FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 13, July, 2026, pp 177 – 191
DOI: https://doi.org/10.33003/fjs-2026-1013-5627

Machine Learning-Based Estimation of Surface $PM_{2.5}$ in Nigeria Using Integration of Satellite, Ground Observations and Reanalysis Data

*[1]Muawiya Sani, [2]Anas Houdou and [3]Rabia Salihu Sa'id

[1]Centre for Atmospheric Research, National Space Research and Development Agency, Kogi State University, Anyigba, Kogi State, Nigeria.
[2]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, United State.
[3]Department of Physics, Bayero University Kano, Kano State, Nigeria.

*Corresponding authors' email: msaniphysics2010@yahoo.com
ORCID iD: https://orcid.org/0009-0006-0682-4600

ABSTRACT

Accurate estimation of surface $PM_{2.5}$ across Nigeria remains challenging because satellite AOD and global reanalysis products alone cannot adequately represent near-surface particulate concentrations owing to complex aerosol-meteorology interactions, regional transport, and limited ground observations for calibration. This study addresses these limitations by developing a multi-source data fusion framework based on the Random Forest (RF) algorithm that integrates low-cost sensor measurements, satellite observations, atmospheric composition, meteorological variables, and reanalysis data to improve $PM_{2.5}$ estimation. Ground-based observations from Purple Air and Clarity sensors were combined with satellite-derived aerosol optical depth, atmospheric trace gases, meteorological variables, and MERRA-2 reanalysis products. The RF model was trained and evaluated using a spatial cross-validation (leave location out) framework and assessed using RMSE, MAE, coefficient of determination (R^2), Index of Agreement (IOA), and correlation coefficient. The RF model consistently outperformed MERRA-2 across all monitoring stations, yielding substantially lower prediction errors (RMSE: 14-35 µg m^{-3} versus 30-126 µg m^{-3}; MAE: 9-31 µg m^{-3} versus 18-89 µg m^{-3}) and stronger agreement with observations (IOA up to 0.68). Whereas MERRA-2 produced large negative R^2 values at several locations, the RF model achieved improved predictive performance, including a positive R^2 of 0.29 in Lagos and higher correlation coefficients across most stations. Feature importance analysis identified relative humidity as the dominant predictor, followed by MERRA-2 $PM_{2.5}$, O_3, NO_2, and AOD, highlighting the combined influence of aerosol hygroscopic growth, atmospheric chemistry, and regional transport. These findings demonstrate that multi-source machine learning data fusion substantially improves surface $PM_{2.5}$ estimation over Nigeria with scalable framework for air quality assessment.

Received: 19 Jun. 2026
Accepted: 23 Jun. 2026
Published: 14 Aug. 2026

Keywords: Aerosol Optical Depth (AOD), $PM_{2.5}$ Random Forest, Machine Learning, Air Quality, West Africa

INTRODUCTION

Air pollution is one of the most significant environmental health risks worldwide, with fine particulate matter ($PM_{2.5}$) identified as a major contributor to global morbidity and premature mortality (Ouarma *et al.,* 2020; Malings *et al.,* 2021; Awokola *et al.,* 2022; Suriano *et al.,* 2024; Kunjir *et al.,* 2025). $PM_{2.5}$ consists of particles with aerodynamic diameters smaller than 2.5 µm that can penetrate deep into the respiratory system and enter the bloodstream, triggering a wide range of adverse health outcomes including cardiovascular disease, chronic obstructive pulmonary disease, lung cancer, and respiratory infections (Yahaya *et al.,* 2023; Zhou *et al.,* 2023). According to World Health Organization, exposure to ambient air pollution is the reason for millions of premature deaths annually (WHO, 2021), making it one of the leading environmental risk factors globally. Large-scale epidemiological assessments such as the Global Burden of Disease Study have consistently demonstrated strong associations between long-term exposure to $PM_{2.5}$ and increased mortality rates across multiple regions of the world (Pendergrass *et al.,* 2022; Amooli *et al.,* 2024). In addition to human health impacts, particulate matter also influences climate processes, atmospheric chemistry, and ecosystem functioning (Geng *et al.,* 2021). Rapid urbanization, industrialization, and increased energy consumption in many developing regions have further intensified the challenges associated with air pollution management (Ouarma *et al.,* 2020; Kunjir *et al.,* 2025). Consequently, understanding the spatial and temporal distribution of $PM_{2.5}$ concentrations has become a central priority in atmospheric science, environmental policy, and public health research.

Nigeria represents one of the most challenging environments for surface $PM_{2.5}$ estimation because air quality is controlled by multiple interacting natural and anthropogenic emission sources that vary considerably across space and time. During the dry Harmattan season, large quantities of mineral dust originating from the Sahara Desert are transported over the country, substantially increasing aerosol loading, particularly across northern Nigeria (Prospero *et al.,* 2002; Horowitz *et al.,* 2017; Ogunjobi & Awoleye, 2019). In addition, seasonal biomass burning associated with agricultural land clearing and wildfires contributes large amounts of fine particulate matter and trace gases to the atmosphere (Liousse *et al.,* 2014;

Marais & Wiedinmyer, 2016; Zhang *et al.,* 2021). Rapid urbanization, increasing vehicular traffic, industrial emissions, power generation, open waste burning, and widespread reliance on biomass fuels for domestic cooking further exacerbate particulate pollution, especially within major urban centers such as Lagos, Kano, Port Harcourt, and Abuja (Aladodo et al., 2022; Amegah & Agyei-Mensah, 2017; B. I. Awokola et al., 2022; Zhou et al., 2023).

These diverse emission sources interact with complex meteorological processes, including seasonal rainfall, atmospheric stability, boundary layer dynamics, and humidity-induced aerosol hygroscopic growth, producing substantial spatial and temporal variability in surface $PM_{2.5}$ concentrations. However, Nigeria possesses only a limited number of long-term regulatory monitoring stations, resulting in sparse observational data for model development and validation (Alani *et al.,* 2019; Petkova *et al.,* 2013; Aladodo *et al.,* 2022). Consequently, most air quality assessments rely on satellite observations, atmospheric reanalysis products, or chemical transport models. While these global products provide valuable regional information, they often inadequately represent localized emission sources, urban pollution hotspots, and the complex aerosol mixtures characteristic of Nigeria because of their coarse spatial resolution and generalized parameterizations. Therefore, $PM_{2.5}$ estimation approaches developed for Europe, North America, or East Asia cannot be directly transferred to Nigeria without incorporating region-specific observational data and locally relevant predictors.

To overcome limitations associated with sparse ground monitoring networks, satellite observations and atmospheric reanalysis datasets have increasingly been used to estimate air pollutant concentrations over large spatial domains (Gelaro *et al.,* 2017). Satellite instruments provide global measurements of atmospheric properties such as aerosol optical depth (AOD), which can serve as an important proxy for particulate matter concentrations when combined with appropriate modeling frameworks (Van Donkelaar *et al.,* 2006; Gupta *et al.,* 2013; Van Donkelaar *et al.,* 2016). In addition to satellite observations, atmospheric reanalysis products integrate satellite data, meteorological observations, and chemical transport models to generate continuous global datasets describing atmospheric composition (Randles *et al.,* 2017). One widely used dataset is the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) produced by NASA. MERRA-2 assimilates satellite observations into a comprehensive atmospheric modeling system to provide global estimates of aerosol properties and particulate matter concentrations (Buchard *et al.,* 2017; Gelaro *et al.,* 2017).

Although satellite observations and atmospheric reanalysis products have greatly improved regional air quality assessment, their direct application to Nigeria remains challenging. Satellite AOD represents the total atmospheric aerosol column rather than near-surface $PM_{2.5}$ concentrations, and the relationship between AOD and surface particulate matter is strongly influenced by aerosol composition, relative humidity, planetary boundary layer height, and vertical aerosol distribution (Hammer *et al.,* 2020; Van Donkelaar *et al.,* 2016). Likewise, the coarse spatial resolution of global reanalysis products such as MERRA-2 often fails to resolve localized emission sources, roadside pollution, industrial activities, and rapidly changing urban environments that characterize many Nigerian cities (Buchard *et al.,* 2017; Gelaro *et al.,* 2017). Furthermore, global models are not specifically calibrated for Nigeria's unique combination of Harmattan dust, biomass-burning aerosols, and anthropogenic emissions, which introduces additional uncertainty into surface $PM_{2.5}$ estimates (Idris et al., 2025). These limitations highlight the need for regionally adapted modelling frameworks capable of integrating multiple observational datasets to improve $PM_{2.5}$ prediction accuracy. To address these limitations, this study develops a multi-source machine learning framework that integrates satellite observations, atmospheric reanalysis products, meteorological variables, atmospheric trace gases, and low-cost sensor measurements within a Random Forest model to improve surface $PM_{2.5}$ estimation across Nigeria.

Unlike conventional statistical approaches, machine learning algorithms can capture the complex nonlinear relationships among aerosol loading, meteorological conditions, atmospheric chemistry, land surface characteristics, and surface $PM_{2.5}$ concentrations (Chen *et al.,* 2018; Di *et al.,* 2019). This capability is particularly important in Nigeria, where $PM_{2.5}$ levels are influenced by seasonally varying Harmattan dust, biomass-burning aerosols, traffic emissions, industrial activities, and rapidly changing meteorological conditions. Among machine learning algorithms, Random Forest (RF) has demonstrated excellent performance because of its robustness to noisy datasets, ability to model nonlinear interactions, resistance to overfitting, and capacity to handle high-dimensional environmental predictors (Breiman, 2001; Chen *et al.,* 2018). Previous studies have successfully applied RF and other machine learning methods to estimate $PM_{2.5}$ in North America, Europe, and Asia using combinations of satellite aerosol optical depth, meteorological variables, land-use characteristics, and ground (Yan *et al.,* 2017; Di *et al.,* 2019; Vignesh *et al.,* 2023; Raheja *et al.,* 2023; Mathew *et al.,* 2023; Abuouelezz *et al.,* 2025; Muhammad & Ahmad, 2025). However, relatively few studies have evaluated such integrated frameworks in Nigeria, where aerosol composition, emission characteristics, and monitoring infrastructure differ substantially from those in regions where most existing models have been developed.

