



# An artificial neural network ensemble approach to generate air pollution maps

S. Van Roode · J. J. Ruiz-Aguilar ·  
J. González-Enrique · I. J. Turias

Received: 28 March 2019 / Accepted: 17 October 2019 / Published online: 7 November 2019  
© Springer Nature Switzerland AG 2019

**Abstract** The objective of this research is to propose an artificial neural network (ANN) ensemble in order to estimate the hourly NO<sub>2</sub> concentration at unsampled locations. Spatial interpolation methods and linear regression models with regularization have been compared to perform the ensemble. The study case is based on the region of the Bay of Algeciras (Spain). This area is very industrialized and presents high concentrations of traffic. Air pollution data has been collected from the monitoring network maintained by the Andalusian Government in the region. On one hand, two totally different methods have been used and compared such as inverse distance weight (IDW) and least absolute shrinkage and selection operator (LASSO) in order to generate maps of pollutant concentration values. On the other hand, an ensemble approach has been developed using the outputs of the previous models. The ensemble is based on an ANN with backpropagation learning. An experimental procedure using cross-validation has been applied in order to compare the different models based on several performance indexes ( $R$  correlation coefficient, MSE, MAE and  $d$  index of fitness) and together to Friedman test and Bonferroni correction. The results reveal that the proposed ensemble approach presents better performance than single models in general terms. A validation

procedure has been conducted using a leave-one-out strategy using each monitoring station. IDW method presents an average value of  $R$  equals 0.72 and a maximum  $R$  equals 0.87, a minimum  $MSE$  equals 78.00, a minimum  $MAE$  equals 5.841 and a maximum  $d$  equals 0.913. LASSO presents an average value of  $R$  equals 0.76 and a maximum  $R$  equals 0.86, a minimum  $MSE$  equals 59.13, a minimum  $MAE$  equals 5.490 and a maximum  $d$  equals 0.900. Finally, the ANN ensemble shows an average value of  $R$  equals 0.77 and a maximum  $R$  equals 0.87, a minimum  $MSE$  equals 54.05, a minimum  $MAE$  equals 4.972 and a maximum  $d$  equals 0.915. The main objective has been to produce adequate atmospheric pollutant concentration maps and, therefore, to obtain estimations for locations that are distinct to the monitoring stations. Another objective has been to have in hand a system to produce robust measurements. This kind of system could be useful for missing data imputation and to find out reading errors (i.e. unexpected deviations or calibration problems) in some of the nodes of a network.

**Keywords** Air pollution maps · Artificial neural networks · Ensemble · IDW · LASSO

S. Van Roode (✉) · J. González-Enrique · I. J. Turias  
Intelligent Modelling of Systems, Department of Computer  
Science Engineering, University of Cádiz, Polytechnic School of  
Engineering, 11202 Algeciras, Spain  
e-mail: steffanie.vanroode@gm.uca.es

J. J. Ruiz-Aguilar  
Intelligent Modelling of Systems, Department of Civil and  
Industrial Engineering, University of Cádiz, Polytechnic School of  
Engineering, 11202 Algeciras, Spain

## Introduction

Nowadays, air pollution is one of the most important environmental problems and, therefore, particular attention should be paid to the promotion of monitoring and control air pollution strategies and systems. It is defined as the presence of certain substances harmful to humans and other living beings in the atmosphere. Air pollution comes from natural causes, such as the eruption of volcanoes, but especially comes from human activities, such as industry and the burning of fossil fuels, such as coal. Due to the great development of current industries, it is becoming increasingly difficult to stop it.

In Europe, the emissions of many atmospheric pollutants have been reduced significantly during the last decades, and as a result air quality has improved throughout the region. However, air pollution concentrations remain very high and air quality problems persist. Most of the European population lives in urban areas where air quality levels are exceeded. Bay of Algeciras region has important population centres with a population of approximately 300,000 inhabitants. The area shows a high concentration of industries from different sectors. This important activity and the current traffic require the necessary control of their environmental impact.

It is widely acknowledged that particles, nitrogen dioxide and ozone are the three pollutants that most affect human health (Donahue 2018; Sellier et al. 2014). The effects produced by long exposures to these pollutants range from respiratory system affections to premature death. Nitrogen dioxide pollutant ( $\text{NO}_2$ ) has been selected in this research since it is one of the most harmful contaminants. It is also one of the main greenhouse gases which has potentially harmful effects on the ecosystem, biodiversity and human livelihood on the planet (Mora et al. 2013). The World Health Organization (WHO) establishes the annual  $\text{NO}_2$  threshold limit cannot exceed  $40 \mu\text{g}/\text{m}^3$  and the hourly  $\text{NO}_2$  threshold limit of  $200 \mu\text{g}/\text{m}^3$ , a value that should not be exceeded more than 18 times per year. 2008/50 EU Directive sets that the Member States should apply air pollution reduction policies to ensure fulfillment of the limit values. Therefore, there is a need to make environmental measurements for planning, risk analysis and decision-making.

Nowadays, a very important task is environmental monitoring. A wireless sensor network consists of a number of spatially distributed sensor nodes for

monitoring environmental conditions. These collected data are sent to a central location. On the one hand, mapping environmental or pollution values are possible due to the volume and the accessibility to the remote sampled data. On the other hand, failures in the network occur and missing data are commonly found. Thus, an interesting approach would be to develop algorithms capable of estimating values in unsampled locations.

Considering the importance of estimating  $\text{NO}_2$  concentrations, the present study has been undertaken to compare the accuracy of two different techniques such as inverse distance weight (IDW) and least absolute shrinkage and selection operator (LASSO) and a new approach based on an ANN ensemble of the results of the individual methods. The main objective of this study is to estimate the air pollution concentrations in unsampled locations from sampled locations.

There are many possibilities to interpolate and to generate maps. Hence, selecting the correct method for a given input data could be a difficult task. Dubois and Galmarini (2005) performed a spatial interpolation comparison exercise where they compared from splines, kriging algorithms, to neural networks. Li and Heap (2014, 2011, 2008) listed different techniques (non-geostatistical interpolators, geostatistical interpolators and combined methods) as possible spatial interpolation or prediction methods. Tadić et al. (2015) also compared geostatistical and machine learning methods for interpolation and tested the effectiveness of hybrids between them.

Spatial interpolation methods are strongly recommended from many different approaches in the study of environmental sciences, as mentioned above. IDW is a spatial interpolation technique (Shepard 1968). It is based directly on the neighbouring sampled values. The value of the study variable in a new location can be derived as a weighted mean inversely proportional to the distance. Kriging (Matheron 1965) is another popular spatial interpolation method where weights are calculated objectively considering statistical concepts. It assumes that the distance or direction between points reflects a spatial correlation that can be used to explain the variability of a phenomenon. Normally statistical methods such as kriging explain phenomena better, but in other cases, deterministic methods such as IDW can perform the same or even better (Hengl et al. 2009). The selection of the most suitable interpolation method depends on the objectives to be achieved and the sample type. The sample size is one of the main aspects to

consider. Kriging method supposes an improvement in samples bigger than 30-point data; nevertheless, IDW produces better results in smaller samples. Thus, kriging has not been carried out in this research. These techniques have been commonly used in the estimation of pollutants (de Mesnard 2013; Gong et al. 2014). Yu et al. (2018) compared different methods, which include IDW, for developing spatiotemporal air pollutant concentration estimates of several pollutants. They concluded that IDW and kriging methods present similar results. Lastly, these methods are suitable for dense monitoring network areas. Gómez-Losada et al. (2019) used hidden Markov models to estimate the exposure of four air pollutants ( $\text{NO}_2$ ,  $\text{PM}_{10}$ ,  $\text{SO}_2$  and  $\text{O}_3$ ). Then, their spatial distribution was performed combining the interpolation results of ordinary kriging and inverse distance weighting.

