Journal of Geoscience and Environment Protection, 2026, 14(5), 137-169
https://www.scirp.org/journal/gep
ISSN Online: 2327-4344
ISSN Print: 2327-4336

# Spatio-Temporal Modelling of Air Pollution Using Earth Observation and Deep Learning Techniques: A Case Study of Nairobi Metropolitan Area

**Samuel Orwa Odoyo[1*], Solomon Mwanjele Mwagha[1], Arthur W. Sichangi[2]**

[1]School of Science and Informatics (SSI), Taita Taveta University, Voi, Kenya
[2]Institute of Geomatics, GIS & Remote Sensing (IGGReS), Dedan Kimathi University of Technology, Nyeri, Kenya
Email: *samodoyo@gmail.com

**How to cite this paper:** Odoyo, S. O., Mwagha, S. M., & Sichangi, A. W. (2026). Spatio-Temporal Modelling of Air Pollution Using Earth Observation and Deep Learning Techniques: A Case Study of Nairobi Metropolitan Area. *Journal of Geoscience and Environment Protection, 14,* 137-169.
https://doi.org/10.4236/gep.2026.145010

**Received:** March 29, 2026
**Accepted:** May 25, 2026
**Published:** May 28, 2026



## Abstract

**Aims:** This study models key air pollutants ($PM_{10}$, CO, HCHO, $CH_4$, $NO_2$, $O_3$, $SO_2$) across the Nairobi Metropolitan Area using a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) framework for the years 2019-2024, and forecasts their 2025 levels using the NeuralProphet time series model. Random Forest Regression Kriging was incorporated to refine spatial outputs. **Study design:** Retrospective modelling and forecasting study based on multi-source Earth Observation and meteorological datasets, validated with Flow 2 Sensor data. Place and Duration of Study: Nairobi Metropolitan Area; January 2019-December 2025. **Methodology:** Pollution data were sourced from Sentinel-5P, with $PM_{10}$ derived using aerosol as a proxy. Weather variables came from MERRA-2 reanalysis. LULC was extracted from Sentinel-2 and SRTM data gave us elevation inputs. CNN extracted spatial features while LSTM captured temporal patterns. NeuralProphet handled future forecasting, and Regression Kriging improved spatial continuity. **Results:** CNN + LSTM model accurately captured spatio-temporal pollution trends between 2019 and 2024, with low validation MAE (0.0456) and RMSE (0.20), while the NeuralProphet model preserved spatial patterns and seasonal dynamics in 2025 forecasts with a validation MAE (0.16) and RMSE (0.20). However, both models underestimated ground-level $PM_{10}$ and $NO_2$ concentrations. Random Forest Regression Kriging was incorporated to refine spatial outputs, with pollutant specific outcomes. For $NO_2$, kriging marginally reduced $R^2$ from 0.998 to 0.996 but improved spatial autocorrelation (Moran's I = 0.994) and reduced RMSE to 1.302 μg/m$^3$. For $PM_{10}$, however, kriging degraded accuracy ($R^2$ declined from 0.915 to 0.782, RMSE increased to 13.805 μg/m$^3$), reflecting the

pollutant's high spatial variability and episodic nature, which limits the suitability of geostatistical interpolation for particulate matter. **Conclusion:** The study demonstrates the effectiveness of combining satellite data and deep learning in modelling and forecasting urban air pollution. CNN + LSTM and NeuralProphet provide a robust, scalable alternative to sparse ground monitoring networks, supporting urban air quality management in rapidly growing African cities. **Key findings:** 1) CNN + LSTM achieved validation MAE of 0.0456 and RMSE of 0.20; 2) NeuralProphet forecasts preserved seasonal dynamics (MAE_val = 0.16); 3) Kriging improved $NO_2$ spatial accuracy but degraded $PM_{10}$ due to episodic variability.

## Keywords

Air Pollution, Nairobi, CNN, LSTM, NeuralProphet, Sentinel-5P, Deep Learning, Forecasting

## 1. Introduction

Air pollution has remained a major global environmental health concern, contributing significantly to premature mortality and disease burden, particularly in urban areas of developing countries. In the Nairobi Metropolitan Area (NMA), which hosts over 10 million residents, air quality had continued to deteriorate due to escalating vehicle emissions, industrial discharges, and rapid land use changes (Oguge et al., 2024). Traditional ground-based air quality monitoring systems, though precise, are often sparse and costly, limiting their effectiveness in capturing the spatial complexity of urban pollution (Snyder et al., 2013).

Recent technological advances in Earth Observation (EO) and Artificial Intelligence (AI)—specifically deep learning—have provided alternative approaches to model and forecast urban air pollution. Sentinel-5P, with its TROPOspheric Monitoring Instrument (TROPOMI), has offered frequent, high-resolution satellite data on atmospheric pollutants such as $NO_2$, $SO_2$, CO, $CH_4$, and HCHO (Mamić et al., 2023). Deep learning models, especially Convolutional Neural Networks (CNNs), have shown strong capabilities in identifying spatial patterns in complex datasets, while Long Short-Term Memory (LSTM) networks are well-suited for capturing temporal dynamics in environmental time series (Tsokov et al., 2022). This study hypothesized that a hybrid CNN + LSTM model can outperform conventional EO-only models in capturing urban pollution dynamics.

This study addressed two specific objectives aimed at advancing urban air quality modelling and forecasting in the Nairobi Metropolitan Area. The first objective was to model historical concentrations of key pollutants—$PM_{10}$, CO, HCHO, $CH_4$, $NO_2$, $O_3$, and $SO_2$—from 2019 to 2024 using a hybrid deep learning framework that combined CNN and LSTM architectures. This corresponded to the research question: *How can a hybrid CNN + LSTM model accurately represent historical spatio-temporal air pollution patterns across the Nairobi Metropolitan*?

The second objective was to forecast the levels of the same pollutants for the year 2025 using Facebook's NeuralProphet—a time-series forecasting model that integrates autoregression, seasonality, and trend decomposition. This aligned with the research question: *What are the projected 2025 air pollution levels for key pollutants using NeuralProphet, and how do these forecasts preserve seasonal and spatial dynamics across the metropolitan area*?

By addressing these objectives, the study demonstrated how deep learning integrated with EO data could overcome the limitations of sparse ground monitoring networks. The modelling and forecasting outputs not only supported environmental policy and public health planning but also established a replicable framework for spatio-temporal pollution monitoring in other rapidly urbanizing regions.

## 2. Materials and Methods

### 2.1. Study Area

The Nairobi Metropolitan Area (**Figure 1**) spans approximately 32,000 km$^2$ and encompasses the counties of Nairobi, Kiambu, Machakos, Kajiado, and Murang'a. Our Area of study was bound by the following coordinates: West Longitude = 35˚57'26.8216"E; North Latitude = 0˚22'46.1686"S; East Longitude = 37˚46'18.7927"E; South Latitude = 2˚12'20.7384"S. This delineation follows the metropolitan framework established by Omwenga (2010). The region is typified by a mix of residential, commercial, industrial, and peri-urban land uses (Cyton, 2023), with diverse terrain and topographic variation. The elevation ranges and climatic diversity across the area, coupled with rapid urban expansion, make it particularly suitable for spatio-temporal modelling of air pollution. The area is home to over 10 million residents, as projected by the Kenya National Bureau of Statistics (KNBS, 2019).

### 2.2. Data Sources

The study utilized a diverse range of multi-source datasets to support the spatio-temporal modelling and forecasting of air pollution across the Nairobi Metropolitan Area (**Table 1**). Air pollution data spanning from 2019 to 2024 were derived from the Sentinel-5P satellite's TROPOspheric Monitoring Instrument (TROPOMI), which provided measurements (*Spatial Resolution of* 3.5 × 5.5 km$^2$) for key atmospheric pollutants including nitrogen dioxide ($NO_2$), sulphur dioxide ($SO_2$), carbon monoxide (CO), methane ($CH_4$), formaldehyde (HCHO), and ozone ($O_3$) (van Geffen et al., 2022). $PM_{10}$ concentrations, which are not directly observed by Sentinel-5P, were estimated using the aerosol index as a proxy (Lorente et al., 2019).

Meteorological variables such as air temperature, humidity, wind speed and direction, and surface pressure were sourced from NASA's MERRA-2 reanalysis dataset, produced by the Global Modelling and Assimilation Office (GMAO) (Gelaro et al., 2017). These weather variables are known to significantly influence the dispersion and concentration of atmospheric pollutants.

Land Use and Land Cover (LULC) information was extracted from Sentinel-2 imagery (*4 bands of 10 m and 2 bands of 20 m spatial resolutions*), from January

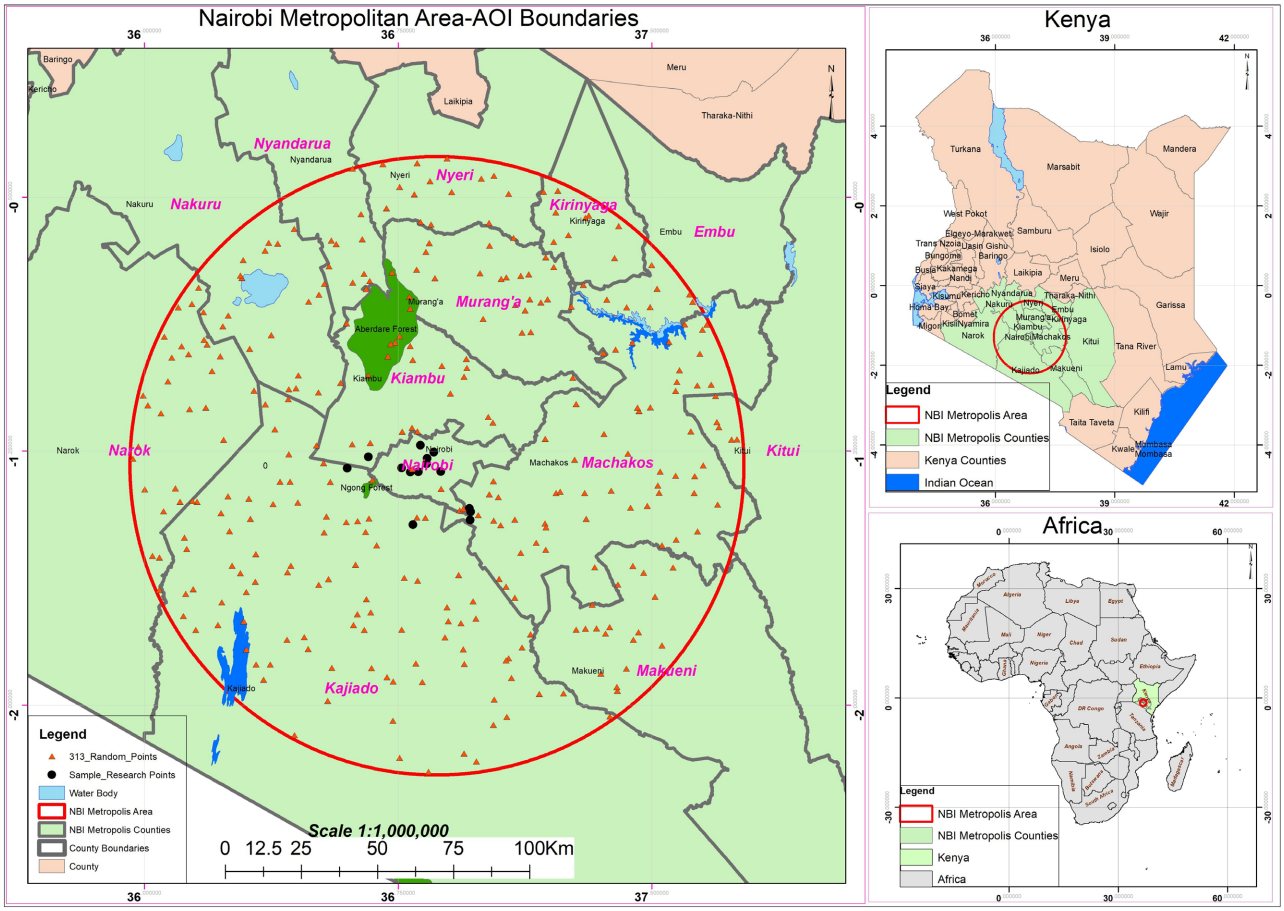



**Figure 1.** Area of study (Circle showing the 32,000 km² Nairobi Metropolitan Area, with Inset1-Kenya & Inset2-Africa. The Orange Triangles are the sampling locations).

**Table 1.** Multi-source earth observation and meteorological datasets.

| S/NO. | Data Source | Details | Units | Duration | Type |
|---|---|---|---|---|---|
| 1 | Sentinel-5P Data Sources | Level 2 Aerosol Index (AER_AI) | Index | 2019-2024 | Time Series |
| 2 | | Level 2 Carbon Monoxide (CO) | Mol/m² | | Time Series |
| 3 | | Level 2 Cloud | Radiometric Fraction | | Time Series |
| 4 | | Level 2 Formaldehyde (HCHO) | Mol/m² | | Time Series |
| 5 | | Level 2 Methane ($CH_4$) | PPB (Parts Per Billion) | | Time Series |
| 6 | | Level 2 Nitrogen Dioxide ($NO_2$) | Mol/m² | | Time Series |
| 7 | | Level 2 Ozone ($O_3$) | Mol/m² | | Time Series |
| 8 | | Level 2 Sulphur Dioxide ($SO_2$) | Mol/m² | | Time Series |
| 9 | NASA POWER MERRA-2 Data Sources | Wind Speed at 10 Meters (WS10M) | m/s | 2019-2024 | Time Series |
| 10 | | Northward Wind at 10 Meters (V10M) | m/s | | Time Series |
| 11 | | Eastward Wind at 10 Meters (U10M) | m/s | | Time Series |
| 12 | | Wind Direction at 10 Meters (WD10M) | Degrees | | Time Series |

**Continued**

| | | | | |
|---|---|---|---|---|
| 13 | | Dew/Frost Point at 2 Meters (T2MDEW) | ˚C | Time Series |
| 14 | | Wet Bulb Temperature at 2 Meters (T2MWET) | ˚C | Time Series |
| 15 | | Temperature at 2 Meters (T2M) | ˚C | Time Series |
| 16 | | Precipitation Corrected (PRECTOTCORR) | mm/day | Time Series |
| 17 | | Relative Humidity at 2 Meters (RH2M) | % | Time Series |
| 18 | | Specific Humidity at 2 Meters (QV2M) | g/kg | Time Series |
| 19 | | Surface Pressure (PS) | kPa | Time Series |
| 20 | Location Data Sources | Latitude (Lat) | Decimal Degrees | Static |
| 21 | | Longitude (Lon) | Decimal Degrees | Static |
| 22 | | Elevation (SRTM) | Meters | Static |
| 23 | Sentinel-2 Satellite Imagery Data | Land Use/Land Cover | Indices/Classes | Static |
| 24 | Flow 2 Data | $PM_{10}$ & $NO_2$ | Ground data | Validation dataset |

1, 2019 to December 31, 2023. This date range was used to create a median composite image for land cover classification, meaning the resulting LULC product represents an averaged condition across those five years (2019-2023) (Phiri et al., 2020). This spatial data was essential in evaluating the role of human settlement, vegetation, and industrial areas in modulating pollution levels. Elevation data were obtained from the Shuttle Radar Topography Mission (SRTM), enabling the incorporation of terrain variability in the modelling framework (Uuemaa et al., 2020).

