sustainability

MDPI

Article

Long Short-Term Memory Approach for Short-Term Air Quality Forecasting in the Bay of Algeciras (Spain)

María Inmaculada Rodríguez-García [1,*], María Gema Carrasco-García [2], Javier González-Enrique [1], Juan Jesús Ruiz-Aguilar [2] and Ignacio J. Turias [1]

[1] Department of Computer Science Engineering, Algeciras School of Engineering and Technology (ASET), University of Cádiz, 11202 Algeciras, Spain
[2] Department of Industrial and Civil Engineering, Algeciras School of Engineering and Technology (ASET), University of Cádiz, 11202 Algeciras, Spain
* Correspondence: inma.rodriguezgarcia@gm.uca.es

Abstract: Predicting air quality is a very important task, as it is known to have a significant impact on health. The Bay of Algeciras (Spain) is a highly industrialised area with one of the largest superports in Europe. During the period 2017–2019, different data were recorded in the monitoring stations of the bay, forming a database of 131 variables (air pollutants, meteorological information, and vessel data), which were predicted in the Algeciras station using long short-term memory models. Four different approaches have been developed to make SO_2 and NO_2 forecasts 1 h and 4 h in Algeciras. The first uses the remaining 130 exogenous variables. The second uses only the time series data without exogenous variables. The third approach consists of using an autoregressive time series arrangement as input, and the fourth one is similar, using the time series together with wind and ship data. The results showed that SO_2 is better predicted with autoregressive information and NO_2 is better predicted with ships and wind autoregressive time series, indicating that NO_2 is closely related to combustion engines and can be better predicted. The interest of this study is based on the fact that it can serve as a resource for making informed decisions for authorities, companies, and citizens alike.

Keywords: air pollution forecasting; LSTM models; deep learning; maritime traffic; ANNs; nitrogen oxides; sulphur dioxide

check for updates

Citation: Rodríguez-García, M.I.; Carrasco-García, M.G.; González-Enrique, J.; Ruiz-Aguilar, J.J.; Turias, I.J. Long Short-Term Memory Approach for Short-Term Air Quality Forecasting in the Bay of Algeciras (Spain). *Sustainability* **2023**, *15*, 5089. https://doi.org/10.3390/su15065089

Academic Editor: Giovanni Gualtieri

Received: 31 January 2023
Revised: 9 March 2023
Accepted: 10 March 2023
Published: 13 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The motivation of this study is to deepen the study of atmospheric pollution in a region with special meteorological characteristics as well as a strong industrial and port environment using the latest trends in machine learning. Air pollution is recognised and proven to be a source of countless health problems including premature deaths (e.g., [1,2]). Salvaraji [1] explained the correlation between air pollution and cardiovascular disease in a study period of ten years, and in [2]'s review, several cohort studies analysing nitrogen oxides and particles were compiled, reporting that the pollutants were most associated with liver cancer. Air pollution particularly affects the most vulnerable populations, such as children, the elderly, and people with respiratory diseases [3]. Childhood is a critical period for brain maturation and mental development, with many factors involved, such as the inhalation of air pollutants in the air, which can affect children's healthy development. Particularly, industries release nitrogen monoxide (NO), which is transformed into nitrogen dioxide (NO_2) by reacting with oxygen in the atmosphere [4]. Nitrogen dioxide is an irritant gas that can cause respiratory problems in people with pre-existing lung diseases such as asthma, chronic bronchitis, or chronic obstructive pulmonary disease (COPD). In the short term, exposure to NO_2 can cause coughing, wheezing, shortness of breath, fatigue, and chest pain. In the long term, repeated exposure to NO_2 can worsen symptoms of chronic respiratory diseases and increase the risk of developing heart disease [5]. Sulphur dioxide (SO_2) is an irritant gas that can cause respiratory problems in people with pre-existing lung

disease, as well as in healthy people exposed to high levels of SO_2. Symptoms of acute SO_2 exposure include coughing, wheezing, shortness of breath, sore throat, and chest pain. In the long term, repeated exposure to SO_2 can worsen symptoms of chronic respiratory diseases and increase the risk of developing cardiovascular disease [6].

Since the Bay of Algeciras is European territory, Directive 2008/50/EC [7] governs the concentrations of each controlled pollutant. Further, the IMO (International Maritime Organization) in their MARPOL Treaty (Annex VI) [8] sets out all the obligations that maritime transport has to fulfill in terms of pollution to achieve the European Green Deal's aim for zero pollution, creating an environment free from harmful pollution by 2050. The municipality of Algeciras, with 122,982 (www.ine.es (20 December 2022)) inhabitants in 2022, makes walking and sports areas available next to the port, providing a strong connection between the harbour and the city.

The Port of Algeciras is located at a strategic point where two continents, the Atlantic Ocean and the Mediterranean Sea converge. For this reason, the Port of Algeciras is a perfect hub port where two container terminals operate, linking hundreds of world ports on a weekly basis. One of the problems that the Port of Algeciras faces is its weak rail connection with other points in Spain and Europe, the main reason for the increase in heavy goods traffic and consequently the increase in air pollution associated with the port. The five largest shipping companies in the world choose the Port of Algeciras to operate. Maritime transport is efficiently irreplaceable when it comes to transporting goods between continents. The duality of vessel and port dominates the transport of goods due to its low cost, huge capacity, and speed. This is why maritime transport moves, in physical terms, more than 80% of international trade [8]. The side effect of the port is that there is evidence that shipping commerce affects air pollution (e.g., [9–12]). High rates of total ship emissions can be dispersed to 400 km inland [13]. Progress on air quality studies and how shipping emissions affect human health are presented by [11], and furthermore, future research on how maritime emissions will impact coastal cities has been studied by [10]. Many researchers have discovered the importance of vessel emissions in port areas (e.g., [14–20]). It is possible to use ship information to make pollution predictions, as ships are a major source of pollution in ports. However, it would require access to a large amount of accurate and up-to-date data on ships' activities, such as their position, engine type, speed, cargo, type of fuel, etc., in order to obtain real data. In addition, it would be necessary to consider other factors, such as weather, which can also affect air pollution. The Automatic Identification System (AIS) can help to obtain or predict the position and number of vessels [21] passing through the Bay of Algeciras.

This article aims to continue with air quality studies in the Bay of Algeciras using novel computational techniques. For example, [22] compared several machine learning methods to forecast air quality, using meteorological parameters and other chemical species, while [12] have studied the estimation of PM, NO_X, and SO_2 and air quality in the Strait of Gibraltar (Spain), utilizing their own ship's energy and emissions model. The study [16] in the port of Hong Kong showed that emissions from ships account for 25% of the total mass.

In the present study, our attention is focused on SO_2 and NO_2 pollutants and the use of new deep learning models, particularly long short-term memory (LSTM), applied to forecast these pollutants in the port city of Algeciras. LSTM models are specifically designed to predict time series [23]. In order to accurately predict certain situations, LeCun [24] invented useful deep learning techniques that were pioneering for later studies and are well-known in the field of computational prediction.

Methods to predict air pollution by encompassing pollution measurements and meteorological values are presented by Korunoski [25]. Masood and Ahmad [26] used deep learning approaches to forecast particulate matter in Delhi (India). Amongst other machine learning techniques, they used LSTM for their predictions using several pollutants and meteorological data as inputs, very similar to what we expose in this work. Recently, several papers have used deep learning techniques (LSTM) to forecast air pollution (e.g., [22,26–32]). A review of the potential of architectures of deep networks to explore nonlinear spatiotem-

poral features is examined by [28]. Wang and Tang [32] have developed a combined model of air quality index forecasting based on a whale algorithm optimised using deep learning and empirical analysis, obtaining the highest prediction accuracy in their models. Drewil and Al-Bahadili [29] studied four pollutants (PM_{10}, $PM_{2.5}$, NO_X, and CO) and chose the best hyperparameters for LSTM to predict the pollution level for 1 day ahead, obtaining faster and more accurate results. Similar to what is done in our research, daily and hourly pollutant concentration forecasting within the next 30 days and 72 h, is proposed in [31]. They used LSTM to obtain the best performance and prediction of air pollutants, proving that a data-driven approach is essential for resolving air pollution. In addition to pollutant and meteorological variables, they use the distances to all point sources and highways to predict air quality using deep learning methods, (LSTM+ARIMA+CNN) [30]. After applying different machine learning and deep learning techniques in their research, [26] found that in the field of air pollution forecasting, using meteorological values and several air pollutants as inputs, the LSTM model was the best-performing technique for the prediction of pollutants. The air quality index (AQI) is obtained by averaging pollutant sensor measurements, and its prediction accuracy is proposed by [32] by means of deep learning techniques. Further, [33] proposed an LSTM-based aggregated model for air pollution forecasting. The relationship between meteorological factors and air pollution is studied in three megacities in China by Zhang [34]. One positive aspect of the Bay of Algeciras is the existence of strong winds, which produce a cleaning effect [35]. Thus, the authors of [36] have previously studied the prediction of winds in the bay using machine learning techniques.

