IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 February 2024, accepted 4 March 2024, date of publication 5 March 2024, date of current version 18 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3374208

RESEARCH ARTICLE

Improving Air Quality Zoning Through Deep Learning and Hyperlocal Measurements

EDUARDO ILLUECA FERNÁNDEZ[1], ANTONIO JESÚS JARA VALERA[2], (Senior Member, IEEE), AND JESUALDO TOMÁS FERNÁNDEZ BREIS[1], (Senior Member, IEEE)

[1]Department of Informatics and Systems, University of Murcia, 30100 Murcia, Spain
[2]Research and Development Department, Libelium LAB, 30562 Ceutí, Spain

Corresponding author: Eduardo Illueca Fernández (eduardo.illueca@um.es)

This work was supported in part by the Fundación Séneca under Grant 21300/FPI/19; in part by Libelium LAB S. L. Región de Murcia, Spain, under Grant 21681/EFPI/21; and in part by European Commission under Grant 101037648 (SOCIO-BEE).

ABSTRACT According to the Air Quality Directive 2008/50/EC, air quality zoning divides a territory into air quality zones where pollution and citizen exposure are similar and can be monitored using similar strategies. However, there is no standardized computational methodology to solve this problem, and only a few experiences in the Comunidad of Madrid based on chemistry transport models. In this study, we propose a methodological improvement based on the application of deep learning. Our method uses the CHIMERE-WRF air quality modelling system and adds a step that uses neural networks architectures to calibrate the simulations. We have validated our method in the Region of Murcia. The results obtained are promising given the values of the Pearson coefficient, obtaining $r = 0.94$ for NO_2 and $r = 0.95$ for O_3, improving 86 % and 29 % the performances reported in the state of the art. In addition, the cluster score improves after applying neural networks, demonstrating that neural networks improve the consistency of clusters compared to the current air quality zoning. This opened new research opportunities based on the use of neural networks for dimension reduction in spatial clustering problems, and we were able to provide recommendations for a new measurement point in the Region of Murcia Air Quality Network.

INDEX TERMS Air quality, artificial neural networks, atmospheric modeling, clustering algorithms, deep learning.

I. INTRODUCTION

Air pollution is the primary environmental health risk that significantly affects morbidity and mortality [1]. It is a key aspect of human health and environmental preservation, and has deteriorated due to increased anthropogenic emissions from different economic sectors [2]. Climate change mitigation is one of the most important challenges related to digital transformation, as outdoor air pollution is a major problem, attributed to 3.7 million deaths globally, and is related to global climate change, acid rain, haze, ozone depletion and crop damage [3]. For this reason, one of the key insights of digital transformation is air quality monitoring, evaluation, prediction and mitigation to create plans for cleaner air. This problem is more serious in cities, which are highly populated and perceived by citizens as unhealthy places. Thus, air quality management is among the most important services in the new smart cities paradigm [4]. The European legislation establishes that member countries should identify different air quality zones by their air quality properties, that is, through a process called air quality zoning [5]. This is a land classification problem where the spatial units are aggregated according to spatial contiguity or adjacency constraints. In this context, these restrictions are defined by the pollutant concentration values of the different spatial units - in this work, we will define the spatial unit as each cell of a grid domain [6]. Defining an air quality dispersion model that generates this information is necessary to assign a value to each cell.

To improve the air that citizens breathe, Directive 2008/50/EC of European Parliament and Council on May 21, 2008, on ambient air quality and cleaner air for Europe

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

indicates that each member state must classify its territory into air quality zones that constitute an independent and homogeneous geographical unit for air quality management (monitoring, implementation, and evaluation of air quality plans) [7]. In addition, population exposure to air pollution within each zone is assumed to be similar [8]. Consequently, exceeding AQ standards and objectives for a pollutant at any AQ monitoring station (AQMS) implies non-compliance with AQ standards and objectives. It is easy to note that methodologies that help stakeholders identify measurement points are essential before implementing an air quality network. This problem is called air quality zoning and consists of identifying areas with similar pollution patterns that can be monitored with the same devices, because it is possible to assume that pollutant concentration values can be extrapolated to the entire zone [9].

In this context, the scientific development of air quality models is essential for zoning based on the evidence provided by data. The chemistry transport models based on an Eulerian approach are the most important ones, and CHIMERE [10] is one of the most relevant at the European level. This is an open-source multi-scale chemistry-transport model designed to produce i) accurate analysis of pollution episodes, ii) daily forecasts of ozone, aerosols and other pollutants, and iii) long-term simulations (entire seasons or years) for emission control scenarios [11]. In this sense, several studies have used chemistry transport models coupled with clustering algorithms to automatise the problem of air quality zoning at the regional level in Madrid [12]. However, that study has several limitations: i) the air quality zoning can only be performed at a regional level due to the computational complexity of the high-resolution simulation, and ii) the process of air quality zoning does not take into account the historical measurement data.

For this reason, we propose a new methodology starting from the automatic clustering proposed in [12]. Thus, this paper aims to validate this new air quality zoning in the Region of Murcia, dealing with the previous limitation by exploiting deep learning technologies and ground measurements. We hypothesised that using artificial neural networks trained with hyperlocal air quality data could improve the accuracy of the predictions, allowing automatic better high-resolution zoning. This has been done by trying different architectures, namely simple artificial neural networks (ANN); long short term memory networks (LSTM-NN); convolutional neural networks and residual neural networks (ResNET). Thus, our proposal will contribute with the following points to state of the art: i) it is the first approach for automatic air quality zoning in the Region of Murcia, an area with an important ecological value, agricultural and livestock activities, and medium-scale industry, being representative of the Mediterranean countries [13]; ii) it is the first application of Deep Learning to adjust simulations with hyperlocal data and iii) we can identify missing measurements points in the implemented air quality networks. We have validated our methods by comparing the proposed air quality zoning with the one proposed by the Spanish Government through the Ministerio para la Transición Ecológica y el Reto Demográfico (MITECO).

The rest of the paper is structured as follows. First, the State of the Art section sets the theoretical background and the work performed in air quality modelling and zoning. Then, the Materials and Methods section describes the methodology used for modelling, the neural networks architecture and the strategy for the air quality zoning. These simulations provide Results, drawing a final air quality zoning for Spain and Ireland, compared with the current proposal. Then, these results are analysed in the Discussion section and compared with state of the art, obtaining recommendations for future researchers. Finally, the main conclusions and future challenges are presented in the Conclusions.

II. STATE OF THE ART

Research on air quality dispersion models has been approached in the literature from two perspectives. The first one aims to model the physical and chemical processes in the atmosphere. These models are named deterministic and typically solve the differential equations that represent the processes controlling the environment and are used for a range of tasks, including developing new scientific understanding and environmental policies [14]. 3D Chemistry Transport Models are the most important and were developed to support decision-making. These models are limited in terms of resolution, mainly because higher model resolutions may provide more accurate and detailed information but with higher computational costs due to the Courant-Friedrichs-Lewy condition [15], being impossible to achieve resolution higher than 1 km x 1 km. Also, the presence of uncertainties in their initial conditions, input variables and parameterisations [16] is the cause of various biases which limit their usefulness for some tasks. In this context, one of the most notable 3D Chemistry Transport Models, CHIMERE, has implemented an online coupled mode with Weather Research Forecast (WRF) in the version available at the time of writing [11].

Other methods are based on statistical and artificial intelligence techniques, including machine learning. Statistical techniques focus on improving the data quality by missing data imputations or signal reconstruction models based on a graph learned from the data, exhibiting good results in O_3, NO_2 and PM10 networks [17]. However, the state-of-the-art statistical models are not enough to capture all the complexity of atmospheric physics.

From the machine learning side, random forest regressors have been used to train a machine learning replacement for the gas-phase chemistry in the GEOS-Chem chemistry model on the timescales of days to weeks [18]. The architectures of the neural networks determine their ability to extract complex nonlinear features across scales from high-dimensional datasets, which is important for

modelling non-linear pollutants such as O_3 or PM2.5 [19]. Neural networks usually obtain better results than traditional machine-learning strategies [20]. Other studies suggest that the gated recurrent unit (GRU) model is slightly better than the RNN and LSTM models for predicting PM10 and PM2.5 concentration [21], [22]. GRU is an improvement of traditional LSTM approaches, and it is based on integrating and screening information in a chronological order, some of which is retained, and the other one is discarded.

Several neural network architectures have been proposed in the air quality dispersion problem context. For example, some studies used recurrent neural networks (RNN) and long short-term memory (LSTM) models to predict air pollution from historical time series pollutant data and meteorological data [23], [24]. For this reason, a new trend is to use the power of deep learning to improve the performance of deterministic models, which could be complemented by machine learning [25]. Deep learning techniques can potentially accelerate atmospheric chemistry computations of CTMs by emulating the chemical mechanism [19]. On the other hand, a simple recurrent 3-layer neural network predicts daily mean concentrations of NO_2, SO_2 and C_2H_6 over Europe based on the Community Multiscale Air Quality model (CMAQ), showing that neural networks can approximate the continuity equation [26]. Other strategies focus on calibrating the model with a backward propagation neural network using ground monitoring data [27]. Last, neural networks have been used to perform high-resolution simulations with low computational cost, using the simulations of the CHIMERE model as a background [15].

According to the literature, there is very limited experience in automated air quality zoning. The most automated approach was performed in Madrid [12], coupling chemistry transport models with clustering algorithms. A Functional Principal Component Analysis (FPCA) was applied to identify the most relevant clustering variables from these statistics and subsequently apply a k-mean cluster analysis. The definition of the number of air quality zones (clusters) is based on three methods: Elbow, Silhouette, and Gap statistics. However, this air quality zoning can only be performed at a regional level, in other words, in a wide area and not on a local/urban scale due to the computational complexity of the high-resolution simulation [28].

The use of neural networks remains unexplored for air quality zoning. It is true that the applications of neural networks for dimensionality reduction have been explored in other fields such as hydrology [29] or medical imaging processing [30]. Other unsupervised strategies for dimensionality reduction includes multiview fuzzy c-means clustering algorithms, a technique used in data analysis and machine learning to group data points into clusters when multiple perspectives or views of data are available [31]. The main advantage of ANN in land classification problems is its contribution to solving the local minima problem present in traditional clustering algorithms, as well as being able to separate complex populations. However, its main limitation is the low explainability of deep learning models, making them not applicable for some use cases. In addition, the computational cost for training is higher [32].

This is the starting point for our work and our main contribution beyond the state of the art is that we propose the first automatic air quality zoning method based on CHIMERE and neural networks.

III. MATERIALS AND METHODS

The overall methodology proposed for this study is based on the work carried out in the Comunidad of Madrid [12], with the addition of deep learning strategies as a scientific contribution in this paper.

In general terms, the methodology is composed of the following steps, as summarised in Figure 1: i) a mesoscale air quality dispersion simulation (CHIMERE-WRF) over the study domain to compute the air quality concentration in each one of the grid points; ii) a calibration using ANN using ground measurements as reference data; iii) a principal component analysis preprocessing to reduce dimensionality; iv) a municipality aggregation algorithm and; v) a clustering approach to classifying each municipality in one air quality zone.

