Downscaling UFS AQM Surface NO_2 to High-Resolution using Ground & Satellite (TEMPO) Observation through Machine Learning Data Fusion

Beiming Tang[1,2], Barry Baker[1], Daniel Tong[1,2], Patrick Campbell[1,2], Youhua Tang[1,2], Wei Li[1,2], Madankui Tao[3], Siqi Ma[2]

[1]NOAA Air Resources Laboratory, Atmospheric Sciences Modeling Division;
[2]George Mason University, Center for Spatial Information Science & Systems;
[3]MIT, Department of Earth, Atmospheric, and Planetary Sciences

1. Background & Objectives

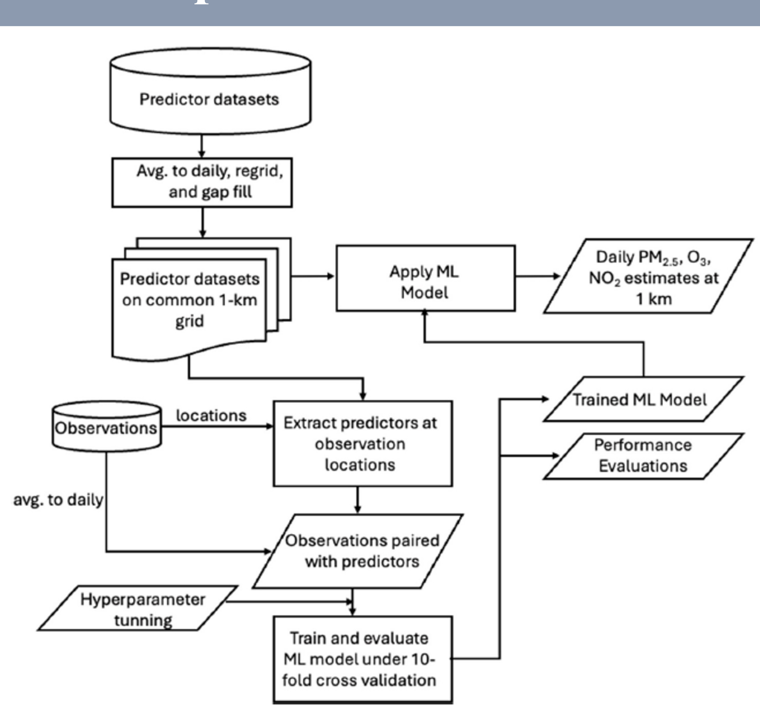
- Each year, more than 6.7 million deaths are attributed to indoor and outdoor air pollution according to most recent Global Burden of Disease study. There is a need for improved estimates of exposure to support air quality management, research and compliance, and analysis of air pollution impacts. However, ground observations are unevenly distributed and focus on urban regions, leaving majority of rural region and developing countries un-represented.
- We proposed an approach to fill gaps in observation networks by machine learning data fusion: including **NOAA UFS AQM** model predictions, tropospheric NO_2 column from **TEMPO**, and high-resolution NEMO anthropogenic emissions. We aim to build a daily, 1-km distribution dataset for NO_2 with potential to include other species in CONUS domain.

2.Representative Inputs

Inputs	Source	Year of Product
CTM	UFS-AQM	2023
NO_2 column	TEMPO	2023
NO_2 column (merge)	TROPOMI	2023
AOD	MODIS MAIAC	2023
NDVI	MODIS	2023
Land use cover	MODIS	2022
Meteorology	ERA5	2023
Anthropogenic Emissions	NEMO	2019
Elevation	SRTM	2022
Population	Land Scan	2022