Despite recent advances in machine learning-based $PM_{2.5}$ estimation, limited research has developed and evaluated integrated multi-source prediction frameworks specifically for Nigeria using low-cost sensor observations, satellite aerosol products, atmospheric composition variables, meteorological parameters, and reanalysis datasets. Consequently, the applicability of globally developed $PM_{2.5}$ estimation models to Nigeria remains uncertain. This study addresses this gap by developing a Random Forest framework that integrates ground-based measurements from Purple Air and Clarity sensors with satellite-derived aerosol optical depth, atmospheric trace gases, meteorological variables, and MERRA-2 reanalysis data to estimate surface $PM_{2.5}$ concentrations across multiple monitoring stations in Nigeria. Model performance is evaluated using group K-fold cross validation framework leaving one location out and compared with MERRA-2 reanalysis estimates to determine whether multi-source machine learning can improve $PM_{2.5}$ prediction in a data-sparse environment. Specifically, the study aims to: (i) develop a Random Forest model for estimating surface $PM_{2.5}$ concentrations in Nigeria; (ii) compare Random Forest predictions with MERRA-2 $PM_{2.5}$ estimates in Nigeria; (iii) evaluate model performance across multiple monitoring stations using standard statistical metrics; and (iv) examine the temporal variability of predicted and observed $PM_{2.5}$ concentrations.

MATERIALS AND METHODS

Study Area

This study focuses on monitoring locations distributed across regions of West Africa, an area characterized by diverse climatic conditions, rapidly expanding urban centers, and increasing anthropogenic emissions. West Africa experiences complex atmospheric dynamics driven by seasonal monsoon circulation, Saharan dust transport, biomass burning, and urban emissions (Jenkins *et al.,* 2023). These processes contribute to significant spatial and temporal variability in particulate matter concentrations across the region(Zhang *et al.,* 2021; Awokola *et al.,* 2022). Rapid urbanization and population growth in many West African cities have resulted in increasing emissions from transportation, industrial activities, open waste burning, and residential fuel combustion (Liousse *et al.,* 2014; Malings *et al.,* 2020; Awokola *et al.,* 2022). At the same time, natural sources such as desert dust originating from the Sahara desert frequently influence regional air quality, particularly during dry seasons (Jenkins *et al.,* 2023). These factors create a highly variable atmospheric environment that presents challenges for accurately estimating particulate matter concentrations using coarse-resolution atmospheric datasets alone. Understanding the distribution and drivers of $PM_{2.5}$ in West Africa is therefore essential for assessing population exposure and supporting air quality management strategies in the region (Liousse *et al.,* 2014). However, the limited availability of long-term ground monitoring networks has hindered comprehensive assessments of particulate matter concentrations across the region(Ogunjobi & Awoleye, 2019; Sharafa *et al.,* 2020; Aladodo *et al.,* 2022). Nigeria for example (Figure 1).

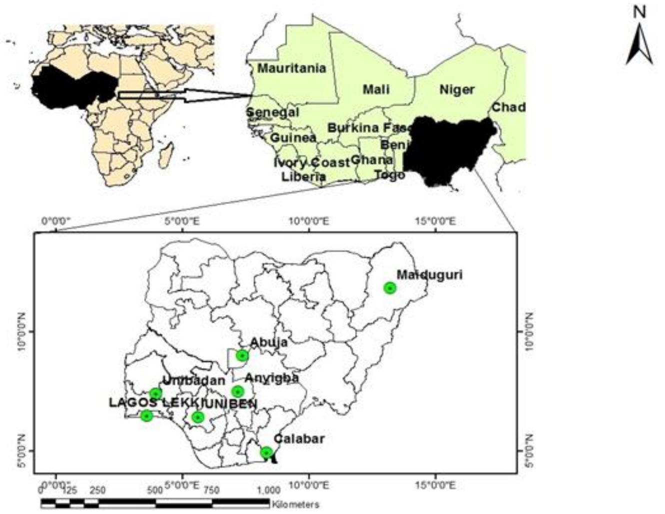


Figure 1: Stations Considered in the Study Over Nigeria, Representing a Region in West Africa

Ground-Based $PM_{2.5}$ Observations

Ground-based particulate matter observations were obtained from 7 monitoring stations across Nigeria using Purple Air and Clarity sensors. These low-cost optical sensors estimate $PM_{2.5}$ concentrations based on light-scattering measurements, which are inherently sensitive to ambient environmental conditions, particularly relative humidity (Ardon-Dryer *et al.,* 2020). The dataset used in this study is $PM_{2.5}$ concentrations (Table 1). Since no additional post-deployment calibration was applied beyond manufacturer-provided corrections, the measurements reflect factory-calibrated outputs. However, the influence of relative humidity on particle hygroscopic growth and optical scattering can be represented as:

$$PM_{2.5}^{obs} = f(PM_{dry}RH) \qquad (1)$$

where $PM_{2.5}^{obs}$ denotes the measured concentration and PM_{dry} represents the dry aerosol mass. Quality control procedures were implemented to ensure data integrity, including the removal of missing and non-physical observations, and records with missing $PM_{2.5}$ values were excluded to maintain consistency in model development (Ardon-Dryer *et al.,* 2020). Low-cost air quality sensors have increasingly been used to complement sparse regulatory monitoring networks, particularly in data-limited regions (Nobell *et al.,* 2023). Although such sensors may exhibit systematic biases relative to reference-grade instruments, previous studies (Castell *et al.,* 2017; Kelly *et al.,* 2017; Nobell *et al.,* 2023; Raheja *et al.,* 2023) demonstrate that manufacturer-calibrated data can still provide valuable insights for air quality assessment when appropriate quality control procedures are applied.

Atmospheric Reanalysis Data

To complement the ground-based observations, this study utilizes particulate matter estimates from the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) reanalysis dataset produced by NASA. The MERRA-2 is a long-term global reanalysis developed by NASA's Goddard Earth Observing System and Global Modeling and Assimilation Office, providing meteorological and aerosol parameters from 1980 to present (Gelaro *et al.,* 2017). It employs the GEOS atmospheric model with resolution of approximately 50 × 65 km (0.5° latitude and 0.625° longitude) (Table 1) and 72 vertical layers extending from the surface to 0.01 hPa (Colarco *et al.,* 2010; Rienecker *et al.,* 2011; Sayeed *et al.,* 2022). Atmospheric conditions are updated using a three-dimensional variational data assimilation system (3D-Var) based on the Grid point Statistical Interpolation scheme with a 6-hour cycle, incorporating observations through incremental analysis update (Rienecker *et al.,* 2011). MERRA-2 assimilates bias-corrected aerosol optical depth (AOD) at 550 nm from multiple satellite and ground-based sources, including MODIS, MISR, AVHRR, and AERONET, though data availability varies over time (Randles *et al.,* 2017). Aerosol species dust, sea salt, black carbon (BC), organic carbon (OC), and sulfate (SO_4) are simulated using the GOCART model coupled with GEOS (Colarco *et al.,* 2010). Surface $PM_{2.5}$ is then diagnostically derived as the sum of these components, expressed in Equation 2: (Buchard *et al.,* 2015; Randles *et al.,* 2017).

$$PM_{2.5} = DUST_{2.5} + SS_{2.5} + BC_{2.5} + OC_{2.5} + 1.375 \text{ X } SO_4 \quad (2)$$

where the sulfate term is scaled to account for ammonium sulfate formation and total $PM_{2.5}$ represents the sum of dust, sea salt, black carbon, organic carbon, and sulfate aerosols (Randles *et al.,* 2017).

Table 1: The General Information of Input Datasets Used in $PM_{2.5}$ Prediction Models

Category	Dataset	Variable	Unit	Resolution
Ground-$PM_{2.5}$ Observations	Purple Air, Clarity	$PM_{2.5}$ (target variable)	ug/m3	Daily
Satellites	MODIS AQUA, OMI	AOD, NO_2, SO_2, O_3	DU, mol/cm3	1^0, 0.25^0 daily
Meteorological data	MERRA-2 reanalysis, GP Models	Temperature, Relative Humidity, Solar Radiation, Wind Speed, Rainfall	^{0}C, %, w/m^2, m/s, mm/day	$0.5\text{x}0.625^0$, 1^0 daily

Methodology

Data Preprocessing

The $PM_{2.5}$ prediction framework integrated multi-source datasets comprising ground-based observations from Purple Air and Clarity (daily $PM_{2.5}$, μg m^{-3}), satellite products from MODIS AQUA and OMI (AOD and trace gases NO_2, SO_2, O_3 in DU and mol cm^{-2} at ~10 km and 0.25° daily resolution), and meteorological variables (temperature, relative humidity, solar radiation, rainfall, and wind speed) obtained from ground sensors, MERRA-2, and GP models (0.5° × 0.625°) as shown in Table 1. Data treatment involved quality control, where extreme non-physical $PM_{2.5}$ values (>500 μg m^{-3}) from low-cost sensors were removed to reduce bias, stabilize regression and machine learning training. Short data gaps (≤4 days) across all datasets were filled using local mean imputation to preserve temporal continuity while minimizing distortion of variability, whereas longer gaps were excluded in line with World Meteorological Organization (WMO) and U.S. Environmental Protection Agency (EPA) recommendations. A structured preprocessing workflow excluded PM_1 and PM_{10} to prevent data leakage, treated invalid atmospheric composition data, and to ensure all features contribute equally to machine learning model Min-Max normalization (equation 3) was applied for numerical stability, it is important because it prevents features with larger magnitudes from dominating others. Feature engineering incorporated physically meaningful predictors, including AOD-relative humidity interaction terms, seasonal cyclic components, autoregressive $PM_{2.5}$ lags (1-, 3-, and 7-day), and MERRA-2 $PM_{2.5}$ for hybrid modeling. Finally, the dataset was partitioned using a time-ordered 80/20 train test split to prevent temporal leakage and ensure strong generalization, supported by exploratory data analysis like statistical summaries, correlations, and spatial visualization to validate data consistency and predictor-response relationships prior to model development.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

where x_{new} is the normalized feature, x_{min}, minimum value of the feature and x_{max} maximum value of the feature.