There is work being done using LSTM networks to predict air pollution. One example is the paper [37], where the authors used an LSTM network to predict the concentration of nitrogen dioxide (NO_2) in Central London. Another example is the paper [38], where the number of features was reduced from 25 to 5, resulting in greater accuracy for a 72-h ozone (O_3) forecast. Deep [39] tackled the air pollution problem using LSTM models in South Korea. In the paper [40], the authors used an LSTM network to predict the concentrations of various pollutants in India and China, using meteorological data as inputs.

Therefore, in recent years, deep learning has become a powerful tool that has enabled breakthroughs in many fields (e.g., [41,42]). Some state-of-the-art models have recently been proposed in the literature in this field. Cheng [42] presented a model to suggest the best locations for air quality monitoring stations by designing an inference model using existing air quality monitoring data in conjunction with urban city traffic measurements. Using these two methods, Cheng's model is able to provide real-time estimates of city-wide air quality, rather than predictions of future air quality. Laña [43] specifically looked for relationships and patterns between meteorological variables, urban traffic variables, and pollution using regression models. Zheng [44] predicted air quality in the next 48 h, using a data-driven method that combines a temporal predictor based on linear regression and a spatial predictor based on artificial neural networks (ANNs). However, this method does not consider the long-term temporal influence, the possible spatial correlation at different scales, and different weather patterns. In the work of [45], data are clustered according to weather patterns, and spatial and temporal correlations are subsequently captured using an ensemble of deep nets. Our work is framed along similar lines as the studies mentioned above. More research related to air pollution in the Bay of Algeciras can be found in [46–50]. A comprehensive statistical analysis and calculation of risks and trends in the bay have been analysed by [47]. The authors have previously developed different papers focused on air pollution forecasting using machine learning [46–50], highlighting the last studies based on deep learning techniques [46,48]. On the other hand, it is worth mentioning the work of [50], comparing several feature selection methods using LSTM techniques to forecast air pollution using only the relevant exogenous variables. In addition, other authors [51,52] mix several deep learning techniques to forecast air pollution.

This manuscript is organised in sections as follows. The introduction is the Section 1 where the state of the art is presented. The Section 2 is the materials and methods section where the database, study area, used methods, and experimental procedure are explained. The Section 3 presents the tables and figures of the obtained results. The Section 4 presents the discussion and, finally, the conclusions that have been drawn.

2. Materials and Methods

2.1. Materials

Database

Algeciras, like many port cities, is surrounded by a road network saturated with road traffic, including the abundant truck traffic of goods, mainly containers. The Port of Algeciras borders residential areas, and it is integrated into the city. The port is potentially a major contributor to the city's air pollution. The Andalusian government (Junta de Andalucía) has established the Air Quality Monitoring and Control Network of Andalusia, where monitoring stations are located in the region (see Figure 1 and Table 1), although only two stations are maintained in the city of Algeciras: "Algeciras" (MS1) and "Rinconcillo" (MS2). MS1 is right on the premises of the Algeciras School of Engineering and Technology (ASET) in the city centre and at a distance of about 500 m from the first mooring lines of the ships, and MS2 is located on the beach of the same name on the outskirts of the city. In addition, the network is completed by five atmospheric sensors, also owned by the Andalusian government (W1–5) where wind speed, wind direction, atmospheric pressure, solar radiation, relative humidity, rainfall, and temperature are collected (see Table 2). All the pollutants observed in this study are recorded indistinctly in several stations, as Table 3 shows. It consists of hourly concentrations ($\mu g/m^3$) of pollutants databases during the period of 2017 to 2019.

Simultaneously, vessel movements within the port are collected by the Algeciras Bay Port Authority using the AIS mentioned above. Either AIS is used and sensors are placed very close to the port or only "background" pollution can only be obtained [21]. Annually, nearly 30,000 ships visit this port. The vessels database is transformed into gross tonnage per hour (GT/h).



(**a**) (**b**)

Figure 1. (**a**) Location of the study area in Europe. Andalusia (the South of Spain). The Bay of Algeciras in Cádiz is highlighted in a bold circle. (**b**) Monitoring stations (MS1–MS16/W1–5) are spread over the Bay.

Table 1. Pollutant monitoring stations (MS) and weather stations (W).

MS Codes	MS Names
MS1	Algeciras
MS2–MS16	The rest of the monitoring stations
W1–5	Weather stations

To analyse these data with fundamentally different temporal scales and types, the data are organised into hourly databases, visualized, and analysed to quantify the variation in local air quality according to shipping and road traffic movements, and the extent to which this is modulated by weather effects. Our database is formed by a total of 131 variables, including hourly concentration data of pollutants, atmospheric values, and vessels in GT/h. Tables 2 and 3 list the 24 meteorological variables and the pollutants recorded in each monitoring station, respectively. Firstly, the databases were merged, preprocessed, outliers removed, and missing data imputed using ANNs as the authors have successfully used in previous research. Preprocessing consisted of the normalisation of each variable. On the other hand, the data for all variables underwent a thorough review, and any values that appeared highly unusual or excessively large or small (including negative values) were identified as outliers and subsequently removed from the database. Finally, an imputation procedure based on ANNs was applied in order to fill in missing values. This procedure was developed by the authors in a previous work and used by them in [46–50,53]. Each variable is estimated using a different ANN trained using available data.

Table 2. Meteorological variables recorded in weather stations in the Bay from 2017 to 2019.

Atmospheric Values	Weather Stations Code	Units
WS	W1–4	Wind speed (km/h)
WD	W1, W3–5	Wind direction (degrees °)
SR	W2,W4	Solar radiation (W/m^2)
T	W1–4	Temperature (°C)
AP	W1–3	Atmospheric relative pressure (hPa)
RF	W1–2, W4–5	Rainfall (l/m^2)
RH	W1–2, W4	Relative humidity (%)

Table 3. Pollutants collected in monitoring stations in the bay from 2017 to 2019.

Pollutants	Monitoring Stations Code	Units
SO_2	MS1–MS16	$\mu g/m^3$
O_3	MS1, MS2, MS3, MS5, MS8, MS9, MS14, MS15	
NO_X	MS1–MS16	
NO	MS1–MS16	
$PM_{2.5}$	MS3, MS4, MS7, MS8, MS9, MS10, MS11, MS13, MS15, MS16	
PM_{10}	MS1, MS2, MS4, MS6, MS8, MS11, MS12, MS13, MS14	
NO_2	MS1–MS16	
Ethylbenzene	MS3, MS9, MS15	
CO	MS1, MS3, MS9, MS15, MS16	
Toluene	MS1, MS3, MS9, MS10, MS15	
Benzene	MS1, MS3, MS9, MS10, MS15	

2.2. Methods

2.2.1. Objective

This work aims to provide reliable forecasts of air quality in a port city by making 1 h and 4 h predictions of pollutant concentrations using deep learning, specifically LSTM. Several approaches are developed to learn whether autoregressive time series, including vessels and wind data, improve the predictions or not. This interest is analysed in short-term forecasts so that citizens can decide, for example, whether to go out to partake in sports in areas close to the port, whether pupils from a nearby school can go outside to play in the playground, or other such decisions.

2.2.2. Quality Indexes

The evaluation of the generalisation capabilities is done by means of Pearson's correlation coefficient (R), the mean absolute error (MAE), and the mean squared error (MSE). The best model is the one with the highest R and the lowest MAE and MSE. Equations (1)–(3) show these indexes, where T are the tested values and P the predicted values.

$$R = \frac{\sum_{j=1}^{N}(T_j - T)\cdot(P_j - P)}{\sqrt{\sum_{j=1}^{N}(T_j - T)^2 \cdot \sum_{j=1}^{N}(P_j - P)^2}} \quad (1)$$

$$MAE = \frac{1}{N}\sum_{j=1}^{N}|P_j - T_j| \quad (2)$$

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(P_j - T_j)^2 \quad (3)$$

The index of agreement (d), introduced by Willmott in 1982 [54], is a measure of agreement between observed and predicted values, ranging from 0 (no agreement) to 1 (perfect agreement). This index quantifies the degree of model prediction error by taking into account both the potential error and the ratio of the mean square error. Equation (4) shows the expression of this index, where O_j represents the vector of observation data, P_j the vector of predicted values, and $\overline{O}$ the vector of the average of observed values. However, it is worth noting that the index d is highly sensitive to extreme values due to its squared nature [55,56].

$$d = 1 - \frac{\sum_{j=1}^{n}(O_j - P_j)^2}{\sum_{j=1}^{n}(|P_j - \overline{O}| + |O_j - \overline{O}|)^2} \ , \ 0 \le d \le 1 \quad (4)$$

2.2.3. Long Short-Term Memory (LSTM)

To accomplish this goal, the utilization of LSTM models has been employed. LSTM was developed by [23], who introduced a highly effective gradient-based technique that eliminates the problem of lengthy processes and insufficient decay error backflow, which are inherent in LSTM. LSTM networks are a specific type of recurrent neural network (RNN) that have the capability to understand and process long-term dependencies within sequential data time units. By introducing memory cells, the memory can be sustained for extended periods. However, traditional RNNs are limited in their capacity to understand longer-term dependencies, as they are usually trained using backpropagation, which can be hindered by the "vanishing" or "exploding" gradient issue. These problems cause the network weights to become too small or too large, restricting the network's effectiveness in applications that require it to learn long-term relationships. To overcome this problem, LSTM networks incorporate additional gates.