LASSO is grouped within linear regression models (Tibshirani 2011; Tibshirani 1996). It uses the L1 regularization technique and automatically performs a variable selection. LASSO improves the performance of common multiple linear regression. Previous authors consider linear regression models as one of the most useful techniques in environmental sciences (Li and Heap 2014; Li and Heap 2011). Regression models are usually used to generate soil characteristic maps such as concentrations of heavy metals and organic matter content (Jiang et al. 2019; Piccini et al. 2018). Other authors (Aznarte 2017; Naughton et al. 2018; Shahbazi et al. 2018; Sharma et al. 2018) use regression techniques for forecasting or to compute imputations of missing data. Beelen et al. (2009) compared the validity of ordinary kriging, universal kriging and regression mapping in order to produce maps of air pollution. Universal kriging produced better results than regression mapping and ordinary kriging for  $\text{NO}_2$ ,  $\text{PM}_{10}$  and  $\text{O}_3$ . They deduced it is possible to develop detailed maps of air pollution. Contreras and Ferri (2016) used different techniques for forecasting air pollution including linear regression and LASSO and, furthermore, the predictions have been interpolated all around the city using IDW and kriging. Ma et al. (2019a) used a standard land use regression to represent the  $\text{NO}_2$  dispersion in a city. This model outperformed IDW and ordinary kriging. They concluded this approach is a robust option for modelling and mapping spatial concentrations of air pollutants in large areas with different designs and configurations. Requia et al. (2019) compared the predictive capabilities of ordinary geostatistical interpolation (ordinary

kriging), hybrid interpolation (empirical bayesian kriging and land use regression) and machine learning techniques (random forest-based regression) for estimating  $\text{PM}_{2.5}$  components. The random forest model reached the best performance, the next was the hybrid model and finally ordinary kriging.

Ensemble methods are machine learning techniques that combine multiple models in order to improve predictions or estimations, reducing variance and reducing bias. These methods are used to improve the stability and accuracy of an experimental procedure. There are different methods to create ensembles such as bagging (Breiman 1996), boosting (Drucker et al. 1994) or stacking (Ting and Witten 1999; Wolpert 1992). In the literature, ensemble methods have been performed obtaining good results predicting and generating maps; for example, Wang and Song (2018) showed that ensemble model is superior to the base model and improves predictions and mapping accuracy; similarly, Healey et al. (2018) perform a stacking approach where an ensemble of maps is used to improve the accuracy using an ANN as a meta-model. ANNs are used in many different fields in order to study non-linear relations, and are commonly used to forecast airborne pollutant concentrations in the atmosphere (Cabaneros et al. 2019; Feng et al. 2015; J. Ma et al. 2019b; Martín et al. 2008; Muñoz et al. 2014; Russo and Soares 2014; Van Roode et al. 2020). Alimissis et al. (2018) use an ANN in order to produce the spatial estimation of  $\text{O}_3$  and determines that ANN outperforms multiple linear regression (MLR) for air pollution forecasting. He and Christakos (2018) proposed a synthetic approach for spatiotemporal  $\text{PM}_{2.5}$  mapping. This approach combines the following method among other: a land use regression and an ANN. Ma et al. (2019b) proposed an ANN approach to interpolate the spatial distribution of  $\text{PM}_{2.5}$ . This work showed that the ANN model achieved better results compared to traditional spatial interpolation techniques such as IDW or kriging. Qi et al. (2019) suggested a hybrid model for  $\text{PM}_{2.5}$  spatiotemporal forecasting based on two types of ANN models.

So far, few studies have been conducted in this area related to air pollution monitoring. Most of the works have been developed by the authors of this paper. In these previous works, the authors focused on the estimation and prediction of several air pollutants in the main cities of the study area, such as Algeciras and La Línea de la Concepción. The different works address estimation and prediction of different atmospheric

pollutants in certain locations and considering spatial, meteorological and traffic relevance (García et al. 2011; González-Enrique et al. 2019a, 2019b, 2019c; ; Martín et al. 2008; Muñoz et al. 2014; Turias et al. 2006, 2017, 2008; Van Roode et al. 2018, 2020).

The remainder of the paper is organized as follows: “Area and data description” gives a brief description of the database and the study area. “Methods” discusses the applied methodology. “Experimental procedure” describes the experimental procedure performed and the results obtained are in “Results and discussion.” Finally, conclusions are drawn in “Conclusions.”

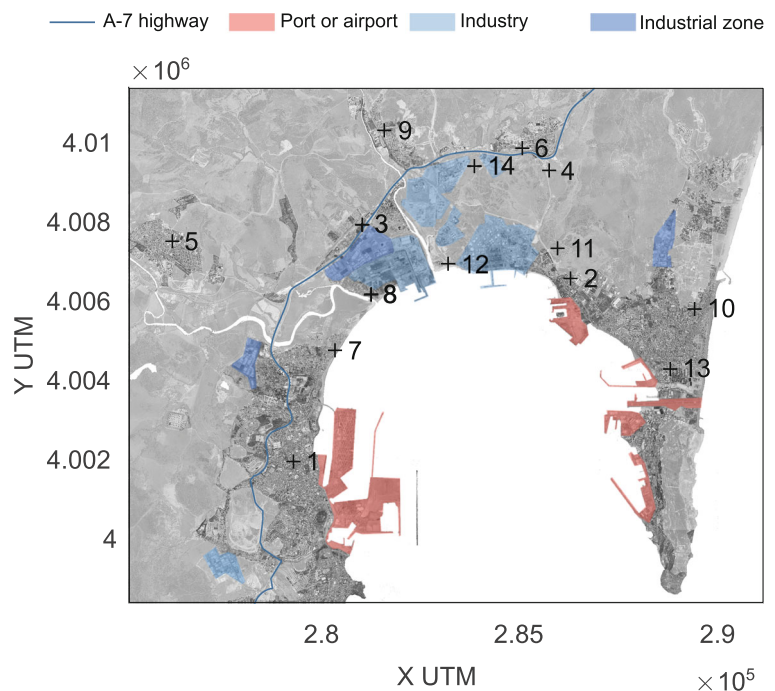
### Area and data description

Bay of Algeciras is the study area in this work which is located at the southern end of Spain (Europe). Its climate is Mediterranean and the predominant winds blow from East and West. The area has important centres with a population of almost 300,000 inhabitants in 2018. It is a very complex scenario. It is one of the most industrialized areas in the South of Spain. There are industries from different sectors such as an oil-refinery, different petrochemical factories and an important steel factory, among other important activities. In addition, it is located at the most important trading port of the

Mediterranean Sea. The Port Bahía de Algeciras is also the fifth in Europe with a total freight traffic of 9,390,000 tons in 2018 (<http://www.apba.es/estadisticas>). The proximity to the African continent (a distance of 14 km between its shores) makes this port a link between continents, involving an important racking of heavy tracks (approximately 325,000 ro-ro vehicles per year). A highway run from southwest to northeast connecting the different population, industry, administration and work centres and becoming the backbone of the bay. The road supports an average daily intensity of almost 70,000 vehicles/day, of which heavy trucks are 4000 (Dirección General de Carreteras 2017). Figure 1 shows the land uses.

Pollution data have been provided by the Environmental Agency of the Andalusian Government (research project TIN2014-58516-C2-2-R supported by MICINN Ministerio de Economía y Competitividad-Spain). Data has been collected with a 1-h sampling resolution during 6 years (2010–2015). The Andalusian Automatic Air Quality Network (Spain) is composed of a series of monitoring stations that measure emission levels of different pollutants. There is a data logging package in these locations, which collected the information of all sensors and sends it, via GPRS or internet, to the Air Quality Data Centre. Our local monitoring network is composed of 14 air pollution monitoring stations

**Fig. 1** Location map. Monitoring network and land uses



distributed throughout the area. Table 1 lists each air pollution monitoring station, its name and its position in UTM coordinates. Figure 1 shows also the monitoring network. No imputation methods have been used.

**Methods**

This study proposes a two-stage ANN ensemble considering the combination of two different methods in order to produce air pollution maps. On the one hand, IDW is a spatial-based interpolation method and, on the other hand, LASSO is a multiple linear regression method. A stacking ensemble has been developed using a BPNN as a meta-model and using the outputs (of the first stage) obtained by the previous models described. Both simple techniques have shown satisfactory results in the estimation of air pollution. Thus, the merging of the two techniques can add the strengths of each, improving the overall performance. Calculations have performed using NO<sub>2</sub> values measured in the air pollution monitoring station (Table 1) and distances between locations. The different methods are described briefly below and more thoroughly in the cited references.