Autoregressive predictors (e.g., lag-1, lag-3, and lag-7 $PM_{2.5}$) were computed exclusively from historical observations preceding each prediction time step. To avoid information leakage, lag variables were generated without using future observations, and model evaluation was conducted using a leave-one-location-out Group K-Fold cross-validation framework in which the validation station was completely excluded from model training. Records lacking sufficient historical observations to compute the required lag values were omitted from model development rather than being filled using information from subsequent observations.

Overview of the Random Forest (RF) Modelling Framework

The modelling framework integrates satellite Aerosol Optical Depth (AOD), low-cost sensor data (Purple Air/Clarity), and meteorological variables (e.g., relative humidity, temperature, wind speed, atmospheric gases) to estimate surface $PM_{2.5}$ concentrations across the stations. The Random Forests (RF) modelling technique is a non-parametric ensemble method that captures the aerosol-meteorology system, allowing it to get the nonlinear relationships between predictors and $PM_{2.5}$. RF constructs an ensemble of decision trees using bootstrap resampling, with random predictor selection at each split to reduce overfitting (Breiman, 2001).

The RF prediction for an input x is:

$$\hat{y}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \quad (4)$$

where $T_b(x)$ is the prediction from tree b and B is the number of trees.

In this study, RF was trained using equation (4) a cross-validation strategy to preserve temporal structure. RF is particularly effective for: nonlinear AOD-$PM_{2.5}$ relationships, humidity-dependent aerosol growth and seasonal dust and biomass-burning dynamics common in West Africa.

The Random Forest model was employed using the scikit-learn machine learning library in Python (Pedregosa *et al.,* 2011) and Hyperparameters were retained at their default scikit-learn settings except where explicitly stated. The model configuration included multiple decision trees, bootstrap aggregation, and feature sampling to improve generalization performance. Predictor variables included atmospheric and environmental variables such as (MODIS satellite AOD, Relative Humidity, Temperature, Wind Speed, Atmospheric gases, MERRA-2 $PM_{2.5}$ estimates and other relevant predictors derived from the dataset.

Cross-Validation Strategy

To ensure better model evaluation and avoid spatial data leakage, a grouped cross-validation strategy was implemented using the Group K Fold approach (Roberts *et al.,* 2017; Ploton *et al.,* 2020). In this framework, all models were used with their default parameters and observations belonging to the same monitoring station are grouped together so that all data from a given station are either included in the training set or the testing set within each fold. This approach is particularly important for environmental datasets that exhibit spatial dependence, as traditional random cross-validation might lead to too optimistic model performance estimates when nearby observations are included in both training and testing subsets

approach (Roberts *et al.,* 2017; Ploton *et al.,* 2020). The grouped cross-validation procedure was implemented using seven folds, allowing the model to be trained and evaluated across multiple station combinations while maintaining independence between training and testing data.
Hence, Model evaluation followed a multi-stage validation framework. During model development, a chronological 80/20 split and Time-Series-Split were used to preserve temporal ordering during model training and hyperparameter optimization. The final model evaluation was performed using a leave-one-location-out Group-K-Fold cross-validation approach, in which one monitoring station was excluded entirely for validation while the remaining stations were used for training. This strategy was selected to assess the spatial generalizability of the proposed framework to unseen monitoring locations. Hyperparameter optimization was performed prior to the final Group-K-Fold evaluation; therefore, a fully nested cross-validation framework was not implemented.

Modeling Evaluation

Model performance was evaluated using root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), Pearson correlation coefficient, and the index of agreement (IOA) using equations 5, 6, 7 and 8 respectively (Wilks, 2006; Willmott & Matsuura, 2005). These metrics provide complementary information regarding model accuracy, error magnitude, and agreement between predicted and observed $PM_{2.5}$ concentrations (Willmott, 1981). While MAE and RMSE quantify the magnitude of prediction errors, R^2 and the correlation coefficient measure the strength of the relationship between predicted and observed values. The index of agreement provides an additional measure of model performance by evaluating how closely predicted values match observations relative to the variability in the observed dataset. Using multiple performance metrics allows for a comprehensive evaluation of the Random Forest model and facilitates comparison with the baseline MERRA-2 $PM_{2.5}$ estimates as widely used in air quality modeling studies (Emery *et al.,* 2017).

i. Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (5)$$

ii. Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \quad (6)$$

iii. Correlation Coefficient (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (7)$$

iv. index of agreement (IOA):

$$d = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (8)$$

v. Seasonal Stability Index (SSI):

$$SSI = \frac{|RMSE_{Dry} - RMSE_{Wet}|}{RMSE_{Annual}} \quad (9)$$

where P_i is the predicted values, O_i is the observed values and n is the total number of measurements at a particular site. In this way, these metrics offer a comprehensive evaluation of model performance by quantifying prediction accuracy, variability explained, and systematic error, which is essential for assessing the reliability of AOD-$PM_{2.5}$ machine learning models and their consistency with ground-based observations and reanalysis products (Willmott & Matsuura, 2005).

RESULTS AND DISCUSSION

Descriptive Statistics of Predictor Variables

Table 2. summarizes the descriptive statistics of the variables used to develop the Random Forest (RF) model for estimating surface $PM_{2.5}$ concentrations. A total of 9,455 observations were used in the modeling framework, providing a sufficiently large dataset for machine learning analysis. The observed atmospheric $PM_{2.5}$ concentration ($PM_{2.5}$_atm) shows substantial variability across the study region, with values ranging from 0 to 332.2 μg m^{-3} and an average concentration of 32.60 μg m^{-3}. These values exceed the annual guideline limits recommended by the World Health Organization, highlighting the potential public health implications of particulate pollution across the study region. Aerosol optical depth (AOD) values exhibit considerable variability, with a mean of 0.46 and a maximum of 3.5, reflecting episodic aerosol loading events such as desert dust transport and biomass burning. Meteorological variables also show significant variability across the dataset. Solar radiation ranged between 42.84 and 324.71 W m^{-2}, while relative humidity varied from 4% to 99.5%, indicating diverse atmospheric conditions influencing aerosol formation and removal processes. Rainfall values ranged from 0 to 220.61 mm, suggesting that wet deposition processes may play an important role in modulating particulate matter concentrations. Atmospheric chemical species also exhibited notable variability. Nitrogen dioxide (NO_2) and ozone (O_3) concentrations varied widely, indicating the influence of urban emissions and photochemical processes on air quality. Wind speed values ranged from 0.96 to 10.89 m s^{-1}, which can significantly affect pollutant dispersion and atmospheric mixing. The reanalysis-derived $PM_{2.5}$ estimates from the MERRA-2 show a mean value of 50.72 μg m^{-3}, which is considerably higher than the observed mean $PM_{2.5}$ concentration suggesting potential biases associated with the coarse spatial resolution of the dataset. Hence, the wide range of values observed in the predictor variables provides sufficient variability for training the Random Forest model.

Table 2: Descriptive Statistics of Datasets Used in the Random Forest Model

Variable	Mean	Std	Min	25%	Median	75%	Max
$PM_{2.5}$_atm (μg m^{-3})	32.60	26.79	0	13.7	26.8	43.3	332.2
AOD (Unitless)	0.46	0.32	-0.016	0.266	0.367	0.551	3.5
Solar Radiation (W m^{-2})	184.64	55.64	42.84	138.04	186.86	230.02	324.71
Relative Humidity (%)	57.85	21.91	4	43.13	62	77	99.5
Rainfall (mm)	5.33	13.49	0	0	0.155	4.115	220.61
NO_2 (DU)	7.86×10^{14}	1.07×10^{15}	-9.15×10^{15}	2.53×10^{14}	7.12×10^{14}	1.30×10^{15}	9.49×10^{15}
SO_2 (DU)	0.031	0.194	-1.12	-0.08	0.025	0.132	1.32
Wind Speed (m s^{-1})	4.45	1.47	0.96	3.40	4.23	5.28	10.89
O_3 (DU)	268.88	10.97	235.2	261.1	269.8	277.15	305.5
MERRA-2 $PM_{2.5}$(μg m^{-3})	50.72	55.74	3.25	15.56	31.05	62.63	540.01

Correlation Analysis

The correlation matrix (Figure 2) provides clear insight into the relationships between $PM_{2.5}$ concentrations and the predictor variables used in the Random Forest model. Moderate positive correlations were observed between observed $PM_{2.5}$ and several environmental variables, including aerosol optical depth (AOD) ($r = 0.25$) and nitrogen dioxide (NO_2) ($r = 0.24$), indicating that increased aerosol loading and anthropogenic emissions may contribute to elevated particulate matter concentrations.

The strongest positive association with $PM_{2.5}$ was observed with reanalysis-derived $PM_{2.5}$ from MERRA-2 dataset ($r = 0.40$), suggesting that large-scale atmospheric processes captured by the reanalysis dataset are partially consistent with ground observations. Also, Negative correlations were observed between $PM_{2.5}$ and relative humidity ($r = -0.43$) as well as ozone ($r = -0.31$). The inverse relationship with relative humidity may reflect enhanced removal processes such as wet deposition or atmospheric dilution during humid conditions. Similarly, the negative correlation between $PM_{2.5}$ and ozone may be associated with photochemical interactions and differing atmospheric formation mechanisms. Solar radiation exhibited moderate correlations with several variables, including longitude ($r = 0.45$) and relative humidity ($r = -0.56$), reflecting regional atmospheric dynamics and seasonal variations. Relative humidity also showed a strong negative correlation with longitude ($r = -0.75$), suggesting spatial gradients in atmospheric moisture across the study area. Hence, the correlation analysis indicates that $PM_{2.5}$ concentrations in the study region are influenced by a combination of aerosol loading, meteorological conditions, and atmospheric chemistry. The relatively moderate correlations among predictors also suggest limited multicollinearity, which is useful for machine learning modeling such as Random Forest.

