



# A systematic comparison of different machine learning models for the spatial estimation of air pollution

Elena Cerezuela-Escudero<sup>1</sup> · Juan Manuel Montes-Sanchez<sup>1</sup> · Juan Pedro Dominguez-Morales<sup>1</sup> · Lourdes Duran-Lopez<sup>1</sup> · Gabriel Jimenez-Moreno<sup>1</sup>

Accepted: 12 October 2023 / Published online: 31 October 2023  
© The Author(s) 2023

## Abstract

Air pollutants harm human health and the environment. Nowadays, deploying an air pollution monitoring network in many urban areas could provide real-time air quality assessment. However, these networks are usually sparsely distributed and the sensor calibration problems that may appear over time lead to missing and wrong measurements. There is an increasing interest in developing air quality modelling methods to minimize measurement errors, predict spatial and temporal air quality, and support more spatially-resolved health effect analysis. This research aims to evaluate the ability of three feed-forward neural network architectures for the spatial prediction of air pollutant concentrations using the measures of an air quality monitoring network. In addition to these architectures, Support Vector Machines and geostatistical methods (Inverse Distance Weighting and Ordinary Kriging) were also implemented to compare the performance of neural network models. The evaluation of the methods was performed using the historical values of seven air pollutants (Nitrogen monoxide, Nitrogen dioxide, Sulphur dioxide, Carbon monoxide, Ozone, and particulate matters with size less than or equal to 2.5  $\mu\text{m}$  and to 10  $\mu\text{m}$ ) from an urban air quality monitoring network located at the metropolitan area of Madrid (Spain). To assess and compare the predictive ability of the models, three estimation accuracy indicators were calculated: the Root Mean Squared Error, the Mean Absolute Error, and the coefficient of determination. FFNN-based models are superior to geostatistical methods and slightly better than Support Vector Machines for fitting the spatial correlation of air pollutant measurements.

**Keywords** Air pollution · Spatial estimation · Air quality monitoring · Feed-forward neural networks · Support vector machines · Kriging

## 1 Introduction

Exposure to air pollution is a risk factor for diseases such as stroke, asthma, cancer and chronic obstructive pulmonary

disease. It is associated with noxious effects on human health and it is especially harmful to vulnerable groups such as children, the elderly and patients with respiratory and cardiovascular diseases [1–4]. Air pollutants not only severely impact public health, but also the climate and ecosystems because several of them are greenhouse gases [1, 5]. Many of the air pollutants are also sources of greenhouse gas emissions. Considering the significance of air quality on health and the environment, the World Health Organization (WHO) has developed guidelines to improve air quality by setting limits on the concentrations of various air pollutants: ozone (O<sub>3</sub>), nitrogen dioxide (NO<sub>2</sub>), sulphur dioxide (SO<sub>2</sub>), carbon monoxide (CO), particulate matter with size less than or equal to 2.5  $\mu\text{m}$  (PM<sub>2.5</sub>) and particulate matter with size less than or equal to 10  $\mu\text{m}$  (PM<sub>10</sub>) [5].

Air quality monitoring is an important task for governments to provide information on potential health risks and determine appropriate environmental management policies.

---

✉ Elena Cerezuela-Escudero  
ecerezuela@us.es

Juan Manuel Montes-Sanchez  
juanmanuelmontes@us.es

Juan Pedro Dominguez-Morales  
jpdominguez@us.es

Lourdes Duran-Lopez  
lduran2@us.es

Gabriel Jimenez-Moreno  
gaji@us.es

<sup>1</sup> Robotics and Technology of Computers Lab (RTC), ETSI Informática, Universidad de Sevilla, Avd. Reina Mercedes s/n, Sevilla 41012, Spain

The development of air pollution sensing technology in the last few decades and the support of government agencies have contributed to build an air quality monitoring network in many urban areas with the aim of analysing and publishing the concentrations of several air pollutants that are potentially harmful to health<sup>1 2</sup>. However, these networks are usually sparsely distributed and the sensor calibration problems that may appear lead to missing and wrong measurements [6–9]. There is an increasing interest in developing air quality modelling methods to minimize measurement errors, predict spatial and temporal air quality, and support more spatially-resolved health effect analyses [8–13].

Air pollution modelling follows two different approaches. The first approach consists in using the deterministic mathematical modelling of atmospheric pollutant dispersion. The second approach, on the other hand, consists in employing statistical models based on historical air quality data, and, in some cases, meteorological and geographic information too.

The deterministic mathematical modelling involves the simulation of pollutant dispersion and transport mechanisms using emission values in industrial and urban areas, physical and chemical processes in the atmosphere, meteorological data, and geographic and topological information. The deterministic methods that are most present in the scientific literature are the “Community Multiscale Air Quality (CMAQ) model” [14, 15], “Weather Research and Forecasting model with Chemistry (WRF-Chem)” [16], and “Nested Air Quality Prediction Modelling System (NAQPMS)” [17, 18]. The deterministic modelling has limitations due to the enormous number of pollution sources and the fact that air distribution is influenced by several complex physical/chemical processes that require many variables.

The statistical approach takes advantage of the spatial and temporal correlations that are present in the air pollution concentration time series and formulates models that simulate these dependencies with a high degree of accuracy. Several methodologies have been developed along this approach, including classical statistics [19], artificial intelligence [8, 9, 11–13, 20, 21] and geostatistical techniques [6, 7, 22].

The application of Artificial Neural Networks (ANNs) has been frequently used to forecast air quality. Some recent articles [7, 13, 20, 21, 23–26] use the historical values of various pollutants to predict the air quality index and/or air pollutant concentrations. Several of them use meteorological data too. Machine and deep learning methods show a more remarkable ability to simulate non-linear systems because of their self-learning, self-organizing, and self-adaptation features.

Instead, fewer studies have used the ANN technique for the spatial estimation of air pollution [6, 8, 11, 12, 27]. The study [8] evaluates the use of a Back-Propagation Neural Networks (BPNN) for modelling the spatial atmospheric pollution of five air pollutants (NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, PM<sub>2.5</sub> and PM<sub>10</sub>). The authors of [6] proposed machine learning and geostatistical methods to predict PM<sub>2.5</sub> pollution levels. Some of these studies applied Deep learning methods to extract complex and non-linear spatiotemporal correlations [11, 12]. The next section further describes the references about the use of machine learning and deep learning methods for predicting air pollution.

The main objective of this research was to develop an ANN-based system for modelling the spatial characteristics of air pollutant concentrations measured at an urban air quality monitoring network. To estimate the air pollutant value at the target site based on the measurements collected at nearby locations, we applied three Feed-Forward Neural Network (FFNN) architectures for regression (with one, two, and three fully connected hidden layers), a Support Vector Machine (SVM) and geostatistical methods. The evaluation of the methods was performed by using the historical values of seven air pollutants (Nitrogen monoxide (NO), NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, CO, PM<sub>2.5</sub>, and PM<sub>10</sub>) collected at the urban air quality monitoring network located at the greater metropolitan area of Madrid (Spain). For assessing and comparing the predictive ability of the models, three estimation accuracy indicators were calculated: the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the coefficient of determination ( $R^2$ ).

The main contributions of this work include the following:

- A broad analysis of the most common pollutants evaluated on a real dataset consisting of samples obtained from the air quality monitoring network deployed in the city of Madrid (Spain).
- A FFNN-based air pollution spatial estimation system able to accurately predict NO<sub>2</sub>, NO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub> and PM<sub>10</sub> concentrations.
- A systematic comparison between two geostatistical models, SVM, and three FFNN architectures (one-hidden FFNN, bi-layer FFNN, and tri-layer FFNN) for evaluating the prediction of the concentration of seven air pollutants (NO<sub>2</sub>, NO, CO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub> and PM<sub>10</sub>).

The present work is structured as follows. The following section reviews literature related to machine learning and deep learning methods applied to air quality prediction. Section 3 presents the study area and the air pollution dataset used in the experiments. In Section 4, the methods applied to estimate air pollutant concentrations are described: Inverse distance weighted (IDW), Ordinary Kriging (OK), SVM, FFNN. In Section 5, every phase of the experiments is pre-

<sup>1</sup> European Air Quality Index: <https://www.eea.europa.eu/themes/air/air-quality-index>. Retrieved October 28, 2023

<sup>2</sup> AIRNOW: <https://www.airnow.gov/aqi/aqi-basics>. Retrieved October 28, 2023

sented. Section 6 describes the results of the experiments performed. In Section 7 we discuss our results. Finally, the conclusions are presented in Section 8.

## 2 Related work

Machine learning methods are widely used to extract non-linear correlations in air pollutant concentration data. Deploying air pollution monitoring stations in urban areas generates a massive amount of collected data, creating databases suitable for statistical analysis. In addition, machine learning methods do not require a deep understanding of the dynamic and chemical processes between air pollutants and other relative atmospheric variables.

