Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 3 (2024) 212

African Scientific Reports

Investigating factors influencing $PM_{2.5}$ threshold exceedance in Nigeria using multivariate logistic regression analysis

Tertsea Igbawua [a,*], Aondongu Alexander Tyovenda[a], Terver Sombo[a], Idugba Mathias Echi[a], Fanan Ujoh[b]

[a] *Department of Physics, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria*
[b] *Centre for the Integrated Delivery of the Built Environment (IDoBE), School of the Built Environment and Architecture, London South Bank University, United Kingdom*

Abstract

The understanding of the drivers of fine particulate matter of diameter less than or equal to 2.5 micrometers ($PM_{2.5}$) in Nigeria will greatly assist policymakers and other stakeholders in developing appropriate air quality management strategies to protect the health of the public. This research analyzed the drivers of $PM_{2.5}$ threshold exceedance (PTE) in Nigeria from 2000 to 2022, employing a multivariate logistic regression (MLR) model. The work assessed aerosol emission variables (AEVs) and meteorology variables (MEVs) to understand their impact on $PM_{2.5}$ concentrations. Results showed a significant spatial heterogeneity in the trend of $PM_{2.5}$ concentrations, with increasing values in all the study areas. The estimated coefficients of black carbon emission (BCEM) and sulfate emission (SUEM) showed a significant positive response, exacerbating the threshold exceedance in Abuja and Lagos, while dust emission (DUEM) had a significant positive response in Kano and Lagos. On the other hand, OCEM had a negative association in all the study locations. The MEV variables of sea level pressure (SLP), relative humidity (RH), planetary boundary layer height (PBLH), and wind speed (WS) had significant negative responses in Abuja, while RH and WS were significant in Kano, and RH, PBLH, and precipitation (P) were significant in Lagos, indicating that the $PM_{2.5}$ concentrations decrease with increasing values of these factors; however, temperature (TMP) had a significant positive response in Lagos but was not significant in Abuja and Kano. The output of the study revealed complex relationships between aerosol emissions, climate, and other environmental variables driving the PTE in Nigeria.

DOI:10.46481/asr.2024.3.2.212

Keywords: $PM_{2.5}$ Threshold exceedance, Logistic regression model, Meteorological variables, Aerosol emission rate, Nigeria

Article History :
Received: 15 May 2024
Received in revised form: 24 July 2024
Accepted for publication: 30 July 2024
Published: 12 September 2024

1. Introduction

Fine particulate matter ($PM_{2.5}$) in the ambient air is a potential risk for public health all over the world. Due to its small diameter of 2.5 microns or $PM_{2.5}$ can transport toxic substances deep into the lower regions of the respiratory tract, inducing potentially harmful health effects [1, 2]. The presence of pollutant chemical, biological, or physical elements in the ambient air modifies the composition of this medium and is known as air pollution [3]. Air pollution can negatively affect health and is associated with poor

*Corresponding author: Tel.: +234-803-695-2266.
Email address: tertsea.igbawua@uam.edu.ng (Tertsea Igbawua)

outcomes including low birth weight, respiratory illnesses, allergic rhinitis, gross motor development delay, lung cancer, increased risks of neurological disorders, and overall poorer health [2, 4–8].

For the monitoring of ambient air, its quality has to be checked and modeled regarding the concentration of pollutants (like $PM_{2.5}$) it contains. Air quality modeling techniques are often employed, particularly when pollutant levels exceed the guidelines or regulatory limits [9]. Several documented studies have reported that meteorological [1, 5, 10–12] and aerosol variables [5] strongly affect $PM_{2.5}$ ambient levels. Xu [12] reported that increased temperature was the most important factor for the increase in $PM_{2.5}$ concentration, while an increase in the normalized difference vegetation index (NDVI) played an important role in the reduction in $PM_{2.5}$ concentration. Also, Xu [12] suggested that precipitation and the green coverage rate of built-up areas cause the $PM_{2.5}$ concentrations to change, whereas others have indicated that the $PM_{2.5}$ concentration changed with the growth in population and gross domestic product (GDP). Sun [1] determined that from 2013–2016, the influencing factors of $PM_{2.5}$ pollution days included wind speed, no precipitation day, relative humidity, population density, construction area, transportation, coal consumption, and green coverage rate. A report by Yang [11] indicated that wind speed was the most important meteorological factor affecting $PM_{2.5}$ and PM_{10}; temperature, air pressure, and relative humidity were also key affecting factors in some seasons. Among the above meteorological factors, wind velocity, rainfall, air temperature, soil temperature, and soil humidity are negatively correlated with $PM_{2.5}$ concentrations [10]. Sulaymon [13] documented and elucidated the influence of meteorological conditions on the formation and dispersion of $PM_{2.5}$ during atmospheric pollution episodes (APE).

The World Health Organization (WHO) has set standards and threshold limits for the management of exposures to air pollution [8]. According to Angulo and Madrid [14], threshold exceedances are essentially the standard for risk assessment. The analysis of environmental phenomena for risk assessment usually involves the construction of indicators related to the structural characteristics of extremal events defined by exceedances over critical thresholds [15]. Threshold exceedances have been studied using Markov chain models [16]. Accordingly, Grineski *et al.* [17] worked on $PM_{2.5}$ threshold exceedances (PTEs) during the prenatal period and the risk of intellectual disability and demonstrated that chronic exposure to fine particulates increases the risks of neurodevelopmental conditions, such as intellectual disability (ID). Rincon *et al.* [18] conducted a study on $PM_{2.5}$ exceedances and identified their sources to help in developing an early warning system. They created a logistic model to predict $PM_{2.5}$ exceedances (≥ 12.5 µg/m^3) and found that both forest fires and heavy traffic played significant roles in elevating $PM_{2.5}$ concentrations.

Increased air pollution has been observed in densely populated Nigerian cities such as Lagos, Abuja, and Port-Harcourt [5]. Results from an air pollution assessment by Sulaymon *et al.* [5] in Lugbe, Abuja, showed that the highest ambient $PM_{2.5}$ concentrations (142 µg/m^3) were recorded in January, while the lowest (84 µg/m^3) were observed in July. Idris *et al.* [19] investigated the effect of meteorological parameters on the dispersion of vehicular emissions in some selected areas in Kano State, Nigeria. The concentration of $PM_{2.5}$ was highest at sampling point 2 (IBB Road), with a value of 184 µg/m^3. In Abuja, Wambebe & Duan [20] investigated air quality levels and also did a health risk assessment of particulate matter (PM), and results showed that the daily averaged concentrations of $PM_{2.5}$ varied from 15.30 µg/m^3 to 70.20 µg/m^3. The top four most polluted locations were found to be above the acceptable (25 µg/m^3) air quality index (AQI) limit stipulated by WHO, which all fell far below the unhealthy AQI value index level. Kanee *et al.* [21] also worked in Abuja, and reports showed that air quality during the dry season was 15–95 µg/m^3 for $PM_{2.5}$ while in the wet season, 09–75 µg/m^3 was observed for $PM_{2.5}$. Sulaymon *et al.* [13] modeled $PM_{2.5}$ in Lagos, and results showed that spatially elevated $PM_{2.5}$ concentrations were found in the northwestern region of Lagos, an urban area with higher anthropogenic emissions. Residential land use contributed the most to total $PM_{2.5}$ (∼40 µg/m^3), followed by industry (∼20 µg/m^3). In all the earlier-mentioned works done in Nigeria, all concentrations exceeded the threshold limits set out by the WHO and local national standards.

