Air Pollution Forecast and The Air Quality Changes in The Beijing-Tianjin-Hebei Region

SHIXIAN SUN

Air Pollution Forecast and The Air Quality Changes in The Beijing-Tianjin-Hebei Region

Shixian Sun

A thesis presented for the degree of

Doctor of Philosophy

Department of Engineering

Durham University

United Kingdom

April 2025

Air pollution forecast and the air quality changes in the Beijing-Tianjin-Hebei region

Shixian Sun

Abstract

Air pollution control and its health impacts are critical concerns. Traditional air pollution prediction often focuses on the Air Quality Index (AQI), which represents the most harmful pollutant but neglects the cumulative effects of others, limiting its accuracy in assessing hazards. To address this, a neural network model is proposed to simultaneously predict three indices: AQI, the Aggregate Air Quality Index (AAQI), and the Health Risk-Based Air Quality Index (HAQI). The model integrates Long Short-Term Memory (LSTM), Multilayer Perceptron (MLP), and Transformer, selecting the best prediction among them to enhance performance and stability. Its effectiveness was validated using data from 12 monitoring stations in Beijing.

In addition, a GCN+LSTM model was developed for AQI prediction, leveraging Graph Convolutional Networks (GCN) to extract spatial correlations from historical data and integrating them with LSTM's temporal modelling capabilities. This approach simplifies spatial data collection, improves prediction accuracy, and outperforms other models, including Long Short-Term Memory, LSTM+Attention, and Gated Recurrent Unit (GRU), in experimental evaluations.

Furthermore, a comprehensive analysis of air quality trends in the Beijing-Tianjin-Hebei region from 2014 to 2023 was conducted. This analysis examined the effects of strict pollution controls, COVID-19, seasonal variations, and associated health risks, providing insights into regional and temporal patterns.

This study contributes by proposing a multi-index prediction framework for air quality assessment, introducing an innovative spatial-temporal model, and offering a decade-long analysis of air pollution trends and their health impacts in the Beijing-Tianjin-Hebei region.

CONTENTS

List of Figures

List of tables

List of Symbols and Abbreviations

$C_{m,t}$	The concentration of pollutant m, p67
n	Health category index, p67
$C_{m,n}$	Upper limit concentrations of the n, p67
ρ	Empirical constant with an optimal value, p67
β_m	Exposure-response coefficient, p68
C_m	Pollutant concentration, p68
$C_{m,0}$	Threshold concentration, p68
C_m^*	Equivalent concentration, p68
f	Forgetting gate, p70
i	Input gate, p71
o	Output gate, p71
C_t	Current cell state, p71
x_t	Current network input, p71
C_t	Unit state, p71
f_t	Output of the forget gate, p71
σ	Sigmoid function, p71
$\widetilde{C_t}$	Candidate cell information, p71
h_t	Hidden state at time step t, p71
x_t	Input at time step t, p71
$tan\ h$	Hyperbolic tangent activation function, p71
W_f	Weight coefficient matrix, p71
X	Input sequence, p75
W_Q	Learnable weight matrices, p75
d_k	Dimension of the key vector, p75
pos	Position, p76

i	Dimension index, p76
d	Model dimension, p77
V	Number of nodes in the graph, p77
E	Set of edges, p77
A	Adjacency matrix of the graph, p77
A_{ij}	Connection relationship between nodes V_i and V_j in graph G, p77
I	Identified matrix of size N*N, p77
$\widetilde{D}$	Diagonal matrix, p77
$\sigma(.)$	Activation function, p79
W^l	Parameter values of layer l, p79
$\mathcal{Z}_t$	Output of the updaters, p79
σ	Sigmoid activation function, p79
W_Z	Weight matrix, p79
h_{t-1}	Hidden state at the previous moment, p79
x_t	Input at the current moment, p79
r_t	Output of the reset gate, p79
W_r	Weight matrix, p79
$\tilde{h}_t$	Candidate hidden state, p79
tanh	Tanh activation function, p79
W	Weight matrix, p79
Q	Query vector, p80
K	Key vector, p80
V	Value vector, p80
d_k	Vector dimension of the Key, p80
$\widehat{y}_i$	Actual values, p81
y_i	Predicted values, p81
$SS_{residual}$	Residual sum of squares, p82
SS_{total}	Total sum of squares, p82

IE	Rate of the health outcome attributable to the exposure, p86
NE	Number of cases attributed to the exposure, p86
N	Population’s size investigated, p86
AQI	Air Quality Index, p67
AAQI	Aggregate Air Quality Index, p68
HAQI	Health Risk-Based Air Quality Index, p68
LSTM	Long Short-Term Memory, p8
MLP	Multilayer Perceptron, p8
GCN	Graph Convolutional Networks, p9
CRU	Gated Recurrent Unit, p36
BTH	Beijing-Tianjin-Hebei region, p61
WHO	World Health Organization, p83
COPD	Chronic obstructive pulmonary disease, p2
ALRTI	Acute lower respiratory tract infection, p2
OECD	Organisation for Economic Co-operation and Development, p2
NAAQS	National Ambient Air Quality Standard, p3
SVM	Support vector machine, p5
CNN	Convolutional neural network, p5
RNN	Recurrent neural network, p5
AP	Attributable proportion, p10
PM	Particulate matter, p16
EPA	Environmental Protection Agency, p18
NAPCA	National Air Pollution Control Administration, p19
ARIMA	Auto Regressive Integrated Moving Average, p25
AR	Autoregressive, p25
MA	Moving average, P25
MAE	Mean absolute error, p25
WT	Wavelet transform, p26

AIC	Akaike Information Criterion, p26
GS	Grid search, p26
RMSE	Root Mean Squared Error, p26
R^2	Coefficient of Determination, p26
SFGM	Seasonal grey model with the fractional order accumulation, p27
GM	Grey model, p27
ALO	Ant Lion Optimizer, p27
SGM	Seasonal grey model, p27
MAPE	Mean absolute percentage error, p27
CMAQ	Community Multiscale Air Quality, p27
AI	Artificial intelligence, p31
GPS	Global positioning system, p31
MSE	Mean square error, p33
PSO	Particle swarm optimization, p33
AdaBoost	Adaptive Boosting model, p33
SSE	Sum of square error, p34
FNN	Fully Connected Neural Network, p35
FB	Fractional deviation, p35
NMSE	Normalized mean square error, p38
RF	Random Forest, p5
DTR	Decision Tree Regression, p33
LR	Linear Regression, p33
SGDR	Stochastic Gradient Descent Regression, p37
SVR	Support Vector Regression, p37
BPNN	Backpropagation neural network, p38
SWT	Stationary wavelets transform, p38
PC	Principal component, p38
ANN	Artificial neural network

MLR	Multiple linear regression model, p38
CEEMDAN	Adaptive noise complete ensemble empirical mode decomposition, p38
GWO	Gray Wolf Optimizer, p39
VMD	Variational mode decomposition, p39
PE	Permutation entropy, p39
Bi-LSTM-AT	Bidirectional long short-term memory network, p39
MIF	Multiple influencing factors, p39
GNSS-ZTD	Global Navigation Satellite System-derived Zenith Total Delay, p39
GOLD	Global Initiative for Chronic Obstructive Pulmonary Disease, p53
RR	Relative risk, p53
ER	Excessive risk, p53
CAAQS	Chinese Ambient Air Quality Standards, p67
FFN	Feed- Forward Neural Network, p75
RNN	Recurrent neural network, p78

Declaration

The work in this thesis is based on research carried out by the author, at the Department of Engineering, Durham University, England. No part of this thesis has been submitted elsewhere for any other degree or qualification, and it is the sole work of the author unless referenced to the contrary in the text.

Sun, S., & Wang, Q. (2022, June). A comparison of air pollution in developed and developing cities: A case study of London and Beijing. Presented at ICAQMM2022:XVI. International Conference on Air Quality Management and Monitoring, Rome, Italy

Acknowledgements

First and foremost, I would like to sincerely thank my supervisor, Dr. Qing Wang, for her guidance and support throughout this journey. Her mentorship has been invaluable, from providing inspiration at the start to helping me overcome challenges along the way.

I am also profoundly grateful to my second supervisor, Professor Dagou Zeze, for his professional advice and timely assistance at critical moments. His expertise and insights have been an invaluable resource during my research.

My heartfelt thanks also go to my two reviewers, Professor Karen Johnson and Professor Paul N. Hughes, for their constructive feedback and warm encouragement during the annual reviews. Their input has been vital in refining and strengthening my work.

To my family, who have never cared how high I soared but only whether I was weary—your love, care, and encouragement have been a constant source of comfort and strength. You have always helped me regain balance and resilience whenever I faced stress or discouragement. I am deeply thankful for your unconditional support and unwavering dedication.

Finally, my heartfelt gratitude goes to my girlfriend, Wang Yu, whose unwavering support and companionship have been my greatest source of strength throughout this journey.

Chapter 1. Introduction

1.1 Background

The term "air pollution" is used to describe the contamination of indoor or outdoor environments by any chemical, physical, or biological factors that alter the natural characteristics of the atmosphere. The principal pollutants in the atmosphere are gases and particulate matter that are harmful to human health and the environment[1]. The sources of air pollution are numerous and diverse. Some are naturally occurring, including ash and gases from volcanic eruptions, smoke from wildfires, and others. Nevertheless, the majority of air pollution is caused by human activity. In particular, the rapid development of industry, automobile emissions, home heating, factory manufacturing, and power generation has led to the production of a considerable quantity of air pollutants, which have a detrimental impact on human health and quality of life[2].

Globally, nine out of 10 people are exposed to polluted air. Furthermore, air pollution represents one of the most significant global health and environmental concerns, not only due to its impact on climate change, but also in terms of its adverse effects on public and individual health, manifested in increased morbidity and mortality rates[3]. The rapid development of industry and globalization have brought numerous benefits to human life. However, these developments have also led to a significant challenge: the problem of air pollution[4]. The World Health Organization (WHO) has estimated that air pollution resulted in 4.2 million premature deaths worldwide in 2019. These deaths were attributed to exposure to fine particulate matter, which has been linked to an increased risk of cardiovascular and respiratory disease, as well as cancer.

Approximately 37% of premature deaths attributable to air pollution are attributed to ischemic heart disease and stroke. A further 18% of premature deaths are attributed to chronic obstructive pulmonary disease (COPD), while 23% are caused by acute lower respiratory tract infection (ALRTI). Additionally, 11% of premature deaths are attributed to respiratory cancer. Those residing in low- and middle-income countries are more adversely affected by outdoor air pollution, with 89% of cases occurring in these countries[5]. Moreover, air pollution can exacerbate respiratory conditions and increase the risk of asthma attacks, leading to a higher incidence of hospital admissions. While all members of society are susceptible to the adverse effects of air pollution, those who are most vulnerable, including children and the elderly, are at a greater risk. Exposure to air pollution can result in respiratory problems and impaired lung development in children. Some studies have indicated a potential association between Alzheimer's disease and air pollution[6]. Furthermore, the detrimental effects of air pollution extend beyond environmental concerns to have a profound impact on economic and social production. The WHO and the Organisation for Economic Co-operation and Development (OECD) [7] have estimated that in 2015, the economic cost of premature death and disability due to air pollution in Europe was close to $1.6 trillion [8]. Air pollution has a detrimental impact on the economy in a multitude of ways. It results in human mortality, impairs the capacity of individuals to engage in productive work, diminishes the resilience of ecosystems to perform the functions that society requires, and incurs financial costs associated with the repair or restoration of damaged ecosystems. In India, the estimated economic losses from lost productivity, premature deaths, and other consequences of air pollution were $95 billion in 2019, representing 3% of the country’s GDP. The cost of air pollution to human health in the UK is estimated to be up to £20 billion per year. China's air pollution situation is similarly concerning. In 2016, 254 of 338 cities at or above the prefecture level failed to meet the ambient air quality standards, representing 75.1% of the total. Air pollution has a profound impact on human health, the environment and the economy.

There are significant discrepancies in the impact of air pollution on developed and developing countries. It is estimated that in excess of 90% of deaths attributable to air pollution occur in low- and middle-income countries. The reasons for this considerable disparity can be attributed to differences in governmental conduct and levels of economic development. In low-income countries, there is a paucity of regulations governing air quality and vehicle emissions, as well as a dearth of control over industrial emissions and vehicle exhaust emissions. This results in the emission of extremely serious pollution levels in these countries[9]. As the world's most populous developing country, China is undergoing a period of rapid economic growth, accompanied by a parallel increase in the severity of its air pollution problem. In 2013, the China Environmental Quality Bulletin concluded that 71 of 74 cities that conducted fine particle monitoring did not meet the National Ambient Air Quality Standard (NAAQS) for China, which is 35 μg/m³. The mean annual concentration of $PM_{2.5}$ in these 74 cities is as high as 72 μg/m³, which is seven times the World Health Organization's air quality guideline (10 μg/m³). The Beijing-Tianjin-Hebei region is one of the most polluted areas in China, with an average annual concentration of $PM_{2.5}$ of 106 μg/m³[10].

The levels of air pollution are determined by the concentration of complex mixtures of air pollutants. Currently, six major pollutants are widely recognized: $PM_{2.5}$, PM_{10}, SO_2, O_3, NO_2, and CO[11]. The concentration of each pollutant and their associated health hazards can vary significantly, making it challenging for the general public to comprehend the health effects solely based on concentrations. Consequently, a standardised index has been developed and promoted with the intention of providing a uniform measure of the health impacts caused by air pollution. This index, known as the Air Quality Index (AQI), is utilised in numerous countries worldwide. The AQI is calculated based on the concentrations of six standard pollutants, with values ranging from 0 to 500. A higher AQI value indicates a more severe pollution situation, with the overall AQI being determined by selecting the highest sub-AQI value among these

pollutants[12]. Although this method offers a more intuitive representation of the extent of air pollution, some researchers have argued that it fails to adequately express the combined effects of multiple pollutants on human health. Therefore, Swamee et al. [13] proposed an aggregate AQI (AAQI) which combines the effects of various sub-pollutants into one comprehensive measure. According to evaluations conducted by Cairncross et al.[14], [15], AAQI proves more effective than traditional AQI in illustrating the health consequences resulting from exposure to environments with multiple sources of pollution. In addition, the pollution index Health risk-based AQI(HAQI), which is based on the health risks associated with exposure to various types of air pollution, is also receiving increasing attention and research[16], [17]. Some studies[18] have indicated that the indices of AQI and HAQI may vary under different concentrations, suggesting that a variety of different pollution indices may be more effective in comprehensively expressing the impact of pollution on health. The prediction performance of air pollution indices has been significantly enhanced through the application of neural networks and other methodologies. However, the majority of these studies have focused on the prediction of the AQI, with less attention paid to the prediction of other indexes. The prediction of different pollution indexes can effectively enhance the ability to predict the impact of air pollution on health. The prediction can more accurately reflect the health effects of multi-pollutant environments.

Traditional methods employed for the prediction of air pollutions include statistical methods such as time series analysis and regression analysis, etc[9] [10]. These methods are typically applicable in stable environmental conditions, although they may be less effective in complex air pollution scenarios. The advent of artificial intelligence has led to an increasing number of scientists employing machine learning methods to predict air pollution these include support vector machines (SVMs)[21] and random forests[22], among others. The capacity of machine learning methods to handle complex data and patterns is a significant advantage. The use of neural network models by researchers to predict air pollution is becoming increasingly prevalent. These models

include long and short memory network (LSTM) models, convolutional neural network (CNN) models, and recurrent neural network (RNN) models, as well as their combination models[13] [14], which have been shown to achieve higher levels of accuracy in their predictions. The occurrence of air pollution is influenced by a multitude of factors. Therefore, the incorporation of additional variables into the predictive model enhances its predictive accuracy. In their investigation of the impact of wind fields on $PM_{2.5}$ diffusion, Li and colleagues[25] have identified a need for further research in this area. Building on this, Zhou and co-authors [26] have developed a hybrid neural network model that includes wind field data and has been found to achieve higher prediction accuracy than traditional models. It is important to note that spatial discrepancies among monitoring stations can influence pollution levels. To address this, Qi et al. [27] have developed a station topology map that details the spatial connections between stations using their spatial distribution. The spatial connections were analysed through graph convolution using a CNN-like operation. However, although the inclusion of spatial information from monitoring stations may enhance the accuracy of predictions, it necessitates extra effort and does not explicitly explain how spatial information affects observations, particularly when monitoring stations are geographically distant yet exhibit similar climatic conditions. Consequently, direct spatial extraction is not universally applicable. Therefore, increasing the accuracy of prediction by parameters related to air pollution, such as spatial information, is helpful. In addition, the generalization of spatial information can enhance the convenience of prediction.

The degree of air pollution is influenced by a multitude of factors. In addition to uncontrollable natural factors such as volcanic eruptions, human factors also have a significant impact on pollution levels[28]. This is particularly evident in some developing countries where there are a large number of high-pollution industries and a dense population, which have a great impact on air pollution[29]. As the most populous developing country in the world, China's air pollution problem has consistently

attracted the attention and concern of the government and people, especially in the Beijing-Tianjin-Hebei region, which has a large population and a significant number of polluting industries, resulting in the poor air quality. In addition, since the outbreak of the COVID-19 pandemic began in early 2020, the region has carried out intermittent restrictions on population activity for a period of three years. These restrictions have served to prevent the spread of the pandemic, while simultaneously reducing the exhaust emissions of numerous industrial production vehicles. Furthermore, this is also a unique instance of a mandatory reduction in human contributions to air pollution. Therefore, many researchers [30], [31], [32], [33] have analysed and studied the impact of human activities on air pollution in detail by comparing the air pollution data before the pandemic with the air pollution situation after the implementation of restrictions on human activities. Nevertheless, the period following the lift of the lockdown remains relatively brief, and the requisite time and data have not yet been gathered to substantiate the impact of the resumption of natural life and production. Consequently, this study has collected data on air pollution for the entire year following the lifting of the lockdown in order to facilitate a comparison with the data collected before and during the lockdown period. This will enable a more comprehensive analysis of the impact of the lockdown on air pollution and the real improvement in air quality.

Although extensive studies have been conducted on air quality assessment and prediction, several important research gaps remain unaddressed.

First, most existing prediction models focus only on a single indicator, typically AQI, while multi-index prediction that incorporates both aggregated pollution effects (AAQI) and health-risk-based metrics (HAQI) has received limited attention. This restricts the ability of current studies to fully capture the health implications of multi-pollutant environments.

Second, although spatial information significantly affects air pollution prediction, current spatial modelling approaches often rely on manually collected geographical or

meteorological data, which limits their practicality and general applicability. More importantly, no studies have explored the extraction of spatial relationships directly from historical pollution data.

Third, while many studies have examined changes in air pollution during the COVID-19 lockdown, there is a lack of comprehensive long-term analysis that compares pre-pandemic, pandemic, and post-pandemic periods within the Beijing–Tianjin–Hebei region, especially using multi-dimensional assessments and health-risk evaluations.

This thesis addresses these gaps by making three key contributions:

(1) Developing a unified deep-learning framework for the simultaneous prediction of AQI, AAQI, and HAQI, providing a more holistic representation of air pollution and health risks.

(2) Proposing a novel GCN–LSTM model that extracts spatial correlations from an abstract feature space, eliminating the need for extensive spatial data collection and improving prediction efficiency and accuracy.

(3) Conducting a decade-long multi-index assessment of air quality and its associated health impacts in the Beijing–Tianjin–Hebei region, covering pre-pandemic, pandemic, and post-pandemic phases.

Together, these contributions advance the field by offering more comprehensive predictive tools and deeper empirical insights into air pollution dynamics.

1.2 Research motivations

In the field of air pollution prediction, several limitations persist. Traditional air pollution forecasts primarily focus on AQI predictions, lacking a comprehensive evaluation of health risks in complex pollution environments. Furthermore, the

collection and utilization of spatial information in air pollution prediction remain challenging, making it difficult to enhance prediction efficiency and performance effectively. Additionally, there has been no thorough and systematic analysis of the long-term changes in air pollution and its health impacts, particularly in relation to the variations before, during, and after the COVID-19 pandemic. Therefore, it is of great significance to develop methods for predicting complex air pollution conditions, to simplify and improve the applicability of spatial information collection in pollution forecasting, and to conduct a comprehensive analysis of long-term air pollution trends over large regions, as well as the impact of COVID-19 on air pollution dynamics.

Based on the above background, the motivation for conducting the research presented in this thesis can be summarised as follows:

- The current sole air pollution index is inadequate for the purpose of air quality monitoring. The incorporation of health-related and multi-pollution-related indicators can significantly aid in providing a comprehensive assessment of the pollution status. Therefore, the implementation of multiple indices' prediction will present a challenge. Additionally, applying neural network models to predict multiple pollution indices while achieving higher predictive performance remains a significant challenge.
- Enhanced air pollution forecasting facilitates the safeguarding of human health and supports governments in devising more precise policies concerning environmental regulation. This enables scientists to gain a comprehensive understanding of the mechanisms underlying the formation of air pollution, which in turn facilitates advancements in environmental protection technology. Air pollution prediction models consider spatial influences to enhance precision. However, the gathering and handling of additional spatial data raises concerns regarding its universal applicability. The challenge thus arises in devising a method to enhance prediction accuracy while reducing the efforts required to collect spatial information.

- The COVID-19 pandemic has had a significant impact on air pollution and human health, particularly due to the implementation of preventive and control measures. In the Beijing-Tianjin-Hebei region of China, data on changes in air pollution and its effects on human health following the outbreak remain limited. Furthermore, since 2014, air pollution control measures in the region have led to notable improvements in air quality over time. A comprehensive analysis of these changes could provide valuable insights into regional air quality improvements, with a focus on spatial variations in pollution levels across different areas, shifts in health-related risks, and post-outbreak air pollution trends. Such a study would not only deepen our understanding of the long-term effectiveness of air pollution control measures but also offer a clearer perspective on the outbreak's impact on air quality and its broader implications for public health.

1.3 Research aims and objectives

This study proposes three main research aims. The first aim is to validate the prediction of different pollution indices and achieve a more comprehensive air pollution prediction. The second aim is to enhance the accuracy and accessibility of air pollution prediction. The third aim is to conduct a comprehensive analysis of the impact of the COVID-19 pandemic on air pollution by combining the pollution change data before and after the pandemic and the results of air pollution control in the Beijing-Tianjin-Hebei region after nearly ten years.

In order to achieve high-performance prediction of multiple pollution indices, this study proposed a combined model using multiple neural network models Long Short-Term Memory (LSTM), Multilayer Perceptron (MLP) and Transformer to simultaneously predict the three pollution indices AQI, AAQI and HAQI. This will achieve

simultaneous prediction of three different pollution indices while taking into account the performance and stability of the prediction. Ultimately, the study aims to achieve efficient and accurate prediction of multiple air pollution indices.

To achieve the second research objective, this study proposes a novel GCN+LSTM model. The model utilizes Graph Convolutional Networks (GCN) to extract nodes from an abstract space that captures the correlations between historical data. This approach enhances the prediction parameters while minimizing the need for extensive spatial information collection. The model offers improved prediction accuracy, greater convenience, and streamlined processes.

To achieve the third research objective, this study collected data on various pollutants in 13 cities in the Beijing-Tianjin-Hebei region over a ten-year period (pre-pandemic, pandemic, and post-pandemic) and conducted correlation analyses of these data. The results were then combined with an assessment of their impacts on human health to provide a comprehensive analysis of air pollution trends. Additionally, the study evaluated the effectiveness of air pollution control measures implemented in the region over the past decade, examining the improvements in air quality and their corresponding impact on public health. By integrating these analyses, the study aims to offer a clearer understanding of both the short- and long-term impacts of the pandemic on air pollution, while also highlighting the significant improvements in air quality and public health outcomes resulting from sustained pollution control efforts.

1.4 Research contributions

The main contributions of this thesis are summarised as follows:

- A multi-pollutant indicator prediction model is proposed, representing a significant advancement in the field of air pollution forecasting. Unlike traditional models that

rely solely on AQI to represent the highest pollutant, this model simultaneously predicts AQI, AAQI, and HAQI, offering a more comprehensive and accurate representation of air pollution and its effects on human health. By considering the AAQI associated with each pollutant and the HAQI linked to health risk factors, the model provides a more complete picture of air quality's impact on public health. Furthermore, the model incorporates three state-of-the-art neural network architectures—LSTM, MLP, and Transformer—whose combination enhances predictive performance, accuracy, and stability, significantly improving both reliability and robustness compared to traditional methods.

- A more accurate air pollution prediction method is proposed, significantly improving the prediction accuracy for various pollutants compared to the comparison model, with improvements ranging from 2% to 16%. The model combines the strengths of neural network architectures, particularly GCN and LSTM, to enhance prediction accuracy, while also incorporating spatial information to further optimize its predictive capabilities. Unlike traditional methods, which require considerable time and resources to collect spatial data, the model derives spatial information from an abstract space extracted from historical data at each monitoring station. This eliminates the need for additional data collection efforts and broadens the model's applicability. As a result, the proposed model offers a novel approach that improves the accuracy, efficiency, and usability of air pollution prediction.
- This study provides a comprehensive analysis of the changes in air pollution in the Beijing-Tianjin-Hebei region of China, with a focus on both the long-term impacts of air pollution control measures implemented since 2014 and the short-term changes induced by the COVID-19 pandemic. By examining air pollution trends before, during, and after the pandemic, the study offers valuable insights into the effects of varying levels of human activity, environmental policies, and other factors on air quality. Furthermore, the study assesses the health impacts of air

pollution by calculating the attributable proportions (APs) of mortality and morbidity associated with key pollutants such as $PM_{2.5}$, PM_{10}, and O_3. The findings highlight the significant improvements in air quality resulting from sustained pollution control efforts, as well as the notable reduction in pollution during the pandemic. Additionally, the study compares air pollution levels and health outcomes across different cities within the region, providing a clear picture of the effectiveness of air quality management strategies over the past decade and the enduring effects of the pandemic. These insights can guide future policies aimed at balancing economic development, air quality improvement, and public health in the region.

