Received: 17 November 2025 | Revised: 28 May 2026 | Accepted: 24 July 2026 | Published online: 22 September 2026

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2026, Vol. 00(00) 1–19
DOI: 10.47852/bonviewJDSIS62028274

BON VIEW PUBLISHING

A Data-Driven and Cost-Effective Framework for Urban Air Quality Monitoring and NO_2 Forecasting in Cheltenham

Johnson Ohakwe[1,*], Derrick Ofori[1], Bhupesh Kumar Mishra[2], William Sayers[1], Timothy Olusakin[1] and John Chisimkwuo[3]

[1] *Department of Data Science, University of Gloucestershire, UK*

[2] *Department of Data Science, University of Hull, UK*

[3] *Department of Statistics, Michael Okpara University of Agriculture Umudike, Nigeria*

Abstract: Effective urban air quality management requires monitoring systems that are reliable, affordable, and fairly distributed across commercial and residential areas. This study examines nitrogen dioxide (NO_2) concentrations in Cheltenham by comparing commercial sites, Boots Corner/High Street, with the residential site London Road across pre-COVID-19, COVID-19, and post-COVID-19 periods. Missing values were handled using mean imputation, and nonparametric Kruskal–Wallis H-tests followed by Dunn's post hoc pairwise comparisons were used to compare monitoring locations. Forecasting performance was assessed using Autoregressive Integrated Moving Average (ARIMA), Holt-Winters exponential smoothing (Holt-Winters), Extreme Gradient Boosting (XGBoost), Gated Recurrent Unit, and Long Short-Term Memory (LSTM) models, with root mean square error (RMSE) as the evaluation metric and bootstrap simulations for validation. Results showed that NO_2 levels at London Road were statistically indistinguishable from Cheltenham citywide levels ($p = 0.74$), suggesting it can act as a cost-effective proxy for average citywide trends, although it should not replace monitoring of short-term fluctuations or localized pollution episodes. In contrast, Boots Corner/High Street differed significantly from citywide levels ($p < 0.05$). For forecasting, Holt-Winters slightly outperformed ARIMA, while LSTM achieved the best overall accuracy, with an RMSE of 2.66 and about 16% lower error than competing models. Bootstrap simulations supported these findings, though performance varied with dataset size. Overall, London Road can support cost-efficient NO_2 surveillance in Cheltenham, while supplementary monitoring remains important. LSTM models should be prioritized for high-accuracy forecasting, with bootstrap validation guiding model selection.

Keywords: air quality, machine learning, nitrogen dioxide, air pollution, data analysis

1. Introduction

Despite improvements in ambient air quality across Europe in recent decades [1], current pollution levels still pose significant health risks [2], particularly in the United Kingdom, where nitrogen dioxide (NO_2) levels frequently exceed European Union thresholds due to the high prevalence of diesel vehicles [3–5]. As NO_2 is a leading contributor to premature mortality worldwide [6], accurately modeling its concentration and informing mitigation strategies is essential for public health and environmental policy [7, 8]. Robust models must capture the complex interactions between emission sources, geography, and meteorology [9–13], drawing on approaches ranging from machine learning (ML) to dispersion and atmospheric chemistry methods such as Gaussian, Lagrangian, and Eulerian models [14]. The public health case for reducing NO_2 is well established: recent systematic reviews and large cohort studies link long-term NO_2 and NO_x exposure to elevated all-cause, cardiovascular, and respiratory mortality [15–17], with excess risk persisting even at concentrations below current regulatory limits [18], and short-term exposures posing particular risks for the elderly and other vulnerable groups [19, 20].

Despite this substantial body of work, Cheltenham's air quality has remained unexamined for over two decades (1994–2022), a notable gap given the town's mix of commercial and residential exposure profiles. This study addresses that gap directly by analyzing NO_2 levels at Boots Corner/High Street (commercial) and London Road (residential), Cheltenham, across the pre-, during-, and post-COVID-19 periods. The aim is twofold: to characterize local NO_2 trends and to assess whether either site's NO_2 attributes are representative of broader patterns across Cheltenham. Identifying such a representative location would offer a cost-effective alternative to citywide monitoring for future air quality studies. The COVID-19 lockdowns provide a useful backdrop for such trend analysis, as satellite and ground-based observations

***Corresponding author:** Johnson Ohakwe, Department of Data Science, University of Gloucestershire, UK. Email: johakwe2@glos.ac.uk

documented substantial NO_2 reductions across cities worldwide, with magnitudes that varied by local mobility and, in some regions, occurred even without formal lockdown measures [21–23]. Low-cost sensor networks have likewise been proposed to increase the spatial density of observations, although their reliability depends on careful calibration and siting relative to reference-grade instruments [24–26]; calibration uncertainty has been shown to grow with distance from reference stations, reinforcing the value of identifying representative monitoring locations [27].

Comparison tests, exploratory and descriptive analysis, time series methods, and ML have each been used productively elsewhere to characterize and forecast air pollution, for example, Long Short-Term Memory (LSTM)-based forecasting in Beijing [28], spatial comparisons between Indian cities [29], and combined statistical-ML approaches that pair exploratory analysis with forecasting for a fuller picture [30–32]. These studies consistently show that no single method is sufficient on its own and that synergy between descriptive, inferential, and ML techniques produces more robust, interpretable predictions [28, 30–32]. More recent studies have extended these approaches with recurrent and hybrid deep learning architectures, in which LSTM networks and their convolutional variants frequently outperform classical baselines such as Autoregressive Integrated Moving Average (ARIMA) [33, 34]; comparative evaluations using root mean square error (RMSE) across Convolutional Neural Network (CNN), LSTM, and Gated Recurrent Unit (GRU) models report that no single architecture dominates across all pollutants [35, 36], while decomposition-based hybrids and systematic hyperparameter optimization offer further accuracy gains [37, 38], including in locations without dedicated monitoring [39].

This study brings that synergy to bear on Cheltenham specifically. Using descriptive and inferential statistics, ML, and Bootstrapping-based simulation, it (i) compares NO_2 levels at Cheltenham's commercial and residential sites against the wider area's patterns; (ii) develops and evaluates predictive models, selecting the best-performing approach via RMSE; and (iii) validates that selection by re-applying both traditional time series models and ML algorithms to three bootstrap-simulated NO_2 datasets, with the overall best model again chosen by RMSE [40].

As shown in Table 1, existing studies have made important contributions to air quality monitoring, pollutant forecasting, and spatial air pollution analysis. However, many prior studies focus on pollutant characterization, statistical comparisons, or predictive modeling in isolation. In contrast, the present study provides an integrated framework for Cheltenham NO_2 analysis by combining location-based comparison, exploratory analysis, time series modeling, ML forecasting, RMSE-based model evaluation, and bootstrap simulation. The novelty of this study lies in its localized Cheltenham case study, its identification of London Road as a potential cost-effective proxy for wider NO_2 surveillance, and its validation of model selection using simulated bootstrap datasets.

The remainder of this paper is organized as follows. Section 2 presents the research methodology, covering data collection and preprocessing, the comparison test, exploratory data analysis (EDA), time series modeling (ARIMA and Holt-Winters exponential smoothing (Holt-Winters)), ML modeling (LSTM, Extreme Gradient Boosting (XGBoost), and GRU), model evaluation, and the bootstrap-based data simulation used to validate model selection. Section 3 reports the corresponding results and discussion, including a comparison with existing studies. Section 4 concludes the paper and outlines directions for future research.

2. Research Methodology

2.1. Research framework

To enhance clarity and provide a structured overview of the study, a research framework is presented in Figure 1. The framework outlines the sequential process from data collection and preprocessing through exploratory analysis, modeling, simulation, and final model evaluation.

2.2. Data collection, preprocessing, and feature engineering

This study utilizes secondary data obtained from Cheltenham Borough Council to analyze monthly NO_2 concentrations (measured in $\mu g/m^3$) across selected locations in Cheltenham between 1994 and 2022. The dataset is publicly available through the council's environmental monitoring reports and is used in accordance with open-access data provisions for research and non-commercial purposes.

The dataset consists of 345 observations (rows) representing monthly aggregated NO_2 readings over the study period. The original dataset includes variables such as monitoring location, date (year and month), and NO_2 concentration levels. From this, key variables selected for analysis include:

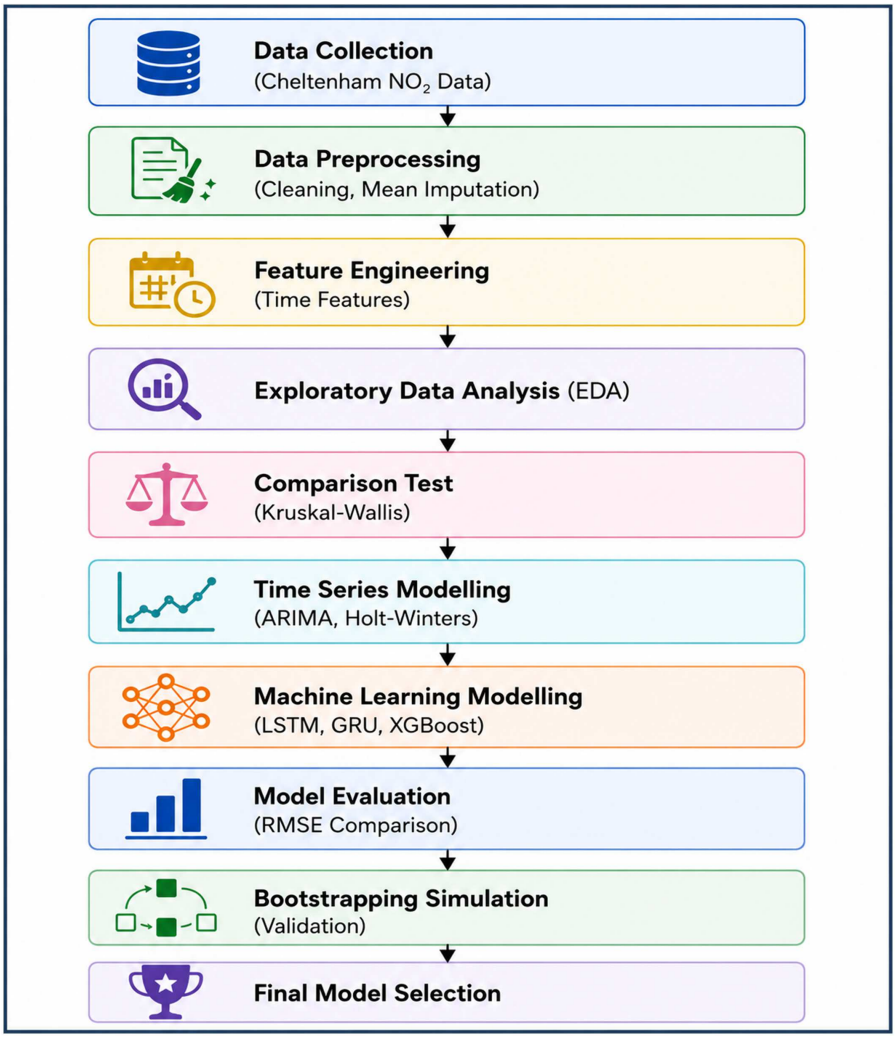
1) Location (Cheltenham, London Road, Boots Corner/High Street)
2) Time (year and month)
3) NO_2 concentration ($\mu g/m^3$)

Data preprocessing involved aggregating NO_2 readings from multiple monitoring sites. For Cheltenham, readings across all available monitoring stations were combined. At the same time, for London Road (sites 81, 104, 212 London Road) and Boots Corner/High Street (including Boots Corner, 452, 422, 443, YMCA, Cutting Room, Restoration, Pisa Pizza, and The Swan High Street), site-specific data were aggregated to represent

Table 1
Comparative summary of related work and distinction from the proposed study

Study	Focus/pollutant	Method used	Key contribution	Limitation/gap
Li et al. [28]	Air pollutant prediction	LSTM	Improved forecasting accuracy	Limited comparative modeling
Marinov et al. [32]	Air quality forecasting	Time series	Forecasting approach	No ML integration
Chen et al. [41]	Multi-pollutant	Random Forest	High-resolution prediction	No cost-effective monitoring insight

Figure 1
Research framework



each area. Annual and monthly mean concentrations were computed to ensure consistency and comparability. Missing values were replaced with column means to maintain statistical integrity. Below is a mathematical representation of a mean across columns:

$$\overline{X}_j = \sum_{j=1}^{N} X_j / N \tag{1}$$

Where:

$\overline{X_j}$ is the mean of NO_2 across the jth columns, X_j are the observations of NO_2 across the jth columns, and N is the total number of NO_2 observations across the jth columns.

