Equation-Based Geoscientific Modeling and Surrogate Modeling

Christopher Tessum

University of Illinois Urbana-Champaign

JuliaEO Workshop

1/5/2026

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

EarthSciML

What is Geoscientific Modeling?

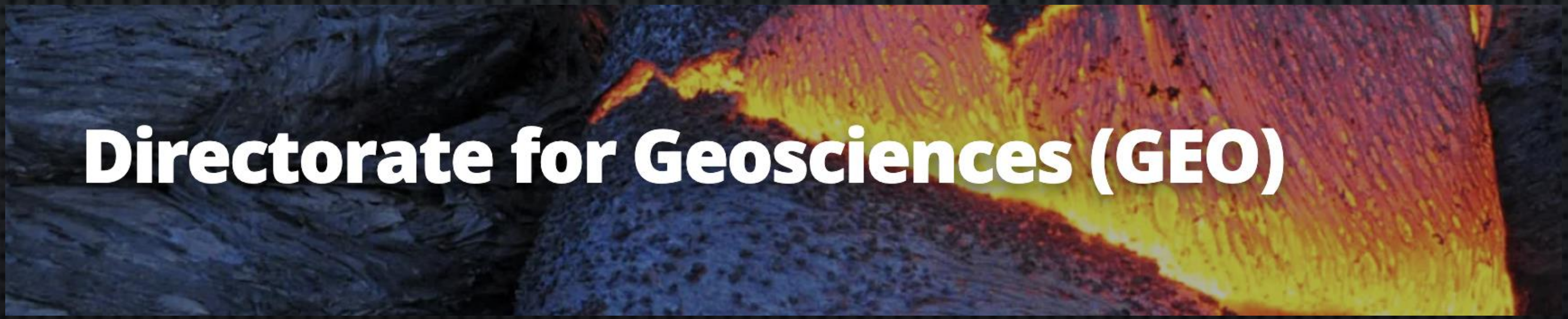
- Creation/use of numerical models of the Earth system and its components

AGU ADVANCING EARTH AND SPACE SCIENCES

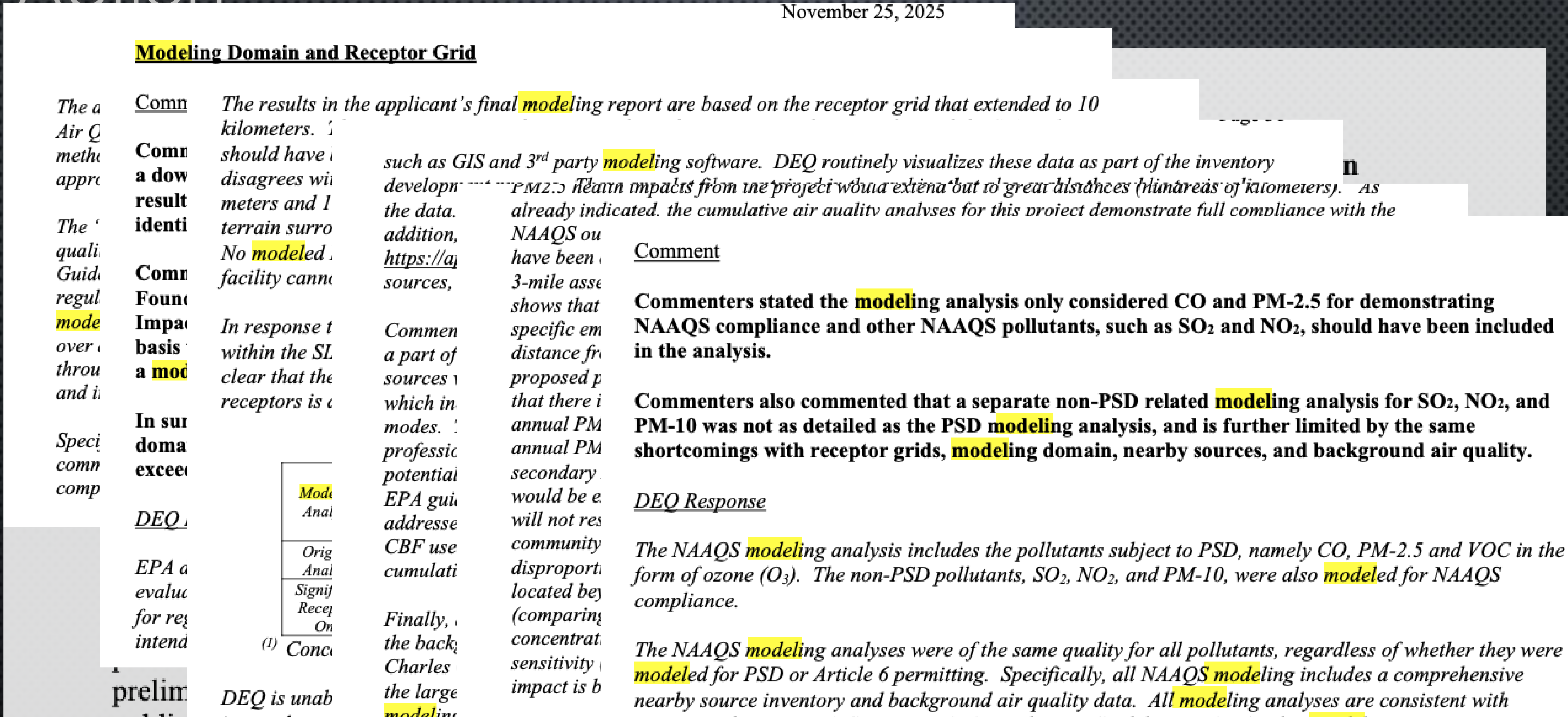
EGU European Geosciences Union

Geoscientific Model Development

ARTICLES & PREPRINTS SUBMISSION POLICIES PEER-REVIEW PROCESS EDITORIAL BOARD ABOUT EGU PUBLICATIONS



Air Quality Modeling: A Front Line for Climate Action



November 25, 2025

Modeling Domain and Receptor Grid

The results in the applicant's final modeling report are based on the receptor grid that extended to 10 kilometers.

such as GIS and 3rd party modeling software. DEQ routinely visualizes these data as part of the inventory

PM2.5 health impacts from the project would extend out to great distances (hundreds of kilometers). As already indicated, the cumulative air quality analyses for this project demonstrate full compliance with the

Comment

Commenters stated the modeling analysis only considered CO and PM-2.5 for demonstrating NAAQS compliance and other NAAQS pollutants, such as SO_2 and NO_2, should have been included in the analysis.

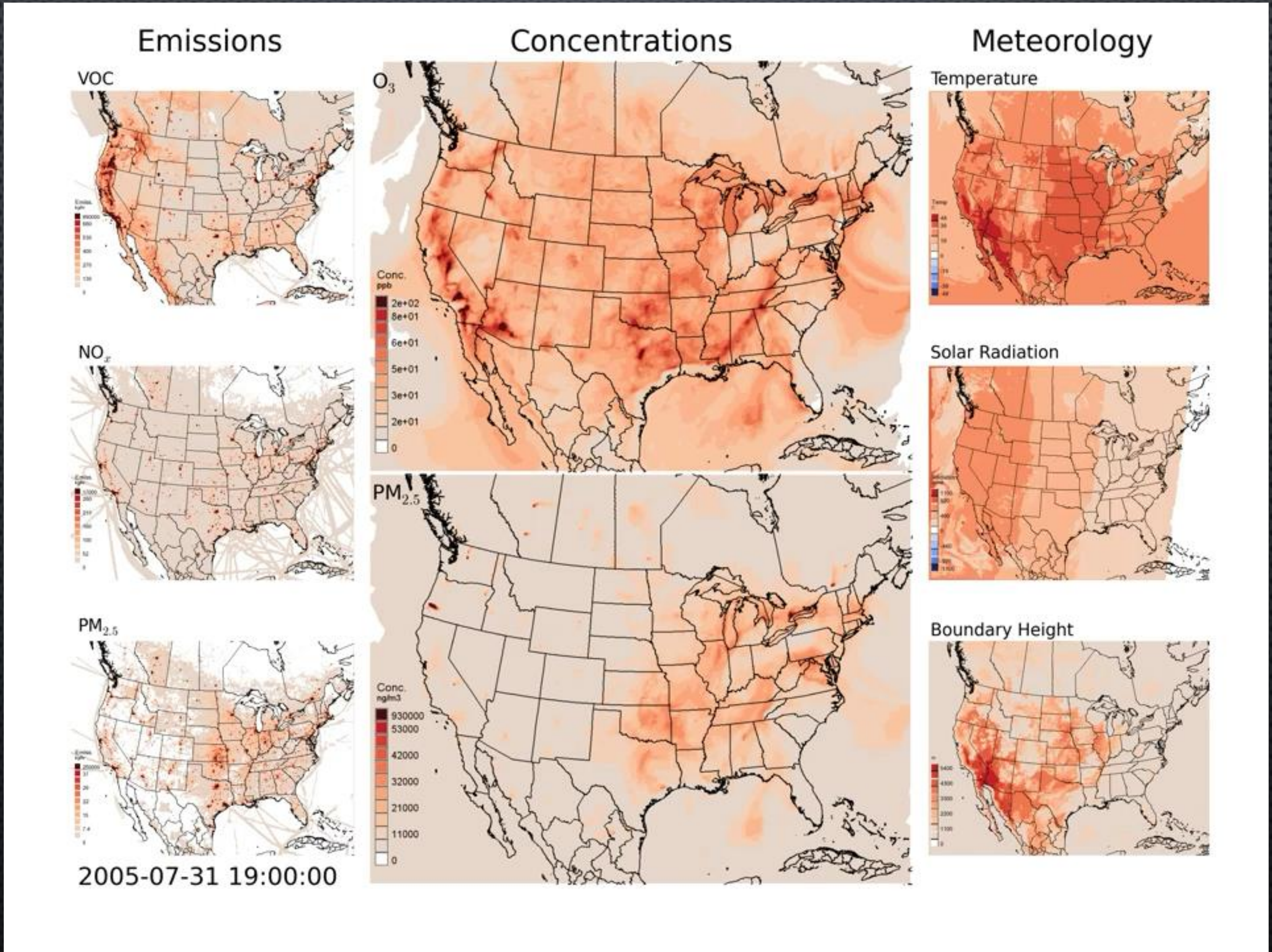
Commenters also commented that a separate non-PSD related modeling analysis for SO_2, NO_2, and PM-10 was not as detailed as the PSD modeling analysis, and is further limited by the same shortcomings with receptor grids, modeling domain, nearby sources, and background air quality.

