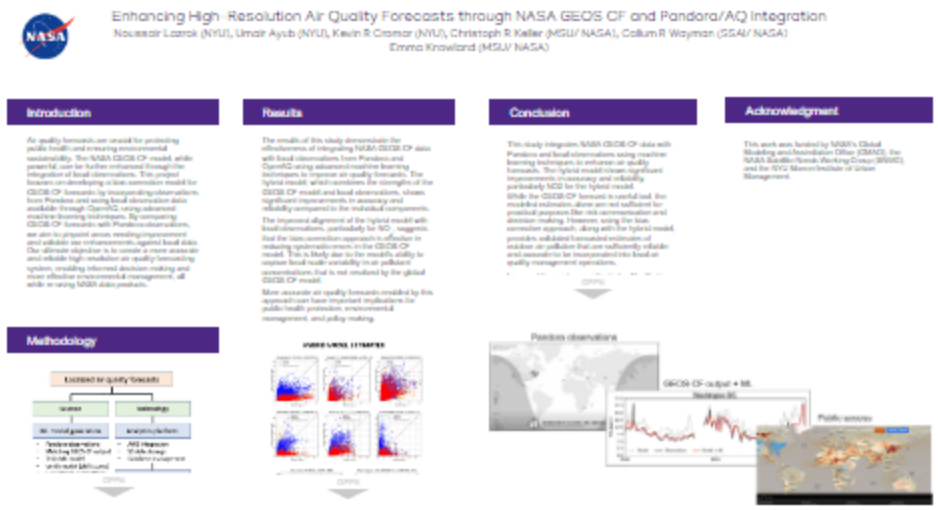
Enhancing High-Resolution Air Quality Forecasts through NASA GEOS CF and Pandora/AQ Integration



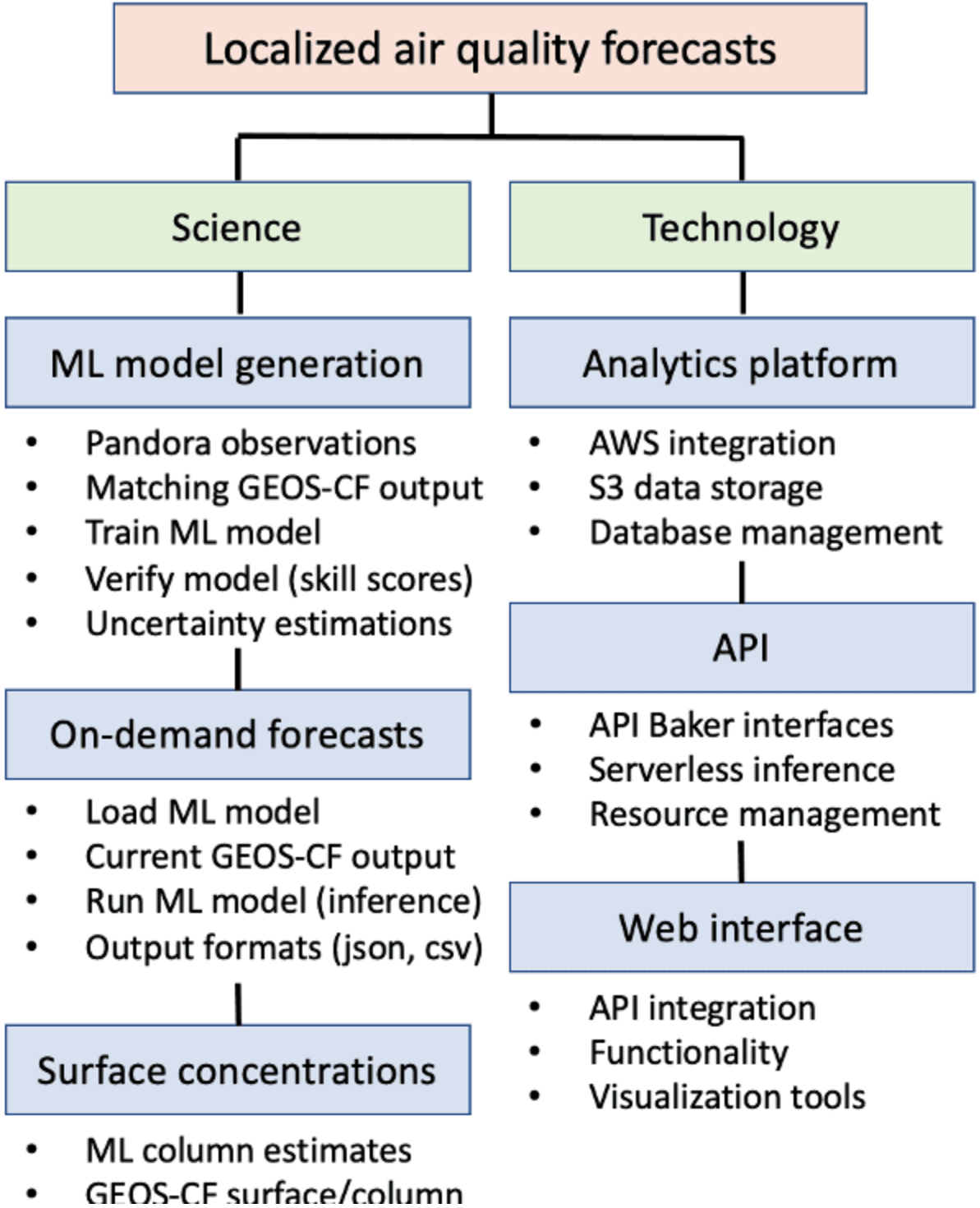
Noussair Lazrak (NYU), Umair Ayub (NYU), Kevin R Cromar (NYU), Christoph R Keller (MSU/ NASA), Callum R Wayman (SSAI/ NASA) and Emma Knowland (MSU/ NASA)

