Adaptive Expert-guided Deep Imbalanced Regression for Global PM2.5 Forecasting with Temporal Convolutional Networks and GEOS-FP Inputs

Junhyeon Seo[1,2], Alqamah Sayeed[3,4], Pawan Gupta[2]

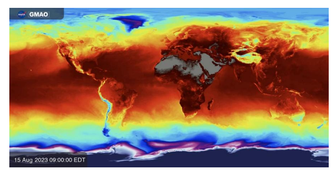
[1] Goddard Goddard Earth Sciences Technology and Research (GESTAR) II, Morgan State University, Baltimore, MD, 21251, USA
[2] Goddard Space Flight Center (GSFC), National Aeronautics and Space Administration (NASA), Greenbelt, MD, 20771, USA
[3] Universities Space Research Association Huntsville (USRA), Huntsville, AL, 35806, USA
[4] Marshall Space Flight Center (MSFC) , National Aeronautics and Space Administration (NASA), Huntsville, AL, 35808, USA

INTRODUCTION

- **Background**
  - **Growing Concern**: Air quality is a significant issue in many urban areas globally, particularly in developing countries experiencing rapid industrialization
  - **Health Impacts**: PM2.5, a harmful atmospheric aerosol, negatively affects human health, leading to lung and heart diseases
  - **DoS Initiative**: The Department of State (DoS) initiated the DoSAir program in collaboration with the US EPA in 2008 to monitor air quality at embassies and consulates worldwide which was stopped in early 2025
  - **Forecasting Gap**: Despite extensive monitoring, the DoSAir program needs air quality forecasting, which is provided by NASA
- **Objective**
  - Improve a City Air quality foREcasting and analysis System (CARES) using advanced machine learning & NASA models, in collaboration with the DoS, and NASA's HAQAST & SNWG programs

DATA and METHODOLOGY

Input & Output



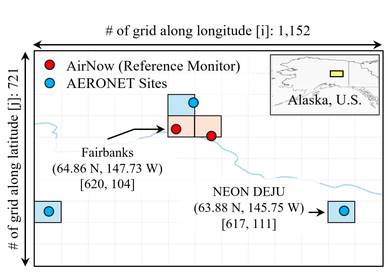
GEOS Forward Processing

GEOS FP Meteorology: Q500/QV10m, T850/T500/T10m, Wind, PS

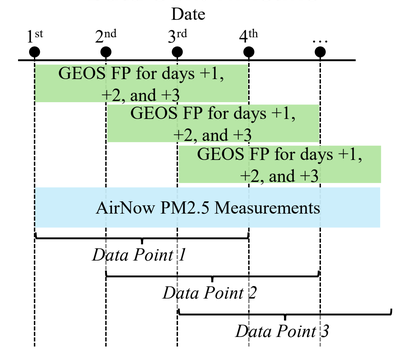
GEOS FP Aerosols: DUSMASS25, OCSMASS, SO2SMASS, SO4SMASS, TOTEXTTAU, BCSMASS, SSSMASS25