1.5 Thesis structure

In accordance with the objectives of the study, the structure of the thesis will comprise the following chapters:

- Chapter 1: This chapter presents the research background of this study. The motivation and research objectives for undertaking this study are presented, along with an analysis of the gaps that exist in current relevant research and the direction of the research. Furthermore, the thesis's structure is outlined.
- Chapter 2: This chapter presents a literature review of the different indicators of pollutants, prediction models and the relationship between the pollution situation and health and human activities. This section comprises a historical overview of each study, an account of its development, and an analysis of the advantages and disadvantages of existing research. Furthermore, through the synthesis of existing research, I identify the shortcomings of existing studies and present my own opinions and proposed avenues for future research.

- Chapter 3: This chapter outlines the methodological and geographic information necessary for the research presented in Chapters 4, 5, and 6. It includes an overview of the principles and structure of the neural network models used in Chapters 4 and 5, as well as the calculation methods for evaluating model performance. Additionally, this chapter details the software and relevant formulas utilized in Chapter 6 to analyze changes in air pollution and associated health risks in the Beijing-Tianjin-Hebei region.
- Chapter 4: This chapter firstly describes the correlation and differences between the three air pollution indices AQI, AAQI and HAQI, then describes the workflow of the prediction model proposed for predicting the three indices and the prediction results, and finally describes the validation of the model's performance after applying the model on data from several monitoring stations.
- Chapter 5: This chapter introduces a novel model integrating GCN and LSTM to achieve high-performance predictions of six major air pollutants. The chapter is organized as follows: first, the correlation between each pollutant and key meteorological parameters is analyzed to highlight the influencing factors of air pollution. Next, the architecture and working principles of the GCN+LSTM model are described in detail. Finally, the model's output results are presented, and its performance and stability are evaluated through comparisons with other models.
- Chapter 6: This chapter analyzes the evolution of air pollution in the Beijing-Tianjin-Hebei region over the past decade, focusing on the impact of governance measures since 2014 and the influence of COVID-19 on pollution trends. The chapter is organized as follows: first, it examines the changes in the frequency of major pollutants. Then, the seasonal and annual variations in air pollution levels in the region are discussed. Finally, the chapter investigates the changes in attribution coefficients related to health risks in the region.
- Chapter 7: This chapter provides a comprehensive summary of the key findings and significance of each study presented in this thesis. Additionally, it highlights

the limitations of the research, discusses their implications, and proposes directions for future research to further advance the field.

Chapter 2. Literature Review

2.1 Introduction

Air pollution has resulted in significant damage to public health, the social environment and other aspects. Consequently, the accurate prediction of air pollution, the analysis of changes in air pollution and the correlation with health are of particular importance. This chapter commences with an introduction to the causes, composition and societal, environmental and health-related consequences of air pollution. Then, the chapter describes the development process of air pollution indices, with a particular focus on some new indices (AAQI, HAQI). Next, the evolution of air pollution prediction methodologies and the introduction of advanced prediction models will be presented. Subsequently, pertinent research on the alterations in air pollution resulting from the impact of the COVID-19 pandemic is presented. At the same time, after describing the development status of each study, the shortcomings of the current research and the direction of research are added.

2.2 The hazards and components of air pollution

2.2.1 Hazards of air pollution

Air pollution represents one of the most significant contemporary hazards, both in terms of its impact on climate change and its association with increased morbidity and mortality from respiratory, cardiovascular and cerebrovascular diseases. The

detrimental effects of air pollution on human health can be classified into two distinct categories: indoor and outdoor pollution hazards. Indoor air pollution is influenced by the influx of outdoor air pollutants, in addition to human activities such as smoking, cleaning, and cooking. It is evident that both indoor and outdoor air pollution have a profound impact on human health and associated illnesses[34]. In recent years, the rapid expansion of industry has coincided with a concurrent improvement in the quality of human life and production. However, this has also resulted in a considerable rise in pollution emissions, which have a deleterious effect on human health and well-being [35].

The World Health Organization (WHO) has estimated that ambient air pollution was responsible for 6.7 million deaths in 2019, representing more than 10% of all-cause mortality. This figure is more than double previous estimates [36], [37]. In particular, the problem of environmental pollution is more serious in some low- and middle-income countries, such as China, which is currently eight times more polluted than the United States and is one of the world's worst polluters of sulphur dioxide and particulate matter[38]. Since the 1990s, with the rapid development of China’s economy, air pollutants, mainly smog, have also increased significantly, and the atmospheric visibility in most parts of China has decreased significantly[39]. In particular, the Beijing-Tianjin-Hebei region, situated in the North China Plain, has witnessed the proliferation of numerous heavy industrial factories, which have collectively contributed to the formation of a dense pollution source in the area. The pollution generated by these sources is in excess of the limits of natural purification. In addition, the region is characterised by a dry climate, low surface vegetation coverage, and the surrounding deserts are prone to severe weather conditions such as sandstorms, which result in fine particulate matter remaining in the air for extended periods, contributing to the formation of haze. Furthermore, the utilisation of coal-fired heating for the purpose of winter heating represents a significant source of pollution in the region. The region is home to two of China’s largest cities, Beijing and Tianjin, which have large

populations and experience year-round vehicle congestion. This has a significant impact on air quality[40].

2.2.2 Components of air pollution

Air pollution is defined as the alteration of indoor or outdoor environments due to the introduction of chemical, physical or biological elements that disrupt the natural balance of the atmosphere[1]. Common air pollutants can be divided into two categories according to their characteristics: gaseous pollutants and particulate matter. Particulate matter (PM) pollution is divided into two categories based on particle size: $PM_{2.5}$ (aerodynamic diameter less than 2.5 microns) and PM_{10} (aerodynamic diameter less than 10 microns). The different sizes of particulate matter result in varying invasion points within the human body. PM_{10} is primarily deposited in the upper respiratory tract, whereas finer $PM_{2.5}$ can penetrate deeper into the alveoli[41]. The main sources of particulate matter pollution include factories, power plants, motor vehicles, construction activities, dust and so forth. These are the two pollutants with the highest proportion among air pollutants. Gaseous pollutants include SO_2, NO_2, CO, and O_3, which are primarily the result of the combustion of fossil fuels. Among these, SO_2 is produced by the combustion of sulfur-containing fossil fuels, while CO is a product of incomplete combustion. The main source of CO is road transportation. Nitrogen dioxide is mainly derived from the combustion of fossil fuels, especially motor fuels[42]. Surface ozone is not emitted directly from sources but is generated in the atmosphere by chemical reactions between, for example, nitrogen oxides and volatile organic compounds. Generation is particularly noticeable under sunny conditions[43], [44].

As illustrated in Table 2-1, the main sources of air pollutants include industrial emissions, motor vehicle exhaust emissions, dust, etc.

Table 2-1: Sources and effects of pollutants [45], [46]

pollutant	Main sources	Main impacts
$PM_{2.5}$	Industrial emissions, automobile exhaust, dust	respiratory illness, reduced visibility, contamination of soil and water
PM_{10}	Industrial emissions, building construction, dust	Respiratory diseases, reduced visibility, contamination of soil and water
SO_2	Emissions from industrial processes, home heating, vehicle emissions.	Irritation to respiratory tract, acid rain
CO	Internal combustion engine emissions, heating and cooking, industrial production	hypoxia, coma
NO_2	automobile exhaust, industrial emissions	asthma, chronic obstructive pulmonary disease
O_3	Nitrogen oxide photochemical reaction, sunlight	Stimulate the respiratory tract and affect plant growth

The primary health risks associated with air pollution are respiratory and cardiovascular diseases, including asthma, respiratory infections, bronchitis, and chronic obstructive pulmonary disease, heart disease and other diseases. Furthermore, the pollutants contribute to the development of cardiovascular and cerebrovascular diseases, which can result in conditions such as asthma, respiratory infections, bronchitis, and chronic obstructive pulmonary disease.

2.3 Indices of air pollution

The detrimental impact of air pollution on human health has led to the development of air quality indexes by countries around the world. These indexes serve to communicate to the public the current level of air pollution or the predicted level of pollution. The various countries of the world have developed their own air quality indexes, which correspond to the air quality standards of their respective countries. The main air quality index (AQI) is calculated by determining the highest value among each pollutant and then assigning it an AQI value. However, the phenomenon of air pollution is the consequence of a multiplicity of pollutants. This index is too simplistic and therefore unable to accurately reflect the actual level of air pollution. Consequently, Kyrkilis et al. proposed the overall air quality index (AAQI) and Eugene et al. proposed the health risk-based air quality index (HAQI)[14], [45]. The introduction of these new indices allows for a more comprehensive assessment of the health risks associated with air pollution.

2.3.1 Development of air quality index (AQI)

The genesis of the Air Quality Index can be traced date back to the 1970s. At that time, there was a growing awareness of the health effects of air pollution and a recognition of the need for a simple and intuitive way to describe air quality conditions. Therefore, the concept of the AQI was established. In 1976, the U.S. Environmental Protection Agency (EPA) developed the Air Quality Index (AQI) system with the intention of providing the public with a more comprehensive understanding of air quality and facilitating the implementation of appropriate protective measures. The system is based on the work of the EPA [46]. In the late 1960s and early 1970s, the National Air

Pollution Control Administration (NAPCA) established a link between the concentrations of different air pollutants and their associated health effects. Initially, the AQI in the United States was primarily based on the main pollutants. The following pollutants are included in the calculation of the AQI: O_3, SO_2, CO, PM_{10} and NO_X[47]. As scientific research has deepened and the impact of air pollutants has become better understood, $PM_{2.5}$ has gradually been included in the calculation of the AQI. The U.S. AQI system has undergone continuous improvement throughout its implementation, becoming a model for other countries and regions seeking to develop their own AQI systems. Over time, various countries and regions have developed AQI standards that are tailored to the specific environmental and health conditions present in their respective locales, and have established corresponding monitoring and reporting mechanisms that facilitate the public's comprehension of and response to the impact of air pollution on health[48].

Although the standards set by various countries are slightly different, they are based on six common pollutants: $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, O_3. The grading standards also adopt a 0 to 500 grading standard, based on the concentration of the pollutant, which is converted into a single index. The Table 2-2 illustrates the correlation between the AQI value and the corresponding pollution level. This classification method is generally adopted by countries. The AQI is calculated by first converting the concentration of each pollutant into its corresponding sub-index using a standard piecewise linear interpolation formula. For each pollutant, the AQI is determined based on its position between breakpoint concentration values specified by national air quality standards[12]. The overall AQI is then obtained by selecting the maximum sub-index value among the six pollutants ($PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3), reflecting the principle that the pollutant posing the highest health risk defines the total air quality condition.

The AQI scoring range of **0–500** (with values above 300 generally grouped as "301+") is designed to correspond to increasing levels of public health concern. These ranges are divided into standard categories—such as Good, Moderate, Unhealthy for Sensitive

Groups, Unhealthy, Very Unhealthy, and Hazardous—each representing a specific level of risk based on epidemiological evidence[18].

Table 2-2: The level corresponding to the AQI value[49].

AQI	Grade	Description of air quality
0-50	Good	Air quality is satisfactory, with minimal to no risk air pollution.
51-100	Moderate	The air quality is deemed acceptable. Nonetheless, certain individuals, especially those with heightened sensitivity to air pollution, may face a slight risk.
101-150	Unhealthy for sensitive groups	Individuals belonging to sensitive groups may encounter health effects, while the general public is less susceptible to such impacts.
151-200	Unhealthy	Some individuals within the general population may experience health effects, whereas members of sensitive groups may face more severe health impacts.
201-300	Very unhealthy	Health alert: The risk of health effects is heightened for all individuals.
301 and above	Hazardous	Health emergency warning: Everyone is at higher risk of being affected.

Lower values (0–50) indicate air quality that poses minimal risk, whereas higher values (300+) represent hazardous conditions where severe health effects are possible even for the general population.

As the study area is in China, the Chinese AQI criteria were used. The Table 2-3 presents the different air pollutant concentration values corresponding to the AQI of China. The concentration corresponding to the AQI_m value of each pollutant. In practice, the maximum value of all AQI_m is employed as the final AQI in order to reflect the prevailing air quality. The AQI is the most widely used indicator for air pollution monitoring globally.

Table 2-3: Chinese AQI category and pollutant breakpoints[50]

Individual index	Units are in μg/m^3 except CO, which is in mg/m^3					
IAQI	$PM_{2.5}$	PM_{10}	SO_2	CO	NO_2	O_3 (8-hour average)
0	0	0	0	0	0	0
50	35	50	50	2	40	100
100	75	150	150	4	80	160
150	115	250	475	14	180	215
200	150	350	800	24	280	265
300	250	420	1600	36	565	800
400	350	500	2100	48	750	-
500	500	600	2620	60	940	-

2.3.2 Development and application of AAQI and HAQI

The application of AQI provides a valuable service to the public in the prevention of

pollution and the monitoring and management of government departments in order to predict future air pollution. It is of great significance in improving air quality and protecting public health. However, the pollutants present in the atmosphere are diverse, and the impact of pollution is also complex. AQI only considers the impact of the largest pollutants, thereby failing to acknowledge the harm caused by other pollutants to human health and society[51]. Luo et al. also observed that AQI fails to account for the differential impact of various pollutants on human health in their study of air pollution characteristics in northwest China[52]. To address this issue, various attempts have been made to introduce indices that consider the combined impact of various pollutants on human health. Swamee et al. [13] proposed the concept of a comprehensive air quality index AAQI, which considers the impact of six pollutants rather than the maximum value among them. In addition, Hu et al. proposed an alternative air quality index HAQI, which replaces AQI. The index is based on the correlation between multiple pollutants and health risks, with consideration given to the impact of different pollutants on health risks[53].

Xu et al. [18] conducted a study investigating the spatiotemporal evolution characteristics of air pollutants in North China and their impact on human health. The authors compared the AQI, AAQI, and HAQI of local pollutant changes from 2016 to 2019. The results demonstrated that AQI and AAQI, as well as AQI and HAQI, exhibit strong correlations, with R^2 values of 0.931 and 0.944. The AAQI and HAQI values are higher than the AQI, indicating that there are differences between single and multiple pollutants. Furthermore, the AQI underestimates air pollution.

To investigate the interrelationship between three air quality indexes, Hu et al.[15] analysed the pollution data of six major Chinese cities (Beijing, Shanghai, Guangzhou, Shijiazhuang, Xi'an, and Wuhan). The experiment used data from 2013 to 2014 to first analyse the comparison of AQI, AAQI and HAQI when AQI>100. The results demonstrate that in all cities, the AAQI is consistently higher than the AQI. When the AQI is between 100 and 150, the AAQI is typically higher than the HAQI. However,

when the AQI is above 200, the HAQI is generally higher. The results indicate that when pollution conditions are particularly severe, the AAQI may underestimate the health risks associated with these high-pollution days, even when the combined impact of all pollutants is taken into account. The findings of this study indicate that when air pollution levels exceed the normal range, the AQI only considers the impact of a single pollutant, which significantly underestimates the health impact of multiple pollutants on the public, particularly vulnerable groups such as the elderly and children, as well as patients with respiratory diseases.

Moreover, Luo et al.[16] conducted a comprehensive analysis of the air pollution situation in northwest China based on AQI, AAQI and HAQI. The area is subject to the impact of severe weather conditions, including sandstorms, throughout the year, which contributes to the elevated level of PM_{10}. The research indicates that when the AQI is elevated, the level of PM_{10} is also elevated. The AQI which considers a single pollutant, significantly underestimates the impact of other pollutants such as PM10 on health. This is because the AQI is calculated using the Total Attributable Proportion (AP). The demonstration indicates that PM10 and O_3 also have a significant impact on health.

In addition to the aforementioned studies on multiple pollution indexes, Janjani et al., Ma et al., Mao et al., and Shen et al. [54], [55], [56], [57] also conducted research on the correlation and health risks of multiple pollution indexes in different regions. The findings demonstrated that the AQI, which considers a single pollutant (the one with the highest AQIm among all pollutants), fails to account for the impact of multiple pollutants on health. In particular, when pollution is severe, it is essential to consider the combined effects and hazards of multiple pollutants in order to more accurately reflect the actual hazards of air pollution.

2.3.3 Key Results and Discussion of Multi-Pollution Index

In summary, as the AQI increases, the values of AAQI and HAQI tend to rise more rapidly in comparison, with the difference between them and the AQI becoming increasingly significant. The use of different indicators allows for a more accurate reflection of the harm of air pollution to human health under different pollution conditions. A single AQI does not account for the combined impact of multiple pollutants. In the field of air pollution forecasting, there is a plethora of forecasting methods[58], with the forecasting accuracy continually improving. However, the forecasting of multiple indicators is overlooked. Often, only the AQI indicator is forecasted, which does not provide an accurate reflection. Consequently, it is of paramount importance to utilise sophisticated models to forecast a multitude of pollution indicators, thereby enabling the prediction outcomes to more accurately reflect the detrimental effects of pollution.

2.4 Air pollution forecast

Air pollution represents a significant environmental concern that has a profound impact on human health, particularly the cardiovascular and respiratory systems. It can even result in mortality[59]. Therefore, the accurate prediction of air pollution concentrations plays a pivotal role in the formulation of effective atmospheric management strategies and the prediction of future conditions. It can also facilitate the prevention and control of air pollution and protect the health of the population[60]. The advent of artificial intelligence has brought about a profound transformation in the way human interact with their environment. Its applications have become pervasive, permeating all aspects of human life. In particular, the field of air pollution prediction has benefited greatly

from the application of artificial intelligence, which has led to a notable improvement in the accuracy of air pollution prediction[61]. This section will commence with an overview of the evolution of air pollution forecasting, before delving into the utilisation of various forecasting methodologies. It will then proceed to examine the integration of artificial intelligence with forecasting models that exhibit high levels of accuracy. Finally, this section will present a summary of the shortcomings and limitations of existing models, as well as potential future development directions.

2.4.1 Traditional air pollution prediction methods

Given the detrimental impact of air pollution on human health, researchers from a multitude of countries have dedicated their energies to the prediction of air pollution. Since 1994, research on air pollution AI has been on the rise, with a notable increase since 2017[62]. Initial forecasts can be divided into two broad categories: numerical methods and statistical methods[63]. Statistical models are based on the analysis of changing patterns of pollutants in the air[64]. The Auto Regressive Integrated Moving Average (ARIMA) model is a representative statistical forecasting model that is employed for time series analysis and forecasting. The ARIMA model combines the concepts of autoregressive (AR) and moving average (MA) to describe trends, seasonality and randomness in time series data.

In the process of predicting air pollution, historical air quality data can be used to construct an ARIMA model, which can then be employed to predict future air pollution levels. In a study conducted by Ujjwal et al. [65], the ARIMA modelling method was employed to predict the daily average concentration of air pollutants in Delhi, India.

The results demonstrates that the predicted values are in close alignment with the observed values, indicating that there is minimal discrepancy between the predicted and

actual results. In addition, the model performance was evaluated through the mean absolute error (MAE), which revealed that the MAEs of O_3, NO, NO_2, and CO were 10.9, 7.3, 8.2, and 249.9, respectively. This demonstrated that, in addition to CO, the prediction performance of the other three pollutants was relatively satisfactory[65].

Furthermore, Wang et al.[66] employed the ARIMA model to predict local $PM_{2.5}$ in Los Angeles, USA. The model was able to successfully predict the development trend of $PM_{2.5}$, indicating that it can achieve better results for relatively stable data.

In order to address the limitations of ARIMA, Erdinc[67] proposes a novel approach that combines ARIMA with a wavelet transform. This integration effectively mitigates the issues associated with ARIMA, such as the presence of substantial noise and a high number of outliers, while also enhancing the precision of the prediction. The model was applied to the monthly PM_{10} prediction in Erzurum, Turkey. The results illustrates that the prediction outcomes of the novel Wavelet Transform (WT)-ARIMA model, when combined with wavelet transform, exhibit greater consistency with the observed values and can provide more accurate predictions (R^2 of 0.997 for WT-ARIMA and 0.780 for ARIMA).

Moreover, Zhao et al.[68] also employed a combination of other models based on the ARIMA model with the objective of enhancing the accuracy of the prediction. This study employs a combination of the Akaike Information Criterion (AIC) and modified grid search (GS). This method addresses the limitation of ARIMA in accurately predicting nonlinear fluctuations, and the integration of AIC and GS mitigates the bias associated with the incomplete normality of the data when AIC is employed in isolation to determine the order. The results are compared with those of related studies and the results of the ordinary ARIMA model. The results demonstrate that the model has superior prediction performance. The results demonstrate that the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2) values have all improved significantly. In particular, the RMSE has improved by

99.23%, the MAE by 99.97%, and R^2 by 118.61%. The model has demonstrated a significant enhancement in both its predictive accuracy and its ability to fit the data.

In addition to RIMA, the grey model is also a reliable statistical prediction model for predicting air pollution. This model has a broad range of applications and can be employed to predict linear, nonlinear, and other types of data. Zhou et al.[69] employed an enhanced grey model for air pollution to predict the seasonal concentration of $PM_{2.5}$ in four cities in eastern China. The cities of Shanghai, Nanjing, Hangzhou, and Hefei were selected for analysis. The resulting RMSE values were 6.2110%, 8.3190%, and 5.6336%, respectively. The result is more accurate than other comparison models (SARIMA, LSSVM, BPNN). Wang et al. [70] also enhanced the grey model and integrated the estimation of economic losses resulting from air pollution on this basis. In the model performance evaluation, the new grey model seasonal grey model with the fractional order accumulation (SFGM) was employed for comparison with three widely used predictive grey models, namely grey model (GM), Ant Lion Optimizer (ALO) - GM, and seasonal grey model (SGM). The Mean Absolute Percentage Error (MAPE) of the three models used for comparison was 22.98%, 20.81% and 18.00%, respectively. In contrast, the new model achieved a MAPE of 8.74%, which is significantly lower than the MAPE of the comparison models. Additionally, the potential economic losses associated with health outcomes were quantified based on the prediction results. The results indicated that premature mortality represented the most significant economic loss, accounting for over 70% of all economic losses attributable to $PM_{2.5}$. Furthermore, numerous scholars have employed fuzzy time series models in the prediction of air pollution. Domańska et al.[71] employed a fuzzy time series model to predict the concentrations of particulate matter (PM_{10}, $PM_{2.5}$, SO_2, NO, CO and O_3) for selected hours in the future. This method incorporates relevant climate factors into the derivation of the formula, thereby enhancing the accuracy of the resulting outcomes. For example, the prediction outcomes for PM_{10} demonstrate that the prediction errors for the subsequent 12, 24 and 36 hours are 11.23%, 20.55%, and 21.37%, respectively.

The numerical prediction of air quality is based on the laws of atmospheric movement, integrating the possible physical and chemical changes between pollutants in the air, and using mathematical methods to establish a model of the diffusion of pollutants in the air to predict pollution in the future. This model takes into account changes in substance concentration[72]. The Community Multiscale Air Quality (CMAQ) model is a widely used tool among researchers for predicting air pollution in various regions. CMAQ is an atmospheric chemistry model employed for the purpose of air quality modelling and prediction. The project is an open-source initiative funded by the U.S. Environmental Protection Agency (EPA)[132] [133], which aims to simulate and predict pollutant concentrations in the atmosphere. This information is intended to assist governments and environmental management agencies in formulating and implementing air quality management policies. The following paragraphs set out the basic principles of CMAQ air pollution forecasting:

Atmospheric dynamics simulation:

The CMAQ model initially simulates the movement and transport processes of atmospheric components through an atmospheric dynamics model. These models consider factors such as wind fields, temperature, and humidity in the atmosphere, as well as the physical parameters of the atmospheric boundary layer and terrain features, in order to simulate the movement and mixing process of air.

Pollutant transport and dispersion simulation:

CMAQ employs aerosol and gas transport models to simulate the transport and dispersion of pollutants in the atmosphere. These models consider processes such as turbulent mixing, turbulent diffusion, and convective diffusion in the atmosphere in order to predict the distribution of pollutants in space and time.

Chemical reaction simulation:

Additionally, CMAQ incorporates an atmospheric chemical reaction model, which is employed to simulate chemical reaction processes in the atmosphere. The aforementioned models incorporate chemical reactions between pollutants, photochemical reactions, wet deposition and other processes, as well as mechanisms for the generation and elimination of pollutants.

Boundary conditions and initial condition settings:

In order to make predictions, it is necessary for CMAQ to provide boundary conditions and initial conditions. Boundary conditions encompass atmospheric motion and pollutant concentrations on the outer boundary, whereas initial conditions represent the atmospheric state and pollutant distribution at the commencement of the simulation.

Model validation and evaluation:

Once the predictions have been completed, the CMAQ model is typically subjected to a process of validation and evaluation. This is undertaken in order to assess the accuracy and reliability of the model. This typically involves a comparison of model predictions with observed data, with the objective of validating the model's simulation capabilities and prediction accuracy.

In essence, the fundamental tenet of CMAQ air pollution prediction is predicated upon the simulation of physical processes including atmospheric dynamics, pollutant transport and chemical reactions. This is achieved through the utilisation of numerical simulation methods, which are employed to predict the concentration of pollutants and the quality of the atmosphere.

Wang[75] and others employed the CMAQ model to simulate the transport of pollutants across the Pacific Ocean. The simulated NO_2 and CO were found to be in good agreement with the results of satellite observations. However, it was also demonstrated that the model underestimated the impact of precipitation. Liu et al.[76] also employed the CMAQ modelling approach to analyse regional air pollution in China. The

generated map allows for the detailed observation of the pollution situation in various locations. The results demonstrate that the predicted surface contaminants exhibit pronounced seasonal and spatial variations. The highest surface concentrations of SO_2, NO_2 and O_3 are observed in January, followed by October, April and July. The results demonstrate that the predicted surface contaminants exhibit strong seasonal and spatial variations, with a small error in the representation of pollution changes in each region.

Moreover, Hu's research demonstrated that the CMAQ model can simulate severe air pollution in China[77]. Mathur subsequently extended the capabilities of CMAQ to simulate the distribution of ozone and particulate matter throughout the northern hemisphere[78]. In addition to CMAQ, two-way coupled modelling systems such as WRF-CMAQ can be employed to predict air pollution. Liu et al.[79] employed this method to investigate the $PM_{2.5}$ aerosol feedback effect in South Korea. This study presents evidence of seasonal variations in aerosol concentrations across different regions of South Korea.