**Ensemble**

It is widely known that individual models usually have worse performance than ensemble models. Ensembles

**Table 1** Location of monitoring network

	Air pollution monitoring station name	X UTM	Y UTM
1	Algeciras EPS	279239	4001847
2	Campamento	286237	4006469
3	Cortillijos	280980	4007826
4	Escuela Hostelería	285698	4009196
5	Colegio Los Barrios	276184	4007408
6	Colegio Carteya	285021	4009758
7	Rinconcillo	280289	4004653
8	Palmones	281205	4006069
9	Estación San Roque	281534	4010206
10	El Zabal	289371	4005695
11	Economato	285910	4007229
12	Guadarranque	283147	4006841
13	La Línea	288757	4004181
14	Madrevieja	283811	4009303

are machine learning techniques that merge different methods. The methodology is used to improve predictions or estimations, reducing the estimated error variances and reducing bias. There are different methods such as bagging (Breiman 1996), boosting (Drucker et al. 1994) or stacking (Ting and Witten 1999; Wolpert 1992). In bagging approach, different random subsets from the original dataset with replacement are generated, the same learning method on each sample is trained later and, finally, outputs of each model are simply weighted. This technique decreases the variance but does not improve the prediction. Boosting is similar to bagging. In boosting approach, different random subsets with replacement are also generated but training models are applied sequentially to each subset considering the previous model performance and, finally, outputs of these models are averagely weighted. In this case, a stacking approach has been used. It is similar to boosting. In the first step, outputs of different models (base-method) are performed and, in a second step, these outputs are used as inputs of another model (meta-model). Outputs of this last model are the final output values of the experiment. These models are able to understand the unstable behaviour of the variables. Thus, base-method outputs are combined to estimate the difficult behaviour of air pollution in this work.

As mentioned above, a stacking technique has been used in this case. This approach introduces the concept of a meta-learner. First, the base-learned models are trained based on a complete training set and then the meta-learned model is trained based on the outputs of the base-learned models as inputs. There are different approaches on which base models and meta-models are used (Sun and Li 2008; Woźniak et al. 2014)

In this paper, two different models have been used as base models and another learner model as a meta-model. LASSO and IDW are the base-learner models. As meta-model, the approach uses an artificial network model (ANN) based on the outputs of the previous models to perform estimates in new locations of the area

**IDW**

A spatial interpolation method is a procedure in which the value of a phenomenon is estimated in locations not sampled from other sampled locations within an area. There are different methods: deterministic, geostatistic or based on experts (Hengl 2009). All are based on the



idea that closer locations are more related than locations that are further away (Burrough and McDonnell 1998).

IDW is an interpolation method based on neighbouring sampled values (Shepard 1968). The estimations are made as a weighted average, as is shown in Eq. (1).

$$Z'_k(s_0) = \sum_{i=1}^n w_i(s_0) \cdot Z(s_i) \quad (1)$$

where  $w_i$  is the weight for neighbour  $i$ ,  $s_0$  is the location of a new point and  $s_i$  the location of a known sample point. The sum of weights must be equal to one to ensure that the interpolation is unbiased. The weights are determined according to the inverse distance from sampled points to the new point, as is shown in the Eq. (2).

$$w_i(s_0) = \frac{1}{\sum_{i=1}^n \frac{d^\beta(s_0, s_i)}{1}}; \quad \beta > 1 \quad (2)$$

where  $d(s_0, s_i)$  is the distance between the new point and the known sample point and  $\beta$  is a parameter that is used to highlight the spatial relationship between points. Farther points will be less important larger  $\beta$ . The parameter has an arbitrary character. Some authors determine which exponent should be chosen depending on what form of pollution you want to study (de Mesnard 2013). The established limits are between 1–4 for air pollution.

There are several criteria to select the number of neighbouring points. The selection of criteria is very important since the result of the interpolation strongly depends on it. One approach is to use all sampled points. Another approach would be to establish the maximum radius of the search from the point to be estimated. The remaining problem is to estimate that maximum distance. In this work, all available sampled points were used to compute the new measurement at a point  $(x, y, z, t)$ .

## LASSO

LASSO is a multiple linear regression method that uses a regularization term in order to avoid overfitting (Tibshirani 1996). An overfitted model is a model that contains more parameters than it really needs to explain the data. If a model is overfitted, the results are too dependent on training data and the model presents a

worse behaviour on new unseen data. Therefore, the models suffer a lack of generalization. Regularization methods correct the overfitting issue. They impose a penalty on the different parameters of the model to reduce the model freedom. In this way, the model improves the generalization capability. The methods presented below use the following form of penalization:

- $L_1$ -norm penalty (used by LASSO)
- $L_2$ -norm penalty (used by Ridge)
- $L_1$ -norm and  $L_2$ -norm penalty (used by Elastic net)

The different methods solve the following optimization problem, as shown in Eq. (3).

$$\min_{\beta} \left[ \frac{1}{2N} \sum_{i=1}^N (y_i - \beta x_i)^2 + \lambda \alpha \sum_{j=1}^k |\beta_j| + \lambda \frac{1-\alpha}{2} \sum_{j=1}^k \beta_j^2 \right] \quad (3)$$

where  $y_i$  is the response variable,  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$  is the input variable vector for  $i = 1, 2, \dots, N$ ,  $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$  is the model coefficient vector for  $j = 1, 2, \dots, k$  and  $\lambda$  is the hyperparameter that controls the penalty term. Lambda is zero or greater than zero. The higher the lambda value, the bigger is the penalty and thus coefficients are reduced. The  $\alpha$  parameter depend on the method. LASSO uses  $\alpha$  equal to 1, Ridge uses  $\alpha$  equal to 0 and Elastic net uses other values of  $\alpha$ .

In this case, LASSO is the method tested. In LASSO, with larger values of lambda, more coefficients are shrunk to zero. Thus, LASSO is also a feature selection method because reducing coefficients to zero remove them from the model.

## ANNS

ANNS are one of the main tools used in machine learning. Its approach tries to simulate the biological nervous system behaviour. They are composed of a number of artificial neurons, or units, arranged in several layers and linked by synaptic weights. A typical ANN architecture includes an input layer, one or more hidden layers and an output layer. Feedforward multi-layer perceptron using backpropagation (Rumelhart et al. 1986) is the most used ANNs model. Backpropagation networks are considered universal approximators (Hornik et al. 1989) and, in this sense, they are able to perform any non-linear relationship between inputs and outputs. In normal performance, the values of the inputs cross the network from front to back. The activation function in each layer determines the output value of each neuron that becomes the input

value for neurons in the next hidden layer connected to it. Backpropagation algorithm calculates the difference between output values and real output values to change the weights of the connections between the artificial neurons. The process is repeated until the error is minimized on the output nodes over all the patterns presented to the neural network. Levenberg-Marquardt is the minimization algorithm used in this work. Overfitting has to be avoided using an adequate experimental procedure as the one explained below.

### Experimental procedure

The aim of this study is to estimate the hourly NO<sub>2</sub> concentration in unsampled locations ( $x, y, t$ ) in order to produce air pollution maps with many objectives, such as missing data imputation and robust measurement. IDW, LASSO and an ANN ensemble were used as estimation methods.

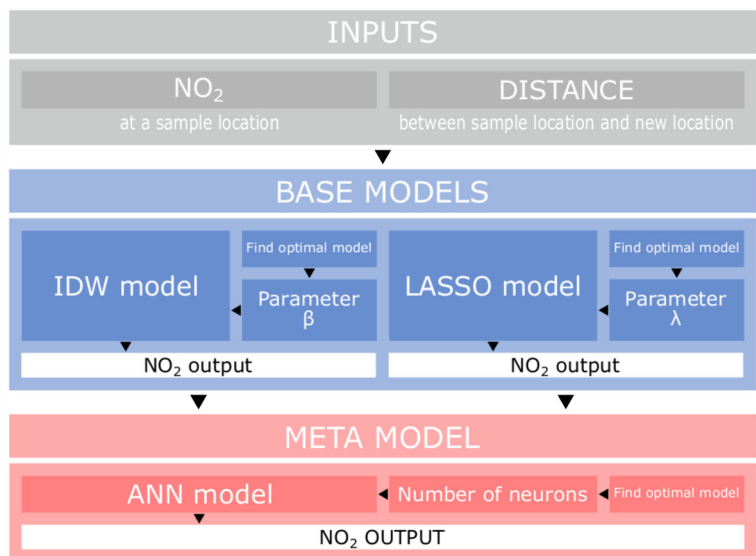
A flowchart that describes the experimental procedure is shown in Fig. 2. The main points of the experimental procedure are described as follows:

- (a) IDW and LASSO have been used as base models. Two variable types have been used: NO<sub>2</sub> concentrations at each monitoring station and distances to each monitoring station have been used to calculate the NO<sub>2</sub> concentration at a certain point.
- (b) Outputs of base models are used as inputs to a meta-model (ensemble). An ANN has been used

as the meta-model. ANN model consists of a backpropagation feedforward multilayer neural network (BPNN). A resampling procedure is used to compute the optimal number of hidden units (using only one hidden layer). Authors have successfully used this procedure in order to guarantee independence of the results and to avoid overfitting (Martín et al. 2008; Muñoz et al. 2014; Turias et al. 2008). The BPNN used contained a single hidden layer and were performed using a different number of hidden neurons (from 1 to 20). Friedman test and Bonferroni correction have been performed to select the optimal number of hidden neurons in the experiment (14 neurons).