This study focused on three locations: the Federal Capital Territory (Abuja), Kano (Kano), and Lagos (Lagos). These states were chosen based on reports of elevated $PM_{2.5}$ readings [3, 13, 19, 21] and the availability of ground-based data. In all these studies, AQI and WHO standards have been exceeded, emphasizing the need to study the factors influencing $PM_{2.5}$ Threshold Exceedance (PTE) in the selected states. The work adopted satellite data for the spatial representation of $PM_{2.5}$ due to the scarcity and very limited availability of ground observation datasets. Satellite data has been invaluable for conducting spatial air pollution research due to its synoptic spatial coverage and low cost. Given these considerations, the purpose of this study was to investigate the effects of meteorological factors and aerosol emission variables on PTE from 2000 to 2022.

2. Materials and methods

2.1. Study area

$PM_{2.5}$ pollution is a great problem in Abuja, Kano and Lagos, with all the cities experiencing high levels of fine particulate matter which are mainly emitted from local sources and transported from farthest sources to the receptor areas of the study [5]. Despite the variation in the geographical locations of the states and their distances from natural sources such as desert and ocean (Figure 1), each state is facing the challenges of air pollution as a result of its high population density. In Kano, its closeness to the desert leads to the abundance of airborne dust particles in the atmosphere which could greatly increase the concentrations of $PM_{2.5}$. Moreover, industrial activities and vehicle emissions also add to the magnitude of air pollution in this highly populated city. Abuja is located in the center of Nigeria and high levels of $PM_{2.5}$ in this city are mainly attributed to urbanization, vehicle emissions, long range transport from farthest sources and industrial activities. Its increasing population and development lead to the increase in the levels

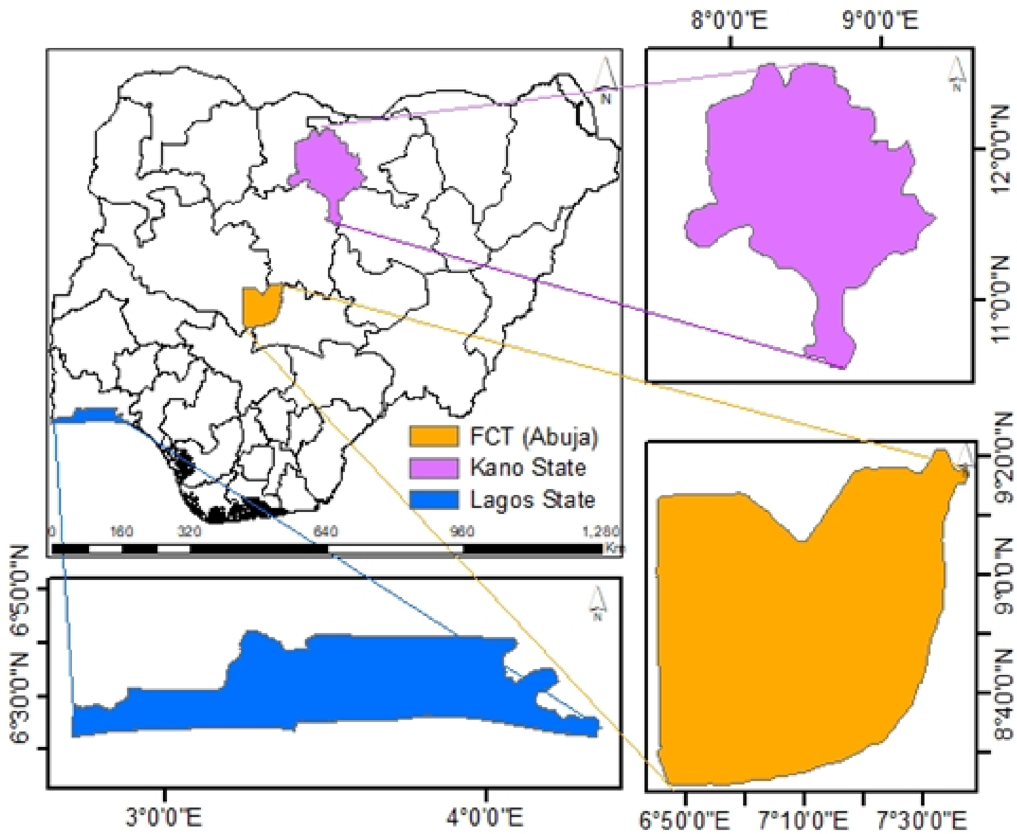

Figure 1: Study area.

of pollution in the city despite the efforts put in place to establish regulations for air quality in the city. Lagos state is nearest to the ocean and therefore gets affected by the oceanic activities. The high population density and large extent of industrial, sea salts, long range transport from farthest sources and commercial activities are the major contributors to the air pollution problems in this highly populated metropolitan city. Vehicle emissions, industrial pollution and biomass burning also contribute to the high concentrations of $PM_{2.5}$ in all the chosen study areas. In the end, regardless of their distinct geographical features and population densities, Abuja, Kano, and Lagos all struggle with excessive $PM_{2.5}$ levels, demanding comprehensive actions to reduce air pollution and protect public health in these major study areas.

2.2. Data

The data sets (Table 1) used in this work include: Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2) [22–24]. European Center for Medium Range Forecast (ECMWF) version 5 (ERA5) [25]. Climatic Research Unit (CRU) of the University of East Anglia. The CRU data are a collection of observation stations from meteorological stations around the world that are gridded into a global high-resolution data set [26]. The ground-based observation data for $PM_{2.5}$ was obtained from the PurpleAir data sensor network. The $PM_{2.5}$ mass concentration [27] was aggregated from the various aerosol species using equation 1.

$$PM_{2.5} = 1.375 \times [SO_4SMASS] + [OCSMASS] + [BCSMASS] + [DUSMASS25] + [SSSMASS25]\,. \tag{1}$$

2.3. Methods

2.3.1. Prediction of $PM_{2.5}$ threshold exceedance using Multivariate Logistic Regression Model (MLRM)

The standard MLRM was used to predict the factors influencing PTE in the study area. The MLRM is given in [28] as:

$$p(y = 1\,|x) = \frac{1}{1 + e^{-(a+b_1x_1+b_2x_2+\cdots+b_nx_n)}}, \tag{2}$$

where $p(y = 1\,|X)$ is the probability of the dependent variable being in class 1, the exp is the base of the natural logarithm, $b_1, b_2 \ldots b_n$ are the coefficients for the independent variables $x_1, x_2 \ldots x_n$ respectively and a is the intercept. The coefficients (Equation (2)) reflect the impact of each independent variable on the likelihood of $PM_{2.5}$ concentrations exceeding the WHO threshold.

Table 1: Data sets used for this study.

Data	Quantity	Resolution	Temporal Span
ERA5	Precipitation (P)	0.5×0.5	2000 – 2022
	Temperature (TMP)	0.5×0.5	2000 – 2022
	Wind speed (WS)	0.5×0.5	2000 – 2022
	Relative humidity (RH)	0.5×0.5	2000 – 2022
MERRA2	Black Carbon surface mass conc. (BCSMASS)	0.5×0.625	2000 – 2022
	Organic Carbon surface mass conc. (OCSMASS)	0.5×0.625	2000 – 2022
	Dust surface mass conc. PM25 (DUSMASS25)	0.5×0.625	2000 – 2022
	Sulfate surface mass conc. (SO_4SMASS)	0.5×0.625	2000 – 2022
	Sea salt surface mass conc. PM25 (SSSMASS25)	0.5×0.625	2000 – 2022
	Air Density	0.5×0.625	2000 – 2022
	Planetary Boundary Layer Height (PBL)	0.5×0.625	2000 – 2022
	Sea Level Pressure (SLP)	0.5×0.625	2000 – 2022
PurpleAir data CRU	$PM_{2.5}$Temperature and Precipitation	Station based Gridded (0.5×0.5)	2000 – 2022

Table 2: Multivariate regression model for aerosol emission rate (μg/m^2/s) Variables (AEV).

