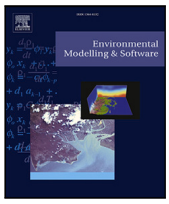
Environmental Modelling and Software 195 (2026) 106734

Contents lists available at ScienceDirect

Environmental Modelling and Software

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



Explainable AI for rule reduction in fuzzy models for air pollution measurement adjustment☆

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

[a] Faculty of Physics and Applied Computer Science, AGH University of Krakow, al. Adama Mickiewicza 30, Kraków, 30-059, Poland
[b] Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, Warszawa, 00-447, Poland
[c] "Enzo Ferrari" Engineering Department, University of Modena and Reggio Emilia, Modena, Italy

ARTICLE INFO

Dataset link: https://zenodo.org/records/13982208

Keywords:
Air quality
Particulate matter
Low-cost sensors
ANFIS
Explainable AI

ABSTRACT

Air quality monitoring using low-cost sensors has become increasingly important, yet their measurements are often inaccurate. Traditional adjustment methods face limitations in both applicability and explainability. This paper presents an explainable artificial intelligence approach for rule reduction in adaptive neuro-fuzzy inference systems, to improve the interpretability and efficiency of fuzzy models for fine particulate matter ($PM_{2.5}$) measurement adjustment. We introduce two novel algorithms, the Binary Activation Method and the Weighted Activation Method, to assess and eliminate redundant rules while maintaining predictive performance, validating the approaches in multiple geographic locations. On average, rule pruning results in an increase in MAE of 0.2 on the training set and 0.1 on the test set. The simplified models retain strong correlation, with Pearson's correlation coefficients ranging from 0.73 to 0.96 in the test set. These results support the development of reliable and interpretable artificial intelligence systems for environmental monitoring.

Software and data availability

Name of the software: ANFIS_PM_adjustment: AQDR - v1.1.0
Developers: Martina Casari, Piotr A. Kowalski
Contact information: martina.casari@unimore.it
First-year available: 2025
Software required: MATLAB
Programming language: MATLAB
Cost: free
Software availability and open data: Martina Casari et al. (2024).
Program size: less than 200 kB
License: CC BY 4.0
Data required for the local installation and use of software is accessed through the cloud.

1. Introduction

Air quality monitoring increasingly relies on dense networks of low-cost sensors, which enable high spatial resolution and participatory science but exhibit significant limitations in accuracy and reliability. Issues such as calibration drift, meteorological sensitivity, and sensor-to-sensor variability lead to biased and inconsistent measurements, making raw outputs unsuitable for decision-making without appropriate adjustment (Castelli et al., 2020; Guo et al., 2022). Addressing these challenges requires models that are not only accurate but also interpretable, so that corrections to raw data can be trusted in sensitive domains like public health and environmental policy. In this study, we explicitly address this challenge by investigating how explainable AI techniques can be applied to adaptive neuro-fuzzy inference systems (ANFIS) models in order to improve the calibration and adjustment of air pollution measurements.

Fuzzy logic, and in particular ANFIS, provide a promising framework for this task. ANFIS integrates the learning ability of neural networks with the interpretability of fuzzy inference, allowing non-linear relationships to be captured while maintaining a rule-based structure (Jang, 1993b). This makes it particularly suitable for sensor calibration tasks, where robustness and transparency are equally important. However, automatically trained ANFIS models often produce large rule bases that reduce interpretability and increase computational costs.

Explainable Artificial Intelligence (XAI) techniques offer a way to overcome these issues by guiding model simplification and enhancing transparency. Methods such as rule pruning, feature attribution,

☆ This article is part of a Special issue entitled: 'Air quality modelling' published in Environmental Modelling and Software.
* Corresponding author.
E-mail addresses: pkowal@agh.edu.pl (P.A. Kowalski), martina.casari@unimore.it (M. Casari), laura.po@unimore.it (L. Po).

https://doi.org/10.1016/j.envsoft.2025.106734
Received 27 May 2025; Received in revised form 11 September 2025; Accepted 6 October 2025
Available online 11 October 2025

and interpretability-driven optimisation have been applied in fields including medicine, environmental modelling, and safety-critical systems (Arrieta et al., 2020; Holzinger et al., 2017). Building on these approaches, we explore how XAI can be integrated with ANFIS to improve model compactness and interpretability in the context of air pollution measurement adjustment.

In this paper, we focus specifically on a real-world case study: applying XAI methods to support rule reduction in ANFIS trained on multi-location air pollution datasets. Unlike studies that rely on synthetic or benchmark data, we demonstrate how explainability can directly improve the reliability and deployment readiness of models used for air quality monitoring.

AI-based approaches to air pollution modelling have been widely investigated, from regression and machine learning methods to deep learning architectures (Castelli et al., 2020; Guo et al., 2022; Wang et al., 2022). In this landscape, hybrid neuro-fuzzy methods provide a valuable balance between predictive performance and interpretability. For example, Casari et al. (2024b) demonstrated the effectiveness of ANFIS for calibrating low-cost $PM_{2.5}$ sensors, comparing different rule-generation strategies and benchmark machine learning techniques.

Interpretability has also emerged as a key focus in environmental modelling. Studies have applied techniques such as SHAP and LIME to identify the importance of meteorological features in air pollution prediction, showing, for instance, the critical role of humidity in shaping sensor responses (Casari et al., 2024a; Casari and Po, 2024c). These works highlight the value of explainability but typically stop short of addressing model simplification. Our study extends this line of research by introducing XAI-based rule reduction methods for ANFIS, explicitly targeting the interpretability–complexity trade-off in environmental applications.

Beyond these contributions, Concas et al. (2021) provided a broad survey of low-cost air quality monitoring, identifying calibration drift and environmental variability as persistent challenges for sensor reliability. Alhasa et al. (2018) further validated ANFIS for the calibration of electrochemical ozone sensors, showing that hybrid fuzzy-neural models consistently outperform linear and standard machine learning approaches when meteorological covariates are incorporated.

From a broader engineering perspective, systematic reviews in Bressane et al. (2024) and Korkidis and Dounis (2023) demonstrate that fuzzy-ML models consistently outperform traditional methods in scenarios marked by uncertainty, highlighting their transparent and interpretable nature in environmental engineering applications.

In the realm of XAI for fuzzy systems and neuro-fuzzy models, Talpur et al. (2023) provide a systematic survey of deep neuro-fuzzy system architectures, optimisation methods, and future directions, underscoring the role of hybrid neuro-fuzzy systems in bridging performance and explainability. Islam et al. (2019) proposes fuzzy integral neural networks (ChIMP/iChIMP) to enable explainable fusion in deep learning, with applications in remote sensing, illustrating how fuzzy aggregation can support interpretability in sensor-based domains.

Kusy et al. addressed the explainability of air pollution prediction models by applying machine learning techniques combined with XAI methods such as SHAP and LIME (Kusy et al., 2022). Their work highlights how interpretability tools can uncover the influence of meteorological and environmental variables on model predictions, demonstrating the practical role of XAI in increasing trust in data-driven air quality forecasting.

There is growing interest in advanced calibration paradigms such as meta-learning. Authors in Yadav et al. (2022) introduced a few-shot calibration method for $PM_{2.5}$ sensors using model-agnostic meta-learning (MAML), enabling sensor calibration with minimal co-deployment with reference monitors.

The motivation of this study is twofold. First, we tackle the challenge of calibrating low-cost sensors by using ANFIS, which naturally combines expert knowledge with data-driven learning. Second, we demonstrate how XAI techniques can improve interpretability by enabling principled rule reduction, ensuring that models remain compact, accurate, and transparent. Importantly, the focus of this paper is not on developing a generic framework for XAI in neuro-fuzzy systems, but on showing how such techniques can be concretely applied to the domain of air quality monitoring as a real-world case study. This combination directly addresses the practical requirements of environmental monitoring, where both predictive reliability and explainability are essential.

This work introduces two entirely novel algorithms for interpreting the relative importance of individual rules within ANFIS. Unlike conventional pruning approaches that rely on heuristic metrics or global model sensitivity, the proposed methods evaluate the contribution of each rule to the model's inference process in a context-sensitive and explainable manner. By grounding the analysis in real-world environmental data, we show that XAI-driven rule reduction can yield compact ANFIS models that retain high calibration performance while significantly improving interpretability.

This behaviour is highly desirable in practical applications, especially in resource-constrained environments. The ability to reduce computational costs, memory requirements, and energy consumption—while maintaining near-original model quality—makes the proposed approach an attractive solution for real-time or embedded deployment of neuro-fuzzy systems. The proposed rule importance evaluation mechanisms also contribute to the broader field of explainable AI, offering insights into internal model functioning that support informed decisions about model optimisation and trustworthiness.

The paper is structured as follows. Section 2 describes the ANFIS architecture and the proposed Binary and Weighted Activation Methods for rule importance evaluation. Section 3 presents experimental validation using real-world datasets, showing how rule reduction enhances interpretability with minimal accuracy loss. Conclusions and future directions are summarised in Section 4.

2. Modelling approach and fuzzy logic explainable techniques

Fuzzy logic (FL) provides a well-established mathematical framework for reasoning under conditions of uncertainty, vagueness, and imprecision (Zadeh et al., 1996; Zadeh, 1978). Unlike classical binary logic, which constrains statements to truth values of either 0 or 1, FL allows for intermediate degrees of truth expressed through membership functions. This property makes fuzzy systems particularly suited to modelling real-world phenomena in which linguistic variables such as *low*, *medium*, and *high* cannot be crisply defined. A fuzzy inference system (FIS) thus offers a structured mechanism for capturing human-like reasoning in a way that remains interpretable to domain experts.