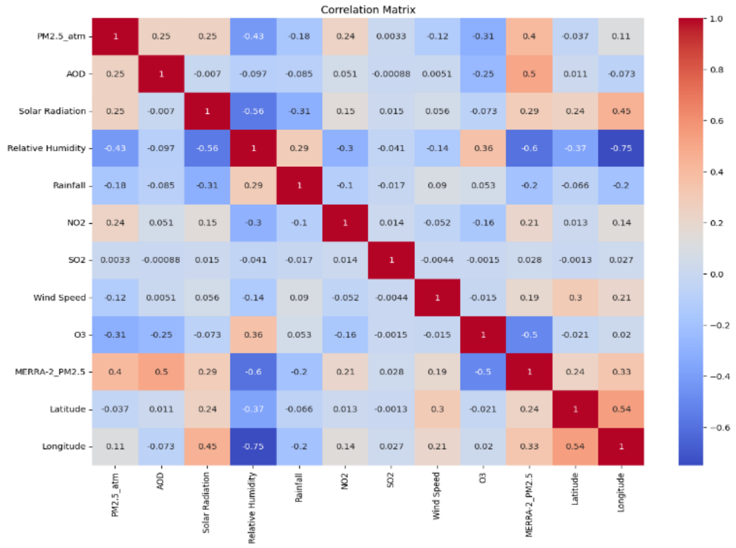
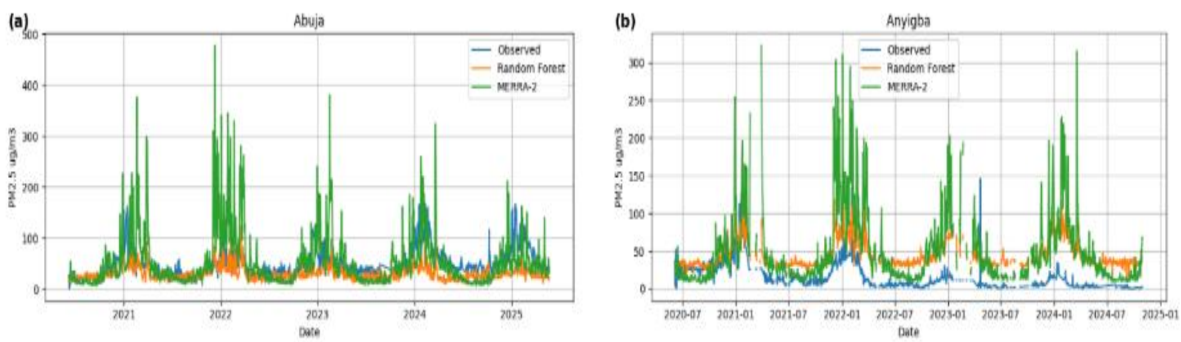


Figure 2: Statistical Correlations Between the Variables Used in the Model Development

Temporal Variability of $PM_{2.5}$ Concentrations

Temporal analysis of observed and predicted $PM_{2.5}$ concentrations reveals pronounced variability across the study period Figure 3. Both the RF model predictions and observational data capture fluctuations in particulate matter levels that may be associated with meteorological conditions, emission patterns, and regional atmospheric transport processes. Seasonal variability is a prominent feature of air quality across West Africa. During the dry season, increased dust transport from the Sahara Desert and biomass burning activities can elevate particulate matter concentrations. Conversely, during the wet season, precipitation and enhanced atmospheric mixing often reduce particulate matter levels through wet deposition and dispersion processes (Knippertz & Todd, 2012).

The RF model successfully reproduced many of these temporal patterns, suggesting that the predictor variables incorporated in the modeling framework adequately represent key environmental drivers of $PM_{2.5}$ variability. The ability of the model to track temporal changes further highlights the potential of machine learning approaches for monitoring air quality dynamics in data-sparse regions (Liousse *et al.,* 2014).



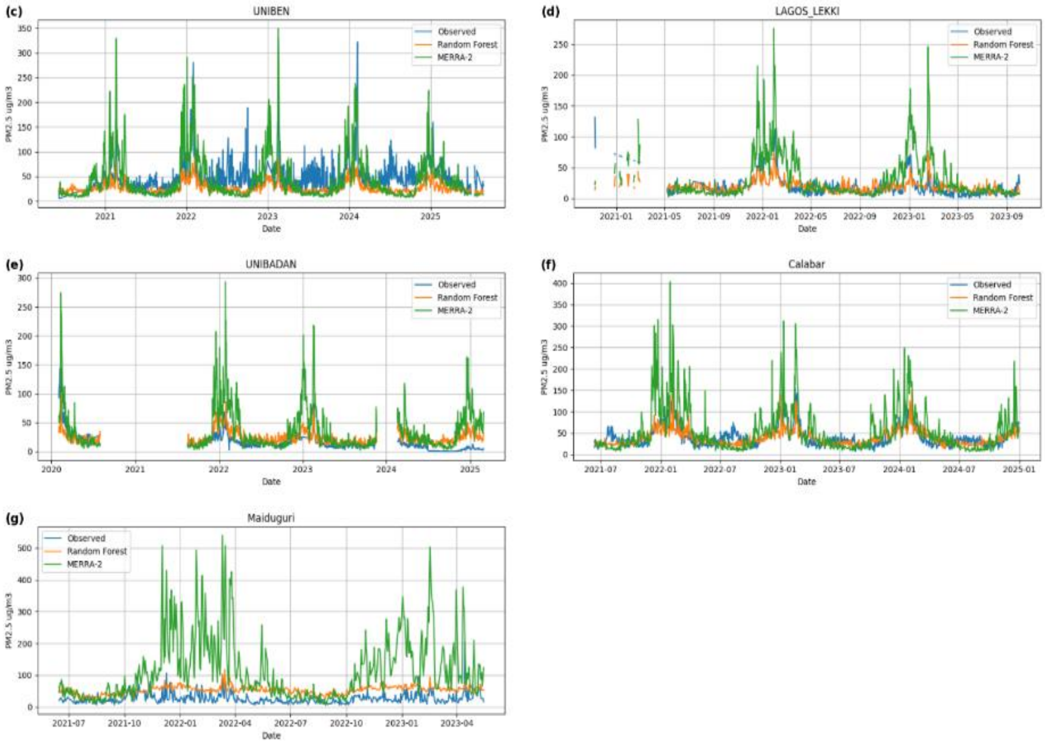


Figure 3: Diurnal Variations Plots Comparison Between Observed, RF Models and MERRA-2 $PM_{2.5}$ for the Stations over Nigeria

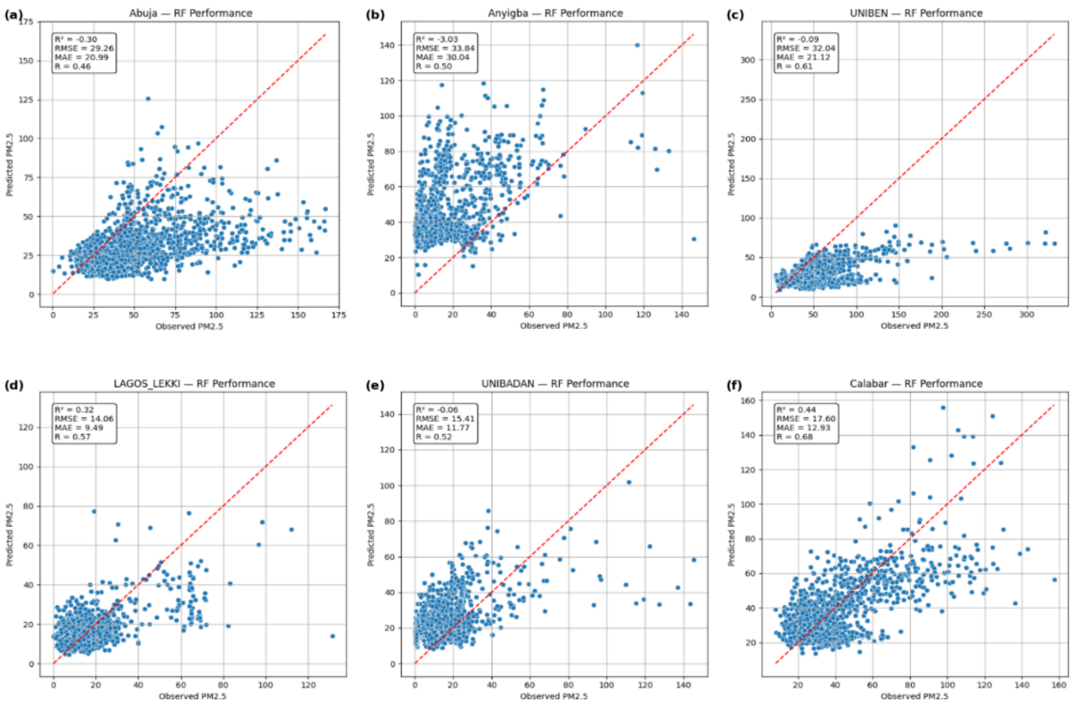


Figure 4: Scattered Plots of the Predicted and the Observed $PM_{2.5}$ Concentrations for the Stations (a, b, c, d, e and f) Representing Abuja, Anyigba, Uniben, Lagos Lekki, Unibadan and Calabar Respectively

Model Performance Across Monitoring Stations

Table 3 presents the performance metrics for the Random Forest model and the MERRA-2 reanalysis dataset across the monitoring stations used in this study. Model performance was evaluated using multiple statistical metrics including RMSE, MAE, R^2, IOA, and r. The model performance results indicated that the Random Forest model generally outperforms the reanalysis-based estimates from the MERRA-2 dataset across most monitoring stations. In particular, the RF model produced lower MAE and RMSE values at nearly all locations, suggesting improved agreement with ground observations. The best model performance was observed in Lagos, where the RF model achieved an R^2 value of 0.29 and relatively low error metrics (RMSE = 14.40 µg m^{-3}) as shown in Figure 4. This result may reflect stronger relationships between predictor variables and $PM_{2.5}$

concentrations in densely populated urban environments where emission sources are more consistent. In contrast, stations such as Maiduguri and Anyigba exhibited weaker model performance, with negative R^2 values indicating that the model struggled to capture the variability in observed $PM_{2.5}$ concentrations. These results may be associated with regional factors such as dust transport from the Sahara, limited training data, or complex atmospheric processes that are not fully represented by the predictor variables since all performance statistics reported in Table 3 were computed from the same final model evaluation used to generate the station-specific scatter plots, ensuring consistency between the quantitative metrics and graphical results. Despite these limitations, the Random Forest model consistently produced smaller prediction errors compared with the MERRA-2 dataset, highlighting the advantages of machine learning approaches for improving air quality estimation at local scales.

