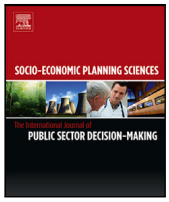
Socio-Economic Planning Sciences 95 (2024) 101958

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

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



Factor-Augmented Autoregressive Neural Network to forecast NOx in the city of Madrid

Gema Fernández-Avilés [a], Raffaele Mattera [b,*], Germana Scepi [c]

[a] Department of Applied Economics, University of Castilla-La Mancha, Toledo, Spain
[b] Department of Social and Economic Sciences, Sapienza University of Rome, Rome, Italy
[c] Department of Economics and Statistics, University of Naples "Federico II", Naples, Italy

ARTICLE INFO

Keywords:
Factor model
Spatio-temporal data
Environmental forecasting
Neural networks
Principal Component

ABSTRACT

Air pollution poses a significant threat to public health and the environment in urban areas worldwide. In the context of urban air quality, nitrogen oxides (NOx), comprising nitrogen dioxide (NO2) and nitric oxide (NO), stand out as key pollutants with well-documented adverse effects. The city of Madrid, as the capital and largest urban center of Spain and the third largest of Europe, is no exception to the challenges posed by NOx pollution. Most of the recent literature on forecasting air pollution, and specifically on NOx, is based on the use of Neural Networks (NN). Little is known about the forecasting ability of factor models in this context. The main aim of this paper is to use Factor-Augmented Autoregressive Neural Networks (FA-ARNN-X) to predict future patterns of NOx pollutants in the territorial monitoring stations of Madrid, using lagged NOx values, meteorological variables and latent factors. The main results indicate that the proposed forecasting model provides statistically more accurate predictions of air pollution than its competing benchmarks and should be used by policymakers for more accurate air pollution monitoring.

1. Introduction

Air pollution is recognized as the foremost environmental hazard affecting human health [1,2]. Prolonged exposure to air pollutants, such as particulate matter and noxious gases, has been linked to a wide range of adverse health effects, including respiratory diseases, cardiovascular problems, and even premature death [3,4]. For example, a 2016 report of the European Monitoring and Evaluation Programme (EMEP) model shows that 412,000 premature adult deaths were attributable to long-term exposure to particulate matter concentrations (in particular to PM2.5), while the estimated impact of exposure to nitrogen dioxide (NOx) was around 71,000 premature deaths per year [5]. Consequently, air pollution monitoring has become an urgent priority in public health policy [6,7].

Aside from the impact on health, however, poor air quality also has a socio-economic impact [8,9]; for example, in terms of reduced labor productivity [10] and tourist flows [11]. The Organization for Economic Cooperation and Development (OECD) estimates that the overall costs associated with increasing air pollution will be about 2% of European GDP in 2060, thus leading to slower economic growth [12]. That said, it should be noted that the relationship between socio-economic development and air pollution is complex and multifaceted. Among other factors, the expansion of transportation systems – particularly those modes of transport heavily reliant on fossil fuels, such as cars, trucks, and airplanes – is a major source of greenhouse gas emissions [13]. In this sense, economic growth, if driven by energy-intensive industries and increased transportation activity, can exacerbate greenhouse gas emissions, further intensifying environmental concerns [14].

Territorial policies, whether enacted at the municipal, provincial, or regional level, have a significant role to play in addressing air pollution, while considering the dynamics of economic development at the relevant territorial scale. Indeed, territorial authorities usually wield substantial influence over urban planning and land use decisions. For example, by advocating for efficient public transportation systems and the creation of green spaces, they can contribute directly to improved air quality and enhanced living conditions for residents. Moreover, territorial authorities have a deeper understanding of the specific air quality challenges within their territorial units. This allows them to implement policies that focus on emissions reduction strategies tailored to the unique circumstances of their territory.

An interesting example about the relevance of local policies is provided by the city of Madrid, which is one of the areas in Spain most severely affected by bad air quality [15]. In 2019, NOx concentrations

* Corresponding author.
E-mail address: raffaele.mattera@uniroma1.it (R. Mattera).

https://doi.org/10.1016/j.seps.2024.101958
Received 16 January 2024; Received in revised form 29 April 2024; Accepted 28 May 2024
Available online 7 June 2024

exceeded the EU legal limit for the 10th consecutive year. To address this, the city implemented "Plan A", which included creating Madrid Central to reduce emissions and traffic. Under 'Plan A', stringent restrictions were placed on high-emission vehicles, with a particular focus on diesel cars. Only vehicles that met specific emissions standards, such as Euro 6 for diesel and Euro 4 for gasoline engines, were granted access to Madrid Central. Additionally, the plan actively promoted the adoption of sustainable transportation alternatives, such as walking, cycling, and public transit. It also advocated for the expansion of bike lanes and the development of pedestrian-friendly zones.

1.1. Motivation

The primary motivation behind this paper is the critical need to provide local governments with accurate air pollution forecasts to facilitate the timely development of environmental-related public policies. This paper discusses the challenge of predicting NOx pollution, a matter of particular importance in urban areas and major cities such as Madrid [16]. This dangerous pollutant is generated by high-temperature combustion, with transportation being the primary source. Population growth and improved economic conditions further escalate transportation activity, and therefore increase the emission of NOx pollutants. Monitoring NOx levels is important because they have a range of harmful effects on human health and vegetation. Prolonged exposure to high levels of NOx can indeed cause irreversible damage to the respiratory system, and it can also hurt vegetation making it more susceptible to disease and frost damage.

Policy makers are expected to come up with robust policy strategies, plans and programmes, and some level of knowledge about the future is needed for this aim (e.g. see [17,18]). Air pollution forecasts play a pivotal role in shaping public health policy [19,20], particularly at the territorial level [21,22]. In this context, predictive models and monitoring systems enable authorities and individuals to make informed decisions regarding outdoor activities, protective measures, and healthcare planning. Timely alerts and forecasts empower communities to reduce their exposure to harmful pollutants [23], thereby potentially preventing a multitude of health problems. Additionally, proactive efforts to manage and reduce air pollution can yield long-term benefits in terms of improved public health and environmental sustainability [24]. Hence, accurate predictions and effective measures to combat air pollution are pivotal components in safeguarding human health and enhancing overall quality of life.

1.2. Contribution

Air pollution time series are characterized by complex and interrelated patterns, including seasonality, trends, cyclic behavior, heteroscedasticity, non stationarity, non-normal distributions, and spatial dependence [25]. These characteristics require specialized statistical techniques for accurate modeling, and forecasting. Due to the complexity of environmental data, previous studies commonly employ neural networks (NNs) and other deep learning methods to predict air pollution (for extensive and recent reviews, see [26–28]).

An NN is a machine learning model inspired by the structure and function of the human brain [29,30]. It consists of interconnected nodes, or artificial neurons, organized into layers. An NN can be used for forecasting air pollution by training on previous air pollution levels and by learning complex patterns that pollutants have with other meteorological and environmental variables. Once trained, the NN can make predictions about future air pollution levels based on input variables such as weather conditions, emissions data, and historical pollution levels. It is a powerful tool for modeling and predicting air quality, especially when dealing with nonlinear and complex relationships in the data.

It is often the case, however, that not all variables affecting air pollution are measured. Furthermore, there could be some unobservable factors that researchers and policymakers fail to consider in their analyses. Unobservable factors – often referred to as "hidden" or "latent" factors – can significantly impact air pollution levels. Some examples of unobservable factors affecting air pollution include emissions from unreported sources, residential heating practices, weather conditions, chemical reactions, natural pollution sources, human behavior, and long-term trends in urbanization and economic development. These hidden variables can lead to variations in pollution levels and require advanced statistical and machine-learning techniques to account for them and improve air quality forecasting and management.

While previous studies make extensive use of NNs when forecasting air pollution, little is known about the forecasting ability of factor models in this context. In sum, this paper makes the following contributions:

(*i*) To the best of our knowledge, this is the first paper to assess the performance of factor-based approaches in forecasting air pollution;
(*ii*) A new Factor-Augmented Autoregressive Neural Network (FA-ARNN) is proposed, where latent common factors are estimated considering the presence of observable factors (FA-ARNN-X);
(*iii*) Latent factors are estimated taking inspiration from heterogeneous panel data models [31]. This methodology has been mainly applied in economics and finance, but has rarely been used for environmental modeling;
(*iv*) A mode ensemble of different predictions is used to boost the accuracy of the proposed approach.

The rest of the paper is structured as follows. Section 2 describes the FA-ARNN model and the approach adopted for the estimation of the latent factors. Section 3.1 shows the data used for the forecasting exercises related to the Madrid area as well as the details related to the experimental set-up. Section 4 highlights the main results and provides a discussion, while Section 5 concludes with some final remarks and future research directions.

2. Methodology

In what follows, we introduce the methodology adopted for forecasting NOx pollution measured in the territorial monitoring stations (MS) of Madrid, using both observable (e.g. meteorological, environmental, etc.) and unobservable variables. In particular, we rely on NNs to account for the highly nonlinear patterns of the type of data considered in the paper.

