Future Generation Computer Systems 136 (2022) 15–33

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Multi-objective evolutionary spatio-temporal forecasting of air pollution

Raquel Espinosa *, Fernando Jiménez, José Palma

Department of Information and Communication Engineering, University of Murcia, Spain

ARTICLE INFO

Article history:
Received 5 February 2022
Received in revised form 11 May 2022
Accepted 21 May 2022
Available online 31 May 2022

Keywords:
Spatio-temporal forecasting
Air pollution
Multi-objective optimization
Evolutionary algorithms
Machine learning
Ensemble learning

ABSTRACT

Nowadays, air pollution forecasting modeling is vital to achieve an increase in air quality, allowing an improvement of ecosystems and human health. It is important to consider the spatial characteristics of the data, as they allow us to infer predictions in those areas for which no information is available. In the current literature, there are a large number of proposals for spatio-temporal air pollution forecasting. In this paper we propose a novel spatio-temporal approach based on multi-objective evolutionary algorithms for the identification of multiple non-dominated linear regression models and their combination in an ensemble learning model for air pollution forecasting. The ability of multi-objective evolutionary algorithms to find a Pareto front of solutions is used to build multiple forecast models geographically distributed in the area of interest. The proposed method has been applied for one-week NO_2 prediction in southeastern Spain and has obtained promising results in statistical comparison with other approaches such as the union of datasets or the interpolation of the predictions for each monitoring station. The validity of the proposed spatio-temporal approach is thus demonstrated, opening up a new field in air pollution engineering.

1. Introduction

Although emissions of harmful substances into the atmosphere have been reduced in recent decades [1], globalization, the burning of fossil fuels and the increase in the number of industries, among others, are some of the main reasons why air quality is failing to improve. The conditions of indoor and outdoor air can be determined by the concentrations of chemicals present in it. High levels of certain compounds such as nitrogen oxides (NO_X), sulfur dioxide (SO_2), ozone (O_3), particulate matter (PM), carbon monoxide (CO) and carbon dioxide (CO_2) are particularly hazardous. These gases can be emitted into the atmosphere both naturally, e.g. volcanic eruptions, forest fires or microbial decaying processes, and anthropogenically, e.g. combustion engines, industrial processes, agricultural activities or farming. In addition, other factors such as noise, light or radiation are also considered pollutants and reduce or contribute to reducing air quality. All of these contaminants are harmful to both human health and ecosystems. Some symptoms that can be caused due to the emission of toxic gases are irritation of the respiratory tract, skin and eyes, coughing, asthma, breathing difficulties or an increased risk of suffering a heart attack. Furthermore, prolonged exposure to these compounds can lead to the development of certain types of cancer or damage to the respiratory or immune systems. Children and the elderly are particularly sensitive to these compounds [2,3]. According to a 2019 report [4], 8.8 million deaths worldwide are caused by indoor and outdoor air pollution, of which 5.5 million deaths are prematurely and come from anthropogenic sources. The *World Health Organization* [5] estimates that the deaths from all sources of air pollution were around 7 million in 2016, 4.2 million corresponded to outdoor air pollution. In ecosystems, they can cause acid rain, eutrophication or accelerate climate change due to the greenhouse effect. For all these reasons, air pollution engineering is vital to improving both human and ecosystem health. *Air pollution engineering* can be divided into two phases [6], on the one hand, control, and on the other hand, engineering. Air quality control is a preventive step and includes those techniques for monitoring and preventing air pollution emissions. Air quality engineering focuses on large-scale, multi-source control strategies, with special emphasis on the physics and chemistry of pollutant interactions in the atmosphere.

This paper focuses on computational modeling for air quality multivariate space–time series forecasting with the aim of preventing and alerting the population and supporting decision making by public authorities. We propose a multi-objective optimization based approach in which multiple *linear regression* (LR) [7] models are identified (forming a *Pareto* front) and a *stacking regression* based *ensemble learning* model is finally built with them. The multi-objective optimization problem has been

* Corresponding author.
E-mail addresses: raquel.espinosa@um.es (R. Espinosa), fernan@um.es (F. Jiménez), jtpalma@um.es (J. Palma).

https://doi.org/10.1016/j.future.2022.05.020

solved in this paper with *multi-objective evolutionary algorithms* (MOEAs) [8] and its purpose is to optimize the *root mean squared error* (RMSE) of multiple LR models built from the data from multiple monitoring stations located in different places of the geography under study. MOEAs are powerful multi-objective optimization techniques that have been widely applied in different fields of engineering such as design [9], mechanical [10], aeronautical and aerospace [11], biomedical [12], chemical [13], nuclear [14], electrical [15], kansei [16], and many others. *Evolutionary algorithms* (EA) [17] have also been applied to air quality optimization [18] and air pollution prediction [19,20].

We have statistically compared different regression algorithms to build the final ensemble learning model from the LR models found by the MOEA. We have compared the multi-objective optimization based spatio-temporal approach with two other spatio-temporal approaches. The first approach is based on the simple union of the datasets obtained from the different monitoring stations to build a single forecast model. The second approach, most used in the literature, is based on the interpolation of the predictions of the models built separately for each monitoring station. We have compared different regression algorithms also in these cases. In all three approaches, the spatio-temporal characteristics of the data are considered. The data on which the experiments have been performed contain air quality information from four monitoring stations (La Aljorra, Alcantarilla, Lorca and Valle de Escombreras) located within the Region of Murcia, Spain. The data has been taken daily for four years, between 2017 and 2020. Predictions have been made for 7-steps ahead, i.e. one week in advance. In this paper, the concentrations to be predicted will be nitrogen dioxide (NO_2).

The most important contributions of the research are listed below:

- We propose a multi-objective optimization based spatio-temporal approach combined with ensemble learning for air pollution forecasting. Simultaneous minimization of the RMSE of LR models for data observed over time at multiple monitoring stations located at different geographic points produces a Pareto front of multiple regression models uniformly distributed over the forecast area, which facilitates a better approximation to the spatio-temporal forecasting problem. These LR models must be properly assembled to make future predictions in any geographical location, for which we have proposed a stacking based ensemble learning method. This approach has not been used previously in the literature, therefore it is a completely novel method.
- We propose the specific components of the MOEA (representation of individuals and fitness function) for solving the proposed multi-objective optimization problem. Different base algorithms for multi-objective optimization are statistically compared.
- We have statistically compared different meta-learners to build the final stacking based ensemble learning model from the non-dominated LR models obtained with the MOEA.
- Our multi-objective optimization based spatio-temporal approach has been compared with two other spatio-temporal approaches based on the union of data from the different monitoring stations and on the interpolation of the predictions, respectively, using different regression algorithms to build the forecast model.
- For comparisons, a multi-criteria decision metric that combines different performance metrics and h-steps ahead predictions of NO_2 is also proposed.
- This is the first time that a study for the prediction of NO_2 in the Region of Murcia has been carried out based on spatio-temporal techniques.

With this background, the paper has been organized as follows: Section 2 describes some basic aspects of continuous multi-objective optimization and presents works related to spatio-temporal forecasting of air pollution published in the last four years; Section 3 formally defines the proposed approaches and methods for air quality prediction; Section 4 describes the experiments performed and their outcomes; Section 5 analyzes the obtained results; Section 6 reviews the threat to the validity of our study; Section 7 draws the main conclusions of this paper and future works.

2. Background

2.1. Continuous multi-objective optimization

A *continuous multi-objective optimization problem* can be mathematically formulated as follows:

$$Min./Max. \quad f_k(\mathbf{x}), \quad k = 1, \ldots, n \tag{1}$$

where $f_k(\mathbf{x})$ are *objective functions*, $\mathbf{x} = \{x_1, x_2, \ldots, x_w\}$ represents the set of *decision variables*, with $x_i \in [l_i, u_i] \subset \mathbb{R}$, where $[l_i, u_i]$ is the domain of the variable x_i, $i = 1, \ldots, w$. Let $\mathcal{F} = \{\mathbf{x} \in [l_i, u_i]^w\}$ be the *search space* of the problem (1). We want to find a subset of solutions $\mathcal{S} \subseteq \mathcal{F}$ called *non-dominated set* (or *Pareto optimal set*). A solution $\mathbf{x} \in \mathcal{F}$ is *non-dominated* if there is no other solution $\mathbf{x}' \in \mathcal{F}$ that dominates $\mathbf{x}$, and a solution $\mathbf{x}'$ *dominates* $\mathbf{x}$ if and only if there exists k ($1 \leq k \leq n$) such that $f_k(\mathbf{x}')$ improves $f_k(\mathbf{x})$, and for every k ($1 \leq k \leq n$), $f_k(\mathbf{x})$ does not improve $f_k(\mathbf{x}')$. In other words, $\mathbf{x}'$ *dominates* $\mathbf{x}$ if and only if $\mathbf{x}'$ is better than $\mathbf{x}$ for at least one objective, and not worse than $\mathbf{x}$ for any other objective. For minimization problems, the set $\mathcal{S}$ of non-dominated solutions of (1) can be formally defined as:

$$\mathcal{S} = \left\{\mathbf{x} \in \mathcal{F} \mid \nexists\, \mathbf{x}' \in \mathcal{F} \mid \mathcal{D}(\mathbf{x}', \mathbf{x})\right\}$$

where:

$$\mathcal{D}(\mathbf{x}', \mathbf{x}) \equiv \exists k,\ 1 \leq k \leq n,\ f_k(\mathbf{x}') < f_k(\mathbf{x}) \wedge \forall k,\ 1 \leq k \leq n,\ f_i(\mathbf{x}') \leq f_i(\mathbf{x})$$

Solving the multi-objective optimization problem consists of finding or approximating all or a representative set of Pareto optimal solutions [21,22].

2.2. Spatio-temporal air pollution forecasting related works

This section presents relevant works published in the last 4 years in the field of spatio temporal forecasting prediction for air pollution or air quality.

Seng et al. [23] propose a *multi-output and multi-index of supervised learning* (MMSL) model based on *long short-term memory* (LSTM) for predicting concentrations of $PM_{2.5}$, SO_2, NO_2, O_3 and CO in the next hours. This will enable the prevention of air pollution, thus improving government decision-making and people's health. Hourly data from four different stations located in Beijing between 2016 and 2017 have been used to obtain both local and global air quality values. *Mean absolute error* (MAE), RMSE and *Pearson's correlation coefficient* (R^2) metrics have been used to measure the performance of the models. MMSL predictions were compared with other *machine learning* (ML) and *deep learning* (DL) techniques.

Lu and Liu [24] develop a weighted-averaged forecasting method. The technique takes into consideration the temporal characteristics for the 48-hour $PM_{2.5}$ forecast along with the spatial characteristics to obtain improved predictions. The data were taken between 2017 and 2018 at stations located at different spots in Taipei. The described method has obtained better RMSE

and MAE values than other air quality prediction methods with which comparisons have been made.

Bui et al. [25] apply a method called STAR based on a multimodal fusion for spatio-temporal prediction of $PM_{2.5}$ and PM_{10}. STAR has an encoder in charge of transforming the data into heat maps for the analysis and capture of their spatial properties. This graphical representation is combined with the temporal features, which are fed into an LSTM. The decoder part is composed of a *convolutional neural network* (CNN)-LSTM model in charge of generating heat maps with the predictions made. A dataset with hourly data taken during 5 years in Seoul has been used for the experimentation while a dataset with data taken during one year in China has been used for the evaluation of the model. This method has been able to reduce the MAE compared to baselines.

Huang et al. [26] use a *spatio-attention embedded recurrent neural network* called SpAttRNN for *air quality index* prediction. This model is able to learn both spatial and temporal dependencies. For this purpose, they have developed a spatio-attention graph cell-based LSTM to capture the relationships between stations. This new cell is combined with an LSTM to extract the spatial and temporal correlations from the data to make predictions. Two datasets with Beijing air quality data from different stations, between 2014 and 2015 and 2017 and 2018, have been used to test the performance of the model. SpAttRNN has been compared with other methods by improving them in terms of RMSE, MAE and *symmetric* MAPE.

Le et al. [27] deal with the spatial and temporal characteristics of the data to predict air quality with a *Convolutional LSTM*. The model has as input 2D air pollution images and other factors such as traffic volume. The dataset used for the experiments contains data taken between 2015 and 2017. The Convolutional LSTM has been compared with other baselines in terms of RMSE with good results.

Zou et al. [28] propose an LSTM with a spatiotemporal attention mechanism for air quality prediction based on an encoder–decoder architecture. The encoder is responsible for assigning weights to different regions according to their importance on the area over which the predictions are made. The decoder is able to obtain the temporal correlation between the future and historical values. The data used to verify the model are from various stations in Beijing taken hourly during 2018. The RMSE of the proposed method performs better than other air quality prediction methods.

Zhang et al. [29] define an LSTM with a graph attention mechanism in order to predict air quality. The LSTM determines the temporal correlation of historical data. The graph attention mechanism has an encoder–decoder architecture for spatial correlation of monitoring stations. Each node uses the attention mechanism to collect similarity information from the feature vectors of its neighbors and uses this information to update its feature vector. The dataset used contains information from several Beijing stations with data taken hourly for one year. This model can be used with transfer learning to make predictions for cities for which insufficient data is available.

Lin et al. [30] use *geo-context based diffusion convolutional recurrent neural network* (GC-DCRNN) for the prediction of $PM_{2.5}$. To take into consideration the spatial component, graphs are used to determine the spatial correlation between stations. On the other hand, GRU is applied for the analysis of the temporal dependencies of the data. In addition, a *Sequence to Sequence* architecture is adopted for multi-step ahead forecasting. GC-DCRNN has been evaluated with two datasets, one from Los Angeles and one from Beijing. Compared to other models GC-DCRNN obtains better values in terms of MAE.

Chae et al. [31] propose a model based on interpolation and CNN for real-time prediction of $PM_{2.5}$ and PM_{10} concentrations. Interpolation allows obtaining air quality predictions in areas that do not have a monitoring station. The dataset used contains data from different stations in South Korea taken between 2018 and 2019. The resulting model obtains good values of both RMSE and R^2 and is able to classify high concentration events of the chemical compounds.

Samal et al. [32] present a model based on CNN, *bidirectional LSTM* and an *Inverse Distance Weighting* interpolation called CNN-BILSTM-IDW for the prediction of PM_{10}. The model uses past data for the prediction of the spatial distribution of the air pollution level. This model has been used in mobile and web applications for visualization of future values of PM_{10}. The dataset used to train the model contains air quality data for Odisha (India) between 2005 and 2015 from different stations.

Le [33] proposes a graph-based and a RNN-based model called Spatiotemporal GCRNNN that allows the representation of the spatial characteristics of the air. For this purpose, the monitoring stations are depicted as a weighted graph. Each graph has been observed at an instant in time. With these spatial and temporal features, predictions are then made. The dataset contains spatiotemporal data for Seoul (Korea) taken between 2015 and 2019. The error metrics used were RMSE, R^2 and a new metric that takes into account the spatiotemporal component of the predictions called spRMSE. Compared to other spatiotemporal prediction models, GCRNN performs better than other spatiotemporal prediction models.