DEQ Response

The NAAQS modeling analysis includes the pollutants subject to PSD, namely CO, PM-2.5 and VOC in the form of ozone (O_3). The non-PSD pollutants, SO_2, NO_2, and PM-10, were also modeled for NAAQS compliance.

The NAAQS modeling analyses were of the same quality for all pollutants, regardless of whether they were modeled for PSD or Article 6 permitting. Specifically, all NAAQS modeling includes a comprehensive nearby source inventory and background air quality data. All modeling analyses are consistent with

Air Quality Modeling (what it can be)



Air Quality Modeling (what it often is)

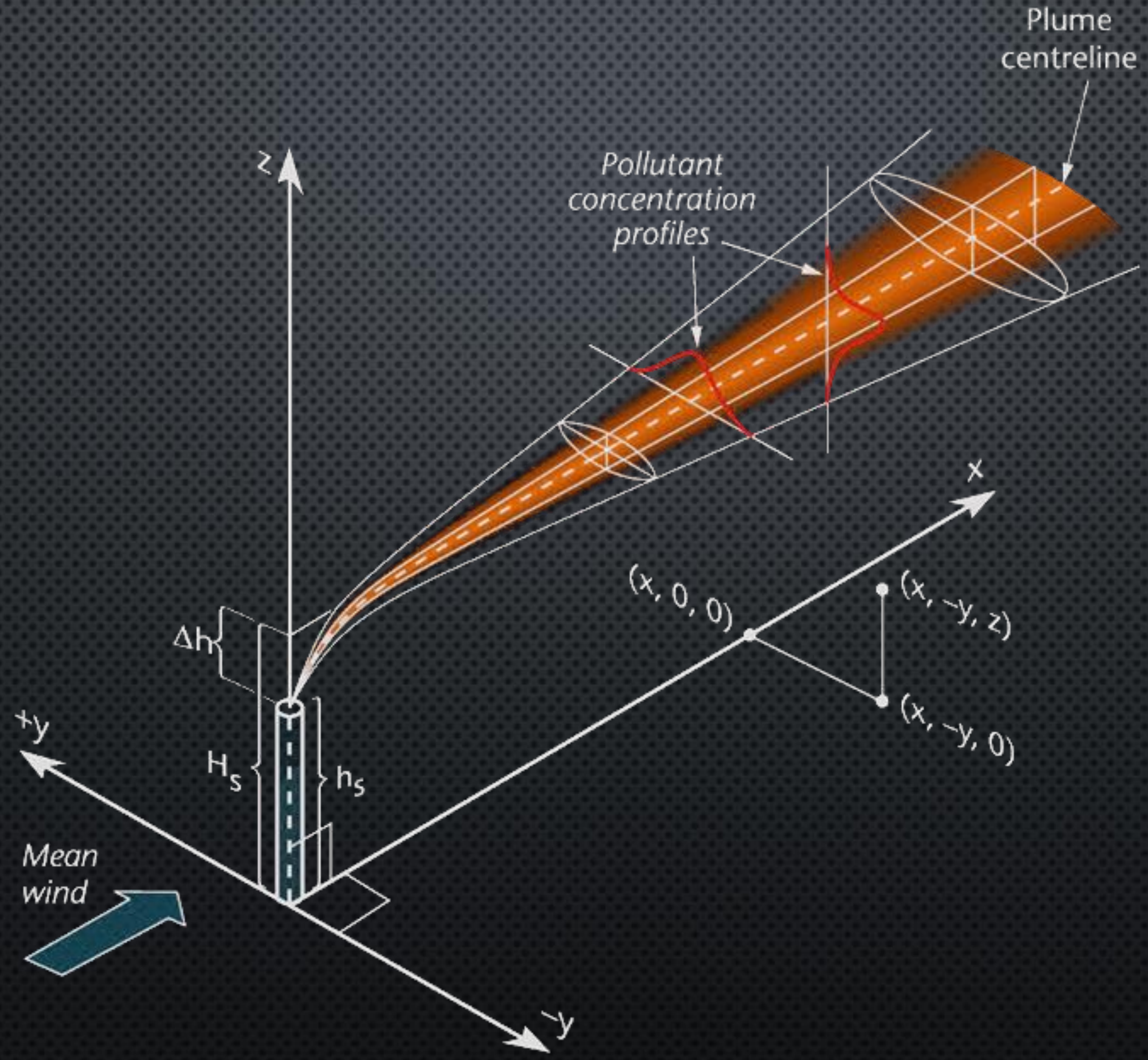


Image credit: epalife.com

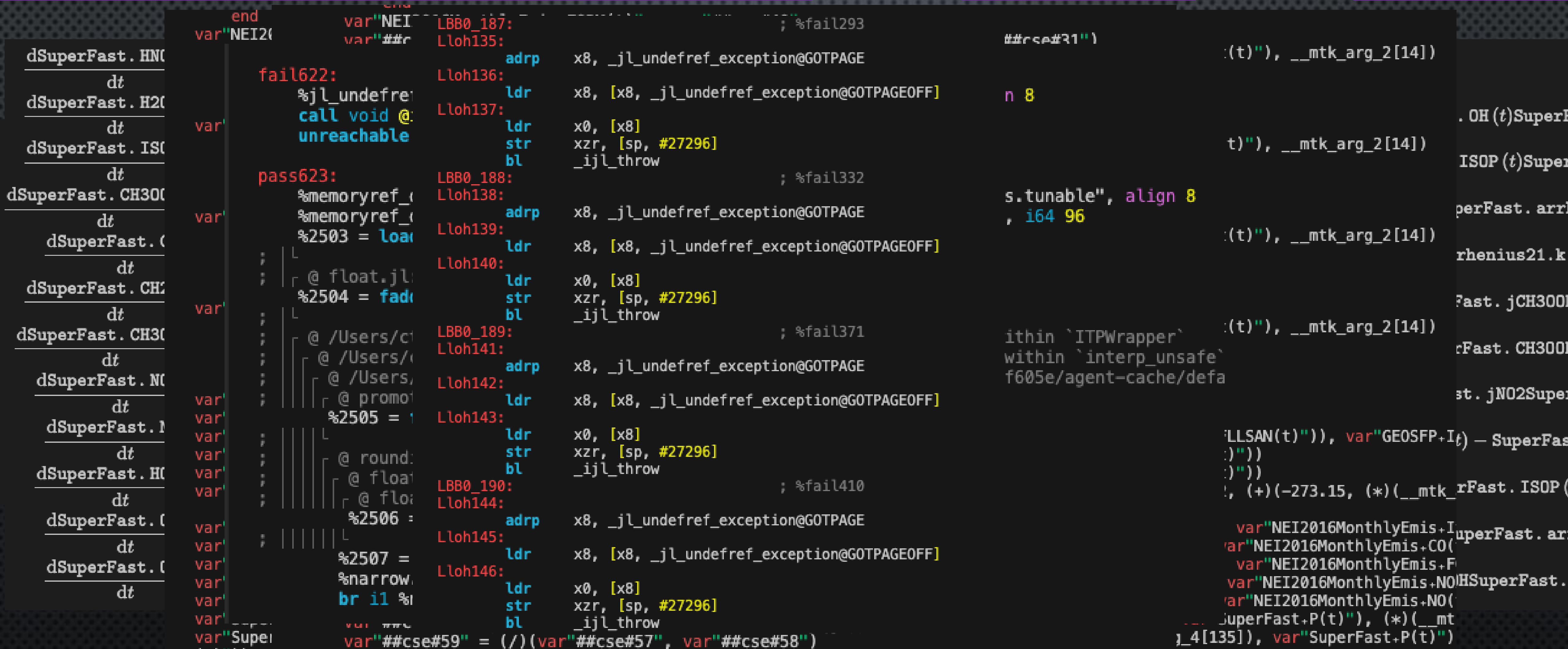
Geoscientists tend to understand this:

$$
\begin{aligned}
\mathtt{ISOP} + \mathtt{OH} &\xrightarrow{\mathtt{arrhenius18.k}(t)} \mathtt{2CH3O2} \\
\mathtt{ISOP} + \mathtt{OH} &\xrightarrow{\mathtt{arrhenius19.k}(t)} \mathtt{ISOP} \\
\mathtt{ISOP} + \mathtt{OH} &\xrightarrow{\mathtt{arrhenius20.k}(t)} \mathtt{ISOP} + \mathtt{0.5OH} \\
\mathtt{ISOP} + \mathtt{O3} &\xrightarrow{\mathtt{arrhenius21.k}(t)} \mathtt{0.87CH2O} + \mathtt{1.86CH3O2} + \mathtt{0.06HO2} + \mathtt{0.05CO} \\
\mathtt{O3} &\xrightarrow{\mathtt{jO32OH}} \mathtt{2OH} \\
\mathtt{H2O2} &\xrightarrow{\mathtt{jH2O2}} \mathtt{2OH} \\
\mathtt{NO2} &\xrightarrow{\mathtt{jNO2}} \mathtt{NO} + \mathtt{O3} \\
\mathtt{CH2O} &\xrightarrow{\mathtt{jH2COa}} \mathtt{2HO2} + \mathtt{CO} \\
\mathtt{CH2O} &\xrightarrow{\mathtt{jH2COb}} \mathtt{CO} \\
\mathtt{CH3OOH} &\xrightarrow{\mathtt{jCH3OOH}} \mathtt{CH2O} + \mathtt{HO2} + \mathtt{OH} \\
\mathtt{arrhenius1.k}\left(t\right) &= \frac{\left(\frac{\mathtt{arrhenius1.K\_300}}{T}\right)^{\mathtt{arrhenius1.b0}} P \mathtt{arrhenius1.A} \mathtt{arrhenius1.a0} \mathtt{arrhenius1.ppb\_unit} e^{\frac{\mathtt{arrhenius1.c0}}{T}}}{T \mathtt{arrhenius1.R}} \\
\mathtt{arrhenius2.k}\left(t\right) &= \frac{\left(\frac{\mathtt{arrhenius2.K\_300}}{T}\right)^{\mathtt{arrhenius2.b0}} P \mathtt{arrhenius2.A} \mathtt{arrhenius2.a0} \mathtt{arrhenius2.ppb\_unit} e^{\frac{\mathtt{arrhenius2.c0}}{T}}}{T \mathtt{arrhenius2.R}}
\end{aligned}
$$

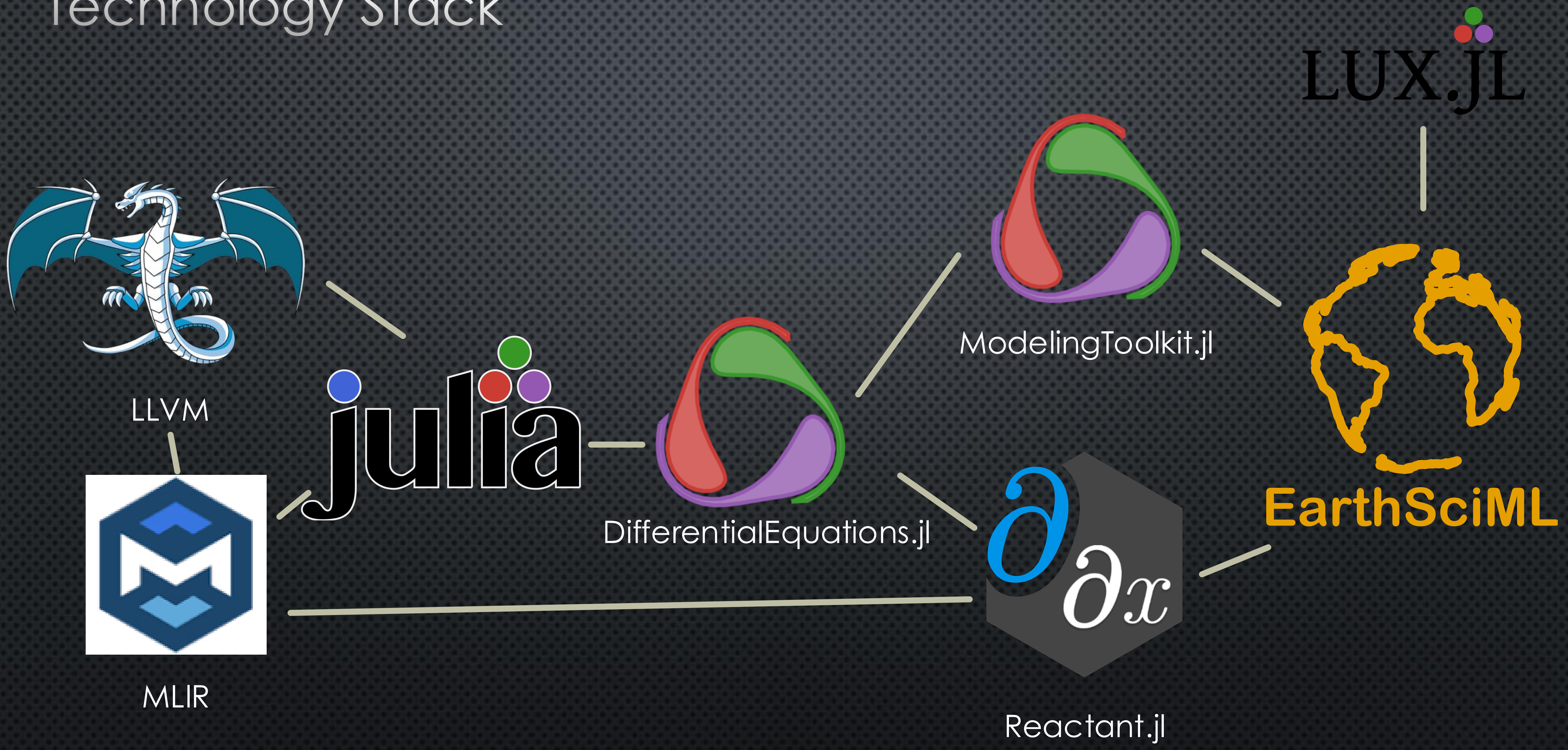
...better than this:

```
#> Generate Library Locations
 setenv CMAQ_LIB    ${lib_basedir}/${system}/${compilerString}
 setenv MPI_DIR     $CMAQ_LIB/mpi
 setenv NETCDF_DIR  $CMAQ_LIB/netcdf
 setenv NETCDFF_DIR $CMAQ_LIB/netcdff
 setenv PNETCDF_DIR $CMAQ_LIB/pnetcdf
 setenv IOAPI_DIR   $CMAQ_LIB/ioapi

#> Create Symbolic Links to Libraries
 if ( ! -d $CMAQ_LIB ) mkdir -p $CMAQ_LIB
 if (   -e $MPI_DIR  ) rm -rf $MPI_DIR
     mkdir $MPI_DIR
     ln -s $MPI_LIB_DIR $MPI_DIR/lib
     ln -s $MPI_INCL_DIR $MPI_DIR/include
 if ( ! -d $NETCDF_DIR )  mkdir $NETCDF_DIR
 if ( ! -e $NETCDF_DIR/lib ) ln -sfn $NETCDF_LIB_DIR $NETCDF_DIR/lib
 if ( ! -e $NETCDF_DIR/include ) ln -sfn $NETCDF_INCL_DIR $NETCDF_DIR/include
 if ( ! -d $NETCDFF_DIR )  mkdir $NETCDFF_DIR
 if ( ! -e $NETCDFF_DIR/lib ) ln -sfn $NETCDFF_LIB_DIR $NETCDFF_DIR/lib
 if ( ! -e $NETCDFF_DIR/include ) ln -sfn $NETCDFF_INCL_DIR $NETCDFF_DIR/include
 if ( ! -d $IOAPI_DIR ) then
    mkdir $IOAPI_DIR
    ln -sfn $IOAPI_INCL_DIR $IOAPI_DIR/include_files
    ln -sfn $IOAPI_LIB_DIR  $IOAPI_DIR/lib
 endif
```