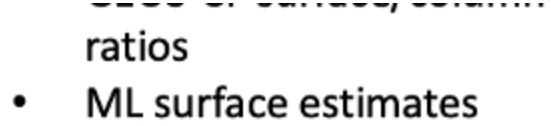
PRESENTED AT:

INTRODUCTION

Air quality forecasts are crucial for protecting public health and ensuring environmental sustainability. The NASA GEOS-CF model, while powerful, can be further enhanced through the integration of local observations. This project focuses on developing a bias-correction model for GEOS-CF forecasts by incorporating observations from Pandora and using local observation data available through OpenAQ, using advanced machine learning techniques. By comparing GEOS-CF forecasts with Pandora observations, we aim to pinpoint areas needing improvement and validate our enhancements against local data. Our ultimate objective is to create a more accurate and reliable high-resolution air quality forecasting system, enabling informed decision-making and more effective environmental management, all while re-using NASA data products.

METHODOLOGY


Localized air quality forecasts
Science
Technology
ML model generation
Pandora observations
Matching GEOS-CF output
Train ML model
Verify model (skill scores)
Uncertainty estimations
On-demand forecasts
Load ML model
Current GEOS-CF output
Run ML model (inference)
Output formats (json, csv)
Surface concentrations
ML column estimates
GEOS-CF surface/column
Analytics platform
AWS integration
S3 data storage
Database management
API
API Baker interfaces
Serverless inference
Resource management
Web interface
API integration
Functionality
Visualization tools



We use a hybrid model "a combination of NASA's GEOS-CF model and Pandora observations with machine learning techniques" to build a high-resolution forecasting model for air pollutants concentration. This approach leverages NASA data products to retrieve meteorological and atmospheric parameters, such as temperature, humidity, wind speed, and atmospheric composition. Pandora observations provide high-resolution measurements of atmospheric composition, including gases and aerosols.The trained machine learning model estimates near real-time air pollutants concentrations at selected locations globally. The model is validated against local monitoring data and uses SHAP Analysis to quantify contributing factors and track performance in extreme conditions. This hybrid approach aims to build a robust and high-resolution forecasting model for air pollutants concentration using existing NASA data products.

For the technology side, we leverage API Baker, a NASA and NAVTECA service, to utilize NASA's computing infrastructure and facilitate real-time data dissemination through Open Science Studio. The API transmits data to our localized forecast online tool.