	Coefficient	Estimate	Std. Error	z	Pr (>|z|)	Sign. Level	AIC
Abuja	(Intercept)	0.247	1.927	0.128	0.89800		135.32
	BCEM	695.841	148.506	4.686	0.00000	***	
	OCEM	-428.646	92.599	-4.629	0.00000	***	
	SUEM	3.561	1.562	2.280	0.02260	*	
Kano	(Intercept)	-3.777	1.814	-2.082	0.03738	*	189.44
	DUEM	7.565	3.517	2.151	0.03151	*	
	BCEM	32.275	10.063	3.207	0.00134	**	
	OCEM	-16.791	7.364	-2.280	0.02260	*	
	SUEM	3.849	1.591	2.419	0.01556	*	
Lagos	(Intercept)	-2.556	0.843	-3.032	0.00243	**	193.76
	BCEM	91.890	17.621	5.215	0.00000	***	
	OCEM	-72.534	17.719	-4.094	0.00004	***	
	SUEM	6.018	1.411	4.265	0.00002	***	
	SSEM	-9.062	1.366	-6.635	0.00000	***	
	DUEM	4.309	1.490	2.892	0.00382	**	

(Significant levels: 0 (***), 0.001(**). 0.01 (*), 0.05 ($^-$))

The binary input variable of the model was designed by setting values of $PM_{2.5}$ based on the World Health Organization's (WHO) threshold limits of 35 $\mu g/m^3$ [8]. Consequently, values of the $PM_{2.5} < 35\ \mu g/m^3$ were set to 0, while values $>= 35\mu g/m^3$ were set to 1 in each study location.

The value of p is the log of odds, and the odds are a function of p which is assumed to be linear [29] (Equation (3)). The relationship between the probability of the dependent variable being in class 1 and the independent variables is expressed through the logit function [30]. This is illustrated by Equation (3), which simplifies the relationship by taking the natural logarithm of the odds ratio [29]. Essentially, it helps to transform the probability of $PM_{2.5}$ exceeding the threshold into a form that can be easily modeled using a linear equation given by:

$$\text{logit}(\log(odds)) = logit\,(p(y = 1\,|X)\,) = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n = ln\left(\frac{p}{1-p}\right). \quad (3)$$

The value of p was converted to another categorical variable using the value of 0.5, which was considered as fitted variable, and compared with the original x values to determine the prediction accuracy, specificity, sensitivity and precision of the model. Similarly, for $p \geq 0.5$ and $p < 0.5$, the value of p was converted to 1 and 0 respectively. The accuracy, sensitivity, specificity, and precision of the model was determined as:

$$\text{accuracy} = \frac{TP + TN}{N}, \quad (4)$$

Table 3: Multivariate Logistic Regression Model (MLRM) for Meteorological Variables (MEV).

	Coefficient	Estimate	Std. Error	z	Pr(>|z|)	Sign. Level	AIC
Abuja	(Intercept)	15.174	4.513	3.362	0.00077	***	108.52
	SLP	-4.996	2.542	-1.966	0.04933	*	
	RH	-12.700	4.674	-2.717	0.00658	**	
	TMP	6.268	4.817	1.301	0.19322		
	PBLH	-6.949	3.307	-2.101	0.03563	*	
	WS	-4.645	1.702	-2.728	0.00636	**	
	P	-3.837	2.597	-1.478	0.13951		
Kano	(Intercept)	20.379	5.656	3.603	0.00031	***	99.13
	SLP	-7.852	5.595	-1.403	0.16054		
	RH	-13.979	4.672	-2.992	0.00277	**	
	TMP	3.448	6.819	0.506	0.61311		
	PBLH	-5.219	5.465	-0.955	0.33966		
	WS	-6.917	2.284	-3.028	0.00246	**	
	P	-3.220	2.796	-1.152	0.24951		
Lagos	(Intercept)	14.571	4.965	2.935	0.00333	**	93.35
	SLP	-2.824	2.652	-1.065	0.28692		
	RH	-11.935	4.909	-2.431	0.01506	*	
	TMP	5.989	3.054	1.961	0.04988	*	
	PBLH	-5.672	2.762	-2.054	0.04001	*	
	WS	-0.710	2.564	-0.277	0.78188		
	P	-21.399	4.243	-5.043	0.00000	***	

(Significant levels: 0 (***), 0.001(**). 0.01 (*), 0.05 (⁻))

Table 4: Confusion matrix of Aerosol Emission Rate (AEV) and Meteorological Variables (MEV) logistic regression model.

Variables	Actual Value	Abuja		Kano		Lagos	
AEV	VALUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE
	0	114	4	41	27	140	17
	1	22	136	19	189	20	99
MEV	VALUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE
	0	107	11	62	6	144	13
	1	10	148	13	195	6	113

$$\text{sensitivity} = \frac{TP}{(TP + FN)}, \tag{5}$$

$$\text{specificity} = \frac{TN}{(TN + FP)}, \tag{6}$$

$$\text{precision} = \frac{TP}{(TP + FP)}, \tag{7}$$

where TN (true negative) is the number of samples actually classified as negative and were correctly predicted as negative, TP (true positive) is the number of samples that are actually classified as positive and were predicted as positive, FN (false negative) is the number of samples that are actually positive but predicted as negative, FP (false positive) is the number of samples that are actually negative but predicted as positive, and N is the total number of observations.

In this study, the whole time series had 276 observations of monthly data records from the years 2000 to 2022. We partitioned the data into 70% and 30% for training and testing, respectively. The multivariate logistic regression algorithm developed in R-Lab was applied in this study. Two classes of response variables were considered. They are meteorological variables (MEV) and aerosol emission variables (AEV). The meteorological variables (MEV) include relative humidity (RH), wind speed (WS), precipitation (P), temperature (TMP), sea level pressure (SLP), and planetary boundary layer height (PBLH). The aerosol emission variables (AEV) include dust emissions (DUEM), black carbon emissions (BCEM), organic carbon emissions (OCEM), sea salt emissions (SSEM), and sulfate emissions (SUEM). The AEV and MEV model PTE in each of the study sites.

