JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Difference Learning for Air Quality Forecasting Transport Emulation

NeurIPS 2023 Workshop: Tackling Climate Change with Machine Learning

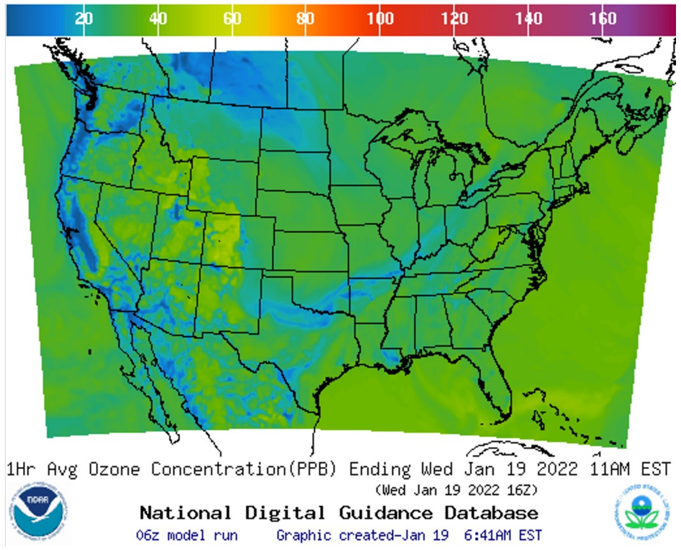
Reed Chen | Researcher, Johns Hopkins Applied Physics Laboratory
Christopher Ribaudo, Jennifer Sleeman (PI),
Chace Ashcraft, Collin Kofroth, Marisa Hughes,
Ivanka Stajner, Kevin Viner, Kai Wang

This work was funded by NOAA grant NA21OAR4310383, SUBAWD003728

Dynamic Transport Emulation

- NOAA provides operational AQ forecast guidance, including ozone and PM 2.5, based on the Unified Forecast System Air Quality (UFS-AQ) model at a 12 km resolution for the continental US (CONUS)
- Modeling the advective transport of chemical and aerosol tracers accounts for 40% of the overall computation
- We aim to develop a machine-learning approach, specifically a 3D U-Net, to emulate the UFS-AQ transport model and reduce computation time

British Columbia Wildfire Service, 2023
Reuters



AQ Forecasting

Data Distribution

- Concentration distributions vary greatly between atmospheric species
- Difference between the input and output of the UFS-AQ model is small
- The advection equation and our ML model are species agnostic
- Can still apply per-species min-max normalization

$$\frac{\partial c}{\partial t} + \nabla \cdot (\boldsymbol{v}c) = 0 \Leftrightarrow \frac{\partial (ac+b)}{\partial t} + \boldsymbol{v} \cdot \nabla (ac+b) = 0$$

Advection continuity equation where *c* is the species concentration, ***v*** is the velocity vector field, and *a* and *b* parameterize affine transformations.