The mean imputation method, which preserves the data's overall distribution, was utilized for missing data, a standard in air pollution studies. The cleaned, aggregated dataset with monthly means of NO_2 concentration (1994–2022) was prepared for further analysis across selected locations. Feature engineering was applied to enhance the models' predictive performance. Additional temporal variables were derived from the date information, including:

1) Day of the year
2) Day of the week
3) Month
4) Quarter

The feature engineering process was limited to variables that could be derived directly from the official dataset provided by Cheltenham Borough Council. Consequently, temporal features such as year, month, quarter, and other calendar-based variables were generated from the sampling dates to capture seasonal and long-term temporal patterns in NO_2 concentrations. Although meteorological variables (e.g., temperature, humidity, wind speed, rainfall, and wind direction) and traffic flow data are recognized as important determinants of NO_2 variability, these variables were not available in the original monitoring dataset and were therefore not incorporated into the predictive models.

2.3. Comparison test

This study uses the Kruskal–Wallis H-test to compare NO_2 levels in Cheltenham between two locations: London Road and Boots Corner/High Street. The Kruskal–Wallis H-test, a non-parametric alternative to one-way analysis of variance, is suitable for comparing two or more independent groups when the assumption of normality is unmet [42]. The test was applied twice: first to compare NO_2 levels between Cheltenham and London Road and

then between Cheltenham and Boots Corner/High Street. The test is mathematically represented as:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{g} n_i \bar{r}_i - 3(N+1) \quad (2)$$

Where:

N represents the overall number of observations across all groups, g denotes the total number of groups, and n_i is the number of observations in group i. The quantity $\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$ refers to the mean rank of the observations within group i.

The Kruskal–Wallis H-test was chosen because the data did not follow a normal distribution, and variances among groups were not homogeneous. The hypotheses of the Kruskal–Wallis H-test are presented below as:

H_0: The population median of all the groups is equal.

Against.

H_1: There is a significant difference among the groups.

The Kruskal–Wallis H-test was executed using a significance level of $\alpha = 0.05$. According to the decision criterion, the null hypothesis (H_0) is rejected when the p-value < 0.05, signifying a significant difference in NO_2 levels. If the p-value ≥ 0.05, the null hypothesis is not rejected, suggesting similarity in NO_2 levels.

After the Kruskal–Wallis H-test is performed, the study then performs a Dunn's post hoc comparison test to ascertain whether London Road and Boots Corner/High Street differ significantly from each other. Dunn's test compares the mean ranks of two groups while accounting for the pooled ranking of all observations.

Suppose there are k independent groups with a total sample size:

$$N = \sum_{i=1}^{k} n_i \quad (3)$$

where n_i is the number of observations in group i.

Let:

R_i = sum of ranks for group i,

$\bar{R}_i = \frac{R_i}{n_i}$ = mean rank of group i.

For any pair of groups i and j, the Dunn test statistic is computed as:

$$Z_{ij} = \frac{\bar{R}_i - \bar{R}_j}{\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \quad (4)$$

When tied observations are present, the variance is adjusted using the tie correction factor:

$$C = 1 - \frac{\sum_{m=1}^{g} (t_m^3 - t_m)}{N^3 - N} \quad (5)$$

Where:

g is the number of groups of tied ranks, t_m is the number of observations in the m-th tied group.

The variance then becomes:

$$Var = C \times \frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right) \quad (6)$$

and the corrected Dunn statistic is:

$$Z_{ij} = \frac{\bar{R}_i - \bar{R}_j}{\sqrt{\text{Var}}} \quad (7)$$

The resulting Z-statistics are compared with the standard normal distribution to obtain pairwise p-values. The hypotheses of the Dunn's post hoc comparison test are presented below as:

H_0: The distributions of NO_2 concentrations are the same at both locations.

Against.

H_1: The distributions of NO_2 concentrations differ between the two locations.

The Dunn's post hoc comparison test was executed using a significance level of $\alpha = 0.05$. According to the decision criterion, the null hypothesis (H_0) is rejected when the p-value < 0.05, signifying a significant difference in NO_2 concentrations. If the p-value ≥ 0.05, the null hypothesis is not rejected, suggesting similarity in NO_2 concentrations.

2.4. Exploratory data analysis

This study uses statistical methods and visualizations to explore NO_2 patterns, detect trends, and identify normal/abnormal readings. Descriptive statistics summarize the data's key characteristics, including mean, standard deviation, min–max values, and 25–75th percentiles. Several visualization approaches were implemented to facilitate interpretation. Mean bar charts were employed to compare average NO_2 levels across different locations or periods. Density plots revealed the distribution of NO_2 readings, indicating the frequency of different concentration ranges. Violin and box plots displayed the distribution and spread of readings, providing insight into the interquartile range, median values, and potential outliers. Line trend plots depicted temporal changes in NO_2 levels. Additionally, map visualizations provided geographical context for the study locations. When interpreted alongside descriptive statistics, these complementary visualizations enable the effective identification of NO_2-level trends and anomalies, informing subsequent inferential analysis.

2.5. Time series modeling

Time series modeling is a statistical method that analyzes chronologically ordered data to uncover data patterns for future predictions [43]. This study uses the ARIMA and Holt-Winters exponential smoothing (Holt-Winters) models to analyze and forecast Cheltenham's 2022 NO_2 levels using existing data.

2.5.1. Autoregressive Integrated Moving Average

The ARIMA model combines elements of the autoregressive (AR) and moving average (MA) models and was originally introduced by Arumugam and Natarajan [44].

The AR model (AR $\{p\}$) is represented as:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \varepsilon_t \quad (8)$$

The MA model (MA $\{q\}$) is represented as:

$$Y_t = \alpha + \varepsilon_t - \varnothing_1 \varepsilon_{t-1} + \varnothing_2 \varepsilon_{t-2} - \ldots - \varnothing_q \varepsilon_{t-q} \quad (9)$$

The ARIMA model is applied when a time series has been differenced one or more times and incorporates both AR and MA components. Therefore, the ARIMA model (p, d, q) is represented as:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \varepsilon_t - \varnothing_1 \varepsilon_{t-1} + \varnothing_2 \varepsilon_{t-2} - \ldots - \varnothing_q \varepsilon_{t-q} \quad (10)$$

Where from (2.8), (2.9), and (2.10):

Y_t is the monthly air quality (NO_2) data on which the model is to be applied, α is the intercept term, ϵ_t is the error term, $\phi_1, \phi_2, \ldots, \phi_q$ are the MA model coefficients, $\beta_1, \beta_2, \ldots, \beta_p$ are the AR model coefficients, and $\beta_1, \beta_2, \ldots, \beta_p$ and $\phi_1, \phi_2, \ldots, \phi_q$ are the ARIMA model coefficients.

The stationarity assumption was verified using the Augmented Dickey–Fuller (ADF) test before implementing the ARIMA model. The ADF test is mathematically represented as:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \ldots + \delta_p \Delta y_{t-p} + \varepsilon_t \quad (11)$$

Where:

Δy_t is the difference in the time series y at time t, α is the intercept, βt is the coefficient on a time trend, γ is the coefficient on y_{t-1}, δ_1, δ_2, ..., δ_p are the coefficients for the lagged differences of the series, and ε_t denotes the error term.

The hypotheses of the ADF test are presented below as:

H_0: Stationary.

Against.

H_1: Non-stationary.

Data stationarity was evaluated using the ADF test. A p-value higher than the significance level ($\alpha = 0.05$) suggests non-stationarity, while a lower p-value indicates stationarity. The ARIMA model was then configured with an order (p, d, q) and a seasonal order (p, d, q, M) to address the data's non-seasonal and seasonal aspects. The p, d, and q values in the order represent AR terms, differencing steps, and MA terms of the non-seasonal component, respectively. For the seasonal aspect, the order (p, d, q, M) introduces an additional seasonal period ($M = 12$), signifying monthly seasonality. In this framework, the parameter p is chosen by examining the partial autocorrelation function (PACF) plot, q is selected using the autocorrelation function (ACF) plot, and d specifies the number of differencing operations required to achieve stationarity.

2.5.2. Holt-Winters exponential smoothing

The Holt-Winters exponential smoothing model, introduced by Holt [45] and Winters [46], was used to analyze the time series of Cheltenham's NO_2 concentrations from 1994 to 2022. This model effectively encapsulates trends and seasonality in data series [47], making it an appropriate choice for our study. The model comprises three components: level, trend, and seasonality. The level component utilizes exponential smoothing to emphasize recent observations, while the trend and seasonality components employ multiplicative and additive methods. The Holt-Winters method involves four equations:

Level equation.

$$L_t = \alpha (Y_t / S_{t-L}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (12)$$

Trend equation.

$$T_t = \beta (L_t - L_{t-1}) + (1 - \beta) S_{t-L} \quad (13)$$

Seasonal equation.

$$S_t = \gamma (Y_t / L_t) + (1 - \gamma) S_{t-L} \quad (14)$$

Forecast equation.

$$F_{t+m} = (L_t + mT_t) S_{t-L+m} \quad (15)$$

Where from Equations (12)–(15):

Y_t is the actual value at time t, L_t is the level component at time t, T_t is the trend component at time t, S_t is the seasonal component at time t, F_{t+m} is the forecast for m periods ahead, L is the length of the seasonal cycle, and α, β, and γ are the smoothing parameters for the level, trend, and seasonal components, respectively. Their values are between 0 and 1.

The Holt-Winters model was set to a 12-month seasonal period to match our data's annual cycle. After configuring the model, it was fitted to the data, optimizing smoothing coefficients to minimize errors. The final model was then employed to decipher data patterns, forecast Cheltenham's future (2022) NO_2 values, and comprehend how trends and seasonality influence NO_2 levels.

2.6. Machine learning modeling

According to Méndez et al. [48], ML methods are widely used for air quality forecasting and air pollution predictions. This study applies ML techniques, including LSTM, GRU, and XGBoost, to predict air quality.

2.6.1. Long Short-Term Memory

This study employs LSTM recurrent neural networks (RNNs) to forecast NO_2 levels. LSTMs are well-suited for time series data, as they can retain information over long sequences and handle varying input lengths effectively [49, 50]. An LSTM unit, commonly used within RNN layers, consists of a cell state and three types of gates: input, output, and forget. These components work together to regulate information flow and maintain memory over time. The following equations describe the internal workings of the LSTM:

Forget Gate.

$$f_t = \sigma (W_f . [h_{t-1}, x_t] + b_f) \quad (16)$$

Input Gate.

$$i_t = \sigma (W_i . [h_{t-1}, x_t] + b_i) \quad (17)$$

$$\widetilde{C}_t = tanh (W_C [h_{t-1}, x_t] + b_C) \quad (18)$$

Cell State.

$$C_t = (f_t \times C_{t-1}) + (i_t \times \widetilde{C}_t) \quad (19)$$

Output Gate.

$$o_t = \sigma (W_o . [h_{t-1}, x_t] + b_o) \quad (20)$$

$$h_t = o_t \times tanh (C_t) \quad (21)$$

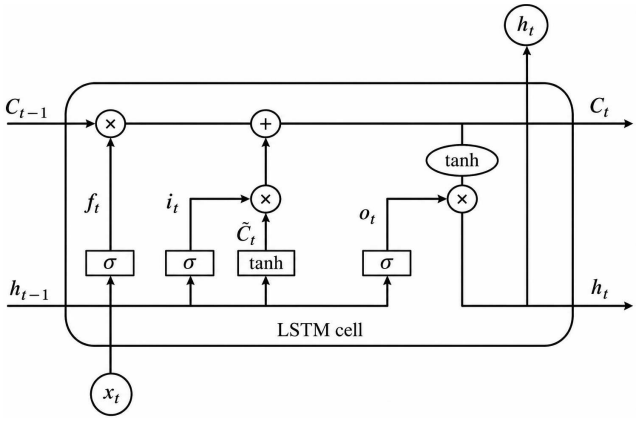
Where from Equations (16)–(21):

σ denotes the sigmoid function, W_f is the weight matrix for the forget gate, h_{t-1} is the output from the previous time step, x_t is the input at the current time step, b_f is the bias of the forget gate, W_i and W_C are the weight matrices for the input gate, b_i and b_C are the bias for the input gate, W_o is the weight matrix for the output gate, and b_o is the bias for the output gate.

These equations together define the LSTM cell as seen in Figure 2, which can be trained via gradient descent and backpropagation through time.