The air quality zoning process involves several steps, each one with specific goals. First, mesoscale air quality dispersion provides continuous concentration values across the domain and historical datasets. Second, neural network calibration addresses bias in chemistry transport models caused by uncertainties in emissions inventories, enhancing the reliability of air quality data [33]. Then, municipality aggregation aims to interpolate data into administrative units for legislative compliance and better interpretation [12]. The next step, first principal component analysis (FPCA), focuses on dimensionality reduction. In the last step, clustering algorithms classify geographical units into different zones based on calculated features [34]. The combined use of these techniques addresses challenges in air quality zoning by automating and standardizing the process, reducing human interaction steps, minimizing discrepancies between regions, and enhancing data quality through neural network calibration.

A. MESOSCALE SIMULATIONS

The mesoscale simulations have been performed with the CHIMERE-WRF coupled model in concrete version v2020r3 [11]. A gridded domain of 15×15 km was filled by cells of 10 km^2 (Figure 2). The meteorology boundary conditions have been downloaded from NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999 (https://rda.ucar.edu/datasets/ds083.2/dataaccess/), and the emission inventories from CAMS Copernicus Atmospheric Monitoring Service (https://eccad3.sedoo.fr/). The emission fluxes in the grid points are computed with the emisurf preprocessor, and the chemical kinetics have been modelled with the MELCHIOR 1 reaction mechanism [35]. The

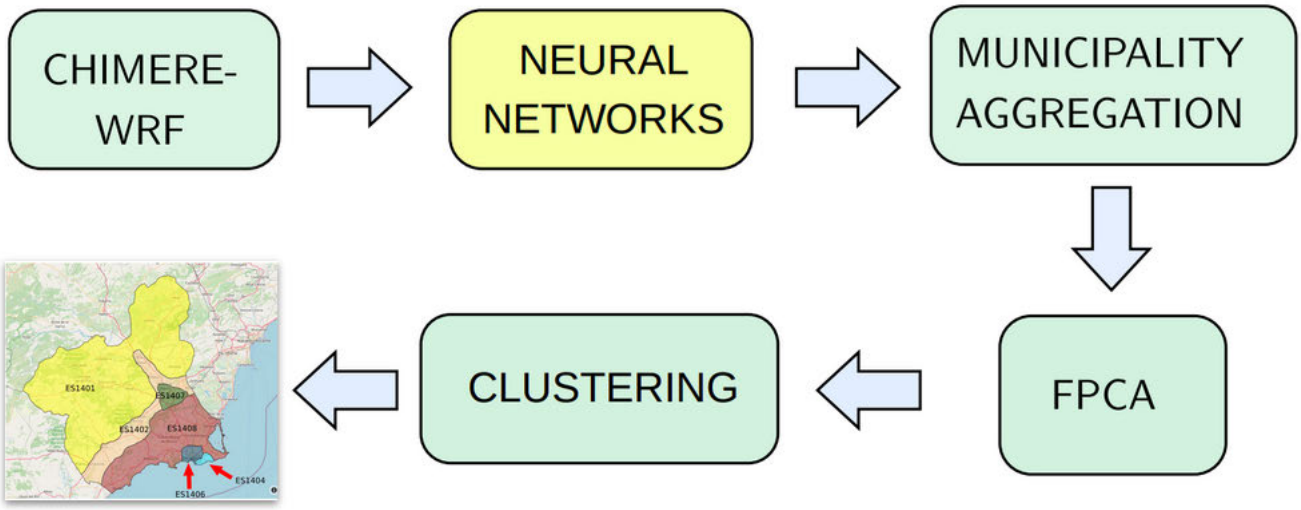


FIGURE 1. Overall methodology for air quality zoning computation, based on [12]. In yellow, the new step that we propose in this paper is based on ANN.

simulation period covers the whole of 2020 with 1 hour timestep.

B. DEEP LEARNING APPROACH

Our main contribution is the use of several neural network architectures for calibration before air quality zoning. The general workflow of the calibration process is presented in Figure 3, in which a model for each pollutant is trained using as feature NO_2, NO, O_3, CO, SO_2, PM10, PM2.5 and temperature computed by CHIMERE-WRF, and the target pollutant as labeled variable. Thus, for each pollutant we resolve a regression problem. Once the model is trained, it is integrated into the architecture presented in Figure 3. Each model takes as input 8 features and returns one, the corrected gas. Then, all these corrections are merged into a single table and used as input for the next step, the FPCA.

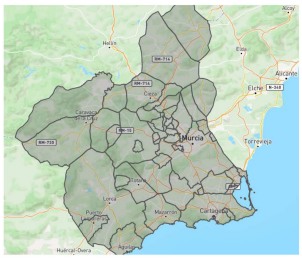
Regarding the hyperparameter tuning process, a k-fold cross validation strategy (Figure 4) has been followed, in which the dataset undergoes a random division into k stratified folds. Each fold serves as a test set once, with the remaining folds temporarily amalgamated to construct a training set for model development [36]. Performance metrics are computed and saved for the test set, and this process is iterated for the total number of generated folds. In general, the number of epochs set for each iteration is 1000, but this number is lower if the training convergence is reached before epoch 1000, understood as the step in which the loss function stops decreasing. The concrete parameters to be tuned depends on the specific architecture, and this section present five different architectures tested: i) simple artificial neural networks with dropout modules (ANN); ii) recurrent neural networks with long short-term memory layers (RNN-LSTM); iii) convolutional neural networks (CNN); and iv) residual neural networks (ResNET). The framework used to define all architectures was tensorflow (https://www.tensorflow.org).

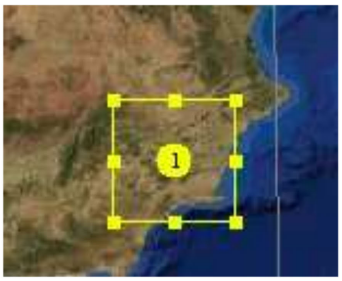
1) ARTIFICIAL NEURAL NETWORK (ANN)

This model consists of a series of layers, starting with a *dense* layer with n units and ReLU activation, which is crucial for capturing complex patterns in the input data. A key step is the introduction of *dropout* module after each *dense* layer, by preventing overfitting and missing values handling, thereby enhancing generalization. The subsequent *LeakyReLU* layers introduce non-linearity with a small negative slope for negative inputs, potentially mitigating the vanishing gradient issue and avoiding collapse in the training process. The architecture proposed concatenates five *Dense-Dropout-LeakyRELU* blocks. By stacking all these layers, the proposed model can learn hierarchical features and intricate relationships within the data with a general and easy replicable architecture. The final Dense layer with a single unit serves as the output layer, producing the regression prediction after passing a sigmoid activation function. The main advantage of the model is its low complexity, crucial for robust performance in regression tasks. However, the simplicity of the model can lead to the inability to learn key process and dependencies of the original dataset. The proposed architecture is schematized in Figure 5.

The hyperparameter tunning process for this architecture is based on exploiting different configuration and probability distributions of the main parameters of the network, the number of neurons (n), the slope of the *LeakyRELU* layer (α), the *learning rate* and the *dropout rate* od the *dropout* module. The *LeakyReLU* α is sampled from a discrete set of values, including [0.1, 0.2, 0.4, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]. The number of neurons in each *dense* layer spans from 1 to 100, enabling the investigation of various model complexities. The learning rate is sampled from a reciprocal distribution between 3e-4 and 3e-2, allowing for exploration across multiple orders of magnitude. The dropout rate (drop) after each Dense layer is selected from the set [0.01, 0.02, 0.03,

MU01 Features	
Resolution (m)	10000 x 10000
Size	15 x 15
Reference Latitude	38.070
Reference Longitude	-1.500
Projection	Lambert





2020

https://rda.ucar.edu/datasets/ds083.2/dataaccess/
https://eccad3.sedoo.fr/

FIGURE 2. Parameters used for the CHIMERE simulations. On the top are the domain dimensions and representation in geographical information systems. On the bottom, the municipalities in the Region of Murcia and the emissions inventories used in this study.

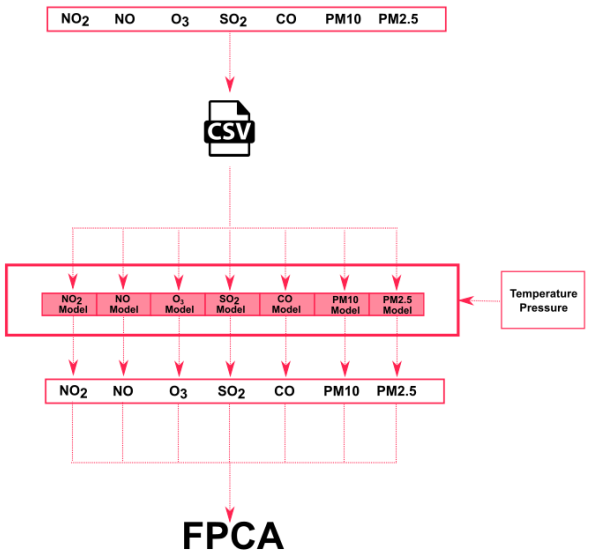


FIGURE 3. Deep learning calibration overview. A model for each pollutant is trained and all the calibrated features as used as input for FPCA.

0.04, 0.05, 0.06, 0.07, 0.08, 0.09]. Adjusting hyperparameters provides flexibility to fine-tune the model's behavior based on the specific characteristics of the data.

2) LONG SHORT TERM MEMORY NEURAL NETWORKS (RNN-LSTM)

The architecture of this neural network is similar to the one proposed by the simple ANN, and it is designed for sequential data processing, focusing on Long Short-Term Memory (LSTM) layers instead of *dense* layers. The model begins with an *LSTM* layer with n units, utilizing a ReLU activation function and an input shape of (1, 8). The choice of *LSTM* layers, each followed by *dropout* regularization and *LeakyReLU* activation, is intended to capture spatiotemporal dependencies and patterns in the input data. As shown in Figure 3, we concatenated seven consecutive blocks *LSTM-Dropout-LeakyRELU*. Each LSTM layer should output sequences rather than a single value, promoting the extraction of sequential information throughout the network. The *LeakyReLU* activation with a specified alpha parameter introduces non-linearity and can mitigate the gradient problem in the training. Stacking multiple LSTM, Dropout, and LeakyReLU blocks facilitates the extraction of hierarchical temporal features. The final layer is a Dense layer with a single unit corresponding to a sigmoid activation function, producing the output for the regression task. The rationale behind this architecture lies in its ability to effectively model complex sequential dependencies and patterns, making it suitable for tasks where understanding temporal and spatial relationships is crucial in air quality zoning problems.

The hyperparameter tunning process is done over the same configurations and parameters proposed by the simple ANN architecture. This implies that the α is sampled from the set [0.1, 0.2, 0.4, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]; the number of neurons in each *LSTM* layer spans from 1 to 100; The learning rate is sampled from a reciprocal distribution between 3e-4 and 3e-2; and the dropout rate is selected from the set [0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09].

