https://doi.org/10.1038/s44407-024-00003-4

Nitrogen dioxide (NO_2) Meteorology and predictability for air quality management using TROPOMI

Prince Junior Asilevi[1] ✉, Enock Nyameasem Dzidzorm[1], Patrick Boakye[2] & Emmanuel Quansah[1]

Nitrogen dioxide (NO_2) is a critical air pollutant and key indicator for air quality. Due to limited monitoring, we leveraged TROPOMI NO_2 and NASA POWER meteorological datasets to evaluate the meteorological drivers on NO_2 tropospheric column concentrations and to develop predictive models for NO_2 levels over Ghana. Employing an 8:2 ratio for model training and testing, NO_2 and meteorology relationships were assessed by seasonality indices and correlation analyses. Results indicate marked seasonal variability in NO_2 columns, prominent during the dry season. Wind speed, relative humidity, and precipitation significantly reduce NO_2, whereas temperature correlated positively in the southern forested zone. Predictive models demonstrate varying efficacy across climatic zones, with mean percentage differences ranging 9.87 to 37.76% and agreement index up to 0.96. The Random Forest and XGBoost models showed outstanding performance, with correlation reaching 0.92. This results presents a scalable methodology for NO_2 monitoring providing insights for air quality management.

Surface nitrogen dioxide (NO_2) is a criteria air pollutant impacting the environment and public health. Major sources of nitrogen oxides (NOx $\equiv NO + NO_2$) include fossil fuel combustion, open biomass burning, and natural sources[1,2]. In the atmosphere, volatile organic compounds (VOCs) are oxidized by hydroxyl radicals (OH), forming peroxy radicals (RO_2 or HO_2), which react with nitric oxide (NO) to produce nitrogen dioxide (NO_2). NO_2 can then be photolyzed to form ozone (O_3). Peroxyacetyl nitrate (PAN) forms when NO_2 reacts with acetyl peroxy radicals, while nitric acid (HNO_3) is formed when NO_2 reacts with OH. O_3 is a significant contributor to smog formation, while nitric acid contributes to acid rain[1,3].

Furthermore, NO_2 can transform into nitrates, contributing to particulate matter pollution, which is linked to respiratory diseases, premature deaths, and hospitalizations, especially in vulnerable populations[4–6]. The World Health Organization (WHO) recently set a stricter annual exposure limit of 10 μg/m³ for NO_2 to mitigate these health risks[7]. Thus, accurate NO_2 monitoring and prediction are vital for effective air quality management.

However, traditional NO_2 monitoring methods are limited by high costs and complex operations, resulting in inadequate spatial data coverage[8,9]. The short atmospheric lifetime of NO_2, coupled with its high spatial and temporal variability, complicates exposure assessments and highlights the need for regional scale monitoring strategies[10,11]. Emerging tools like low-cost sensors (LCSs) offer affordable alternatives but often require frequent calibration due to accuracy issues[8,12–14]. Satellite remote sensing provides extensive spatial coverage, with instruments like NASA's Aura satellite and the Sentinel 5 Precursor used to study tropospheric NO_2 column trends over large areas[15,16]. While these methods improve spatial resolution, combining satellite data with ground based measurements and predictive models can significantly improve forecasting surface NO_2 forecasting.

Considering the influence of weather conditions on NO_2 dispersion and transformations, leveraging meteorological data for NO_2 prediction could afford an innovative and cost effective approach[6,17]. This study evaluates the potential of using local climatology to predict tropospheric NO_2 column concentrations, accounting for seasonal patterns and spatial distribution, providing a scalable solution for NO_2 monitoring in areas with limited measurement infrastructure.

Results and discussions

Annual Mean Spatial NO_2 distribution

Significant variability in annual mean and year-to-year NO_2 distribution reaching $\sim 4 \times 10^{15}$ moleculescm^{-2} from 2019 to 2023 is depicted in Fig. 1a–e, attributive to changes in anthropogenic activities, climate variability, and biomass burning events[18,19]. Generally, NO_2 concentrations are highest over the most urbanized and industrialized capital i.e. Accra (5.559° N, 0.197° W). These areas are characterized by dense traffic, industrial activities, and higher population densities, contributing to elevated NO_2 emissions. Following 2019, particularly in 2020 (Fig. 1b) and 2021 (Fig. 1c), a

[1]Department of Meteorology and Climate Science, Kwame Nkrumah University of Science and Technology, PMB, Kumasi, Ghana. [2]Department of Chemical Engineering, Kwame Nkrumah University of Science and Technology, PMB, Kumasi, Ghana. ✉e-mail: pjasilevi@knust.edu.gh

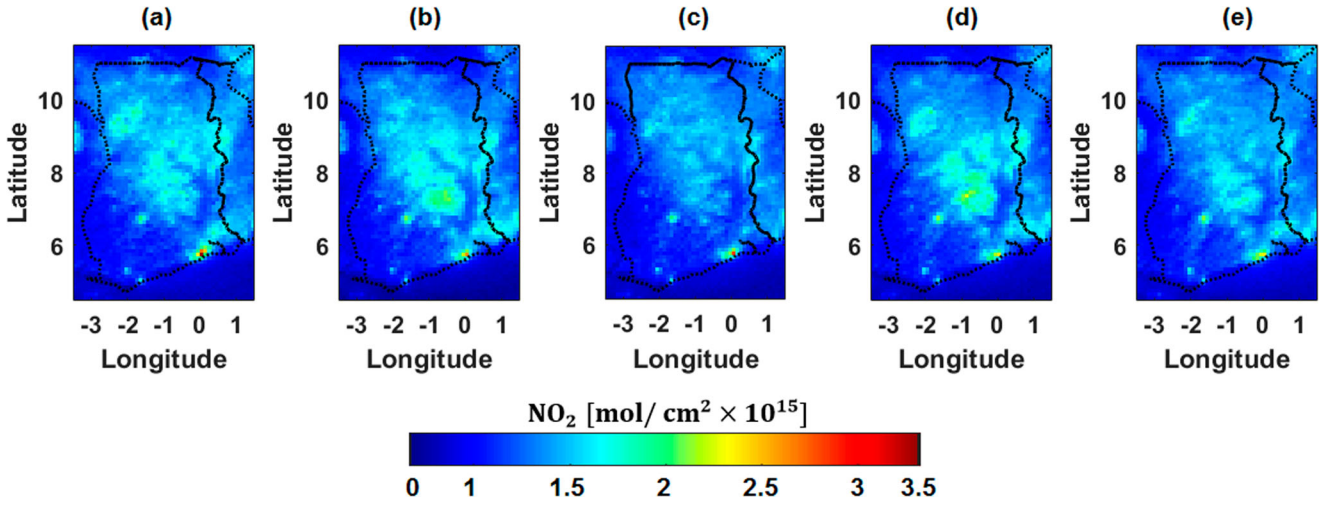

Fig. 1 | Annual mean NO_2 distribution over Ghana during 2019 to 2023. a–e Spatial distribution of the annual mean NO_2 column over Ghana for the years 2019–2023. Red areas indicate regions of higher NO_2 concentrations.