This research implemented an LSTM model using Python's TensorFlow and Keras libraries, with a fixed random seed to ensure reproducibility. The model architecture included an input

Figure 2
Structure of Long Short-Term Memory (LSTM) cell



layer that accepted sequences with five features, followed by an LSTM layer with 64 units. A dense layer with 8 neurons and ReLU activation introduced nonlinearity, while a final dense layer with a single neuron and linear activation produced continuous predictions of NO_2 concentrations. The model was optimized using the Adam algorithm with a learning rate of 0.01 to minimize the Mean Squared Error (MSE) between predicted and actual values. RMSE was selected as the primary evaluation metric due to its interpretability in the original measurement units. To mitigate overfitting, a checkpoint mechanism was employed to retain the best-performing model iteration during training. The model was trained for 100 epochs with a batch size of 32, using a train-validation split to monitor performance on unseen data. The final evaluation was conducted on a separate test dataset consisting of NO_2 measurements from Cheltenham.

2.6.2. Extreme Gradient Boosting

This research utilized the XGBoost model, renowned for efficiency, flexibility, and performance, in predictive modeling [51]. XGBoost's objective function mathematically represents its ability to fit training data and make predictions. It is presented as:

$$\boldsymbol{L(\varnothing) = \sum_{i} l(\hat{y}_i, y_i) + \sum_{k} \Omega(f_k)} \tag{22}$$

Where:

$L(\varnothing)$ denotes the objective function, $\sum_i l(\hat{y}_i, y_i)$ represents the loss component, and $\sum_k \Omega(f_k)$ corresponds to the regularization term.

Optimizing the loss function aims to improve predictive accuracy by minimizing errors, while optimizing the regularization term reduces variance and enhances prediction stability.

The specific model utilized in this research is the XGB Regressor, a part of the XGBoost library designed explicitly for regression tasks. This model has been configured with the following parameters optimized for this specific task:

1) Booster: "gbtree," meaning tree-based models are used as base learners.
2) Learning rate: 0.01, which shrinks the weights associated with features after each boosting round, helping to prevent overfitting.
3) Max depth: 3, which restricts the size of each tree and helps mitigate overfitting by limiting model complexity.
4) N estimators: 1000, representing the number of boosting stages to perform. Gradient boosting is robust to overfitting, so a large number of trees usually yields better performance.
5) Early stopping rounds: 100, which halts training if the validation score fails to improve over 100 consecutive iterations, helping to prevent overfitting.
6) Objective: "reg:linear," indicating that the task is a linear regression.

The XGBoost model was then fitted to the Cheltenham training NO_2 data, and the model's performance was evaluated. The power of XGBoost lies in its scalability in all scenarios. The system is not only fast but also highly efficient.

2.6.3. Gated Recurrent Unit

The GRU adapts the LSTM architecture, addressing some of its limitations [49]. It maintains the capacity to remember long-term dependencies while having a simpler architecture. The equations that govern a GRU are as follows:

Update Gate.

$$\boldsymbol{z_t = \sigma(W_z.[h_{t-1}, x_t] + b_z)} \tag{23}$$

Reset Gate.

$$\boldsymbol{r_t = \sigma(W_r.[h_{t-1}, x_t] + b_r)} \tag{24}$$

Current Memory Content.

$$\boldsymbol{\hat{h}_t = tanh(W.[r_t h_{t-1}, x_t] + b)} \tag{25}$$

Final Memory at Current Time Step.

$$\boldsymbol{h_t = (1 - z_t) h_{t-1} + (z_t \hat{h}_t)} \tag{26}$$

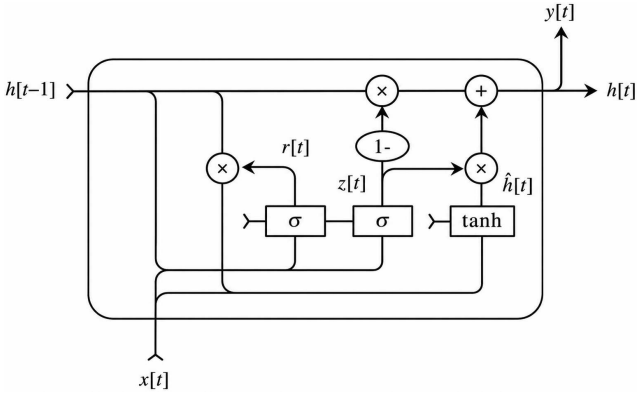
Where from Equations (18)–(21):

σ denotes the sigmoid function, W_z is the weight matrix for the update gate, h_{t-1} is the output from the previous time step, x_t is the input at the current time step, b_z is the bias of the update gate, W_r is the weight matrix for the reset gate, br is the bias for the reset gate, W_h is the weight matrix for the current memory content, and b_h is the bias for the current memory content.

These equations together define the GRU cell as seen in Figure 3, which can be trained via gradient descent and backpropagation through time.

The main distinction between GRUs and LSTMs is that a GRU uses two gates (reset and update), while an LSTM relies on three gates (input, forget, and output).

Figure 3
Architecture of the Gated Recurrent Unit (GRU) cell



This research employed the GRU model, implemented in TensorFlow, to analyze time series data. The model architecture commenced with an input layer specifying the input shape, followed by a 64-unit GRU layer designed to capture temporal dependencies. Two successive dense layers followed this GRU layer; the final dense layer featured a single neuron with a linear activation function for output prediction. The GRU layer contained 13,632 trainable parameters, while the subsequent dense layers contained 520 and 9 parameters, resulting in 14,161 trainable parameters. The Adam optimizer was used with a learning rate of 0.01 and the MSE loss function, both of which are suitable for regression tasks. Model performance was assessed using the RMSE metric.

The model was trained for 100 epochs. Cheltenham's NO_2 data was split into training and validation sets for this training. Model checkpoint callbacks were utilized during training to save the model weights that yielded the best performance on the validation set. After training concluded, these optimal weights were restored.

2.7. Model evaluation

This paper sought the best model for predicting Cheltenham's NO_2 levels, gauged by its RMSE, which measures predicted accuracy against actual data. It is mathematically represented as:

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{27}$$

where y_i are the actual values, $\hat{y}_i$ are the predicted values, and N is the total number of observations.

The evaluation commenced with established time series models, specifically ARIMA and Holt-Winters, chosen for their effectiveness in capturing and forecasting temporal patterns. These models were assessed using RMSE. Subsequently, ML models, namely, LSTM, XGBoost, and GRU, were evaluated, with their performance also measured by RMSE. The most accurate forecaster for Cheltenham's NO_2 levels was identified by comparing RMSE values across all evaluated models; lower RMSE values indicated superior predictive accuracy.

2.8. Data simulation using Bootstrapping

Bootstrapping in data simulation is a resampling technique used to estimate properties (such as the mean, variance, and confidence intervals) of a dataset by repeatedly sampling with replacement. It is especially valuable for small datasets when the true population distribution is uncertain or unspecified [40]. This study simulates three (3) samples of Cheltenham NO_2 data using Bootstrapping for different bootstrap replicates, and both the traditional time series models and the ML algorithms are used to model and predict these simulated NO_2 Cheltenham datasets. Based on their RMSE values, the overall best model for the Cheltenham air quality data is selected for validation. Bootstrapping for this study has the following steps:

1) Define the Original Dataset: Let the original dataset be $\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}\}$, with X being the NO_2 data for Cheltenham and n the number of observations.
2) Resampling with Replacement: Generate a bootstrap sample $\boldsymbol{X^*} = \{\boldsymbol{x_1^*}, \boldsymbol{x_2^*}, \ldots, \boldsymbol{x_n^*}\}$ by randomly selecting n observations from X with replacement. This means each x_i^* is independently drawn from X, and some observations may appear multiple times in X^*.
3) Repeat Resampling: Repeat the resampling process B times (B = 100, 250, 400) to generate B bootstrap samples:

$$\boldsymbol{X_1^*}, \boldsymbol{X_2^*}, \ldots, \boldsymbol{X_B^*}$$

4) Calculate Statistics for Each Bootstrap Sample: For each bootstrap sample $\boldsymbol{X_b^*}$, calculate the mean. The mean of the b_{th} bootstrap sample is:

$$\bar{\boldsymbol{x}}_{\boldsymbol{b}}^* = \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_{bi}^*} \tag{28}$$

This gives a distribution of bootstrap statistics:

$$\{\bar{\boldsymbol{x}}_{\boldsymbol{1}}^*, \bar{\boldsymbol{x}}_{\boldsymbol{2}}^*, \ldots, \bar{\boldsymbol{x}}_{\boldsymbol{B}}^*\}$$

5) Estimate the Population Statistic: Use the bootstrap distribution to approximate the population mean and quantify its uncertainty. The bootstrap-based estimate of the mean is:

$$\bar{\boldsymbol{x}}^* = \frac{\boldsymbol{1}}{\boldsymbol{B}} \sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \bar{\boldsymbol{x}}_{\boldsymbol{b}}^* \tag{29}$$

The standard error of the mean is:

$$\boldsymbol{SE^*} = \sqrt{\frac{\boldsymbol{1}}{\boldsymbol{B-1}} \sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \left(\bar{\boldsymbol{x}}_{\boldsymbol{b}}^* - \bar{\boldsymbol{x}}^*\right)^{\boldsymbol{2}}} \tag{30}$$

6) Confidence Intervals: Use the bootstrap distribution to build a confidence interval for the statistic. For a 95% confidence interval, take the 2.5th and 97.5th percentiles of the bootstrap distribution as the lower and upper bounds:

$$\boldsymbol{CI_{95\%}} = \left[\bar{\boldsymbol{x}}_{\boldsymbol{(0.025)}}^*, \bar{\boldsymbol{x}}_{\boldsymbol{(0.975)}}^*\right]$$

7) Simulate New Data: To simulate new NO_2 data, the study randomly draws observations from the bootstrap samples or the original dataset X with replacement.

To simulate B (B = 100, 250, 400) new observations, draw B values from X with replacement:

$$\boldsymbol{X_{sim}} = \{\boldsymbol{x_{sim,\,1}}, \boldsymbol{x_{sim,\,2}}, \ldots, \boldsymbol{x_{sim,\,B}}\} \tag{31}$$

where each is independently drawn from X.

The original Cheltenham NO_2 dataset consisted of 345 monthly observations collected between 1994 and 2022. In accordance with the classical bootstrap methodology, each bootstrap sample retained the same sample size (n = 345) as the original dataset and was generated by sampling with replacement.

3. Results and Discussion

3.1. Data collection, preprocessing, and feature engineering

Following data collection and preprocessing, the final NO_2 data for Cheltenham, London Road, and Boots Corner/High Street were obtained. To improve the model's performance, additional temporal features were added, including month, quarter, weekday, and day-of-year indicators.

3.2. Comparison test

Table 2 shows the results of NO_2 level comparisons between Cheltenham and two specific locations: London Road and Boots Corner/High Street. The Kruskal–Wallis H-test was used to analyze differences between the broader Cheltenham area and these sample locations.

Upon examining the results presented in Table 2, where the calculated p-value (0.74) exceeds the predetermined significance level of 0.05, the Kruskal–Wallis H-test fails to reject the null hypothesis (H_0), suggesting no detectable difference in NO_2 levels between Cheltenham and London Road.

Furthermore, this study investigates findings from a comparative analysis of Cheltenham as the population and Boots Corner/High Street as sample two, employing the Kruskal–Wallis H-test.

Table 2
Kruskal–Wallis: Cheltenham vs London Road

H-statistics	0.1113
p-value	0.7386

Based on the statistical findings presented in Table 3, where the calculated p-value (2.6983×10^{-28}) falls well below the predetermined significance level of 0.05, our analysis leads to the rejection of the null hypothesis. This indicates a statistically significant difference in NO_2 readings between Cheltenham and Boots Corner/High Street.

Table 3
Kruskal–Wallis: Cheltenham vs Boots Corner/High Street

H-statistics	121.6903
p-value	2.6983×10^{-28}

Following the Kruskal–Wallis analysis, Dunn's post hoc pairwise comparison test with Holm adjustment was conducted to determine whether NO_2 concentrations differed significantly between the London Road and Boots Corner/High Street monitoring locations. As presented in Table 4, the analysis yielded an adjusted p-value of 7.72×10^{-23}, which is substantially lower than the significance threshold of 0.05. Therefore, the null hypothesis of no difference between the two monitoring locations was rejected. This finding provides strong statistical evidence that the distribution of NO_2 concentrations at London Road differs significantly from that observed at Boots Corner/High Street.

Table 4
Dunn's post hoc test: London Road vs Boots Corner/High Street

p-value	7.72×10^{-23}

3.3. Exploratory data analysis

The EDA encompasses a comprehensive examination of descriptive statistics and a comparative assessment of NO_2 levels across distinct areas within Cheltenham, namely, the commercial zones of Boots Corner/High Street and the residential area of London Road. Additionally, it includes an analytical exploration for identifying normal and abnormal NO_2 levels at these specified locations, as visualized in Figure 4.