For model validation, ground-level observations were collected using Flow 2 portable air quality sensors deployed at strategically selected locations across the study area. These sensors provided crucial *in situ* data for calibrating and validating the deep learning and geostatistical models (Plume Labs, 2023). A comprehensive summary of these datasets is provided in **Table 1**.

#### 2.2.1. Conversion of Satellite Column Densities to Surface-Level Concentrations

All Sentinel-5P TROPOMI Level-2 products are originally provided as total vertical column densities (mol/m$^2$). However, air quality monitoring and health impact assessments require surface-level concentrations (μg/m$^3$). Therefore, each pollutant was converted from column density to ground-level concentration using pollutant-specific physical conversion formulas incorporating local meteorological conditions (temperature, pressure) and terrain/elevation. No raw column densities (mol/m$^2$) were used as direct inputs to the CNN + LSTM model; all modelling was performed on surface-converted units (μg/m$^3$).

The general conversion workflow for each pollutant consisted of five steps:

1) Column to volume mixing ratio: Convert mol/m² to mol/m³ using an effective tropospheric height (H = 10,000 − Elevation meters), where elevation is derived from SRTM data.

2) Mass concentration: Convert mol/m³ to g/m³ using the pollutant's molecular weight.

3) Microgram scaling: Convert g/m³ to μg/m³ (multiply by 106,106).

4) Temperature and pressure adjustment: Apply the ideal gas law correction factor:

$$\frac{P_{Local}}{P_{STP}} = \frac{T_{STP}}{T_{Local} + 293.15}$$

where $P_{STP}$ = 101.325 kPa, $T_{STP}$ = 273.15 and Local P and T are from MERRA-2.

5) Vertical profile integration: Apply an exponential decay profile from the surface, where the column density is expressed as:

Column = $\int_0^H C_o \cdot e^{-z/k} dz = C_o \cdot k \cdot \left(1 - e^{-H/k}\right)$, with $C_o$ being the surface concentration, $z$ the height above ground, and $k$ the pollutant-specific e-folding height (scale height). Solving for surface concentration yields

$C_O = Column \Big/ \left[ k \cdot \left(1 - e^{-H/k}\right) \right]$.

Pollutant-specific parameters are summarized in Table 2.

Table 2. Pollutant-specific conversion parameters from sentinel-5P column density to surface concentration (μg/m³).

| Pollutant | Molecular Weight (g/mol) | E-folding height $k$ (m) | Justification for Vertical Profile |
|---|---|---|---|
| $NO_2$ | 46.0055 | 100 | Steep vertical gradient from surface combustion sources (vehicular/industrial). An additional empirical enhancement factor of 1.5 was applied based on calibration against Flow 2 ground measurements ($R^2$ = 0.998 after adjustment), consistent with urban $NO_2$ inversion studies (Lorente et al., 2019; van Geffen et al., 2022). |
| $SO_2$ | 64.066 | 10 | Very shallow boundary layer concentration from industrial and domestic fuel combustion sources (Lin et al., 2019; Zhang et al., 2024). |
| CO | 28.01 | 100 | Urban pollution tracer with moderate vertical gradient from incomplete combustion (Seinfeld & Pandis, 2016). |
| HCHO | 30.03 | 100 | Photochemical oxidation product with near-surface maximum in urban |
| $CH_4$ | 16.04 | 0.4 | Well-mixed long-lived greenhouse gas with negligible vertical gradient in the troposphere (Seinfeld & Pandis, 2016). The small e-folding height effectively produces a uniform vertical profile. |
| $O_3$ | 48 | N/A (surface fraction = 0.2) | Unlike primary pollutants, $O_3$ is a secondary pollutant with maximum concentrations in the upper troposphere/lower stratosphere. Based on ozonesonde profiles over East Africa (Thompson et al., 2021), surface $O_3$ typically represents 10% - 25% of the total column. A fixed surface fraction of 0.2 (20%) was applied, empirically calibrated against ground measurements. |

### 2.2.2. $PM_{10}$ Derivation Workflow

Since Sentinel-5P does not directly measure particulate matter, $PM_{10}$ surface concentrations were estimated from the Aerosol Index (AI) using a linear scaling ap-

proach based on the WHO 24-hour air quality guideline limit of 50 μg/m$^3$:

1) The minimum and maximum AI values across the entire 2019-2024 time series were extracted for the study area.

2) The AI was normalized to the [0, 1] range using min-max scaling.

3) Normalized values were multiplied by 50 μg/m$^3$ to obtain surface $PM_{10}$ estimates.

This method assumes a linear relationship between the Aerosol Index and surface $PM_{10}$, and that the maximum observed AI in the study period corresponds to the WHO guideline limit. These assumptions are acknowledged as limitations in Section 5.

#### 2.2.3. Final Input Units to All Models

All models (CNN + LSTM for historical modelling and NeuralProphet for forecasting) were trained and evaluated using the surface-concentrations in μg/m$^3$ for all pollutants. Therefore, all reported MAE and RMSE values in Sections 3 and 4 are expressed in original physical units (μg/m$^3$) and are directly interpretable against regulatory thresholds. No additional normalization was applied to the target variables before model training.

### 2.3. Modelling (CNN + LSTM Hybrid Framework)

The modelling architecture (**Figure 2**) was designed to capture both spatial and temporal patterns in air pollution data from 2019 to 2024. A hybrid deep learning framework combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks was implemented.

CNN layers were used to extract spatial features (Du et al., 2018) from gridded Earth Observation (EO) data, including pollutant concentrations, terrain elevation, and land use/land cover (LULC). Input data were spatially aligned and resampled to a uniform grid resolution (1 km$^2$). Spatial inputs included**:** Sentinel-5P ($NO_2$, CO, $SO_2$, $CH_4$, HCHO, $O_3$, $PM_{10}$ via aerosol index), Sentinel-2 LULC maps (2020 baseline), SRTM elevation rasters, Geographic coordinates (Latitude, Longitude).

LSTM layers captured temporal dependencies across multiple meteorological and atmospheric time series (O'Donncha et al., 2021). Temporal features were structured into daily sequential data windows, with input variables including: Temperature (T2M), Humidity (RH2M), Wind Speed/Direction (WS10M, WD10M, V10M, U10M), Precipitation (PRECTOTCORR), Pressure (PS), and time indicators being (Year, Month, Day).

The model was trained per pollutant using an 80:20 training-validation split (Wang et al., 2025). The LSTM component processed sequential meteorological dynamics, while CNN processed spatial frames. Feature fusion occurred before the final dense layers. The Adam optimizer minimized the Mean Absolute Error (MAE), while model performance was evaluated using MAE, RMSE, $R^2$, and correlation (Du et al., 2018).

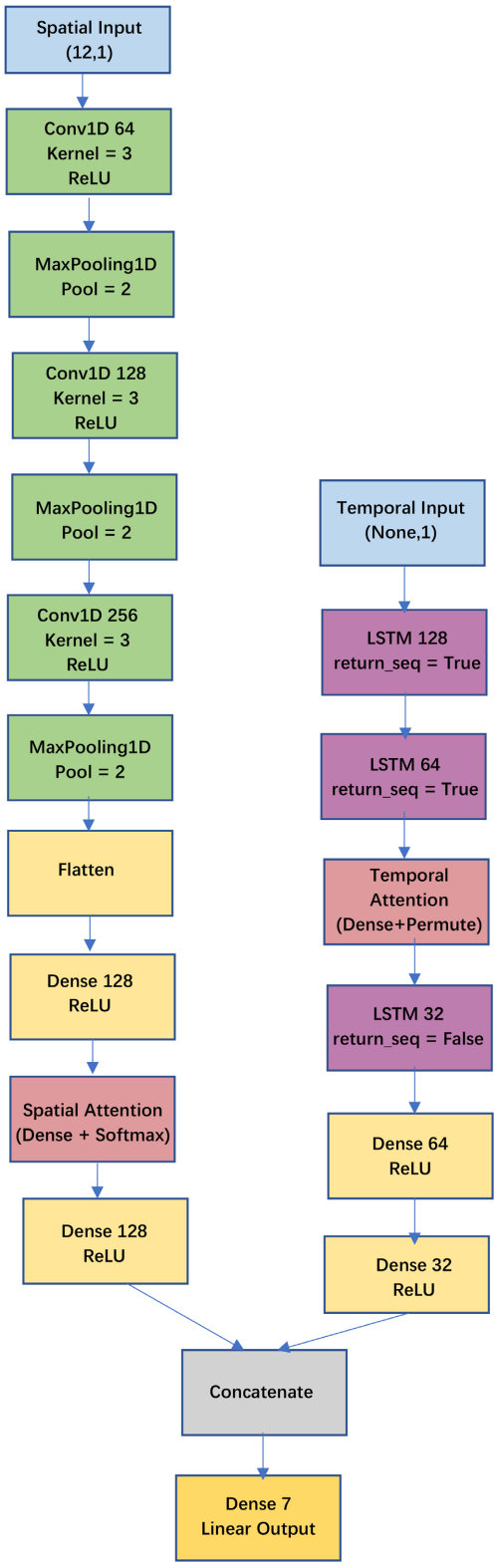



**Figure 2.** CNN + LSTM modelling architecture.

The spatial stream of the model utilized a CNN architecture comprising three 1D convolutional layers with 64, 128, and 256 filters respectively (*each with a kernel size*

*of 3 and ReLU activations*), interspersed with max pooling. A spatial attention mechanism was applied to emphasize important spatial features before merging.

The temporal stream used a stacked LSTM configuration with 128 and 64 units, followed by a temporal attention layer and a final LSTM with 32 units to capture time-dependent patterns. Outputs from both branches were concatenated and passed through fully connected layers to predict seven target pollutants.

Terrain features (elevation) from SRTM were used to detect topographic influences (Wen et al., 2022). A total of 327 georeferenced sampling points across the Nairobi Metropolitan Area were used, and spatial joins were conducted to integrate EO data and ground validation. Ground points were used both in the training of CNN spatial layers and in regression kriging validation of prediction surfaces.

#### 2.3.1. CNN + LSTM Architecture: Spatial and Temporal Branches

The hybrid deep learning framework combines a Convolutional Neural Network (CNN) for spatial feature extraction and a Long Short-Term Memory (LSTM) network for temporal dependency modelling. The model was implemented as a single multi-output architecture predicting all seven pollutants simultaneously ($PM_{10}$, $NO_2$, $SO_2$, $O_3$, CO, $CH_4$, HCHO), rather than training separate models per pollutant. This approach allows the model to learn inter-pollutant correlations and improves computational efficiency.

**Spatial Branch (CNN with 1D Convolutions):**

The spatial input consists of a 1D feature vector of 12 spatial variables: longitude, latitude, elevation, and 9 one-hot encoded Land Use/Land Cover (LULC) classes derived from Sentinel-2. These features are reshaped to (samples, 12, 1) to serve as input to a 1D convolutional network. The use of 1D convolutions (rather than 2D) is appropriate because the spatial data are represented as a vector of co-located attributes at each sampling point, not as an image grid requiring 2D kernels.

The spatial branch architecture comprises three Conv1D layers with 64, 128, and 256 filters respectively, each with a kernel size of 3 and ReLU activation, followed by MaxPooling1D layers (pool size = 2). The output is flattened and passed through two dense layers (128 and 64 units) with ReLU activation. A spatial attention mechanism is applied after the first dense layer, computing *softmax* weights across spatial features to emphasize the most relevant variables for prediction.

**Temporal Branch (LSTM with Attention):**

The temporal input consists of 30 meteorological and time-based features (including temperature, humidity, wind speed/direction, pressure, precipitation, and cyclical encodings of month, day, and holiday). The temporal branch uses a stacked LSTM architecture with three layers (128, 64, and 32 units), with return_sequences = True for the first two layers to maintain the temporal dimension for attention. A temporal attention layer (custom TemporalAttention class) computes softmax weights across time steps, allowing the model to focus on the most relevant historical periods for prediction. The final LSTM layer (return_sequences = False) compresses the temporal information into a fixed-length vector, which is passed through two dense layers (64 and 32 units).

**Feature Fusion and Output:**

The outputs of the spatial and temporal branches are concatenated and passed to a final dense layer with **7** neurons and linear activation, producing simultaneous predictions for all seven target pollutants. The model was compiled using the Adam optimizer with a Huber loss function (robust to outliers) and tracked Mean Absolute Error (MAE) as a primary metric.

This study uses 1D convolutions because the spatial data at each sampling point are organized as a 1D feature vector (12 attributes × 1 channel). If the data were organized as a 2D raster grid (e.g., pixels with spatial neighbourhood structure), 2D convolutions would be appropriate. For point-based spatial features with no inherent neighbourhood topology beyond the feature vector itself, 1D convolutions are the correct architectural choice (Du et al., 2018; Du et al., 2018; Tsokov et al., 2022).

### 2.3.2. Training-Validation Split and Temporal Integrity

To ensure rigorous evaluation of the CNN + LSTM model's ability to generalize to unseen time periods, a chronological split was applied to the time series data spanning 2019-2024. The full dataset was first sorted by date using *df_encoded.sort_values("Date").reset_index(drop = True).* The first 80% of the temporal sequence (January 2019 to December 2023) was allocated to training, and the remaining 20% (January 2024 to December 2024) was reserved for validation. This was implemented using *train_test_split* with *shuffle = False*, which preserves temporal order and prevents random shuffling. Consequently, no future data were used to predict past observations—a critical requirement for time series validation (Bergmeir & Benítez, 2012).

NeuralProphet chronological split: For the NeuralProphet forecasting model (Section 2.4), the built-in *split_df* method was used with *valid_p = 0.2*. This method automatically preserves temporal order, taking the last 20% of dates for validation and the first 80% for training. Thus, both models employed temporally consistent validation schemes.

Inverse scaling of metrics: Prior to training, all target variables (seven pollutant concentrations) were normalized to the range [0, 1] using *MinMaxScaler* to stabilize gradient descent and improve convergence. After model prediction, the scaled outputs were inverse-transformed back to original physical units (μg/m$^3$ for all pollutants) using the *inverse_transform* method of the fitted scaler. Therefore, all reported MAE, RMSE, $R^2$, and bias values in Section 3 are expressed in original units (μg/m$^3$) and are directly interpretable against air quality guidelines and regulatory thresholds.

Spatial considerations: All 327 georeferenced sampling locations were present in both the training and validation sets. While this allows the model to learn location-specific patterns, it does not test spatial generalizability to entirely unmonitored locations. Future work should consider spatial holdout validation (e.g., reserving specific geographic points for testing) to assess model transferability across space (Meyer et al., 2019). The spatial interpolation and kriging steps (Section 2.6) par-

tially address this limitation by improving spatial continuity of predictions.

Validation data usage: The validation set (last 20% of dates) was used exclusively for final model evaluation. No validation data were used for early stopping or hyperparameter tuning. The model architecture and hyperparameters were fixed before evaluating on the validation set, ensuring an unbiased estimate of generalization error.

