Ecological Informatics 83 (2024) 102781

Contents lists available at ScienceDirect

Ecological Informatics

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

Optimisation of the adaptive neuro-fuzzy inference system for adjusting low-cost sensors PM concentrations

Martina Casari [a,*], Piotr A. Kowalski [b,c], Laura Po [a]

[a] Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Via P. Vivarelli, 10, Modena 41125, Italy
[b] Faculty of Physics and Applied Computer Science, AGH University of Krakow, al. Adama Mickiewicza 30, Kraków 30-059, Poland
[c] Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, Warszawa 01-447, Poland

ARTICLE INFO

Keywords:
Air quality
Particulate matter
Adjustment
ANFIS model
Fuzzy logic

ABSTRACT

Driven by the urgent necessity for accurate environmental data in urban settings, this research leverages the Adaptive Neuro-Fuzzy Inference System (ANFIS) as a machine learning-based approach to refine SPS30 low-cost sensor data influenced by hygroscopicity in Turin, Italy. Employing ANFIS offers several advantages: it enhances clarity regarding the correspondence between output and input values and rules, improves system interpretability, and facilitates the representation of linguistic variables and rules, thereby encouraging domain experts' involvement in enhancing the system's performance as needed. This paper illustrates the utility of ANFIS in adjusting the detected particulate matter (PM) concentration and compares its effectiveness with other established machine-learning techniques, including linear regression, decision trees, random forest, SVR and a multilayer perceptron (MLP). These methods are chosen as benchmarks owing to their established effectiveness in calibration procedures.

We propose certain preprocessing steps for detecting and rectifying anomalies, alongside introducing two distinct data-splitting methodologies. Additionally, a discussion about feature selection is presented to elucidate the impact of specific features on performance enhancement. The efficacy of ANFIS in refining PM data is demonstrated through a comparative assessment, where it outperforms all the established machine-learning techniques. Notably, incorporating only PM2.5, relative humidity and temperature as features yields optimal performance while mitigating overfitting issues. The paper also explores various ANFIS configurations, including two distinct optimization algorithms, and investigates the impact of the number and type of membership functions on the fuzzy system's performance. Our study highlights the potential of the Adaptive Neuro-Fuzzy Inference System as a versatile and effective tool for addressing real-world challenges in environmental sensing.

1. Introduction

In environmental science, accurate pollution measurement has emerged as a critical concern, especially when employing low-cost monitoring stations (Bachechi et al., 2024; Brilli et al., 2021; deSouza, 2022; Gerboles et al., 2017). Such a situation arises because the increasing impact of human activities on air quality necessitates precise and reliable data in order to formulate effective mitigation strategies and safeguard public health. This article delves into the pivotal issue of obtaining accurate pollution measurements from low-cost stations, highlighting the significance of addressing this challenge to advance our understanding of environmental dynamics and enhance monitoring capabilities.

As technology continues to evolve, the integration of artificial intelligence (AI) has revolutionized the way we approach complex tasks, including environmental monitoring (Karthika, 2023; Kusy et al., 2022; Mahdavinejad et al., 2018; Neo et al., 2023). Neural networks, fuzzy logic, swarm intelligence and other AI methodologies have proven instrumental in handling vast and intricate datasets, offering novel insights into pollution patterns and trends. Developments in AI technologies and their application to environmental science shed light on their potential to transform our ability to monitor, measure and predict air pollution with unprecedented accuracy (Chianese et al., 2019; Dun et al., 2022; Gokul et al., 2023; Kowalski et al., 2020, 2022; Navares and Aznarte, 2020).

Contemporary investigations on intelligent methods applied to air

* Corresponding author.
E-mail addresses: martina.casari@unimore.it (M. Casari), pkowal@agh.edu.pl (P.A. Kowalski), laura.po@unimore.it (L. Po).

https://doi.org/10.1016/j.ecoinf.2024.102781
Received 22 March 2024; Received in revised form 17 August 2024; Accepted 17 August 2024
Available online 22 August 2024
1574-9541/

pollution monitoring, measurement and prediction present very promising results (Mitreska Jovanovska et al., 2023). Leveraging advanced algorithms and computational models, these methods go beyond traditional approaches, offering a more nuanced and adaptive approach to environmental data analysis. By combining the power of AI with ecological science practices, researchers can unlock new dimensions of understanding, enabling real-time monitoring and timely interventions to mitigate the adverse effects of pollution on both ecosystems and human health (Bak et al., 2012; Mueller et al., 2023; Saeed et al., 2024; Tran et al., 2023; Wang et al., 2024).

Fuzzy logic, a computational paradigm that mimics human decision-making under uncertainty, has found diverse and impactful applications across various facets of environmental sciences (Do et al., 2022; Pham et al., 2024; Pouw and Kwiatkowska, 2013; Sheehan and Gough, 2016). Its adaptability to handle imprecise and vague information makes it a valuable tool in addressing the inherent complexity and uncertainty prevalent in environmental systems (Borri et al., 1998; Khatua et al., 2020).

One of the applications of fuzzy logic lies in environmental modelling, where it serves as a bridge between traditional deterministic models and the unpredictable nature of ecological processes. By incorporating fuzzy logic, researchers can capture the nuances of environmental variables that resist precise quantification, providing a more realistic representation of the intricate relationships within ecosystems. This approach enhances the accuracy of predictive models, contributing to more effective decision-making in areas such as climate change projections, land-use planning and biodiversity conservation (Biber et al., 2021; Caniani et al., 2016; D'Aniello, 2023; Rahman, 2020).

Furthermore, fuzzy logic plays a crucial role in the field of air and water quality monitoring (Barzegar et al., 2023; Güler Dincer and Akkuş, 2018; Manzar et al., 2022; Trach et al., 2022). Environmental data, often characterized by inherent uncertainties and variations, can be challenging to interpret accurately. Fuzzy logic-based systems excel in processing and analyzing this data, offering a robust framework to account for imprecision in pollutant measurements. This methodology proves particularly beneficial in discerning pollution levels near regulatory thresholds, aiding in timely interventions and ensuring compliance with environmental standards (WHO et al., 2021).

In environmental risk assessment, fuzzy logic provides a nuanced approach to evaluating the potential impacts of contaminants (Shwetank et al., 2019). Traditional risk assessment methods often rely on deterministic assumptions, neglecting the variability in exposure scenarios and ecological response values, promoting sustainable environmental management. In the field of air monitoring, fuzzy logic has been successfully applied to predict Air Quality Index (AQI) levels, focusing on gases and particulate matter pollutants, particularly PM10, or encompassing gases alone (for a detailed exploration of the literature on the air quality topic, please refer to Section 2).

In addition to Fuzzy logic, extensive research has been dedicated to calibrating and adjusting particulate matter (PM) data obtained from low-cost sensors. The burgeoning use of these sensors in environmental monitoring has prompted the development of techniques aimed at enhancing the accuracy and reliability of collected data (Okafor et al., 2020; Popescu et al., 2024; Rivera-Muñoz et al., 2022). Researchers have explored a spectrum of approaches, ranging from conventional methods like simple linear regression to cutting-edge technologies such as complex deep neural networks. These calibration and adjustment techniques are crucial for mitigating the inherent limitations of low-cost sensors, among others, sensitivity to environmental conditions and potential measurement inaccuracies). As the demand for cost-effective monitoring solutions grows, refining the methodologies for calibrating and adjusting data from low-cost sensors becomes paramount, ensuring the credibility of environmental measurements and bolstering the effectiveness of pollution assessment efforts.

The motivation behind this study stems from the growing importance of ensuring accurate pollution measurement in environmental science, particularly with regard to data derived from low-cost monitoring stations. As human activities continue to exert increasing pressure on air quality, there is a pressing need for precise and reliable data to inform effective mitigation strategies and protect public health (Nakhjiri and Kakroodi, 2024). The inadequacies of traditional monitoring methods and the proliferation of low-cost sensor technology underscore the significance of addressing this challenge. Using advanced methodologies such as the Adaptive Neuro-Fuzzy Inference System to adjust sensor data at low-cost, this study aims to deepen understanding of environmental dynamics and improve monitoring capabilities. This, in turn, will facilitate more informed decision-making in environmental management.

This research offers a novel approach by integrating fuzzy logic principles with neural network structures to address the complexities of low-cost sensor data adjustment. While previous studies have explored various machine learning techniques for pollution measurement and prediction, applying ANFIS to adjust low-cost sensor data represents a unique contribution to the field. By harnessing the interpretability and adaptability of fuzzy logic, coupled with the learning capabilities of neural networks, this research seeks to overcome the limitations of existing methods and provide a more robust framework for environmental data analysis. Integrating fuzzy logic with advanced machine learning methodologies exemplifies a forward-thinking approach towards environmental monitoring and underscores the potential of interdisciplinary research in addressing pressing environmental challenges. The novelty of this research lies in several key aspects: the utilization of low-cost PM2.5 data gathered from six SPS30 Sensirion sensors co-located with a reference station, providing a comprehensive dataset for analysis; the innovative methodology combining fuzzy logic with neural networks for air quality adjustment, offering a unique approach to calibrating low-cost sensor data; a comparative analysis between machine learning techniques and ANFIS, which sheds light on the effectiveness and advantages of the proposed methodology; an exploration into data and model optimisation that contributes to improved accuracy and reliability in ANFIS applied to air quality PM data, advancing the state-of-the-art in environmental data analysis and monitoring techniques.

The manuscript is structured as follows: The exploration of literature within the domain is meticulously detailed in Section 2. Section 3 elaborates on the utilized sensors, procedures for data collection, intricacies of the study site, and the methodological framework employed. Delving into the nuances of model configuration, data preprocessing, and dataset partitioning, Section 4 offers a discourse. In Section 5, a discerning comparative analysis is presented, contrasting the efficacy of Machine Learning against that of the Fuzzy Inference System approaches. Additionally, this section provides into an in-depth examination of the findings concerning ANFIS configuration, particularly in the realms of feature selection and model optimisation. Finally, Section 6 draws the manuscript to a close, encapsulating the principal findings and ramifications of the study.

2. Related work

Fuzzy logic has been widely applied in the domain of air quality assessment, as demonstrated by several exemplary studies. In Nihalani et al. (2020) the authors investigate air quality indices (AQI), crucial indicators of local air quality, which consider key pollutants such as sulphur dioxide, nitrogen dioxide, ground-level ozone, carbon monoxide and particulates. The study proposes a consistent method for determining AQI using a fuzzy logic system, enhancing the accuracy of air quality evaluation. Similarly, Ángel Olvera-García et al. (2016) introduces a novel evaluation model integrating fuzzy inferences with an Analytic Hierarchy Process, resulting in a new air quality index. By assessing environmental parameters and assigning individual weights to pollutants, this model offers improved air quality assessments, as evidenced by experiments conducted in Mexico City (Mexico). The published work Debnath et al. (2018) presents an integrated model based on interval type-2 fuzzy reasoning and analytic hierarchy process for air quality assessment in urban areas. This approach effectively models uncertainties involved in air quality classification, as demonstrated by its application in the Kolkata Metropolitan area (West Bengal, India). Moreover, fuzzy logic application extends to predictive modelling and forecasting of air pollution levels. The publication of Suganya and Meyyappan (2023) explores the development of a hybrid Adaptive Neuro-Fuzzy Inference System (ANFIS) combined with a recurrent neural network (RNN) for accurate air pollution prediction. This approach considers multiple climatic factors and pollutant concentrations, providing valuable insights for daily health monitoring and governmental decision-making. By leveraging the ability of fuzzy logic to handle complex and uncertain data, such predictive models offer invaluable tools for anticipating future air quality trends and implementing proactive mitigation measures. ANFIS is also employed in another study (Prasad et al., 2016), focusing on forecasting daily air pollution concentrations of five major air pollutants (sulphur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO), ozone (O3) and particulate matters (PM10)) in the atmosphere of a megacity (Howrah, West Bengal, India).