In the multilayer feed-forward network, each layer of nodes receives inputs from the previous layers. The outputs of the nodes in one layer are inputs to the next layer and the inputs to each node are combined using a weighted linear combination. The result is then modified by a nonlinear function before being output [29,30].

Let us consider the case of a simple network architecture, where the NOx time series are the output to predict with the network, given a set of meteorological (observable) variables. Considering a single layer network [32], an NN for an nth time series y_{nt} (i.e. NOx concentrations at time t in the nth MS), given P observable inputs x_{pt} (i.e. meteorological variables observed in Madrid), can be written as follows:

$$y_{nt} = \beta_0 + \sum_{h=1}^{H} \beta_h F\left(\gamma_{0h} + \sum_{p=1}^{P} \gamma_{hp} x_{pt}\right) + \varepsilon_{nt}, \tag{1}$$

with $\boldsymbol{w} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ being the network weights with $\boldsymbol{\beta} = \left[\beta_1, \ldots, \beta_H\right]$, $\gamma = \left[\gamma_{11}, \ldots, \gamma_{HP}\right]$ for the output and the hidden layers, respectively. β_0 and γ_{0p} are the biases of each neuron, which for each neuron acts similarly to the intercept in a regression. H is the number of hidden nodes in the network and $F(\cdot)$ is a nonlinear transfer function, which is usually either the sigmoid logistic or the hyperbolic tangent function. A visual example of a network with a single hidden layer and $H = 3$ hidden nodes with $P = 4$ inputs is shown in Fig. 1.

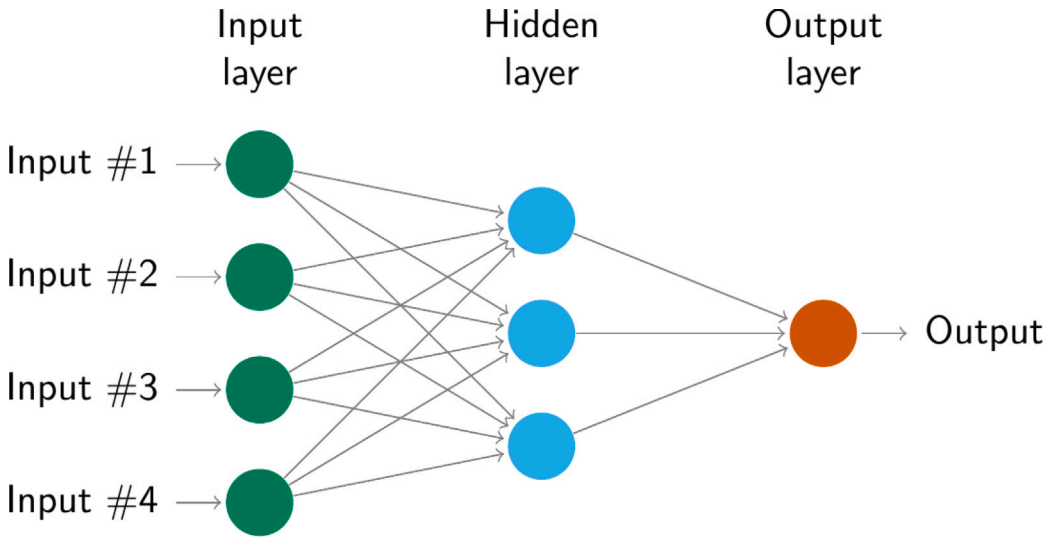


Fig. 1. Example of an NN with 1 hidden layer and $H = 3$ hidden nodes with output y_{nt} and $P = 4$ inputs x_{pt} $(p = 1, \dots, 4)$ [33].

With time series data, lagged values of the time series can be used as inputs to an NN. We call this an Autoregressive Neural Network or ARNN model. We consider feed-forward networks with one hidden layer, and the notation is $ARNN(L, H)$ to indicate there are L $(l = 1, \dots, L)$ lagged inputs (in our case the level of NOx L days before) and H $(h = 1, \dots, H)$ nodes in the single hidden layer; that is,

$$y_{nt} = \beta_0 + \sum_{h=1}^{H} \beta_h F\left(\gamma_{0l} + \sum_{l=1}^{L} \gamma_{hl} y_{nt-l}\right) + \varepsilon_{nt}, \tag{2}$$

with additional inputs given by the L lags of the output series. In practice, an ARNN is a feed-forward neural network which includes the lagged values of the target variable as inputs. However, it is reasonable to assume that other factors affect the patterns of the time series y_{nt} along with its lags. If the goal is to predict N time series and we assume that these factors are latent, it is possible to recover the unobservable factors from the information of the set of N multivariate time series [34].

Following this intuition, Babikir and Mwambi [35] propose augmenting (5) for the presence of K unobservable factors explaining the set of N output time series. In other words, given Y_t, the vector of the N time series of NOx pollutants in the nth MS at time t, Babikir and Mwambi [35] consider the following factorial structure

$$Y_t = \lambda F_t + \zeta_t, \quad \forall t = 1, \dots, T, \tag{3}$$

with F_t being the vector of low-dimensional K latent factors and λ the associated loading. Babikir and Mwambi [35] propose estimating the latent factors F_t using the Principal Component Estimator (PCE), which in our setting means the latent factors are obtained from the set of NOx time series for all the N MS in Madrid. Then, to predict the NOx pollutants in each nth monitoring station with the Babikir and Mwambi [35] approach, we consider the following model:

$$y_{nt} = \beta_0 + \sum_{h=1}^{H} \beta_h F\left(\gamma_{0l} + \sum_{l=1}^{L} \gamma_{hl} y_{nt-l} + \sum_{k=1}^{K} \beta_{hk} f_{kt}\right) + \varepsilon_{nt}, \tag{4}$$

with β_{hk} is the parameter in the hth node associated with the kth factor f_{kt}. The authors call the model (4) a Factor-Augmented Neural Network (FA-NN). In this paper, taking inspiration from heterogeneous panel data literature, we consider a different perspective and we propose a different Factor-Augmented model.

2.1. The Factor-Augmented Autoregressive Neural Network (FA-ARNN)

A natural extension of the pure ARNN consists in augmenting the model (2), where only NOx time series lags are considered, for the presence of P additional observable inputs, such as meteorological variables which can be associated with the levels of the NOx pollutants in the MS. If we denote such observable variables x_{pt}, the ARNN can be written as follows:

$$y_{nt} = \beta_0 + \sum_{h=1}^{H} \beta_h F\left(\gamma_{0l} + \sum_{l=1}^{L} \gamma_{hl} y_{nt-l} + \sum_{p=1}^{P} \gamma_{hp} x_{pt}\right) + \varepsilon_{nt}, \tag{5}$$

In what follows, we argue that the model (5) can even be extended considering the presence of K latent factors along with the L lags and the P observable inputs. In this way, we assume that NOx levels in a given nth monitoring station can be predicted considering its past observations (y_{nt-l}), meteorological variables (x_{pt}) and some latent factors (f_{kt}) given by variables affecting pollutant levels but not measured in the dataset. This leads to the Factor-Augmented Autoregressive Neural Network (FA-ARNN)

$$y_{nt} = \beta_0 + \sum_{h=1}^{H} \beta_h F\left(\gamma_{0l} + \sum_{l=1}^{L} \gamma_{hl} y_{nt-l} + \sum_{p=1}^{P} \gamma_{hp} x_{pt} + \sum_{k=1}^{K} \beta_{hk} f_{kt}\right) + \varepsilon_{nt}. \tag{6}$$

To estimate (6), we need to estimate both the number of factors K and then separately identify loadings and factors.

Given the presence of P meteorological variables affecting the panel of NOx levels Y_t, we cannot consider the PCE approach directly, for example, discarding any information about the existing relationship between the meteorological variables and the NOx levels. This would lead to a severe bias in the estimation of the latent factors. Instead, we consider the following factorial structure:

$$Y = X\gamma + F\Lambda' + Z, \tag{7}$$

where Y is $T \times N$ and X is a three-dimensional matrix with P sheets $(T \times N \times P)$, the pth sheet of which is associated with the pth element of $\gamma (i = 1, 2, \dots, P)$. The product $X\gamma$ is $T \times N$ and Z is $T \times N$. Note that (7) is a common setting adopted by previous studies (e.g. [31,34,36]), and assumes that the panel of N units is explained by a set of observable factors and that the residuals follow a factorial structure. In fact, we can immediately recognize that (7) implies a panel data of N cross-sectional units observed for T times. Therefore, to estimate the latent factor we can use the heterogeneous panel data model of Bai [31].

Identification is not possible without restrictions. The normalizations $F'F/T = I_K$ and $\Lambda'\Lambda$ diagonal are commonly used [34,36] for this purpose. However, sufficient variation in X is also needed in this setting. The usual identification condition for γ is that the matrix $(NT)^{-1} \sum_{n=1}^{N} \mathbf{x}_n' M_F \mathbf{x}_n$ is of full rank, where $M_F = I_T - F\left(F'F\right)^{-1} F'$. Since F is not observable and is estimated, a stronger condition is

required (see [31], for details). These sets of restrictions uniquely determine Λ and F.