## 2.4. Forecasting (NeuralProphet)

NeuralProphet was employed to forecast daily pollution levels for the period October 2024 to December 2025 (**Figure 3**); (note that this could have been made even for the year 2035); across all 327 georeferenced sampling locations in the AOI. Each forecast was generated per pollutant and per spatial point, preserving both geographic coordinates and land cover context.

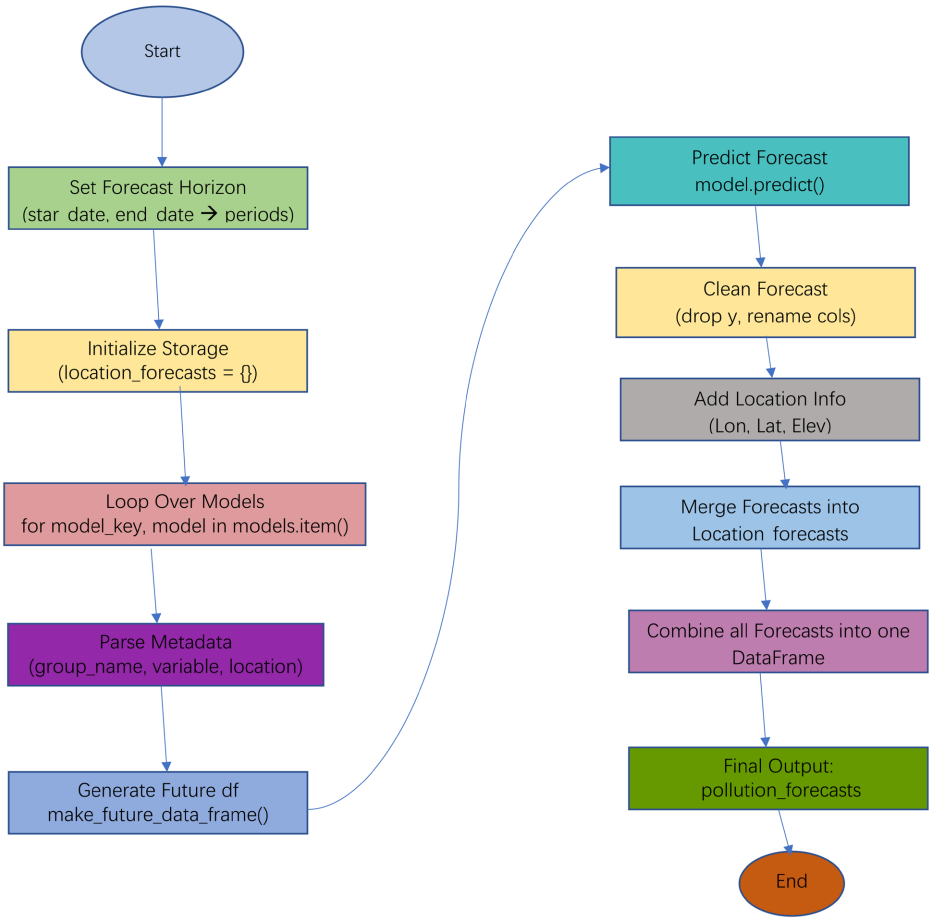



**Figure 3.** NeuralProphet modelling workflow.

The model incorporated trend, seasonality, autoregression, and lagged regressors (NeuralProphet Documentation [n.d.], 2024). For spatial consistency, the historical CNN + LSTM outputs were first aggregated at the grid cell level and tagged with location-specific variables that are Latitude & Longitude, Elevation and LULC

Classes.

Forecasting was conducted per point, resulting in spatially disaggregated forecasts for each pollutant at each location. These were then visualized as spatially explicit heatmaps, allowing observation of future high-risk zones. Furthermore, forecast outputs were used as inputs for spatial autocorrelation analysis (Moran's I and Geary's C) and Regression Kriging, to assess and improve the spatial continuity of the predictions. This helped detect patterns of forecast over-/under-prediction and ensured that forecast values aligned with regional pollutant dispersion dynamics.

Spatial accuracy of forecasts was evaluated by comparing forecasted $NO_2$ and $PM_{10}$ values with ground-level Flow 2 observations at the same location points (coordinates), using $R^2$, MAE, RMSE, and spatial correlation indices.

For each of the 327 locations, a separate NeuralProphet model was trained using only the historical pollution time series at that location. No spatial regressors (latitude, longitude, elevation, LULC) were included in the forecasting model, ensuring spatial independence between locations. The data for each location were explicitly sorted by date before any split. The split_df method was used with valid_p = 0.2, which chronologically reserves the last 20% of dates for validation and the first 80% for training. This prevents temporal leakage, as the model never sees future data during training.

The NeuralProphet models were configured with yearly, weekly, and daily seasonality enabled. No autoregressive terms were used (n_lags = 0, the default), meaning predictions were based solely on trend and seasonal decomposition rather than short-term lagged dependencies. Models were trained for 50 epochs with a batch size of 10 and a learning rate of 0.1, using the default Huber loss function. The validation metrics (MAE_val = 0.16, RMSE_val = 0.20) reported in Section 3.2 represent the average performance across all locations and pollutants on the withheld chronological validation period (January 2024 to December 2024) (**Figure 20**).

## 2.5. Ground-Based Validation Data: Flow 2 Sensor Deployment and Protocol

To validate the satellite-derived and deep learning modelled outputs, ground-level measurements of $PM_{10}$ and $NO_2$ were collected using a single Plume Labs Flow 2 portable air quality sensor. This section describes the deployment protocol, data processing, and validation framework.

### 2.5.1. Sensor Specifications and Deployment Strategy

The Flow 2 sensor is a handheld, battery-operated device capable of measuring $PM_{10}$ (μg/m$^3$), $PM_{2.5}$ (μg/m$^3$), $NO_2$ (ppb and μg/m$^3$), and volatile organic compounds (VOC). The sensor was operated following Plume Labs guidelines for portable deployment (**Table 3**). No pre-deployment calibration was performed due to the absence of regulatory reference stations in Kenya—a limitation common to low-cost sensor studies in data-sparse regions (Abera et al., 2021).

**Table 3.** Flow-2 sensor model & deployment.

| Parameter | Details |
|---|---|
| Sensor model | Plume Labs Flow 2 (portable handheld) |
| Number of sensors | 1 |
| Total unique sampling locations | 14,229 |
| Harmonized validation points | 327 (spatially interpolated from 14,229) |
| Sampling mode | Opportunistic mobile (driving transects) |
| Recording interval | 1 second |
| Temporal aggregation | Daily averages |
| Calibration | None (no reference stations in Kenya) |
| Quality control | Outlier removal based on literature ranges |
| Pollutants validated | $PM_{10}$ and $NO_2$ |
| Validation use | Independent (not used in model training except $NO_2$ 1.5 factor) |

Data collection occurred across the Nairobi Metropolitan Area using opportunistic mobile sampling while driving along major roads and traversing residential, commercial, and peri-urban areas. The sensor recorded measurements at 1-second intervals, generating a dense set of georeferenced points. A total of 14,229 unique locations were sampled intermittently, with most sites sampled for approximately one minute per visit. The full dataset covers multiple days across the study period, though the exact temporal coverage per location was uneven due to the mobile sampling design.

#### 2.5.2. Spatial Interpolation to Common Grid Points

The raw 14,229 Flow 2 points did not spatially coincide with the 327 fixed sampling locations used for Sentinel-5P and MERRA-2 extraction. To enable pointwise comparison, the Flow 2 measurements were first spatially interpolated across the entire area of interest (circular domain of approximately 101 km radius) using Inverse Distance Weighting (IDW) combined with a k-Nearest Neighbors (k-NN) strategy. This approach transformed the heterogeneous mobile measurements into a harmonized gridded surface. Values were then extracted at the 327 predefined grid points that matched the satellite and meteorological sampling locations enabling direct point pairwise comparison.

#### 2.5.3. Quality Control and Temporal Aggregation

Quality control was performed through outlier removal based on physically plausible ranges informed by previously published air quality studies in East African urban environments (Oguge et al., 2024). No co-location with regulatory monitors was possible due to the absence of such infrastructure in Kenya. The 1-second instantaneous measurements were aggregated to daily averages at each of the 327 locations to match the temporal resolution of the satellite-derived and modelled outputs.

#### 2.5.4. Validation Framework: Independent Use Only

Crucially, the Flow 2 ground observations were used exclusively for independent

validation of the CNN + LSTM modelled outputs and NeuralProphet forecasts. No Flow 2 data were used in model training, calibration of the CNN + LSTM architecture, or fitting of the NeuralProphet parameters. The sole exception was the empirical 1.5 enhancement factor for $NO_2$ (Section 2.2.1), which was calibrated against the Flow 2 dataset. All other model parameters were derived solely from satellite and meteorological data.

The validation framework employed three comparative pairwise analyses at the 327 harmonized points:

1) Flow 2 vs. Sentinel-5P (raw satellite)

2) Flow 2 vs. CNN + LSTM (modelled historical)

3) Flow 2 vs. NeuralProphet (forecasted)

Quantitative metrics included Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), bias, Pearson correlation coefficient, and $R^2$. Additionally, spatial autocorrelation (Moran's I and Geary's C) was computed for the Flow 2 data to characterize the spatial structure of ground-truth pollution patterns ($PM_{10}$ Moran's I = 0.3721, $p < 0.0001$; $NO_2$ Moran's I = 0.3192, $p < 0.0001$), confirming moderate spatial clustering consistent with urban pollution gradients (Table 4 **(a)-(b)**).

Table 4. (a). Three comparatives pairwise analyses at the 327 harmonized points for $PM_{10}$, (b). Three comparatives pairwise analyses at the 327 harmonized points for $NO_2$.

(a)

| Dataset | Moran's I | p-value (I) | Geary's C | p-value (C) |
|---|---|---|---|---|
| Plume Lab's Flow 2 | 0.3721 | 0.0001 | 0.6482 | 0.0001 |
| Sentinel-5P | 0.0217 | 0.0001 | 0.9428 | 0.0004 |
| CNN + LSTM Deep Learning | 0.0346 | 0.0001 | 0.9487 | 0.0004 |
| Facebook NeuralProphet | 0.6495 | 0.0001 | 0.2815 | 0.0001 |

(b)

| Dataset | Moran's I | p-value (I) | Geary's C | p-value (C) |
|---|---|---|---|---|
| Plume Lab's Flow 2 | 0.3192 | 0.0001 | 0.6662 | 0.0001 |
| Sentinel-5P | 0.1517 | 0.0001 | 0.8428 | 0.0004 |
| CNN + LSTM Deep Learning | 0.9937 | 0.0001 | 0.0163 | 0.0001 |
| Facebook NeuralProphet | 0.9969 | 0.0001 | 0.0032 | 0.0001 |

## 3. Results

### 3.1. CNN Performance (2019-2024)

#### 3.1.1. Spatial Patterns and Performance (CNN Component)

The CNN + LSTM model effectively captured spatio-temporal air pollution dynamics across the Nairobi Metropolitan Area from 2019 to 2024, achieving low validation errors (Validation MAE = 0.046, Validation Loss = 0.0013) (Figure 4). The CNN layers extracted spatial features from EO data (Sentinel-5P, SRTM, LULC), while the LSTM units modeled temporal dependencies across time-period in the study.

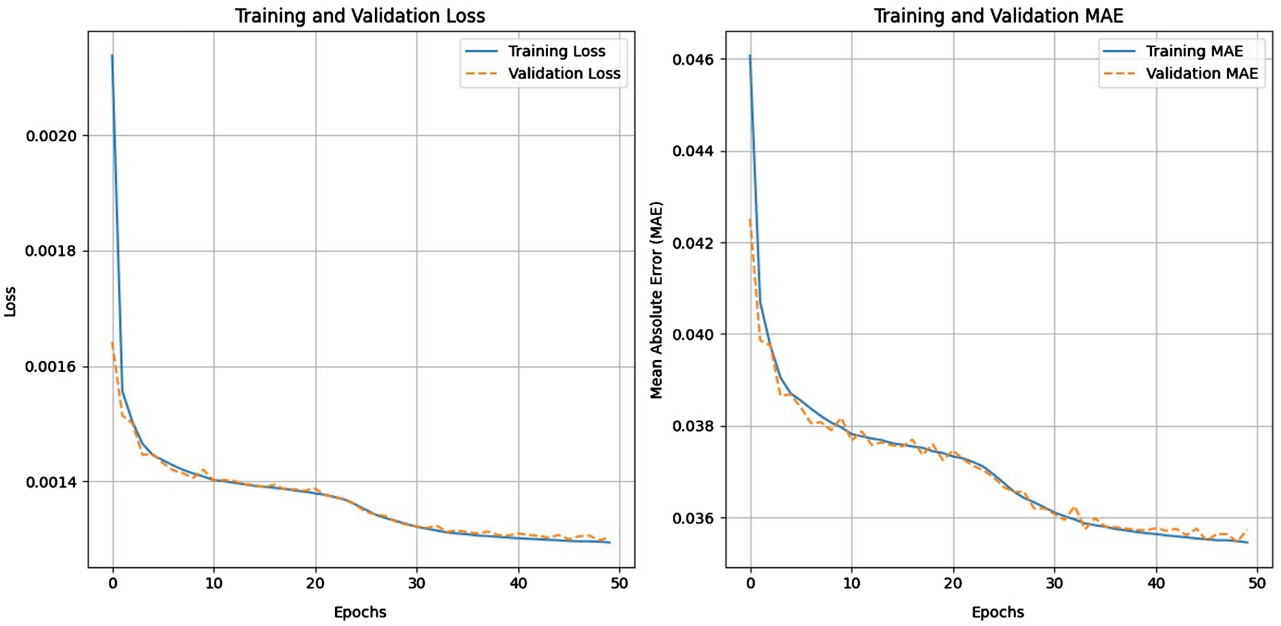



**Figure 4.** Plots of CNN + LSTM prediction modelling metrics (Plots showing Loss and MAE for training & validation datasets).

Despite strong general performance, surface-level $PM_{10}$ and $NO_2$ concentrations were systematically overestimated or underestimated in peri-urban zones and areas with limited sensor coverage. These residual errors were geographically patterned, suggesting influence from LULC and terrain mismatches. Figure 5 and Figure 6 show a comparison of the spatial distribution between Sentinel-5P & Modelled data for $PM_{10}$ and $NO_2$ respectively.

Moran's I and Geary's C confirmed strong spatial clustering in CNN + LSTM-predicted $NO_2$ values (*Moran's I* = 0.994, *Geary's* C = 0.016) and very weak spatial structure in $PM_{10}$ (*Moran's I* = 0.035, *Geary's C* = 0.949). This difference reflects the diffuse emission sources of $NO_2$ (urban and industrial zones) vs. the localized, episodic nature of $PM_{10}$ (e.g., road dust, construction) (Figures 7-10).