ANNs have been improved through years of research and applications, bringing more evolved versions to air pollution prediction. For example, [27] proposed a BPNN to estimate the hourly concentrations of NO<sub>2</sub> in unsampled locations at Algeciras Bay (Spain) using the historical values of NO<sub>2</sub> concentrations at fourteen monitoring stations and the distances to each monitoring station. The prediction system has a first stage that applies the IDW interpolation and a multiple linear regression method to produce air pollution maps that are used as input of the the BPNN. The highest result achieved was an  $R^2$  value of 0.76. They also evaluated the methods separately, showing the BPNN as the best prediction method in most monitoring stations. Instead, [6] used several methods to obtain daily estimates of PM<sub>2.5</sub> concentration across the contiguous US, and the results showed a better predictive performance of spatial statistical models over machine learning methods. The 829 monitor stations take measurements every 1, 3, or 6 days, with only approximately 15% of monitors sampling daily, implying an irregular and sparse dataset. In other recent research, [8] applied a BPNN for spatially estimating pollutant concentrations in the metropolitan city of Athens in Greece. Five pollutants were estimated, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> and SO<sub>2</sub>, and the  $R^2$  values for O<sub>3</sub> and PM<sub>10</sub> were above 0.87, and for NO<sub>2</sub>, PM<sub>2.5</sub>, and SO<sub>2</sub> were 0.76, 0.69 and 0.55, respectively.

Deep learning has become increasingly widely used in air quality prediction because of its ability to extract complex and non-linear spatiotemporal correlations on large datasets. Many researchers employed a Long Short Term Memory (LSTM) network for modelling the complex and non-linear temporal correlation of the historical values of the pollutants [9, 20, 25, 28, 29]. LSTM is an enhanced version of the Recurrent Neural Network for handling long-time sequence data. The recent studies [11, 24, 30] have proposed combined models for air quality prediction based on Convolutional Neural Networks (CNNs) to extract the spatial characteristics, and LSTMs to predict future air pollution concentrations. [24] developed a CNN-LSTM method for predicting the next

day's daily average PM<sub>2.5</sub> concentration in Beijing City. [30] proposed an attention-based CNN-LSTM multilayer structure to predict the PM<sub>2.5</sub> concentration in the next 72 hours at Beijing-Tianjin-Hebei region. This research analysed the historical air quality and meteorological data of 100 monitoring stations for spatiotemporal correlation. [11] proposed spatiotemporal forecasting models of Beijing's Air Quality Index. Four methods (CNN, LSTM, CNN-LSTM, BPNN) were evaluated to extract the spatiotemporal characteristics of air quality concentration data (hourly PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and CO) and the relations with meteorological and spatiotemporal data. The method that showed the best performance in next-hour forecasting was the CNN-LSTM method. [12] developed a spatiotemporal air quality prediction model based on LSTMs. The input data were Beijing's historical concentrations of PM<sub>2.5</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and CO and meteorological data. The output was the concentration sequence of PM<sub>2.5</sub>, CO, NO<sub>2</sub>, O<sub>3</sub>, and SO<sub>2</sub> at four monitoring sites. The model's prediction accuracy is high, as shown by the best  $R^2$  value of the four analysed sites: 0.939 for PM<sub>2.5</sub>, 0.847 for CO, 0.875 for NO<sub>2</sub>, 0.935 for O<sub>3</sub>, and 0.809 for SO<sub>2</sub>. Some researchers took advantage of the CNN's ability to process sequence-structure data for air pollutant concentration prediction. For example, [31] used a five-layer CNN to extract the temporal correlation from historical observation data and predict the ozone concentration in the next 24 hours in an urban area. [7] applied the IDW method to interpolate air quality and weather data collected in South Korea and then used the interpolation as input of the CNN to predict PM<sub>2.5</sub> and PM<sub>10</sub> concentrations. The results show an effective prediction performance with an  $R^2$  higher than 0.97.

In our research, we have deployed and compared six air quality prediction methods present in the recent literature: spatial-based interpolation (IDW), geostatistical model (Kriging), machine learning (SVM and FFNN), and deep learning (FFNN with 2 - 3 hidden layers). Each method is evaluated for extracting spatial dependencies of seven air pollutants: NO<sub>2</sub>, NO, CO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>. Therefore, we have broadly compared different methods to predict the concentrations of several air pollutants that are generated by various sources and that have different behaviours. The analysis of so many air pollutants is rare in the literature. The urban area selected has an air quality monitoring network with a spatial density and a sample time higher or similar to recent studies.

## 3 Study area and dataset

The area of study is the city of Madrid, which is the capital of Spain and the largest and most populated metropolitan area of the country. Madrid's province population has grown

from 6,446 million in 2016 to 6,751 million in 2021<sup>3</sup>. The elevation at its centre (40° 25' N, 3° 41' W) is 657 m. Madrid area's expected mean temperature changes from 9.8 °C in January to 32.1 °C in July, experiencing cold winters and hot summers. Spring and autumn are the seasons with more expected rainy days, while the summer months are usually dry and sunny<sup>4</sup>.

Madrid's air pollution levels are high, although, since the activation of air quality policies in 2011, those levels were effectively reduced [32]. This study analysed hourly time series of four air pollutants (NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, and SO<sub>2</sub>) monitoring in Madrid urban area during the period from 2001 to 2017 by a two-stage method: first, a Hidden Markov Model was used to characterize the air pollution at temporal scales. Then, the spatial distribution was analysed by combining the interpolation results of Ordinary Kriging and Inverse Distance Weighting. [32] concludes that the air pollution spatial analysis is challenging to assess due to meteorological and physical factors and the regional contributions originated in adjacent municipalities. Not only is human activity responsible for bad air quality, but also other climate events like Saharan dust intrusions have an impact by rising PM levels [32, 33]. The research [34] examined the effects of local road traffic, meteorological conditions, and temporal variables on air pollution in Madrid. Its result showed that air pollutant levels were weakly linked to local vehicular emissions because various elements affect the pollutant concentration, mainly meteorological agents, topography, tree and shrub presence, building distribution, and water streams like rivers.

This study uses hourly air quality data measured between January 2016 and August 2018 by the air monitoring network of Madrid. Madrid's city council operates an air quality monitoring network from 2001 and publishes both real-time and historical air quality data<sup>5</sup>. We used the Dataset "Air Quality in Madrid"<sup>6</sup> that contains the processed data from the files offered by the Madrid's city council with a structured organization based on timestamp and standard format data. It consists of a file for each year where each row is timestamped and the columns are the different measures performed at that point in time in a certain station. In addition, the information

regarding each station (identifier, name, address, coordinates and elevation) is available in another file. The measurements of many pollutants are available (NO, NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, CO, PM<sub>2.5</sub>, PM<sub>10</sub>, toluene, benzene, methane), but not every station is equipped with all air pollution sensors, which have been increasing over the years.

The study area and the distribution of 24 air quality monitoring stations deployed in Madrid are shown in Fig. 1. The maximum distance between two stations is 7 kilometres, and the average distance between stations at the urban centre is 3. Table 1 contains the coordinates, elevation and the air pollutant measured in each air quality monitoring station between 2016 and 2018. We selected the most harmful air pollutants according to WHO and several government agencies like the U.S. Environmental Protection Agency<sup>7</sup>, European Environment Agency<sup>8</sup> and Health Canada and Environment Canada<sup>9</sup>. The criteria for the selection of the monitoring sites are the hourly data availability along with homogeneous spatial data coverage for each air pollutant.

Table 2 presents the measuring units and the range of measured concentrations for each pollutant between 2016 and 2018.

## 4 Methods

In this work, we present a system based on FFNN architecture for regression to predict the air pollutant concentration in a specific location based on the measurements obtained from nearby monitored locations. To compare our proposal, we applied the following geostatistical and machine learning methods: Inverse Distance Weighting (IDW), Ordinary Kriging (OK), and Support Vector Machine (SVM).

### 4.1 Inverse distance weighting and kriging methods

Nearly all spatial interpolation methods share the same general estimation formula, which is as follows:

$$Z(x_0) = \sum_{i=1}^n (w_i z(x_i)) \quad (1)$$

where  $Z$  is the estimated value at the point of interest  $x_0$ ,  $z$  is the observed value at the sampled point  $x_i$ ,  $w_i$  is the weight

<sup>3</sup> Instituto Nacional de Estadística (National Statistics Institute): <https://www.ine.es/dynt3/inebase/index.htm?padre=1689&capsel=9041>. Retrieved October 13, 2023.

<sup>4</sup> Madrid, Retiro: Madrid, Retiro - Agencia Estatal de Meteorología - AEMET. Gobierno de España: <https://www.aemet.es/serviciosclimaticos/datosclimatologicos/valoresclimatologicos?l=3195&k=28>. Retrieved October 13, 2023.

<sup>5</sup> Open data portal of the Madrid City Council: <https://datos.madrid.es/portal/site/egob>. Retrieved October 13, 2023.