The descriptive statistics in Table 5 provide valuable insights into NO_2 levels across different areas of Cheltenham over three distinct periods: pre-COVID-19, COVID-19, and post-COVID-19. By comparing these statistics, trends, and similarities, potential cost-effective alternatives for future air quality studies in Cheltenham can be identified. Starting with Cheltenham's NO_2 levels, a decline in the mean from 32.70 μg/m^3 in the pre-COVID-19 period to 26.45 μg/m^3 during COVID-19 and finally to 24.10 μg/m^3 in the post-COVID-19 period can be observed. This decrease is mirrored in the samples from London Road and Boots Corner/High Street, where NO_2 levels also decreased over these periods, as indicated in Figure 5.

NO_2 levels in Cheltenham and London Road showed similar central tendencies, particularly in the pre-COVID period, where the mean concentrations were very close: 32.70 μg/m^3 for Cheltenham and 32.47 μg/m^3 for London Road. This supports the interpretation that London Road broadly reflects the average NO_2 level observed across Cheltenham. However, London Road exhibited greater dispersion during the pre-COVID period, with a standard deviation of 9.29 μg/m^3 compared with 7.14 μg/m^3 for Cheltenham. This indicates that although London Road may be representative in terms of mean concentration, it captures a wider range of short-term fluctuations than the wider Cheltenham aggregate. Therefore, London Road should be interpreted as a useful proxy for broad NO_2 trends and central tendency, rather than a complete substitute for citywide monitoring of variability or localized pollution episodes.

Standard deviations, which measure data dispersion, further qualify this interpretation. Cheltenham's standard deviation decreased from 7.14 μg/m^3 in the pre-COVID period to 4.86 μg/m^3 during COVID and 5.18 μg/m^3 in the post-COVID period. London Road and Boots Corner/High Street also showed a general reduction in variability across the observed periods, as visually represented in Figure 6. However, London Road's higher pre-COVID standard deviation indicates that its NO_2 readings were more dispersed than the Cheltenham aggregate before COVID-19. This suggests that London Road is more reliable as a proxy for overall trends and average NO_2 levels than for representing the full variability of Cheltenham's air quality.

These similarities suggest that variability in air quality levels tends to decrease over the observed periods in both Cheltenham and the samples.

Minimum and maximum values offer insights into the range of NO_2 levels. Cheltenham's range has a minimum of 13 μg/m^3 in the pre-COVID period and a maximum of 36.30 μg/m^3 in the post-COVID period. London Road's range is narrower, with a minimum value of 11.30 μg/m^3 and a maximum of 39.00 μg/m^3. In contrast, Boots Corner/High Street's range is much wider, with a minimum value of 5.10 μg/m^3 and a maximum of 42.20 μg/m^3, as illustrated in Figure 7.

These differences indicate that Boots Corner/High Street experiences more significant fluctuations in NO_2 levels, making it a potentially useful location for detailed studies of air quality variations.

Comparing NO_2 levels between Cheltenham's commercial area (Boots Corner/High Street) and residential area (London Road) revealed distinct trends across the COVID periods. Pre-COVID, Boots Corner consistently showed higher abnormal NO_2 concentrations than London Road, indicating commercial zones' greater susceptibility to air pollution from traffic and business activities. During COVID, Boots Corner experienced substantial

Figure 4
Map of Cheltenham showing the locations of interest

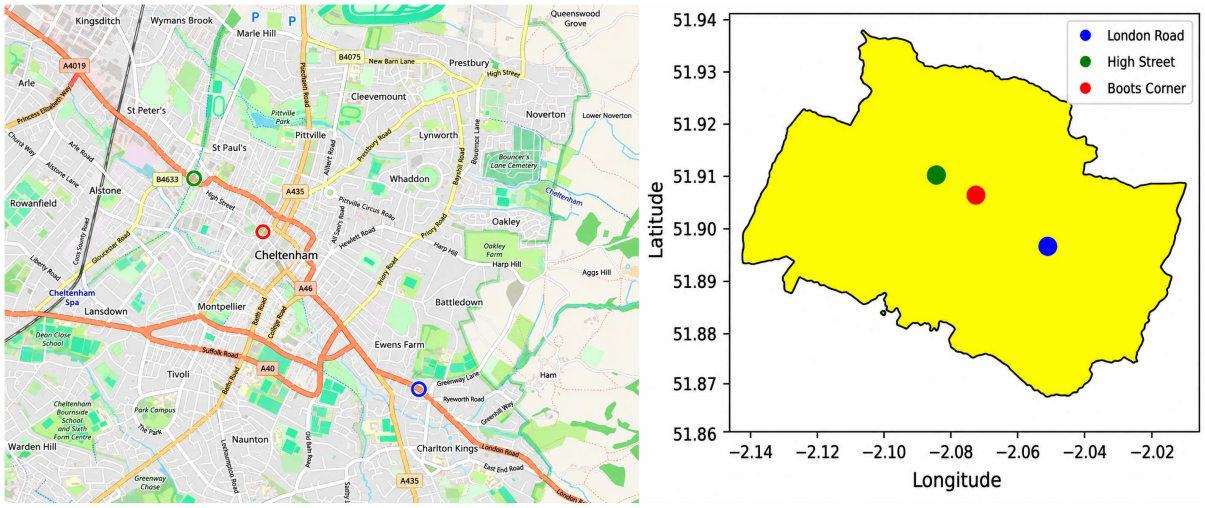


Table 5
Descriptive statistics of Cheltenham NO_2 for pre-COVID, COVID, and post-COVID period

Location	Period	Count	Mean	Std	25%	50%	75%	Min	Max
Cheltenham's	Pre-COVID	314	32.70	7.14	28.10	32.00	37.00	13.00	58.00
NO_2 ($\mu g/m^3$)	COVID	22	26.45	4.86	22.80	26.80	29.30	17.70	34.30
	Post-COVID	9	24.10	5.18	21.60	22.10	24.60	19.50	36.30
London Road's	Pre-COVID	314	32.47	9.29	26.13	33.10	38.53	11.30	63.20
NO_2 ($\mu g/m^3$)	COVID	22	30.17	4.34	27.10	31.05	33.70	20.80	36.40
	Post-COVID	9	27.59	4.71	24.20	26.40	28.90	23.50	39.00
Boots Corner /	Pre-COVID	314	40.89	9.90	35.03	39.65	46.40	5.10	97.90
High Street's	COVID	22	28.51	6.26	25.80	29.25	31.58	16.80	43.20
NO_2 ($\mu g/m^3$)	Post-COVID	9	28.57	5.78	25.20	26.40	28.40	23.60	42.20

Figure 5
Bar chart illustrating the mean values of NO_2 by locations and period

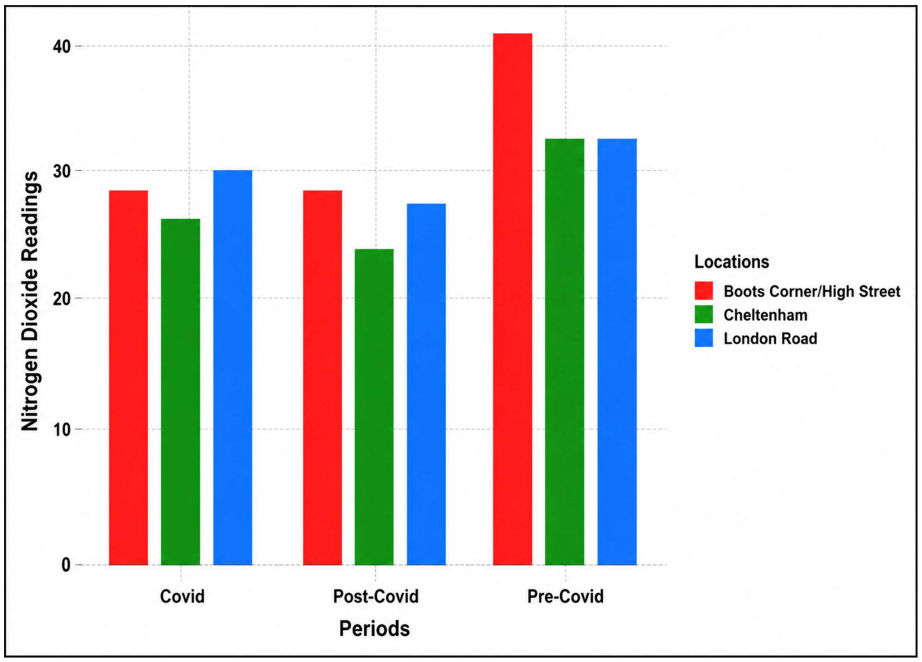


Figure 6
Bar chart showing the standard deviation of NO_2 by locations and period

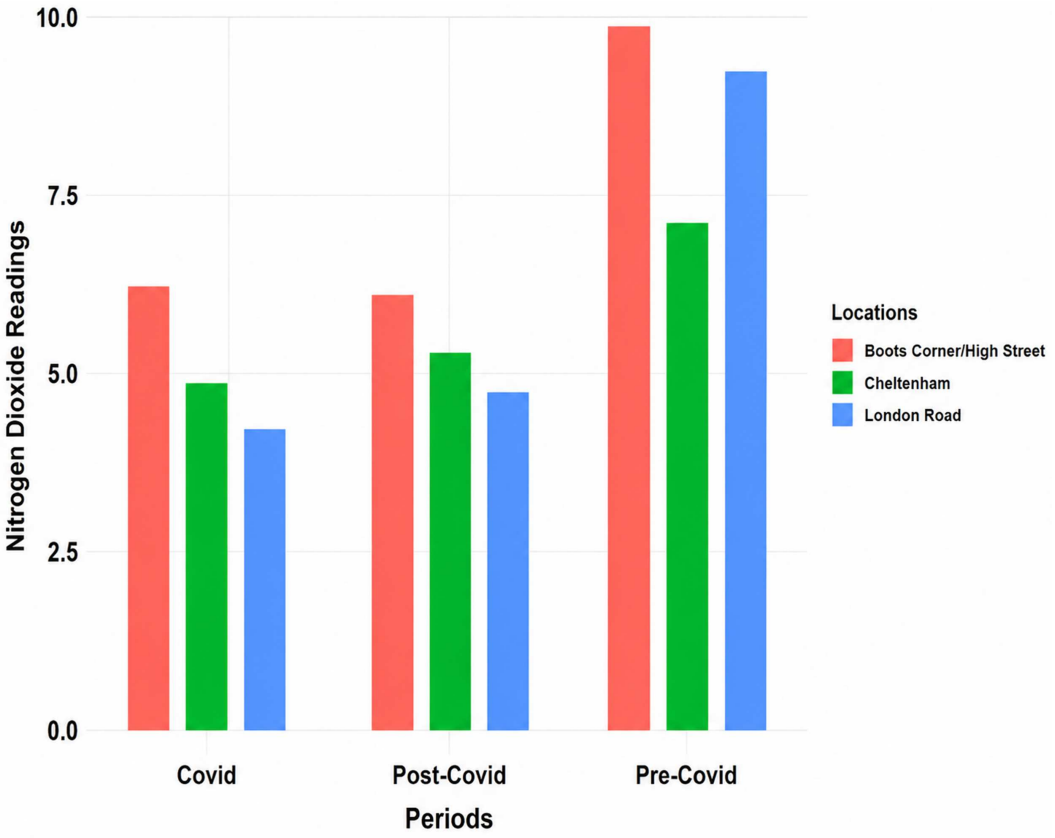
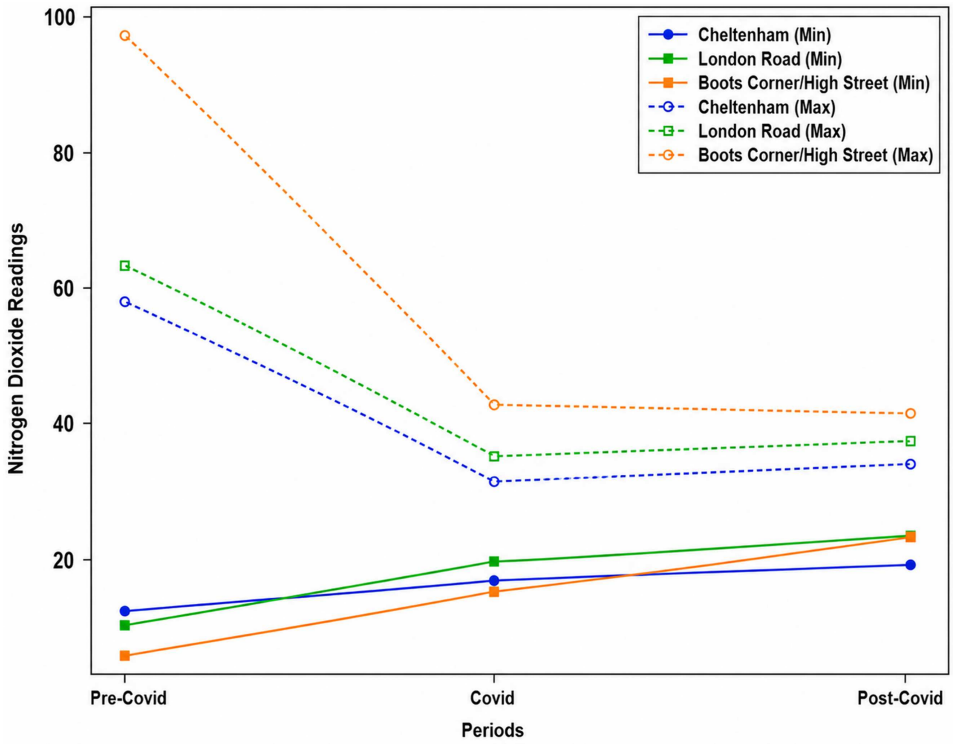


Figure 7
Line plot showing the minimum and maximum values of NO_2 by locations and period



NO_2 reductions approaching normal thresholds, while London Road showed slight decreases. Notably, Boots Corner's initial COVID month recorded above-normal levels (43.20 μg/m^3), suggesting possible noncompliance with early stay-at-home orders. Post-COVID, London Road maintained lower NO_2 levels than Boots Corner, which remained within the normal range except for the first month after restrictions lifted (42.20 μg/m^3).