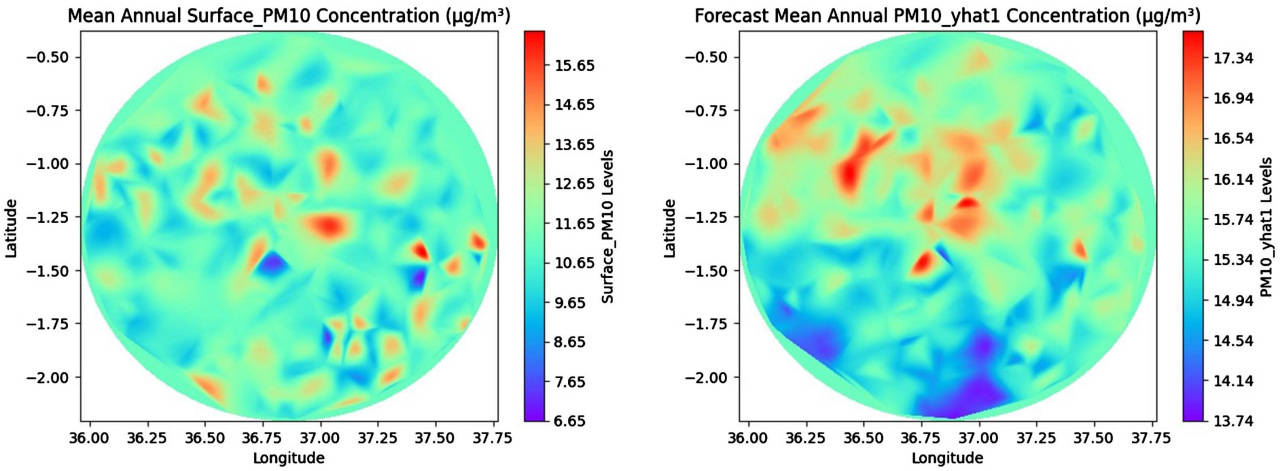



**Figure 5.** Spatial maps showing visual comparison of observed and forecast $PM_{10}$ concentrations levels (The Forecasts exhibited higher concentrations especially on the Northen part of the AOI and lower concentrations in the southern part of the AOI).

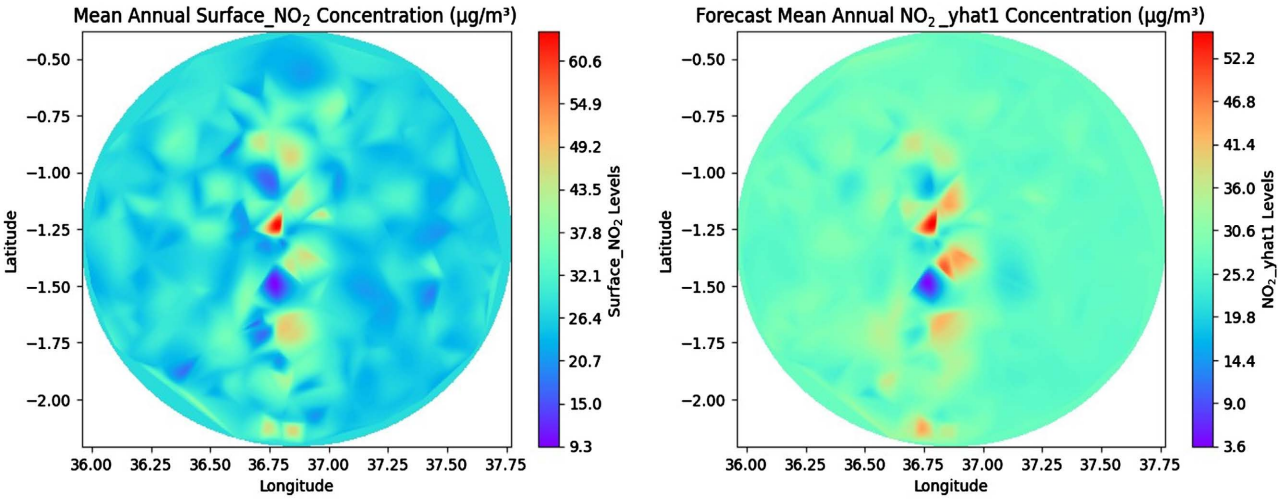



**Figure 6.** Spatial maps showing visual comparison of observed and forecast $NO_2$ concentrations levels. The spatial distribution pattern of the $NO_2$ forecast was similar to the observed with forecast values generally exhibiting higher concentrations.

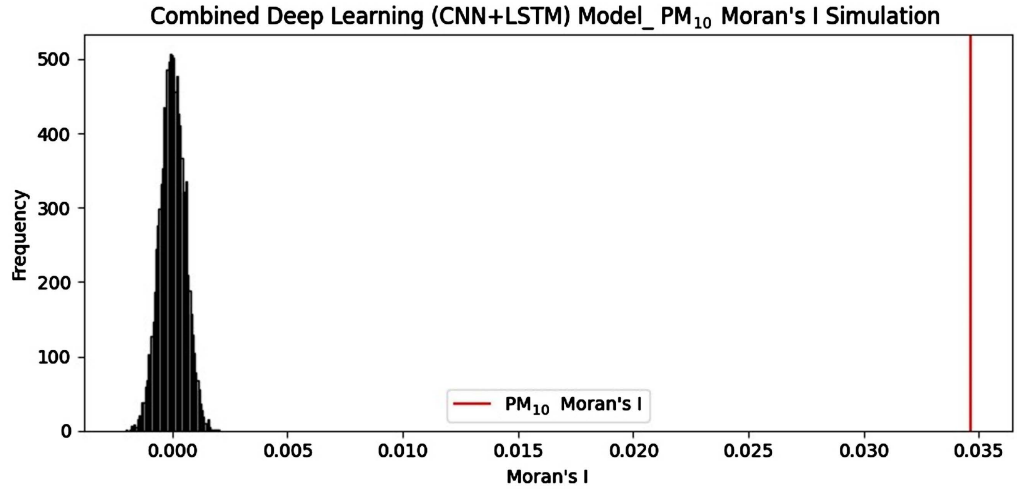



**Figure 7.** Combined (CNN + LSTM) deep learning ($PM_{10}$ Moran's I & P-Values).

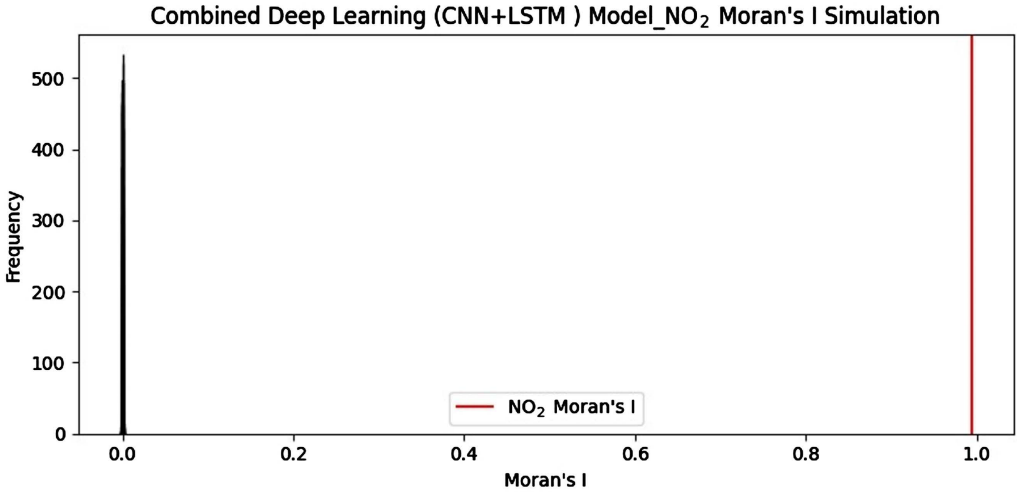



**Figure 8.** Combined (CNN + LSTM) deep learning ($NO_2$ Moran's I & P-Values).

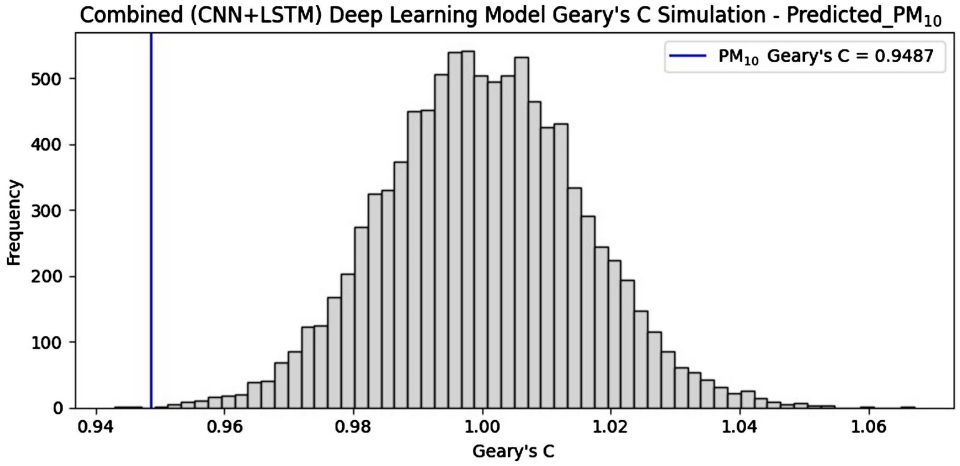



**Figure 9.** Combined (CNN + LSTM) deep learning model ($PM_{10}$ Geary's C & P-Values).

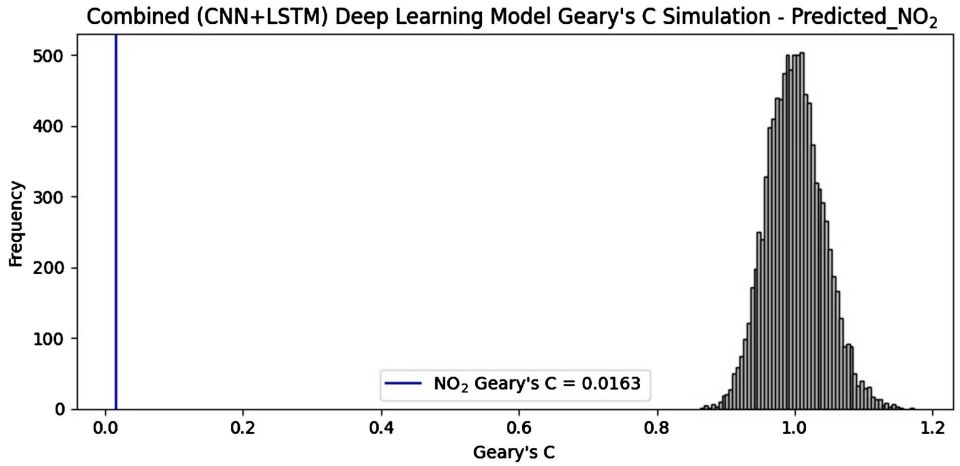
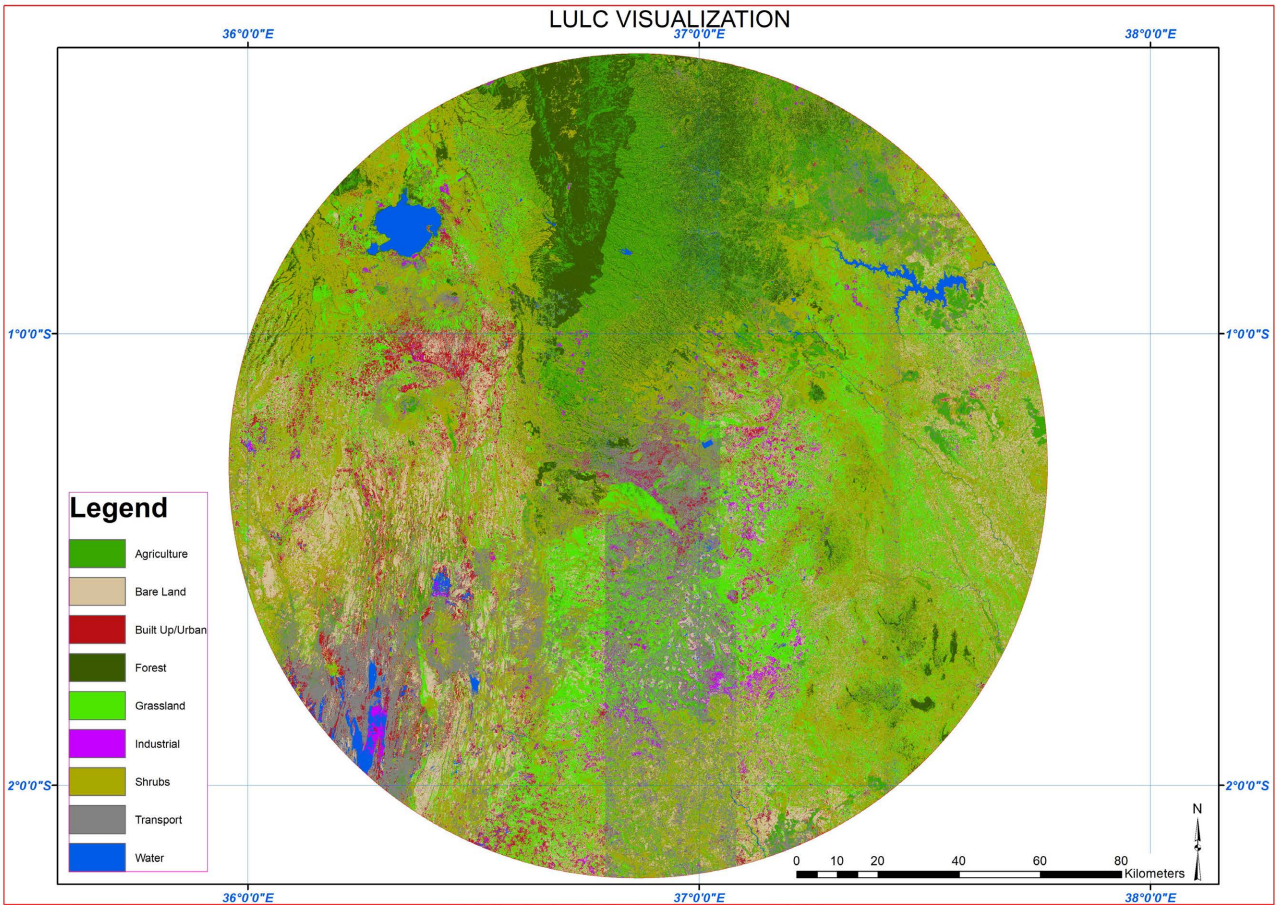



**Figure 10.** Combined (CNN + LSTM) deep learning model ($NO_2$ Geary's C & P-Values).

Visual correlation maps (Figure 11 and Figure 12) showed that $NO_2$ and $SO_2$ concentrations were the highest in built-up and transport corridors, while $CH_4$ and $O_3$ were dominant in agricultural/vegetated areas.