A review of the literature reveals that the statistical method has several advantages over the other method. It is relatively simple and easy to use, requires minimal computational resources, and can handle small data sets. Additionally, it is widely regarded as a practical approach. However, the statistical method is subject to a number of constraints, including assumptions about the distribution and relationship of the data, a relatively narrow scope of use, and limitations in its ability to handle complex nonlinear relationships and dynamic changes. Statistical methods typically yield one or a few potential prediction outcomes. The prediction accuracy and stability are relatively low, and the method is therefore unsuitable for data analysis. The quality requirements are high, and the accuracy and completeness of the data must be ensured to prevent inaccurate prediction results[64], [80]. The numerical method has the advantage of being based on the principles of atmospheric science and chemistry, considering atmospheric dynamics and other factors. It has a more accurate simulation effect, is suitable for air quality prediction at different scales, can simulate well from local to

global scales, and can be customized for different regions, time periods and forecasting tasks. However, it is evident that the method also has certain limitations. Firstly, it necessitates a considerable amount of computational effort. The implementation of numerical methods necessitates the availability of substantial computational resources, including high-performance computing hardware, substantial storage capacity, and the ability to handle high-resolution data. The calculation cost is high, and the input data must meet specific requirements, which necessitates the use of a considerable amount of data. In addition, the design of multiple models and parameters using numerical methods is more complex[63], [81], [82].

2.4.2 Improvements in air pollution prediction through by artificial intelligence

Artificial intelligence (AI) is a technology that enables computers and machines to simulate human intelligence and problem-solving abilities. AI can be employed in isolation or in conjunction with other technologies (e.g. sensors, geolocation, robotics) to perform tasks that would otherwise require human intelligence or intervention. Examples of such applications include digital assistants, global positioning system (GPS) navigation, self-driving cars and generative AI tools (such as Chat GPT). As a field of computer science, artificial intelligence encompasses machine learning and deep learning[83]. These disciplines involve the development of artificial intelligence algorithms, which are based on the decision-making processes of the human brain, that can "learn" from available data and make increasingly accurate classifications or predictions over time. In addition, machine learning and deep learning are sub-disciplines of artificial intelligence, with deep learning being a sub-discipline of machine learning. Both machine learning and deep learning employ neural networks to facilitate the acquisition of knowledge from a vast quantity of data. These neural

networks are program structures that emulate the decision-making process of the human brain[84]. They are comprised of interconnected nodes that extract features from data and make predictions about the data's representation. The distinction between machine learning and deep learning lies in the type of neural networks employed and the extent of human intervention incorporated into the design. Classical machine learning algorithms employ neural networks comprising an input layer, one or two hidden layers, and an output layer. Typically, these algorithms are limited to supervised learning, and the data must be structured or labelled by human experts in order for the algorithm to extract features from the data. In contrast, deep learning algorithms utilise deep neural networks, which comprise an input layer, three or more hidden layers, and an output layer. The existence of multiple layers enables the unsupervised learning process to extract features from large, unlabelled, and unstructured data sets[85].

The implementation of accurate and reliable air pollution forecasts can assist environmental protection departments in the implementation of effective control measures, thereby facilitating the improvement of air quality. Conventional methods for forecasting air pollution are either computationally expensive or exhibit poor prediction accuracy and range. The advent of artificial intelligence has markedly enhanced the precision of forecasts and diminished the financial outlay associated with them. This is because AI is capable of effectively capturing nonlinear and high-order interactions between pollutant concentration time series and predictor variables, and also requires less domain expertise and resources as a data-driven approach. Moreover, neural networks accept multi-dimensional data, enabling them to process not only air pollutant concentration data but also to consider the impact of other factors, such as meteorological data, on air pollutant concentration. Therefore, the utilisation of artificial intelligence in the field of air quality prediction is becoming increasingly prevalent[86].

2.4.3 Application of Machine Learning in air quality prediction

The exponential growth of artificial intelligence has led to the development of increasingly sophisticated AI models, which are being employed in a multitude of fields, particularly in the domain of air pollution prediction. This has resulted in a notable enhancement in the precision of prediction. The initial models to be widely adopted were those based on machine learning. The efficacy of various methods, including support vector machine (SVM), random forest, decision tree regression, and linear regression, has been demonstrated in relevant research[87]. Among these, SVM has undergone a significant evolution since its initial proposal by Chervonenkis[88] in 2013. In the field of air pollution prediction, Nilesh et al.[89] employed SVM to predict the AQI in Ahmedabad, India. The data set comprised daily values of various pollutants over a seven-year period. These were subjected to preprocessing prior to the application of SVM. Next, start training the model. The outcomes were contrasted with those of the statistical model ARIMA. The results demonstrate that both the coefficient of determination (R^2) and the mean square error (MSE) are superior to those of the ARIMA model. The SVM demonstrated robust predictive capabilities.

The objective of researchers has consistently been the development of more accurate and effective prediction methods. Methods to enhance prediction efficacy encompass the incorporation of pertinent parameters, the optimization of model performance, and other analogous techniques. In order to enhance the prediction capacity of $PM_{2.5}$, Liu et al. incorporated meteorological variables associated with air pollution into the SVM model, subsequently employing the particle swarm optimization (PSO) algorithm to optimize the SVM classification parameters. This study evaluated the model performance by predicting the number of days of $PM_{2.5}$ pollution level and compared it with the Adaptive Boosting (AdaBoost) model and the ANN model. Results indicated that the PSO-SVM exhibits high accuracy, particularly in the context of minimal

pollution, and demonstrates the strongest predictive capacity. To further evaluate the performance of the models, the researchers compared the precision, recall and F1-score of the three models. Results indicated that the ANN model exhibits superior recall, the PSO-SVM model demonstrates the most optimal overall performance[90].

Liang et al.[21] also employed SVM to predict the Malaysian air pollution index. In this study, the researchers investigated the impact of three kernel functions on the performance of the SVM model: the linear kernel function, the polynomial kernel function and the radial basis function. The aim was to identify the optimal kernel function. Finally, the sum of square errors (SSE), mean square error (MSE) and coefficient of determination (R^2) are employed as performance criteria. The results indicated that the SVM model with linear kernel function exhibits the highest SSE, MSE and the lowest R^2 value, while the SVM model with RBF kernel function exhibits the lowest SSE, MSE, and the highest R^2 value. This indicates that the RBF kernel function can predict the air pollution index with greater accuracy overall.

To evaluate the overall performance of the machine learning model, Abdul et al.[91] simultaneously evaluated the performance of four machine learning models including random forest, decision tree regression, linear regression and support vector machine in Malaysia and compared them. The research results demonstrate that among machine learning algorithms, random forest exhibits the highest performance, followed by linear regression, support vector machine and decision tree (R^2 are 0.958, 0.805, 0.739 and 0.676, respectively).

A comparison of the performance of the above-related literature reveals that machine learning models are more adept at handling complex data, are more flexible, and have higher prediction accuracy than traditional statistical models. However, machine learning is not without its limitations. For example, some feature extraction necessitates manual intervention, which incurs additional labour costs. In addition, machine learning models are susceptible to fluctuations in data volume and data quality. In the

event that the data volume is insufficient or the data quality is inadequate, this may have an adverse impact on the performance of the model, resulting in suboptimal processing efficiency for complex data. consequently, neural network models with enhanced learning capabilities are increasingly employed by researchers to predict air pollution, and their advantages are readily apparent[92], [93], [94].

2.4.4 Application of neural network model in air pollution prediction

Neural network models offer significant advantages over traditional prediction models. Due to its exceptional autonomous learning capabilities, it is not necessary to possess in-depth knowledge of the dynamics between air pollution concentration levels and other explanatory variables. Furthermore, it can also reduce the high computational costs associated with the large number of parameters involved[95], [96]. Therefore, researchers continue to investigate and enhance the predictions of neural network models, with the performance of these predictions also undergoing continual improvement[97].

The Fully Connected Neural Network (FNN) is a classic neural network model. In a study by Fabio et al.[98], the model was employed to debug and predict $PM_{2.5}$ and PM_{10}, with a comparison being made with the statistical model multiple linear regression MLR. The evaluation criteria employed include the correlation coefficient(R), fractional deviation (FB), and normalized mean square error (NMSE). The results indicated that the FNN exhibits superior predictive performance for each pollutant in comparison to the statistical model MRL. Furthermore, research by Shivang[99] and Alimissis[100] has demonstrated that neural network models, such as ANN and FNN exhibit superior performance compared to traditional prediction models. More than 90% of the evaluation indicators of the neural network model are better than those of the traditional prediction model.

In addition, Patricio et al.[101] employed another neural network model, MLP, to predict PM_{10} in Santiago, Chile, and compared it with a linear model. The findings indicate that the prediction outcomes of the neural network model MLP on data from multiple monitoring sites in Chile are superior to those of the linear model (average error of MLP and Linear model are 0.16 and 0.18 respectively).

Moreover, the Long Short-Term Memory (LSTM) model proposed by Hochreiter[102] has been employed by numerous researchers in the field of air pollution prediction exhibiting remarkable performance. For instance, Zhou et al.[103] employed an enhanced model based on LSTM to predict the primary air pollutants $PM_{2.5}$, PM_{10}, and NO_2 in Taipei, resulting in stable and precise predictions, improvement rates range from 2.18% to 13.91%. Furthermore, Wu et al.[104] employed a novel model based on the LSTM model to predict the daily air quality index of two cities, Beijing and Guilin. Comparative experiments demonstrated the effectiveness of the model in predicting air quality in different cities, with a high accuracy rate, RMSE values improved by 39.3% to 52.25%. The adaptation of the model to factors related to air pollution can effectively enhance the prediction accuracy. For example, Zhang et al.[105] integrated data from air quality monitoring stations, monitoring stations in nearby industrial areas and external pollution source monitoring stations to create a combined prediction model based on the LSTM algorithm. This aggregated LSTM model was developed by integrating data from different monitoring stations and was compared with the SVR, GBTR, and LSTM models to assess its performance. The performance of the model is evaluated using MAE, RMSE and MAPE. The MAE error of the proposed model is 0.74, 0.34 and 2.32 lower than the LSTM, SVR and GBT. The findings indicate that as the number of hours in the future predicted increases, the aggregated LSTM model demonstrates a growing advantage.

A multitude of neural network models exist. In addition to LSTM, Gated Recurrent Unit (GRU) is also an excellent recurrent neural network. For example, Zhou et al.[106] employed GRU to predict $PM_{2.5}$ in Beijing, with the outcome indicating that GRU

exhibited a high degree of predictive accuracy. Shilpa et al.[107] also predicted NO_2 pollution changes in India after optimizing GRU for air pollution conditions and compared it with other models. This indicates that GRU exhibits satisfactory performance in forecasting diverse air pollutants.

Extreme Gradient Boosting (XGBoost) exhibits efficient parallel computing capabilities and is robust to the presence of missing values. Given the prevalence of missing data in air pollution datasets, XGBoost is an optimal choice for air pollution prediction[108]. In a study conducted by Pan et al.[109] used XGBoost was employed as a prediction model for hourly $PM_{2.5}$ concentration in Tianjin. Additionally, the study compared the prediction results with those obtained from random forest, multiple linear regression, decision tree regression, and support vector machine regression models. The results demonstrate that the XGBoost algorithm outperforms other methods. In a study conducted by Li et al.[110]. XGBoost, Random Forest (RF), Decision Tree Regression (DTR), Linear Regression (LR), Stochastic Gradient Descent Regression (SGDR), Support Vector Regression (SVR), and WRF-Chem were employed in a simultaneous optimization of pollutant concentration in Beijing. The performance of the various models was evaluated using the R, RMSE and MAE metrics. The R, RMSE and MAE of the XGBoost reaches 0.97, 20.80 and 11.49 respectively. It is much higher than comparison model.

The results of the preceding table demonstrate that the XGBoost algorithm exhibits the most optimal performance, with the highest correlation coefficient and the smallest error. The results indicates that the neural network model has been demonstrated to be an effective technology, exhibiting high efficiency and high prediction accuracy. In comparison to traditional prediction methods, it is capable of achieving superior accuracy in the context of non-linear relationships, and its predictive capabilities are markedly superior. Although neural network models are effective in improving prediction accuracy, a single prediction model has inherent limitations when confronted with more complex problems. The hybrid model effectively addresses this issue. The

term ‘hybrid model’ encompasses a combination of different algorithms or methods, as well as the integration of the respective advantages of each algorithm to achieve enhanced performance[20], [111].

In essence, simple combination models typically integrate feature extraction methodologies with predictors. For example, Bai et al.[112] integrated wavelet technology and a backpropagation neural network (BPNN) to create a W-BPNN model for forecasting the concentration of daily air pollutants. The model initially applies the stationary wavelet transform (SWT) to decompose the historical time series of daily air pollutant concentrations into different scales. This process represents the wavelet coefficients of the air pollutant concentrations. Subsequently, the wavelet coefficients are employed to train the BPNN model at each scale. In comparison to the BPNN, the W-BPNN model demonstrated a reduction in the MAPE for PM_{10}, SO_2 and NO_2 by 15.989%, 6.253% and 18.486% respectively. The results demonstrate that the decomposition of mixed air pollutant concentration features into multiple simple features through SWT can effectively enhance the prediction level. Dhirendra et al.[113] also adopted a similar approach, combining feature extractors and prediction methods. A novel model for NO_2 prediction is presented. The model initially employs principal component (PC) analysis to identify the interrelationships between distinct predictor variables pertaining to meteorological conditions and pollutants. Subsequently, the significant variables are utilised as output parameters, and finally an artificial neural network (ANN) is employed to predict air pollution. A comparison of the R, NMSE and other evaluation indicators with the multiple linear regression model (MLR) revealed that the PCA-ANN model demonstrated superior performance. The R-value for PCA-ANN model is 0.89, while that for the MLR model is 0.69.

As research in the field of air quality prediction continues to deepen, the input data required for air pollution prediction increases, with the result that the accuracy and comprehensiveness of the prediction become more important. Therefore, many researchers[114], [115], [116], [117], [118] have proposed hybrid prediction models,

which combine the strengths of multiple neural network models to jointly process input data, thereby markedly enhancing the prediction capacity of air pollution[20]. Wu et al.[119] proposed a composite model comprising multiple processors, namely CEEMDAN-PE-GWO-VMD-MIF-BiLSTM-AT. The model initially employs adaptive noise complete ensemble empirical mode decomposition (CEEMDAN) to decompose the hourly $PM_{2.5}$ series. Secondly, the Gray Wolf Optimizer (GWO) algorithm is employed to optimize the parameters of variational mode decomposition (VMD). The GWO-VMD is employed to optimize the parameters of VMD with the objective of decomposing the maximum permutation entropy (PE) of the first decomposition. Then, a bidirectional long short-term memory network (Bi-LSTM-AT) was employed to construct hourly $PM_{2.5}$ concentration predictions by integrating multiple influencing factors (MIF), including air pollution factors, meteorological parameters and Global Navigation Satellite System-derived Zenith Total Delay (GNSS-ZTD). The model was subjected to testing in four Chinese cities: the model was tested in Beijing, Wuhan, Urumqi and Lhasa, and compared with other models. The findings indicate that the model exhibits enhanced capacity for hourly $PM_{2.5}$ concentration prediction. The average RMSE, MAE, MAPE and R^2 of hourly $PM_{2.5}$ for the four cities were 1.5737, 1.3025, 0.0518 and 0.9961. The advantage of the CNN model lies in the extraction of features[120], while the advantage of the LSTM model lies in the mining of time series data[121]. Therefore, Li et al.[122] integrated these two models to create a novel model CNN-LSTM, and applied it to the prediction of $PM_{2.5}$ concentration in Beijing. In order to test the efficacy of the proposed model, four distinct models were developed: univariate LSTM, multivariate LSTM, univariate CNN-LSTM and multivariate CNN-LSTM. Finally, two indicators, MAE and RMSE, were employed to evaluate the model. The MAE of Multivariate CNN-LSTM reaches 13.9697. The results indicate that the multivariate CNN-LSTM model exhibited the most optimal performance. This indicates that the multivariate composite model exhibits enhanced processing capabilities for data sets comprising a substantial volume of data and multitude of

features. Moreover, Unjin et al.[123] applied the CNN-LSTM model to the prediction of O_3 in Beijing. The experimental results are compared with those obtained using the MLP and LSTM models. The results demonstrate that the performance indicators RMSE, MAE, and MAPE of the CNN-LSTM model are reduced by 83% in comparison to MLP and by 35% in comparison to LSTM. The research by Huang et al.[124] also demonstrated that the hybrid model CNN-LSTM exhibited superior performance in predicting air pollution compared to other models. A multitude of composite models exist. For example, Chen et al.[125] have proposed a hybrid model combining CNN and Transformer. The model initially extracts valuable information in the feature dimension from the CNN layer, compensating for the limited ability of the Transformer encoder to mine information from multivariate data sets. This results in a CNN-Transformer model with higher prediction accuracy, the short-term forecast RMSE value is 7.75 and the long-term forecast RMSE value is 16.27, both of which are the most superior. In addition, Zhang et al.[126] integrated convolutional neural networks, bidirectional long and short-term memory neural networks, and attention mechanisms to develop a novel CNN-BiLSTM-Attention model. In this model, CNN is employed to extract effective spatial features from all factors, while BiLSTM is utilized to address the vanishing and exploding gradient issues in a time series manner and identify temporal features in both directions of the hidden layer. Finally, an attention mechanism is employed to analyze the relative importance of all features and to assign corresponding weights to each feature. Each model is capable of leveraging its respective strengths to enhance the precision of $PM_{2.5}$ concentration forecasting. The results were compared with multiple models and demonstrated that the model exhibited high accuracy in predicting multiple times (MAE:2.366, RMSE:3.095, R^2: 0.960).

The preceding literature demonstrates that a composite model can compensate for the limitations of individual models, thereby creating a more comprehensive and robust model. Composite models are capable of processing larger data sets with greater efficiency and accuracy, with higher accuracy rates[127]. This is of significant benefit

in the context of air pollution prediction in complex scenarios. In addition, the above-mentioned literature indicates that numerous parameters are associated with fluctuations in air pollution, and the incorporation of additional parameters can enhance the precision of prediction. Among these, the most commonly relevant parameters are meteorological parameters. It can be demonstrated that air pollution is diffusive[128]. Therefore, incorporating spatial information as parameters into the model will enhance the accuracy of prediction[129].

The monitoring of air pollution in urban areas is typically conducted by means of a network of monitoring stations distributed across the city. The data collected by these stations is subject to a certain degree of correlation. Therefore, Zhao et al.[130] proposed a fully connected neural network model LSTM-FC for the purpose of predicting $PM_{2.5}$ concentration. The model simultaneously predicts the data of a designated monitoring station and the data of multiple monitoring stations in the vicinity of the designated monitoring station. It employs a full-connector neural network to integrate the data of the surrounding monitoring stations after analysis and integration. The final prediction results are obtained from the data predicted by the designated monitoring station. The results of this study were compared with those obtained using an ANN and LSTM, with MAE and RMSE serving as the evaluation indicators.

The MAE, RMSE reached 23.97 and 35.85 respectively. The results demonstrate that the LSTM-FC model, which incorporates data from surrounding monitoring stations, exhibits superior performance compared to other models.

Furthermore, Qi et al. [27]utilized data from air pollution monitoring stations across the Beijing-Tianjin-Hebei region to predict air pollution in the same region. They incorporated the spatial information of each station as parameters in a GCN-LSTM model. This model effectively integrates spatial information from each site, thereby enhancing prediction accuracy. The MAE and RMSE of the 24-hour prediction of GC-

LSTM reached 23.10 and 35.92, respectively, which are both better than other models. It is evident that the performance of the GCN-LSTM model that incorporates spatial information has been considerably enhanced in comparison to the traditional model.

The results demonstrate that the model incorporating spatial information as a parameter exhibits greater strength than the general composite model and the single neural network model.

2.4.5 Selection of specific models

- LSTM-MLP-Transformer for multi-indices prediction

When simultaneously predicting multiple air pollution indices (AQI, AAQI, and HAQI), the task is inherently more challenging than traditional single-index prediction because these indices differ in their calculation methods, weighting schemes, and sensitivities to individual pollutants. As a result, multi-index prediction places greater demands on the model's ability to handle heterogeneity across targets. The model was selected based on the following criteria:

1. Complementarity in modeling temporal patterns
2. Robustness across indices with different sensitivities
3. Avoidance of excessive model complexity
4. Transparency and interpretability
5. Empirical performance on validation data

LSTM serves as a robust baseline model due to its proven ability to capture temporal dependencies in air quality time series. The LSTM model, through its gating mechanisms, can retain important long-term dependency information while filtering out

historical information that contributes little to the prediction task. Owing to its distinctive memory cell structure, LSTM can effectively capture both short-term fluctuations and long-term trends in air pollutant concentrations. Consequently, LSTM is particularly well suited for air pollution prediction problems characterized by complex temporal dynamics and nonlinear relationships. As reviewed in Section 2.4.4, a substantial body of existing studies[130] [27] [119] [123] has adopted LSTM as a baseline model for air pollution forecasting, further demonstrating its suitability and effectiveness in this field.

The Multi-Layer Perceptron (MLP) is one of the most fundamental forms of feedforward artificial neural networks and consists of an input layer, one or more hidden layers, and an output layer. Each layer is composed of multiple neurons that perform a weighted linear transformation followed by a nonlinear activation function, enabling the network to approximate complex nonlinear mappings between inputs and outputs. Owing to its universal approximation capability, MLP can effectively model highly nonlinear relationships without requiring explicit assumptions about the underlying data distribution[194].

The Transformer is a neural network architecture originally proposed for sequence modelling tasks and is based entirely on the self-attention mechanism, without relying on recurrent or convolutional structures. By computing attention weights between all elements in an input sequence, the Transformer can explicitly model both short-range and long-range dependencies in time-series data. This parallel attention mechanism enables efficient learning of global temporal patterns and reduces the limitations associated with recurrent architectures, such as gradient vanishing and limited long-term memory[125]. In air pollution prediction, pollutant concentrations are influenced by complex temporal dynamics and long-term trends, as well as by interactions among multiple meteorological and emission-related factors. The self-attention mechanism allows the Transformer to dynamically assign different importance to historical observations at different time steps, thereby capturing non-stationary and nonlinear

temporal relationships in pollutant concentration series. In addition, the multi-head attention structure enables the model to learn diverse temporal dependency patterns simultaneously, which is particularly beneficial for complex environmental time-series data.

In this study, LSTM is selected as the baseline model due to its strong capability in modeling temporal dependencies and capturing both short-term fluctuations and long-term trends in air pollution time series. MLP, owing to its simple and computationally efficient structure, demonstrates higher efficiency and competitive prediction accuracy when applied to relatively simple data patterns. Transformer-based models, on the other hand, exhibit strong performance in handling complex temporal relationships and large-scale datasets; however, their predictive stability may be compromised when the available data volume is insufficient.

Therefore, by adopting LSTM as the baseline model and integrating MLP and Transformer components, the proposed framework is capable of achieving efficient and stable prediction across small-, medium-, and large-scale datasets, thereby enhancing its adaptability to multi-pollutant air quality prediction tasks.

- GCN+LSTM for prediction of six main pollutants

Air pollutant concentrations exhibit strong temporal persistence due to atmospheric processes, as well as spatial dependence caused by pollutant transport, regional emissions, and meteorological conditions. Therefore, the selection of a prediction model must account not only for strong temporal dependencies but also for pronounced spatial dependencies. Only models that are capable of effectively capturing and integrating both temporal and spatial information can achieve higher prediction accuracy. Therefore, the model was selected based on the following criteria:

1. Ability to model long-term temporal dependencies

2. Capability to capture irregular spatial relationships
3. Suitability for multi-pollutant, multi-site prediction
4. Computational efficiency and stability
5. Proven effectiveness in air pollution forecasting literature

The performance of LSTM models in air pollution prediction has been extensively validated in the literature [27] [130] [123]. LSTM is capable of capturing both short- and long-term dependencies and exhibits strong stability when modelling nonlinear time-series data. As a result, it is well suited to serve as a baseline model and as a benchmark for evaluating whether the proposed hybrid models achieve meaningful performance improvements.

Graph Convolutional Networks (GCNs) are designed to model data with graph-structured representations, where nodes denote entities and edges represent relationships between them. Unlike traditional convolutional neural networks that operate on regular grid structures, GCNs can effectively capture irregular and non-Euclidean spatial dependencies, making them particularly suitable for modelling interactions among air quality monitoring stations. Unlike CNNs that rely on fixed local receptive fields, GCNs allow spatial dependencies to be defined in a data-driven manner through graph adjacency matrices. Therefore, GCN serves as a highly effective and well-suited extractor of spatial information.

For predicting the six major pollutants, hybrid models have demonstrated strong performance[119] [122] [123]. In particular, models that incorporate spatial information—such as GCN-LSTM—provide more accurate predictions by leveraging both temporal and spatial dependencies. Therefore, when adopting GCN to improve predictive accuracy, simplifying the process of spatial information extraction could further enhance both model performance and applicability. Finally, to comprehensively evaluate model performance, the proposed model should be compared against classical

LSTM, simplified GRU, and hybrid models such as LSTM-Attention. These proposed models will also be validated in subsequent chapters.

2.4.6 Summary of the performance improvement of neural network models for air pollution prediction

As evidenced by the literature presented above, the prediction of air pollution has evolved from the initial statistical model to the current use of neural network models. In particular, the final composite model integrates the advantages of different neural network models in order to enhance the accuracy of the prediction of air pollution. The enhanced prediction performance is of great significance for the prevention of air pollution and the formulation of related air pollution control policies. In particular, the incorporation of spatiotemporal information that is highly correlated with air pollution has been demonstrated to further enhance the prediction performance of air pollution. However, the incorporation of spatiotemporal data introduces a further layer of complexity to the modelling process and the collection of data, necessitating additional work. If the requisite workload can be reduced while incorporating spatiotemporal information, the process will be more straightforward, and the accuracy of the predictions will be enhanced. Based on the advantages of different models in existing research, the feasibility models for future research are proposed.

2.5 The impact of COVID-19 on air pollution and human health

At the beginning of 2020, the COVID-19 emerged on a global scale. In order to prevent

the rapid spread of the virus and protect public health, governments around the world took unprecedented measures, such as restricting road and air travel and reducing the movement of people. The restrictions imposed as a result of the pandemic led to a significant decline in air pollutant emissions in numerous countries during this period. These measures present a unique opportunity for researchers to assess the impact of anthropogenic emissions of air pollutants on air quality[131]. This section will first introduce the global impact of the spread of COVID-19. We will then proceed to present the relevant literature on the changes in various pollutants during the COVID-19 pandemic. This will be followed by a discussion of the relevant literature on the comparative analysis of air pollution changes before and during the pandemic. In addition, we will examine the interrelationship between the spread of the virus and the deterioration of air quality and human health. Finally, this section will present some research gaps in the current study.