Furthermore, fuzzy logic plays a crucial role in the development of air quality monitoring systems, particularly those integrated with Internet of Things (IoT) technologies. The paper of Fahim et al. (2023) presents a novel IoT-based weather station device capable of measuring air quality parameters in real-time. Utilizing fuzzy inference systems to categorise parameter data and generate an air quality index (AQI), this system provides accessible and actionable environmental information for various stakeholders, including farmers, urban planners and the general public. Combining fuzzy logic with IoT technology, such monitoring systems offer scalable and cost-effective solutions for continuous environmental monitoring and management. Fuzzy logic also finds application in comparative studies aiming to assess the effectiveness of different air quality indices and evaluation methodologies. Hamedian et al. (2016) exemplifies this by comparing traditional air quality indices with those generated using fuzzy inference systems and clustering techniques. The research "Fuzzy Inference of Air Quality – A case study of Vadodara City" (West Bengal, India) compares conventional AQI with fuzzy AQI, demonstrating the capability of fuzzy inference systems to manage data ambiguity and interpret complex conditions. Such comparative analyses contribute to refining existing air quality assessment frameworks and identifying areas for improvement, ultimately enhancing the accuracy and reliability of environmental monitoring efforts. Additionally, through these exemplary articles, fuzzy logic emerges as a powerful tool for improving air quality assessment, offering more consistent and reliable methods for evaluating environmental parameters and safeguarding public health.

Fuzzy logic's versatility and robustness make it an indispensable tool in various aspects of air quality assessment, ranging from predictive modelling and real-time monitoring to comparative analyses and evaluation methodologies. As environmental challenges continue to evolve, integrating fuzzy logic with advanced technologies and methodologies promises to enhance further our ability to monitor, manage and mitigate the impacts of air pollution on human health and the environment.

Continuing the exploration of techniques for enhancing particulate matter data accuracy from low-cost sensors, a comprehensive overview of such research endeavours has been compiled and is presented in Table 1. This extensive compilation encapsulates the diverse methodologies for calibrating and adjusting data obtained from these sensors. The table is a valuable resource for researchers and practitioners seeking insights into the spectrum of techniques, ranging from traditional statistical methods to sophisticated artificial intelligence approaches. By presenting a synthesis of the findings from various studies, the contents of Table 1 facilitate a holistic understanding of the evolving landscape of methodologies employed to refine and optimise the precision of low-cost sensor data in environmental monitoring applications.

Table 1
State of the art of machine learning calibration algorithms for low-cost air quality sensors.

Study Title	Methodology and Approach	R^2 Values	RMSE Values (μg/m^3)
Machine learning techniques to improve the field performance of low-cost air quality sensors (Bush et al., 2022)	RF regression during 7 months	0.91	–
Performance Assessment of a Low-Cost PM2.5 Sensor for a near Four-Month Period in Oslo, Norway (Liu et al., 2019)	Statistical corrections for RH and T using MLR and RF models.	Site 1: 0.80 Site 2: 0.79 Site 3: 0.76	Site 1: 0.80 Site 2: 0.79 Site 3: 0.76
Evaluation of nine machine learning regression algorithms for calibration of low-cost PM2.5 sensor (Kumar and Sahu, 2021)	Best performances using kNN, RF and GB among MLR, Lasso regression, Ridge regression, SVR, MLP and Regression Tree.	Train: 0.99 Test: 0.97 (kNN) 0.96 (RF) 0.95 (GB)	–
Improving accuracy of air pollution exposure measurements: Statistical correction of a municipal low-cost airborne particulate matter sensor network (Considine et al., 2021)	Long-term dataset: RF model considering time-varying covariates and arterial road length. On-the-fly correction: MLR.	Long-term: 0.75 On-the-fly: 0.78	Long-term: 2.9 On-the-fly: 3.1
Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study (Zusman et al., 2020)	Region-specific multivariate linear regression calibration models for diverse particle sources and meteorological conditions.	Site 1: 0.74 Site 2: 0.95	Site 1:2.46 Site 2: 0.84
Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea (Lim et al., 2019)	Land Use Regression (LUR) models using LR, RF and a stacked ensemble (SE).	LR: 0.63 RF: 0.73 SE: 0.80	–
Evaluation and calibration of low-cost particulate matter sensors for respirable coal mine dust monitoring (Feng et al., 2023)	A two-layer correction model was introduced, incorporating top-performing models (KNN, RF, ET, XGBoost) and temperature/humidity data.	Sensor 1: 0.97 Sensor 2: 0.98	Sensor 1: 80 Sensor 2: 91
AirMLP: A Multilayer Perceptron Neural Network for Temporal Correction of PM2.5 Values in Turin (Italy) (Casari et al., 2023a)	MLP	0.932	
Calibration of Low-Cost Particle Sensors by Using Machine-Learning Method (Chen et al., 2018)	LR, SVR and Feedforward NN.	Uncalibrated: 0.618 LR: 0.728 SVR: 0.85 FNN: 0.905	–
Assessment and Calibration of a Low-Cost PM2.5 Sensor Using Machine Learning (HybridLSTM Neural	HybridLSTM model combining a deep neural network and an LSTM	Raw data 0.59 MLR: 0.80 DNN: 0.90 HybridLSTM: 0.93	–

(continued on next page)

Table 1 (*continued*)

Study Title	Methodology and Approach	R^2 Values	RMSE Values (μg/m^3)
Network) (Park et al., 2021)			
GAMMA: A universal model for calibrating sensory data of multiple low-cost air monitoring devices (Nguyen et al., 2024)	Deep learning approach based on the GAN structure.	0.928	3.51
Applying machine learning for large-scale field calibration of low-cost PM2.5 and PM10 air pollution sensors (Adong et al., 2022)	kNN, SVR, multivariate linear regression, ridge regression, lasso, elastic net regression, XGBoost, MLP, RF and gradient boosting.	Factory calibrated: 0.52 SVM: 0.84 Lasso: 0.86 Elastic net: 0.86 Ridge: 0.87 MLR: 0.87 KNN: 0.89 XGBoost: 0.92 GB: 0.92 MLP: 0.92 RF: 0.92	Factory calibrated: 18.6 SVM: 10.4 Lasso: 9.7 Elastic net: 9.7 Ridge: 9.5 MLR: 9.5 KNN: 8.9 XGBoost: 7.6 GB: 7.2 MLP: 7.2 RF: 7.2
Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods (Si et al., 2020)	Calibration methods including SLR, MLR, NN and XGBoost.	–	SLR: 4.91 MLR: 4.65 XGBoost: 4.19 NN: 3.91

3. Materials and data acquisition

3.1. SPS30 sensor

The sensors utilized in this study are SPS30 sensors developed by Sensirion, a leading manufacturer of environmental sensing solutions. Each sensor is housed within a device called *Arianna*, which is deployed by Wiseair S.r.l., a Milan-based startup dedicated to fostering awareness and understanding of air quality issues across Italy and beyond (Wiseair, n.d.). The *Arianna* device incorporates an SPS30 laser-scattering sensor for detecting particulate matter, as well as relative humidity and temperature sensors, also developed by Sensirion.

This sensor is frequently used in literature (Jaafar et al., 2024; Koziel et al., 2024a; Koziel et al., 2024b) because it is MCERTS-certified, it has an affordable price, has good electrical parameters, reduced physical dimensions and uses contamination-resistance technology exploiting Sensirion technology. Furthermore, the SPS30 sensor demonstrated high linearity for PM2.5 ($R^2 = 0.95$) (Nguyen et al., 2021). The accuracy of the SPS30 sensor was over 95% for PM1.0 mass concentrations below 100 $\mu g/m^3$, but this accuracy decreased to approximately 77% for PM1.0 mass concentrations above 100 $\mu g/m^3$. For PM2.5 mass concentrations, the accuracy remained relatively stable, ranging from 81% to 96%. In laboratory experiments conducted at 20 $^\circ C$ and 40% relative humidity, the SPS30 sensors generally overestimated PM1.0 and PM2.5 measurements compared to GRIMM reference instruments (AQ-SPEC, n.d.). The given specifications have been reported in Table 2 and the recommended operation condition in Table 3. Furthermore, the SPS30 has a lifetime of ten years in continuous operation, with a start-up time of 30 s. It has a built-in fan to facilitate air transportation. SPS30 sensors operate based on optical particle counting (OPC) principles utilizing laser scattering. Ambient particles are directed to a measurement cell containing a light source and a photodetector. When particles interact with light, some of it scatters to the photodetector. The collected signal is processed to obtain real-time particle count and mass concentration values, which are expressed in units of particles per cubic centimetre ($\#/cm^3$) and micrograms per cubic meter ($\mu g/m^3$), respectively. It is worth noting that the PM4.0 and PM10 outputs of Sensirion's PM sensors are estimated from measurements of PM0.5, PM1.0 and PM2.5. These estimates take into account typical aerosol profiles rather than being solely based on real raw data events from larger particles.

Even if the sensor itself has a sampling frequency of 1 ± 0.04 seconds the data collected are transmitted to the Wiseair server at intervals of every 15 min. Nevertheless, the frequency of data transmission may vary depending on the battery life of the device, which is charged by a small solar panel located on the device's surface. The air enters the device without the aid of a pump, passing through a grid designed to prevent the entry of insects and larger particles.

Table 2
SPS30 specifications.

Parameter	Conditions	Value	Units
Mass concentration range	–	0 to 1000	$\mu g/m^3$
Mass concentration size range	PM1.0	0.3 to 1.0	μm
	PM2.5	0.3 to 2.5	μm
	PM4	0.3 to 4.0	μm
	PM10	0.3 to 10.0	μm
Mass concentration precision for PM1 and PM2.5	0 to 100 $\mu g/m^3$ 100 to 1000 $\mu g/m^3$	± 10 ± 10	$\mu g/m^3$ $\%m.v.$
Number concentration precision for PM0.5, PM1 and PM2.5	0 to 1000 $\#/cm^3$	± 100	$\#/cm^3$
	1000 to 3000 $\#/cm^3$	± 10	$\%m.v.$
Lifetime	24 h/day operation	>10	years
Maximum long-term mass concentration precision limit drift	0 to 100 $\mu g/m^3$ 100 to 1000 $\mu g/m^3$	± 1.25 ± 1.25	$\mu g/m^3/year$ $\%m.v.$ / year

Table 3
SPS30 recommended operating conditions.

Parameter	Recommended Operating Conditions
Temperature	10 to 40 $^\circ C$
Relative humidity	20 to 80 %

3.2. Data collection

The data utilized for this study were collected in Turin, Italy, spanning from June 2022 to September 2023 (the dataset is accessible through Zenodo Casari et al. (2023b)). This dataset comprises information gathered by six SPS30 low-cost sensors strategically positioned alongside a Tecora reference station located at an altitude of 4 m along the station fence (Arpa-Piemonte, n.d.).

In the experimental setup, it is important to note that all six low-cost sensors were not operational simultaneously. Instead, they were deployed over different periods. Table 4 provides a detailed description

Table 4
SPS30 Turin sensors validity periods.

sensor_id	min_valid_at	max_valid_at
ari-1727	2021-06-30 23:00:00	2022-07-05 10:00:00
ari-1952	2022-02-07 08:00:00	2023-02-01 11:00:00
ari-1953	2022-02-07 09:00:00	2022-09-16 12:00:00
ari-2049	2022-07-25 13:00:00	2023-09-25 22:00:00
ari-1885	2022-09-26 08:00:00	2023-09-25 22:00:00
ari-2074	2022-12-16 13:00:00	2023-09-25 22:00:00

of the running period for each sensor.

The information directly collected from the *Arianna* devices, includes measurements of four different particulate matter, relative humidity and temperature. These primary measurements are then supplemented with additional meteorological features by linking them to online available data sources. The features encompass:

- PM1, PM2.5, PM4, and PM10 mass concentrations ($\mu g/m^3$): these represent different size fractions of particulate matter suspended in the air. It is worth noting that the classification of particulate matter into different-size fractions follows a hierarchical structure. In this hierarchy, larger diameter categories, such as PM10, encompass the masses of smaller ones, including PM4, PM2.5 and PM1, and so forth.
- Relative humidity (%): RH measures the amount of moisture in the air, for instance, a relative humidity of 80% indicates that the air is holding 80% of the maximum water vapour it could hold at that particular temperature.
- Temperature (°C): temperature reflects the thermal conditions of the environment.
- Wind speed (m/s): Wind speed denotes the rate at which air is moving.
- Atmospheric pressure (hPA): Atmospheric pressure represents the weight of the air above the sensor.