Since its inception, fuzzy logic has been applied successfully in diverse fields, including process control, image analysis, artificial intelligence, and information systems, where ambiguous or noisy data must be interpreted. Recent advances have also extended its theoretical foundations: algebraic contributions on fuzzy classical two-absorbing secondary sub-modules (Hanoon and Ibrahim, 2024), fuzzy soc-small two-absorbing modules (Hanoon et al., 2024), and optimisation with metaheuristic algorithms in fuzzy environments (Sutikno, 2023) illustrate the breadth of current research directions.

A standard fuzzy inference system can be described in terms of five interconnected components: fuzzification, a knowledge base, an inference engine, aggregation, and defuzzification. In the fuzzification stage, crisp numerical inputs are mapped to fuzzy sets through membership functions (MF). Given a domain of input values X and a specific observation $x \in X$, a fuzzy set A can be defined as:

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

where $\mathrm{MF}(x)$ is a membership degree between 0 and 1 indicating the extent to which the observation belongs to a category. For instance, a temperature of 25 °C might simultaneously belong to the categories

Table 1
Commonly used defuzzification methods in fuzzy inference systems.

Method	Definition	Advantages/Limitations
Centroid (Centre of Gravity)	$x^* = \frac{\int x\,\mu(x)\,dx}{\int \mu(x)\,dx}$	Most widely used; provides smooth and balanced outputs. Computationally more demanding due to integration.
Bisector of area	Finds x^* that divides the fuzzy set area into two equal parts	Simple conceptually; less sensitive to extreme values. May produce discontinuous outputs in asymmetric sets.
Mean of Maxima (MOM)	$x^* = \frac{1}{n}\sum_{i=1}^{n} x_i$, where x_i are points with maximal membership degree	Very efficient; easy to implement. However, ignores the overall shape of the membership function.
Smallest/Largest of maxima	Chooses the smallest (SOM) or largest (LOM) x_i with maximal membership	Extremely simple; but highly sensitive to noise and not representative of the fuzzy set as a whole.

warm (0.8) and *hot* (0.3). Membership functions are typically chosen in triangular, trapezoidal, Gaussian, or sigmoid forms, depending on the problem domain.

The knowledge base encodes expert knowledge in the form of `IF--THEN` rules. Each rule combines antecedents (conditions on the input variables) and consequents (system responses). Logical operators are implemented using Zadeh's fuzzy operators:

$$\text{NOT } x = 1 - \text{MF}(x), \quad x \text{ AND } y = \min(\text{MF}(x), \text{MF}(y)),$$
$$x \text{ OR } y = \max(\text{MF}(x), \text{MF}(y)).$$

Although the number of possible rules can grow rapidly with the dimensionality of the input space, in practice the complexity is more strongly influenced by the granularity of the membership functions rather than the absolute number of rules (Gegov et al., 2017).

The inference engine evaluates the rules using the fuzzified input variables and produces fuzzy conclusions. Rules can be weighted to adjust their relative influence on the output. The results are aggregated into a single fuzzy output set, which represents the system response. Finally, the defuzzification stage converts this fuzzy set back into a crisp value suitable for practical interpretation. Among several methods, the centroid technique, which computes the centre of mass of the output fuzzy set, is the most widely adopted. A concise comparison of commonly used defuzzification strategies, including centroid, bisector, and mean of maxima, and their respective characteristics, is presented in Table 1.

This modelling approach not only accommodates imprecise and uncertain data but also produces interpretable structures in the form of rules and membership functions. As such, fuzzy inference systems offer a natural foundation for developing explainable techniques in hybrid architectures such as the adaptive neuro-fuzzy inference system, which integrates fuzzy reasoning with neural network learning.

ANFIS, originally proposed by Jang and Sun in 1994 (Jang, 1993a), represents a powerful framework for modelling nonlinear, uncertain, or imprecise data by integrating fuzzy inference with neural network learning. This hybrid structure enhances transparency while adapting to data-driven patterns, enabling the modelling of complex nonlinear relationships with both accuracy and interpretability (Sayyaadi, 2021; Chanal et al., 2022).

In practice, ANFIS begins with the construction of a fuzzy inference system based on initial membership functions and rule structures. Through supervised training, commonly using a hybrid learning algorithm that combines backpropagation with least-squares estimation (Karaboga and Kaya, 2019), the system iteratively adjusts both the membership functions and the rule parameters to minimise error and better fit the observed data. A key feature of this architecture is that, even after optimisation, the fuzzy rules remain accessible and understandable, thus preserving interpretability, a property of critical importance in high-stakes applications where model outputs must be verified and certified.

ANFIS can be viewed from two complementary perspectives: either as a fuzzy inference system enriched with neural network learning capabilities, or as a neural network that embeds linguistic information contributed by domain experts. This dual interpretation underscores the framework's versatility. Fig. 1 presents the canonical five-layer feedforward ANFIS architecture, while Fig. 2 illustrates the membership functions at the fuzzification stage.

The operation of ANFIS can be formalised as follows. In the first layer, fuzzification is performed by adaptive nodes equipped with membership functions. For instance, bell-shaped functions are defined as

$$\mu_{A_i}(x) = \frac{1}{1 + \left|\frac{x - c_i}{a_i}\right|^{2b_i}},$$

where $\{a_i, b_i, c_i\}$ are premise parameters to be tuned during training. In the second layer, the firing strength of each fuzzy rule is computed as the product of the corresponding membership degrees, i.e.

$$w_i = \mu_{A_1}(x) \times \mu_{B_1}(y),$$

which is equivalent to a fuzzy logical *AND*. The third layer normalises these firing strengths across all rules:

$$\overline{w_i} = \frac{w_i}{\sum_{j=1}^{N} w_j}.$$

In the fourth layer, each rule contributes a linear function of the inputs weighted by its normalised firing strength:

$$f_i = \overline{w_i}(p_i x + q_i y + r_i),$$

where $\{p_i, q_i, r_i\}$ are consequent parameters updated during learning. Finally, in the fifth layer, all rule outputs are aggregated into the system output:

$$f = \sum_{i=1}^{N} \overline{w_i} f_i.$$

This layered representation illustrates how ANFIS retains the transparency of fuzzy inference while enhancing flexibility through data-driven training. It thus occupies a unique position among neuro-fuzzy systems, combining interpretability with predictive accuracy.

While ANFIS provides both accuracy and interpretability, automatically generated models often become too complex due to an excessive number of rules. To address this, we introduce two novel explainability-driven rule reduction strategies, presented in the following section.

2.1. Fuzzy rule usage analysis

To assess the contribution of individual fuzzy rules during model training, two complementary approaches were used to evaluate their activation frequency: a *Baseline Activation Method* (BAM) and a *Weighted*

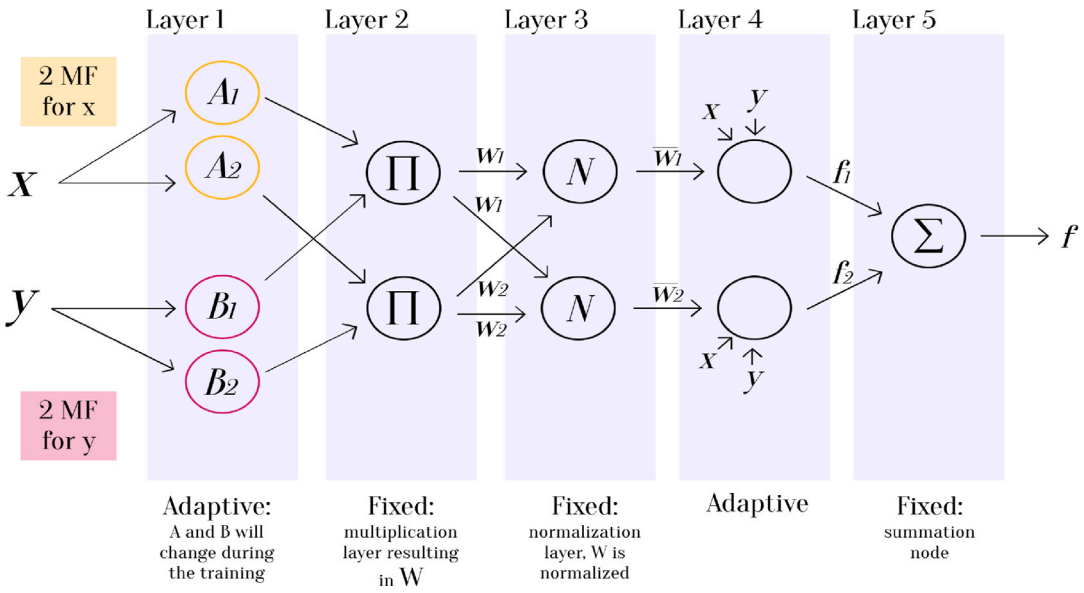


Fig. 1. The canonical ANFIS architecture consisting of five computational layers.

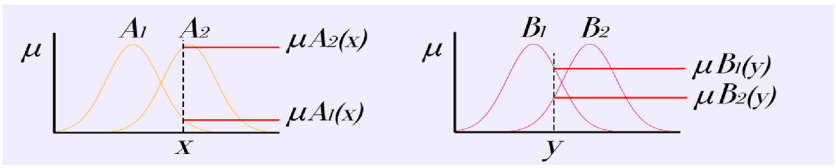


Fig. 2. Membership functions at the fuzzification layer for a two-input system.

Activation Method (WAM). Both techniques aim to quantify the participation of each rule in the inference process, informing subsequent decisions about rule pruning and system simplification.

Let $\alpha_r^{(i)}$ denote the activation level of rule r for the ith training sample. A rule is considered *active* for a sample if $\alpha_r^{(i)} > 0$. The evaluation process is structured into four main steps:

1. **Rule Evaluation:** For each input sample, the activation level $\alpha_r^{(i)}$ of every fuzzy rule is computed. This involves evaluating the membership degrees of each antecedent and taking their minimum to determine the rule activation.
2. **Activation Counting:** Rules with non-zero activation are considered active and contribute to the usage statistics.
3. **Frequency Computation:** After processing the entire dataset, usage counters are normalised by the total number of training samples to obtain relative activation frequencies.
4. **Rule Pruning**: applies a common iterative procedure (detailed in Section 2.1.3) that progressively removes the least used rules, regardless of whether BAM or WAM is used for the frequency computation.