AirNow & OpenAQ PM2.5

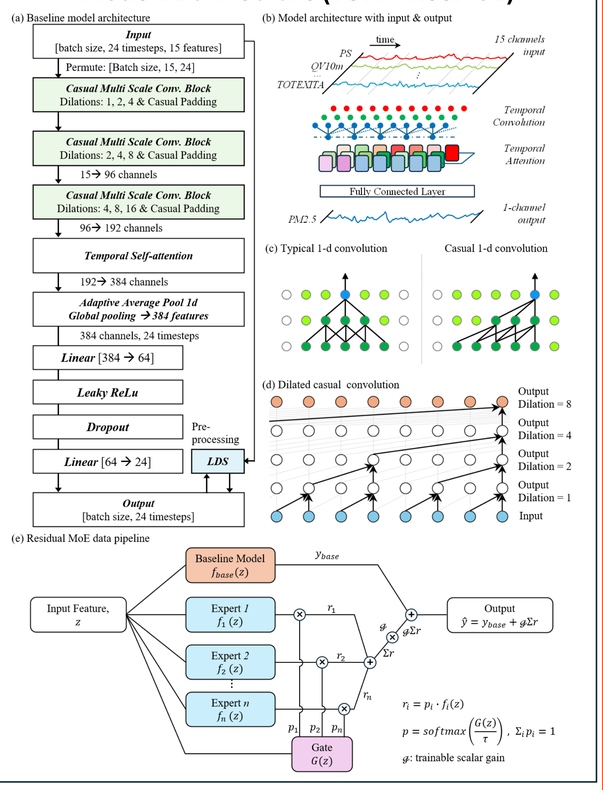
Data/Forecasting Location



Data Gen. Timeline

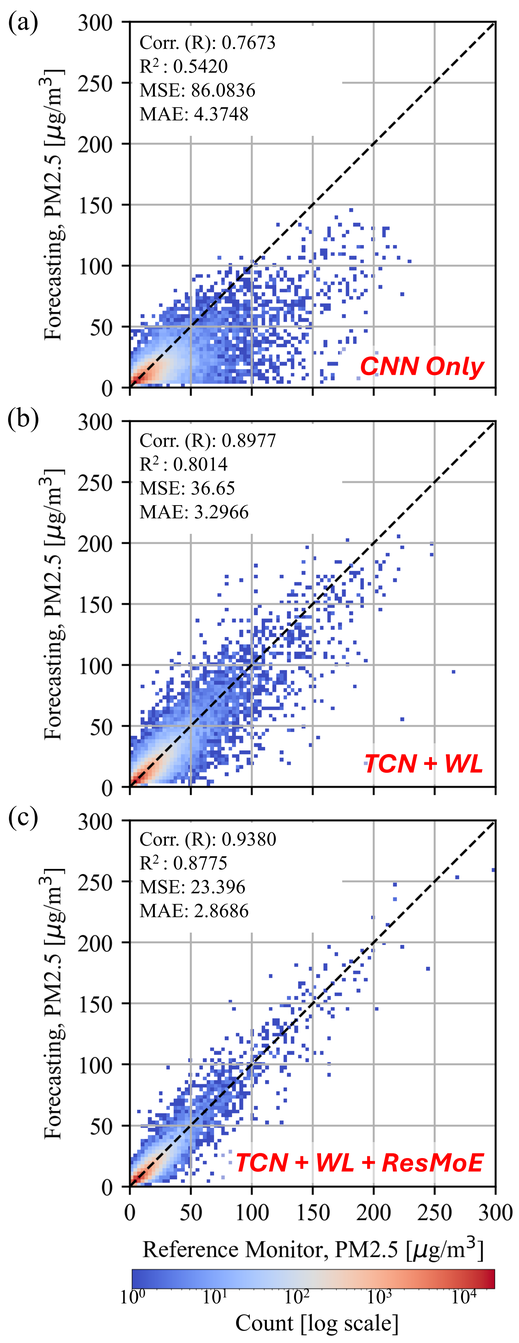


Model Architecture (TCN + ResMoE)