Table 3: Random Forest Model Performance in Comparison to MERRA-2 $PM_{2.5}$

Station	Model	RMSE	MAE	R^2	IOA	Correlation
Abuja	RF	29.26	20.99	-0.30	0.54	0.46
	MERRA-2	50.21	29.51	-2.83	0.47	0.37
Anyigba	RF	33.84	30.04	-0.301	0.49	0.50
	MERRA-2	55.95	37.26	-10.00	0.37	0.43
UNIBEN	RF	32.04	21.12	-0.09	0.53	0.61
	MERRA-2	36.10	25.18	-0.38	0.75	0.61
Lagos	RF	14.06	9.49	0.32	0.68	0.57
	MERRA-2	30.33	17.78	-2.16	0.64	0.61
Unibadan	RF	15.41	11.77	-0.06	0.64	0.52
	MERRA-2	37.10	22.52	-5.15	0.50	0.53
Calabar	RF	17.60	12.93	0.44	0.49	0.68
	MERRA-2	55.95	37.26	-10.00	0.37	0.43
Maiduguri	RF	33.17	28.98	-3.00	0.40	0.20
	MERRA-2	125.68	88.56	-56.43	0.21	0.47

Model Performance Across Monitoring Stations

The performance of the Random Forest (RF) model was evaluated across all monitoring stations and compared with $PM_{2.5}$ estimates from the MERRA-2 reanalysis product (Table 3). Overall, the RF model consistently outperformed MERRA-2 in terms of prediction accuracy, yielding substantially lower Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values across all stations. For example, in Abuja, the RF model achieved an RMSE of 29.26 $\mu g\ m^{-3}$ and an MAE of 20.99 $\mu g\ m^{-3}$, compared with 50.21 $\mu g\ m^{-3}$ and 29.51 $\mu g\ m^{-3}$, respectively, for MERRA-2. Similar improvements were observed in Lagos, where the RF model reduced the RMSE from 30.33 to 14.06 $\mu g\ m^{-3}$ and the MAE from 17.78 to 9.49 $\mu g\ m^{-3}$, while in Unibadan, the RMSE decreased from 37.10 to 15.41 $\mu g\ m^{-3}$ and the MAE from 22.52 to 11.77 $\mu g\ m^{-3}$. The largest improvement was recorded in Maiduguri, where the RF model reduced the RMSE from 125.68 to 33.17 $\mu g\ m^{-3}$ and the MAE from 88.56 to 28.98 $\mu g\ m^{-3}$, demonstrating the substantial advantage of the machine learning framework over the coarse-resolution reanalysis product.

Although the RF model consistently reduced prediction errors, its coefficient of determination (R^2) varied across monitoring stations, reflecting differences in local atmospheric conditions and emission characteristics. Positive predictive skill was achieved in Calabar ($R^2 = 0.44$) and Lagos ($R^2 = 0.32$), indicating that the model successfully explained a meaningful proportion of the observed $PM_{2.5}$ variability. In contrast, moderate negative R^2 values were obtained for Abuja (−0.30), UNIBEN (−0.09) and Unibadan (−0.06), while larger negative values occurred in Anyigba (−3.01) and Maiduguri (−3.00), suggesting that $PM_{2.5}$ variability at these locations is influenced by complex atmospheric processes that remain difficult to capture. Nevertheless, the RF model consistently produced markedly higher R^2 values than MERRA-2, whose performance was particularly poor in Maiduguri ($R^2 = -56.43$) and Anyigba ($R^2 = -10.00$), indicating limited capability for representing local pollution dynamics.

The Index of Agreement (IOA) and Pearson correlation coefficient further demonstrate the superiority of the RF model. The highest agreement was observed in Lagos (IOA = 0.68; r = 0.57) and Calabar (IOA = 0.49; r = 0.68), while UNIBEN recorded the strongest correlation (r = 0.61) despite a relatively modest IOA (0.53). Overall, RF achieved moderate to strong correlations ranging from 0.20 to 0.68, whereas MERRA-2 generally produced weaker predictive skill despite occasionally exhibiting comparable correlation values. An exception was UNIBEN, where MERRA-2 showed a higher IOA (0.75) while maintaining the same correlation coefficient (r = 0.61) as the RF model, suggesting that the reanalysis product can reproduce broader regional aerosol behaviour in some coastal environments despite its larger prediction errors.

The comparatively poorer performance of MERRA-2 is primarily attributable to its coarse spatial resolution (approximately 0.5° × 0.625°) and its representation of atmospheric processes at regional rather than local scales. Such resolution limits its ability to resolve highly heterogeneous emission sources, including urban traffic, industrial activities, domestic combustion, and land-use variability. In contrast, the RF model integrates ground-based observations with satellite-derived aerosol optical depth (AOD), atmospheric composition, and meteorological variables, enabling it to capture complex nonlinear relationships and local pollution characteristics more effectively. Overall, these results demonstrate that although both approaches have inherent limitations, the Random Forest framework provides more accurate and locally representative estimates of surface $PM_{2.5}$ across Nigeria's diverse environmental settings. These findings agree with previous studies (e.g., Hu *et al.,* 2017; Chen *et al.,* 2018), which demonstrated that ensemble machine learning algorithms substantially improve air quality estimation by integrating multi-source environmental datasets.

Average Model Performance Metrics Across the Seven Monitoring Stations

A comparative evaluation between the Random Forest (RF) model and the MERRA-2 reanalysis product demonstrates the superior predictive capability of the machine learning framework across the monitoring stations (Table 4). On average, the RF model achieved substantially lower prediction errors, with a mean MAE of 19.47 μg m⁻³ and RMSE of 25.77 μg m⁻³, compared with 36.87 μg m⁻³ and 55.62 μg m⁻³, respectively, for MERRA-2. These results indicate that the RF model reduced both systematic and random prediction errors by approximately one-half relative to the reanalysis product, demonstrating a much closer agreement with the ground-based $PM_{2.5}$ observations.

The coefficient of determination further illustrates the relative strengths of the two approaches. Although the mean R^2 of the RF model (−0.53) indicates that predictive performance varied considerably among stations, it represents a substantial improvement over the MERRA-2 mean R^2 (−11.57), which reflects very poor skill in reproducing local $PM_{2.5}$ variability. Positive R^2 values obtained at Lagos (0.32) and Calabar (0.44) further demonstrate that the RF model successfully captured the dominant sources of $PM_{2.5}$ variability in some environments, whereas MERRA-2 produced negative R^2 values at every monitoring station. Similarly, the RF model achieved a higher average Index of Agreement (IOA = 0.54) than MERRA-2 (0.47), indicating stronger agreement with observed $PM_{2.5}$ concentrations. The average correlation coefficient was also slightly higher for the RF model (r = 0.51) than for MERRA-2 (r = 0.49). Although the difference in correlation is modest, the substantially lower RMSE and MAE obtained by the RF model demonstrate that it more accurately reproduces both the magnitude and temporal variability of $PM_{2.5}$ concentrations. This suggests that correlation alone is insufficient for evaluating model performance and should be interpreted alongside error-based performance metrics.

The superior performance of the RF model is primarily attributable to its ability to integrate multiple environmental predictors, including satellite-derived aerosol optical depth, atmospheric trace gases, meteorological variables, and background aerosol information from MERRA-2, to capture complex nonlinear relationships governing $PM_{2.5}$ variability. In contrast, the coarse spatial resolution of the MERRA-2 reanalysis limits its capacity to resolve local emission sources and fine-scale atmospheric processes, particularly within heterogeneous urban environments. Consequently, the RF framework provides more reliable estimates of surface $PM_{2.5}$ across Nigeria than the standalone global reanalysis product. These findings are consistent with previous studies (e.g., Buchard *et al.,* 2017; Gelaro *et al.,* 2017; Hammer *et al.,* 2020), which reported that machine learning approaches substantially improve $PM_{2.5}$ estimation by downscaling coarse-resolution atmospheric products using multi-source environmental observations.

Table 4: Comparison of Accuracy RF and MERRA-2 $PM_{2.5}$

Metric	RF Model $PM_{2.5}$	MERRA-2_$PM_{2.5}$
MAE	19.47	36.87
RMSE	25.77	55.62
R^2	-0.53	-11.57
IOA	0.54	0.47
Correlation(r)	0.51	0.49

Feature Importance Analysis

An important advantage of the Random Forest (RF) algorithm is its ability to quantify the relative importance of predictor variables contributing to $PM_{2.5}$ estimation. Feature importance in RF is derived from the reduction in prediction error (e.g., mean decrease in impurity) attributed to each variable across the ensemble of decision trees. Variables that consistently reduce model uncertainty receive higher importance scores, reflecting stronger influence on predicted $PM_{2.5}$ concentrations. As shown in Figure 5, relative humidity (RH) emerged as the most influential predictor (importance = 0.275), highlighting the critical role of hygroscopic growth in modulating particulate matter concentrations. This is followed by MERRA-2 $PM_{2.5}$ (0.167), indicating that reanalysis products, despite their coarse resolution, still provide valuable background information for local-scale modeling. Trace gases such as ozone (O_3; 0.109) and nitrogen dioxide (NO_2; 0.086) also contributed significantly, reflecting their association with atmospheric chemistry and pollution sources.

Satellite-derived aerosol optical depth (AOD; 0.107) demonstrated strong importance, confirming its relevance as a proxy for columnar aerosol loading in $PM_{2.5}$ estimation. Meteorological variables, including solar radiation (0.080), wind speed (0.073), and rainfall (0.037), exhibited moderate to lower importance, suggesting their secondary but still meaningful role in influencing dispersion, removal, and formation processes. Sulfur dioxide (SO_2; 0.066) contributed modestly, likely reflecting its role as a precursor for secondary aerosol formation.