Saez and Barceló [34] predict air quality in Catalonia (Spain) through a *hierarchical Bayesian spatiotemporal* model at a low computational cost. The dataset contains hourly data of compound concentrations from diverse monitoring stations between 2011 and 2020. The error metrics to determine the performance have been *mean absolute percentage error* (MAPE) and RMSE. Except for $PM_{2.5}$ the model was able to improve spatial predictions.

Zhang [35] provides two models that consider spatial and temporal correlations for air quality prediction. On the one hand, the *Spatio-temporal Orthogonal Cube* model identifies the spatial and temporal components and links them together to generate a prediction framework. On the other hand, *Spatio-temporal Dynamic Advection* model that learns the spatial and temporal characteristics. Real data have been used to validate the proposed approaches, which improve on other existing methods.

Zhao et al. [36] propose *Relevance Data Cube* (RDC) a 3D structure based on clustering, time sliding windows and correlation to measure air quality. In addition, a spatio-temporal model is created with RDC to predict air quality based on CNN for the spatial component and LSTM for the temporal component. The proposed model has been used for the prediction of $PM_{2.5}$ concentration in the air. The dataset contains air quality data from various monitoring stations in Beijing and Tianjin during 2014 and 2015. The proposed model has been compared in terms of R^2, MAE, RMSE and MAPE with other ML and DL techniques obtaining the best results.

Deb and Tsay [37] use a spatio-temporal model that averages the values of spatial and temporal characteristics for air pollution data. The dataset used contains information on hourly concentrations of $PM_{2.5}$ collected at various monitoring stations across Taiwan between 2006 and 2015. The model is able to obtain consistent results and make accurate pollution predictions.

Lin et al. [38] propose an architecture based on GRU for the development of five models for the prediction of $PM_{2.5}$ taking into account spatial and temporal factors. In order to integrate the predictive models built, an ensemble learning forecasting metamodel based on multiple LR is also introduced. The data, taken between 2013 and 2018, are from 67 monitoring stations located in different areas of Taiwan. RMSE, MAE and absolute percent error less than 3% have been used to evaluate the proposed methods. For all three metrics, better results have been obtained compared to other ensemble learning models.

2.3. Conclusions of related works

As shown above, most of the recent work focuses on air quality prediction based on spatio-temporal characteristics using DL algorithms, especially LSTM. The spatial component in several papers is modeled from graphs, while the temporal component is implicit in the input data, as they are time series. Spatio-temporal attention mechanisms, interpolation techniques and ensemble learning have also been used in the literature. Most of these air quality studies have been conducted in Beijing and other parts of Asia. Additionally, the compound to be predicted in most cases is $PM_{2.5}$. To the best of our knowledge, there are no multi-objective evolutionary spatio-temporal prediction methods for air pollution or quality forecasting. Then, we propose a novel spatio-temporal approach based on multi-objective evolutionary algorithms for the identification of multiple non-dominated LR models and their subsequent combination in an ensemble learning model. The spatial component is included in the input data of the models. Unlike most of the works seen previously the compound to be predicted is NO_2 and the data have been collected within the Region of Murcia, Spain. The main reason for focusing our work on NO_2 is that Murcia is among the cities with the highest NO_2 levels [39]. Although Murcia city has achieved the highest reduction of NO_2 levels in the last year, the NO_2 levels remain high (4th highest), considering OMS recommendations (10 micrograms/m3). This situation leads to a normative change which, among other measures, obliges cities over 50.000 habitants to establish low emission areas. At the regional level, the presence of heavy industries, such as Escombreras Oil Refinery[1] and SABIC petrochemical[2] and the massive use of nitrogen fertilizers, makes pollution due to nitrogen oxides the subject of special monitoring.

3. A multi-objective optimization based spatio-temporal approach for air pollution forecasting

In this section we describe the proposed spatio-temporal approach to air pollution forecasting. Let $D_k = \{\mathbf{d}_1^k, \ldots, \mathbf{d}_{r_k}^k\}$ $(k = 1 \ldots, n)$ be normalized datasets with r_k instances. Each instance $\mathbf{d}_t^k = \{d_t^{k1}, \ldots, d_t^{kw}, o_t^k\}$, $t = 1, \ldots, r_k$, has w input attributes (after of sliding window transformation), and one output attribute $o_t^k \in [0, 1]$. Each D_k dataset has been obtained with observations at the air quality monitoring station E_k, $k = 1, \ldots, n$, which contain information on air quality and geographic coordinates (latitude and longitude) of the monitoring station.

The general scheme of this approach is shown graphically in Fig. 1. In this approach, a multi-objective optimization problem is defined where each objective consists of minimizing the RMSE of a LR model learned with the data from a monitoring station. Therefore, there are as many objectives as there are monitoring stations. Since the problem is a multi-objective optimization problem as in (1), the solution consists of a Pareto front formed by multiple LR models that allow making predictions in geographically distributed locations among the different monitoring stations, including the locations of the monitoring stations themselves. Once the Pareto front of the optimization problem has been identified, an ensemble learning model is built to combine the prediction models that form it. In the following subsections, the multi-objective optimization problem proposed in this paper for spatio-temporal forecasting is formulated mathematically, the mechanism used to combine the multiple LR models that form the Pareto front is described, and finally the main components of an MOEA to solve the multi-objective optimization problem are shown.

[1] https://cartagena.repsol.es/

[2] https://cartagena.sabic.com/es

3.1. Multi-objective optimization problem formulation

We define the following multi-objective optimization problem:

$$\text{Minimize} \quad \mathcal{F}_{D_k}(\mathbf{p}) \quad k = 1, \ldots, n \tag{2}$$

where $\mathbf{p} = \{p_0, p_1, \ldots, p_w\}$ is the set of decision variables, with $p_i \in [l, u] \subset \mathbb{R}$, $i = 0, \ldots, w$. The vector $\mathbf{p}$ represents the parameter set that defines a LR model as follows:

$$\hat{y}(\mathbf{x}) = p_0 + \sum_{i=1}^{w} p_i x_i \tag{3}$$

The vector $\mathbf{x} = (x_1, \ldots, x_w)$ represents an input data with w attributes. The function $\mathcal{F}_{D_k}(\mathbf{p})$ is the RMSE of the LR model $\hat{y}(\mathbf{x})$ which is defined by the parameter set $\mathbf{p}$:

$$\mathcal{F}_{D_k}(\mathbf{p}) = \sqrt{\frac{\sum_{t=1}^{r_k} \left(\hat{y}\left((d_t^{k1}, \ldots, d_t^{kw})\right) - o_t^k\right)^2}{r_k}} \tag{4}$$

The optimization problem (2) is an instance[3] of optimization problem (1), so the solution consists of a Pareto optimal set. In this paper, we will use meta-heuristic methods, specifically MOEAs, to find an approximation $\mathcal{S}' = \{s_1, \ldots, s_m\}$ of the true Pareto front $\mathcal{S}$. Each non-dominated solution s_j, $j = 1, \ldots, m$, represents an LR model.

3.2. Building the ensemble learning model

An ensemble [40] consists of the union of two or more base models to improve the predictions that each individual model would make, achieving a better generalization and robustness that a single model will reach. Some of the most common types of ensemble are *bagging* [41], *boosting* [42], or *stacking* [43]. However, not all ensemble learning methods can be applied in the context of our methodology. The bagging method is based on the random sampling of the dataset and therefore it is not applicable to assemble the LR models obtained with the MOEA. With the boosting method, each model tries to correct the errors of the previous models, so it is also not applicable in the space-temporal forecasting context of our methodology. Then, we use the stacking technique to combine the models and thus create the ensemble. Stacking method uses a meta-learner to combine the multiple predictions made by different base models trained with learning algorithms. The meta-learner is usually a regression model since it is a simple model that allows a better interpretation of the predictions made. However, any learning algorithm can be used for meta-learner training. The following process is performed to build the ensemble learning model with the set of LR models $\mathcal{S}' = \{s_1, \ldots, s_m\}$:

1. Evaluate each input data $(d_t^{k1}, \ldots, d_t^{kw})$, $t = 1, \ldots, r_k$, $k = 1, \ldots, n$ (a total of $r = \sum_{k=1}^{n} r_k$ instances) in each LR model s_j, $j = 1, \ldots, m$:

$$\hat{y}_{s_j}\left((d_t^{k1}, \ldots, d_t^{kw})\right) = p_0^j + \sum_{i=1}^{w} p_i^j d_t^{ki}, j = 1, \ldots, m, \; t = 1, \ldots, r_k, \; k = 1, \ldots, n \tag{5}$$

2. Build a dataset D with the evaluations of (5) and the corresponding observations o_t^k, $t = 1, \ldots, r_k$, $k = 1, \ldots, n$ as shown in Fig. 2.
3. Finally, a prediction model is built using the dataset D as training data and some learning algorithm for regression.

[3] Note that the names of the variables in the optimization problem (2) have been conveniently modified with respect to the names used in the optimization problem (1) for an adaptation to the context of LR.

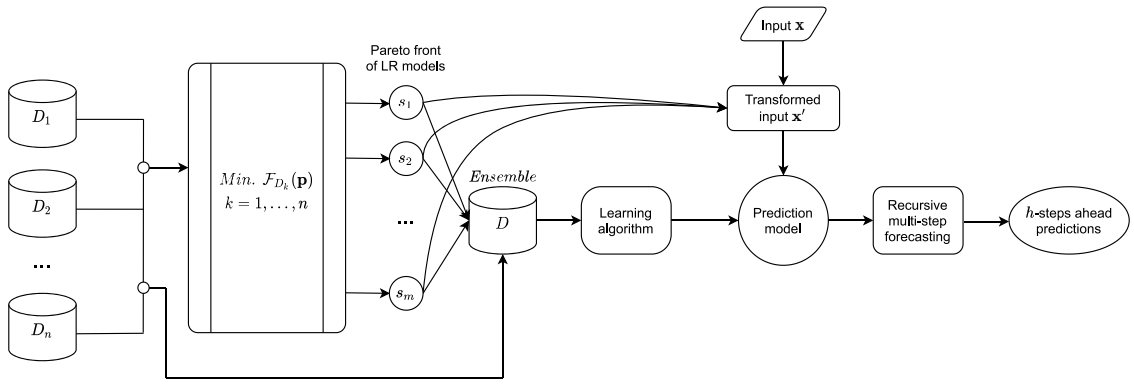


Fig. 1. Multi-objective optimization based spatio-temporal approach.

$$D = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ r_1 \\ r_1{+}1 \\ r_1{+}2 \\ \vdots \\ r_1{+}r_2 \\ \vdots \\ r_1{+}r_2{+}\ldots{+}r_{n-1}{+}1 \\ r_1{+}r_2{+}\ldots{+}r_{n-1}{+}2 \\ \vdots \\ r_1{+}r_2{+}\ldots{+}r_n \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \ldots & m \end{array} \\ \begin{pmatrix} \hat{y}_{s_1}\left((d_1^{11},\ldots,d_1^{1w})\right) & \hat{y}_{s_2}\left((d_1^{11},\ldots,d_1^{1w})\right) & \ldots & \hat{y}_{s_m}\left((d_1^{11},\ldots,d_1^{1w})\right) \\ \hat{y}_{s_1}\left((d_2^{11},\ldots,d_2^{1w})\right) & \hat{y}_{s_2}\left((d_2^{11},\ldots,d_2^{1w})\right) & \ldots & \hat{y}_{s_m}\left((d_2^{11},\ldots,d_2^{1w})\right) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{s_1}\left((d_{r_1}^{11},\ldots,d_{r_1}^{1w})\right) & \hat{y}_{s_2}\left((d_{r_1}^{11},\ldots,d_{r_1}^{1w})\right) & \ldots & \hat{y}_{s_m}\left((d_{r_1}^{11},\ldots,d_{r_1}^{1w})\right) \\ \hat{y}_{s_1}\left((d_1^{21},\ldots,d_1^{2w})\right) & \hat{y}_{s_2}\left((d_1^{21},\ldots,d_1^{2w})\right) & \ldots & \hat{y}_{s_m}\left((d_1^{21},\ldots,d_1^{2w})\right) \\ \hat{y}_{s_1}\left((d_2^{21},\ldots,d_2^{2w})\right) & \hat{y}_{s_2}\left((d_2^{21},\ldots,d_2^{2w})\right) & \ldots & \hat{y}_{s_m}\left((d_2^{21},\ldots,d_2^{2w})\right) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{s_1}\left((d_{r_2}^{21},\ldots,d_{r_2}^{2w})\right) & \hat{y}_{s_2}\left((d_{r_2}^{21},\ldots,d_{r_2}^{2w})\right) & \ldots & \hat{y}_{s_m}\left((d_{r_2}^{21},\ldots,d_{r_2}^{2w})\right) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{s_1}\left((d_1^{n1},\ldots,d_1^{nw})\right) & \hat{y}_{s_2}\left((d_1^{n1},\ldots,d_1^{nw})\right) & \ldots & \hat{y}_{s_m}\left((d_1^{n1},\ldots,d_1^{nw})\right) \\ \hat{y}_{s_1}\left((d_2^{n1},\ldots,d_2^{nw})\right) & \hat{y}_{s_2}\left((d_2^{n1},\ldots,d_2^{nw})\right) & \ldots & \hat{y}_{s_m}\left((d_2^{n1},\ldots,d_2^{nw})\right) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{s_1}\left((d_{r_n}^{n1},\ldots,d_{r_n}^{nw})\right) & \hat{y}_{s_2}\left((d_{r_n}^{n1},\ldots,d_{r_n}^{nw})\right) & \ldots & \hat{y}_{s_m}\left((d_{r_n}^{n1},\ldots,d_{r_n}^{nw})\right) \end{pmatrix} \end{array} \begin{array}{c} output \\ \begin{pmatrix} o_1^1 \\ o_2^1 \\ \vdots \\ o_{r_1}^1 \\ o_1^2 \\ o_2^2 \\ \vdots \\ o_{r_2}^2 \\ \vdots \\ o_1^n \\ o_2^n \\ \vdots \\ o_{r_n}^n \end{pmatrix} \end{array}$$

Fig. 2. Dataset D for stacking based ensemble learning.

3.3. Multi-step forecasting process

Once the ensemble learning model has been built, the process for the h-steps ahead prediction of a new input data $\mathbf{x} = (x_1, \ldots, x_w)$ requires the following steps:

1. Transform the input data $\mathbf{x}$ (with w air quality attributes) to a data $\mathbf{x}'$ with ensemble format (with m attributes corresponding to the evaluation of $\mathbf{x}$ in the m LR models of the ensemble):

$$\mathbf{x} = (x_1, \ldots, x_w) \rightarrow \mathbf{x}' = (\hat{y}_{s_1}(\mathbf{x}), \ldots, \hat{y}_{s_m}(\mathbf{x})) \tag{6}$$

where:

$$\hat{y}_{s_j}(\mathbf{x}) = p_0^j + \sum_{i=1}^{w} p_i^j x_i \tag{7}$$

2. Perform multi-step forecasting. This task consists of predicting the next h values of the output time series, where $h > 1$ denotes the *forecasting horizon*. We use the *recursive strategy* [44] to perform multi-step forecasting. The recursive strategy trains first a one-step model (in our case, the ensemble learning model) and then uses it recursively for returning a multi-step prediction.