- (c) IDW is a deterministic model and hence weights are estimated empirically but it is necessary to determine the parameter  $\beta$ . A leave-one-out cross-validation has performed to determine this parameter, as explained below. A trial-error search method has been used to find out the best value of beta resulting in a value equal to 2.3. It was proved that this value is within the limits previously established in order to interpolate air pollution.
- (d) In the case of LASSO and ANN, it is necessary to determine the parameters of each model. For that purpose, the sample has been divided into a subset for training-validation and another subset for testing. The testing subset is 25% of the data selected randomly. The procedure has been carried out 4 times. A  $k$ -fold cross-validation is used with these training-validation subsets to

**Fig. 2** The process described in the experimental procedure



select the best model. In the case of LASSO, the best lambda is selected, and in the case of ANNs, the number of hidden neurons is selected. The method relies on dividing the original sample into  $k$  equal-sized subsamples. A single subsample is removing for validation, the model and the remaining  $k-1$  subsamples are used for training data. The process is repeated  $k$  times; in this way, each subset is used once for validation. In this case,  $k$  equals to 2 has been used. Once the best model has been obtained, the model is re-trained with the whole no-test subset (training-validation). Finally, the model is tested with the reserved test subset. It is repeated 2 times due to the randomness of the experiment.

- (e) A leave-one-out (LOO) validation strategy has been performed to assess the accuracy of all methods. The method relies on removing one data point from the data set and estimating the value of this point with the remaining known values. In this case, the procedure is carried out for each sample station attempting to estimate its value from the remaining sample stations.
- (f) As a result, NO<sub>2</sub> concentrations have been estimated for the whole time series and the parameters  $R$ ,  $MSE$ ,  $MAE$  and  $d$  have been calculated for unseen data not used for training and validation purposes in order to evaluate the generalization capabilities of the models. These performance indexes are shown in Eqs. (4-7).

$$R = \frac{\sum_{i=1}^N (Z_i - \bar{Z})(Z'_i - \bar{Z})}{\sqrt{\sum_{i=1}^N (Z_i - \bar{Z})^2 \sum_{i=1}^N (Z'_i - \bar{Z})^2}} \tag{4}$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (Z'_i - \bar{Z})^2 \tag{5}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Z'_i - Z_i| \tag{6}$$

$$d = 1 - \frac{\sum_{i=1}^N (Z'_i - Z_i)^2}{\sum_{i=1}^N (|Z_i - \bar{Z}| + |Z'_i - \bar{Z}|)} \tag{7}$$

where  $Z$  is observed value and  $Z'$  is the estimated value. The  $R$  index has a value between + 1 and - 1. A value of  $\pm 1$  indicates total linear regression between  $Z$  and  $Z'$ , and 0 indicates no linear regression. The index  $MSE$  represents the average squared difference between the estimated values and the estimator (Lehmann and Casella 1998). The index  $MAE$  represents the average absolute error between the estimated value and the real value. Finally, the index  $d$  varies between 0 and 1. A value of 1 indicates a perfect match and 0 indicates no agreement at all (Willmott 1981).

Spatial estimation is a process of applying an algorithm to an array of grid nodes (point-to-point spatial estimation). The grid represents the domain of interest. In this case, base models have been performed to estimate the value in each node (step (a)) and then meta-model has been performed to produce a final value in each node (step (b)), as shown Fig. 3.

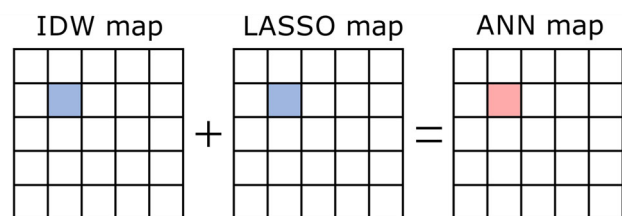
### Results and discussion

The main objective of this work has been to estimate air pollution concentration in unsampled locations in order to produce air pollution maps. Furthermore, the present study has been undertaken to compare the performance of the proposed two-stage approach with individual IDW and LASSO techniques.

As mentioned above, parameters  $R$ ,  $MSE$ ,  $MAE$  and  $d$  have been calculated once the data have been estimated for the whole time series. Table 2 shows the aforementioned performance indexes and, furthermore, their respective mean and standard deviation are obtained considering of each model.

**Fig. 3** Spatial estimation in an unsampled location using estimations from the base models first (step (a)) and using a meta-model (ANN) later (step (b))

• (x, y)





**Table 2** Results of the model performance indexes (*R*, *MSE*, *MAE* and *d*) for each monitoring station. Best values in bold

station	R			MSE			MAE			d		
	IDW	LASSO	ANN	IDW	LASSO	ANN	IDW	LASSO	ANN	IDW	LASSO	ANN
1	0.572	0.659	<b>0.715</b>	425.32	263.79	<b>228.98</b>	14.802	12.657	<b>11.169</b>	0.675	0.741	<b>0.794</b>
2	0.711	0.734	<b>0.736</b>	140.21	<b>114.00</b>	116.71	8.211	<b>7.532</b>	7.547	0.819	0.846	<b>0.851</b>
3	0.700	<b>0.737</b>	0.735	151.10	<b>128.45</b>	130.01	9.191	<b>7.766</b>	7.879	0.811	<b>0.846</b>	<b>0.846</b>
4	0.736	0.757	<b>0.770</b>	164.82	138.03	<b>131.99</b>	8.702	8.474	<b>8.036</b>	0.825	0.835	<b>0.854</b>
5	<b>0.620</b>	0.727	<b>0.755</b>	116.22	75.37	<b>66.98</b>	7.309	6.220	<b>5.474</b>	0.772	0.846	<b>0.858</b>
6	0.757	0.804	<b>0.831</b>	139.88	75.14	<b>66.65</b>	8.119	5.943	<b>5.464</b>	0.844	0.879	<b>0.901</b>
7	<b>0.865</b>	0.861	<b>0.871</b>	78.00	83.95	<b>75.61</b>	<b>5.841</b>	6.383	5.933	0.913	0.900	<b>0.915</b>
8	0.797	0.818	<b>0.819</b>	132.73	112.18	<b>111.27</b>	7.377	7.003	<b>6.761</b>	0.861	0.884	<b>0.889</b>
9	0.797	0.805	<b>0.830</b>	93.67	59.13	<b>54.05</b>	7.145	5.747	<b>4.972</b>	0.848	0.895	<b>0.903</b>
10	<b>0.840</b>	0.785	0.810	81.42	99.98	<b>85.73</b>	6.559	7.087	<b>6.551</b>	0.901	0.829	<b>0.872</b>
11	0.728	0.725	<b>0.736</b>	134.62	108.92	<b>107.62</b>	8.338	6.581	<b>6.568</b>	0.823	0.835	<b>0.846</b>
12	<b>0.667</b>	<b>0.667</b>	0.663	209.12	<b>204.22</b>	206.07	<b>9.447</b>	9.790	9.650	0.775	<b>0.802</b>	0.799
13	0.776	0.772	<b>0.781</b>	235.42	175.63	<b>166.27</b>	10.509	9.931	<b>9.243</b>	0.787	0.820	<b>0.844</b>
14	<b>0.635</b>	0.777	<b>0.781</b>	111.30	63.92	<b>62.81</b>	7.361	5.490	<b>5.277</b>	0.788	0.878	<b>0.880</b>
mean	0.729	0.759	0.774	158.13	121.62	115.05	8.494	7.615	7.180	0.817	0.845	0.861
std	0.085	0.056	0.055	88.72	58.18	54.06	2.185	2.005	1.850	0.059	0.042	0.036
max	0.865	0.861	0.871	78.00	59.13	54.05	5.841	5.490	4.972	0.913	0.900	0.915