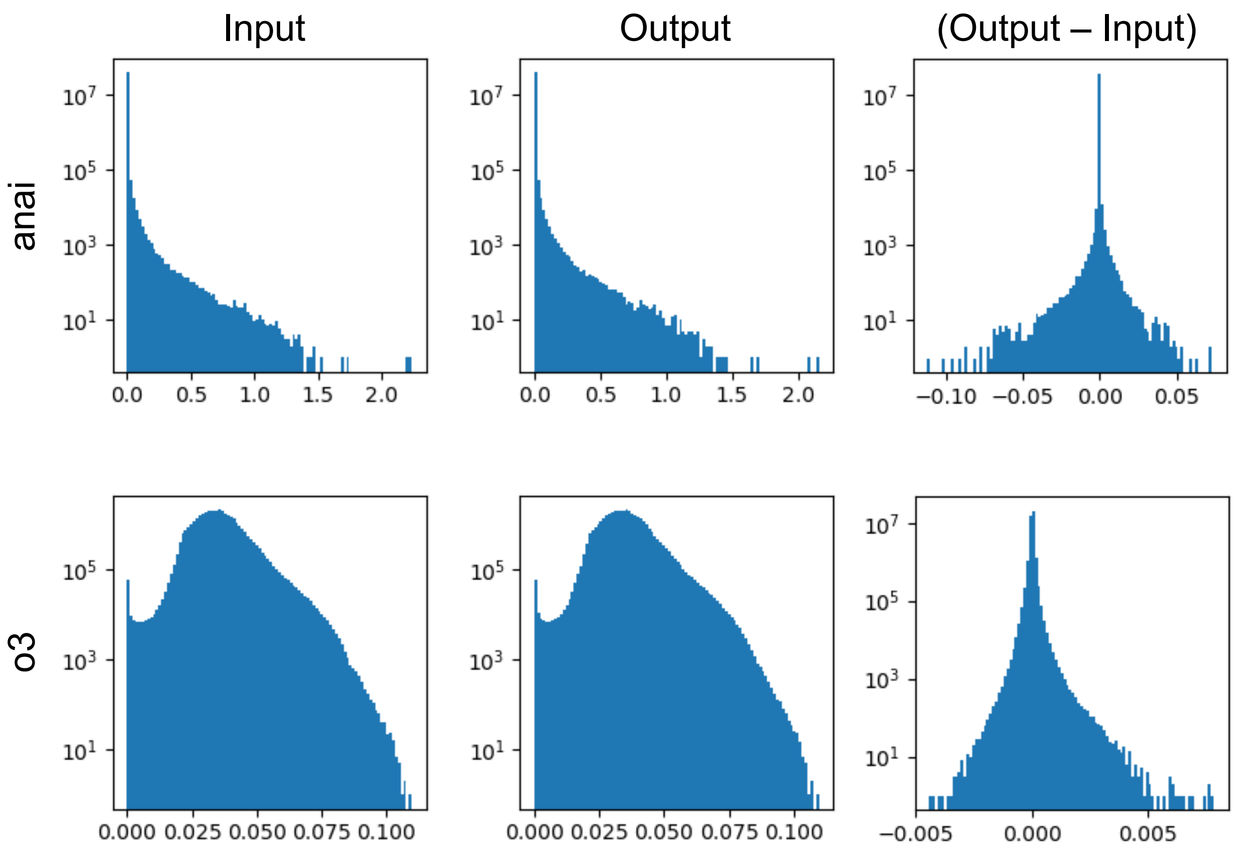


Figure 1: Concentration distributions of *anai* (a PM2.5 constituent) and ozone in μg/m³ and ppm respectively. Column 3 is the distribution of the difference between the Input and Output data. Note that the y-axis is log-scaled.

Difference Learning

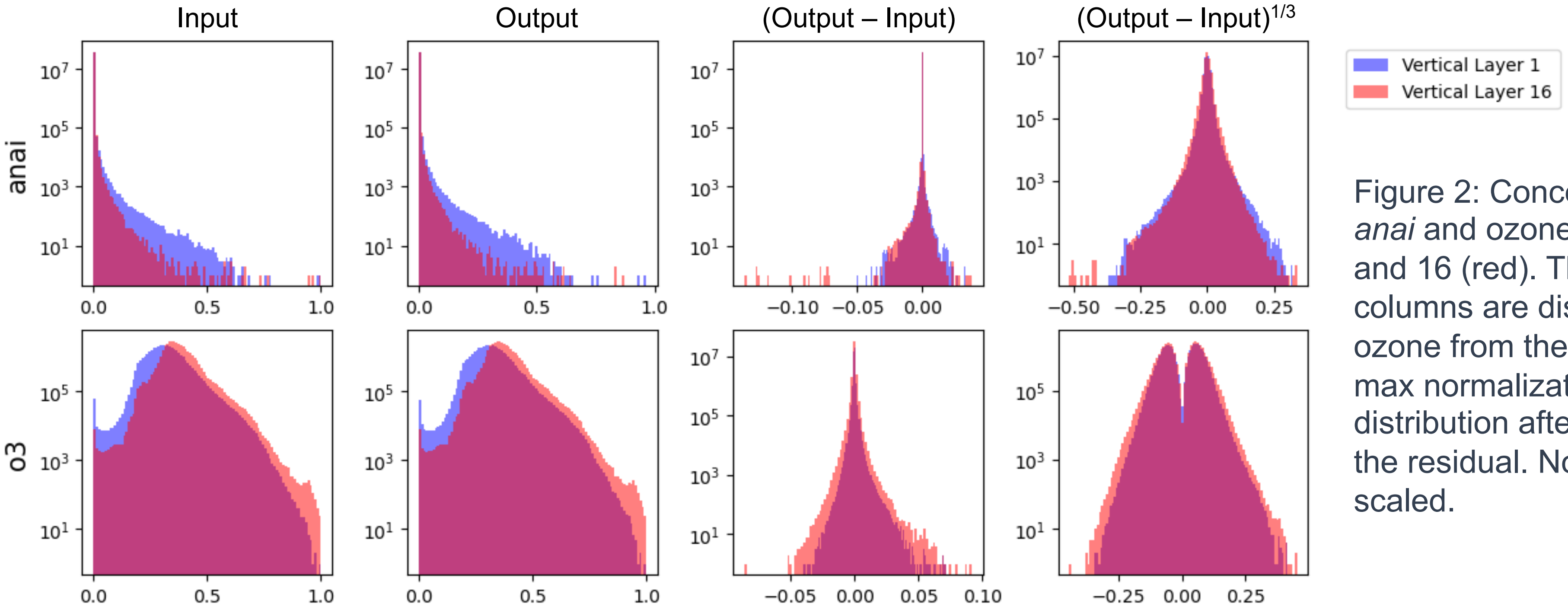


Figure 2: Concentration distributions of *anai* and ozone at vertical levels 1 (blue) and 16 (red). The Input and Output columns are distributions of *anai* and ozone from the UFS-AQ model after min-max normalization. Column 4 is the distribution after taking the cube-root of the residual. Note that the y-axis is log-scaled.