In the following sections, detailed examinations of the BAM (Section 2.1.1) and WAM (Section 2.1.2) approaches are provided, along with the iterative procedure used to progressively remove rules and evaluate their impact on model performance, which is described as a common component in Section 2.1.3.

2.1.1. Binary Activation Method

In the Binary Activation Method, each rule is treated equally once activated, regardless of its actual activation strength. This method simply counts how many times each rule is triggered across the dataset. This method is useful when interpretability and simplicity are prioritised over a nuanced understanding of rule strength.

If the rule r is active for the sample i, i.e., $\alpha_r^{(i)} > 0$, its counter is incremented as:

$$\text{UsageCounter}_r \leftarrow \text{UsageCounter}_r + 1. \tag{1}$$

After all N samples have been processed, the relative frequency of rule activation is given by:

$$f_r^{(d)} = \frac{\text{UsageCounter}_r}{N}. \tag{2}$$

The straightforward approach to rule reduction involves iteratively removing the least frequently activated rules based on their usage counts; in this way, the system can be simplified with minimal expected loss in performance. The resulting value f_r corresponds to the list that, when sorted, is used for rule pruning (as represented by `ruleUsageCounter` in line 1 of Algorithm 1).

Its main advantage lies in its ease of implementation. However, the method also presents limitations, as it implicitly assumes that rules activated infrequently are unimportant, an assumption that may not hold in cases where rarely triggered rules play a crucial role in exceptional but meaningful scenarios.

To represent the BAM for the entire analysis over each dataset, the BAM was independently normalised using min–max scaling for each dataset. Let $f_r^{(d)}$ denote the binary activation count of a rule r in the dataset d, and let $f_{\min}^{(d)}$ and $f_{\max}^{(d)}$ denote the minimum and maximum rule counts in that dataset, respectively. The min–max normalisation is defined as:

$$\tilde{f}_r^{(d)} = \frac{f_r^{(d)} - f_{\min}^{(d)}}{f_{\max}^{(d)} - f_{\min}^{(d)}}, \tag{3}$$

where $\tilde{f}_r^{(d)}$ is the normalised activation frequency of the rule r in the dataset d.

After computing $\tilde{f}_r^{(d)}$ for all datasets, the final value for each rule is obtained by averaging the normalised frequencies in all D datasets:

$$\bar{f}_r = \frac{1}{D}\sum_{d=1}^{D} \tilde{f}_r^{(d)}. \tag{4}$$

This process yields a dataset-agnostic representation of rule usage, facilitating fair comparison and further analysis.

2.1.2. Weighted Activation Method

In the second, more refined approach, the Weighted Activation Method, the actual activation strength of each rule was accumulated across the dataset. Instead of incrementing by one, the counter for each rule was increased by its activation value. This provides a better estimate of how strongly each rule contributes to the model's output.

Formally, if $\alpha_r^{(i)} > 0$, then:

$$\text{UsageCounter}_r \leftarrow \text{UsageCounter}_r + \alpha_r^{(i)}. \tag{5}$$

This method emphasises the influence of highly activated rules and results in a normalised usage profile that reflects both frequency and strength of activations:

$$f_r^{(d)} = \frac{\sum_{i=1}^{N} \alpha_r^{(i)}}{N}. \tag{6}$$

As for BAM, the resulting value $f_r^{(d)}$ corresponds to the list that is sorted during the pruning procedure (i.e., `ruleUsageCounter` in line 1 of Algorithm 1).

This weighted approach addresses the key limitation of the binary count method. While both techniques yield useful insights into rule relevance, the binary method may disproportionately highlight rules that are frequently activated with low intensity. In contrast, the weighted method provides a more nuanced perspective by emphasising rules that consistently exhibit strong activations, offering a better foundation for model refinement and pruning.

2.1.3. Rule pruning procedure

The pruning procedure described here is applied identically regardless of whether the rule usage statistics are obtained through BAM or WAM. In both cases, fuzzy rules are ordered according to their activation statistics, and the least frequently activated rules are progressively removed. At each iteration, the model is re-evaluated to monitor changes in performance. The main steps of the procedure are outlined in Algorithm 1. To provide a deeper understanding, each step of the algorithm is detailed below:

1. **Sorting the Rules:** (Line 1 of Algorithm 1) The fuzzy rules are first ranked based on their activation frequency. Rules that are triggered more frequently are considered more important. This ranking is obtained by sorting the `ruleUsageCounter` array in descending order:
2. **Initialisation:** (Prior to the Loop in Algorithm 1) Arrays or variables are prepared to track model performance metrics during each iteration, including the Pearson Correlation Coefficient (PCC), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) for both training and test datasets.
3. **Iterative Rule Pruning and Model Tuning:** (Loop in Algorithm 1) The algorithm iteratively reduces the rule base, retrains the system, and evaluates performance as follows:
   - **Rule Subset Selection:** (Lines 2–3 of Algorithm 1) In each iteration $k \in \{1, \ldots, N-1\}$, where N is the total number of rules, the top $N-k$ most frequently used rules are retained:
   - **FIS Reduction:** (Block 4 of Algorithm 1) A new reduced FIS is created by keeping only the rules and output membership functions associated with the selected indices:

Algorithm 1: Reduced Model Tuning

```
   // Sorting the Rules
1  sortedFrequencies, sortedIndices ←
     sort(ruleUsageCounter, descending)
   // Iterative Rule Pruning and Model Tuning
2  for iter ← 1 to numRules - 1 do
       // Rule Subset Selection
3      selectedRulesIdx ← sortedIndices[1 : end - iter +
         1]
       // FIS Reduction
4      reducedFIS ← copy of outputFIS
5      reducedFIS[Rules] ←
         outputFIS.Rules[selectedRulesIdx]
6      reducedFIS.Output.MembershipFunctions ← output-
         FIS.Output.MembershipFunctions[selectedRulesIdx]
       // ANFIS-Based Tuning
7      opt ← create tuning options with method "anfis"
8      in, out, rule ← get tunable settings from reducedFIS
9      retrainedFIS ← tune reducedFIS using [in; out], x, y,
         and opt
       // Model Evaluation
10     resultTrain ← evaluate retrainedFIS using
         completeTrainTable and currentCombination
```

   - **ANFIS-Based Tuning:** (Block 5 of Algorithm 1) The reduced FIS is then retrained using ANFIS to fine-tune the system parameters (input and output membership functions). This step compensates for the removed rules by reoptimizing the FIS:
   - **Model Evaluation:** (Line 6 of Algorithm 1) The retrained FIS is evaluated on the training dataset to measure its performance:
4. **Performance Tracking:** (Throughout Algorithm 1) At each iteration, performance metrics are recorded to monitor how the gradual reduction in rule base size affects predictive accuracy. This allows for identifying a balance between system complexity and performance, enabling the selection of an optimally reduced model.

This process enables a gradual simplification of the fuzzy model, identifying the minimal rule set that retains acceptable performance. It is especially useful for reducing computational complexity and improving interpretability without significantly degrading predictive quality.

3. Numerical verification of proposed XAI procedures

3.1. Datasets

The datasets used in this study were acquired from multiple sources, including three Italian datasets obtained through a collaborative effort with Regional Agencies for Environmental Protection (ARPA) and low-cost sensors (LCS) supplied bysupplied by Wiseair SRL (Wiseair, 2024), a Milan-based startup focused on data-driven solutions for optimising commuting patterns and generating measurable impacts on workplace productivity. In these deployments, SPS30 sensors (Sensirion, 2023) were co-located with ARPA reference stations (RS) across three urban environments: Reggio Emilia, Trento, and Aosta. The raw sensor data were recorded at a temporal resolution of 15 min, while the RS measurements were sampled every hour. To increase geographical diversity, open-source datasets from Calgary (Canada) (Si et al., 2020; Si, 2019) and Lima (Peru) (Villanueva et al., 2023) were also incorporated, selected due to the availability of aligned and co-located LCS and RS. The simultaneous availability of data from both co-located RS and LCS is essential, as it allows for the calibration and adjustment of low-cost sensor readings. The RS data serve as a reliable benchmark, enabling

Table 2
Dataset dimensions and temporal coverage.

City	Train samples	Test samples	Total	Date range
Aosta	4534	1530	6064	2024-02–2024-06
Reggio Emilia	3440	1147	4587	2024-09–2024-12
Trento	1586	532	2118	2024-11–2025-02
Lima	3201	1078	4279	2021-11–2022-01
Calgary	2268	758	3026	2018-12–2019-04

the development of models that improve the accuracy and reliability of LCS measurements.

Data preprocessing. To ensure consistency across locations, both LCS and RS data were temporally aligned and resampled to an hourly resolution. As shown in Table 2, which reports sample sizes and temporal coverage for each dataset, a monthly stratified 75/25 train-test split was applied. This approach systematically allocated 25% of each month's observations to the test set, ensuring seasonal representativeness.

Table 2 summarises the number of training and test samples, along with the total sample size and temporal coverage for each location.

The preprocessing pipeline included the following key steps:

- **Missing data imputation**: Gaps in the time series were filled using linear interpolation
- **Data quality control**: Measurements exceeding physically plausible thresholds were discarded
- **Feature selection**: The input of the model is created by including:
  - Low-cost sensor $PM_{2.5}$ measurements
  - Meteorological parameters (relative humidity, temperature, pressure)
- **Target variable**: $PM_{2.5}$ concentrations from RSs
- **Standardisation**: All features were standardised after the train-test split.