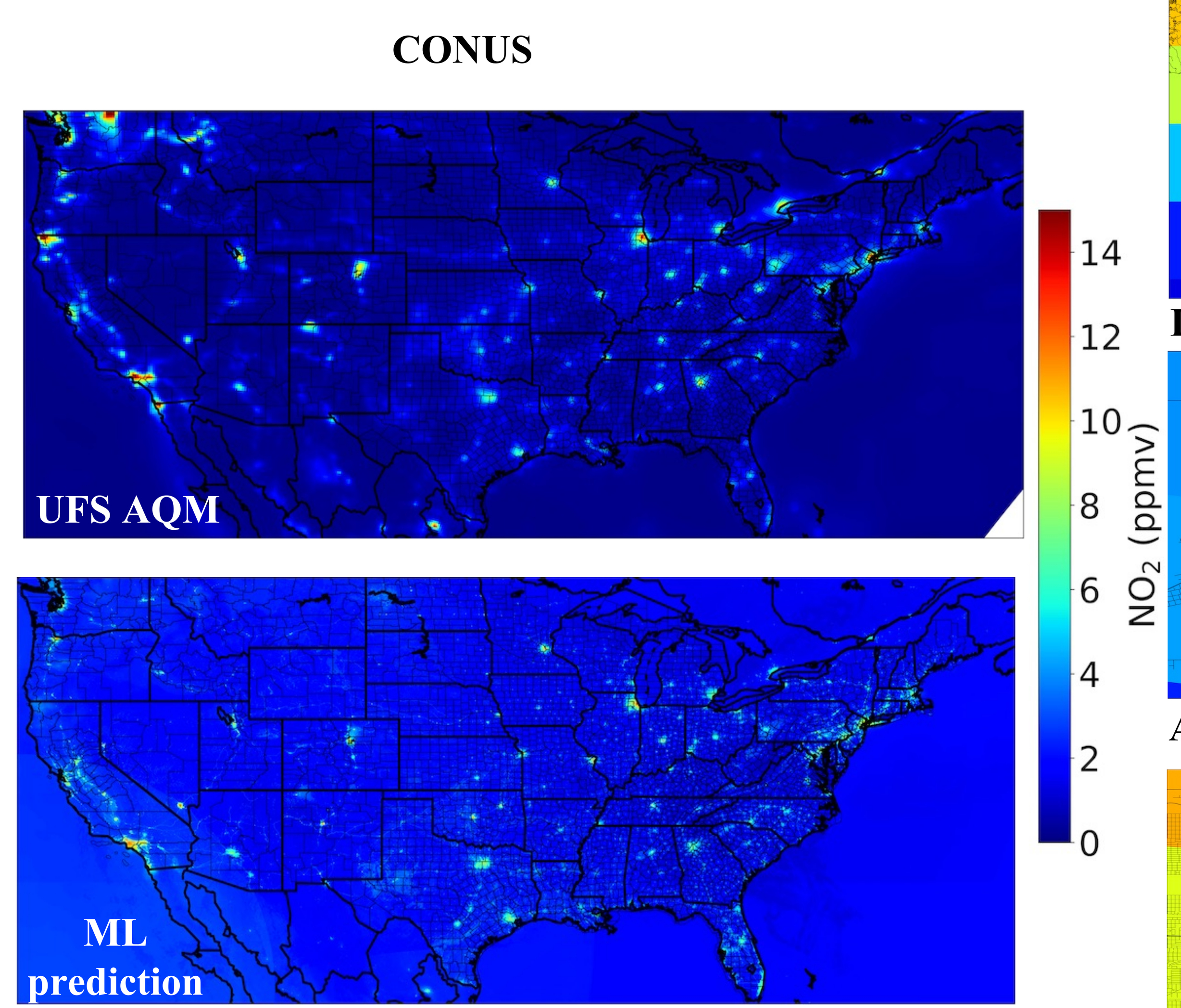
- Chemical Transport Model output from NOAA UFS-AQM.
- Tropospheric NO_2 column is from TEMPO, gap-filling with TROPOMI NO_2.