- Normalize the residual with a species-agnostic method
- Applying a cube-root transformation to the residual increases the spread of the distribution and facilitates training

Experimental Setup

- Dataset
  - Chemical species concentrations and meteorological variables across CONUS were generated from the UFS-AQ model
  - Data was collected for 7 days between Sept 1, 2020 and Oct 1, 2020
  - The CONUS was divided into 24 patches, and only the lowest 16 vertical atmospheric layers were used
  - Patches were classified as extreme or nonextreme based on the EPA's Air Quality Index thresholds
  - 394,214 train, 98,554 validation, 369,576 test patches
- Model
  - 3D U-Net (90,310,657 parameters)
- Training
  - 1 GPU: Tesla V100
  - 20 epochs

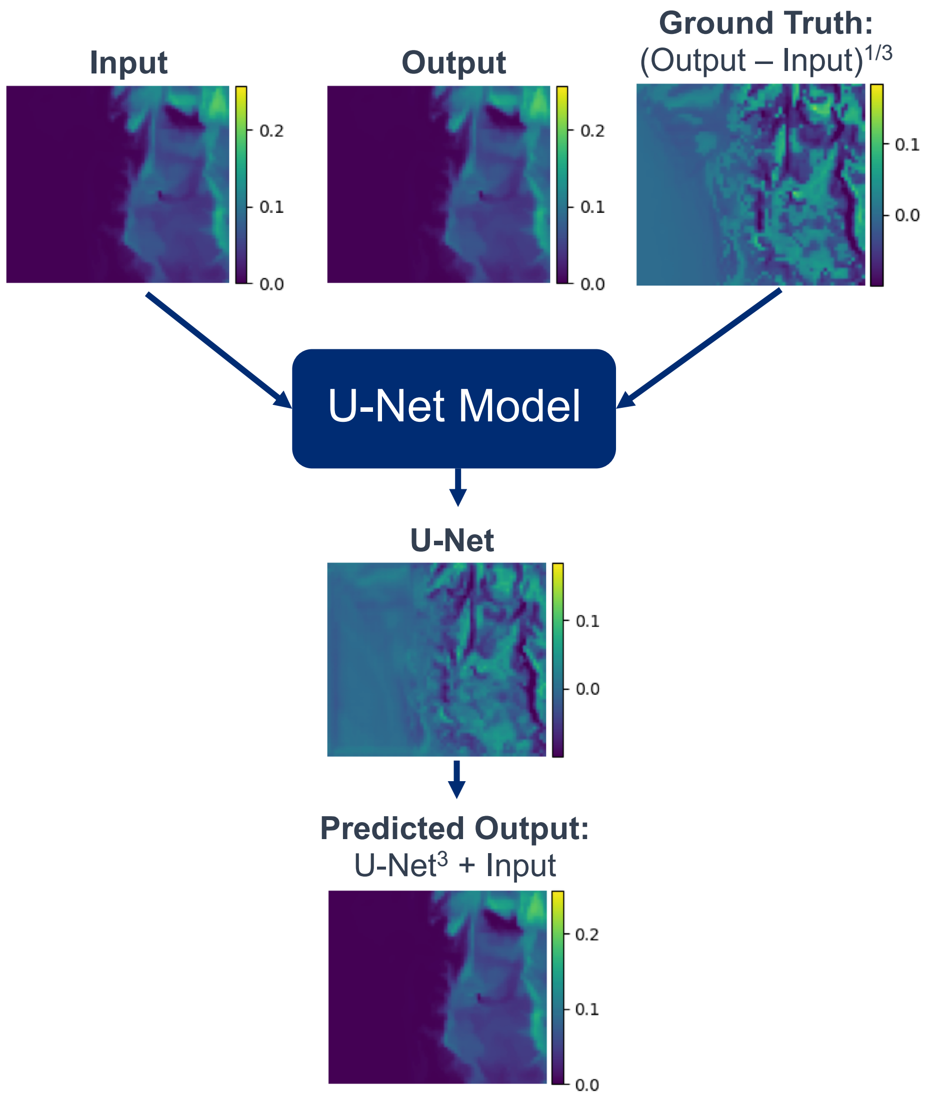


Figure 3: The U-Net learns the mapping between Input and (Output − Input)$^{1/3}$. The Predicted Output is the U-Net's prediction in the min-max normalized space.