Fig. 3 presents one-month excerpts of $PM_{2.5}$ sensor data collected from three representative locations: Aosta, Trento, and Calgary. Each subplot compares measurements from two types of sensors: LCS in blue and RS in black.

- **Aosta** (Fig. 3(a)): February 2024 data exhibit multiple peaks in $PM_{2.5}$ concentration, reaching 50–60 $\mu g/m^3$. LCS and RS measurements follow similar temporal trends, although LCS readings are consistently slightly higher during peaks.
- **Trento** (Fig. 3(b)): Data from November to December 2024 show greater variability, with LCS occasionally recording values above 150 $\mu g/m^3$. Although general trends are aligned, LCS readings tend to overestimate during high-pollution episodes.
- **Calgary** (Fig. 3(c)): In December 2018 – January 2019, $PM_{2.5}$ levels remained relatively low (mostly under 20 $\mu g/m^3$), with only a few episodic peaks. During such peaks, the discrepancy between LCS and RS measurements becomes more pronounced, indicating increased sensor sensitivity under elevated pollution conditions.

This illustration provides a broader overview of the available data, highlighting how location significantly influences $PM_{2.5}$ levels even when using the same low-cost sensor model. In particular, it contrasts Aosta and Trento, where the same sensor (SPS30) was deployed but under different environmental conditions, resulting in notably different $PM_{2.5}$ patterns. Moreover, it shows that Aosta and Calgary, despite using different low-cost sensors (SPS30 and PMS5003, respectively), exhibit similar $PM_{2.5}$ concentration ranges, suggesting that location plays a prominent role along with the sensor type in determining the observed values.

3.2. Results and discussion

This section reports the results of fuzzy rule pruning experiments, conducted using the two proposed strategies to evaluate the usage of rules. The goal is to investigate to what extent each method enables rule reduction while maintaining the predictive performance of the FIS.

3.2.1. Binary Activation Method Results

This section reports the results obtained using the BAM, as exposed in Section 2.1.1, where rules are evaluated solely based on how frequently they are triggered, regardless of the strength of their activation. The focus is on how performance metrics, such as Mean Absolute Error and the Pearson correlation coefficient, evolve as less frequently activated rules are iteratively removed from the system.

Fig. 4 presents a heatmap of 48 fuzzy rules, each defined by linguistic labels for the variables `pm2p5_x`, `relative_humidity`, `temperature`, and `pressure`, along with their corresponding mean Activation Count Normalised (ACN), ranging from 0 to 1. The first column of the table represents the unique identifier assigned to each rule. Each row corresponds to a fuzzy rule defined by a conjunction (i.e., logical AND) of conditions on the input features, which form the rule's antecedent. The consequent of each rule, representing the predicted value of *pm2p5_y* (i.e., the $PM_{2.5}$ concentration measured by the reference sensor), is not shown, as it is not informative for interpretability. The figure also reports the mean ACN value for each rule, computed using the Binary Activation Method, reflecting the relative importance or frequency of activation of each rule across the dataset.

The ACN is computed to enable a fair comparison of rule usage frequencies across multiple datasets (e.g., Aosta, Trento, etc.), following the procedure detailed in Section 2.1.1 and formalised in Eqs. (1) and (2).

Rules exhibiting high activation values (ACN > 0.75) are predominantly associated with low `pm2p5_x` levels in combination with high or very high `relative_humidity`, where the `temperature` level is equally distributed. For instance, Rule 13 (`LOW`, `HIGH`, `LOW`, `HIGH`) and Rule 17 (`LOW`, `VERY HIGH`, `LOW`, `HIGH`) achieve ACN values of 0.835 and 0.800, respectively. These configurations suggest that under conditions of low particulate concentration, elevated humidity, and pressure significantly contribute to rule activation.

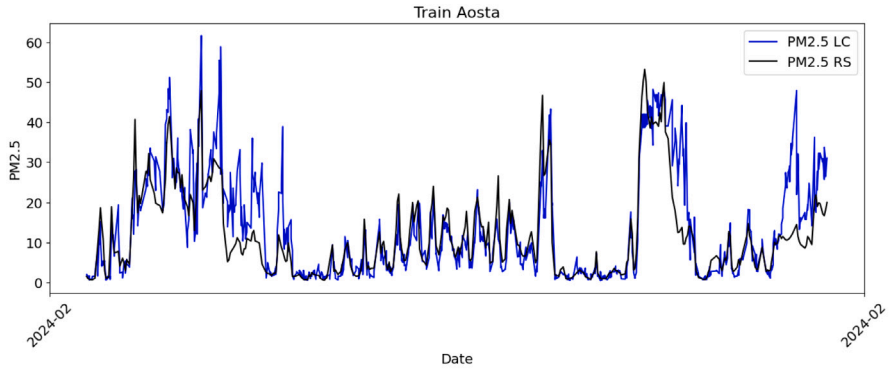
Conversely, rules with low activation values (ACN < 0.1) are typically linked to low or high `pm2p5_x` values combined with very low `relative_humidity`, regardless of temperature and pressure. Rules 24 to 27 (e.g., `HIGH`, `VERY LOW`, `LOW`, `LOW`) demonstrate minimal activation (ACN ≈ 0), suggesting these environmental states are either rare or considered non-significant by the system.

Medium activation values (ACN between 0.3 and 0.7) are commonly found in rules involving `LOW`, `MEDIUM` or `SATURATED` humidity levels, indicating moderately relevant conditions. Notably, the role of `temperature` is less decisive, as both low and high values appear across the activation spectrum, suggesting a more nuanced influence in the rule base. In contrast, high `pressure` is frequently linked to rules with elevated ACN values compared to low pressure, whereas low `pressure` is more typical of rules with lower activations.

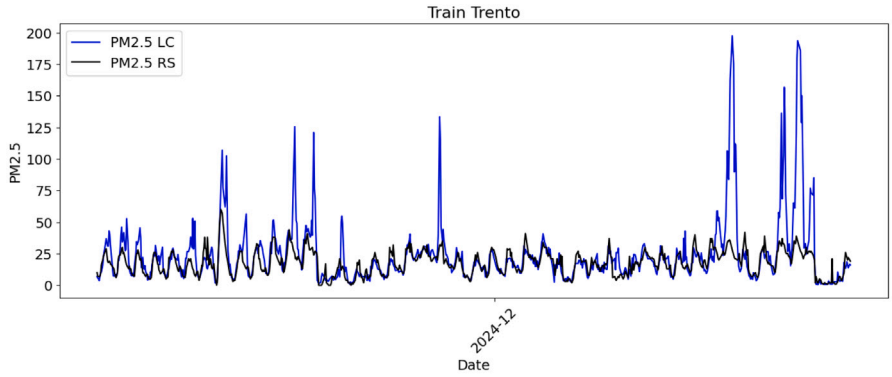
Overall, the rule activation pattern highlights the importance of `relative_humidity` and `pressure` in shaping the system's interpretation of environmental conditions, particularly in conjunction with low `pm2p5_x` levels. These insights may support the design of more targeted monitoring strategies or adaptive decision-making mechanisms in air quality or ambient sensing applications.

While Fig. 4 provides a numerical summary of normalised average activations across datasets, Fig. 5 shows a visual representation of rule activations per dataset.

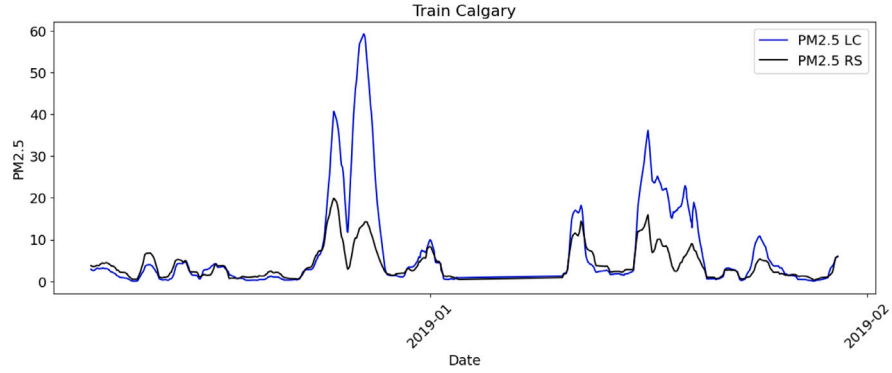
The plot reveals two distinct macro areas where rules are predominantly active across the datasets, with variations in their importance.



(a) Zoomed view of one month of data from Aosta.



(b) Zoomed view of one month of data from Trento.



(c) Zoomed view of one month of data from Calgary.

Fig. 3. One-month excerpts of sensor data for three selected locations: Aosta, Trento, and Calgary.

For example, Calgary and Trento activate distinct sets of rules, while Aosta shows a broader distribution of active rules compared to the others. Trento and Reggio display similar activation patterns in the initial phase, though Reggio shows a lower activation count in the second macro area. The middle and upper sections of the plot indicate that certain scenarios are less frequently observed or activated. These rules are primarily the ones pruned by the system without significantly impacting overall performance.

These activation patterns, highlighted by the black line representing the mean, align with the global average values shown in Fig. 4, as also reflected by its colour scale.

Fig. 6 illustrates the progression of the MAE during the rule pruning process for both the training and test datasets. The MAE is computed separately for each dataset (i.e., for each location) at every pruning step; the final plot summarises this information by reporting the average MAE across all locations, along with the range of variation. The close overlap between the two curves indicates consistent model behaviour across data splits. The solid lines represent the average MAE across all locations, while the shaded areas denote the range of variation observed for each dataset. The MAE remains unchanged in the initial pruning iterations, suggesting that removing the least frequently activated rules does not degrade performance. However, beyond a certain point, further pruning leads to a noticeable increase in MAE, indicating that rules with substantive predictive value are being removed, thereby affecting the system's accuracy.