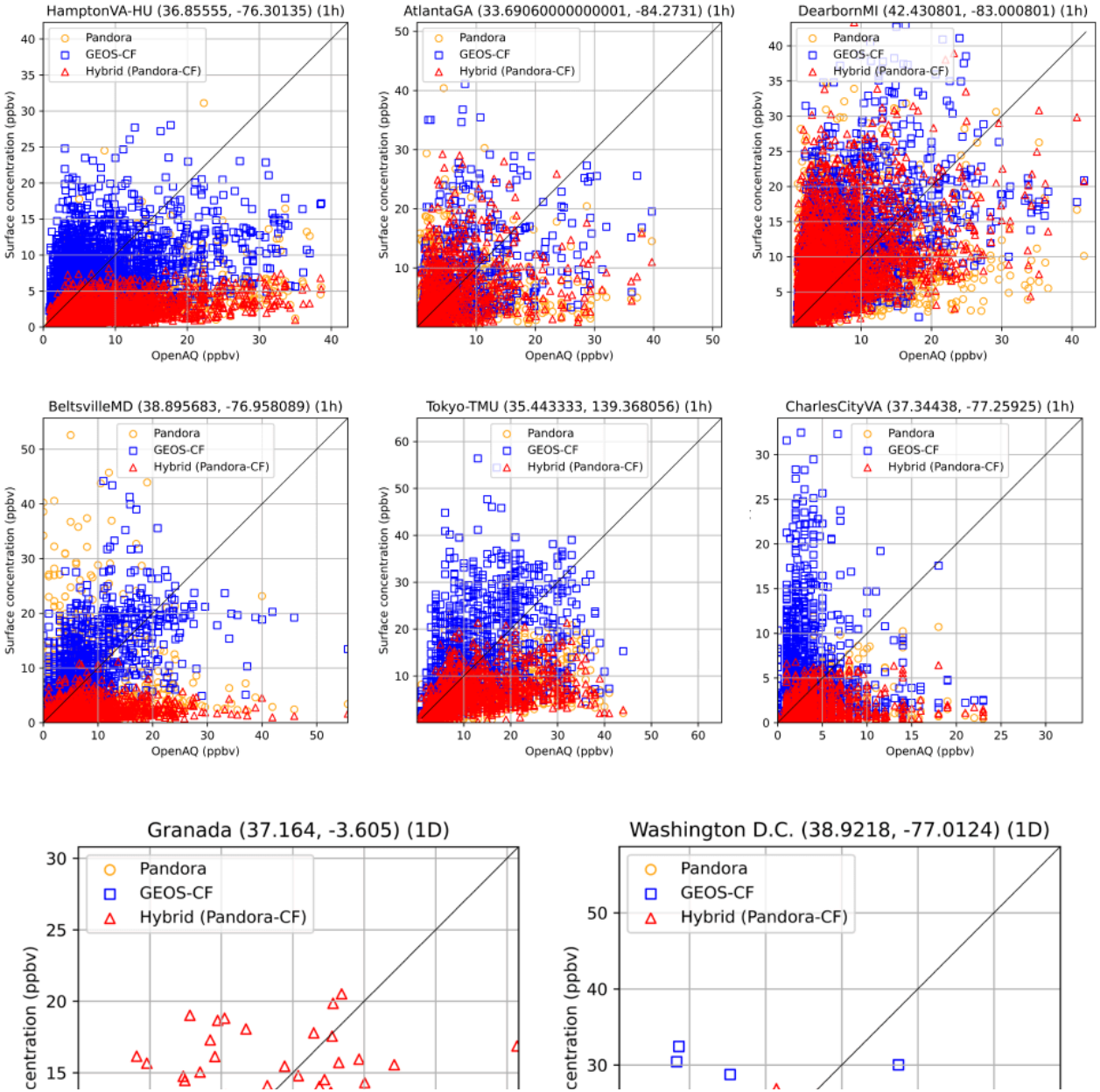
RESULTS

The results of this study demonstrate the effectiveness of integrating NASA GEOS-CF data with local observations from Pandora and OpenAQ using advanced machine learning techniques to improve air quality forecasts. The hybrid model, which combines the strengths of the GEOS-CF model and local observations, shows significant improvements in accuracy and reliability compared to the individual components.

The improved alignment of the hybrid model with local observations, particularly for NO_2, suggests that the bias-correction approach is effective in reducing systematic errors in the GEOS-CF model. This is likely due to the model's ability to capture local-scale variability in air pollutant concentrations that is not resolved by the global GEOS-CF model.

