Quantifying Power Plant NO_x Emissions Using Deep Learning

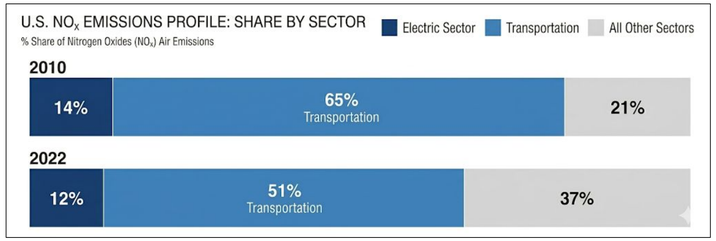
Cohen Research
An Atmospheric Chemistry Laboratory

Students: Pranav Walimbe, Ryuto Tsuruoka | Advisors: Prof. Ron Cohen, Hikari Murayama

College of Chemistry
UNIVERSITY OF CALIFORNIA, BERKELEY

Background

- Power plants are among the largest emitters of nitrogen oxides (NO_x) globally (640,000+ tons in the U.S. alone in 2023)
- Many nations outside U.S., E.U. lack plant emissions transparency due to inconsistent mandates and reporting failures
- TEMPO is a NASA geostationary satellite launched in 2023 that measures hourly trace gas concentrations over the U.S. with state-of-the-art precision (~3km per pixel)
- We present one of the first deep learning frameworks to quantify NO_x emissions from TEMPO images on 40 U.S. coal power plants

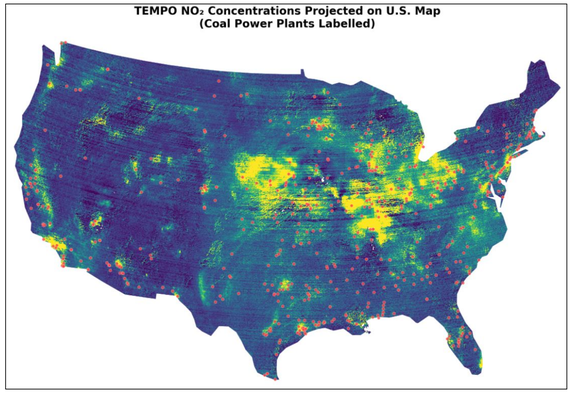


Source: ERM Group "Benchmarking Air Emissions" (2023)

Data Collection

Data Sources:

- Plant-level emissions: ~7 GB of plant-level hourly NO_x emissions
- Satellite data: ~7 TB of hourly NO_2 vertical column density (TEMPO)
- Wind data: ~2.5 GB of hourly ERA5 atmospheric reanalysis



Power Plant Filtering:

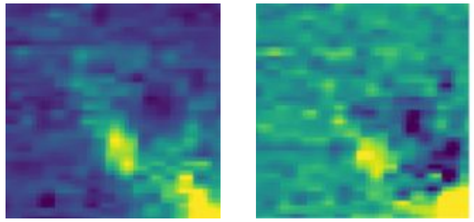
- Strictly used stand-alone coal plants for better plume visibility (kept ~10% of total plants)

Dataset Creation

- Extracted 72km x 72km patches around plants
- Quality filtering: distance to nearby major city, cloud cover, plume visibility (kept ~30% of samples)
- Dataset split sizes: 17.5K train, 3.5K test, 3.5K validation
- Performance: used multiprocessing on Savio HPC to parallel process ~7TB of satellite images, achieved ~8x speedup

Model Inputs:

NO_2 image | NO_2 (t) - NO_2(t-1) image | Tabular features



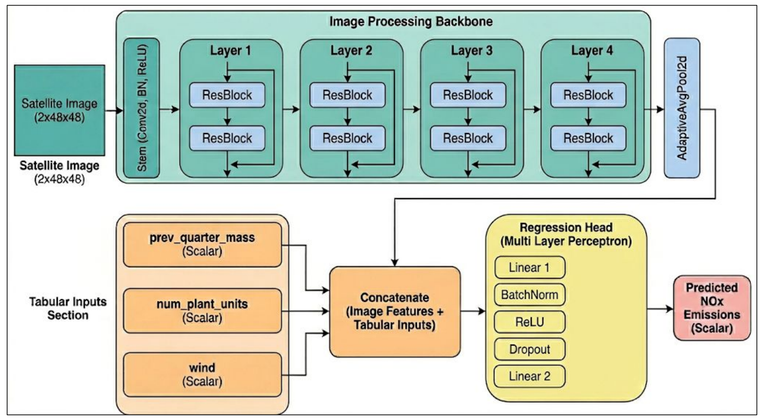
- Wind east-west
- Wind north-south
- Num plant units
- Prev quarter NO_x

Labels: hourly NO_x mass emissions (lb/hr) by power plant

Modeling

- Custom ResNet Convolutional Neural Network (CNN)
- Cross-compared against custom DenseNet CNN
- Trained on A40 GPUs using Savio HPC

Model Architecture (ResNet):

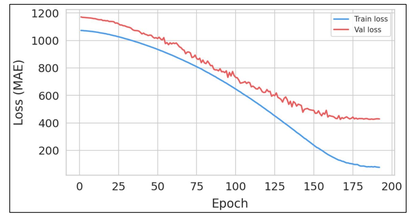


Results

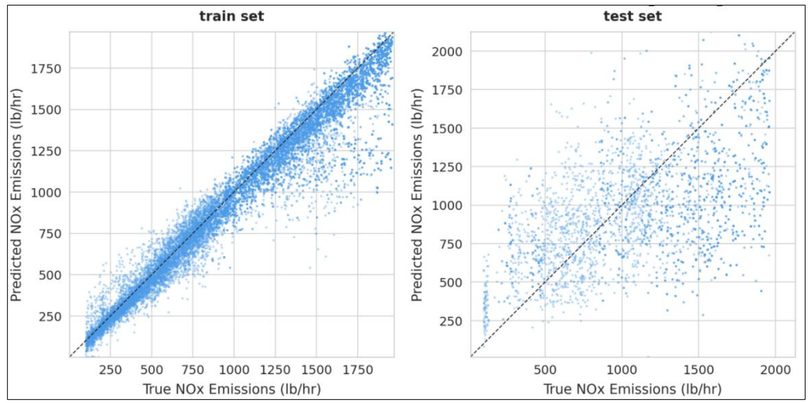
Performance Comparison:

		Mean Absolute Error (MAE, lb/hr)	
		Model	
		ResNet	DenseNet
Split	Train	84.22	114.56
	Test	373.13	360.02
	Validation	429.06	433.56

Loss Curve (ResNet):



Predictions Scatterplot (ResNet):



Conclusion & Future Work

- We were primarily limited by (1) coarse resolution of TEMPO images, (2) number of dataset samples that fit our filtering requirements, and (3) temporal misalignment between labels (hourly sum) and images (intra-hour snapshot)
- Using the extensive filtering discussed, we were able to achieve ~35% improvement in performance
- Future work will explore improved feature sets (e.g. correlated trace gases, atmospheric variables), denoising images, and additional models for the task

Acknowledgements

We'd like to thank Professor Ron Cohen and Hikari Murayama for their mentorship throughout the course of the project