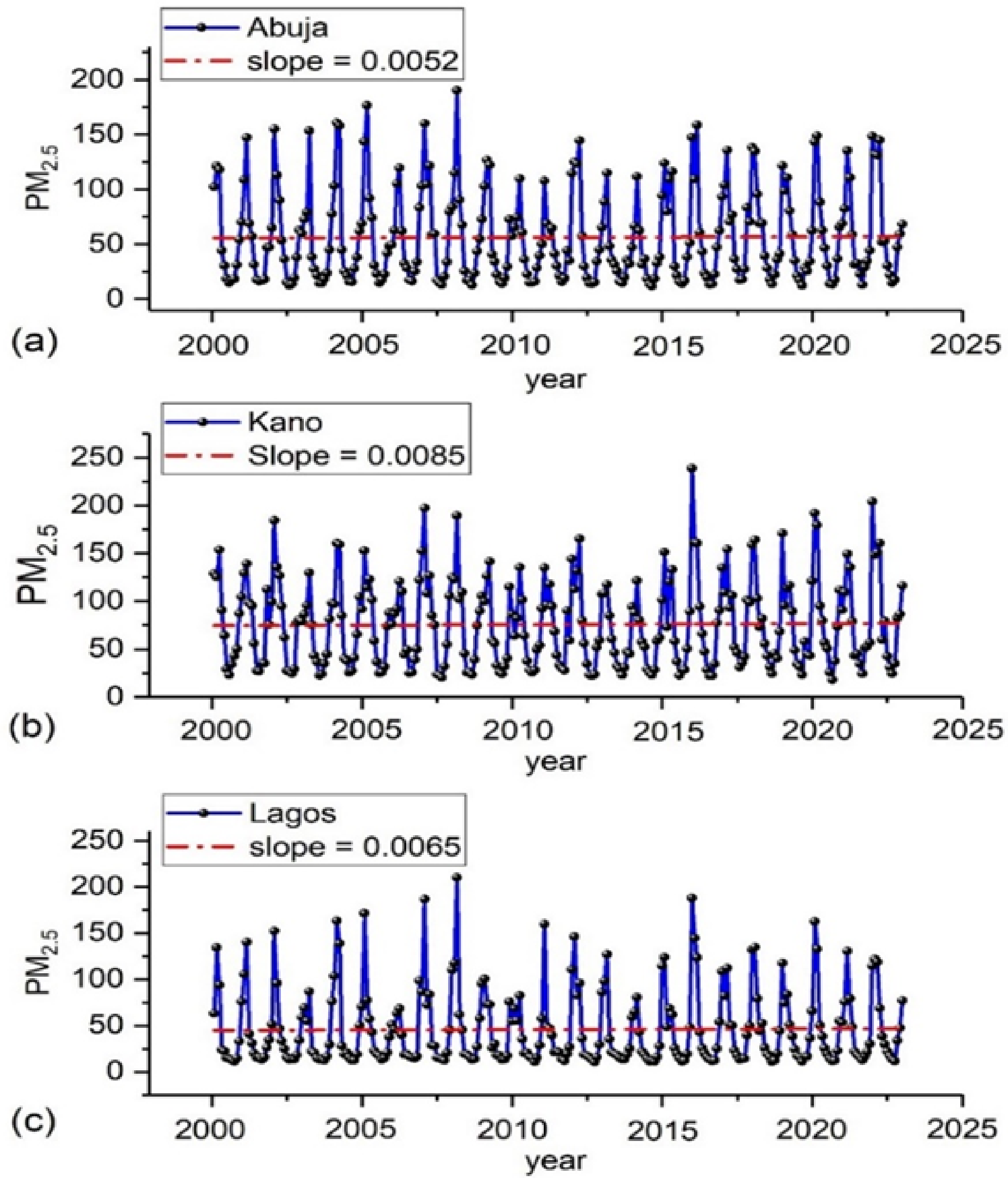

Figure 2: Annual distribution of $PM_{2.5}$ from 2000 to 2022.

3. Results and discussion

3.1. Spatial and temporal distribution of $PM_{2.5}$ in the study area

Figure 2 shows the annual average concentration of $PM_{2.5}$ (μg/m^3) from 2000–2022, in all the study sites. The findings showed significant monthly variations in the aerosol concentration of $PM_{2.5}$ in all the states. The maximum value was recorded in Kano State and the minimum in Abuja and Lagos State. The linear regression slopes, which indicate the consistent increase in the concentration of $PM_{2.5}$ over the months, were 0.0052, 0.0085, and 0.0065 μg/m^3/month in Abuja, Kano, and Lagos, respectively.

A similar distribution pattern is seen for BC and OC, with high loads in the southwest and south-south states and moderate to low loads in the north-central and northwest states (Figures 3a and 3b). A high distribution of SU (Figure 3c) is seen in the south, north, and north-central states, with scattered low values in the northwest states, especially around Sokoto State. The distribution of sea salt (Figure 3d) is confined mainly to states close to the ocean (the Atlantic and Gulf of Guinea). The distribution of DU is similar to that of $PM_{2.5}$, with high values in northern areas and low values in southern areas, as shown in Figure 3e.

The results presented in Figure 4 illustrate the mean concentrations of various aerosol species, the percentage composition of aerosol concentrations in each state, and the percentage composition of aerosol species at the study sites. Dust (DU) has the highest percentage concentration of $PM_{2.5}$ across all states, with values of 42.8 μg/m³ in Abuja, 61.8 μg/m³ in Kano, and 30.8 μg/m³ in

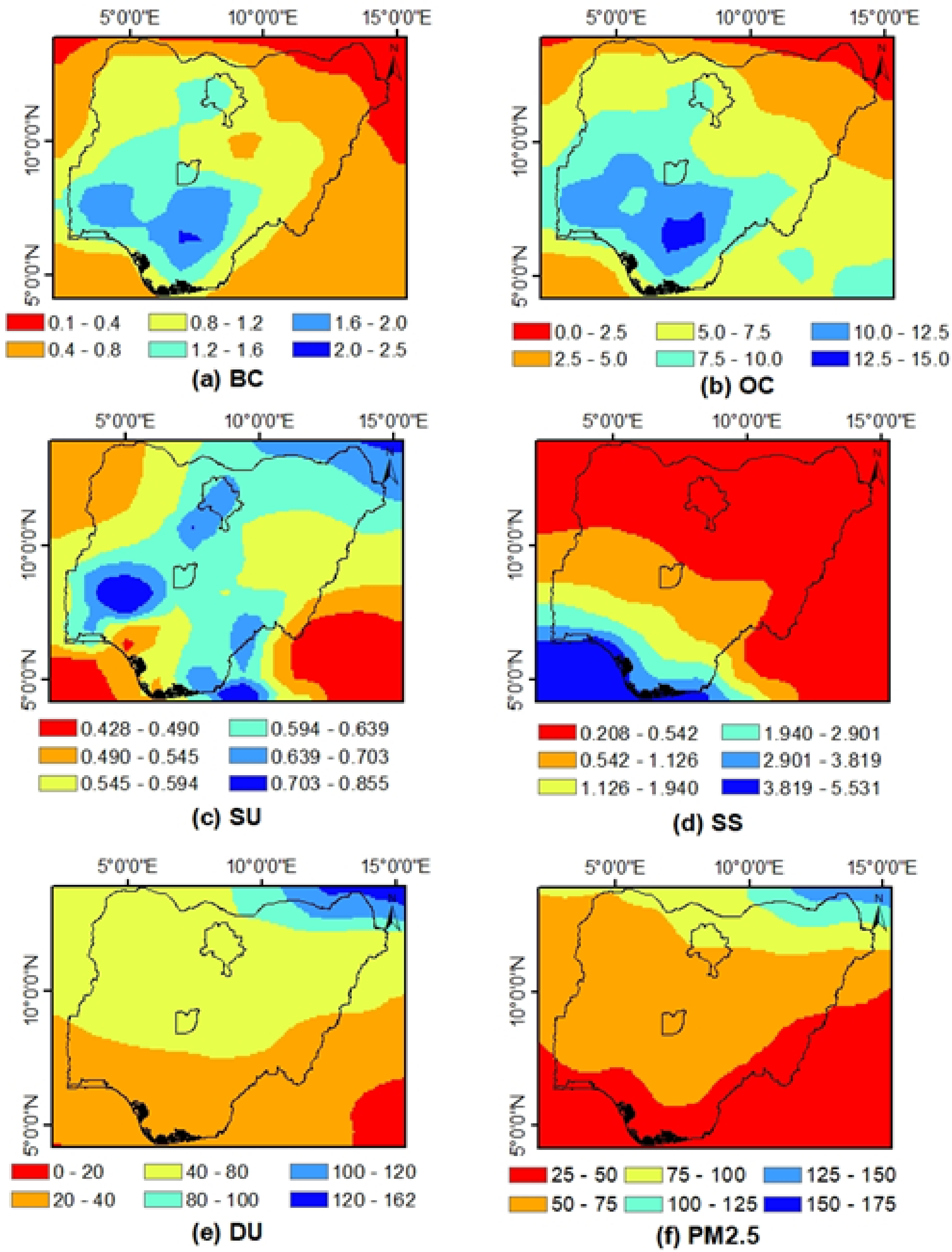

Figure 3: Mean spatial distribution of (a) Black Carbon (BC), (b) Organic Carbon (OC), (c) Sulfate (SU), (d) Sea Salt (SS), (e) Dust (DU), and (f) $PM_{2.5}$

Lagos. Next, OC concentrations were 8.9 µg/m³ in Abuja, 7.3 µg/m³ in Kano, and 8.5 µg/m³ in Lagos. Sea salt (SS) concentrations were lower, measuring 0.7 µg/m³ in Abuja, 0.3 µg/m³ in Kano, and 3.8 µg/m³ in Lagos. Kano had the highest concentration of DU, followed by Abuja and Lagos. The percentage composition of BC, DU, OC, SS, and SU in $PM_{2.5}$ over Abuja was 2, 79, 16, 1, and 1%, respectively. In Kano, it was 2, 87, 10, 0, and 1%, respectively. For Lagos, it is 2, 69, 19, 8, and 1%, respectively. Sulaymon *et al.* [5] reported that out of the highest and lowest ambient $PM_{2.5}$ concentrations recorded in winter (142 µg/m^3) and summer (84 µg/m^3), respectively, it was suggested that dust contributed up to 40.5% of the total $PM_{2.5}$ mass.