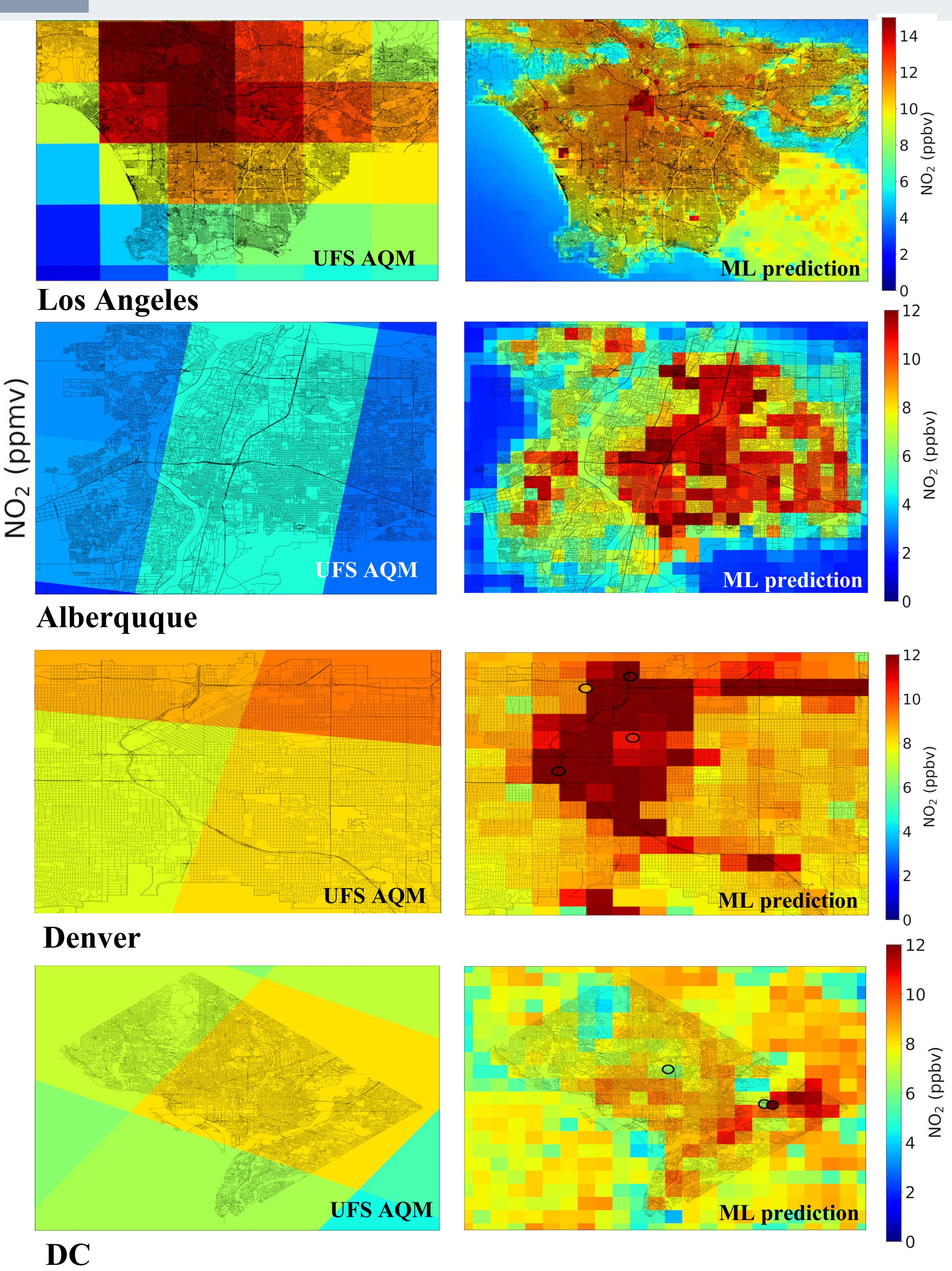
3.Model Development



- We choose Aug-2023 to build the machine learning model. Aug, 2023 was chosen to represent AEROMMA (https://csl.noaa.gov/projects/aeromma/) field campaign.
- We have total of 3309 data for 112 sites within CONUS.
- All inputs were re-grid to daily, 1km before 2D-interpolated to ground station locations. The interpolated data were used for training the model. A random forest algorithm is applied for building ML model
- The developed ML model were applied back to inputs daily, 1km matrix to generate spatial 1km daily estimates.

