$PM_{2.5}$ 3-day Forecasting at U.S. Embassy Locations using Machine Learning

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

in collaboration with Department of States, Air Quality Team

[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

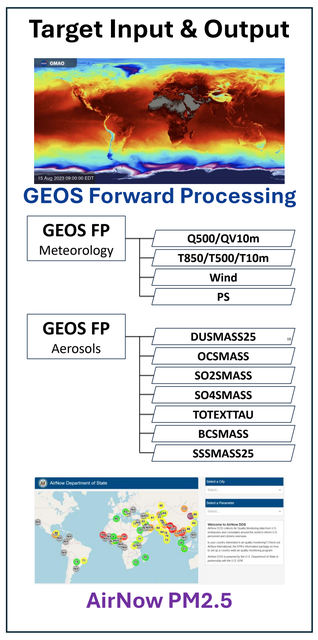
[4] Marshall Space Flight Center (MSFC) , National Aeronautics and Space Administration (NASA), Huntsville, AL, 35808, USA

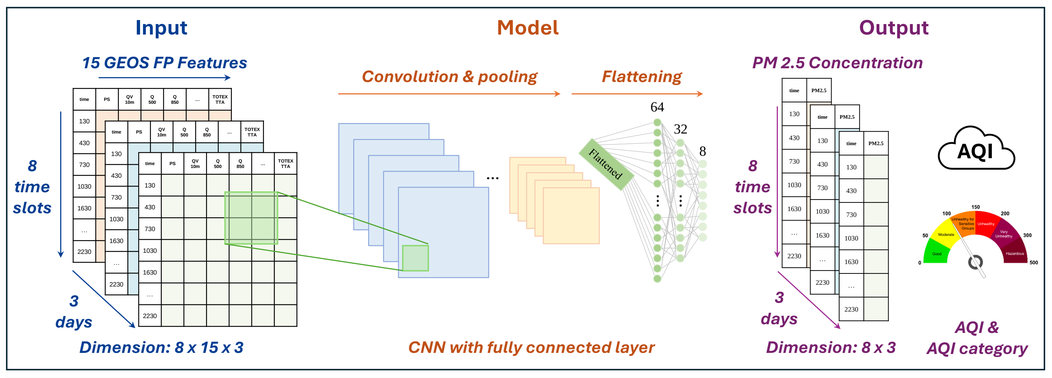
Forecasting API

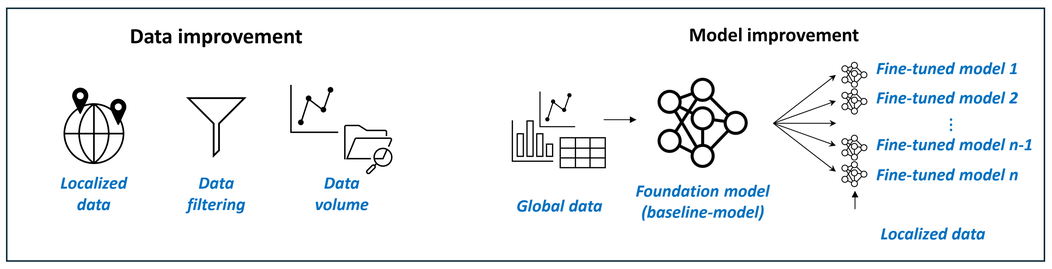
INTRODUCTION

- **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 over 80 embassies and consulates worldwide
  - **Forecasting Gap**: Despite extensive monitoring, the DoSAir program needs air quality forecasting, which is provided by this HAQAST project.
- **Objective**
  - Develop a machine learning-based forecast system named City Air quality foREcasting and analysis System (CARES), in collaboration with the U.S. Department of State and NASA's HAQAST
  - Utilize historical, satellite, and model data, including GEOS Forward Processing (FP) data as input to the CARES