3.2. *Multivariate Logistic Regression Model (MLRM) to investigate influencing factors of $PM_{2.5}$ threshold exceedance*

Table 2 shows the Multivariate Regression Model for Aerosol Emission Variables (AEV), which considers the effects of aerosol emission rates and meteorological variables on PTE, where the z-value is obtained by dividing the estimate by the standard error.

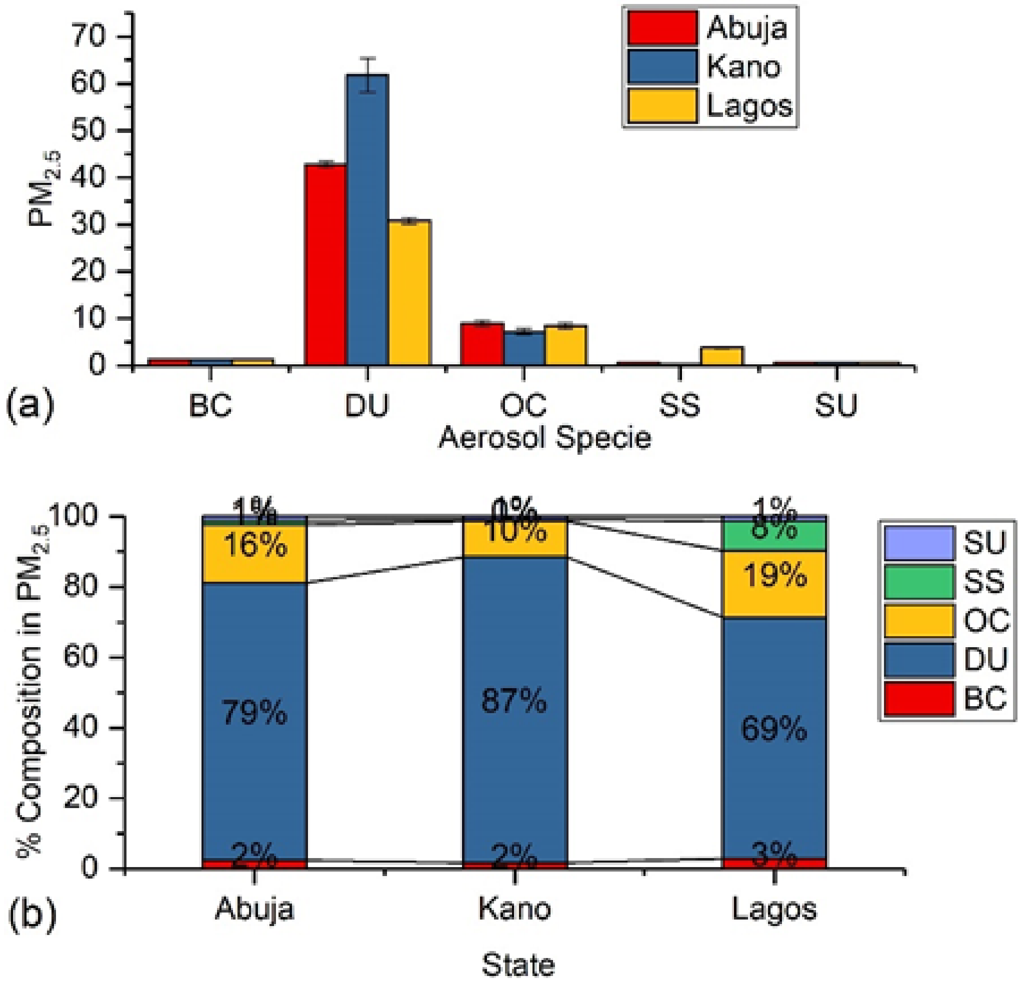

Figure 4: (a) Mean aerosol concentration species with standard deviation (STD) as error bar. (b) Percentage composition of aerosol species in study sites.

For Abuja, both mean DUEM and SSEM variables were zero in all months, while Kano's SSEM was zero in all months for aerosol emission rates; hence, they were removed from the model. However, OCEM was negatively correlated with PTE in Abuja, Kano, and Lagos, with MLRM coefficients of -428.646, -16.791, and -72.534, respectively. Also, SSEM was significantly negative for PTE in Lagos (coefficient: -9.062). On the other hand, BCEM, SUEM, and DUEM were positively correlated with PTE in Abuja and Lagos. The model with the lowest AIC, which represents the best fit, occurred in Abuja. Figure 4 shows that the ambient concentration of DU aerosol is highest in each state; however, the emission rate of DU (DUEM) is zero in Abuja. This suggests that the contribution of DU to ambient $PM_{2.5}$ in Abuja is primarily due to long-range transport from remote DU sources. The 120-h backward trajectories results showed that there was evidence of long-range regional transport of $PM_{2.5}$ into the Lugbe area during the sampling period [5].

Table 3 shows the Multivariate Logistic Regression Model (MLRM) for Meteorological Variables (MEV), with the z-value obtained by dividing the estimate by the standard error. The result showed a negative relationship between all the variables except temperature and PTE in all states. SLP, RH, PBLH, and WS were significantly negative for PTE in Abuja, while in Kano, only RH and WS were significantly negative for PTE. Likewise, RH, PBLH, and P were significantly negative for PTE in Lagos. This implies that when SLP, RH, PBLH, WS, and P are increased, $PM_{2.5}$ is expected to decrease in the study areas. On the other hand, TMP was significantly positive in Lagos but insignificantly positive in Abuja and Kano. This means that TMP over Lagos, when increased, directly increases PTE. The lowest AIC, which indicates the best fit for the model, was recorded in Lagos. Table 4 displays the confusion matrix of the emission rate and meteorological variables in the logistic regression model (MLRM) using Equations 4–7. The summary gave the correct and incorrect predictions made during modeling. The matrix was therefore used to estimate the Performance of the models in terms of accuracy, Sensitivity, Specificity and Precision as displayed in Table 5. From the results, the AEV model showed the highest precision in Abuja, while the MEV model showed the highest precision in Kano and Lagos.