More accurate air quality forecasts enabled by this approach can have important implications for public health protection, environmental management, and policy-making.

HYBRID MODEL ESTIMATES



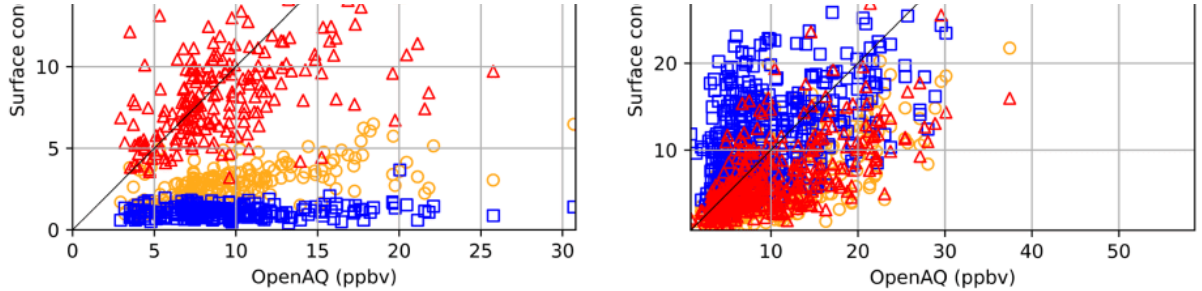


Figure 1: PANDORA and hybrid forecasts correlation with local observation

The scatter plots in Figure 1 demonstrate the relationship between observations from OpenAQ and surface concentrations estimated by Pandora, GEOS-CF, and the hybrid model. Each plot shows the distribution of data points, with the x-axis representing OpenAQ observations and the y-axis representing the estimated surface concentrations.

The hybrid model (red triangles) generally shows improved alignment with OpenAQ observations compared to the GEOS-CF (blue squares) and Pandora (yellow circles) estimates, indicating reduced bias and improved accuracy. This is particularly evident in NO_2 plots, where the hybrid model data points are more tightly clustered around the 1:1 line, suggesting better agreement with the reference observations.