3.4. A multi-objective evolutionary algorithm for spatio-temporal air quality forecasting

In this paper, we propose the use of MOEAs to solve the multi-objective optimization problem (2). MOEAs are meta-heuristics inspired by Darwin's mechanisms of natural selection, crossover, and mutation. MOEAs facilitate the optimization of complex optimization problems with several objective functions that may conflict with each other. The use of MOEAs as search methods allows, starting from an initial population, to converge to a set of diverse solutions as close as possible to the Pareto front. The components of the proposed MOEA are as follows:

Representation. We use *floating-point chromosome representation* of length $w+1$ for the individuals in the population. Therefore, an individual $\mathcal{I}$ is represented by a vector of real number as follows:

$$I = \{p_0^I, \ldots, p_w^I\}$$

where $p_i^I \in [-1, 1]$, $i = 0, \ldots, w$, represent the coefficients of a LR model including the term independent term.

Initial population. The coefficients of the individuals in the population are randomly generated in the interval $[-1,1]$ with a uniform distribution.

Fitness functions. The fitness functions for an individual I of the population correspond to the functions $\mathcal{F}_{D_k}(I)$, $k = 1, \ldots, n$, defined in (2), where $\mathcal{F}$ is the RMSE, the individual I represents the vector of coefficients of an LR model, and D_k is the dataset of the monitoring station E_k, $k = 1, \ldots, n$.

Variation operators. We use the variation operators *simulated binary crossover* [45] and *polynomial mutation* [46]. These operators are set by default in the *Platypus platform*[4] for MOEAs with floating-point representation.

Mechanisms of selection, sampling and generational replacement. These three mechanisms are specific to the particular MOEA used for multi-objective optimization. Concretely, we have selected for the experiments the *non-dominated sorting genetic algorithm II* (NSGA-II [47]), the *multi-objective evolutionary algorithm based on decomposition* (MOEA/D [48]) and the *strength Pareto evolutionary algorithm 2* (SPEA2 [49]). These MOEAs are three of the most popular algorithms used in the literature for multi-objective optimization.

4. Experiments and results

In this section, we show the experiments carried out and the results. Section 4.1 describes the air quality datasets used in the experiments and their preprocessing. In the first set of experiments, described in Section 4.2, different MOEAs are statistically compared. For the best MOEA, the results of the hypervolume evolution are shown, as well as the Pareto front found for a multi-objective optimization problem with three monitoring stations in south-eastern Spain. Section 4.3 describes the second set of experiments, in which the effectiveness of the multi-objective optimization based spatio-temporal approach is tested by comparing it with a simple approach consisting only in the union of the datasets from the different monitoring stations into a single dataset on which to apply some regression algorithm. Finally, in the third set of experiments, the multi-objective optimization based spatio-temporal approach is compared with an interpolation-based spatio-temporal approach.

4.1. Air quality databases

Data are extracted from the *Autonomous Community of the Region of Murcia*[5] (Spain), which provides information on air quality in the Region of Murcia thanks to *Consejería de Agua, Agricultura, Ganadería, Pesca y Medio Ambiente*. The data collected comes from four different monitoring stations located in the Region of Murcia. These stations are located in the towns of La Aljorra (longitude −1.06588, latitude 37.692500), Alcantarilla (longitude −1.232139, latitude 37.974500), Lorca (longitude −0.926556, latitude 37.687833) and Valle de Escombreras (longitude −1.702778, latitude 37.574444), as shown in Fig. 3. The data have been measured daily between 2017 and 2020, therefore each dataset has 1461 instances. Initially, the datasets had 19 attributes: *Date*, *Latitude*, *Longitude*, NO, NO_2, SO_2, O_3, *TMP* (temperature), *HR* (relative humidity), NO_X, *DD* (wind direction), *PRB* (atmospheric pressure), *RS* (solar radiation), *VV* (wind speed), C_6H_6, C_7H_8, *XIL*, PM_{10}, *Noise*.

For the initial preprocessing, all attributes with more than 25% missing values were first removed from all datasets. These attributes were: C_6H_6, C_7H_8, *XIL*, *HR*, *PRB* and *Noise*. *Date* attribute has also been removed as it is a string that does not provide relevant information to the problem. Table 1 shows a summary of the final attributes for La Aljorra dataset. Linear interpolation has been applied to deal with the remaining missing values.

For each dataset, a lagged transformation of the input variables[6] has to be done with *sliding window* method [50] in order to remove time dependencies in the data. A window size of 7 has been selected, representing one week forecast. The rows with missing values resulting from the sliding window transformation have been removed. Therefore, each transformed dataset have 1454 rows and 73 columns. Finally, the datasets have been normalized and split into 70% for training and 30% for testing, preserving the order of the instances.

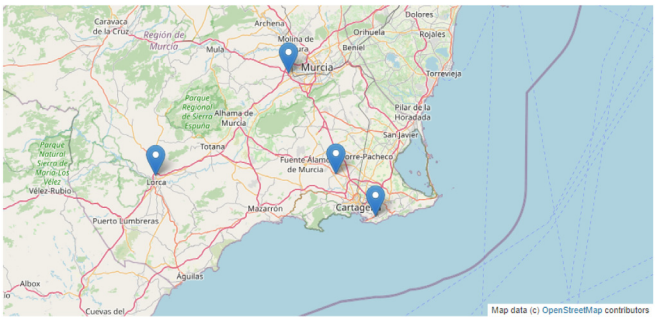


Fig. 3. Map with the localization of the monitoring stations.

Table 1
Summary of attributes of the La Aljorra dataset.

Attribute	Count	Mean	Std	Min	Max
Latitude (φ)	1461	37.69250	0.00	37.69250	37.69250
Longitude (λ)	1461	−1.06588	0.00	−1.06588	−1.06588
NO	1272	4.38	2.43	1.00	31.00
NO_2	1272	14.64	8.19	2.00	58.00
SO_2	1299	9.21	3.45	2.00	23.00
O_3	1403	57.94	15.25	19.00	112.00
TMP	1409	19.59	5.74	4.00	32.00
NO_X	1272	21.21	11.20	3.00	104.00
DD	1409	192.11	104.45	0.00	360.00
RS	1409	182.32	82.74	13.00	338.00
VV	1409	1.22	0.43	1.00	3.00
PM_{10}	1381	26.27	12.98	5.00	168.00

Table 2
Summary of the results of the MOEAs, 10,000 evaluations, 30 runs.

Algorithm	Best	Worse	Average	SD
NSGA-II	0.35393	0.22136	0.30934	0.03169
MOEA/D	0.33819	0.17268	0.26805	0.04673
SPEA2	0.25006	0.14090	0.19320	0.02874

4.2. Comparison of multi-objective evolutionary algorithms

We have compared the NSGA-II, MOEA/D and SPEA2 algorithms to check which of them performs better in a 3-objective optimization problem with training data (70%) from Lorca, Alcantarilla and Valle de Escombreras monitoring stations. These three MOEAs were run with population size of 100, 100 generations (10,000 evaluations of the objective function in total), 30 runs, a crossover probability of 1.0 and a mutation probability of 1.0. *Hypervolume indicator* [51] is used to make the comparisons. For the calculation of the hypervolume, the minimum and maximum of each objective function have been set to 0 and 1, respectively, since the data are normalized. We have performed a Mann–Whitney U statistical test to make a win–loss ranking of the algorithms and select the best one. Table 2 shows the best, worst, mean hypervolume and standard deviation of each MOEA, and Table 3 shows the number of times that each MOEA has been better with a statistically significant difference than another MOEA (win), the number of times that it has been worse (loss) and the difference (win – loss), ordered from greater to lesser difference.

[4] https://platypus.readthedocs.io/en/latest/

[5] https://sinqlair.carm.es/calidadaire/redvigilancia/redvigilancia.aspx

[6] Note that the lagged transformation has not been applied to φ and λ since they are not temporal attributes.

Table 3
Ranking of the MOEAs with 100,00 evaluations and 30 runs sorted from best to worse.

MOEA	Win	Loss	Win − Loss
NSGA-II	2	0	2
MOEA/D	1	1	0
SPEA2	0	2	−2

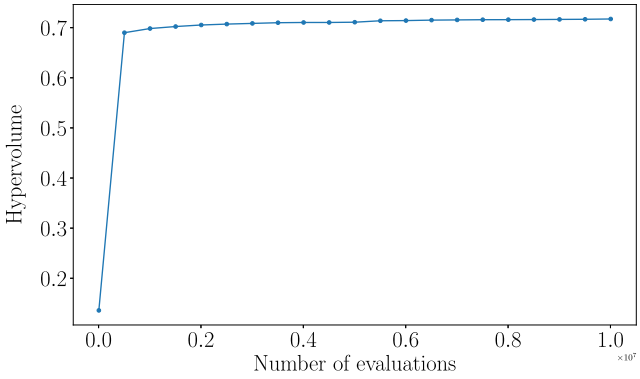


Fig. 4. Hypervolume evolution with NSGA-II, 10 million evaluations.

Once NSGA-II has been identified as the best MOEA, it has been run through 10 million evaluations and crossover or mutation probabilities of 1.0 to find the final solution. Fig. 4 shows the evolution of the hypervolume over the number of evaluations, and Fig. 5 shows the Pareto front obtained, visualized both in 3D and in 2D for each pair of objectives.

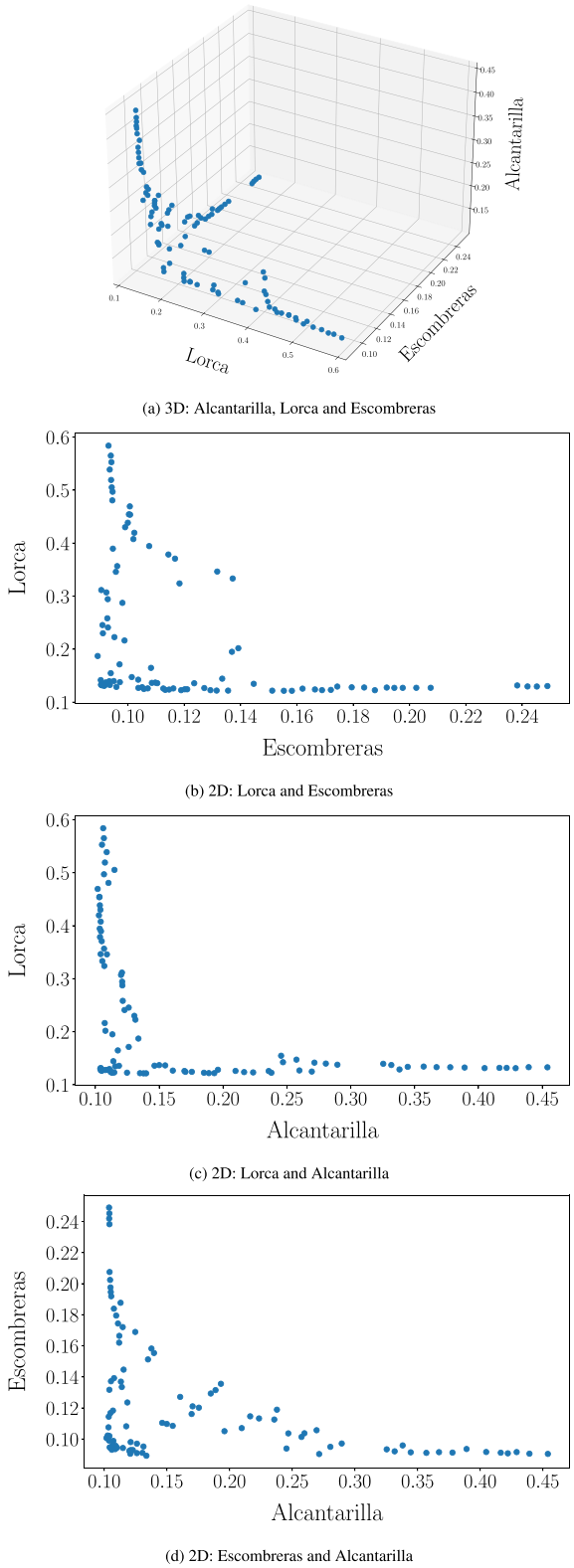
4.3. Testing the effectiveness of the multi-objective optimization based spatio-temporal approach

The objective of this second set of experiments is to test the effectiveness of the multi-objective optimization spatio-temporal approach proposed in this paper for air pollution forecasting. To do this, the proposed approach is compared with a second simpler spatio-temporal approach in which the multi-objective search and the construction of the ensemble learning model have been eliminated. This second approach consists of building a dataset D resulting from the union of the datasets D_k, $k = 1 \ldots, n$:

$$D = \bigcup_{k=1}^{n} D_k \tag{8}$$

and then apply some regression algorithm and recursive multi-step forecasting to make h-steps ahead predictions. Fig. 6 shows the general scheme of this other approach.

To make the comparisons, statistical tests are applied with a 10-fold cross-validation, 10 repetitions (a total of 100 prediction models). The statistical tests have been performed using the dataset D obtained in each of the two space-temporal approaches (which we have called *Ensemble* and *Union* in Figs. 1 and 6 respectively) and the learning algorithms LR, *random forest* (RF) [52], *support vector machine* (SVM) [53], *multi-layer perceptron* (MLP) [54], *k-nearest neighbors* (kNN) [55], *quasi-recurrent neural networks* (QRNN) [56] and *ZeroR* [57]. We used *paired t-test* for these experiments. The results of the paired t-tests are used to make a win–loss ranking, both for the datasets and for the learning algorithms, thus allowing to determine which are the best. The Table 4 shows the evaluation and results of the paired t-test using the *Ensemble* (baseline) and *Union* datasets with the different learning algorithms. An asterisk means that the model is statistically worse than the baseline. Table 5 shows the



(a) 3D: Alcantarilla, Lorca and Escombreras

(b) 2D: Lorca and Escombreras

(c) 2D: Lorca and Alcantarilla

(d) 2D: Escombreras and Alcantarilla

Fig. 5. Pareto front in 3D and 2D with NSGA-II, 10 million evaluations.

number of times that each dataset has been better with a statistically significant difference than the other (win), the number of times that it has been worse (loss) and the difference (win – loss), ordered from greater to lesser difference. Table 6 shows the

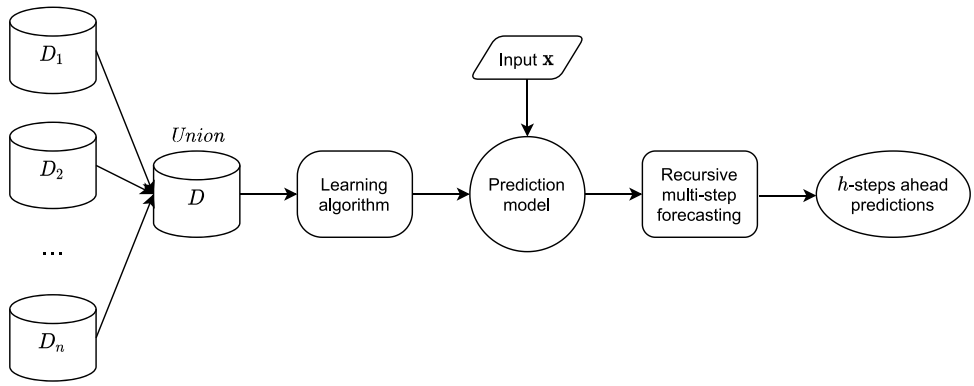


Fig. 6. Dataset union spatio-temporal approach.