The main elements of an LSTM network consist of a stream input layer and an LSTM layer. The input stream layer feeds sequential or time series data into the network. Figure 2 illustrates the architecture of a straightforward LSTM network for regression. The network commences with an input layer of sequences, followed by an LSTM layer. The network concludes with a fully connected layer and an output regression layer.



Figure 2. LSTM network architecture.

Figure 3 depicts the architecture of a general LSTM where H_{i1}, $H_{i2} \ldots H_{iS}$ are vectors of hidden units and X_{i1}, $X_{i1} \ldots$ X_{iS} are the vectors of features. This diagram illustrates the flow of X, which are the time series with C channels (or features), and S represents the length through an LSTM layer. All the vectors X_{iS} represent the number of time steps.

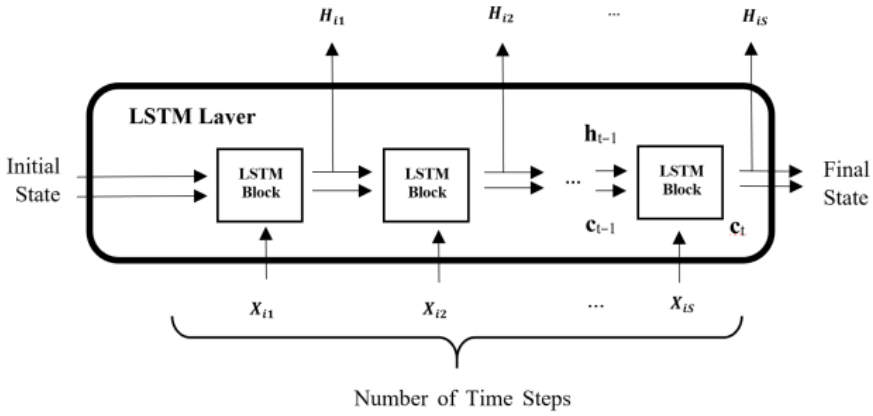


Figure 3. General LSTM architecture.

The initial LSTM block uses the first time unit and the initial grid state to calculate the first output and updated cell state. At each subsequent time unit t, the block utilizes the current network state (c_{t-1}, h_{t-1}) and the next time unit in the sequence to calculate the output and updated cell state c_t. The LSTM's hidden state at time t captures the output of the layer for this time unit, while the cell state stores information learned from the previous time units. The layer modifies the cell state by adding or removing information at each time unit using specialised gates. These gates allow the LSTM to selectively control what information is exported as output and passed on to the next hidden state h_{t-1}. To capture long-term relationships, the LSTM employs additional gates.

The LSTM architecture comprises a sequential arrangement of four neural networks and specialised memory cells for storing information. To manage the stored memory, three distinct gates are employed: the forget, update, and output gates (as depicted in Figure 4). The gates receive two inputs, x_t (input at a specific time) and h_{t-1} (previous cell output), which are then subjected to weight matrix multiplication and addition of bias. The resulting output is then passed through an activation function that generates a binary value. If the output value is 0 for a particular cell state, then the information is forgotten, whereas an output value of 1 means that the information is retained for future use. The input gate (i) regulates the useful information that enters the cell state by filtering the values through a sigmoid function. An additional activation function, using the inputs h_{t-1} and x_t, is used to control the amount of information that is allowed through. Finally, the controlled values are multiplied by the values of the vector to generate an output, which is then passed to the next cell. Any information that is deemed no longer useful is erased by the forget gate.

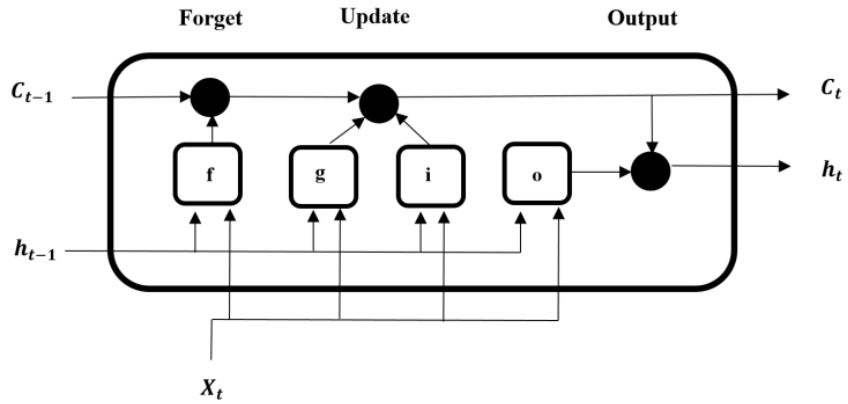


Figure 4. This diagram depicts the data flow during time step t and visualizes the mechanisms by which the gates modify the hidden states of the cell. x_t: input, h_t: hidden state; C_t: cell state, f: forget gate, g: memory cell, i: input gate, o: output gate.

In Figure 4, the output (also referred to as the hidden state) and the cell state at time unit t are determined by h_t and c_t, respectively. The sigmoid function (σ) is used in every gate to obtain only positive values, while negative features are discarded. A value of 0 indicates that the gate has blocked the current value, and 1 means that the gate is allowing the features to pass through it. Equations (5)–(7), shown below, are the conventional LSTM equations for the gates, where i_t represents the input gate, f_t is the forget gate, and o_t represents the output gate. The weights for the respective neurons and gates are denoted by w_x, where x_t represents the inputs at the current timestamp and h_{t-1} represents the outputs of the previous LSTM block at timestamp $t-1$. Finally, b_x represents the biases for the respective gates.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{5}$$

$$f_t = \sigma\Big(w_f[h_{t-1}, x_t] + b_f\Big) \tag{6}$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{7}$$

Furthermore, each unit within the LSTM cell has its equation shown in Equations (8)–(10). Equation (10) is the cell state or memory equation at t, and $\widetilde{c}_t$ is the candidate cell state, where h_t is the final output. The candidate must be calculated using Equation (8) to obtain c_t using Equation (9), the memory vector for the current t.

$$\widetilde{c}_t = \sigma(w_c[h_{t-1}, x_t] + b_c) \tag{8}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \widetilde{c}_t \tag{9}$$

$$h_t = o_t \cdot \sigma\big(c^t\big) \tag{10}$$

2.2.4. LSTM Hyperparameters

Hyperparameters are adjustable settings that allow the training process of a model to be controlled. In this LSTM approach, the hyperparameters are calculated following the methodology used by [57] and refined years later by [58]:

- *L*2 Regularisation factor (*L*2*r* or λ) is a weight loss non-negative scaling which multiplies the regularisation function $\Omega(w)$ and accompanies the weight loss function $\mathrm{E}(\theta)$, (see Equation (11)) where $\mathrm{E}_R(\theta)$ is the regularised weight loss function, w is the weight vector, and n is the number of terms (Equation (12)).