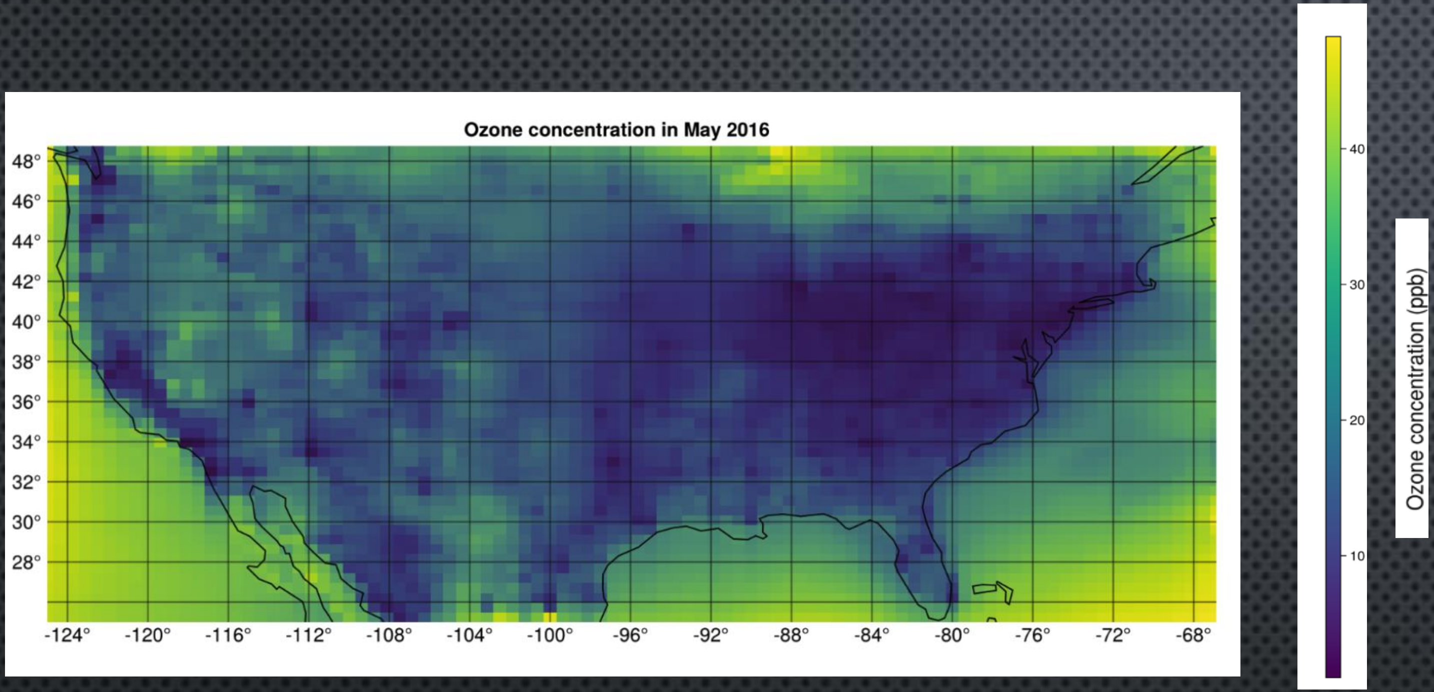
Equation-Based Model Development

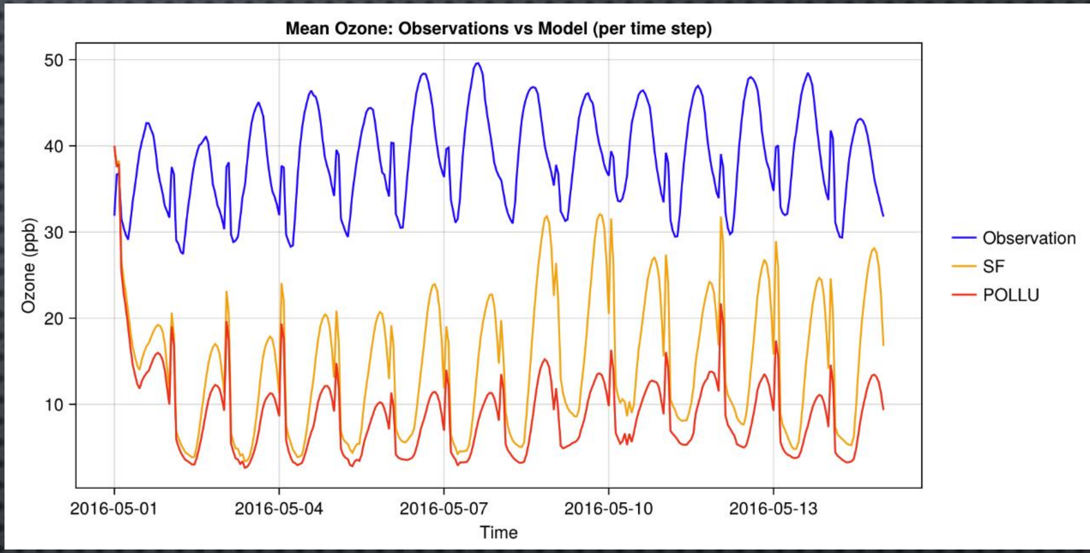


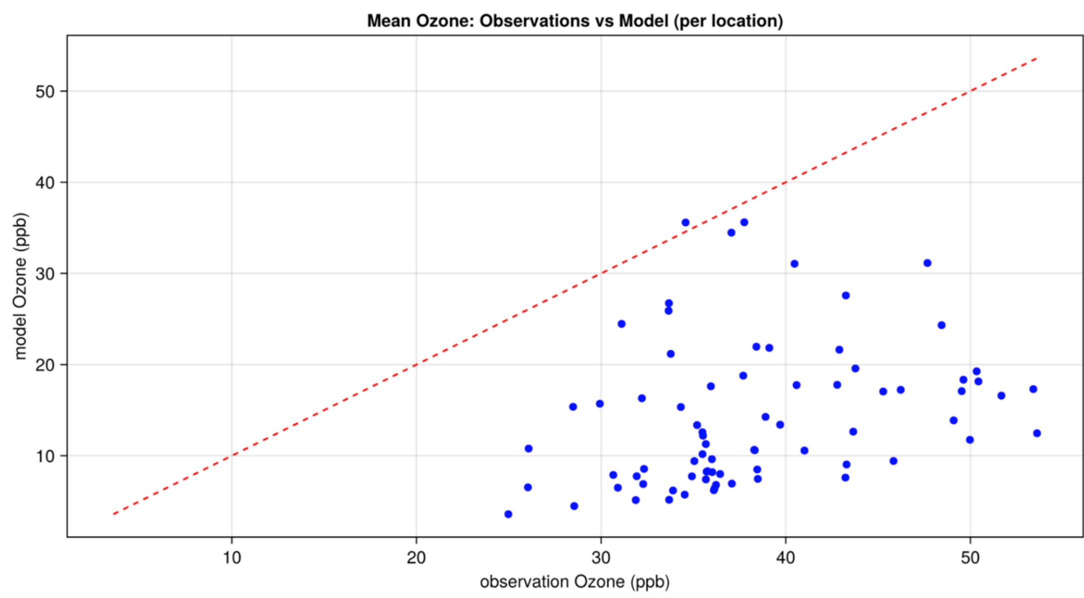
Technology Stack



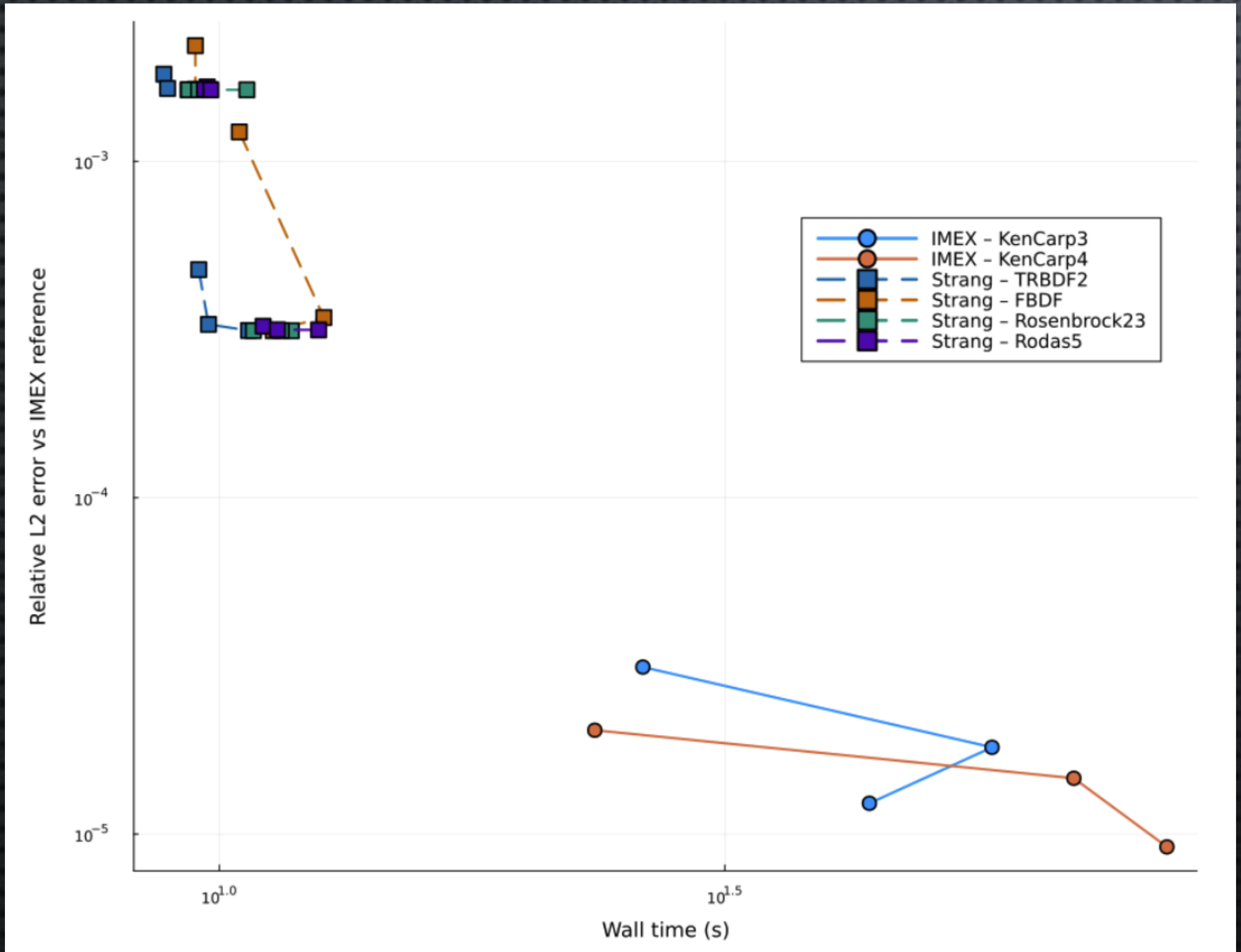
Equation-Based Model Development

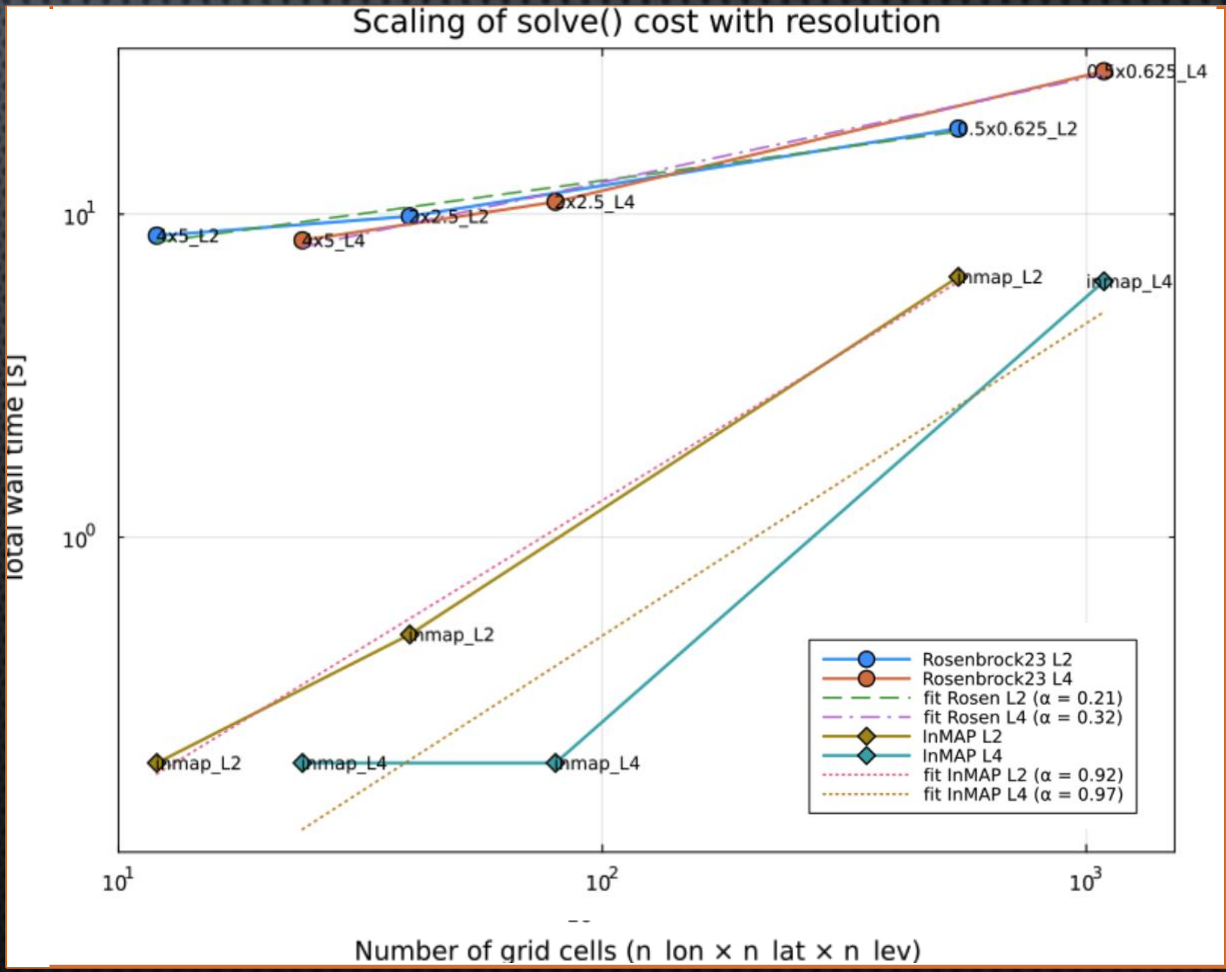






Computational Performance





Scenario Analysis

30% reduction in emissions from the "on-road" sector

- Base model, shared between the 2 scenarios

```julia
using EarthSciMLBase, GasChem, AtmosphericDeposition, EarthSciData
using EnvironmentalTransport, ModelingToolkit, OrdinaryDiffEq
using DiffEqCallbacks
using Dates, Plots, NCDatasets
using ProgressLogging # Needed for progress bar. Use `TerminalLoggers` if in a terminal.

domain = DomainInfo(
    DateTime(2016, 5, 1),
    DateTime(2016, 5, 2);
    lonrange = deg2rad(-115):deg2rad(2.5):deg2rad(-68.75),