3) CONVOLUTIONAL NEURAL NETWORKS (CNN)

The third architecture proposed involves the use of three instances of a convolutional neural network (CNN) followed by concatenation to capture diverse and hierarchical features from the input data, as shown in Figure 7. The input layer is defined with a shape of (8, 1), indicating a one-dimensional convolutional operation over sequences of length 8. Each instance of the model consists of multiple convolutional

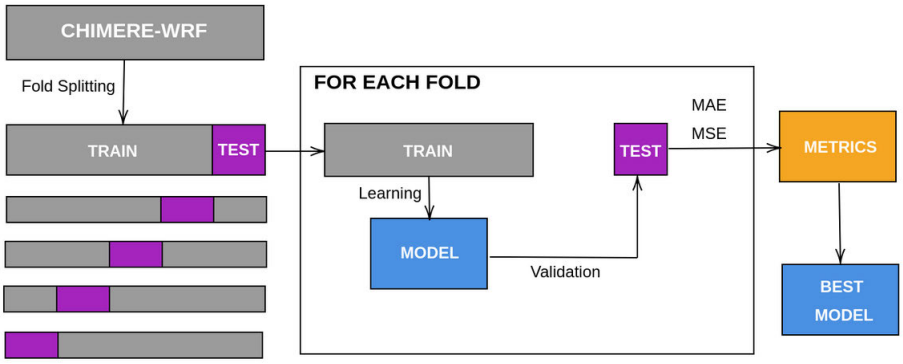


FIGURE 4. The dataset is randomly split into k folds. Each one is used as a test set once, while the other folds are temporarily combined to form a training dataset. Performance metrics on the test set are calculated and stored, and the best metrics are selected at the end of the process.

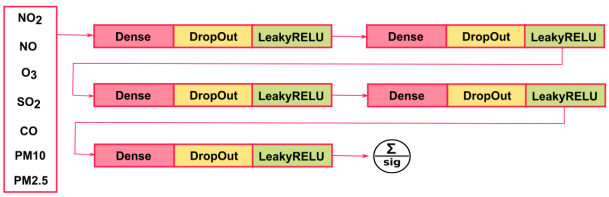


FIGURE 5. Simple artificial neural network architecture (ANN).

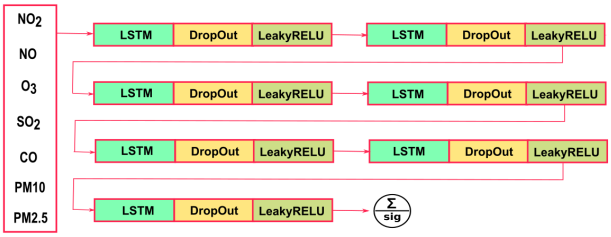


FIGURE 6. Long short term memory recurrent neural networks architecture (LSTM-RNN).

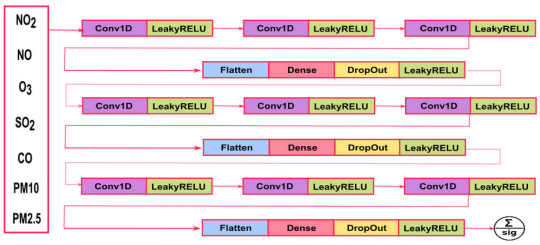


FIGURE 7. Convolutional neural network architecture (CNN).

layers with ReLU activation and LeakyReLU activation, fostering non-linearity while preventing the gradient collapse. The use of three separate instances allows the model to learn distinct patterns from different perspectives. The Flatten layer is employed to transform the output of the convolutional layers into a flat vector, facilitating subsequent processing. Following the flattened representation, dense layers with ReLU activation, dropout for regularization, and LeakyReLU activation are incorporated to extract high-level features. The final layer of each instance uses a dense layer with a sigmoid activation function. The concatenation of the three model instances enables the neural network to combine the diverse representations learned by each instance, promoting a more comprehensive understanding of the input data. This architecture is especially beneficial for capturing complex patterns and relationships in spatial data, making it suitable for tasks such as preparing the clustering classification algorithms.

For this architecture, the hyperparameter Tunning configuration differs a bit of the proposed for the first two architecture, to be suitable for the CNN architecture. First, the LeakyReLU α uses the same configuration, based on the set [0.1, 0.2, 0.4, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]. The same occurs for the learning rate, sampled from a reciprocal distribution between 3e-4 and 3e-2m, and the dropout rate is also selected from the set [0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09]. However, we have added two new hyperparameters. The first one is the number of filters in each convolutional layer, spaning from 1 to 20, providing flexibility in extracting diverse features. Then, filter size is chosen from the set [1, 2, 3], enabling the model to capture features at different scales and avoiding collapese byh removing more dimension that are available in the dataset.

4) RESIDUAL NEURAL NETWORKS (RESNET)

The most complex architecture applied in this work is ResNET, which is described in Figure 8. This ResNET architecture is designed for one-dimensional data, specifically for regression tasks. The core component is the residual block, a fundamental building block of residual neural networks. Each residual block consists of two *CNN* layers with batch normalization and ReLU activation, followed by a residual connection that bypasses one of the *CNN* layers. The purpose of this residual connection is to mitigate the vanishing gradient problem and facilitate the training of deep networks, as well as improve robustness of the models. The parameter

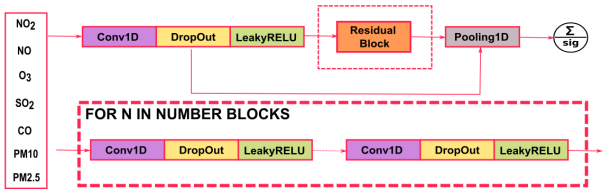


FIGURE 8. Residual neural network architecture (ResNET).

number of blocks control the number of residual blocks that can integrated into the architecture. The overall architecture begins with an initial *CNN* layer with batch normalization and ReLU activation. Subsequently, a *dropout* module is applied, providing the model with the capability to learn hierarchical features efficiently. Then, a *pooling* layer utilizes global average pooling to aggregate spatial information, and a *dense* layer with a sigmoid activation function is employed for regression tasks. This architecture's strength lies in its ability to effectively train deep networks, leveraging residual connections to capture intricate patterns and features in sequential data while mitigating optimization challenges associated with depth to suit the characteristics of the specific regression task.

Regarding the hyperparameter tunning, the configuration for the ResNet architecture involves exploring various key parameters. The kernel size and stride parameters are sampled from discrete sets [1, 2, 3, 4, 5, 6, 7, 8, 9], allowing for flexible adjustment of filter sizes and downsampling rates. The number of residual blocks is selected from [1, 2, 3, 4, 5, 6, 7, 8, 9], offering control over the network depth. The filters parameter, determining the number of filters in the initial convolutional layer, ranges from 1 to 100, enabling exploration of model complexity. Learning rates are sampled from a reciprocal distribution between 3e-4 and 3e-2, and the dropout rate is again selected from the set [0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09], facilitating the identification of an optimal configuration to enhance the ResNet model's performance in terms of convergence, accuracy, and generalization for specific regression tasks.

C. FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

Functional Principal Component Analysis (FPCA) is an adaptation of the classical PCA based on the principles of functional data science [37]. The reason for using this strategy is that FPCA has been used for spatial meteorological problems [38]. In this case, the preprocessing is done in two sequential steps: i) first we perform a basis transformation of the original data before applying FPCA and ii) we apply the FPCA itself. Basis representations are periodic functions, and a useful representation for functions that belong (or can be reasonably projected) to the space spanned by a finite set of basis functions. At the end, we obtain a new dataset with a reduced number of features - principal components - that are a linear transformation of the original features. The framework used to perform FPCA analysis was scikit-fda (https://fda.readthedocs.io/en/latest/index.html)

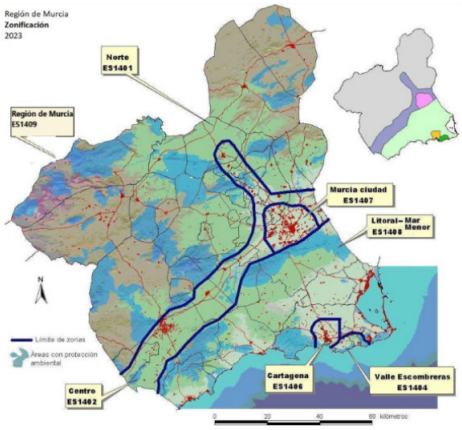


FIGURE 9. Reference air quality zoning at the beginning of 2023 (January). Adapted from SINQLAIR.

D. AGREGATION METHODOLOGY

In this study, the individuals to be classified are the 45 municipalities of the Region of Murcia, and not the grid point of the simulation domain, to produce more interpretable results that are comparable with the current air quality zoning. For this reason, it is necessary to interpolate the grid values to the municipalities. This is done by computing the surface weighted mean of all the grids inside a municipality. This is done after applying FPCA to reduce the computational cost of this process. The framework used to perform the geometric and surface calculation is shapely (https://shapely.readthedocs.io/en/stable/manual.html).

E. CLUSTERING AND HOMOGEINITY

The zoning assessment is done by a clustering analysis using the k-means algorithm suggested by Borge et al. [12]. Cluster analysis is an unsupervised classification technique that gathers individuals or observations with homogeneous characteristics - in this case, municipalities and air quality. This analysis maximises the similarity of the individuals within a group and the difference between groups and has been widely used for clustering classification with various applications in the environmental field [39]. The Elbow rule has chosen the optimal number of clusters using the Silhouette score as a metric, with the restriction that the number of zones should be greater than the current zoning. The framework used to perform k-means and compute metrics was sklearn.

To validate zone consistency, the homogeneity score was computed using Murcia's current air quality zoning as a reference [40]. At the beginning of 2023, the air quality zoning proposed by Murcia public authorities was made available at the SINQLAIR platform (https://sinqlair.carm.es/calidadaire/) (see Figure 9).

In a complementary way, the *evaluomeR* package [41] has been used to compute the optimal number of clusters, allowing to evaluate of the reliability of features by analysing the stability and goodness of the classifications of such metrics. The strength of this method is that we can obtain a detailed metric for each feature and how it contributes to the classification [42]. The clustering computes two metrics:

- Stability: this analysis permits the estimate of whether the clustering is meaningfully affected by small variations in the sample [43]. First, a clustering using the k-means algorithm is carried out. The user can provide the value of k. Then, the stability index is the mean of several bootstrap replicates' Jaccard coefficient [44] values. The values are in the range [0,1] (Unstable: [0, 0.60], Doubtful: [0.60, 0.75], Stable: [0.75, 0.85], Highly Stable: [0.85, 1]).
- Quality: the goodness of the classifications is assessed by validating the clusters generated. For this purpose, we use the Silhouette width as the validity index. This index computes and compares the quality of the clustering outputs found by the different metrics, thus enabling the measurement of the goodness of the classification for both instances and metrics. More precisely, this goodness measurement assesses how similar an instance is to other instances from the same cluster and how dissimilar it is to the rest of the clusters. The average on all the instances quantifies how the instances are appropriately clustered. Kaufman and Rousseeuw [45] suggested the interpretation of the global Silhouette width score as the effectiveness of the clustering structure. The values are in the range [0,1] (There is no substantial clustering structure: [-1, 0.25]; The clustering structure is weak and could be artificial: [0.25, 0.50], There is a reasonable clustering structure: [0.50, 0.70]; A strong clustering structure has been found: [0.70, 1]).

Last, the clustering validation has been performed with the Kruskal-Wallis test, and we defined a clustering score as the $-log_{10}$ of the P-value (P) obtained - it should be taken into account that a P-value under 0.05 indicates a significant difference between the median of the clusters. This clustering score is computed for each pollutant.

$$ClusteringScore = -log_{10}(P)$$

F. COMPUTATIONAL RESOURCES

The CHIMERE-WRF system was run on an OVH server with an Intel(R) Xeon(R) CPU E3-1245 V2 @ 3.40GHz - 8 Cores operating system with 31 GB of RAM and 1 TB of storage. The operating system is Ubuntu 18.04 × 86_64. This machine was used also for the training of the neural network models.