3.2.1. Hygroscopicity effect

When it comes to low-cost sensors for measuring particulate matter concentration, it is crucial to take into account the impact of hygroscopicity. Hygroscopicity refers to particles' ability to attract and retain water molecules from the surrounding air. In the context of low-cost laser scattering sensors, this phenomenon can affect the accuracy of particle concentration measurements, as the particles' mass increases with absorbed water.

In Fig. 1, it is observable that with varying RH ranges, the difference between the concentrations detected by the low-cost sensor and the reference station increases with higher RH. Hygroscopicity depends on the location of interest and the device used for PM concentration. A prior investigation is necessary when dealing with the air quality PM concentration dataset, due to the possibility of correcting the hygroscopicity problem (Casari and Po, 2024; Patel et al., 2023).

In the current study, the dataset exhibits the effects of hygroscopicity due to the temporal and spatial context. According to official climate data, the average annual relative humidity in Turin is 73.8%, with a minimum of 67% in March and a maximum of 78% in October, November, and December. Given the study's objective to adjust inaccurate PM concentration detection, ANFIS appears to be a successful method due to its capability of fuzzifying variables into ranges. The hygroscopicity effect varies by particle type, but a threshold of 70% is generally considered a starting point where particle sizes begin to increase, as shown in Fig. 1 (Won et al., 2021).

As demonstrated in Section 5.3.2 of the Results, ANFIS effectively addresses the issue of hygroscopicity affecting the dataset in this case study.

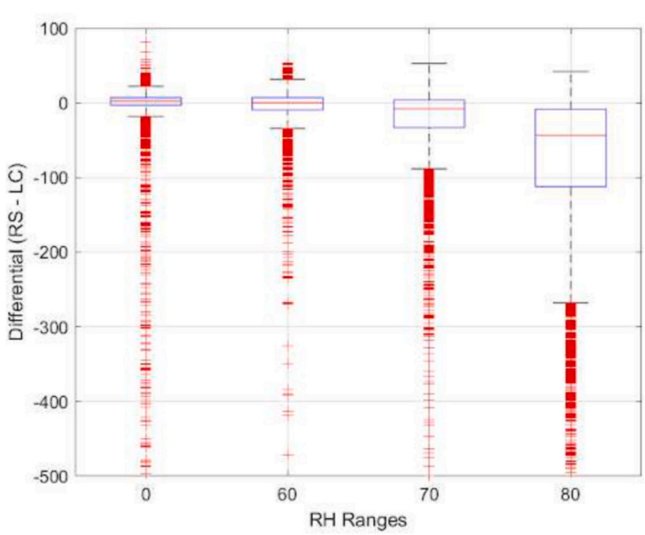


Fig. 1. Difference between PM2.5 mass concentrations detected by the reference station (RS) and the low-cost (LC) SPS30 sensor, categorized by relative humidity (RH) ranges.

4. Methodology and model configuration

4.1. ML methods and fuzzy inference systems

To provide insights about the Adaptive Neuro-Fuzzy Inference System applied to the problem of air quality, the objective is to juxtapose established models such as linear regression (Weisberg, 2005), decision trees (Loh, 2011), random forests (Johansson et al., 2014), SVR (Cherkassky and Ma, 2004), and MLP neural networks (Ivakhnenko and Lapa, 1967) employed as baseline models. The neural network includes an input layer, a batch normalization layer, seven subsequent dense layers of 1500 neurons each (all with ReLU activations), and a final output layer (Casari et al., 2023a). The ANFIS model represents a novel approach for adjusting low-cost PM data; in this way, these machine-learning techniques were selected as benchmarks due to their proven efficacy in calibration procedures.

Understanding the operation of the ANFIS model requires exploring the principles of fuzzy logic and Fuzzy Inference Systems (FIS). These concepts are explored in more detail in the next subsections.

4.1.1. Fuzzy inference system

Fuzzy logic introduced by Lotfi A Zadeh in 1965 (Zadeh, 1978; Zadeh et al., 1996) differs from conventional logic, which operates in binary terms (TRUE or FALSE), through offering a paradigm shift by representing truth not as a binary state but rather as a continuum of truthfulness, ranging in a limited space. This approach allows for a more nuanced representation of real-world phenomena that may not have clear-cut boundaries and do not strictly adhere to Boolean truth. Instead of absolute certainty, fuzzy logic implements degrees of certainty, often represented as linguistic terms (e.g., low, fair, high or certainly yes, possibly yes, possibly no, certainly no).

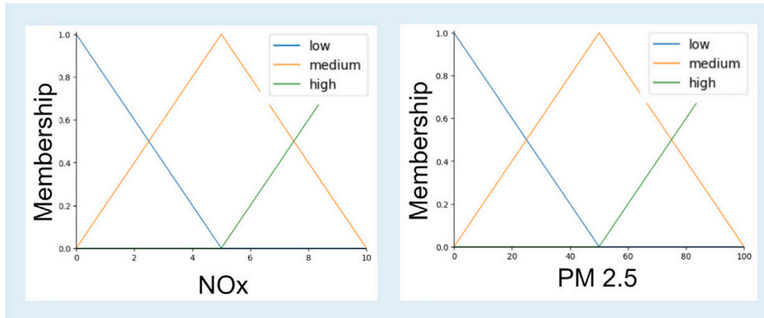
A Fuzzy Inference System (Mamdani and Assilian, 1975) consists of several essential components. An illustration is included in Fig. 2 to enhance the explanation of the subsequent stages:

1. Fuzzification module: This initial stage involves the transformation of crisp (numeric) input values into fuzzy variables, often represented linguistically, through assigned membership functions (MFs) (Fig. 2a). The resulting output is represented as a vector from the fuzzification process, which is subsequently used during the inference step (Fig. 2b). Generally, given a universe X in which the variable x is defined, the fuzzy set A in X comprises ordered pairs, as expressed by:

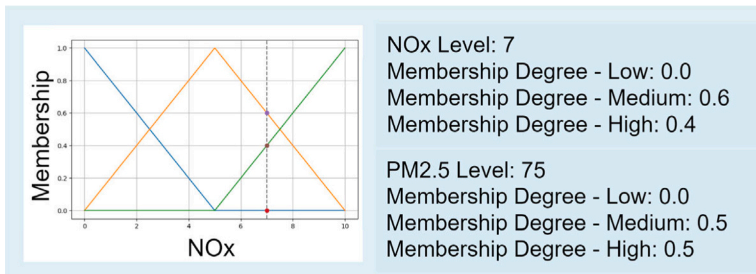
$$A = \{(x, \mathrm{MF}) | x \in X\}$$

Here, the MF represents the membership function that maps each element of X to a membership value between 0 and 1. The MF can assume various shapes, depending on which best describes the universe under consideration, including linear, Gaussian, sigmoid, quadratic and cubic polynomials, or simpler forms composed of straight lines like triangular, trapezoidal, linear ascending or linear descending.

2. Knowledge Base: The knowledge base of a FIS comprises a set of expert-provided rules in the form of IF-THEN statements (Fig. 2c).



(a) Input variables membership functions



(b) Fuzzification

Rule 1: IF NOx[low] AND PM2.5[low] THEN AQI Level[good]

Rule 2: IF NOx[medium] OR PM2.5[medium] THEN AQI Level[moderate]

Rule 3: IF NOx[high] OR PM2.5[high] THEN AQI Level[unhealthy]

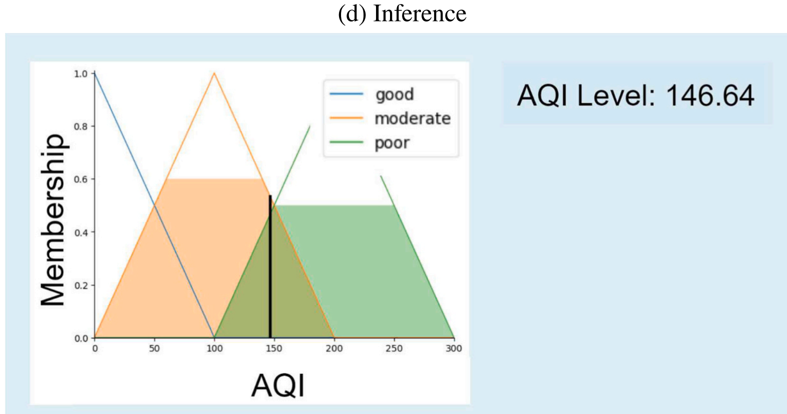
(c) Rules

Rule 1 inference: min(NOx[0.0], PM2.5[0.0]) AQI Level[0.0]

Rule 2 inference: max(NOx[0.6], PM2.5[0.5]) AQI Level[0.6]

Rule 3 inference: max(NOx[0.4], PM2.5[0.5]) AQI Level[0.5]

(d) Inference



(e) Aggregation and defuzzification

Fig. 2. The subfigures illustrate the sequential steps and components involved in the fuzzy inference process.

Each rule specifies conditions (antecedents) based on input variables and corresponding actions (consequents) based on output variables. Each rule can comprise logical operators (*AND*, *OR*, and *NOT*) when combining multiple states regarding different variables. The Boolean logic operators *AND*, *OR*, and *NOT* are typically defined within the scope of fuzzy logic, as operators of minimum, maximum and complement; in this case, they are also called *Zadeh operators* (Zadeh, 1965) and are defined as follows:

$$\mathrm{NOTx} = 1 - \mathrm{MF(x)} \qquad \mathrm{xANDy} = min(\mathrm{MF(x)}, \mathrm{MF(y)}) \qquad \mathrm{xORy} = max(\mathrm{MF(x)}, \mathrm{MF(y)})$$

It is worth noting that as the number of input fuzzy variables increases, the number of rules typically grows, often showing exponential expansion. While the sheer number of rules might suggest the system's complexity, it is crucial to recognize that a system with fewer membership functions per variable could be more complex, especially when incorporating more variables (Gegov et al., 2017).

3. Inference engine: The inference engine simulates the human reasoning process by performing fuzzy inferences based on the inputs and IF-THEN rules. Each rule may carry a weight, typically ranging from 0 to 1, to increase or decrease its effect, or all rules can be assigned a weight of 1 to have equal importance. Each involved variable is assigned a degree derived from its membership function (Fig. 2d), and the resulting rule output is inferred.
4. Aggregation of inference outputs: In an FIS, decisions are made by testing all the rules, and the outputs of these rules need to be combined. This aggregation process merges the fuzzy sets representing the output of each rule into a single fuzzy set (Fig. 2e).
5. Defuzzification module: The fuzzy set obtained from the inference engine is converted into a crisp value through defuzzification. Defuzzification is necessary to derive a single output value from the set, one common method is centroid calculation, which determines the centre of the area under the aggregate fuzzy set (Fig. 2e).

Fuzzy Logic has been used in different fields and it has been proved to deal with the uncertainty and subjectivity of environmental problems adequately, as in Ocampo-Duque et al. (2006). Breaking down the Fuzzy Inference System into systematic steps enhances user understanding, providing clarity on how the output corresponds to input values and rules. This improves the system's interpretability, building trust in its decision-making process. Additionally, the utilization of linguistic variables and rules facilitates domain expert involvement, empowering them to refine the system's performance as necessary.

4.1.2. Adaptive neuro-fuzzy inference system

The Adaptive Neuro-Fuzzy Inference System integrates fuzzy reasoning principles with the structural characteristics of neural networks, enabling it to learn and adapt from data dynamically (Chanal et al., 2022; Sayyaadi, 2021).