Table 4
RMSE evaluation, 10-fold cross-validation, 10 repetitions, and statistical test of datasets (baseline: *Ensemble*).

Algorithm	Ensemble	Union	
LR	0.108788	0.108791	
RF	0.107863	0.108488	
SVM	0.108686	0.109298	
MLP	0.109847	0.113576	*
kNN	0.142134	0.157161	*
QRNN	0.107168	0.109023	*
ZeroR	0.153978	0.153978	

Table 5
Win–loss statistical test of datasets.

Dataset	Win	Loss	Win − Loss
Ensemble	3	0	3
Union	0	3	−3

Table 6
Win–loss statistical test of algorithm.

Algorithm	Win	Loss	Win − Loss
RF	6	0	6
LR	5	0	5
SVM	5	0	5
QRNN	4	0	4
MLP	4	4	0
kNN	1	10	−9
ZeroR	0	11	−11

Table 7
Summary of the statistical test of datasets.

	Ensemble	Union
Ensemble	–	6 (3)
Union	0 (0)	–

Table 8
Summary of the statistical test of algorithms.

	LR	RF	SVM	MLP	kNN	QRNN	ZeroR
LR	–	0 (0)	1 (0)	2 (1)	2 (2)	1 (0)	2 (2)
RF	2 (0)	–	2 (0)	2 (2)	2 (2)	1 (0)	2 (2)
SVM	1 (0)	0 (0)	–	2 (1)	2 (2)	0 (0)	2 (2)
MLP	0 (0)	0 (0)	0 (0)	–	2 (2)	0 (0)	2 (2)
kNN	0 (0)	0 (0)	0 (0)	0 (0)	–	0 (0)	1 (1)
QRNN	1 (0)	1 (0)	2 (0)	2 (0)	2 (2)	–	2 (2)
ZeroR	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)	0 (0)	–

win–loss ranking for the different algorithms in this case. Tables 7 and 8 summarize the results of the paired t-tests applied to the datasets and the algorithms, respectively. An entry a (b) in these tables represents the number of datasets (or learning algorithms) that the column has been worse than row (a) and statistically worse than row (b).

4.4. Comparison with the interpolation based spatio-temporal approach

Finally, we compare our approach with another widely used in the literature [31,32], an interpolation-based spatio-temporal approach. To accomplish this, we first build a regression model with each dataset D_k, $k = 1, \ldots, n$. The set of prediction models $\{\delta_1, \ldots, \delta_n\}$ is used to make h-steps ahead predictions with recursive multi-step forecasting. Subsequently, an interpolation of these forecasts is applied to estimate the predictions of new observations made at any geographical point. Fig. 7 describes graphically the scheme of this spatio-temporal approach. The interpolation method is based on an *inverse distance weighting* (IDW) function [58]. This technique assigns more weight to those predictions that are geographically closer to the point to be interpolated, since they will have a higher influence in that region. Finding the interpolated value $\mathcal{U}$ at a geographical observation $\mathbf{x}$ based on the predictions $\mathbf{y} = (\hat{y}_{\delta_1}(\mathbf{x}), \ldots, \hat{y}_{\delta_n}(\mathbf{x}))$ can be formulated as follows:

$$\mathcal{U}(\mathbf{x}) = \begin{cases} \dfrac{\sum_{k=1}^{n} \mathcal{W}_k(\mathbf{x}) \cdot \hat{y}_{\delta_k}(\mathbf{x})}{\sum_{k=1}^{n} \mathcal{W}_k(\mathbf{x})} & \text{if } d(\mathbf{x}, E_k) \neq 0 \text{ for all } k, \\ \hat{y}_{\delta_k}(\mathbf{x}) & \text{if } d(\mathbf{x}, E_k) = 0 \text{ for some } k, \end{cases} \tag{9}$$

where:

$$\mathcal{W}_k(\mathbf{x}) = \frac{1}{d(\mathbf{x}, E_k)^p} \tag{10}$$

is the IDW function, and $d(x, E_k)$ is the *haversine distance* [59], considering the curvature of the Earth, between the geographical location of observation $\mathbf{x}$ and the monitoring station E_k. In the IDW function, p is the rate at which the weights decrease, called the *power parameter*. We have used $p = 2$ in the experiments.

To compare our multi-objective optimization based approach with the interpolation-based approach we have performed 7-steps ahead predictions with the with learning algorithms that have never lost in the win–loss statistical test ranking (RF, LR, SVM and QRNN), and the RMSE, MAE and *correlation coefficient* (CC) of the prediction models are used to compute their *goodness*. The lower the goodness value, the better the results. The goodness of a prediction model is calculated as the weighted mean of the mean RMSE, MAE, and CC values of the model in the h-steps ahead predictions. In this work, the weights used are the same for the three metrics RMSE, MAE and CC. Tables 9, 10, 11 and 12 show the results of the evaluation on the training set of 7-steps ahead predictions for the spatio-temporal approaches with RF, LR SVM and QRNN, respectively. Tables 13, 14, 15 and 16 show the

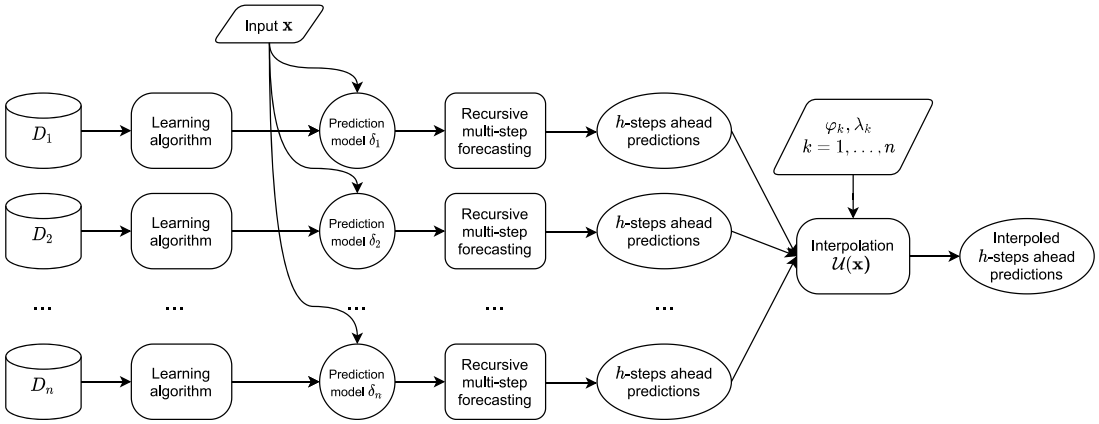


Fig. 7. Interpolation based spatio-temporal approach.

Table 9
Results of the evaluation of models on training set with RF.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.0646	0.0818	0.0927	0.1001	0.1057	0.1097	0.1159
	MAE	0.0472	0.0596	0.0680	0.0738	0.0778	0.0815	0.0864
	CC	0.9162	0.8529	0.8015	0.7612	0.7276	0.7022	0.6606
Interpolation	RMSE	0.0604	0.0704	0.0762	0.0796	0.0826	0.0848	0.0891
	MAE	0.0425	0.0488	0.0526	0.0549	0.0569	0.0586	0.0627
	CC	0.9339	0.9056	0.8854	0.8726	0.8603	0.8520	0.8335

Table 10
Results of the evaluation of models on training set with LR.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.1244	0.1262	0.1264	0.1264	0.1266	0.1273	0.1326
	MAE	0.0943	0.0957	0.0959	0.0959	0.0959	0.0966	0.1007
	CC	0.5936	0.5779	0.5759	0.5757	0.5746	0.5685	0.5187
Interpolation	RMSE	0.1244	0.1262	0.1264	0.1265	0.1266	0.1273	0.1326
	MAE	0.0943	0.0957	0.0959	0.0959	0.0960	0.0966	0.1007
	CC	0.5935	0.5778	0.5757	0.5756	0.5744	0.5683	0.5185

Table 11
Results of the evaluation of models on training set with SVM.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.1060	0.1095	0.1139	0.1248	0.1287	0.1357	0.1503
	MAE	0.0810	0.0835	0.0860	0.0920	0.0938	0.0977	0.1079
	CC	0.7275	0.7053	0.6742	0.5969	0.5681	0.5129	0.3933
Interpolation	RMSE	0.0911	0.0930	0.0950	0.0968	0.0990	0.1036	0.1109
	MAE	0.0754	0.0770	0.0782	0.0796	0.0809	0.0837	0.0888
	CC	0.8080	0.7986	0.7884	0.7786	0.7664	0.7406	0.6960

Table 12
Results of the evaluation of models on training set with QRNN.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.0805	0.0814	0.0830	0.0855	0.0855	0.0876	0.0963
	MAE	0.1064	0.1077	0.1097	0.1126	0.1126	0.1154	0.1261
	CC	0.7249	0.7170	0.7030	0.6829	0.6822	0.6630	0.5807
Interpolation	RMSE	0.0802	0.0805	0.0806	0.0809	0.0810	0.0816	0.0835
	MAE	0.1056	0.1059	0.1061	0.1064	0.1067	0.1075	0.1099
	CC	0.7386	0.7375	0.7367	0.7352	0.7338	0.7295	0.7169

results of the evaluation on La Aljorra test set of 7-steps ahead predictions for the spatio-temporal approaches with RF, LR, SVM and QRNN, correspondingly. The goodness of the RF, LR, SVM and QRNN models is shown in Table 21 for the four approaches. Figs. B.8, B.9, B.10 and B.11 depict the RMSE, MAE and CC on training and test sets of 7-steps ahead predictions with the RF, LR, SVM and QRNN, correspondingly. The times series of the 7-steps-ahead predictions for NO_2 on La Aljorra test set with RF, LR, SVM and QRNN model can be found in Figs. B.12, B.13, B.14 and B.15, respectively. A study of overfitting has also been carried out for RF, LR, SVM and QRNN models, as shown in Tables 17, 18, 19 and 20. In the event that the result is greater than 1, it is considered that there could be overfitting in the data, in the event that were less than 1 there could be underfitting in the data. The closer the values are to 1, the better, since it implies that the results of train and test are similar.

Table 13
Results of the evaluation of models on La Aljorra test set with RF.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.1185	0.1270	0.1428	0.1600	0.1641	0.1888	0.2386
	MAE	0.1003	0.1077	0.1218	0.1356	0.1389	0.1607	0.2062
	CC	0.6545	0.6240	0.5678	0.4873	0.4703	0.3849	0.2648
Interpolation	RMSE	0.1409	0.1520	0.1590	0.1644	0.1692	0.1737	0.1793
	MAE	0.1117	0.1223	0.1296	0.1354	0.1411	0.1466	0.1531
	CC	0.2994	0.3009	0.2989	0.3041	0.3094	0.3118	0.3163

Table 14
Results of the evaluation of models on La Aljorra test set with LR.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.1118	0.1207	0.1229	0.1233	0.1207	0.1261	0.1408
	MAE	0.0890	0.0960	0.0980	0.0984	0.0961	0.1014	0.1154
	CC	0.5410	0.4863	0.4745	0.4724	0.4860	0.4581	0.3830
Interpolation	RMSE	0.1175	0.1207	0.1212	0.1296	0.1322	0.1346	0.1352
	MAE	0.0919	0.0936	0.0939	0.1014	0.1027	0.1035	0.1035
	CC	0.3199	0.2673	0.2393	0.2420	0.2574	0.2057	0.1874

Table 15
Results of the evaluation of models on La Aljorra test set with SVM.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.1097	0.1190	0.1340	0.1388	0.1379	0.1454	0.1659
	MAE	0.0893	0.0962	0.1071	0.1106	0.1093	0.1158	0.1330
	CC	0.5597	0.4964	0.3989	0.3662	0.3733	0.3368	0.2473
Interpolation	RMSE	0.1553	0.1589	0.1580	0.1639	0.1639	0.1666	0.1701
	MAE	0.1298	0.1342	0.1334	0.1398	0.1401	0.1424	0.1455
	CC	0.3307	0.3117	0.3102	0.3173	0.3133	0.3051	0.2920

Table 16
Results of the evaluation of models on La Aljorra test set with QRNN.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.1025	0.1073	0.1227	0.1366	0.1397	0.1579	0.1987
	MAE	0.1214	0.1269	0.1436	0.1591	0.1626	0.1815	0.2267
	CC	0.6389	0.6183	0.5549	0.4965	0.4836	0.4157	0.2566
Interpolation	RMSE	0.1198	0.1208	0.1213	0.1224	0.1225	0.1243	0.1274
	MAE	0.0946	0.0958	0.0965	0.0979	0.0980	0.0996	0.1030
	CC	0.3037	0.3058	0.3092	0.3153	0.3214	0.3246	0.3192

Table 17
Overfitting ratio of models on train and La Aljorra test set with RF.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	1.8344	1.5526	1.5405	1.5984	1.5525	1.7211	2.0587
	MAE	2.1250	1.8070	1.7912	1.8374	1.7853	1.9718	2.3866
	CC	1.3998	1.3668	1.4116	1.5621	1.5471	1.8244	2.4947
Interpolation	RMSE	2.3328	2.1591	2.0866	2.0653	2.0484	2.0483	2.0123
	MAE	2.6282	2.5061	2.4639	2.4663	2.4798	2.5017	2.4418
	CC	3.1192	3.0096	2.9622	2.8695	2.7805	2.7325	2.5292

Table 18
Overfitting ratio of models on train and La Aljorra test set with LR.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	0.8987	0.9564	0.9723	0.9755	0.9534	0.9906	1.0618
	MAE	0.9438	1.0031	1.0219	1.0261	1.0021	1.0497	1.1460
	CC	1.0596	0.9969	0.9786	0.9746	0.9979	0.9527	0.8726
Interpolation	RMSE	0.9445	0.9564	0.9589	1.0245	1.0442	1.0573	1.0196
	MAE	0.9745	0.9781	0.9791	1.0574	1.0698	1.0714	1.0278
	CC	1.0261	1.0224	1.0213	0.9458	0.9348	0.9333	0.9729

5. Analysis of results and discussion

The following analysis follows from the results obtained with the experiments:

- In the search for an appropriate MOEA to solve the multi-objective optimization problem proposed in this paper, the NSGA-II algorithm has shown better performance than the MOEA/D and SPEA2 algorithms in terms of hypervolume. The hypervolume metric measures the space dominated by the set of non-dominated solutions found by a MOEA, and is a valid indicator for both optimality and diversity. The best, worst and mean hypervolume found by NSGA-II in 30 runs, as well as the standard deviation, is better than that

Table 19
Overfitting ratio of models on train and La Aljorra test set with SVM.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	1.0345	1.0869	1.1769	1.1126	1.0714	1.0712	1.1037
	MAE	1.1024	1.1521	1.2449	1.2017	1.1658	1.1857	1.2323
	CC	1.2998	1.4209	1.6902	1.6301	1.5218	1.5230	1.5905
Interpolation	RMSE	1.7051	1.7084	1.6634	1.6931	1.6558	1.6081	1.5338
	MAE	1.7224	1.7422	1.7054	1.7561	1.7317	1.7015	1.6387
	CC	2.4436	2.5618	2.5418	2.4541	2.4459	2.4272	2.3834

Table 20
Overfitting ratio of models on train and La Aljorra test set with QRNN.