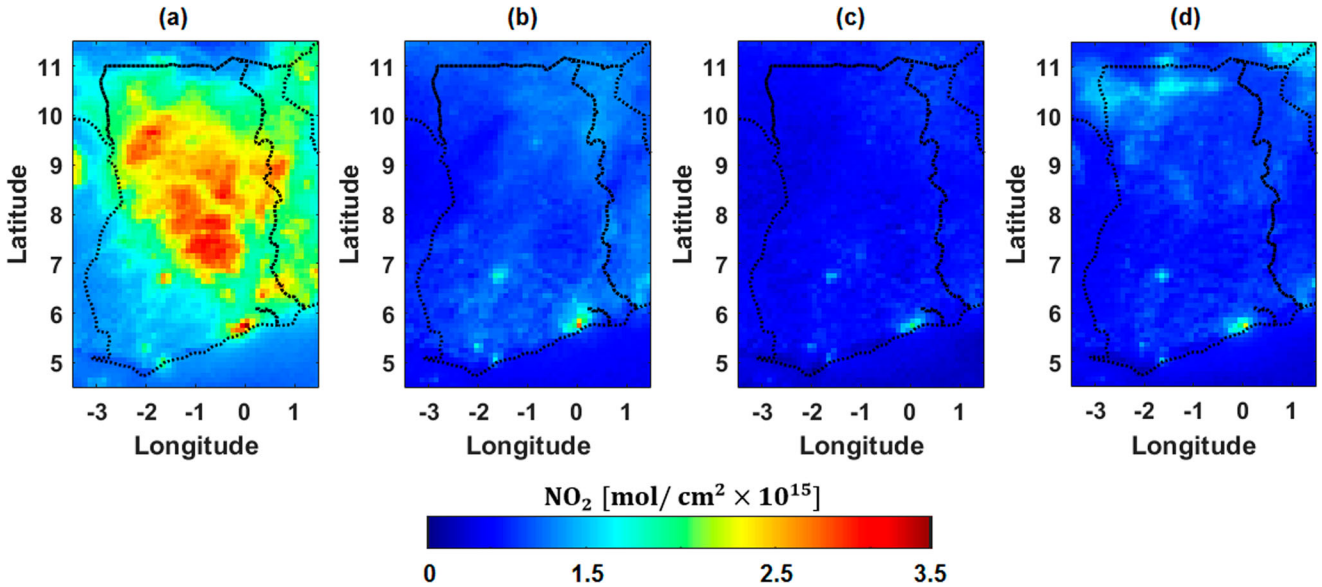

Fig. 2 | Seasonal mean NO_2 distribution across Ghana from 2019 to 2023. The seasonal distribution of NO_2 across Ghana, highlighting variations during different seasons over the period from 2019 to 2023. **a** Mean NO_2 concentrations for December–February (DJF). **b** Mean NO_2 concentrations for March–May (MAM). **c** Mean NO_2 concentrations for June–August (JJA). **d** Mean NO_2 concentrations for September–November (SON). Red areas denote higher NO_2 concentrations, indicating potential hotspots for emission sources or stagnant atmospheric conditions.

relatively suppressed concentration is noticed, being partially explained by the impact of the COVID-19 pandemic, which led to reduced vehicular traffic, industrial slowdown, and a general decrease in fossil fuel combustion activities due to lockdown measures and reduced economic activities.

The spatial distribution of NO_2 also shows regional variability, with higher concentrations generally detected in southern Ghana, where major cities like Accra and Kumasi are located. The coastal regions around Accra, consistently show higher NO_2 levels all-years-around likely due to the combined effects of traffic emissions, industrial activities, and maritime transport influences. The inland areas, particularly the northern regions, such as Tamale (9.403° N, 0.842° W) and Wa (10.060° N, 2.509° W), display comparatively lower NO_2 concentrations, reflecting the lower population density and less intensive industrial and vehicular emissions. The observed higher concentrations in urban areas are consistent with the known relationship between NO_2 and urban heat islands, where increased temperatures can enhance photochemical reactions, leading to higher levels of secondary pollutants like ozone (O_3) and PANs, as well as NO_2 itself[20,21].

Monthly and Seasonal NO_2 evolution

The seasonal NO_2 distribution depicted in Fig. 2a–d highlight distinct spatiotemporal variability for December to February (DJF), March to May (MAM), June to August (JJA), and September to November (SON) during 2019 to 2023. NO_2 column concentrations are at the peak during the DJF season across the study area, with pronounced levels around the highly urbanized and industrial regions like Accra. The elevated NO_2 concentrations during DJF is attributive to increased combustion activities, including open biomass burning, which is notably prevalent in the Savannah and arid areas during the dry season[19]. Furthermore, the DJF period in West Africa is characterized by stable atmospheric conditions, reduced convective activity, and minimal precipitation. These meteorological conditions limit the dispersion of pollutants, facilitating the accumulation of tropospheric column NO_2[22].

This seasonal pattern aligns with the MODIS burnt area map,[19] which indicates that the DJF dry season corresponds to a period of heightened biomass burning and increased NO_2 emissions across the northern

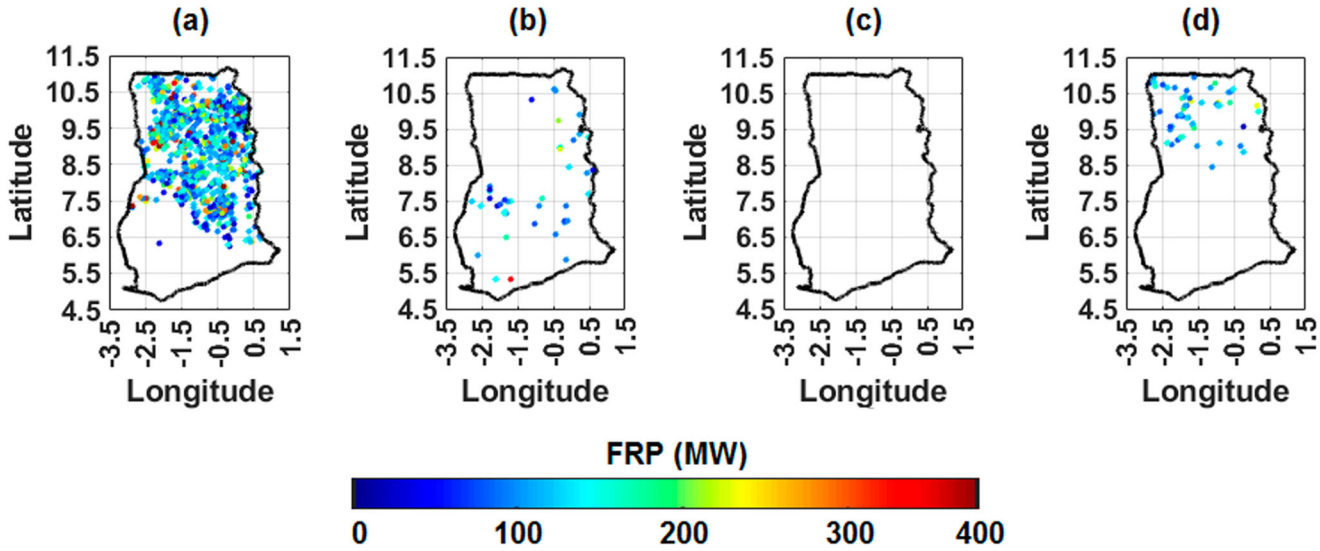

Fig. 3 | Seasonal distribution of 100% confidence MODIS FRP over Ghana. Fire activity patterns observed through MODIS Fire Radiative Power (FRP) for the seasons **a** DJF, **b** MAM, **c** JJA, and **d** SON during the period 2019–2022.

Savannah region, thereby contributing to elevated pollution levels. Moreover, the lack of convective mixing and a stable boundary layer during this dry season inhibit the vertical transport of pollutants, exacerbating the accumulation of NO_2. The transition to the rainy season during MAM (Fig. 2b) marks a significant shift in NO_2 distribution dynamics. Increased precipitation and convective activity characteristic of this period enhance pollutant dispersion, leading to reduced NO_2 column concentrations, particularly in the northern and central parts. Rainfall is an effective sink for NO_2 by facilitating wet deposition, while enhanced convection supports vertical mixing, thus diminishing NO_2 concentrations. However, despite the overall reduction, urban hotspots in the southern part, including Accra, continue to exhibit relatively elevated NO_2 levels, driven by persistent vehicular and industrial emissions.

The JJA season, being the peak rainy season, is marked by the lowest NO_2 concentrations (Fig. 2c). The frequent rain events, convective activities, and increased atmospheric instability characteristic of this period effectively remove NO_2 from the atmosphere through washout processes and enhance dispersion. Moreover, the wet and humid atmospheric conditions during JJA are less conducive to the formation and accumulation of NO_2, resulting in uniformly low values across all regions, including major urban areas.

During the SON season, the rains gradually subside (Fig. 2d), and there is a notable resurgence in NO_2 concentrations, particularly from the northern regions. The decline in rainfall frequency and the onset of more stable atmospheric conditions during this period reduce the efficiency of pollutant removal processes, facilitating a gradual accumulation of NO_2. Biomass burning activities, which tend to resume as conditions become drier, also play a significant role in the increasing NO_2 concentrations observed towards the end of the year. The seasonal transition observed during SON is likely influenced by regional meteorological dynamics, such as the Harmattan winds, which are prevalent during the dry down period and affect pollutant distribution and transport[19,21].

Figure 3 illustrates the fire activity observed by the MODIS Fire Radiative Power (FRP). The seasonal NO_2 distribution from December to February (Fig. 2a) aligns with increased fire activity observed in the MODIS FRP data (Fig. 3a). High NO_2 concentrations in the northern half during the dry season correspond with elevated FRP values, indicating significant NO_2 emissions from biomass burning. Conversely, lower NO_2 levels from March to August coincide with reduced fire activity, underscoring the key role of biomass burning in NO_2 variability, particularly during the dry season[19,22].