Given the presence of K factors, to estimate factors and loadings starting from (7), we consider the following least squares objective function:

$$\mathcal{L}(\gamma, F, \Lambda) = \sum_{n=1}^{N} \left(Y_n - \gamma X_n - F\lambda_n\right)' \left(Y_n - \gamma X_n - F\lambda_n\right), \tag{8}$$

s.t. $F'F/T = I_K, \quad \Lambda'\Lambda = \text{diag}.$

The estimator for γ given F is:

$$\hat{\gamma} = \left(\sum_{n=1}^{N} X_n' M_F X_n\right)^{-1} \sum_{n=1}^{N} X_n' M_F Y_n, \tag{9}$$

so that the residuals $\hat{Z} = Y - \hat{\gamma}X$ have a pure factor structure that can be written as $\hat{Z} = F\Lambda' + U$. At this stage, we can thus use PCE [34] to estimate the common factors in $\hat{Z}$.

Briefly, the estimator for F consists of the K eigenvectors associated with the K largest eigenvalues of the matrix $\hat{Z}\hat{Z}'$ multiplied by $\sqrt{T}$, while the estimator for Λ is:

$$\hat{\Lambda} = \hat{Z}\hat{F}/T. \tag{10}$$

In sum, given F we estimate γ, and given γ we estimate F. This suggests that, unlike Babikir and Mwambi [35], an iterative procedure is needed in our setting. We also note that, when $N \gg T$, we can also consider estimating Λ as the K eigenvectors associated with the K largest eigenvalues of the matrix $\hat{Z}'\hat{Z}$ multiplied by $\sqrt{N}$ and the factor as $\hat{F} = \hat{Z}\hat{\Lambda}/N$.

As a result, to estimate the K factors we adopt the Bai [31] algorithm, which can be summarized as follows:

(*i*) Set the number of factors K and the maximum number of iterations;
(*ii*) For each $n = 1, \ldots, N$ estimate (9) assuming no factors and take the residuals;
(*iii*) Given the matrix of residuals $\hat{Z}$, extract the common factor with PCE (10);
(*iv*) Re-estimate (9) considering the extracted factors and loadings;
(*v*) Repeat steps (*iii*) and (*iv*) until convergence.

Two remaining issues are the estimation of K, which is fixed at the beginning, and the convergence definition. We propose estimating the K factors using the procedures developed in [36,37], while for the convergence we minimize the least square loss, $\mathcal{L}\left(\hat{\gamma}, \hat{F}, \hat{\Lambda}\right)$ iteratively.

2.2. Forecasting with a FA-ARNN model and network ensemble

Once the K latent factors have been estimated, the one-step-ahead forecast for the given model can be obtained as follows:

$$\hat{y}_{nt+1} = \hat{\beta}_0 + \sum_{h=1}^{H} \hat{\beta}_h F\left(\hat{\gamma}_{0l} + \sum_{l=1}^{L} \hat{\gamma}_{hl} y_{nt+1-l} + \sum_{p=1}^{P} \hat{\gamma}_{hp} x_{pt+1} + \sum_{k=1}^{K} \hat{\beta}_{hk} f_{kt+1}\right). \tag{11}$$

Given a network architecture consisting of a single hidden layer, L lags, K factors and H hidden nodes, the set of network parameters is trained in-sample with the goal of making predictions for the out-of-sample dimension.

In general, in a pure ex-ante forecasting framework, we do not know in advance the values of the predictors. The latent factors, in particular, are estimated and reconstructed considering the in-sample observation while the observable factors are obtained by means of measurements. A simple solution to this issue is to assume a random walk (RW) forecast for the factors such that $f_{kt+1} = f_{kt}$. A similar approach is indeed used when the value of the observable factors is not known or when the development of a nested forecasting approach is not suitable, thus $x_{pt+1} = x_{pt}$. In this regard, we note that there is huge empirical literature demonstrating that RW – although simple – is very effective in out-of-sample forecasting of many real-life phenomena (e.g. see [38–40]).

Another issue in forecasting with FA-ARNN lies in the stability of the predictions obtained with different random initializations of the parameters in the network. Each different initialization leads to different optimal solutions, and thus different forecasts. A simple approach to limit this weakness lies in the ensemble of the different networks' results [41].

As shown by many authors (e.g. [32,41–43]), the combination of forecasts resulting from an ensemble of NNs has been shown to outperform the use of a single best-trained network model. Moreover, as shown by Naftaly et al. [41], the use of an ensemble removes the need to identify the best initialization.

For each nth time series (we omit the n subscript for ease of reading), given a set of M different networks' predictions:

$$\hat{Y}_{t+1} = [\hat{y}_{1t+1}, \ldots, \hat{y}_{mt+1}, \ldots, \hat{y}_{Mt+1}]',$$

different ensemble operators can be considered, such as the average ensemble or the median ensemble, where the mean and median of the vector $\hat{Y}_{t+1}$ are taken as the resulting forecast. However, following a more recent approach, we propose the use of a mode ensemble as proposed in [32]. Briefly, the mode ensemble is based on Kernel density estimation, which is a non-parametric approach commonly used to estimate the probability density function of a random variable. Given that the distribution of $\hat{Y}_{t+1}$ is unknown, we can approximate it using the kernel density estimator:

$$\hat{p}_{th}(x) = (Mb)^{-1} \sum_{m=1}^{M} \phi\left(\frac{x - \hat{y}_{mt+1}}{b}\right), \tag{12}$$

where $\hat{y}_{mt+1}$ is the mth observation of the M-dimensional vector of forecasts at $t+1$, which we assume to be i.i.d drawn from univariate distribution with unknown density at any given point x. The function $\phi(\cdot)$, satisfying the property $\int \phi(x)dx = 1$, is called kernel and $b > 0$ is its bandwidth. A commonly chosen kernel is the Gaussian kernel. The mode ensemble is proven to be more effective than the mean and median ensembles and is more robust to outliers [32]. With the mode ensemble, moreover, we eliminate the problem of the best training of the network as well as the local optima issue, as we combine M alternative results obtained by the same network model.

3. Data and forecasting experiment

3.1. Data

This study focuses on NOx as it is a dangerous pollutant that must be kept under control. The problem of air pollution is very serious, especially in urban areas and large cities such as Madrid. Since NOx is emitted as a result of high-temperature combustion, transportation is a major contributor to air pollution (NOx is emitted directly by vehicles, especially diesel vehicles). In turn, the increase in transportation is a consequence of a growing population and better economic conditions [15].

In this paper, we aim to forecast NOx pollutants in the MS in the city of Madrid. To do so, we employ two types of data: (i) the NOx concentration time series data that are available at MS level, and (ii) the meteorological data – which we call environmental observable factors – that are known to be useful for predicting air pollution and are specific to the city of Madrid as a whole.

The NOx concentration time series have been collected from the Atmosphere Pollution Monitoring System of the municipality of Madrid.[1]

[1] We remind the reader of the website www.munimadrid.es.

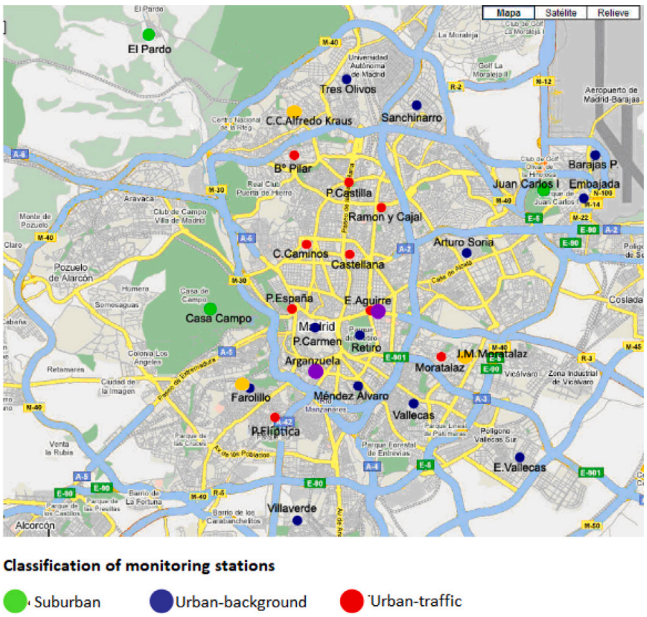


Fig. 2. Location of MS in Madrid (2011–2019).
Source: Own elaboration from www.munimadrid.es.

In particular, we consider the daily average concentrations calculated from the measures registered at 10-min levels for 23 MS in Madrid. Daily averages are used because the time scale is the same as for the observable environmental factors. The meteorological variables are collected from 'climaemet', an R package that serves as an interface for downloading the climate data of the Spanish Meteorological Agency (AEMET) directly from R using their API [44].

According to the European Environmental Agency (EEA), the MS considered are classified as urban-traffic, urban-background, and suburban. In Madrid, the type of MS considered for each level are the following:

- Urban-traffic (UT): Escuelas Aguirre, Avda. Ramón *y* Cajal, Moratalaz, Cuatro Caminos, Barrio del Pilar, Castellana, Plaza Castilla, Plaza Elíptica, Plaza España.
- Urban-background (UB): Arturo Soria, Villaverde, Farolillo, Barajas Pueblo, Plaza del Carmen, Puente de Vallecas, Mendez Alvaro, Retiro, Ensanche de Vallecas, Urbanizacion Embajada, Sanchinarro and Tres Olivos.
- Suburban (S): El Pardo, Juan Carlos I and Casa de Campo.