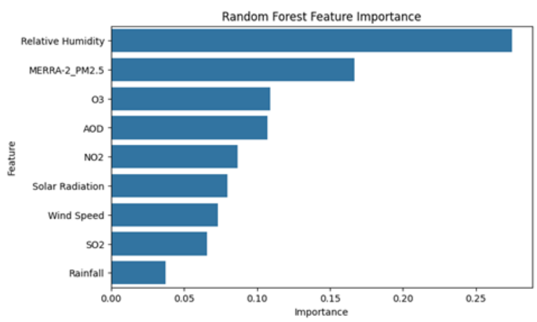


Figure 5: The Feature Importance Between the Variables Used in the RF Model Development

The overall feature importance results indicate that $PM_{2.5}$ variability in the study region is primarily driven by humidity effects, background aerosol conditions, and atmospheric composition, with meteorological dispersion factors playing a supporting role. These findings are consistent with established atmospheric processes, where hygroscopic growth, chemical transformation, and regional transport jointly control particulate matter dynamics. Importantly, the inclusion of diverse predictors enhances model interpretability and provides insight into the dominant environmental drivers of air pollution across the region.

Implications for Air Quality Monitoring in Data-Sparse Regions

The results of this study underscore the value of integrating machine learning approaches with multi-source environmental datasets to improve air quality estimation in regions with limited monitoring infrastructure. In many parts of West Africa, including Nigeria, the scarcity of regulatory-grade monitoring stations constrains the ability to accurately assess population exposure to air pollution. By combining ground-based observations from Purple Air and Clarity with satellite observations, meteorological variables, and reanalysis products such as MERRA-2, the RF model provides improved spatial and temporal representation of $PM_{2.5}$ concentrations. The superior performance of the RF model relative to MERRA-2 highlights the importance of data fusion and nonlinear modeling techniques for capturing local pollution dynamics that are not resolved by coarse-resolution global datasets. This approach enables more accurate characterization of air quality, which is essential for environmental policy formulation, health risk assessment, and sustainable urban planning. Furthermore, the framework developed in this study is flexible and can be extended to incorporate additional datasets, such as higher-resolution satellite products, emission inventories, and advanced meteorological reanalysis. Such integration has the potential to further enhance predictive accuracy and expand spatial coverage, particularly in under-monitored regions.

Seasonal and Ecological Variability of Model Performance

Seasonal evaluation across ecological zones revealed pronounced spatial heterogeneity in $PM_{2.5}$ model performance. In Table 5. the Sahel region, represented by Maiduguri, predictive skill was highest during the Dry season (RMSE = 11.50 µg m⁻³; R^2 = 0.45), while Wet-season performance deteriorated significantly (RMSE = 29.41 µg m⁻³), indicating strong sensitivity to convective atmospheric processes. In the Savannah region (Abuja and Anyigba), Dry-season conditions exhibited strong predictive capability (R^2 up to 0.83), reflecting stable aerosol–meteorology relationships under Harmattan influence. However, Wet-season performance showed substantial degradation, with negative R^2 values observed despite reductions in RMSE, suggesting increased atmospheric variability and nonlinear behaviour. The Guinea Coast region demonstrated comparatively stable performance across seasons. Wet-season RMSE values were consistently lower (e.g., Calabar: 6.87 µg m⁻³; UNIBEN: 12.04 µg m⁻³), and R^2 remained positive, indicating improved model reliability under monsoonal conditions. Coastal influence, particularly in Lagos (LAGOS LEKKI P-I), contributed to enhanced Wet-season predictability (R^2 = 0.40), although Dry-season performance could not be evaluated due to data limitations. Therefore, these results highlight a clear ecological gradient, with Dry-season dominance in the Sahel, transitional instability in the Savannah, and Wet-season enhancement in the Guinea Coast.

Table 5: Seasonal Performance of Best Models Across Ecological Zones

Station	Region	Dry RMSE	Wet RMSE	Dry R^2	Wet R^2	Best	Seasonal Behaviour
Maiduguri	Sahel	11.50	29.41	0.45	0.17	RF	Strong Dry dominance
Abuja	Savannah	17.62	8.73	0.57	-0.14	RF	Wet RMSE improves, R^2 unstable
Anyigba	Savannah	3.18	1.92	0.83	-0.35	RF	Wet RMSE improves, R^2 unstable
UNIBEN	Guinea Coast	12.53	12.04	0.61	0.33	RF	Stable, slight Wet advantage
Unibadan	Guinea Coast	6.59	3.34	-7.38	0.15	OLS	Wet dominates strongly
Calabar	Guinea Coast	7.97	6.87	0.77	0.30	RF	Wet better
LAGOS	Guinea Coast	NA	4.79	NA	0.40	OLS	Wet-only dataset

Seasonal Model Performance Across Ecological Zones

As shown in Table 6, strong regional contrasts are evident across ecological zones. In the Sahel (Maiduguri), model performance is clearly dominated by the Dry season, with RMSE improving from 29.41 µg m⁻³ (Wet) to 11.50 µg m⁻³ (Dry) and R^2 decreasing from 0.45 (Dry) to 0.17 (Wet). This is further confirmed by large negative ΔRMSE values in Table 8 (−13.52 for OLS; −17.91 for RF), indicating strong Dry-season superiority.

In the Savannah (Abuja and Anyigba), Wet-season RMSE improves significantly (e.g., Abuja: ΔRMSE = +8.26 (OLS) and +9.05 (RF); Anyigba: +3.63 (OLS) and +0.42 (RF), Table 4.8). However, despite this reduction in error, predictive skill declines, with R^2 dropping to negative values in the Wet season. This indicates that models capture mean conditions better but fail to explain variability.

In the Guinea Coast (UNIBEN, Unibadan, Calabar, Lagos), performance is more stable across seasons. Wet-season improvements are moderate but consistent (e.g., UNIBEN: ΔRMSE = +1.39 (OLS), +0.48 (RF); Unibadan: +3.25 (OLS); Calabar: +1.11 (RF), Table 4.8), and R^2 remains positive (e.g., Lagos Wet-season R^2 = 0.40), indicating reliable predictive performance.

Table 6: Seasonal Sensitivity (ΔRMSE = Dry − Wet)

Station	Model	ΔRMSE
Abuja	OLS	+8.261525
Abuja	RF	+9.047501
Anyigba	OLS	+3.628531
Anyigba	RF	+0.423786
UNIBEN	OLS	+1.392040
UNIBEN	RF	+0.482392
Unibadan	OLS	+3.247828
Calabar	RF	+1.106210
Maiduguri	OLS	−13.524918
Maiduguri	RF	−17.910110

Note: Positive Indicate Wet Season Performs Better and Negative Indicate Dry Season Performs Better

Seasonal Stability of Model Performance (SSI)

The Seasonal Stability Index (SSI) (Table 7) highlights a clear north–south gradient in model stability. The Sahel (Maiduguri) shows extreme instability (SSI = 1.56), reflecting large seasonal performance differences. The Savannah exhibits moderate variability, with Abuja (SSI ≈ 0.68) showing instability and Anyigba (SSI ≈ 0.15) relatively stable. In contrast, the Guinea Coast is the most stable region, with very low SSI values (UNIBEN ≈ 0.04; Calabar ≈ 0.16). This means that SSI reveals a clear North–South gradient in model stability. The Sahel exhibits extreme instability (SSI = 1.56), indicating strong sensitivity to seasonal changes driven by shifts from dust-dominated to convective regimes. The Savannah shows moderate instability, reflecting its transitional nature between northern dust influence and southern monsoonal systems. In contrast, the Guinea Coast demonstrates high stability (low SSI values), with minimal seasonal variation in model performance. This pattern confirms that model reliability increases from inland semi-arid regions to coastal humid environments, where atmospheric conditions are more consistent.

Table 7: Seasonal Stability Index (SSI)

Station	Region	SSI	Stability
Maiduguri	Sahel	1.56	Extremely unstable
Abuja	Savannah	0.68	Moderate
Anyigba	Savannah	0.15	Stable
UNIBEN	Guinea Coast	0.04	Highly stable
Calabar	Guinea Coast	0.16	Stable
Unibadan	Guinea Coast	0.65	Unstable
LAGOS	Guinea Coast	NA	Not computable

Although the proposed Random Forest framework demonstrated improved predictive performance relative to the MERRA-2 reanalysis product, model performance was not uniform across all monitoring stations and seasons. In particular, negative R^2 values observed at some locations indicate reduced predictive skill under certain environmental conditions. The seasonal and ecological analyses further demonstrate that model reliability varies across atmospheric regimes, with comparatively stable performance observed in the Guinea Coast and greater variability in the Sahel and Savannah, especially during the Wet season. Consequently, model predictions intended for exposure assessment or policy applications should be interpreted alongside these regional and seasonal performance characteristics.

The seasonal evaluation demonstrates that model performance is strongly influenced by regional atmospheric regimes. The superior Dry-season performance observed in the Sahel reflects the relatively stable aerosol conditions associated with Harmattan dust transport, whereas increased convective activity during the Wet season reduces predictive skill. In contrast, the Guinea Coast exhibits comparatively stable performance due to the moderating influence of maritime and monsoonal processes. These findings demonstrate that the proposed framework is capable of capturing seasonal variability across Nigeria's major ecological zones, although future studies should further investigate its performance during individual Harmattan dust outbreaks and biomass-burning events.

CONCLUSION

This study developed a Random Forest (RF) machine learning framework to estimate surface $PM_{2.5}$ concentrations across multiple monitoring stations in Nigeria by integrating low-cost sensor observations with satellite-derived aerosol optical depth (AOD), atmospheric trace gases, meteorological variables, and MERRA-2 reanalysis data. The RF model demonstrated strong predictive capability, capturing a substantial proportion of the observed variability in particulate matter concentrations while maintaining relatively low prediction errors. Comparison with the MERRA-2 reanalysis dataset showed that the RF model achieved better agreement with ground observations, highlighting the limitations of coarse-resolution atmospheric products in representing local-scale air quality and spatially heterogeneous emission sources. The findings confirm that machine learning-based data fusion provides a robust approach for improving $PM_{2.5}$ estimation in regions where conventional monitoring networks remain sparse.