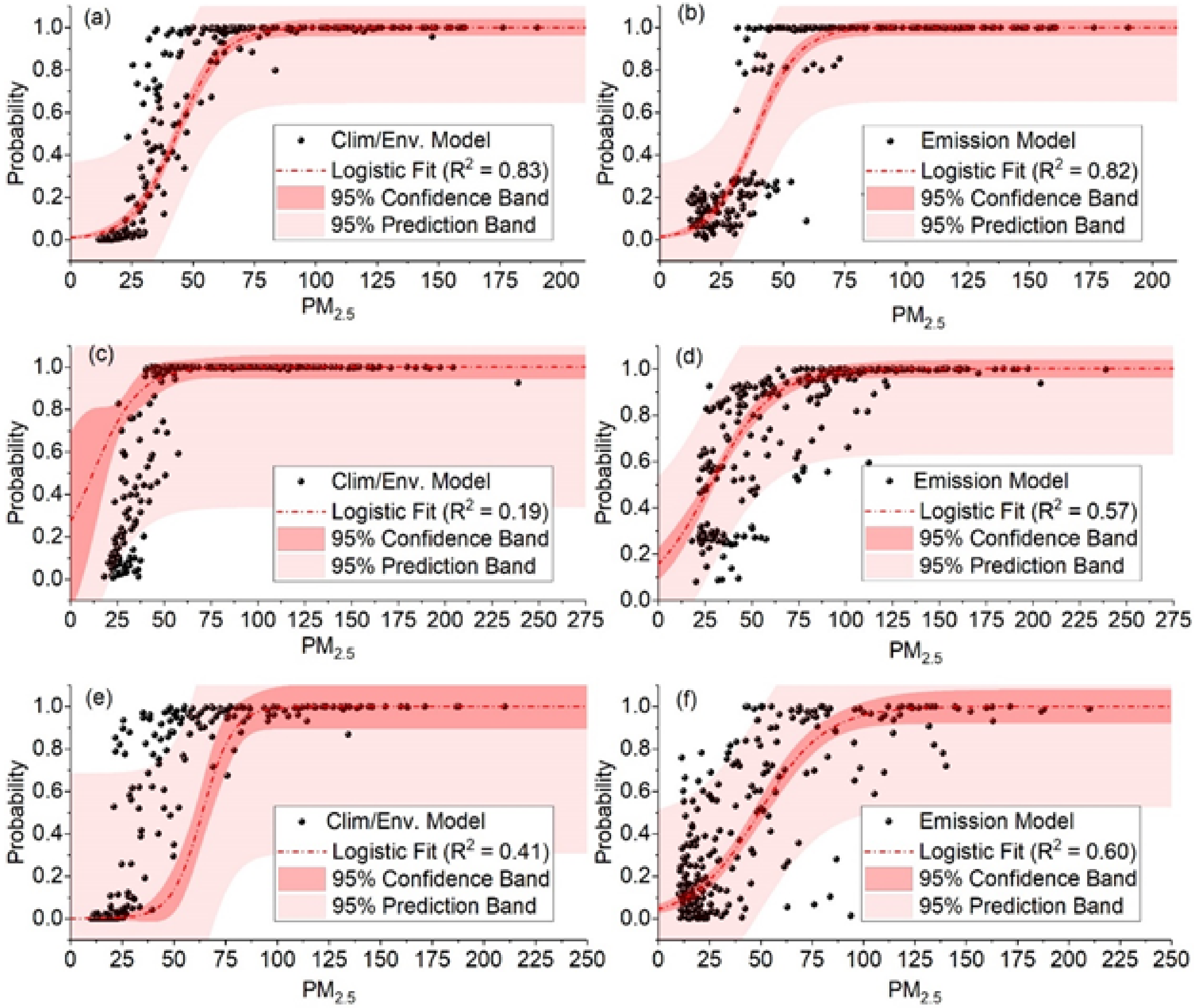

Figure 5: Predicted Probabilities versus $PM_{2.5}$ concentration showing Logistic fit of AEV and MEV model with confidence and prediction bands.

Zhang and Jiang [10] reported that meteorological factors including wind velocity, rainfall, air temperature, soil temperature, and soil humidity have a negative correlation with $PM_{2.5}$ concentration, but in the current work, wind speed and precipitation have a negative correlation with $PM_{2.5}$ while temperature has a positive correlation with $PM_{2.5}$. Xu *et al.* [12] also confirmed in their report that TMP strongly correlates with an increase in $PM_{2.5}$.

Figure 5 shows the predicted probability of the outcome (usually coded as 1 or 0) for MEV and AEV in each state on the y-axis and the dependent variable ($PM_{2.5}$) on the x-axis. Each point on the plot is an observation in the data set, and its position on the plot is the predicted probability of the outcome based on the Logistic Regression Model (LRM). The results showed that high clusters of $PM_{2.5}$ values < 35 µg/m^3 were associated with lower predicted probabilities close to 0. Similarly, $PM_{2.5}$ values >= 35 µg/m^3 had high clusters close to the outcome coded as 1, which tend to have higher predicted probabilities.

3.3. Validation of satellite-derived data with observation data

3.3.1. ERA-5 precipitation and temperature with ground-based CRU observation Data

The performance evaluation of precipitation and temperature over the three cities in Nigeria using Taylor diagrams is represented in Figure 6(a)–(f). The metrics used by these diagrams include the standard deviation, the root mean square error (RMSE), and the correlation coefficient (r). The reference variable (REF), illustrated in the Taylor diagrams (Figure 6), was the ground observation data used for validating precipitation, temperature, and $PM_{2.5}$. For temperature over Abuja, the RMSE of 4.38 shown on the diagram