To improve spatial continuity, residuals from a Random Forest baseline were kriged using a variogram-based geostatistical model. The results revealed a pollutant-specific trade-off (Table 5). For $NO_2$, kriging marginally reduced $R^2$ (from 0.998 to 0.996) but improved spatial autocorrelation (Moran's I = 0.994, up from 0.319 for raw Flow 2 data) and reduced RMSE from 1.695 to 1.302 μg/m$^3$. For $PM_{10}$, however, kriging degraded accuracy: $R^2$ dropped from 0.915 to 0.782, RMSE increased from 8.600 to 13.805 μg/m$^3$, and bias increased from −0.004 to −4.136. This decline reflects $PM_{10}$'s high spatial variability and episodic nature (e.g., localized dust events, construction plumes, road resuspension), which violate the spatial stationarity assumption underlying kriging interpolation (Addiena A Rahim et al., 2023). Consequently, while kriging is beneficial for smoothly varying pollutants like $NO_2$, it is not recommended for $PM_{10}$ in this context without additional ground truthing or alternative spatial smoothing methods.



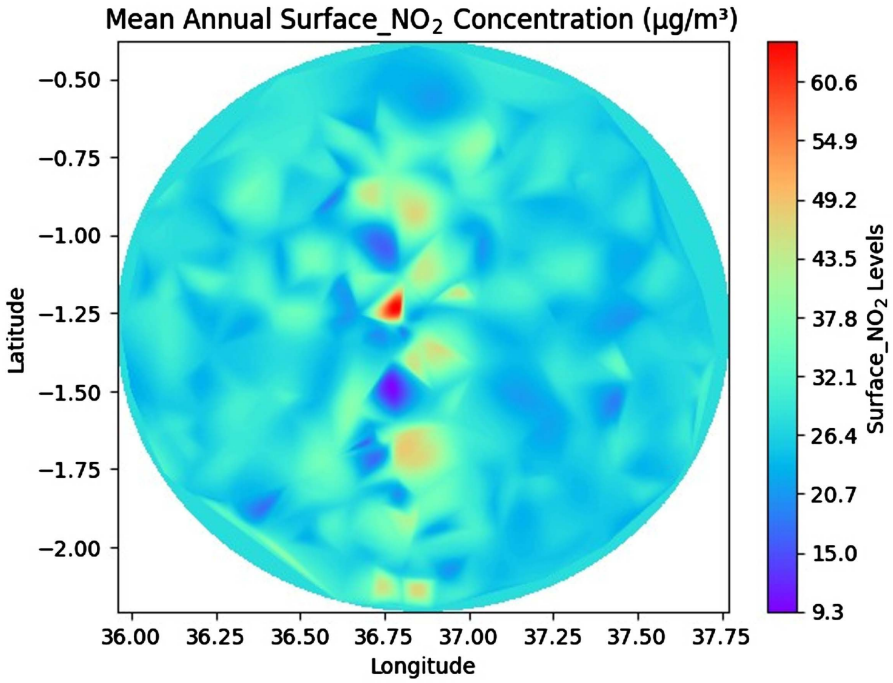
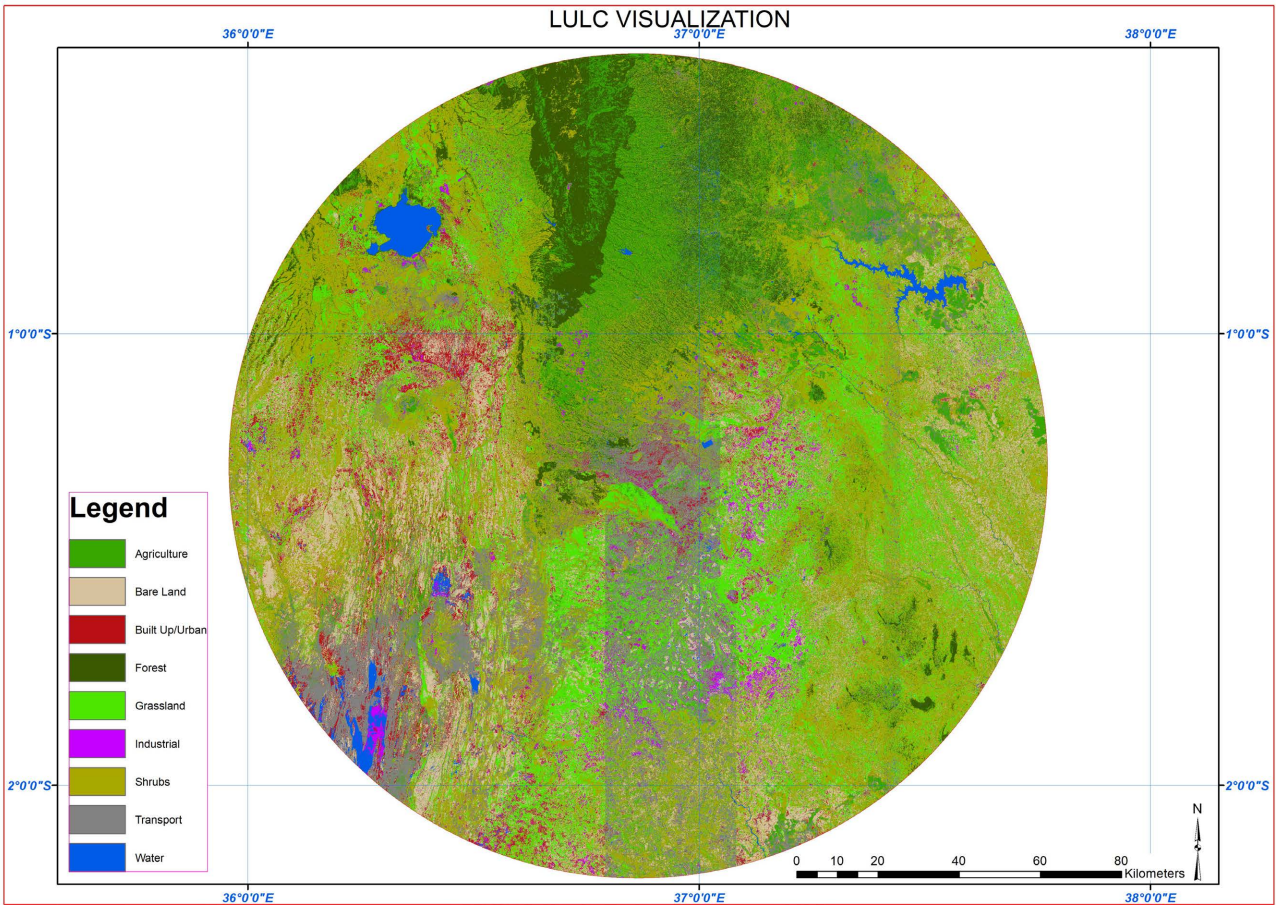




**Figure 11.** Spatial maps showing the relationship between LULC & $NO_2$ (Land use/land cover and mean annual surface $NO_2$ showing higher concentration aligning with the Northern Corridor Transport route. Thus, $NO_2$ seems to be correlated to the transport Land Use/Land Cover & Urban Areas).



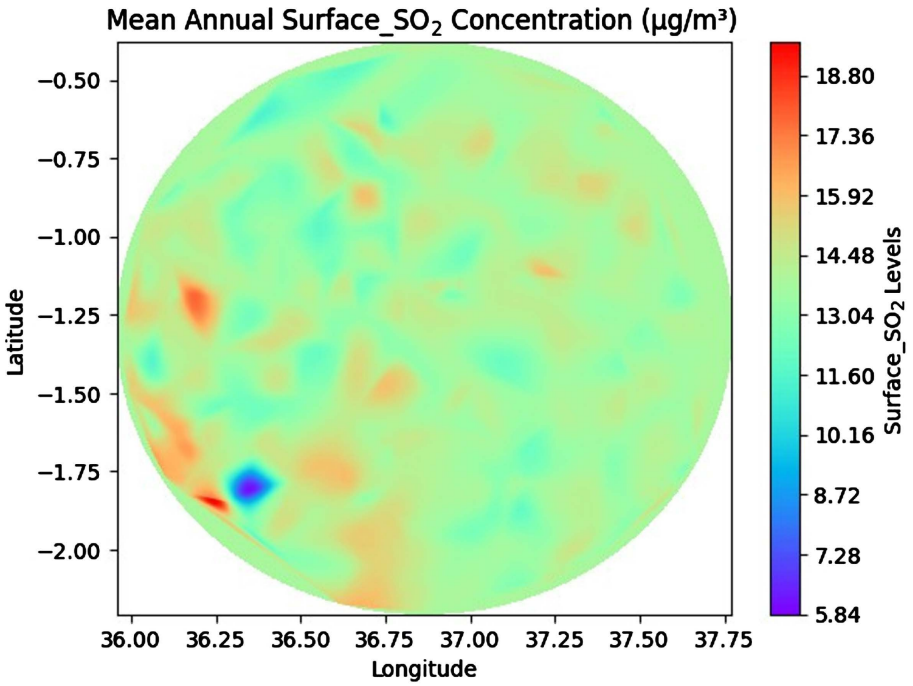




**Figure 12.** Spatial maps showing the relationship between LULC & $SO_2$ (Land use/land cover and mean annual surface $SO_2$ showing higher concentration distribution on the lower left side of the study area which is associated with the shrubs land cover with a distinct lowest concentration over Lake Magadi).

Table 5. CNN + LSTM model accuracies (Historical 2019-2024).

| Model Type | Pollutant | MSE | $R^2$ | RMSE | MAE | Bias | Corr. |
|---|---|---|---|---|---|---|---|
| Random Forest Only | $PM_{10}$ | 73.953 | 0.915 | 8.600 | 3.510 | −0.004 | 0.957 |
| RF + Kriging | $PM_{10}$ | 190.572 | 0.782 | 13.805 | 4.768 | −4.136 | 0.922 |
| Random Forest Only | $NO_2$ | 0.693 | 0.998 | 0.832 | 0.519 | −0.000 | 0.999 |
| RF + Kriging | $NO_2$ | 1.695 | 0.996 | 1.302 | 0.580 | 0.455 | 0.998 |

Harmonized Flow 2-CNN + LSTM comparisons at spatial points showed spatially patterned underpredictions in the CBD and overpredictions in high-altitude, vegetated areas.

#### 3.1.2. Temporal Accuracy and Trends (LSTM Component)

The LSTM layers were instrumental in capturing temporal dependencies and seasonal cycles from meteorological variables (such as temperature, humidity, wind speed, pressure) and pollutant time series. The model learned lagged relationships between weather changes and pollutant concentration surges.

Time series plots (Figures 13-19) show strong agreement between LSTM predictions and observed satellite values. The $O_3$, $CH_4$, and HCHO, especially during seasonal transitions (March-May, October-December).

$NO_2$ and $PM_{10}$ exhibited high temporal variance, and while trends were captured, short-term spikes (e.g., dust events, traffic peaks) led to occasional prediction errors. The LSTM's learning was aided by sequential input structuring helping it identify multi-day pollutant buildup patterns during dry periods.

The plots of CO show that CO was overestimated, while $O_3$ and $CH_4$ predictions agreed perfectly with the original Sentinel-5P satellite data.

### 3.2. Forecasts for 2025 (NeuralProphet)

NeuralProphet was employed to forecast air pollution levels for the year 2025 and successfully achieved temporal trends and seasonal dynamics (Figure 20) with (validation MAE = 0.160, RMSE = 0.200) for 50 epochs (Table 6). However, as with the historical models, NeuralProphet also underestimated $PM_{10}$ and $NO_2$ concentrations compared with Ground Based observations, particularly in under-monitored zones. Forecast validation using Random Forest showed moderate performance: for $NO_2$, $R^2 = 0.811$ and RMSE = 9.028 μg/m$^3$; and for $PM_{10}$, $R^2 = 0.786$ and RMSE = 13.655 μg/m$^3$. Post-processing with Regression Kriging slightly reduced $R^2$ for both pollutants but improved correlation for $NO_2$ (to 0.939), indicating better spatial structural alignment (Table 7).

Figures 21-27 show the Time Series NeuralProphet Forecast Plots from 8$^{th}$ October 2024 to 31$^{st}$ December 2025 for the 7 Pollutants.

The results of NeuralProphet forecast model were run through Random Forest and Kriging Regression modelling to improve model accuracy and the resultant metrics are as shown in Table 7:

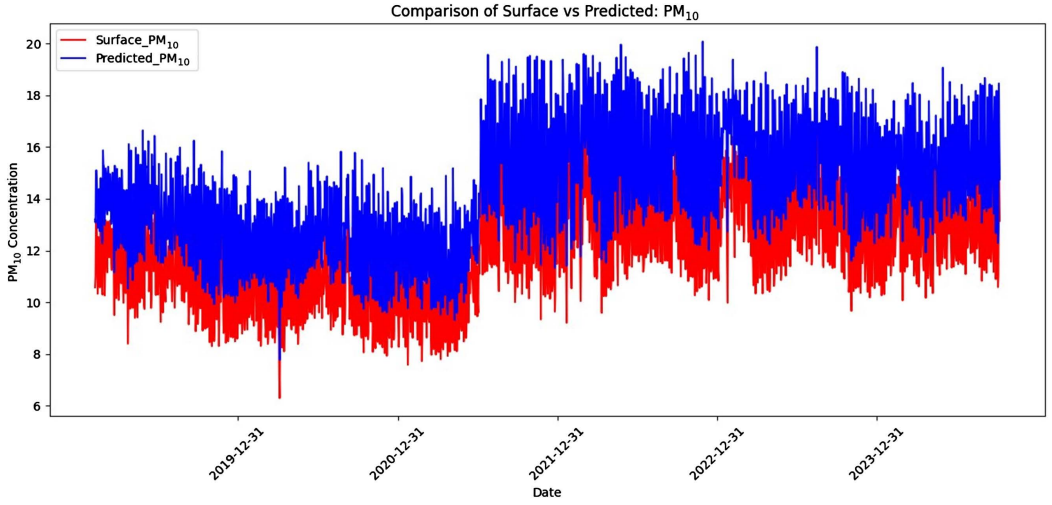



**Figure 13.** $PM_{10}$ Time series plots {CNN + LSTM modelling output (Blue) and Sentinel-5P Datasets (Red)}.

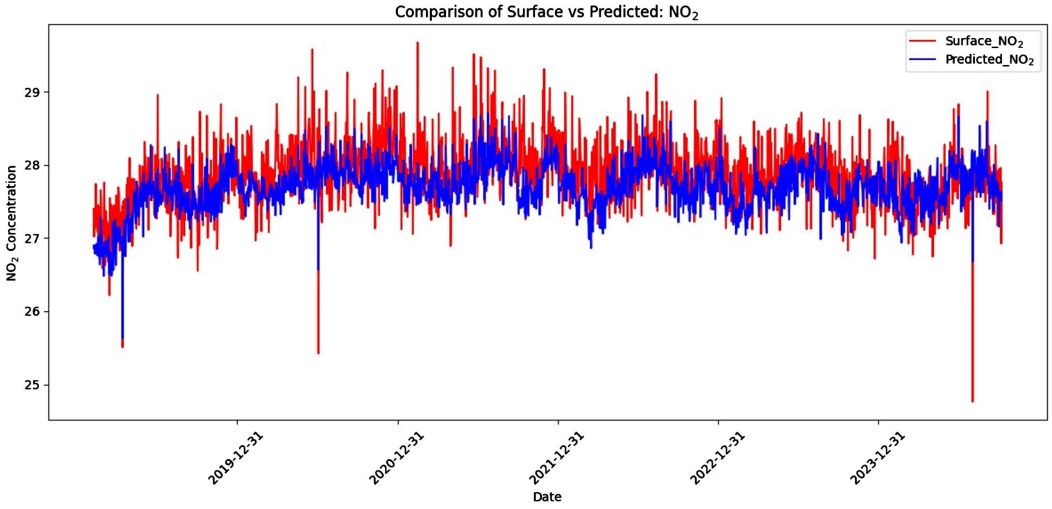



**Figure 14.** $NO_2$ Time series plots {CNN + LSTM modelling output (Blue) and Sentinel-5P Datasets (Red)}.

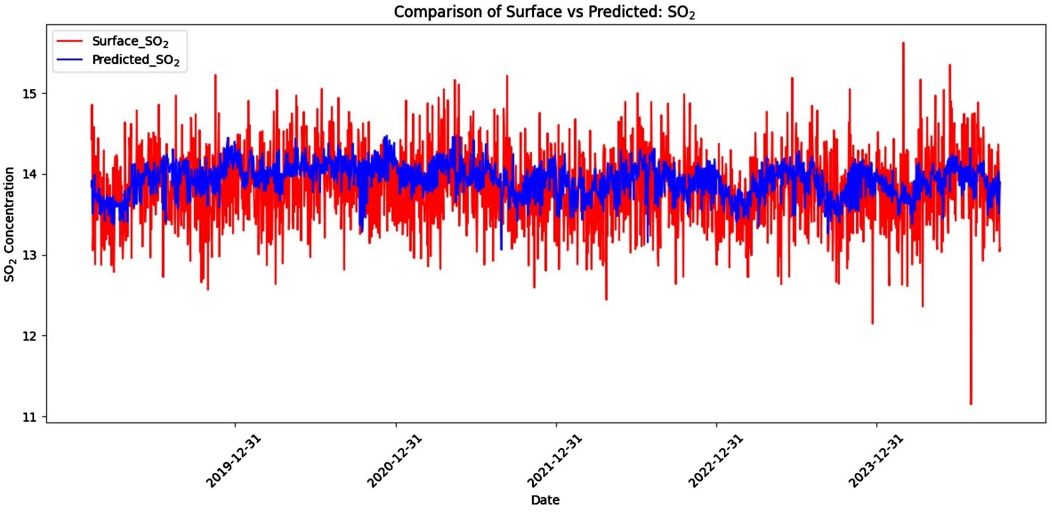



**Figure 15.** $SO_2$ Time series plots {CNN + LSTM modelling output (Blue) and Sentinel-5P Datasets (Red)}.

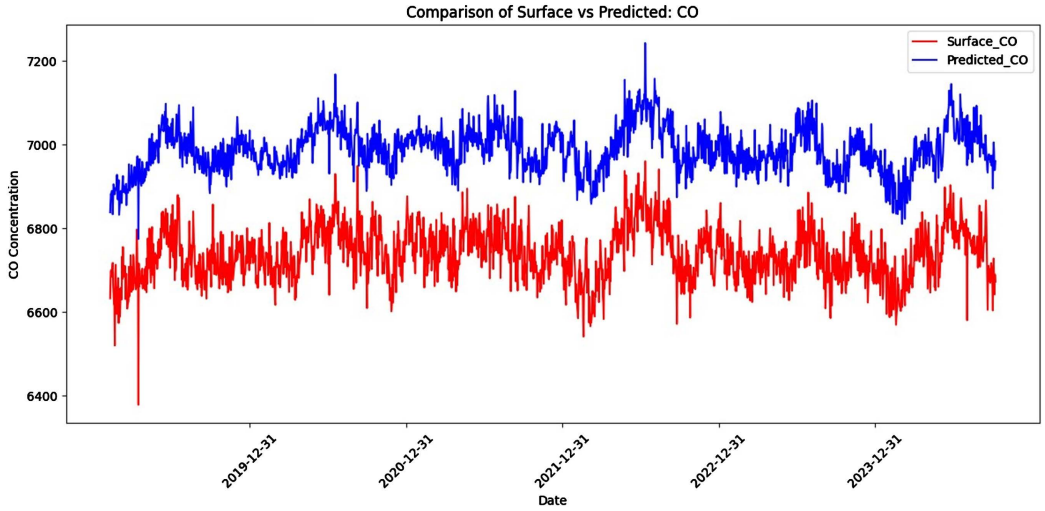



**Figure 16.** CO time series plots {CNN + LSTM modelling output (Blue) and Sentinel-5P Datasets (Red)}.

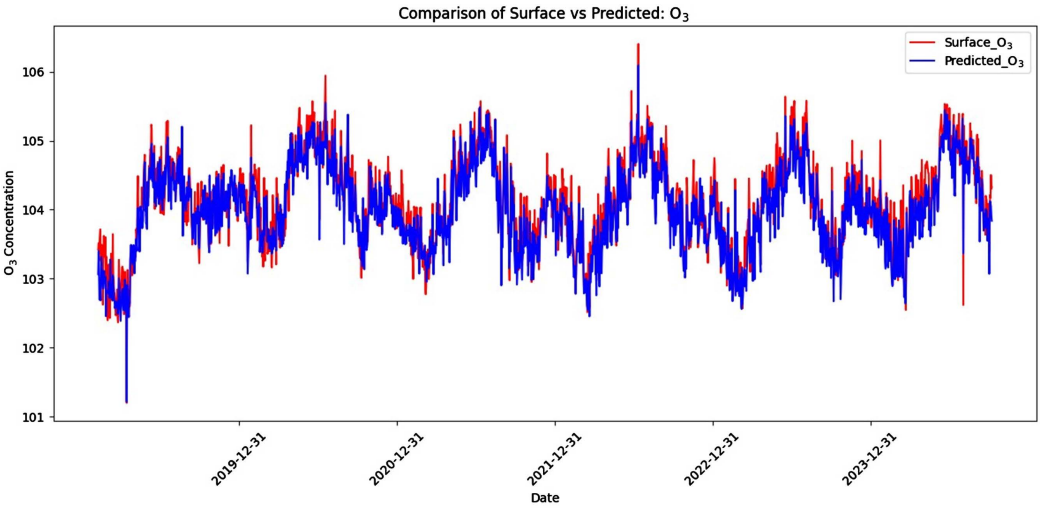



**Figure 17.** $O_3$ time series plots {CNN + LSTM modelling output (Blue) and Sentinel-5P Datasets (Red)}.

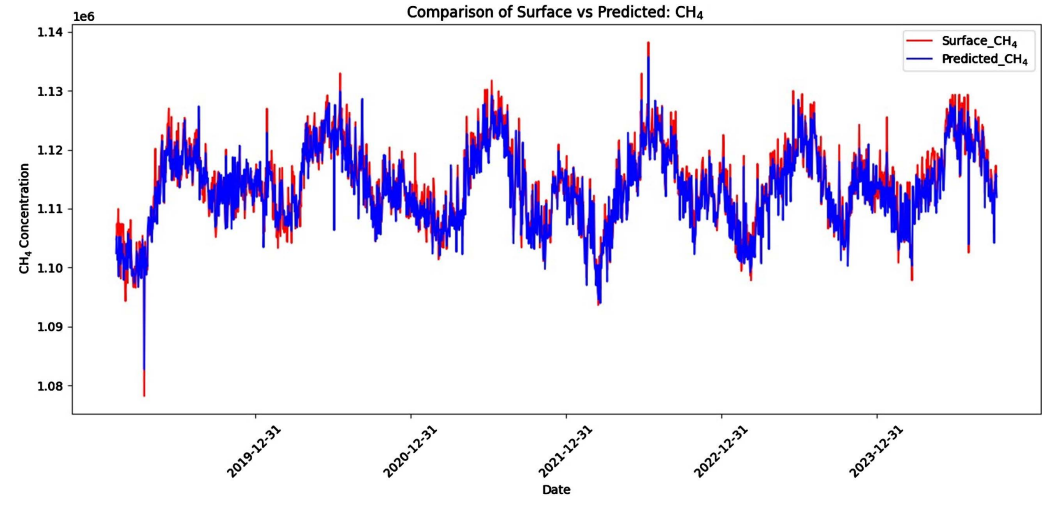



**Figure 18.** $CH_4$ time series plots {CNN + LSTM modelling output (Blue) and Sentinel-5P Datasets (Red)}.

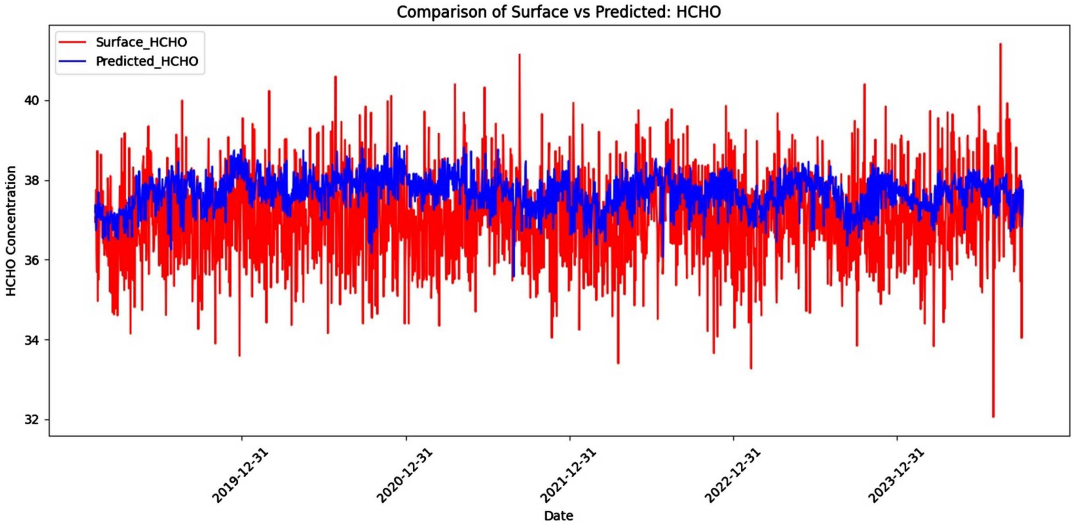



Figure 19. HCHO time series plots {CNN + LSTM modelling output (Blue) and Sentinel-5P Datasets (Red)}.

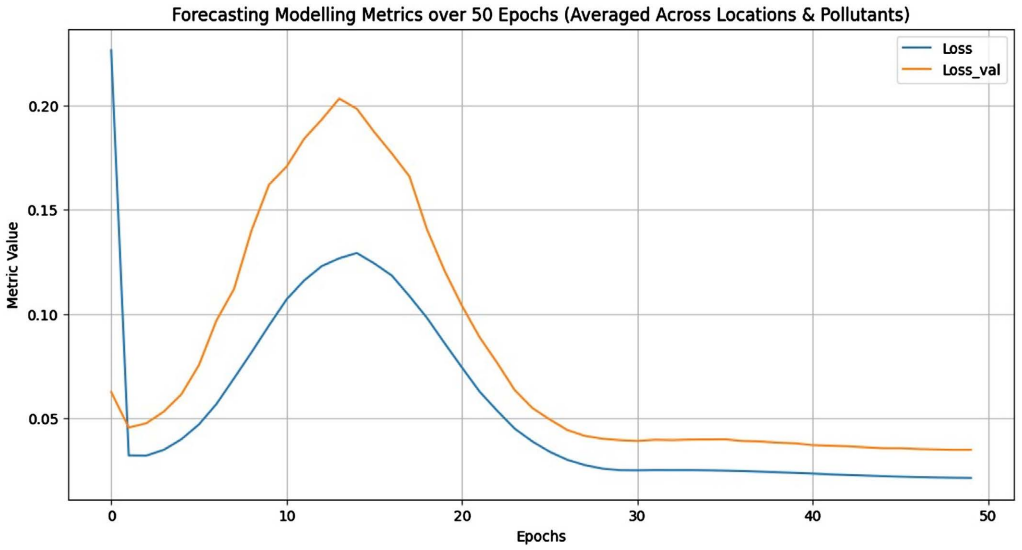



Figure 20. NeuralProphet model training metrics over 50 epochs (averaged across 327 locations & 7 pollutants showing summation of all MAEs & MAE_Vals).

Table 6. Results of neuralprophet forecasting metrics.

| S/No. | Metric | Value |
|---|---|---|
| 1 | MAE_val | 0.16 |
| 2 | RMSE_val | 0.20 |
| 3 | Loss_val | 0.13 |
| 4 | RegLoss_val | 0.00 |
| 5 | Epochs | 50.00 |
| 6 | MAE | 0.18 |
| 7 | RMSE | 0.22 |
| 8 | Loss | 0.11 |
| 9 | RegLoss | 0.00 |

**Table 7.** NeuralProphet model regression kriging accuracies (Forecast 2025).

| Model Type | Pollutant | MSE | $R^2$ | RMSE | MAE | Bias | Corr. |
|---|---|---|---|---|---|---|---|
| Random Forest Only | $PM_{10}$ | 186.454 | 0.786 | 13.655 | 8.921 | −0.417 | 0.894 |
| RF + Kriging | $PM_{10}$ | 261.682 | 0.700 | 16.177 | 9.719 | 4.094 | 0.890 |
| Random Forest Only | $NO_2$ | 81.504 | 0.811 | 9.028 | 6.638 | −0.060 | 0.926 |
| RF + Kriging | $NO_2$ | 85.292 | 0.803 | 9.235 | 6.732 | −1.005 | 0.939 |