<sup>6</sup> Kaggle Air Quality in Madrid (2001-2018): <https://www.kaggle.com/datasets/decide-soluciones/air-quality-madrid>. Retrieved October 13, 2023.

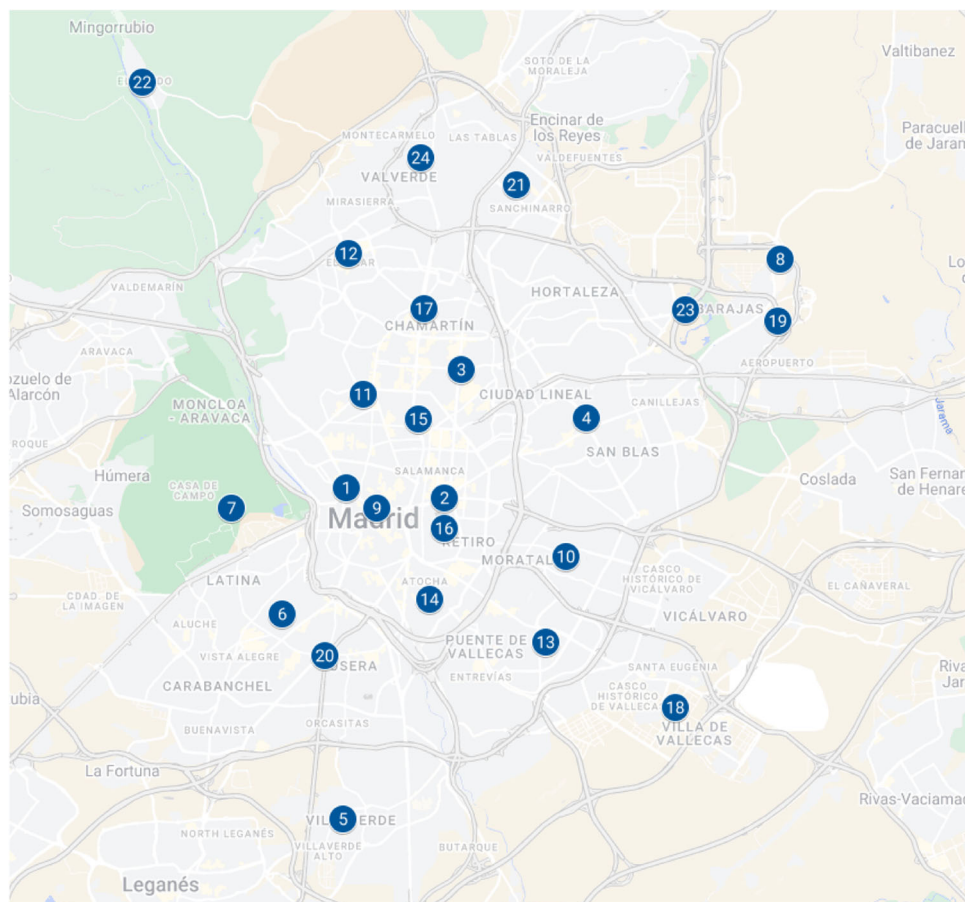
<sup>7</sup> AIRNOW: <https://www.airnow.gov/aqi/aqi-basics>. Retrieved October 13, 2023

<sup>8</sup> European Air Quality Index: <https://www.eea.europa.eu/themes/air/air-quality-index>. Retrieved October 13, 2023

<sup>9</sup> Canada Air Quality Health Index: <https://www.canada.ca/en/environment-climate-change/services/air-quality-health-index/about.html>. Retrieved October 13, 2023



**Fig. 1** Area of study and position of the different air monitoring stations. Map image is obtained from Google Maps



assigned to the sampled point, and  $n$  represents the number of sampled points used for the estimation. The difference between the methods depends on the formula to calculate the weights. The two most commonly used interpolation methods in the literature are IDW and OK [35, 36]. The IDW method uses the following expression for the weight:

$$w_i = \frac{\frac{1^p}{d_i}}{\sum_{i=1}^n \frac{1^p}{d_i}} \quad (2)$$

where  $d_i$  is the distance between  $x_0$  and  $x_i$ , and  $p$  is an exponent that determines the influence of values closest to the interpolated point, while the weight for OK is estimated by minimizing the variance of the prediction errors. It is assumed that the data are part of an intrinsic function  $z(x)$  with the sample variogram [37]. The sample variogram is fitted with specific known positive defined functions. The most common functions are linear, spherical, exponential, and Gaussian.

In our experiments, we tested the IDW model for  $p=1$  and  $p=2$  and, for the implementation of the OK method, we applied four function models for fitting the sample variogram: spherical, exponential, Gaussian and bounded linear.

## 4.2 Support vector machine

SVM is a popular machine learning tool for classification, but it can also be used for regression analysis [38, 39]. SVMs aim to provide a nonlinear function to map a given training data set  $D: (x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$  to a high dimensional feature space. In this space, a hyperplane is optimized to be within a certain threshold of the selected data, called the support vectors, and the hyperplane is used for predicting regression.

A linear epsilon-insensitive ( $\epsilon$ ) SVM was used for regression, which is also known as L1 loss. The set of training data included predictor variables and observed response values. The goal was to find a function  $z(x)$  that deviates from observed values no greater than  $\epsilon$  for each training point  $x$ , and that is as flat as possible at the same time. The training of the SVM with epsilon-insensitive loss function was performed by using quadratic programming for minimising the objective-function.

## 4.3 Feed-forward neural network for regression

ANNs are massively parallel interconnected networks of simple, hierarchically organized elements (artificial neurons)

**Table 1** Air quality monitoring stations, coordinates, and measured pollutants (from 2016 to 2018)

Station ID	Longitude	Latitude	Height (m)	Measured pollutant
1	-3.7122472	40.4238528	635	NO2, NO, CO, SO2
2	-3.6823194	40.4215638	670	NO2, NO, CO, SO2, O3, PM2.5, PM10
3	-3.6773555	40.4514750	708	NO2, NO
4	-3.63923333	40.4400472	693	NO2, NO, CO, O3
5	-3.7133222	40.3471389	604	NO2, NO, SO2, O3,
6	-3.7318527	40.3947805	630	NO2, NO, CO, SO2, O3, PM10
7	-3.7473472	40.4193555	642	NO2, NO, CO, SO2, O3, PM2.5, PM10
8	-3.5800305	40.4769278	621	NO2, NO, O3
9	-3.7031722	40.4192083	659	NO2, NO, CO, O3
10	-3.6453056	40.4079472	685	NO2, NO, CO, SO2, PM10
11	-3.7071278	40.4455444	698	NO2, NO, SO2, PM2.5, PM10
12	-3.7115417	40.4782278	674	NO2, NO, CO, O3
13	-3.6515222	40.3881527	677	NO2, NO, SO2, PM10
14	-3.686825	40.3981138	599	NO2, NO, PM2.5, PM10
15	-3.6903667	40.4398972	676	NO2, NO, PM2.5, PM10
16	-3.6825833	40.4144444	662	NO2, NO, O3
17	-3.68876945	40.4655722	728	NO2, NO, PM2.5, PM10
18	-3.6121167	40.3729333	627	NO2, NO, O3,
19	-3.5807472	40.4625305	618	NO2, NO, PM10
20	-3.7187278	40.3849639	604	NO2, NO, CO, O3
21	-3.6605028	40.4942083	700	NO2, NO, CO, SO2, PM10
22	-3.7746111	40.5180583	615	NO2, NO, O3
23	-3.60907222	40.4652500	660	NO2, NO, O3
24	-3.6897611	40.5005889	715	NO2, NO, O3, PM10

that attempt to interact with the environment in the same way as the biological nervous system [40]. The output of such an artificial neuron can be calculated using the (3).

$$y = f\left(\sum_{i=1}^n (w_i x_i)\right) + b \tag{3}$$

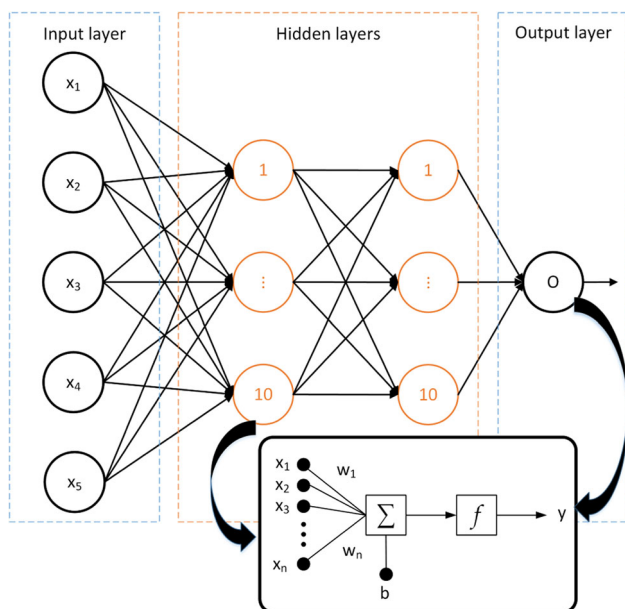
where  $x_i$  are the inputs,  $n$  the number of inputs,  $w_i$  the synaptic weights,  $b$  the threshold and  $f$  the activation function. The most commonly used activation functions are linear, sigmoid, and hyperbolic tangent. Artificial neurons are arranged in several layers and connected by synaptic weights.