Figure 4 presents the seasonality evaluation results, comparing NO_2 seasonality at selected stations using the Seasonality Index (SI) and monthly NO_2 column concentrations. Figure 4a shows Kete Krachi and Wa exhibit the highest SI values, signifying substantial seasonal variability in NO_2 levels. This variability connects strongly with seasonal biomass burning prominent during the dry season, as well as shifts in local meteorology. This trend is further corroborated by Fig. 4b, showing elevated NO_2 concentrations during the dry months (December to February). The high standard deviations for Kete Krachi (9.50×10^{14} mol/cm^2) and Wa (7.76×10^{14} mol/cm^2) further support this substantial variability, indicating significant month-to-month changes in NO_2 levels. In contrast, Axim has the lowest SI value among all the stations, indicating limited seasonal variation in NO_2 levels. Axim is a densely forested area situated in the rain hub of Ghana, and least urbanized and industrialized.

Thus consistently low NO_2 levels reflects the absence of large scale anthropogenic activities, such as industry or dense vehicular emissions. This supported by Fig. 4b, where Axim shows minimal variation throughout the year, and by its low standard deviation (2.01×10^{14} mol/cm^2). Whereas Kumasi and Accra show comparatively lower SI values than Kete Krachi and Wa (Fig. 4a), there is noticeable monthly variability (Fig. 4b). In particular, Accra exhibits pronounced increases in NO_2 concentrations during the dry months, suggesting that seasonal effects such as harmattan and biomass burning contribute to variations in NO_2. The standard deviations for Accra (5.89×10^{14} mol/cm^2) and Kumasi (5.60×10^{14} mol/cm^2) indicate considerable variability, reinforcing the observation that these urban centers experience dry seasonal peaks linked to both continuous and episodic emission sources.

The SI metric provides a useful indication of the degree of seasonality, while Fig. 4b shows that Accra, Kumasi, and Kete Krachi all exhibit peaks in NO_2 concentrations during the dry season. Thus, while the SI values for Accra and Kumasi are lower compared to Kete Krachi and Wa, indicating a blend of continuous emissions and seasonal influences, the monthly data in Fig. 4b highlight significant dry season peaks in these cities. This suggests that moderate SI values can represent meaningful seasonal changes, especially in urban regions where NO_2 sources are a mixture of continuous industrial emissions and episodic contributions like seasonal biomass burning. This highlights the significant impact of both meteorological conditions and seasonal biomass burning activities on NO_2 distribution in Ghana.

NO_2 and Meteorology relationship

To establish the meteorological conditions associated with the mean NO_2 concentration distribution as a proxy for NO_2 prediction, a correlation

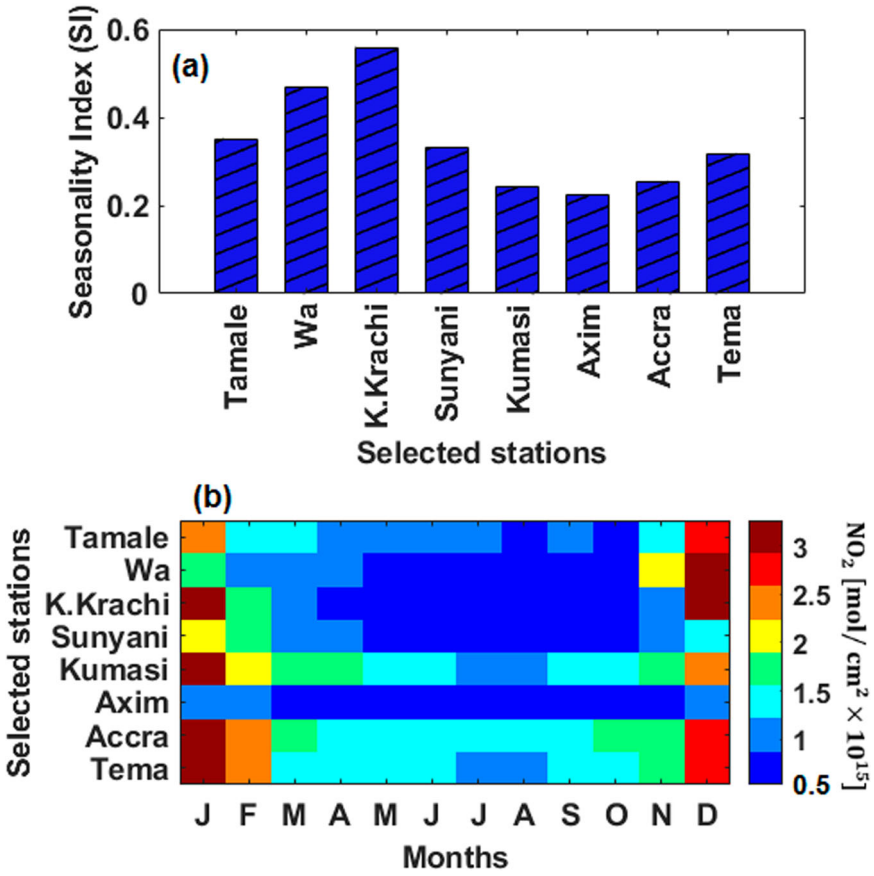

Fig. 4 | NO_2 seasonality patterns for Ghana from 2019 to 2023. a Seasonality index comparison for selected stations and **b** time-series evaluation of monthly NO_2 column concentrations.

matrix for NO_2 at selected sites representative of the climatic zones with fourteen (14) climatological parameters is presented in Fig. 5.

In the Savannah zone, NO_2 concentrations show a negative correlation with average temperature, particularly in Wa (−0.33) and Tamale (−0.20). This suggests that elevated temperatures may enhance photochemical reactions, leading to increased NO_2 conversion into other species such as ozone. Moreover, higher temperatures can support stronger convective activity, resulting in more effective vertical redistribution of NO_2. The minimum temperature exhibits a strong negative correlation with NO_2 in both Wa and Tamale (−0.75 and −0.82, respectively), indicating that lower nocturnal temperatures might contribute to increased NO_2 levels, possibly due to temperature inversions that trap pollutants near the surface[23]. The negative correlation with relative humidity in Wa and Tamale (−0.51 and −0.79) aligns with the expectation that higher moisture levels, often linked with precipitation and cloud cover, reduce NO_2 through wet deposition processes. Wind speed at both 2 meters and 10 meters shows a positive correlation in Wa (0.25 and 0.37) but a stronger positive correlation in Tamale (0.42 and 0.49), suggesting that higher wind speeds contribute to the dispersal of NO_2. The negative correlation with wind direction in both locations (−0.70 and −0.88) may reflect the influence of the prevailing monsoonal reversing wind patterns on pollutant distribution. Additionally, the negative correlation with cloud amount and precipitation in both stations suggests that these factors are crucial in reducing NO_2 levels in the Savannah zone.

In the Transition zone (Sunyani and Kete Krachi), NO_2 positively correlates with temperature (0.61 for average temperature and 0.69 for maximum temperature), suggesting that higher temperatures coincide with increased emissions, likely due to more intensive anthropogenic activities during warmer months. This increase in emissions appears to outweigh the photochemical loss of NO_2 that typically occurs at higher temperatures, leading to a net rise in observed NO_2 levels. However, the minimum temperature still shows a strong negative correlation (−0.80 in Sunyani and −0.88 in Kete Krachi), similar to the Savannah zone, indicating that cooler nights might lead to the accumulation of NO_2. The negative correlation with relative humidity is nearly perfect in Sunyani (−0.99), suggesting that high humidity significantly reduces NO_2 levels, likely through condensation and deposition processes. Wind speed shows a weak or negative correlation with NO_2, indicating that in the Transition zone, wind might not be as effective in dispersing NO_2 as it is in the Savannah zone. The strong negative correlation with wind direction, precipitation, and cloud amount reinforces the idea that these meteorological factors play a significant role in controlling NO_2 concentrations.

The Forest zone (Kumasi and Axim) exhibits a strong positive correlation between NO_2 and temperature, particularly in Kumasi (0.56 for average temperature and 0.65 for maximum temperature). This suggests that higher temperatures coupled with increased vehicular emissions and biomass burning, are linked to elevated NO_2 levels. However, the minimum temperature again shows a weaker negative correlation in Axim (−0.07), implying that nighttime cooling is less effective at trapping NO_2 in the Forest zone compared to other zones. The correlation with relative humidity is strongly negative in both Kumasi and Axim (−0.92 and −0.95), indicating that higher humidity, often associated with rain, reduces NO_2 concentrations significantly. Wind speed shows a consistent negative correlation with NO_2 in both stations, with Axim having a stronger correlation, suggesting that wind plays an important role in dispersing pollutants in the Forest zone. However, Axim shows a weak correlation with precipitation and solar radiation, reflective of local influencing factors such as proximity to the coast and vegetation cover.