$$\mathrm{E}_R(\theta) = \mathrm{E}(\theta) + \lambda\Omega(w) \tag{11}$$

$$L_2Reg = \sum_1^n w^2 = w_1^2 + w_2^2 + \ldots + w_n^2 \tag{12}$$

- Learning rate (*Lr*) in LSTM networks refers to the amount by which the optimiser's learning rate is reduced during training. *Lr* is a learning rate schedule that decreases over time as learning progresses. In the beginning, the rate is higher in order to be able to perform the weight distribution more correctly and bias the LSTM cells. As the model improves the fit of the training data, its updates must be smaller to fine-tune the model parameters. The speed of learning rate reduction is determined by the reduction factor [58,59].
- Learn rate drop factor (*Lrdrop*), a scalar from 0 to 1, is used to reduce the learning rate by a certain factor at specific intervals during training, helping to converge more efficiently and quickly by reducing the learning rate as the network approaches a good solution [56,58,59].
- Minibatch size (*Minib*) is where, instead of using all training data in each iteration, the dataset is divided into minibatches to update the weights of the network in one training iteration, increasing its efficiency. It is a small set of training examples of positive integers that helps to avoid overfitting [56,58,59].
- Gradient decay factor (*Grdec* or β_1) is the rate at which the gradient decays over time. It is a scalar value between 0 and 1, with a value closer to 1, indicating that the LSTM should "remember" the previous state more strongly, and a value closer to 0, indicating that the LSTM should "forget" the previous state more quickly. The gradient decay factor is usually determined during training by backpropagation and is used to update the weights of the LSTM [56,58,59].
- Squared gradient decay factor (*SqGrdec* or β_2) is used to adjust the decay gradient during the training stage in an LSTM network. It is a non-negative scalar value less than 1. Its function is to multiply the squared gradient of the weights by the decay factor, helping the training process converge faster and more efficiently by reducing the gradient step size while the network approaches the optimal solution. To scale the learning rate of the enhancer, β_2 is used. The small value of β_2, multiplied by the squared gradient of the weights at each step of the optimisation process, slows down the process while the network searches for the right solution, adjusting the parameters of the network to minimize its output deviation [56,58,59].
- Drop factor (*Dropf*) is used to reduce overfitting in neural networks by applying it to the input and/or output of LSTM cells, making this technique more robust to changes in input or output data. *Dropf* is the probability of excluding a unit by setting its output to zero during training. It is a scalar between 0 and 1, indicating that a percentage of units will be randomly excluded at each training step [56,58,59].

2.2.5. Bayesian Optimisation (BO)

Bayesian optimisation provides an alternative strategy to calculate the hyperparameters in an experiment returning different results while evaluating the same hyperparameter. A range of values for each hyperparameter is specified, and BO searches for a combination of hyperparameters that optimises the selected metric, reducing the error by adding an additional regulatory term [60,61]. BO is a regularisation method that incorporates uncertainty into the network weights by placing an a priori probability distribution over the network weights, whereby the network finds the weights that maximise the likelihood of the given data and prevents overfitting [62]. This method is usually computationally expensive and requires powerful computers, so β_1 and β_2 are more commonly used [56,58,59].

2.2.6. Experimental Approaches

A preliminary study consisted of calculating predictions of SO_2 and NO_2 using the entire database (130 variables) in order to perform forecasts at 1 h ahead and 4 h ahead. Using all variables, the prediction results obtained at 1 h and 4 h ahead were not as good as desired, obtaining mean square error rates in the interval (35–78) for SO_2 at and (118–280) for NO_2. Therefore, three different approaches have been proposed in this manuscript. The first one uses the time series of each pollutant. The second approach is composed of an

autoregressive arrangement of the time series data. Equation (13) shows the form of this autoregressive organisation. The third approach also uses exogenous variables such as the vessel database and wind information (direction and speed). Equation (14) represents this autoregressive arrangement where x represents the exogenous variables. In each of these approaches, a cross-validation (4-CV) procedure has been applied to calculate the different measures of performance for test sets (selecting 75% of the samples for training and the 25% for test).

In the case of an s-hour ahead forecasting, different values of s and n can be selected. In this work, s was selected to be 1 and 4, in order to achieve short-term predictions. In the case of a 1 h ahead prediction, s is equal to 1, and the value of n has been settled down to 8, as we have seen in previous work [50] that no better results are obtained by using windows longer than 8 h in the past. In the case of 4 h ahead prediction, s is equal to 4, and the value of n has been settled down to 3. Therefore, two autoregressive window sizes of 8 h and 12 h were employed.

$$y(t - n{\cdot}s) \dots y(t - s)y(t) \rightarrow y(t + s) \tag{13}$$

$$y(t - n{\cdot}s) \dots y(t - s)y(t)x(t - n{\cdot}s) \dots x(t - s)x(t) \rightarrow y(t + s) \tag{14}$$

A multicomparison procedure based on a Friedman test and the Bonferroni method [63] has been used to select the best model. The Friedman test is a non-parametric statistical test used to compare the differences between two or more related groups. It is similar to the repeated measures ANOVA, but it does not assume normality of the data. The Bonferroni method is a statistical technique used to adjust for the problem of multiple comparisons in hypothesis testing. When performing multiple hypothesis tests, the probability of observing at least one significant result due to chance alone increases, leading to an increased likelihood of making a Type I error (false positive).

3. Results

The experiments have been carried out at the Algeciras location for two different air pollutants, SO_2 and NO_2. Further, in order to prevent overfitting, different hyperparameters have been considered using BO. In each experiment, 25% of the total database (N) was used as separate unseen data (data test), while the remaining 75% was used as a training test. In each experiment, the last N/4 data samples have been selected as a test set. The performance indexes R, MAE, and MSE have been computed, as has the additional parameter d.

A number of LSTM topologies and hyperparameters have been tested. To this end, 1-step-ahead and 4-step-ahead prediction models have been developed to predict the next value of a time series. The results of the preliminary study are collected in Table 4, where lower R and MSE values are observed than in the following experiments.

Table 4. Preliminary study. SO_2 and NO_2 concentrations with 131 exogenous variables.

Parameters	1 h Ahead		4 h Ahead	
	SO_2	**NO_2**	**SO_2**	**NO_2**
d	0.8209	0.9273	0.2811	0.7902
Grdec	0.9031	0.8132	0.9980	0.9091
L2r	5.2956×10^{-5}	1.3755×10^{-4}	9.3367×10^{-5}	5.7315×10^{-5}
Lr	0.0019	0.0018	0.0015	0.0019
Lrdrop	0.1320	0.0702	0.1278	0.1675
MAE	2.5376	7.51142	3.8624	12.6110
Minib	128	64	64	128
MSE	34.3969	118.1229	77.6043	282.3092
Optimal neurons	18	2	82	9
R	0.7707	0.8718	0.2133	0.6690
SqGrdec	0.9526	0.9630	0.9047	0.9008
Dropf	0.0613	0.0685	0.5547	0.1038

After running the experiments and doing a search for neurons going from 0 to 50, 200, and 600 neurons at a 1 h and 4 h prediction horizon, the results have been collected in Tables 5–16. The main parameters are included in the tables, and it can be seen that, e.g., in SO_2 prediction at 1 h ahead, although more neurons are used, the results do not improve significantly. As the models are virtually identical, in the case of NO_2 prediction at 1 h ahead the main point that is observed is that the best models have far fewer neurons.

We have included results from a randomised resampling experiment with 20 replicates, where one test set is chosen and one validation set is chosen. Results are provided for a separate independent test set. The best model of a set of three Tables in each case is highlighted in bold. It should be noted that the results in each table for each prediction horizon, by varying the neurons, produces results with different numbers of neurons. However, in each table, it has been verified with the Friedman test that the models are very similar.

For each pollutant and prediction horizon, a comparison has been made between different scenarios. On the one hand, we looked for significant differences using different numbers of neurons. Configurations with a maximum of 50, 200, and 600 neurons were tested in order to try to find the best comparative configuration. For each experiment, 20 replicates were performed using Bayesian regularisation in each replicate. In addition, an experiment was conducted in which the inputs have been changed: (i) with only the time series (no exogenous variables) (Equation (13)), (ii) with the time series presented in an autoregressive way as foreseen in Equation (14), and (iii) including as input case (ii) and in addition introducing wind speed, wind direction, and the hourly time series of ships, also presented in an autoregressive way (also Equation (14)). For each pollutant and prediction horizon, three tables of results are presented, compressing a large number of experiments, upon whose results a multiple comparison test has been performed using a Friedman test on the values of R, MAE, MSE and d. Subsequently, a post hoc test such as the Bonferroni test is used to find groups of equivalent models. In each group of equivalent models, the model with the lowest number of neurons and complexity was chosen, following the criterion of Occam's razor [64] (e.g., in Table 6, since experiments show similar results for 50, 200, and 600 neurons, i.e., a difference of one ten-thousandth of a millimetre, the model with the lowest number of neurons was selected, from 572 to 19. The best model for SO_2 1 h ahead has larger values for the hyperparameters *Grdec*, *L2r*, *Lr*, and *Lrdrop*, and smaller values for *SqGrdec* and *Dropf*.

Furthermore, Figures 5–8 represent the results of the best models in each of the run simulations. All figures show the actual values versus those predicted by the models. Only test examples are represented, i.e., examples that have not been used in the training and validation of the models. Therefore, we can see how well the models generalise. The results for R, MAE, and MSE are averaged over 20 replications. As can be seen, the best results in the case of SO_2 are obtained when an autoregressive scheme is used and in the case of NO_2 when wind information (speed and direction) and ships in the port area are also included.

Table 5. SO_2 1 h ahead predictions. Time series data without exogenous variables.