3.2.2. *Weighted Activation Method Results*

This section presents the results of the rule pruning process driven by the WAM, see Section 2.1.2 for details. Unlike the binary strategy,

ACN 0.0 0.2 0.4 0.6 0.8 1.0

Rule Number	pm2p5_x	relative_humidity	temperature	pressure	ACN
0	LOW	VERY LOW	LOW	LOW	0.039
1	LOW	VERY LOW	LOW	HIGH	0.043
2	LOW	VERY LOW	HIGH	LOW	0.048
3	LOW	VERY LOW	HIGH	HIGH	0.051
4	LOW	LOW	LOW	LOW	0.389
5	LOW	LOW	LOW	HIGH	0.395
6	LOW	LOW	HIGH	LOW	0.395
7	LOW	LOW	HIGH	HIGH	0.400
8	LOW	MEDIUM	LOW	LOW	0.663
9	LOW	MEDIUM	LOW	HIGH	0.677
10	LOW	MEDIUM	HIGH	LOW	0.631
11	LOW	MEDIUM	HIGH	HIGH	0.638
12	LOW	HIGH	LOW	LOW	0.809
13	LOW	HIGH	LOW	HIGH	0.835
14	LOW	HIGH	HIGH	LOW	0.766
15	LOW	HIGH	HIGH	HIGH	0.783
16	LOW	VERY HIGH	LOW	LOW	0.788
17	LOW	VERY HIGH	LOW	HIGH	0.800
18	LOW	VERY HIGH	HIGH	LOW	0.766
19	LOW	VERY HIGH	HIGH	HIGH	0.774
20	LOW	SATURATED	LOW	LOW	0.357
21	LOW	SATURATED	LOW	HIGH	0.355
22	LOW	SATURATED	HIGH	LOW	0.347
23	LOW	SATURATED	HIGH	HIGH	0.345

Rule Number	pm2p5_x	relative_humidity	temperature	pressure	ACN
24	HIGH	VERY LOW	LOW	LOW	0.000
25	HIGH	VERY LOW	LOW	HIGH	0.001
26	HIGH	VERY LOW	HIGH	LOW	0.003
27	HIGH	VERY LOW	HIGH	HIGH	0.003
28	HIGH	LOW	LOW	LOW	0.247
29	HIGH	LOW	LOW	HIGH	0.249
30	HIGH	LOW	HIGH	LOW	0.245
31	HIGH	LOW	HIGH	HIGH	0.247
32	HIGH	MEDIUM	LOW	LOW	0.478
33	HIGH	MEDIUM	LOW	HIGH	0.488
34	HIGH	MEDIUM	HIGH	LOW	0.445
35	HIGH	MEDIUM	HIGH	HIGH	0.450
36	HIGH	HIGH	LOW	LOW	0.633
37	HIGH	HIGH	LOW	HIGH	0.646
38	HIGH	HIGH	HIGH	LOW	0.594
39	HIGH	HIGH	HIGH	HIGH	0.600
40	HIGH	VERY HIGH	LOW	LOW	0.668
41	HIGH	VERY HIGH	LOW	HIGH	0.670
42	HIGH	VERY HIGH	HIGH	LOW	0.649
43	HIGH	VERY HIGH	HIGH	HIGH	0.649
44	HIGH	SATURATED	LOW	LOW	0.303
45	HIGH	SATURATED	LOW	HIGH	0.301
46	HIGH	SATURATED	HIGH	LOW	0.292
47	HIGH	SATURATED	HIGH	HIGH	0.291

Fig. 4. Fuzzy inference rules based on `pm2.5_x` (low-cost sensors), relative humidity, temperature, and pressure. Each rule is accompanied by its corresponding mean activation value normalised, indicating its relevance across datasets.

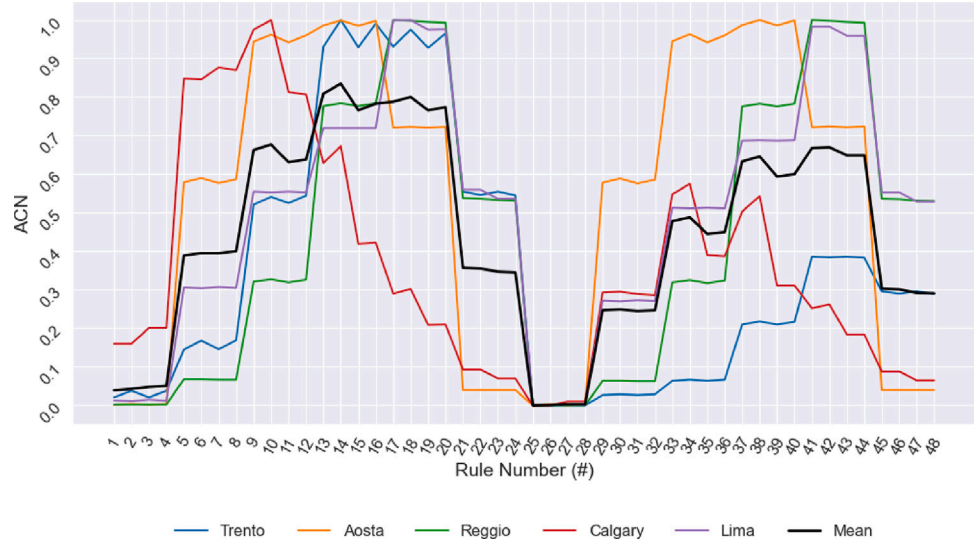


Fig. 5. Normalised activation frequency of the rules in each dataset compared to the activation count normalised (ACN) by location (per rule).

this approach considers the intensity of the activation of each rule, providing a more nuanced basis for pruning. The evaluation again focuses on the evolution of key performance metrics as rules with lower cumulative activation are progressively excluded.

As shown in Fig. 7, adopting a more refined rule evaluation strategy helps the model maintain stronger performance during the initial stages of pruning. In comparison to Fig. 6, the MAE in Fig. 7 remains lower and more stable across a greater number of pruning iterations. This indicates that WAM is more effective in preserving model accuracy during rule reduction. As a result, this approach supports the construction of a more compact FIS, allowing a larger number of rules to be removed without significantly compromising performance.

While this trend is visually evident in the figure, where performance remains stable over successive pruning steps, it is further substantiated

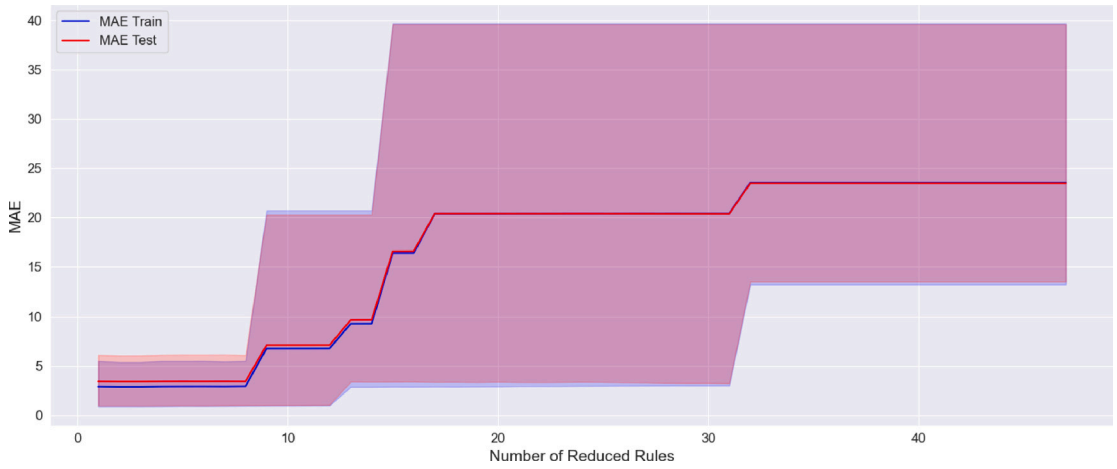


Fig. 6. MAE variation as a function of the number of reduced rules across locations, based on BAM.

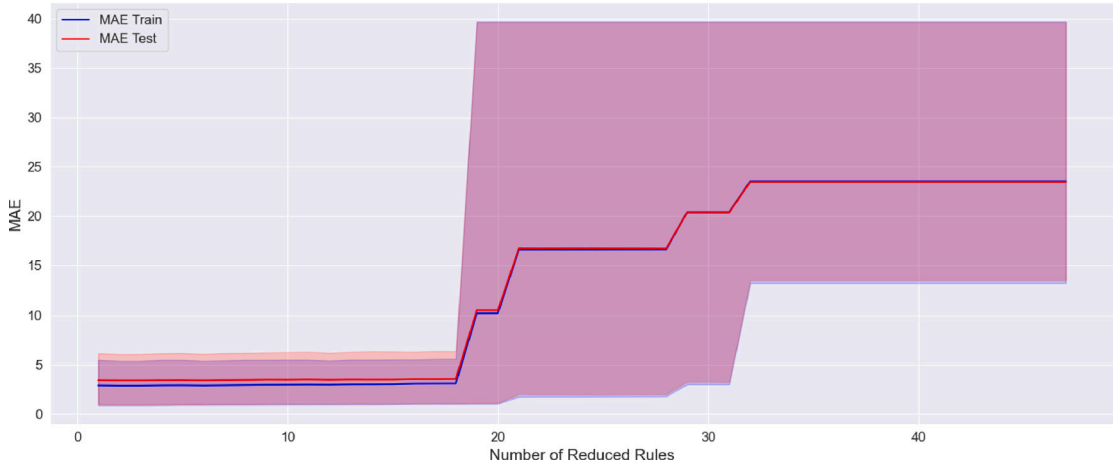


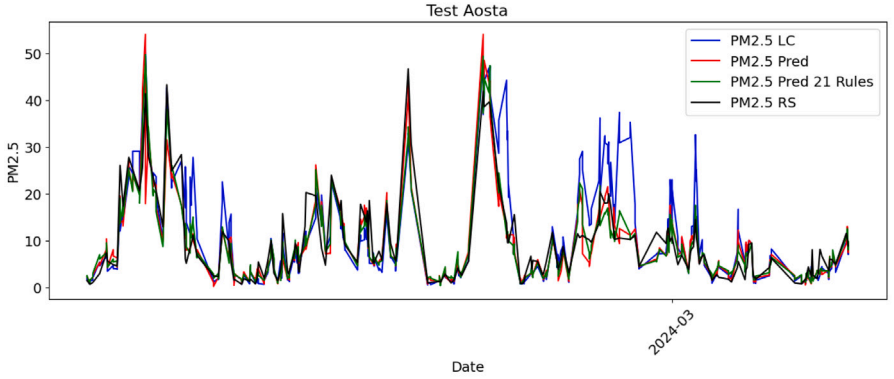
Fig. 7. MAE variation as a function of the number of reduced rules across locations, based on WAM.