    latrange = deg2rad(25):deg2rad(2):deg2rad(53.7),
    levrange = 1:15,
    dtype = Float64)

geosfp = GEOSFP("0.5x0.625_NA", domain; stream=false)
emis = NEI2016MonthlyEmis("mrggrid_withbeis_withrwc", domain; stream=false)
dt = 300.0 # Operator splitting timestep

base_model = couple(
    SuperFast(),
    FastJX(),
    DrydepositionG(),
    Wetdeposition(),
    AdvectionOperator(dt, upwind1_stencil, ZeroGradBC()),
    emis,
    geosfp,
    domain
)
```

- Components that will be unique to each scenario

a. Business as usual scenario

```julia
bau_outfile = ("RUNNER_TEMP" ∈ keys(ENV) ? ENV["RUNNER_TEMP"] : tempname()) * "bau_output.nc" #
bau_model = couple(base_model, NetCDFOutputter(bau_outfile, 3600.0))
```

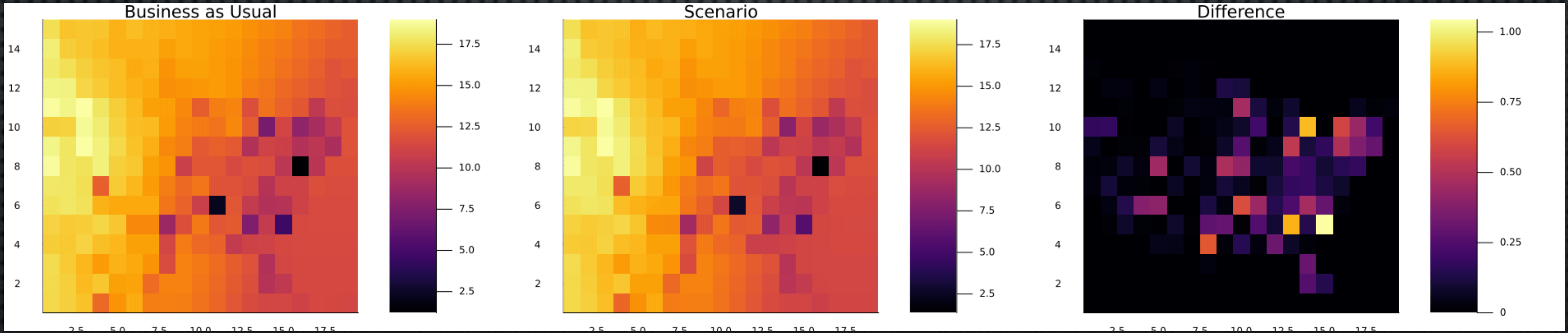
b. 30% reduction in the on-road emissions scenario