IV. RESULTS

This section provides an overview of the results and analyses performed, organised into several key subsections. In the first subsection, Hyperlocal Measurement Data, we present the empirical data collected from various monitoring stations. This offers insights into real-time air quality conditions in the Region of Murcia Air Quality Network. Next, in the CHIMERE Simulations subsection, we delve into the outcomes of numerical simulations using the CHIMERE-WRF model, shedding light on the spatiotemporal dynamics of air pollutants. Then, Deep Learning Calibration presents the neural network models employed and their performance. Last, the Air Quality Zoning subsection discusses the clustering results and the effect of neural networks on these outcomes.

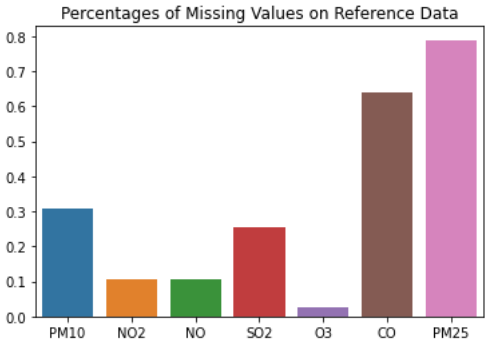


FIGURE 10. Percentage of missing values in the data collected by the Region of murcia air quality network.

A. HYPERLOCAL MEASUREMENT DATA

Figure 10 shows the results of an exploratory analysis of the 2020 historical data collected from the Region of Murcia Air Quality Network. A computation of missing values was done to understand the reliability of the trained models. These results showed a high difference between pollutants. It should be noted that CO and PM2.5 showed a high percentage of missing values - 64 % and 79 %, respectively - while O3 and NO_x presented a high degree of completeness - 2 % and 10 % of missing values, respectively. Last, PM10 and SO2 presented 31 % and 26 % of not imputed measurements, respectively. For this reason, the inclusion of dropout modules in the architectures presented in section III-B was a wise decision because they helped to improve the robustness of the training process in relation to the high percentage of missing values in some datasets.

B. MESOSCALE SIMULATION PERFORMANCE

Table 1 presents the results of a validation study. It consists of several columns, including the name of the gas under evaluation, the Pearson Coefficient, the Determination Coefficient, and the Mean Absolute Error. Each row corresponds to a different gas computed by CHIMERE. The Pearson Coefficient quantifies the linear correlation between the CHIMERE-WRF model and the reference station data, while the Determination Coefficient assesses the goodness of fit. The Mean Absolute Error represents the difference between the model and the reference station values. The data in the

TABLE 1. Validation CHIMERE-WRF against reference station (hourly mean values).

Gas	r	R^2	MAE
NO_2	0.27	0.07	11.0 $\mu g/m^3$
NO	0.09	0.01	4.80 $\mu g/m^3$
SO_2	0.14	0.02	6.56 $\mu g/m^3$
O_3	0.61	0.37	20.8 $\mu g/m^3$
CO	0.17	0.02	0.19 mg/m^3
$PM10$	0.02	0.00	13.9 $\mu g/m^3$
$PM2.5$	0.11	0.01	7.91 $\mu g/m^3$

TABLE 2. Validation CHIMERE-WRF against reference station (daliy mean values).

Gas	r	R^2	MAE
NO_2	0.68	0.47	10.6 $\mu g/m^3$
NO	0.67	0.46	4.78 $\mu g/m^3$
SO_2	0.03	0.01	6.59 $\mu g/m^3$
O_3	0.71	0.52	18.7 $\mu g/m^3$
CO	0.25	0.06	0.19 mg/m^3
$PM10$	0.02	0.01	10.7 $\mu g/m^3$
$PM2.5$	0.04	0.01	5.02 $\mu g/m^3$

table suggests varying levels of agreement between the model and the reference station data, with different gases showing different degrees of correlation and accuracy. For instance, the best performance corresponds to O_3, with $r = 0.61$, $R^2 = 0.37$ and $MAE = 20.8 \mu g/m^3$. The rest of the pollutants show low performance, with correlations under 0.30 and R^2 close to 0.

Table 2 presents the same results computed for daily means, the standard used in the literature to validate these models. In general terms, the correlations with observations are higher for daily values. In contrast, the MAE levels are similar to those obtained in hourly validation. In detail, a better performance for O_3, NO_2 and NO_x is observed, with $r = 0.71$, $r = 0.68$ and $r = 0.67$, respectively. Regarding determination coefficients, the values are $R^2 = 0.52$, $R^2 = 0.47$ and $R^2 = 0.46$ for O_3, NO_2 and NO_x, respectively, with errors of 18.7, 10.6 and 4.78 $\mu g/m^3$. In addition, a slight increase in CO performance is appreciated, with r = 0.25, R2 = 0.06 and MAE = 0.19 mg/m^3.

C. ARCHITECTURE SELECTION

This section compares the performance of the different architectures tested to select the one that fits better for each pollutant. First, Table 3 presents the determination coefficient (R^2) for each one of the architectures and pollutants. It is observed that ResNET presents the best performance for all pollutants except for PM10. In concrete, the computed R^2 values were equal to 0.62, 0.57, 0.20, 0.73, 0.54 and 0.35 for NO_2, NO, SO_2, O_3, CO and PM2.5, respectively. The best model for PM10 was simple neural networks (ANN) with $R^2 = 0.42$. In contrast, the other architectures exhibited similar performance, obtaining the lowest values for LSTM-RNN for SO_2, CO and PM10. For CNN the poorest performance was for NO_2, NO, O_3 and PM2.5.

TABLE 3. Comparison between different architectures (R^2).

Gas	ANN	LSTM-RNN	CNN	ResNET
NO_2	0.41	0.44	0.40	0.62
NO	0.16	0.18	0.10	0.57
SO_2	0.15	0.15	0.16	0.20
O_3	0.69	0.68	0.66	0.73
CO	0.42	0.00	0.34	0.54
$PM10$	0.42	0.16	0.18	0.36
$PM2.5$	0.23	0.27	0.15	0.35

TABLE 4. Comparison between different architectures (MAE). The units are $\mu g/m^3$, except for CO that are expresed in mg/m^3.

Gas	ANN	LSTM-RNN	CNN	ResNET
NO_2	7.51	7.35	7.22	6.51
NO	5.26	4.13	4.16	2.71
SO_2	3.02	2.90	2.98	2.92
O_3	13.0	13.7	12.9	12.2
CO	0.32	0.32	0.32	0.32
$PM10$	9.34	11.4	10.7	9.81
$PM2.5$	5.74	5.72	6.40	5.69

In addition, similar results are obtained if we use the mean absolute error (MAE) as metric. Table4 illustrates the error values for each one of the architectures. Again, the best results were obtained for NO_2, NO, SO_2, O_3 and PM2.5 with MAE scores equal to 6.51, 2.71, 2.92, 12.2, 9.81 and 5.69 $\mu g/m^3$. The ANN architecture fitted better for PM10, with $MAE = 9.34 \mu g/m^3$. For CO, all the architectures presented the same values - 0.32 mg/m^3 - probably because the precision in the simulation was lower - CO is in the milligrams order of magnitude. However, for MAE the poorest performance was obtained by the ANN architectures for NO_2, NO, SO_2 and O_3 and PM2.5 for the CNN architecture.

D. ACCURACY OF NEURAL NETWORK MODELS

Table 5 shows the results of a hyperparameter tuning process for the neural networks used to predict various gas concentrations. The hyperparameters included are the ones described by the ResNET architecture including the number of blocks, filters, kernel size, learning rate and stride for each model corresponding to the pollutants NO_2, NO, SO_2, O_3, CO and $PM2.5$. As highlighted in Table 3 and Table 4, the selected architecture for PM10 is ANN, and not ResNET. For this reason, the hyperparameters used are different, in concrete alpha, drop rate, neurons, and learning rate. Different configurations for each model are observed, highlighting that each pollutant should be modelled separately.

Table 6 assesses the performance of CHIMERE-WRF-NN in modelling various air quality parameters when compared to data from ground measurements in concrete hourly mean values. The Pearson Coefficient represents the strength and direction of the linear correlation with values ranging from 0.45 to 0.85, being observed as the best performance for O_3, NO and NO_2. With similar trends, the Determination Coefficient (R-squared), indicating the proportion of variance explained, varying from 0.20 to 0.73, and the Mean Absolute

TABLE 5. Hyperparameters selected for each model.

Model	Blocks	Filters	Kernel Size	Learning Rate	Stride
NO_2	9	64	3	0.002	5
NO	3	19	9	0.003	1
SO_2	1	98	96	0.003	4
O_3	9	67	3	0.0009	5
CO	2	98	5	0.002	7
$PM2.5$	5	40	7	0.003	5
Model	**Alpha**	**Drop Rate**	**Neurons**	**Learning Rate**	-
$PM10$	0.1	0.01	66	0.0005	-

TABLE 6. Results of the comparison CHIMERE-WRF + neural networks versus reference station (hourly mean values).

Gas	r	R^2	MAE
NO_2	0.79	0.62	6.51 $\mu g/m^3$
NO	0.76	0.57	2.71 $\mu g/m^3$
SO_2	0.45	0.20	2.92 $\mu g/m^3$
O_3	0.85	0.73	12.2 $\mu g/m^3$
CO	0.73	0.54	0.32 mg/m^3
$PM10$	0.64	0.42	9.81 $\mu g/m^3$
$PM2.5$	0.59	0.35	5.69 $\mu g/m^3$

TABLE 7. Results of the comparison CHIMERE-WRF + neural networks versus reference station (daily mean values).

Gas	r	R^2	MAE
NO_2	0.94	0.89	2.17 $\mu g/m^3$
NO	0.95	0.90	0.72 $\mu g/m^3$
SO_2	0.55	0.30	0.74 $\mu g/m^3$
O_3	0.97	0.94	4.87 $\mu g/m^3$
CO	0.93	0.86	0.11 mg/m^3
$PM10$	0.87	0.76	3.59 $\mu g/m^3$
$PM2.5$	0.71	0.50	1.87 $\mu g/m^3$

Error (MAE), reflecting prediction accuracy, with values between 2.71 $\mu g/m^3$ and 12.2 $\mu g/m^3$, as well as 0.32 mg/m^3 for CO. These metrics collectively offer insights into the neural network performance, revealing varying degrees of correlation and accuracy for different gases but a general improvement compared to results without neural networks.

Regarding daily mean values (Table 7), neural networks also improves the results, and the highest scores are obtained. In concrete, this improvement is quite significant in the case of MAE. The performance for O_3 is quite good, with $r = 0.97$, $R^2 = 0.94$ and $MAE = 4.87 \mu g/m^3$. The same occurs for NO_2 -$r = 0.94$ and $R^2 = 0.89$-, NO - $r = 0.95$ and $R^2 = 0.90$ -,PM10 - $r = 0.87$ and $R^2 = 0.76$ - and CO - $r = 0.93$ and $R^2 = 0.86$. On the other hand, PM2.5 and SO_2 present moderate performance with $R^2 = 0.50$ and $R^2 = 0.30$, respectively.