by the results in Table 3. To assess the effect of rule pruning on model performance, the table compares outcomes obtained using the complete rule set with those from a reduced configuration, in which pruning was halted just before a notable decline in accuracy. This quantitative comparison reinforces the observation that conservative pruning can achieve model simplification with minimal performance loss. For Aosta, Trento, and Calgary, the pruned system retains only the most relevant rules, as identified by WAM. Despite substantial reductions in rule count, in some cases removing more than half of the original rules, the models maintain comparable performance in terms of PCC, MAE, and RMSE. For example, in Aosta, the system with only 21 rules still achieves a PCC above 0.87 on the test set and shows only marginal differences in error metrics compared to the full 48-rule configuration. A similar trend is observed in Calgary, where the reduced configuration (29 rules) yields PCC values of 0.9686 and 0.9579 for training and test sets, respectively, only slightly lower than with the full rule set, and well above 0.95. This supports the idea that the pruning strategy effectively identifies and preserves high-value rules. Notably, removing just one additional rule at this stage causes a sharp increase in MAE and RMSE, demonstrating the importance of selecting an optimal pruning threshold.

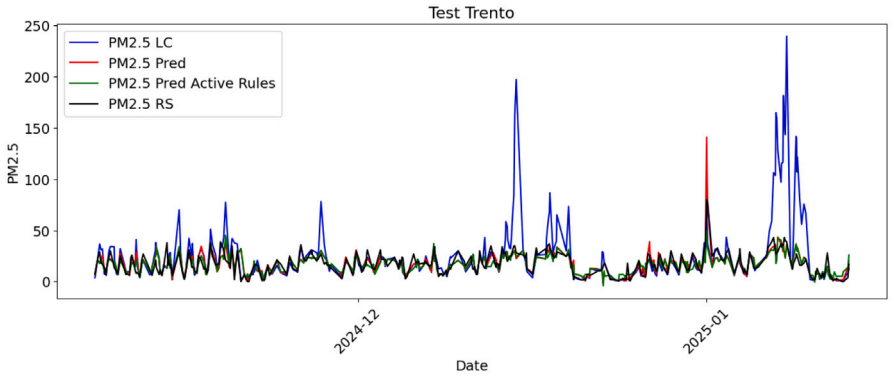
In contrast, the datasets from Reggio Emilia and Trento show greater sensitivity to rule reduction. Although simplified models continue to perform well on training data, they exhibit larger performance drops when tested on new data, though the increase in MAE never exceeds 1 unit. This pattern indicates that some datasets contain fewer redundant rules, making aggressive pruning more challenging and potentially harmful to model performance. However, even under these conditions, the models maintain strong predictive capabilities, demonstrating the robustness and effectiveness of the proposed approach. The results for Lima indicate that even with a reduction to just 18 rules, the model retains strong predictive power, with improved test set PCC and reduced MAE compared to the full configuration, suggesting that pruning may also act as a regularisation mechanism, improving generalisation.

Fig. 8 presents a two-week excerpt of $PM_{2.5}$ data for the selected cities. The blue line represents the LCS measurements, the black line corresponds to the RS, while the red and green lines indicate the predictions from the original (non-reduced) and reduced FIS, respectively, with the latter based on the configurations in Table 3. In all cases, both the full and reduced FIS models closely follow the RS data. In Aosta, the reduced system shows negligible deviation from the original, indicating that rule pruning does not significantly impact performance. In Trento, the reduced FIS appears to better handle peak smoothing, as some overestimated spikes seen in the full model are mitigated. After rule reduction, the predicted $PM_{2.5}$ values align more closely with the reference station measurements, indicating improved accuracy in capturing peak behaviour. In contrast, in Calgary, a slight underestimation is observed in the reduced system, where predicted values tend to be lower than the ground truth during certain peaks, suggesting a minor loss of information due to pruning.

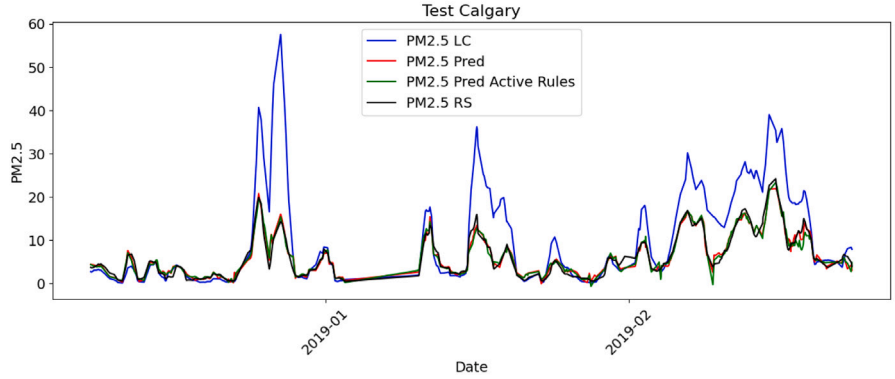
To assess the stability of WAM during the rule reduction process, Fig. 9 presents the PCC for each location, plotted for both the training set (solid line with dots) and the test set (dotted line with crosses).



(a) Zoomed view of two weeks of data from Aosta.



(b) Zoomed view of two weeks of data from Trento.



(c) Zoomed view of two weeks of data from Calgary.

Fig. 8. Two-week comparison of sensor data from three locations (Aosta, Trento, Calgary) showing: original LCS and RS measurements, full fuzzy model predictions, and predictions using only the most active rules before performance drop.

Since the Pearson coefficient eventually drops to zero for all locations, only the first 20 iterations are shown for clarity. Notably, for Trento, the metric reaches zero before this threshold. This is linked to the one more iteration, equals to one less rules, dropping performance a lot.

The plot illustrates the method's ability to preserve model performance during pruning. In some cases, such as Reggio Emilia, an increasing gap between training and testing curves indicates a decline in generalisation. Conversely, in locations like Aosta and Lima, test performance improves throughout the pruning process, suggesting enhanced generalisation by removing rarely activated rules. Calgary exhibits consistently stable behaviour, while Trento shows a gradual decline in performance, albeit with an initial improvement peaking at iteration 8. This is because one additional iteration, corresponding to the removal of just one more rule, leads to a significant drop in performance.

3.2.3. BAM and WAM comparison

The experimental evaluation demonstrates that the proposed Explainable AI strategies, particularly WAM, enable significant pruning of fuzzy rules without compromising model accuracy. This capability is crucial for air quality monitoring applications, where interpretable, lightweight models are preferred for real-time and resource-constrained environments. The predictions of the reduced models align well with those of the full models in all cities and, in some cases, exhibit improved robustness during peak events (as shown in Fig. 8). This highlights the practical benefits of pruning for deployment-ready models that preserve interpretability and reliability while reducing computational overhead.

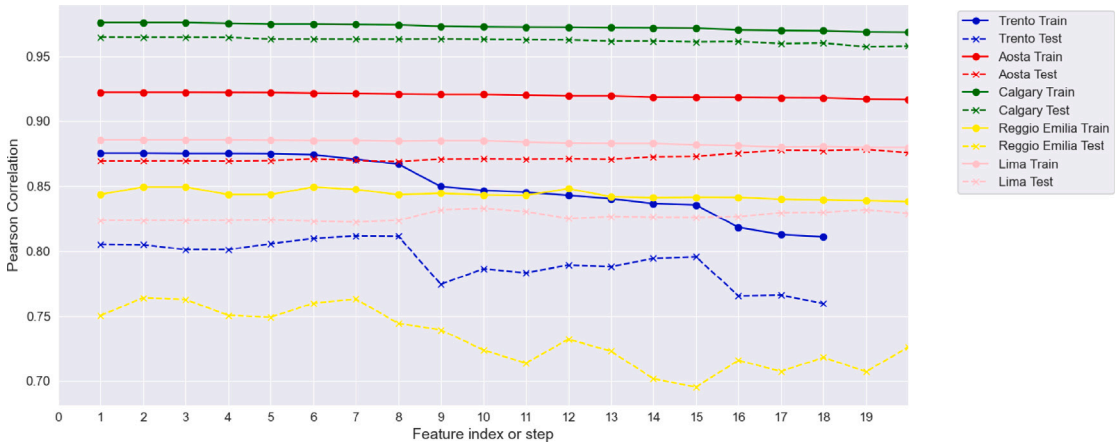


Fig. 9. Pearson correlation coefficient across different locations for both training and testing sets during the rule reduction process.

Table 3
Performance comparison of FISs using all rules versus a reduced rule set, just before a significant drop in predictive accuracy.