Initialize a new emissions component only containing the on-road emissions with a scale factor of -0.3

```julia
@named scenario_emis = NEI2016MonthlyEmis("onroad", domain; scale=-0.3, stream=false)
scenario_emis = EarthSciMLBase.copy_with_change(scenario_emis, discrete_events=[]) # Workaround


scenario_outfile = ("RUNNER_TEMP" ∈ keys(ENV) ? ENV["RUNNER_TEMP"] : tempname()) * "scenario_ou
scenario_model = couple(base_model, scenario_emis, NetCDFOutputter(scenario_outfile, 3600.0))
```

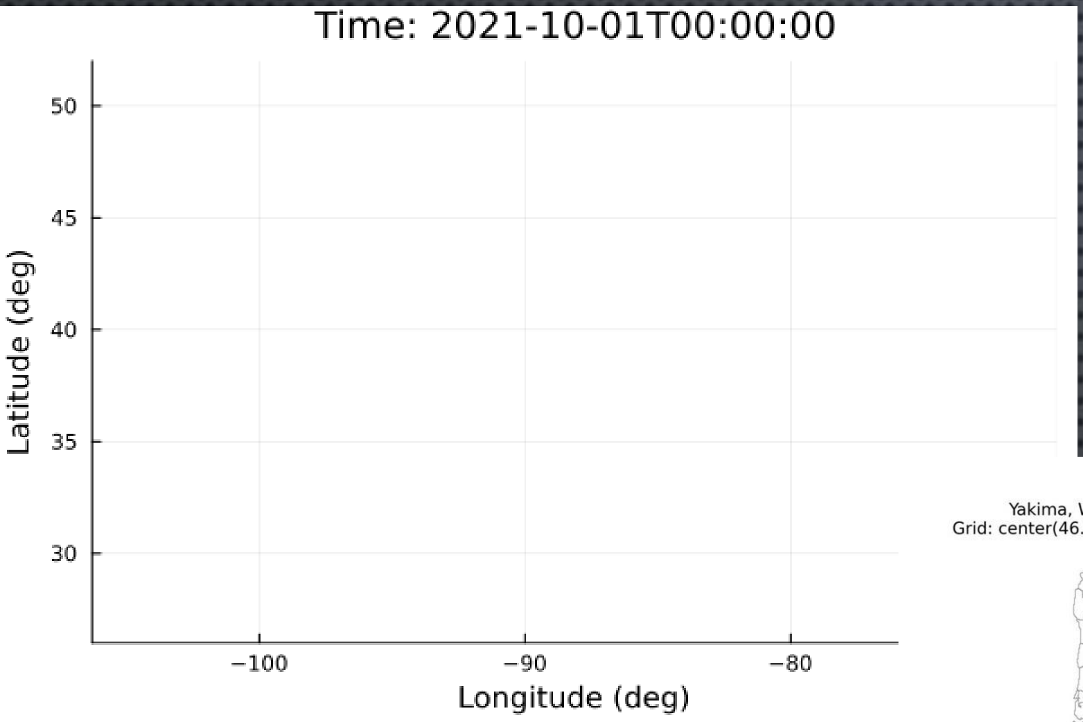
Scenario Analysis

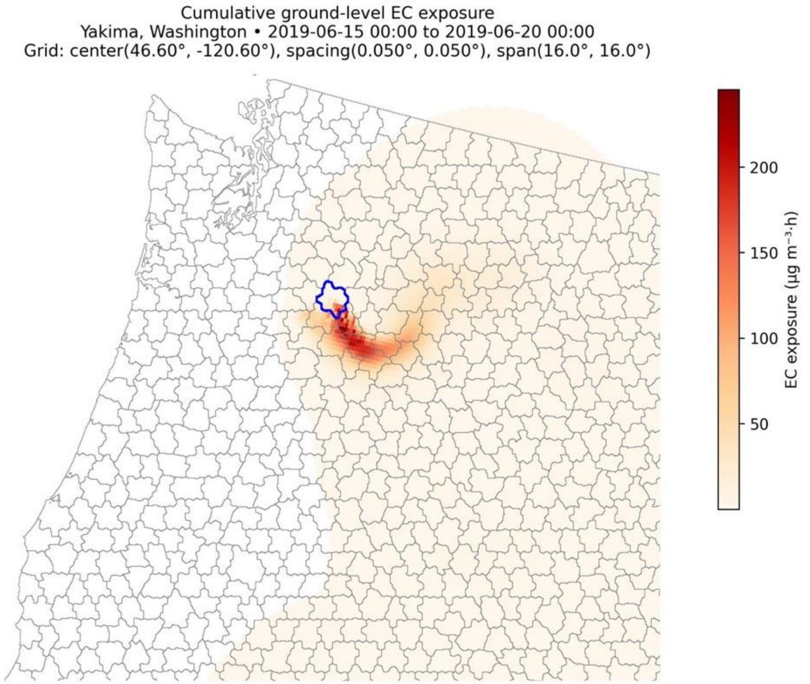
```
st = SolverStrangSerial(Rosenbrock23(), dt, callback=PositiveDomain(save=false))

bau_prob = ODEProblem(bau_model, st)
sol = solve(bau_prob, SSPRK22(); dt=dt, progress=true, progress_steps=1,
    save_on=false, save_start=false, save_end=false, initialize_save=false)


scenario_prob = ODEProblem(scenario_model, st)
sol = solve(scenario_prob, SSPRK22(); dt=dt, progress=true, progress_steps=1,
    save_on=false, save_start=false, save_end=false, initialize_save=false)
```



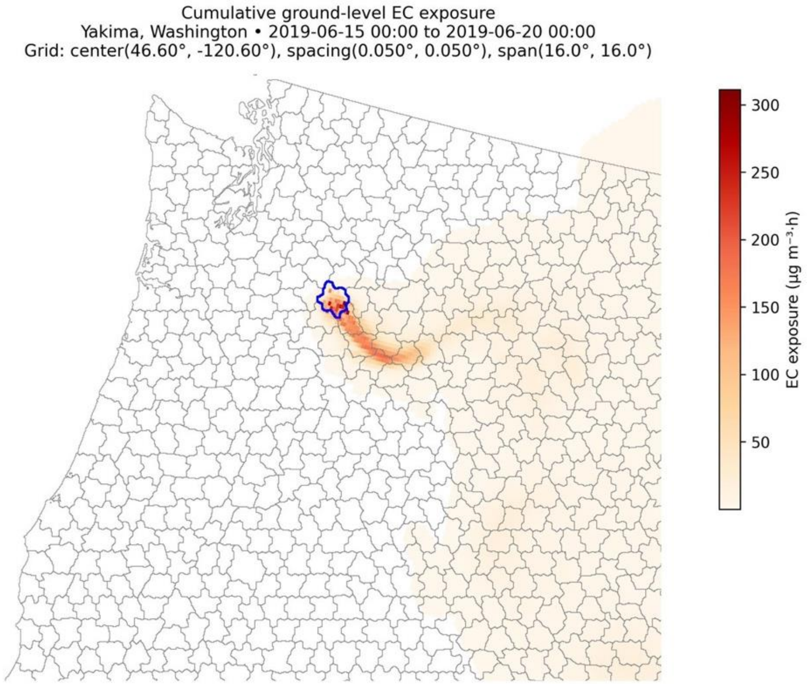
Animation 1. Ground level ozone concentrations at each timestep