An examination of the descriptive statistics indicated that Cheltenham's air quality trends closely matched those of London Road, suggesting it as a proxy for future studies. In contrast, the greater variability at Boots Corner provides useful data for analyzing temporal changes. These results emphasize the need for strategic site selection in air quality research, depending on study goals and resource availability.

3.4. Time series modeling results

This section presents the combined results of the ARIMA model and Holt-Winters exponential smoothing model for comparative analysis.

3.4.1. Autoregressive Integrated Moving Average

The analysis centered on the performance outcomes derived from the ARIMA model, constructed using NO_2 data from Cheltenham to forecast Cheltenham's NO_2 levels for 2022. The first step involved testing the ARIMA series for stationarity using the ADF test. Afterward, ACF and PACF plots were examined to determine suitable values for the p and q terms in the ARIMA model.

Since the p-value from Table 6 is 0.31375, which exceeds the predetermined significance level of 0.05, the null hypothesis is not rejected. Therefore, the Cheltenham NO_2 series is inferred to be non-stationary. In response to this observation, a different procedure is applied, followed by a subsequent stationarity test.

Table 6
ADF test result for Cheltenham NO_2 data

ADF test statistic	–1.93943
p-value	0.31375

Since the p-value in Table 7 is 9.65402×10^{-24}, which is below the predetermined significance threshold of 0.05, the null hypothesis is rejected. Therefore, the differenced Cheltenham NO_2 series is inferred to be stationary.

Table 7
ADF test result for Cheltenham NO_2 data after differencing

ADF test statistic	–8.93382
p-value	$9.65402e^{-24}$

Figure 8 facilitated the determination of the p and q parameters for the initial ARIMA model, providing essential guidance for model specification. Consequently, the corresponding p, d, and q values for the model were established as 1, 1, and 2.

The performance evaluation of the ARIMA model, as presented in Table 8, highlights the RMSE score for the NO_2 levels in Cheltenham. The obtained RMSE score is 3.24, indicating the average magnitude of the model's prediction errors, as seen in Figure 9. This metric is a crucial indicator of the model's

Figure 8
ACF/PACF combined plot for Cheltenham NO_2 data

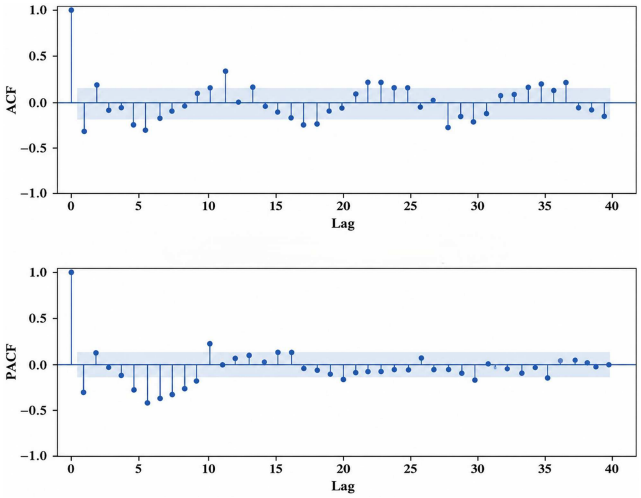
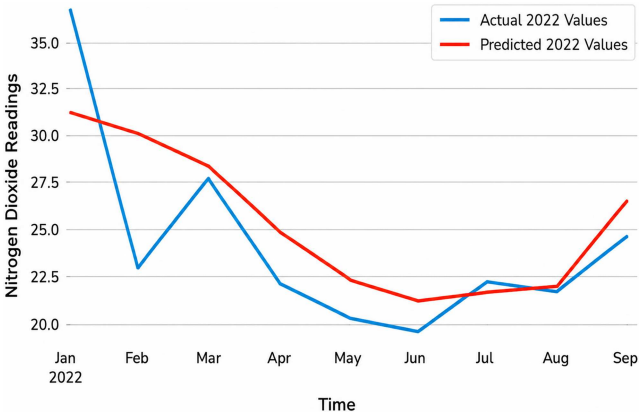


Table 8
ARIMA model performance

Model	RMSE score
ARIMA	3.24

Figure 9
Line plot showing Cheltenham's actual 2022 NO_2 data against Cheltenham's predicted 2022 NO_2 data using the ARIMA model



accuracy, with lower values indicating better predictive performance. In the study of NO_2 levels in Cheltenham, the RMSE of 3.24 indicates relatively favorable performance, given the observed data patterns.

3.4.2. Holt-Winters exponential smoothing

The application of the Holt-Winters model to analyze the NO_2 data of Cheltenham revealed three pivotal components: level, trend, and seasonality, as illustrated in Figure 10. Examination of Cheltenham's NO_2 data revealed noteworthy trends in 2010 and 2011. Additionally, discernible seasonal patterns were identified. These observations substantiated the adoption of both multiplicative and additive methods in formulating the Holt-Winters models. The recognition of such elements enhances the model's comprehensiveness, facilitating a more nuanced understanding of the underlying dynamics in Cheltenham's NO_2 data.

Figure 10
Level, trend, seasonality, and residual plot for Cheltenham NO_2 data

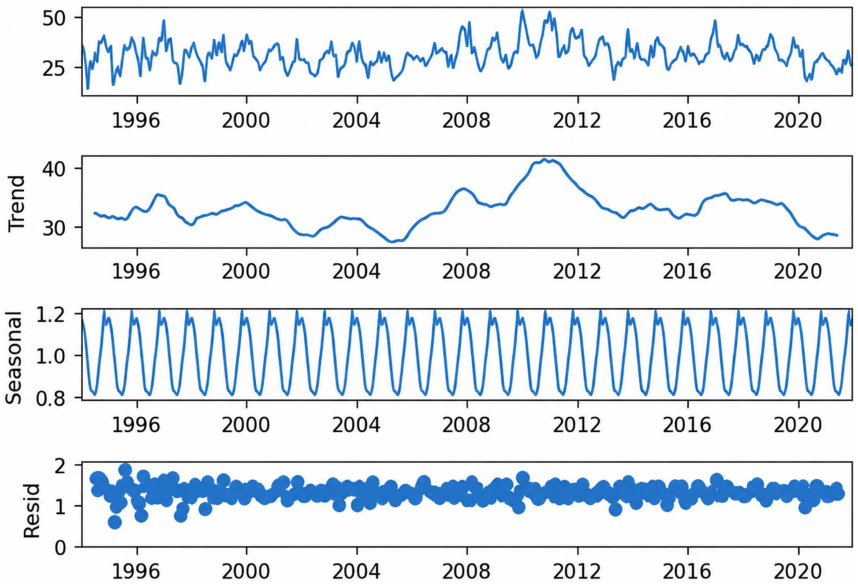
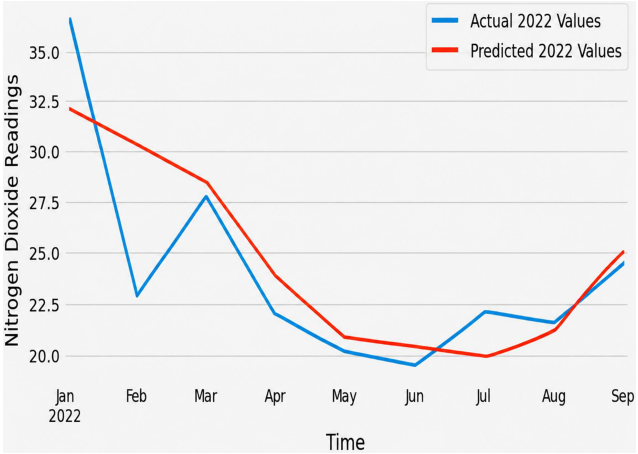


Table 9 reports the performance of the Holt-Winters model using RMSE for the Cheltenham NO_2 dataset. The obtained RMSE value of 3.08 provides a numerical summary of how well the model reproduces the observed NO_2 concentrations and indicates the average magnitude of the residuals (prediction errors) between the observed and predicted values. A lower RMSE score generally indicates a better fit of the model to the data, suggesting that the Holt-Winters model, in this instance, demonstrates a relatively accurate predictive capability for Cheltenham's NO_2 levels, as presented in Figure 11.

Table 9
Holt-Winters model performance

Model	RMSE score
Holt-Winters	3.08

Figure 11
Line plot showing Cheltenham's actual 2022 NO_2 data against Cheltenham's predicted 2022 NO_2 data using the Holt-Winters model



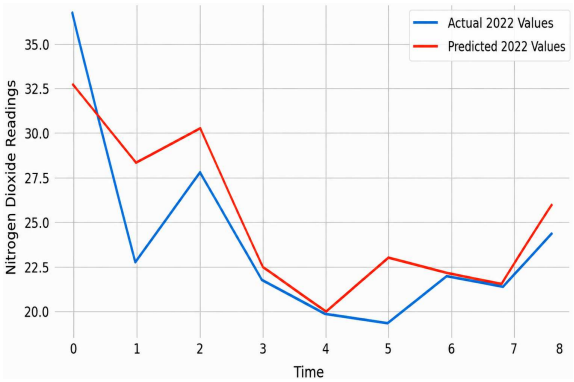
3.5. Machine learning modeling results

The results outlined below are the performance of the ML algorithms. The ML algorithms are LSTM, XGBoost, and GRU.

3.5.1. Long Short-Term Memory

The performance evaluation results for the LSTM model, as illustrated in Figure 12, showed an RMSE of 2.66. The RMSE score signifies commendable performance by the LSTM model, indicating its predictive accuracy.

Figure 12
Line plot showing Cheltenham's actual 2022 NO_2 data against Cheltenham's predicted 2022 NO_2 data using the LSTM model



3.5.2. Extreme Gradient Boosting

The performance evaluation results for the XGBoost model, as illustrated in Figure 13, showed an RMSE of 3.18. The RMSE score signifies commendable performance by the XGBoost model, indicating its predictive accuracy.

3.5.3. Gated Recurrent Unit

The performance evaluation of the GRU model showed an RMSE score of 3.19. The graphical representation of this performance is depicted in Figure 14.

Figure 13
Line plot showing Cheltenham's actual 2022 NO_2 data against Cheltenham's predicted 2022 NO_2 data using the XGBoost model

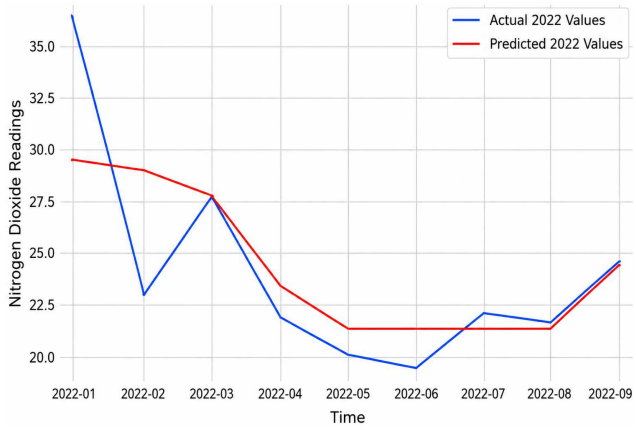
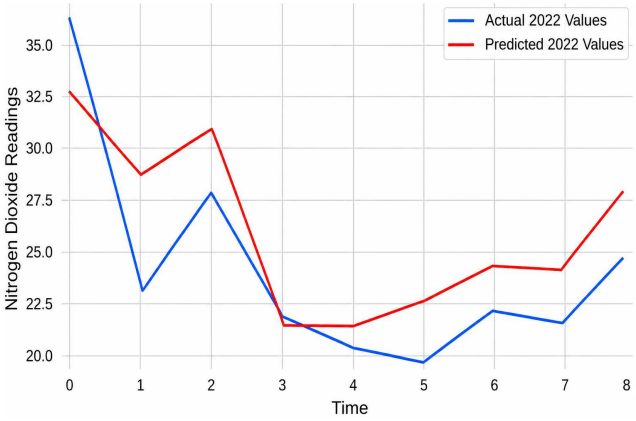


Figure 14
Line plot showing Cheltenham's actual 2022 NO_2 data against Cheltenham's predicted 2022 NO_2 data using the GRU model



3.6. Model evaluation

The examination of the outcomes stemming from the evaluation of both conventional time series models and ML algorithms is outlined below.

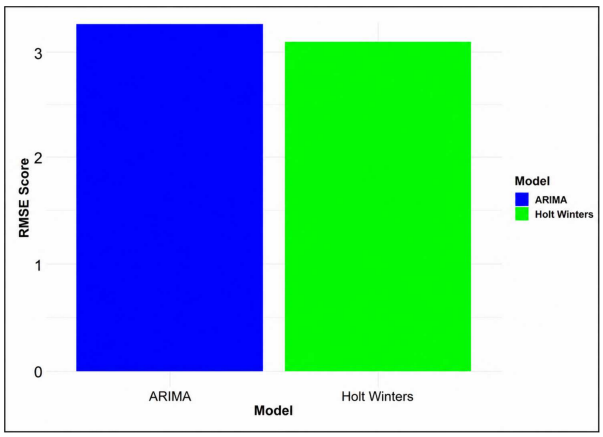
3.6.1. Time series model evaluation

The outcomes from the model evaluation, as presented in Table 10, shed light on the performance of two prominent time series forecasting models: ARIMA and Holt-Winters exponential smoothing. The evaluation metric employed in this analysis is the RMSE. The ARIMA model yielded an RMSE of 3.24, indicating the average magnitude of forecasting errors when applied to the time series data under consideration. In contrast, the Holt-Winters model demonstrated only a marginal improvement, with an RMSE of 3.08, as shown in Figure 15. The difference in RMSE scores between these models indicates that, for this specific evaluation metric, the Holt-Winters model outperformed the ARIMA model in forecasting accuracy.

Table 10
Model evaluation for time series models

Model	RMSE score
ARIMA	3.24
Holt-Winters	3.08

Figure 15
A bar chart showing the RMSE scores for time series models



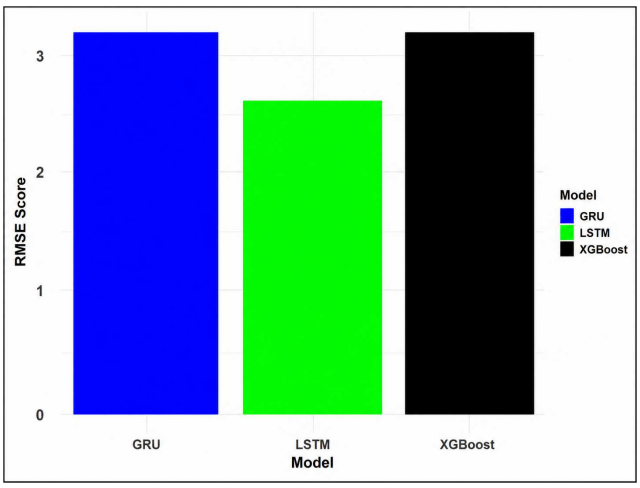
3.6.2. Machine learning model evaluation

Table 11 presents the RMSE scores for three ML models: LSTM, XGBoost, and GRU. Among the evaluated models, the LSTM achieved the lowest RMSE score of 2.66, demonstrating superior performance in minimizing prediction error. This result aligns with the well-documented effectiveness of LSTM networks in handling sequential data by capturing long-term dependencies. The LSTM's ability to retain relevant information across extended sequences likely contributed to its lower error rate compared to the other models. The XGBoost and GRU models yielded similar RMSE scores of 3.18 and 3.19, respectively. XGBoost, a tree-based ensemble learning method, has been widely recognized for its robustness and efficiency in structured data tasks. However, its slightly higher RMSE compared to LSTM suggests that it may not be as effective in modeling temporal dependencies within the dataset. On the other hand, GRU, a variant of LSTM with a simplified gating mechanism, performed slightly worse than LSTM, possibly due to its simpler architecture, which may have impacted its ability to capture intricate temporal patterns. The results indicate that LSTM outperformed the other models in this evaluation, reinforcing its suitability for predictive modeling in time series tasks, as presented in Figure 16. The performance gap between LSTM and GRU suggests that while both recurrent architectures are effective, LSTM's enhanced memory capabilities provide an advantage in complex sequences. Additionally, the higher RMSE of XGBoost highlights the challenges of applying non-recurrent and convolutional architectures to sequential data, particularly when long-term dependencies are crucial.

Table 11
Model evaluation for ML models

Model	RMSE score
LSTM	2.66
XGBoost	3.18
GRU	3.19

Figure 16
A bar chart showing the RMSE scores for ML models



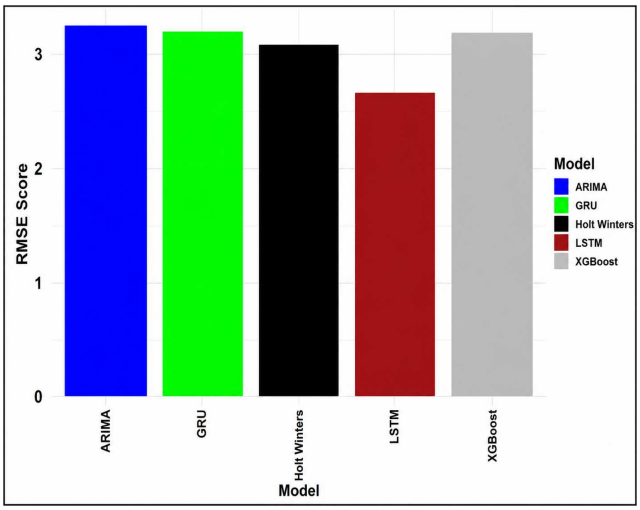
3.6.3. Time series and ML model evaluation

The differences in RMSE scores among these models, as illustrated in Figures 15 and 16, highlight varying degrees of prediction accuracy. The ML model, particularly LSTM, outperforms traditional time series models, underscoring the potential advantages of leveraging more advanced techniques for air quality prediction. These findings emphasize that forecasting models must be chosen to match the unique properties of time series data and the specific aims of the prediction problem. Conventional approaches such as ARIMA and Holt-Winters can serve as reference models. Still, the LSTM, as an ML method, delivers superior predictive performance, as reflected in its lower RMSE (see Figure 17).

3.7. Data simulation using Bootstrapping

Bootstrapping is used to simulate 3 NO_2 Cheltenham datasets for different bootstrap sample sizes (BS = 100, 250, and 400), as presented in Figure 18.

Figure 17
A bar chart showing the RMSE scores for time series and ML models



Both traditional time series models and advanced ML models were used to fit the simulated datasets, and RMSE was calculated for model evaluation.

Table 12 presents the RMSE scores for five predictive models, ARIMA, Holt-Winters, LSTM, XGBoost, and GRU, applied to three different NO_2 Cheltenham datasets.

For the first simulated dataset ($B = 100$), GRU and XGBoost performed the worst, both yielding RMSEs above 1.00 (1.10 and 1.09, respectively). ARIMA, Holt-Winters, and LSTM performed similarly, with RMSE scores of 1.01, 0.99, and 0.96, respectively. These results indicate that while traditional time series models (ARIMA and Holt-Winters) performed well, the RNNs (LSTM and GRU) offered reasonable accuracy.

For the second simulated dataset ($B = 250$), LSTM achieved the lowest RMSE of 0.39, closely followed by Holt-Winters at 0.40. ARIMA, GRU, and XGBoost performed slightly worse but remained within a narrow RMSE range (0.40–0.43). This relatively uniform distribution of RMSE scores suggests that the dataset exhibits a structure that is well captured by all models, with LSTM demonstrating a marginal advantage.

For the third simulated dataset ($B = 400$), the RMSE scores were relatively close across models, with Holt-Winters, LSTM, and XGBoost all achieving the lowest score of 0.46. ARIMA and GRU followed with slightly higher RMSE values of 0.49 and 0.48, respectively. The convergence of RMSE values at $B = 400$ suggests that the advantage of deep learning becomes less pronounced when the simulated dataset is larger and more stable. In smaller bootstrap samples, LSTM can benefit from its ability to model complex temporal dependencies and nonlinear patterns. However, as the bootstrap sample size increases, random fluctuations are averaged out, and the simulated series becomes smoother, with clearer trend and seasonal structure. Under these conditions, Holt-Winters can perform comparably because it is specifically designed to capture level, trend, and seasonality in time series data. Therefore, the reduced performance gap at $B = 400$ is likely related not simply to increased data volume but to the lower relative noise and more regular structure produced by larger bootstrap resampling. This finding indicates that deep learning models may offer the greatest advantage when the data contain stronger nonlinearities or irregular temporal dependencies, whereas traditional time series models remain competitive when the series is smoother and dominated by trend-seasonal components.

The results indicate that no single model consistently outperforms the others across all datasets, though LSTM and Holt-Winters demonstrated strong predictive capabilities in multiple cases. Overall, LSTM showed the strongest performance across the bootstrap simulations, particularly at $B = 100$ and $B = 250$ as presented in Figure 19. However, at $B = 400$, its advantage diminished, with Holt-Winters and XGBoost achieving the same RMSE. This indicates that model performance depends on the structure and variability of the simulated dataset and that traditional time series models can remain competitive when the data exhibit stable trend-seasonal behavior.

3.8. Comparison with existing studies

The findings of this study are broadly consistent with, and extend, prior work on air quality prediction, while also revealing important methodological and dataset-level differences that qualify any direct numerical comparison.

Li et al. [28] developed an LSTM-extended architecture for hourly PM2.5 forecasting across 12 monitoring stations in Beijing (2014–2016), explicitly incorporating spatial correlations between

Figure 18
Histogram plot for both NO_2 Cheltenham data and bootstrapped simulated NO_2 datasets

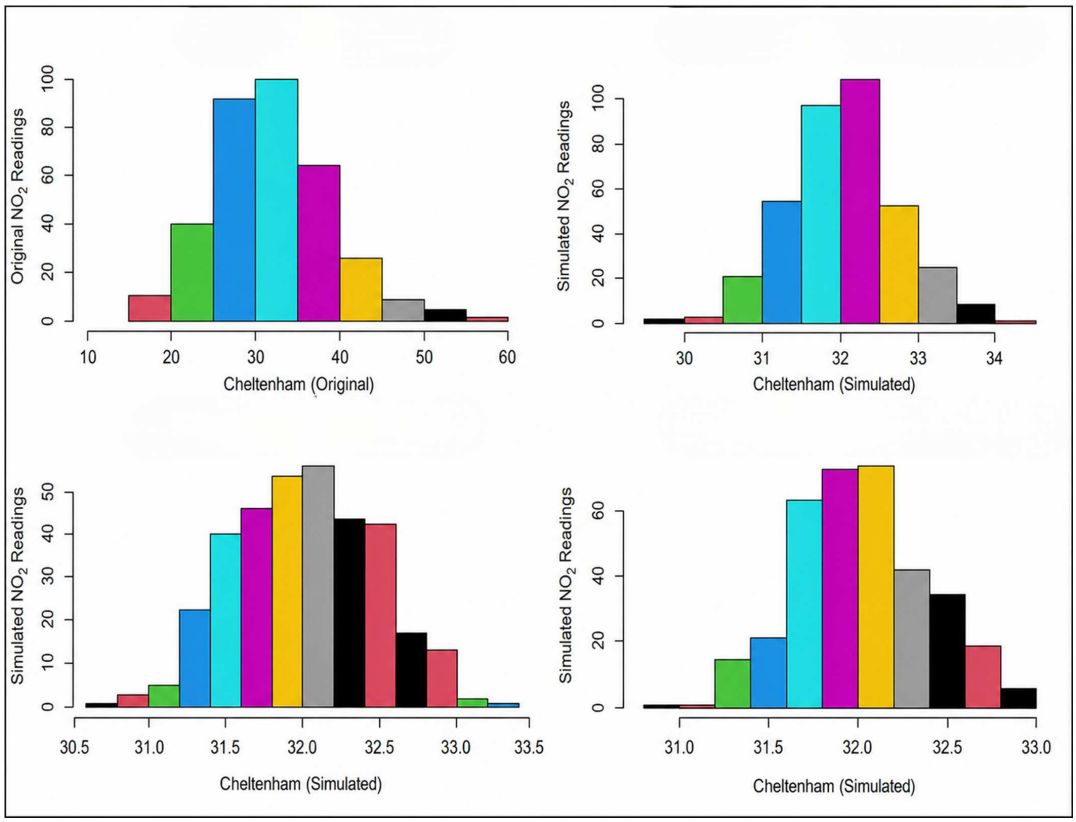
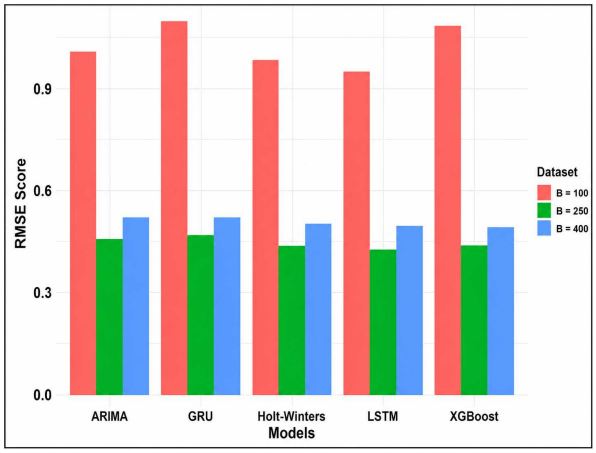