One interesting result is that the effect of neural networks provides valuable insights into the effect of correlations between predictors. It is important to note that clustering algorithms work better when fed by non-correlated predictors. This is why clustering is normally coupled with PCA strategies as preprocessors. In this case, Figure 11 shows the correlation matrix between CHIMERE predictors

TABLE 8. Optimal *k* value according to evalueomer methodology. on the top, values without neural networks and in the bottmo values with neural networks.

Metric	k (OS)	Stability	Quality	k(OQ)	Stability	Quality	k
PCA 1	6	0.78	0.66	10	0.68	0.68	5
PCA 2	5	0.82	0.70	5	0.82	0.70	6
PCA 3	6	0.76	0.67	6	0.76	0.67	6
PCA 4	6	0.89	0.68	6	0.89	0.68	6
PCA 5	6	0.77	0.57	11	0.68	0.58	6
Metric	k (OS)	Stability	Quality	k(OQ)	Stability	Quality	k
PCA 1	7	0.71	0.62	7	0.71	0.62	7
PCA 2	5	0.90	0.66	5	0.90	0.66	5
PCA 3	7	0.77	0.59	11	0.69	0.63	7
PCA 4	5	0.91	0.68	6	0.81	0.69	6
PCA 5	5	0.75	0.58	10	0.68	0.60	5

before and after applying neural networks. It is observed that before applying neural networks, there are strong correlations between some predictors, especially particles - $r = 0.93$ between PM10 and PM2.5 - and some gases - $r = 0.63$ between NO_2 and O_3. In contrast, the correlations between predictors decrease after applying neural networks in concrete for particles with a $r = 0.082$ between PM10 and PM2.5.

E. AIR QUALITY ZONING

The results regarding Air Quality Zoning were obtained after classifying each of the municipalities in the function of the simulated values with CHIMERE. The first step was computing the optimal number of clusters according to the Silhouette Score, a process shown in Figure 7, for data with and without neural network calibration. Regarding the results provided by the *evaluomeR* method, the main conclusions provide an optimal number of clusters equal to 5-6 with and without neural networks, and the clustering performance is quite similar. The results are presented in detail in Table 7, where $k(OS)$ is the optimal number of clusters according to the stability score, $k(OQ)$ is the optimal number of clusters according to the quality score, and k is the global optimal number of clusters. We obtained an optimal number of clusters of 5, 6, 6, 6 and 6 for PCA 1, PCA 2, PCA 3, PCA 4 and PCA 5, respectively (without neural networks), and 7, 5, 7, 6 and 5 for PCA 1, PCA 2, PCA 3, PCA 4 and PCA 5, respectively (with neural networks).

In this context, the number of optimal clusters was selected by performing a qualitative analysis of the *stability* and the *quality* according to the ranges defined in the methodology section. It is true that the stability and cluster quality slightly improve after applying neural networks, except for PCA 2, achieving a transition from a *stable cluster* to a *highly stable cluster*, and for PCA 5 with a transition from *stable cluster* to *doubtful*. On the other hand, the qualitative interpretation of quality does not change as all the metrics fits in the stable cluster category and there is a reasonable cluster structure - it can be discussed that PCA 2 shows lower qualitative values at its limit without neural networks. Thus, the optimal k value for all metrics - despite PCA 5 - is k = 5, k = 6 or

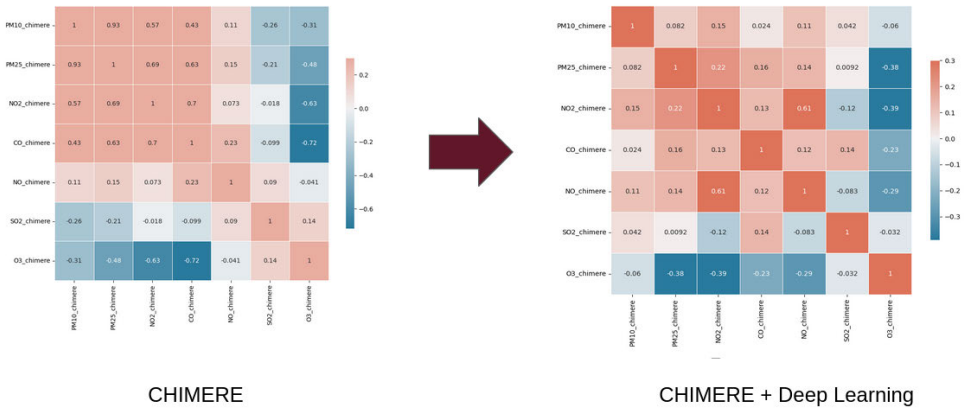


FIGURE 11. Correlation according to Pearson test between CHIMERE features before and after applying neural networks.

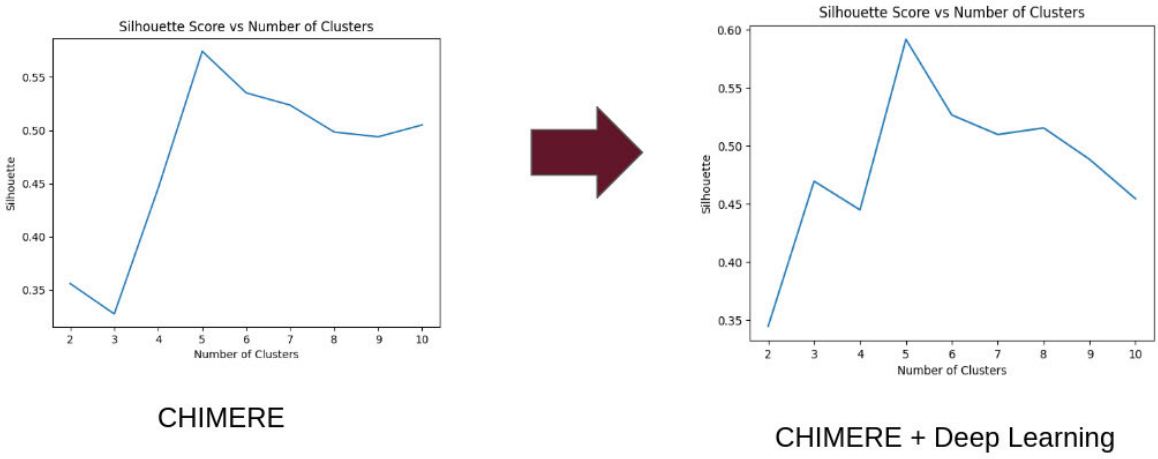


FIGURE 12. Determination of the optimal number of clusters for data without neural networks and with neural networks.

k = 7. However, the selected value was k=6 to obtain the number of clusters of the current zoning, as the qualitative interpretation of the cluster does not change. Without neural networks, the Silhouette Score equals 0.53, and the Davies Bouldin Score equals 0.60. On the other hand, with neural networks, we obtained a Silhouette Score equal to 0.54 and a Davies Bouldin Score equal to 0.60, as shown in Figure 12.

According to the results provided by both methods, the clustering is computed for k = 6. This is done for the simulations without and with neural networks, as shown in Figure 13. In each case, six groups with a similar geographical distribution are doneneural. In addition, the homogeneity of the clusters increased after applying neural networks, as is observed in the Tukey post hoc analysis values.

Assessing the performance of a K-means clustering algorithm using boxplots is a technique for understanding clustering consistency. Boxplots allow analysis of the distribution of distances between data points and their respective cluster centroids, gaining insights into the dispersion and separation of clusters. A well-defined clustering solution will exhibit boxplots with tight, compact boxes and small whiskers, indicating that most data points are close to their centroids and the clusters are distinct. Figure 14 presents the boxplots for each one of the principal components among the six clusters. Differences between the medians in all components were observed except for PCA 4, where there was more similarity between the clusters. In addition, there were outliers in PCA 1, PCA 3, PCA 4 and PCA 5 in different clusters.

In contrast, Figure 15 presents the results with neural networks, and it shows differences in the distribution in PCA 2 and a reduction in the number of outliers in PCA 3 and PCA 5. The rest of the trends are similar, with similarities between clusters in PCA 4 and apparent differences between medians in all the components.

A larger difference is highlighted when quantifying clustering difference through Kruskal-Wallis tests. In all cases (without and with neural networks), the P-value obtained is lower than 0.05 for all components, indicating a significant

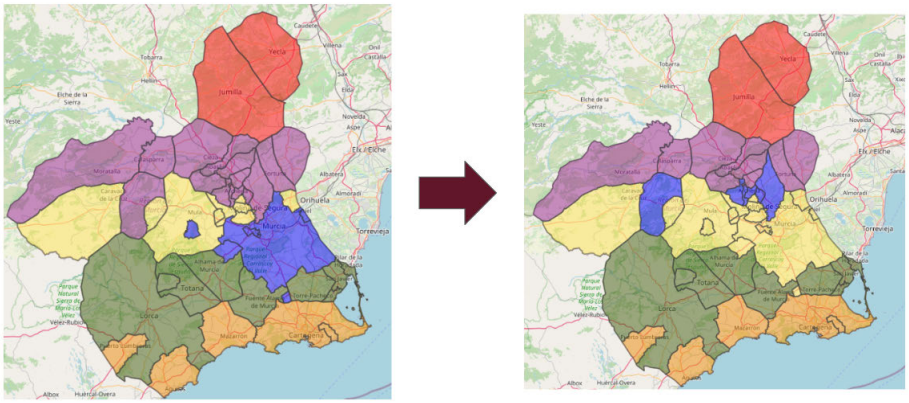


FIGURE 13. Comparison of the Zoning final results without neural networks and with neural networks.

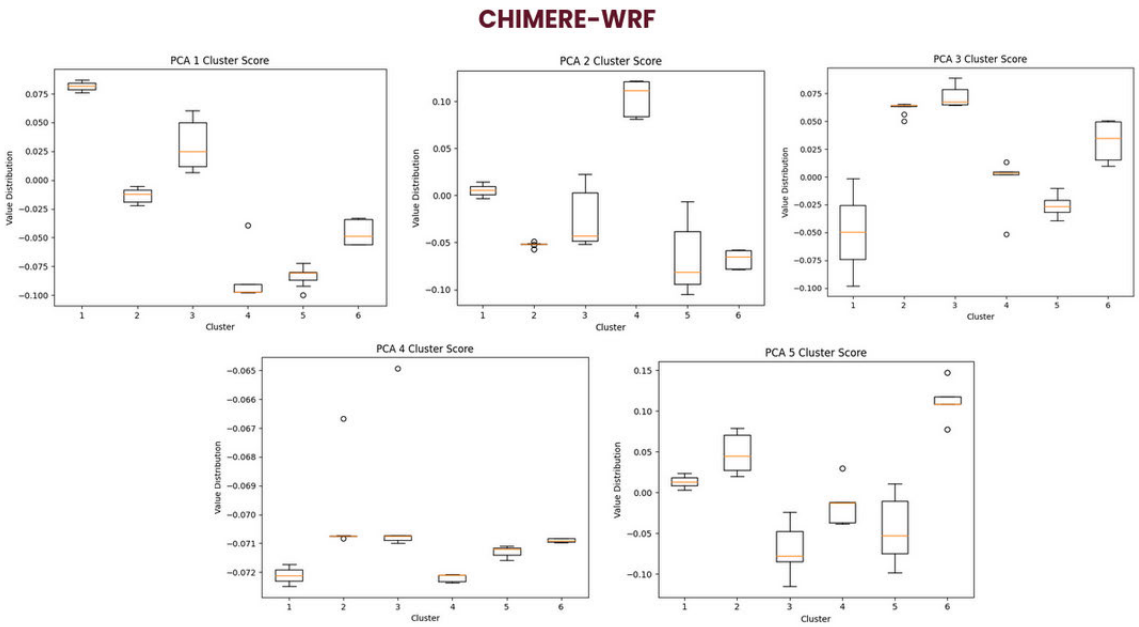


FIGURE 14. Distribution of boxplots of each principal component among different clusters without neural networks.

difference, at least between the two clusters. The neural networks approach achieves better results regarding cluster score, defined as the -log(p-value), in concrete in PCA 2 and PCA 6, indicating that the clusters are more independent than with the classical approach (without neural networks), as observed in Table 9. Similar results were observed for the H-Score, with a higher value for the neural networks approach.