Initially, ANFIS constructs a FIS with a basic framework, lacking a comprehensive understanding of membership functions or rules. However, it iteratively refines and optimises these rules and functions to minimise output error or to enhance the explanation of complex system behaviours. This optimisation is achieved through the adjustment or tuning of membership function parameters using hybrid learning algorithms or backpropagation techniques applied to specific input-output data patterns (Karaboga and Kaya, 2019). Through this integration, ANFIS effectively constructs fuzzy *IF-THEN* rules and membership functions, enabling accurate modelling of input-output relationships.

The resulting model remains highly interpretable, with easily understandable rules. This characteristic is particularly beneficial for systems where verification and certification play a crucial role.

4.2. Data preprocessing

The dataset employed in this study underwent meticulous preprocessing to enhance data quality and consistency. Key preprocessing steps include the following:

1. Standardizing data frequency to 1 h: The granularity of the data obtained from the reference station is at one-hour intervals. To ensure consistency, the data derived from the SPS30 sensors has been resampled hourly using the nearest approximation method. In the context of Python resampling, the nearest approximation involves assigning each new timestamp the value of the existing data point closest to it in time, ensuring a synchronized temporal alignment between the low-cost sensor and the reference station data.
2. Outlier reduction beyond 3 standard deviations: Data points exceeding 3 standard deviations from the mean were deemed outliers and subsequently set as *null*. This step aids in eliminating data stemming from potentially malfunctioning instrumentation.
3. Interpolation utilizing kNN technique: Missing data points were imputed using the k-nearest neighbours (kNN) interpolation technique. This method leverages the entire feature vector to estimate *null* values, with the parameter *k* set to 5, ensuring a robust estimation of missing data points.
4. Left-side median cleaning smoothing technique: To further refine the dataset, a left-side median cleaning technique was applied for smoothing purposes. This involves using a window of preceding hours to the current data point. If the data point deviates above or below the median by a specified threshold, it is adjusted to the median value of the window, promoting data consistency and reducing noise.
5. Normalization: The final step involved normalizing the dataset across all features. This normalization process was carried out after splitting the dataset into training and test sets as required.

4.3. Dataset splitting

In order to ensure the integrity of the data separation and avoid overfitting issues, a specific approach was adopted. Instead of randomly splitting the data into training and test sets, each day's data was treated as a separate batch. This approach was chosen due to the temporal nature of the data, which exhibited daily variations in PM values. By grouping the data into daily batches, the risk of including overly similar data points in both sets was mitigated.

Two distinct strategies were employed to handle the data batches:

- Sequential approach: the first 75% of days from each month were allocated to the training set, while the remaining 25% were assigned to the test set.
- Random approach: 75% of random days from each month were allocated to the training set, the remaining to the test set.

Both strategies were evaluated to determine their effectiveness in training the model and their ability to accurately generalize to unseen data. This comparison provided insights into the optimal data separation method for the fuzzy system, informing its subsequent implementation and evaluation.

5. Results and discussion

The overall performance of the different models tested is detailed first, highlighting their performance using metrics such as R^2 (Eq. 1) and RMSE (Eq. 2).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n}} \tag{2}$$

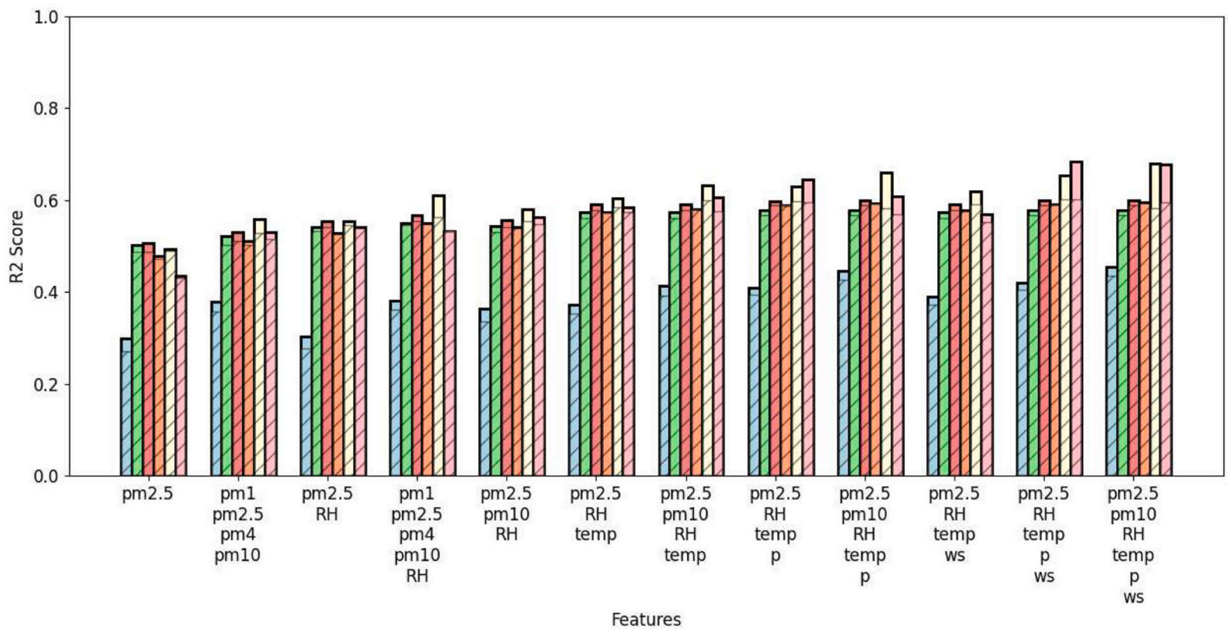
This is followed by an exploration of the preprocessing and dataset-splitting outcomes. Subsequently, the exploration shifts to the ANFIS fuzzy system configuration, encompassing aspects such as the type of membership functions utilized, the number of membership functions allocated to each feature, and the optimisation algorithm employed for ANFIS training.

5.1. Overall methods results

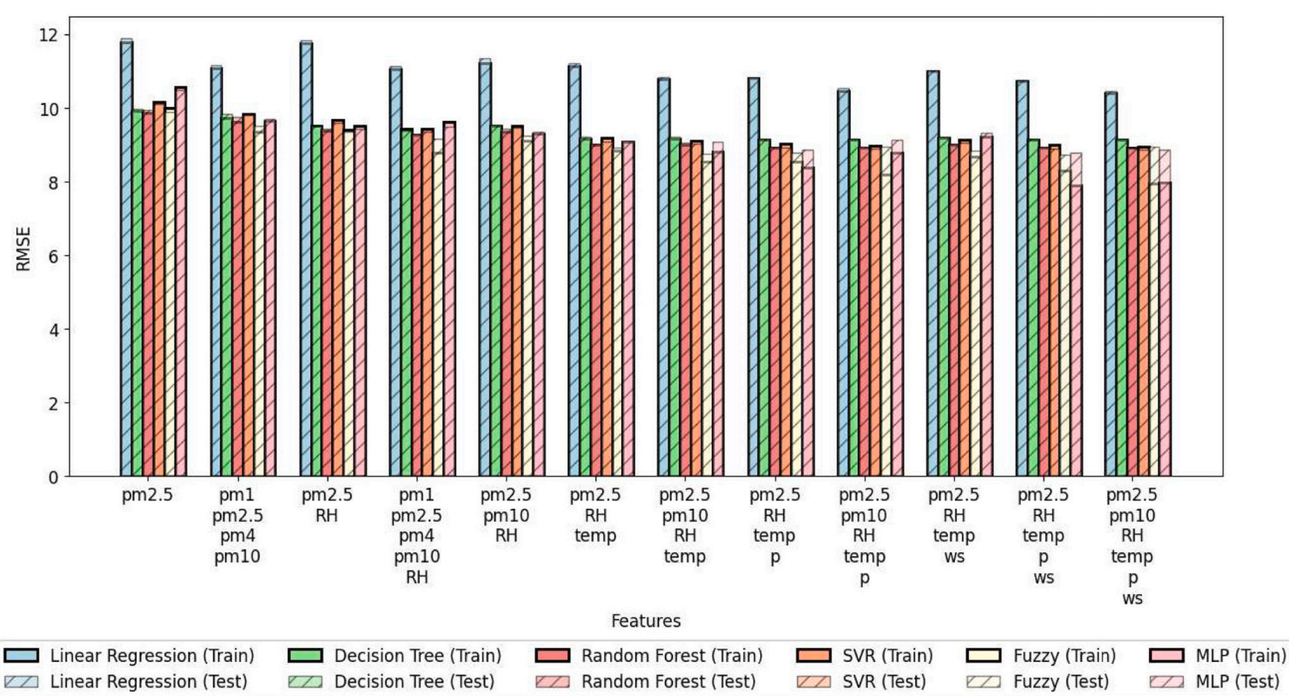
The study results showcase the effectiveness of various methods in adjusting PM2.5 data acquired from low-cost sensors, compared to the reference station data. These methods include linear regression, decision trees, random forests, support vector machines regression, the Fuzzy Adaptive Neuro-Fuzzy Inference System and an MLP neural network.

Fig. 3 displays the R^2 and RMSE scores of the models trained on the preprocessed data, with both train and test values shown. The ANFIS system demonstrated promising results, particularly concerning the inclusion of PM2.5, PM10, RH, and temperature, with no notable performance improvements observed when additional features were included in the model. Additionally, a tendency for overfitting was observed in the fuzzy system with increased features.

In contrast, linear regression consistently lagged, never surpassing an R^2 of 0.5. Among the models, random forest performed relatively better



(a) R^2



(b) RMSE

2

Fig. 3. Comparative R^2 and RMSE scores for various models (including the Fuzzy Inference System) across different features (*pm*1, *pm*2.5, *pm*4, *pm*10, *RH* for relative humidity, *temp* for temperature, *p* for pressure, and *ws* for wind speed).

apart from the fuzzy ANFIS system.

The MLP neural network exhibits comparable performance in terms of test results, surpassing the R^2 of the ANFIS method only when utilizing the full set of features. Furthermore, the RMSE tends to be slightly higher compared to ANFIS. Nevertheless, ANFIS was chosen for its interpretability and explainability, which can be advantageous in certain scenarios. Even if the NN had higher performance, there would still be cases where ANFIS is preferable due to its transparency and ease of understanding.

In general, compared to the studies proposed in Table 1, classical machine learning methods performed worse in this study, possibly due to the greater complexity of the data. The SPS30 sensor's significant hygroscopicity effect necessitates consideration when working with this data. Models trained on data without this effect may have an advantage.

The results obtained in our study when compared to the findings of Prasad et al. (2016), exhibit consistency, with a decrease in performance observed when using fewer features. It is worth noting that the comparison is drawn between hourly data in our analysis and daily data in theirs. Additionally, the variation in performance may be influenced by factors such as the types of sensors utilized and the specific context of sensor deployment. Furthermore, it is important to highlight that while their work focuses on forecasting, ours is centred on data adjustment. These factors collectively contribute to a nuanced understanding of the comparative results and underscore the importance of contextual considerations in interpreting research findings. Nevertheless, the fact that the results are comparable despite these variations is a promising outcome, suggesting the robustness and generalizability of the ANFIS method across different contexts and methodologies.

5.2. Preprocessing and dataset splitting

As discussed in Section 4.2, this study went under a meticulous preprocessing phase, which included the removal of 3 standard deviations to eliminate gross anomalies and the application of interpolation using k-nearest neighbours to fill in the missing data. Subsequently, a one-sided median cleaning procedure has been employed with a window size of 4 h to smooth the data.

The choice of the window size was crucial, as it influenced the smoothing process. After experimenting with different window sizes (see Table 5), a window size of 4 h was identified as yielding the highest R^2 score on the test set, indicating a superior fit to the data. This optimized window size was then used for all subsequent analyses.

As elaborated in Section 4.3, a critical consideration was the dataset-splitting methodology. Fig. 4 illustrates that the sequential method appeared to yield better results compared to the random method, particularly concerning the test set. Therefore. the sequential method was selected as the preferred splitting approach for all subsequent analyses.

