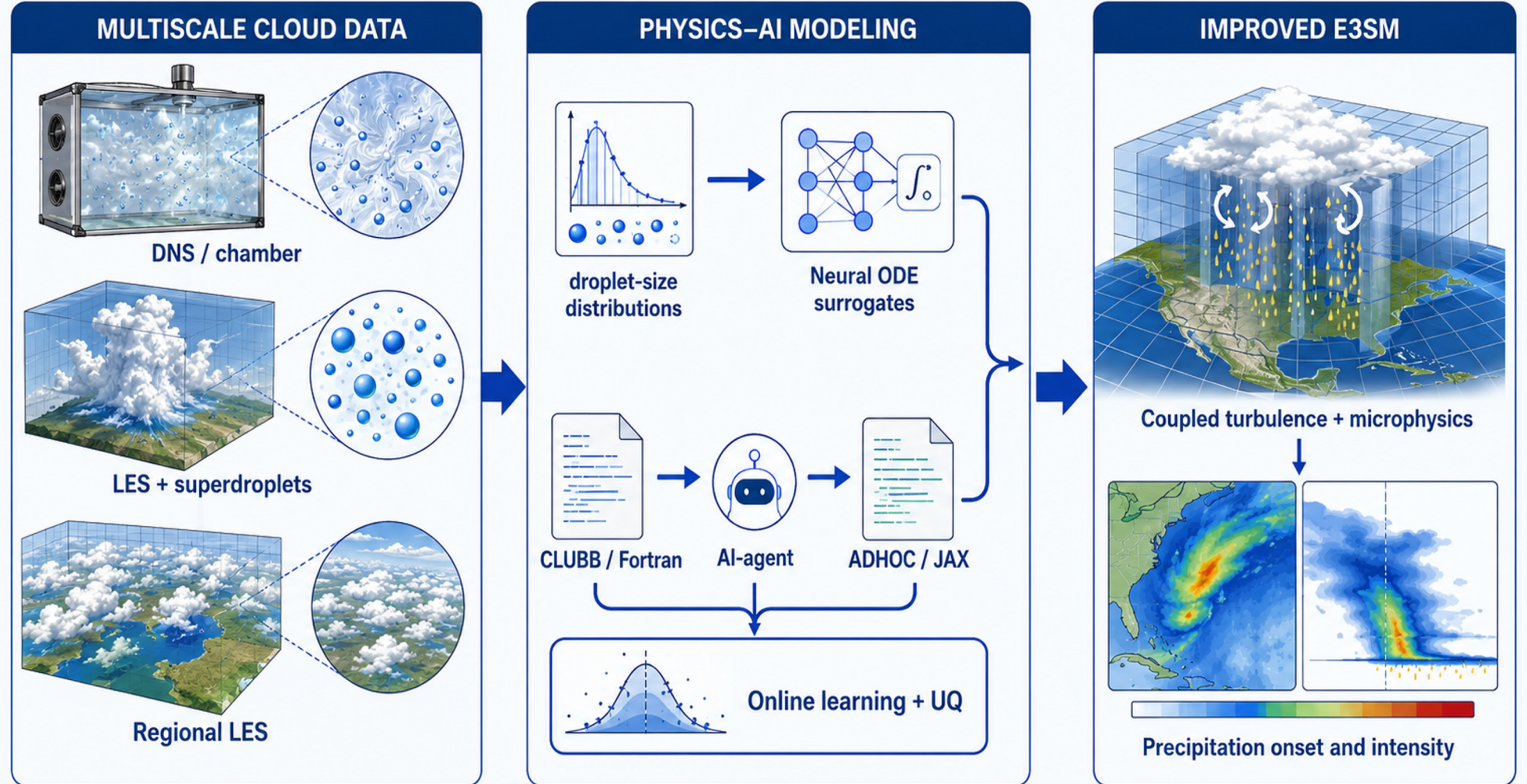
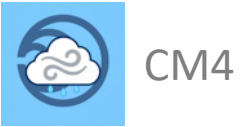
Search arms