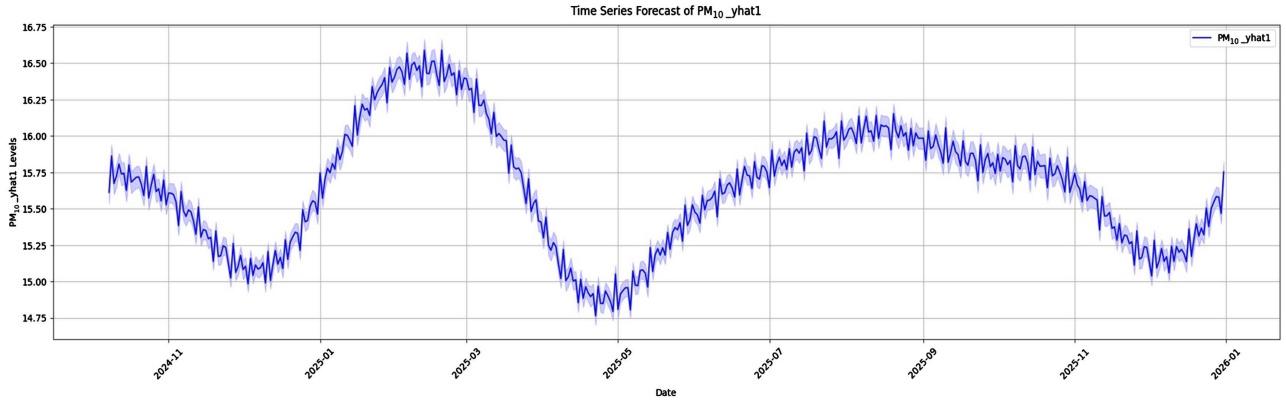



**Figure 21.** Forecast $PM_{10}$ concentrations levels

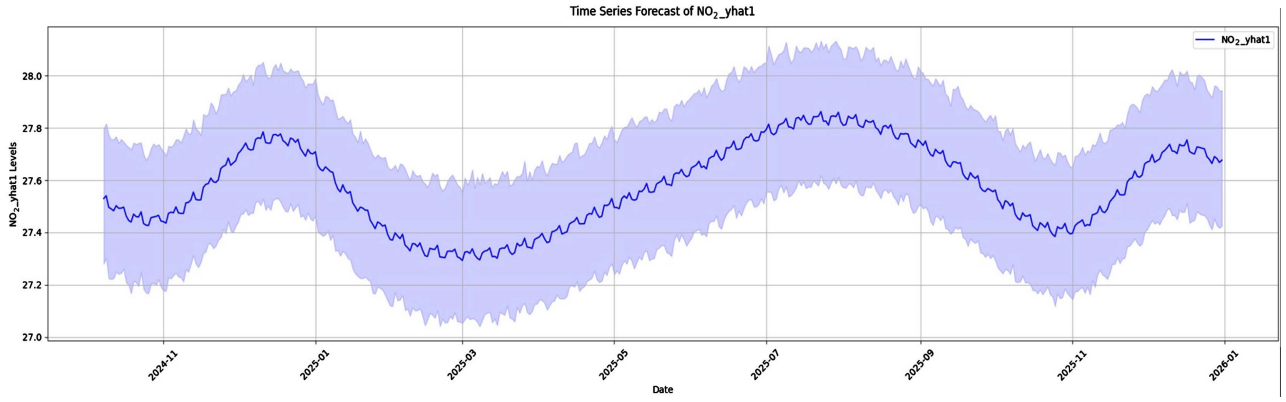



**Figure 22.** Forecasts $NO_2$ concentration levels.

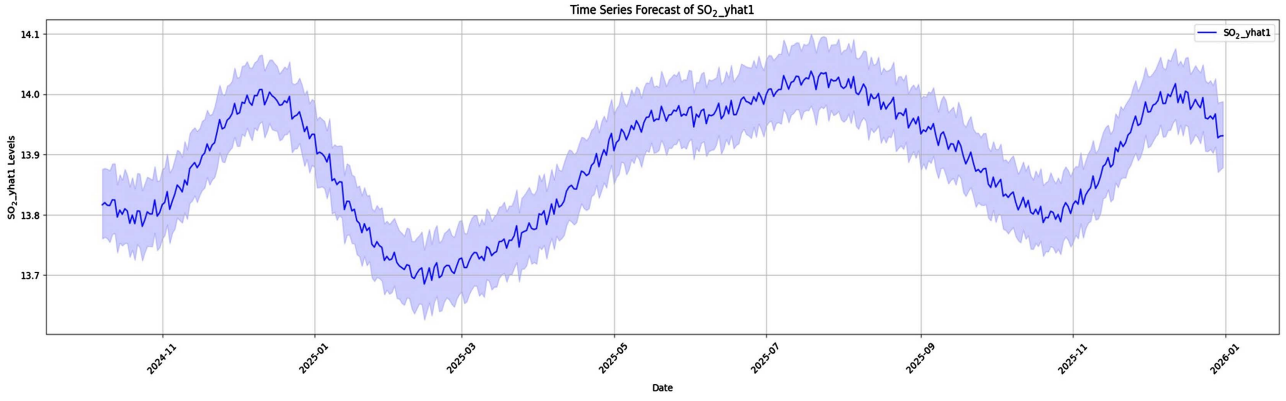



**Figure 23.** Forecasts $SO_2$ concentration levels.

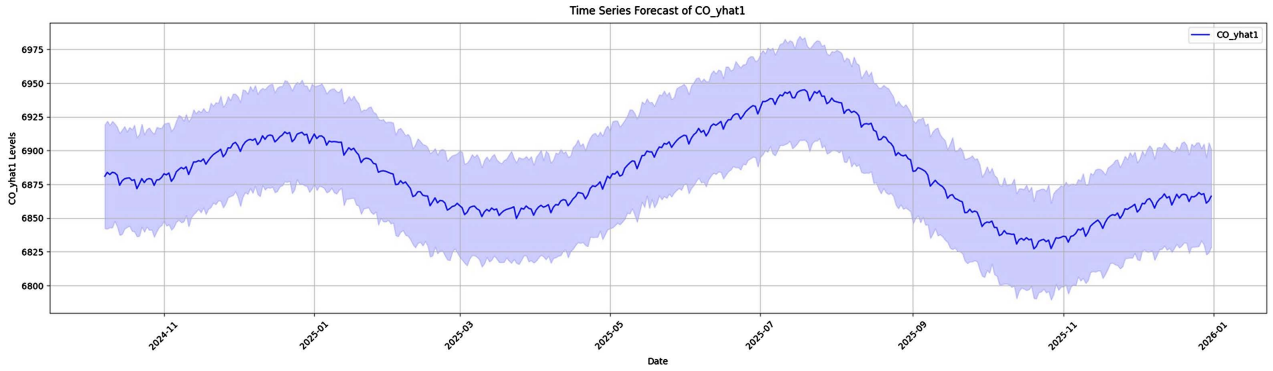



**Figure 24.** Forecast CO concentration levels.

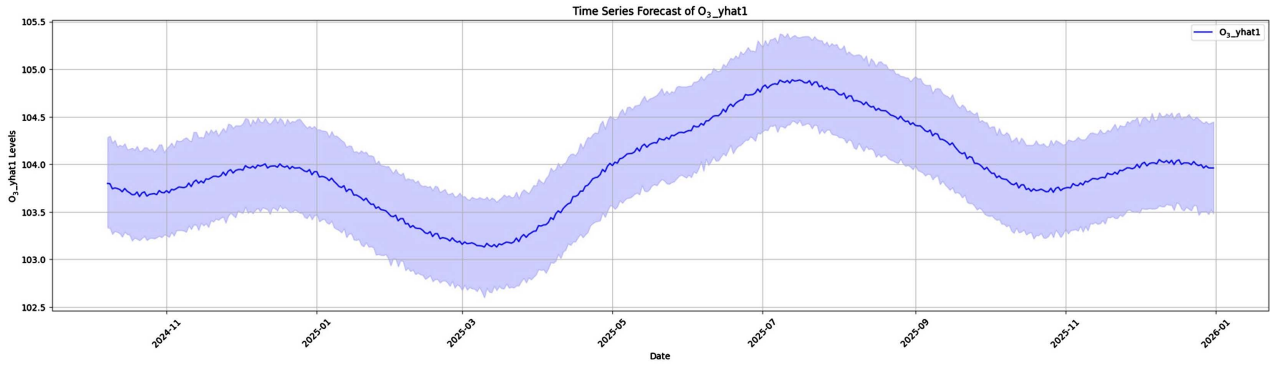



**Figure 25.** Forecast $O_3$ concentration levels.

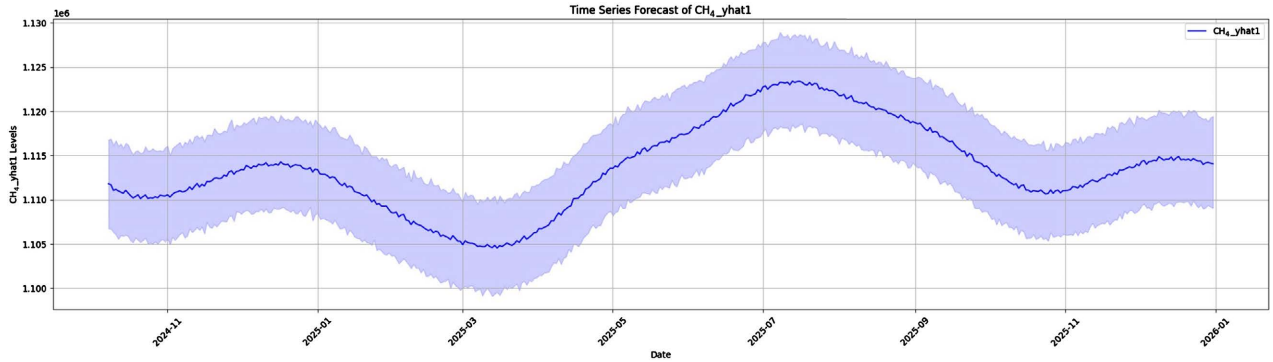



**Figure 26.** Forecast $CH_4$ concentration levels.

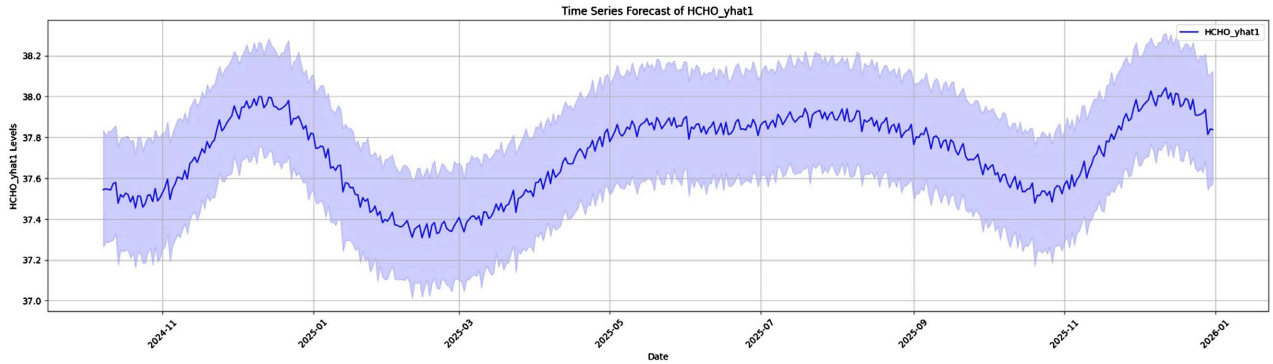



**Figure 27.** Forecast Formaldehyde (HCHO) concentration levels.

## 4. Discussion

This study demonstrates the viability of integrating deep learning models with Earth Observation (EO) and meteorological data to model and forecast urban air pollution across the Nairobi Metropolitan Area. Using a CNN + LSTM architecture, the research successfully captured both spatial and temporal pollution dynamics for the period 2019-2024, achieving strong model performance with a validation MAE of 0.0456 and RMSE of 0.20. The use of CNN layers enabled effective extraction of spatial patterns from gridded EO datasets, while LSTM units modelled temporal dependencies across seasons and years. These findings are consistent with prior studies that have shown the superiority of CNN-LSTM hybrid frameworks in environmental and spatio-temporal prediction tasks (Du et al., 2018; Tsokov et al., 2022)

The application of NeuralProphet to forecast air pollution levels into 2025 demonstrated its utility in preserving seasonal and trend-based signals within the time series. The model achieved validation metrics comparable to CNN + LSTM (MAE = 0.16, RMSE = 0.20), while offering the added benefit of interpretability through its decomposition of trend, seasonality, and holiday effects. These capabilities make NeuralProphet a practical forecasting tool for decision-makers who require explainable outputs for planning interventions.

Despite the strong performance of both models, notable underpredictions were observed for ground-level $PM_{10}$ and $NO_2$ concentrations—particularly in peri-urban areas with limited sensor calibration. To mitigate this, Regression Kriging was applied post-modelling to enhance spatial accuracy. For $NO_2$, Kriging improved spatial correlation and alignment, producing an $R^2$ of 0.996 and RMSE of 1.302 $\mu g/m^3$. This outcome underscores the compatibility of geostatistical interpolation with deep learning outputs, particularly for pollutants that exhibit smoother dispersion characteristics. However, the same approach resulted in diminished accuracy for $PM_{10}$ ($R^2 = 0.782$, RMSE = 13.805 $\mu g/m^3$), likely due to the pollutant's high spatial variability and episodic behaviour, which are difficult to interpolate with kriging methods.

The variability in performance between $NO_2$ and $PM_{10}$ highlights the importance of pollutant-specific modelling strategies and the need for dense, well-distributed ground sensors in complex urban environments. Furthermore, the limitations of EO-based proxies, such as using aerosol index for $PM_{10}$ estimation, may have contributed to the observed prediction biases. This finding supports earlier work by Widya et al., (2020) who noted that $PM_{10}$ & $NO_2$ models benefit from higher spatial density of ground sensors.

Overall, this study supports the integration of EO, deep learning, and geostatistics for urban air quality modelling. It also emphasizes the importance of *in situ* sensor data for validating and correcting model outputs—particularly for pollutants that are influenced by localized anthropogenic activities. The methodology is replicable and scalable, offering a valuable framework for cities in developing regions facing rapid urbanization and limited monitoring infrastructure.

## 5. Limitations

Several limitations of this study should be acknowledged to contextualize the findings and guide future research.

Limitations of the ground validation dataset (Flow 2 deployment): Several limitations of the Flow 2 deployment should be acknowledged. First, the sensor was not calibrated against a reference instrument, which may introduce absolute concentration biases. Second, the opportunistic mobile sampling (driving transects) preferentially captures roadside and transportation corridor concentrations, potentially oversampling emission hotspots relative to background residential areas. Third, the interpolation from 14,229 mobile points to 327 fixed grid points introduces smoothing uncertainty. Fourth, the single sensor unit could not provide simultaneous multi-location measurements, limiting the ability to capture synoptic spatial patterns. Despite these limitations, the Flow 2 dataset represents the most comprehensive ground-level air quality measurements available for the Nairobi Metropolitan Area to date and provides a valuable independent benchmark for evaluating satellite-driven models.

Column-to-surface mismatch in Sentinel-5P products: TROPOMI measures total vertical column densities (mol/m$^2$), not ground-level concentrations. While we applied physical conversion formulas incorporating temperature, pressure, elevation, and pollutant-specific vertical profiles (Section 2.2.1), these conversions rely on assumptions about the vertical distribution of each pollutant. For $O_3$, a fixed surface fraction of 20% was assumed based on tropical climatology, but this fraction varies seasonally and with meteorological conditions. For $NO_2$, an empirical enhancement factor of 1.5 was applied, calibrated against Flow 2 data; however, this factor may not generalize to other regions or time periods.

$PM_{10}$ derivation from Aerosol Index: Sentinel-5P does not directly measure particulate matter. $PM_{10}$ was estimated using a linear min-max scaling of the Aerosol Index to the WHO 24-hour guideline limit of 50 μg/m$^3$. This method assumes: 1) a linear relationship between AI and surface $PM_{10}$, 2) that the maximum observed AI in 2019-2024 corresponds to the WHO limit, and 3) that aerosol composition and optical properties are constant across the study area. These assumptions introduce uncertainty, particularly during dust events or biomass burning episodes when aerosol properties differ from urban pollution aerosols. This limitation likely contributed to the lower $PM_{10}$ prediction accuracy ($R^2$ = 0.782 after kriging) compared to $NO_2$ ($R^2$ = 0.996).