Configuration	Split	PCC	MAE	RMSE
Aosta				
All Rules (48)	Train	0.9223	1.6736	2.6114
	Test	0.8696	2.0413	3.2580
Reduced (21 Rules)	Train	0.9106	1.7864	2.7935
	Test	0.8779	1.9954	3.1432
Reggio Emilia				
All Rules (48)	Train	0.8436	5.4748	7.6765
	Test	0.7503	6.1145	9.7517
Reduced (27 Rules)	Train	0.8381	5.6002	7.7980
	Test	0.7263	6.3621	10.257
Trento				
All Rules (48)	Train	0.8750	3.5950	4.8806
	Test	0.8057	4.5680	6.5460
Reduced (31 Rules)	Train	0.8110	4.2068	5.8977
	Test	0.7597	4.9832	6.7890
Calgary				
All Rules (48)	Train	0.9757	0.8969	1.2455
	Test	0.9644	0.9763	1.3801
Reduced (29 Rules)	Train	0.9686	1.0344	1.4142
	Test	0.9579	1.0988	1.4988
Lima				
All Rules (48)	Train	0.8855	2.8203	4.1711
	Test	0.8237	3.3935	5.1643
Reduced (18 Rules)	Train	0.8674	2.9891	4.4676
	Test	0.8417	3.2463	4.8197

One key finding is that WAM consistently outperforms BAM, both in terms of preserving predictive accuracy and ensuring stability during pruning. By incorporating the strength of rule activations, WAM offers a more nuanced view of rule relevance, avoiding the premature elimination of rules that may be rarely triggered but highly informative when activated. This leads to a smoother degradation curve (as observed in Fig. 7) and extends the number of pruning iterations in which the performance remains stable.

The impact of pruning varies between locations, reflecting differences in data characteristics and environmental conditions. For example, Aosta and Calgary exhibit high tolerance to rule reduction, maintaining strong correlation and low error even with over 40% fewer rules. In contrast, Reggio Emilia and Trento show greater sensitivity to pruning, possibly due to higher noise levels or lower redundancy in their input distributions. Lima presents an interesting case where pruning not only reduces model complexity but also slightly improves generalisation, suggesting that removing noisy or inconsistent rules can have a regularising effect.

WAM-based pruning can maintain high correlation levels in both training and test sets across locations, at least up to an optimal threshold (as shown in Fig. 9). The sudden drops observed after excessive pruning iterations further justify the need for early stopping based on performance monitoring.

Comparison with existing XAI methodologies. XAI techniques, introduced in Section 1, are increasingly used to explain models that predict air pollutant concentrations. Among these, SHAP and LIME are the most common, especially for $PM_{2.5}$ prediction using models like Random Forest and LightGBM (Sunder et al., 2023; Aman et al., 2025; Bekkar et al., 2024; Chakraborty et al., 2024). SHAP is popular due to its solid theoretical background, ability to give both global and local insights, and efficient performance with tree-based models. LIME is often used with SHAP to help identify the key features affecting predictions.

In contrast, FIS are naturally interpretable by design. Each rule clearly defines a combination of input conditions, and the frequency at which a rule is activated provides a straightforward measure of its importance. This removes the need for additional explanation tools. However, fuzzy systems do not typically reveal the extent to which each feature contributes to a prediction or explain individual cases in detail; in our case, we tried to address this limitation through our design.

As noted in Cao et al. (2024), fuzzy rules are among the most widely used explanation techniques in sequence-related tasks, especially in healthcare. They offer strong interpretability and are easy to follow, but they are closely tied to the specific model structure, which limits their flexibility. In contrast, SHAP works with any model, but is harder to interpret and requires more computing resources. Heatmaps sit in the middle, depending on how they are applied.

Fuzzy systems are especially good at showing the reasoning process using IF–THEN rules, which helps make the logic clear. They also deal well with uncertainty and vague data, making them suitable for applications where human understanding is important.

Combining fuzzy systems with XAI tools like SHAP could be a useful strategy. It would bring together the understandable structure of fuzzy rules with the detailed insights from SHAP, helping to check rule coverage and uncover more complex patterns.

Overall, fuzzy systems continue to be among the most interpretable methods, especially when clarity and simplicity are a priority.

3.2.4. Complexity of proposed algorithms

Both the BAM and WAM exhibit the same time complexity. In both cases, the overall computational cost is determined primarily by the product of the number of input samples and the number of fuzzy rules. The latter is a fixed characteristic of the examined FIS. Among the various steps involved, the most computationally demanding component

is the sorting sub-procedure, which is necessary for ordering the rules based on their usage statistics. For many standard sorting algorithms, this step has a time complexity of $O(n^2)$, but more efficient algorithms such as QuickSort can reduce this to $O(n \log_2 n)$.

To better illustrate the real-world impact of reducing computational complexity from $O(n^2)$ to $O(n \log^2 n)$, it is helpful to consider tasks that are both fundamental to computer science and central to contemporary AI. Two such examples are sorting large datasets, a classic computational problem, and the attention mechanism (Niu et al., 2021), which is used extensively in Large Language Models (Naveed et al., 2023) (LLMs), such as GPT and BERT (Topal et al., 2021).

Consider an input size of $n = 16{,}384$, which is representative of real-world scenarios such as sorting high-volume logs or processing long textual sequences in transformer-based models. For algorithms with $O(n^2)$ complexity, such as bubble sort or standard dense attention, the number of required operations grows to approximately 268 million ($n^2 = 2^{14} \cdot 2^{14} = 2^{28}$). In contrast, methods that achieve $O(n \log^2 n)$ complexity, such as optimised sorting routines or recent efficient attention implementations (e.g., Flash Attention or kernel-based variants) require only about 3.6 million operations ($n \log^2 n \approx 16{,}384 \cdot 14^2 \approx 3.2 \times 10^6$). This corresponds to a reduction of more than 98% in the number of operations.

Empirical results further underscore this difference. On an NVIDIA V100 GPU, the execution time for an $O(n^2)$ algorithm at this input size is approximately 350 ms per run. In contrast, the $O(n \log^2 n)$ alternative completes the same task in roughly 25 ms, yielding a 14-fold speedup. This gain becomes even more pronounced as input sizes grow, highlighting the scalability advantage of the improved complexity.

Therefore, while the base operations scale linearly with the size of the dataset and the number of rules, the sorting step introduces a nonlinear overhead that should be considered when applying the procedure to systems with a large number of rules. Nevertheless, with the computational capabilities of modern hardware, sorting even several thousand elements can be performed efficiently and within a relatively short time frame, making it a minor concern in most practical applications.

Such reductions in computational cost are not merely theoretical. They enable the practical deployment of large-scale models and data processing systems, especially in latency-sensitive or resource-constrained environments. As demands for model size and sequence length continue to rise, the benefits of more efficient algorithms become increasingly critical.

4. Conclusions

The consistent performance observed with the proposed method underscores the effectiveness of rule pruning, particularly when guided by the Weighted Activation Method, in preserving high adjustment accuracy while significantly reducing model complexity. This reduction leads to a more compact and computationally efficient system, with minimal loss in predictive performance, even across diverse geographic locations.

This study introduces and validates two Explainable AI techniques for fuzzy rule pruning in adaptive neuro-fuzzy inference systems, applied to the task of $PM_{2.5}$ measurement adjustment from low-cost sensors: Binary Activation Method and Weighted Activation Method.

Among the two, WAM proves to be the more effective strategy, enabling the construction of compact, interpretable, and computationally efficient models while maintaining high predictive accuracy. The proposed pruning procedure demonstrates that up to 40%–60% of rules can be safely removed in certain cases without significant loss of performance. In some scenarios, pruning even improves generalisation, acting as a form of regularisation.

The method's robustness across diverse geographic locations and datasets supports its applicability in real-world environmental monitoring, particularly where transparency, low latency, and adaptability are critical. By promoting semantic clarity and reducing complexity, this approach contributes to the growing demand for trustworthy and explainable AI systems in high-impact domains.

Fig. 5 illustrates how different cities exhibit varying rule activation patterns, reflecting local data characteristics, yet shared activation regions suggest some generalisable structure across locations.

The pruning strategy effectively reduces the number of fuzzy rules, simplifying the model without sacrificing generalisability. As shown in Fig. 9, performance remains comparable to the full model, with MAE deviations remaining below 0.5 points. In Lima, pruning improves model performance, as detailed in Table 3.

Among the two rule evaluation strategies considered, WAM consistently outperforms the Binary Activation Method, providing more stable behaviour during pruning and better accuracy in the early reduction stages. This is visible comparing Figs. 6 and 7, where WAM maintains superior robustness. These characteristics make WAM a more reliable and interpretable technique for guiding rule reduction in an adaptive neuro-fuzzy inference system.

Given its low computational cost, high explainability, and demonstrated stability, WAM represents a valuable tool for environmental monitoring and data-driven decision-making.

Future work will explore combining data from multiple locations during training to provide more information for improving the output adjustment at individual locations. By incorporating data that yields similar rule rankings, this approach can augment the amount of training data available, particularly benefiting the performance of low-cost sensors with limited local data. This direction also opens the opportunity to assess the generalisation capability of the ANFIS method by including data from locations not involved in the co-location process but sharing similar environmental contexts. Ultimately, this could reduce the time required for co-locating low-cost sensors near reference stations.

CRediT authorship contribution statement

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

Declaration of competing interest

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

Acknowledgements

This work was conducted as part of the AIQS project (AI-enhanced Air Quality Sensor for Optimising Green Routes), under the project code DIP_AIQS_PO_2025_PNRR_ECOS_SK4AF_E93C22001100001. AIQS was funded through a closed call within the initiative "Ecosystem for Sustainable Transition in Emilia-Romagna" (ECOSISTER), financed under the National Recovery and Resilience Plan (PNRR) – Mission 4 "Education and Research", Component 2 "From Research to Business", Investment 1.5 "Creation and strengthening of innovation ecosystems, building territorial R&D leaders" – funded by the European Union – *NextGenerationEU* (Grant Agreement No. 0001052, dated 23/06/2022 – Project ECS_00000033 – CUP E93C22001100001).