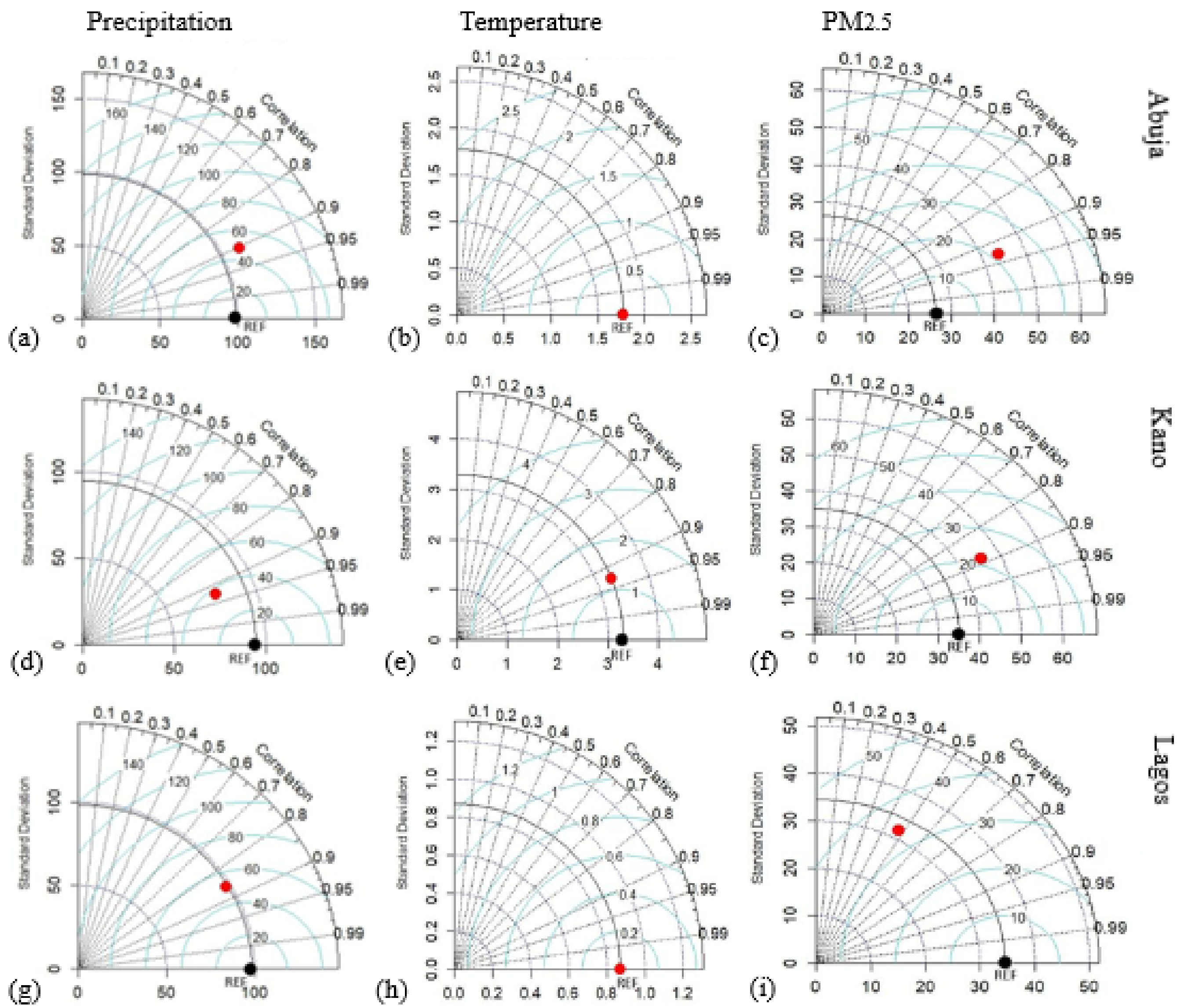

Figure 6: Comparison of satellite precipitation, temperature and $PM_{2.5}$ data with in situ observation data. The ground observation data for climate (P and TMP) and $PM_{2.5}$ are CRU and PurpleAir data, respectively.

Table 5: Model Evaluation for AEV and MEV logistic regression model.

Variables	Metric	Abuja	Kano	Lagos
AEV	Accuracy	90.58	83.33	86.59
	Sensitivity	86.07	90.86	83.19
	Specificity	96.61	60.29	89.17
	Precision	97.14	87.50	85.34
MEV	Accuracy	92.39	93.12	93.12
	Sensitivity	93.67	93.75	94.95
	Specificity	90.67	91.17	91.72
	Precision	93.08	97.01	89.68

indicates that, on average, there is a difference of 4.38 °C between the satellite and the observed data, and the correlation coefficient of 0.99 implies a very strong positive linear correlation. For Kano, the larger RMSE value of 6.1 °C indicates the satellite data is larger than the observation data by this amount, or vice versa on average. However, the correlation coefficient of 0.93 still denotes a very strong relationship. Lagos, as shown in the bottom row, has a much lower RMSE of 2.6 °C, indicating that, on average, predictions for this city are much more accurate. Lagos also has the same very high correlation coefficient of 0.99 as Abuja, indicating strong agreement between satellite and actual values.

For satellite precipitation data, distinctive patterns emerged. For Abuja and Kano, the relatively high RMSE of up to 40 mm/month suggests substantial prediction errors; however, the correlation coefficient of 0.91 (Abuja) and 0.92 (Kano) indicated a strong positive linear relationship. Conversely, our findings in Lagos show a higher RMSE of up to 50 mm/month, suggesting larger prediction errors compared to both Abuja and Kano. However, the remarkably high correlation coefficient of 0.99 indicates an exceptionally strong positive linear relationship, suggesting that despite higher RMSE, the satellite performance closely aligns with the observed values.

3.3.2. Comparison of $PM_{2.5}$ satellite data with ground-based PurpleAir observation data

The validation statistics are shown in Figure 6. It can be noticed that the statistics are quite different for each city. For Abuja, the model showed an RMSE of 22.0 and a correlation coefficient of 0.93. For Kano, the RMSE was 35.8 and the correlation coefficient was 0.89. For Lagos, the RMSE was 34.34 and the correlation coefficient was 0.48. The statistics give a good insight into how the data performed and how well it represents the $PM_{2.5}$ concentrations in the study locations. The model-predicted $PM_{2.5}$ concentrations showed different levels of skill for the three states.

4. Conclusion

A multivariate logistic regression analysis was carried out for the period 2000–2022 to examine the influencing factors for $PM_{2.5}$ threshold exceedance (PTE) in Nigeria. The results showed that BCEM and SUEM were significant at the level of $P < 0.01$ for PTE in Abuja and Lagos, implying that they can potentially enhance the concentrations of $PM_{2.5}$ pollution in the states. Besides, DUEM was significant at the level of $P < 0.01$ for PTE in Kano and Lagos. On the other hand, OCEM was negatively correlated with PTE at the level of $P < 0.01$ for all three states of Abuja, Kano, and Lagos. Therefore, OCEM is not the main cause of the high concentrations of $PM_{2.5}$ in the study states. Meteorological variables (MEV) including sea level pressure (SLP), relative humidity (RH), planetary boundary layer height (PBLH), and wind speed (WS) were significant at the levels of $P < 0.05$ or $P < 0.01$ for PTE in Abuja, and the negative correlations were opposite to the change trends of $PM_{2.5}$. This implies that the increase of those MEVs resulted in a decrease in $PM_{2.5}$ concentrations in the study states. Only RH and WS were significant at the level of $P < 0.05$ for PTE in Kano, and RH, PBLH, and P were significant at the level of $P < 0.01$ for PTE in Lagos. This indicated that the increase of RH, PBLH, and P resulted in a decrease in $PM_{2.5}$ concentrations in the study states. TMP was significant at the level of $P < 0.01$ for PTE in Lagos, but its partial regression coefficient was not significant for Abuja and Kano. This implies that higher TMP in Lagos directly contributes to the increase of PTE in the city. In general, the results emphasize the urgent need for periodic, systematic, and continuous assessment of air quality in the cities of Nigeria and thus should be considered in policy implementation to reduce emissions in the country, as Nigeria joined the Conference of Parties (COP) 21 Paris Agreement in 2015 to tackle climate change.

Acknowledgement

We sincerely thank the anonymous reviewers for their insightful comments and suggestions, which significantly enhanced our manuscript. We also appreciate the support and guidance from the editorial team throughout the review process.

References

[1] R. Sun, Y. Zhou, J. Wu & Z. Gong, "Influencing factors of $PM_{2.5}$ pollution: Disaster points of meteorological factors", International Journal of Environmental Research and Public Health **16** (2019) 3891. https://doi.org/10.3390/ijerph16203891.

[2] T. Amnuaylojaroen & N. Parasin, "Future health risk assessment of exposure to $PM_{2.5}$ in different age groups of children in Northern Thailand", Toxics **11** (2023) 291. https://doi.org/10.3390/toxics11030291.

[3] A. U. Abidin, F. B. Maziya, S. H. Susetyo, M. Yoneda & Y. Matsui, "Exposure particulate matter ($PM_{2.5}$) and health risk assessment on informal workers in landfill site, Indonesia", Environmental Challenges **13** (2023) 100795. https://doi.org/10.1016/j.envc.2023.100795.

[4] S. Gao, Z. Cong, H. Yu, Y. Sun, J. Mao, H. Zhang & Z. Bai, "Estimation of background concentration of PM in Beijing using a statistical integrated approach", Atmospheric Pollution Research **10** (2018) 858. https://doi.org/10.1016/j.apr.2018.12.014.

[5] I. D. Sulaymon, X. Mei, S. Yang, S. Chen, Y. Zhang, P. K. Hopke, J. J. Schauer & Y. Zhang, "$PM_{2.5}$ in Abuja, Nigeria: Chemical characterization, source apportionment, temporal variations, transport pathways and the health risks assessment", Atmospheric Research **273** (2019) 104833. https://doi.org/10.1016/j.atmosres.2019.104833.

[6] J. Liu, G. Kiesewetter, Z. Klimont, J. Cofala, C. Heyes, W. Schöpp, T. Zhu, G. Cao, A. G. Sanabria, R. Sander, F. Guo, Q. Zhang, B. Nguyen, I. Bertok, P. Rafaj & M. Amann, "Mitigation pathways of air pollution from residential emissions in the Beijing-Tianjin-Hebei region in China", Environmental International **125** (2019) 236. https://doi.org/10.1016/j.envint.2018.09.059.