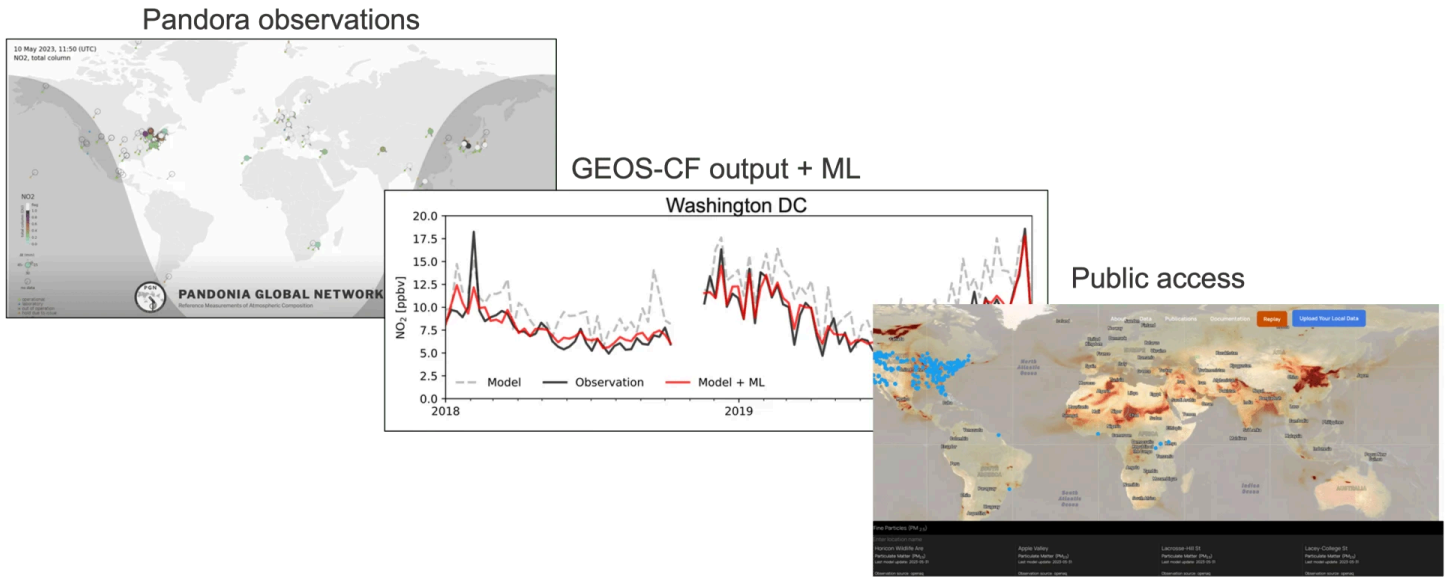
CONCLUSION

This study integrates NASA GEOS-CF data with Pandora and local observations using machine learning techniques to enhance air quality forecasts. The hybrid model shows significant improvements in accuracy and reliability, particularly NO2 for the hybrid model.
While the GEOS-CF forecast is useful tool, the modeled estimates alone are not sufficient for practical purposes like risk communication and decision-making. However, using the bias-correction approach, along with the hybrid model, provides validated forecasted estimates of outdoor air pollution that are sufficiently reliable and accurate to be incorporated into local air quality management operations.

Improved forecasts can mitigate health effects and support effective environmental management at the local level. Implementation efforts are already underway to make use of these improved forecasts for local risk communication purposes. Future research should expand this approach to other regions and pollutants, and incorporate additional data sources to enhance accuracy and reliability.

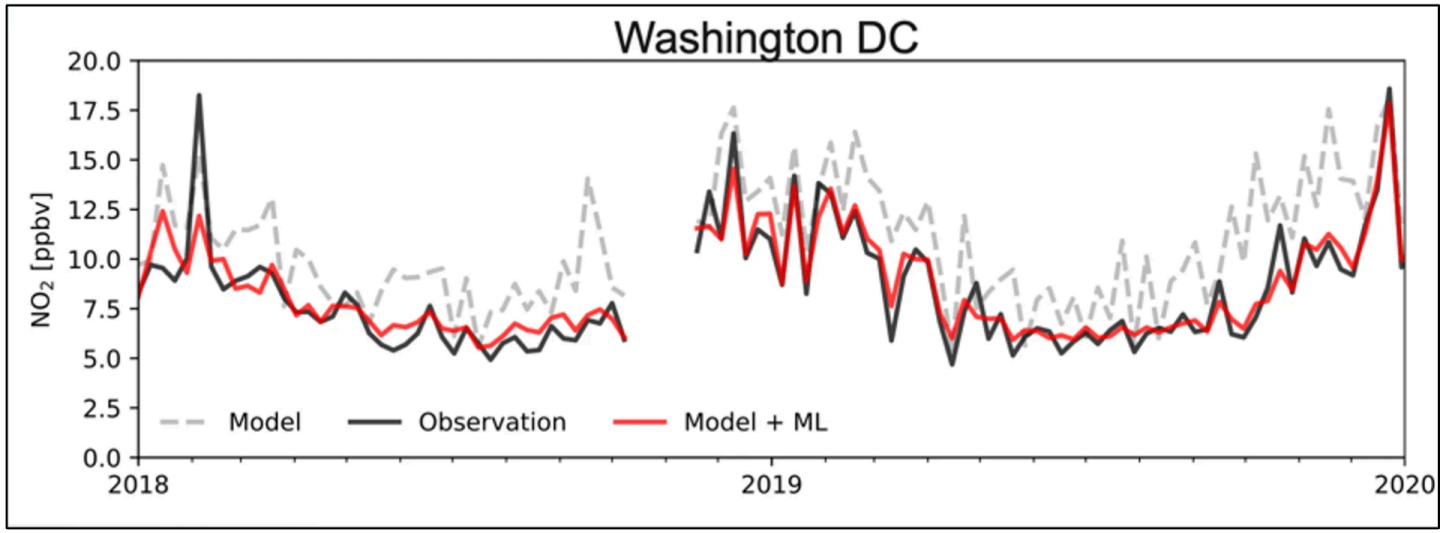
ACKNOWLEDGMENT

This work was funded by NASA's Global Modeling and Assimilation Office (GMAO), the NASA Satellite Needs Working Group (SNWG), and the NYU Marron Institute of Urban Management.