DATA and METHODOLOGY



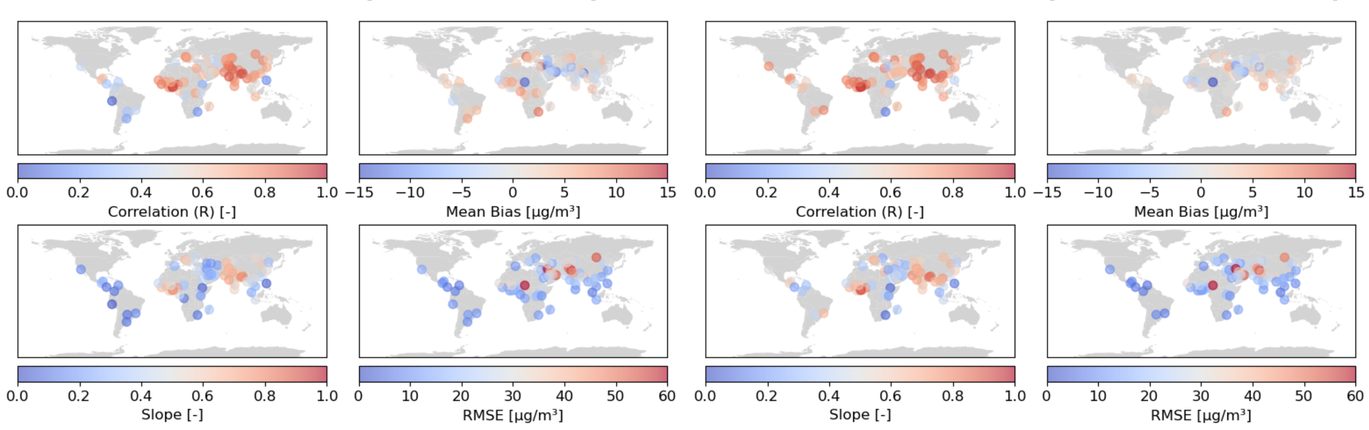




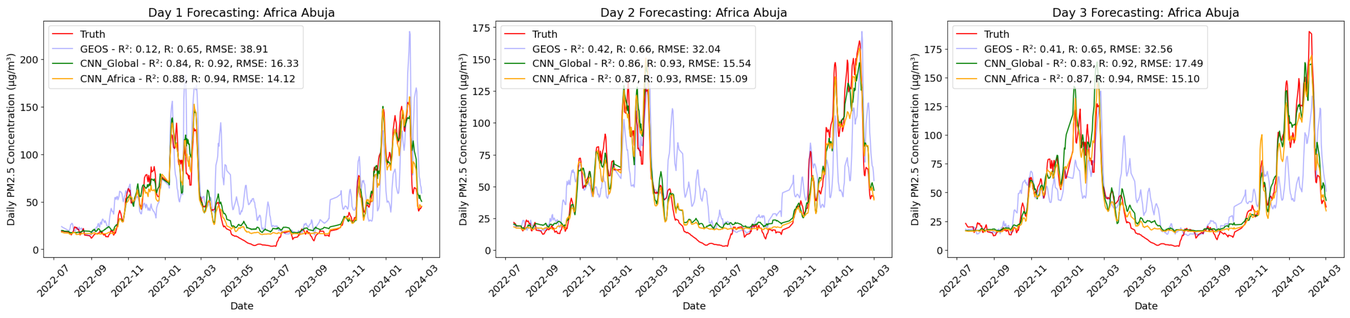
- **Dataset** (**Historical:** July 2022 to February 2024, **Independent Validation**: March 2024)
  - Input Data: Utilizes GEOS FP aerosol and meteorological parameters
  - Output Data: PM2.5 measurements from AirNow DoS
- **Model development**
  - Utilized a 10-fold cross-validation approach to train a Convolutional Neural Network (CNN) for capturing high nonlinearity between input and output data
  - Developed a baseline model using global data and fine-tuned it with local data to create continent-specific models

RESULTS and APPLICATIONS

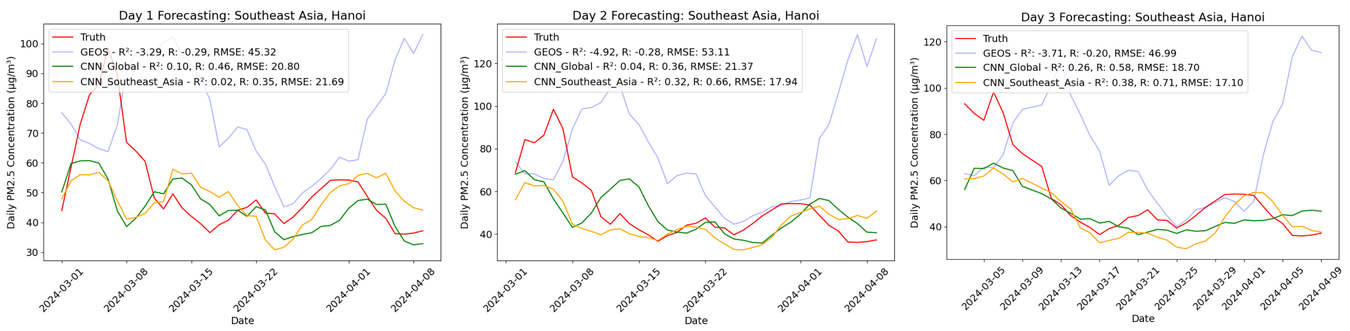
- **Daily PM2.5 Forecasting Statistical Comparison (historical data set)**