[7] T. Ayua, A. Tyovenda, T. Sombo, E. V. Tikyaa & T. Igbawua, "Fine particulate matter and heavy metals pollution status in ambient air of some selected industrial sites in Northern Nigeria", Journal of Geoscience and Environment Protection **8** (2020) 1. https://doi.org/10.4236/gep.2020.88001.

[8] J. Soares, A. González Ortiz, A. Gsella, J. Horálek, D. Plass & S. Kienzler, "Health risk assessment of air pollution and the impact of the new WHO guidelines (Eionet Report – ETC HE 2022/10)", European Topic Centre on Human Health and Environment. Retrieved from https://iris.who.int/bitstream/handle/10665/345329/9789240034228-eng.pdfonMarch30,2024.

[9] E. M. M. Zahran, “Calibration of background concentrations versus the use of grid sources in air pollution dispersion modelling”, Asian Journal of Natural and Applied Sciences **4** (2015) 1. http://www.ajsc.leena-luna.co.jp/AJSCPDFs/Vol.4(1)/AJSC2015(4.1-01).pdf.

[10] Y. Zhang & W. Jiang, “Pollution characteristics and influencing factors of atmospheric particulate matter ($PM_{2.5}$) in Chang-Zhu-Tan area”, IOP Conference Series: Earth and Environmental Science **108** (2018) 042047. https://doi.org/10.1088/1755-1315/108/4/042047.

[11] H. Yang, Q. Peng, J. Zhou, G. Song & X. Gong, “The unidirectional causality influence of factors on $PM_{2.5}$ in Shenyang city of China”, Scientific Reports **10** (2020) 1. https://doi.org/10.1038/s41598-020-65391-5.

[12] X. Xu, K. Shi, Z. Huang & J. Shen, “What factors dominate the change of PM2.5 in the world from 2000 to 2019? A study from multi-source data”, International Journal of Environmental Research and Public Health **3** (2023) 2282. https://doi.org/10.3390/ijerph20032282.

[13] I. D. Sulaymon, Y. Zhang, P. K. Hopke, F. Ye, K. Gong, J. Mao, & J. Hu, “Modeling PM2.5 during severe atmospheric pollution episode in Lagos, Nigeria: Spatiotemporal variations, source apportionment, and meteorological influences”, Journal of Geophysical Research: Atmospheres **128** (2023) 2022JD038360. https://doi.org/10.1029/2022JD038360.

[14] J. M. Angulo & A. E. Madrid, “Structural analysis of Spatio-temporal threshold exceedances”, Environmetrics **21** (2009) 415. https://doi.org/10.1002/env.1018.

[15] A. E. Madrid, J. M. Angulo & J. Mateu, ”Spatial threshold exceedance analysis through marked point processes.” Environmetrics **23** (2011) 108. https://doi.org/10.1002/env.1141.

[16] R. L. Smith, J. A. Tawn & S. G. Coles, “Markov chain models for threshold exceedances”, Biometrika **84** (1997) 249. http://www.jstor.org/stable/2337455.

[17] S. E. Grineski, R. A. Renteria, T. W. Collins, A. V. Bakian, D. Bilder, J. A. VanDerslice, A. Frase, J. Gomez & K. D. Ramos, “$PM_{2.5}$ threshold exceedances during the prenatal period and risk of intellectual disability”, Journal of Exposure Science & Environmental Epidemiology **34** (2024) s41370-023-00634-w. https://doi.org/10.1038/s41370-024-00647-0.

[18] G. Rincon, G. Morantes Quintana, A. Gonzalez, Y. Buitrago, J. C. Gonzalez, C. Molina & B. Jones, “PM2.5 exceedances and source appointment as inputs for an early warning system”, Environmental Geochemistry and Health **44** (2022) 4569. https://doi.org/10.1007/s10653-021-01189-2.

[19] M. Idris, A. G. Muhammad, R. S. Said & D. O. Akpootu, “Effect of meteorological parameters on the dispersion of vehicular emission in some selected areas in Kano State, Nigeria”, Bayero Journal of Pure and Applied Sciences **13** (2022) 517. https://doi.org/10.1234/bjpas.2022.13.1.517.

[20] N. M. Wambebe & X. Duan “Air quality levels and health risk assessment of particulate matters in Abuja Municipal Area, Nigeria”, Atmosphere **11** (2020) 817. https://doi.org/10.3390/atmos11080817.

[21] R. Kanee, A. Adeyemi, D. Edokpa & P. Ede, “Particulate matter-based air quality index estimate for Abuja, Nigeria: Implications for health”, Journal of Geoscience and Environment Protection **8** (2020) 313. https://doi.org/10.4236/gep.2020.85019.

[22] Global Modeling and Assimilation Office (GMAO), “tavgM_2d_aer_Nx: MERRA-2 2D, Monthly Mean, Time-Averaged, Aerosol Diagnostics (0.625x0.5, L42), version 5.12.4.”, Greenbelt, MD, USA: Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC), 2015. https://doi.org/10.5067/FH9A0MLJPC7N. (Accessed February 23, 2023)

[23] Global Modeling and Assimilation Office (GMAO), “tavgM_2d_flx_Nx: MERRA-2 2D, Monthly Mean, Time-Averaged, Surface Flux Diagnostics (0.625x0.5, L42), version 5.12.4.”, Greenbelt, MD, USA: Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC), 2015. https://doi.org/10.5067/0JRLVL8YV2Y4. (Accessed March 5, 2023)

[24] Global Modeling and Assimilation Office (GMAO), “instM_3d_asm_Np: MERRA-2 3D IAU State, Meteorology Instantaneous, Monthly Mean (p-coord, 0.5x0.625, L42), version 5.12.4.”, Greenbelt, MD, USA: Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC). https://doi.org/10.5067/2E096JV59PK7. (Accessed July 31, 2024)

[25] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee & J. -N. Thépaut, “ERA5 Monthly Averaged Data on Single Levels from 1940 to Present”, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2023. https://doi.org/10.24381/cds.f17050d7. (Accessed: Aug. 28, 2023)

[26] I. Harris, T. J. Osborn, P. Jones & D.H. Lister, “Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset”, Sci Data **34** 7 (2020) 109. https://doi.org/10.1002/joc.3711.

[27] J. L. Hand, S. A. Copeland, D. E. Day, A. M. Dillner, H. Indresand, W. C. Malm, C. E. McDade, C. T. Moore, M. L. Pitchford, B. A. Schichtel & J. G. Watson, “Spatial and seasonal patterns and temporal variability of haze and its constituents in the United States: Report V June 2011”, Interagency Monitoring of Protected Visual Environments **35** (2011) 3119977. https://vista.cira.colostate.edu/Improve/spatial-and-seasonal-patterns-and-temporal-variability-of-haze-and-its-constituents-in-the-united-states-report-v-june-2011/.

[28] L. M. Tesfaw, & H. M. Fenta, “Multivariate logistic regression analysis on the association between anthropometric indicators of under-five children in Nigeria: NDHS 2018”, BMC Pediatrics **21** (2021) 193. https://doi.org/10.1186/s12887-021-02657-5.

[29] G. Bordeaux & A. Couto, “A binary logistic regression model to identify key aspects that enhance global hub airports status”, J. Air Transport Res. Soc. **3** (2024) 100035. https://doi.org/10.1016/j.jatrs.2024.100035.

[30] G. Zeng, “A comprehensive study of coefficient signs in weighted logistic regression”, Heliyon **10** (2024) e35040. https://doi.org/10.1016/j.heliyon.2024.e35.