TABLE 9. Validation CHIMERE-WRF + neural networks against reference station (daily mean values).

Component	H-Score	H-Score (NN)	Cluster Score	Cluster Score (NN)
PCA 1	40.41	41.29	6.91	7.08
PCA 2	30.30	37.56	4.89	6.33
PCA 3	39.43	35.00	6.71	5.82
PCA 4	34.98	34.20	5.82	5.65
PCA 5	34.17	39.06	5.66	6.63

V. DISCUSSION

Regarding air quality zoning, recent literature studies have proposed automatic approaches based on mathematical and statistical methods [12]. In this context, this paper has taken this methodology and has extended it by proposing neural networks as an additional step to calibrate air quality simulations. The results are promising and open new research lines in this recent topic.

First, we have stated that the main hypothesis of this work is that neural networks would improve the performance of air quality zoning strategies. Considering the results presented, we consider that the hypothesis is demonstrated,

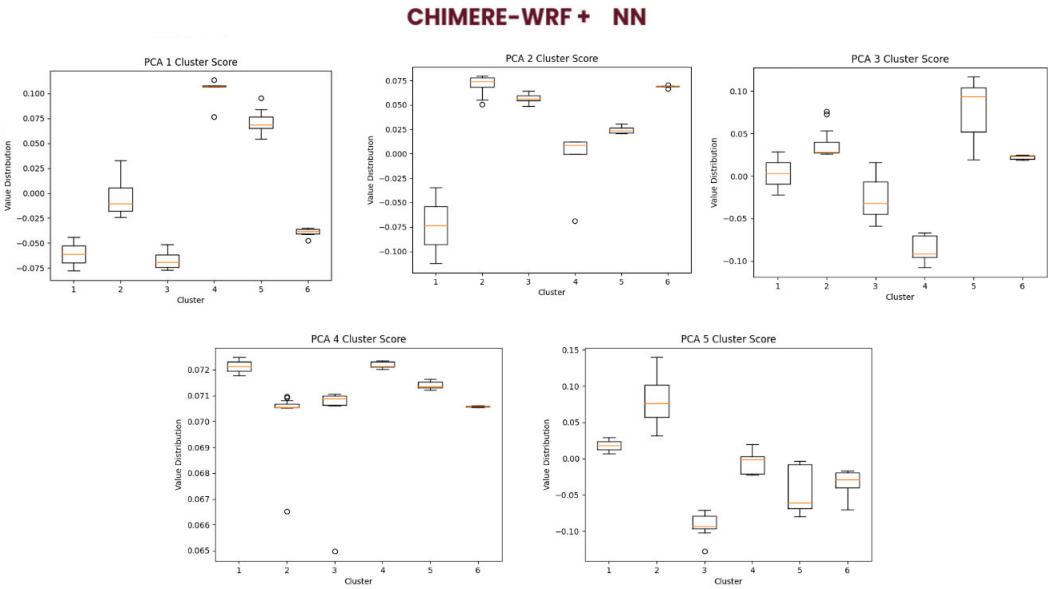


FIGURE 15. Distribution of boxplots of each principal component among different clusters with neural networks.

as the clustering proposed in Figure 7 is more coherent with official air quality zoning and presents a better score in terms of homogeneity - a better cluster score for all the components. This draws new scenarios for neural networks in deep learning applied for air quality. Currently, the most important approach for neural networks is for models and ground measurements calibration forecasting [46], [47], but understanding the role of neural networks in dimensional reduction is needed. Thus, we consider that the role of Convolutional Neural Networks (CNN) in air quality zoning should be studied in future steps [48].

In addition, it is also important to compare our approach based on deep learning with state of the art solutions. The most similar approach is the one proposed in [12], which applies an air quality zoning in Madrid. The main strength of our solution in comparison to this one is that we apply deep learning techniques to improve data quality of chemistry transport models before applying clustering. The air quality zoning applied in Madrid is based on traditional classification techniques, which does not correct the uncertainty associated with air quality dispersion models [12], [16]. However, there are approaches in the state of the art that use ANN for land classification problems, especially in the field of land cover [49]. There, all the classification work relies on the ANN, and no clustering techniques are applied. This has the limitation of understanding, because clustering is an explainable model. This is critical in air quality zoning, in which an interpretation of the clustering results is needed. In contrast, the problems related to traditional clustering algorithms are still present in our approach, because of the local minima problem in the clustering step. The local minima problem is the convergence to a minimum number of clusters producing counterintuitive results. As our solution combines ANN for calibration with k-means, the solution could converge to a non-optimal solution. This idea strengthens the need for human interpretation of the clustering results [50].

Furthermore, the results proposed in this study provide valuable insights that can be used to review the current air quality network in the Region of Murcia. For this reason, we propose a new air quality zoning based on the clusters obtained after applying neural networks, following the recommendation of functional air quality zoning and the European regulation [13]. This proposal is reflected in Figure 16, in which we can compare the current zoning - proposed by MITECO - and new zoning - obtained after applying clustering to the neural networks results. The main difference is the creation of a new zone in the north of the region, with a new monitoring point. This is quite important because, at this moment, the ES1401 (yellow) covers more than 50 % of the territory with a unique monitoring point, so it is impossible to define all the exposure in this zone [51]. Thus, the new zoning is based on splitting this huge area into two different ones and proposing a new monitoring point in Cieza or Jumilla. However, this zoning assessment should be done with environmental experts, urban planners and public authorities to consider all the important factors.

In contrast, this study presents some limitations that should be addressed in future iterations. The first one is the accuracy of the base simulations, which is moderate for O_3 and NO_x and poor for the rest of the pollutants, as observed in Table 10. If we compare these values with the validation performed by CHIMERE developers [11], the performance is similar in literature for O_3 - 0.71 in our validation versus 0.75 in Paris - and NO_2 - 0.68 in our validation versus 0.51 in Paris.

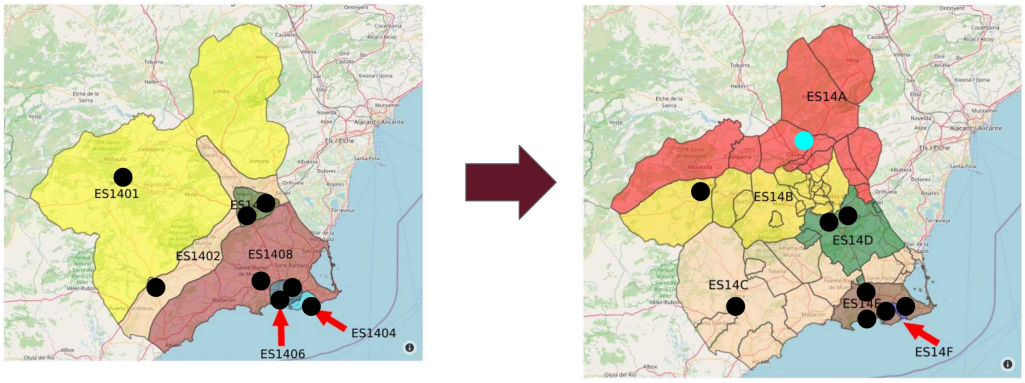


FIGURE 16. New proposed zoning and air quality station for the Region of Murcia Air Quality Network. On the left is the proposed zoning by MITECO with the official codes, and on the right, the zoning is proposed using the automatic methodology with new assigned codes. In blue is the proposed monitoring point.

TABLE 10. CHIMERE performance versus state of the art [11].

Gas	CHIMERE STOA v2017	CHIMERE STOA v2020	CHIMERE v2020	CHIMERE v2020 + NN
NO_2	0.51	0.51	0.68	0.95
O_3	0.75	0.75	0.71	0.97
$PM10$	0.33	0.50	0.02	0.87
$PM2.5$	0.50	0.66	0.04	0.71

This comparison has been performed for daily mean values, as there are no references for hourly values in [11], and only for the gases validated in this study - NO, SO_2 and CO are omitted in [11] as they argue the performance of CHIMERE is not good enough. We obtained in our validation $r = 0.67$, $r = 0.25$ and $r = 0.02$ for NO, CO and SO_2, supporting the CHIMERE developer's assessment.

These results are interesting because we provided an independent validation of the CHIMERE model, similar to the one presented in the literature in the release of the v2020 model for O_3 and NO_2, but poorer for particles. There are two reasons for this performance: i) the CHIMERE model has been only validated and developed mainly in the Île de France region, so the performance might be different in other locations, and ii) the quality of the data collected by the Region of Murcia Air Quality Network is lower in comparison with the Paris network used in terms of number of devices and missing values, as will be discussed later. As the percentage of missing values were high for particles - 30 % for PM10 and 79 % for PM2.5 -, the validation of these two pollutants should not be considered significant.

The performance changes after applying neural networks calibration, obtaining better performance for all pollutants, increasing all the values of r and R^2 and improving the performance reported by CHIMERE [11], as indicated in Table 6. For O_3, a Pearson Coefficient of 0.97 is obtained, improving our performance by 37 and official performance by 29 %. The same occurs for NO_2, improving our performance by 40 % and official performance by 86 % with an $r = 0.95$. In contrast, for PM10, the performance is $r = 0.87$ and for PM2.5, $r = 0.71$, but the elevated number of missing values indicates that it is necessary to be careful in interpreting these results. NO and CO present good correlations - $r = 0.95$ and $r = 0.93$, respectively - but the number of missing values for CO is around 60 %. At the same time, for NO, it is possible to infer an improvement as the time series for training is complete. For SO_2, the performance is improved after neural network calibration.

The improvement of the correlations against hyperlocal data brings valuable practical insights into our proposal. The first one is the reliability of the simulated data, because a higher correlation means that the trends in concentration values are closer to the real ones [11]. This is important in the context of air quality zoning, because a high correlation generates clusters closer to reality if the concentration values are more accurate. This is an improvement of the solution for the land classification problem, because the predicted zones are more similar to the real distribution over the Region of Murcia, meaning better zoning. In addition, accurate models can provide simulations closer to the real measurements, allowing to estimate air quality concentrations in locations where it is not possible to install sensors or reference stations, improving the resolution of the network [52]. Last, better data increases citizens' trustability in air quality measurements, a key aspect in air quality regulations, allowing them to gain insights into the air quality in individual countries, regions and cities, and become more engaged regarding climate change mitigation [53].