Limited suitability of kriging for $PM_{10}$: As shown in Section 3.1.1, Regression Kriging degraded $PM_{10}$ accuracy ($R^2$ dropped from 0.915 to 0.782; RMSE increased from 8.60 to 13.81 μg/m$^3$) because $PM_{10}$ exhibits high spatial variability and episodic behavior (e.g., construction dust, road resuspension, localized sources) that violates the spatial stationarity assumption underlying variogram-based interpolation. Kriging remains useful for smoothly varying pollutants like $NO_2$ (Moran's I = 0.994) but is not recommended for $PM_{10}$ without dense ground monitoring or alternative spatial smoothing methods.

Temporal and spatial validation scope: The CNN + LSTM model was validated

on the last 20% of dates (2024) but using the same 327 spatial points in both training and validation. This tests temporal generalizability but not spatial generalizability to unmonitored locations. Future work should employ spatial holdout validation (e.g., reserving specific geographic points entirely from training) to assess model performance in areas without any historical monitoring.

NeuralProphet simplification: The forecasting model used no autoregressive terms (n_lags = 0), meaning predictions were based solely on trend and seasonality decomposition without short-term lagged dependencies. While this captures annual and weekly cycles, it may miss day-to-day persistence effects (e.g., multi-day pollution episodes). Future work should explore including lagged terms (n_lags > 0) and exogenous regressors (e.g., forecasted meteorology) to improve short-term forecast accuracy.

Single-city scope: The models were developed and validated only for the Nairobi Metropolitan Area. While the framework is designed to be replicable, the specific conversion parameters (e.g., $NO_2$ enhancement factor of 1.5, $O_3$ surface fraction of 20%) may not transfer directly to other cities with different emission profiles, meteorology, or topography. Transfer learning or regional re-calibration would be required for application elsewhere.

## 6. Conclusion

This study demonstrates the effectiveness of hybrid deep learning models in capturing and forecasting urban air pollution dynamics when integrated with satellite-based Earth Observation (EO) data and validated through ground sensor networks. The CNN + LSTM architecture successfully modelled spatio-temporal variations in key pollutants such as $NO_2$ and $O_3$, while NeuralProphet offered explainable and seasonally consistent forecasts for future concentrations. The integration of Regression Kriging improved spatial resolution for $NO_2$, which exhibits strong spatial autocorrelation (Moran's I = 0.994). However, for $PM_{10}$, kriging degraded predictive accuracy due to the pollutant's high spatial variability and episodic emission sources. This trade-off underscores that geostatistical post-processing is not universally beneficial; its effectiveness depends critically on the underlying spatial structure of the target pollutant. Future work should explore alternative spatial smoothing techniques for $PM_{10}$, such as random forests with spatial cross-validation, geographically weighted regression, or ensemble methods that combine multiple interpolation approaches.

These results underscore the potential of scalable, data-driven frameworks for supporting proactive environmental policy and public health planning in rapidly urbanizing regions. However, challenges remain in forecasting pollutants like $PM_{10}$, which are highly variable and influenced by localized sources. Future work should prioritize the real-time assimilation of *in situ* sensor data, as well as the adoption of ensemble learning strategies to enhance model robustness, especially for coarse or episodic pollutants. Ultimately, this approach lays the groundwork for replicable, cost-effective air quality monitoring systems aligned with global sustainability and urban resilience goals (**Table 8**).

Table 8. Random forest Krigging Trade-off.

| Metric | $NO_2$ (RF only) | $NO_2$ (RF + Kriging) | $PM_{10}$ (RF only) | $PM_{10}$ (RF + Kriging) |
|---|---|---|---|---|
| $R^2$ | 0.998 | 0.996↓ | 0.915 | 0.782↓ |
| RMSE | 0.832 | 1.302↑ | 8.6 | 13.805↑ |
| Bias | 0 | 0.455↑ | −0.004 | −4.136↑ |

Arrows indicate direction of change (↓ = worse, ↑ = worse for error metrics).

These findings suggest that satellite-driven deep learning models can support low-cost air quality early warning systems in data-sparse regions like sub-Saharan Africa, where regulatory monitoring infrastructure is limited.

## Acknowledgements

First and foremost, I am deeply grateful to the LORD Almighty God for the gift of life, His constant protection, and divine provision throughout this long and often torturous journey of my study. Without His grace and sustaining power, this work would not have been possible.

I sincerely thank my beloved wife Susan and our children for their unwavering support, patience, and understanding—especially during the many late nights spent in study and research. Your love and encouragement kept me going.

Special appreciation goes to my dedicated supervisors and academic mentors. I am profoundly thankful to Dr. Solomon Mwanjele, whose guidance on Geoprogramming and overall structure of this thesis was instrumental in shaping the direction and quality of this research. I also extend my heartfelt thanks to Dr. Arthur Sichangi, who diligently followed up with me through each chapter and section, offering valuable insights and support.

I would also like to express my sincere gratitude to Dr. Nashon Juma Adero for encouraging me to enroll in the Msc. Geoinformatics program at Taita Taveta University and for introducing me to Systems Thinking. His support went beyond mentorship—he facilitated the shipment of the Flow 2 air quality data collector that played a crucial role in this research.

I remain indebted to all my lecturers at Taita Taveta University for their contributions. In particular, I thank Dr. Ngesa for the rigorous training in Geostatistics and Research Methods, and Dr. Mika Siljander for the comprehensive Geoinformatics course materials and practical GIS training that greatly enriched my academic experience. Dr. Grace for the overall coordination of the program, Pro. Makokha for in-depth analysis of modelling techniques (Abera et al., 2021).

Finally, I appreciate the Taita Taveta University administration for accepting my application and granting me the opportunity to pursue this course. This academic journey has been a transformative experience, and I thank all who contributed to its success.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References


Abera, A., Friberg, J., Isaxon, C., Jerrett, M., Malmqvist, E., Sjöström, C. et al. (2021). Air Quality in Africa: Public Health Implications. *Annual Review of Public Health, 42,* 193-210. https://doi.org/10.1146/annurev-publhealth-100119-113802

Addiena A Rahim, N. A., Noor, N. M., Mohd Jafri, I. A., Ul-Saufie, A. Z., Ramli, N., Abu Seman, N. A. et al. (2023). Variability of PM10 Level with Gaseous Pollutants and Meteorological Parameters during Episodic Haze Event in Malaysia: Domestic or Solely Transboundary Factor? *Heliyon, 9,* e17472. https://doi.org/10.1016/j.heliyon.2023.e17472

Bergmeir, C., & Benítez, J. M. (2012). On the Use of Cross-Validation for Time Series Predictor Evaluation. *Information Sciences, 191,* 192-213. https://doi.org/10.1016/j.ins.2011.12.028

Cyton (2023). *Nairobi Metropolitan Area Residential Report 2023.* Cyton.

Du, S. D., Li, T., Yang, Y., & Horng, S. J. (2018). *Deep Air Quality Forecasting Using Hybrid Deep Learning Framework.* http://arxiv.org/pdf/1812.04783

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L. et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate, 30,* 5419-5454. https://doi.org/10.1175/jcli-d-16-0758.1

KNBS (2019). *2019 Kenya Population and Housing Census.* Kenya National Bureau of Statistics.

Lin, C., Chen, Y., Liu, C., Chen, W., Seinfeld, J. H., & Chou, C. C. K. (2019). Satellite-derived Correlation of $SO_2$, $NO_2$, and Aerosol Optical Depth with Meteorological Conditions over East Asia from 2005 to 2015. *Remote Sensing, 11,* Article 1738. https://doi.org/10.3390/rs11151738

Lorente, A., Boersma, K. F., Eskes, H. J., Veefkind, J. P., van Geffen, J. H. G. M., de Zeeuw, M. B. et al. (2019). Quantification of Nitrogen Oxides Emissions from Build-Up of Pollution over Paris with Tropomi. *Scientific Reports, 9,* Article No. 20033. https://doi.org/10.1038/s41598-019-56428-5

Mamić, L., Gašparović, M., & Kaplan, G. (2023). Developing $PM_{2.5}$ and $PM_{10}$ Prediction Models on a National and Regional Scale Using Open-Source Remote Sensing Data. *Environmental Monitoring and Assessment, 195,* Article No. 644. https://doi.org/10.1007/s10661-023-11212-x

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of Spatial Predictor Variable Selection in Machine Learning Applications—Moving from Data Reproduction to Spatial Prediction. *Ecological Modelling, 411,* Article ID: 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

NeuralProphet Documentation (2024). *Tutorial 3: Seasonality—NeuralProphet 1.0.0rc8 Documentation.* https://neuralprophet.com/tutorials/tutorial03.html

O'Donncha, F., Hu, Y., Palmes, P., Burke, M., Filgueira, R., & Grant, J. (2021). A Spatio-Temporal LSTM Model to Forecast across Multiple Temporal and Spatial Scales. *Ecological Informatics, 69,* Article ID: 101687. https://doi.org/10.1016/j.ecoinf.2022.101687

Oguge, O., Nyamondo, J., Adera, N., Okolla, L., Okoth, B., Anyango, S. et al. (2024). Fine Particulate Matter Air Pollution and Health Implications for Nairobi, Kenya. *Environmental Epidemiology, 8,* e307. https://doi.org/10.1097/ee9.0000000000000307

Omwenga, M. (2010). *Nairobi—Emerging Metropolitan Region: Development Planning and Management Opportunities and Challenges.* https://www.isocarp.net/data/case_studies/1662.pdf

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing, 12,* Article 2291. https://doi.org/10.3390/rs12142291

Plume Labs (2023). *Be Empowered against Air Pollution.*

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. New York Academy of Sciences Series.* John Wiley & Sons. https://www.wiley.com/en-gb/Atmospheric+Chemistry+and+Physics%3A+From+Air+Pollution+to+Climate+Change%2C+3rd+Edition-p-9781118947401

Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S. W. et al. (2013). The Changing Paradigm of Air Pollution Monitoring. *Environmental Science & Technology, 47,* 11369-11377. https://doi.org/10.1021/es4022602

Thompson, A. M., Stauffer, R. M., Wargan, K., Witte, J. C., Kollonige, D. E., & Ziemke, J. R. (2021). Regional and Seasonal Trends in Tropical Ozone from SHADOZ Profiles: Reference for Models and Satellite Products. *Journal of Geophysical Research: Atmospheres, 126,* e2021JD034691. https://doi.org/10.1029/2021jd034691

Tsokov, S., Lazarova, M., & Aleksieva-Petrova, A. (2022). A Hybrid Spatiotemporal Deep Model Based on CNN and LSTM for Air Pollution Prediction. *Sustainability, 14,* Article 5104. https://doi.org/10.3390/su14095104

Uuemaa, E., Ahi, S., Montibeller, B., Muru, M., & Kmoch, A. (2020). Vertical Accuracy of Freely Available Global Digital Elevation Models (ASTER, AW3D30, MERIT, TandemX, SRTM, and NASADEM). *Remote Sensing, 12,* Article 3482. https://doi.org/10.3390/rs12213482

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J. et al. (2022). Sentinel-5P TROPOMI $NO_2$ Retrieval: Impact of Version V2.2 Improvements and Comparisons with OMI and Ground-Based Data. *Atmospheric Measurement Techniques, 15,* 2037-2060. https://doi.org/10.5194/amt-15-2037-2022

Wang, Y., He, H., Huang, H., Yang, J., & Peng, Z. (2025). High-Resolution Spatiotemporal Prediction of $PM_{2.5}$ Concentration Based on Mobile Monitoring and Deep Learning. *Environmental Pollution, 364,* Article ID: 125342. https://doi.org/10.1016/j.envpol.2024.125342

Wen, Y., Xiao, J., Yang, J., Cai, S., Liang, M., & Zhou, P. (2022). Quantitatively Disentangling the Geographical Impacts of Topography on $PM_{2.5}$ Pollution in China. *Remote Sensing, 14,* Article 6309. https://doi.org/10.3390/rs14246309

Widya, L. K., Hsu, C., Lee, H., Jaelani, L. M., Lung, S. C., Su, H. et al. (2020). Comparison of Spatial Modelling Approaches on $PM_{10}$ and $NO_2$ Concentration Variations: A Case Study in Surabaya City, Indonesia. *International Journal of Environmental Research and Public Health, 17,* Article 8883. https://doi.org/10.3390/ijerph17238883

Zhang, H., Li, A., Wang, S., Qin, M., Hu, Z., & Xu, J. (2024). $O_3$ Sensitivity and Vertical Distribution of Summertime HCHO, $NO_2$, and $SO_2$ in Shihezi, China. *Atmospheric Pollution Research, 15,* Article ID: 102113. https://doi.org/10.1016/j.apr.2024.102113

## Definitions, Acronyms, Abbreviations

**Spatio-Temporal Modelling:** An analytical technique that captures variations across both space and time, often used to understand how environmental phenomena like air pollution change geographically and over periods.

**Earth Observation (EO):** The collection of information about Earth's physical, chemical, and biological systems using remote sensing technologies, particularly from satellite platforms such as Sentinel-5P and Sentinel-2.

**Deep Learning (DL):** A subset of machine learning involving neural networks with multiple layers (e.g., CNNs and LSTMs) that automatically learn features from data.

**CNN (Convolutional Neural Network):** A type of deep learning model particularly well-suited for extracting spatial features from data such as images or satellite-derived grids.

**LSTM (Long Short-Term Memory):** A type of recurrent neural network designed to model sequential data and capture long-range temporal dependencies, often used for time-series forecasting.

**NeuralProphet:** A neural network-based time series forecasting tool developed on top of Facebook's Prophet model. It integrates trend, seasonality, and autoregressive components for prediction.

**Regression Kriging:** A spatial interpolation method that combines regression modelling of a dependent variable on auxiliary variables with kriging of the residuals to improve prediction accuracy.

**Geary's C:** A spatial autocorrelation metric that measures the extent to which similar values cluster together in space; more sensitive to local differences than Moran's I.

**Moran's I:** A measure of spatial autocorrelation indicating whether spatial patterns are clustered, dispersed, or random.

**Variogram:** A fundamental geostatistical tool that describes the degree of spatial dependence between sample data over distance.

**Sentinel-5P/TROPOMI:** A satellite platform under ESA's Copernicus program equipped with the TROPOspheric Monitoring Instrument (TROPOMI) used to measure atmospheric gases including $NO_2$, CO, and $O_3$.

**MERRA-2:** NASA's Modern-Era Retrospective Analysis for Research and Applications Version 2—a reanalysis product providing meteorological data for climate and atmospheric studies.

**Flow 2:** A portable air quality monitoring device by Plume Labs used to collect ground-level pollution data (e.g., $PM_{10}$, $NO_2$) used in this study for model validation.

**Aerosol Index (AER_AI):** A satellite-derived index that measures the presence of absorbing aerosols (e.g., smoke, dust) in the atmosphere, used here as a proxy for $PM_{10}$.

**Google Earth Engine (GEE):** A cloud-based platform that enables large-scale geospatial data processing and analysis, especially from EO datasets.

**Rectified Linear Unit (ReLU):** a commonly used and computationally efficient activation function in deep learning that returns the input value when it is positive, and outputs zero when it is negative or zero.

**Root Mean Squared Error (RMSE):** A common measure of model prediction error that gives higher weight to large errors.

**Mean Absolute Error (MAE):** The average of absolute differences between predicted and observed values—a linear error metric.

**Spatial Autocorrelation:** A measure of the degree to which spatial observations resemble each other over a given distance.