Finally, the Coastal zone (Accra and Tema) reveals strong positive correlations between NO_2 and temperature, with both cities showing correlations above 0.60 for average temperature and maximum temperature. This may reflect the integrated effects of urban heat and local emissions,

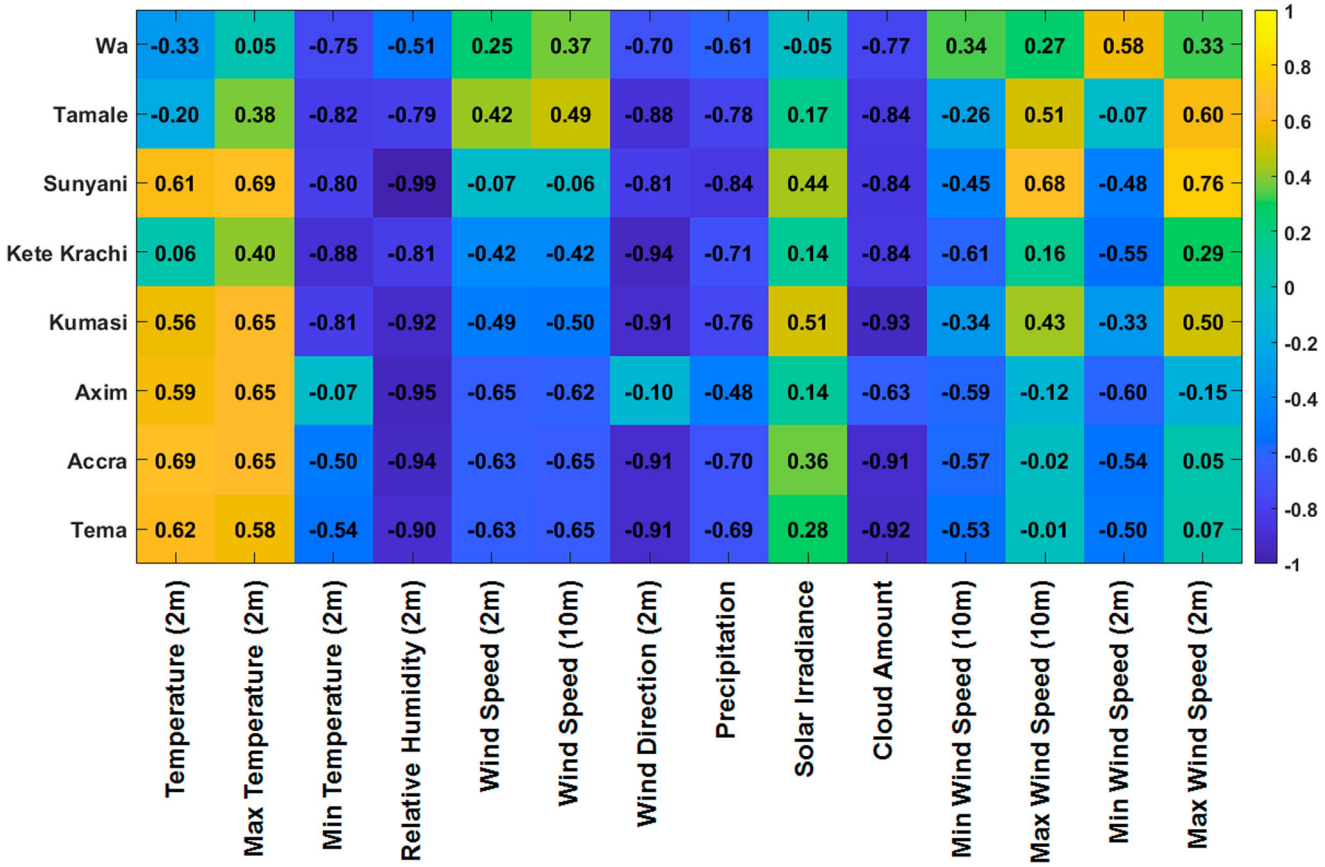

Fig. 5 | Correlation between NO_2 and meteorological parameters at selected sites. Matrix displaying the correlation coefficients between NO_2 concentrations and local meteorological variables for selected locations in Ghana.

where higher temperatures coincide with increased NO_2 levels. The negative correlation with minimum temperature, though weaker than in other zones, suggests that nighttime cooling may help reduce NO_2 concentrations but is less effective in coastal cities. The correlation with relative humidity is strongly negative (−0.94 in Accra and −0.90 in Tema), indicating that higher moisture levels, which are common in coastal areas, play a critical role in NO_2 removal[23]. Wind speed and direction also show strong negative correlations with NO_2, especially in Tema, highlighting the importance of coastal breezes in pollutant dispersion. The moderate positive correlation between NO_2 and solar radiation in Accra suggests that photolytic processes might also influence NO_2 levels in this region.

Meteorology based NO_2 Predictability

The predictive performance of the selected models for NO_2 columns across different climatic zones of Ghana (Savannah, Transition, Forest, and Coastal) is evaluated in Figs. 6, 7, and Table 1. The evaluation leverages statistical metrics including Mean Percentage Difference (MPD), Pearson's correlation coefficient (r), and Willmott's index of agreement (d) to assess how well the models can replicate observed NO_2 concentrations as captured by the TROPOMI instrument.

In the Savannah zone (Fig. 6a), all models exhibit reasonable accuracy, with MPD values ranging from 28.59 to 35.69%. Linear regression, Ridge regression, and XGBoost performed particularly well, each achieving MPD values under 30%, indicating smaller differences from the actual values. Pearson's correlation coefficient (r) for these models remains high (0.84–0.89), with Willmott's index (d) ranging between 0.87 and 0.92. This suggests that the models effectively captured the temporal variability of NO_2, though some bias remains in the seasonal prediction, particularly during the dry season, likely due to variable contributions from biomass burning. The Transition zone (Fig. 6b) presents a unique challenge. MPD values are notably higher than those in other zones, particularly for the Gradient Boosting model (37.76%), indicating challenges in accurately representing NO_2 concentrations. The correlation coefficients (r) are relatively consistent (0.81–0.89), but Willmott's index (d) ranges between 0.85 and 0.93, showing that while temporal dynamics are somewhat captured, higher residuals may be indicative of complex interactions between emission sources and local meteorology that are less effectively modeled.

In the Forest zone (Fig. 6c), Random Forest, XGBoost, and Gradient Boosting models show improved performance compared to the simpler linear and ridge models. The Random Forest and XGBoost models have the lowest MPD values (10.11% and 9.87% respectively), demonstrating excellent predictive capability. High r values (0.92 and 0.91) and d values (0.92 and 0.91) also reflect their capacity to accurately capture temporal NO_2 patterns, particularly during transitions between the wet and dry seasons.

The Coastal zone (Fig. 6d) yields the lowest MPD values for Ridge Regression (13.84%) and XGBoost (15.35%), indicating that these models have relatively better predictive accuracy in this environment, likely due to more stable, continuous emission sources such as traffic and industry. However, the correlation coefficients (r) for all models in this zone are lower (0.66–0.75), and d values are also lower compared to the Forest zone, suggesting that capturing the variability in NO_2 emissions along the coast is inherently challenging. The influence of both anthropogenic and maritime factors leads to greater complexity, which may reduce model accuracy[24,25].