2.5.1 Global impact of the spread of COVID-19

The coronavirus was an ongoing and devastating pandemic that had spread rapidly around the globe due to its high contagiousness. On 11 March 2020, the World Health Organization (WHO) declared the virus a pandemic. This has resulted in the implementation of various forms of lockdown in almost all regions of the globe. At least five pandemics have occurred in this century, including the 2009 H1N1 influenza pandemic, the 2014 polio outbreak, the 2014 Ebola virus pandemic, the 2016 Zika virus outbreak, and the 2019 Ebola virus pandemic, among others. Notably, none of these incidents resulted in a global lockdown. Instead, the emergence of the COVID-19 has become the most significant public health crisis of the 21st century[132].

The virus is highly contagious and can be transmitted through droplets and close contact. Aerosols generated by coughing, sneezing, and breathing can be inhaled through the

nose or mouth and penetrate into the human respiratory system. The symptoms of individuals infected with COVID-19 range from mild to severe pneumonia. Those who are clingy to others or have pre-existing respiratory problems are at greater risk pf developing severe pneumonia. It is inevitable that any pandemic will cause irreparable damage to lives and society [133], [134], [135].

2.5.2 Impact of COVID-19 on air pollution

The main sources of ambient air pollution include industrial production, stationary power generation, emissions from automobiles, aircrafts, ships and other transportation vehicles[136]. The global outbreak of COVID-19 in early 2020 has had a significant impact on global society, prompting governments of various countries to implement restrictions on social activities. As a consequence of the imposition of strict restrictions, there was a significant reduction in population movement and production, as well as in the general level of activity. In such exceptional circumstances, there is an opportunity to assess the impact of restrictions on people's daily activities on air pollution.

Zander et al. [137] conducted an analysis of the changes in NO_2 and $PM_{2.5}$ in 34 countries and regions in 2020 compared with the same period in 2019. The data was collected by air quality stations. The results indicated a 19% reduction in NO_2 concentration and a 29% decline in $PM_{2.5}$ concentration. Notably, in developing countries such as China and India, NO_2 concentrations decreased by 16 $\mu g \cdot m^{-3}$ in China and 15 $\mu g \cdot m^{-3}$ in India. In addition, the study revealed that global NO_2 and $PM_{2.5}$ abnormalities related to the blockade returned to normalcy following the blockade’s lifting. Furthermore, beginning in 2020, China's NO_2 concentration has been observed to be 10% to 30% lower than the same period in 2019[138]. The NO_2 concentrations in Bergamo, Italy, and Barcelona, Spain, decreased by 47% and 55%, respectively, in comparison to the same period in 2019[139]. Jesse et al.[140] conducted a study on the

changes in air pollution in the United States during the COVID-19 pandemic. The study sought to ascertain the differences in NO_2 and $PM_{2.5}$ concentrations during the period of the COVID-19 pandemic, comparing data from 2017 to 2019 and 2020. The results indicate that during the period of the COVID-19 pandemic, the concentration of NO_2 decreased by 25.5%. The city's NO_2 concentration exhibited a pronounced decline, reaching a level of 26%. The closure of non-essential businesses at any point during the study period was associated with a statistically significant reduction in NO_2 levels. Jesse and colleagues conducted a study to examine the impact of the COVID-19 pandemic on air pollution in the United States. The study assessed the differences in NO_2 and $PM_{2.5}$ concentrations during the COVID-19 pandemic from 2017 to 2019 and 2020. The results indicate that during the period of the global pandemic caused by the novel coronavirus, NO_2 levels decreased by 25.5%. The city's NO_2 levels exhibited a pronounced decline, reaching a value of 26%. The closure of non-essential businesses at any point during the period under consideration resulted in a statistically significant decrease in NO_2 levels. $PM_{2.5}$ levels decreased by 11.3%, yet there was no discernible difference between the current and historical data in rural areas. The decline in $PM_{2.5}$ is less pronounced than that of NO_2, potentially due to the complex nature of $PM_{2.5}$ sources, which are often attributable to multiple non-traffic sources. In summary, the overall concentrations of pollutants have decreased significantly during the course of the COVID-19 pandemic.

In a study conducted by Alvaro et al. [141], the air pollution levels in 11 representative cities in Spain were compared during the period of the COVID-19 pandemic with the same period in 2019. A regression model was employed to analyse the changes in five pollutants. Furthermore, the study considered the influence of meteorological variables, including temperature, precipitation, wind, sunshine hours. The results indicate that the lockdown has led to a notable reduction in air pollution, particularly with regard to NO_2. Moreover, the levels of other pollutants have also declined, with ozone levels either decreasing to a low extent or even increasing. It is possible that the reduction in NO

emissions has resulted in a relatively high content of O_3 due to the ability of O_3 to synthesise NO_2 with man-made NO emissions.

Zhang et al. [142] conducted a study on the three-year changes in air quality in Chengdu, China, with a focus on aerosol particles. The data on emission sources, including vehicle emissions, dust, biomass burning, and coal burning, were subjected to analysis. The results demonstrate a reduction in the concentrations of $PM_{2.5}$, PM_{10}, SO_2, and NO_2 between 2020 and 2022. Specifically, the concentrations of these pollutants decreased by 25.6%, 24.7%, 28.8%, and 38.5% respectively, while the O_3 concentration increased by 11.0%. In addition, the contributions of particulate matter from biomass combustion and coal combustion decreased by 12.1% and 0.9% respectively, while those from vehicle emissions and dust decreased by 22.2% and 11.0% respectively.

Moreover, Chu et al.[143] analysed the changes in air pollution during the COVID-19 pandemic in Wuhan, the most severely affected city in China, as well as in Hubei Province and throughout China. The results indicate that, in comparison to the same period in 2019, $PM_{2.5}$ level in Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei) at the beginning of 2020 decreased by 53%, 50%, and 30%, respectively. The reductions in NO_2 and PM_{10} are also positively correlated with the decline in $PM_{2.5}$. Concentrations of SO_2 and CO also exhibited a decline, although the magnitude of this decline was considerably less than that observed for $PM_{2.5}$ and NO_2. These observations indicate that emissions from stationary sources, such as coal-fired power plants and steel production, have not declined to the same extent as traffic. In addition, the study revealed that the most significant decreases in $PM_{2.5}$ and NO_2 concentrations were observed in the North China Plain and the Yangtze River Delta. In contrast, O_3 concentrations exhibited an increase in the majority of regions across China. Furthermore, further analysis of the impact of the control measures implemented in response to the COVID-19 pandemic on air quality revealed that, once the pandemic was controlled and the lockdown was lifted, the order of production and living was essentially restored, and the concentration of pollutants returned to normal levels. This

indicates that the concentration of air pollutants is contingent upon the implementation of control measures.

2.5.3 Changes in air quality before and after COVID-19

The analysis of data collected during the period of the COVID-19 pandemic can provide insight into the impact of the lockdown on air quality. The changes in air quality following the lifting of the lockdown can be more comprehensively evaluated in order to gain a more accurate understanding of the changes in air quality before and after the lockdown.

China has implemented a rigorous control strategy to contain the spread of the pandemic, which has resulted in a stabilisation of the national pandemic situation from May 2020 to the first half of 2021. Uzair et al. [144] conducted an analysis of the early stages of the pandemic, from January to May 2019, and the active period of the pandemic, from January 2020. A comparative analysis was conducted on the changes in air pollution from January to May 2021 during the period of relative calm following the initial stages of the pandemic. The study revealed that following the implementation of stringent lockdown measures in Jiangsu, there was a 72% reduction in PM_{10}, a 22% decline in NO_2, a 56% decrease in $PM_{2.5}$, a 39% fall in SO_2, and an increase of 21% in O_3. Prior to the onset of the COVID-19 pandemic, the average concentration of $PM_{2.5}$ in the majority of cities in Jiangsu was found to exceed China's environmental standards. However, during the active phase of the pandemic, the average concentration of $PM_{2.5}$ in these same cities was observed to be below the aforementioned standards. In addition, the government's control of air pollution resulted in a gradual decrease in air pollution from 2017 to 2021 (Figure 2-1). In particular, the downward trend became more pronounced in 2020, and then during the period of relative calm in 2021, the concentration of air pollution exhibited an upward trend. The average concentration of

$PM_{2.5}$ decreased by 18% during the period of elevated activity associated with the COVID-19 pandemic, while only a 2% increase was observed during the subsequent period of relative calm. PM_{10} exhibited a 19% decline during the active phase of the COVID-19 pandemic, while a 16% increase was observed during the subsequent period of relative calm. The changes observed in O_3 are particularly noteworthy. During the active period of the COVID-19 pandemic, there was an increase of 11% in the concentration of ozone, while during the subsequent quiet period, there was a 20% decrease. This evidence suggests that the lockdown measures implemented during the COVID-19 pandemic had a significant impact on the reduction of air pollution. During the normal period of the pandemic, there was an increase in the level of pollution. However, when compared with the early stages of the pandemic, there was still a slight decrease.

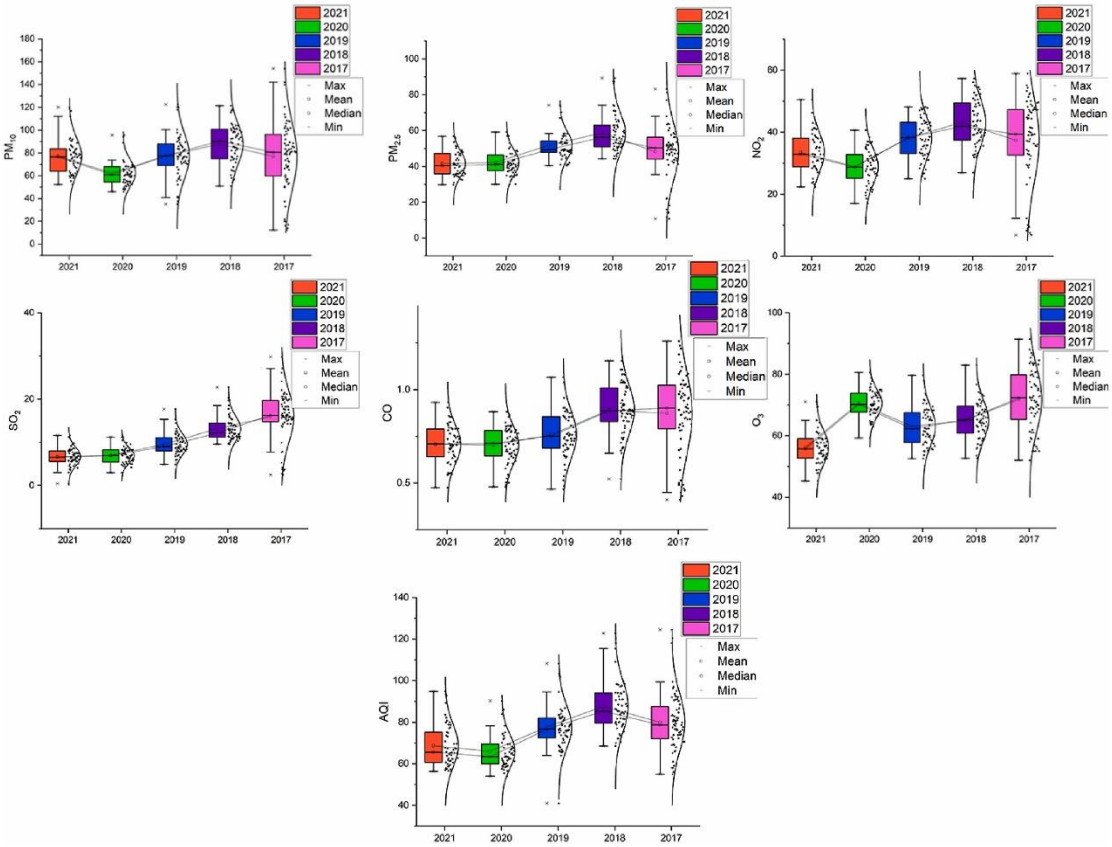


Figure 2-1: Yearly change in the concentration of pollutants in Jiangsu China 2017–2021(figure created by Uzair et al.)[144].

In a similar study, Pragya et al.[145] examined changes in air pollution levels in India before, during and after the nationwide lockdown. The study examined air pollution in the Dewas region over the course of five years, from 2018 to 2022. The findings indicated that the local pollution levels during the blockade period were markedly distinct from those observed prior to and following the lockdown. During the lockdown period, the concentration levels of particulate matter and gaseous pollutants were found to be significantly reduced. Similarly, Leonardo et al. [146] examined the changes in air pollution across all Italian provinces during the lockdown period. Italy experienced abnormally low temperatures during the lockdown period, which resulted in an increase in the concentration of PM_{10} rather than a reduction. This in conjunction with local heating led to an increase in wood burning, which compensated for the reduction in PM_{10} emissions from vehicle emissions. The concentration of PM_{10} emissions increased by 30% in urban areas and 20% in rural areas. NO_2, which is closely related to liquidity, exhibited a sharp decline in almost all provinces, with a reduction of approximately 50%. This indicates that mobility restrictions have a significant impact on vehicle emissions.

In evaluating the impact of lockdowns on air quality, it is essential to consider the influence of local air pollution control measures on air quality. In their assessment of air quality improvement in China's Greater Bay Area, Wang et al.[147] not only analysed the improvement level of air pollution during the lockdown period in 2020 compared with the previous year, but also evaluated the impact of other factors such as air pollution control measures on changes in air quality. The results of the simulation indicated a reduction in the concentrations of SO_2, NO_2, PM_{10} and $PM_{2.5}$ by 1.38 $\mu g/m^3$ (3.5%), 6.57 $\mu g/m^3$ (7.4%), 2.96 $\mu g/m^3$ (2.3%) and 1.82 $\mu g/m^3$ (2.0%) $\mu g/m^3$ respectively in 2020. In contrast, only O_3 exhibited an inverse trend, with a 5.52 $\mu g/m^3$ increase (-2.4%). In addition to the shutdown of vehicles and industrial sectors due to lockdowns, which reduced pollution, local control measures also made a greater contribution. As illustrated in the Figure 2-2, the contribution rate of control measures

to the reduction of SO_2 reached 59.2%, NO_2 was 52.2%, O_3 was 48.1%, and PM_{10} is 61.9%, and $PM_{2.5}$ is 64.0%. The results demonstrate that the implementation of lockdowns has significantly enhanced air quality, with pollution control measures also playing a pivotal role. The implementation of the lockdown has resulted in a notable enhancement of the overall quality of the air. However, it is important to note that certain meteorological conditions may also result in elevated levels of pollution during periods of lockdown.

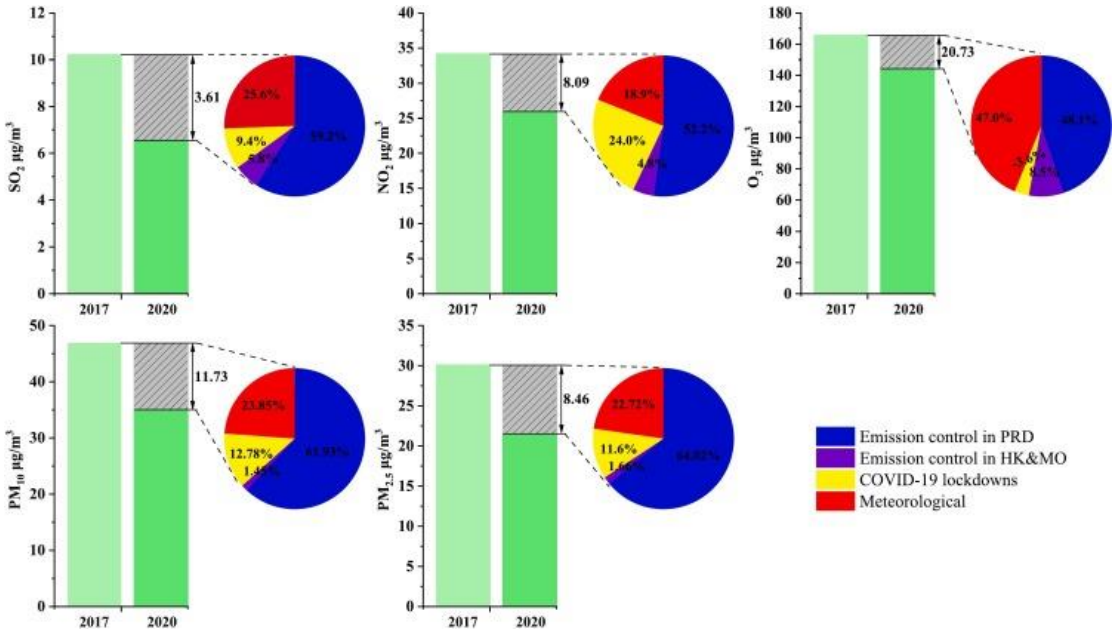


Figure 2-2: The Integrated portion of the contributions of control measures, meteorology, and the COVID-19 lockdowns[147].

Such conditions include persistent low wind speeds, frequent inversions, and convergence of wind directions. In North China, the concentration of $PM_{2.5}$ once exceeded 80 $\mu g/m^3$ during the lockdown period, and the concentration of other pollutants also increased. It was not until the wind direction changed and the wind speed increased that the situation improved[148].

2.5.4 Health effect of air pollution

In addition to the environmental, economic, and food safety risks that air pollution presents, it also has a significant impact on human health[149]. Since the 1980s, epidemiological studies have consistently demonstrated a correlation between particulate matter and mortality. It has been estimated that high levels of air pollution have resulted in the deaths of hundreds of thousands of people in Europe and the United States[86] [87]. There is a strong association between particulate matter and morbidity and mortality from respiratory and cardiovascular diseases. Long-term exposure to pollution is associated with an increased risk of developing lung cancer, stroke, heart disease and chronic obstructive pulmonary disease[152], [153], [154].

In 2012, the World Health Organization (WHO) [91] estimated that cardiovascular disease, respiratory disease and lung cancer accounted for 72% and 28% of deaths respectively. In accordance with pertinent studies[155], [156], [157], inhalation of elevated concentrations of $PM_{2.5}$ has been demonstrated to induce oxidative stress, cause inflammation and stimulate the autonomic nervous system. This potential damage can persist for decades, increasing the risk of cardiovascular disease and leading to premature death. Additionally, numerous studies[158], [159], [160], [161] have demonstrated that particulate matter exposure is associated with alterations in both systolic and diastolic blood pressure. Long-term exposure may result in the development of hypertension. Furthermore, elevated $PM_{2.5}$ levels have been associated with an increased risk of dyslipidaemia, particularly among the elderly and overweight individuals[162]. In experiments on mice, mice exposed to $PM_{2.5}$ exhibited increased blood vessel inflammation and increased macrophage inflammation, which subsequently led to the development of atherosclerosis in the mice[163]. relevant literature[164], [165], [166] also indicates that short-term exposure can cause

myocardial infarction and stroke, while long-term exposure can accelerate the progression of subclinical diseases and lead to premature death from cardiovascular disease. In the analysis, a 10μg/m^3 increase in $PM_{2.5}$ was associated with a 0.84% increase in mortality from cardiovascular causes[167].

The detrimental effects of air pollution on the respiratory system are considerable. For example, Liu et al.[168] examined the relationship between asthma and the concentration of various pollutants in Wuhan, Hubei, China. Their findings indicated that short-term exposure to $PM_{2.5}$, NO_2 and O_3 was significantly associated with asthma mortality. In addition, chronic obstructive pulmonary disease (COPD) is also significantly impacted by air pollution. The World Health Organization (WHO) [169] has estimated that 8% of global COPD deaths in 2012 were attributable to air pollution. Xu et al. [170]conducted an impact assessment on the relationship between daily changes in $PM_{2.5}$ and cause-specific mortality in Beijing. The findings indicated that for each 10μg/m^3 increase in $PM_{2.5}$, the COPD mortality rate would increase by 0.96%. Furthermore, the impact of particle size on COPD mortality is also a factor. The results demonstrated that the impact of $PM_{2.5}$ on COPD mortality varies depending on the particle size. A survey conducted in Shanghai revealed a strong correlation between the mortality rate from COPD and the size of particulate matter. This correlation was found to increase as the particle size decreased[171]. In recent years, O_3 has also emerged as a significant factor influencing COPD mortality. A study conducted in the United States has demonstrated that long-term exposure to O_3 significantly increases the risk of death from COPD[172]. A review by the Scientific Committee of the Global Initiative for Chronic Obstructive Pulmonary Disease (GOLD) [173] found that both gaseous and particulate components of air pollution may contribute to COPD. According to GOLD, it is not possible to identify an absolutely safe level of ambient air pollution. Furthermore, the relationship between air pollution levels and respiratory events is super-linear. Some instances of severe air pollution have been observed to significantly elevate the risk of morbidity and mortality in patients with COPD, particularly in

children whose lungs are still developing. This may result in the onset of developmental COPD.

In order to facilitate the quantification of the impact of air pollution on health, the WHO has developed a software tool, AirQ+. The software can be employed to assess the impact of diverse pollutants on health in any urban area. This encompasses the attributable proportion (AP) and relative risk (RR) of overall mortality and specific diseases, such as cardiovascular and respiratory conditions [174]. In a study by Muhammad et al. [150], the AirQ+ software was employed to assess the impact of O_3, NO_2, and SO_2 on the prevalence of COPD in the atmosphere. The main findings are as follows.

Table 2-4: Attributive proportion (AP) values of air pollutants for different human health disease in the $AirQ_{2.2.3}$ model (**Bold** represents larger values)[18].

Estimated AP(CI:95%) (%)	$PM_{2.5}$	PM_{10}	SO_2	CO	NO_2	O_3
Total	1.87(1.53-2.21)	3.12(2.74-3.40)	1.11(0.98-1.25)	1.40(1.09-1.70)	4.19(3.85-4.53)	2.52(2.01-3.03)
Respiratory	2.50(1.48-3.54)	3.12(2.26-3.87)	1.61(1.14-2.06)	-	5.19(1.25-6.07)	3.78(2.57-4.96)
Cardiovascular	2.16(1.63-2.64)	4.14(3.59-4.70)	1.17(0.96-1.37)	1.79(1.34-2.25)	4.68(4.09-5.23)	2.36(1.54-3.13)

The results demonstrate that short-term exposure to O_3, NO_2, and SO_2 exceeds 10, with the AP of COPD in one year being 100,000. The prevalence of COPD among residents

was found to be 1.96%, 0.74% and 0.52%.

Xu et al.[18] employed the AirQ+ to assess the impact of air pollutants on public health in North China. The study quantified the mortality rates attributable to each pollutant. The Table 2-4 illustrates the attributable proportion (AP) value of each pollutant to different human diseases. The results indicate that PM_{10} has a significant negative impact on human health, with an excessive risk (ER) that is higher than that of other pollutants. It is evident that other pollutants, namely SO_2, NO_2, CO, and O_3 continue to exert a significant influence on human health at lower concentrations.

2.5.5 Changes of air pollution in the Beijing-Tianjin-Hebei region in recent years

The Beijing-Tianjin-Hebei region has experienced a high prevalence of severely polluted weather in recent years. In 2013, more than 80 million people in northern cities were affected by severe $PM_{2.5}$ pollution, resulting in the closure of highways, flight delays and health threats[175]. In light of this, China has devised a comprehensive plan to regulate air pollutant emissions, with the objective of mitigating the deterioration of the atmospheric environment. In 2013, the Chinese government released an action plan for air pollution prevention and control[176], which was followed by a series of subsequent measures designed to control pollution. These measures have significantly improved the air pollution situation[177].

Deng et al.[178] conducted a study on the changes in $PM_{2.5}$ concentrations in the Beijing-Tianjin-Hebei region from 2015 to 2020. The results illustrates that the $PM_{2.5}$ concentration in the Beijing-Tianjin-Hebei region decreased from 77 $\mu g/m^3$ in 2015 to 45 $\mu g/m^3$ in 2020, representing a reduction of over 41%. The concentration of $PM_{2.5}$

exhibits pronounced seasonal fluctuations, with elevated levels observed during the winter months, reduced levels during the summer months, and a gradual decline from spring to autumn. In comparison to 2015, the concentration of $PM_{2.5}$ exhibited a 48% reduction in spring, a 45% reduction in summer, a 38% reduction in autumn, and a 39% reduction in winter, respectively. From a monthly perspective, the monthly declines exceeded 12%. This indicates that a series of policies designed to prevent air pollution are playing an important role.

The relocation of highly polluting enterprises from Beijing to Hebei has resulted in an increase in pollution levels in Hebei, which are now higher than those in Beijing. Results indicated that between 2013 and 2015, while the air quality in Hebei was consistently worse than that in Beijing, it exhibited a discernible improvement trend under the implementation of rigorous air pollution control measures. However, the proximity of Beijing and Hebei means that Beijing's air quality will continue to be affected by Hebei. A 100 $\mu g/m^3$ increase in the $PM_{2.5}$ concentration in an area 100 kilometres away from Beijing will result in a 10.97$\mu g/m^3$ increase in the $PM_{2.5}$ concentration in Beijing. The data indicate that the pollution in Beijing and the entire Beijing-Tianjin-Hebei region is closely related[179].

2.5.6 Summary and analysis of regional pollution analysis and health-related impacts

The above review presents the findings of research and analysis conducted by various researchers in recent years on the changes in air pollution levels in different regions. It demonstrates that air pollution levels have been reduced as a result of the implementation of relevant control measures. A gradual downward trend in pollutants

has been observed in various locations, providing valuable information for future pollution control. In addition, as the global spread of the COVID-19 pandemic has unfolded, numerous researchers have conducted extensive research on changes in air pollution that have occurred during this unique period. The findings indicate that reducing human activities can significantly reduce air pollution, offering a unique opportunity for future research. Moreover, the software AirQ+, developed by the WHO, is of great significance in quantifying the impact of air pollution on human health and in displaying the harm of pollution to health in a more intuitive manner. Therefore, numerous researchers employ this software to assess the detrimental impact of air pollution on human health across various periods, thereby facilitating the safeguarding of human health and the reduction of medical costs. Nevertheless, there is a paucity of studies in the later stages of the COVID-19 pandemic, and there is a dearth of research on the changes in air pollution following the end of the COVID-19 lockdown and the return to normalcy in human life. This will facilitate a more comprehensive understanding of the impact of reducing human activities on air pollution. In conjunction with the specific risks to human health, the impact of air pollution on human health before and after the lockdown can be quantified.