Method	Metric	1-step ahead	2-steps ahead	3-steps ahead	4-steps ahead	5-steps ahead	6-steps ahead	7-steps ahead
MOEA + Ensemble	RMSE	1.2723	1.3173	1.4789	1.5972	1.6332	1.8014	2.0638
	MAE	1.1414	1.1786	1.3094	1.4138	1.4438	1.5731	1.7968
	CC	1.1346	1.1596	1.2669	1.3754	1.4107	1.5949	2.2630
Interpolation	RMSE	1.4933	1.5004	1.5042	1.5131	1.5114	1.5237	1.5266
	MAE	0.8964	0.9047	0.9101	0.9201	0.9186	0.9267	0.9378
	CC	2.4317	2.4118	2.3829	2.3315	2.2834	2.2476	2.2458

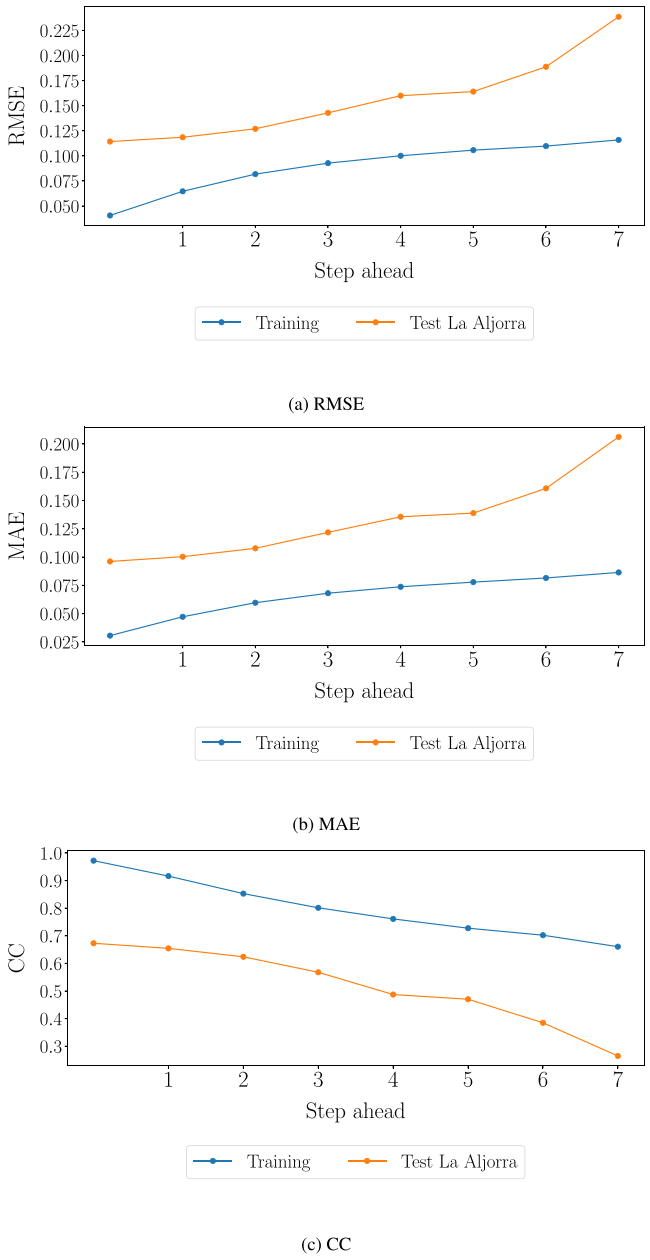
Table 21
Goodness of the predictions models with RF, LR, SVM and QRNN.

Models	Goodness RF	Goodness LR	Goodness SVM	Goodness QRNN
Training MOEA + Ensemble	0.130603	0.218094	0.206317	0.173174
Training Interpolation	0.081890	0.218155	0.136968	0.151813
La Aljorra Test MOEA + Ensemble	0.269400	0.250435	0.282549	0.267744
La Aljorra Test Interpolation	0.339744	0.326777	0.329596	0.302130

Table A.22
Abbreviations.

Abbreviation	Meaning
CC	Correlation Coefficient
CNN	Convolutional Neural Network
CO	Carbon Monoxide
CO_2	Carbon Dioxide
DL	Deep Learning
EA	Evolutionary Algorithm
GC-DCRNN	Geo-Context base Diffusion Convolutional Recurrent Neural Network
IDW	Inverse Distance Weighting
kNN	k-Nearest Neighbors
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
MMSL	Multi-output and Multi-index of Supervised Learning
MOEA	Multi-Objective Evolutionary Algorithm
MOEA/D	Multi-Objective Evolutionary Algorithm based no Decomposition
NO_2	Nitrogen dioxide
NO_X	Nitrogen oxides
NSGA-II	Non-dominated Sorting Genetic Algorithm II
O_3	Ozone
QRNN	Quasi-Recurrent Neural Networks
PM	Particulate Matter
PM_{10}	Particulate Matter with a diameter of less than 10 micrometers
$PM_{2.5}$	Particulate Matter with a diameter of less than 2.5 micrometers
R^2	Pearson's correlation coefficient
RDC	Relevance Data Cube
RF	Random Forest
RMSE	Root Mean Squared Error
SO_2	Sulfur Dioxide
SPEA2	Strength Pareto Evolutionary Algorithm 2
SVM	Support Vector Machine
WS	Window Size

obtained by MOEA/D and SPEA2, with statistically significant differences. This allows us to conclude that NSGA-II is better than MOEA/D and SPEA2 for the multi-objective optimization problem considered. However, this problem is a three-objective optimization problem, and the performance test should be performed again in the case of having a *many-objective optimization* problem (more than three objectives). There are specialized MOEAs, such as NSGA-III [60] or IBEA [61] that will provide better results for

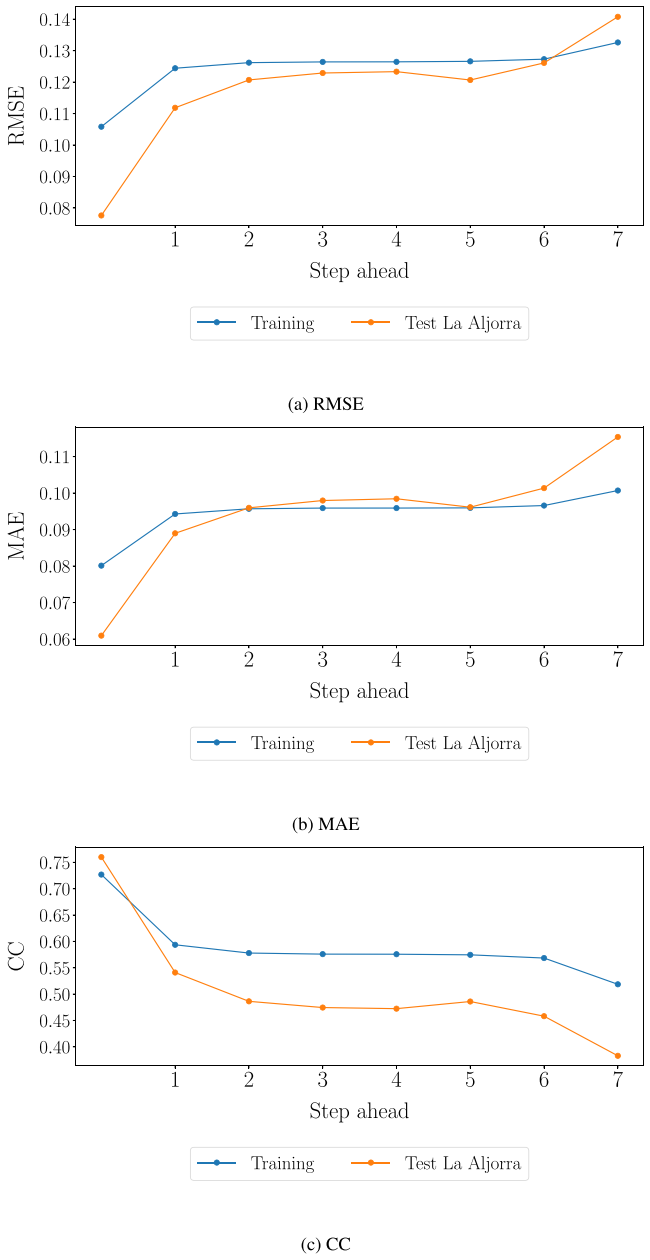


(a) RMSE

(b) MAE

(c) CC

Fig. B.8. Results of the evaluation of the model built with the multi-objective optimization based spatio-temporal approach and RF.



(a) RMSE

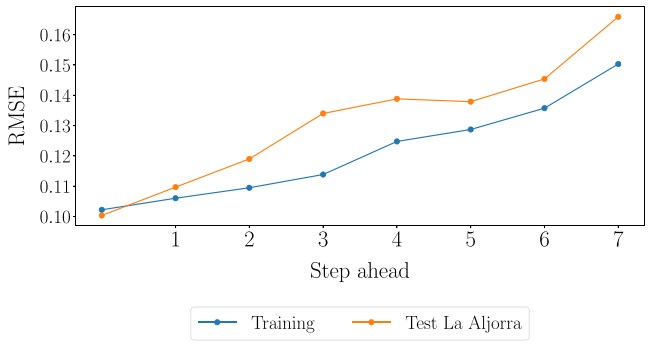
(b) MAE

(c) CC

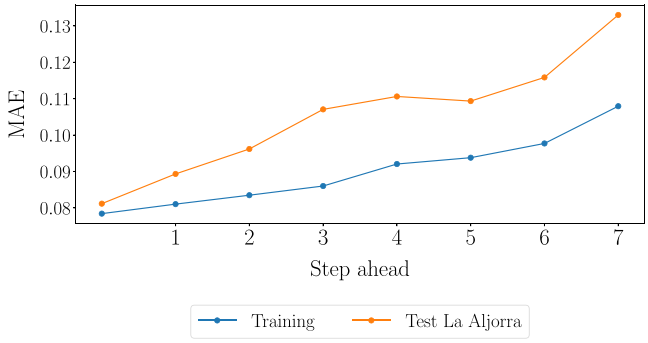
Fig. B.9. Results of the evaluation of the model built with the multi-objective optimization based spatio-temporal approach and LR.

many-objective optimization, as well as more appropriate performance metrics, e.g. *inverted generational distance* [62].

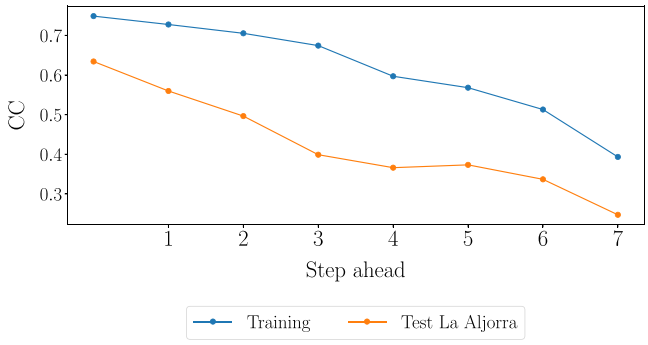
- We have statistically verified that the identification of a Pareto front of LR models distributed in the study area to then build an ensemble learning regression model is better than building a regression model with the simple union of the datasets that delimit the study area. It is to be expected that the objectives defined in problem (2) are conflicting objectives, since normally the contamination levels in each monitoring station do not oscillate in the same way. In the case study of this work, we can verify that the 3 objectives are conflicting since the MOEA identifies a Pareto front with the 100 individuals of the population distributed in the 3 dimensions (as can be seen in Fig. 5). Therefore, assuming that the objectives are conflicting, the number of LR models on the Pareto front can be controlled by the size of the population in the MOEA. It is also expected that the higher this number, the greater the generalization error of the ensemble learning model. We have also statistically verified that RF, LR, SVM and QRNN are the best regression algorithms to build ensemble learning models in the spatio-temporal scenario studied in this paper, beating regression algorithms such as MLP or kNN. If we look at the goodness metric in Table 21 we can see that, in test set, the best model is LR. Although QRNN is statistically worse than LR, the best evaluation of RMSE in 10-fold cross-validation (10 repetitions) is obtained with the QRNN model with the ensemble approach, as shown in Table 4. Moreover, QRNN is the second best model in test in terms of goodness.
- The spatio-temporal approach based on multi-objective optimization has also been shown to be better than the interpolation method used by other authors for spatio-temporal



(a) RMSE

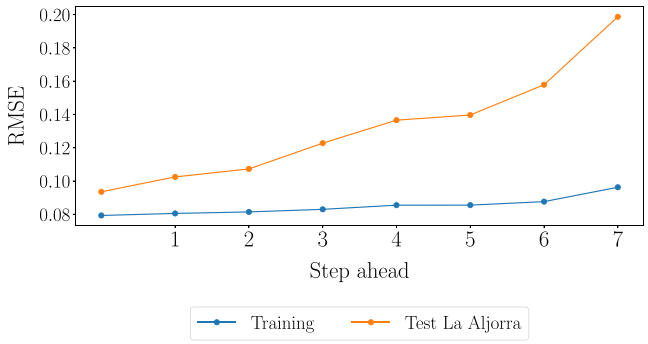


(b) MAE

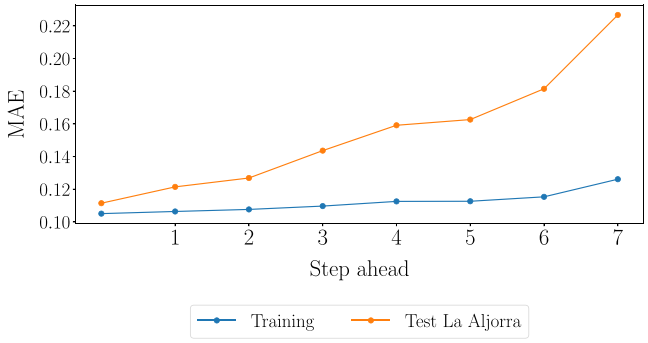


(c) CC

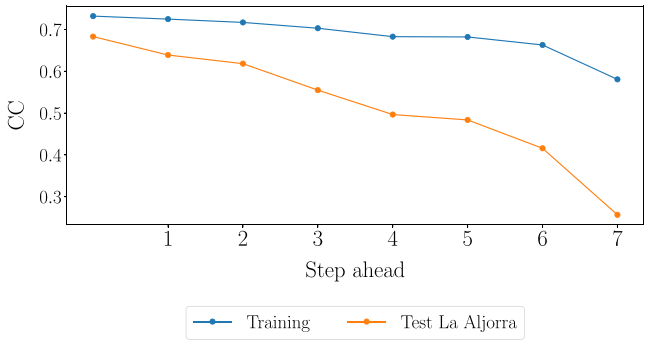
Fig. B.10. Results of the evaluation of the model built with the multi-objective optimization based spatio-temporal approach and SVM.