Parameters	**Results**		
Maximum neurons	50	200	600
d	0.8915	0.8919	0.8951
Grdec	0.9213	0.9988	0.8183
L2r	5.4822×10^{-5}	1.6018×10^{-4}	5.1116×10^{-5}
Lr	0.0019	9.7171×10^{-4}	6.5588×10^{-4}
Lrdrop	0.0930	0.1377	0.0918
MAE	2.0879	2.0414	2.0403
Minib	128	128	64
MSE	25.0679	25.0049	24.8045
Optimal neurons	7	141	36
R	0.8312	0.8320	0.8320
SqGrdec	0.9750	0.9875	0.9768
Dropf	0.1072	0.4646	0.0619

Table 6. SO_2 1 h ahead predictions using autoregressive time series. The best model of the set of three Tables 5–7 is highlighted in bold.

Parameters	**Results**		
Maximum neurons	**50**	200	600
d	**0.9005**	0.8977	0.9007
Grdec	**0.9461**	0.8488	0.8418
L2r	$\mathbf{6.1300 \times 10^{-5}}$	5.5342×10^{-5}	5.5180×10^{-5}
Lr	**0.0018**	0.0013	9.2116×10^{-4}
Lrdrop	**0.1700**	0.0683	0.1133
MAE	**2.0438**	2.0243	2.0016
Minib	**64**	128	64
MSE	**23.7903**	23.9695	23.7806
Optimal neurons	**19**	54	572
R	**0.8396**	0.8395	0.8397
SqGrdec	**0.9205**	0.9434	0.9123
Dropf	**0.0828**	0.1273	0.1124

Table 7. SO_2 1 h ahead predictions using autoregressive time series with wind and vessels.

Parameters	**Results**		
Maximum neurons	50	200	600
d	0.8932	0.8919	0.8871
Grdec	0.9058	0.8360	0.8498
L2r	5.0021×10^{-5}	5.2238×10^{-5}	5.4045×10^{-5}
Lr	0.0019	0.0015	0.0018
Lrdrop	0.1645	0.0950	0.1991
MAE	2.0219	2.0392	2.0165
Minib	64	64	128
MSE	24.7903	25.1374	25.2364
Optimal neurons	10	16	59
R	0.8338	0.8315	0.8344
SqGrdec	0.9487	0.9818	0.9106
Dropf	0.0526	0.0534	0.2728

Table 8. NO_2 1 h ahead predictions. Time series data without exogenous variables.

Parameters	**Results**		
Maximum neurons	50	200	600
d	0.9262	0.9266	0.9263
Grdec	0.8024	0.9312	0.8008
L2r	6.4848×10^{-5}	5.5884×10^{-5}	6.1008×10^{-5}
Lr	0.0018	0.0018	0.0016
Lrdrop	0.0508	0.0585	0.0591
MAE	7.2310	7.2224	7.2928
Minib	32	16	128
MSE	115.4112	115.3174	115.7281
Optimal neurons	4	6	12
R	0.8734	0.8734	0.8732
SqGrdec	0.9443	0.9788	0.9505
Dropf	0.0811	0.1005	0.0918

Table 9. SO_2 1 h ahead predictions using autoregressive time series.

Parameters	**Results**		
Maximum neurons	50	200	600
d	0.9292	0.9285	0.9274
Grdec	0.8010	0.8244	0.9144
L2r	6.4437×10^{-5}	5.2499×10^{-5}	6.2772×10^{-5}
Lr	0.0013	6.5391×10^{-4}	0.0019
Lrdrop	0.0668	0.1574	0.1578
MAE	6.8833	7.0865	7.0572
Minib	128	64	128
MSE	112.9641	113.7278	113.2927
Optimal neurons	26	57	27
R	0.8749	0.8751	0.8753
SqGrdec	0.9149	0.9151	0.9793
Dropf	0.1072	0.0786	0.1220

Table 10. NO_2 1 h ahead predictions using autoregressive time series with wind and vessels. The best model of the set of three Tables 8–10 is highlighted in bold.

Parameters	**Results**		
Maximum neurons	**50**	200	600
d	**0.9327**	0.9326	0.9321
Grdec	**0.9234**	0.9161	0.8753
L2r	$\mathbf{5.3357 \times 10^{-5}}$	5.4322×10^{-5}	5.6660×10^{-5}
Lr	**0.0015**	0.00128	0.0013
Lrdrop	**0.1128**	0.0836	0.1715
MAE	**6.6893**	6.6498	6.6644
Minib	**256**	128	256
MSE	**109.6412**	109.3660	109.6794
Optimal neurons	**49**	104	115
R	**0.8790**	0.8794	0.8790
SqGrdec	**0.9480**	0.9651	0.9031
Dropf	**0.0807**	0.0537	0.1655

Table 11. SO_2 4 h ahead predictions. Time series data without exogenous variables.

Parameters	**Results**		
Maximum neurons	50	200	600
d	0.6988	0.6931	0.7121
Grdec	0.9922	0.9985	0.8008
L2r	5.4534×10^{-5}	5.2772×10^{-5}	7.0673×10^{-5}
Lr	0.0018	0.0019	0.0018
Lrdrop	0.1601	0.1168	0.0759
MAE	3.4586	3.6347	3.6021
Minib	64	64	64
MSE	48.4230	48.6198	47.8566
Optimal neurons	19	149	185
R	0.6393	0.6406	0.6415
SqGrdec	0.9359	0.9913	0.9472
Dropf	0.1076	0.1577	0.2264

Table 12. SO_2 4 h ahead predictions using autoregressive time series. The best model of the set of three Tables 11–13 is highlighted in bold.

Parameters	**Results**		
Maximum neurons	50	200	**600**
d	0.8958	0.8949	**0.8966**
Grdec	0.9525	0.8282	**0.8921**
L2r	6.0155×10^{-5}	5.1761×10^{-5}	$\mathbf{5.4341 \times 10^{-5}}$
Lr	$9.9426 \cdot 10^{-4}$	0.0015	**0.0018**
Lrdrop	0.1251	0.1329	**0.1447**
MAE	2.0186	2.0426	**2.0588**
Minib	8	64	**32**
MSE	24.4654	24.4692	**24.3246**
Optimal neurons	6	10	**46**
R	0.8352	0.8357	**0.8362**
SqGrdec	0.9697	0.9978	**0.9108**
Dropf	0.0804	0.0809	**0.1596**

Table 13. SO_2 4 h ahead predictions using autoregressive time series with wind and vessels.

Parameters	**Results**		
Maximum neurons	50	200	600
d	0.8900	0.8919	0.8908
Grdec	0.9010	0.8360	0.8385
L2r	5.4647×10^{-5}	5.2238×10^{-5}	7.9642×10^{-5}
Lr	0.0019	0.0015	0.0018
Lrdrop	0.1216	0.0950	0.1997
MAE	2.0108	2.0392	2.0396
Minib	128	64	64
MSE	25.2105	25.1374	24.9333
Optimal neurons	50	16	16
R	0.8317	0.8315	0.8335
SqGrdec	0.9013	0.9818	0.9598
Dropf	0.1095	0.0534	0.1564

Table 14. NO_2 4 h ahead predictions. Time series data without exogenous variables.

Parameters	**Results**		
Maximum neurons	50	200	600
d	0.7183	0.7161	0.7056
Grdec	0.9525	0.8123	0.9069
L2r	5.2742×10^{-5}	5.3231×10^{-5}	1.8764×10^{-4}
Lr	0.0019	0.0018	0.0016
Lrdrop	0.1040	0.1693	0.1884
MAE	13.4623	13.2543	13.3251
Minib	64	16	128
MSE	313.2493	309.8679	312.2236
Optimal neurons	2	2	4
R	0.6069	0.6067	0.6008
SqGrdec	0.9012	0.9953	0.9146
Dropf	0.0530	0.0557	0.0675

Table 15. NO_2 4 h ahead predictions using autoregressive time series.

Parameters	**Results**		
Maximum neurons	50	200	600
d	0.9293	0.9290	0.9294
Grdec	0.8286	0.9160	0.9920
L2r	5.0361×10^{-5}	6.5346×10^{-5}	1.4932×10^{-4}
Lr	0.0019	0.0019	0.0017
Lrdrop	0.1473	0.1499	0.1242
MAE	6.8670	6.9318	6.8960
Minib	64	64	256
MSE	112.5655	112.9569	113.2348
Optimal neurons	12	25	137
R	0.8754	0.8751	0.8747
SqGrdec	0.9788	0.9791	0.9930
Dropf	0.0912	0.1673	0.3204

Table 16. NO_2 4 h ahead predictions using autoregressive time series with wind and vessels. The best model of the set of three Tables 14–16 is highlighted in bold.