5.3. Exploration of the ANFIS configuration

Following the dataset preprocessing and splitting phases, the ANFIS was trained and tested using different configurations of the membership functions shape, and number. In addition, the optimisation algorithm was tested between GridSearch and SubtractiveClustering.

5.3.1. Membership functions

One of the key aspects of the ANFIS system is the type and number of membership functions used for each feature. Different types of membership functions, such as Gaussian, triangular, and trapezoidal, were experimented with to determine their impact on the adjustment process. Additionally, the number of membership functions for each feature was systematically varied to assess its impact on the performance of the ANFIS fuzzy system. The features considered were PM1, PM2.5, PM4, PM10, RH, temperature, pressure, and wind speed, which were kept in the same order throughout the trials. The number of membership functions tried were:

- Run 1: *[3,3,3,3,2,3,3,3]*, where only RH was set to 2.
- Run 2: *[2, 6, 2, 2, 2, 3, 3, 3]*, with the number of PM features reduced apart from PM2.5, which was set to 6.
- Run 3: *[2, 6, 2, 2, 2, 2, 3, 3]*, with the number of temperature-related features reduced.
- Run 4: *[2, 6, 2, 2, 2, 2, 2, 2]*, with the number of membership functions reduced to 2 for all features except PM2.5.

Each configuration was tested, and the performance of the ANFIS was evaluated to determine the optimal number of membership functions for each feature, as shown in Fig. 5.

In general, reducing the number of membership functions helped to avoid overfitting. Consequently, in the final ANFIS configuration, the Run 4 setup is retained.

The results, depicted in Fig. 6a, reveal that the triangular membership function consistently yielded the most stable performance, with an R^2 score never dropping below 0.3. This robust performance is further illustrated in the zoomed-in view provided in Fig. 6b. It is worth noting that Prasad et al. (2016) also found triangular functions to be optimal for air quality data, where they restricted the number of membership functions to 3, thus reducing computational costs, suggesting a consistent pattern across studies.

5.3.2. Optimisation algorithms

During the training of ANFIS, optimisation algorithms play a crucial role in efficiently handling the numerous combinations required for optimisation. Rather than exhaustively attempting every combination, these algorithms aim to identify the optimal solution by intelligently sampling only a small subset of the entire solution space. In this study, two distinct optimization algorithms have been explored: GridSearch (Pontes et al., 2016) and SubtractiveClustering (Chen, 2013).

While GridSearch rigorously explores the entire parameter space to find the best solution, SubtractiveClustering dynamically adjusts to the data distribution, providing a more flexible and potentially stable optimisation approach.

It was found that the GridSearch algorithm generally exhibited better performance in terms of optimizing the fuzzy system. However, when dealing with a larger number of selected features, the SubtractiveClustering algorithm demonstrated greater stability, see Fig. 7.

5.3.3. Illustrative result

An illustrative example is presented in Fig. 8 using three variables: PM2.5, relative humidity and temperature, with membership functions.

Table 5
Performance metrics on training and test sets changing one-sided median cleaning window.

Time Interval	Training Set				Test Set			
	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
2 h	0.5590	6.7183	90.1888	9.4317	−0.0988	6.7124	238.1817	12.7064
3 h	0.5581	6.7391	90.4346	9.4403	0.4080	6.5927	113.9166	10.0674
4 h	0.5556	6.7721	90.9021	9.4618	0.5074	6.5321	88.3200	9.2406
5 h	0.5519	6.8135	91.7828	9.5046	0.5002	6.5818	89.9019	9.3091
12 h	0.5234	7.0669	97.2473	9.7971	0.4661	6.8248	94.4673	9.5959

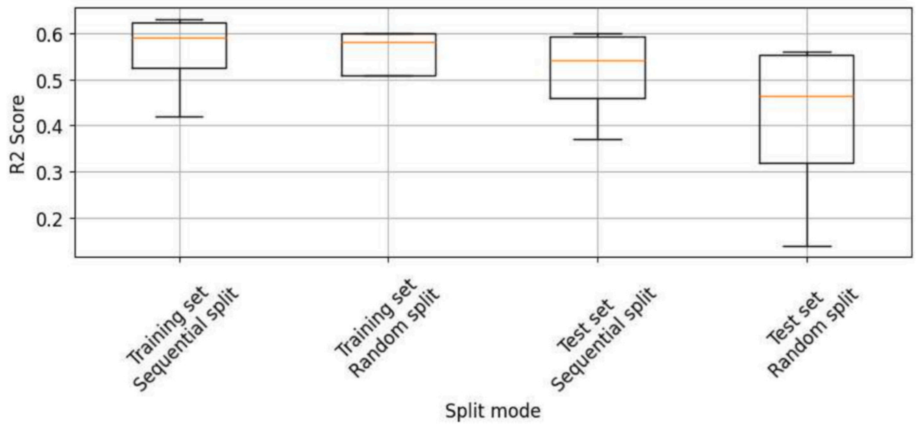


Fig. 4. R^2 scores for training and test sets obtained by the fuzzy method using sequential and random splits.

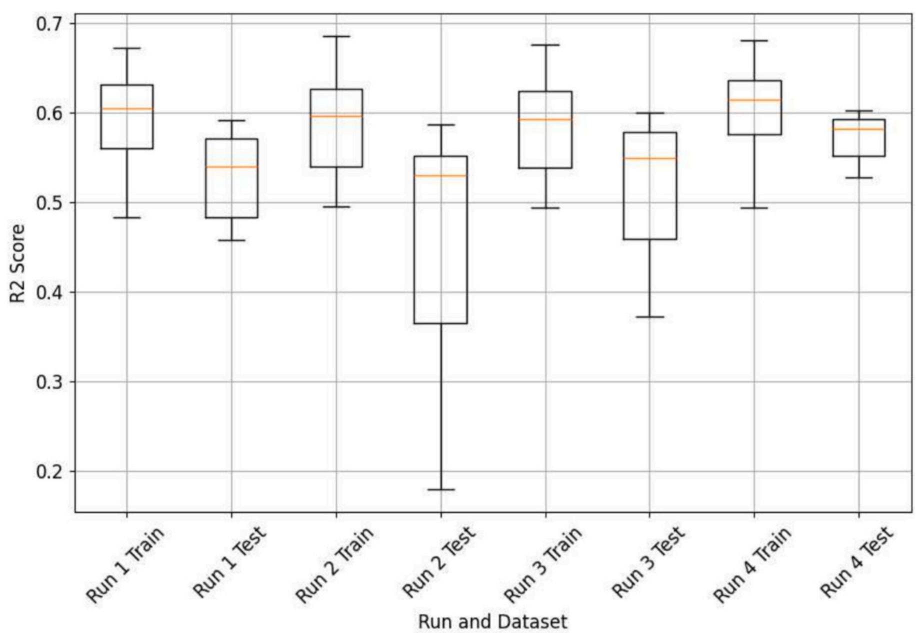


Fig. 5. The obtained R^2 scores for both the training and test sets across varying numbers of membership functions.

Triangular membership functions are chosen for each variable. It is noted that the number of membership functions retained is crucial to avoid overfitting. In this case, 6 membership functions are selected for PM2.5 to accurately capture various behaviours across different PM ranges, while relative humidity and temperature each employ 2 membership functions.

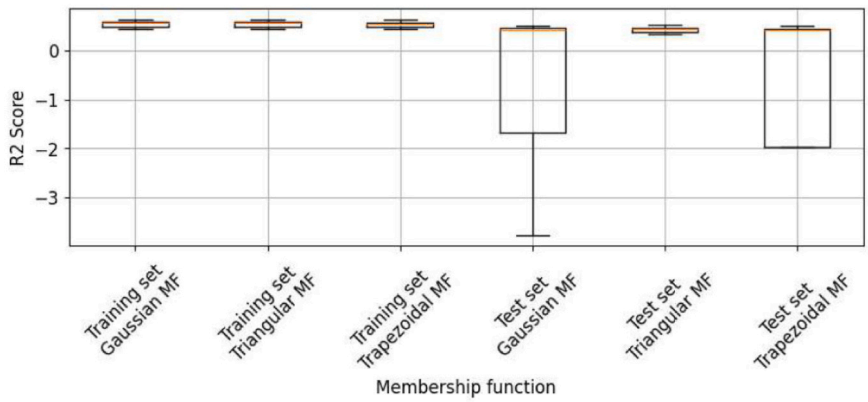
Fig. 9 illustrates time series data obtained from a low-cost sensor, showcasing PM2.5 readings from the low-cost sensor itself, the reference station, and predictions generated by the ANFIS model for both the training and test sets. This comparison provides significant insights into the performance and accuracy of the ANFIS model in predicting PM2.5 levels. Notably, both figures demonstrate that ANFIS can mitigate the hygroscopic effect and replicate the behaviour of the reference station in both training and test sets. These results are satisfactory and provide insights into the effectiveness of ANFIS when applied to PM2.5 hourly data in a high RH context.

5.3.4. Advantages and disadvantages of fuzzy logic

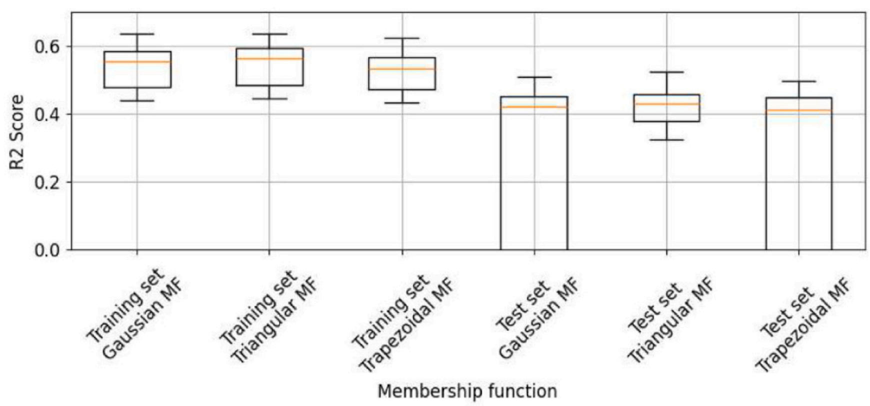
Fuzzy logic modelling, while highly versatile and adaptable, presents limitations that are crucial to consider. One of the primary drawbacks is its inherent subjectivity; defining fuzzy sets and rules relies on human judgment, which can introduce bias and inconsistency. This subjectivity can lead to ambiguity and a lack of precision, particularly in complex systems where clear, objective data might be preferable. Additionally, as the number of rules and variables increases, fuzzy logic systems can become quite complex, making them difficult to manage and optimise effectively.

On the other hand, fuzzy logic offers significant advantages that make it a powerful tool in many applications. Its ability to handle uncertainty and imprecise information allows it to mimic human reasoning more closely than traditional binary logic systems. This makes fuzzy logic particularly useful in situations requiring human-like decision-making, such as in control systems, robotics, and consumer electronics. The flexibility of fuzzy logic enables it to adapt to new data and changing conditions without requiring extensive reprogramming, saving time and resources in dynamic environments. Furthermore, its interpretability and ease of integration with other AI techniques, such as neural networks, enhance its potential for creating robust, adaptive systems.

Applying the Adaptive Neuro-Fuzzy Inference System (ANFIS) to adjust low-cost sensor PM concentrations highlights both the strengths and limitations of AI methods in environmental monitoring. Unlike



(a) No y-axis limitation



(b) y-axis limited between 0 and 0.7

Fig. 6. In (a), a boxplot displays the training and test set R^2 scores for different membership function types. No y-axis limit is imposed in this subfigure. In (b), a similar boxplot is shown, but with the y-axis lower limit set to 0 to provide a clearer representation of R^2 scores above 0.