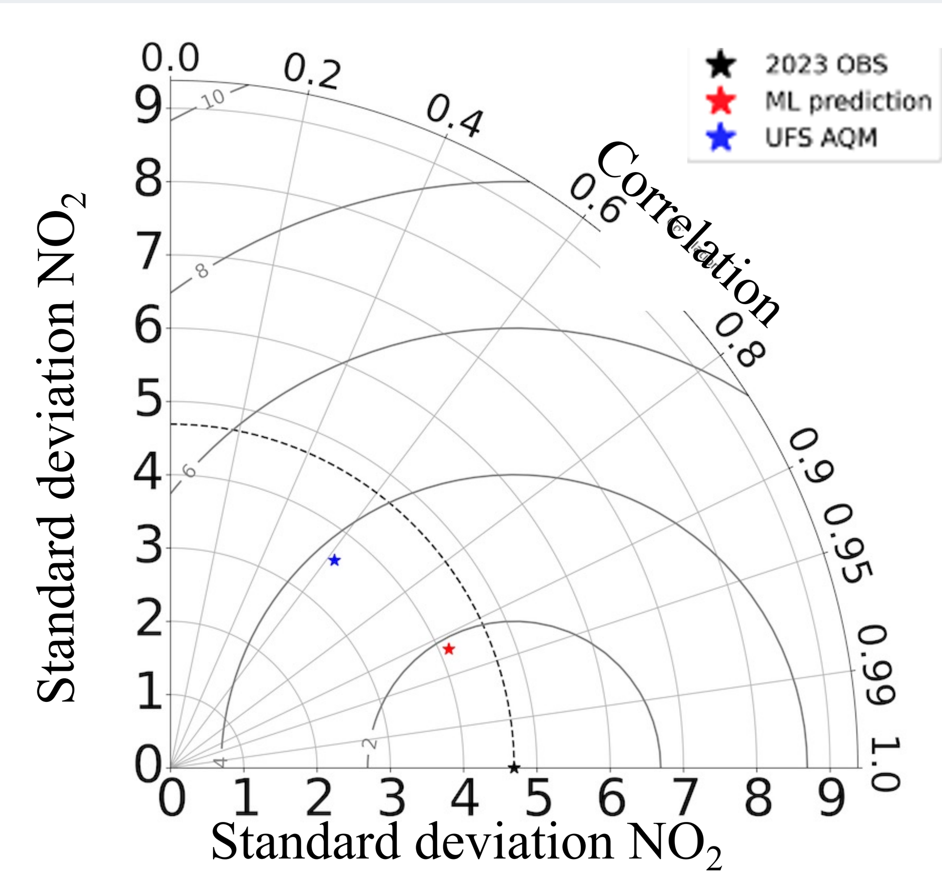
4. Result Spatial Distribution



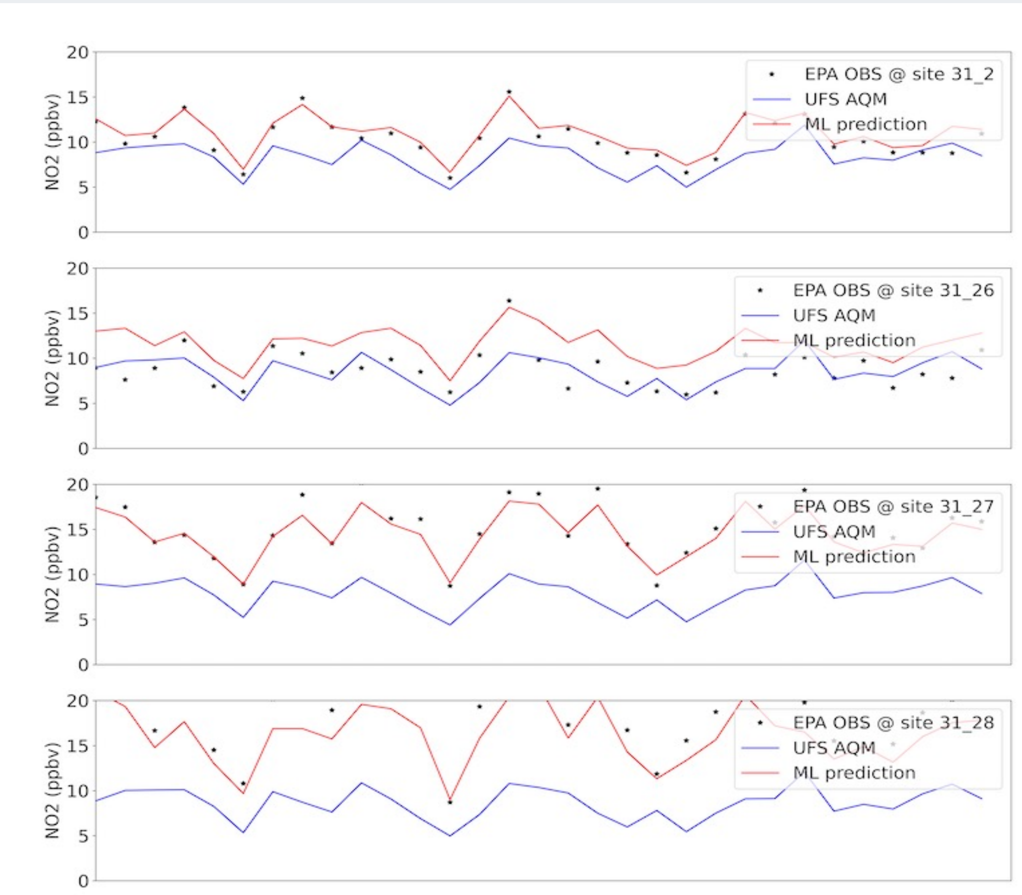


- Machine learning (ML) prediction 1km Downscaled from Original UFS-AQM 13km was plotted against EPA AQS observation in CONUS domain
- Our high-resolution model enables to demonstrate sub-city scale surface NO_2 distribution as examples shown in LA, Denver, Alberquerque, and DC.
- Background "square" color indicate ML model prediction, while "dots" indicate ground observation from EPA AQS
- Daily product averaged to monthly mean being plotted