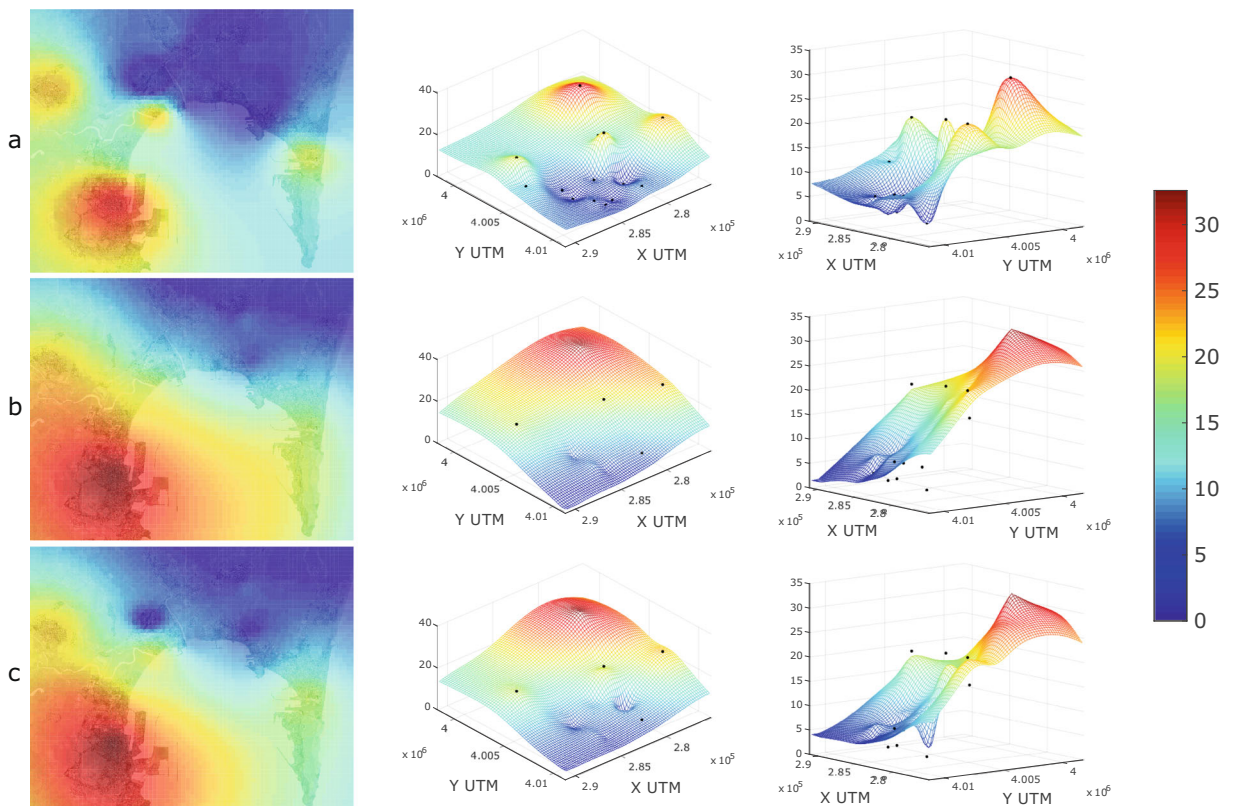
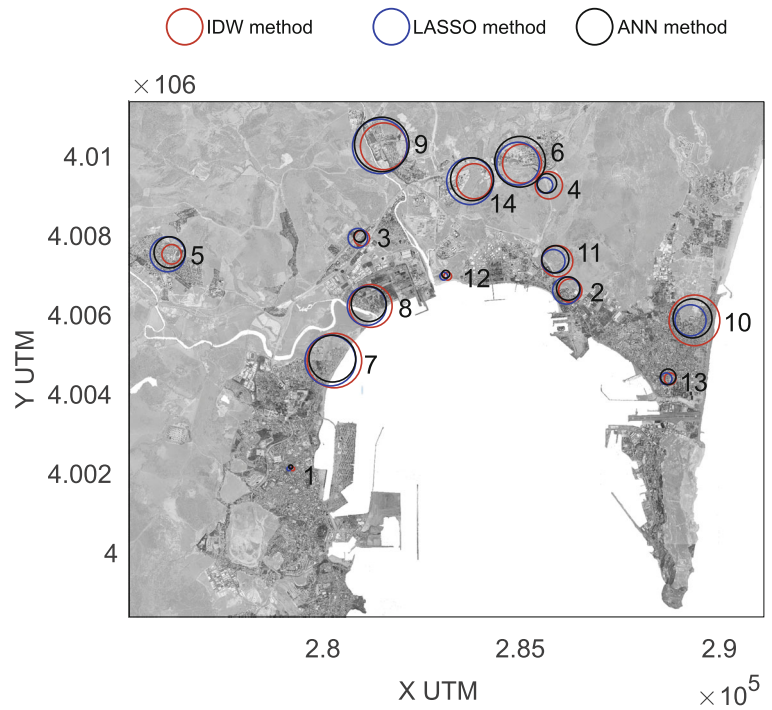
The performance indexes have been calculated for each station in order to study the spatial behaviour of models. Each station has been left out using a LOO procedure. The proposed two-stage model and the two individual models (IDW and LASSO) have been carried out and an estimation of the concentration of NO<sub>2</sub> was computed for each station. Performance indexes have been computed using the model results and the real concentration values measured at each station. Also, the average of each model performance indexes has been calculated. The different models show similar average values but the ANN ensemble always presents better results, then LASSO and finally IDW. Likewise, the standard deviation of the performance indexes of each model has been calculated. The standard deviation of the performance indicates how stable a model is. For example, IDW presents maximum *R* and maximum *d* values greater than LASSO but its averages values are lower than LASSO values. The ANN ensemble presents the lower standard deviations.

The general results of each model are commented below. Simple models (LASSO and IDW) present similar performances. ANN ensemble model has outperformed simple models for all parameters. This model presents the most stable results since the standard deviation of its performance indexes are the lowest. IDW presents an average *R* equals 0.729, LASSO an average *R* equals 0.759 and the proposed ANN ensemble an average *R* equals 0.774. IDW presents an average *MSE* equals 158.13, LASSO an average *MSE* equals 121.62 and the ANN ensemble an average *MSE* equals 115.05. IDW presents an average *MAE* equal to 8.494, LASSO an average *MAE* equals 7.615 and ANN

ensemble an average *MAE* equals 7.180. IDW presents an average index *d* equals 0.817, LASSO an average *d* equals 0.845 and ANN ensemble an average *d* equals 0.861. The best results of the performance indexes for each monitoring station are marked in bold in Table 2.

A colour performance ranking has been performed as shown in Table 2. The ranking sorts the results of each index considering the three methods. The ranking shows a range of colours from red to blue where intense red means the worst result and intense blue means the best result obtained. Stations 6, 7, 8, 9 and 10 show the best results while stations 1, 12 and 13 show the worst results. The general performances have been represented graphically according to the spatial distribution of monitoring stations, as shown in Fig. 4. Figure 4 represents the overall performance of each of the models with different sized circles according to the spatial distribution of monitoring stations, as mentioned above. Smaller circles represent the worst general results and larger circles represent the best general results according to the rest of the stations. All methods present similar behaviour according to the monitoring station locations. As discussed above, stations 1 and 12 present the worst results. Station 1 is far from the rest of the stations and thus the methods extrapolate to calculate the value at that point. IDW model tends to an average value at points located outside the area defined by the sampling sites. LASSO and ANN ensemble tend to limit values in the points located outside the area defined by sampling locations (see Fig. 5). Station 12 is located in the emission focus so it records the highest concentration peaks. It is surrounded by other monitoring stations and the whole area is highly influenced by easterly and westerly

**Fig. 4** General model performances according to the spatial distribution of monitoring station. The size of the circles presents the overall performance of each model



**Fig. 5** NO<sub>2</sub> concentration values ( $\mu\text{g}/\text{m}^3$ ) in  $t$  using the following methods: **a** IDW, **b** LASSO and **c** ANN ensemble. Maps (2D) and different perspectives (3D) are shown in order to understand the different model performances.

winds. Stations 6, 7, 8, 9 and 14 located around the main focus (station 12) present the best average performance.