(a) RMSE



(b) MAE

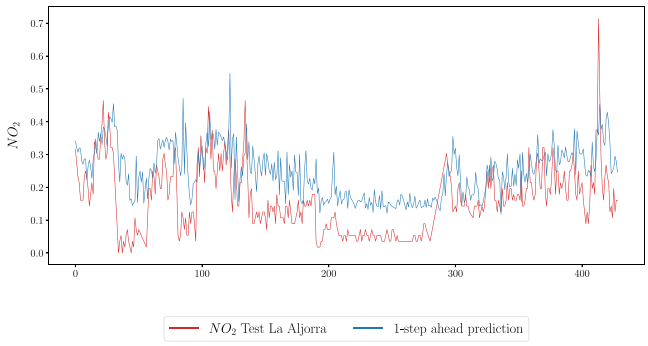


(c) CC

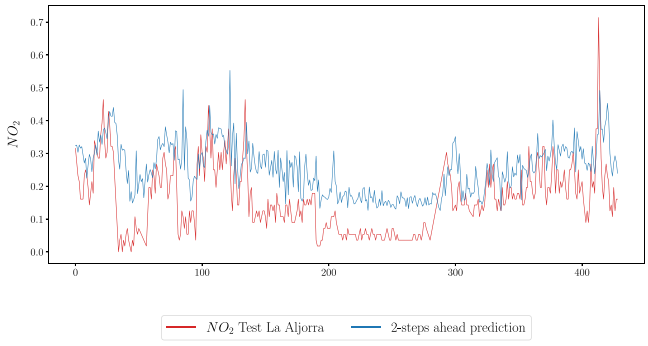
Fig. B.11. Results of the evaluation of the model built with the multi-objective optimization based spatio-temporal approach and QRNN.

forecasting. The goodness metric proposed in this paper has allowed us to compare both approaches. This metric is an aggregation of the RMSE, MAE and CC metrics in 7-steps ahead predictions. Although the interpolation method presents better results in the training set with RF, the results in the test set with the data from La Aljorra station (which have not been seen by the training algorithms) are substantially worse than those obtained with our method. This indicates a better generalization error of our method, which is one of the most important properties in modeling forecasting systems. We have also verified in the experiments that the CC on the test set with the interpolation method is low positive (on average), while with the method proposed in this work the CC is moderately positive, which indicates a greater strength of the relationship between the predicted and observed variables in the model built with our approach.

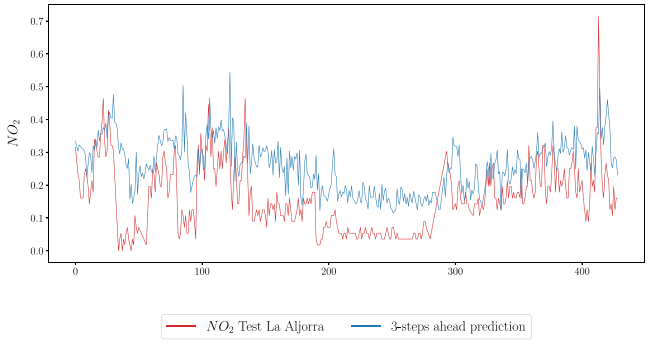
- The main disadvantage of the proposed method compared to the union and interpolation methods is the required runtime. The proposed method runtime is the sum of the time required by the MOEA plus the time required by the ensemble method. The NSGA-II algorithm requires an $O(np^2)$ runtime to do the non-dominated sort, where n is the number of objectives and p is the population size, and a runtime $O(wtn)$ to evaluate each individual, where w is the number of attributes and $t = \max_k\{r_k\}$ is the maximum number of instances in the n datasets. Therefore, the total runtime of NSGA-II is $O(gnp^2 + gwtnp)$, where g is the number of generations of the MOEA. To this runtime we must add the runtime $\mathcal{C}(r, m)$ required to build the ensemble model, where m is the number of LR models found by the MOEA, $r = \sum_{k=1}^{n} r_k$ is the sum of the instances of the n datasets, and $\mathcal{C}(x, y)$ is the asymptotic function of the runtime required by a given regression algorithm to build a model from



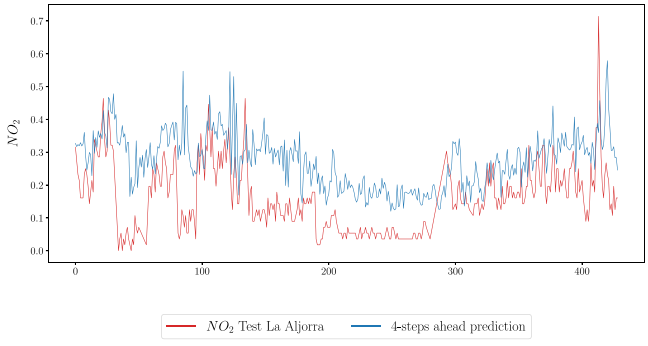
(a) 1-step ahead



(b) 2-steps ahead

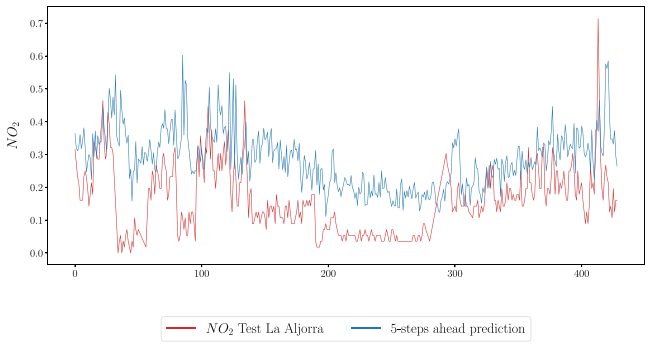


(c) 3-steps ahead

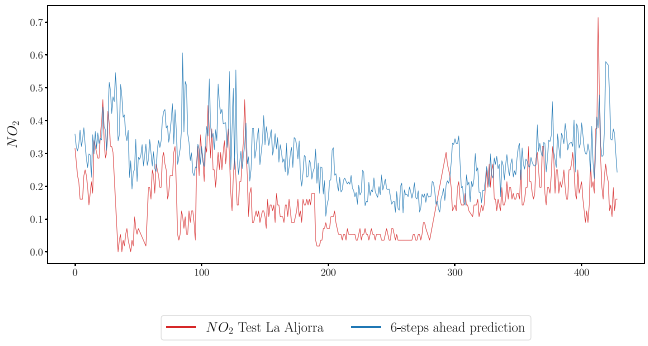


(d) 4-steps ahead

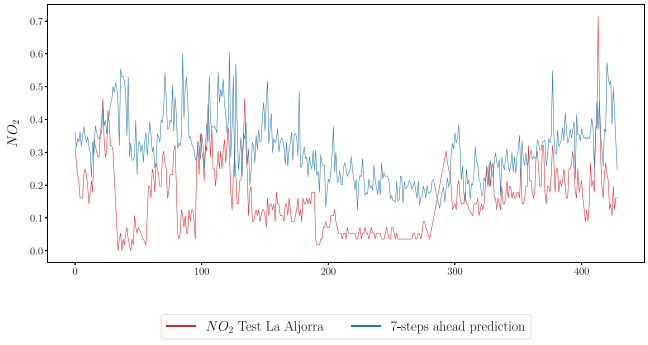
Fig. B.12. Times series of the 7-steps ahead predictions for NO_2 of the model built with the multi-objective optimization based spatio-temporal approach and RF.



(e) 5-steps ahead



(f) 6-steps ahead



(g) 7-steps ahead

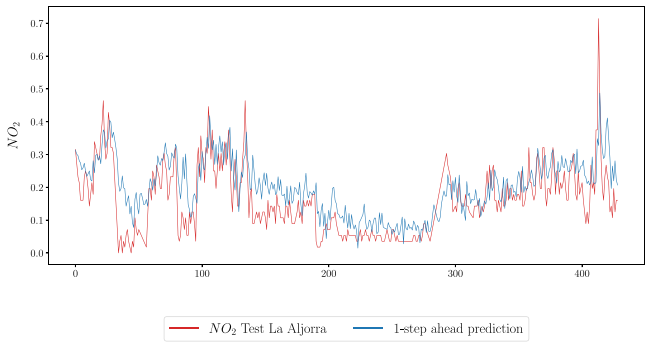
Fig. B.12. (*continued*).

a dataset with x instances and y attributes. In summary, the runtime required by the proposed approach, assuming that we use NSGA-II, is $O(gnp^2 + gwtnp + \mathcal{C}(r, m))$. On the other hand, the runtime required by the dataset union approach is $O(\mathcal{C}(r, w))$, and the time required by the interpolation approach is $O(n\mathcal{C}(t, w))$. Obviously, when $m \geq w$ the time of the proposed approach is always higher than the time required by the union dataset approach. In the opposite case ($m < w$), the time the proposed approach will be less or greater than the other two approaches depending on the number of generations g evolved by the MOEA and the number of monitoring stations n (objectives of the problem). However, the runtime is not a big drawback because the training phase is often an off-line process, and current advances in high-performance computing can greatly alleviate this drawback. In the case of online training processes, the use of incremental evolutionary algorithms [63] is convenient to carry out the proposed approach.

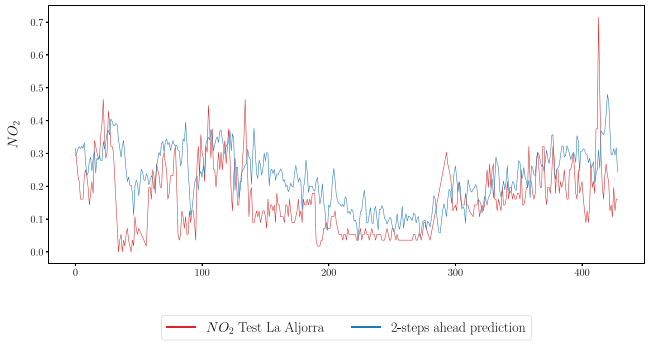
6. Threat to validity

Our analysis and its discussion may suffer from some threat to validity which we have tried to mitigate:

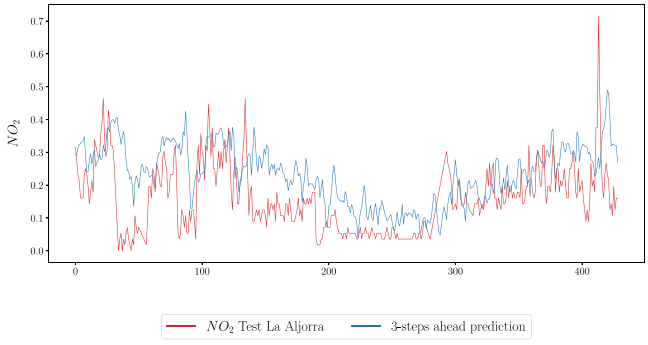
- *External validity* refers to the extent to which the results are generalizable. First, the regional government has a protocol



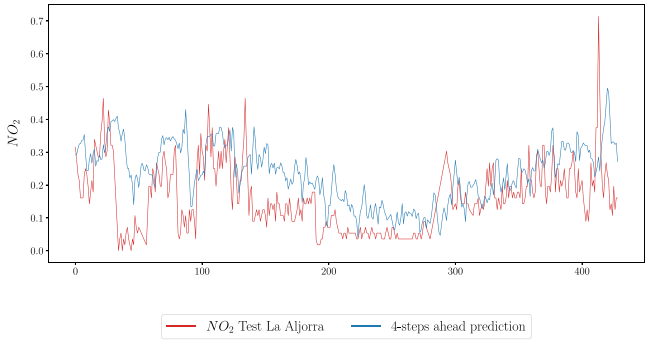
(a) 1-step ahead



(b) 2-steps ahead

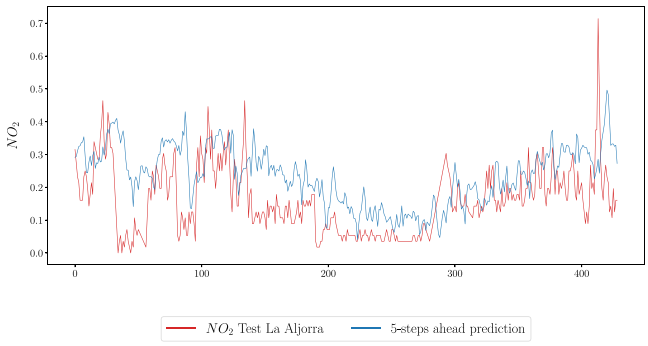


(c) 3-steps ahead

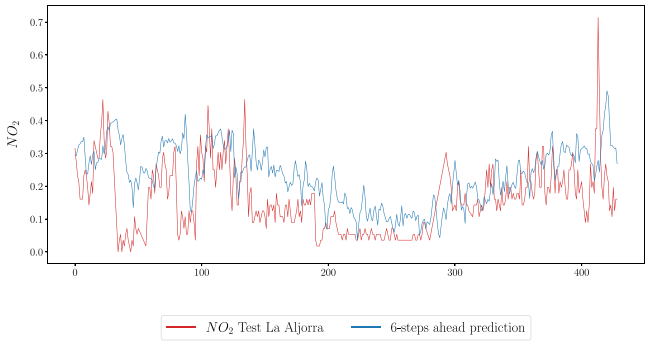


(d) 4-steps ahead

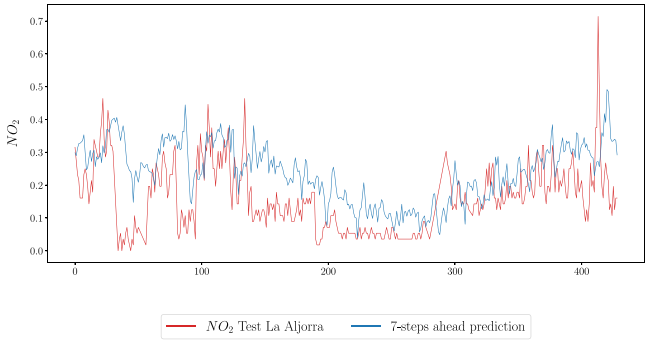
Fig. B.13. Times series of the 7-steps ahead predictions for NO_2 of the model built with the multi-objective optimization based spatio-temporal approach and LR.



