Interpretable machine learning for high-speed, high-fidelity GEOS-Chem model simulations with uncertainty quantification

PI: **Christopher Tessum**, Assistant Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

Collaborators:

Daven Henze, Associate Professor, Mechanical Engineering, University of Colorado Boulder

Christopher Rackauckas: Instructor, Applied Mathematics, MIT; Sr. Research Analyst, Pharmacy, U. Maryland

Nicole Riemer, Professor and Associate Head, Atmospheric Sciences, University of Illinois at Urbana-Champaign

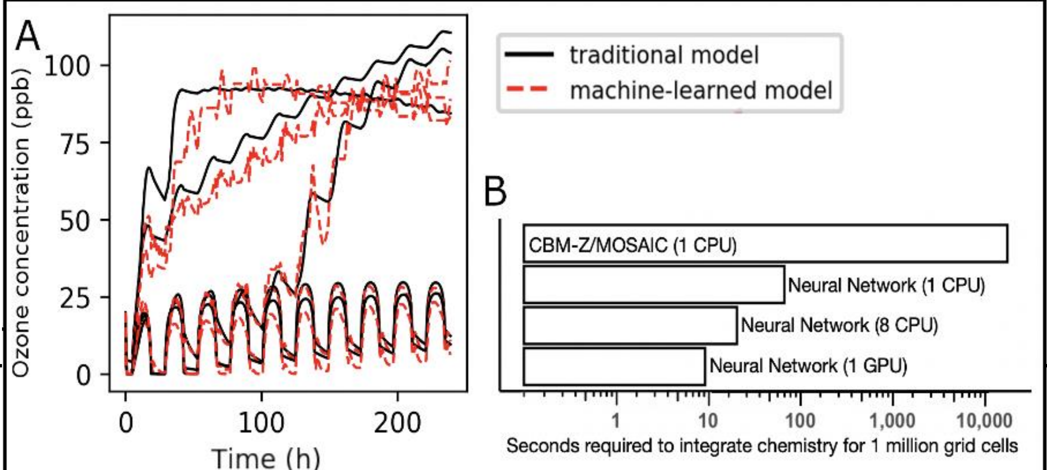
Research objectives

1. Implement machine-learned advection and chemistry operators into the GEOS-Chem model.
2. Create and evaluate probabilistic versions of the same learned operators that allow uncertainty quantification.
3. Fine tune overall the overall model by using online training during full model simulations to minimize prediction error against observational data.

Comparison to SOA: We expect ~100× speedup compared to current GEOS-Chem. Existing models cannot make probabilistic predictions or adjust chemistry parameters to improve agreement with observations.

Innovation: Project represents a paradigm shift from manual to automated model reduction for full-physics models.

TRL (varies by task): Start: 1–2; End: 3–4: **Operational testing by end of project**



We will build on results of preliminary work which has shown that by building physical constraints into machine-learning models we can create high-fidelity surrogate models (A) that run orders-of-magnitude faster than the reference model (B).

Approach

Task 1: Create machine-learned transport interpolation schemes using neural differential equations with spatial and temporal coarse-graining

Task 2: Rewrite part of GEOS-Chem mediation layer in the Julia language

Task 3: Implement the machine learned transport and (created through separate project) chemistry operators in GEOS-Chem and test against observations

Task 4: Fine tune the machine-learned models by minimizing full-model error in predicting observations

Task 5: Create probabilistic versions of the transport and chemistry operators and test against observations

Potential impact: The project will advance the state of knowledge on how machine learning techniques that leverage mechanistic model structure can be used in operational settings rather than just in small-scale experiments. This has the potential to expand the Pareto-frontier between accuracy and computational cost in geophysical modeling, resulting both in more accurate "digital twins" of Earth as well as in computationally tractable simulations for decision-support analysis.