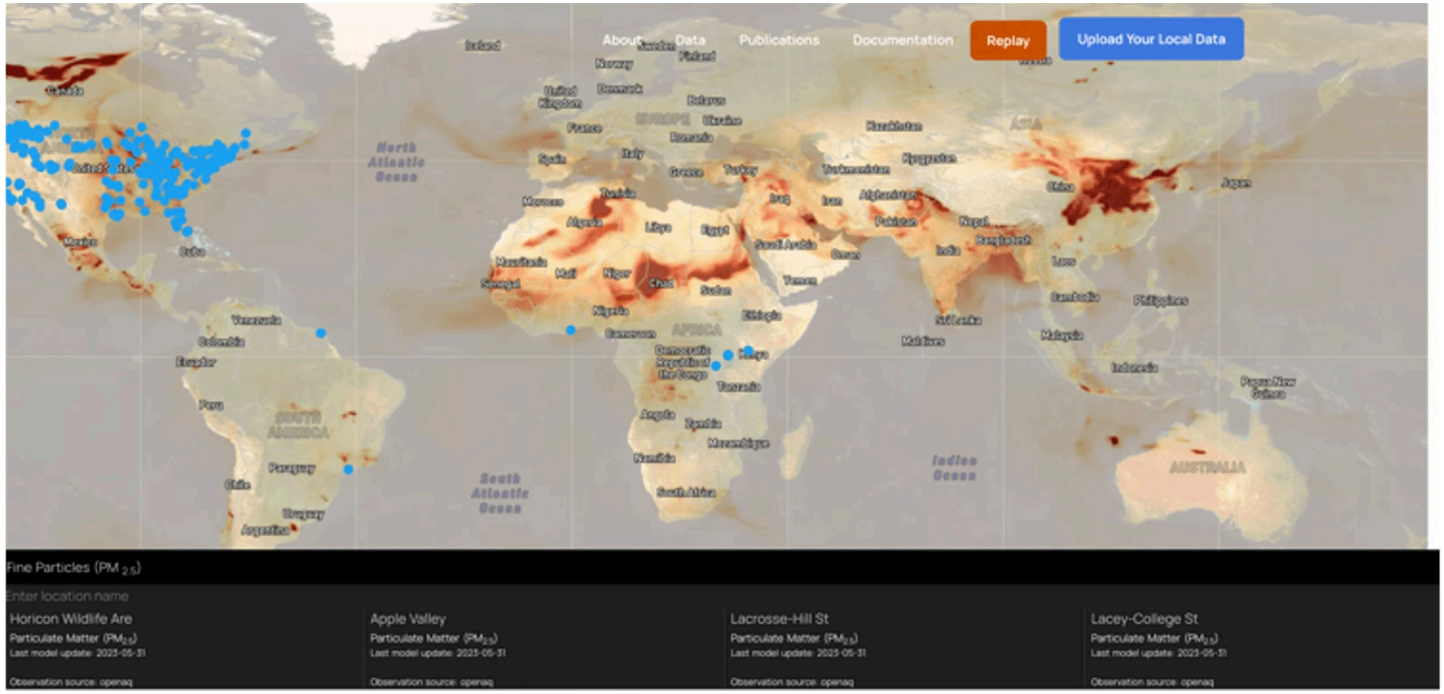


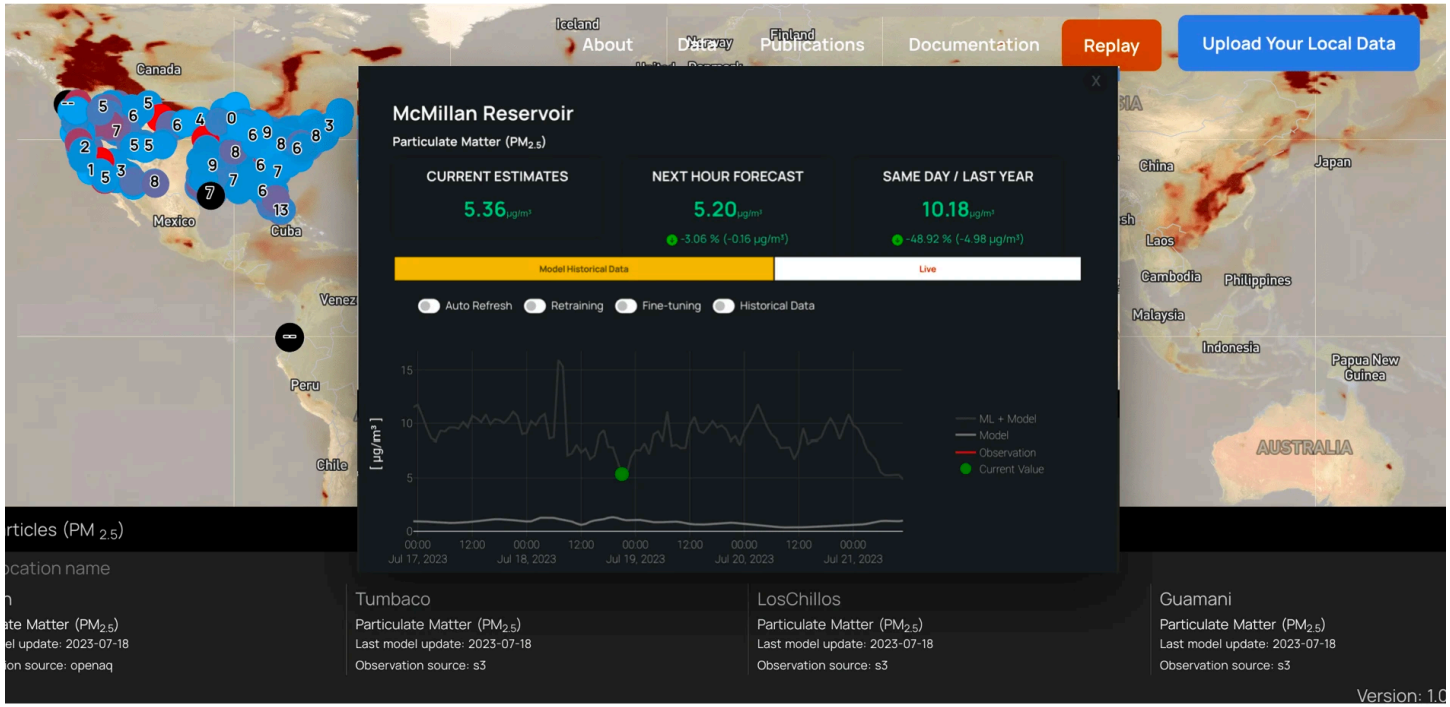
Pandora observations

GEOS-CF output + ML



Online tool for public access

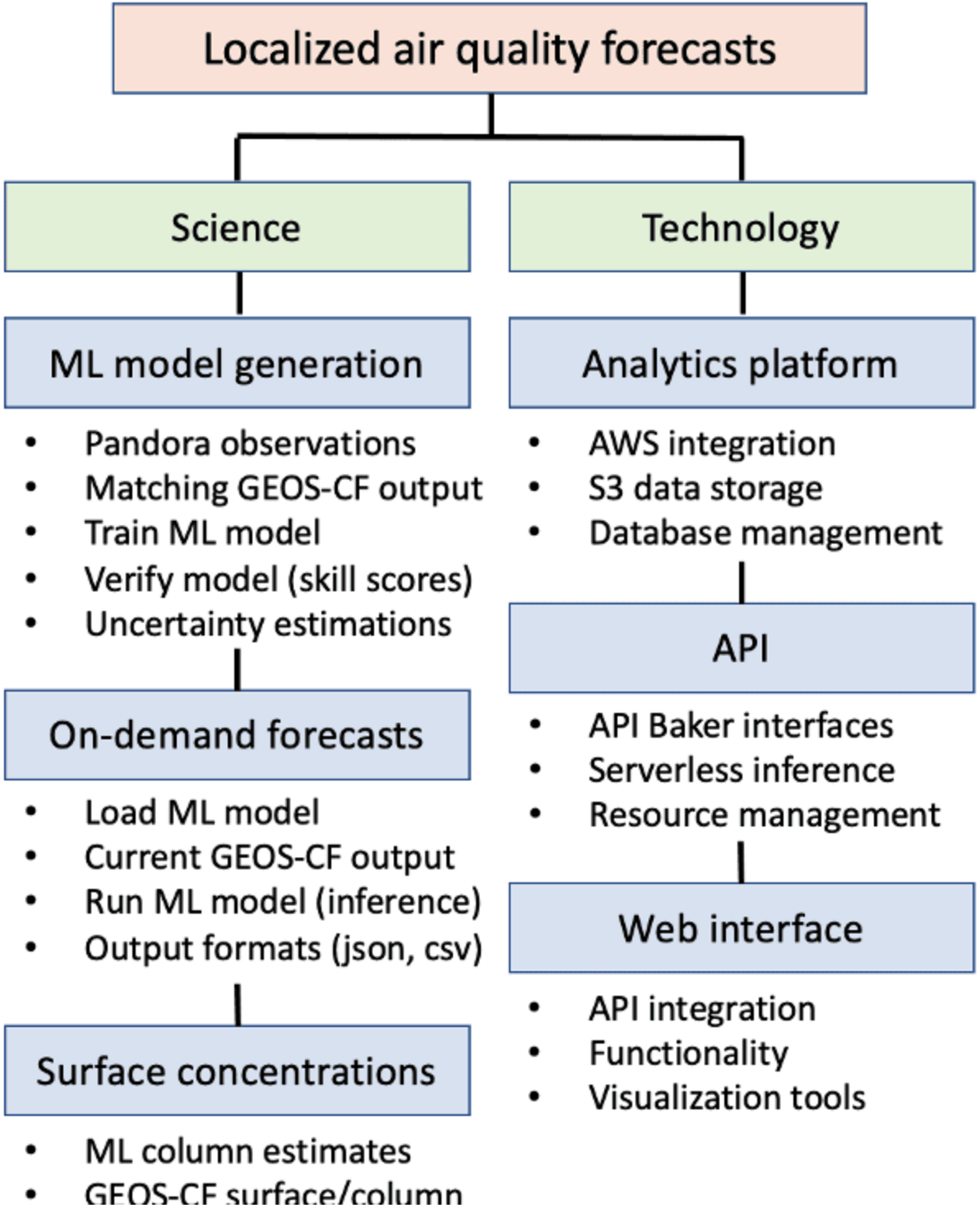



About
Publications
Documentation
Replay
Upload Your Local Data
McMillan Reservoir
Particulate Matter (PM2.5)
CURRENT ESTIMATES
5.36µg/m³
NEXT HOUR FORECAST
5.20µg/m³
-3.06 % (-0.16 µg/m³)
SAME DAY / LAST YEAR
10.18µg/m³
-48.92 % (-4.98 µg/m³)
Model Historical Data
Live
Auto Refresh
Retraining
Fine-tuning
Historical Data
ML + Model
Model
Observation
Current Value
Jul 17, 2023
Jul 18, 2023
Jul 19, 2023
Jul 20, 2023
Jul 21, 2023
Tumbaco
Particulate Matter (PM2.5)
Last model update: 2023-07-18
Observation source: s3
LosChillos
Particulate Matter (PM2.5)
Last model update: 2023-07-18