In addition, this work was partially supported by the "Excellence Initiative – Research University" program at AGH University of Krakow, Poland and by a grant for statutory activity from the Faculty of Physics and Applied Computer Science of AGH. We gratefully acknowledge the Regional Agency for Environmental Protection of the Aosta Valley,

Arpae Emilia-Romagna, and the Provincial Agency for Environmental Protection of Trento for providing data from their reference monitoring stations. These stations were co-located with low-cost sensors donated by Wiseair Srl, whose contribution of sensors is also sincerely appreciated.

Data availability

https://zenodo.org/records/13982208.

References

Alhasa, K.M., Hussein, T., Shraideh, S.A., Alshawabkeh, L., Al-Hunaiti, L., Al-Addous, A., 2018. Calibration of a low-cost air quality sensor using an adaptive neuro-fuzzy inference system. Sensors 18 (9), 3021. http://dx.doi.org/10.3390/s18093021.

Aman, N., Panyametheekul, S., Pawarmart, I., Sudhibrabha, S., Manomaiphiboon, K., 2025. A visibility-based historical PM2. 5 estimation for four decades (1981–2022) using machine learning in thailand: Trends, meteorological normalization, and influencing factors using SHAP analysis. Aerosol Air Qual. Res. 25 (1), 1–17.

Arrieta, A.B., et al., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115.

Bekkar, A., Hssina, B., Abekiri, N., Douzi, S., Douzi, K., 2024. Real-time AIoT platform for monitoring and prediction of air quality in southwestern Morocco. PLoS One 19 (8), e0307214.

Bressane, A., Garcia, A.J.d.S., Castro, M.V.d., Xerfan, S.D., Ruas, G., Negri, R.G., 2024. Fuzzy machine learning applications in environmental engineering: Does the ability to deal with uncertainty really matter? Sustainability 16 (11), 4525.

Cao, J., Zhou, T., Zhi, S., Lam, S., Ren, G., Zhang, Y., Wang, Y., Dong, Y., Cai, J., 2024. Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. Inform. Sci. 662, 120212. http://dx.doi.org/10.1016/j.ins.2024.120212, URL https://www.sciencedirect.com/science/article/pii/S0020025524001257.

Casari, M., Arigliano, A., Po, L., 2024a. A comparative study of lightgbm on air quality data across multiple locations. 3762, pp. 505–509, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85205575146&partnerID=40&md5=a3f01042cd811312daad16b0bf5963b1,

Casari, M., Kowalski, P.A., Po, L., 2024b. Optimisation of the adaptive neuro-fuzzy inference system for adjusting low-cost sensors PM concentrations. Ecol. Inform. 83, 102781. http://dx.doi.org/10.1016/j.ecoinf.2024.102781, URL https://www.sciencedirect.com/science/article/pii/S1574954124003236.

Casari, M., Po, L., 2024c. MitH: A framework for mitigating hygroscopicity in low-cost PM sensors. Environ. Model. Softw. 173, 105955. http://dx.doi.org/10.1016/j.envsoft.2024.105955, URL https://www.sciencedirect.com/science/article/pii/S1364815224000161.

Castelli, M., Clemente, F.M., Popovič, A., Silva, S., 2020. A machine learning approach to predict air quality in california. Complexity 2020, 1–10.

Chakraborty, S., Misra, B., Dey, N., 2024. Explainable artificial intelligence (XAI) for air quality assessment. In: Design Studies and Intelligence Engineering. IOS Press, pp. 333–341.

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. http://dx.doi.org/10.1016/B978-0-12-819723-3.00099-8, URL https://www.sciencedirect.com/science/article/pii/B9780128197233000998.

Concas, F., Mineraud, J., Lagerspetz, E., Varjonen, S., Liu, X., Puolamäki, K., Nurmi, P., Tarkoma, S., 2021. Low-cost outdoor air quality monitoring and sensor calibration: A survey and critical analysis. ACM Trans. Sens. Netw. (TOSN) 17 (2), 1–44.

Gegov, A., Sanders, D., Vatchova, B., 2017. Aggregation of inconsistent rules for fuzzy rule base simplification. Int. J. Knowl.-Based Intell. Eng. Syst. 21 (3), 135–145. http://dx.doi.org/10.3233/KES-170358, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85027501153&doi=10.3233%2fKES-170358&partnerID=40&md5=2e77c100b39626bc31999b9da12ad987, All Open Access, Green Open Access.

Guo, R., Qi, Y., Zhao, B., Pei, Z., Wen, F., Wu, S., Zhang, Q., 2022. High-resolution urban air quality mapping for multiple pollutants based on dense monitoring data and machine learning. Int. J. Environ. Res. Public Heal. 19 (13), 8005.

Hanoon, W.H., Ibrahim, A.S., 2024. Fuzzy classical two-absorbing second (secondary) sub-modules and strongly classical two-absorbing second sub-modules. Babylon. J. Math. 2024, 85–94. http://dx.doi.org/10.58496/BJM/2024/011.

Hanoon, W.H., Salman, M.B., Salem, K.A., 2024. Fuzzy soc–small two-absorbing modules and related concepts. Babylon. J. Math. 2024, 198–204. http://dx.doi.org/10.58496/BJM/2024/020.

Holzinger, A., et al., 2017. What do we need to build explainable AI systems for the medical domain?. arXiv preprint arXiv:1712.09923.

Islam, M.A., Anderson, D.T., Pinar, A.J., Havens, T.C., Scott, G., Keller, J.M., 2019. Enabling explainable fusion in deep learning with fuzzy integral neural networks. IEEE Trans. Fuzzy Syst. 28 (7), 1291–1300.

Jang, J.-S., 1993a. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. 23 (3), 665–685. http://dx.doi.org/10.1109/21.256541.

Jang, J.-S.R., 1993b. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. 23 (3), 665–685.

Karaboga, D., Kaya, E., 2019. Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. Artif. Intell. Rev. 52, 2263–2293.

Korkidis, P., Dounis, A., 2023. Intelligent fuzzy models: WM, ANFIS, and patch learning for the competitive forecasting of environmental variables. Sustainability 15 (10), 8032.

Kusy, M., Kowalski, P.A., Szwagrzyk, M., Konior, A., 2022. Machine learning techniques for explaining air pollution prediction. In: 2022 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8.

Martina Casari, Kowalski, P.A., Po, L., 2024. ANFIS_PM_adjustment: AQDR - v1.1.0. http://dx.doi.org/10.5281/zenodo.10854496.

Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., Mian, A., 2023. A comprehensive overview of large language models. ACM Trans. Intell. Syst. Technol..

Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62.

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. http://dx.doi.org/10.1016/B978-0-12-816656-7.00008-7, URL https://www.sciencedirect.com/science/article/pii/B9780128166567000087.

Sensirion, 2023. Sensirion PM Sensors Datasheet SPS30. https://www.sensirion.com/media/documents/8600FF88/64A3B8D6/Sensirion_PM_Sensors_Datasheet_SPS30.pdf, (Accessed 26 May 2025).

Si, M., 2019. Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine learning methods. http://dx.doi.org/10.5281/zenodo.3473833.

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. Atmospheric Meas. Tech. 13 (4), 1693–1707. http://dx.doi.org/10.5194/amt-13-1693-2020, URL https://amt.copernicus.org/articles/13/1693/2020/.

Sunder, M.S.S., Tikkiwal, V.A., Kumar, A., Tyagi, B., 2023. Unveiling the transparency of prediction models for spatial PM2.5 over Singapore: Comparison of different machine learning approaches with explainable artificial intelligence. AI 4 (4), 787–811. http://dx.doi.org/10.3390/ai4040040, URL https://www.mdpi.com/2673-2688/4/4/40.

Sutikno, T., 2023. Fuzzy optimisation and metaheuristic algorithms. Babylon. J. Math. 2023, 59–65. http://dx.doi.org/10.58496/BJM/2023/012.

Talpur, N., Abdulkadir, S.J., Alhussian, H., Hasan, M.H., Aziz, N., Bamhdi, A., 2023. Deep neuro-fuzzy system application trends, challenges, and future perspectives: A systematic survey. Artif. Intell. Rev. 56 (2), 865–913.

Topal, M.O., Bas, A., van Heerden, I., 2021. Exploring transformers in natural language generation: Gpt, bert, and xlnet. arXiv preprint arXiv:2102.08036.

Villanueva, E., Espezua, S., Castelar, G., Diaz, K., Ingaroca, E., 2023. Smart multi-sensor calibration of low-cost particulate matter monitors. Sensors 23 (7), http://dx.doi.org/10.3390/s23073776, URL https://www.mdpi.com/1424-8220/23/7/3776.

Wang, L., Chen, X., Xia, Y., Jiang, L., Ye, J., Hou, T., Wang, L., Zhang, Y., Li, M., Li, Z., Song, Z., Jiang, Y., Liu, W., Li, P., Zhang, X., Yu, S., 2022. Operational data-driven intelligent modelling and visualization system for real-world, on-road vehicle emissions—A case study in Hangzhou City, China. Sustainability 14 (9), 5434.

Wiseair, 2024. Wiseair - Transforming home-work mobility into strategic value. URL https://wiseair.vision/, (Accessed 26 May 2025).

Yadav, K., Arora, V., Kumar, M., Tripathi, S.N., Motghare, V.M., Rajput, K.A., 2022. Few-shot calibration of low-cost air pollution (pm _{2.5}) sensors using meta learning. IEEE Sensors Lett. 6 (5), 1–4.

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

Zadeh, L.A., Fu, K., Shimura, M., 1996. An introduction to fuzzy logic applications in intelligent systems. Proc. IEEE 83 (3), 345–377.