2.6 Summary of the literature review

The aforementioned literature offers a comprehensive overview of the current state of multi-index air pollution prediction, advancements in improving air pollution prediction accuracy, the effects of the COVID-19 pandemic on air quality, and the health impacts of air pollution.

In the context of multi-index air pollution prediction, the current state of the art incorporates two key indicators: AAQI, which reflects the comprehensive impact of air

pollution, and HAQI, which specifically accounts for its effects on human health. These indicators offer a more comprehensive representation of the true impact of air pollution. However, there is currently no evidence of the existence of multiple indicators. Consequently, future research should aim to develop a simultaneous prediction of multiple pollution indicators, which would assist sensitive groups in comprehending the comprehensive impact of air pollution on human health and in preventing it.

In addition, the rapid development of artificial neural networks has greatly assisted in enhancing the efficacy of air pollution prediction. In particular, the incorporation of the GCN model, which extracts spatial information, has markedly enhanced the predictive capacity of air pollution models. However, the incorporation of spatial information as additional parameters introduces a more complex data collection process, necessitating additional collection work and introducing inconvenience. Therefore, if the collection of spatial information could be simplified or even omitted through the implementation of certain methodologies, this would result in a simpler, universal, and highly performant air pollution prediction method for air pollution prediction. The provision of more accurate data will prevention and control of air pollution.

Regarding the impact of the COVID-19 pandemic on air pollution, numerous studies have documented the changes in air pollution levels during lockdown periods and their corresponding health impacts. However, there is a lack of data on air pollution trends following the conclusion of pandemic lockdowns, as well as on the associated health risks. By analyzing post-lockdown changes in air pollution and related health hazards, a more comprehensive evaluation can be achieved, offering insights into the actual improvements in air quality and reductions in health risks brought about by policy interventions and other efforts.

Chapter 3. Methodology

3.1 Introduction

The aim of this study is to realise the enhancement of air pollution in terms of prediction and how air pollution changes under the influence of air pollution control and COVID-19. Achieving these aims will help people to prevent and manage air pollution concerns. On this basis, this study extends three main research directions: 1. To achieve the prediction of several different AQIs(AQI, AAQI, HAQI) through neural network models; 2. To add spatial information as a training parameter to the neural network prediction model to improve the prediction of AQIs for six major pollutants; 3. To add the data of the air quality changes after the end of the control and regulation of COVID-19 (after the end of 2022), and to analyse the comprehensively Changes in air quality and improvements in health risks associated with air pollution in 13 cities in the Beijing-Tianjin-Hebei region since the beginning of control in 2014.

In order to achieve the above aims and directions, this chapter describes the data sources, data analysis methods, predictive modelling structures and methods, etc., required to achieve these.

The chapter is organized as follows. Section 3.2 describes the research area. Section 3.3 describes the sources of the data required for the correspondence study. Section 3.4 describes the calculation of the three air pollution indices AQI, AAQI and HAQI. Section 3.5 describes the principles of the neural network model used for the study. Section 3.6 describes the software and tools used for research. Section 3.7 summarizes the chapter.

3.2 Research area

3.2.1 Research area for chapters 4, 5

Beijing, the capital of China with a permanent population of over 20 million and is located in the north of the North China Plain. The city centre is situated at 116°20' east longitude and 39°56' north latitude. The terrain is characterised by high elevations in the northwest and low in the southeast, with mountains surrounding the city on three sides. Such a terrain surrounded by mountains is prone to the accumulation of pollutants on the surface, which in turn affects wind direction and speed. The climate is classified as warm temperate zone semi-humid semi-arid monsoon climate[180]. In addition, the Beijing-Tianjin-Hebei region is recognised for experiencing some of the most severe ~~worst~~ air pollution in China. This is largely attributed to the region's rapid industrialization, dense population and complex emission sources, which collectively pose significant challenges to air quality [181]. Moreover, the Beijing-Tianjin-Hebei region is a megacity cluster comprising the capital, Beijing, the industrialized city of Tianjin and surrounding areas in Hebei province. Given the region’s high population density and economic activity, the combined effects of industrial activity, transport and residential emissions make it an important case study in air pollution. In terms of policy, the Chinese government has implemented a number of stringent air pollution control measures in the Beijing with the aim of addressing the significant pollution problem. The area has valuable implications as a testing ground for policy interventions on air pollution levels[182]. Furthermore, Beijing has a relatively complete air pollution detection system, which facilitates the acquisition of pertinent data on air pollution and meteorological conditions[183]. Beijing is a city with a large population and a

developed economy. It is also affected by serious air pollution in surrounding areas. Therefore, this study selects Beijing as the area for air pollution prediction.

3.2.2 Research area for chapter 6

The region chosen for this study is the Beijing-Tianjin-Hebei (BTH) region (Figure 3-1) of China, since the BTH region is one of the most heavily polluted areas in China[184]. In addition, the region contains the capital city of Beijing, mega-city Tianjin, which are two large cities with large populations, and the province of Hebei, which is one of the most heavily industrialised regions in China, with a large number of heavily polluting factories in this province and so on[185]. In addition, the region is also a key area in China's air pollution control. The region has both mega-cities with fewer heavy industries but large populations, as well as large, medium and small cities with a lot of heavy industries. Therefore, it is important to study the air pollution control and improvement in this region.

The Beijing-Tianjin-Hebei (BTH) region is the largest urban agglomeration in North China, with a total population of 110 million in 2020[186].The 13 cities include the capital city of Beijing, the municipality directly under the central government of Tianjin, and the 11 cities of Baoding, Cangzhou, Chengde, Handan, Hengshui, Langfang, Qinhuangdao, Tangshan, Shijiazhuang, Xingtai, and Zhangjiakou in Hebei province. The distribution of the cities in the BTH region is illustrated in Figure 3-1, with the two megacities of Beijing and Tianjin are adjacent to each other and are surrounded by Hebei Province.

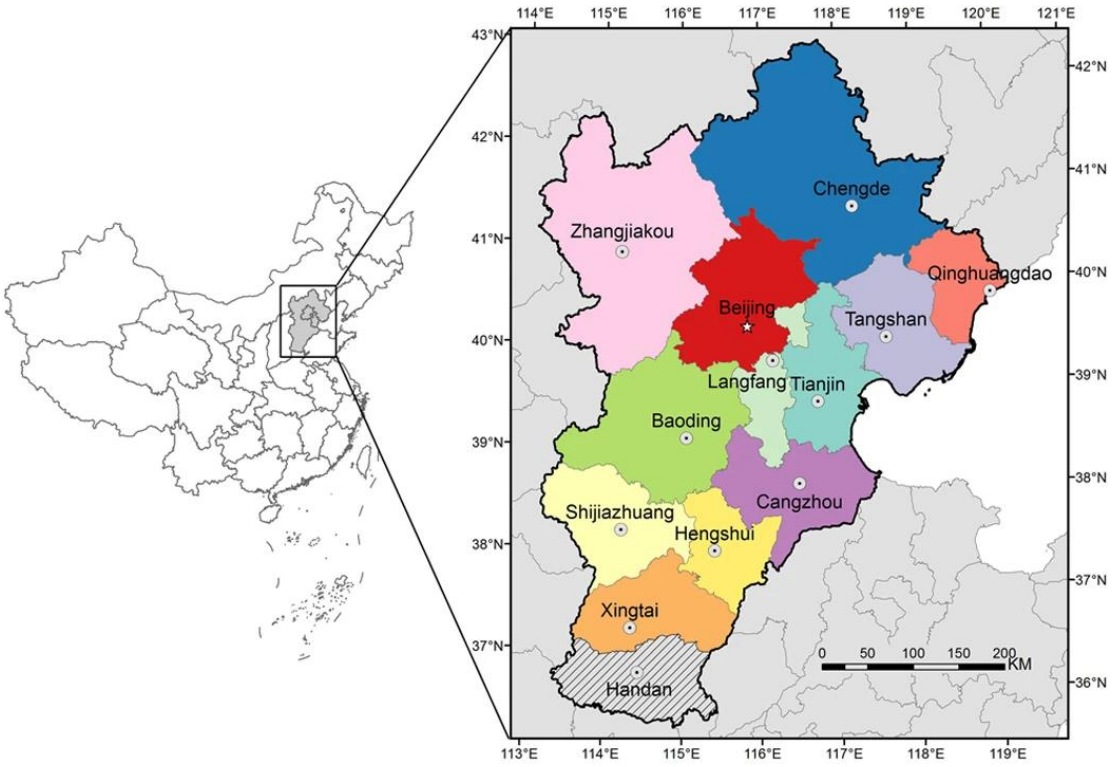


Figure 3-1: Geographic location of cities in the Beijing-Tianjin-Hebei region[187].

3.3 Sources of the data

3.3.1 Data sources for chapter 4, 5

The air pollution prediction models in Chapters 4 and 5 use the same data from the UCI machine learning repository[188], which encompasses hourly data for six pollutants and weather data from 12 monitoring stations, as illustrated in Figure 3-2. The data was collected in Beijing from March 2013 to February 2017 (the coordinates of the monitoring stations are provided in Table 3-2). The pollutants included in the study are $PM_{2.5}$, PM_{10}, CO, O_3, SO_2 and NO_2. Furthermore, the weather data encompass temperature, pressure, dew point, precipitation, and wind speed.

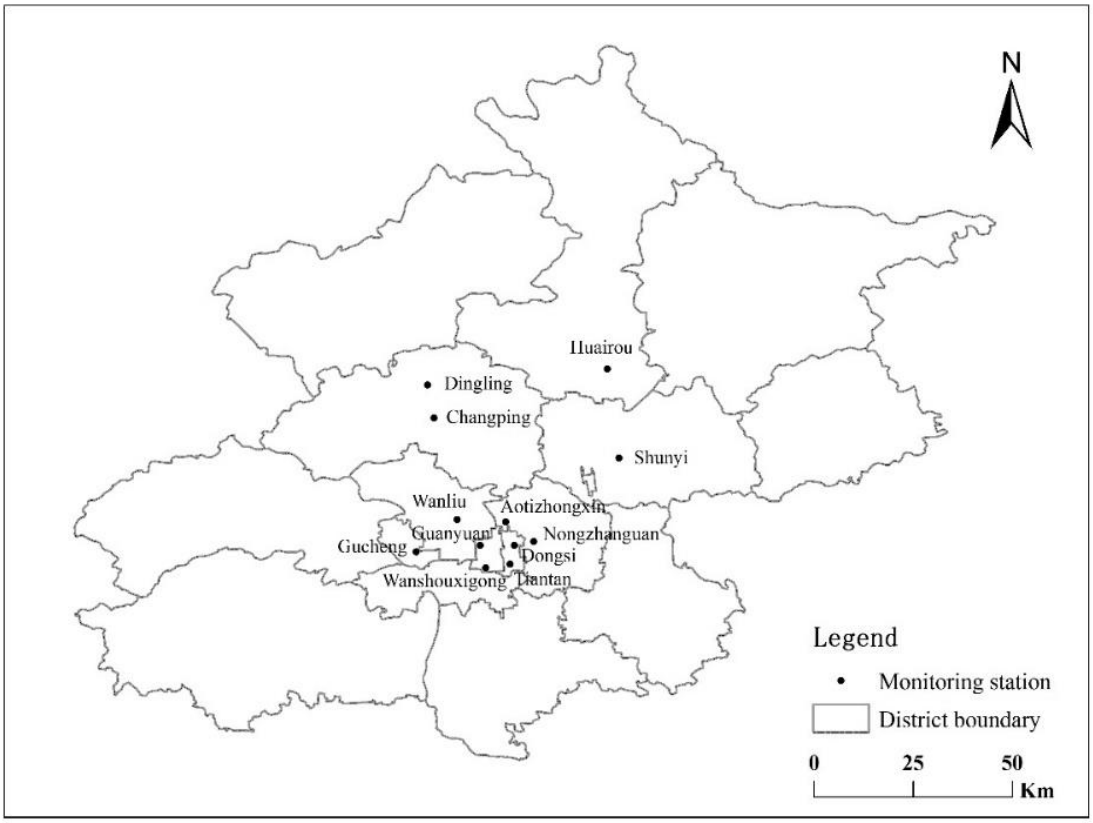


Figure 3-2: Location map of pollution monitoring stations in Beijing

Table 3-1 presents the unit, quantity of data, mean value, maximum and minimum statistics of each variable. As can be seen from the table, the average values of PM_{10} and $PM_{2.5}$ among the pollutants are the highest and the second highest respectively, with the highest values exceeding 500, reaching 584.67 and 567.42 respectively, which are the main pollutants. In addition, there are considerable discrepancies between the maximum and minimum values of each pollutant, indicating that each pollutant exhibits considerable fluctuations. These fluctuations may be influenced by various factors, including season and climate[189]. According to data and literature, air quality varies greatly in different seasons. In addition, meteorological conditions also affect air quality. For example, in windy or rainy weather, air pollution will be diluted and air quality will be relatively good. With regard to temperature, the average temperature difference is relatively modest, at 13.53℃. However, the highest and the lowest temperature

Table 3-1: Statistics of various parameters contained in the data (including major pollutants and meteorological data)

VARIABLES	UNIT	QUANTITY OF DATA	AVERAGE	MAX	MIN
$PM_{2.5}$	ug/m^3	17532	79.86631	567.4167	3
PM_{10}	ug/m^3	17532	104.7132	584.6667	4.83333
SO_2	ug/m^3	17532	15.85952	158.3182	0.714
NO_2	ug/m^3	17532	50.54861	201.2857	2
CO	mg/m^3	17532	1238.675	9109.091	100
O_3	ug/m^3	17532	57.31071	524.8696	0.2142
TEMPRATURE	℃	17532	13.53071	32.36667	-15.6833
PRESSURE	hPa	17532	1010.754	1041.025	985.7875
DEW POINT	℃	17532	2.481017	27.61667	-33.0208
RAIN	mm	17532	0.064448	11.10833	0
WIND SPEED	m/s	17532	1.73023	7.2375	0.17083

differences are notable, at nearly 47℃. With regard to the air pressure and dew point, there is a minimal difference between the highest and lowest values, and the air pressure is in a stable state. Furthermore, the rainfall was relatively low, averaging only

0.064mm, and the average wind speed was at a low level of 1.73 m/s.

Table 3-2 presents the geographical coordinates of 12 air pollution monitoring stations in Beijing (Due to the serious lack of data from two stations, Wanliu and Gucheng, only data from the other 10 stations were selected for chapter 4.). The locations of these stations are distributed across the city, providing a comprehensive overview of the city's overall pollution situation.

Table 3-2: Monitoring stations coordinates

Monitoring stations	Latitude	Longitude
Aotizhongxin	116.397	39.982
Changping	116.234	40.217
Dingling	116.22	40.292
Dongsi	116.417	39.929
Guanyuan	116.339	39.929
Gucheng	116.194	39.914
Huairou	116.628	40.328
Nongzhanguan	116.461	39.937
Shunyi	116.655	40.127
Tiantan	116.407	39.886
Wanliu	116.287	39.987
Wanshouxigong	116.352	39.878

3.3.2 Data sources for chapter 6

The data for the study of air pollution changes in the BTH area in Chapter 6 are from

https://aqicn.org/data-platform/register/ [190], which included data on six major pollutants in 13 cities in the Beijing-Tianjin-Hebei region. The data period is from 1 January 2014 to 19 July 2023, covering all stages of the Beijing-Tianjin-Hebei region from severe air pollution to gradual improvement, and it also includes complete data before the COVID-19, during the COVID-19 control period, and after the COVID-19 control is completely lifted (The control period for COVID-19 ranges from approximately January 2020 to December 2022). The six main pollutants are $PM_{2.5}$, PM_{10}, CO, O_3, SO_2 and NO_2. Given the extensive temporal coverage of the data set, a number of missing and anomalous values were identified. These were addressed during the data preprocessing stage, with outliers and evidently erroneous data points removed and missing values linearly interpolated.

3.4.1 Calculation of the three air pollution indices AQI, AAQI and HAQI

The Air Quality Index (AQI) is a common indicator used to assess the public health impacts of six criteria pollutants ($PM_{2.5}$, PM_{10}, CO, O_3, SO_2 and NO_2) The study was located in Beijing, China, and therefore the AQI was calculated with reference to the Chinese Ambient Air Quality Standards (CAAQS)[191]. The methods of calculating the three indices, namely the AQI, the AAQI, and the HAQI, are described separately in the following sections.

- AQI

The AQI reflects air quality by taking the maximum value of the AQI_m for all pollutants as the final AQI, and the formula for calculating the AQI_m for each pollutant and AQI are as follows:

$$\mathrm{AQI}_m = \frac{AQI_{m,n} - AQI_{m,n-1}}{(C_{m,n} - C_{m,n-1})} \times \left(C_{m,t} - C_{m,n-1}\right) + AQI_{m,n-1,}\, n > 1 \quad (3.1)$$

$$\mathrm{AQI}_m = AQI_{m,1} \frac{C_{m,t}}{C_{m,1}}, n = 1 \quad (3.2)$$

Where $C_{m,t}$ is the concentration of pollutant m, and n is the health category index. $C_{m,n}$ and $C_{m,n-1}$ are the upper limit concentrations of the n and n-1 health categories, including $C_{m,t}$. $AQI_{m,n}$ and $AQI_{m,n-1}$ are the AQI values corresponding to $C_{m,n}$ and $C_{m,n-1}$ in pollutant m, respectively.

$$\mathrm{AQI} = \max\left(AQI_1, AQI_2, \ldots, AQI_p,\right), p = 1,2,\cdots\cdots,6 \quad (3.3)$$

- AAQI

AAQI is calculated from the combined impact of multiple pollutants, and its calculation formula is as follows:

$$AAQI = \left(\sum_i^n (AQI_i)^\rho\right)^{\frac{1}{\rho}} \quad (3.4)$$

Where ρ is an empirical constant with an optimal value that remains an unsolved problem[45]. Previous research has shown that the value of ρ should be between 2 and 3. When Hu et al. used four different values (1.5, 2.0, 2.5, 3.0) to test the ratio of AAQI and AQI, they found that the change in the ratio was not obvious, and finally adopted 2 as the value[15]. Therefore, in the present study, the 2.0 value of ρ was used. For better comparison with the AQI, the range of AAQI values will be the same as the AQI, also 0-500.

- HAQI

HAQI is related to excess risk to health. To define AQI based on health risk, the total excess risk (ER) of exposure to multiple pollutants based on the Cairncross[14] concept can be used. In general, the relative risk (RR) for each pollutant is estimated from health effects studies using equations.

$$\mathrm{RR}_m = \mathrm{e}^{[\beta_m(C_m - C_{m,0})]}, C_m > C_{m,0} \quad (3.5)$$

Where β_m is the exposure-response coefficient, which represents the additional health risk (such as mortality) per unit of pollutant increase (such as $PM_{2.5}$ of 1 μg/m^3). C_m is the pollutant concentration, $C_{m,0}$ is the threshold concentration, and when the pollutant concentration is lower than the threshold concentration, there is no adverse health effect (RR = 1).

The ER_m of each pollutant m is calculated as follows:

$$ER_m = RR_m - 1 \quad (3.6)$$

Total ER is calculated as follows:

$$ER_{total} = \sum_{m=1}^{n} ER_m = \sum_{m=1}^{n} (RR_m - 1) \quad (3.7)$$

Therefore, the higher the ER_{total}, the higher the risk of the reaction. The researchers converted ER_{total} into an index between 0 and 10 to represent the risk of air pollution to human health. This index cannot be directly compared with AAQI, HAQI and AQI because they have different scales, AQI and AAQI have scales from 0 to 500. Therefore, the equivalent concentration C_m^* of a standard pollutant is defined as the concentration level when the ER of the pollutant is equal to ER_{total}, such as the following formula:

$$RR_m^* = ER_{total} + 1 = \mathrm{e}^{[\beta_m(C_m^* - C_{m,0})]} \quad (3.8)$$

$$C_m^* = \frac{\ln(RR_m^*)}{\beta_m} + C_{m,0} \quad (3.9)$$

Where β_m and $C_{m,0}$ are the β_m and $C_{m,0}$ values of the ith pollutant, respectively. Because RR_m^* is the total risk of exposure to multiple pollutants, the equivalent concentration represents the comprehensive health risk of all standard pollutants, and then C_m^* is used to determine the HAQI based on the equivalent concentration of standard pollutant m. Equation :

$$\mathrm{HAQI}_m = \frac{AQI_{m,n} - AQI_{m,n-1}}{(C_{m,n} - C_{m,n-1})} \times (C_m^* - C_{m,n-1}) + AQI_{m,n-1}, n > 1 \quad (3.10)$$

$$\mathrm{HAQI}_m = AQI_{m,1}\frac{C_m^*}{C_{m,1}},\ n = 1 \tag{3.11}$$

The composite HAQI is determined as the maximum of all sub-HAQIs, as shown in Equation 3.12:

$$\mathrm{HAQI} = \max\left(HAQI_1, HAQI_2, \dots, HAQI_p,\right), p = 1,2,\cdots\cdots,6 \tag{3.12}$$

3.4.2 Physical interpretation of AQI, AAQI and HAQI

Before presenting the model architecture, it is necessary to clarify the physical meaning and formulation logic of the AQI, AAQI, and HAQI. These indices differ not only in their mathematical expressions but also in the aspects of air pollution they are designed to capture. The AQI reflects the severity of the dominant pollutant, the AAQI represents the overall pollution burden across multiple pollutants, and the HAQI quantifies the associated health risks using exposure–response relationships. Understanding these distinctions is essential because the model aims to predict all three indices simultaneously, and their differing sensitivities to pollutant concentrations directly influence feature selection, model design, and performance evaluation. The following section therefore provides a concise explanation of each index and the rationale behind its specific formula and range.

The AQI provides a standardized representation of ambient air quality by converting pollutant concentrations into a unified scale. Physically, it reflects the relative severity of pollution levels and their potential impacts on human health. Moreover, the AQI formula is based on a piecewise linear transformation, where each pollutant concentration is mapped to a corresponding sub-index according to predefined breakpoints. These breakpoints are set using epidemiological evidence on health risks associated with different pollution levels. Furthermore, the overall AQI is defined as the maximum of the individual pollutant sub-indices, highlighting the principle that the

most harmful pollutant determines the overall health risk. Its range (commonly 0–301+) provides a clear risk categorization from “Good” to “Hazardous,” allowing the public and regulators to interpret air quality conditions intuitively[48], [116].

The AAQI is designed to capture the cumulative effects of multiple pollutants. Its formula typically involves a power-mean or root-mean-square aggregation, ensuring that pollutants with higher sub-indices contribute more strongly to the final value while still retaining information about intermediate pollutants. This approach acknowledges that exposure to a mixture of pollutants may pose health risks that cannot be fully captured by a single “dominant” pollutant. The resulting index reflects the overall air pollution load, not merely the most extreme component. Its value range generally aligns with that of AQI for comparability, but its magnitude better represents multi-pollutant episodes[53], [56].

The HAQI extends beyond pollutant concentration levels by incorporating exposure–response (E–R) functions derived from epidemiological studies. Physically, the HAQI quantifies the estimated health risk attributable to current pollution levels, often expressed using the relative risk (RR) and excess risk (ER) of mortality or morbidity for each pollutant. The HAQI formula converts the accumulated health risks into a standardized index that resembles the AQI scale, enabling direct comparison with traditional indices. Unlike AQI or AAQI, the HAQI reflects actual health outcomes rather than merely concentration thresholds[56].

In general, all three air pollution indices use standardized scales (consistent with AQI categories) to ensure that policymakers and the public can understand the risk levels. And AQI uses concentration-response breakpoints, AAQI integrates multiple pollutants to reflect mixture effects and HAQI directly quantifies epidemiological risk, bridging atmospheric science and public health.

3.5 Models and indices for air pollution prediction

3.5.1 Models for multi-indexes prediction

Compared with traditional single-index prediction, forecasting multiple indices is inherently more complex and therefore requires higher model stability. For this reason, LSTM—one of the most widely used and high-performing architectures for time-series prediction—is adopted as the baseline model. To enhance the model's applicability across diverse datasets, an MLP and a Transformer are incorporated to compensate for LSTM's limitations: the MLP improves efficiency on simpler datasets, while the Transformer strengthens performance when handling more complex data patterns. The resulting composite model, consisting of LSTM, MLP, and Transformer, integrates the strengths of all three architectures, yielding a more stable and generalizable framework for multi-indices air quality prediction. Moreover, compared with alternative architectures such as CNN-based or statistical models, the LSTM-MLP-Transformer combination offers a balanced trade-off between temporal modelling strength, nonlinear approximation capability, computational efficiency and adaptability to different index characteristics.

This model also has strong generalizability to other geographical regions. None of its components depend on region-specific assumptions; they are entirely data driven. By retraining the model with local pollutant data, meteorological inputs, and index values, the architecture can adapt to different climates, emission profiles, and pollution compositions. Furthermore, the model adds MLP and Transformer to LSTM to improve its adaptability, enabling it to adapt well to both simple and complex environmental data.

The structure and characteristics of the three neural network models LSTM, MLP and Transformer are described below.

- Long short-term memory network (LSTM)

LSTM is a neural network prediction model that has demonstrated strong performance in time series modelling and prediction. It is characterised by three main components.

The three types of LSTM gating units are the forgetting gate 'f', the input gate 'i' and the output gate 'o' as shown in Figure 3-3.

The forgetting gate is used to control the information to be discarded in the current cell state C_t. The state update formula of the forgetting gate is as follows:

$$f_t = \sigma(W_f[h_{t+1}, x_t] + b_f) \tag{3.13}$$

The input gate controls how much of the current network input x_t will be retained in the unit state C_t. The state update rules of the input gate are as follows:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i); \tag{3.14}$$

$$\widetilde{C_t} = \tanh(W_c[h_{t-1}, x_t] + b_c); \tag{3.15}$$

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C_t} \tag{3.16}$$

The output gate regulates the quantity of information that can be exported from the current state of the cell. The update rules of the output gate are given by:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o); \tag{3.17}$$

$$h_t = o_t \times \tanh(C_t) \tag{3.18}$$

Where, f_t is the output of the forget gate; σ represents the sigmoid function, which will generate the vector between [0,1] based on the input. $\widetilde{C_t}$ describes candidate cell

information; C_t is the updated cell state at the current time step t; C_{t-1} is the cell state from the previous time step; h_t represents the hidden state at time step t; x_t is the input at time step t and $tan\ h$ the hyperbolic tangent activation function. W_f, W_i, W_0, W_c represent the weight coefficient matrix in the LSTM cell state update process while b_f, b_i, b_0, b_c are the bias matrix in the status update process [102], [192].