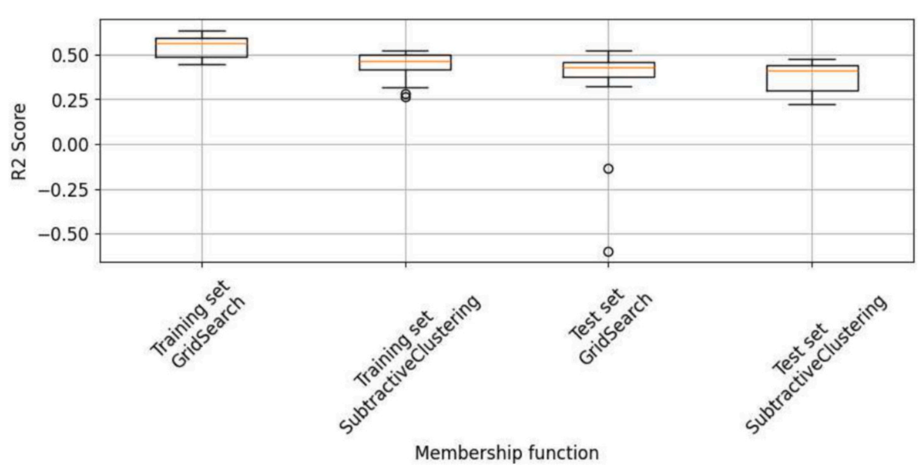


Fig. 7. Training and test set R^2 scores obtained using the GridSearch and SubtractiveClustering optimisation algorithms.

deterministic approaches that offer rigorous proof of correctness, AI methods, including ANFIS, rely on test procedures involving random selection and repeated validation using various datasets. This introduces an inherent uncertainty, as the lack of formal proof means that the reliability of the results is heavily dependent on the quality and representativeness of the test data. However, fuzzy logic, central to ANFIS, provides a bridge between AI's complex computations and human interpretability by mimicking the way humans perceive and process information. This human-like reasoning capability allows for greater transparency and understanding of how decisions are made within the system.

Despite these advantages, relying on fuzzy logic can also be seen as a drawback, as its interpretability can lead to subjective conclusions that may not always align with objective accuracy. Moreover, the scalability

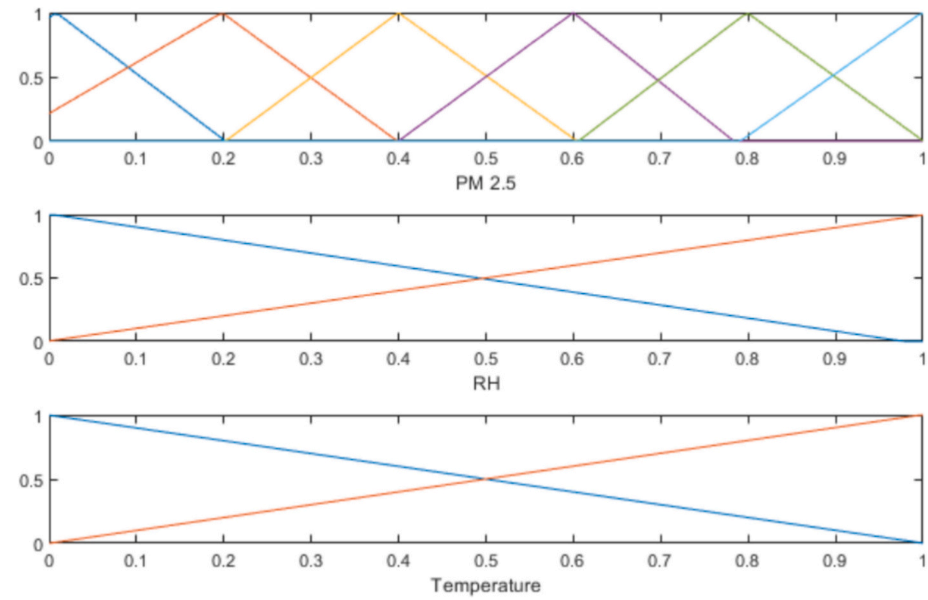
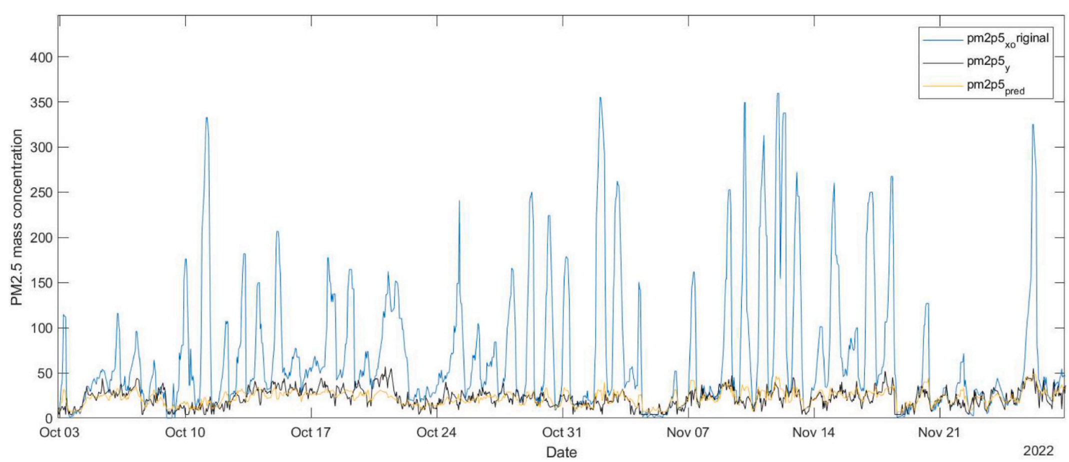
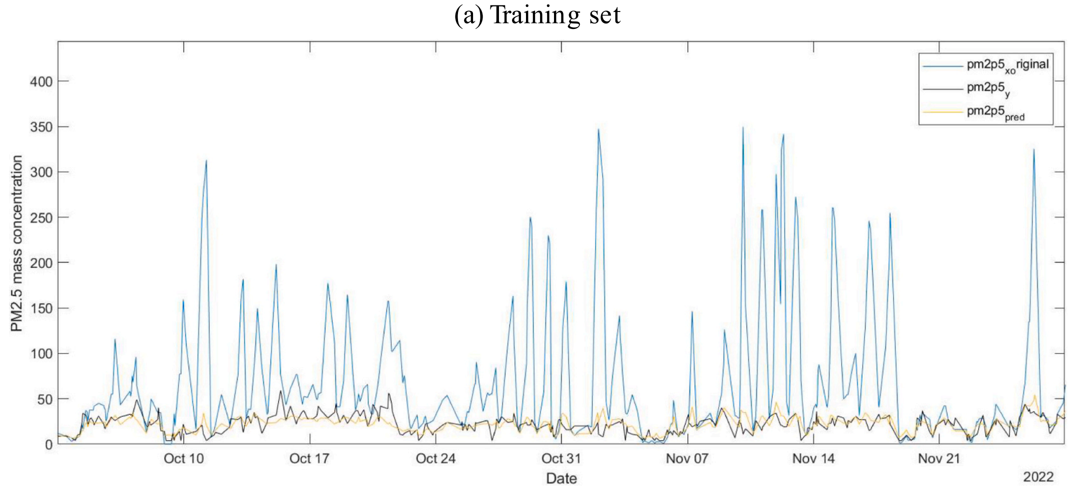


Fig. 8. Membership functions for the variables PM2.5, relative Humidity and temperature within the Fuzzy Inference System (FIS).



(a) Training set



(b) Test set

Fig. 9. Comparison of time series data depicting PM2.5 levels obtained from a low-cost sensor (blue line), a reference station (black line), and predictions generated by the Fuzzy Inference System (yellow line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the ANFIS algorithm is a notable advantage, allowing for two primary approaches: repeating the entire optimisation process or expanding the rule base with new, interesting cases specific to different locations or devices. This flexibility is beneficial for adapting to diverse environmental conditions and sensor characteristics, but it also necessitates careful management to avoid overfitting and maintain generalizability. While ANFIS and fuzzy logic introduce complexity and potential for subjective bias, their adaptability and interpretability make them valuable tools for fine-tuning sensor data for improved monitoring of air quality.

6. Conclusion

This study underscores the effectiveness of the Adaptive Neuro-Fuzzy Inference System in ameliorating low-cost sensor data. The Fuzzy system's rules provide valuable insights into the adjustment process, making it a promising approach for addressing real-world challenges in environmental sensing, without forgetting to discuss the problems associated with this method used in such a context.

The comparison with other ML and artificial neural networks methods gives a more precise insight into how much ANFIS applies to the problem in use, in particular, it is possible to observe that in terms of R^2 and RMSE, the performance are intriguing, having the ANFIS method surpassing the other methods for almost all features sets.

With respect to the ANFIS method parameters, the exploration of the various membership function types revealed that the triangular membership function exhibited the most stable performance in the system. Furthermore, reducing the number of membership functions resulted in a reduction in overfitting.

Our investigation into two optimisation algorithms, GridSearch and SubtractiveClustering, for training the ANFIS fuzzy system unveiled that while GridSearch generally outperformed in terms of R^2 score, SubtractiveClustering demonstrated greater stability. This emphasises the importance of selecting an appropriate optimisation algorithm tailored to the specific dataset's characteristics.

Moving forward, further research could focus on refining the ANFIS model by exploring different membership function types for different features and unequal range distributions. Additionally, integrating domain knowledge and expert insights into the adjustment process could enhance the fuzzy system's interpretability and robustness.

Overall, this study underscores the potential of the ANFIS fuzzy system as a versatile tool for adjusting low-cost sensor data in environmental monitoring applications. By harnessing its interpretability and adaptability, researchers and practitioners can gain deeper insights into complex environmental phenomena, enabling informed decision-making to tackle environmental challenges effectively.

CRediT authorship contribution statement

Martina Casari: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Piotr A. Kowalski:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Laura Po:** Writing – review & editing, Supervision.

Declaration of competing interest

None.

Data availability

The data and MATLAB software utilized in this study can be accessed through Casari et al. (2023b) and Casari and Kowalski (2024), respectively.

References

Adong, P., Bainomugisha, E., Okure, D., Sserunjogi, R., 2022. Applying machine learning for large scale field calibration of low-cost pm2.5 and pm10 air pollution sensors. Appl. AI Lett. 3 https://doi.org/10.1002/ail2.76. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85135942842&doi=10.1002%2fail2.76&partnerID=40&md5=4605a386235c1ea47272b1bffaa98662.

Ángel Olvera-García, M., Carbajal-Hernández, J.J., Sánchez-Fernández, L.P., Hernández-Bautista, I., 2016. Air quality assessment using a weighted fuzzy inference system. Eco. Inform. 33, 57–74. URL: https://www.sciencedirect.com/science/article/pii/S1574954116300322.

AQ-SPEC,. Aiq quality sensor performance evaluation center. https://www.aqmd.gov/aq-spec/evaluations/criteria-pollutants/field. Accessed: July 13, 2024.

Arpa-Piemonte,. Torino rubino - reference station. http://www.sistemapiemonte.it/ambiente/srqa/stazioni/pdf/226.pdf. Accessed: January 13, 2024.

Bachechi, C., Rollo, F., Po, L., 2024. Hypeair: a novel framework for real-time low-cost sensor calibration for air quality monitoring in smart cities. Eco. Inform. 81, 102568. URL: https://www.sciencedirect.com/science/article/pii/S1574954124001109.

Bak, J., Doytchinov, S., Fisher, R., Forsius, M., Grennfelt, P., Harmens, H., Hettelingh, J. P., Holmberg, M., Jenkins, A., Krzyzanowski, M., et al., 2012. Impacts of Air Pollution on Ecosystems, Human Health and Materials under Different Gothenburg Protocol Scenarios. Technical Report, United Nations Economic Commission for Europe(UNECE).

Barzegar, Y., Gorelova, I., Bellini, F., D'Ascenzo, F., 2023. Drinking water quality assessment using a fuzzy inference system method: A case study of rome (italy). Int. J. Environ. Res. Public Health 20. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85167742121&doi=10.3390%2fijerph20156522&partnerID=40&md5=cecdc3fc37576459baab321177489965, doi: 10.3390/ijerph20156522. all Open Access, Gold Open Access, Green Open Access.