Table 12
Model evaluation for Cheltenham NO_2 simulated datasets

RMSE score					
	Models				
B	ARIMA	Holt-Winters	LSTM	GRU	XGBoost
100	1.01	0.99	0.96	1.10	1.09
250	0.42	0.40	0.39	0.43	0.40
400	0.49	0.46	0.46	0.48	0.46

Figure 19
Bar chart of RMSE values for the various models applied to three simulated NO_2 datasets from Cheltenham



stations alongside temporal dependencies, and benchmarked it against Auto Regressive Moving Average, Support Vector Regression, Time Delay Neural Network, and Spatiotemporal Deep Learning models. Their Long Short Term Memory Extended (LSTME) model outperformed all baselines and reported an Mean Absolute Percentage Error (MAPE) of approximately 31% for longer (13–24 h) forecast horizons, illustrating the general difficulty of extending LSTM accuracy beyond short horizons. This aligns directionally with the present study's finding that LSTM (RMSE = 2.66) outperformed ARIMA (RMSE = 3.24), Holt-Winters (RMSE = 3.08), XGBoost (RMSE = 3.18), and GRU (RMSE = 3.19) by an average of approximately 16%. However, a direct quantitative comparison of improvement magnitude is not possible: Li et al. report MAPE rather than RMSE, forecast PM2.5 rather than NO_2, use hourly rather than the present study's temporal resolution, and incorporate a spatial (multi-station) architecture rather than the single-site univariate framework used here. The comparison is therefore best read as convergent evidence for LSTM's advantage in modeling temporal dependencies, rather than a like-for-like effect-size comparison.

Marinov et al. [32] applied ARIMA to forecast CO, NO_2, O_3 and PM2.5 at five Sofia monitoring stations (2015–2019) across multiple temporal granularities (1–24 h), reporting mean absolute errors (MAE) ranging from 1.52 to 5.12 at 1-h granularity and from 2.53 to 14.12 at 3-h granularity. While MAE and RMSE are not interchangeable, RMSE penalizes larger errors more heavily; the lower end of their reported error range is broadly comparable in magnitude to the ARIMA RMSE of 3.24 obtained in this study, suggesting similar baseline ARIMA performance despite the difference in city, pollutant mix, and granularity. Marinov et al. did not test any ML alternative to ARIMA, so their study offers no basis for comparing traditional and deep learning approaches directly; the present study's finding that LSTM improves on ARIMA by approximately 18% therefore represents a novel contribution relative to their single-model framework.

Chen et al. [41] took a markedly different approach, using random forest models to predict NO_2, PM2.5, PM10 and O_3 across all of Great Britain at 1 km spatial resolution and daily temporal resolution (2006–2013), integrating multiple predictor sources including satellite and meteorological data. Their out-of-bag NO_2 RMSE of 9.71 $\mu g/m^3$ at daily resolution is substantially higher than the RMSE values obtained in the present study (2.66–3.24), but this is attributable to fundamental differences in modeling objective rather than model quality: Chen et al. predict spatially interpolated concentrations across an entire country from sparse monitoring stations, a task with inherently greater spatial uncertainty, whereas this study forecasts future values at fixed, densely observed local sites. Chen et al. also report that accuracy improves substantially at monthly and annual aggregation, consistent with the general pattern that averaging over time reduces RMSE, a pattern also visible in this study's bootstrap results, where higher replicate counts (B = 400) reduced RMSE to as low as 0.39–0.46. Beyond the scale difference, Chen et al.'s use of random forest with rich exogenous predictors (traffic, land use, satellite-derived variables) contrasts with the present study's more constrained set of forecasting models trained on the pollutant series itself; this suggests a natural direction for future work, namely, incorporating comparable meteorological and traffic covariates to test whether the LSTM advantage persists once exogenous predictors are added.

Taken together, these comparisons indicate that this study's methodological contribution is not simply a marginal accuracy gain over any single prior model, but an integration that existing studies lack; none of Li et al., Marinov et al., or Chen et al. combines spatial representativeness testing, multi-model temporal forecasting, and bootstrap-based robustness validation within a single framework. The practical outcome of identifying London Road as a viable low-cost proxy for citywide NO_2 trends has no direct precedent in the reviewed literature, none of which tested site-level representativeness as a monitoring strategy.

4. Conclusion

This research analyzed NO_2 levels in Cheltenham across commercial (Boots Corner/High Street) and residential (London Road) sites, combining Kruskal–Wallis H-tests with traditional (ARIMA, Holt-Winters) and ML (LSTM, XGBoost, GRU) forecasting models, and validated the results through bootstrap simulation.

The Kruskal–Wallis H-test detected no statistically significant difference between citywide NO_2 levels and those at London Road ($p = 0.74$), whereas Boots Corner/High Street differed significantly from citywide levels ($p < 0.05$). Furthermore, Dunn's post hoc pairwise comparison confirmed that NO_2 concentrations differed significantly between London Road and Boots Corner/High Street ($p < 0.001$). Together, these findings indicate that London Road is broadly representative of Cheltenham's overall NO_2 profile, while the commercial site captures a distinct pollution regime. All locations exhibited decreasing NO_2 concentrations across pre-COVID-19, COVID-19, and post-COVID-19 periods, with commercial areas showing higher concentrations and greater fluctuations than residential zones. Commercial districts recorded more abnormal readings pre-COVID-19, approaching normal thresholds during lockdowns. These findings indicate that London Road can serve as a cost-effective proxy for monitoring broad citywide NO_2 trends and average concentration levels, while Boots Corner/High Street provides valuable insights into commercial pollution dynamics and extreme variations. However, the higher pre-COVID standard deviation at London Road compared with Cheltenham indicates that London Road captures greater short-term variability than the citywide aggregate. Therefore, its representativeness should be interpreted primarily in terms of central tendency and temporal trend, rather than as a complete substitute for monitoring variability, exceedance events, or localized pollution episodes. In predicting Cheltenham's 2022 NO_2 levels, LSTM demonstrated superior performance (RMSE: 2.66) compared to other approaches. Among traditional time series methods, Holt-Winters (RMSE: 3.08) slightly outperformed ARIMA (RMSE: 3.24), likely due to its better capture of seasonal patterns. XGBoost and GRU yielded similar results (RMSE: 3.18 and 3.19, respectively). LSTM's effectiveness stems from its specialized architecture for modeling long-term dependencies in time series data, making it particularly suitable for environmental pollution forecasting. The clear performance hierarchy reveals the advantages of deep learning architectures over conventional techniques for capturing complex temporal patterns in environmental data. This performance gap underscores LSTM's value for air quality prediction applications. Bootstrap simulation analysis across three datasets (replicates of 100, 250, and 400) revealed varying model performance. LSTM excelled with B = 100 (RMSE = 0.96) and B = 250 (RMSE = 0.39), while for B = 400, Holt-Winters, LSTM, and XGBoost shared optimal performance (RMSE = 0.46). No single model dominated all simulations, indicating that model choice should be guided by dataset characteristics rather than a fixed default. Taken together, the findings of this study carry three practical implications for urban air quality management. First, residential monitoring at London Road can support lower-cost citywide NO_2 surveillance by approximating Cheltenham's average NO_2 profile and broad temporal decline. Nevertheless, because London Road showed higher pre-COVID variability than the citywide aggregate, supplementary monitoring remains important where the objective is to capture short-term fluctuations, exceedance events, or micro-scale spatial differences. Second, where high-accuracy forecasting is required, for example, in early-warning systems or health-impact modeling, LSTM should be the preferred choice, given its consistent approximately 16% average improvement in RMSE over competing methods. Third, because no single algorithm dominated across all replicate sizes, bootstrap-based validation should accompany model selection in similar urban monitoring studies. Compared with existing studies that typically focus on either statistical analysis or ML in isolation [28, 32, 41], this study provides an integrated framework that combines spatial comparison, temporal modeling, and simulation-based validation, thereby improving both predictive accuracy and robustness. Future work should extend this framework in three directions: (i) incorporating additional pollutants

(e.g., $PM_{2.5}$, O_3, NO_x) and meteorological covariates such as wind speed, humidity, temperature, and traffic flow covariates to improve predictive accuracy; (ii) testing the transferability of the London Road-as-proxy finding in other UK and international cities with comparable urban morphologies; (iii) hyperparameter tuning such as grid search or random search should be employed for model prediction improvement or optimization; and (iv) exploring hybrid architectures (e.g., LSTM-Holt-Winters ensembles, attention-augmented LSTMs, or Transformer-based models) to determine whether the residual error in LSTM forecasts can be further reduced. Such extensions would strengthen the evidence base for streamlined ML-assisted air quality monitoring as a scalable strategy for sustainable urban planning, public health protection, and climate-resilient city design.

Acknowledgment

The authors wish to express their heartfelt thanks to their advisor for consistent and invaluable guidance during this study. They also acknowledge their colleagues' contributions to their collaboration. Finally, the authors are profoundly grateful to their friends and families, whose constant encouragement was essential to bringing this work to completion.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available from the Cheltenham Borough Council at https://doi.org/10.5281/zenodo.21478965.

Author Contribution Statement

Johnson Ohakwe: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Derrick Ofori:** Methodology, Data curation, Writing – review & editing. **Bhupesh Kumar Mishra:** Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **William Sayers:** Conceptualization, Validation, Resources, Writing – review & editing, Supervision, Project administration. **Timothy Olusakin:** Formal analysis, Methodology, Writing – original draft. **John Chisimkwuo:** Validation, Formal analysis, Supervision.

References

[1] Beloconi, A., & Vounatsou, P. (2021). Substantial reduction in particulate matter air pollution across Europe during 2006-2019: A spatiotemporal modeling analysis. *Environmental Science & Technology*, *55*(22), 15505–15518. https://doi.org/10.1021/acs.est.1c03748

[2] Chen, Z.-Y., Petetin, H., Méndez Turrubiates, R. F., Achebak, H., García-Pando, Pérez, C, & Ballester, J. (2024). Population exposure to multiple air pollutants and its compound episodes in Europe. *Nature Communications*, *15*(1), 2094. https://doi.org/10.1038/s41467-024-46103-3

[3] Wilson, S., Farren, N. J., Rose, R. A., Wilde, S. E., Davison, J., Wareham, J. V., . . . , & Carslaw, D. C. (2023). The impact on passenger car emissions associated with the promotion and demise of diesel fuel. *Environment International*, *182*, 108330. https://doi.org/10.1016/j.envint.2023.108330

[4] Benavides, J., Guevara, M., Snyder, M. G., Rodríguez-Rey, D., Soret, A., Pérez García-Pando, C., & Jorba, O. (2021). On the impact of excess diesel NOX emissions upon NO_2 pollution in a compact city. *Environmental Research Letters*, *16*(2), 024024. https://doi.org/10.1088/1748-9326/abd5dd

[5] Font, A., Hedges, M., Han, Y., Lim, S., Bos, B., Tremper, A. H., & Green, D. C. (2024). Air quality on UK diesel and hybrid trains. *Environment International*, *187*, 108682. https://doi.org/10.1016/j.envint.2024.108682

[6] Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., . . . , & Yan, C. (2022). Pollution and health: A progress update. *The Lancet Planetary Health*, *6*(6), e535–e547. https://doi.org/10.1016/S2542-5196(22)00090-0

[7] McCarron, A., Semple, S., Braban, C. F., Swanson, V., Gillespie, C., & Price, H. D. (2023). Public engagement with air quality data: Using health behaviour change theory to support exposure-minimising behaviours. *Journal of Exposure Science & Environmental Epidemiology*, *33*(3), 321–331. https://doi.org/10.1038/s41370-022-00449-2