In this work, three feed-forward, fully connected neural networks were used for regression. The structure includes an input layer, one or more hidden layers, and an output layer. The input layer takes information (predictor data) from the

domain and passes it to all the neurons from the first hidden layer. As the first hidden layer is fully connected to the input layer, each subsequent layer is connected to all the neurons from the previous layer. Each neuron of a fully connected layer multiplies the input by the synaptic weight and then adds the multiplication results with the bias. The sum is passed through an activation function. The final fully connected layer produces the network’s output (predicted response values). The proposed architecture for the fully connected layered neural network with two hidden layers is shown in Fig. 2. An enlarged diagram of a single artificial neuron is presented separately to show its five components -inputs, synaptic weights, sum, bias, and activation function. We chose two activation functions for hidden layers: a Rectified Linear Unit (ReLU) and a sigmoid function. These functions are described in (4) and (5), respectively. Accord-

**Table 2** Measuring units and range of measured concentrations for each pollutant (years 2016 to 2018)

NO2 ( $\mu\text{g}/\text{m}^3$ )	NO ( $\mu\text{g}/\text{m}^3$ )	CO ( $\text{mg}/\text{m}^3$ )	SO2 ( $\mu\text{g}/\text{m}^3$ )	O3 ( $\mu\text{g}/\text{m}^3$ )	PM2.5 ( $\mu\text{g}/\text{m}^3$ )	PM10 ( $\mu\text{g}/\text{m}^3$ )
[1, 349]	[1, 973]	[0.1, 4.9]	[1, 98]	[1, 193]	[0, 96]	[0, 318]



**Fig. 2** Fully interconnected bi-layer FFNN with 5 input nodes, two hidden layers with 10 neurons and the output layer with one neuron. An enlarged diagram of a single artificial neuron is presented separately to show its five components -inputs, synaptic weights, sum, bias, and activation function

ing to the regression problem, the activation function of the output layer is the linear function  $f(x) = x$ .

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4)$$

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

The training is based on the limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm (LBFGS) [41], where the mean squared error (MSE) is minimized.

We proposed three FFNN architectures for extracting the spatial characteristics of the air pollution concentration: one hidden layer neural network, a bi-layer neural network, and a tri-layer neural network. Regarding the current application, the number of monitoring sites for predicting each pollutant defines the number of input nodes, and the output layer consists of a fully connected neuron to return the prediction. The number of neurons in each hidden layer is configurable. In order to compare the same structure for different air pollutants, we determined 10 neurons per hidden layer for each architecture.

## 5 Experiments

Figure 3 shows the diagram of the experimental workflow. It was followed for the evaluation of several prediction models to estimate each air pollutant.

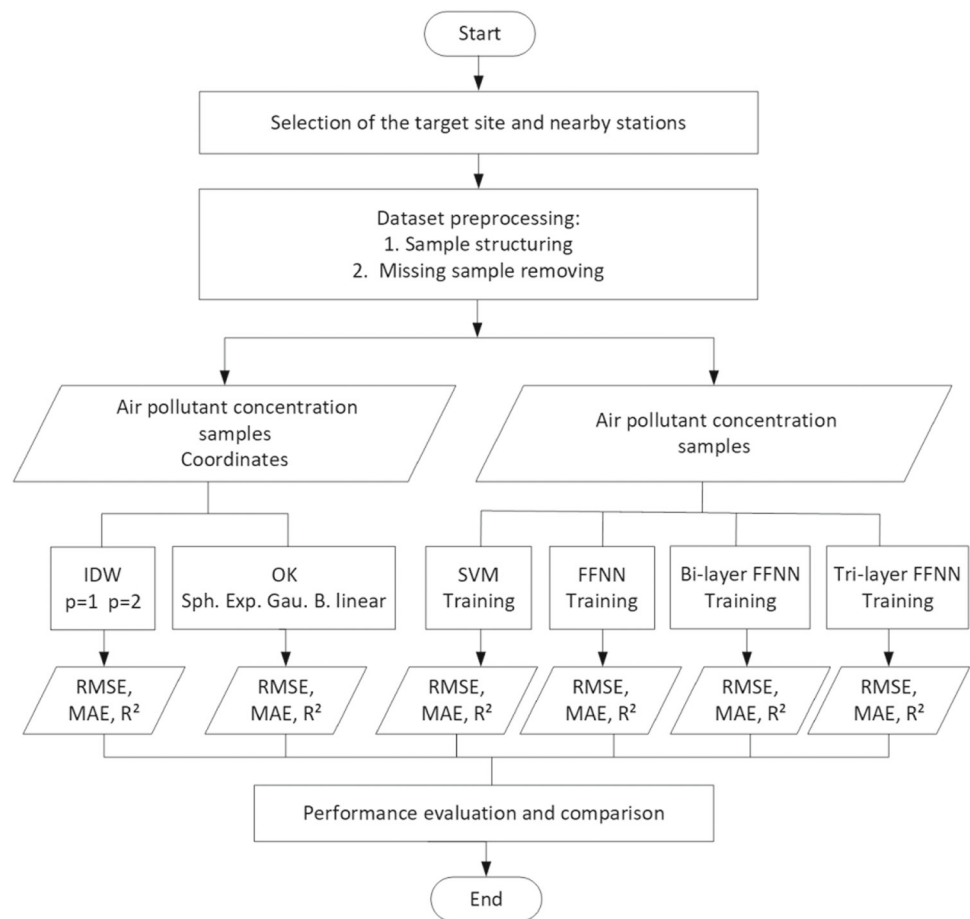
Firstly, we selected the target location, which should be a point with historical concentrations of the air pollutant to be estimated. We chose target stations based on their position close to the centre of Madrid and the five stations closest to the target site. However, our decision was constrained by the availability of air pollution sensors in each station. Table 3 shows the monitoring station selected as the target site and five nearby monitoring stations used to estimate the air pollutant concentrations. The average distance between the target and the selected stations is 3.26 kilometres. The coordinates and distribution of the air quality monitoring stations are shown in Table 1 and in Fig. 1, respectively.

Secondly, we processed the dataset to get the air pollutant concentrations from the selected stations. Each sample included the air pollutant concentrations for each station taken simultaneously. The samples with some missing values were removed. The input values for IDW and OK models are the latitudes, longitudes, and pollutant concentration at the five nearby stations. However, the input data for SVM and FFNN are the historical values of the pollutants from six stations, for training, and the pollutant concentration at the five nearby stations, for predicting.

Then, the predictive models were designed. We developed two IDW models ( $p=1$  and  $p=2$ ) for evaluating the influence of the distance in the prediction. When  $p=2$ , the method is known as the inverse distance squared weighted interpolation. To implement the OK method, four function models were applied to fit the sample variogram: spherical, exponential, Gaussian, and bounded linear. We proposed a linear  $\epsilon$ -SVM for regression and three different FFNN architectures: one hidden layer neural network, a bi-layer neural network, and a tri-layer neural network. The number of monitoring sites for predicting each pollutant defines the number of input nodes, a parameter set to five in our system. The output layer consists of a fully connected neuron to return the prediction. The number of neurons in each hidden layer is a configurable parameter, and we set it to ten to compare the same structure for different air pollutants. We tested two activation functions for hidden layers: a Rectified Linear Unit (ReLU) and a sigmoid function.

In the case of the SVM and FFNN models, the next phase is training. In both cases, we used 80% of the samples to train the neuronal network model and the remaining 20% to test the performance of the trained model with new data. Figure 4 shows the data and processes of training stage. The results are the trained FFNN (or SVM) and the difference and correlation measures. The IDW and OK models do not require training since the estimated value is calculated from the simultaneous measurements taken at the nearby stations and the distances to the target site. We used the test set to validate the SVM and FFNN models and the whole dataset to evaluate the IDW and OK models. The accuracy of each model is based on the comparison of the observed

**Fig. 3** Experimental workflow followed for the prediction of each air pollutant and the evaluation of the proposed models



and predicted concentrations and the statistical analysis of the residual values.