Figure 7 shows the scatterplots comparing predicted NO_2 concentrations to TROPOMI observations across all models and zones. Across all subplots (a–e), the highest consistency between predicted and actual values is observed for the Random Forest and XGBoost models, as indicated by the proximity of scatter points to the 1:1 line and higher r values. Conversely,

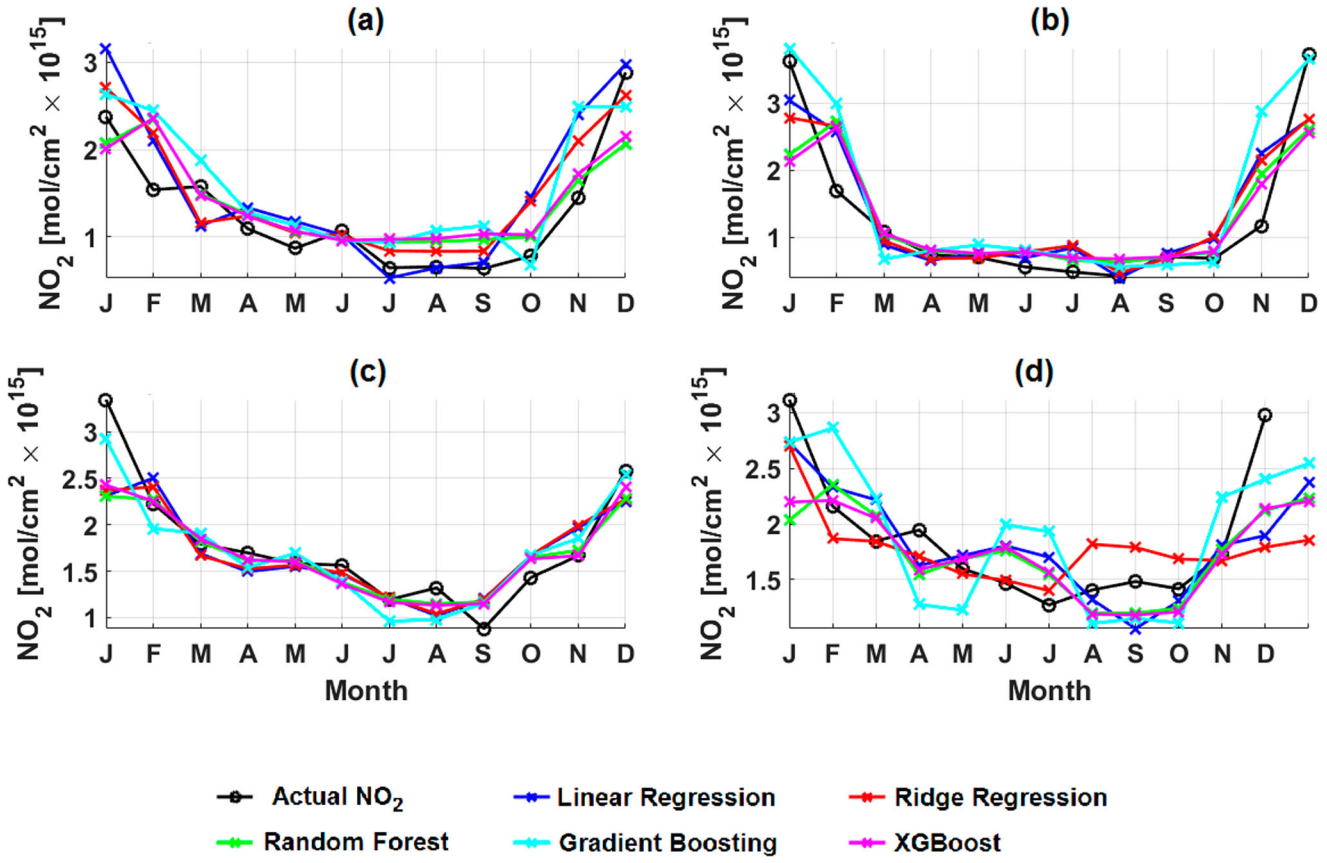

Fig. 6 | Model performance comparison for NO_2 prediction across climatic zones. Predicted NO_2 concentrations across **a** Savannah, **b** Transition, **c** Forest, and **d** Coastal Savannah climatic zones using various meteorology-based models.

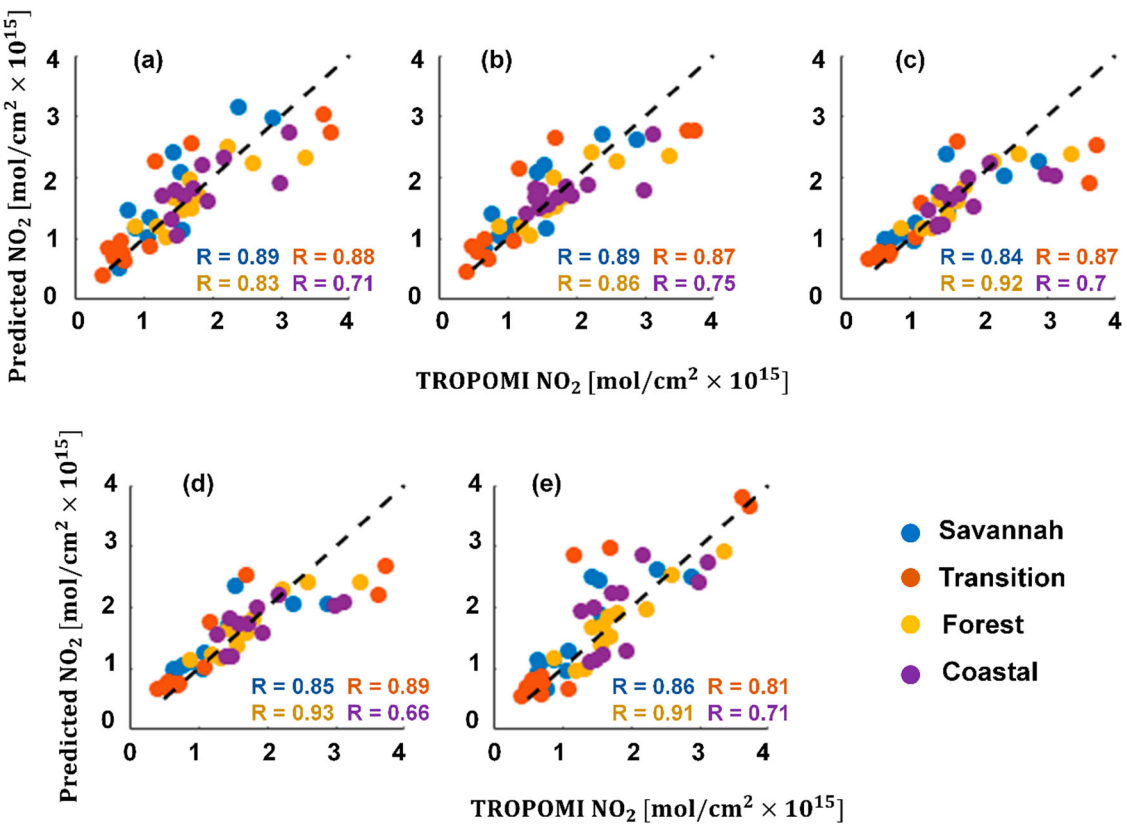

Fig. 7 | Scatterplots of predicted NO_2 versus TROPOMI column NO_2. Performance comparison of **a** Linear regression, **b** Ridge regression, **c** Random forest, **d** Gradient boosting, and **e** XGBoost models in predicting NO_2 concentrations across all four climatic zones.

Table 1 | Statistical summary of NO_2 prediction model performance

Model	Savanna			Transition			Forest			Coastal		
	MPD	r	d	MPD	r	d	MPD	r	d	MPD	r	d
Linear	28.84	0.89	0.91	31.77	0.88	0.93	14.74	0.83	0.88	17	0.71	0.81
Ridge	28.59	0.89	0.92	33.02	0.87	0.91	13.96	0.86	0.89	13.84	0.75	0.77
Random Forest	29.01	0.84	0.87	28.41	0.87	0.89	10.11	0.92	0.92	16.51	0.7	0.74
Gradient Boosting	35.69	0.85	0.87	37.76	0.89	0.92	13.73	0.93	0.96	27.17	0.66	0.81
XGBoost	28.86	0.86	0.89	28.65	0.81	0.85	9.87	0.91	0.91	15.35	0.71	0.73

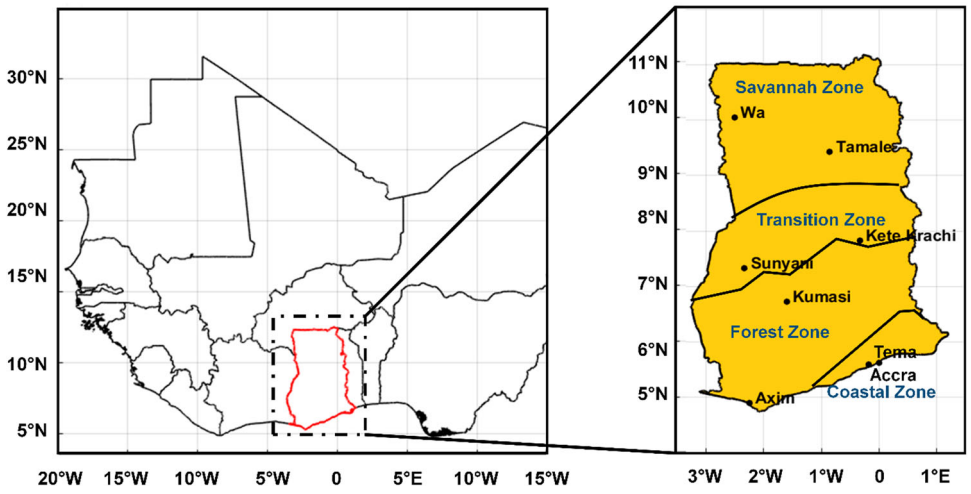

Fig. 8 | Geographical location of the study area, Ghana, in West Africa. Map showing Ghana's position in West Africa, highlighting the four main climatic zones (Savannah, Transition, Forest, and Coastal Savannah) and selected synoptic stations.