Maps and different perspectives generated by each model are shown for an instant of time in Fig. 5. Simple models (LASSO and IDW) have similar performances in general but it can be noticed that surfaces generated by each model have completely different behaviour. On the one hand, IDW can fit better the local trends of the data and the result values are always within the range of the database domain. Then, estimated and measured values match exactly at each monitoring location. On the other hand, LASSO shows a global trend for the whole area and sometimes the output values can fall outside the range of given values. Estimated and measured values do not match exactly at monitoring locations. Therefore, and in order to improve the results, a staking ensemble has been proposed. The objective here has been to develop a model capable of weighing the outputs of simple models. The idea is that ANN learns the relationship between estimated value errors (obtained with IDW and LASSO) and real values. ANN ensemble model generates a surface that presents an intermediate (and more general) behaviour. The ensemble takes the LASSO global trend but is influenced by local peaks generated by IDW. As has been commented above, this model presents a better performance against a simple model.

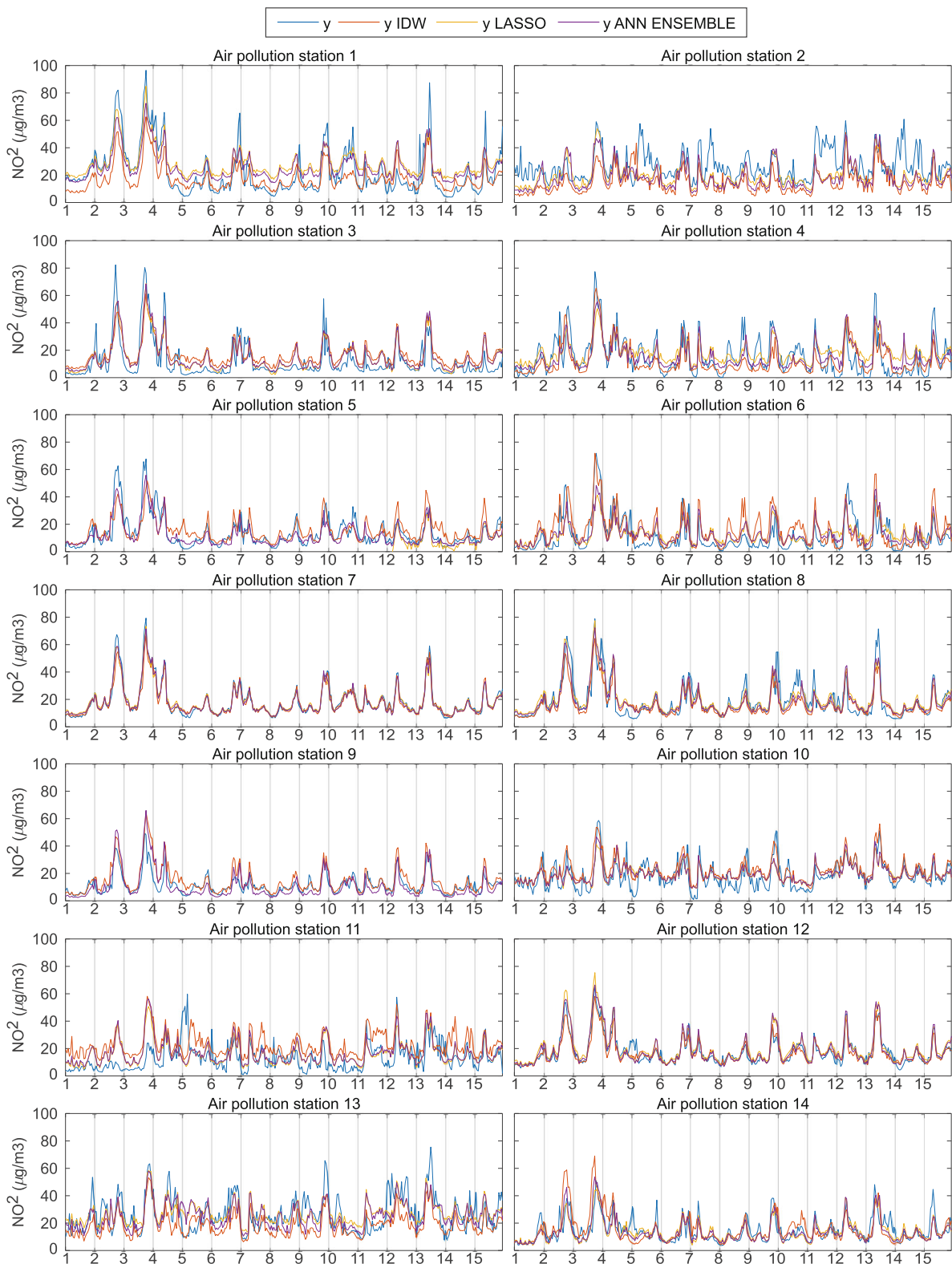
Figure 6 shows the measured values versus estimated values by the three methods for any period. The measured and estimated values for January 2015 have been represented for all air pollution monitoring stations. This period has been selected randomly in order to visualize results. Note that models are able to fit adequately the real values. Models are quite adapted to the different peaks and valleys that the real time series present.

As mentioned in Section 1, Yu et al. (2018) compared different methods. They obtained better-suited results for secondary pollutants than for primary pollutants; for example, average  $R^2 > 0.85$  for  $PM_{2.5}$  but  $R^2 < 0.35$  for  $NO_2$ . Ma et al. (2019a) used a regression model for a city obtaining an  $R^2$  of 0.86 and RMSE of  $3.53 \mu g m^{-3}$  considering a leave-one-out cross-validation. A value of  $R^2$  slightly higher than 0.7 was obtained in the work of Requía et al. (2019) estimating  $PM_{2.5}$  concentrations. Also, Qi et al. (2019) using a hybrid model obtained an  $R^2$  of 0.72 for  $PM_{2.5}$  estimations. In this work,  $NO_2$  estimations reach a maximum average  $R^2$  value of 0.76.

## Conclusions

As commented above, the main objective of this work has been to propose an ANN ensemble in order to estimate the hourly  $NO_2$  concentration at unsampled locations. In addition, spatial interpolation methods (IDW) and linear regression model with regularization (LASSO) have been compared and then combined using an ANN as a meta-learner. The following concluding remarks can be made from the results discussed above:

- IDW presents  $R$  values from 0.572 to 0.865,  $MSE$  values from 78.00 to 425.32,  $MAE$  values from 5.841 to 14.802 and  $d$  values from 0.675 to 0.913. LASSO presents  $R$  values from 0.659 to 0.861,  $MSE$  values from 59.13 to 425.32,  $MAE$  values from 5.490 to 12.657 and  $d$  values from 0.741 to 0.900. ANN ensemble presents  $R$  values from 0.663 to 0.871,  $MSE$  values from 54.05 to 228.98,  $MAE$  values from 4.972 to 11.169 and  $d$  values from 0.794 to 0.915.
- The standard deviation of the results has been calculated in order to indicate how stable the models are. ANN ensemble is the most stable model due to its standard deviation is lower. Comparing the base models, LASSO presents better average performance indexes than IDW. Nevertheless, IDW presents maximum  $R$  and maximum  $d$  values greater than LASSO and LASSO presents minimum  $MSE$  and minimum  $MAE$  values lower than IDW. LASSO is most stable than IDW considering their standard deviation.
- ANN ensemble outperforms base models in all stations. The model shows an important improvement at station 1, which is the station located further away and that presents the worst results. The ensemble model was able to increase  $R$  by 25% and  $d$  by 18% and was also able to decrease  $MSE$  by 46% and  $MAE$  by 24.5%. The improvement was not so substantial in the rest of the stations.
- The results obtained with the proposed ANN ensemble were very satisfactory achieving an  $R$  coefficient above 0.80 in most of monitoring stations. The results were always computed for test or unseen data within a cross-validation procedure and a LOO-station strategy.