Regarding the comparison with other neural networks architecture present in the state of the art, our solution improves for NO_2, where an $R^2 = 0.75$ has been obtained using LSTM-GRU compared to a $R^2 = 0.89$ for ResNET [23]. In contrast, there are better performances for SO_2 in the literature using RNN with $R^2 = 0.56 - 0 - 62$ [26]. Regarding CNN, they have been used to forecast air qulity index with a $R^2 = 0.71$ [54]. One of the key applications of CNN is particulate matter modelling, with Pearson coefficients ranging between $0.67 - 0.93$, compared to the 0.87 obtained for PM10 in our approach [55].

Last, convolutional networks (CNN) have been used for O_3 forecasting obtaining Pearson coefficient ranging from 0.74-0.80, while our ResNET model provide a correlation of 0.85 and 0.97 for hourly and daily means [56].

However, it is necessary to analyse this performance in detail for the following reasons: i) the neural networks have been trained with data from Murcia, so the replicability of the models developed in other regions is not guaranteed; ii) the number of measurement points is low and with a non-homogeneous distribution among the domain; and iii) the number of missing values presented in some pollutants time series. These overfitting or unrealistic estimations are deduced from the correlations shown in Figure 5 after applying neural networks, with low correlations between particles and other linked pollutants, suggesting artificial values - good for clustering purposes but not for real modelling goals. For this reason, the following steps should be addressed in future research: i) expand the area of study to a wider area with more monitoring points, ensuring a homogeneous dataset with low missing values and a significant number of ground hyperlocal stations over all the territory to be studied and ii) design different architectures for calibration and dimensions processing in the function of the application and goals of the study.

Finally, a remark should be made regarding the quality of the measurements. It is easy to note that hyperlocal data plays a key role in this methodology, essential to guarantee the completeness of the data. Some of the results proposed in this paper are based on the data provided by public authorities - with special emphasis on the training of neural networks. Thus, the proposed methodology should be validated with more complete data in other locations, especially concerning particles where the number of missing values is around 80 % for PM2.5. To mitigate this, there are approaches based on time series reconstruction using GAN models [57] or LSTMN to improve resolution through data fusion [58]. However, the idea is to act on the measurements. This can also be highlighted as an insight of the paper, and future research should dedicate part of the effort to providing high-quality measurements [52].

VI. CONCLUSION

In this work, we have described and evaluated an automatic air quality zoning methodology in the Region of Murcia, demonstrating that the use of neural networks combined with the CHIMERE-WRF model improves the homogeneity and the consistency of the clusters computed and compared with the current air quality zoning proposed by public authorities. Moreover, the results provide important insights and action points for the Region of Murcia Air Quality Network, highlighting the need for a new air quality zone and a new measurement point covering the region's north. Additionally, we have shown that neural networks improves the correlations of simulations compared to hyperlocal measurements.

However, the results obtained open new research possibilities and highlight different limitations that should be addressed in different studies: i) expand the area of study to collect more geographically representative data; ii) understand the replication of the neural network models in other locations as they have been trained with data from Murcia and iii) improve the quality of the reference datasets provided by public authorities, especially regarding the number of missing values for some pollutants. On the other hand, more research is needed in the field of automatic air quality zoning to understand the role of neural networks in dimensionality reduction and provide different architectures for calibration and preprocessing. Last, this methodology should include interaction and expert feedback to include human knowledge in the interpretation and zoning assessment.

REFERENCES

[1] A. Keswani, H. Akselrod, and S. C. Anenberg, "Health and clinical impacts of air pollution and linkages with climate change," *NEJM Evidence*, vol. 1, no. 7, Jun. 2022, Art. no. EVIDra2200068, doi: 10.1056/evidra2200068.

[2] A. J. Cohen et al., "Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the global burden of diseases study 2015," *Lancet*, vol. 389, no. 10082, pp. 1907–1918, May 2017, doi: 10.1016/S0140-6736(17)30505-6.

[3] A. C. Rai, P. Kumar, F. Pilla, A. N. Skouloudis, S. Di Sabatino, C. Ratti, A. Yasar, and D. Rickerby, "End-user perspective of low-cost sensors for outdoor air pollution monitoring," *Sci. Total Environ.*, vols. 607–608, pp. 691–705, Dec. 2017, doi: 10.1016/j.scitotenv.2017.06.266.

[4] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014, doi: 10.1109/JIOT.2014.2306328.

[5] R. Ignaccolo, S. Ghigo, and S. Bande, "Functional zoning for air quality," *Environ. Ecological Statist.*, vol. 20, no. 1, pp. 109–127, Mar. 2013, doi: 10.1007/s10651-012-0210-7.

[6] R. P. Haining, *Spatial Data Analysis: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[7] E. Lezina and M. Misyurev, "Ambient air quality measurements in a large city: Existing solutions, new opportunities and challenges," *J. Phys. Conf. Ser.*, vol. 2192, no. 1, Mar. 2022, Art. no. 012031.

[8] P. Maranzano, "Air quality in Lombardy, Italy: An overview of the environmental monitoring system of ARPA Lombardia," *Earth*, vol. 3, no. 1, pp. 172–203, Feb. 2022, doi: 10.3390/earth3010013.

[9] P. Sicard, E. Agathokleous, A. De Marco, E. Paoletti, and V. Calatayud, "Urban population exposure to air pollution in Europe over the last decades," *Environ. Sci. Eur.*, vol. 33, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/s12302-020-00450-2.

[10] L. Menut, B. Bessagnet, D. Khvorostyanov, M. Beekmann, N. Blond, A. Colette, I. Coll, G. Curci, G. Foret, A. Hodzic, S. Mailler, F. Meleux, J.-L. Monge, I. Pison, G. Siour, S. Turquety, M. Valari, R. Vautard, and M. G. Vivanco, "CHIMERE 2013: A model for regional atmospheric composition modelling," *Geosci. Model Develop.*, vol. 6, no. 4, pp. 981–1028, Jul. 2013, doi: 10.5194/gmd-6-981-2013.

[11] L. Menut, B. Bessagnet, R. Briant, A. Cholakian, F. Couvidat, S. Mailler, R. Pennel, G. Siour, P. Tuccella, S. Turquety, and M. Valari, "The CHIMERE v2020r1 online chemistry-transport model," *Geosci. Model Develop.*, vol. 14, no. 11, pp. 6781–6811, Nov. 2021.

[12] R. Borge, D. Jung, I. Lejarraga, D. de la Paz, and J. M. Cordero, "Assessment of the Madrid region air quality zoning based on mesoscale modelling and k-means clustering," *Atmos. Environ.*, vol. 287, Oct. 2022, Art. no. 119258, doi: 10.1016/j.atmosenv.2022.119258.

[13] M. Doval-Miñarro and M. C. Bueso, "A comparative study of air pollutant concentrations before the COVID-19 pandemic and in the new normal in the Región de murcia (Spain)," *Atmosphere*, vol. 14, no. 1, p. 147, Jan. 2023, doi: 10.3390/atmos14010147.

[14] M. Gavrilescu, "Theoretical predictive air quality models," *The Quality of Air*, vol. 72, pp. 97–115, May 2016, doi: 10.1016/bs.coac.2016.03.019.

[15] B. Bessagnet, M. Beauchamp, L. Menut, R. Fablet, E. Pisoni, and P. Thunis, "Deep learning techniques applied to super-resolution chemistry transport modeling for operational uses," *Environ. Res. Commun.*, vol. 3, no. 8, Aug. 2021, Art. no. 085001, doi: 10.1088/2515-7620/ac17f7.

[16] M. Schaap, C. Cuvelier, C. Hendriks, B. Bessagnet, J. M. Baldasano, A. Colette, P. Thunis, D. Karam, H. Fagerli, A. Graff, R. Kranenburg, A. Nyiri, M. T. Pay, L. Rouïl, M. Schulz, D. Simpson, R. Stern, E. Terrenoire, and P. Wind, "Performance of European chemistry transport models as function of horizontal resolution," *Atmos. Environ.*, vol. 112, pp. 90–105, Jul. 2015, doi: 10.1016/j.atmosenv.2015.04.003.

[17] D. Iskandaryan, F. Ramos, and S. Trilles, "Graph neural network for air quality prediction: A case study in Madrid," *IEEE Access*, vol. 11, pp. 2729–2742, 2023, doi: 10.1109/ACCESS.2023.3234214.

[18] C. A. Keller and M. J. Evans, "Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-chem chemistry model v10," *Geosci. Model Develop.*, vol. 12, no. 3, pp. 1209–1225, Mar. 2019, doi: 10.5194/gmd-12-1209-2019.

[19] Q. Liao, M. Zhu, L. Wu, X. Pan, X. Tang, and Z. Wang, "Deep learning for air quality forecasts: A review," *Current Pollut. Rep.*, vol. 6, no. 4, pp. 399–409, Dec. 2020, doi: 10.1007/s40726-020-00159-z.

[20] J. Wang, L. Jin, X. Li, S. He, M. Huang, and H. Wang, "A hybrid air quality index prediction model based on CNN and attention gate unit," *IEEE Access*, vol. 10, pp. 113343–113354, 2022, doi: 10.1109/ACCESS.2022.3217242.

[21] S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, and M. N. Asghar, "Comparative analysis of machine learning techniques for predicting air quality in smart cities," *IEEE Access*, vol. 7, pp. 128325–128338, 2019, doi: 10.1109/ACCESS.2019.2925082.

[22] G. Huang, X. Li, B. Zhang, and J. Ren, "PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition," *Sci. Total Environ.*, vol. 768, May 2021, Art. no. 144516, doi: 10.1016/j.scitotenv.2020.144516.

[23] G. Cican, A.-N. Buturache, and R. Mirea, "Applying machine learning techniques in air quality prediction—A bucharest city case study," *Sustainability*, vol. 15, no. 11, p. 8445, May 2023, doi: 10.3390/su15118445.

[24] V. Athira, P. Geetha, R. Vinayakumar, and K. P. Soman, "DeepAirNet: Applying recurrent networks for air quality prediction," *Proc. Comput. Sci.*, vol. 132, pp. 1394–1403, 2018, doi: 10.1016/j.procs.2018.05.068.

[25] P. D. Ivatt and M. J. Evans, "Improving the prediction of an atmospheric chemistry transport model using gradient-boosted regression trees," *Atmos. Chem. Phys.*, vol. 20, no. 13, pp. 8063–8082, Jul. 2020, doi: 10.5194/acp-20-8063-2020.

[26] A. Vlasenko, V. Matthias, and U. Callies, "Simulation of chemical transport model estimates by means of a neural network using meteorological data," *Atmos. Environ.*, vol. 254, Jun. 2021, Art. no. 118236, doi: 10.1016/j.atmosenv.2021.118236.

[27] Q. Di, P. Koutrakis, and J. Schwartz, "A hybrid prediction model for $PM_{2.5}$ mass and components using a chemical transport model and land use regression," *Atmos. Environ.*, vol. 131, pp. 390–399, Apr. 2016, doi: 10.1016/j.atmosenv.2016.02.002.

[28] P. Thunis, A. Miranda, J. M. Baldasano, N. Blond, J. Douros, A. Graff, S. Janssen, K. Juda-Rezler, N. Karvosenoja, G. Maffeis, A. Martilli, M. Rasoloharimahefa, E. Real, P. Viaene, M. Volta, and L. White, "Overview of current regional and local scale air quality modelling practices: Assessment and planning tools in the EU," *Environ. Sci. Policy*, vol. 65, pp. 13–21, Nov. 2016, doi: 10.1016/j.envsci.2016.03.013.