Results

- Overall RMSE of 0.0115
- Estimated 2.6 seconds to produce predictions for all species over the entirety of CONUS
- Mass conservation of *asvpo1j* (a PM2.5 constituent) with a mass difference of 0.0741%

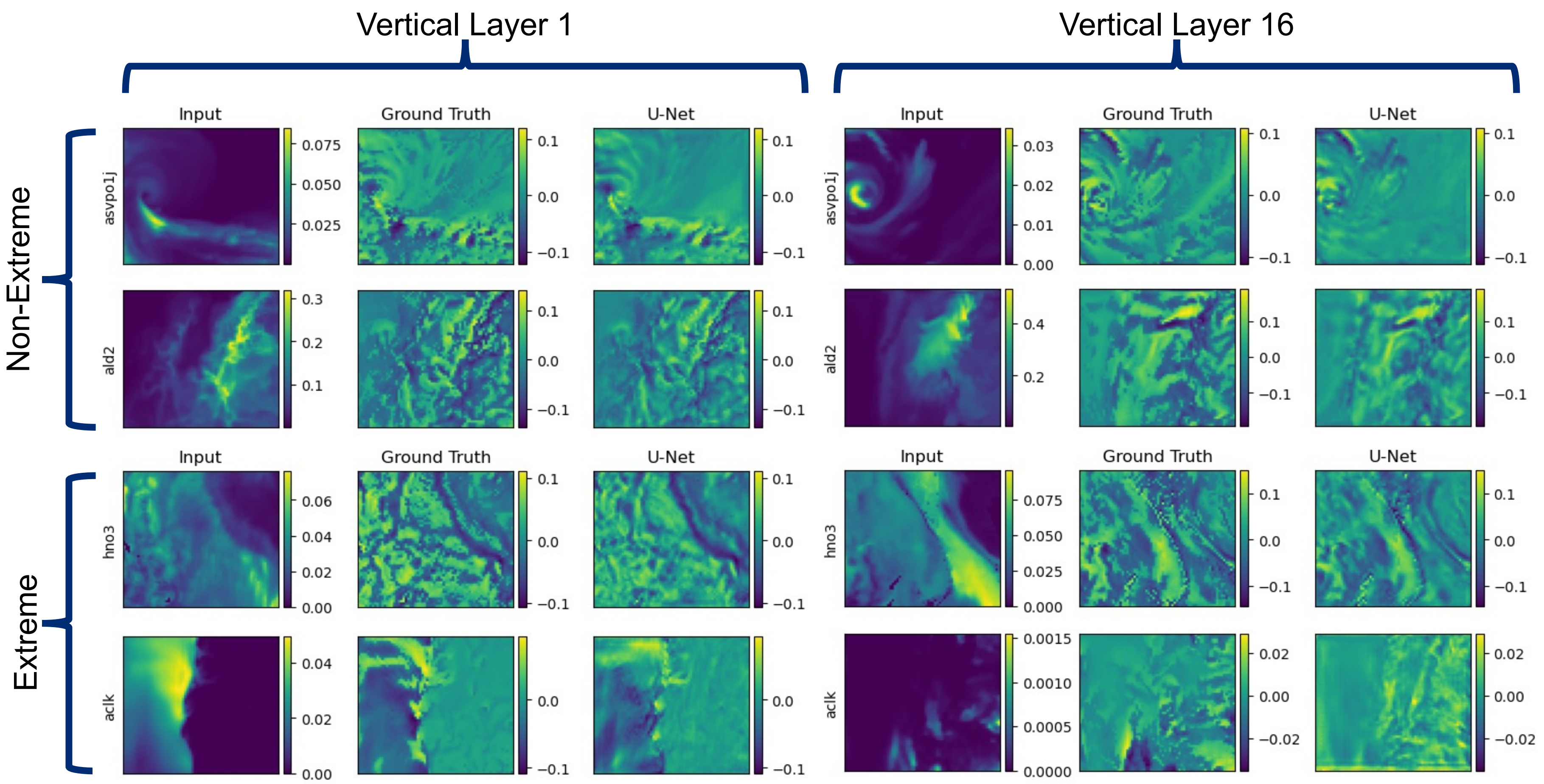


Figure 4: U-Net predictions for different species, vertical layers, and extreme/non-extreme events.

Table 1: U-Net RMSEs calculated in the min-max normalized space.

Non-Extreme	Extreme	All Data
0.00838	0.0129	0.0115

Conclusion

- Our model emulates the per-timestep advective transport of atmospheric chemical species
- With an overall RMSE of 0.0115 and estimated prediction time of 2.6 seconds, this ML method exhibits significant potential for integration into the NOAA operational air-quality environment
- Ultimately, we aim to develop an ML model which efficiently emulates advective transport over large time-scales for implementation in the UFS-AQ model

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY