Article

Short-Term Hourly Ozone Concentration Forecasting Using Functional Data Approach

Ismail Shah [1,2,*,†], **Naveed Gul** [2,†], **Sajid Ali** [2,†] **and Hassan Houmani** [3,*,†]

1   Department of Statistical Sciences, University of Padua, 35121 Padua, Italy
2   Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan;
    naveedgul@stat.qau.edu.pk (N.G.); sajidali@qau.edu.pk (S.A.)
3   Department of Economics, School of Business, Lebanese International University-LIU, Beirut 146404, Lebanon
*   Correspondence: ismail.shah@unipd.it (I.S.); hassan.houmani@liu.edu.lb (H.H.)
†   These authors contributed equally to this work.

Abstract: Air pollution, especially ground-level ozone, poses severe threats to human health and ecosystems. Accurate forecasting of ozone concentrations is essential for reducing its adverse effects. This study aims to use the functional time series approach to model ozone concentrations, a method less explored in the literature, and compare it with traditional time series and machine learning models. To this end, the ozone concentration hourly time series is first filtered for yearly seasonality using smoothing splines that lead us to the stochastic (residual) component. The stochastic component is modeled and forecast using a functional autoregressive model (FAR), where each daily ozone concentration profile is considered a single functional datum. For comparison purposes, different traditional and machine learning techniques, such as autoregressive integrated moving average (ARIMA), vector autoregressive (VAR), neural network autoregressive (NNAR), random forest (RF), and support vector machine (SVM), are also used to model and forecast the stochastic component. Once the forecast from the yearly seasonality component and stochastic component are obtained, both are added to obtain the final forecast. For empirical investigation, data consisting of hourly ozone measurements from Los Angeles from 2013 to 2017 are used, and one-day-ahead out-of-sample forecasts are obtained for a complete year. Based on the evaluation metrics, such as R^2, root mean squared error (RMSE), and mean absolute error (MAE), the forecasting results indicate that the FAR outperforms the competitors in most scenarios, with the SVM model performing the least favorably across all cases.

Keywords: ozone concentration; functional data analysis; forecasting; autoregressive; machine learning

MSC: 37M10; 81T80; 46N30; 62J05

1. Introduction

Air pollution refers to the existence of detrimental materials such as gases, solids, or liquid particles in the atmosphere, leading to harmful consequences for human health, the environment, and ecosystems. These materials are termed pollutants and can have diverse origins, arising from sources like industrial operations, transportation, natural occurrences, and human activities. Typical pollutants include ozone (O_3), nitric oxide (NO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), particulate matter 2.5 ($PM_{2.5}$), and particulate matter 10 (PM_{10}), among others. Air pollution, especially ground-level O_3, is a significant global concern affecting public health and the environment (Seinfeld and Pandis 2016).

In the United States of America (USA), the environmental protection agency (EPA) estimates that mobile sources such as cars, buses, planes, trucks, and trains account for about half of the cancer risk and more than 70% of the non-cancer health effects associated

with air toxics (Munshed et al. 2023). Stationary sources such as power plants, oil refineries, industrial facilities, and factories emit large amounts of pollutants from particular sites called point sources. Area sources are made up of several minor pollution contributors that combined have a significant impact on air quality. Area sources include agricultural areas, urban regions, and wood-burning fireplaces. Natural sources like wind-blown dust, wildfires, and volcanic activity can also impact air quality. Figure 1 displays the different sources described above that contribute significantly to air pollution. The collective impact of these sources on the environment poses a serious challenge for researchers and policymakers (Lotrecchiano et al. 2020; Mavroidis and Ilia 2012).
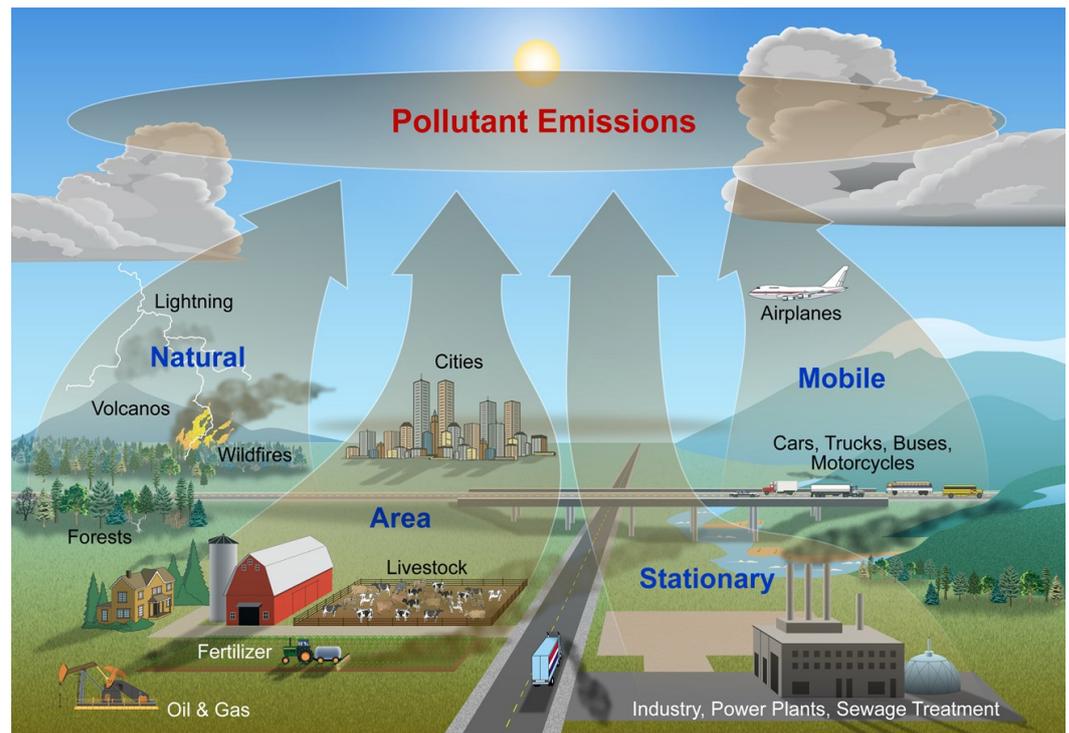
Figure 1. Different sources of air pollution. Source: (Suraki 2013).

Exposure to ground-level O_3 can have a range of adverse effects on human health. For example, O_3 exposure is associated with an elevated risk of chronic respiratory and cardiovascular ailments, as indicated by numerous studies (World Health Organization 2020). Even short-term exposure to O_3 can perturb lung mucociliary function, consequently weakening resistance to bacterial infections (Sujith et al. 2017). Alarming statistics reveal that approximately 16,400 premature deaths within the European Union (EU) are attributed to O_3 pollution (Nuvolone et al. 2018). Beyond human health, O_3 poses a substantial threat to vegetation and biodiversity. Its capacity to impair crops, forests, and other forms of vegetation manifests in the form of diminished photosynthesis, reduced carbon sequestration, and losses in biodiversity. Moreover, it induces visible leaf injuries (Paoletti and Manning 2007). This concern extends to various regions, including parts of southern Europe, where a multitude of grassland areas face the risk of high O_3 levels, impacting plant composition, seasonal flowering, and seed production for diverse natural species (Mills et al. 2016).

Ozone concentrations are important for air quality management and public health, and accurate prediction of these concentrations can aid in decision-making processes. Due to the importance of this topic, researchers in the past have proposed different methods and techniques to model and forecast O_3 concentrations. For example, Hong et al. (2023) presented a novel approach for predicting hourly O_3 levels using deep learning techniques called the multi-order difference embedded long short-term memory (MDELSTM) method, using a Los Angeles air quality dataset. The performance of the proposed model

is compared with partial least squares (PLS), gated recurrent unit (GRU), long short-term memory (LSTM), multilayer perceptron (MLP), and stationary difference embedded LSTM (SDELSTM). Based on the results, the MDELSTM model demonstrates superior prediction performance among the models considered. Machine learning techniques are popular for forecasting O_3 concentrations (Eslami et al. 2020; Hashim et al. 2022; Yafouz et al. 2022). For instance, Yafouz et al. (2022) conducted a study to predict the tropospheric O_3 concentration using different machine learning models. These models included linear regression (LR), support vector regression (SVR), tree regression (TR), Gaussian process regression (GPR), ensemble regression (ER), and artificial neural networks (ANNs). Using data from Peninsular, Malaysia, the results indicated that the LR, SVR, GPR, and ANN performed better in terms of having a high R^2.

Many researchers have investigated the performance of classical time series models for forecasting O_3 concentrations (Salazar et al. 2019). For example, Kumar et al. (2004) studied the autoregressive integrated moving average (ARIMA) model to forecast the daily surface O_3 concentration. Using a real dataset, the ARIMA (1,0,1) model demonstrated the best performance in predicting the maximum daily O_3 concentration, with a MAPE of 13.14%. Aneiros-Pérez et al. (2004) forecast daily maximum O_3 concentrations in Toulouse, France, using nonlinear models based on kernel estimators. They also included exogenous variables in their analysis. The study found that the functional additive model (FAM), with its back-fitting kernel approach, produced the lowest quadratic errors and explained the highest percentage of variability in the data compared to other models. Due to the challenges posed by O_3 to human health, researchers are actively proposing different models to model and forecast O_3 concentration time series data (Arsić et al. 2020; Gao et al. 2018; Ghoneim et al. 2017; Rahman and Nasher 2024; Su et al. 2020).

This research contributes to the literature on O_3 forecasting by applying functional data analysis (FDA) methods, which can capture the dynamic and complex features of the O_3 concentration as a function of time. FDA methods have been used in various fields such as bio-statistics, econometrics, and environmental science, but are less explored in the context of O_3 forecasting (Jan et al. 2022; Shah et al. 2022). This study proposes a novel time series model based on FDA, which treats each day as a single functional observation with 24 discrete points. The performance of the FDA model is compared with classical time series and machine learning models using standard accuracy metrics. The results of this study provide insights into the advantages and limitations of different forecasting methods and suggest ways to improve the accuracy and reliability of O_3 forecasts. This study also has practical implications for policymakers, environmental managers, and public health officials, who can use the O_3 forecasts to implement timely and effective measures to mitigate the adverse effects of O_3 pollution.

The rest of the article is organized in the following manner. Section 2 provides details about the general modeling framework and the details of different models used in this research work. Analysis and results based on empirical investigation are provided in Section 3. Finally, Section 4 concludes the study.

2. Methodology

This section explores the predictive models used in this study to model and forecast O_3 concentration data. To this end, our proposed approach FAR is outlined in more detail. Several parametric and non-parametric competing models like ARIMA, VAR, SVM, NNAR, and RF are also briefly described in this section. Before going into detail, we revisit some of the preliminaries of functional data analysis.

2.1. Functional Data Analysis

James O. Ramsay coined the term "functional data analysis" to describe this field of study (Ramsay 1982). FDA is at the forefront of modern statistical computing, ushering transformative changes across diverse fields. The rapid advancement of technology allows for quicker and more accurate measurement equipment, facilitating the collection of

continuous data spanning time, space, or other continua. This paradigm shift challenges classical statistical assumptions, particularly the conventional belief that the number of data points should surpass the number of variables in a dataset. In response to this evolution, FDA emerged as a field treating data as functional, where each curve is regarded as a single observation, rather than a collection of discrete data points. The FDA method leverages information within and across sample units to incorporate derivatives, smoothness, and other inherent features of the functional data structure. The process of converting data to functional objects involves collecting information at discretized points, typically equally spaced, and implementing suitable basis functions to eliminate common noise while accurately representing each data curve.

Basis functions are pivotal in FDA, serving as the fundamental components for modeling intricate functional observations. These functions are expressed as a linear combination of coefficients C and known basis functions ϕ. The representation of a functional observation $y(j)$ is given by

$$y(j) = \sum_{k=1}^{K} C_k \phi_k(j) \tag{1}$$

where C_k denotes parameters, $\phi_k(j)$ represents the known basis functions, and K is the number of basis functions used. A more simplified representation using matrix notation can be written as

$$\mathbf{y} = \mathbf{c}^{\top} \phi = \phi^{\top} \mathbf{c}$$

Several common basis functions, including the Fourier basis, B-spline basis, polynomial principal components, and exponential basis are used to cater to different data characteristics. The selection depends on the nature of the data, e.g., the Fourier basis is generally suitable for periodic data and the B-spline basis is preferred for non-periodic data. In our research, we chose the B-spline basis system for its applicability to non-periodic data.

The B-spline basis functions are polynomial segments that are joined at certain knots or breakpoints. The compact support property, where each basis function is positive over a limited number of adjacent intervals, increases the computational efficiency. The order of a B-spline determines the degree of polynomial segments while the knot sequence governs their placement. The strategic placement of knots, whether equally spaced or tailored to data characteristics, plays a vital role in achieving accurate and meaningful representations. Coincident knots are employed strategically, offering flexibility to induce specific characteristics like derivative discontinuities.

2.2. Functional Autoregressive Model

The FAR model used in this study is an extension of the classical AR model for functional data. The FAR model of order 1, FAR(1), operates within the framework of a separable Hilbert space denoted by H. It specifically considers the Hilbert space $L^2[0,1]$, although the idea is applicable to other L^2-spaces. An autoregressive Hilbertian process of order 1, ARH(1), which is also called FAR(1), is identified in this context as a sequence Y_t of H-random variables. The process is strictly stationary and satisfies the AR equation:

$$Y_t(j) - \mu(j) = \psi(Y_{t-1}(j) - \mu(j)) + \epsilon_t(j) \quad, \quad j \in J \tag{2}$$

where $\mu(j)$ is the mean function, ψ is a bounded linear operator, and $\epsilon(j)$ is the shock or innovation term. The AR operator ψ is assumed to be a compact Hilbert–Schmidt, symmetric, and positive bounded linear operator from $L^2[0,1]$ to itself. The compactness property permits a decomposition using orthonormal bases and real numbers, and the operator is said to be nuclear if certain eigenvalue conditions are met. Within the functional framework, we will drop the compact support (j) for simplicity.

2.2.1. Operators in the Hilbert Space ($L^2[0,1]$)

Operators in the Hilbert space $L^2[0,1]$ are defined by a norm and are bounded linear operators and obtained from the inner product of the space H. This norm is the supremum of the operator's norm over all unit vectors y in H. It is represented by the symbol $\|\psi\|_L$ and written as

$$\|\psi\|_L = \sup_{\|y\| \leq 1} \|\psi(y)\|$$

An operator ψ is termed compact with respect to orthonormal bases v_j and f_j in H and a sequence λ_j approaching zero if it can be expressed as

$$\psi(y) = \sum_{j=1}^{\infty} \lambda_j \langle y, v_j \rangle f_j$$

If the sum of the squares of an operator's sequence λ_j is finite, the operator is Hilbert–Schmidt, i.e.,

$$\sum_{j=1}^{\infty} \lambda_j^2 < \infty$$

The space of Hilbert–Schmidt operators S is separable and is equipped with an inner product and associated norm:

$$\langle \psi_1, \psi_2 \rangle_S = \sum_{1 <= i,j <= \infty} \langle \psi_1(g_i), h_j \rangle \langle \psi_2(g_i), h_j \rangle$$

$$\|\psi\|_S = \left(\sum_j \lambda_j^2 \right)^{1/2}$$

Operators are classified as symmetric if $\langle \psi(x), y \rangle = \langle x, \psi(y) \rangle$, and positive if $\langle \psi(y), y \rangle \geq 0$. A symmetric positive Hilbert–Schmidt operator admits the decomposition

$$\psi(y) = \sum_{j=1}^{\infty} \lambda_j \langle y, v_j \rangle v_j$$

A compact operator is nuclear if the sum of the absolute values of its sequence λ_j is finite:

$$\sum_j |\lambda_j| < \infty$$

The norms of these operators follow the relationship $\| \cdot \|_N \geq \| \cdot \|_S \geq \| \cdot \|_L$. If ψ is an integral operator in L^2 defined by $\psi(x)(t) = \int \psi(t,s)x(s)ds$, where $\psi(.,.)$ is a real kernel, it is a Hilbert–Schmidt operator if and only if $\int \int \psi^2(t,s)dtds < \infty$. The model is non-parametric, as ψ represents an infinite-dimensional parameter.

2.2.2. Estimation of the Operator ψ

Estimating the AR operator ψ in the FAR model involves addressing specific assumptions for obtaining a stationary solution. Two assumptions are considered to ensure the existence of a stationary solution. The first assumption requires the existence of an integer $j_0 \geq 1$ such that $\|\psi^{j_0}\|_L < 1$, while the second assumption necessitates the existence of $a > 0$ and $0 < b < 1$ such that $\|\psi^j\|_L \leq ab^j$ for all $j \geq 0$. These assumptions, under certain conditions, guarantee a unique strictly stationary solution, as proven in Bosq (2000).

It is important to note that estimating ψ cannot rely on likelihood due to the non-existence of the Lebesgue measure in non-locally compact spaces and the concept of density is not available for the functional data. Instead, the classical method of moments is employed. The estimation of ψ is represented as $\psi = C\Gamma^{-1}$, where $\Gamma = E(Y_t \otimes Y_t)$ and $C = E(Y_t \otimes Y_{t+1})$ are the covariance and cross-covariance operators of the process, and $\otimes$

is the Kronecker product. The sample versions of these operators are denoted as $\hat{\Gamma}_t$ and $\hat{C}_t$, respectively.

To simplify notation, it is assumed that the mean of the process $E(Y_t) = 0$ is known. The sample versions of the covariance and cross-covariance operators, denoted as $\hat{\Gamma}_t$ and $\hat{C}_t$, are given by $\hat{\Gamma}_t = \frac{1}{n} \sum_{t=0}^{n-1} Y_t \otimes Y_t$ and $\hat{C}_t = \frac{1}{n} \sum_{t=0}^{n-1} Y_t \otimes Y_{t+1}$.

The covariance operator Γ is a symmetric, positive definite, and compact operator. It can be decomposed into eigenfunctions and eigenvalues, denoted as λ_j and ν_j, respectively. However, Γ^{-1} is not a bounded operator. To address this, a practical solution is proposed involving the consideration of the first p most important empirical functional principal components as substitutes for unknown population principal components, as given by

$$\Gamma_t^{-1}(y) = \sum_{j=1}^{p} \lambda_j^{-1} \langle y, \hat{v}_j \rangle \hat{v}_j = \Gamma_t^{\dagger}(y)$$

From the ARH(1) equation, when multiplying by Y_n we obtain the relation

$$Y_t \otimes Y_{t+1} = Y_t \otimes (\psi Y_t) + Y_t \otimes \epsilon_{t+1} = \psi Y_t \otimes Y_t + Y_t \otimes \epsilon_{t+1}$$

By the definitions of the covariance and cross-covariance operators of ARH(1) and using $E(\epsilon) = 0$, we have $C = \psi \Gamma$ and $\psi = C \Gamma^{-1}$. The estimate of ψ is then given by

$$\hat{\psi}_n(y) = \frac{1}{n-1} \sum_{k=1}^{n-1} \sum_{j=1}^{p} \sum_{i=1}^{p} \hat{\lambda}_j^{-1} \langle y, \hat{v}_j \rangle \langle Y_k, \hat{v}_j \rangle \langle Y_{k+1}, \hat{v}_i \rangle \hat{v}_i$$

The last term is derived by performing an additional smoothing step on Y_{k+1} and $\hat{v}_j$. The empirical eigenfunctions are known to converge asymptotically to the population eigenfunctions.

Once the estimator $\hat{\psi}$ of the population parameter ψ is obtained, it is crucial to assess its optimality in estimating the true parameter. For the FAR parameter ψ, Didericksen et al. (2012) demonstrated that the estimator is optimal in terms of MSE and MAE, as its prediction error is comparable to the infeasible predictor $\psi(y)$ for an appropriately chosen p.

2.3. Autoregressive Integrated Moving Average (ARIMA) Models

Time series data analysis is a vital tool for comprehending and forecasting temporal patterns. The autoregressive moving average (ARMA) model combines elements of autoregressive (AR) and moving average (MA) models, providing a framework for modeling univariate time series data. For a univariate time series Y_t, an ARMA(p, q) model can be written as

$$Y_t = C + \sum_{r=1}^{p} \phi_r Y_{t-r} + \sum_{l=1}^{q} \Phi_l Z_{t-l} + Z_t \tag{3}$$

Equation (3) contains an intercept term C, AR parameters ϕ_r ($r = 1, 2, \ldots, p$), MA parameters Φ_l ($l = 1, 2, \ldots, q$), and a white noise term $Z_t \sim N(0, \sigma_z^2)$. The ARMA models are well-suited for stationary time series data. However, differencing is required to achieve stationarity for non-stationary data, and this is where the ARIMA model comes into play. The ARIMA model is an extension of ARMA specifically tailored for non-stationary time series (Shumway et al. 2000). It involves differencing the data to attain stationarity, which can be expressed as

$$Y_t^d = \sum_{r=1}^{p} \phi_r Y_{t-r}^d + \sum_{l=1}^{q} \Phi_i Z_{t-l} + Z_t \tag{4}$$

where Y_t^d represents the dth difference of the series, ϕ_r ($r = 1, 2, \ldots, p$) and Φ_l ($l = 1, 2, \ldots, q$) are the parameters of the AR and MA components, respectively, and $Z_t \sim N(o, \sigma_z^2)$.

Identifying the appropriate ARIMA model is a critical step and it involves determining the order of differencing '*d*' and the numbers of AR (*p*) and MA (*q*) terms. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are valuable tools in this process. The ACF indicates the series correlation with itself at different lags, whereas the PACF reveals autocorrelation at a lag *k* with intervening data deleted. For parameter estimation, generally, the maximum likelihood estimation (MLE) is used (Shumway et al. 2000). This study investigated different models and found that the ARIMA(5,0,0), which is a pure autoregressive model with five lagged observations, fits the data well and provides white noise errors, thus being suitable for forecasting.

2.4. Vector Autoregressive

A vector autoregressive (VAR) model is a powerful and commonly used time series analysis method that enables us to capture the dynamic interactions between variables over time. It extends the concept of univariate AR models to a multivariate situation. Because of their flexibility and forecasting accuracy, VAR models are extremely useful for understanding and forecasting complicated real-world behavior. They can effectively express variable interdependence.

A VAR model is built on a set of equations that describe the evolution of various time series variables. Each variable in the model is treated as a linear function of its own lagged values as well as the lagged values of all other variables in the model. Suppose $\mathbf{Y_t}$ is a vector of univariate time series, then a VAR model can be written as

$$\mathbf{Y_t} = C + \mathbf{\Phi_1 Y_{t-1}} + \mathbf{\Phi_2 Y_{t-2}} + \cdots + \mathbf{\Phi_p Y_{t-p}} + \epsilon_t \tag{5}$$

where $\mathbf{Y_t}$ is an $n \times 1$ vector representing the current values of n distinct response time series variables at time t. The $n \times 1$ vector C consists of constant offsets that serve as intercepts, accounting for the baseline level of the variables. The $\mathbf{\Phi}_j$ matrices, ranging from $j = 1$ to p, are $n \times n$ matrices of AR coefficients. These matrices capture the relationships between the variables at different time lags; specifically, the impact of past values up to lag p on the current values. The parameter "p" defines the order of the VAR model, determining the maximum number of past periods considered in each equation and influencing the model's ability to capture system dynamics. Finally, ϵ_t represents an $n \times 1$ vector of "white noise" terms, introducing randomness into the model and accounting for unexplained variability not captured by the lagged variables, thereby enhancing the model's accuracy in describing the underlying time series data.

The choice of the order is a critical step and is typically determined through cross-validation techniques and information criteria such as AIC, BIC, and HQ. As we estimate models for 365 days, the value of the selected p can vary. However, in most cases, a VAR of order five, i.e., VAR(5), is suitable as it generally provides whitened residuals and produces lower AIC values compared to other orders of the model. Estimating and inferring parameters in VAR models is typically achieved by using the MLE or ordinary least squares (OLS) techniques. These methods rely on certain assumptions, such as the error terms having a conditional mean of zero, stationary variables, and the absence of perfect multicollinearity.

2.5. Artificial Neural Networks

Artificial neural networks (ANNs) are a foundational component of machine learning, particularly in the realm of deep learning. They mirror the human brain's structure and functionality, comprising interconnected nodes arranged in layers: input, hidden, and output. These nodes, akin to artificial neurons, process information through weighted connections, activating based on a threshold to transmit data across layers. The ANNs are renowned for their proficiency in rapid data classification, clustering, and various applications such as speech and image recognition, with Google's search algorithm being a prominent example of their implementation.

ANNs constitute a multi-layered architecture comprising essential components featuring an input layer, hidden layers, and an output layer. The input layer receives diverse

data formats while the hidden layer, functioning as a "distillation layer", extracts pertinent patterns to enhance network efficiency by recognizing crucial information and discarding redundancy. The output layer processes transformed information from the hidden layers to produce the final output. The most commonly employed ANNs for predictive tasks are MLPs with hidden layers, utilizing a three-layered network interconnected by acyclic connections. Nodes are considered as manufacturing components and the architecture allows for more than one hidden layer.

Neural Network Autoregressive

The neural network autoregressive (NNAR) model is a potent deep learning structure widely employed in various fields like natural language processing, time series forecasting, and speech recognition. It utilizes past data to predict future values, employing a recursive approach where each time step's input is the prior prediction. This model handles non-linear relationships within intricate datasets and can be constructed using feedforward or recurrent neural networks.

The NNAR(*p*, *k*) architecture involves "*p*" lagged inputs for forecasting and includes "*k*" nodes in the hidden layer. It shares similarities with ARIMA models but operates without constraints for stationarity. The NNAR model equation comprises weighted node connections, nonlinear activation functions, AR dependencies, exogenous impacts, and error terms, and can be written as

$$Y_{it} = f\left(\alpha_1 \sum_{j=1}^{N} w_{ij} Y_{j(t-1)} + \lambda Z_{i(t-1)}^{T}\right) + \epsilon_t \tag{6}$$

where Y_{it} represents the univariate response at node i and time t, w_{ij} indicates the connection strength between nodes i and j for $i, j = 1, \ldots, N$, $\lambda Z_{i(t-1)}^{T}$ captures the exogenous impact, and $f()$ is an unidentified smoothed link function. Estimation methods for the NNAR include profile least squares estimation and local linear approximation to model the unknown link function and optimize parameter estimates. This model iterates through historical inputs to generate multi-step forecasts in time series analysis, showing adaptability and robustness in handling intricate data patterns. In our case, a simple autoregressive of order one with one node in the hidden layer, NNAR(1,1), is used.

2.6. Support Vector Machine

Support vector machines (SVMs) are powerful supervised learning models widely employed for data analysis in classification, regression, and outlier detection tasks. They excel in handling both linear and nonlinear data separations, making them versatile in various domains such as text classification, image recognition, gene expression analysis, and even time series prediction. Notably, these machines were meticulously designed not only for effective ranking but also for efficiently simplifying training set outcomes. This distinct quality has led to the widespread adoption of SVM techniques, especially in forecasting time series.

An SVM commences with the input data, composed of labeled feature vectors, where each vector corresponds to one of two classes. Extending beyond binary classification, SVMs can also tackle multi-class classification problems. These feature vectors are then transformed into a higher-dimensional space using various kernel functions, such as the linear, polynomial, or radial basis functions. This transformation is pivotal as it calculates the similarity between different feature vectors, allowing the SVM to address nonlinear separable data.

Key Elements and Optimization in the SVM

- Training data and class labels: The training data contains assigned class labels (y_i) denoted within $\{-1, 1\}$ for binary classification, serving as an anchor for the model's learning process.

- Margin and decision boundary: The SVM aims to establish a hyperplane maximizing the margin between classes, mathematically expressed as

$$w^T x + b = 0,$$

  where w denotes the weight vector and b is the bias.
- Support vectors and decision boundary: Essential in shaping the decision boundary, support vectors are the data points lying within or on the margin. They play a defining role in establishing the decision boundary.
- Optimization and decision function: The optimization process fine-tunes key parameters like the weight vector, bias, and regularization factor to craft the decision function.
- Optimization (mathematical model): The hyperplane determination involves an optimization problem:

$$\text{Minimizing}(\frac{1}{2}\|w\|^2 + C \sum \xi_i)$$

  under the constraints $y_i(w^T x_i + b) \geq 1 - \xi_i$ with $\xi_i \geq 0$, where C denotes the regularization parameter and ξ_i are slack variables permitting misclassifications within margin boundaries.
- Decision function and prediction: The decision function anticipates the class label for a new data point x, utilizing the transformed feature vector generated via the kernel function. This function is represented as $f(x) = \text{sign}(w^T \phi(x) + b)$, where $\phi(x)$ denotes the transformed input feature vector.

In this work, the SVM model is an epsilon-regression type with a radial basis function kernel. It uses two lagged values of ozone concentration as inputs and is parameterized with a cost of 1.0, gamma of 0.5, and epsilon of 0.1. The model is supported by 1191 support vectors to ensure accurate predictions.

2.7. Random Forest

The random forest (RF) algorithm, introduced by Leo Breiman (Breiman 2001), is an important machine learning method suitable for both classification and regression tasks. It employs an ensemble of decision trees, each trained on a distinct subset of data created through bagging. Bagging, or bootstrap aggregating, involves forming diverse training sets for each tree by random sampling with replacement. A distinctive feature of the RF is the additional randomness introduced during tree construction. Instead of opting for the optimal feature for node splitting, it considers a random subset of features, thereby enhancing model diversity. This unique characteristic contributes to RF's robustness against overfitting and its effectiveness in handling complex datasets.

Random forest constructs a "forest" of decision trees through bagging. Each decision tree is descriptive, with a root node containing the variable of interest and leaf nodes representing predicted outcomes. The trees grow without pruning and predictions for new observations are made by aggregating results using $f(x) = \frac{1}{K} \sum_{i=1}^{K} f_i(x)$, where $f(x)$ is the RF regression predictor, K is the number of trees, and $f_i(x)$ represents the individual regression trees.

Working Principle of Random Forests

- Tree construction:

1. Random subsampling (bootstrap aggregating): The algorithm selects a subset of training data with a replacement (bootstrap sample) and trains a decision tree on each subset. Samples are obtained by randomly selecting observations with replacements from the original dataset.

2. Random feature selection: At each decision tree node, a random subset of features is chosen for splitting, reducing the correlation between trees and enhancing ensemble diversity. The number of features considered at each node is typically the square root of the total features for classification and one-third for regression.

- Prediction and aggregation:

3. OOB error and feature importance: Post-training, RF provides two indices: the out-of-bag (OOB) error and the importance value of each feature. The OOB error measures prediction accuracy on data not included in the bootstrap sample, while feature importance is determined using the OOB dataset, aiding in variable selection.

4. Voting: Once all decision trees are trained, predictions are made by aggregating individual tree results. Majority voting is used for classification, and averaging for regression. The algorithm's performance is assessed through OOB error, providing an estimate of the RF's generalization error by $\hat{y} = \frac{1}{n_{\text{tree}}} \sum_{j=1}^{n_{\text{tree}}} y_j$, where $\hat{y}$ is the final predicted value, n_{tree} is the number of trees in RF, and y_j is the predictive value of the jth tree.

- Introducing randomness:

The RF introduces randomness by selecting a subset of features for node splitting, enhancing model performance and reducing overfitting. The equation for the RF regression predictor with randomness in feature selection is $f(x) = \frac{1}{K} \sum_{i=1}^{K} f_i(x, \theta_i)$ where θ_i are independent random vectors with the same distribution, representing the randomly selected features.

- Hyperparameters and feature importance:

The RF has hyperparameters such as the number of trees ($n_{\text{estimators}}$), maximum features for splitting a node max_features, and minimum samples required to split an internal node (min_samples_leaf). These can be tuned for optimal performance. The RF assesses feature importance by measuring the accuracy decrease using OOB error estimation during feature selection.

Finally, our RF model uses four lagged values of ozone concentration as predictors in this work. It was fine-tuned through 5-fold cross-validation, achieving optimal performance with the number of variables tried at each split (mtry) set to 2.

2.8. General Modeling Framework

The main goal of this study is to forecast O_3 concentrations and compare the proposed model, FAR(1), with other traditional time series and machine learning models. To explain in detail, for the O_3 concentration series denoted by $X_{t,h}$ for the tth day ($t = 1, 2, \ldots, n$) and the hth hour in a day ($h = 1, 2, \ldots, 24$), the dynamics of the O_3 concentration can be modeled as

$$X_{t,h} = D_{t,h} + Y_{t,h} . \tag{7}$$

This means that the O_3 series $Y_{t,h}$ is decomposed into two components. The first one is $D_{t,h}$, which is the deterministic component that contains long-term dynamics, and the second one is $Y_{t,h}$, which is the stochastic component that captures short-term variations. The deterministic component consists of yearly seasonality, which is modeled and forecast using the generalized additive modeling (GAM) technique using smoothing splines. More precisely, the yearly seasonality, which is represented by the series (1, 2,..., 365, 1, 2,..., 365, 1, 2,..., 365, 1, 2,..., 366, 1, 2,..., 365), is modeled using a smoothing spline, the one-day-ahead forecast for $y_{t+1,h} = y_{t,h}$, as this component represents long-term dynamics for our forecast horizon. Hence, the one-day-ahead forecast for the deterministic component is obtained as follows:

$$\hat{D}_{t+1,h} = \hat{D}_{t,h} . \tag{8}$$

To compute the stochastic component, we subtract the deterministic component $D_{t,h}$ from $X_{t,h}$. Mathematically, this is represented as

$$Y_{t,h} = X_{t,h} - \hat{D}_{t,h} . \tag{9}$$

The modeling and forecasting of the stochastic component is performed using FAR(1) and five alternate competing models. To this end, the stochastic component $Y_{t,h}$ is converted to a matrix of dimensions "days × hours", where each row represents a day and each

column represents an hour. In the case of FAR(1), this matrix is then transformed into a functional object using the B-spline basis function given in Equation (1), that results in $Y_t(j)$, and the FAR(1) model is applied thereafter. In the case of the VAR model, the matrix of hourly series is considered and modeled as given in Section 2.4. Finally, for all other models, the hourly series is modeled and forecast separately considering each hourly series as a univariate time series.

Once both components, deterministic and stochastic, are modeled and forecast separately, the final one-day-ahead out-of-sample forecasts are obtained by combining the forecasts of both components. In summary, the following procedure is used to obtain one-day-ahead out-of-sample forecasts from different models.

- For an hourly time series, decompose the O_3 concentration time series into deterministic and stochastic components, as given in Equation (7).
- Using Equation (8), obtain a one-step-ahead forecast for each hour that will result in a 24 h one-day-ahead forecast.
- Obtain the stochastic component using Equation (9). Use the models described in Section 2 to obtain a one-day-ahead forecast.
- Compute the final one-day-ahead out-of-sample forecasts by combining the forecasts from deterministic and stochastic components.
- Repeat the above steps for 365 days using a rolling window to obtain the one-day-ahead out-of-sample forecasts for the entire year.

Figure 2 shows the flowchart of the proposed modeling framework.
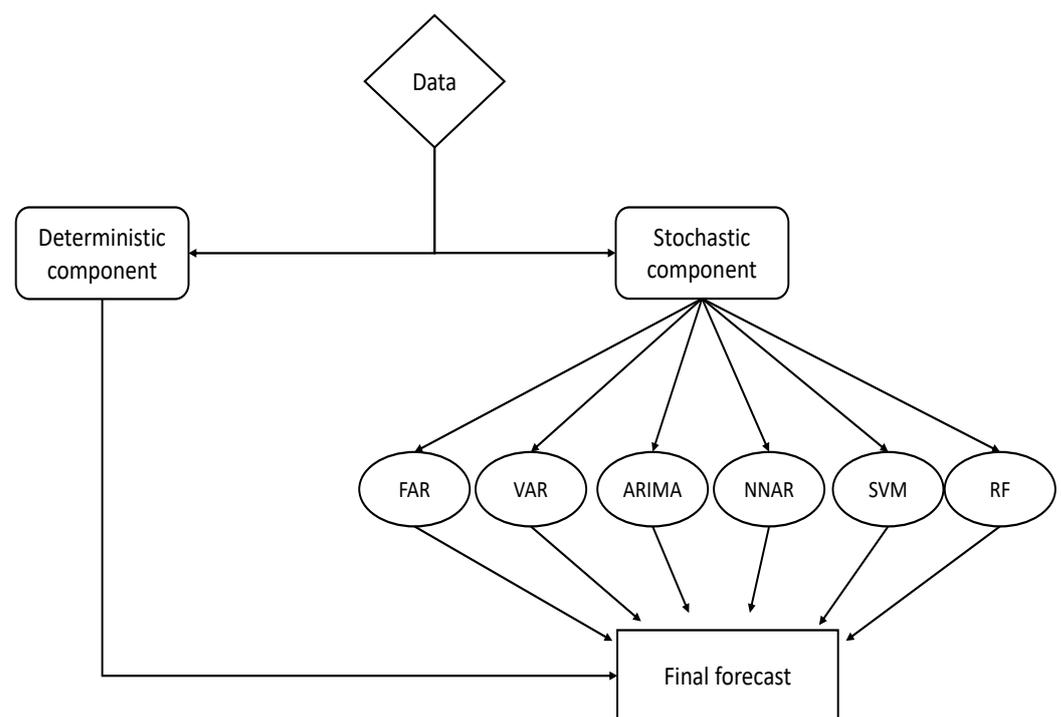
Figure 2. Flowchart of the proposed general modeling framework.

3. Analysis and Results

3.1. Data Overview and Preprocessing

The data for this study originate from Los Angeles, an area highly prone to ground-level O_3 pollution due to its unique geography, surrounded by mountains and bordered by the sea, resulting in a Mediterranean climate with high temperatures and minimal rainfall. The dataset utilized for our research was sourced exclusively from O_3 concentration measurements taken at the North Main Street monitoring site (-118.22688 E, 34.06659 N) as part of the Environmental Protection Agency's (EPA's) air quality system in Los Angeles[1].

This dataset spans the years from 2013 to 2017, capturing a comprehensive record of O_3 levels in the region. While the EPA's air quality system typically collects a wide range of meteorological and air pollution data, we have focused specifically on the O_3 concentration data for our study, thereby refining our dataset to provide a precise and specialized foundation for our research. The dataset is plotted in Figure 3, where the red line separates the model estimation and out-of-sample forecasting periods. This figure shows considerable variation in O_3 concentration throughout the year.
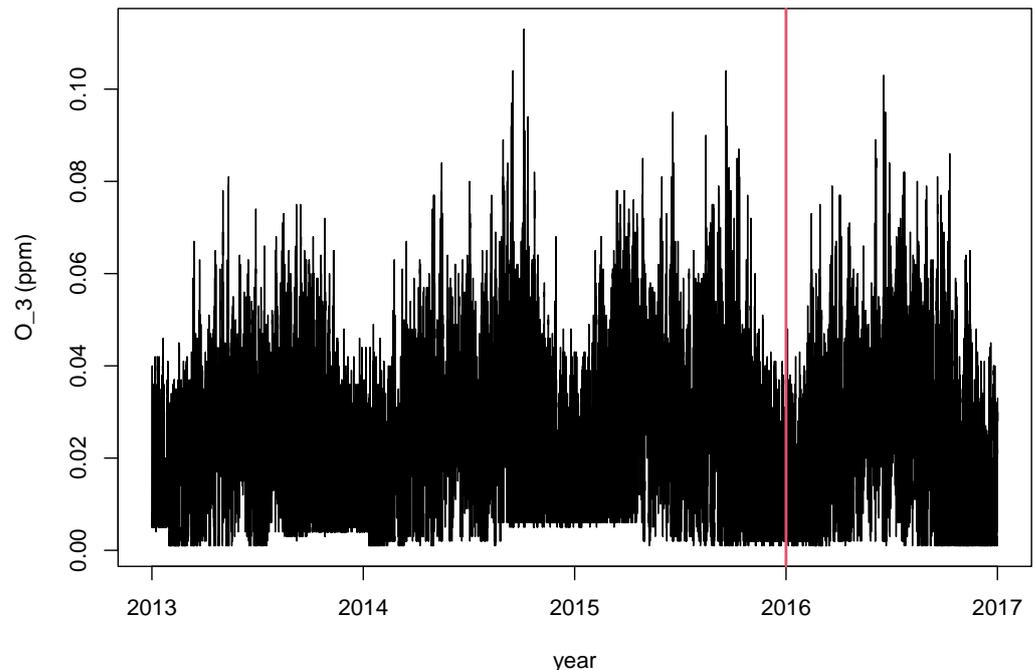
Figure 3. O_3 concentration time series for Los Angeles. The red line divides the estimation and out-of-sample forecasting periods.

Missing values are common in air quality datasets and we have a very small number of missing values in our dataset. To address this, linear interpolation is applied to estimate missing values by assuming a linear relationship between neighboring known data points. This method calculates the value of a missing data point based on the values of the nearest data points before and after it in the dataset. This method is commonly used in various fields, such as time series analysis, geographic information systems, and data analysis, to interpolate and fill in missing data points. The dataset was organized with 80% of the data, spanning from 2013 to 2016, allocated for model estimation, while the remaining 20%, which corresponds to the data for 2017, was reserved for one-day-ahead out-of-sample forecasts.

The descriptive statistics in Table 1 reveal that the smallest observed value is 0.0010 parts per million (ppm), while the largest value stands at 0.1160 ppm. The mean O_3 concentration is calculated at 0.0239 ppm. Notably, the distribution appears to be slightly positively skewed, evident from the mean being marginally higher than the median (0.0220 ppm). Moreover, the standard deviation of 0.0180 ppm signifies a reasonable level of variability around the mean O_3 concentration value. Additionally, the first quartile, at 0.0070 ppm, and the third quartile, at 0.0360 ppm, help us to understand where the majority of values cluster within the dataset.

Table 1. Descriptive statistics of O_3 concentration time series.

Statistic	Value
Minimum	0.0010
1st Quartile	0.0070
Median	0.0220
Mean	0.0239
3rd Quartile	0.0360
Maximum	0.1160
Standard Deviation	0.0180

3.2. Forecasting Accuracy Metrics

The accuracy of our forecasting models is measured using three standard error measures, i.e., mean absolute error (MAE), root mean squared error (RMSE), and R-squared (R^2). These error measures are common descriptive statistics that show how close the predictions are to the actual values. Mathematically, they are described as

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} \left| X_{t,h} - \hat{X}_{t,h} \right|$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(X_{t,h} - \hat{X}_{t,h} \right)^2}$$

$$R^2 = \frac{\sum_{t=1}^{N} (\hat{X}_{t,h} - \bar{X}_h)^2}{\sum_{t=1}^{N} (X_{t,h} - \bar{X}_h)^2}$$

where

- N is the number of observations in the out-of-sample forecast period.
- $X_{t,h}$ and $\hat{X}_{t,h}$ represents the actual and forecast O_3 values of the tth ($t = 1, 2, \ldots 365$) day and hth ($h = 1, 2, \ldots 24$) hour.
- $\bar{X}_h$ denotes the mean O_3 value of the hth hour.

3.3. Results

A comparison of the forecasting accuracy and model performance using R^2, MAE, and RMSE as the evaluation criteria is shown in Table 2 and depicted in Figure 4. The results indicate that the FAR(1) model achieved the best performance, as it had the highest R^2 value and the lowest error values. The second-best model was the VAR model, followed by the ARIMA, NNAR, and RF models. The SVM model had the worst performance among the six models.

Table 2. One-day-ahead out-of-sample forecast results for O_3 concentration.

Method	MAE	RMSE	R^2
FAR(1)	**0.0066**	**0.0090**	**0.7971**
ARIMA(5,0,0)	0.0077	0.0101	0.7429
VAR(5)	0.0069	0.0092	0.7827
NNAR	0.0078	0.0102	0.7375
SVM	0.0085	0.0116	0.6693
RF	0.0078	0.0106	0.7136
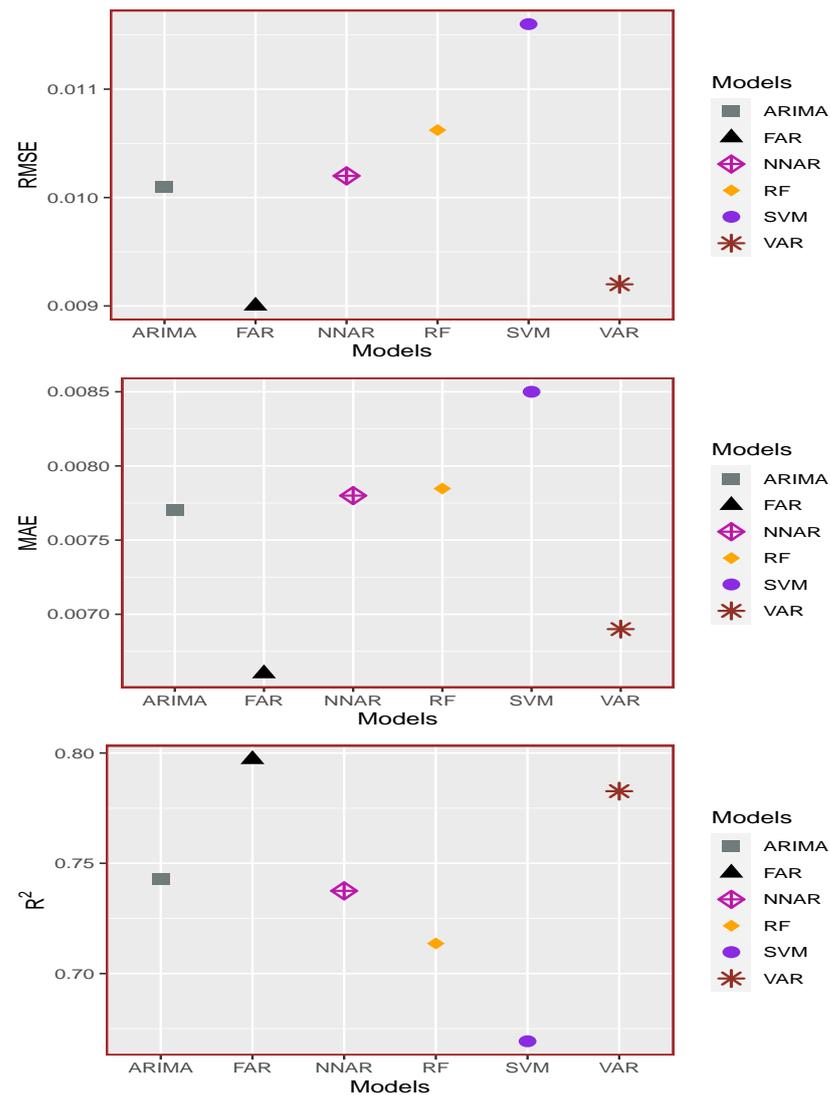
Figure 4. One-day-ahead out-of-sample forecasts for O_3 concentration: (**top**) RMSE, (**middle**) MAE, and (**bottom**) R^2 values.

The O_3 concentration forecasting performance for each model for different days of the week is summarized in Table 3 and illustrated in Figure 5. These errors represent the average errors for the whole year when calculated for different days of the week. From these results, one can notice that the proposed model outperforms the other models by producing the lowest RMSE and MAE values for every day of the week. The forecast errors are different for each day of the week, with the lowest values on Wednesday, for which the O_3 concentration is more stable than the rest of the days, and the highest values on Saturday and Sunday. The SVM model is the worst among the six models, as it has the highest RMSE and MAE values for every day.

Table 4 compares the hourly O_3 forecasting errors of six different models using the RMSE and MAE metrics. The results show that the FAR(1) is the most accurate for most hours, except for the first five hours, when the VAR model performs better. The O_3 levels and the errors change throughout the day due to various factors such as weather, traffic, and emissions. The errors are usually higher during the peak hours (14–18) and the late hours (21–24), and lower during the off-peak hours (1–4). The VAR model is slightly better than the ARIMA, NNAR, and SVM models, but still worse than the FAR(1) model. Again, the SVM model has the highest errors for most hours and is the least accurate model. These results are also depicted in Figure 6.
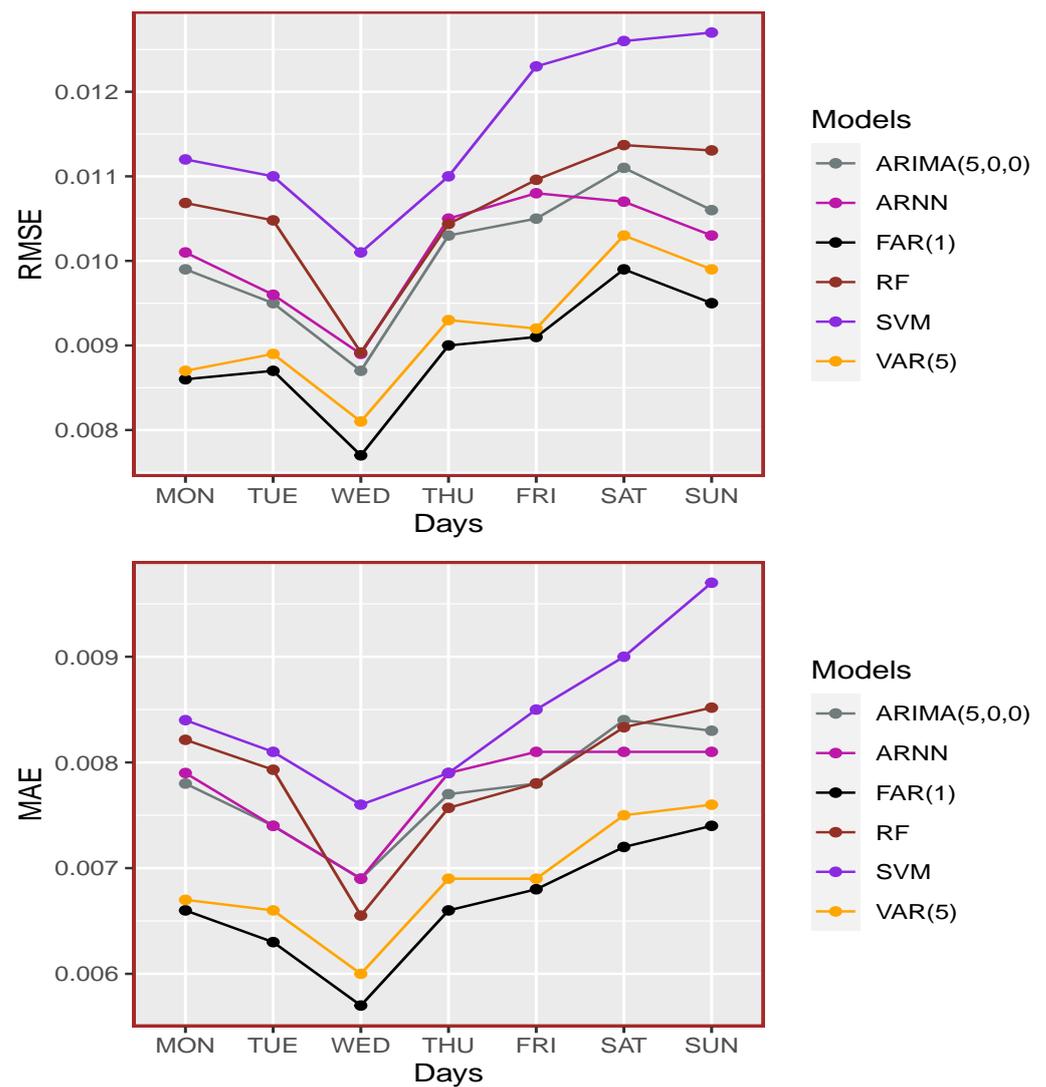
Figure 5. Day-specific RMSE and MAE values for O_3 concentration forecasting.

Table 3. Day-specific forecasting errors of all models.

| Method	Error	Day of the Week						
		Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
FAR(1)		**0.0086**	**0.0087**	**0.0077**	**0.0090**	**0.0091**	**0.0099**	**0.0095**
ARIMA		0.0099	0.0095	0.0087	0.0103	0.0105	0.0111	0.0106
VAR	RMSE	0.0087	0.0089	0.0081	0.0093	0.0092	0.0103	0.0099
NNAR		0.0101	0.0096	0.0089	0.0105	0.0108	0.0107	0.0103
RF		0.0107	0.0105	0.0089	0.0104	0.0110	0.0114	0.0113
SVM		0.0112	0.0110	0.0101	0.0110	0.0123	0.0126	0.0127
FAR(1)		**0.0066**	**0.0063**	**0.0057**	**0.0066**	**0.0068**	**0.0072**	**0.0074**
ARIMA		0.0078	0.0074	0.0069	0.0077	0.0078	0.0084	0.0083
VAR	MAE	0.0067	0.0066	0.0060	0.0069	0.0069	0.0075	0.0076
NNAR		0.0079	0.0074	0.0069	0.0079	0.0081	0.0081	0.0081
RF		0.0082	0.0079	0.0066	0.0076	0.0078	0.0083	0.0085
SVM		0.0084	0.0081	0.0076	0.0079	0.0085	0.0090	0.0097
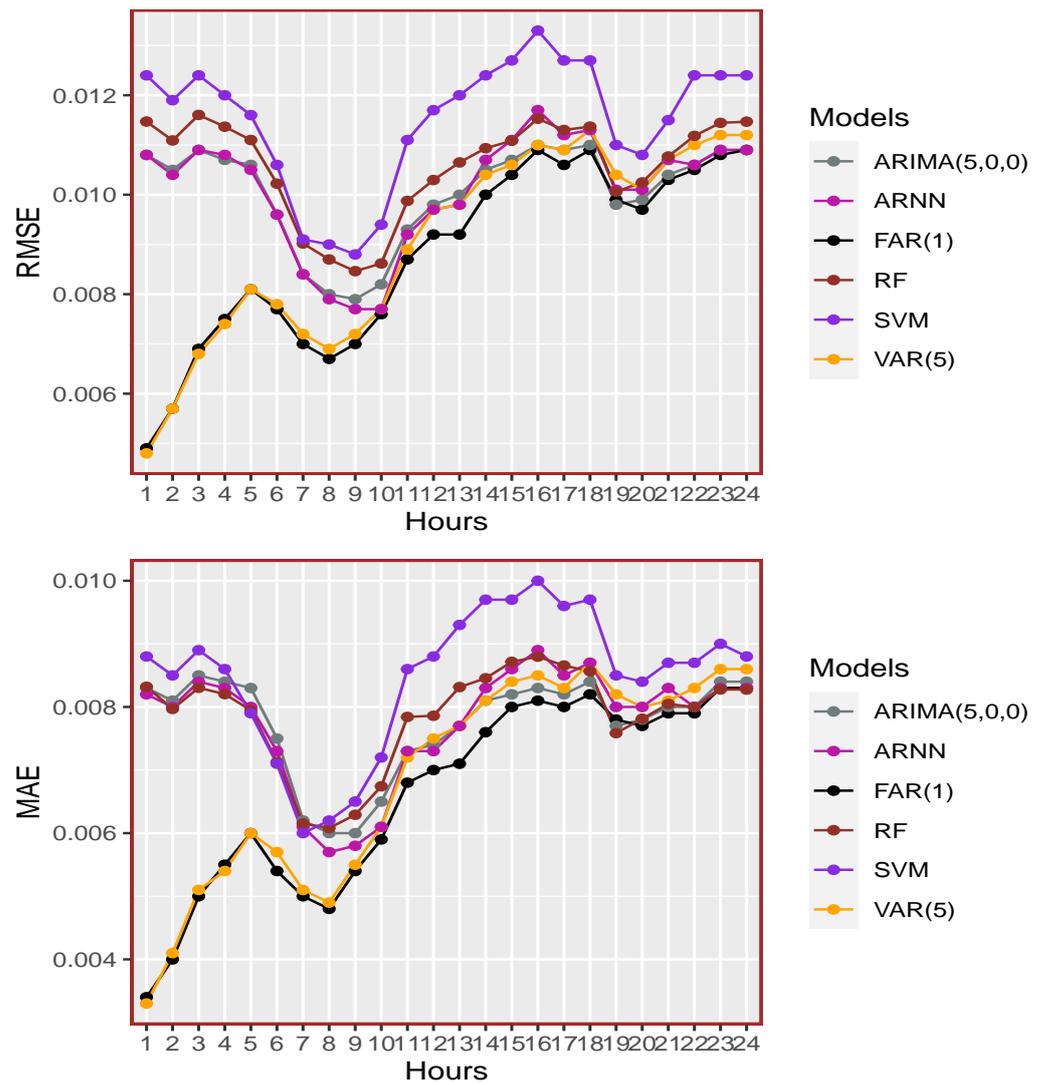
Figure 6. One-day-aheadout-of-sample hour-specific RMSE and MAE for O_3 concentration.

Table 4. Hour-specific forecasting errors for O_3 concentration.

Hour	Errors	FAR(1)	ARIMA	VAR	NNAR	RF	SVM
1	RMSE	0.0049	0.0108	**0.0048**	0.0108	0.0115	0.0124
	MAE	0.0034	0.0083	**0.0033**	0.0082	0.0083	0.0088
2	RMSE	**0.0057**	0.0105	**0.0057**	0.0104	0.0111	0.0119
	MAE	**0.0040**	0.0081	0.0041	0.0080	0.0080	0.0085
3	RMSE	0.0069	0.0109	**0.0068**	0.0109	0.0116	0.0124
	MAE	**0.0050**	0.0085	0.0051	0.0084	0.0083	0.0089
4	RMSE	0.0075	0.0107	**0.0074**	0.0108	0.0114	0.0120
	MAE	0.0055	0.0084	**0.0054**	0.0083	0.0082	0.0086
5	RMSE	**0.0081**	0.0106	**0.0081**	0.0105	0.0111	0.0116
	MAE	**0.0060**	0.0083	**0.0060**	0.0080	0.0079	0.0079
6	RMSE	**0.0077**	0.0096	0.0078	0.0096	0.0102	0.0106
	MAE	**0.0054**	0.0075	0.0057	0.0073	0.0071	0.0071

Table 4. *Cont.*

Hour	Errors	FAR(1)	ARIMA	VAR	NNAR	RF	SVM
7	RMSE	**0.0070**	0.0084	0.0072	0.0084	0.0090	0.0091
	MAE	**0.0050**	0.0062	0.0051	0.0061	0.0062	0.0060
8	RMSE	**0.0067**	0.0080	0.0069	0.0079	0.0087	0.0090
	MAE	**0.0048**	0.0060	0.0049	0.0057	0.0061	0.0062
9	RMSE	**0.0070**	0.0079	0.0072	0.0077	0.0085	0.0088
	MAE	**0.0054**	0.0060	0.0055	0.0058	0.0063	0.0065
10	RMSE	**0.0076**	0.0082	0.0077	0.0077	0.0086	0.0094
	MAE	**0.0059**	0.0065	0.0061	0.0061	0.0067	0.0072
11	RMSE	**0.0087**	0.0093	0.0089	0.0092	0.0099	0.0111
	MAE	**0.0068**	0.0073	0.0072	0.0073	0.0078	0.0086
12	RMSE	**0.0092**	0.0098	0.0097	0.0097	0.0103	0.0117
	MAE	**0.0070**	0.0074	0.0075	0.0073	0.0079	0.0088
13	RMSE	**0.0092**	0.0100	0.0098	0.0098	0.0106	0.0120
	MAE	**0.0071**	0.0077	0.0077	0.0077	0.0083	0.0093
14	RMSE	**0.0100**	0.0105	0.0104	0.0107	0.0109	0.0124
	MAE	**0.0076**	0.0081	0.0081	0.0083	0.0085	0.0097
15	RMSE	**0.0104**	0.0107	0.0106	0.0111	0.0111	0.0127
	MAE	**0.0080**	0.0082	0.0084	0.0086	0.0087	0.0097
16	RMSE	**0.0109**	0.0110	0.0110	0.0117	0.0115	0.0133
	MAE	**0.0081**	0.0083	0.0085	0.0089	0.0088	0.0100
17	RMSE	**0.0106**	0.0109	0.0109	0.0112	0.0113	0.0127
	MAE	**0.0080**	0.0082	0.0083	0.0085	0.0087	0.0096
18	RMSE	**0.0109**	0.0110	0.0113	0.0113	0.0114	0.0127
	MAE	**0.0082**	0.0084	0.0087	0.0087	0.0086	0.0097
19	RMSE	**0.0099**	0.0098	0.0104	0.0101	0.0101	0.0110
	MAE	**0.0078**	0.0077	0.0082	0.008	0.0076	0.0085
20	RMSE	**0.0097**	0.0099	0.0101	0.0101	0.0102	0.0108
	MAE	**0.0077**	0.0078	0.0080	0.0080	0.0078	0.0084
21	RMSE	**0.0103**	0.0104	0.0107	0.0107	0.0108	0.0115
	MAE	**0.0079**	0.0080	0.0081	0.0083	0.0080	0.0087
22	RMSE	**0.0105**	0.0106	0.0110	0.0106	0.0112	0.0124
	MAE	**0.0079**	0.0080	0.0083	0.0080	0.0080	0.0087
23	RMSE	**0.0108**	0.0109	0.0112	0.0109	0.0114	0.0124
	MAE	**0.0083**	0.0084	0.0086	0.0083	0.0083	0.0090
24	RMSE	**0.0109**	0.0109	0.0112	0.0109	0.0115	0.0124
	MAE	**0.0083**	0.0084	0.0086	0.0083	0.0083	0.0088

Finally, Table 5 lists the month-specific RMSE and MAE values for the models used in the study to forecast the one-day-ahead O_3 concentration, which are further depicted in Figure 7. These errors are calculated by averaging the errors month-wise for the year 2017. The results show that the forecasting errors are not uniform across the months. The errors tend to be higher in February, March, April, and October, which implies that O_3 is more challenging to predict in these months. Again, the FAR(1) model performs the best in most months, except for April, where the VAR model is slightly better, and February, where both models have equal errors. Again, the SVM model has the worst performance in all months, which suggests that it is not good for O_3 forecasting. The ARIMA, NNAR, and RF models have comparable performance but are inferior to the FAR(1) models. The FAR(1) model has the lowest RMSE and MAE values in December, when the O_3 level is the lowest.
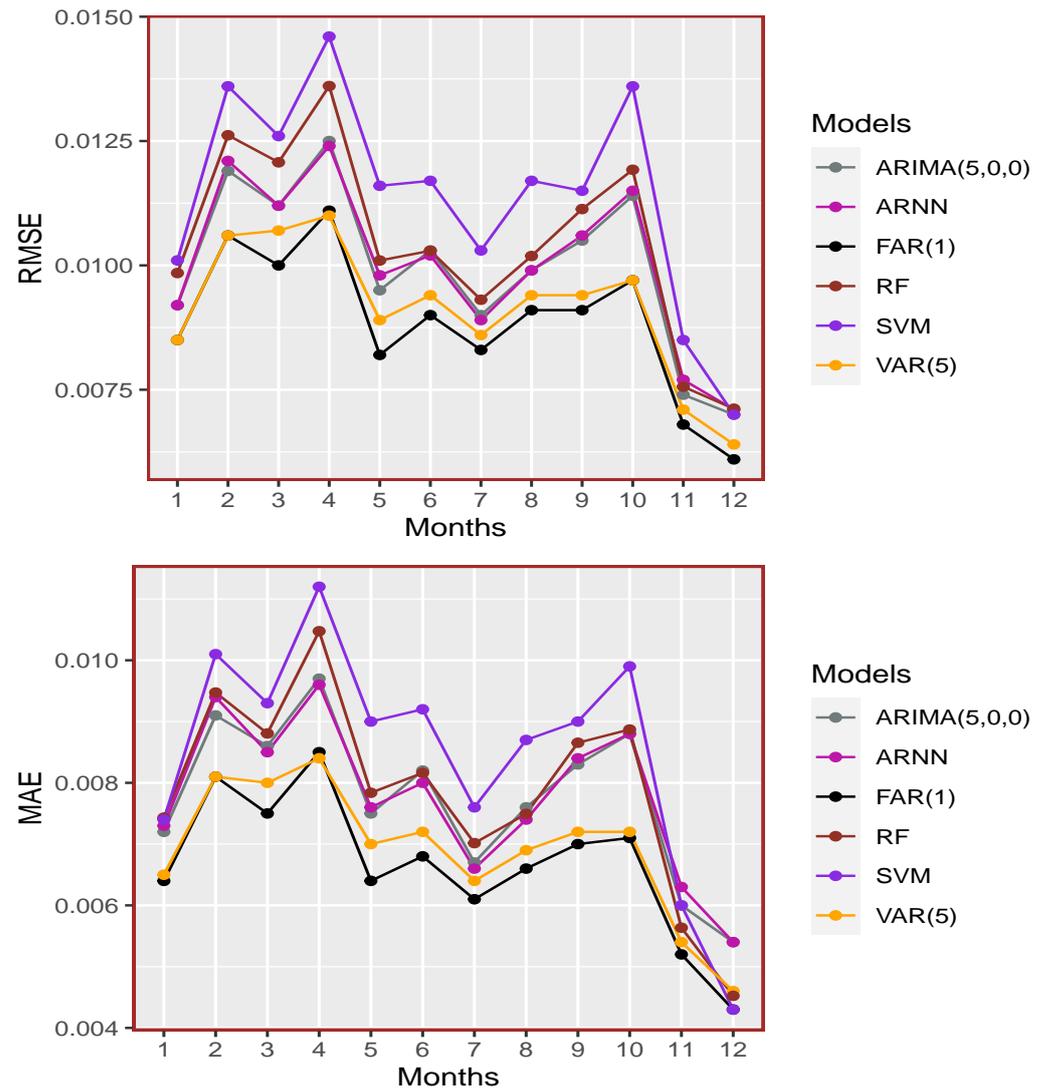
Figure 7. Out-of-sample month-specific RMSE and MAE for O_3 concentration.

Table 5. Month-specific forecasting errors of all models.

Month	Errors	FAR(1)	ARIMA	VAR	NNAR	RF	SVM
January	RMSE	**0.0085**	0.0092	0.0085	0.0092	0.0098	0.0101
	MAE	**0.0064**	0.0072	0.0065	0.0073	0.0074	0.0074
February	RMSE	**0.0106**	0.0119	**0.0106**	0.0121	0.0126	0.0136
	MAE	**0.0081**	0.0091	**0.0081**	0.0094	0.0095	0.0101
March	RMSE	**0.0100**	0.0112	0.0107	0.0112	0.0121	0.0126
	MAE	**0.0075**	0.0086	0.0080	0.0085	0.0088	0.0093
April	RMSE	0.0111	0.0125	**0.0110**	0.0124	0.0136	0.0146
	MAE	0.0085	0.0097	**0.0084**	0.0096	0.0105	0.0112
May	RMSE	**0.0082**	0.0095	0.0089	0.0098	0.0101	0.0116
	MAE	**0.0064**	0.0075	0.0070	0.0076	0.0078	0.0090
June	RMSE	**0.0090**	0.0103	0.0094	0.0102	0.0103	0.0117
	MAE	**0.0068**	0.0082	0.0072	0.0080	0.0082	0.0092

Table 5. *Cont.*

Month	Errors	FAR(1)	ARIMA	VAR	NNAR	RF	SVM
July	RMSE	**0.0083**	0.0090	0.0086	0.0089	0.0093	0.0103
	MAE	**0.0061**	0.0067	0.0064	0.0066	0.0070	0.0076
August	RMSE	**0.0091**	0.0099	0.0094	0.0099	0.0102	0.0117
	MAE	**0.0066**	0.0076	0.0069	0.0074	0.0075	0.0087
September	RMSE	**0.0091**	0.0105	0.0094	0.0106	0.0111	0.0115
	MAE	**0.0070**	0.0083	0.0072	0.0084	0.0087	0.0090
October	RMSE	**0.0097**	0.0114	0.0097	0.0115	0.0119	0.0136
	MAE	**0.0071**	0.0088	0.0072	0.0088	0.0089	0.0099
November	RMSE	**0.0068**	0.0074	0.0071	0.0077	0.0076	0.0085
	MAE	**0.0052**	0.0060	0.0054	0.0063	0.0056	0.0060
December	RMSE	**0.0061**	0.0070	0.0064	0.0071	0.0071	0.0070
	MAE	**0.0043**	0.0054	0.0046	0.0054	0.0045	0.0043

4. Conclusions

Air pollution, particularly ground-level O_3, is a global issue with severe implications for both human health and ecosystems. It reduces agricultural output, exacerbates global warming as a greenhouse gas, and affects respiratory health, causing symptoms such as coughing, chest tightness, and worsening asthma. Given the severity of these consequences, accurate forecasting of ground-level O_3 concentrations is crucial. This study aims to use a functional approach to model ground-level O_3 concentrations, a method less explored in the literature. The performance of this approach is compared with traditional time series and machine learning models, including ARIMA, VAR, NNAR, RF, and SVM. Hourly data from Los Angeles, collected from 2013 to 2017, are used, with 80% allocated for model estimation and the remaining 20% for one-day-ahead out-of-sample forecasting. The performance of the model is evaluated using R^2, MAE, and RMSE.

The results revealed that the FAR(1) model performs better than other models included in the study by producing lower forecasting errors. The second-best model was the VAR model, followed by the ARIMA, NNAR, and RF models. The SVM model performed the worst among all models. These findings highlight the importance of choosing a suitable model for accurate O_3 forecasting, which is crucial for mitigating the impacts of air pollution. In future research, including exogenous variables, including temperature, humidity, wind speed, solar radiation, and other meteorological factors, may improve the forecasting accuracy. As the current study is based on the Los Angeles dataset, the performance of the proposed approach could be assessed by conducting a study on other site datasets.

Note

1   https://aqs.epa.gov/aqsweb/airdata (accessed on 12 March 2023).

References

Aneiros-Pérez, Germán, Herve Cardot, Graciela Estévez-Pérez, and Philippe Vieu. 2004. Maximum ozone concentration forecasting by functional non-parametric approaches. *Environmetrics* 15: 675–85. [CrossRef]

Arsić, Milica, Ivan Mihajlović, Djordje Nikolić, Živan Živković, and Marija Panić. 2020. Prediction of ozone concentration in ambient air using multilinear regression and the artificial neural networks methods. *Ozone: Science & Engineering* 42: 79–88.

Bosq, Denis. 2000. *Linear Processes in Function Spaces: Theory and Applications*. Berlin/Heidelberg: Springer Science & Business Media, vol. 149.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45: 5–32. [CrossRef]

Didericksen, Devin, Piotr Kokoszka, and Xi Zhang. 2012. Empirical properties of forecasts with the functional autoregressive model. *Computational Statistics* 27: 285–98. [CrossRef]

Eslami, Ebrahim, Yunsoo Choi, Yannic Lops, and Alqamah Sayeed. 2020. A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Computing and Applications* 32: 8783–97. [CrossRef]

Gao, Meng, Liting Yin, and Jicai Ning. 2018. Artificial neural network model for ozone concentration estimation and monte carlo analysis. *Atmospheric Environment* 184: 129–39. [CrossRef]

Ghoneim, Osama, Doreswamy, and Busnur Rachotappa Manjunatha. 2017. Forecasting of ozone concentration in smart city using deep learning. Paper presented at the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, September 13–16. pp. 1320–26.

Hashim, NurIzzah M., Norazian Mohamed Noor, Ahmad Zia Ul-Saufie, Andrei Victor Sandu, Petrica Vizureanu, György Deák, and Marwan Kheimi. 2022. Forecasting daytime ground-level ozone concentration in urbanized areas of malaysia using predictive models. *Sustainability* 14: 7936. [CrossRef]

Hong, Fugui, Cheng Ji, Jingzhi Rao, Chang Chen, and Wei Sun. 2023. Hourly ozone level prediction based on the characterization of its periodic behavior via deep learning. *Process Safety and Environmental Protection* 174: 28–38. [CrossRef]

Jan, Faheem, Ismail Shah, and Sajid Ali. 2022. Short-term electricity prices forecasting using functional time series analysis. *Energies* 15: 3423. [CrossRef]

Kumar, Krishan, A. K. Yadav, M. P. Singh, H. Hassan, and V. K. Jain. 2004. Forecasting daily maximum surface ozone concentrations in brunei darussalam—An arima modeling approach. *Journal of the Air & Waste Management Association* 54: 809–14.

Lotrecchiano, Nicoletta, Daniele Sofia, Aristide Giuliano, Diego Barletta, and Massimo Poletto. 2020. Pollution dispersion from a fire using a gaussian plume model. *International Journal of Safety and Security Engineering* 10: 431–39. [CrossRef]

Mavroidis, Ilias, and M. Ilia. 2012. Trends of nox, no2 and o3 concentrations at three different types of air quality monitoring stations in Athens, Greece. *Atmospheric Environment* 63: 135–47. [CrossRef]

Mills, Gina, Harry Harmens, Serena Wagg, Katrina Sharps, Felicity Hayes, David Fowler, Mark Sutton, and Bill Davies. 2016. Ozone impacts on vegetation in a nitrogen enriched and changing climate. *Environmental Pollution* 208: 898–908. [CrossRef] [PubMed]

Munshed, Mohammad, Jesse Van Griensven Thé, and Roydon Fraser. 2023. Methodology for mobile toxics deterministic human health risk assessment and case study. *Atmosphere* 14: 506. [CrossRef]

Nuvolone, Daniela, Davide Petri, and Fabio Voller. 2018. The effects of ozone on human health. *Environmental Science and Pollution Research* 25: 8074–88. [CrossRef] [PubMed]

Paoletti, Elena, and William J Manning. 2007. Toward a biologically significant and usable standard for ozone that will also protect plants. *Environmental Pollution* 150: 85–95. [CrossRef] [PubMed]

Rahman, Azizur, and N. M. Refat Nasher. 2024. Forecasting hourly ozone concentration using functional time series model—A case study in the coastal area of bangladesh. *Environmental Modeling & Assessment* 29: 125–34.

Ramsay, James O. 1982. When the data are functions. *Psychometrika* 47: 379–96. [CrossRef]

Salazar, Ledys, Orietta Nicolis, Fabrizio Ruggeri, Jozef Kisel'ák, and Milan Stehlík. 2019. Predicting hourly ozone concentrations using wavelets and arima models. *Neural Computing and Applications* 31: 4331–40. [CrossRef]

Seinfeld, John H., and Spyros N. Pandis. 2016. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Hoboken: John Wiley & Sons.

Shah, Ismail, Izhar Muhammad, Sajid Ali, Saira Ahmed, Mohammed M. A. Almazah, and A. Y. Al-Rezami. 2022. Forecasting day-ahead traffic flow using functional time series approach. *Mathematics* 10: 4279. [CrossRef]

Shumway, Robert H., David S. Stoffer, and David S. Stoffer. 2000. *Time Series Analysis and Its Applications*. Berlin/Heidelberg: Springer, vol. 3.

Su, Xiaoqian, Junlin An, Yuxin Zhang, Ping Zhu, and Bin Zhu. 2020. Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods. *Atmospheric Pollution Research* 11: 51–60. [CrossRef]

Sujith, Babu, Meena Sehgal, Karthik L. Balajee, and Rizwan Suliankatchi. 2017. Characteristics of the ozone pollution and its health effects in india. *International Journal of Medicine and Public Health* 7: 56–60.

Suraki, Mohsen Yaghoubi. 2013. Nfc: A surveillance system for preventing and controlling air pollution. Paper presented at the 2013 7th International Conference on Application of Information and Communication Technologies, Baku, Azerbaijan, October 23–25. pp. 1–4.

World Health Organization. 2020. Air Pollution. Available online: https://www.who.int/health-topics/air-pollution#tab=tab_1 (accessed on 21 September 2020).

Yafouz, Ayman, Nouar AlDahoul, Ahmed H. Birima, Ali Najah Ahmed, Mohsen Sherif, Ahmed Sefelnasr, Mohammed Falah Allawi, and Ahmed Elshafie. 2022. Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction. *Alexandria Engineering Journal* 61: 4607–22. [CrossRef]