Biber, P., Schwaiger, F., Poschenrieder, W., Pretzsch, H., 2021. A fuzzy logic-based approach for evaluating forest ecosystem service provision and biodiversity applied to a case study landscape in southern germany. Eur. J. For. Res. 140, 1559–1586. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117606930&doi=10.1007%2fs10342-021-01418-4&partnerID=40&md5=748166606b69a2b136efe1c95cb1cb15. https://doi.org/10.1007/s10342-021-01418-4. all Open Access, Green Open Access, Hybrid Gold Open Access.

Borri, D., Concilio, G., Conte, E., 1998. A fuzzy approach for modelling knowledge in environmental systems evaluation1this paper describes research coordinated by a. magnaghi, university of Florence, "theories, methods and experiences of spatial analysis and ecological planning", ministry of scientific research, research program of national interest and relevant interest for science development. The research unit (ai applications in spatial planning) of the polytechnic of bari is directed by d. borri. The paper is the result of team work. The introduction was written by e. conte; the next nine sections are by g. concilio; and the conclusion is by d. borri.1. Comput. Environ. Urban. Syst. 22, 299–313. URL: https://www.sciencedirect.com/science/article/pii/S0198971598000453.

Brilli, L., Carotenuto, F., Andreini, B.P., Cavaliere, A., Esposito, A., Gioli, B., Martelli, F., Stefanelli, M., Vagnoli, C., Venturi, S., Zaldei, A., Gualtieri, G., 2021. Low-cost air quality stations' capability to integrate reference stations in particulate matter dynamics assessment. Atmosphere 12. https://doi.org/10.3390/atmos12081065. URL: https://www.mdpi.com/2073-4433/12/8/1065.

Bush, T., Papaioannou, N., Leach, F., Pope, F.D., Singh, A., Thomas, G.N., Stacey, B., Bartington, S., 2022. Machine learning techniques to improve the field performance of low-cost air quality sensors. Atmospheric. Meas. Tech. 15, 3261–3278. URL: https://amt.copernicus.org/articles/15/3261/2022/.

Caniani, D., Labella, A., Lioi, D.S., Mancini, I.M., Masi, S., 2016. Habitat ecological integrity and environmental impact assessment of anthropic activities: a gis-based fuzzy logic model for sites of high biodiversity conservation interest. Ecol. Indic. 67, 238–249. URL: https://www.sciencedirect.com/science/article/pii/S1470160X16300590.

Casari, M., Kowalski, P.A., 2024. Anfis Low-Cost Pm Adjustment. https://doi.org/10.5281/zenodo.10854497.

Casari, M., Po, L., 2024. Mith: a framework for mitigating hygroscopicity in low-cost pm sensors. Environ. Model Softw. 173, 105955. URL: https://www.sciencedirect.com/science/article/pii/S1364815224000161.

Casari, M., Po, L., Zini, L., 2023a. Airmlp: a multilayer perceptron neural network for temporal correction of pm2.5 values in Turin. Sensors 23. https://doi.org/10.3390/s23239446. URL: https://www.mdpi.com/1424-8220/23/23/9446.

Casari, M., Po, L., Zini, L., 2023b. Low-Cost pm Data. https://doi.org/10.5281/zenodo.10037781.

Chanal, D., Yousfi Steiner, N., Petrone, R., Chamagne, D., Péra, M.C., 2022. Online diagnosis of pem fuel cell by fuzzy c-means clustering. In: Cabeza, L.F. (Ed.), Encyclopedia of Energy Storage. Elsevier, Oxford, pp. 359–393. URL: https://www.sciencedirect.com/science/article/pii/B9780128197233000998. https://doi.org/10.1016/B978-0-12-819723-3.00099-8.

Chen, M.Y., 2013. A hybrid anfis model for business failure prediction utilizing particle swarm optimization and subtractive clustering. Inf. Sci. 220, 180–195.

Chen, C.C., Kuo, C.T., Chen, S.Y., Lin, C.H., Chue, J.J., Hsieh, Y.J., Cheng, C.W., Wu, C. M., Huang, C.M., 2018. Calibration of low-cost particle sensors by using machine-learning method. In: In: 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), pp. 111–114. https://doi.org/10.1109/APCCAS.2018.8605619.

Cherkassky, V., Ma, Y., 2004. Practical selection of svm parameters and noise estimation for svm regression. Neural Netw. 17, 113–126.

Chianese, E., Camastra, F., Ciaramella, A., Landi, T., Staiano, A., Riccio, A., 2019. Spatio-temporal learning in predicting ambient particulate matter concentration by multi-

layer perceptron. Eco. Inform. 49, 54–61. URL: https://www.sciencedirect.com/science/article/pii/S1574954118301870.
Considine, E.M., Reid, C.E., Ogletree, M.R., Dye, T., 2021. Improving accuracy of air pollution exposure measurements: statistical correction of a municipal low-cost airborne particulate matter sensor network. Environ. Pollut. 268, 115833. URL: https://www.sciencedirect.com/science/article/pii/S0269749120365222.
D'Aniello, G., 2023. Fuzzy logic for situation awareness: a systematic review. J. Ambient. Intell. Humaniz. Comput. 14, 4419–4438. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85148348909&doi=10.1007%2fs12652-023-04560-6&partnerID=40&md5=63b0ac595a61d6a5c908994f99157faf.
Debnath, J., Majumder, D., Biswas, A., 2018. Air quality assessment using weighted interval type-2 fuzzy inference system. Eco. Inform. 46, 133–146. URL: https://www.sciencedirect.com/science/article/pii/S1574954118300359.
deSouza, P.N., 2022. Key concerns and drivers of low-cost air quality sensor use. Sustainability 14. https://doi.org/10.3390/su14010584. URL: https://www.mdpi.com/2071-1050/14/1/584.
Do, A.N.T., Tran, H.D., Ashley, M., 2022. Employing a novel hybrid of ga-anfis model to predict distribution of whiting fish larvae and juveniles from tropical estuaries in the context of climate change. Eco. Inform. 71, 101780. URL: https://www.sciencedirect.com/science/article/pii/S1574954122002308.
Dun, A., Yang, Y., Lei, F., 2022. Dynamic graph convolution neural network based on spatial-temporal correlation for air quality prediction. Eco. Inform. 70, 101736. URL: https://www.sciencedirect.com/science/article/pii/S1574954122001868.
Fahim, M., El Mhouti, A., Boudaa, T., Jakimi, A., 2023. Modeling and implementation of a low-cost iot-smart weather monitoring station and air quality assessment based on fuzzy inference model and mqtt protocol. Model. Earth Syst. Environ. 9, 4085–4102.
Feng, Z., Zheng, L., Zhang, X., Liu, J., Xue, N., Wang, M., 2023. Evaluation and calibration of low-cost particulate matter sensors for respirable coal mine dust monitoring. Aerosol Sci. Technol. 0, 1–12. https://doi.org/10.1080/02786826.2023.2288609 doi: 10.1080/02786826.2023.2288609, arXiv:doi: 10.1080/02786826.2023.2288609.
Gegov, A., Sanders, D., Vatchova, B., 2017. Aggregation of inconsistent rules for fuzzy rule base simplification. Int. J. Knowledge-Based Intell. Eng. Syst. 21, 135–145. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85027501153&doi=10.3233%2fKES-170358&partnerID=40&md5=2e77c100b39626bc31999b9da12ad987. https://doi.org/10.3233/KES-170358. all Open Access, Green Open Access.
Gerboles, M., Spinelle, L., Borowiak, A., 2017. Measuring Air Pollution with Low-Cost Sensors.
Gokul, P., Mathew, A., Bhosale, A., Nair, A.T., 2023. Spatio-temporal air quality analysis and pm2.5 prediction over Hyderabad city, India using artificial intelligence techniques. Eco. Inform. 76, 102067. URL: https://www.sciencedirect.com/science/article/pii/S1574954123000961.
Güler Dincer, N., Akkuş, Özge, 2018. A new fuzzy time series model based on robust clustering for forecasting of air pollution. Eco. Inform. 43, 157–164. URL: https://www.sciencedirect.com/science/article/pii/S1574954117302030.
Hamedian, A.A., Javid, A., Zarandi, S.M., Rashidi, Y., Majlesi, M., 2016. Air quality analysis by using fuzzy inference system and fuzzy c-mean clustering in Tehran, Iran from 2009–2013. Iran. J. Public Health 45, 917.
Ivakhnenko, A., Lapa, V., 1967. Cybernetics and Forecasting Techniques. Modern Analytic and Computational Methods in Science and Mathematics. American Elsevier Publishing Company. URL: https://books.google.it/books?id=rGFgAAAAMAAJ.
Jaafar, W., Xu, J., Farrar, E., Jeong, C.H., Ganji, A., Evans, G., Hatzopoulou, M., 2024. Challenges and opportunities of low-cost sensors in capturing the impacts of construction activities on neighborhood air quality. Build. Environ. 254, 111363. URL: https://www.sciencedirect.com/science/article/pii/S0360132324002051.
Johansson, U., Boström, H., Löfström, T., Linusson, H., 2014. Regression conformal prediction with random forests. Mach. Learn. 97, 155–176.
Karaboga, D., Kaya, E., 2019. Adaptive network based fuzzy inference system (anfis) training approaches: a comprehensive survey. Artif. Intell. Rev. 52, 2263–2293.
Karthika, D., 2023. A study on artificial intelligence for monitoring smart environments. Mater. Today: Proc. 80, 2009–2013. URL: https://www.sciencedirect.com/science/article/pii/S2214785321043911. https://doi.org/10.1016/j.matpr.2021.06.046. sI: 5 NANO 2021.
Khatua, A., Jana, S., Kar, T.K., 2020. A fuzzy rule-based model to assess the effects of global warming, pollution and harvesting on the production of hilsa fishes. Eco. Inform. 57, 101070. URL: https://www.sciencedirect.com/science/article/pii/S1574954120300200.
Kowalski, P.A., Sapała, K., Warchałowski, W., 2020. Pm10 forecasting through applying convolution neural network techniques. Int. J. Environ. Impacts 3, 31–43. https://doi.org/10.2495/EI-V3-N1-31-43.
Kowalski, P.A., Kusy, M., Szwagrzyk, M., Izodorczyk, J., 2022. Estimation of atmospheric boundary layer values in the context of the daily prediction of pm10 air pollution. Int. J. Environ. Impacts 5, 271–283. https://doi.org/10.2495/EI-V5-N3-271-283.
Koziel, S., Pietrenko-Dabrowska, A., Wojcikowski, M., Pankiewicz, B., 2024a. Efficient calibration of cost-efficient particulate matter sensors using machine learning and time-series alignment. Knowl.-Based Syst. 295, 111879. URL: https://www.sciencedirect.com/science/article/pii/S0950705124005136.
Koziel, S., Pietrenko-Dabrowska, A., Wojcikowski, M., Pankiewicz, B., 2024b. Field calibration of low-cost particulate matter sensors using artificial neural networks and affine response correction. Measurement 230, 114529. URL: https://www.sciencedirect.com/science/article/pii/S0263224124004147. https://doi.org/10.1016/j.measurement.2024.114529.
Kumar, V., Sahu, M., 2021. Evaluation of nine machine learning regression algorithms for calibration of low-cost pm2.5 sensor. J. Aerosol Sci. 157, 105809. URL: https://www.sciencedirect.com/science/article/pii/S0021850221005401.
Kusy, M., Szwagrzyk, M., Kowalski, P.A., Konior, A., 2022. Machine learning techniques for explaining air pollution prediction. In: The IEEE World Congress on Computational Intelligence (IEEE WCCI), International Joint Conference on Neural Networks (IJCNN-IEEE 2022), Padova (Italy), pp. 1–8. https://doi.org/10.1109/IJCNN55064.2022.9891994.
Lim, C.C., Kim, H., Vilcassim, M.R., Thurston, G.D., Gordon, T., Chen, L.C., Lee, K., Heimbinder, M., Kim, S.Y., 2019. Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. Environ. Int. 131, 105022. URL: https://www.sciencedirect.com/science/article/pii/S0160412019304854.
Liu, H.Y., Schneider, P., Haugen, R., Vogt, M., 2019. Performance assessment of a low-cost pm2.5 sensor for a near four-month period in Oslo, Norway. Atmosphere 10. https://doi.org/10.3390/atmos10020041. URL: https://www.mdpi.com/2073-4433/10/2/41.
Loh, W.Y., 2011. Classification and regression trees. Wiley Interdiscip. Rev. Data Mining Knowledge Disc. 1, 14–23.
Mahdavinejad, M.S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., Sheth, A.P., 2018. Machine learning for internet of things data analysis: a survey. Digit Commun. Netw. 4, 161–175. URL: https://www.sciencedirect.com/science/article/pii/S235286481730247X. https://doi.org/10.1016/j.dcan.2017.10.002.
Mamdani, E.H., Assilian, S., 1975. An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man-Machine Stud. 7, 1–13.
Manzar, M.S., Benaafi, M., Costache, R., Alagha, O., Mu'azu, N.D., Zubair, M., Abdullahi, J., Abba, S., 2022. New generation neurocomputing learning coupled with a hybrid neuro-fuzzy model for quantifying water quality index variable: a case study from Saudi Arabia. Eco. Inform. 70, 101696. URL: https://www.sciencedirect.com/science/article/pii/S1574954122001467.
Mitreska Jovanovska, E., Batz, V., Lameski, P., Zdravevski, E., Herzog, M.A., Trajkovik, V., 2023. Methods for urban air pollution measurement and forecasting: challenges, opportunities, and solutions. Atmosphere 14. https://doi.org/10.3390/atmos14091441. URL: https://www.mdpi.com/2073-4433/14/9/1441.
Mueller, N., Westerby, M., Nieuwenhuijsen, M., 2023. Health impact assessments of shipping and port-sourced air pollution on a global scale: a scoping literature review. Environ. Res. 216, 114460. URL: https://www.sciencedirect.com/science/article/pii/S001393512201787X.
Nakhjiri, A., Kakroodi, A.A., 2024. Air pollution in industrial clusters: a comprehensive analysis and prediction using multi-source data. Eco. Inform. 80, 102504. URL: https://www.sciencedirect.com/science/article/pii/S1574954124000463.
Navares, R., Aznarte, J.L., 2020. Predicting air quality with deep learning lstm: towards comprehensive models. Eco. Inform. 55, 101019. URL: https://www.sciencedirect.com/science/article/pii/S1574954119303309.
Neo, E.X., Hasikin, K., Lai, K.W., Mokhtar, M.I., Azizan, M.M., Hizaddin, H.F., Razak, S. A., Yanto, 2023. Artificial intelligence-assisted air quality monitoring for smart city management. PeerJ Comp. Sci. 9 https://doi.org/10.7717/peerj-cs.1306. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85160785869&doi=10.7717%2fpeerj-cs.1306&partnerID=40&md5=aeebda26a18264c43ecc52dde08d5d22. all Open Access, Gold Open Access, Green Open Access.
Nguyen, N.H., Nguyen, H.X., Le, T.T., Vu, C.D., 2021. Evaluating low-cost commercially available sensors for air quality monitoring and application of sensor calibration methods for improving accuracy. Open J. Air Pollut. 10 https://doi.org/10.4236/ojap.2021.101001.
Nguyen, A.D., Phung, T.H., Nguyen, T.D., Pham, H.H., Nguyen, K., Nguyen, P.L., 2024. Gamma: a universal model for calibrating sensory data of multiple low-cost air monitoring devices. Eng. Appl. Artif. Intell. 128 https://doi.org/10.1016/j.engappai.2023.107591. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85179178744&doi=10.1016%2fj.engappai.2023.107591&partnerID=40&md5=91b91c04d2e75aa0aff98625fb046ec7.
Nihalani, S.A., Moondra, N., Khambete, A., Christian, R., Jariwala, N., 2020. Air quality assessment using fuzzy inference systems. In: Advanced Engineering Optimization Through Intelligent Techniques: Select Proceedings of AEOTIT 2018. Springer, pp. 313–322.
Ocampo-Duque, W., Ferré-Huguet, N., Domingo, J.L., Schuhmacher, M., 2006. Assessing water quality in rivers with fuzzy inference systems: a case study. Environ. Int. 32, 733–742. URL: https://www.sciencedirect.com/science/article/pii/S0160412006000407.
Okafor, N.U., Alghorani, Y., Delaney, D.T., 2020. Improving data quality of low-cost iot sensors in environmental monitoring networks using data fusion and machine learning approach. ICT Express 6, 220–228. URL: https://www.sciencedirect.com/science/article/pii/S2405959520300862. https://doi.org/10.1016/j.icte.2020.06.004.
Park, D., Yoo, G.W., Park, S.H., Lee, J.H., 2021. Assessment and calibration of a low-cost pm2.5 sensor using machine learning (hybridlstm neural network): feasibility study to build an air quality monitoring system. Atmosphere 12. https://doi.org/10.3390/atmos12101306. URL: https://www.mdpi.com/2073-4433/12/10/1306.
Patel, M.Y., Vannucci, P.F., Kim, J., Berelson, W.M., Cohen, R.C., 2023. Towards a universal hygroscopic growth calibration for low-cost $pm_{2.5}$ sensors. EGUsphere 2023, 1–14. URL: https://egusphere.copernicus.org/preprints/2023/egusphere-2023-1701/. https://doi.org/10.5194/egusphere-2023-1701.
Pham, V.T., Do, T.A.T., Tran, H.D., Do, A.N.T., 2024. Classifying forest cover and mapping forest fire susceptibility in dak nong province, Vietnam utilizing remote sensing and machine learning. Eco. Inform. 79, 102392. URL: https://www.sciencedirect.com/science/article/pii/S1574954123004211.
Pontes, F.J., Amorim, G., Balestrassi, P.P., Paiva, A., Ferreira, J.R., 2016. Design of experiments and focused grid search for neural network parameter optimization. Neurocomputing 186, 22–34.