Parameters	**Results**		
Maximum neurons	50	**200**	600
d	0.9318	**0.9320**	0.9329
Grdec	0.8636	**0.8990**	0.8000
L2r	5.1211×10^{-5}	$\mathbf{5.6252 \times 10^{-5}}$	5.6384×10^{-5}
Lr	0.0017	**0.0017**	0.0019
Lrdrop	0.0775	**0.0723**	0.1764
MAE	6.6689	**6.7096**	6.6630
Minib	128	**128**	512
MSE	109.3849	**109.3647**	109.9590
Optimal neurons	43	**32**	211
R	0.8792	**0.8789**	0.8788
SqGrdec	0.9738	**0.9955**	0.9288
Dropf	0.0675	**0.0722**	0.1039

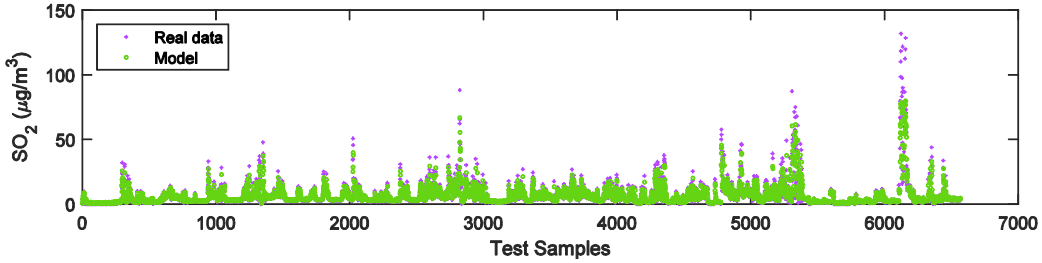


Figure 5. SO_2 1 h ahead forecasting. LSTM best model with 19 neurons and autoregressive time series with R = 0.8396 and MSE = 23.7903.

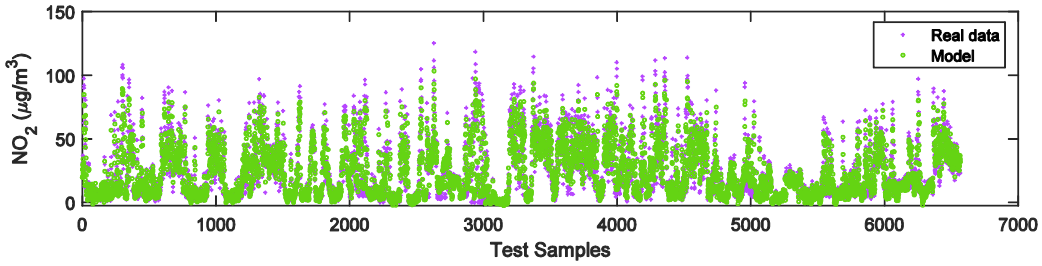


Figure 6. NO_2 1 h ahead forecasting. LSTM best model with 49 neurons and autoregressive time series with wind and vessels, and R = 0.8790 and MSE = 109.6412.

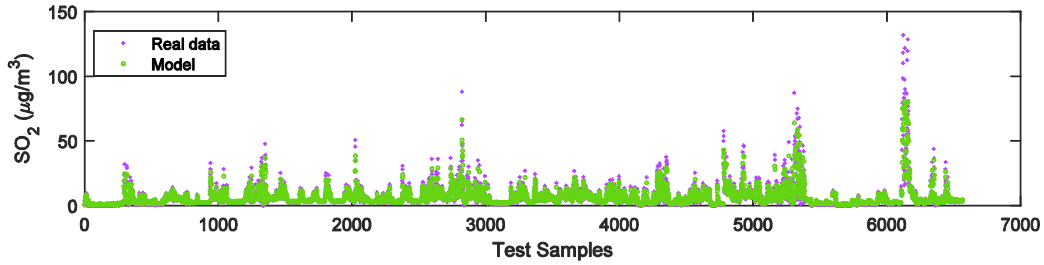


Figure 7. SO_2 4 h ahead forecasting. LSTM best model with 46 neurons and autoregressive time series with R = 0.8362 and MSE = 24.3246.

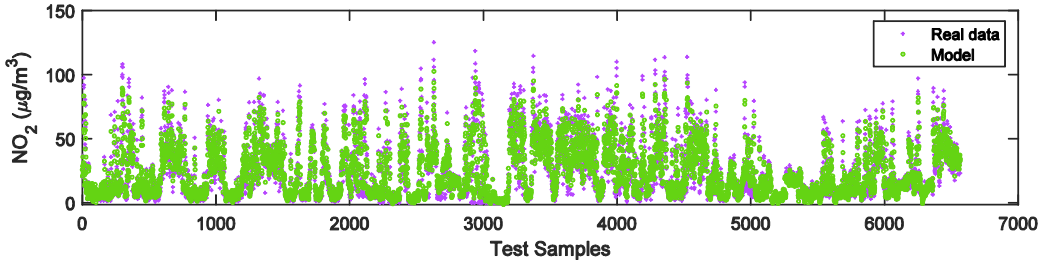


Figure 8. NO_2 4 h ahead forecasting. LSTM best model with 32 neurons and autoregressive time series with wind and vessels, and R = 0.8789 and MSE = 109.3647.

4. Discussion

Tables 5–7 present the best model for SO_2 1 h ahead predictions. In this case, a configuration with 19 neurons using only the autoregressive time series as input shows larger values of the hyperparameters *Grdec*, *L2r*, *Lr*, and *Lrdrop*, and smaller values of *SqGrdec* and *Dropf*, reaching R values of 0.8396 and MSE values of 23.7903. In all simulations of this scenario, the minibatch size is moderate (64, in this case) indicating that no very long-distance relationships seem to exist. In most models, it occurs in this way.

Tables 8–10 show the case of NO_2 1 h ahead predictions. The best model corresponds to a configuration of 49 neurons, with the largest *minibatch* size equal to 256, which seems

to indicate that there is interesting information at a larger distance. The rest of the different hyperparameters show variability between the different cases. The best configuration includes wind and ships information as inputs, in addition to the autoregressive time series.

Tables 11–13 illustrate SO_2 4 h ahead forecasting. In this case, it is also observed that the best model does not have a high *minibatch* value (32). Nevertheless, in this scenario, the best model used a database arranged in an autoregressive way, as in Equation (14), where $s = 4$. Therefore, the *minibatch* of 32 would mean a time span of $32 \times 4 = 128$ h, approximately 5 days. The values of $R = 0.8352$ are almost equivalent to those obtained for 1 h ahead predictions. The best configuration also includes the autoregressive organisation of the time series at 8 h, without using wind or ship information.

Tables 14–16 exhibit the case of NO_2 4 h ahead forecasting. In this case, the best model was found using a configuration that includes wind and ships as inputs, in addition to the autoregressive time series and a large *minibatch* size equal to 128, which means that in the case of NO_2, relevant information is found at a large distance, obtaining $R = 0.8790$.

In all four cases, the best models are found with configurations where an autoregressive scheme is used as input, which indicates that it is interesting to provide the LSTMs with more information than just the time series itself.

Some of the limitations of this study are related to limitations of LSTM networks, such as, in particular, the computational complexity, since LSTM networks are computationally intensive and require a lot of time and resources to train on large datasets; further, the training difficulty can be difficult to perform due to their highly nonlinear nature and a large number of parameters and hyperparameters. On the other hand, in the scope of future work, the authors plan to use exogenous information as well as bidirectional LSTM networks.

5. Conclusions

The objective of this study is to develop precise forecasting models for predicting the levels of air pollutant concentration in the Bay of Algeciras area, Spain. Two distinct prediction horizons ($t + 1$, $t + 4$) were established to evaluate forecasting capabilities. Initially, a forecasting methodology based on the entire set of exogenous variables (130) was employed. However, the results were not satisfactory. Therefore, three alternative approaches were suggested concerning the input data. The first option utilized a univariate time series dataset that contained only the data obtained from each monitoring station. The second approach employed an autoregressive scheme of the time series by using the data from the past 8 h for 1 h ahead forecasting, and 3 samples in the past in the case of 4 h ahead forecasting. Finally, the third approach added exogenous features, including wind speed, wind direction, and vessel data together with the time series data in an autoregressive arrangement.

The following main conclusions emerge from the development of this study:

- LSTMs have proven to be effective tools for the prediction of atmospheric pollutants.
- We have additionally proven that the use of lagged information with an autoregressive input data entry scheme is an effective tool for the prediction of atmospheric pollutants.

With this work, we have shown that short-term pollution forecasts have been effectively produced using LSTMs so that citizens, administrations, and companies can use this tool as an aid to decision-making. One of the future applications of this research is to implement high-resolution local prediction systems and thus, for example, to know the level of pollution in schools and be aware of its importance for children's health. The idea is to create an area of sensors in certain critical areas to monitor and predict pollutants and advise on air quality, so that decisions can be made and solutions found. We need to create a healthier environment for the next generations.