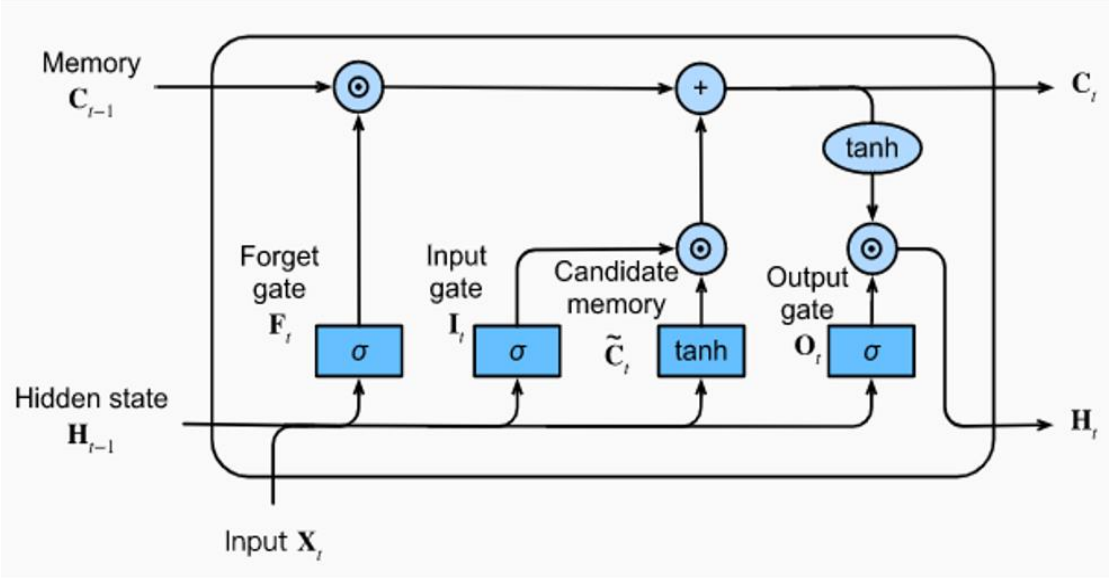


Figure 3-3: The structure of the LSTM model that is composed of a Forget Gate module, an Input Gate module and an Output Gate module[193].

The structure of LSTM is depicted in Figure 3-3. The fundamental operational principle of LSTM is to regulate the transmission state through gate control to retain important information and forget unimportant information. This enables the data to be trained in a more flexible manner.

- Multilayer perceptron (MLP)

Multilayer perceptron is a deep learning model based on feedforward neural network,

which consists of multiple neuron layers, each of which is fully connected to the previous layer. Multilayer perceptron can be used to solve various machine learning problems such as classification, regression and clustering. Figure 3-4 shows the MLP structure.

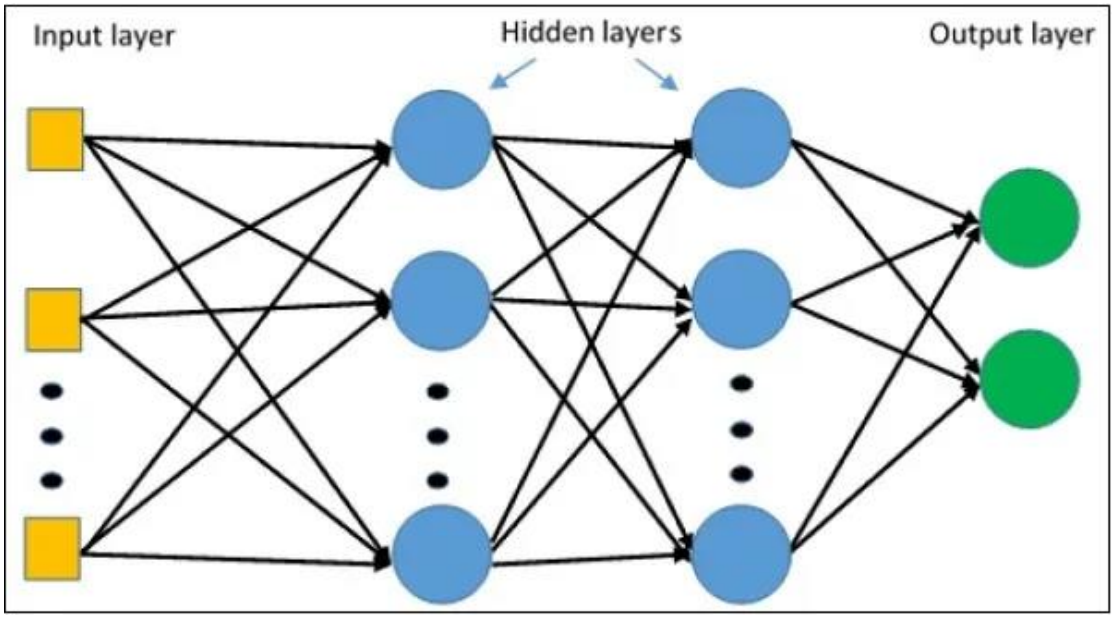


Figure 3-4: The structure of the MLP model that is composed of an input layer module, hidden layer module and an Output layer module[196]

Each neuron layer of multilayer perceptron consists of many neurons, where the input layer receives input features, the output layer gives the final prediction results, and the middle-hidden layer is used to extract features and perform nonlinear transformations. Each neuron receives the output of the previous layer, performs weighting and activation function operations, and obtains the output of the current layer. Through continuous iterative training, multilayer perceptron can automatically learn the complex relationship between input features and predict new data[194].

In addition, MLP has many advantages. MLP has strong expression ability and can handle nonlinear problems and high-dimensional data. It can be trained by backpropagation algorithm to automatically learn features and patterns. MLP also has

good generalization ability[195].

- Transformer

The Transformer model is a deep learning model that uses an attention mechanism, which assigns different weights to different parts of the input data according to their importance. It overcomes the limitations of traditional recurrent neural networks through self-attention (Self-attention allows the Transformer model to capture complex relationships between elements in a sequence, weigh their importance, and process them in parallel, making it highly efficient and capable of handling long-range dependencies) and parallel computing. Figure 3-5 shows the Transformer structure.

The transformer consists of an encoder and a decoder, each of which contains multiple identical layers. The hierarchical structure of the encoder and decoder is as follows:

- Encoder:
  - Multi-Head Self-Attention Mechanism
  - Feed- Forward Neural Network (FFN)
  - Residual Connection and Layer Normalization
- Decoder
  - Multi-Head Self-Attention Mechanism
  - Encoder-Decoder Attention
  - Feed- Forward Neural Network
  - Residual Connection and Layer Normalization

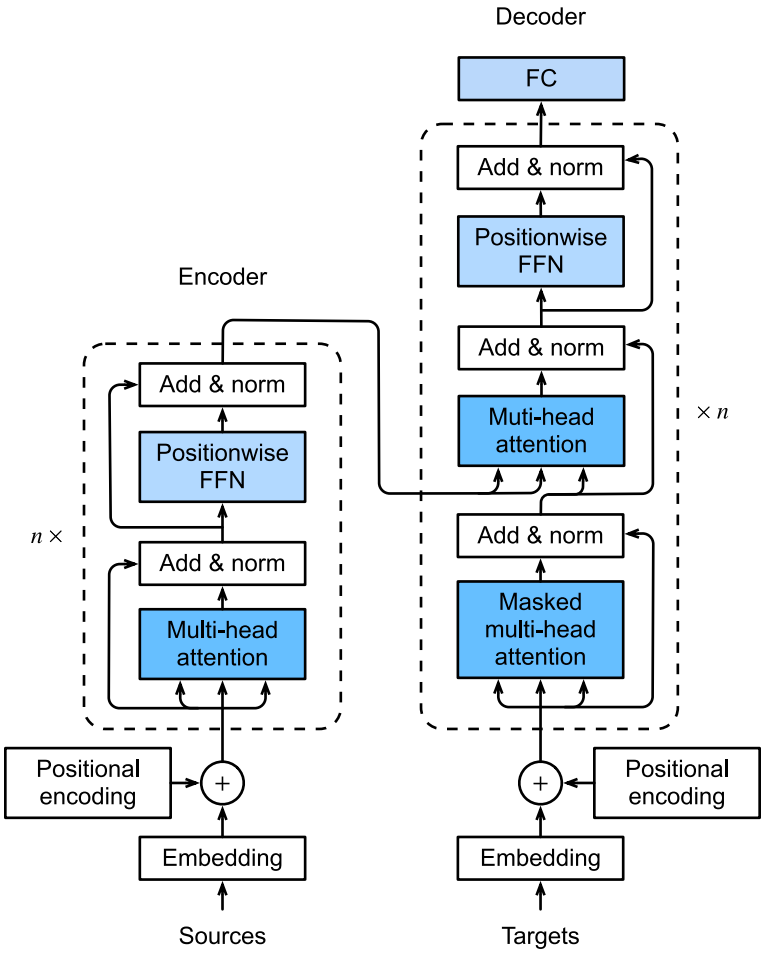


Figure 3-5: The structure of the Transformer model that is composed of an input layer module, an Encoder layer module, an Decoder layer module and an Output layer module[197]

The self-attention mechanism is the core of the transformer, allowing the model to focus on different positions in the input sequence when processing the input at each position. This mechanism enables the model to capture long-distance dependencies. The specific steps include:

Calculate the Query (Q), Key (K), Value (V) vector:

$$Q = XW_Q,\ K = XW_K,\ V = XW_V \tag{3.19}$$

Where X is the input sequence and W_Q, W_K, W_V are learnable weight matrices.

Calculate the attention score:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}}) \quad (3.20)$$

Where d_k is the dimension of the key vector.

The multi-head attention mechanism divides attention into h heads, and each head is calculated by equation 3.21. The outputs of all heads need to be connected to get the final value.

The multi-head attention calculation is as follows:

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W_O \quad (3.21)$$

Where $head_i = Attention(QW_{Q_i}, KW_{K_i}, VW_{V_i})$ (3.22)

Each encoder and decoder layer also contains a feed-forward neural network (FFN) applied to the output at each position.

The FFN calculation formula is as follows:

$$FFN(x) = \max(0, xW_1 + b_1)\, W_2 + b_2 \quad (3.23)$$

In addition, the residual connection allows the input and output of each sub-layer to be added together, alleviating the gradient vanishing problem of deep networks. Then the layer is normalized to ensure that the input distribution is stable.

The calculation is as follows:

$$Output = LayerNorm(x + Sublayer(x)) \quad (3.24)$$

Since the Transformer does not use recurrence or convolution, the model has no sequential information, therefore, Positional Encoding is needed to introduce position information into the input sequence:

$$PE(pos, 2i) = \sin(pos/10000^{2i/d}) \quad (3.25)$$

$$PE(pos, 2i+1) = \cos\left(pos/10000^{2i/d}\right) \quad (3.26)$$

Where pos is the position, i is the dimension index, and d is the model dimension.

Through these basic principles, Transformer can efficiently process sequence data, capture long-distance dependencies, perform parallel computing, and improve training speed and effectiveness[198], [199].

3.5.2 Neural network models used in Chapter 5

The air pollution prediction analysis in Chapter 5 primarily aims to improve forecasting performance by incorporating spatial parameters. Accordingly, LSTM is selected as the base model due to its strong ability to process time-series data, while GCN—with its powerful capacity for extracting graph-structured spatial relationships—is integrated to supplement the temporal model and enhance predictive accuracy. To rigorously evaluate the effectiveness of this GCN–LSTM architecture, it is compared against several widely used benchmark models. The standard LSTM serves as the fundamental baseline, enabling verification that the inclusion of spatial information provides added value beyond temporal modelling alone. The LSTM–Attention model is included because attention mechanisms strengthen long-range dependency learning; comparing against this variant helps determine whether spatial information contributes more effectively than temporal attention. Additionally, the GRU model is selected as a representative recurrent architecture with competitive performance and lower structural complexity, ensuring that any improvements are attributable to the introduction of spatial dependencies rather than the choice of recurrent unit. Together, these comparisons allow for a comprehensive and fair assessment of the predictive advantages offered by the proposed GCN–LSTM framework. Compared with alternative spatiotemporal models such as CNN-LSTM, ST-GCN, or Transformer-

based architectures, the GCN–LSTM remains computationally efficient and less demanding in terms of training data while still providing strong interpretability. CNN-based models assume regular grid structures, which are unsuitable for irregular monitoring networks, while Transformers require substantially larger datasets and offer limited advantages for hourly environmental data where long-range dependencies are already well modelled by LSTM units. The GCN–LSTM thus offers a balanced trade-off between model complexity, computational cost, and predictive performance.

The model also demonstrates strong generalization capability and can be readily extended to other geographical regions. As long as historical data from monitoring stations are available, spatial dependencies can be extracted using the GCN module, while the LSTM component can be retrained on the corresponding historical time series. This modular design ensures that the model is not constrained by the characteristics of any specific city or region. Given the presence of relevant spatial relationships and time-series observations, the GCN–LSTM framework can adapt to varying climates, emission patterns, and monitoring network configurations, making it a robust and scalable approach for air quality forecasting.

The LSTM model has been introduced in Chapter 3 section (3.2.4), therefore the GCN, GRU and Attention models will be introduced in the following sections.

- Graph Convolutional Network (GCN) model's theory and structure

GCN is frequently employed to extract features and information from graph networks. The fundamental principle underlying this process is now outlined. For a graph G= (V, E, A), an input signal X and an output signal Y, the processing method f adopted by a graph convolutional neural network is defined as:

$$Y = f(X, A) \tag{3.27}$$

Where, V is the number of nodes in the graph, $V = \{v_i\}_{i=1}^{N}$, and E represents the set of

edges. A is the adjacency matrix of the graph, A∈ R^{N*N}, and the element A_{ij} in matrix A represents the connection relationship between nodes V_i and V_j in graph G. The forward propagation formula of graph convolution is:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{l}W^{l}) \tag{3.28}$$

Where, $\tilde{A}$=A+I, I is an identified matrix of size N*N; $\tilde{D}$ is a diagonal matrix, $\tilde{D}_{ii}=\sum_j \tilde{A}_{ij}$; $H^l \in R^{N*D}$, representing the output value of layer L , where H^0=X; $\sigma(.)$ stands for activation function; W^l indicates the parameter values of layer L [200].

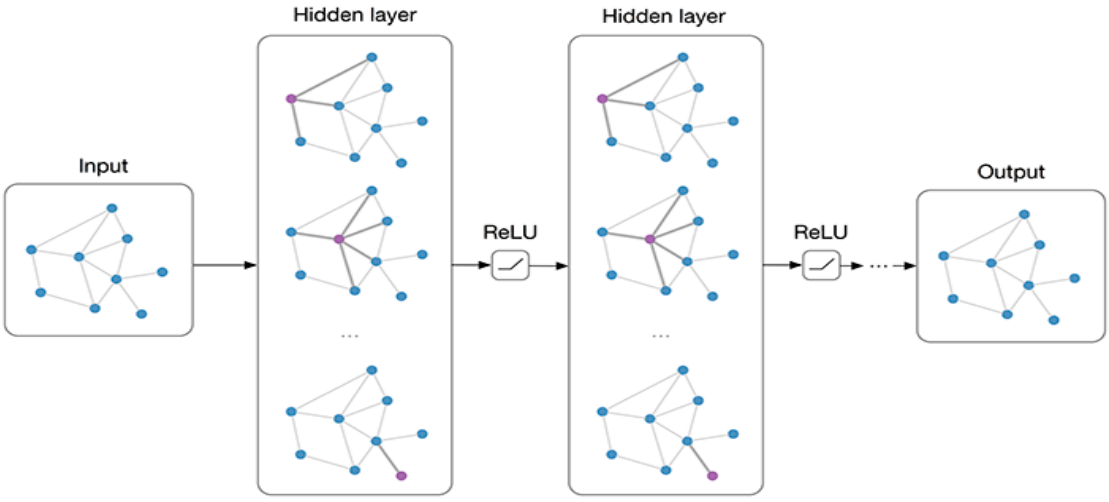


Figure 3-6: Structure of a GCN model that is composed of an Input module, a Hidden layer and an Output module [201].

GCN acts as a feature extractor that can derive characteristics from graph data. These characteristics are subsequently used for tasks such as node classification, graph classification, and link prediction. Figure 3-6 shows the basic structure of the GCN model. In this study, GCN will be used to extract the node information from the abstract image formed by the correlation between variables.

- Gated Recurrent Unit (GRU) model's theory and structure

GRU is a variant of recurrent neural network (RNN) for sequence data processing, proposed by Cho et al. in 2014[202]. Compared to traditional RNNs, GRU performs better in processing long sequence data and can solve the long-term dependency problem more effectively. Compared to LSTM, GRU is simplified by introducing only update gates and reset gates. In GRU, the update gate (or input gate) determines how much of the input and previously ground output to pass to the next unit, while the reset gate is used to determine how much past information to forget. The current memory content ensures that only relevant information needs to be passed to the next iteration, which is determined by the weights W. The structure of the GRU is shown in Figure 3-7. The main operations in GRU are controlled by the following equations.

Update gate:

$$\mathcal{Z}_t = \sigma(W_Z * [h_{t-1}, x_t]) \tag{3.29}$$

Where, $\mathcal{Z}_t$ is the output of the updaters, σ is the sigmoid activation function, W_Z is the weight matrix, h_{t-1} is the hidden state at the previous moment, and x_t is the input at the current moment.

Reset gate:

$$r_t = \sigma(W_r * [h_{t-1}, x_t]) \tag{3.30}$$

Where, r_t is the output of the reset gate and W_r is the weight matrix.

Candidate hidden state:

$$\tilde{h}_t = \tanh\,(W * [r_t * h_{t-1}, x_t]) \tag{3.31}$$

Where, $\tilde{h}_t$ is the candidate hidden state, tanh is the tanh activation function, *denotes element-by-element multiplication, and W is the weight matrix.

Hide status updates:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{3.32}$$

GRU model has a simpler structure, high computational efficiency, and outperforms LSTM on some tasks. it has a wide range of applications in natural language processing, time series prediction, and other fields.

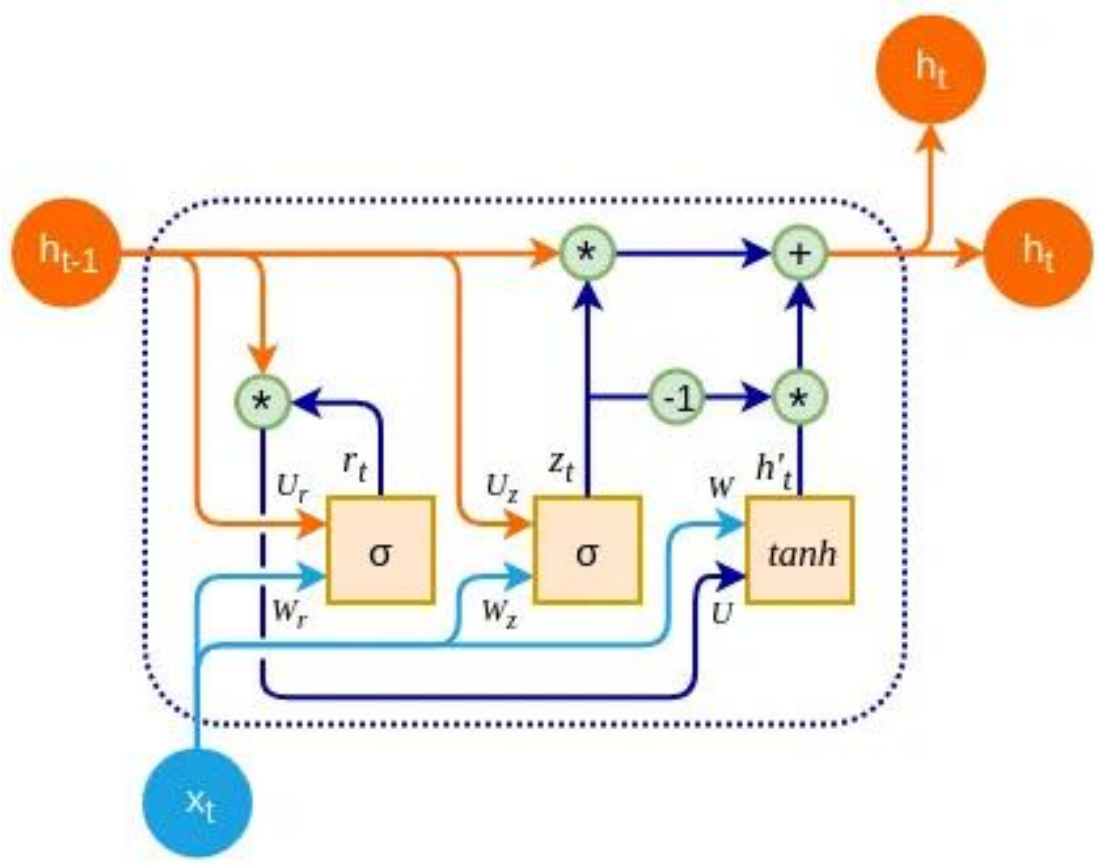


Figure 3-7: Structure of a GRU model that is composed of an Update gate, a Reset gate[203].

- Improvement of LSTM model by Attention mechanism

The Attention mechanism is a powerful technique in the field of deep learning, originally introduced by Bahdanau et al. in 2014 [204] to improve machine translation models. Attention weight is used to measure the correlation between Query and Key, and the commonly used calculation method is Scaled Dot-Product Attention. The

calculation formula is as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.33}$$

Where Q is the query vector, which represents the information that needs to be focused on at the current moment, K is the key vector, which represents the features of each input element, and V is the value vector, which represents the actual content of each input element. d_k is the vector dimension of the Key, and $\frac{1}{\sqrt{d_k}}$ is the scaling factor, which is used to prevent the over-sizing of the dot product result.

The integration of the attention mechanism with LSTM has been demonstrated to markedly enhance the efficacy of sequence modelling tasks, particularly in the context of processing extensive sequential data. While LSTM can address long-term dependencies, when confronted with exceedingly lengthy sequences, the model's attention may remain concentrated on the most proximate time step, resulting in the neglect of information situated at the far end of the range. The attention mechanism is therefore capable of addressing this issue, providing a more detailed level of control and consequently enhancing the performance of the model[205].

3.5.3 Evaluation indexes

After the model predicts air pollution, some indicators are needed to evaluate whether the model's performance meets the standard. Therefore, we use the indicators Mean Absolute Error (MAE), Mean Squared Error (MSE) and Coefficient of Determination (R^2), which are often used in many air pollution prediction studies[206], [207], [208], [209], to evaluate the prediction performance.

MAE is a statistical measure that provides a more accurate representation of the discrepancy between predicted and actual values. It is calculated as the average of the

absolute errors between the predicted and actual data points. MSE is a metric that calculates the average of the sum of squared errors between the predicted and original data points. Both MAE and MSE are inversely proportional to the accuracy of the prediction. R^2 is a statistical measure that assesses the degree of fit between the predicted and actual values. The closer R^2 is to 1, the higher the degree of fit. 0.7-1 is usually considered to be a good explanatory power of the model; 0.4-0.7 indicates a moderate explanatory power of the model, which may need to be improved, but in complex areas this may be sufficient; and below 0.3-0.4 indicates a weak explanatory power of the model, which may be substandard[210]. The formulae for these evaluation indices are given below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\widehat{y_i} - y_i|; \tag{3.34}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2; \tag{3.35}$$

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} \tag{3.36}$$

Where, $\widehat{y_i}$ and y_i are the actual and predicted values, respectively. $SS_{residual}$ is the residual sum of squares and SS_{total}, the total sum of squares [211], [212].

3.6 The software and tools used for the research

3.6.1 The software and tools used for air pollution predictions

The experiments of air pollution prediction in chapter 4 and 5 were conducted by Python (version 3.6) where packages such as Pandas, Numpy, Torch, Sklearn, Matplotlib, Math, Pdb were mainly used to complete the prediction of multiple indices. These packages provide the necessary processes such as computational analysis and

image generation necessary for the experiments.

3.6.2 The software and tools used for air pollution changes

The research in chapter 6 examines a large area and a wide period of data, and in order to analyse the changes in air pollution in the BTH area in a more comprehensive way, a variety of methods such as statistical analysis, time trend analysis, spatial analysis, and health risk assessment have been used.

3.6.2.1 Methods of analysis

- Descriptive statistical analysis

In the process of data analysis, descriptive statistical analysis represents a crucial means of understanding and interpreting data. Descriptive statistical analysis enables the understanding of the concentration trend, degree of dispersion, shape of distribution and other information pertaining to the data, thereby providing the basis for further data analysis and modelling[213].

To address the complexity of air pollution changes, this study provides a comprehensive analysis of the frequency of major air pollutants, the number of times air quality standards are met, the annual quarterly changes in pollutants, and the health risks with changes in air quality.

- Time trend analysis and data visualisation

In order to analyse the time trend of pollution changes, discounted graphs and bar charts were plotted based on the statistical data. In addition, annual and quarterly maps of

pollutant concentrations in the cities of the BTH region were prepared using ArcGIS, which enabled the visualisation of the trends in pollutant concentrations and facilitated the identification of pollution hotspots. Finally, the potential health risks associated with each pollutant were evaluated using the AirQ+ software, with reference to the health risk thresholds for air pollutants set forth by the World Health Organization (WHO) and the Chinese national standards[214].

3.6.2.2 Software tools

The data were processed and analysed using Python, employing data processing and analysis libraries such as Pandas, NumPy, SciPy and stats models. The graphs were processed and generated using Python matplotlib, Microsoft Word and Excel. The analysis of health-related risks was conducted using the AirQ+ software developed by the WHO.

- Python

Python statistics libraries are comprehensive, popular and widely used tools. The combination of several Python libraries, including NumPy, SciPy, pandas and Matplotlib, enables the handling of outliers, data trends, correlation analysis, and the visualisation and analysis of data[215]. Therefore, Python is used as the main tool for data analysis in this study.