Fig. 2 is a map showing the location of the analyzed municipal MS.

We consider NOx time series starting from 1st January 2011. Indeed, in 2011 Madrid city's Air Quality Surveillance Network aligned with European Directive 2008/50/EC pertaining to ambient air quality and promoting a cleaner atmosphere in Europe. This alignment involved adjusting station numbers and parameters to address current pollution issues. The analyzed time period ends on 31st December 2019 to exclude data from the COVID-19 pandemic. The restrictions on mobility due to COVID-19 had a significant effect in terms of reducing emissions and improving air quality [45,46]. Fig. 3 shows the time series for the NOx pollutants for all the analyzed MS during the period between 1st January 2011 and 31st December 2019, for a total of $T = 3287$ temporal observations, while Fig. 4 shows the time series distribution.

Overall, Fig. 3 shows that NOx levels in all the MS follow similar patterns. Some stations such as El Pardo, which is located outside the city center, show lower levels than other stations. Unfortunately, some MS exceed the threshold throughout most of the analyzed period.

Table 1
Number of times that the 40 μg/m^3 per day limit has been exceeded by MS (full sample).

Stations	Type	Limit exceedance (frequency)
Escuelas Aguirre	UT	3050
Avda. Ramon y Cajal	UT	2519
Arturo Soria	UB	1992
Villaverde	UB	1986
Farolillo	UB	1749
Casa de Campo	S	895
Barajas Pueblo	UB	1901
Plaza del Carmen	UB	2613
Moratalaz	UT	2124
Cuatro Caminos	UT	2594
Barrio del Pilar	UT	2152
Puente de Vallecas	UB	1963
Mendez Alvaro	UB	1827
Castellana	UT	2101
Retiro	UB	1269
Plaza de Castilla	UT	2663
Ensanche de Vallecas	UB	1840
Urbanizacion Embajada	UB	2277
Plaza Elíptica	UT	3028
Sanchinarro	UB	1562
El Pardo	S	500
Juan Carlos I	S	1095
Tres Olivos	S	1407

Table 1 provides a more accurate picture of the differences registered in the analyzed MS, showing how many times the limit has been exceeded in each station.

The capital has failed to comply with the Annual Limit Value (ALV) for NOx from 2010 to 2019. The European Union sets the maximum allowed for the annual average at 40 μg/m^3 (Directive 2008/50/CE and RD 102/2011). In 2016, the limit was surpassed in nine stations; in 2017, in 15 stations; in 2018, in seven; and in 2019 in two.

It comes as no surprise that MS far from the city center (Suburban) show the lowest frequency of exceedance, e.g. El Pardo and Casa de Campo, while those placed in Urban Traffic are associated with the highest frequency. Forecasting NOx values for these stations, therefore, represents an important problem from an environmental policy perspective. Fig. 4 highlights the differences among the analyzed MS in terms of NOx time series distribution, through violin plots.

From the distributions, we again see evidence that different MS types give rise to different distributions of the NOx patterns. Suburban stations, for example, show on average less variability in their distribution compared to other types. At the same time, however, Urban Traffic stations show the largest variability and peaks. For example, El Pardo shows the one with lowest distribution variability distribution, while Plaza de Castilla is the station with the largest variation in the sample.

As meteorological observable factors, we consider the average, maximum and minimum temperature; daily sunshine hours; average wind speed; wind gusts; and rainfall. Fig. 5 shows the temporal pattern of the temperature data and the distribution of the other meteorological variables. Note that the values of these variables refer to the entire Madrid area, so the observable factors are common to all MS.

It should be noted, however, that the all the time series are characterized by a strongly deterministic seasonal component. Therefore, we consider seasonally adjusted time series for the rest of the analysis and forecasting exercise, as it is more appropriate for forecasting [47]. Fig. 6 shows the correlation between the environmental observable factors and the NOx time series, taken as the average NOx values across the MS. The selected factors are highly (either negatively or positively) correlated with the average NOx values observed in the Madrid area, and all the correlation coefficients are statistically significant at the 95% level, suggesting that all of them can be considered ex-ante as useful predictors for NOx pollutants.

The main descriptive statistics associated with the NOx time series for all the MS and the descriptive statistics for the observable factors are shown in Table 2.

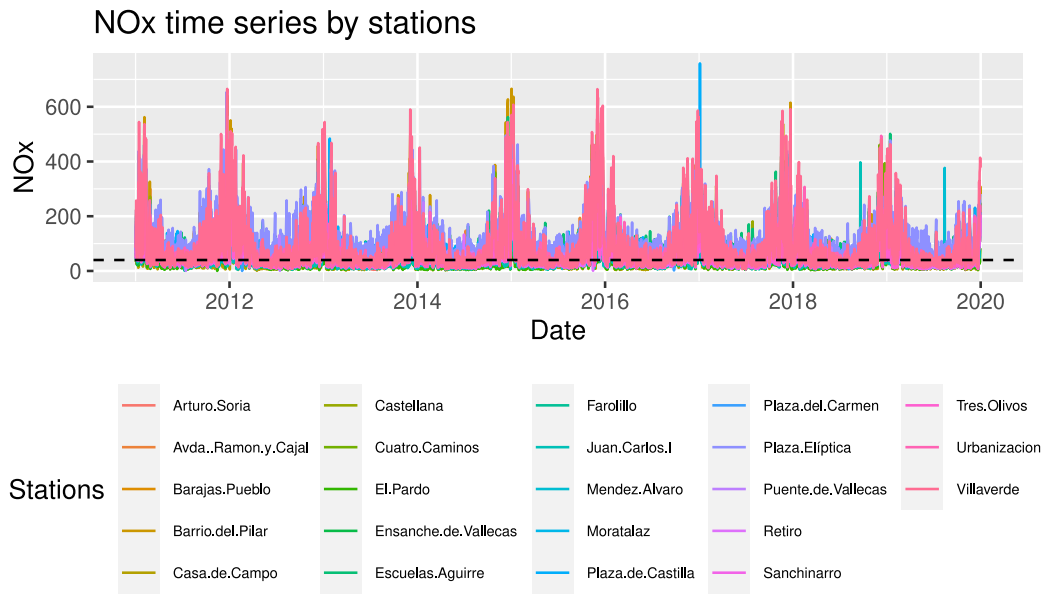


Fig. 3. NOx time series for the $n = 23$ territorial MS in Madrid.

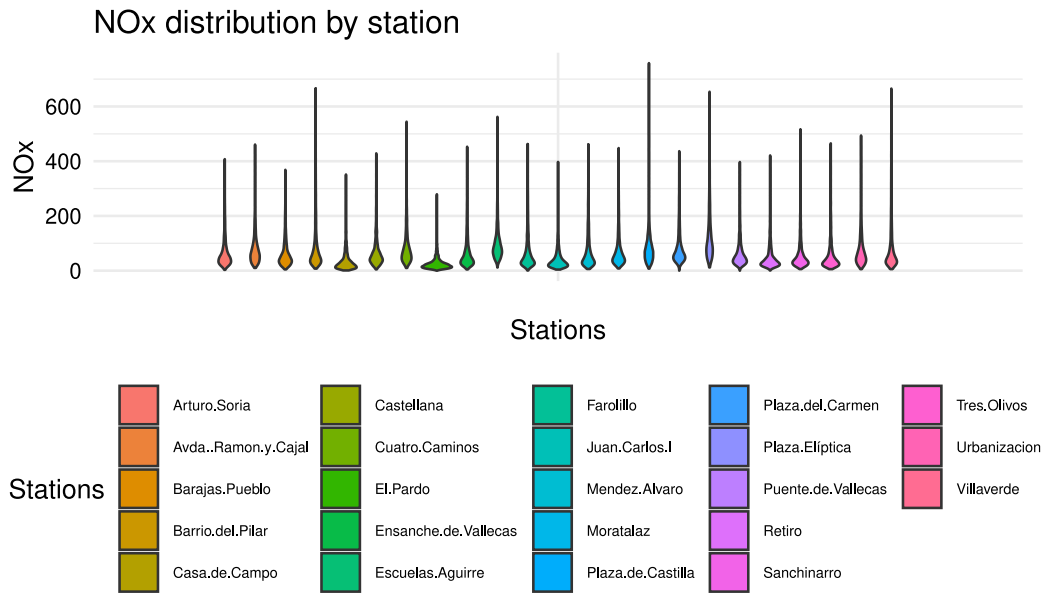


Fig. 4. NOx time series for the $n = 23$ territorial MS in Madrid.

3.2. Experimental set-up

In what follows, we compare the out-of-sample forecasting accuracy of four alternative models used for forecasting air pollution (NOx) in the whole set of 23 territorial MS in Madrid.