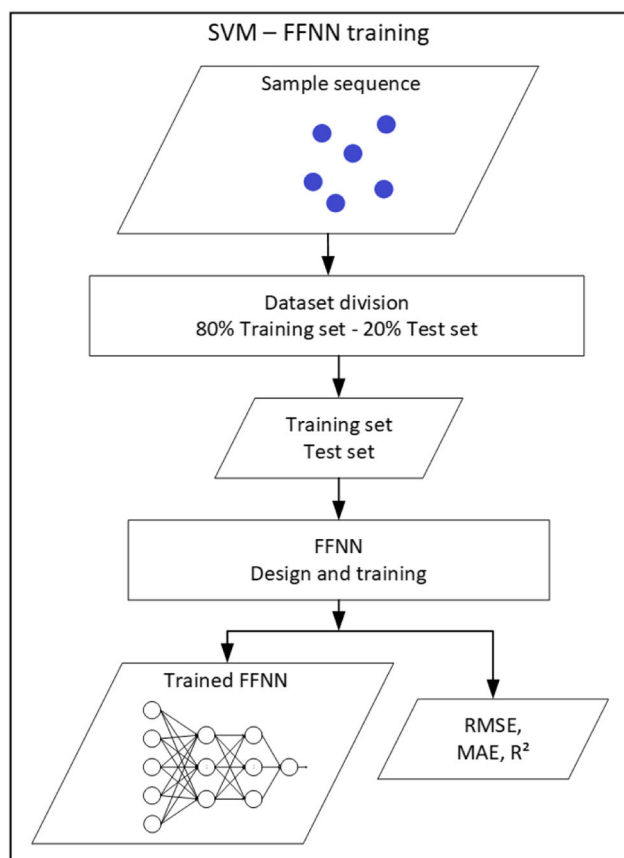
Finally, we determined the mean performance of each model by using a set of difference and correlation measures: Mean Absolute Error (MAE), Root Mean Square Error

(RMSE), and the coefficient of determination ( $R^2$ ).  $R^2$  is the proportion of variation in the dependent variable that is predicted by the statistical model (range from 0 to 1). Therefore,  $R^2$  provides information about the model's goodness of fit. The following equations determine these error measures

**Table 3** The selected monitoring stations and distances to the target site for each air pollutant under evaluation

Pollutant	Target site	Monitoring stations and distance to target site				
NO <sub>2</sub> and NO	15	1	2	3	9	11
		2.57 km	2.15 km	1.69 km	2.54 km	1.55 km
CO	2	1	6	9	10	20
		2.55 km	5.15 km	1.79 km	3.49 km	5.11 km
SO <sub>2</sub>	2	1	6	10	11	13
		2.55 km	5.15 km	3.49 km	3.39 km	4.54 km
O <sub>3</sub>	9	2	4	6	16	20
		1.79 km	5.90 km	3.64 km	1.82 km	4.02 km
PM <sub>2.5</sub>	15	2	7	11	14	17
		2.15 km	5.35 km	1.55 km	4.65 km	2.85 km
PM <sub>10</sub>	15	2	10	11	14	17
		2.15 km	5.22 km	1.55 km	4.65 km	2.85 km





**Fig. 4** The training phase flowchart of SVM and FFNN methods

between observed ( $x_i$ ) and predicted ( $y_i$ ) values, where  $n$  represents the number of sampled points used for the estimation and  $\mu_x$  is the mean of observed values.

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \mu_x)^2} \quad (8)$$

This process was followed for the seven selected pollutants: NO, NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, CO, PM<sub>2.5</sub> and PM<sub>10</sub>.

## 6 Results

Six methods for modelling the spatial characteristics of the air pollutant concentrations were implemented and evaluated for each of the seven air pollutants (NO<sub>2</sub>, NO, CO,

SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>). The experiments used historical pollutant data from monitoring stations of Madrid city's air quality monitoring network collected from January 1, 2016, to December 31, 2018. Among the different model variants that were tested for the IDW and OK methods, only the best ones are shown in the results. We proposed three FFNN architectures: a FFNN with one fully connected layer (FCL), two FCLs, and three FCLs. Except in the prediction of SO<sub>2</sub> concentration, we chose to apply the ReLU function in the hidden layers because it has a greater accuracy than the sigmoid one.

Tables 4, 5, 6, 7, 8, 9 and 10 present the assessment of the statistical (IDW and OK) and machine learning (SVM and FFNN) methods by RMSE, MAE and  $R^2$ . RMSE and MAE are in the same units of pollutant concentrations, and  $R^2$  ranges from 0 to 1.

The results for predicting NO<sub>2</sub> and NO showed a better predictive performance of machine learning methods over spatial statistical models. The bi-layer FFNN method presents maximum  $R^2$  and minimum RMSE for estimating NO<sub>2</sub> and NO air pollutants. The proposed models for estimating the CO concentrations result in low accuracy. Neither method is appropriate to extract the spatial characteristics of the CO concentrations. A possible reason for this result may be that CO is a primary pollutant whose concentration is closely linked to local combustion phenomena, aggravated in some cases by low dispersion due to thermal inversion or lack of wind. Heavy traffic or traffic jams in a specific area can cause high local measurements of CO without hardly affecting neighbouring areas. A higher density of measurement stations could be needed to get better results for CO because its concentration relative to the distance from the pollution source can decrease quickly. In the case of SO<sub>2</sub>, the accuracy of the machine learning methods is vastly superior to that of the spatial statistical models. FNN methods experiment an exceptional improvement when increasing the number of hidden layers. The prediction models of O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub> present similar performances, with slightly higher accuracy for bi-layer FFNN methods.

The bi-layer FFNN method exhibits the best result for most of the air pollutants under evaluation, except for the prediction of SO<sub>2</sub>, where the best performance is led by the tri-layer NN. The highest coefficient of determination (0.9) is reached by the bi-layer FFNN method for the prediction of NO<sub>2</sub> and PM<sub>10</sub>. In most cases, we could rank the methods according to their performance (from best to worst) as follows: bi-layer FFNN, tri-layer FFNN, FFNN, SVM, IDW, and OK. The IDW and OK methods present similar results and have the lowest accuracy for all the evaluated air pollutants. The SVM models exhibit lower predictive ability than the FFNN methods for fitting air pollution concentration.

**Table 4** Overall performance metrics for the prediction of NO<sub>2</sub>

	IDW p=2	OK B. linear	SVM	FFNN		
				1 FCL	2 FCL	3 FCL
RMSE ( $\mu\text{g}/\text{m}^3$ )	14.06	14.79	10.09	8.67	8.61	8.65
MAE ( $\mu\text{g}/\text{m}^3$ )	10.51	11.15	6.86	6.15	6.04	6.05
R <sup>2</sup>	0.84	0.83	0.86	0.89	0.90	0.89

**Table 5** Overall performance metrics for the prediction of NO

	IDW p=2	OK B. linear	SVM	FFNN		
				1 FCL	2 FCL	3 FCL
RMSE ( $\mu\text{g}/\text{m}^3$ )	25.16	26.31	15.75	14.67	14.30	14.81
MAE ( $\mu\text{g}/\text{m}^3$ )	12.96	13.73	6.54	5.86	5.71	5.71
R <sup>2</sup>	0.77	0.76	0.79	0.82	0.83	0.82

**Table 6** Overall performance metrics for the prediction of CO

	IDW p=1	OK Expon.	SVM	FFNN		
				1 FCL	2 FCL	3 FCL
RMSE ( $\mu\text{g}/\text{m}^3$ )	0.178	0.179	0.177	0.162	0.164	0.162
MAE ( $\mu\text{g}/\text{m}^3$ )	0.104	0.105	0.093	0.090	0.089	0.089
R <sup>2</sup>	0.51	0.51	0.49	0.57	0.57	0.57

**Table 7** Overall performance metrics for the prediction of SO<sub>2</sub>

	IDW p=2	OK Gaussian	SVM	FFNN		
				1 FCL	2 FCL	3 FCL
RMSE ( $\mu\text{g}/\text{m}^3$ )	6.15	6.15	3.77	2.94	2.51	2.43
MAE ( $\mu\text{g}/\text{m}^3$ )	4.70	4.70	2.87	2.13	1.73	1.62
R <sup>2</sup>	0.14	0.12	0.49	0.69	0.78	0.79