[8] Larkin, A., Anenberg, S., Goldberg, D. L., Mohegh, A., Brauer, M., & Hystad, P. (2023). A global spatial-temporal land use regression model for nitrogen dioxide air pollution. *Frontiers in Environmental Science*, *11*, 1125979. https://doi.org/10.3389/fenvs.2023.1125979

[9] Macintyre, H. L., Mitsakou, C., Vieno, M., Heal, M. R., Heaviside, C., & Exley, K. S. (2023). Impacts of emissions policies on future UK mortality burdens associated with air pollution. *Environment International*, *174*, 107862. https://doi.org/10.1016/j.envint.2023.107862

[10] Zhao, C., Lin, Z., Yang, L., Jiang, M., Qiu, Z., Wang, S., . . . , & Chen, Z. (2025). A study on the impact of meteorological and emission factors on $PM_{2.5}$ concentrations based on machine learning. *Journal of Environmental Management*, *376*, 124347. https://doi.org/10.1016/j.jenvman.2025.124347

[11] Yan, Y., Ren, P., & Meng, Q. (2024). Quantitative evaluation of the synergistic effects of multiple meteorological parameters on air pollutants based on generalized additive models. *Urban Climate*, *55*, 101965. https://doi.org/10.1016/j.uclim.2024.101965

[12] Koçak, E. (2025). Comprehensive evaluation of machine learning models for real-world air quality prediction and health risk assessment by $AirQ^+$. *Earth Science Informatics*, *18*(3), 447. https://doi.org/10.1007/s12145-025-01941-7

[13] Enciso-Díaz, W. C., Zafra-Mejía, C. A., & Hernández-Peña, Y. T. (2025). Global trends in air pollution modeling over cities under the influence of climate variability: A review. *Environments*, *12*(6), 177. https://doi.org/10.3390/environments12060177

[14] Kozhevnikova, M. F., & Levenets, V. V. (2023). Modeling the distribution of radionuclides in the air and on the soil surface.

East European Journal of Physics, (2), 191–200. https://doi.org/10.26565/2312-4334-2023-2-20

[15] Kasdagli, M.-I., Orellano, P., Pérez Velasco, R., & Samoli, E. (2024). Long-term exposure to nitrogen dioxide and ozone and mortality: Update of the WHO Air Quality Guidelines systematic review and meta-analysis. *International Journal of Public Health*, *69*, 1607676. https://doi.org/10.3389/ijph.2024.1607676

[16] Chen, X., Qi, L., Li, S., & Duan, X. (2024). Long-term NO_2 exposure and mortality: A comprehensive meta-analysis. *Environmental Pollution*, *341*, 122971. https://doi.org/10.1016/j.envpol.2023.122971

[17] Yang, S., Li, M., Guo, C., Requia, W. J., Zare Sakhvidi, M. J., Lin, K., . . . , & Yang, J. (2025). Associations of long-term exposure to nitrogen oxides with all-cause and cause-specific mortality. *Nature Communications*, *16*(1), 1730. https://doi.org/10.1038/s41467-025-56963-y

[18] Stafoggia, M., Oftedal, B., Chen, J., Rodopoulou, S., Renzi, M., Atkinson, R. W., . . . , & Janssen, N. A. H. (2022). Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: Results from seven large European cohorts within the ELAPSE project. *The Lancet Planetary Health*, *6*(1), e9–e18. https://doi.org/10.1016/S2542-5196(21)00277-1

[19] Yun, H., Ahn, S., Oh, J., Kang, C., Kim, A., Kwon, D., . . . , & Lee, W. (2024). Short-term exposure to outdoor nitrogen dioxide and respiratory mortality, with high-risk populations: A nationwide time-stratified case-crossover study. *BMC Public Health*, *24*(1), 3484. https://doi.org/10.1186/s12889-024-21048-w

[20] Schwarz, M., Peters, A., Stafoggia, M., de'Donato, F., Sera, F., Bell, M. L., . . . , & Breitner, S. (2024). Temporal variations in the short-term effects of ambient air pollution on cardiovascular and respiratory mortality: A pooled analysis of 380 urban areas over a 22-year period. *The Lancet Planetary Health*, *8*(9), e657–e665. https://doi.org/10.1016/S2542-5196(24)00168-2

[21] Cooper, M. J., Martin, R. V., Hammer, M. S., Levelt, P. F., Veefkind, P., Lamsal, L. N., . . . , & McLinden, C. A. (2022). Global fine-scale changes in ambient NO_2 during COVID-19 lockdowns. *Nature*, *601*(7893), 380–387. https://doi.org/10.1038/s41586-021-04229-0

[22] Zheng, M., Liu, F., & Wang, M. (2025). Assessing the COVID-19 lockdown impact on global air quality: A transportation perspective. *Atmosphere*, *16*(1), 113. https://doi.org/10.3390/atmos16010113

[23] Wong, Y. J., Shiu, H.-Y., Chang, J. H.-H., Ooi, M. C. G., Li, H.-H., Homma, R., . . . , & Nik Sulaiman, N. M. (2022). Spatiotemporal impact of COVID-19 on Taiwan air quality in the absence of a lockdown: Influence of urban public transportation use and meteorological conditions. *Journal of Cleaner Production*, *365*, 132893. https://doi.org/10.1016/j.jclepro.2022.132893

[24] Bagkis, E., Hassani, A., Schneider, P., DeSouza, P., Shetty, S., Kassandros, T., . . . , & Khan, J. (2025). Evolving trends in application of low-cost air quality sensor networks: Challenges and future directions. *npj Climate and Atmospheric Science*, *8*(1), 335. https://doi.org/10.1038/s41612-025-01216-4

[25] Zimmerman, N. (2022). Tutorial: Guidelines for implementing low-cost sensor networks for aerosol monitoring. *Journal of Aerosol Science*, *159*, 105872. https://doi.org/10.1016/j.jaerosci.2021.105872

[26] Schneider, P., Vogt, M., Haugen, R., Hassani, A., Castell, N., Dauge, F. R., & Bartonova, A. (2023). Deployment and evaluation of a network of open low-cost air quality sensor systems. *Atmosphere*, *14*(3), 540. https://doi.org/10.3390/atmos14030540

[27] Zhivkov, P., & Fidanova, S. (2026). Machine learning calibration transfer for low-cost air quality sensors: Distance-based uncertainty quantification in a hybrid urban monitoring network. *Atmosphere*, *17*(4), 335. https://doi.org/10.3390/atmos17040335

[28] Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, *231*, 997–1004. https://doi.org/10.1016/j.envpol.2017.08.114

[29] Pandey, M., George, M. P., Gupta, R. K., Gusain, D., & Dwivedi, A. (2021). Impact of COVID-19 induced lockdown and unlock down phases on the ambient air quality of Delhi, capital city of India. *Urban Climate*, *39*, 100945. https://doi.org/10.1016/j.uclim.2021.100945

[30] Biancardi, M., Zhou, Y., Kang, W., Xiao, T., Grubesic, T., Nelson, J., & Liang, L. (2024). Exploring spatiotemporal dynamics, seasonality, and time-of-day trends of $PM_{2.5}$ pollution with a low-cost sensor network: Insights from classic and spatially explicit Markov chains. *Applied Geography*, *172*, 103414. https://doi.org/10.1016/j.apgeog.2024.103414

[31] Cao, R., Li, B., Wang, Z., Peng, Z.-R., Tao, S., & Lou, S. (2020). Using a distributed air sensor network to investigate the spatiotemporal patterns of $PM_{2.5}$ concentrations. *Environmental Pollution*, *264*, 114549. https://doi.org/10.1016/j.envpol.2020.114549

[32] Marinov, E., Petrova-Antonova, D., & Malinov, S. (2022). Time series forecasting of air quality: A case study of Sofia City. *Atmosphere*, *13*(5), 788. https://doi.org/10.3390/atmos13050788

[33] Hameed, S., Islam, A., Ahmad, K., Belhaouari, S. B., Qadir, J., & Al-Fuqaha, A. (2023). Deep learning based multimodal urban air quality prediction and traffic analytics. *Scientific Reports*, *13*(1), 22181. https://doi.org/10.1038/s41598-023-49296-7

[34] Rabie, R., Asghari, M., Nosrati, H., Emami Niri, M., & Karimi, S. (2024). Spatially resolved air quality index prediction in megacities with a CNN-Bi-LSTM hybrid framework. *Sustainable Cities and Society*, *109*, 105537. https://doi.org/10.1016/j.scs.2024.105537

[35] Kırelli, Y. (2025). Air quality forecasting in urban environments: A deep learning approach. *Düzce University Journal of Science and Technology*, *13*(4), 1445–1454. https://doi.org/10.29130/dubited.1591784

[36] Asklany, S. A., Mohammed, D., Youssef, I. K., Nawaz, M., & Al Malwi, W. (2025). Forecasting urban air quality in Paris using ensemble machine learning: A scalable framework for environmental management. *PLOS One*, *20*(11), e0336897. https://doi.org/10.1371/journal.pone.0336897

[37] Huang, M.-L., Chamnisampan, N., & Ke, Y.-R. (2025). A deep learning model integrating EEMD and GRU for air quality index forecasting. *Atmosphere*, *16*(9), 1095. https://doi.org/10.3390/atmos16091095

[38] Eren, B., Erden, C., Atalı, A., & Ozdemir, S. (2025). A comparative analysis of hyperparameter optimization using LSTM-based deep learning models for urban air quality

predictions. *Ain Shams Engineering Journal*, *16*(12), 103786. https://doi.org/10.1016/j.asej.2025.103786

[39] Illescas-Martinez, F., Garcia, L., Garcia-Sanchez, A.-J., Asorey-Cacheda, R., & Garcia-Haro, J. (2025). Air quality forecasting in non-monitored urban areas through machine and deep-learning models. *Expert Systems with Applications*, *284*, 127749. https://doi.org/10.1016/j.eswa.2025.127749

[40] Song, E., Lam, H., & Barton, R. R. (2024). A shrinkage approach to improve direct bootstrap resampling under input uncertainty. *INFORMS Journal on Computing*, *36*(4), 1023–1039. https://doi.org/10.1287/ijoc.2022.0044

[41] Chen, J., Zhu, S., Wang, P., Zheng, Z., Shi, S., Li, X., . . . , & Meng, X. (2024). Predicting particulate matter, nitrogen dioxide, and ozone across Great Britain with high spatiotemporal resolution based on random forest models. *Science of the Total Environment*, *926*, 171831. https://doi.org/10.1016/j.scitotenv.2024.171831

[42] Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

[43] Moraffah, R., Sheth, P., Karami, M., Bhattacharya, A., Wang, Q., Tahir, A., . . . , & Liu, H. (2021). Causal inference for time series analysis: Problems, methods and evaluation. *Knowledge and Information Systems*, *63*(12), 3041–3085. https://doi.org/10.1007/s10115-021-01621-0

[44] Arumugam, V., & Natarajan, V. (2023). Time series modeling and forecasting using autoregressive integrated moving average and seasonal autoregressive integrated moving average models. *Instrumentation Mesure Métrologie*, *22*(4), 161–168. https://doi.org/10.18280/i2m.220404

[45] Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

[46] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324–342. https://doi.org/10.1287/mnsc.6.3.324

[47] Kramar, V., & Alchakov, V. (2023). Time-series forecasting of seasonal data using machine learning methods. *Algorithms*, *16*(5), 248. https://doi.org/10.3390/a16050248

[48] Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: A survey. *Artificial Intelligence Review*, *56*(9), 10031–10066. https://doi.org/10.1007/s10462-023-10424-4

[49] Yunita, A., Pratama, M. I., Almuzakki, M. Z., Ramadhan, H., Akhir, E. A. P., Firdausiah Mansur, A. B., & Basori, A. H. (2025). Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models. *MethodsX*, *15*, 103462. https://doi.org/10.1016/j.mex.2025.103462

[50] Chandra, R., Goyal, S., & Gupta, R. (2021). Evaluation of deep learning models for multi-step ahead time series prediction. *IEEE Access*, *9*, 83105–83123. https://doi.org/10.1109/ACCESS.2021.3085085

[51] Pham, X. T. T., & Ho, T. H. (2021). Using boosting algorithms to predict bank failure: An untold story. *International Review of Economics & Finance*, *76*, 40–54. https://doi.org/10.1016/j.iref.2021.05.005

How to Cite: Ohakwe, J., Ofori, D., Mishra, B. K., Sayers, W., Olusakin, T., & Chisimkwuo, J. (2026). A Data-Driven and Cost-Effective Framework for Urban Air Quality Monitoring and NO_2 Forecasting in Cheltenham. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS62028274