# **Project Manager: Motamedi, ali**

## **Literature Review**

### **Introduction**

Air quality is one of the major factors that significantly affects human health, which ultimately effects environment, and the economy. Urbanization and industrial activities which have led to rise in the levels of pollutants and major of them are Carbon Monoxide (CO), Nitrogen Oxides (NOx), and Benzene (C6H6), which pose challenges for environmental policies and public health initiatives. The corpus of research that is now available for analyzing air quality and the effects of different contaminants is examined in this overview of the literature. It highlights the main ideas, theories, and topics that guide the project's comprehension of pollutant variance, how it interacts with the environment, and how to create predictive models for managing air quality.

Several seminal studies have characterized the adverse effects of criteria air pollutants in urban environments. These include ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), particulate matter (PM10 and PM2.5), carbon monoxide (CO), and volatile organic compounds such as benzene (C6H6) (Kelly and Fussell, 2015). Chronic exposure to these pollutants is linked to increased mortality from stroke, heart disease, lung cancer, and chronic respiratory diseases.

Hoek et al. (2013) conducted a meta-analysis of studies globally and determined that COPD and lung cancer incidences increased by 2-3% per 10 μg/m3 rise in PM2.5 levels. NO2 exposure also elevates asthma symptoms in children and the risk of respiratory infections (WHO, 2018). Benzene is classified as a Group 1 carcinogen, with links to leukemia at chronic exposures over 10 μg/m3 (WHO, 2010). Monitoring and analyzing the trends in these pollutants is critical for developing interventions to protect public health.

Numerous studies have analyzed air pollution levels, trends, and health impacts in cities worldwide. Vetter et al. (2020) conducted a trend analysis of nitrogen oxides (NOx) and particulate matter across 421 cities in Europe and North America.

### **Key Themes and Concepts**

#### **Air Pollution and Health Outcomes:**

A vast amount of literature has established a strong correlation between air pollution and adverse health outcomes. For instance, studies have shown that exposure to high levels of NOx and CO can lead to respiratory and cardiovascular diseases (WHO, 2021). Benzene, a carcinogenic compound, has been linked to increased risk of leukemia (Rinsky et al., 1987). This project draws on such findings to underscore the importance of analysing air quality data to mitigate health risks associated with air pollution.

#### **Environmental Conditions and Pollutant Behaviour:**

Environmental conditions, including temperature, humidity, and atmospheric pressure, play a critical role in the dispersion and chemical transformation of pollutants. Research by Jacob and Winner (2009) highlights how temperature can influence the photochemical reactions of pollutants, affecting their concentration levels. This project aims to explore these relationships further by analysing how environmental conditions impact pollutant behaviour over time.

#### **Predictive Modelling of Air Quality**

The use of statistical and machine learning techniques to predict air quality indices based on historical data and environmental parameters is a growing field of research. Studies such as those by Grivas and Chaloulakou (2006) have utilized time series analysis and regression models to forecast air pollution levels. The development of accurate predictive models is crucial for effective air quality management and is a key objective of this project.

### **Theoretical Frameworks**

#### **Time Series Analysis:**

Time series analysis is a fundamental method for analysing air quality data, allowing for the identification of long-term trends and seasonal variations in pollutant levels. Autoregressive Integrated Moving Average (ARIMA) models and their variations are commonly used for this purpose (Box et al., 2015). This project will employ time series analysis to examine trends in air quality data and assess the impact of environmental factors.

#### **Machine Learning for Environmental Data:**

Most of the Machine learning models, including decision trees, random forests, and neural networks, have been applied to environmental data analysis. These models are programmed to handle complex nonlinear relationships between various variables, making them suitable for predicting air quality indices (Breiman, 2001; LeCun et al., 2015). The project will explore the use of machine learning techniques to develop predictive models for air quality.

### **Methodological Considerations**

#### **Data Cleaning and Preparation:**

The accuracy of any analysis or predictive model heavily relies on the quality of the data. As such, literature emphasizes the importance of data cleaning and preparation, including handling missing values and removing outliers (Wickham & Grolemund, 2017). This project will follow recommended practices for data cleaning to ensure the reliability of the analysis.

#### **Exploratory Data Analysis (EDA):**

EDA is crucial step in this study for understanding the distribution and relationships between variables in a dataset. It involves visualizing data, identifying patterns, and testing hypotheses (Tukey, 1977). This project will conduct a thorough EDA to gain insights into the distribution of pollutants and their correlations with environmental conditions.

### **Conclusion**

In this literature review we have identified key themes, concepts, and theoretical frameworks relevant to the analysis of air quality data. It underscores the importance of understanding the relationship between pollutants and environmental conditions, the health implications of air pollution, and the potential of predictive modelling in managing air quality. By drawing on existing research, this project aims to contribute to the development of effective strategies for improving air quality and reducing the health risks associated with air pollution. The comprehensive analysis of air quality trends and pollutant behavior, utilizing statistical and machine learning techniques, will provide actionable insights for environmental policy-making and public health advisories.