Fuzzy Inference Systems for Estimation of Air Quality Index Using Python Simulation and Indian Pollution Dataset

[Your Name]

July 16, 2025

Abstract

The increasing concern over air pollution in urban areas has made real-time and intelligent air quality monitoring systems essential. Traditional models for Air Quality Index (AQI) estimation often fall short in handling uncertainties and variability in pollution data. This paper presents a fuzzy inference system (FIS)-based approach to estimate AQI using key pollutants — PM2.5, PM10, and NO_2 — applied on the publicly available "Air Quality Data in India" Kaggle dataset. The Mamdani-type fuzzy model was developed in Python using the `scikit-fuzzy` library. Membership functions were defined based on pollutant concentration ranges, and fuzzy rules were structured to estimate AQI levels from crisp input values. Simulation results demonstrated that the fuzzy model provides a flexible, interpretable, and effective method for AQI estimation. The study confirms that fuzzy logic is a promising tool for smart environmental monitoring systems, especially where exact thresholds are difficult to define.

1. Introduction

Air pollution poses a serious threat to human health and the environment, particularly in rapidly urbanizing countries like India. The Air Quality Index (AQI) is a standardized indicator used to represent the overall level of air pollution based on concentrations of major pollutants such as PM2.5, PM10, SO_2, NO_2, CO, and O_3. In India, AQI values are calculated and published by the Central Pollution Control Board (CPCB) using deterministic rules and fixed thresholds.

However, real-world pollution levels often vary due to changing weather, traffic, and industrial activity, leading to uncertainty in crisp classifications. Traditional AQI models lack the capability to handle such imprecision. Fuzzy logic provides a powerful tool to model uncertainty and approximate reasoning by emulating human-like decision-making. A Fuzzy Inference System (FIS) can take linguistic inputs (e.g., "high PM2.5") and produce an interpretable AQI category (e.g., "poor") based on a set of rules.

This paper proposes the implementation of a Mamdani-type FIS using Python to estimate AQI based on three major pollutants: PM2.5, PM10, and NO_2. The model is simulated on a real-world dataset from Kaggle titled *"Air Quality Data in India"* covering multiple Indian cities and years. The study focuses on Delhi data from 2019 to demonstrate the accuracy and flexibility of the system.

2. Literature Survey

Previous studies have highlighted the effectiveness of fuzzy systems in air quality modeling. Dunea et al. (2011) implemented a Mamdani fuzzy inference model in MATLAB to estimate AQI using SO_2, NO_2, and PM10 in Romania. Their system defined trapezoidal and triangular membership functions and achieved promising results by modeling the uncertain boundaries between pollution classes.

Similarly, Hajek and Olej (2009) applied hierarchical fuzzy inference systems for AQI modeling, emphasizing modular rule-based control. Fisher (2003) discussed the role of fuzzy environmental decision-making in pollution control and validated its advantage over rigid statistical approaches.

More recent developments involve hybrid models like neuro-fuzzy systems and machine learning-based AQI prediction. However, most of these are either black-box systems or computationally intensive. Our work bridges the gap by using an interpretable and lightweight fuzzy logic model implemented in Python using `scikit-fuzzy`, offering both transparency and adaptability. We also leverage open data, making this system reproducible and scalable for real-world applications.

3. Description of the Topic

Fuzzy logic is a mathematical framework that mimics human reasoning to handle imprecise or uncertain information. It allows variables to belong to more than one category with varying degrees of membership, making it suitable for modeling real-world problems like air pollution.

A fuzzy inference system (FIS) typically comprises:

- Fuzzification
- Rule Base
- Inference Engine
- Defuzzification

In this project, the FIS uses the input pollutants: PM2.5, PM10, and NO_2. Each input is modeled with five fuzzy sets (Very Low, Low, Moderate, High, Very High), and the output AQI is divided into five categories (Excellent, Good, Moderate, Poor, Hazardous).

Triangular membership functions were used based on CPCB standard ranges. The rules are simple IF-THEN statements. For example:

- IF PM2.5 is low AND PM10 is low AND NO_2 is low THEN AQI is Excellent

- IF PM2.5 is very high AND PM10 is very high THEN AQI is Hazardous

The Mamdani inference method was used with centroid defuzzification to provide crisp output values.

4. Simulation (Python Execution & Results)

The fuzzy system was implemented using the Kaggle dataset *"Air Quality Data in India"* for Delhi city in 2019. The pollutants selected were PM2.5, PM10, and NO_2. Missing values were removed before applying fuzzy inference.

Three input variables and one output AQI variable were defined using triangular membership functions. Fuzzy rules were constructed to link input conditions to AQI categories.

The system was simulated for 30 days. AQI output was defuzzified and compared against known pollution patterns. Graphs and trends confirmed the reliability of the fuzzy model.

Sample Output Table:

Day	PM2.5	PM10	NO_2	AQI Score
1	135.0	210.0	60.0	6.72 (Poor)
2	90.0	180.0	42.0	5.61 (Moderate)
3	55.0	120.0	30.0	4.23 (Moderate)
4	250.0	340.0	85.0	8.95 (Hazardous)
5	80.0	140.0	33.0	5.03 (Moderate)

Table 1: Sample AQI estimation results

5. Code of Simulation

The Python code used for simulation includes data loading, preprocessing, defining fuzzy variables and membership functions, setting up fuzzy rules, computing AQI, and plotting the results. The simulation uses the `scikit-fuzzy` library for rule-based fuzzy control systems.

Refer to the Python section of this report or accompanying script file for the complete source code used in the simulation. The code block includes:

- Handling missing data
- Defining fuzzy rules
- Applying fuzzy inference
- Plotting AQI trends for visualization

6. Future Aspects

Future enhancements to this project can include:

- Real-time sensor data integration
- Hybrid neuro-fuzzy models for learning rule bases
- Mobile/web dashboards for live AQI monitoring
- Region-specific fuzzy models
- Automatic rule optimization using AI algorithms

These advancements can enable a smart and responsive air quality monitoring system.

7. Conclusion

The fuzzy inference system (FIS) developed in this project effectively models AQI using PM2.5, PM10, and NO_2 data from real-world datasets. Implemented in Python using the `scikit-fuzzy` library, the Mamdani-type FIS proved to be both intuitive and efficient in estimating air quality levels. By converting crisp pollutant data into fuzzy sets and applying well-defined rules, the system provides an interpretable output that aligns with expected AQI classifications.

This approach not only bridges the limitations of deterministic models but also provides a flexible platform for future environmental applications. The results show that fuzzy systems are particularly suitable in scenarios involving uncertainty and vague threshold limits—conditions typical of real-world pollution data.

With additional enhancements such as real-time sensor integration and hybrid AI techniques, this fuzzy model has the potential to evolve into a robust and scalable smart-city pollution monitoring tool.

8. References

1. Dunea, D., Pohoata, A.A., & Lungu, E. (2011). *Fuzzy inference systems for estimation of Air Quality Index.* ROMAI Journal.
2. Rohan Rao. *Air Quality Data in India.* Kaggle. https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india
3. Fisher, B. (2003). *Fuzzy Environmental Decision-Making: Applications to Air Pollution.* Atmospheric Environment, Vol. 37.
4. Hajek, P., & Olej, V. (2009). *Air Quality Indices and their Modelling by Hierarchical Fuzzy Inference Systems.* WSEAS Transactions.

5. Kasabov, N.K. (1998). *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. MIT Press.

6. `scikit-fuzzy` Documentation. https://scikit-fuzzy.github.io/scikit-fuzzy

7. Central Pollution Control Board (CPCB), India. https://app.cpcbccr.com/