Gradient Boosting shows larger discrepancies, which can be attributed to a higher sensitivity to hyperparameters and the risk of overfitting during model training.

Table 1 present summaries of the statistical results comparing the predicted and TROPOMI derived NO_2 based on the 20% monthly meteorology and NO_2 2022 test dataset.

The evaluations presented, highlight that while the selected models especially with Random Forest and XGBoost provide improved predictability, there remain notable limitations related to model generalization across different climatic zones. The Forest zone benefits the most from these models, whereas the Transition and Coastal zones continue to pose significant challenges, primarily due to variability in seasonal meteorological conditions and diverse emission sources. In general, the assessment present the opportunity for further refinement in both model selection and hyperparameter tuning, especially for regions with complex emission and meteorological dynamics.

In summary, this study analyzed NO_2 distribution across Ghana (2019–2023) and evaluated statistical and machine learning models for predicting NO_2 concentrations, using TROPOMI NO_2 data and NASA POWER meteorology. Significant spatiotemporal variability was found, mainly due to anthropogenic activities, climate variability, and biomass burning. Elevated NO_2 levels were observed in urban areas like Accra, particularly during the dry season, influenced by emissions from traffic, industry, and biomass burning. Random Forest and XGBoost models showed strong predictive performance, especially in the Forest zone, while Transition and Coastal zones posed challenges due to complex meteorological and emission factors. Gradient Boosting was sensitive to hyperparameters and prone to overfitting, highlighting the need for careful tuning. Machine learning models offer potential for improved NO_2 predictions in regions with limited monitoring, but further model refinement and better representation of regional emissions and meteorology are needed to enhance prediction accuracy.

Methods

Study area geography and climatology

The study focuses on Ghana in Sub-Sahara West Africa, located at latitudes 4.5° N and 11.5° N and longitudes 3.5° W and 1.5° E, highlighted by the dashed outline in Fig. 8. This geographic location positions it within the tropical climate zone, characterized by distinct wet and dry seasons. The climate varies significantly across the country, influenced by the interplay between the dry, dust laden Harmattan winds from the Sahara Desert to the north and the moist maritime air from the Gulf of Guinea to the south[26,27].

Ghana's climate can be categorized into four main zones: the Savannah, the Transition, the Forest, and Coastal Savannah (Fig. 8). The Savannah zone, located in the northern part of the country, experiences an unimodal rainfall pattern from May to October. The Transition zone, located centrally, marks a shift between the unimodal rainfall in the north and the bimodal rainfall pattern found in the southern Forest zone. The Forest and Coastal Savannah zones, covering the southern part of Ghana, experience two distinct rainy seasons, from March to July and September to November, interspersed with short dry spells. These variations in rainfall patterns, temperature, and humidity across the country present diverse environmental conditions for atmospheric NO_2 dispersion and transformation[28,29]. Selected cities representative of the climate zones and synoptic stations under the Ghana Meteorological Agency, relevant for discussions are highlighted as black dots in Fig. 8.

The climatology is further characterized by its temperature regime, with average temperatures ranging between 21 °C and 32 °C. The coastal areas typically exhibit higher humidity levels and more stable temperatures due to the moderating influence of the Atlantic Ocean. In contrast, the northern Savannah region experiences more extreme temperatures and lower humidity levels, contributing to different atmospheric chemical processes and pollutant dispersion dynamics[23,30,31]. For example,Ogen[23] showed that positive vertical airflow indicative of downward air movement suppressed the dispersion of NO_2 over Western Europe.

Understanding NO_2 distribution requires considering both regional meteorological conditions and localized factors influencing air quality. This study leverages this geographic and climatic diversity to assess NO_2 dynamics, using a combination of satellite data and meteorological parameters to enhance predictive modeling capabilities for air quality management in Ghana.

TROPOMI Nitrogen Dioxide (NO_2) data

The TROPOspheric Monitoring Instrument (TROPOMI), launched on 13 October 2017 as part of the European Space Agency's Sentinel 5 Precursor satellite mission, provides high resolution measurements of atmospheric NO_2 column. Operating in a sun synchronous, low earth orbit (825 km altitude) with a 13:30 local solar time equator crossing, TROPOMI achieves daily global coverage with a 2600 km swath width and a pixel resolution reduced to 3.5×5.6 km² at nadir after August 2019[21,32]. This passive optical sensor relies on solar UV visible radiation, capturing radiance in the 405–465 nm spectral window to derive NO_2 column amounts through differential optical absorption spectroscopy[33].

The NO_2 retrieval process involves converting top-of-atmosphere spectral radiances to slant column densities, followed by separating the tropospheric and stratospheric components using the TM5-MP model and applying air mass factors (AMF) to estimate the vertical column content. The AMF calculation, which influences the retrieval's accuracy, is sensitive to factors such as surface reflectance, NO_2 vertical profiles, and atmospheric scattering. While there is some significant bias (ranging from 20 to 40%) in urban areas due to uncertainties in air mass factor (AMF) calculations, as well as additional uncertainties from other factors such as fitting errors, stratospheric assumptions, and surface reflectance, TROPOMI's data remains valuable for analyzing trends in NO_2 column[34–37].

For this study, we utilized monthly averaged NO_2 data from the NASA Health And Air Quality Applied Science Team (HAQAST) Level 3 GLOBAL TROPOMI NO_2 collections (Version 2.4) at a $0.1° \times 0.1°$ resolution from 2019 to 2023, which provides good spatial and temporal coverage suitable for integrating with NASA's climatology products to evaluate NO_2 meteorology, for modeling and prediction of NO_2 levels in Ghana. The data is accessible at: https://search.earthdata.nasa.gov/search?q=HAQ_TROPOMI_NO2_GLOBAL_M_L3.

The MODIS burnt area data presented by Kugbe[19] indicates that biomass burning significantly contributes to NO_2 levels over Ghana. To further highlight the impact of biomass burning on NO_2 distribution, we retrieved Fire Radiative Power (FRP) data from NASA's Fire Information for Resource Management System (FIRMS). The FRP data, derived from MODIS Aqua/Terra, was used for comparison with NO_2 data, covering the period from 2019 to 2022.

Meteorological Data from NASA POWER

A collage of recent 30 years of climatological reanalysis products comprising 14 parameters was retrieved from the National Aeronautics and Space Administration Prediction of Worldwide Energy Resource (NASA POWER) agro-climatological online data viewer service system at 0.5° grid area (https://power.larc.nasa.gov/data-access-viewer/). The products consisted of Temperature at 2 m (°C), Relative Humidity at 2 m (%), Wind Direction at 2 m (Degrees), Wind Speed at 2 m (m/s), Wind Speed at 10 m (m/s), Maximum Temperature at 2 m (°C), Minimum Temperature at 2 m (°C), Maximum Wind Speed at 2 m (m/s), Minimum Wind Speed at 2 m (m/s), Cloud Amount (%), Maximum Wind Speed at 10 m (m/s), Minimum Wind Speed at 10 m (m/s), Total Rainfall (mm), and All Sky Surface Shortwave Downward Irradiance (MJ/m²/day).

A detailed description of NASA POWER and extensive validation of the cloud and radiation products over Ghana were presented by Asilevi Junior[25] and Quansah,[38] showing Willmott's index of agreement between 0.7 and 0.99 ± 0.01. Alongside, several works have represented NASA POWER's suitability and adaptability for the region of interest[26,39–41].

In this work, the spatial coverage of NASA POWER data was adapted as the regional climatological framework to evaluate the NO_2 meteorology and subsequently develop the meteorology based NO_2 prediction.