K. Lamb, “Can LLM Agents Improve Imperfect Parameterizations in ESMs?”, Accepted, *NeurIPS Sim2Sci Workshop*, 2026

K. Lamb, “Automating Parameterization Development for Earth System Models with Agentic AI”, in prep.

Next: Understand how turbulence impacts collision-coalescence

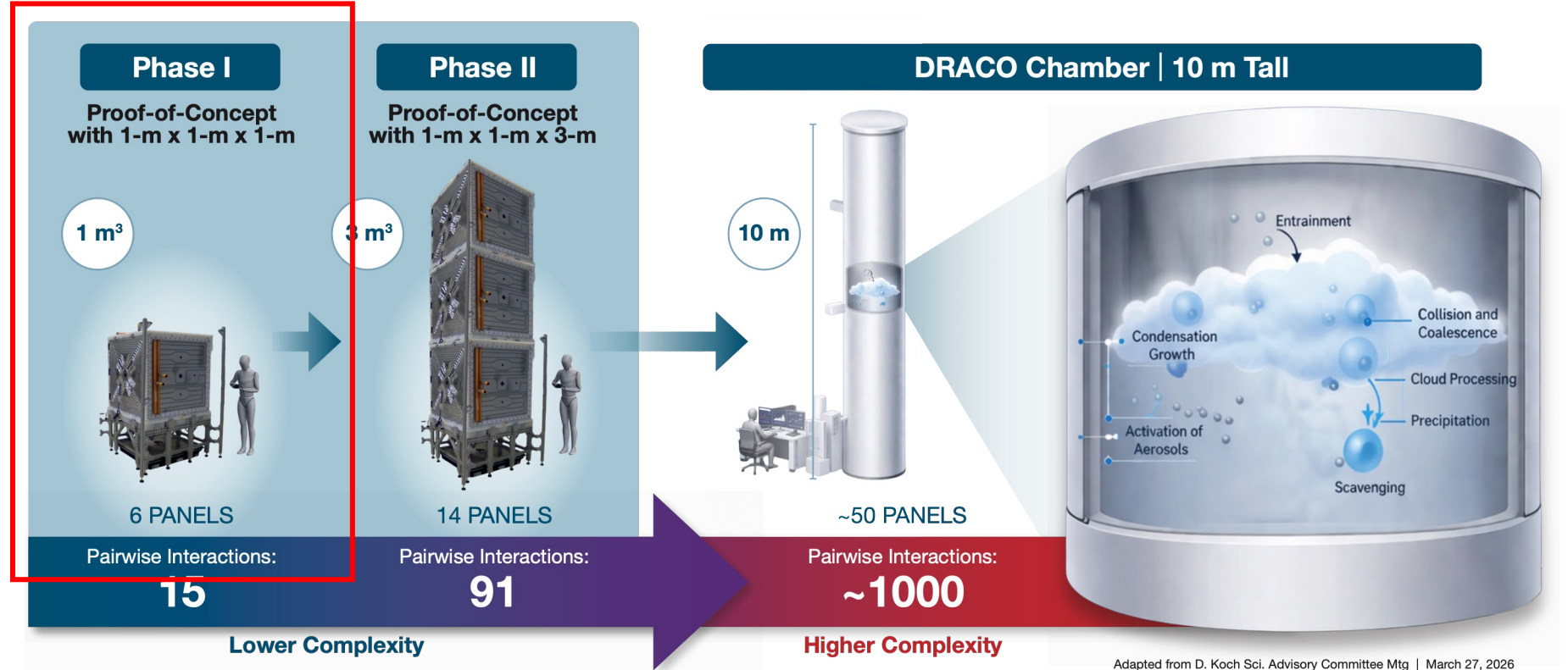


Next: directly integrate laboratory observations into ESM development cycle



Genesis Mission

Cutting-edge AI-enabled laboratory experiments in microphysics and turbulence



- Large-Scale cloud chambers demand AI-driven control of high-dimensional chamber dynamics & coupled cloud processes.
- AI-enabled cloud chamber experiments can accelerate progress in closing critical gaps in predicting precipitation formation.



Brookhaven
National Laboratory



Stony Brook
University

Toward more adaptive & accessible ESM development

Much faster iteration

years → weeks/months

Reproducible & systematic

consistent, auditable evaluation

Multi-modal

Better connect observations and models



Outstanding Challenges :

Creativity

Can AI generate truly novel scientific discoveries?

"LLM's Can't Jump"

Scientific validity

avoid benchmark overfitting; discovery vs. curve-fitting

Trust & reliability

hallucinated or subtly wrong code; physically invalid yet 'successful' fits