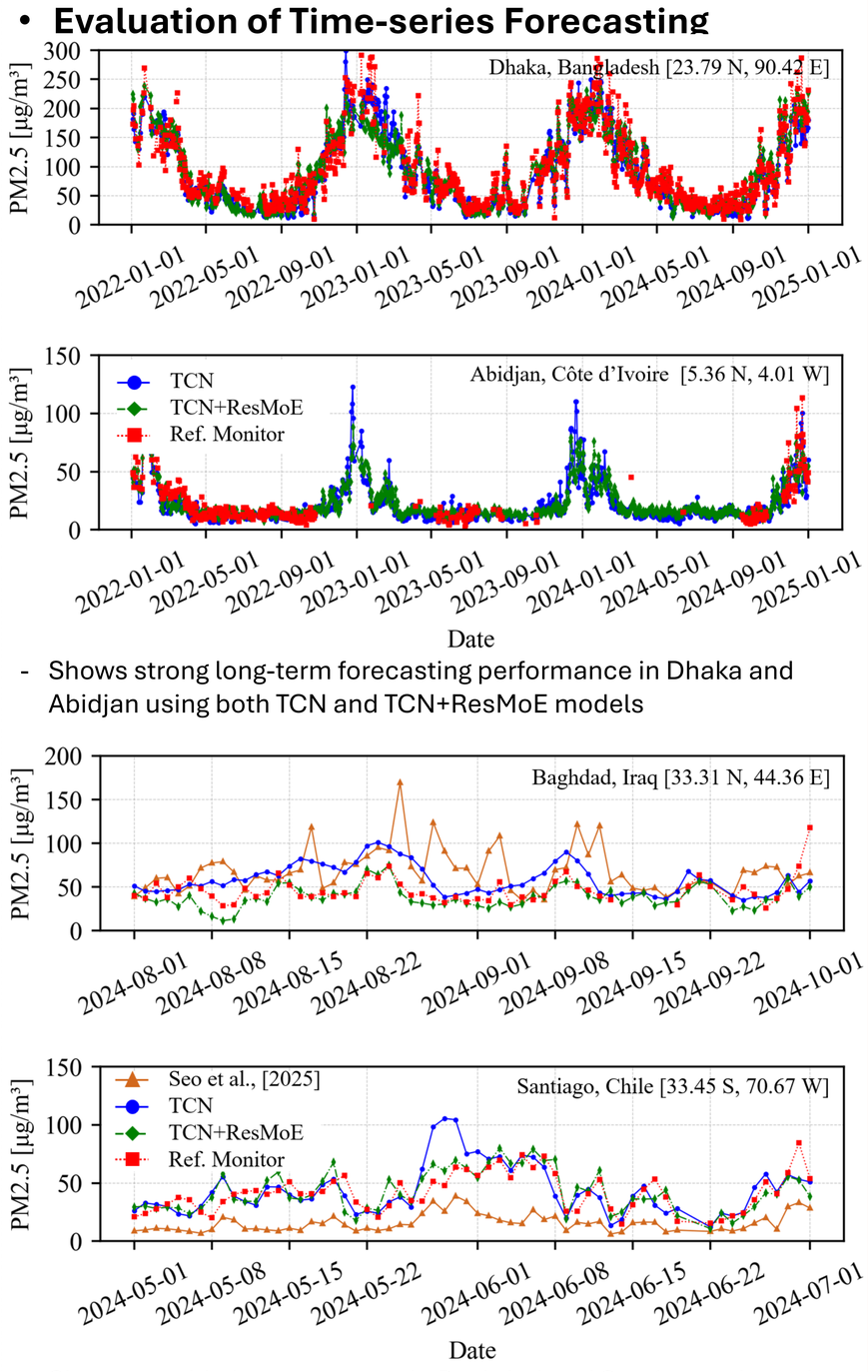


RESULTS

Scatter–Performance Comp.



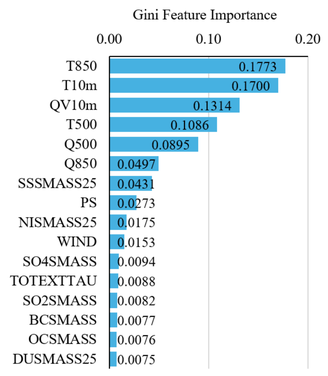
Evaluation of Time-series Forecasting



- Shows strong long-term forecasting performance in Dhaka and Abidjan using both TCN and TCN+ResMoE models
- Shows notable improvements in Baghdad and Santiago, where the TCN + ResMoE model greatly reduced errors and better captured regional meteorological patterns compared to the baseline CNN