Seasonal and ecological evaluations further demonstrated that model performance varied systematically across Nigeria's major climatic zones. The Sahel exhibited the strongest predictive performance during the Dry (Harmattan) season but showed pronounced degradation during the Wet season, reflecting the influence of seasonal convective processes. The Savannah displayed improved Wet-season RMSE but reduced predictive skill, indicating increased atmospheric complexity and nonlinear behaviour during the monsoon period. In contrast, the Guinea Coast showed comparatively stable performance throughout the year, with generally lower Wet-season prediction errors and consistently positive R^2 values. Seasonal Sensitivity (ΔRMSE) and the Seasonal Stability Index (SSI) further revealed a clear north–south gradient in model stability, with the Sahel exhibiting the greatest seasonal variability, the Savannah showing intermediate variability, and the Guinea Coast demonstrating the highest stability. These results confirm that ecological setting and seasonal atmospheric dynamics strongly influence $PM_{2.5}$ predictability across Nigeria.

Temporal analysis showed that both observed and modelled $PM_{2.5}$ concentrations exhibit substantial variability associated with seasonal meteorology and regional emission sources. The ability of the Random Forest model to reproduce these temporal patterns across contrasting ecological environments demonstrates its robustness for air quality assessment and monitoring. Feature importance analysis identified relative humidity, background aerosol conditions represented by MERRA-2 $PM_{2.5}$, atmospheric trace gases (particularly O_3 and NO_2), and aerosol optical depth as the dominant predictors $PM_{2.5}$ variability, emphasizing the combined influence of aerosol hygroscopic growth, atmospheric chemistry, regional transport, and meteorological processes.

In summary, this study demonstrates that integrating multi-source atmospheric datasets within a Random Forest framework substantially improves surface $PM_{2.5}$ estimation compared with standalone reanalysis products. Beyond improving prediction accuracy, the study provides new insight into the seasonal and ecological controls governing $PM_{2.5}$ predictability across Nigeria. The proposed framework offers valuable support for air quality assessment and environmental decision-making. However, the observed spatial and seasonal variability in model performance indicates that predictions should be interpreted with greater confidence in regions and seasons exhibiting higher predictive skill, while greater caution is warranted in areas characterized by lower model performance. Incorporating formal uncertainty quantification and prediction confidence estimates will further enhance the operational applicability of the proposed framework in future studies. With the increasing availability of low-cost air-quality sensors, satellite observations, and atmospheric reanalysis products across the continent, the proposed multi-source Random Forest framework provides a scalable and transferable approach for supporting operational air-quality forecasting, population exposure assessment, and evidence-based environmental policy. Its application has the potential to strengthen air-quality management, public health protection, and sustainable environmental planning not only in Nigeria and West Africa but also across other data-sparse regions of Africa where conventional monitoring infrastructure remains limited.

REFERENCES

Abuouelezz, W., Ali, N., Aung, Z., Altunaiji, A., Shah, S. B., & Gliddon, D. (2025). Exploring PM2.5 and PM10 ML forecasting models: A comparative study in the UAE. Scientific Reports, 15(1), 9797. https://doi.org/10.1038/s41598-025-94013-1

Aladodo, S. S., Akoshile, C. O., Ajibola, T. B., Sani, M., Iborida, O. A., & Fakoya, A. A. (2022). Seasonal Tropospheric Aerosol Classification Using AERONET Spectral Absorption Properties in African Locations. Aerosol Science and Engineering, 6(3), 246–266. https://doi.org/10.1007/s41810-022-00140-x

Alani, R. A., Ayejuyo, O. O., Akinrinade, O. E., Badmus, G. O., Festus, C. J., Ogunnaike, B. A., & Alo, B. I. (2019). The level PM2.5 and the elemental compositions of some potential receptor locations in Lagos, Nigeria. Air Quality, Atmosphere & Health, 12(10), 1251–1258. https://doi.org/10.1007/s11869-019-00743-3

Amegah, A. K., & Agyei-Mensah, S. (2017). Urban air pollution in Sub-Saharan Africa: Time for action. Environmental Pollution, Part A, 220, 738–743. https://doi.org/10.1016/j.envpol.2016.09.042.

Amooli, J. A., Hackman, K. O., Nana, B., & Westervelt, D. M. (2024). Fine particulate air pollution estimation in Ouagadougou using satellite aerosol optical depth and meteorological parameters. Environmental Science: Atmospheres, 4(9), 1012–1025. https://doi.org/10.1039/D4EA00057A

Ardon-Dryer, K., Dryer, Y., Williams, J. N., & Moghimi, N. (2020). Measurements of PM2.5 with PurpleAir under atmospheric conditions. Atmospheric Measurement Techniques, 13(10), 5441–5458. https://doi.org/10.5194/amt-13-5441-2020

Awokola, B. I., Okello, G., Dobson, R., Amusa, G. A., Johnson, O., Erhart, A., Mortimer, K., Jewell, C., Semple, S., & Ma3 Study Group. (2022). Measuring Air Quality for Advocacy in Africa (MA3): Ambient PM 2.5 Concentrations Over One-Year in 15 Locations in Eight Sub-Saharan African Countries Using Low-Cost Sensors. D25. The View Ahead: Exposure Assessment, Communication, And Early Life Impact, A5170–A5170. https://doi.org/10.1164/ajrccm-conference.2022.205.1_MeetingAbstracts.A5170

Awokola, B., Okello, G., Johnson, O., Dobson, R., Ouédraogo, A. R., Dibba, B., Ngahane, M., Ndukwu, C., Agunwa, C., Marangu, D., Lawin, H., Ogugua, I., Eze, J., Nwosu, N., Ofiaeli, O., Ubuane, P., Osman, R., Awokola, E., Erhart, A., … Semple, S. (2022). Longitudinal Ambient PM2.5 Measurement at Fifteen Locations in Eight Sub-Saharan African Countries Using Low-Cost Sensors. Atmosphere, 13(10), 1593. https://doi.org/10.3390/atmos13101593

Breiman, L. (2001). Random Forests (Machine Learning, Vol. 45). Kluwer Academic Publishers.

Buchard, V., Da Silva, A. M., Colarco, P. R., Darmenov, A., Randles, C. A., Govindaraju, R., Torres, O., Campbell, J., & Spurr, R. (2015). Using the OMI aerosol index and absorption aerosol optical depth to evaluate the NASA MERRA Aerosol Reanalysis. Atmospheric Chemistry and Physics, 15(10), 5743–5760. https://doi.org/10.5194/acp-15-5743-2015

Buchard, V., Randles, C. A., Da Silva, A. M., Darmenov, A., Colarco, P. R., Govindaraju, R., Ferrare, R., Hair, J., Beyersdorf, A. J., Ziemba, L. D., & Yu, H. (2017). The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part II: Evaluation and Case Studies. Journal of Climate, 30(17), 6851–6872. https://doi.org/10.1175/JCLI-D-16-0613.1

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., & Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environment

International, 99, 293–302. https://doi.org/10.1016/j.envint.2016.12.007

Chen, Z.-Y., Zhang, T.-H., Zhang, R., Zhu, Z.-M., Ou, C.-Q., & Guo, Y. (2018). Estimating PM2.5 concentrations based on non-linear exposure-lag-response associations with aerosol optical depth and meteorological measures. Atmospheric Environment, 173, 30–37. https://doi.org/10.1016/j.atmosenv.2017.10.055

Colarco, P., Silva, A. da, Chin, M., & Diehl, T. (2010). Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth. *Journal Of Geophysical Research*, 115(D14207). https://doi.org/10.1029/2009JD012820

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M. B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L. J., & Schwartz, J. (2019). An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environment International, 130, 104909. https://doi.org/10.1016/j.envint.2019.104909

Emery, C., Liu, Z., Russell, A. G., Odman, M. T., Yarwood, G., & Kumar, N. (2017). Recommendations on statistics and benchmarks to assess photochemical model performance. Journal of the Air & Waste Management Association, 67(5), 582–598. https://doi.org/10.1080/10962247.2016.1265027

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Da Silva, A. M., Gu, W., … Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Journal of Climate, 30(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Tong, D., Zheng, B., Peng, Y., Huang, X., He, K., & Zhang, Q. (2021). Tracking Air Pollution in China: Near Real-Time PM2.5 Retrievals from Multiple Data Sources. Atmospheric and Oceanic Physics. https://doi.org/10.48550/arXiv.2103.06520

Gupta, P., Khan, M., da Silva, A., & Patadia, F. (2013). MODIS aerosol optical depth observations over urban areas in Pakistan: Quantity and quality of the data for air quality monitoring. Atmospheric Pollution Research, 4(1), 43–52. https://doi.org/10.5094/apr.2013.005

Hammer, M. S., Van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., Levy, R. C., Garay, M. J., Kalashnikova, O. V., Kahn, R. A., Brauer, M., Apte, J. S., Henze, D. K., Zhang, L., Zhang, Q., Ford, B., Pierce, J. R., & Martin, R. V. (2020). Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998–2018). Environmental Science & Technology, 54(13), 7879–7890. https://doi.org/10.1021/acs.est.0c01764

Horowitz, H. M., Garland, R. M., Thatcher, M., Landman, W. A., Dedekind, Z., Van Der Merwe, J., & Engelbrecht, F. A. (2017). Evaluation of climate model aerosol seasonal and spatial variability over Africa using AERONET. Atmospheric Chemistry and Physics, 17(22), 13999–14023. https://doi.org/10.5194/acp-17-13999-2017

Idris, M., Sani, M., & Aliyu, R. (2025). Comparison OF Particulate Matter (PM2.5) Ground Data And Satellite Data in Kano State, Nigeria. Fudma Journal of Sciences (FJS), 9(12), 621–628. https://doi.org/10.33003/fjs-2025-0912-4318

Jenkins, G. S., Freire, S. M., Ogunro, T., Niang, D., Andrade, M., Drame, M. S., Huvi, J. B., Pires, E. E. S., Toure, E. N., & Camara, M. (2023). COVID-19 New Cases and Environmental Factors During Wet and Dry Seasons in West and Southern Africa. GeoHealth, 7(8), e2022GH000765. https://doi.org/10.1029/2022GH000765

Kelly, K. E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R., & Butterfield, A. (2017). Ambient and laboratory evaluation of a low-cost particulate matter sensor. Environmental Pollution, 221, 491–500. https://doi.org/10.1016/j.envpol.2016.12.039.

Knippertz, P., & Todd, M. C. (2012). Mineral dust aerosols over the Sahara: Meteorological controls on emission and transport and implications for modeling. Reviews of Geophysics, 50(1), 2011RG000362. https://doi.org/10.1029/2011RG000362

Kunjir, G. M., Tikle, S., Das, S., Karim, M., Roy, S. K., & Chatterjee, U. (2025). Assessing particulate matter (PM2.5) concentrations and variability across Maharashtra using satellite data and machine learning techniques. Discover Sustainability, 6(1), 238. https://doi.org/10.1007/s43621-025-01082-3

Liousse, C., Assamoi, E., Criqui, P., Granier, C., & Rosset, R. (2014). Explosive growth in African combustion emissions from 2005 to 2030. Environmental Research Letters, 9(3), 035003. https://doi.org/10.1088/1748-9326/9/3/035003

Malings, C., Knowland, K. E., Keller, C. A., & Cohn, S. E. (2021). Sub-City Scale Hourly Air Quality Forecasting by Combining Models, Satellite Observations, and Ground Measurements. Earth and Space Science, 8(7), e2021EA001743. https://doi.org/10.1029/2021EA001743

Malings, C., Westervelt, D. M., Hauryliuk, A., Presto, A. A., Grieshop, A., Bittner, A., Beekmann, M., & R. Subramanian. (2020). Application of low-cost fine particulate mass monitors to convert satellite aerosol optical depth to surface concentrations in North America and Africa. Atmospheric Measurement Techniques, 13(7), 3873–3892. https://doi.org/10.5194/amt-13-3873-2020

Marais, E. A., & Wiedinmyer, C. (2016). Air Quality Impact of Diffuse and Inefficient Combustion Emissions in Africa (DICE-Africa). 4(50(19)), 10739–10745. https://doi.org/10.1021/acs.est.6b02602

Mathew, A., Gokul, P. R., Raja Shekar, P., Arunab, K. S., Ghassan Abdo, H., Almohamad, H., & Abdullah Al Dughairi, A. (2023). Air quality analysis and PM2.5 modelling using machine learning techniques: A study of Hyderabad city in India. Cogent Engineering, 10(1), 2243743. https://doi.org/10.1080/23311916.2023.2243743

Muhammad, N. A. A. A., & Ahmad, M. Y. (2025). PM2.5 Concentration Prediction Model in Jakarta Area Using

Random Forest Algorithm. Journal of Computation Physics and Earth Science, 5(1), 31–39. https://journal.physan.org/index.php/jocpes/index

Nobell, S., Majumdar, A., Mukherjee, S., Chakraborty, S., Chatterjee, S., Bose, S., Dutta, A., Sethuraman, S., Westervelt, D., Sengupta, S., Basu, R., & McNeill, V. F. (2023). Validation of In-field Calibration for Low-Cost Sensors Measuring Ambient Particulate Matter in Kolkata, India [Preprint]. Chemistry. https://doi.org/10.26434/chemrxiv-2023-8lhrq-v2

Ogunjobi, K. O., & Awoleye, P. O. (2019). Intercomparison and Validation of Satellite and Ground-Based Aerosol Optical Depth (AOD) Retrievals over Six AERONET Sites in West Africa. Aerosol Science and Engineering, 3(1), 32–47. https://doi.org/10.1007/s41810-019-00040-7

Ouarma, I., Nana, B., Haro, K., Béré, A., & Koulidiati, J. (2020). Assessment of Pollution Levels of Suspended Particulate Matter on an Hourly and a Daily Time Scale in West African Cities: Case Study of Ouagadougou (Burkina Faso). Journal of Geoscience and Environment Protection, 08(11), 119–138. https://doi.org/10.4236/gep.2020.811007

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. Machine Learning in Python.

Pendergrass, D. C., Zhai, S., Kim, J., Koo, J.-H., Lee, S., Bae, M., Kim, S., Liao, H., & Jacob, D. J. (2022). Continuous mapping of fine particulate matter (PM2.5) air quality in East Asia at daily 6 × 6 km2 resolution by application of a random forest algorithm to 2011–2019 GOCI geostationary satellite data. Atmospheric Measurement Techniques, 15(4), 1075–1091. https://doi.org/10.5194/amt-15-1075-2022

Petkova, E., Jack, D. W., Volavka-Close, N. H., & Kinney, P. L. (2013). Particulate matter pollution in African cities. Air Quality, Atmosphere and Health, 6, 603–614. https://doi.org/10.1007/s11869-013-0199-6

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature Communications, 11(1), 4540. https://doi.org/10.1038/s41467-020-18321-y

Prospero, J. M., Ginoux, P., Torres, O., Nicholson, S. E., & Gill, T. E. (2002). Environmental Characterization Of Global Sources of Atmospheric Soil Dust Identified With The Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) Absorbing Aerosol PRODUCT. Reviews of Geophysics, 40(1). https://doi.org/10.1029/2000RG000095

Raheja, G., Nimo, J., Appoh, E. K.-E., Essien, B., Sunu, M., Nyante, J., Amegah, M., Quansah, R., Arku, R. E., Penn, S. L., Giordano, M. R., Zheng, Z., Jack, D., Chillrud, S., Amegah, K., Subramanian, R., Pinder, R., Appah-Sampong, E., Tetteh, E. N., … Westervelt, D. M. (2023). Low-Cost Sensor Performance Intercomparison, Correction Factor Development, and 2+ Years of Ambient PM 2.5 Monitoring in Accra, Ghana. Environmental Science & Technology, 57(29), 10708–10720. https://doi.org/10.1021/acs.est.2c09264

Randles, C. A., Da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., & Flynn, C. J. (2017). The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation. Journal of Climate, 30(17), 6823–6850. https://doi.org/10.1175/JCLI-D-16-0609.1

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., … Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. Journal of Climate, 24(14), 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Ecography—2016—Roberts—Cross-validation strategies for data with temporal spatial hierarchical or phylogenetic.pdf. Ecography, 40, 913–929. https://doi.org/10.1111/ecog.02881

Sayeed, A., Lin, P., Gupta, P., Tran, N. N. M., Buchard, V., & Christopher, S. (2022). Hourly and Daily PM2.5 Estimations Using MERRA-2: A Machine Learning Approach. Earth and Space Science, 9(11), e2022EA002375. https://doi.org/10.1029/2022EA002375

Sharafa, S. B., Aliyu, R., Ibrahim, B. B., Tijjani, B. I., Darma, T. H., Gana, U. M., Ayedun, F., & Sulu, H. T. (2020). Model prediction and climatology of aerosol optical depth (τ550) and angstrom exponent (α470-660) over three aerosol robotic network stations in Sub-Saharan Africa using moderate resolution imaging spectroradiometer data. Nigerian Journal of Technology, 39(1), 255–268. https://doi.org/10.4314/njt.v39i1.29

Suriano, D., Akinnusotu, A., Abulude, F. O., & Oluwagbayide, S. D. (2024). Assessment of Particulate Matter (PM2.5) and Air Quality Index (AQI) in Eight Locations of Lagos State, Nigeria. https://doi.org/10.20944/preprints202409.1991.v1

Van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., & Winker, D. M. (2016). Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environmental Science & Technology, 50(7), 3762–3772. https://doi.org/10.1021/acs.est.5b05833

Van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM 2.5 using aerosol optical depth determined from satellite remote sensing. Journal of Geophysical Research, 111(D21), D21201. https://doi.org/10.1029/2005JD006996

Vignesh, P. P., Jiang, J. H., & Kishore, P. (2023). Predicting PM2.5 Concentrations Across USA Using Machine Learning. Earth and Space Science, 10(10), e2023EA002911. https://doi.org/10.1029/2023EA002911

Wilks, D. S. (2006). Statistical methods in the atmospheric sciences (2nd ed). Academic Press.

Willmott, C. J. (1981). ON THE VALIDATION OF MODELS. Physical Geography, 2(2), 184–194. https://doi.org/10.1080/02723646.1981.10642213

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30, 79–82. https://doi.org/10.3354/cr030079

Yahaya, T., Umar, F. M., Zanna, A. M., Abdulmalik, A., Ibrahim, B. A., Bilyaminu, M., & Joseph, A. (2023). Concentrations and health risks of particulate matter (PM2.5) and associated elements in the ambient air of Lagos, Southwestern Nigeria. Bio-Research, 21(3), 2141–2149. https://doi.org/10.4314/br.v21i3.9

Yan, X., Shi, W., Li, Z., Li, Z., Nana, L., Wenji, Z., Wang, H., & Yu, X. (2017). Satellite-based PM2.5 estimation using fine-mode aerosol optical thickness over China. Atmospheric Environment, 170, 290–302. https://doi.org/10.1016/j.atmosenv.2017.09.023.

Zhang, D., Du, L., Wang, W., Zhu, Q., Bi, J., Scovronick, N., Naidoo, M., Garland, R. M., & Liu, Y. (2021). A machine learning model to estimate ambient PM2.5 concentrations in industrialized highveld region of South Africa. Remote Sensing of Environment, 266, 112713. https://doi.org/10.1016/j.rse.2021.112713

Zhang, Y., Li, Z., Bai, K., Wei, Y., Xie, Y., Zhang, Y., Hong, J., Xu, H., Guang, J., Lv, Y., Li, K., & Li, D. (2021). Satellite remote sensing of atmospheric particulate matter mass concentration: Advances, challenges, and perspectives. Fundamental Research, 1(3), 240–258. https://doi.org/10.1016/j.fmre.2021.04.007

Zhou, Q., Nizamani, M. M., Zhang, H.-Y., & Zhang, H.-L. (2023). The Air We Breathe: An In-Depth Analysis of PM2.5 Pollution in 1312 Cities from 2000 to 2020. 4. https://doi.org/10.21203/rs.3.rs-2740958/v1
.