Popescu, S.M., Mansoor, S., Wani, O.A., Kumar, S.S., Sharma, V., Sharma, A., Arya, V.M., Kirkham, M., Hou, D., Bolan, N., Chung, Y.S., 2024. Artificial intelligence and iot driven technologies for environmental pollution monitoring and management. Front. Environ. Sci. 12 https://doi.org/10.3389/fenvs.2024.1336088. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85186579907&doi=10.3389%2ffenvs.2024.1336088&partnerID=40&md5=0175a8e84806e8e5e19bafc814ab6ed1. all Open Access, Gold Open Access.

Pouw, F.A., Kwiatkowska, M., 2013. An overview of fuzzy-logic based approaches to ecology: Addressing uncertainty. In: 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS). IEEE, pp. 540–545.

Prasad, K., Gorai, A.K., Goyal, P., 2016. Development of anfis models for air quality forecasting and input optimization for reducing the computational cost and time. Atmos. Environ. 128, 246–262. URL: https://www.sciencedirect.com/science/article/pii/S1352231016300073.

Rahman, M.A., 2020. Improvement of rainfall prediction model by using fuzzy logic. Am. J. Clim. Chang. 9, 391–399.

Rivera-Muñoz, L., Giraldo-Forero, A., Martinez-Vargas, J., 2022. Deep matrix factorization models for estimation of missing data in a low-cost sensor network to measure air quality. Eco. Inform. 71, 101775. URL: https://www.sciencedirect.com/science/article/pii/S1574954122002254.

Saeed, N., Malik, H., Naeem, A., Bashir, U., 2024. Incorporating big data and iot in intelligent ecosystems: state-of-the-arts, challenges and opportunities, and future directions. Multimed. Tools Appl. 83, 20699–20741. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85170388222&doi=10.1007%2fs11042-023-16328-3&partnerID=40&md5=ac09f3803dd84029c63b311245ad54e6.

Sayyaadi, H., 2021. Chapter 8 - real-time optimization of energy systems using the soft-computing approaches. In: Sayyaadi, H. (Ed.), Modeling, Assessment, and Optimization of Energy Systems. Academic Press, pp. 479–527. URL: https://www.sciencedirect.com/science/article/pii/B9780128166567000087. https://doi.org/10.1016/B978-0-12-816656-7.00008-7.

Sheehan, T., Gough, M., 2016. A platform-independent fuzzy logic modeling framework for environmental decision support. Eco. Inform. 34, 92–101. URL: https://www.sciencedirect.com/science/article/pii/S1574954116300425.

Shwetank, Suhas, Chaudhary, J.K., 2019. Estimation of groundwater contamination using fuzzy logic: a case study of Haridwar, India. Groundw. Sustain. Dev. 8, 644–653. URL: https://www.sciencedirect.com/science/article/pii/S2352801X17302205.

Si, M., Xiong, Y., Du, S., Du, K., 2020. Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods. Atmos. Measure. Tech. 13, 1693–1707. URL: https://amt.copernicus.org/articles/13/1693/2020/. https://doi.org/10.5194/amt-13-1693-2020.

Suganya, S., Meyyappan, T., 2023. Prediction of the level of air pollution using adaptive neuro-fuzzy inference system. Multimed. Tools Appl. 82, 37131–37150. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85150231771&doi=10.1007%2fs11042-023-15046-0&partnerID=40&md5=59a03b250f0ad5a3cc87d84f8fb565d9.

Trach, R., Trach, Y., Kiersnowska, A., Markiewicz, A., Lendo-Siwicka, M., Rusakov, K., 2022. A study of assessment and prediction of water quality index using fuzzy logic and ann models. Sustainability 14. https://doi.org/10.3390/su14095656. URL: https://www.mdpi.com/2071-1050/14/9/5656.

Tran, H.M., Tsai, F.J., Lee, Y.L., Chang, J.H., Chang, L.T., Chang, T.Y., Chung, K.F., Kuo, H.P., Lee, K.Y., Chuang, K.J., et al., 2023. The impact of air pollution on respiratory diseases in an era of climate change: a review of the current evidence. Sci. Total Environ. 166340.

Wang, X., Yao, W., Luo, Q., Yun, J., 2024. Spatial relationship between ecosystem health and urbanization in coastal mountain city, Qingdao, China. Eco. Inform. 79, 102458. URL: https://www.sciencedirect.com/science/article/pii/S1574954123004879.

Weisberg, S., 2005. Applied Linear Regression, vol. 528. John Wiley & Sons.

WHO, et al., 2021. WHO Global Air Quality Guidelines: Particulate Matter (PM2. 5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization.

Wiseair,. Web Site. https://wiseair.vision/. Accessed: March 2024.

Won, W.S., Oh, R., Lee, W., Ku, S., Su, P.C., Yoon, Y.J., 2021. Hygroscopic properties of particulate matter and effects of their interactions with weather on visibility. Sci. Rep. 11 https://doi.org/10.1038/s41598-021-95834-6. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85112722798&doi=10.1038%2fs41598-021-95834-6&partnerID=40&md5=cc5cf7875496fab8e0872c9d0b88c08c. all Open Access, Gold Open Access.

Zadeh, L., 1965. Fuzzy sets. Inf. Control. 8, 338–353. URL: https://www.sciencedirect.com/science/article/pii/S001999586590241X.

Zadeh, L., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. 1, 3–28. URL: https://www.sciencedirect.com/science/article/pii/0165011478900295. https://doi.org/10.1016/0165-0114(78)90029-5.

Zadeh, L.A., Fu, K., Shimura, M., 1996. An introduction to fuzzy logic applications in intelligent systems. Proc. IEEE 83, 345–377.

Zusman, M., Schumacher, C.S., Gassett, A.J., Spalt, E.W., Austin, E., Larson, T.V., Carvlin, G., Seto, E., Kaufman, J.D., Sheppard, L., 2020. Calibration of low-cost particulate matter sensors: model development for a multi-city epidemiological study. Environ. Int. 134, 105329. URL: https://www.sciencedirect.com/science/article/pii/S0160412019321920.