Observation source: s3
Guamani
Particulate Matter (PM2.5)
Last model update: 2023-07-18
Observation source: s3
Canada
Mexico
Cuba
Peru
Chile
China
Japan
Laos
Cambodia
Philippines
Malaysia
Indonesia
Papua New Guinea
AUSTRALIA

TRANSCRIPT

ABSTRACT

This work aims to improve air quality forecasts by integrating NASA's GEOS-CF model with local observations from Pandora and OpenAQ using advanced machine learning techniques. The hybrid model combines the strengths of the GEOS-CF model and local observations to provide more accurate and reliable high-resolution air quality forecasting. The results demonstrate significant improvements in accuracy and reliability, particularly for NO2. The hybrid model's improved alignment with local observations suggests that it can capture local-scale variability in air pollutant concentrations not resolved by the global GEOS-CF model. The enhanced forecasts can have important implications for public health protection, environmental management, and policy-making.


Localized air quality forecasts
Science
Technology
ML model generation
Pandora observations
Matching GEOS-CF output
Train ML model
Verify model (skill scores)
Uncertainty estimations
On-demand forecasts
Load ML model
Current GEOS-CF output
Run ML model (inference)
Output formats (json, csv)
Surface concentrations
ML column estimates
GEOS-CF surface/column
Analytics platform
AWS integration
S3 data storage
Database management
API
API Baker interfaces
Serverless inference
Resource management
Web interface
API integration
Functionality
Visualization tools

ratios

- ML surface estimates