Author Contributions: Conceptualization, M.I.R.-G., I.J.T. and J.J.R.-A.; data curation, M.I.R.-G., J.G.-E. and M.G.C.-G.; formal analysis, M.I.R.-G., J.G.-E. and I.J.T.; funding acquisition, I.J.T.; investigation, M.I.R.-G. and J.G.-E.; methodology, M.I.R.-G., M.G.C.-G. and I.J.T.; project administration,

J.J.R.-A. and I.J.T.; software, M.I.R.-G., M.G.C.-G., J.G.-E. and I.J.T.; resources M.I.R.-G., J.G.-E. and I.J.T.; supervision J.J.R.-A. and I.J.T.; validation, M.I.R.-G.; visualization, M.I.R.-G.; writing—original draft, M.I.R.-G. and I.J.T.; writing—review and editing, M.I.R.-G., J.J.R.-A. and I.J.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the research project RTI2018-098160-B-I00 supported by MICINN. 'Programa Estatal de I+D+i Orientada a los Retos de la Sociedad'. This research is supported by 'Plan Propio de la Universidad de Cádiz'.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Pollutant and atmospheric data used in this work have been kindly provided by the Andalusian Regional Government and the vessels database by the Algeciras Bay Port Authority.

Conflicts of Interest: The authors do not have relevant conflict of interest to declare regarding the content of this article.

References

1. Salvaraji, L.; Avoi, R.; Jeffree, M.S.; Saupin, S.; Toha, H.R.; Shamsudin, S.B. Effects of ambient air pollutants on cardiovascular disease hospitalization admission. *Glob. J. Environ. Sci. Manag.* **2023**, *9*, 157–172. [CrossRef]
2. Gan, T.; Bambrick, H.; Tong, S.; Hu, W. Air pollution and liver cancer: A systematic review. *J. Environ. Sci.* **2023**, *126*, 817–826. [CrossRef] [PubMed]
3. Blanc, N.; Liao, J.; Gilliland, F.; Zhang, J.O.; Berhane, K.; Huang, G.; Yan, W.; Chen, Z. A systematic review of evidence for maternal preconception exposure to outdoor air pollution on Children's health. *Environ. Pollut.* **2023**, *318*, 120850. [CrossRef]
4. Magee, J.L.; Miller, J.M. Kinetics and Mechanism of the Reaction between Nitric Oxide and Oxygen in the Gas Phase. *J. Phys. Chem.* **1971**, *75*, 2312–2317. [CrossRef]
5. Fernández-Pampillón, J.; Palacios, M.; Núñez, L.; Pujadas, M.; Artíñano, B. Potential ambient NO_2 abatement by applying photocatalytic materials in a Spanish city and analysis of short-term effect on human mortality. *Environ. Pollut.* **2023**, *323*, 121203. [CrossRef]
6. Huang, F.; Pan, B.; Wu, S.; Guo, X.; Li, G.; Chen, Y. Short-term exposure to sulfur dioxide and daily mortality in 17 Chinese cities: The China air pollution and health effects study (CAPES). *Environ. Res.* **2017**, *159*, 1–7.
7. EU. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Off. J. Eur. Union* **2008**, *152*, 1–44.
8. IMO (International Maritime Organization). *The International Convention for the Prevention of Pollution from Ships*; Marine Pollution (MARPOL): London, UK, 2021; annex VI.
9. Moreno-Gutiérrez, J.; Calderay, F.; Saborido, N.; Boile, M.; Rodríguez, R.; Durán-Grados, V. Methodologies for estimating shipping emissions and energy consumption: A comparative analysis of current methods. *Energy* **2015**, *86*, 603–616. [CrossRef]
10. Monteiro, A.; Russo, M.; Gama, C.; Borrego, C. Shipping emissions and their impact on air quality in urban coastal areas: Present and future scenarios. *WIT Trans. Built Environ.* **2019**, *186*, 145–151. [CrossRef]
11. Contini, D.; Merico, E. Recent advances in studying air quality and health effects of shipping emissions. *Atmosphere* **2021**, *12*, 92. [CrossRef]
12. Durán-Grados, V.; Rodríguez-Moreno, R.; Calderay-Cayetano, F.; Amado-Sánchez, Y.; Pájaro-Velázquez, E.; Nunes, R.A.O.; Alvim-Ferraz, M.; Sousa, S.I.V.; Moreno-Gutiérrez, J. The Influence of Emissions from Maritime Transport on Air Quality in the Strait of Gibraltar (Spain). *Sustainability* **2022**, *14*, 12507. [CrossRef]
13. González, Y.; Rodríguez, S.; Guerra García, J.C.; Trujillo, J.L.; García, R. Ultrafine particles pollution in urban coastal air due to ship emissions. *Atmos. Environ.* **2011**, *45*, 4907–4914. [CrossRef]
14. Lu, G.; Brook, J.R.; Rami Alfarra, M.; Anlauf, K.; Richard Leaitch, W.; Sharma, S.; Wang, D.; Worsnop, D.R.; Phinney, L. Identification and characterization of inland ship plumes over Vancouver, BC. *Atmos. Environ.* **2006**, *40*, 2767–2782. [CrossRef]
15. Corbett, J.J.; Wang, C.; Winebrake, J.; Green, E.H. Allocation and forecasting of global ship emissions. In *Clean Air Task Force*; University of Delaware: Boston, MA, USA, 2007; p. 26. Available online: https://www.researchgate.net/publication/241579973_Allocation_and_Forecasting_of_Global_Ship_Emissions (accessed on 30 January 2023).
16. Yau, P.S.; Lee, S.C.; Cheng, Y.; Huang, Y.; Lai, S.C.; Xu, X.H. Contribution of ship emissions to the fine particulate in the community near an international port in Hong Kong. *Atmos. Res.* **2013**, *124*, 61–72. [CrossRef]
17. Liu, T.K.; Sheu, H.Y.; Tsai, J.Y. Sulfur dioxide emission estimates from merchant vessels in a Port area and related control strategies. *Aerosol Air Qual. Res.* **2014**, *14*, 413–421. [CrossRef]
18. Fan, Q.; Zhang, Y.; Ma, W.; Ma, H.; Feng, J.; Yu, Q.; Yang, X.; Ng, S.K.W.; Fu, Q.; Chen, L. Spatial and seasonal dynamics of ship emissions over the Yangtze river delta and east China sea and their potential environmental influence. *Environ. Sci. Technol.* **2016**, *50*, 1322–1329. [CrossRef]