**Fig. 6** Measured hourly values ( $y$ ) vs. estimated hourly values (IDW, LASSO and ANN ensemble) for January 2015.  $x$  axis represents the days of January



- The base models present very similar performances (similar values of  $R$ , MSE, MAE and  $d$ ) but generate completely different surfaces. IDW generates an abrupt surface strongly influenced by local values of sampling stations while LASSO generates a smooth surface of average values.
- The meta-model (ANN ensemble) presents the best and the most robust results. The model creates a mixture surface from the other two previous surfaces. The new surface presents an intermediate and better behaviour.
- The interpolation performed using IDW produced values always within the range of the database domain; however, the interpolation through LASSO and ANN models can produce values outside the range of given values.
- Models have very good results within the area defined by sampling locations. Nevertheless, the performance of the model decreases for points furthest from sampling location centroids. Stations 6, 7, 8, 9 and 10 show the best results while stations 1, 12 and 13 show the worst. The models have a good performance when points are in the interior domain (as an interpolation inside the area formed by the considered stations) while the performance decreases when the points are located outside the domain (extrapolation).
- Base methods have only considered two variable types, NO<sub>2</sub> concentrations and distances as inputs in order to compute NO<sub>2</sub> concentrations at a certain location  $(x, y, t)$ . IDW is a very simple method that produces very good yields. It is a deterministic model and hence weights are estimated empirically. LASSO is a more complex method. It considers the NO<sub>2</sub> concentration at monitoring stations and the distances between those stations in order to generate the model.
- Meta-model considers only two independent variables, the IDW output and the LASSO output, which are composed to a final solution using the proposed ensemble approach (NO<sub>2</sub> estimates) which is finally compared to the NO<sub>2</sub> real value.
- The monitoring network density affects the models predicting ability. Obviously, the more points, the more information and better interpolation results.
- The results have been promising in order to obtain air pollution map concentrations and to be used as a decision support system to prevent people, administrations or companies.

Therefore, the results obtained are very promising in order to provide new insights into the research. Thus, this approach could be used to a robust tool with the power of disposal of concentration measurements at any point  $(x, y, t)$  of a study area. This knowledge will be certainly useful to further researches in order to make as a decision support tool, to analyse risks or to design a new monitoring network.

**Acknowledgements** This work is part of the research project TIN2014-58516-C2-2-R supported by (MICINN Ministerio de Economía y Competitividad-Spain). The data have been kindly provided by Junta de Andalucía (Spain).

## References

- Alimissis, A., Philippopoulos, K., Tzani, C. G., & Deligiorgi, D. (2018). Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmospheric Environment*, *191*, 205–213. <https://doi.org/10.1016/J.ATMOENV.2018.07.058>.
- Aznarte, J. L. (2017). Probabilistic forecasting for extreme NO<sub>2</sub> pollution episodes. *Environmental Pollution*, *229*, 321–328. <https://doi.org/10.1016/J.ENVPOL.2017.05.079>.
- Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., de Hoogh, K., & Briggs, D. J. (2009). Mapping of background air pollution at a fine spatial scale across the European Union. *Science of the Total Environment*, *407*, 1852–1867. <https://doi.org/10.1016/j.scitotenv.2008.11.048>.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123–140. <https://doi.org/10.1007/BF00058655>.
- Burrough, P.A., McDonnell, R.A., 1998. Principles of geographical information systems. Oxford Univ. Press.
- Cabaneros, S. M., Calautit, J. K., & Hughes, B. R. (2019). A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling and Software*, *119*, 285–304. <https://doi.org/10.1016/J.ENVSOFT.2019.06.014>.
- Conteras, L., & Ferri, E. (2016). Wind-sensitive interpolation of urban air pollution forecasts. *Procedia Computer Science*, *80*, 313–323. <https://doi.org/10.1016/j.procs.2016.05.343>.
- de Mesnard, L. (2013). Pollution models and inverse distance weighting: some critical remarks. *Computational Geosciences*, *52*, 459–469. <https://doi.org/10.1016/j.cageo.2012.11.002>.
- Dirección General de Carreteras, Ministerio de Fomento, 2017. Mapa de tráfico 2017.
- Donahue, N.M., 2018. Air pollution and air quality, in: Green Chemistry. Elsevier, pp. 151–176. <https://doi.org/10.1016/B978-0-12-809270-5.00007-8>
- Drucker, H., Cortes, C., Jackel, L. D., LeCun, Y., & Vapnik, V. (1994). Boosting and other ensemble methods. *Neural Computation*, *6*, 1289–1301. <https://doi.org/10.1162/neco.1994.6.6.1289>.



- Dubois, G., & Galmarini, S. (2005). Introduction to the spatial interpolation comparison (SIC) 2004 exercise and presentation of the datasets. *Applied GIS, 1*, 1–11. <https://doi.org/10.2104/ag050009>.
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., & Wang, J. (2015). Artificial neural networks forecasting of PM<sub>2.5</sub> pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment, 107*, 118–128. <https://doi.org/10.1016/j.atmosenv.2015.02.030>.
- García, E.M., Rodríguez, M.L.M., Jiménez-Come, M.J., Espinosa, F.T., Domínguez, I.T., 2011. Prediction of peak concentrations of PM<sub>10</sub> in the area of Campo de Gibraltar (Spain) using classification models. pp. 203–212. [https://doi.org/10.1007/978-3-642-19644-7\\_22](https://doi.org/10.1007/978-3-642-19644-7_22)
- Gómez-Losada, Á., Santos, F. M., Gibert, K., & Pires, J. C. M. (2019). A data science approach for spatiotemporal modelling of low and resident air pollution in Madrid (Spain): implications for epidemiological studies. *Computers, Environment and Urban Systems, 75*, 1–11. <https://doi.org/10.1016/J.COMPENVURBSYS.2018.12.005>.
- Gong, G., Mattevada, S., & O'Bryant, S. E. (2014). Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environmental Research, 130*, 59–69. <https://doi.org/10.1016/J.ENVRES.2013.12.005>.
- González-Enrique, J., Ruiz-Aguilar, J.J., Moscoso-López, J.A., Van Roode, S., Urda, D., Turias, I.J., 2019a. A genetic algorithm and neural network stacking ensemble approach to improve NO<sub>2</sub> level estimations. Springer International Publishing, pp. 856–867. [https://doi.org/10.1007/978-3-030-20521-8\\_70](https://doi.org/10.1007/978-3-030-20521-8_70)
- González-Enrique, J., Turias, I. J., Ruiz-Aguilar, J. J., Moscoso-López, J. A., & Franco, L. (2019b). Spatial and meteorological relevance in NO<sub>2</sub> estimations: a case study in the Bay of Algeciras (Spain). *Stochastic Environmental Research and Risk Assessment, 33*, 801–815. <https://doi.org/10.1007/s00477-018-01644-0>.
- González-Enrique, J., Turias, I. J., Ruiz-Aguilar, J. J., Moscoso-López, J. A., Jerez-Aragonés, J., & Franco, L. (2019c). Estimation of NO<sub>2</sub> concentration values in a monitoring sensor network using a fusion approach. *Fresenius Environmental Bulletin, 28*, 681–686.
- He, J., & Christakos, G. (2018). Space-time PM<sub>2.5</sub> mapping in the severe haze region of Jing-Jin-Ji (China) using a synthetic approach. *Environmental Pollution, 240*, 319–329. <https://doi.org/10.1016/J.ENVPOL.2018.04.092>.
- Healey, S. P., Cohen, W. B., Yang, Z., Kenneth Brewer, C., Brooks, E. B., Gorelick, N., Hernandez, A. J., Huang, C., Joseph Hughes, M., Kennedy, R. E., Loveland, T. R., Moisen, G. G., Schroeder, T. A., Stehman, S. V., Vogelmann, J. E., Woodcock, C. E., Yang, L., & Zhu, Z. (2018). Mapping forest change using stacked generalization: an ensemble approach. *Remote Sensing of Environment, 204*, 717–728. <https://doi.org/10.1016/J.RSE.2017.09.029>.
- Hengl, T., 2009. A practical guide to geostatistical mapping, *JCR Scientific and Technical Research Series*.
- Hengl, T., Minasny, B., & Gould, M. (2009). A geostatistical analysis of geostatistics. *Scientometrics, 80*, 491–514. <https://doi.org/10.1007/s11192-009-0073-3>.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*, 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- Jiang, X., Zou, B., Feng, H., Tang, J., Tu, Y., & Zhao, X. (2019). Spatial distribution mapping of Hg contamination in subclass agricultural soils using GIS enhanced multiple linear regression. *Journal of Geochemical Exploration, 196*, 1–7. <https://doi.org/10.1016/J.GEXPLO.2018.10.002>.
- Lehmann, E.L., Casella, G., 1998. Theory of point estimation, 2nd ed. ed.
- Li, J., Heap, A.D., 2008. A review of spatial interpolation methods for environmental scientists, Record 200. ed.
- Li, J., & Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecological Informatics, 6*, 228–241. <https://doi.org/10.1016/j.ecoinf.2010.12.003>.
- Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: a review. *Environmental Modelling and Software, 53*, 173–189. <https://doi.org/10.1016/j.envsoft.2013.12.008>.
- Ma, J., Ding, Y., Cheng, J. C. P., Jiang, F., & Wan, Z. (2019a). A temporal-spatial interpolation and extrapolation method based on geographic long short-term memory neural network for PM<sub>2.5</sub>. *Journal of Cleaner Production, 237*, 117729. <https://doi.org/10.1016/J.JCLEPRO.2019.117729>.
- Ma, X., Longley, I., Gao, J., Kachhara, A., & Salmond, J. (2019b). A site-optimised multi-scale GIS based land use regression model for simulating local scale patterns in air pollution. *Science of the Total Environment, 685*, 134–149. <https://doi.org/10.1016/J.SCITOTENV.2019.05.408>.
- Martín, M. L., Turias, I. J., González, F. J., Galindo, P. L., Trujillo, F. J., Puntonet, C. G., & Gorziz, J. M. (2008). Prediction of CO maximum ground level concentrations in the Bay of Algeciras, Spain using artificial neural networks. *Chemosphere, 70*, 1190–1195. <https://doi.org/10.1016/j.chemosphere.2007.08.039>.
- Matheron, G., 1965. Les variables régionalisées et leur estimation, une application de la théorie de fonctions aléatoires aux sciences de la nature. Masson.
- Mora, C., Frazier, A. G., Longman, R. J., Dacks, R. S., Walton, M. M., Tong, E. J., Sanchez, J. J., Kaiser, L. R., Stender, Y. O., Anderson, J. M., Ambrosino, C. M., Fernandez-Silva, I., Giuseffi, L. M., & Giambelluca, T. W. (2013). The projected timing of climate departure from recent variability. *Nature, 502*, 183–187. <https://doi.org/10.1038/nature12540>.
- Muñoz, E., Martín, M. L., Turias, I. J., Jimenez-Come, M. J., & Trujillo, F. J. (2014). Prediction of PM<sub>10</sub> and SO<sub>2</sub> exceedances to control air pollution in the Bay of Algeciras, Spain. *Stochastic Environmental Research and Risk Assessment, 28*, 1409–1420. <https://doi.org/10.1007/s00477-013-0827-6>.
- Naughton, O., Donnelly, A., Nolan, P., Pilla, F., Misstear, B. D., & Broderick, B. (2018). A land use regression model for explaining spatial variation in air pollution levels using a wind sector based approach. *Science of the Total Environment, 630*, 1324–1334. <https://doi.org/10.1016/J.SCITOTENV.2018.02.317>.
- Piccini, C., Marchetti, A., Riviaccio, R., & Napoli, R. (2018). Multinomial logistic regression with soil diagnostic features and land surface parameters for soil mapping of Latium (Central Italy). *Geoderma, 168*, 1–11. <https://doi.org/10.1016/J.GEODERMA.2018.09.037>.