Same Equations, Different Model





Puff Model

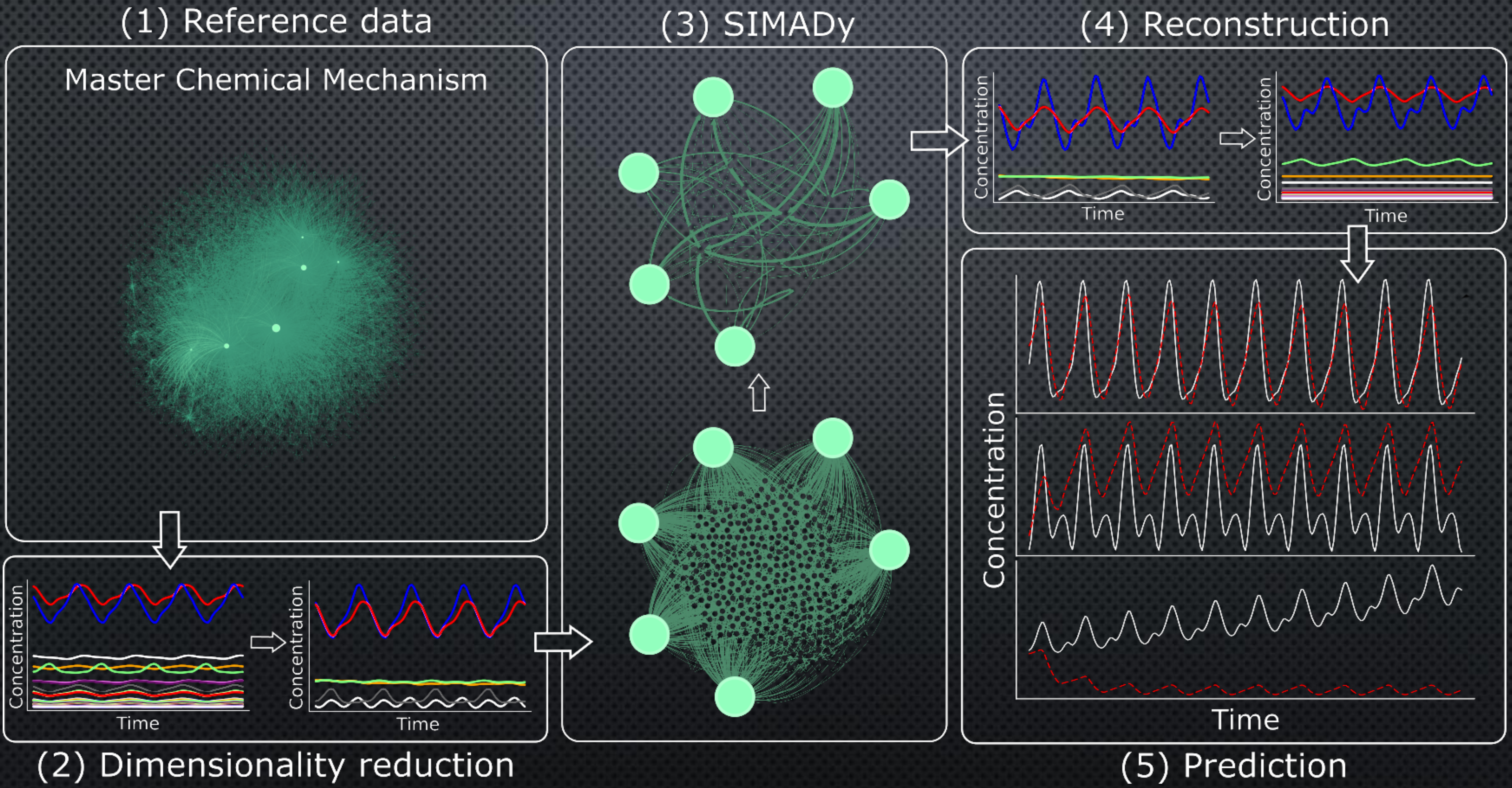


Hysplit

Surrogate Modeling

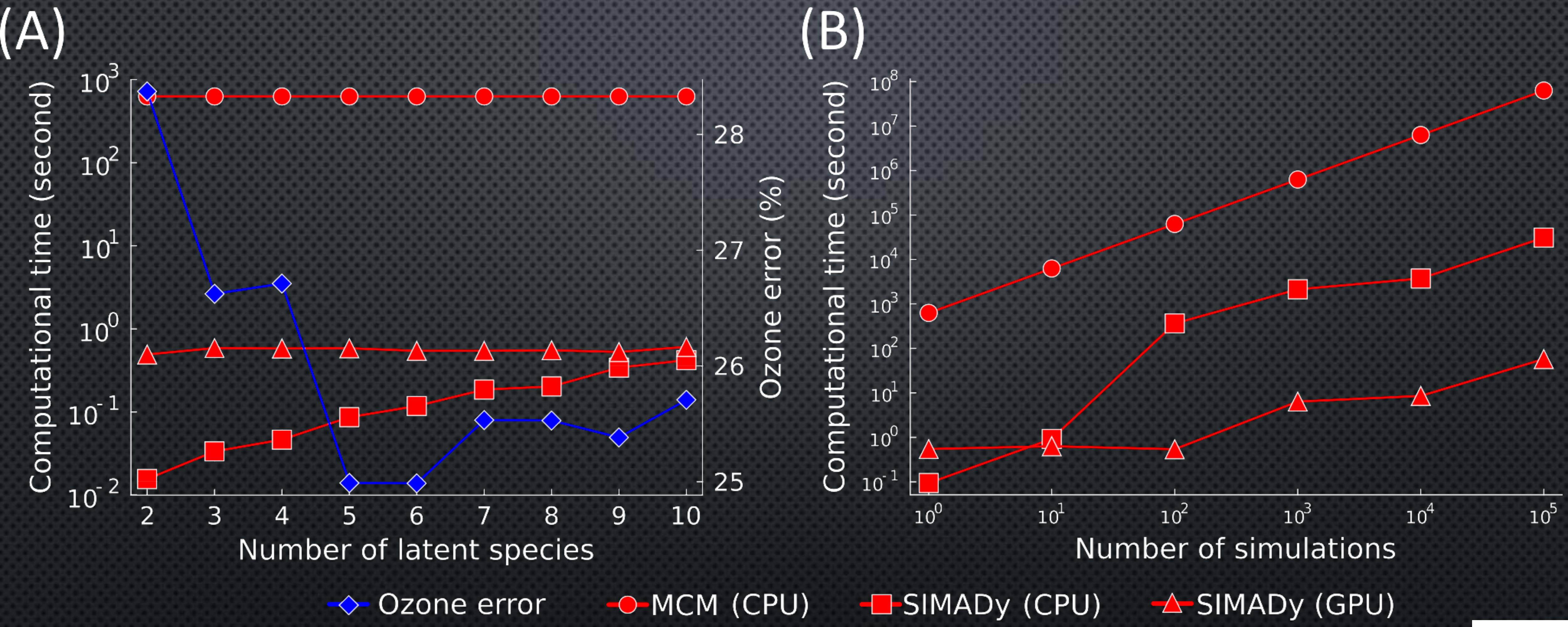
What if we need the model to be *faster*?

Machine-Learned Atmospheric Chemistry



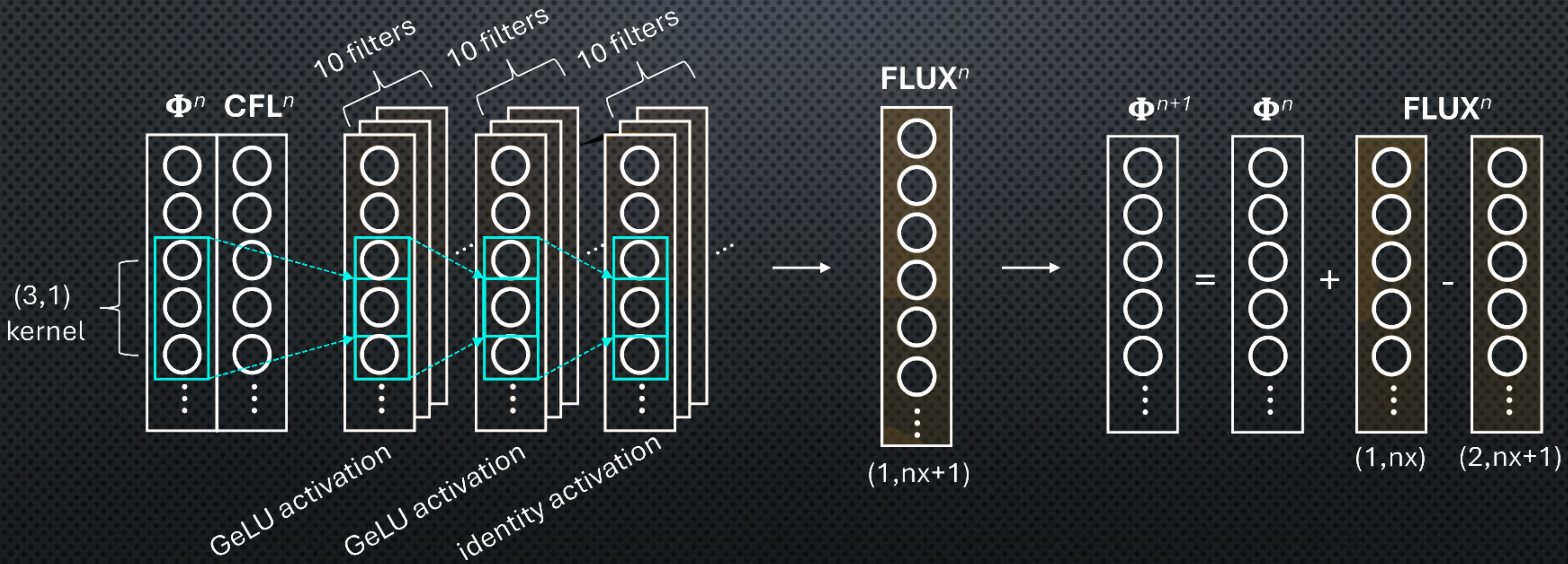
Yang, X., Guo, L., & Tessum, C. W. (2025). Atmospheric chemistry surrogate modeling with sparse identification of mass action dynamics. *Journal of Geophysical Research: Machine Learning and Computation*, 2, e2025JH000731. **https://doi.org/10.1029/2025JH000731**