19. Zhang, Y.; Yang, X.; Brown, R.; Yang, L.; Morawska, L.; Ristovski, Z.; Fu, Q.; Huang, C. Shipping emissions and their impacts on air quality in China. *Sci. Total Environ.* **2017**, *581–582*, 186–198. [CrossRef]
20. Nunes Rafael, A.O.; Alvim-Ferraz, M.C.M.; Martins, F.G.; Calderay-Cayetano, F.; Durán-Grados, V.; Moreno-Gutiérrez, J.; Jalkanen, J.-P.; Hannuniemi, H.; Sousa, S.I.V. Shipping emissions in the Iberian Peninsula and the impacts on air quality. *Atmos. Chem. Phys.* **2020**, *20*, 9473–9489. [CrossRef]
21. Widyantara, I.M.O.; Hartawan, I.P.N.; Karyawati, A.A.I.N.E.; Er, N.I.; Artana, K.B. Automatic identification system-based trajectory clustering framework to identify vessel movement pattern. *IAES Int. J. Artif. Intell.* **2023**, *12*, 1–11. [CrossRef]
22. Kujawska, J.; Kulisz, M.; Oleszczuk, P.; Cel, W. Machine Learning Methods to Forecast the Concentration of PM_{10} in Lublin. *Energies* **2022**, *15*, 6428. [CrossRef]
23. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
24. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
25. Korunoski, M.; Stojkoska, B.R.; Trivodaliev, K. Internet of Things Solution for Intelligent Air Pollution Prediction and Visualization. In Proceedings of the IEEE EUROCON 2019—18th International Conference on Smart Technologies, Novi Sad, Serbia, 1–4 July 2019. [CrossRef]
26. Masood, A.; Ahmad, K. Data-driven predictive modeling of $PM_{2.5}$ concentrations using machine learning and deep learning techniques: A case study of Delhi, India. *Environ. Monit. Assess.* **2022**, *195*, 60. [CrossRef] [PubMed]
27. Liao, Q.; Zhu, M.; Wu, L.; Pan, X.; Tang, X.; Wang, Z. Deep Learning for Air Quality Forecasts: A Review. *Curr. Pollut. Rep.* **2020**, *6*, 399–409. [CrossRef]
28. Ban, W.; Shen, L. $PM_{2.5}$ Prediction Based on the CEEMDAN Algorithm and a Machine Learning Hybrid Model. *Sustainability* **2022**, *14*, 16128. [CrossRef]
29. Drewil, G.I.; Al-Bahadili, R.J. Air pollution prediction using LSTM deep learning and metaheuristics algorithms. *Meas. Sens.* **2022**, *24*, 100546. [CrossRef]
30. Gunasekar, S.; Retna Kumar, G.J.; Kumar, Y.D. Sustainable optimized LSTM-based intelligent system for air quality prediction in Chennai. *Acta Geophys.* **2022**, *70*, 2889–2899. [CrossRef]
31. Waseem, K.H.; Mushtaq, H.; Abid, F.; Abu-Mahfouz, A.M.; Shaikh, A.; Turan, M.; Rasheed, J. Forecasting of Air Quality Using an Optimized Recurrent Neural Network. *Processes* **2022**, *10*, 2117. [CrossRef]
32. Wang, W.; Tang, Q. Combined model of air quality index forecasting based on the combination of complementary empirical mode decomposition and sequence reconstruction. *Environ. Pollut.* **2023**, *316*, 120628. [CrossRef]
33. Chang, Y.S.; Chiao, H.T.; Abimannan, S.; Huang, Y.P.; Tsai, Y.T.; Lin, K.M. An LSTM-based aggregated model for air pollution forecasting. *Atmos. Pollut. Res.* **2020**, *11*, 1451–1463. [CrossRef]
34. Zhang, H. Relationships between meteorological parameters and criteria air pollutants in three megacities in China. *Environ. Res.* **2015**, *140*, 242–254. [CrossRef] [PubMed]
35. Rodríguez-García, M.I.; González-Enrique, J.; Moscoso-López, J.A.; Ruiz-Aguilar, J.J.; Rodríguez-Lopez, J.C.; Turias, I.J. Comparison of maritime transport influence of SO_2 levels in Algeciras and Alcornocales Park (Spain). *Transp. Res. Procedia* **2021**, *58*, 591–598. [CrossRef]
36. Ruiz-Aguilar, J.J.-E.; Turias, I.; González-Enrique, J.; Urda, D.; Elizondo, D. A permutation entropy-based EMD–ANN forecasting ensemble approach for wind speed prediction. *Neural Comput. Appl.* **2021**, *33*, 2369–2391. [CrossRef]
37. Mclean, S.; Kaiser, J.; Ben Hughes, B. Spatial estimation of outdoor NO_2 levels in Central London using deep neural networks and a wavelet decomposition technique. *Ecol. Model.* **2020**, *424*, 109017. [CrossRef]
38. Freeman, B.S.; Graham Taylor, G.; Gharabaghi, B.; Thé, J. Forecasting air quality time series using deep learning. *J. Air Waste Manag. Assoc.* **2018**, *68*, 866–886. [CrossRef]
39. Deep, B.; Mathur, I.; Joshi, N. An approach to forecast pollutants concentration with varied dispersion. *Int. J. Environ. Sci. Technol.* **2022**, *19*, 5131–5138. [CrossRef]
40. Samal, K.; Panda, A.; Babu, K.; Das, S. An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach. *Sustain. Cities Soc.* **2021**, 70. [CrossRef]
41. Urda, D.; Jerez, J.M.; Turias, I.J. Data dimension and structure effects in predictive performance of deep neural networks. In *New Trends in Intelligent Software Methodologies, Tools and Techniques: Proceedings 17th International Conference SoMeT_18*; IOS Press: Amsterdam, The Netherlands, 2018; pp. 303–361.
42. Cheng, W.; Shen, Y.; Zhu, Y.; Huang, L. A Neural Attention Model for Urban Air Quality Inference: Learning the Weights of Monitoring Stations. *AAAI Conf. Artif. Intell.* **2018**, *32*, 2151–2158. [CrossRef]
43. Laña, I.; Del Ser, J.; Padro, A.; Velez, M.; Casanova-Mateo, C. The role of local urban traffic and meteorological conditions in air pollution: A data-based case study in Madrid, Spain. *Atmos. Environ.* **2016**, *145*, 424–438. [CrossRef]
44. Zheng, Y.; Yi, X.; Li, M.; Li, R.; Shan, Z.; Chang, E.; Li, T. Forecasting fine-grained air quality based on big data. In Proceedings of the Twenty-First ACM SIGKDD. International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; ACM: New York, NY, USA, 2015; pp. 2267–2276.
45. Wang, J.; Song, G. A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction. *Neurocomputing* **2018**, *314*, 198–206. [CrossRef]
46. Moscoso-López, J.; González-Enrique, J.; Urda, D.; Ruiz-Aguilar, J.J.; Turias, I.J. Hourly pollutants forecasting using a deep learning approach to obtain the AQI. *Log. J. IGPL* **2022**, jzac035. [CrossRef]

47. Rodríguez-García, M.I.; González-Enrique, J.; Moscoso-López, J.A.; Ruiz-Aguilar, J.J.; Turias, I.J. Air pollution relevance analysis in the Bay of Algeciras (Spain). *Int. J. Environ. Sci. Technol.* **2022**. [CrossRef] [PubMed]
48. Turias, I.J.; González, F.J.; Martin, M.L.; Galindo, P.L. Prediction models of CO, SPM and SO_2 concentrations in the Campo de Gibraltar Region, Spain: A multiple comparison strategy. *Environ. Monit. Assess.* **2008**, *143*, 131–146. [CrossRef] [PubMed]
49. González-Enrique, J.; Turias, I.J.; Ruiz-Aguilar, J.J.; Moscoso-López, J.A.; Franco, L. Spatial and meteorological relevance in NO_2 estimations: A case study in the Bay of Algeciras (Spain). *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 801–815. [CrossRef]
50. González-Enrique, J.; Ruiz-Aguilar, J.J.; Moscoso-López, J.A.; Urda, D.; Deka, L.; Turias, I.J. Artificial Neural Networks, Sequence-to-Sequence LSTMs, and Exogenous Variables as Analytical Tools for NO_2 (Air Pollution) Forecasting: A Case Study in the Bay of Algeciras (Spain). *Sensors* **2021**, *21*, 1770. [CrossRef]
51. Mengara, A.G.; Park, E.; Jang, J.; Yoo, Y. Attention-Based Distributed Deep Learning Model for Air Quality Forecasting. *Sustainability* **2022**, *6*, 3269. [CrossRef]
52. Xayasouk, T.; Lee, H.; Lee, G. Air Pollution Prediction Using Long Short-Term Memory (LSTM) and Deep Autoencoder (DAE) Models. *Sustainability* **2020**, *12*, 2570. [CrossRef]
53. González-Enrique, J.; Ruiz-Aguilar, J.J.; Moscoso-López, J.A.; Urda, D.; Turias, I.J. A comparison of ranking filter methods applied to the estimation of NO_2 concentrations in the Bay of Algeciras (Spain). *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1999–2019. [CrossRef]
54. Willmott, C.J. On the Validation of Models. *Phys. Geogr.* **1982**, *2*, 184–194. [CrossRef]
55. Legates, D.R.; McCabe, G.J., Jr. Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. *Water Resour. Res.* **1999**, *35*, 233–241. [CrossRef]
56. Willmott, C.J.; Robeson, S.M.; Matsuura, K. A refined index of model performance. *Int. J. Climatol.* **2011**, *32*, 2088–2094. [CrossRef]
57. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
58. Willmott, C.J.; Ackleson, S.G.; Davis, R.E.; Feddema, J.J.; Klink, K.M.; Legates, D.R.; O'Donnell, J.; Rowe, C.M. Statistics for the Evaluation and Comparison of Models. *J. Geophys. Res.* **1985**, *90*, 8995–9005. [CrossRef]
59. Willmott, C.J. On the evaluation of model performance in physical geography. In *Spatial Statistics and Models*; Springer: Dordrecht, The Netherlands, 1984; pp. 443–460.
60. Aoki, M. On Some Convergence Questions in Bayesian Optimization Problems. *IEEE Trans. Autom. Control* **1965**, *10*, 180–182. [CrossRef]
61. Akin, Y.; Cansu, Z.; Oktay, H. Air Pollution Modelling with Deep Learning: A Review. *Int. J. Environ. Pollut. Environ. Model.* **2018**, *1*, 58–62.
62. Močkus, J.B.; Močkus, L.J. Bayesian approach to global optimization and application to multiobjective and constrained problems. *J. Optim. Theory Appl.* **1991**, *70*, 157–172. [CrossRef]
63. Hochberg, Y.; Tambane, A.C. *Multiple Comparison Procedures*; Wiley: New York, NY, USA, 1987.
64. Bargagli Stoffi, F.J.; Cevolani, G.; Gnecco, G. Simple Models in Complex Worlds: Occam's Razor and Statistical Learning Theory. *Minds Mach.* **2022**, *32*, 13–42. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.