- Qi, Y., Li, Q., Karimian, H., & Liu, D. (2019). A hybrid model for spatiotemporal forecasting of PM<sub>2.5</sub> based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, *664*, 1–10. <https://doi.org/10.1016/J.SCITOTENV.2019.01.333>.
- Requia, W. J., Coull, B. A., & Koutrakis, P. (2019). Evaluation of predictive capabilities of ordinary geostatistical interpolation, hybrid interpolation, and machine learning methods for estimating PM<sub>2.5</sub> constituents over space. *Environmental Research*, *175*, 421–433. <https://doi.org/10.1016/J.ENVRES.2019.05.025>.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). *Learning internal representations by error propagation*, in: *Parallel Distributed Processing* (pp. 318–362). Cambridge: MIT Press.
- Russo, A., & Soares, A. O. (2014). Hybrid model for urban air pollution forecasting: a stochastic spatio-temporal approach. *Mathematical Geoscience*, *46*, 75–93. <https://doi.org/10.1007/s11004-013-9483-0>.
- Sellier, Y., Galineau, J., Hulin, A., Caini, F., Marquis, N., Navel, V., Bottagisi, S., Giorgis-Allemand, L., Jacquier, C., Slama, R., & Lepeule, J. (2014). Health effects of ambient air pollution: do different methods for estimating exposure lead to different results? *Environment International*, *66*, 165–173. <https://doi.org/10.1016/J.ENVINT.2014.02.001>.
- Shahbazi, H., Karimi, S., Hosseini, V., Yazgi, D., & Torbatian, S. (2018). A novel regression imputation framework for Tehran air pollution monitoring network using outputs from WRF and CAMx models. *Atmospheric Environment*, *187*, 24–33. <https://doi.org/10.1016/J.ATMOSENV.2018.05.055>.
- Sharma, N., Taneja, S., Sagar, V., & Bhatt, A. (2018). Forecasting air pollution load in Delhi using data analysis tools. *Procedia Comput. Sci.*, *132*, 1077–1085. <https://doi.org/10.1016/J.PROCS.2018.05.023>.
- Shepard, D. (1968). *A two-dimensional interpolation function for irregularly-spaced data* (pp. 517–524). New York: Proceedings of the 1968 ACM National Conference. <https://doi.org/10.1145/800186.810616>.
- Sun, J., & Li, H. (2008). Listed companies' financial distress prediction based on weighted majority voting combination of multiple classifiers. *Expert Systems with Applications*, *35*, 818–827. <https://doi.org/10.1016/j.eswa.2007.07.045>.
- Tadić, J. M., Ilić, V., & Biraud, S. (2015). Examination of geostatistical and machine-learning techniques as interpolators in anisotropic atmospheric environments. *Atmospheric Environment*, *111*, 28–38. <https://doi.org/10.1016/J.ATMOSENV.2015.03.063>.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *58*, 267–288.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *73*, 273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>.
- Ting, K. M., & Witten, I. H. (1999). Issues in stacked generalization. *Journal of Artificial Intelligence Research*, *10*, 271–289.
- Turias, I. J., González, F. J., Martín, M. L., & Galindo, P. L. (2006). A competitive neural network approach for meteorological situation clustering. *Atmospheric Environment*, *40*, 532–541. <https://doi.org/10.1016/j.atmosenv.2005.09.065>.
- Turias, I. J., González, F. J., Martín, M. L., & Galindo, P. L. (2008). Prediction models of CO, SPM and SO<sub>2</sub> concentrations in the Campo de Gibraltar Region, Spain: a multiple comparison strategy. *Environmental Monitoring and Assessment*, *143*, 131–146. <https://doi.org/10.1007/s10661-007-9963-0>.
- Turias, I. J., Jerez, J. M., Franco, L., Mesa, H., Ruiz-Aguilar, J. J., Moscoso, J. A., & Jiménez-Come, M. J. (2017). Prediction of carbon monoxide (CO) atmospheric pollution concentrations using meteorological variables. *WIT Transactions on Ecology and the Environment*, *211*(9), 137–145. <https://doi.org/10.2495/AIR170141>.
- Van Roode, S., Ruiz-Aguilar, J. J., González-Enrique, J., Moscoso-López, J. A., & Turias, I. J. (2018). Using geostatistical modelling for analysis of air pollution and its relation with road traffic in Bay of Algeciras (Spain). XIII Congreso de Ingeniería del Transporte (CIT). *Gijón*.
- Van Roode, S., Ruiz-Aguilar, J. J., González-Enrique, J., Turias, I. J., 2020. A hybrid approach for short-term NO<sub>2</sub> forecasting: case study of Bay of Algeciras (Spain) 190–198. [https://doi.org/10.1007/978-3-030-20055-8\\_18](https://doi.org/10.1007/978-3-030-20055-8_18)
- Wang, J., & Song, G. (2018). A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing*, *314*, 198–206. <https://doi.org/10.1016/j.neucom.2018.06.049>.
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, *2*, 184–194.
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*, 241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1).
- Woźniak, M., Graña, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, *16*, 3–17. <https://doi.org/10.1016/j.inffus.2013.04.006>.
- Yu, H., Russell, A., Mulholland, J., Odman, T., Hu, Y., Chang, H. H., & Kumar, N. (2018). Cross-comparison and evaluation of air pollution field estimation methods. *Atmospheric Environment*, *179*, 49–60. <https://doi.org/10.1016/J.ATMOSENV.2018.01.045>.

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