More precisely, let us call $\widehat{y}_{nt+1}$ the forecast at time $t+1$ for the nth MS of a given model. The four competing models considered in the paper are: (1) the RW model, where $\widehat{y}_{nt+1} = y_{nt}$; (2) the Feed-Forward Neural Network (NN-X), where environmental observable factors are used as inputs, Eq. (1); (3) the Autoregressive Neural Network (ARNN) of order L, Eq. (2); (4) the Factor-Augmented Autoregressive Neural Network (FA-ARNN) as proposed in [35], where L lags and K factors are included in Eq. (4); (5) the Factor-Augmented Autoregressive Neural Network with environmental observable factors (FA-ARNN-X), which, as shown in Eq. (6), takes advantage of K factors estimated with the Bai [31] approach and L autoregressive lags. We then evaluate a suitable network architecture for all the considered models, the number of autoregressive lags L to include and the number of factors K.

To choose the network architecture, following baseline indications, we consider a single hidden layer network for all the models and set the number of hidden layers based on the rule of thumb of 2/3 of the total number of inputs plus one [29,30]. In our setting, given P environmental covariates, it means that $H = [2/3(P+K+L)]+1$. Moreover, as explained in Section 2.2, we consider a set of $M = 20$ different network models for each of the four analyzed models that are then used to get 20 alternative one-step-ahead predictions [32]. The final-step-ahead prediction, for each model, is given by the mode ensemble of the M alternative networks based on different optimizations.

For the out-of-sample forecasting experiment, a rolling-window procedure is adopted. The sample is divided into a train and a test sample, and the neural networks' parameters are estimated considering training sample observations. We need to implement a rolling-window procedure to estimate the values of the latent factors within each recursion. The training set consists of 95% of the observations to ensure the accurate training of the network parameters, while we leave the last six months for the out-of-sample testing set. Also the number of lags L

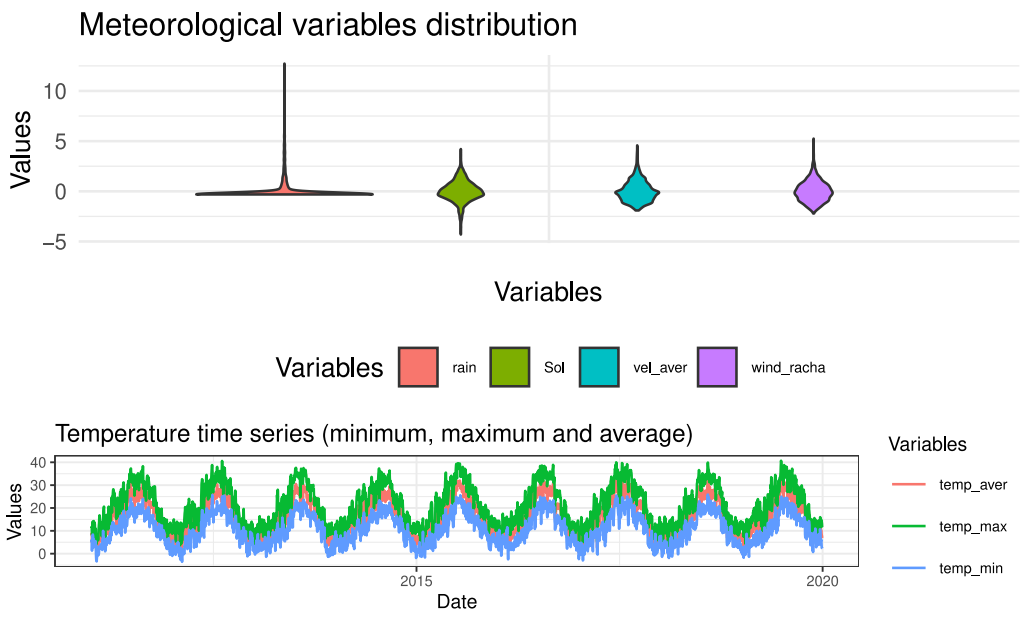


Fig. 5. Meteorological variables included in the forecasting experiment: temperature time series and distribution of (standardized) variables.

Table 2
Descriptive statistics.

	Type	Mean	St. Dev.	Min	Max
Monitoring stations (NOx):					
Escuelas Aguirre	UT	100.627	64.094	11.063	559.907
Avda. Ramon y Cajal	UT	81.506	61.815	9.538	458.771
Arturo Soria	UB	64.848	53.809	2.467	405.257
Villaverde	UB	95.327	109.014	5.386	662.151
Farolillo	UB	65.379	60.180	0.246	460.336
Casa de Campo	S	35.869	37.471	−0.576	348.984
Barajas Pueblo	UB	61.586	49.010	4.122	365.950
Plaza del Carmen	UB	81.477	60.549	−0.870	434.402
Moratalaz	UT	70.135	58.082	7.503	445.862
Cuatro Caminos	UT	87.544	71.213	9.008	541.754
Barrio del Pilar	UT	85.846	88.971	6.535	663.573
Puente de Vallecas	UB	64.085	51.924	−0.573	394.474
Mendez Alvaro	UB	71.106	67.459	5.697	459.020
Castellana	UT	63.835	49.194	4.210	426.953
Retiro	UB	48.786	46.359	−0.359	418.294
Plaza de Castilla	UT	83.610	56.691	6.831	755.853
Ensanche de Vallecas	UB	69.441	66.289	2.921	450.267
Urbanizacion Embajada	UB	76.366	60.944	5.156	491.350
Plaza Elíptica	UT	124.875	87.466	11.271	652.223
Sanchinarro	UB	58.061	54.717	5.293	513.556
El Pardo	S	24.957	21.253	−1.234	276.823
Juan Carlos I	S	40.813	37.191	3.542	395.226
Tres Olivos	S	52.324	49.426	4.394	463.114
Observable environmental factors:					
Daily aver. temp.		14.987	7.371	−0.038	31.354
Daily min. temp.		9.881	6.225	−3.876	23.993
Daily max. temp.		20.091	8.684	2.322	39.137
Rainfall		0.081	3.301	−7.019	39.354
Averg. wind speed		0.882	0.916	−1.314	5.329
Wind direction		8.071	3.169	1.210	24.653
Daily sunshine hours		941.879	5.797	916.902	966.287

for each station and the number of factors K to include in the factor augmented approaches is selected considering training sample data. For the selection of the number of lags, we rely on the automatic procedure explained in [48] based on the BIC criterion. Therefore, we consider a different number of lags L for the $N = 24$ MS. For the number of lags, we adopt the procedure explained in [36].

Given the estimated parameters in the training set and the factor estimation at each recursion, a one-step-ahead forecast is produced as explained in Section 2. Then, a new observation is included in the sample while the oldest one is removed. The procedure is repeated until no new observation is available.

To evaluate forecasting accuracy, we rely on the Mean Absolute Error (MAE):

$$\mathrm{MAE}_n = \frac{1}{T-s}\sum_{t=s+1}^{T}\left|\hat{y}_{nt} - y_{nt}\right|, \tag{13}$$

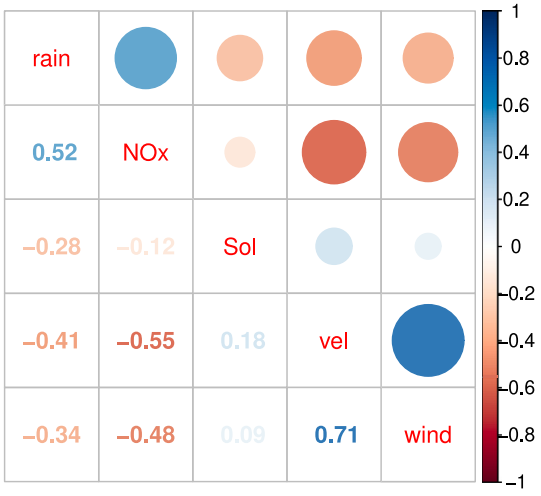


Fig. 6. Unconditional correlations between observable environmental factors and (average) NOx values in Madrid area. Seasonally adjusted data.

with s being the starting temporal data point of the out-of-sample testing set. We use MAE because this accuracy metric is much more robust to outliers and in such settings provides a more reliable comparison across different methods [49].

Finally, given absolute forecasting error loss, we also provide a comparison across the methods using the predictive accuracy test of Diebold and Mariano [50]. In sum, given two competing models j and j' with errors for the nth monitoring station (we omit n subscript for ease of reading) at time t:

$$e_{jt} = g\left(\hat{y}_{nt,j} - y_{nt}\right), \tag{14}$$

and

$$e_{j't} = g\left(\hat{y}_{nt,j'} - y_{nt}\right), \tag{15}$$

where $\hat{y}_{nt,j}$ is the forecast obtained with the j-th model and $g(\cdot)$ is a generic loss function, we consider the following loss differential

$$d_{nt} = e_{jt} - e_{j't} \tag{16}$$

Under the null hypothesis of the Diebold and Mariano [50] test, the two models j and j' have the same predictive accuracy. To test the validity of this hypothesis in practice, we can consider the following linear regression [51]

$$d_{nt} = \iota + \eta_t \tag{17}$$

with $\iota = [1, \ldots, 1]'$ and η_t error terms, and conduct inference on the constant term considering HAC robust standard errors to account for possible auto-correlation and heteroscedasticity in the error terms differential series. Under the null hypothesis, given the stationarity of the error differential d_t, this test is asymptotically normally distributed [50, 51].