Discussion and Applications

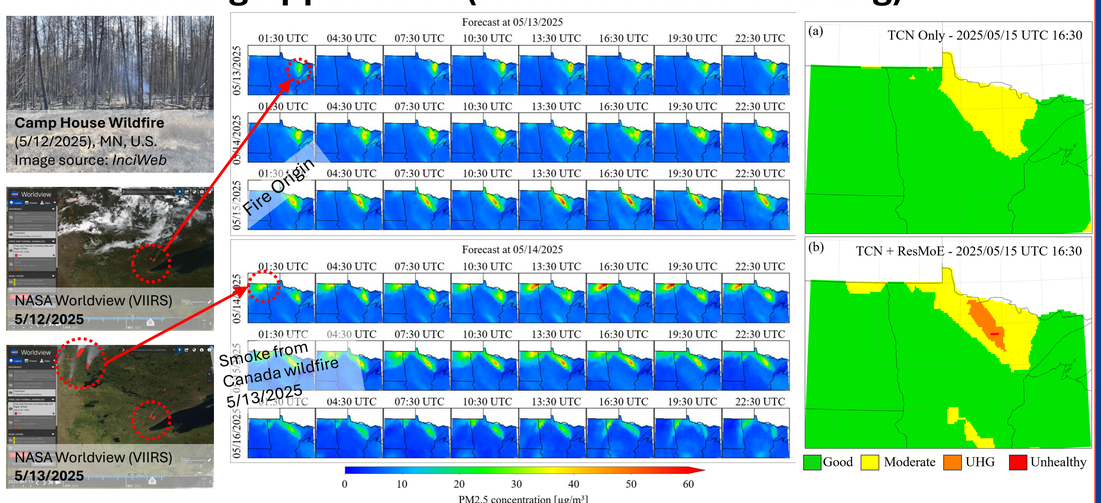
Experts Model Selection



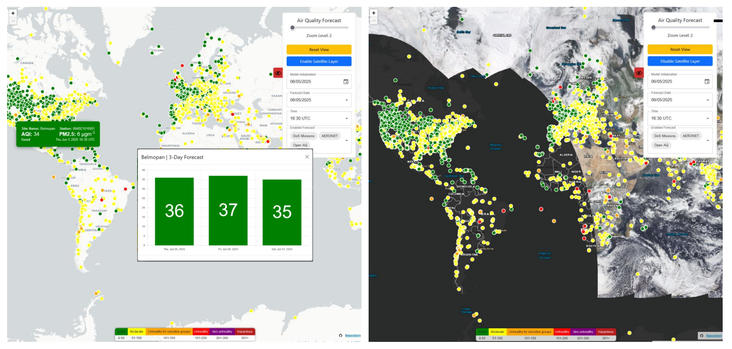
Expert	Physical Interpretation	Relative Active Strength	Representative Active Sites
1	Humid coastal stagnation; hygroscopic growth enhances scattering and AOD	+200–400% (moisture and sea-salt related variables strongly enhanced)	Accra (Ghana), Ho Chi Minh City (Vietnam)
2	Urban emission and photochemical reactions enhance secondary aerosols	+150–350% (SO_2/NO_2 precursors and sulfate highly activated)	Beijing (China), New Delhi (India), Cairo (Egypt)
3	Strong marine advection produces coarse sea-salt and transported dust	+180–300% (wind and sea-salt variables strongly positive)	Colombo (Sri Lanka), Manila (Philippines)
4	Transitional or mixed meteorology (dry–wet interface, seasonal burning)	+10–30% (broad but weak activation across most variables)	Manaus (Brazil), Kinshasa (DR Congo)

- The feature importance derived from a random-forest classifier retrained using the input features of the gating network
- The ResMoE model captures four key aerosol–meteorology regimes, linking distinct and transitional atmospheric patterns worldwide

Forecasting Application (Wildfire Plume Tracking)



Enabling Scalable Air Quality Forecast Distribution



Operational Forecasting for 270 U.S. Embassies, 1,800 AERONET sites, 5,000+ OpenAQ stations, and 800 African cities with population >100K

Conclusions

- Global deep learning framework integrates satellite-assimilated meteorology and aerosols for accurate PM2.5 forecasts
- Temporal and imbalance-aware modeling captures both routine and extreme events while ensuring physical interpretability
- Future focus includes regional adaptation, expert specialization, and improved real-time responsiveness

Acknowledgement

- This study was supported by the NASA HAQAST and the Satellite Needs Working Group (SNWG). We thank the HAQAST team and the broader air-quality community for their valuable guidance on both research and operational matters. GEOS data were obtained from the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center via the NASA Center for Climate Simulation's online portal.

Reference

- Seo J, Sayeed A, Park S, Kerekes J, Christel SM, Tran MT, Gupta, P. PM2.5 forecasting at U.S. Embassy and Consulates worldwide using NASA model powered by machine learning, *Earth Space Sci*. **12**, e2025EA004210 (2025). https://doi.org/10.1029/2025EA004210