- AirQ+

The quantification of the impacts of air pollution on public health has significant important implications for the formulation of policy, the implementation of public health initiatives and the development of strategies for the prevention of pollution. Accordingly, the WHO has developed the software AirQ+ for the purpose of calculating

the impact of air pollution on human health. The software is capable of handling both long-term and short-term exposure to ambient air pollution caused by a multitude of pollutants, and is applicable to any city, country, or region. All calculations performed by AirQ+ are based on methods and concentration-response functions that have been established in epidemiological studies. It is possible to assess the extent to which a given health outcome can be attributed to specific air pollutants, as well as to evaluate how health effects may change in response to projected future changes in air pollution levels[216].

- ArcGIS

ArcGIS is a software program designed for use by organizations in the creation, management, and analysis of spatial data. The GIS is employed for the processing of cartographic materials. The software enables researchers to visualise data with geographic relevance, such as the behaviour of air pollution in different areas and its temporal changes. The visual representation may be presented in a variety of colours, according to the concentration levels of pollutants[217].

3.6.2.3 Health risk analysis

This study used AirQ+ software to assess the health effects associated with air pollution exposure over a ten-year period in the BTH area. The main coverage was on the attributable proportions (AP) of all CAUSES, RESPIRATORY and CIRCULATORY in adults over 30 years of age.

AP is calculated as follows:

$$AP = SUM\{(RR(c) - 1) * p(c)\}/SUN(RR(c) * p(c)) \quad (3.37)$$

Where:

AP is the attributable proportion of the health outcome, assuming a proven causal relation between the exposure and the health outcome, and assuming no major confounding factors affect this association.

Relative risk (RR) for a given health outcome is defined as the risk of an event or developing a disease relative to exposure. It is calculated by the ratio of the probability of the event occurring in the exposed group to the probability if the event occurring in the non-exposed group:

$$RR = \frac{Probability\ of\ event\ when\ exposed}{Probability\ of\ event\ when\ non-exposed}. \quad (3.38)$$

R(c) represents the proportion of the population exposed.

If the baseline frequency of the health outcome in the population under investigation is known, the rate attributable to the exposure can be calculated:

$$IE = I * AP \quad (3.39)$$

Where IE is the rate of the health outcome attributable to the exposure and I is the baseline frequency of the health outcome in the population under investigation.

Knowing the size of the population, the number of cases attributable to the exposure can be estimated as follows:

$$NE = IE * N \quad (3.40)$$

NE is the number of cases attributed to the exposure and N is the population's size investigated.

In this study, the air quality data of 13 cities in BTH region for the years 2014 and 2019 to 2023 were used as input data, and the total and respiratory AP covered four major pollutants $PM_{2.5}$, PM_{10}, NO_2, O_3, and $PM_{2.5}$ was selected for the attribution ratio of the circulatory.

3.7 Chapter summary

This chapter presented a comprehensive overview of the methodological framework employed in this study to achieve the research objectives. The methodology was designed to address three key aspects: multi-index air pollution prediction, spatial information-enhanced prediction model development, and long-term air quality change and health risk analysis in the Beijing-Tianjin-Hebei (BTH) region.

Firstly, the research area and data sources were introduced, explaining the selection of Beijing for air pollution prediction due to its comprehensive monitoring system and high-quality data. The BTH region, with severe pollution and significant policy interventions, was chosen to study long-term pollution changes.

Secondly, three air quality indices (AQI, AAQI, and HAQI) were detailed to evaluate air pollution from different perspectives. AQI reflects the highest pollutant concentration, AAQI considers the combined effects of multiple pollutants, and HAQI focuses on health risks. These indices form the basis for multi-pollutant index prediction in Chapter 4.

Thirdly, the predictive model framework included LSTM, MLP, Transformer, and GCN+LSTM models, each with distinct advantages. LSTM, MLP, and Transformer were used for multi-pollutant index prediction, while GCN+LSTM combined spatial and temporal features. GRU and LSTM+Attention models were introduced for performance comparison. Evaluation metrics such as MAE, MSE, and R^2 ensured rigorous model assessment.

Finally, statistical analyses were conducted using Python libraries, with ArcGIS for visualizing spatial pollution changes. Health risk assessments were performed using AirQ+ software combined with Python.

Chapter 4. Prediction of multi-pollution indices

4.1 Introduction

The variations in the levels of different pollutants in the air are complex, so a single AQI cannot meet the needs of people to guard against air pollution. In this chapter, a model that can simultaneously predict the three indices simultaneously is developed to achieve a comprehensive prediction of air pollution. The three indices are the traditional AQI, the AAQI which combines the weights of all pollutant concentrations, and the HAQI which considers the health-related impacts. In addition, in order to improve the accuracy of the prediction, the prediction model adopts a new model which combines three models, namely, the Long Short-term Memory (LSTM), the Multilayer Perceptron (MLP), and the Transformer, to predict air pollution, and automatically selects the result with the best performance. In this study, the correlations and differences between the three different indices were first evaluated, then the model was built and validated using data from different monitoring stations in Beijing, and finally the results were validated and analyzed using the commonly used Mean Absolute Error (MAE), Mean Squared Error (MSE) and Coefficient of Determination (R^2). Overall, the main contributions are as follows:

- Comprehensively analyzed the correlation and trend of three different air pollution indices: AQI, AAQI and HAQI.
- Predictions for two air pollution indices AAQI and HAQI in addition to the traditional AQI are presented.
- The applicability and predictive stability of the model were validated through predictions based on actual data from multiple air pollution monitoring stations in

Beijing.

- Three different prediction models are combined to make the prediction performance of the model more stable and accurate.

The rest of the chapter is organized as follows. Section 4.2 describes the correlation analysis of three pollution indices: AQI, AAQI and HAQI. Section 4.3 describes the entire operational process and methodology of multi-index forecasting. Section 4.4 describes the results of the forecasts and analyses and discusses the results. Section 4.5 summarizes the chapter.

4.2 Analysis of the correlations and trends of the AQI, AAQI and HAQI

The three air pollution indices are calculated in different ways, so this study first analysed the correlation between the three indices AQI, AAQI and HAQI, which also provided the necessary conditions for the subsequent predictions. The data were selected from daily air pollution data for Beijing throughout 2014, including the AQIm of six major pollutants as well as the corresponding total AQIs, AAQIs and HAQIs. The correlation between the three different air pollution indices is shown in Figure 4-1. As can be seen from the Figure 4-1, there is a strong correlation between AAQI and AQI (R^2=0.961), indicating that AAQI increases gradually with AQI. And the value is almost always higher than AQI, from the average value calculated, AAQI is 25% higher than AQI, which shows that AQI cannot fully reflect the overall harm of air pollution, in addition to the highest value of pollutants in the polluted weather, other pollutants besides the main pollutant will also bring greater harm. HAQI and AQI (R^2=0.965) correlation is slightly higher than the AAQI, again showing an increasing trend as the

AQI increase. The HAQI is also higher than the AQI, by an average of 20%. Moreover, the AAQI and HAQI overlap more in the 100-200 range, with similar values for the two pollution indices. In the interval beyond 200, the AAQI values start to be higher than the HAQI, which may be since at high pollution concentrations, the concentration of each pollutant is higher, which in turn leads to a higher HAQI.

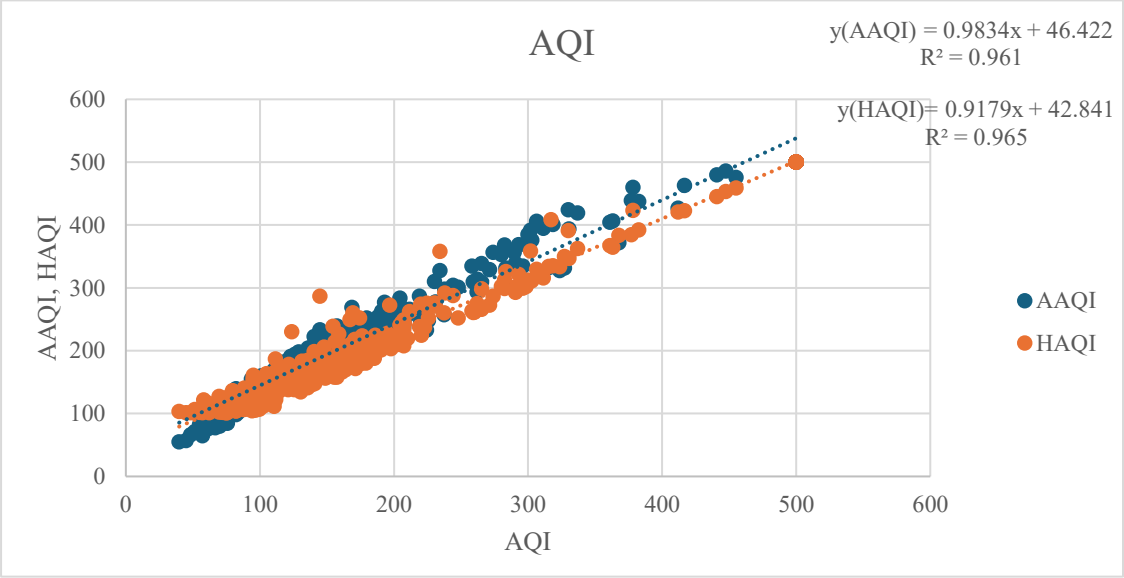


Figure 4-1: Scatterplot of correlation between AAQI, HAQI and AQI. Where blue points are AAQI and orange points are HAQI. The x-axis is the value of AQI, and the y-axis is the value of AAQI and HAQI corresponding to AQI.

AQI, AAQI and HAQI are calculated based on the pollution values of each pollutant, therefore, the correlation of each pollutant with the three pollution indices is also important to achieve the prediction of the three pollution indices. The correlation of the six main pollutants with the AQI, AAQI and HAQI will be evaluated next.

This study first analyses the correlation of $PM_{2.5}$. As can be seen from the Figure 4-2, $PM_{2.5}$ has the high correlation with the AQI-total ($R^2 = 0.8331$), which indicates that $PM_{2.5}$ is the main pollutant most of the time, and the main pollutant is directly related

to the value of the AQI-total. the AAQI has the highest correlation with $PM_{2.5}$ (R^2 = 0.8522), which indicates that $PM_{2.5}$ has a major influence in the AAQI index, which takes into account the combined effects of different pollutants, and that the value of the AAQI is more influenced by $PM_{2.5}$. The correlation between $PM_{2.5}$ and HAQI, on the other hand, is relatively low (R^2=0.7883), although it is still the highest compared to other pollutants. This suggests that the RR to health for each pollutant is more balanced. In addition, the values of the three indices occasionally exceed those of $PM_{2.5}$ within the 0–100 range. This discrepancy arises because, in certain instances, pollutants other than $PM_{2.5}$ dominate as the primary pollutants, resulting in significantly higher levels.

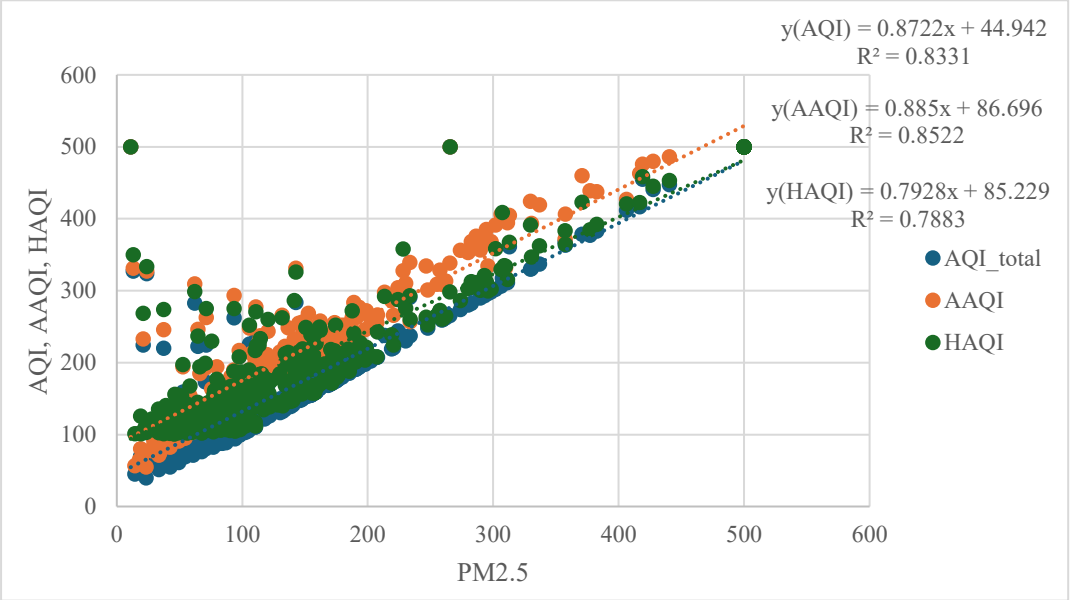


Figure 4-2: Scatter plots of correlation between AQI, AAQI and HAQI with $PM_{2.5}$. Where the x-axis is the index of $PM_{2.5}$ and the y-axis is the index of AQI, AAQI and HAQI. The blue point is the value of AQI-total, the orange point is the value of AAQI and the grey point is the value of HAQI.

Notably, the AAQI and HAQI tend to exhibit elevated values more frequently, further highlighting the broader hazards posed by air pollution. In the range above 100, the

trend of each pollution index shows a big difference, the value of AQI_total is more stable and basically the same as $PM_{2.5}$, while HAQI is slightly higher than $PM_{2.5}$ in general, and the value of AAQI is much higher than $PM_{2.5}$, which shows that other pollutants also have a big influence on AAQI. Therefore, $PM_{2.5}$ should be used as a key characteristic variable when predicting the air pollution index.

Next, Figure 4-3, PM_{10}, which is the second major pollutant, will be analysed. The trend of correlation between PM_{10} and the three pollution indices is more similar to that of $PM_{2.5}$, with all showing an increasing trend. The correlation coefficients R^2 are not quite the same, with PM_{10} having the highest correlation with the AQI (R^2=0.7665), followed by the HAQI (R^2=0.754), the AAQI (R^2=0.7189) being the lowest, and all being lower than $PM_{2.5}$ overall. This suggests that the impact of PM_{10} on the AQI is relatively large and has a higher weighting among the individual pollutants.

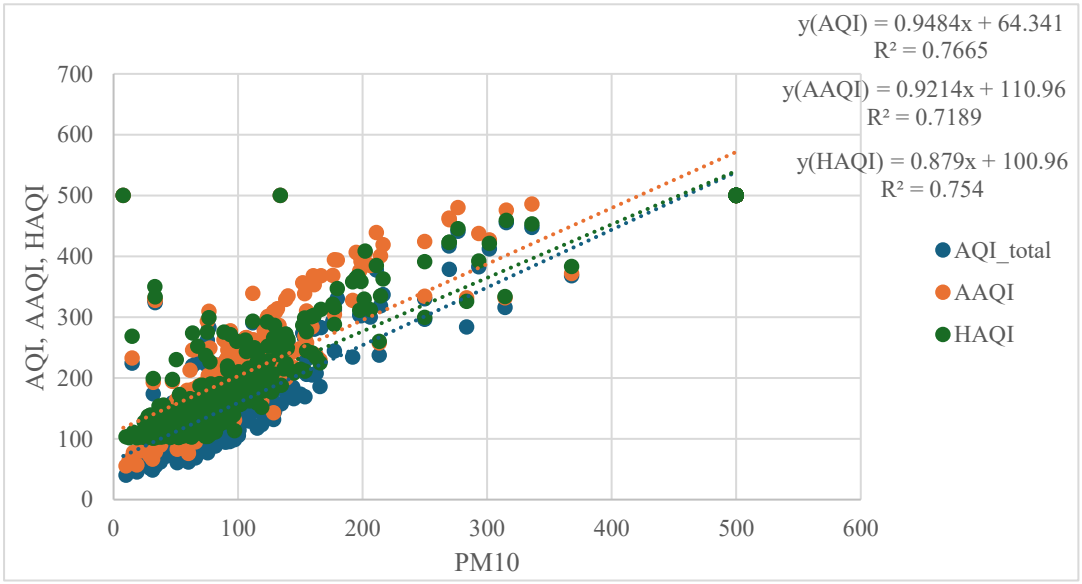


Figure 4-3: Scatter plots of correlation between AQI, AAQI and HAQI with PM_{10}. Where the x-axis is the index of PM_{10} and the y-axis is the index of AQI, AAQI and HAQI. The blue point is the value of AQI-total, the orange point is the value of AAQI and the grey point is the value of HAQI.

In addition, the distribution of the three indices shows a tendency to be heavily concentrated in the 0-200 interval, with relatively few times when AQI of PM_{10} exceeds 200, both AAQI and HAQI show higher values than those of AQI. When predicting the pollution index, the weight of PM_{10} should be increased when the pollution indices are between 0 and 200, especially for the prediction of AAQI and HAQI, where the correlation is stronger.

Gaseous pollutants usually make up a particularly low percentage of air pollution. The correlation of Figure 4-4, SO_2 shows that the correlation of SO_2 with the AQI (0.6423), HAQI (0.6135), and AAQI (0.5110) is not too high and most of the time, the value of SO_2 is in the healthy range of a little under 100.

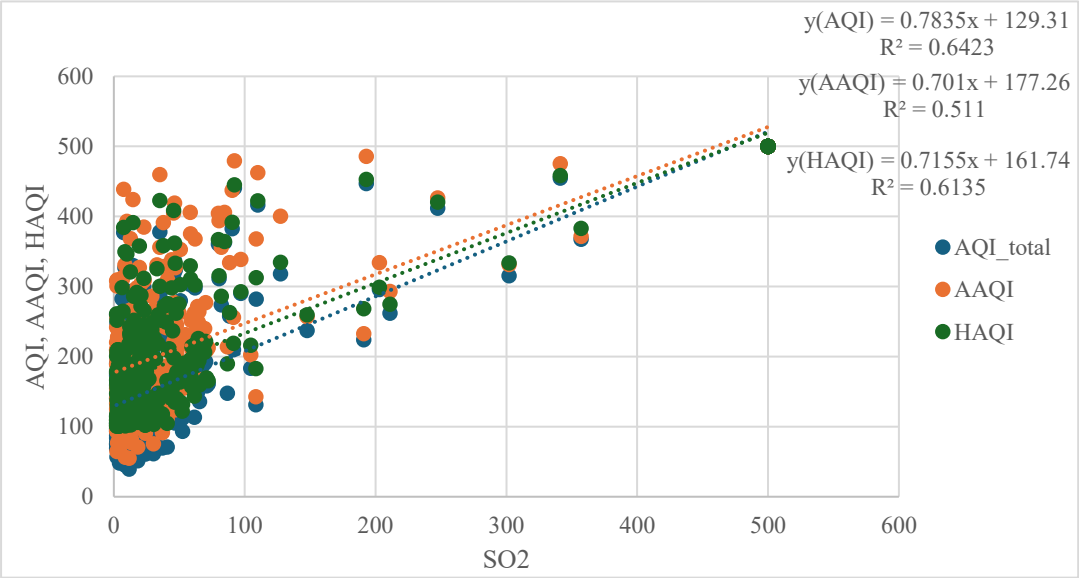


Figure 4-4: Scatter plots of correlation between AQI, AAQI and HAQI with SO_2. Where the x-axis is the index of SO_2 and the y-axis is the index of AQI, AAQI and HAQI. The blue point is the value of AQI-total, the orange point is the value of AAQI and the grey point is the value of HAQI.

The correlation of Figure 4-5 CO shows a similar situation to that of SO_2. However, these two pollutants are more harmful at high levels and should not be ignored. Figure 4-6 NO_2 behaves differently, although being a gaseous pollutant, most of the time pollution values below 100 are in the healthy range. However, due to the large amount of exhaust emissions from motor vehicles, there are many times when this pollutant is above the 100 range. The R^2 with the AQI and HAQI is around 0.7. The lower correlation with AAQI (0.5854) may be due to more emission of other pollutants along with NO_2[218]. Gaseous pollutants SO_2, CO and NO_2 have little impact on the pollution indices in most cases, so their weights should be appropriately reduced. However, when the concentration of pollutants is high, the impact of gaseous pollutants on the air pollution indices will increase, and at this time, consideration will be given to increasing their weight in pollution forecasts.

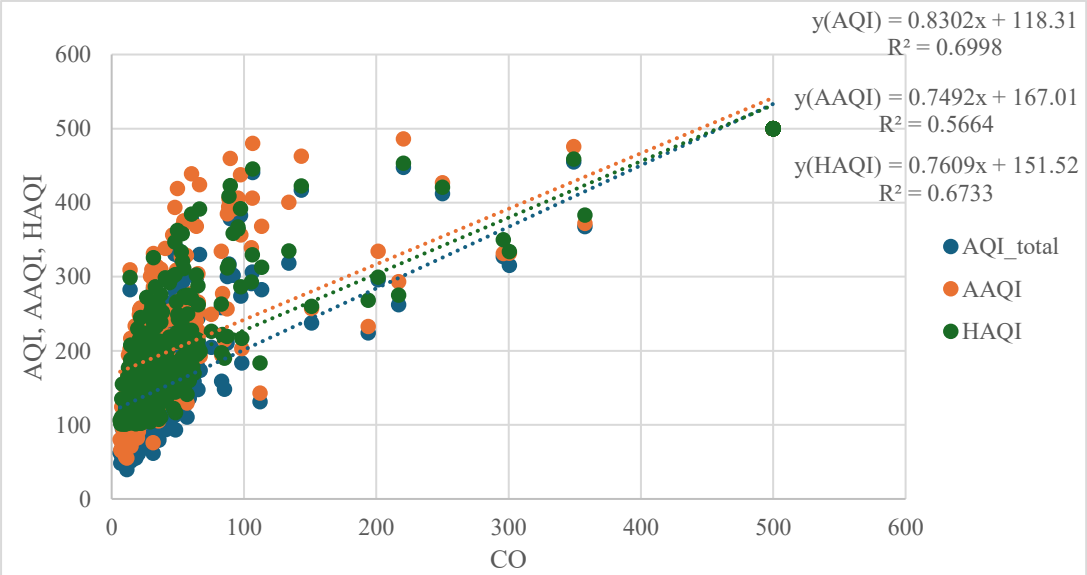


Figure 4-5: Scatter plots of correlation between AQI, AAQI and HAQI with CO. Where the x-axis is the index of CO and the y-axis is the index of AQI, AAQI and HAQI. The blue point is the value of AQI-total, the orange point is the value of AAQI and the grey point is the value of HAQI.

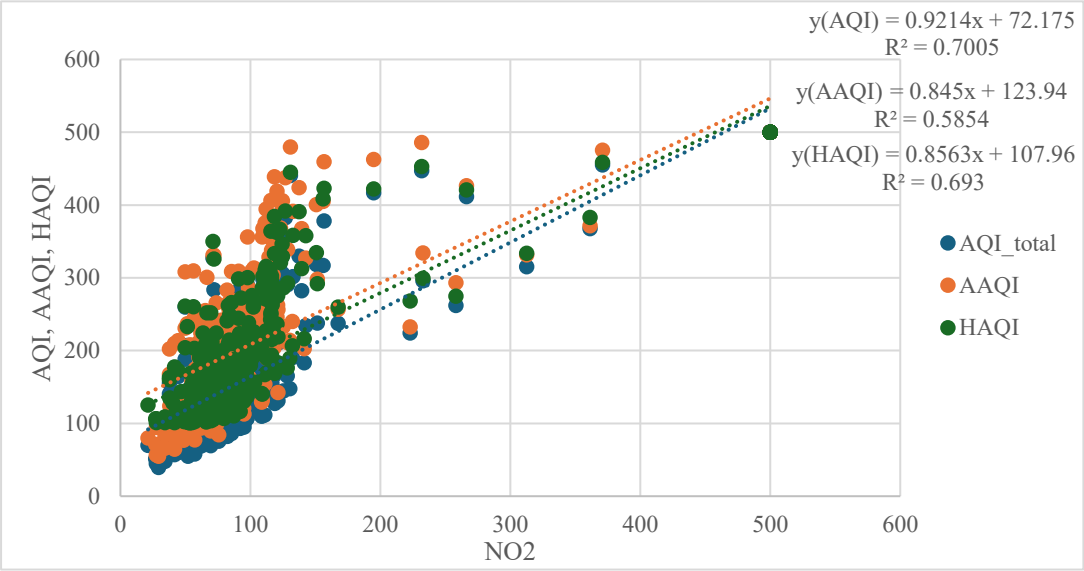


Figure 4-6: Scatter plots of correlation between AQI, AAQI and HAQI with NO_2. Where the x-axis is the index of NO_2 and the y-axis is the index of AQI, AAQI and HAQI. The blue point is the value of AQI-total, the orange point is the value of AAQI and the grey point is the value of HAQI.

Ozone is unique among pollutants in that it has the lowest correlation with all three pollution indices and only the R^2 with the AQI exceeds 0.5. In addition, most of the time the AQI, AAQI, and HAQI are spread out in the healthy range of Figure 4-7 O_3 (AQI < 100). This suggests that O_3 is not the dominant pollutant most of the time and the concentrations are low. However, there are a few occasions when the ozone concentration index is high and close to the total AQI, which is due to the fact that ozone is mainly affected by the strong summer sunlight, and is produced by photochemical reactions of nitrogen oxides and volatile organic compounds in the presence of sunlight. Most of the time when O_3 is a major pollutant occurs in the summer months, when the overall pollution is very low and there is plenty of sunlight.