4. Results

Given the experimental setup, we split the data into training and testing sets. In the training set we choose, for each MS, the number of lags to include in the ARNN models. The optimal lag selection according to the BIC criterion is reported in Table 3.

We note that for all the MS we choose a number of lags between 1 and 3 days. This finding is consistent with previous studies on air pollution modeling and forecasting, showing that just a few lags are needed for accurate predictions in autoregressive models (see [52], for a review on this issue). The different number of lags implies a different size of the NN for the MS, some of which are deeper than others.

Table 3
Selected number of autoregressive lags L (BIC criterion).

Stations	Number of lags L
Escuelas Aguirre	1
Avda. Ramon y Cajal	1
Arturo Soria	3
Villaverde	3
Farolillo	3
Casa de Campo	3
Barajas Pueblo	1
Plaza del Carmen	2
Moratalaz	1
Cuatro Caminos	1
Barrio del Pilar	2
Puente de Vallecas	1
Mendez Alvaro	2
Castellana	1
Retiro	1
Plaza de Castilla	3
Ensanche de Vallecas	1
Urbanizacion Embajada	1
Plaza Elíptica	3
Sanchinarro	2
El Pardo	1
Juan Carlos I	1
Tres Olivos	1

Then, we estimate the number of factors, K, in the sample of MS. Fig. 7 shows the results obtained with two different procedures. The first one is based on the Bai [31] approach, which also includes the presence of observable factors, while the second one is based on the Stock and Watson [34] approach without observable factors.

Fig. 7 shows that for the training sample estimates, the latent factor estimated with the approach adopted in [35] is much more volatile than the one estimated with the Bai [31] approach discussed in Section 2. The higher volatility of the latent factors estimated with the Babikir and Mwambi [35] approach compared to our proposal, can probably be attributed to the absence of the environmental observable factors. Indeed, omitting variables could lead to a more biased estimate of the latent factors.

Considering the number of lags in Table 3 and the values of the $K = 1$ factors identified in Fig. 7, we train $N = 23$ alternative NN models, and in particular $M = 20$ for each MS to implement the mode forecast ensemble proposed by Kourentzes et al. [32]. Moreover, as explained in Section 3.2, we train four different NN models: NN-X, where environmental observable factors are used as inputs, Eq. (1); the ARNN of order L, Eq. (2); the FA-ARNN considering the red common factor in Fig. 7; and the FA-ARNN-X with the blue common factor shown in Fig. 7. Therefore, we train a total of 1840 networks. The parameters are then adopted to obtain one-step-ahead forecasts, with the latent factors re-estimated following a rolling-window approach, as explained in Section 3.2.

The out-of-sample forecasting results, in terms of MAE loss function (13), are shown in Table 4 for all the MS. Also reported at the bottom of the table are the results comparing the different NN methods, including the RW benchmark, considering average forecasting error.

First of all, a very important result to highlight in Table 4 is that the use of autoregressive lags is crucial to obtain accurate out-of-sample forecasts for NOx pollutants. In fact, the NN-X model performs poorly in all MS.

However, among the other models employing autoregressive lags, the one proposed in this paper (FA-ARNN-X) provides, on average, the most accurate out-of-sample prediction. This suggests that the use of observable factors alone is not useful in forecasting, and that the use of latent factors is important to achieve more accurate predictions. The ARNN model, which employs autoregressive lags only, provides an MAE of 19.26, while the FA-ARNN-X model proposed here provides an MAE of 17.41 on average.

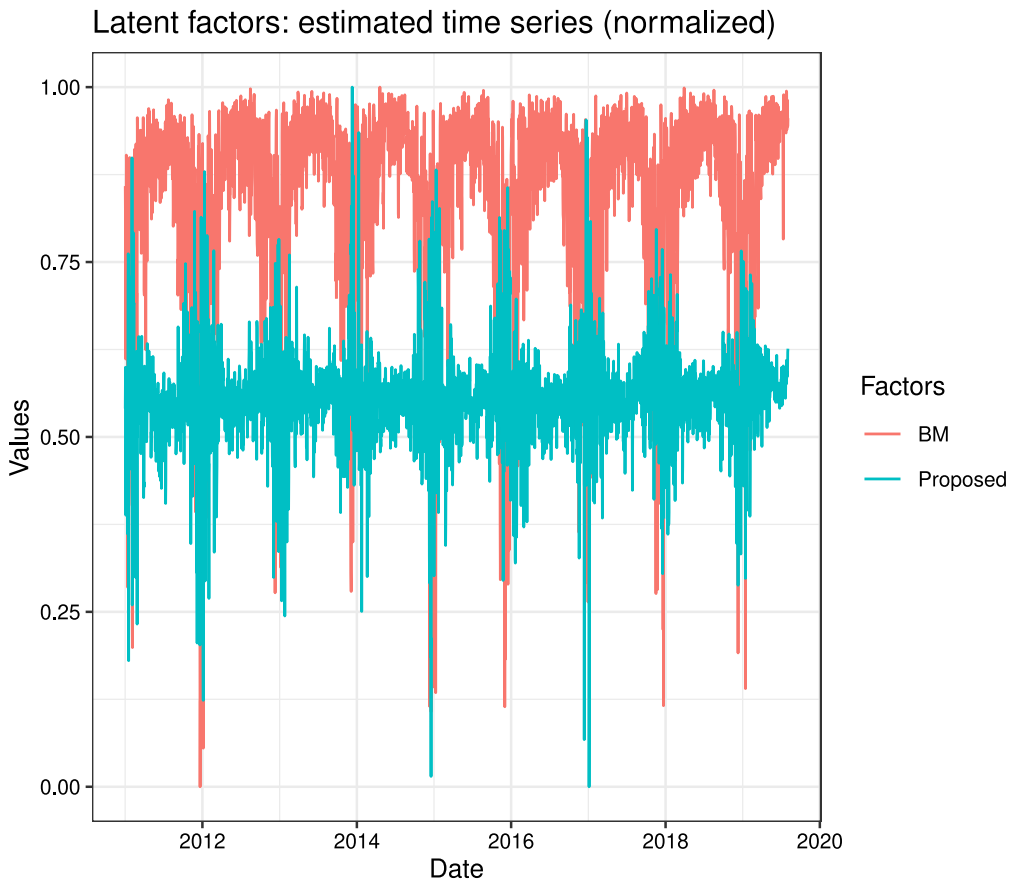


Fig. 7. Normalized time series of the estimated latent factors with Bai [31] (our approach, in blue) and Stock and Watson [34] ([35], in red).

It is interesting to note, however, that the way the latent factors are estimated has an important effect on the performance of the factor-augmented ARNN models. In fact, the Babikir and Mwambi [35] approach provides an MAE of 22.15 on average, which is also lower than the value of the base ARNN model. Looking at this result, one might think that the inclusion of factors reduces the forecasting accuracy of an ARNN, but the FA-ARNN-X results suggest the opposite. In other words, when there are observable factors, these should be included in the estimation procedure for the latent factors.

Another interesting result is that the RW model represents, on average, the second-best option among the four considered NN models. We can thus claim that the use of properly estimated latent factors is required for the accurate prediction of air pollution in general. Finally, we notice that the use of the selected meteorological covariates, motivated however by previous studies (e.g. see [53,54]), seems to exploit its potential while combined with the latent factors. The use of such covariates, indeed, seems to be important for the correct estimation of the factors.

We then evaluate whether the differences highlighted in Table 4 are statistically significant by means of the Diebold and Mariano [50] predictive accuracy test. The results, in terms of p-values, are shown in Table 5.

Table 5 shows the pairwise comparison between the proposed FA-ARNN-X model and all the other competing models. Under the null hypothesis, the model in the column has the same predictive accuracy as the FA-ARNN-X model. Interestingly, considering average errors, we reject the null hypothesis of the test in all cases. Thus, we can claim that the superior accuracy of the FA-ARNN-X model, in terms of absolute errors, is statistically significant.

5. Discussion and final remarks

This study suggests that our proposed factor-augmented autoregressive neural network (FA-ARNN-X), built by referring to previous literature and tailored to the city of Madrid, can be considered a new tool for predicting NOx emissions that accounts for different meteorological variables and latent factor variables. As far as we know, this is the first time that factor-based techniques have been translated to the area of air pollution control in the form of an FA-ARNN-X model for predicting NOx concentrations.

Therefore, we contribute a high-performance strategy to the literature on air pollution control, extended with a meteorological latent factor variable capturing the impact of the meteorological conditions, which in turn have proven to be relevant drivers of air pollutant concentrations. This is undoubtedly an important finding for municipal forecasting teams, municipal authorities, and policy makers in charge of environmental issues.