**Table 8** Overall performance metrics for the prediction of O<sub>3</sub>

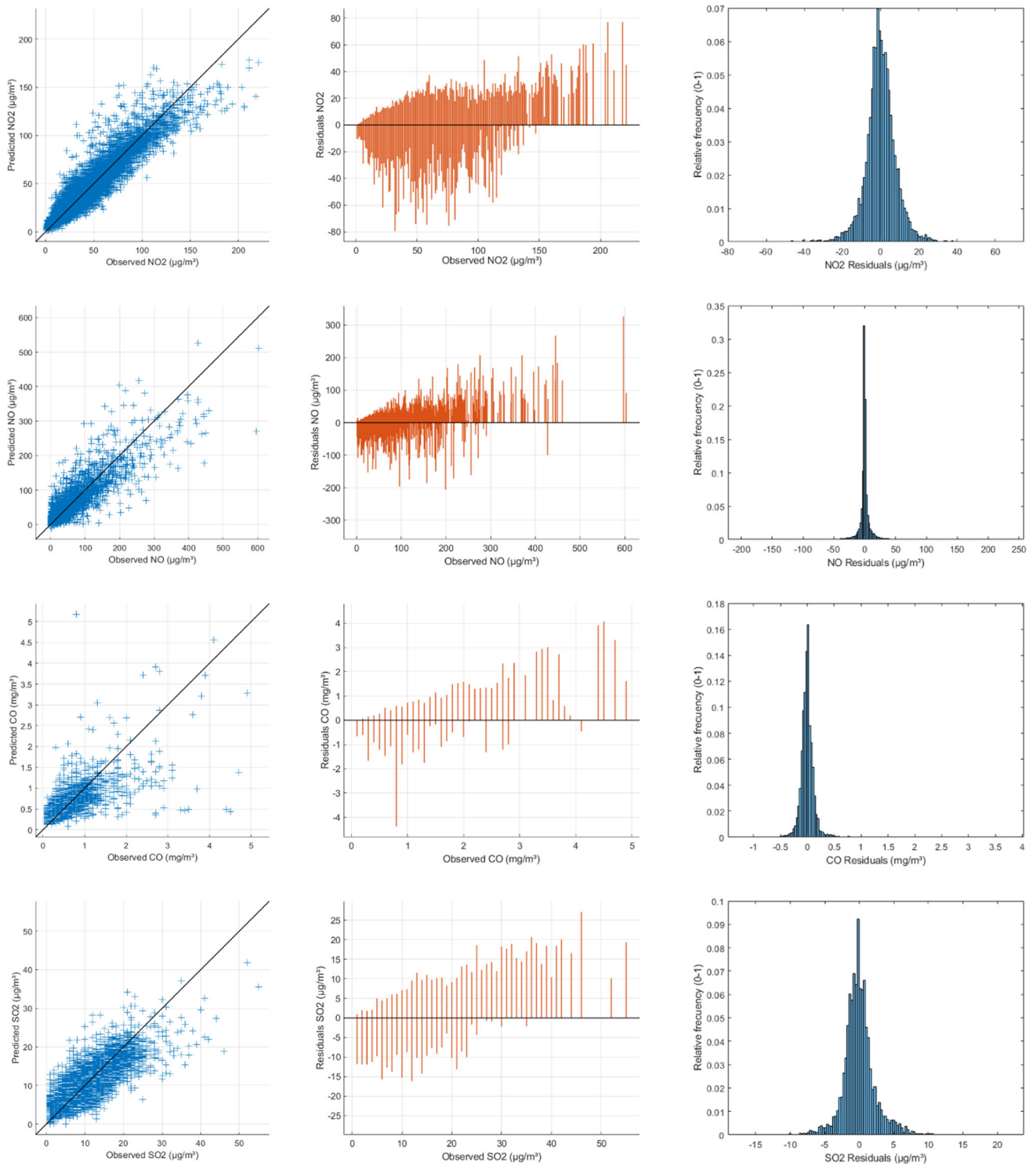
	IDW p=2	OK B. linear	SVM	FFNN		
				1 FCL	2 FCL	3 FCL
RMSE ( $\mu\text{g}/\text{m}^3$ )	14.35	14.63	10.79	9.60	9.49	9.48
MAE ( $\mu\text{g}/\text{m}^3$ )	10.74	10.89	7.98	7.19	7.06	7.06
R <sup>2</sup>	0.84	0.84	0.85	0.88	0.88	0.88

**Table 9** Overall performance metrics for the prediction of PM<sub>2.5</sub>

	IDW p=1	OK Expon.	SVM	FFNN		
				1 FCL	2 FCL	3 FCL
RMSE ( $\mu\text{g}/\text{m}^3$ )	3.84	3.91	3.60	3.53	3.49	3.52
MAE ( $\mu\text{g}/\text{m}^3$ )	2.59	2.65	2.35	2.34	2.18	2.33
R <sup>2</sup>	0.73	0.72	0.74	0.75	0.75	0.75

**Table 10** Overall performance metrics for the prediction of PM<sub>10</sub>

	IDW p=1	OK Expon.	SVM	FFNN		
				1 FCL	2 FCL	3 FCL
RMSE ( $\mu\text{g}/\text{m}^3$ )	5.63	5.53	4.78	4.50	4.52	4.49
MAE ( $\mu\text{g}/\text{m}^3$ )	3.79	3.74	3.01	2.89	2.90	2.89
R <sup>2</sup>	0.88	0.89	0.90	0.91	0.91	0.91



**Fig. 5** FFNN prediction performance: (Left) scatter plots showing observed versus predicted concentrations for the pollutants based on the 20% of data used for validation. (Centre) Residual plot of the difference between the predicted and true responses. (Right) Residual histograms showing the relative probability of error values

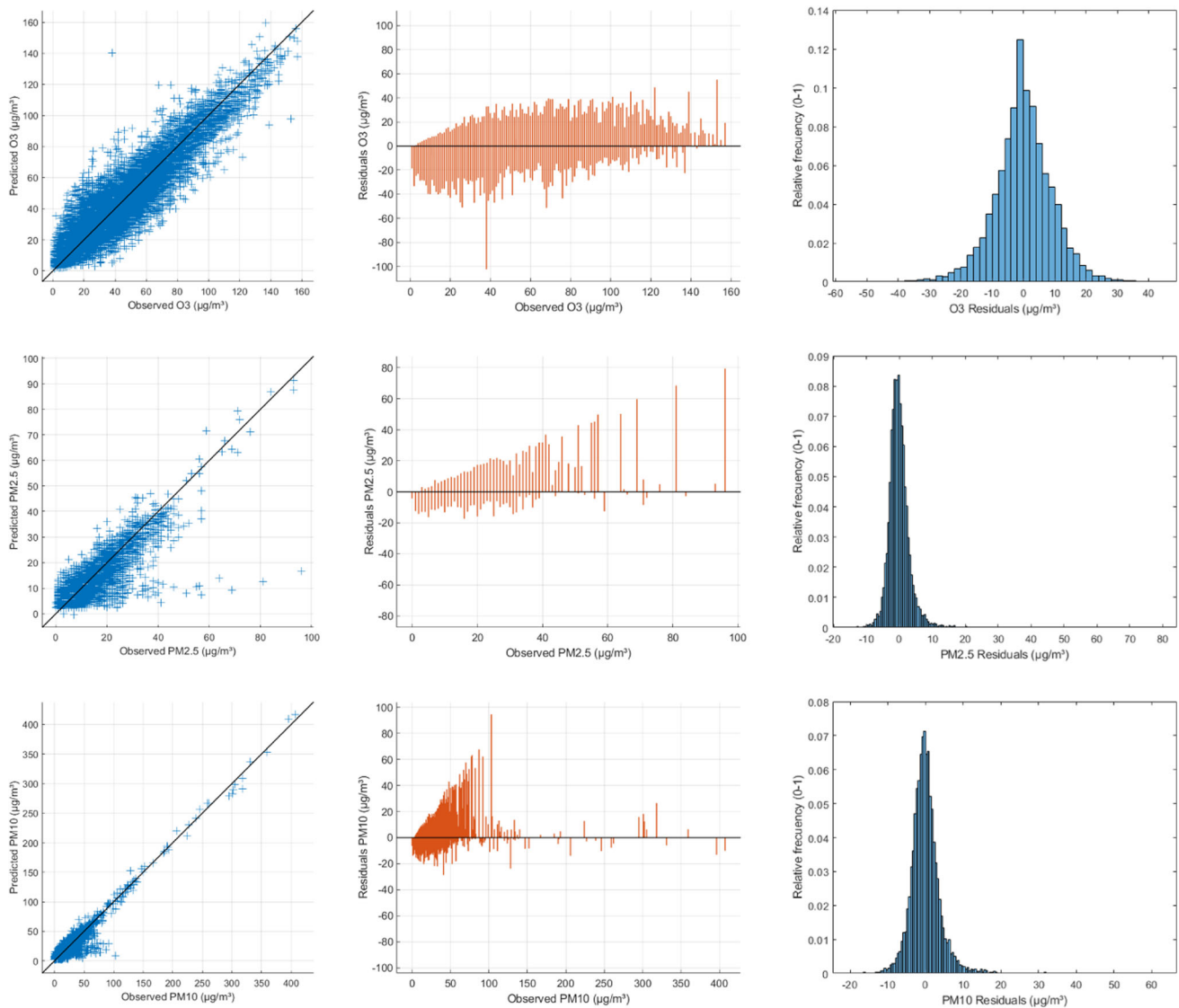


Fig. 5 continued

Figure 5 shows the bi-layer FFNN performance for the seven air pollutants (NO, NO<sub>2</sub>, CO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub> and PM<sub>10</sub>), based on the 20% of data used for validation. In the left column, the scatter plots of observed versus predicted concentrations for each air pollutants are shown. In all cases, low dispersion is observed along the diagonal of the optimum prediction. The centre column shows the residual plots that display the difference between the predicted and measured values. The right column contains the residual histogram showing the relative probability of predictive errors. A general remark is that many error values are around zero and residuals distributions are in accordance with the mean model performance metrics.