Seasonality Index (SI) and NO_2 Predictive Models

Based on the periodicity of NO_2 data, the Seasonality Index (SI) by Walsh and Lawler[42] was adapted as a test tool to ascertain the strength of NO_2 seasonality. This is to ensure that NO_2 has a predictable pattern. The SI is a computational tool developed to quantify the intra-annual distribution of climatic data, which several authors have used for rainfall seasonality studies[43–45]. In the context of NO_2, the SI can be adapted to measure the degree to which NO_2 columns fluctuate throughout the year. A higher SI indicates that NO_2 columns are concentrated in a few months of the year (high seasonality), while a lower SI suggests that NO_2 levels are more evenly distributed across the year (low seasonality). The SI is computed by Eq. 1, where R_i is the total NO_2 concentration for year i, x_{in} is the NO_2 concentration in month n for year i.

$$SI_i = \frac{1}{R_i}\sum_{n=1}^{12}\left|x_{in} - \frac{R_i}{12}\right| \quad (1)$$

The prediction of NO_2 column concentrations leverages a combination of statistical and machine learning models to integrate satellite derived NO_2 column data from TROPOMI with climatological data from NASA POWER. The primary goal is to establish a predictive framework that captures the dynamic interactions between local meteorological conditions and NO_2 levels, providing a scalable solution applicable for monitoring air quality in regions with limited ground based measurements. To achieve this, the study utilizes multiple modeling approaches, each tailored to capture different aspects of the relationship between NO_2 column concentrations and meteorology.

Linear regression is first employed to establish a foundational model that describes the linear relationship between NO_2 columns and meteorological predictors, such as temperature, relative humidity, wind speed, precipitation, solar radiation, and cloud cover. The linear regression model is represented mathematically by Eq. 2:

$$NO_2 = \beta_O + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \quad (2)$$

where β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients for the predictors $X_1, X_2, \ldots, X_n$, and ε is the error term. While linear regression provides a baseline understanding, it may not fully capture the complex, nonlinear interactions between meteorological variables and NO_2 levels[46].

To address potential nonlinearities and multicollinearity among predictors, ridge regression is applied as an extension of linear regression. Ridge regression introduces a regularization parameter that penalizes the size of coefficients, thus reducing model complexity and mitigating overfitting[47,48]. The ridge regression objective function is:

$$\min_\beta\left(\sum_{i=1}^{n}(y_i - X_i \cdot \beta^2) + \lambda\sum_{j=1}^{p}\beta_j^2\right) \quad (3)$$

where λ is a regularization parameter controlling the degree of shrinkage applied to the coefficients. This regularization helps stabilize the estimates when predictors are highly correlated, ensuring more reliable predictions.

For capturing complex, nonlinear interactions between the predictors, random forests are employed. Random forests are an ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction of the individual trees. Each tree in the ensemble is built from a random subset of the data, allowing for diverse decision boundaries and improving the model's generalization capabilities[49].

Gradient boosting is another ensemble learning method applied in this study to further refine NO_2 predictions. Unlike random forests, gradient boosting sequentially builds a series of decision trees, where each new tree

corrects the errors made by the previous ones. The method minimizes a specified loss function by adding new models that predict the residuals of prior models, leading to highly accurate predictions[50]. The model is represented by Eq. 4:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (4)$$

where $F_m(x)$ is the current model, $F_{m-1}(x)$ is the model from the previous step, $h_m(x)$ represents the new decision tree added at stage m, and γ_m is the learning rate. This iterative approach allows the model to focus progressively on the hardest-to-predict data points, enhancing overall performance.

XGBoost, an optimized version of gradient boosting, is also utilized for its overall computational efficiency and improved accuracy. XGBoost incorporates both first and second order derivatives in its optimization process, which allows for faster convergence and better handling of complex data patterns[51,52]. The function representation at step t is:

$$f_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = f_i^{t-1} + f_t(x_i) \quad (5)$$

where $f_k(x_i)$ and $f_t(x_i)$ are the prediction values for the k and t iterations of the XGBoost model respectively; $f_i^{(t)}$ and f_i^{t-1} are the prediction values for the *t*th and *t-1*th iterations of the *i*th sample; x_i is the input variable. Additionally, XGBoost includes regularization techniques to further prevent overfitting, making it particularly suitable for predicting NO_2 concentrations for diverse geographical regions.

The models were trained on a dataset comprising monthly meteorological variables and corresponding NO_2 column concentrations, capturing the local climatology's influence on pollutant dispersion and transformation processes. Specifically, the data was split such that 80% was used for training and 20% was used solely for validation, ensuring that at least 20% of the data was not seen by the model during training. This approach allows for an accurate evaluation of the model's ability to predict NO_2 concentrations when exposed to unseen data, reflecting real world performance. The training dataset comprised the years 2019–2021, while the validation dataset comprised the year 2022. The integration of meteorological observations represents a novel approach to NO_2 monitoring, offering a scalable solution for air quality management.

Statistical and correlation tools

To assess the correlation between NO_2 column and meteorology, as well as the performance of the prediction models, Root mean square difference (RMSD), Mean bias difference (MBD), and Pearson's correlation coefficient (r) for n observations presented in Eqs. 6 to 8 are used where RD is the residual difference between predicted $NO_{2,predicted}$ and observed $NO_{2,Observed}$, and σ is the standard deviation[53,54].

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(RD)^2} \quad (6)$$

$$MBD = \frac{1}{n}\sum_{i=1}^{n}(RD) \quad (7)$$

$$r = \frac{\sum_{i=1}^{n}(NO_{2,Observed} - \sigma_{Observed})(NO_{2,predicted} - \sigma_{Predicted})}{(n-1)\sigma_{Observed}\sigma_{predicted}} \quad (8)$$

The RMSD is a standard statistical measure of the variation margin between observed and predicted data, for which zero in the ideal case and a smaller metric indicative of a low marginal difference[54]. The MBD offers a statistical measure of the spread of estimation from observation, with near zero MBD values being desirable. The Pearson's correlation coefficient (r) was used to measure the strength of NO_2 and local climatology correlation[25].

Data availability

The NASA Health And Air Quality Applied Science Team (HAQAST) Level 3 GLOBAL TROPOMI NO_2 collections (Version 2.4) is available at https://search.earthdata.nasa.gov/search?q=HAQ_TROPOMI_NO2_GLOBAL_M_L3 and the NASA POWER agro-climatological data us available at https://power.larc.nasa.gov/data-access-viewer/). All resulting computational and statistical datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 3 September 2024; Accepted: 12 December 2024;
Published online: 12 March 2025

References

1. Zhu, R. et al. A novel approach to deriving the fine-scale daily NO2 dataset during 2005–2020 in China: Improving spatial resolution and temporal coverage to advance exposure assessment. *Environ Res.* **249**, 118381 (2024).
2. Wan, N. et al. Estimation of biomass burning emission of NO 2 and CO from 2019–2020 Australia fires based on satellite observations. *Atmos Chem.Phys.* **23**, 711–724 (2023).
3. Vîrghileanu, M., Săvulescu, I., Mihai, B. A., Nistor, C. & Dobre, R. Nitrogen Dioxide (NO2) Pollution monitoring with Sentinel-5P satellite imagery over Europe during the coronavirus pandemic outbreak. *Remote Sens* **12**, 3575 (2020).
4. Chianese, E., & Riccio, A. Long-term variation in exposure to NO2 concentrations in the city of Naples, Italy: Results of a citizen science project. *Sci Total Environ*. vol. 931, 172799, (2024).
5. Chaloulakou, A., Mavroidis, I. & Gavriil, I. Compliance with the annual NO2 air quality standard in Athens. required NOx levels and expected health implications. *Atmos Environ.* **42**, 454–465 (2008).
6. Rava, M. et al. A predictive model for the home outdoor exposure to nitrogen dioxide. *Sci.Total Environ.* **384**, 163–170 (2007).
7. W. H. Organization, Air Quality Guidelines–Update 2021," WHO Regional Office for Europe, Copenhagen, Denmark, (2021).
8. Miech, J. A. et al. Calibration of low-cost NO2 sensors through environmental factor correction. *Toxics* **9**, 281 (2021).
9. Zhu, Y. et al. Spatial and temporal representativeness of point measurements for nitrogen dioxide pollution levels in cities. *Atmos Chem Phys.* **20**, 13241–13251 (2020).
10. Korhonen, A. et al. Influence of spatial resolution on population PM 2.5 exposure and health impacts. *AIR QUAL ATMOS HLTH.* **12**, 705–718 (2019).
11. Fenech, S. et al. The influence of model spatial resolution on simulated ozone and fine particulate matter for Europe: implications for health impact assessments. *Atmos Chem Phys.* **18**, 5765–5784 (2018).
12. Westervelt, D. M., Isevulambire, P. K., Yombo Phaka, R., Yang, L. H. & Raheja, G. Milly Low-Cost Investigation into Sources of PM2. *ACS ES&T Air* **1**, 43–51 (2023).
13. Yang, L. H. et al. &amp Investigating the sources of urban air pollution using low-cost air quality sensors at an urban atlanta site. *Environ Sci Technol.* **56**, 7063–7073 (2022).
14. Wei, P. et al. Development and evaluation of a robust temperature sensitive algorithm for long term NO2 gas sensor network data correction. *Atmos Environ.* **230**, 117509 (2020).
15. Xu, J., Zhang, Z., Zhao, X. & Cheng, S. Downward trend of NO2 in the urban areas of Beijing-Tianjin-Hebei region from 2014 to 2020: Comparison of satellite retrievals, ground observations, and emission inventories. *Atmos Environ.* **295**, 119531 (2023).
16. Tzortziou, M. et al. Intimately tracking NO2 pollution over the New York City-Long Island Sound land-water continuum: An integration of shipboard, airborne, satellite observations, and models. *Sci Total Environ.* **897**, 165144 (2023).
17. Bo, M., Salizzoni, P., Pognant, F., Mezzalama, R. & Clerico, M. A combined citizen science-modelling approach for NO2 assessment in Torino urban agglomeration. *Atmos.* **11**, 72 (2020).