[29] T. Hussain and H. Shouno, "Explainable deep learning approach for multi-class brain magnetic resonance imaging tumor classification and localization using gradient-weighted class activation mapping," *Information*, vol. 14, no. 12, p. 642, Nov. 2023, doi: 10.3390/info14120642.

[30] D. Hussain, T. Hussain, A. A. Khan, S. A. A. Naqvi, and A. Jamil, "A deep learning approach for hydrological time-series prediction: A case study of gilgit river basin," *Earth Sci. Informat.*, vol. 13, no. 3, pp. 915–927, Sep. 2020, doi: 10.1007/s12145-020-00477-2.

[31] I. Hussain, K. P. Sinaga, and M.-S. Yang, "Unsupervised multiview fuzzy C-means clustering algorithm," *Electronics*, vol. 12, no. 21, p. 4467, Oct. 2023, doi: 10.3390/electronics12214467.

[32] S. Talukdar, P. Singha, S. Mahato, S. Pal, Y.-A. Liou, and A. Rahman, "Land-use land-cover classification by machine learning classifiers for satellite observations—A review," *Remote Sens.*, vol. 12, no. 7, p. 1135, Apr. 2020, doi: 10.3390/rs12071135.

[33] J. Liu and J. Xing, "Identifying contributors to $PM_{2.5}$ simulation biases of chemical transport model using fully connected neural networks," *J. Adv. Model. Earth Syst.*, vol. 15, no. 2, Feb. 2023, Art. no. e2021MS002898, doi: 10.1029/2021ms002898.

[34] X. Li, X. Meng, X. Ji, J. Zhou, C. Pan, and N. Gao, "Zoning technology for the management of ecological and clean small-watersheds via k-means clustering and entropy-weighted TOPSIS: A case study in Beijing," *J. Cleaner Prod.*, vol. 397, Apr. 2023, Art. no. 136449, doi: 10.1016/j.jclepro.2023.136449.

[35] R. Atkinson, D. L. Baulch, R. A. Cox, R. F. Hampson, J. A. Kerr, M. J. Rossi, and J. Troe, "Evaluated kinetic, photochemical and heterogeneous data for atmospheric chemistry: Supplement V. IUPAC subcommittee on gas kinetic data evaluation for atmospheric chemistry," *J. Phys. Chem. Reference Data*, vol. 26, no. 3, pp. 521–1011, May 1997, doi: 10.1063/1.556011.

[36] P. Kubben, A. Dekker, and M. Dumontier, *Fundamentals of Clinical Data Science*. Cham, Switzerland: Springer, 2019.

[37] H. L. Shang, "A survey of functional principal component analysis," *AStA Adv. Stat. Anal.*, vol. 98, no. 2, pp. 121–142, Apr. 2014.

[38] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202.

[39] R. Borge, J. Lumbreras, S. Vardoulakis, P. Kassomenos, and E. Rodriguez, "Analysis of long-range transport influences on urban PM_{10} using two-stage atmospheric trajectory clusters," *Atmos. Environ.*, vol. 41, no. 21, pp. 4434–4450, Jul. 2007, doi: 10.1016/j.atmosenv.2007.01.053.

[40] A. Jaeger and D. Banks, "Cluster analysis: A modern statistical review," *WIREs Comput. Statist.*, vol. 15, no. 3, May 2023, doi: 10.1002/wics.1597.

[41] J. Bernabé-Díaz, M. Franco-Nicolás, J. M. Vivo-Molina, M. Quesada-Martínez, A. Duque-Ramos, and J. T. Fernández-Breis, "Evaluomer: Evaluation of bioinformatics metrics," *Bioconductor*, 2023.

[42] M. Franco, J. M. Vivo, M. Quesada-Martínez, A. Duque-Ramos, and J. T. Fernández-Breis, "Evaluation of ontology structural metrics based on public repository data," *Briefings Bioinf.*, vol. 21, no. 2, pp. 473–485, Mar. 2020, doi: 10.1093/bib/bbz009.

[43] G. W. Milligan and R. Cheng, "Measuring the influence of individual data points in a cluster analysis," *J. Classification*, vol. 13, no. 2, pp. 315–335, Sep. 1996, doi: 10.1007/bf01246105.

[44] P. Jaccard, "Distribution de La flore Alpine dans le Bassin des Dranses etdans quelques régions voisines," *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 241–272, 1901.

[45] A. S. Hadi, L. Kaufman, and P. J. Rousseeuw, "Finding groups in data: An introduction to cluster analysis," *Technometrics*, vol. 34, no. 1, p. 111, Feb. 1992.

[46] H. A. D. Nguyen and Q. P. Ha, "Wireless sensor network dependable monitoring for urban air quality," *IEEE Access*, vol. 10, pp. 40051–40062, 2022, doi: 10.1109/ACCESS.2022.3166904.

[47] M. M. Rahman, M. Shafiullah, S. M. Rahman, A. N. Khondaker, A. Amao, and M. H. Zahir, "Soft computing applications in air quality modeling: Past, present, and future," *Sustainability*, vol. 12, no. 10, p. 4045, May 2020, doi: 10.3390/su12104045.

[48] Y. Zhang, H. G. Soon, D. Ye, J. Y. H. Fuh, and K. Zhu, "Powder-bed fusion process monitoring by machine vision with hybrid convolutional neural networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5769–5779, Sep. 2020, doi: 10.1109/TII.2019.2956078.

[49] N. Wambugu, Y. Chen, Z. Xiao, M. Wei, S. A. Bello, J. M. Junior, and J. Li, "A hybrid deep convolutional neural network for accurate land cover classification," *Int. J. Appl. Earth Obser. Geoinf.*, vol. 103, Dec. 2021, Art. no. 102515, doi: 10.1016/j.jag.2021.102515.

[50] H. Yuan, C. Van Der Wiele, and S. Khorram, "An automated artificial neural network system for land use/land cover classification from landsat TM imagery," *Remote Sens.*, vol. 1, no. 3, pp. 243–265, Jul. 2009, doi: 10.3390/rs1030243.

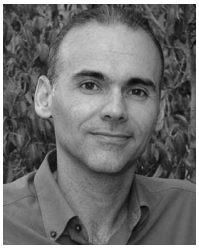
[51] S. Khomenko, E. Pisoni, P. Thunis, B. Bessagnet, M. Cirach, T. Iungman, E. P. Barboza, H. Khreis, N. Mueller, C. Tonne, K. de Hoogh, G. Hoek, S. Chowdhury, J. Lelieveld, and M. Nieuwenhuijsen, "Spatial and sector-specific contributions of emissions to ambient air pollution and mortality in European cities: A health impact assessment," *Lancet Public Health*, vol. 8, no. 7, pp. e546–e558, Jul. 2023, doi: 10.1016/s2468-2667(23)00106-8.

[52] F. Qamar, A. L. Pierce, and G. Dobler, "Covariance in policy diffusion: Evidence from the adoption of hyperlocal air quality monitoring programs by U.S. cities," *Cities*, vol. 138, Jul. 2023, Art. no. 104363, doi: 10.1016/j.cities.2023.104363.

[53] P. Diviacco, M. Iurcev, R. J. Carbajales, A. Viola, and N. Potleca, "Design and implementation of a crowdsensing-based air quality monitoring open and FAIR data infrastructure," *Processes*, vol. 11, no. 7, p. 1881, Jun. 2023.
[54] R. Yan, J. Liao, J. Yang, W. Sun, M. Nong, and F. Li, "Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114513, doi: 10.1016/j.eswa.2020.114513.
[55] E. Sharma, R. C. Deo, R. Prasad, A. V. Parisi, and N. Raj, "Deep air quality forecasts: Suspended particulate matter modeling with convolutional neural and long short-term memory networks," *IEEE Access*, vol. 8, pp. 209503–209516, 2020, doi: 10.1109/ACCESS.2020.3039002.
[56] E. Eslami, Y. Choi, Y. Lops, A. Sayeed, and A. K. Salman, "Using wavelet transform and dynamic time warping to identify the limitations of the CNN model as an air quality forecasting system," *Geosci. Model Develop.*, vol. 13, no. 12, pp. 6237–6251, Dec. 2020, doi: 10.5194/gmd-13-6237-2020.
[57] Y. Luo, Y. Zhang, X. Cai, and X. Yuan, "E^{2}GAN: End-to-end generative adversarial network for multivariate time series imputation," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 473–485, doi: 10.24963/ijcai.2019/429.
[58] X. Yu, M. S. Wong, C.-H. Liu, and R. Zhu, "Synergistic data fusion of satellite observations and in-situ measurements for hourly $PM_{2.5}$ estimation based on hierarchical geospatial long short-term memory," *Atmos. Environ.*, vol. 286, Oct. 2022, Art. no. 119257, doi: 10.1016/j.atmosenv.2022.119257.

EDUARDO ILLUECA FERNÁNDEZ was born in Cieza, in 1996. He received the bachelor's degree in biochemistry and the M.Sc. degree in bioinformatics from the University of Murcia (UMU), in 2018 and 2019, respectively, where he is currently pursuing the Ph.D. degree in industrial engineering. He is the author of the Bioconductor package org.Mxanthus.db: Genome-wide annotation for Myxococcus xanthus DK 1622. He has presented his work at several congresses. His research interests include machine learning and artificial intelligence solutions for air quality monitoring and air quality forecasting. He is being awarded for his academic performance and received a Research Initiation Fellowship from UMU, from November 2019 to December 2019, participating in the project of Development of Semantic Bioinformatics Tools for the Bacterium Myxococcus Xanthus. In 2020, he received a fellowship from the Seneca Foundation to pursue the Ph.D. degree with UMU and HOP Ubiquitous.

ANTONIO JESÚS JARA VALERA (Senior Member, IEEE) received the Ph.D. degree (cum laude) from the University of Murcia (UMU), Spain, and the M.B.A. degree from the ENAE Business School, UCAM, in 2012, where he did his entrepreneurship formation. The Ph.D. results present a novel way to connect objects to internet-enabled platforms in an easy, secure, and scalable way. He is currently the Director of the Research and Development Department, Libelium, and the Chair of the Data Quality and IoT in IEEE. As a part of Libelium, he is focused on the smart cities market with solutions for citizen engagements and environmental monitoring (air quality sensors). He was selected and mentored by the acceleration program FIWARE. He has also participated in more than 100 international events about the Internet of Things as a speaker of more than 100 international publications/articles (5000 citations and H-index of 37). He holds several patents in the IoT domain. He received the Entrepreneurship Awards from the ENAE (sponsored by SabadellCAM Financial Services), emprendeGo (sponsored by Spanish Government), and the IPSO Alliance Award (Sponsored by Google) for his disruptive innovation in the IoT.



JESUALDO TOMÁS FERNÁNDEZ BREIS (Senior Member, IEEE) received the bachelor's degree in computer engineering and the Ph.D. degree in computer science from the University of Murcia (UMU), Spain, in 1999 and 2003, respectively. He is currently a Full Professor with the Faculty of Computer Science, UMU. Since 2004, he has been leading research projects related to semantic web technologies. He is the co-founder of the spin-off Longseq applications. He is also a member of the IMIB-Arrixaca Bio-Health Research Institute. His current research interests include the application of semantic technologies for the development of learning health systems and the development of quality assurance methods for ontologies and terminologies.