## 7 Discussion

The main objective of this work was estimating air pollution concentration in a certain location based on the measurements taken at nearby stations for addressing the missing values and detecting uncalibrated sensors. We have developed several statistical (IDW and OK) and machine learning (SVM, FFNN, bi-layer FFNN, tri-layer FFNN) methods to model the spatial characteristics of the concentrations of seven air pollutants (NO<sub>2</sub>, NO, CO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>) measured by Madrid's air quality monitoring network. The models were evaluated and compared using MAE, RMSE, and  $R^2$  as accuracy indicators.

IDW and OK's statistical models reached  $R^2$  greater than 0.75 for NO<sub>2</sub>, NO, O<sub>3</sub>, and PM<sub>10</sub>. The prediction accuracy of the FFNN methods is better than the results of IDW and OK for all analysed air pollutants. There is a more significant accuracy difference between geostatistical models and FFNN methods for predicting NO<sub>2</sub>, NO, and SO<sub>2</sub> than O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>. The results show the effectiveness of the bi-layer FFNN model to fit spatial correlation of the air pollution concentration of NO<sub>2</sub>, NO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub> PM<sub>10</sub>. For the prediction of SO<sub>2</sub>, the tri-layer FFNN model improves the bi-layer FFNN accuracy. Therefore, the proposed systems have a direct application to provide missing values of the air monitoring network and can be used to detect uncalibrated sensors. Neither method is appropriate to extract the spatial characteristics of the CO concentrations. A possible reason for this result may be that the CO concentration is closely linked to local combustion phenomena. A higher density of measurement stations could be needed to get better results for CO.

As mentioned in Section 2, the models proposed by [12] and [7] reach a very high  $R^2$  value regarding the rest of the recent research analysed in this section. The proposed bi-layer FFNN-based system reaches a better  $R^2$  value of 0.9 for the prediction of NO<sub>2</sub>. The prediction of SO<sub>2</sub> is less frequently analysed in the related recent studies and gets worse performance. [12] presents the highest R<sup>2</sup> value of 0.81. In this work, the tri-layer FFNN system reaches an  $R^2$  value of 0.79. [12] proposed a model based on an LSTM network with  $n$  hidden layers and a fully connected layer. Adding a fully connected layer improves the performance of the prediction model, according to the results presented in [12] and this work. The input data of the models proposed by [12] and [7] are air pollution concentrations collected at the monitoring stations and meteorological data. A possibility of future improvement is introducing an LSTM to extract temporal correlation and adding meteorological variables such as temperature, dew point, pressure, wind direction, and wind speed to data input.

Several ablation experiments were carried out to evaluate the effect of reducing components of the proposed neural network architecture. The effectiveness of hidden layer size was examined by two implementations of bi-layer FFNN with seventy and fifty percent reduction of neurons in hidden layers. For most of the pollutants analyzed, the performance

of the structure with a seventy percent reduction is similar to the full bi-layer FFNN, obtaining the same value of the R<sup>2</sup> metric and a slight degradation of MAE and RMSE. In the case of SO<sub>2</sub>, the R<sup>2</sup> value decreases by 0.02. In contrast, reducing the number of neurons in the hidden layers to fifty percent decreases the system's accuracy for all cases, as shown in Table 11 by the R<sup>2</sup> metric. The performance for PM<sub>10</sub> prediction decreases markedly with an R<sup>2</sup> value of 0.61. To verify the influence of the features on the proposed system performance, the farthest station measurements were removed from the input. For predicting NO<sub>2</sub> and NO, station 1 was eliminated. For CO and PM<sub>10</sub>, station 10. For SO<sub>2</sub>, O<sub>3</sub>, and PM<sub>2.5</sub>, stations 13, 4, and 7, respectively. The third entry in Table 11 shows the performance results for the system with four inputs instead of five. The R<sup>2</sup> value decreases for predicting all pollutants, with the most significant difference for SO<sub>2</sub>. These ablation experiments were also performed for the one-hidden FFNN and tri-layer FFNN systems, obtaining similar results to the bi-layer FFNN.

## 8 Conclusion

In this study, three FFNN architectures (with one, two, and three fully connected hidden layers) were implemented and evaluated for modelling the spatial correlation of the concentrations of seven air pollutants: NO<sub>2</sub>, NO, CO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>. A comparison with other exposure modelling approaches has been presented: an SVM model and two geostatistical models (IWD and OK). The input dataset is historical pollutant measurements collected by Madrid's air quality monitoring network from January 1, 2016, to December 31, 2018.

The performance results reveal that bi-layer FFNN and tri-layer FFNN systems are suitable for the spatial prediction of NO<sub>2</sub>, NO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub> concentration with an accuracy of ( $R^2$ ) 0.9, 0.83, 0.79, 0.88, 0.75, 0.91, respectively. The comparison results show that FFNN models are superior to geostatistical methods and slightly better than Support Vector Machines for fitting the spatial correlation of air pollutant measurements (NO<sub>2</sub>, NO, CO, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>) collected at nearby locations (less than 3.5 kilometres). For the prediction of NO<sub>2</sub> and SO<sub>2</sub> concen-

**Table 11** Effect of decreasing FFNN architecture

	NO <sub>2</sub>	NO	CO	SO <sub>2</sub>	O <sub>3</sub>	PM <sub>2.5</sub>	PM <sub>10</sub>
Full bi-layer FFNN	0.9	0.83	0.57	0.78	0.88	0.75	0.91
Half bi-layer FFNN	0.89	0.82	0.56	0.75	0.87	0.74	0.61
4-input bi-layer FFNN	0.89	0.82	0.55	0.71	0.87	0.74	0.9

The table compares R<sup>2</sup> metric for the performance of bi-layer FFNN in two ablation experiments: reducing by half the number of neurons in the hidden layers and removing an input feature



trations, the bi-layer FFNN and tri-layer FFNN models get a similar accuracy to the recent studies where the BPNN and deep neural network were developed.

In future work, we expect to introduce an LSTM neural network to extract the temporal correlation of air pollution concentration. Also, we will extend the system input with meteorological variables such as temperature, dew point, pressure, wind direction, and wind speed to data input to evaluate the prediction system performance.

**Funding** Funding for open access publishing: Universidad de Sevilla/CBUA. This work was supported by the Spanish grant (with support from the European Regional Development Fund) SANEVEC (TED2021-130825B-I00) and by AIRMOV (P009-21/E03).

**Data availability** The datasets analysed during the current study are available in the Air Quality in Madrid (2001-2018) repository <https://www.kaggle.com/datasets/decide-soluciones/air-quality-madrid>