(e) 5-steps ahead



(f) 6-steps ahead



(g) 7-steps ahead

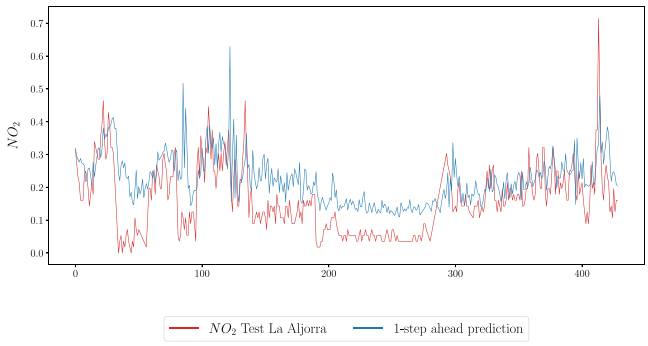
Fig. B.13. (*continued*).

whereby data must undergo quality control before being made public. Secondly, we have tried to make the data used as representative as possible. In terms of the temporal dimension, the temporal window selected is wide enough to capture the seasonality component, as well as the variability inherent in different and changing climatic conditions. Regarding the spatial dimension, the selected monitoring stations represent different urban and rural environments.

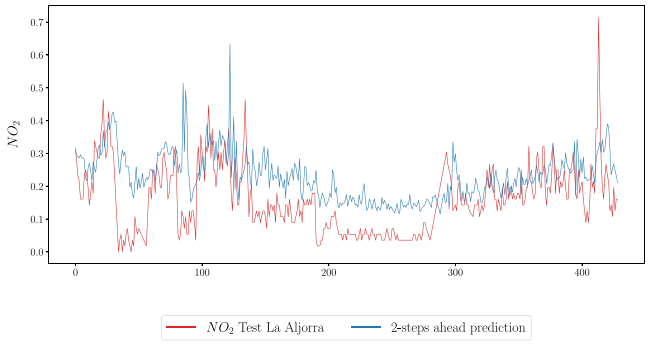
- *Internal Validity* is related to the correctness of the experimental design. To this end, the methodology used for model construction and selection is based on our previous work [64,65], validated after a rigorous peer-review process. On the other hand, special attention has been paid to ensure the reproducibility of the experiments. In terms of coding, to reduce threats due to software failures, platforms sufficiently tested by the scientific community have been used. The Platypus platform was used for the multi-objective evolutionary optimization techniques. ML models have been built using Scikit-learn. This has ensured the correct implementation of the different ML techniques.

7. Conclusions and future work

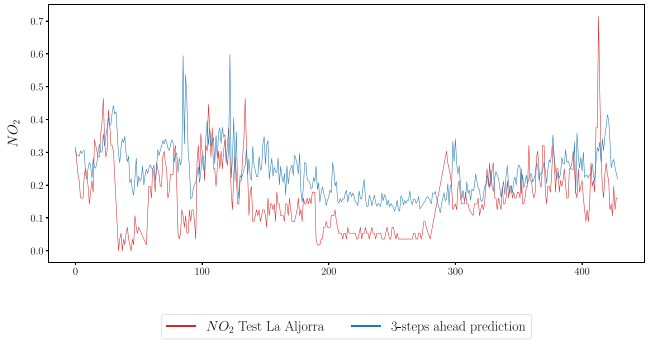
Modeling is very useful for air pollution forecasting. Likewise, in cases where the information and warning thresholds are



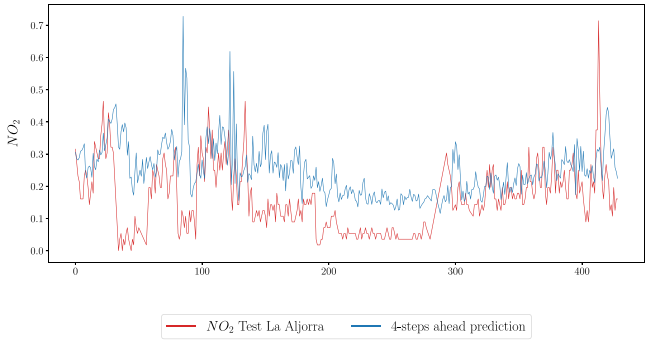
(a) 1-step ahead



(b) 2-steps ahead

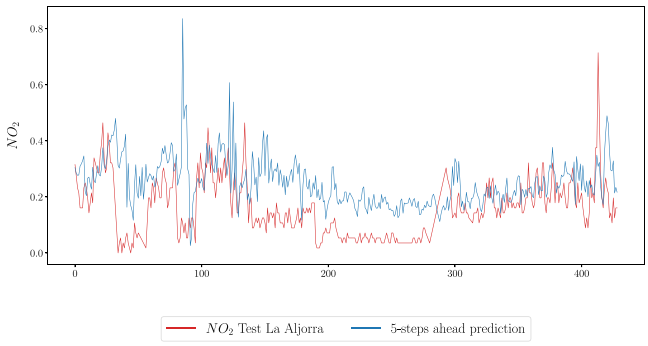


(c) 3-steps ahead

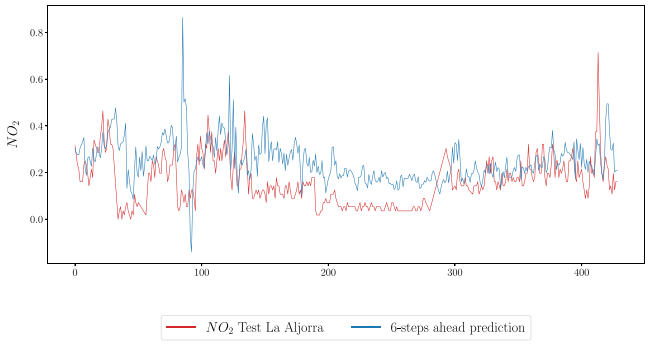


(d) 4-steps ahead

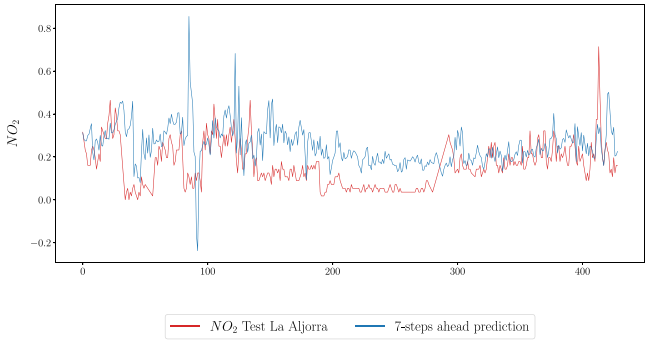
Fig. B.14. Times series of the 7-steps ahead predictions for NO_2 of the model built with the multi-objective optimization based spatio-temporal approach and SVM.



(e) 5-steps ahead



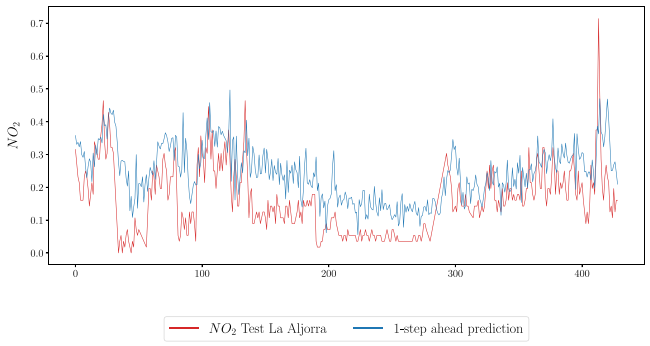
(f) 6-steps ahead



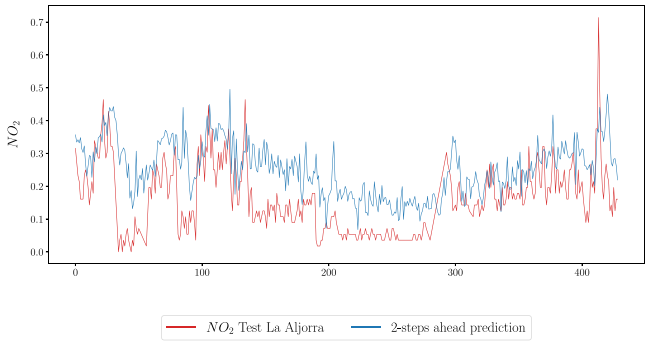
(g) 7-steps ahead

Fig. B.14. (*continued*).

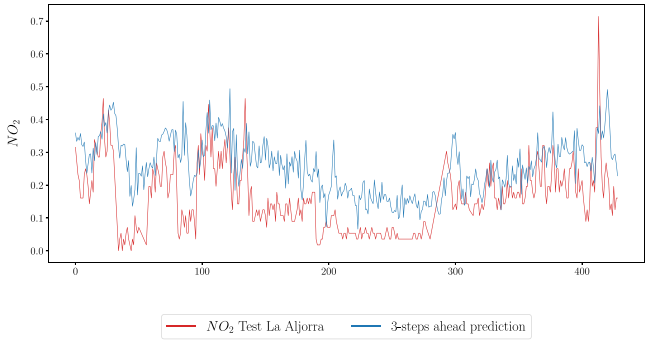
exceeded, modeling is a tool that makes it possible to report on the expected levels and take measures in the application of short-term action plans (pollution protocols). Air pollution depends on both temporal changes and spatial relationships, which makes its prediction a challenge in current research. In this work, we have proposed a method based on multi-objective optimization for space-temporal forecasting of air pollution. The proposed approach builds an ensemble learning model from multiple non-dominated linear models that form the Pareto front of an optimization problem that contains as many objectives as monitoring stations. The multi-objective optimization problem has been solved in this paper with a MOEA, specifically with NSGA-II. We conducted experiments on four real-world air pollution datasets extracted from monitoring stations located in the south of Spain, three of which have been used for training and one for testing. The experiments carried out demonstrate the effectiveness of the proposed method to find short-term spatio-temporal forecast models. Specifically, it has been configured for predictions in the next seven days. The results have been compared with other space-temporal forecasting methods proposed in the literature, in particular, the interpolation method with IDW function. The proposed method has shown better generalization error and better mean scores using the RMSE, MAE and CC performance metrics in 7-step-ahead predictions. This is due to the fact that the number of prediction models is significantly higher in the



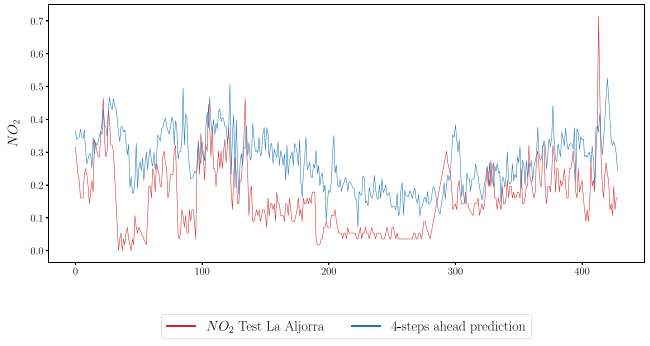
(a) 1-step ahead



(b) 2-steps ahead

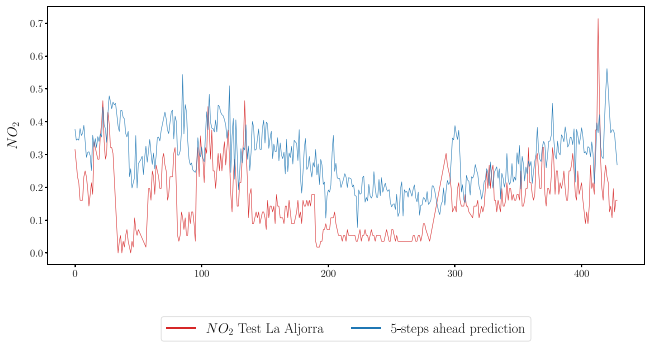


(c) 3-steps ahead

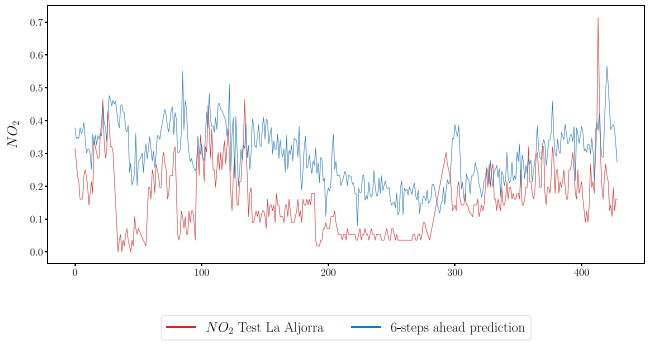


(d) 4-steps ahead

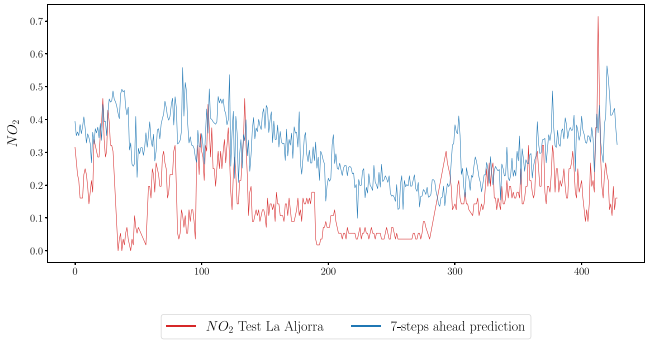
Fig. B.15. Times series of the 7-steps ahead predictions for NO_2 of the model built with the multi-objective optimization based spatio-temporal approach and QRNN.



(e) 5-steps ahead



(f) 6-steps ahead



(g) 7-steps ahead

Fig. B.15. (*continued*).

proposed approach than in the interpolation-based approach, allowing a greater prediction capacity in the new observation areas, as shown by the results of the study. However, the superiority of the proposed method is at the expense of a longer runtime, fundamentally due to the runtime required by the MOEA for an acceptable convergence of the Pareto front.