We focused on the most important area of Spain, the city of Madrid. With its dense population, extensive transportation network, and industrial activities, the city represents a complex urban environment where NOx emissions from various sources converge. The consequences of prolonged exposure to high NOx levels include respiratory and cardiovascular diseases, decreased lung function, and a heightened risk of premature mortality. Moreover, NOx emissions contribute to the formation of ground-level ozone and particulate matter, exacerbating air quality issues.

This paper contributes to the existing literature and provides policy makers with information and tools for air pollution control. Indeed, this city has undergone tremendous economic change, which has caused a rise in population density and increased mobility of people and goods. Such is the interest of the authorities and citizens in air quality that

Table 4
Out-of-sample forecasting accuracy (MAE loss). Models are in columns and MS are in rows. Best model is in bold.

Stations	Type	NN-X	ARNN	FA-ARNN	FA-ARNN-X	RW
Escuelas Aguirre	UT	76.1314	23.4098	28.7967	22.5998	22.8559
Avda. Ramon y Cajal	UT	74.5913	26.9172	31.7812	25.1037	25.6120
Arturo Soria	UB	66.4097	19.3938	22.5408	18.0001	19.7982
Villaverde	UB	123.0084	28.9016	35.5697	27.1729	29.4459
Farolillo	UB	71.7923	20.6246	23.9645	19.3296	21.0165
Casa de Campo	S	42.7393	13.4961	15.0856	12.0261	13.3591
Barajas Pueblo	UB	61.3119	22.6754	27.5242	23.8999	23.8817
Plaza del Carmen	UB	69.1432	20.0037	23.3846	17.9684	19.6065
Moratalaz	UT	69.8669	21.5638	24.9114	19.9532	21.8303
Cuatro Caminos	UT	85.2074	24.2566	29.8417	22.4546	23.1333
Barrio del Pilar	UT	99.8831	24.9284	29.1461	22.9118	23.5874
Puente de Vallecas	UB	60.1968	20.7459	23.3886	20.5087	21.9503
Mendez Alvaro	UB	76.6373	27.1721	26.8512	23.7268	26.8268
Castellana	UT	61.7800	20.0146	21.3992	18.6862	19.9572
Retiro	UB	56.8456	16.4101	17.3804	14.5920	16.2290
Plaza de Castilla	UT	73.8861	21.2728	24.2422	22.1461	21.0458
Ensanche de Vallecas	UB	76.8233	20.8751	23.4541	18.9038	20.3093
Urb. Embajada	UB	75.5113	24.6355	28.3863	23.4654	25.0118
Plaza Elíptica	UT	104.7676	34.3940	42.7900	32.7098	35.2679
Sanchinarro	UB	66.5549	19.2385	20.6788	17.8861	18.6428
El Pardo	S	24.2557	13.0297	24.9940	24.5308	8.9812
Juan Carlos I	S	44.2650	15.8118	17.1670	16.0361	16.2083
Tres Olivos	UB	63.6194	17.3786	20.7699	15.1268	15.7446
Average errors		69.7592	19.2605	22.1570	**17.4165**	19.0065

Table 5
Results of the Diebold and Mariano [50] predictive accuracy test (p-values). Under the null hypothesis, the model in the column has predictive accuracy that is better than or equal to the FA-ARNN model.

Stations	Type	NN	ARNN	FA-NN	RW
Escuelas Aguirre	UT	0.00	0.27	0.02	0.43
Avda. Ramon y Cajal	UT	0.00	0.13	0.05	0.36
Arturo Soria	UB	0.00	0.13	0.03	0.05
Villaverde	UB	0.00	0.17	0.01	0.10
Farolillo	UB	0.00	0.11	0.02	0.06
Casa de Campo	S	0.00	0.02	0.01	0.04
Barajas Pueblo	UB	0.00	0.83	0.07	0.51
Plaza del Carmen	UB	0.00	0.03	0.00	0.07
Moratalaz	UT	0.00	0.10	0.02	0.05
Cuatro Caminos	UT	0.00	0.08	0.00	0.32
Barrio del Pilar	UT	0.00	0.10	0.01	0.32
Puente de Vallecas	UB	0.00	0.43	0.07	0.14
Mendez Alvaro	UB	0.00	0.01	0.08	0.02
Castellana	UT	0.00	0.13	0.11	0.10
Retiro	UB	0.00	0.03	0.03	0.04
Plaza de Castilla	UT	0.00	0.71	0.20	0.81
Ensanche de Vallecas	UB	0.00	0.06	0.01	0.12
Urb. Embajada	UB	0.00	0.20	0.03	0.10
Plaza Elíptica	UT	0.00	0.18	0.00	0.08
Sanchinarro	UB	0.00	0.13	0.07	0.25
El Pardo	S	0.98	1.00	0.38	1.00
Juan Carlos I	S	0.00	0.59	0.22	0.43
Tres Olivos	UB	0.00	0.01	0.00	0.26
Average errors		0.00	0.10	0.03	0.07

the city of Madrid has been divided into five zones for the purposes of Air Quality. Moreover, the Madrid City Council launched the Madrid 360 Environmental Sustainability Strategy in September 2019 to reduce the city's polluting emissions and comply with the air quality limits established in Directive 2008/50/EC of the European Parliament and of the Council, May 21, 2008.

Furthermore, continuous traffic restriction measures are continually being implemented to control NOx levels in accordance with Madrid 360 municipality regulation https://www.madrid360.es/medio-ambiente/zonas-de-bajas-emisiones/. It includes the creation of Low Emission Zones (LEZ) to improve environmental protection in the 21 districts. The entire municipal area has been declared LEZ, although circulation restrictions are to be applied progressively from January 1, 2022 to 2025.

In light of these efforts, our FA-ARNN-X model can be considered an excellent forecasting tool that can be harnessed by the Madrid municipalities. It provides better predictions than the competing models: the Feed-Forward Neural Network (NN-X), where environmental observable factors are used as inputs; the Autoregressive Neural Network (ARNN) of order P; the Factor-Augmented Autoregressive Neural Network (FA-ARNN) considering the red common factor in Fig. 7; and the Factor-Augmented Autoregressive Neural Network with environmental observable factors (FA-ARNN-X) with the blue common factor shown in Fig. 7. Moreover, its results for the most polluted MS are always better in terms of average errors.

Much remains to be done in the field of air pollution control. From the perspective adopted in this article, two promising research avenues are (i) to include variables that take into account the fleet and type of vehicle and (ii) to improve the way of estimating the factors. Both scenarios constitute a major challenge in the area of air quality forecasting. About the use of additional variables, our results show that the inclusion of relevant variables is particularly useful for the correct estimation of the latent factors, which lead to more accurate forecasts in out-of-sample. About the factors estimation process, we highlight that recent papers, such as Andreini et al. [55], Dixon and Polson [56] and Kelly et al. [57], propose theoretical frameworks allowing for the estimation of the factors within the neural network. The application of similar techniques, which share both pros and cons, for environmental problems is an interesting direction for future research, especially given the nonlinearities of environmental time series. In the end, we notice that the proposed approach can also be used for forecasting other pollutants where unobserved factors may play a role in obtaining more accurate forecasts.

Funding

No funding was received for this work.

CRediT authorship contribution statement

Gema Fernández-Avilés: Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Raffaele Mattera:** Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Germana Scepi:** Writing – original draft, Visualization, Validation, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

No conflict of interest exists. We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Data availability

Data will be made available on request.

Acknowledgments

This work has been partially funded by the University of Castilla-La Mancha, Department of Applied Economics I (Code: 00421I126).