- **Daily PM2.5 Forecasting, Abuja (Nigeria, [9.041, 7.477]) - Historical Data**

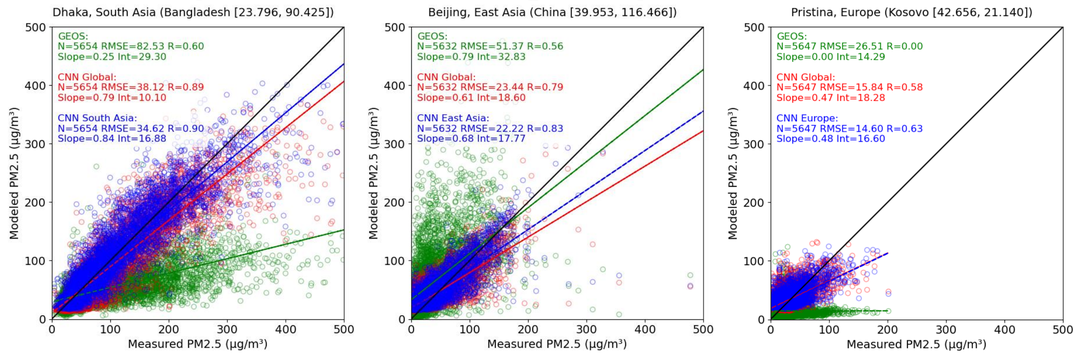


- **Daily PM2.5 Forecasting, Hanoi (Vietnam, [21.0219, 105.819]) – Independent Data**



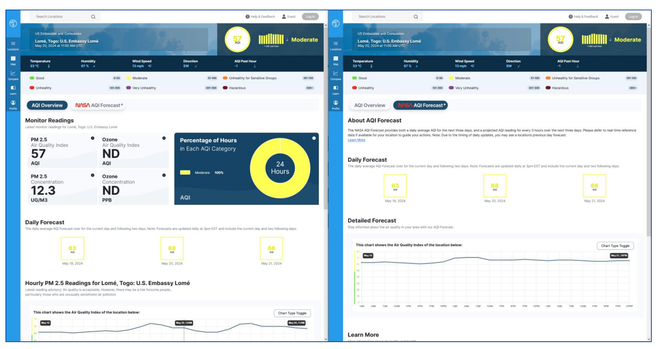
- **Baseline Model Accuracy**: High forecasting accuracy in West Africa, Central, and South Asia due to abundant data
- **Improved Accuracy**: Fine-tuned models showed increased three-day forecasting accuracy, with correlation R values improving by over 50% in South America and Europe
- **Independent Validation**: Fine-tuned models outperformed baseline models and GEOS estimates in independent validation

- **3 hourly PM2.5 Forecasting Comparison**



- **High performance (South Asia, West Africa)**: High forecasting accuracy due to sufficient data capturing available features
- **Moderate performance (Central Asia, East Asia)**: Slight deviation from global trends; accuracy improved with fine-tuning
- **Poor performance Group (Europe, South America)**: Limited data fails to reflect input-output relationship; minimal improvement with additional fine-tuning

- **ZephAir Application with NASA AQI Forecast**



ZephAir Web App

- Our team provides 3-hour PM2.5 forecasts for three days to 269 U.S. embassy and consulate sites for the DoS
- The ZephAir application offers the public 3-hour and daily PM2.5 concentration and AQI data for health-related decisions

- **Conclusions**
  - Developed a forecasting model using advanced CNN and GEOS FP input, achieving a maximum Correlation R of 0.93.
  - Fine-tuning overcame regional limitations, but some areas still have low accuracy due to data scarcity, requiring more data for improvement.