It is our intention in the near future to integrate feature selection into the multi-objective optimization based spatio-temporal approach proposed in this paper, taking advantage of our previous experience in feature selection based time series forecasting [64,66]. For this, a new objective is added to the optimization problem with which the cardinality of the attribute subsets in the LR models is minimized, thus eliminating redundant and inconsistent attributes and improving the predictive capacity of the models. On the other hand, other types of prediction models could be considered, using a more complex representation for the individuals of the evolutionary algorithm, such as deep neural network architectures, although this could involve excessive computational training time to obtain acceptable results. Anyway, the possibility of using other types of prediction models in the MOEA is being analyzed for future work. Finally, the scalability of the proposed approach in terms of the number of objectives (adding monitoring stations) and the number of attributes (increasing the window size) will be analyzed in future works.

CRediT authorship contribution statement

Raquel Espinosa: Conception and design of the study, Acquisition of data, Analysis and interpretation of data, Point by point revision of reviewer's comments. **Fernando Jiménez:** Drafting the article or revising it critically for important intellectual content, Final approval of the version to be submitted, Point by point revision of reviewer's comments. **José Palma:** Drafting the article or revising it critically for important intellectual content, Final approval of the version to be submitted; point by point revision of reviewer's comments.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially funded by the SITSUS project (Ref: RTI2018-094832-B-I00), given by the Spanish Ministry of Science, Innovation and Universities (MCIU), the Spanish Agency for Research (AEI) and by the European Fund for Regional Development (FEDER).

Appendix A. Abbreviations

See Table A.22.

Appendix B. Performance of prediction models

See Figs. B.8–B.15.

References

[1] C. Guerreiro, A.G. Ortiz, F. de Leeuw, M. Viana, J. Horálek, Air Quality in Europe-2016 Report, Publications Office of the European Union, 2016.
[2] H.L. Brumberg, C.J. Karr, A. Bole, S. Ahdoot, S.J. Balk, A.S. Bernstein, L.G. Byron, P.J. Landrigan, S.M. Marcus, A.L. Nerlinger, et al., Ambient air pollution: Health hazards to children, Pediatrics 147 (6) (2021).
[3] M. Bentayeb, M. Simoni, D. Norback, S. Baldacci, S. Maio, G. Viegi, I. Annesi-Maesano, Indoor air pollution and respiratory health in the elderly, J. Environ. Sci. Health A 48 (14) (2013) 1783–1789.
[4] J. Lelieveld, K. Klingmüller, A. Pozzer, R.T. Burnett, A. Haines, V. Ramanathan, Effects of fossil fuel and total anthropogenic emission removal on public health and climate, Proc. Natl. Acad. Sci. 116 (15) (2019) 7192–7197.
[5] WHO | Ambient (outdoor) air pollution, 2021, URL https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health/.
[6] R. Flagan, J. Seinfeld, Fundamentals of Air Pollution Engineering, in: Dover Civil and Mechanical Engineering, Dover, 2012.
[7] N. Matloff, Statistical Regression and Classification: from Linear Models to Machine Learning, in: Chapman & Hall/CRC Texts in Statistical Science, CRC Press, 2017.
[8] K. Deb, Multi-Objective Optimization using Evolutionary Algorithms, John Wiley & Sons, Inc., 2001.
[9] W. Gong, Z. Cai, L. Zhu, An efficient multiobjective differential evolution algorithm for engineering design, Struct. Multidiscip. Optim. 38 (2009) 137–157.
[10] K. Deb, T. Goel, Multi-objective evolutionary algorithms for engineering shape design, in: Evolutionary Optimization, Springer, 2003, pp. 147–175.
[11] A. Arias-Montano, C. Coello, E. Mezura-Montes, Multiobjective evolutionary algorithms in aeronautical and aerospace engineering, Evol. Comput. IEEE Trans. on 16 (2012) 662–694.
[12] M. Judy, K. Ravichandran, K. Murugesan, A multi-objective evolutionary algorithm for protein structure prediction with immune operators, Comput. Methods Biomech. Biomed. Eng. 12 (4) (2009) 407–413.
[13] G. Hernandez-Rodriguez, F. Morales-Mendoza, L. Pibouleau, C. Azzaro-Pantel, S. Domenech, A. Ouattara, Multi-objective genetic algorithms for chemical engineering applications, in: Applications of Metaheuristics in Process Engineering, Springer, 2014, pp. 343–371.
[14] R. Stewart, T. Palmer, B. DuPont, A survey of multi-objective optimization methods and their applications for nuclear scientists and engineers, Prog. Nucl. Energy 138 (2021) 103830.
[15] M. Abido, Multiobjective evolutionary algorithms for electric power dispatch problem, IEEE Trans. Evol. Comput. 10 (3) (2006) 315–329.
[16] M.-D. Shieh, Y. Li, C.-C. Yang, Comparison of multi-objective evolutionary algorithms in hybrid Kansei engineering system for product form design, Adv. Eng. Inf. 36 (2018) 31–42.
[17] D. Dasgupta, Z. Michalewicz, Evolutionary Algorithms in Engineering Applications, Springer Publishing Company, Incorporated, 2014.
[18] P.R. Bhardwaj, D. Pruthi, Evolutionary techniques for optimizing air quality model, Procedia Comput. Sci. 167 (2020) 1872–1879.
[19] W. Lu, H. Fan, S. Lo, Application of evolutionary neural network method in predicting pollutant levels in downtown area of Hong Kong, Neurocomputing 51 (2003) 387–400.
[20] M. Castelli, I. Gonąlves, L. Trujillo, A. Popovič, An evolutionary system for ozone concentration forecasting, Inf. Syst. Front. 19 (2017).
[21] Y. Collette, P. Siarry, Multiobjective Optimization: Principles and Case Studies, Springer Berlin Heidelberg, 2004.
[22] M. Ehrgott, Multicriteria Optimization, in: Lecture notes in economics and mathematical systems, Springer, 2005.
[23] D. Seng, Q. Zhang, X. Zhang, G. Chen, X. Chen, Spatiotemporal prediction of air quality based on LSTM neural network, Alex. Eng. J. 60 (2) (2021) 2021–2032.
[24] E.H.-C. Lu, C.-Y. Liu, A spatial-temporal approach for air quality forecast in urban areas, Appl. Sci. 11 (11) (2021).
[25] T.-C. Bui, J. Kim, T. Kang, D. Lee, J. Choi, I. Yang, K. Jung, S.K. Cha, STAR: SPatio-temporal prediction of air quality using a multimodal approach, 2020.
[26] Y. Huang, J.J.-C. Ying, V.S. Tseng, Spatio-attention embedded recurrent neural network for air quality prediction, Knowl.-Based Syst. 233 (2021) 107416.
[27] V. Le, T. Bui, S. Cha, Spatiotemporal deep learning model for citywide air pollution interpolation and prediction, 2019, CoRR abs/1911.12919, arXiv:1911.12919.
[28] X. Zou, J. Zhao, D. Zhao, B. Sun, Y. He, S. Fuentes, X. Yang, Air quality prediction based on a spatiotemporal attention mechanism, Mob. Inf. Syst. 2021 (2021).
[29] K. Zhang, X. Zhang, H. Song, H. Pan, B. Wang, Air quality prediction model based on spatiotemporal data analysis and metalearning, Wirel. Commun. Mob. Comput. 2021 (2021) 9627776:1–9627776:11.
[30] Y. Lin, N. Mago, Y. Gao, Y. Li, Y.-Y. Chiang, C. Shahabi, J.L. Ambite, Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning, in: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, in: SIGSPATIAL, vol. 18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 359–368.
[31] S. Chae, J. Shin, S. Kwon, S. Lee, S. Kang, D. Lee, PM10 And PM2.5 real-time prediction models using an interpolated convolutional neural network, Sci. Rep. 11 (1) (2021).
[32] K.K.R. Samal, K.S. Babu, S.K. Das, Spatio-temporal prediction of air quality using distance based interpolation and deep learning techniques, EAI Endorsed Trans. Smart Cities 5 (14) (2021).
[33] V.-D. Le, Spatiotemporal graph convolutional recurrent neural network model for citywide air pollution forecasting, 2021, TechRxiv.
[34] M. Saez, M.A. Barceló, Spatial prediction of air pollution levels using a hierarchical Bayesian spatiotemporal model in catalonia, Spain, 2021, MedRxiv.
[35] L. Zhang, D. Li, Q. Guo, J. Pan, Deep spatio-temporal learning model for air quality forecasting, Int. J. Comput. Commun. Control 16 (2) (2021).
[36] G. Zhao, G. Huang, H. He, H. He, J. Ren, Regional spatiotemporal collaborative prediction model for air quality, IEEE Access 7 (2019) 134903–134919.
[37] S. Deb, R.S. Tsay, Spatio-temporal models with space-time interaction and their applications to air pollution data, Statist. Sinica (2019).
[38] C.-Y. Lin, Y.-S. Chang, S. Abimannan, Ensemble multifeatured deep learning models for air quality forecasting, Atmospheric Pollut. Res. 12 (5) (2021) 101045.
[39] La calidad del aire en la Región de Murcia durante 2021, URL https://www.ecologistasenaccion.org/wp-content/uploads/2022/01/Informe-calidad-aire-region-murciana-2021.pdf.
[40] J. Moreira, C. Soares, A. Jorge, J. Sousa, Ensemble approaches for regression: A survey, ACM Comput. Surv. 45 (2012) 10:1–10:40.
[41] L. Breiman, Bagging predictors, Mach. Learn. 24 (2004) 123–140.
[42] Y. Freund, R.E. Schapire, Experiments with a new boosting algorithm, in: In Proceedings of the Thirteenth International Conference on Machine Learning, Morgan Kaufmann, 1996, pp. 148–156.
[43] L. Breiman, Stacked regressions, Mach. Learn. 24 (2004) 49–64.

[44] G. Bontempi, S.B. Taieb, Y. Borgne, Machine learning strategies for time series forecasting, in: EBISS, 2013, pp. 62–77.
[45] K. Deb, R.B. Agrawal, Simulated binary crossover for continuous search space, Complex Systems 9 (1995).
[46] K. Deb, M. Goyal, A combined genetic adaptive search (GeneAS) for engineering design, Comput. Sci. Inf. 26 (1996) 30–45.
[47] K. Deb, A. Pratab, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.
[48] Q. Zhang, H. Li, MOEA/D: A multiobjective evolutionary algorithm based on decomposition, IEEE Trans. Evol. Comput. 11 (6) (2007) 712–731.
[49] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: IMproving the strength Pareto evolutionary algorithm, TIK-Rep. 103 (2001).
[50] J. Brownlee, Time series forecasting as supervised learning, 2020, Machine Learning Mastery, URL https://machinelearningmastery.com/time-series-forecasting-supervised-learning/.
[51] E. Zitzler, L. Thiele, M. Laumanns, C. Fonseca, V. Grunert da Fonseca, Performance assessment of multiobjective optimizers: An analysis and review, IEEE Trans. Evol. Comput. 7 (2002) 117–132.
[52] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[53] S. Shevade, S. Keerthi, C. Bhattacharyya, R.M. Karuturi, Improvements to SMO algorithm for SVM regression, Neural Netw. IEEE Trans. on 11 (2000) 1188–1193.
[54] R. Vang-Mata, Multilayer Perceptrons : Theory and Applications, Computer Science, Technology and Applications, Nova Science Publishers, New York, 2020.
[55] D. Aha, D. Kibler, Instance-based learning algorithms, Mach. Learn. 6 (1991) 37–66.
[56] J. Bradbury, S. Merity, C. Xiong, R. Socher, Quasi-recurrent neural networks, 2016, arXiv preprint arXiv:1611.01576.
[57] How to estimate a baseline performance for your machine learning models in weka, URL https://machinelearningmastery.com/estimate-baseline-performance-machine-learning-models-weka/.
[58] J. Li, A Review of Spatial Interpolation Methods for Environmental Scientists, Geoscience Australia, Record 2008/23, 2008, p. 137.
[59] G. Van Brummelen, Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry, Princeton University Press, 2013.
[60] K. Deb, H. Jain, An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints, IEEE Trans. Evol. Comput. 18 (4) (2014) 577–601.
[61] E. Zitzler, S. Künzli, Indicator-based selection in multiobjective search, in: Conference on Parallel Problem Solving from Nature, PPSN VIII, 2004 pp. 832–842.
[62] Y. Sun, G.G. Yen, Z. Yi, IGD Indicator-based evolutionary algorithm for many-objective optimization problems, IEEE Trans. Evol. Comput. 23 (2) (2019) 173–187.
[63] N. Mansour, M. Awad, k. Fakih, Incremental genetic algorithm, Int. Arab J. Inf. Technol. 3 (2006) 42–47.
[64] F. Jiménez, J. Palma, G. Sánchez, D. Marín, M. Francisco Palacios, M. Lucía López, Feature selection based multivariate time series forecasting: An application to antibiotic resistance outbreaks prediction, Artif. Intell. Med. 104 (2020) 101818.
[65] R. Espinosa, J. Palma, F. Jiménez, J. Kaminska, G. Sciavicco, E. Lucena-Sánchez, A time series forecasting based multi-criteria methodology for air quality prediction, Appl. Soft Comput. 113 (2021) 107850.
[66] A. González-Vidal, F. Jiménez, A.F. Gómez-Skarmeta, A methodology for energy multivariate time series forecasting in smart buildings based on feature selection, Energy Build. 196 (2019) 71–82.

Raquel Espinosa received the B.S. degree from University of Murcia, Murcia, Spain, in 2019 and the M.S. degree from University of Murcia, Spain, in 2020. She is currently working toward the Ph.D. degree with the Department of Information Engineering and Communications, University of Murcia, Spain. Her research interests include Big Data, Deep Learning, Machine Learning, Evolutionary Computation and Data Mining.

Fernando Jiménez received an M.D. in Computer Science in 1991 from the University of Granada. He obtained a Ph.D. in 1996 from the University of Murcia for his research on Evolutionary Computation applied to Fuzzy Transportation Problems. He is Full Professor in the Department of Communications and Information Engineering at the University of Murcia. He has more than 100 publications in prestigious international journals, conferences and book chapters. His research interests include Evolutionary Computation, Multi-objective Constrained Optimization, Soft Computing, Evolutionary Fuzzy Systems, Data Mining, Big Data and Deep Learning.

José Palma received a B.S. degree from the University of Las Palmas de Gran Canaria, Spain, in 1990 and a Ph.D. degree from the University of Murcia, Spain, in 1999, both on Computer Science. He is Associate Professor of Computer Science in the Department of Information Engineering and Communications and the Computer Science School at the University of Murcia since 2000, but were being teaching in this department as Aggregate Professor since 1996. Prior to joining University of Murcia, he worked for 6 years in the Department of Computer Science and System at the University of Las Palmas de Gran Canaria. He has authored/co-authored various journal articles, book chapters and congress papers. His research activity is focused on medical informatics, specifically in intelligent data analysis, medical knowledge-based systems, clinical knowledge management and ambient intelligent applications. The main techniques used in this research are machine learning, fuzzy-logic, knowledge-engineering methodologies, ontologies and temporal reasoning.