5. Evaluation and Temporal Skills

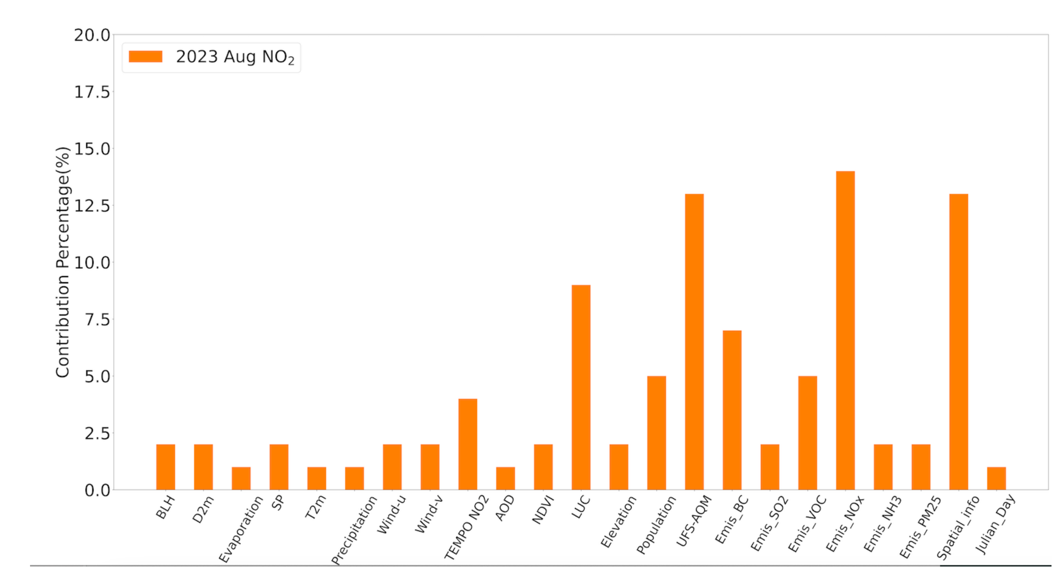


$PM_{2.5}$	R	RMSE (ppbv)	STD model (ppbv)	STD obs (ppbv)
ML model	0.92	1.88	4.13	4.70
Original UFS-AQM	0.62	3.94	3.61	4.70



- Taylor Diagram show ML improvement from original UFS-AQM (blue to red)
- Time Series in representative cities (Denver) show ML model temporal skills in capture peak and valleys for NO_2

6. Inputs Contribution



- Contribution plot indicate **UFS-AQM, Emission of NO_X, spatial input, and land use cover** information are among the top contributor

7. Study Limitation and Future Plan

- Some of the inputs are obsolete (anthropogenic emission is from year 2019). Need to upgrade to most recent version
- Adapt algorithm into a forecast mode or "NRT Downscaled Forecast" that uses forecasted AOD, weather parameters, and concentrations.
- Adapt AOD gap filling algorithm to use GEFS-Aerosol predictions and NESDIS VIIRS products.
- Adapt inputs to use NOAA inputs (perhaps RRFS/HRRR analysis)
- Reprocess historical NOAA NAQFC data to provide high resolution downscaled reanalysis products.
- With MODIS reaching end of life the algorithms and training need to be adapted to use VIIRS products.
- Low-cost sensors from Purple Air and Open-AQ provide a denser network that could improve the ML bias and spatial distribution across the US.

8. Conclusion

- Developed daily, downscaled, reanalysis 1km resolution NO_2 estimates data based on NOAA UFS-AQM and TEMPO retrieval.
- Ten-fold cross-validation indicates satisfying results against routine observations from ground stations.

9. References and Acknowledgements

Related References:
Tang et al. , (2024) "Ozone, nitrogen dioxide, and PM2.5 estimation from observation-model machine learning fusion over S. Korea: Influence of observation density, chemical transport model resolution, and geostationary remotely sensed AOD". Atmospheric Environment. https://www.sciencedirect.com/science/article/pii/S1352231024002784

Ma and Tong., (2022) Neighborhood Emission Mapping Operation (NEMO): A 1-km anthropogenic emission dataset in the United States. Sci Data 9, 680 (2022). https://doi.org/10.1038/s41597-022-01790-9

Acknowledgements:
This study is sponsored by NOAA IIJA award NA22NES4050023D/119812-Z7646201 and ORAU's subaward 2023-1007-01 with CISESS.

Contacts:
Dr. Beiming Tang (beiming.tang@noaa.gov)
Dr. Barry Baker (barry.baker@noaa.gov)