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E (2020) Environmental and health impacts of air pollution: a review. *Front Public Health* 8:14
- Wolf K, Hoffmann B, Andersen ZJ, Atkinson RW, Bauwelinck M, Bellander T, Brandt J, Brunekreef B, Cesaroni G, Chen J et al (2021) Long-term exposure to low-level ambient air pollution and incidence of stroke and coronary heart disease: a pooled analysis of six european cohorts within the elapse project. *The Lancet Planetary Health* 5(9):e620–e632
- Tainio M, Andersen ZJ, Nieuwenhuijsen MJ, Hu L, De Nazelle A, An R, Garcia LM, Goenka S, Zapata-Diomedes B, Bull F et al (2021) Air pollution, physical activity and health: A mapping review of the evidence. *Environ Int* 147:105954
- Hu F, Guo Y (2021) Health impacts of air pollution in china. *Front Environ Sci & Eng* 15:1–18
- de la Salud OM, Weltgesundheitsorganisation WH, Organization EC (2021) for Environment, WHO global air quality guidelines: particulate matter (PM<sub>2.5</sub> and PM<sub>10</sub>), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide (World Health Organization, 2021)
- Berrocal VJ, Guan Y, Muyskens A, Wang H, Reich BJ, Mulholland JA, Chang HH (2020) A comparison of statistical and machine learning methods for creating national daily maps of ambient pm<sub>2.5</sub> concentration. *Atmos Environ* 222:117130
- Chae S, Shin J, Kwon S, Lee S, Kang S, Lee D (2021) Pm<sub>10</sub> and pm<sub>2.5</sub> real-time prediction models using an interpolated convolutional neural network. *Sci Rep* 11(1):11952
- Tzani CG, Alimisis A, Koutsogiannis I (2021) Addressing missing environmental data via a machine learning scheme. *Atmosphere* 12(4):499
- Han P, Mei H, Liu D, Zeng N, Tang X, Wang Y, Pan Y (2021) Calibrations of low-cost air pollution monitoring sensors for co, no<sub>2</sub>, o<sub>3</sub>, and so<sub>2</sub>. *Sensors* 21(1):256
- Araujo LN, Belotti JT, Alves TA, de Souza Tadano Y, Siqueira H (2020) Ensemble method based on artificial neural networks to estimate air pollution health risks. *Environ Model & Softw* 123:104567
- Yan R, Liao J, Yang J, Sun W, Nong M, Li F (2021) Multi-hour and multi-site air quality index forecasting in beijing using cnn, lstm, cnn-lstm, and spatiotemporal clustering. *Expert Syst Appl* 169:114513
- Seng D, Zhang Q, Zhang X, Chen G, Chen X (2021) Spatiotemporal prediction of air quality based on lstm neural network. *Alex Eng J* 60(2)
- Huang Y, Ying JJC, Tseng VS (2021) Spatio-attention embedded recurrent neural network for air quality prediction. *Knowl-Based Syst* 233:107416
- Thongthammachart T, Araki S, Shimadera H, Eto S, Matsuo T, Kondo A (2021) An integrated model combining random forests and wrf/cmaq model for high accuracy spatiotemporal pm<sub>2.5</sub> predictions in the kansai region of japan. *Atmos Environ* 262:118620
- Appel KW, Bash JO, Fahey KM, Foley KM, Gilliam RC, Hogrefe C, Hutzell WT, Kang D, Mathur R, Murphy BN et al (2021) The community multiscale air quality (cmaq) model versions 5.3 and 5.3. 1: system updates and evaluation. *Geosci Model Dev* 14(5):2867–2897
- Wang P, Wang P, Chen K, Du J, Zhang H (2022) Ground-level ozone simulation using ensemble wrf/chem predictions over the southeast united states. *Chemosphere* 287:132428
- Wang T, Li J, Pan J, Ji D, Kim Y, Wu L, Wang X, Pan X, Sun Y, Wang Z et al (2022) An integrated air quality modeling system coupling regional-urban and street models in beijing. *Urban Climate* 43:101143
- Kong L, Tang X, Zhu J, Wang Z, Li J, Wu H, Wu Q, Chen H, Zhu L, Wang W et al (2021) A 6-year-long (2013–2018) high-resolution air quality reanalysis dataset in china based on the assimilation of surface observations from cnmc. *Earth Syst Sci Data* 13(2):529–570
- Baklanov A, Zhang Y (2020) Advances in air quality modeling and forecasting. *Global Trans* 2:261–270
- Liu DR, Hsu YK, Chen HY, Jau HJ (2021) Air pollution prediction based on factory-aware attentional lstm neural network. *Comput* 103:75–98
- Maleki H, Sorooshian A, Goudarzi G, Baboli Z, Tahmasebi Birgani Y, Rahmati M (2019) Air pollution prediction by using an artificial neural network model. *Clean Techn Environ Policy* 21:1341–1352
- Webster R, Oliver MA (2007) *Geostatistics for environmental scientists*. Wiley
- Du S, Li T, Yang Y, Horng SJ (2019) Deep air quality forecasting using hybrid deep learning framework. *IEEE Trans Knowl Data Eng* 33(6):2412–2424
- Pak U, Ma J, Ryu U, Ryom K, Juhyok U, Pak K, Pak C (2020) Deep learning-based pm<sub>2.5</sub> prediction considering the spatiotemporal correlations: A case study of beijing, china. *Sci Total Environ* 699:133561

25. Zhang B, Zhang H, Zhao G, Lian J (2020) Constructing a pm<sub>2.5</sub> concentration prediction model by combining auto-encoder with bi-lstm neural networks. *Environ Model & Softw* 124:104600
26. Arsov M, Zdravevski E, Lameski P, Corizzo R, Koteli N, Gramatikov S, Mitreski K, Trajkovik V (2021) Multi-horizon air pollution forecasting with deep neural networks. *Sensors* 21(4):1235
27. Van Rooode S, Ruiz-Aguilar J, González-Enrique J, Turias I (2019) An artificial neural network ensemble approach to generate air pollution maps. *Environ Environ Monit Assess* 191:1–15
28. Zhang B, Zou G, Qin D, Lu Y, Jin Y, Wang H (2021) A novel encoder-decoder model based on read-first lstm for air pollutant prediction. *Sci Total Environ* 765:144507
29. Cordova CH, Portocarrero MNL, Salas R, Torres R, Rodrigues PC, López-Gonzales PC (2021) Air quality assessment and pollution forecasting using artificial neural networks in metropolitan lima-peru. *Sci Rep* 11(1):24232
30. Zhu J, Deng F, Zhao J, Zheng H (2021) Attention-based parallel networks (apnet) for pm<sub>2.5</sub> spatiotemporal prediction. *Sci Total Environ* 769:145082
31. Sayeed A, Choi Y, Eslami E, Lops Y, Roy A, Jung J (2020) Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Netw* 121:396–408
32. Gómez-Losada Á, Santos FM, Gibert K, Pires JC (2019) A data science approach for spatiotemporal modelling of low and resident air pollution in madrid (spain): Implications for epidemiological studies. *Comput Environ Urban Syst* 75:1–11
33. Linares C, Díaz J, Negev M, Martínez GS, Debono R, Paz S (2020) Impacts of climate change on the public health of the mediterranean basin population-current situation, projections, preparedness and adaptation. *Environ Res* 182:109107
34. Laña I, Del Ser J, Padró A, Vélez M, Casanova-Mateo C (2016) The role of local urban traffic and meteorological conditions in air pollution: A data-based case study in madrid, spain. *Atmos Environ* 145:424–438
35. Webster R, Oliver MA (2007) *Geostatistics for environmental scientists*. Wiley
36. Li J, Heap AD (2014) Spatial interpolation methods applied in the environmental sciences: A review. *Environ Model & Softw* 53:173–189
37. Wackernagel H (2003) *Multivariate geostatistics: an introduction with applications*. Springer, Sci & Bus Media
38. Hu K, Sivaraman V, Bhargubanda H, Kang S, Rahman A (2016) In 2016 IEEE SENSORS, IEEE, pp 1–3
39. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
40. Kohonen T (1988) *An introduction to neural computing*. *Neural Netw* 1(1):3–16
41. Nocedal J, Wright SJ (1999) *Numerical optimization*. Springer

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



**Elena Cerezuela-Escudero** received the B.S. degree in Computer Engineering in 2010 and Ph.D. degrees in Industrial Computing (specialized neuromorphic systems) in 2015 from the University of Seville, Spain. Since 2015, she is an Assistant Professor in the Computer Architecture and Technology Department at University of Seville, Spain. Since 2010, she is a member of the Robotics and Technology of Computers Lab. at the University of Seville and she has participated in several EU and Spanish research project in the area of neuromorphic systems. Her current research interests include vision/auditory neuromorphic systems, neural networks, signal processing, and their hardware implementations, in the field of robotics and smart cities.



**Juan Manuel Montes-Sanchez** received the B.S. degree in Electronic Engineering in 2015 and M.S. degree in Computer and Network Engineering in 2017. He has been an employee of the technology-based company COBER S.L. since 2016, and a part-time lecturer at Universidad de Sevilla since 2021, currently working in both industry and academic environments. He has participated in several EU and national research projects. His research is focused on industrial

applications of embedded AI and neuromorphic systems, which is also the topic of his ongoing PhD studies.



**Juan P. Dominguez-Morales** received the B.S. degree in computer engineering, the M.S. degree in computer engineering and networks, and the Ph.D. degree in computer engineering (specializing in neuromorphic audio processing and spiking neural networks) from the University of Seville, in 2014, 2015 and 2018, respectively. His Ph.D. was granted with a research grant from the Spanish Ministry of Education and Science. Since February 2023, he has been working as Associate Professor in the same university. His research interests include

neuromorphic engineering, spiking neural networks, deep learning and computational pathology.



**Lourdes Duran-Lopez** received the B.S. degree in Biomedical Engineering, the M.S. degree in Biomedical Research, and the Ph.D. degree in industrial computerized, robotic and neuromorphic systems from the University of Seville (Sevilla, Spain), in 2016, 2017 and 2021, respectively. Since 2023, she has been working as Assistant Professor in the Robotics and Technology of Computers Lab. from the University of Seville.

Her research interests include image processing, medical image analysis, computer-aided diagnosis systems and deep learning, particularly, convolutional neural networks.



**Gabriel Jimenez-Moreno** received the M.S. degree in physics (electronics) and the Ph.D. degree from the University of Seville, Seville, Spain, in 1987 and 1992, respectively. He was granted a Fellowship from the Spanish Science and Technology Commission (CICYT). Currently, he is an Associate Professor of computer architecture at the University of Seville. He participated in the creation of the Department of Computer Architecture (also at University of Seville). He is the

author of various papers and research reports on robotics, rehabilitation technology, and computer architecture. His research interests include neural networks, vision processing systems, embedded systems, computer interfaces, and computer architectures.