References

[1] Lee KK, Bing R, Kiang J, Bashir S, Spath N, Stelzle D, Mortimer K, Bularga A, Doudesis D, Joshi SS, et al. Adverse health effects associated with household air pollution: a systematic review, meta-analysis, and burden estimation study. Lancet Glob Health 2020;8(11):e1427–34.
[2] Rodgers M, Coit D, Felder F, Carlton A. Assessing the effects of power grid expansion on human health externalities. Soc-Econ Plan Sci 2019;66:92–104.
[3] Fuller R, Landrigan PJ, Balakrishnan K, Bathan G, Bose-O'Reilly S, Brauer M, Caravanos J, Chiles T, Cohen A, Corra L, et al. Pollution and health: a progress update. Lancet Planet Health 2022;6(6):e535–47.
[4] Portnov BA, Dubnov J, Barchana M. Studying the association between air pollution and lung cancer incidence in a large metropolitan area using a kernel density function. Soc-Econ Plan Sci 2009;43(3):141–50.
[5] European Environment Agency. New initiative to measure outdoor air quality at schools across Europe. Technical report, European Environment Agency; 2019.
[6] Dong J, Li S, Xing J, Sun Y, Yang J, Ren L, Zeng X, Sahu SK. Air pollution control benefits in reducing inter-provincial trade-associated environmental inequality on PM2.5 related premature deaths in China. J Clean Prod 2022;350:131435.
[7] Torkayesh AE, Alizadeh R, Soltanisehat L, Torkayesh SE, Lund PD. A comparative assessment of air quality across European countries using an integrated decision support model. Soc-Econ Plan Sci 2022;81:101198.
[8] Giacalone M, Mattera R, Nissi E. Well-being analysis of Italian provinces with spatial principal components. Soc-Econ Plan Sci 2022;84:101377.
[9] Zhang Z, Zhang G, Su B. The spatial impacts of air pollution and socio-economic status on public health: Empirical evidence from China. Soc-Econ Plan Sci 2022;83:101167.
[10] Aguilar-Gomez S, Dwyer H, Graff Zivin J, Neidell M. This is air: The "nonhealth" effects of air pollution. Annu Rev Resour Econ 2022;14:403–25.
[11] Chien F, Zhang Y, Sharif A, Sadiq M, Hieu MV. Does air pollution affect the tourism industry in the USA? Evidence from the quantile autoregressive distributed lagged approach. Tour Econ 2023;29(5):1164–80.
[12] OECD. The economic consequences of outdoor air pollution. Technical report, OECD Publishing; 2020.
[13] Lamb WF, Wiedmann T, Pongratz J, Andrew R, Crippa M, Olivier JG, Wiedenhofer D, Mattioli G, Al Khourdajie A, House J, et al. A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. Environ Res Lett 2021;16(7):073005.
[14] Chen F, Chen Z. Cost of economic growth: Air pollution and health expenditure. Sci Total Environ 2021;755:142543.
[15] Laureti T, Montero J-M, Fernández-Avilés G. A local scale analysis on influencing factors of NOx emissions: Evidence from the community of madrid, Spain. Energy Policy 2014;74:557–68.
[16] Montero J-M, Fernández-Avilés G, Laureti T. A local spatial STIRPAT model for outdoor NOx concentrations in the community of Madrid, Spain. Mathematics 2021;9(6):677.
[17] Pagan AR, Robertson J. Forecasting for policy. A companion to economic forecasting. 2002, p. 152–78.
[18] Van der Steen M. Anticipation tools in policy formulation: Forecasting, foresight and implications for policy planning. In: Handbook of policy formulation. Edward Elgar Publishing; 2017, p. 182–97.
[19] Lemos MC, Finan TJ, Fox RW, Nelson DR, Tucker J. The use of seasonal climate forecasting in policymaking: lessons from Northeast Brazil. Clim Change 2002;55:479–507.
[20] Green KC, Armstrong JS, Soon W. Validity of climate change forecasting for public policy decision making. Int J Forecast 2009;25(4):826–32.
[21] Montero J-M, Fernández-Avilés G. Functional kriging prediction of atmospheric particulate matter concentrations in Madrid, Spain: Is the new monitoring system masking potential public health problems? J Clean Prod 2018;175:283–93.
[22] Varotsos CA, Mazei Y, Saldaev D, Efstathiou M, Voronova T, Xue Y. Nowcasting of air pollution episodes in megacities: a case study for Athens, Greece. Atmos Pollut Res 2021;12(7):101099.
[23] Sanchis-Marco L, Montero J-M, Fernández-Avilés G. An extended CAViaR model for early-warning of exceedances of the air pollution standards. The case of PM10 in the city of Madrid. Atmos Pollut Res 2022;13(4):101355.
[24] Gao M, Yang H, Xiao Q, Goh M. COVID-19 lockdowns and air quality: Evidence from grey spatiotemporal forecasts. Soc-Econ Plan Sci 2022;83:101228.
[25] Taylor CJ, Pedregal DJ, Young PC, Tych W. Environmental time series analysis and forecasting with the Captain toolbox. Environ Model Softw 2007;22(6):797–814.
[26] Bai L, Wang J, Ma X, Lu H. Air pollution forecasts: An overview. Int J Environ Res Public Health 2018;15(4):780.
[27] Masood A, Ahmad K. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. J Clean Prod 2021;322:129072.
[28] Zhang B, Rong Y, Yong R, Qin D, Li M, Zou G, Pan J. Deep learning for air pollutant concentration prediction: A review. Atmos Environ 2022;119347.
[29] Aggarwal CC, et al. Neural networks and deep learning. Springer 2018;10(978):3.
[30] Heaton J. Introduction to neural networks with Java. Heaton Research, Inc.; 2008.
[31] Bai J. Panel data models with interactive fixed effects. Econometrica 2009;77(4):1229–79.
[32] Kourentzes N, Barrow DK, Crone SF. Neural network ensemble operators for time series forecasting. Expert Syst Appl 2014;41(9):4235–44.
[33] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2018.
[34] Stock JH, Watson MW. Forecasting using principal components from a large number of predictors. J Amer Statist Assoc 2002;97(460):1167–79.
[35] Babikir A, Mwambi H. Factor augmented artificial neural network model. Neural Process Lett 2017;45:507–21.
[36] Bai J, Ng S. Determining the number of factors in approximate factor models. Econometrica 2002;70(1):191–221.
[37] Hallin M, Liška R. Determining the number of factors in the general dynamic factor model. J Amer Statist Assoc 2007;102(478):603–17.
[38] Fama EF. Random walks in stock market prices. Financ Anal J 1995;51(1):75–80.
[39] Meese RA, Rogoff K. Empirical exchange rate models of the seventies: Do they fit out of sample? J Int Econ 1983;14(1–2):3–24.
[40] Moosa I, Burns K. The random walk as a forecasting benchmark: drift or no drift? Appl Econ 2016;48(43):4131–42.
[41] Naftaly U, Intrator N, Horn D. Optimal ensemble averaging of neural networks. Network: Comput Neural Syst 1997;8(3):283.
[42] Crone SF, Hibon M, Nikolopoulos K. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. Int J Forecast 2011;27(3):635–60.
[43] Hansen LK, Salamon P. Neural network ensembles. IEEE Trans Pattern Anal Mach Intell 1990;12(10):993–1001.
[44] Pizarro M, Hernangómez D, Fernández-Avilés G. Climaemet: Climate AEMET tools. 2021, http://dx.doi.org/10.5281/zenodo.5205573, URL: https://hdl.handle.net/10261/250390.
[45] Briz-Redón Á, Belenguer-Sapiña C, Serrano-Aroca Á. Changes in air pollution during COVID-19 lockdown in Spain: a multi-city study. J Environ Sci 2021;101:16–26.
[46] Querol X, Massagué J, Alastuey A, Moreno T, Gangoiti G, Mantilla E, Duéguez JJ, Escudero M, Monfort E, Garcia-Pando CP, et al. Lessons from the COVID-19 air pollution decrease in Spain: now what? Sci Total Environ 2021;779:146380.
[47] Giacalone M, Mattera R, Nissi E. Economic indicators forecasting in presence of seasonal patterns: Time series revision and prediction accuracy. Qual Quant 2020;54:67–84.
[48] Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. J Stat Softw 2008;27:1–22.
[49] McCracken MW. Robust out-of-sample inference. J Econometrics 2000;99(2):195–223.
[50] Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econom Statist 2002;20(1):134–44.
[51] Diebold FX. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. J Bus Econom Statist 2015;33(1):1.
[52] Kaur J, Parmar KS, Singh S. Autoregressive models in environmental forecasting time series: a theoretical and application review. Environ Sci Pollut Res 2023;30(8):19617–41.
[53] Cogliani E. Air pollution forecast in cities by an air pollution index highly correlated with meteorological variables. Atmos Environ 2001;35(16):2871–7.
[54] Hrust L, Klaić ZB, Križan J, Antonić O, Hercog P. Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. Atmos Environ 2009;43(35):5588–96.
[55] Andreini P, Izzo C, Ricco G. Deep dynamic factor models. 2023, arXiv preprint arXiv:2007.11887.

[56] Dixon M, Polson N. Deep fundamental factor models. SIAM J Financial Math 2020;11(3):SC26–37.
[57] Kelly B, Kuznetsov B, Malamud S, Xu TA. Large (and deep) factor models. 2024, arXiv preprint arXiv:2402.06635.

Gema Fernández-Avilés is Full Professor of Applied Economics (Statistics) at the University of Castilla-La Mancha, Spain. Prof. Fernández-Avilés's research lies broadly in economics and data science in the area of air pollution control and real estate prices forecasting, using spatial and spatio-temporal statistical methodology. Her doctoral thesis was awarded as the best doctoral thesis on Regional Economy in the Community of Madrid in 2010. She is the director of Máster in Data Science and Business Analytics (with R software) and the president of R-Quixote Association. Currently, she has co-edited the book "Fundamental of data science with R".

Raffaele Mattera holds a PhD in Econometrics at the University of Naples "Federico II" and is Assistant Professor of Statistics at the Sapienza University of Rome, Italy. His main research interest is in computational statistics and predictive analytics, with a focus on clustering and forecasting of temporal, spatio-temporal and network data. He is a member of the International Institute of Forecasters, the Italian Statistical Society and the Spatial Econometrics Association.

Germana Scepi is Associate Professor of Statistics at the University of Naples "Federico II", Italy. Her research activity in recent years covers different areas ranging from Multidimensional Data Analysis to Data Mining, with special reference to Temporal Data Mining and Textual Data Analysis. Within these areas, her research mainly focused on Cluster Analysis, Conjoint Analysis, High-Frequency Time Series Analysis, and Analysis of documents from large databases. Application areas include Customer Satisfaction, Environmental and Financial Time Series and, finally, Document Analysis from Corporate Annual Reports.