Machine-Learned Atmospheric Chemistry

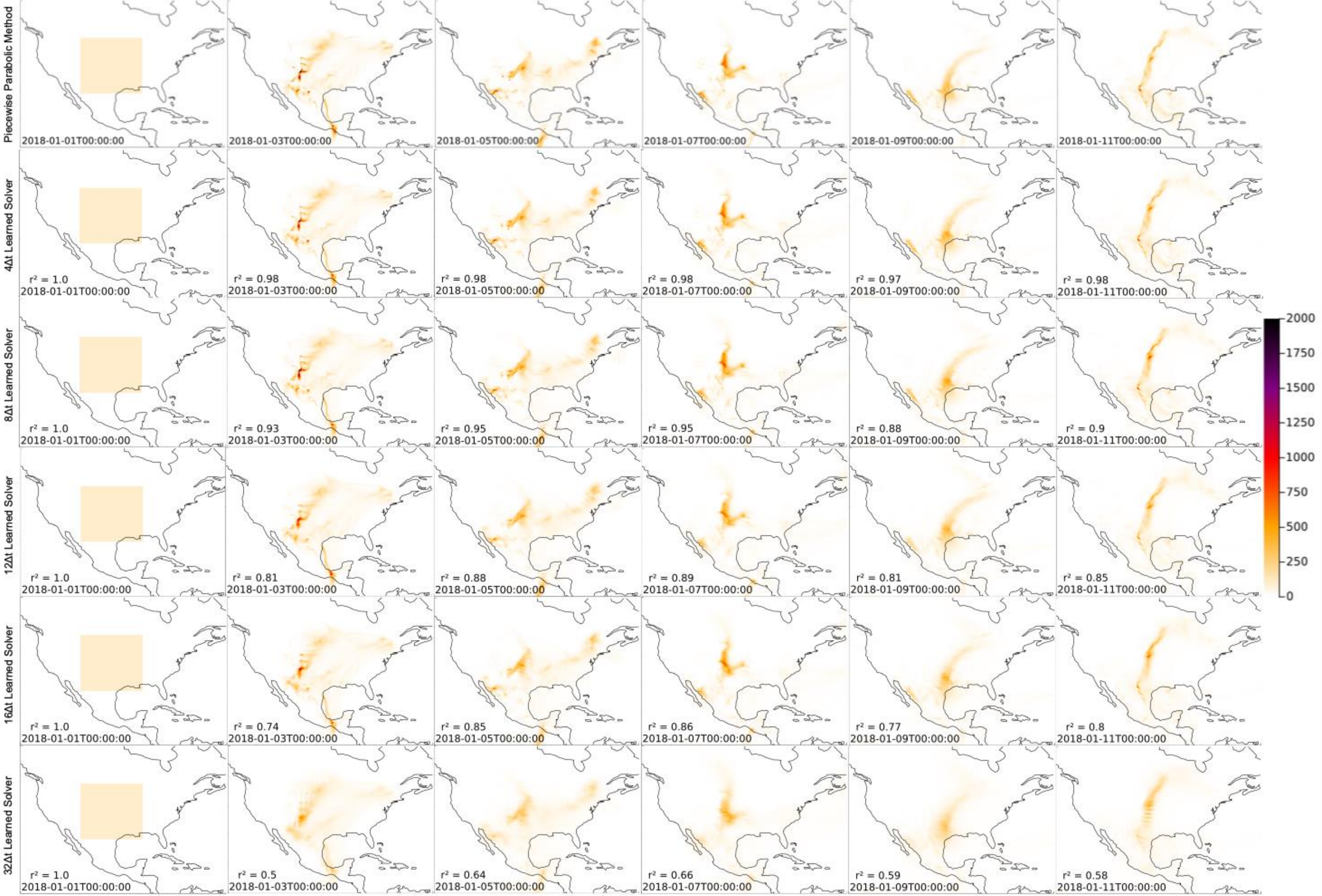


Yang, X., Guo, L., & Tessum, C. W. (2025). Atmospheric chemistry surrogate modeling with sparse identification of mass action dynamics. *Journal of Geophysical Research: Machine Learning and Computation*, 2, e2025JH000731. **https://doi.org/10.1029/2025JH000731**

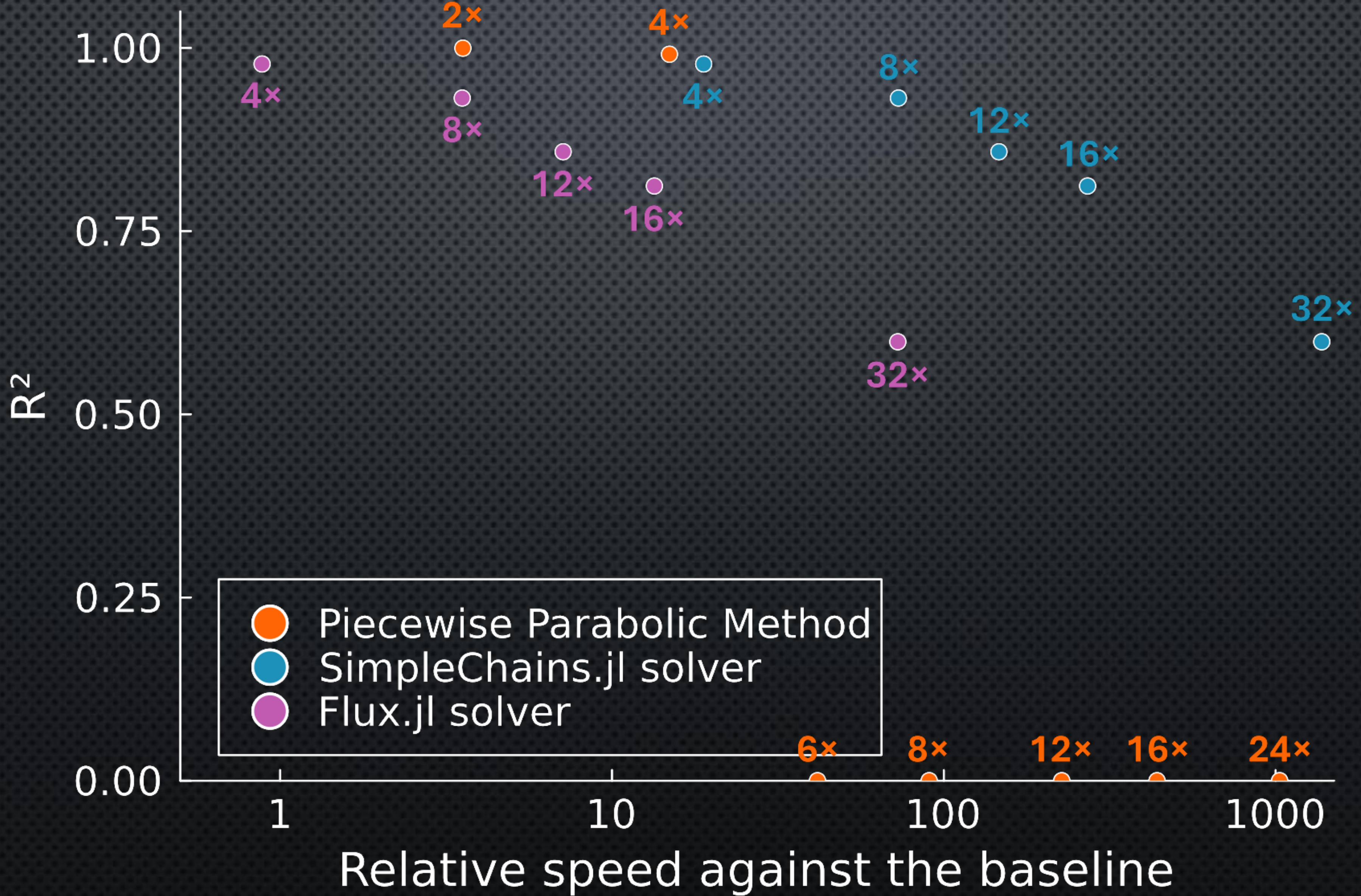
Machine-Learned Advection



Machine-Learned Advection



Machine-Learned Advection



Optimization

- https://earthsci.dev/dev/examples/optimization/
- We are trying to optimize against Earth observation data.

Conclusion

1. Through the use of equation compilation and machine learning, we can make air quality models faster, more realistic, and easier to use.

2. Most of these techniques are general and could apply to other areas of (geo)science.






EarthSciML

https://earthsci.dev

This work was supported by an Early Career Faculty grant from NASA’s Space Technology Research Grants Program and by Assistance Agreements No. R835873 and RD-84001201-0 awarded by the U.S. Environmental Protection Agency, as well as agreements RISE 24-25760 and RISE 25-24740 awarded by the NSF. It has not been formally reviewed by EPA. The views expressed in this document are solely those of authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication.

Potential Discussion Topic

- Julia Infrastructure for satellite data, e.g. similar to: TorchGeo