18. Griffin, D., et al. Biomass burning nitrogen dioxide emissions derived from space with TROPOMI: methodology and validation. *Atmos. Meas Tech Discuss*. vol. 2021, 1-44, (2021).
19. Kugbe, J. X., Mathias, F., Desta, T. L., Denich, M. & Vlek, P. L. Annual vegetation burns across the northern savanna region of Ghana: period of occurrence, area burns, nutrient losses and emissions. *Nutr Cycling Agroecosyst.* **93**, 265–284 (2012).
20. Jamei, E. et al. Investigating the impacts of COVID-19 lockdown on air quality, surface Urban Heat Island, air temperature and lighting energy consumption in City of Melbourne. *Energy Strat Rev.* **44**, 100963 (2022).
21. Goldberg, D. L. et al. TROPOMI NO2 in the United States: A detailed look at the annual averages, weekly cycles, effects of temperature, and correlation with surface NO2 concentrations. *Earth's futur.* **9**, e2020EF001665 (2021).
22. Wimberly, M. C., Wanyama, D., Doughty, R., Peiro, H. & Crowell, S. Increasing fire activity in African tropical forests is associated with deforestation and climate change. *Geophys Res Lett.* **51**, e2023GL106240 (2024).
23. Ogen, Y. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. *Sci Total Environ.* **726**, 138605 (2020).
24. Wang, J. et al. Nitrogen oxides (NO and NO2) pollution in the Accra metropolis: Spatiotemporal patterns and the role of meteorology. *Sci. Total Environ.* **803**, 149931 (2022).
25. Asilevi Junior, P. et al. Development of High Resolution Cloud Cover Climatology Databank Using Merged Manual and Satellite Datasets over Ghana, West Africa. *Atmos.-Ocean* **60**, 566–579 (2022).
26. Asilevi, P. J. et al. Bias-corrected NASA data for aridity index estimation over tropical climates in Ghana, West Africa. *J Hydrol Reg. Stud.* **51**, 101610 (2024).
27. Ballo, A., Omotosho, J. B., Klutse, N. A. B., Abiodun, B. J. & Coulibaly, A. The influence of quasi-biennial oscillation on West African Rainfall. *MODEL EARTH SYST ENV.* **8**, 1461–1480 (2022).
28. Yamba, E. I. et al. Revisiting the agro-climatic zones of Ghana: a re-classification in conformity with climate change and variability. *PLoS Climate* **2**, e0000023 (2023).
29. Antwi-Agyei, P., Dougill, A. J., Doku-Marfo, J. & Abaidoo, R. C. Understanding climate services for enhancing resilient agricultural systems in Anglophone West Africa: The case of Ghana. *Clim Serv.* **22**, 100218 (2021).
30. Dwamena, H. A., Tawiah, K., & Akuoko Kodua, A. S. The effect of rainfall, temperature, and relative humidity on the yield of cassava, yam, and maize in the Ashanti region of Ghana. *Int. J Agron*., vol. 2022, no. 1, 9077383, (2022).
31. Asilevi, P. J., Quansah, E., Amekudzi, L. K., Annor, T. & Klutse, N. A. B. Modeling the spatial distribution of Global Solar Radiation (GSR) over Ghana using the Ångström-Prescott sunshine duration model. *Scientific African* **4**, e00094 (2019).
32. Veefkind, J. P. et al. TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens Environ.* **120**, 70–83 (2012).
33. Van Geffen, J. et al. S5P TROPOMI NO 2 slant column retrieval: Method, stability, uncertainties and comparisons with OMI. *Atmos Meas Tech.* **13**, 1315–1335 (2020).
34. Judd, L. M. et al. Evaluating Sentinel-5P TROPOMI tropospheric NO 2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound. *Atmos Meas Tech.* **13**, 6113–6140 (2020).
35. Verhoelst, T. et al. Ground-based validation of the Copernicus Sentinel-5p TROPOMI NO 2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks. *Atmos Meas Tech.* **14**, 481–510 (2021).
36. Lorente, A. et al. Structural uncertainty in air mass factor calculation for NO 2 and HCHO satellite. *Atmos Meas Tech.* **10**, 759–782 (2017).
37. Martin, R. V., et al. An improved retrieval of tropospheric nitrogen dioxide from GOME. *Geophys Res Atmos*. vol. 107, no. D20, pp. ACH-9, (2002).
38. Quansah, A. D. et al. Assessment of solar radiation resource from the NASA-POWER reanalysis products for tropical climates in Ghana towards clean energy application. *Sci Rep.* **12**, 10684 (2022).
39. Osei-Mensah, E., et al. Assessing Ghana's renewable energy potential and path to clean electricity based on the levelized cost of electricity from solar and wind. *J Therm Anal Calorim*. 1-13, (2024).
40. Arfasa, G. F., Sekyere, E. O. & Doke, D. A. Temperature and precipitation trend analysis using the CMIP6 model in the Upper East region of Ghana. *All Earth* **36**, 1–14 (2024).
41. Junior, P. A., Quansah, E. & Dogbey, F. Satellite-based estimates of photosynthetically active radiation for tropical ecosystems in Ghana—West Africa. *Trop Ecol.* **63**, 615–625 (2022).
42. Walsh, R. P. D. & Lawler, D. M. Rainfall seasonality: description, spatial patterns and change through time. *Weather* **36**, 201–208 (1981).
43. Stosic, T. et al. Changes in Rainfall Seasonality in Serbia from 1961 to 2020. Theoretical and Applied Climatology. *Theor Appl Climatol.* **155**, 4123–4138 (2024).
44. Abebe, B. A., Grum, B., Degu, A. M. & Goitom, H. Spatio-temporal rainfall variability and trend analysis in the Tekeze-Atbara river basin, northwestern Ethiopia. *Meteorol Appl.* **29**, e2059 (2022).
45. R. S. Teegavarapu. Changes and trends in precipitation extremes and characteristics: Links to climate variability and change. *In Trends and changes in hydroclimatic variables*, 91-148, (2019).
46. Su, X., Yan, X. & Tsai, C. L. Linear regression. *WIREs Comp Stats.* **4**, 275–294 (2012).
47. Grover, R. & Sharma, S. Impact of Climate Change on Rainfall Pattern by using Ridge Regression Analysis. *In 2024 International Conference on Computational Intelligence and Computing Applications (ICCICA)* **1**, 558–563 (2024).
48. McDonald, G. C. Ridge regression. *WIREs Comp Stats.* **1**, 93–100 (2009).
49. Smarra, F. et al. Data-driven model predictive control using random forests for building energy optimization and climate control. *Appl Energy* **226**, 1252–1272 (2018).
50. Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review. *Artif Intell Rev.* **54**, 1937–1967 (2021).
51. Dong, J. et al. Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions of China. *Eng Appl Artif. Intell.* **117**, 105579 (2023).
52. Fan, J. et al. Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China. *Energy Convers Manag.* **164**, 102–111 (2018).
53. Kim, J., Han, K. & Hahn, J. W. Selective dual-band metamaterial perfect absorber for infrared stealth technology. *Sci Rep.* **7**, 1–9 (2017).
54. Chai, T. & Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geosci Model Dev.* **7**, 1247–1250 (2014).

Acknowledgements

We are thankful for the NASA HAQAST Level 3 GLOBAL TROPOMI NO_2 and the NASA POWER climatological reanalysis dataset obtained from the NASA Langley Research Center POWER Project funded through the NASA Earth Science Directorate Applied Science Program.

Author contributions

PAJ conceptualized the work, retrieved the data, and developed draft manuscript. END assisted in draft manuscript writing. PB and EQ supervised and shaped the work, interpreted the results, and were instrumental in

revising the manuscript write-up. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Prince Junior Asilevi.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.