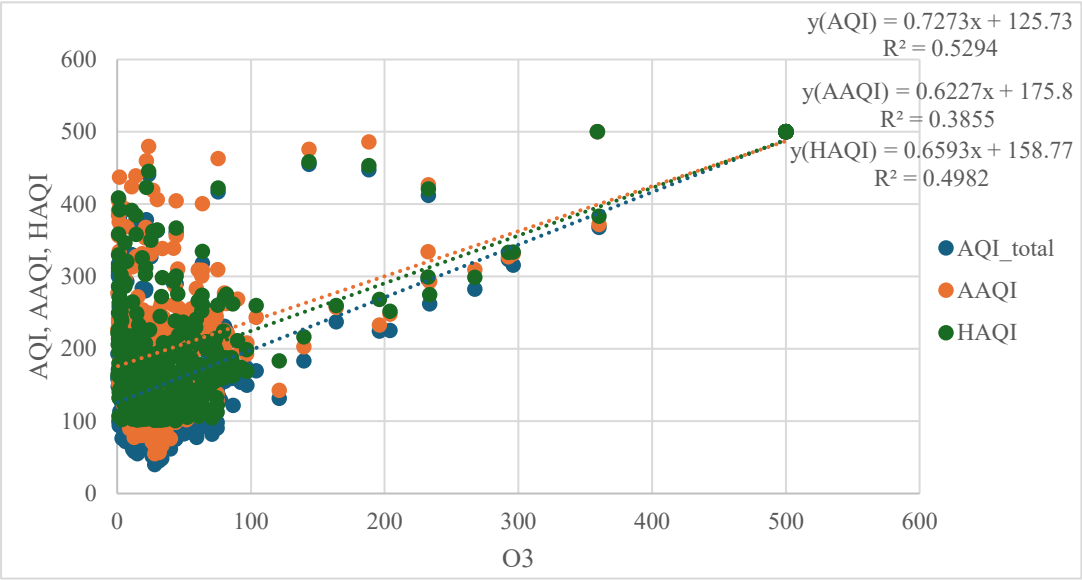


Figure 4-7: Scatter plots of correlation between AQI, AAQI and HAQI with O_3. Where the x-axis is the index of O_3 and the y-axis is the index of AQI, AAQI and HAQI. The blue point is the value of AQI-total, the orange point is the value of AAQI and the grey point is the value of HAQI.

Most of the time, O_3 has little impact on the prediction of the air pollution indices, and its weight is low. Only in the summer when the overall air quality is good, O_3 has a slightly greater impact on the air pollution index due to the large influence of sunlight intensity and temperature. Therefore, the weight of O_3 should be combined with temperature changes to allocate the weight.

The correlation between the air pollution indices AQI, AAQI and HAQI is extremely high, exceeding 0.95, which provides feasibility for the simultaneous prediction of the three indices in the experiment. At the same time, the different correlations between each pollutant and the pollution indices provide support for the weight allocation of input variables in the experimental setting. For example, particulate pollutants should have a higher weight, the weight of gaseous pollutants can be appropriately increased when the pollution index is high, and the weight of O_3 can be appropriately increased in summer.

4.3 Experimental settings

The experiments were conducted by Python (version 3.6) where packages such as Pandas, Numpy, Torch, Sklearn, Matplotlib, Math, Pdb were mainly used to complete the prediction of multiple indices. These packages provide the necessary processes such as computational analysis and image generation necessary for the experiments.

In this study, a multi-pollution index prediction model consisting of several neural network models was developed. The model enables simultaneous prediction of AQI, AAQI and HAQI with predicted values including pollution concentrations for the next 1, 3 and 6 hours. In order to improve the prediction performance, three models, LSTM, MLP and Transformer, are fused together, and then, the performance of each of the three models is evaluated by the MSE, MAE and R^2 indicators, and the prediction result of the model with the best performance is given as the final result. The structure of the model is presented in Figure 4-8 and contains the main steps of the prediction: (1) pre-processing of the data; (2) predictions from multiple models; (3) evaluation and selection of the best model.

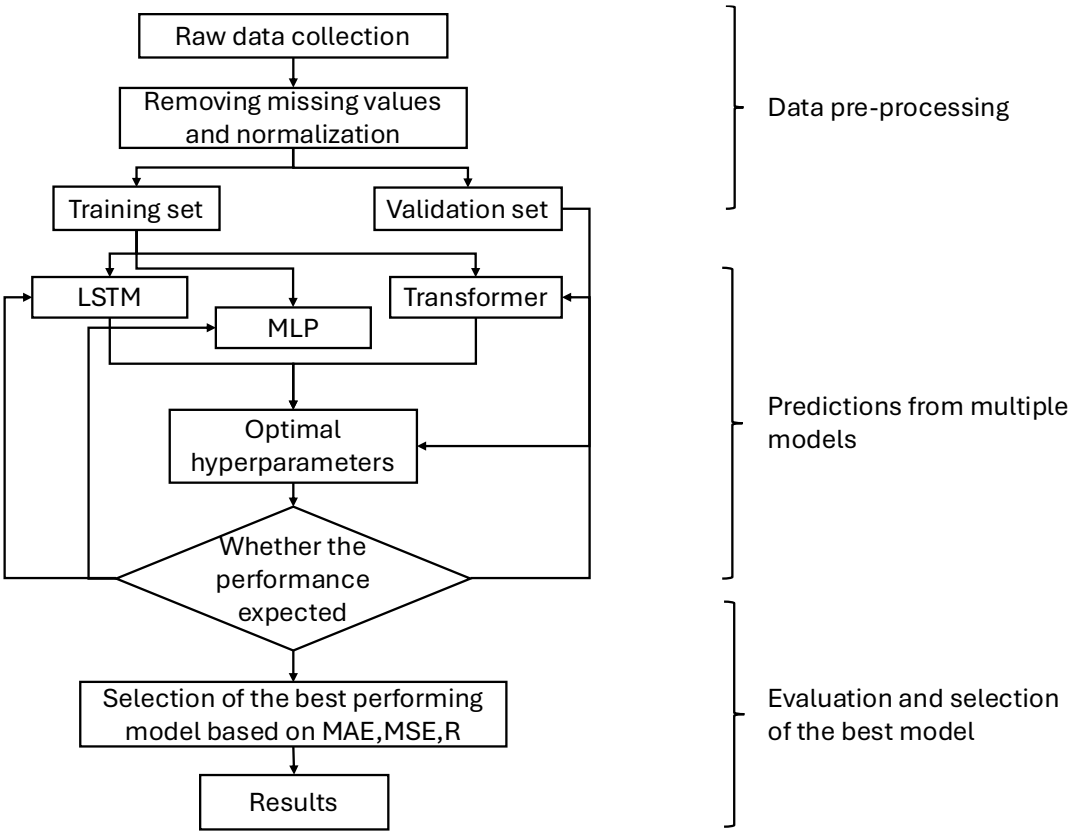


Figure 4-8: The structure of the prediction model with multiple pollution indices. The model consists of LSTM, MLP and Transformer.

There are missing or anomalous data in the model from raw data. To be able to improve the data quality and model performance, the missing values and anomalous data were firstly removed and then the data were normalised. Next the model was divided into a training set and a test set in a ratio of 8 to 2 to ensure reliable evaluation of the model. After the data processing is completed, it comes to the training step, firstly, the data in the training set is added to the three models LSTM, MLP, and Transformer to train them separately, and compared with the test set to adjust the parameters. The final settings of the parameters are shown in Table 4-1. The history step is set to 10 and the future step is set to 6, so that the data of the past ten hours can be used to predict the data of the future 1, 3 and 6 hours. In addition, other parameters such as Epochs are set to 100, Batch Size is set to 256, and Learning Rate is set to 0.001, which ultimately gives the

best parameters and prediction performance. After all the training is done, the performance of the three models will be evaluated using MAE, MSE and R^2 and the best performing model will be selected as the final model. The model is predicted for each of the ten air pollution monitoring stations.

Table 4-1: The parameters of the models.

Parameter	Value
Time interval (hour)	1
Training set	80%
Test set	20%
Prediction length	1, 3, 6
History length	10
Training epochs	100
Batch size	256
Learning rate	0.001
Number of data	35065

4.4 Results and discussion

This section shows the prediction results of AQI, AAQI and HAQI for 10 different monitoring stations in Beijing. In the experiment, the pollution indices were predicted for the next 1 hour, 3 hours and 6 hours and the predicted results were evaluated using the evaluation metrics R^2, MAE and MSE. R^2, MAE, and MSE are used together

because they provide a concise yet comprehensive evaluation of air pollution prediction performance. R^2 reflects how well the model captures overall variability and trends (The closer the R^2 value is to 1, the better the agreement between the predicted values and the observed values, indicating a stronger model fit), MAE measures the average prediction error in an intuitive and robust manner, and MSE emphasizes larger errors, which is important for assessing extreme pollution events. Together, they offer a balanced assessment of accuracy, stability, and risk sensitivity.

In the study by Mehdi et al.[219], the prediction data was from only one monitoring station, which may not be able to verify the stability of the model. For example, when the data from the monitoring station is abnormal, it may affect the performance and parameter settings of the model. This study verified the data from 10 monitoring stations distributed in different locations, making the model of this study more universal and more stable. In addition, to balance readability and accuracy, all results were rounded to three decimal places. This ensures sufficient precision to assess variations and differences in the data, while avoiding reduced readability caused by excessive decimal places. The overall results show that the predictions at the different monitoring stations are close to each other, which indicates that the model is applicable to different locations and has a stable performance.

Table 4-2 shows the prediction results of the Transformer model for three different indices, AQI, AAQI, and HAQI, at forecasting horizons of 1, 3, and 6 hours. For the 1-hour horizon, the model achieves relatively strong predictive accuracy, with R^2 values ranging from 0.758 to 0.867 and the lowest MAE and MSE among all horizons. In particular, AAQI obtains the highest performance (R^2 = 0.867), indicating that the model is more capable of capturing short-term fluctuations in sensitive pollutant components. AQI also shows satisfactory accuracy (R^2 = 0.787), while HAQI presents comparatively lower performance, which is expected due to its composite nature and higher inherent variability.

When the prediction horizon increases to 3 hours, all metrics demonstrate a noticeable decline. The R^2 values for AQI, AAQI, and HAQI decrease to 0.680, 0.753, and 0.630, respectively, accompanied by an increase in MAE and MSE. This suggests that the Transformer model becomes less responsive to the evolving temporal dynamics of air pollutants over medium-term intervals. Although AAQI remains the most predictable indicator, the overall reduction in accuracy indicates growing uncertainty as the forecasting window lengthens.

For the 6-hour horizon, the model exhibits a substantial loss of predictive capability. The R^2 values drop to 0.511 for AQI, 0.533 for AAQI, and 0.417 for HAQI, with correspondingly larger MAE and MSE values. This indicates that the Transformer model's predictions for AQI and AAQI remain at a satisfactory level.

Overall, the results show that the Transformer model performs well in 1- and 3-hour forecasts, and while it is acceptable in 6-hour forecasts, it is unstable. The consistent ranking of AAQI > AQI > HAQI across all horizons further highlights the differing levels of temporal volatility among these indices.

Table 4-2: Prediction results of AQI, AAQI and HAQI by Transformer at Aotizhongxin air pollution monitoring station.

Forecast Horizon	**Index**	**R^2**	**MAE**	**MSE**
1-hour	AQI	0.787	0.066	0.006
	AAQI	0.867	0.056	0.005
	HAQI	0.758	0.078	0.011
3-hour	AQI	0.680	0.071	0.010
	AAQI	0.753	0.073	0.011
	HAQI	0.630	0.087	0.018
6-hour	AQI	0.511	0.088	0.017
	AAQI	0.533	0.105	0.023
	HAQI	0.417	0.117	0.030

Table 4-3 shows the 1-hour, 3-hour and 6-hour prediction results of the MLP model for three pollution indices: AQI, AAQI and HAQI. For 1-hour horizon, the model achieves strong accuracy, with R^2 values of 0.850 for AQI and 0.880 for AAQI, accompanied by the lowest MAE and MSE across all horizons. These results suggest that the MLP is capable of effectively capturing sort-term nonlinear relationships in air quality data. In contrast, HAQI exhibits lower performance (R^2 = 0.758), reflecting its higher variability and the increased difficulty associated with predicting a composite health-weighted index.

At the 3-hour horizon, all three indices show a marked reduction in accuracy. The R^2 values drop to 0.709 for AQI, 0.722 for AAQI, and 0.584 for HAQI, with corresponding increases in MAE and MSE. This degradation indicates that the MLP model becomes less capable of maintaining predictive stability over medium-range intervals, likely due to its limited ability to model longer-term temporal dependencies. AAQI remains the most predictable index, and the results for all three indices are still good.

Table 4-3: Prediction results of AQI, AAQI and HAQI by MLP at Aotizhongxin air pollution monitoring station.

Forecast Horizon	**Index**	**R^2**	**MAE**	**MSE**
1-hour	AQI	0.850	0.045	0.004
	AAQI	0.880	0.048	0.005
	HAQI	0.758	0.068	0.009
3-hour	AQI	0.709	0.060	0.008
	AAQI	0.722	0.074	0.012
	HAQI	0.584	0.093	0.018
6-hour	AQI	0.500	0.085	0.016
	AAQI	0.543	0.101	0.021
	HAQI	0.428	0.114	0.028

For the 6-hour horizon, the predictive capability declines substantially. The R^2 for AQI

decreases to 0.500, while AAQI and HAQI drop to 0.543 and 0.428, respectively. As the prediction time increases, the model's predictive performance decreases significantly. however, the results are still within acceptable limits.

Table 4-4 indicates that LSTM model exhibits strong predictive capability for short-term air quality forecasting with performance gradually declining as the forecasting horizon increases. For the 1-hour horizon, the model achieves the highest accuracy among all tested horizons, with R^2 values of 0.904 for AQI and 0.916 for AAQI, accompanied by low MAE and MSE values. These results highlight the LSTM's strength in capturing temporal dependencies and short-term pollutant dynamics. However, HAQI remains more challenging to predict ($R^2 = 0.797$), which is consistent with its greater variability and the complexity associated with modelling health-weighted composite indices.

Table 4-4: Prediction results of AQI, AAQI and HAQI by LSTM at Aotizhongxin air pollution monitoring station.

Forecast Horizon	**Index**	**R^2**	**MAE**	**MSE**
1-hour	AQI	0.904	0.034	0.003
	AAQI	0.916	0.037	0.003
	HAQI	0.797	0.062	0.008
3-hour	AQI	0.746	0.057	0.008
	AAQI	0.769	0.067	0.010
	HAQI	0.634	0.088	0.017
6-hour	AQI	0.551	0.082	0.015
	AAQI	0.566	0.100	0.021
	HAQI	0.453	0.114	0.027

As the forecasting horizon extends to 3 hours, the model's performance declines noticeably. The R^2 for AQI and AAQI drops to 0.746 and 0.769, respectively, indicating

reduced explanatory power in medium-term predictions. HAQI again shows a larger performance drop ($R^2 = 0.634$), accompanied by increased MAE and MSE values.

For the 6-hour horizon, the predictive accuracy decreases substantially. R^2 values fall to 0.551 for AQI, 0.566 for AAQI, and 0.453 for HAQI, with MAE values exceeding 0.08 for all indices. Although the R^2 value shows a slight decrease, it still falls within an acceptable range. Moreover, HAQI remains a relatively challenging index to predict due to its inherent complexity.

Table 4-2, Table 4-3 and Table 4-4 show the model predictions for AQI, AAQI and HAQI for Aotizhongxin monitoring station. The predicted results were selected from the best of the results predicted by the three prediction models. The LSTM (Table 4-4) is the best in all indicators in the 1-hour prediction. Compared with the AQI prediction results of Doreswamy et al.[221] using Linear Regression(R^2=0.793), Lasso(R^2=0.626), Ridge (R^2=0.793) and other models, the fit between the predicted values and the true values of the results of this study is much higher than these models. In the three-hour prediction, LSTM predicted the best results for all three metrics for AQI and AAQI, followed by MLP (Table 4-3), and Transformer (Table 4-2) was the worst. In the prediction of AQI, MLP is higher than Transformer, and in the prediction of AAQI, Transformer is higher than MLP. In the prediction of HAQI, LSTM has the best in two metrics, which are R^2=0.634, MSE=0.017, and MAE, which has the best of Transformer as 0.087 and Transformer's R^2 and MSE are also better than MLP's. In the six-hour prediction, LSTM is still the best, with all the indices higher than MLP and Transformer, and MLP's indicators are better than Transformer's except for AQI's R^2 and AAQI's MSE, which are in the second place. Therefore, at the Aotizhongxin monitoring station, the prediction result of LSTM was chosen as the final prediction result by evaluating the indices. The prediction results of other monitoring stations were also automatically selected as the final prediction results with the best performance according to this

method. In general, the performance of a single model may vary under different conditions. By integrating the three models, we can leverage their respective strengths in different scenarios, thereby enhancing both the stability and accuracy of predictions. This has been verified in multiple studies by Mao[220], Chen[117], [125] and others.

Taking the Aotizhongxin monitoring station as an example, the R^2 of the one-hour (Table 4-5) predicted AQI and AAQI reached 0.904 and 0.916 respectively, which is a very high level of fit, while the R^2 of the HAQI was lower, but reached 0.797, which is also a high level of fit. In the three-hour (Table 4-6) prediction, the R^2 of AQI and AAQI reached 0.746 and 0.769, respectively, and the R^2 of HAQI was relatively low, also reaching 0.634. This indicates that the prediction for the three-hour period is qualified. However, in the six-hour (Table 4-7) prediction, the value of R^2 drops a lot again, in which the R^2 of AQI and AAQI are 0.551 and 0.566 respectively, and the R^2 of HAQI is even lower than 0.5 to reach 0.453. Taken together, the model's performance is outstanding in the one-hour prediction, and the performance is relatively excellent in the three-hour prediction, and in the six-hour prediction, the R^2 of AQI and AAQI is even lower than 0.5 to reach 0.453 have passable prediction performance and HAQI has poorer prediction. It also shows that the longer the forecasting period, the more volatile the data becomes, especially for the HAQI, which is highly influenced by ER and RR, and that the model has a slightly more difficult time predicting the performance of the model for a six-hour period[221].

For the one-hour (Table 4-5) AQI forecast, six of the ten monitoring stations had results above 0.9, which is extremely high. Of the stations with less than 0.9, Tiantan had the lowest at 0.857, but it also shows that the model's performance in predicting the AQI is extremely high. The prediction performance of AAQI is higher than that of AQI, only three stations have R^2 lower than 0.9, and the lowest one is still Tiantan with 0.887. The prediction performance of HAQI is almost better than the former two, the R^2 is

distributed between 0.6 and 0.8, the highest one is 0.803 in Guanyuan, and the lowest one is 0.626 in Dingling. Taken together, the prediction performance is still high, although it fluctuates slightly.

Table 4-5: 1-hour predictions of AQI, AAQI and HAQI at ten air pollution monitoring stations in Beijing.

Station	Index	R^2	MAE	MSE
Aotizhongxin	AQI	0.904	0.034	0.003
	AAQI	0.916	0.037	0.003
	HAQI	0.797	0.062	0.008
Huairou	AQI	0.903	0.026	0.001
	AAQI	0.932	0.028	0.002
	HAQI	0.673	0.050	0.006
Changping	AQI	0.909	0.031	0.002
	AAQI	0.919	0.036	0.003
	HAQI	0.795	0.049	0.006
Nongzhanguan	AQI	0.890	0.033	0.002
	AAQI	0.919	0.037	0.003
	HAQI	0.772	0.060	0.008
Dingling	AQI	0.902	0.023	0.002
	AAQI	0.932	0.026	0.002
	HAQI	0.626	0.040	0.004
Shunyi	AQI	0.908	0.031	0.002
	AAQI	0.929	0.033	0.002
	HAQI	0.786	0.058	0.007
Dongsi	AQI	0.904	0.034	0.003
	AAQI	0.916	0.037	0.003
	HAQI	0.797	0.062	0.008
Tiantan	AQI	0.857	0.036	0.003
	AAQI	0.887	0.039	0.004
	HAQI	0.734	0.068	0.010
Guanyuan	AQI	0.867	0.035	0.003
	AAQI	0.891	0.040	0.004
	HAQI	0.803	0.058	0.007
Wanshouxigong	AQI	0.873	0.033	0.003
	AAQI	0.898	0.036	0.004
	HAQI	0.764	0.061	0.008

In addition, in the prediction of the three-hour (Table 4-6) AQI, all sites show high levels of R^2, distributed in the range of 0.6 to 0.8, except for Dingling, which has a low R^2 of 0.490.

Table 4-6: 3-hour predictions of AQI, AAQI and HAQI at ten air pollution monitoring stations in Beijing.

Station	Index	R^2	MAE	MSE
Aotizhongxin	AQI	0.746	0.057	0.008
	AAQI	0.769	0.067	0.010
	HAQI	0.634	0.088	0.017
Huairou	AQI	0.687	0.043	0.005
	AAQI	0.744	0.051	0.006
	HAQI	0.325	0.066	0.011
Changping	AQI	0.784	0.047	0.004
	AAQI	0.791	0.058	0.007
	HAQI	0.615	0.068	0.011
Nongzhanguan	AQI	0.673	0.060	0.010
	AAQI	0.711	0.072	0.013
	HAQI	0.533	0.095	0.021
Dingling	AQI	0.490	0.041	0.005
	AAQI	0.645	0.051	0.007
	HAQI	0.479	0.053	0.009
Shunyi	AQI	0.743	0.051	0.006
	AAQI	0.763	0.061	0.009
	HAQI	0.578	0.086	0.017
Dongsi	AQI	0.746	0.057	0.008
	AAQI	0.769	0.067	0.010
	HAQI	0.634	0.088	0.017
Tiantan	AQI	0.656	0.059	0.008
	AAQI	0.695	0.070	0.011
	HAQI	0.506	0.091	0.018
Guanyuan	AQI	0.722	0.052	0.007
	AAQI	0.746	0.064	0.009
	HAQI	0.624	0.081	0.015
Wanshouxigong	AQI	0.704	0.055	0.008
	AAQI	0.724	0.066	0.011
	HAQI	0.547	0.092	0.019

This indicates an overall stable forecast for the three-hour AQI. In the prediction of AAQI, the prediction performance is higher than that of AQI, and only two stations, Dingling and Tiantan, have an R^2 lower than 0.7, with a minimum of 0.645 at Dingling, which indicates that the model has a high performance in predicting the three-hourly AAQI.

The prediction performance of HAQI is similar to that of one hour and is also the worst of the three pollution indicators. While the overall performance is superior in the range of 0.5 to 0.7, the R^2 for Huairou is as low as 0.325 (This may be related to incomplete or unstable data at the monitoring station), which is a poor performance[222]. Overall, the model's ability to predict the AQI, AAQI and HAQI for three hours is better, with only poor performance on individual monitoring stations.

As the prediction gets longer, it becomes more difficult to predict. In the prediction for the next six hours (Table 4-7) , the R^2 of AQI is mostly concentrated in the range of 0.5 to 0.6, only two monitoring stations, Shunyi and Tiantan, are below 0.5, and Shunyi is the lowest with 0.379, which indicated that the predicted results were qualified. The prediction performance of AAQI is still the best among the three indicators, only one station, Shunyi, has an R^2 lower than 0.5, 0.460. The values of the rest of the stations are all higher than 0.5 and higher than that of AQI. This indicates that the model's six-hour prediction performance for AAQI is satisfactory. The prediction performance of HAQI is still the worst, only six stations have R^2 higher than 0.4, which is a poor prediction performance.

Wen's study also predicted air quality at different times in the future. The results showed that the R^2 of AQIs predicted for the next hour were 0.83, 0.76, 0.72, 0.79, and 0.79, the R^2 of AQIs predicted for the next three hours were 0.60, 0.51, 0.37, 0.56, and 0.52, and the R^2 of AQIs predicted for the next six hours were 0.45, 0.40, 0.18, 0.39, and 0.33[223]. Compared with the prediction results of this study, in the one-hour prediction, the R^2 of the prediction results of AQI and AAQI in this study exceeded 0.9, and the R^2

of the prediction of HAQI was slightly lower, 0.7971, which also reached above its average level.

Table 4-7: 6-hour predictions of AQI, AAQI and HAQI at ten air pollution monitoring stations in Beijing.

Station	Index	R^2	MAE	MSE
Aotizhongxin	AQI	0.551	0.082	0.015
	AAQI	0.566	0.100	0.021
	HAQI	0.453	0.114	0.027
Huairou	AQI	0.510	0.062	0.009
	AAQI	0.517	0.078	0.013
	HAQI	0.309	0.082	0.017
Changping	AQI	0.566	0.067	0.010
	AAQI	0.569	0.084	0.015
	HAQI	0.439	0.087	0.018
Nongzhanguan	AQI	0.517	0.077	0.013
	AAQI	0.544	0.094	0.019
	HAQI	0.401	0.109	0.024
Dingling	AQI	0.379	0.057	0.008
	AAQI	0.460	0.072	0.011
	HAQI	0.350	0.067	0.012
Shunyi	AQI	0.517	0.076	0.013
	AAQI	0.539	0.092	0.018
	HAQI	0.399	0.108	0.025
Dongsi	AQI	0.551	0.082	0.015
	AAQI	0.566	0.100	0.021
	HAQI	0.453	0.114	0.027
Tiantan	AQI	0.486	0.076	0.012
	AAQI	0.500	0.094	0.018
	HAQI	0.378	0.110	0.025
Guanyuan	AQI	0.566	0.072	0.011
	AAQI	0.579	0.089	0.016
	HAQI	0.486	0.103	0.022
Wanshouxigong	AQI	0.523	0.077	0.013
	AAQI	0.541	0.094	0.019
	HAQI	0.437	0.109	0.025

This shows that the one-hour prediction of multiple pollution indices in this study is successful. In the three-hour prediction, the R^2 values of the three pollution indices in this study are higher than the results of the three-hour prediction in Wen's study. In the six-hour prediction, Wen's research results also showed a relatively ordinary performance, but the R^2 of the prediction results of multiple pollution indices in this study were better than its performance. This shows that the 1, 3, 6-hour prediction results for the three pollution indices proposed in this study have reached a high level.

The above results and analyses show that the prediction model proposed in this study achieves a satisfactory level of prediction of AQI, AAQI and HAQI at 1, 3 and 6 hours. For the six-hourly predictions, the AQI and AAQI predictions were satisfactory, while the HAQI predictions showed different performances at different stations, with half of the stations being more satisfactory and half less satisfactory. The application of the model to ten sites demonstrated the excellent performance of the model for 1- and 3-hour forecasts, and the predicted performance fluctuates for 6 hours, but predictions are satisfactory.

Figure 4-10, Figure 4-11, Figure 4-11 show the performance of the Aotizhongxin monitoring station model for 6-hour AQI, AAQI and HAQI prediction, where the blue line is the predicted value, and the red line is the real value[224]. In the prediction of AQI, the predicted value and the true value have a high degree of match, especially when the fluctuation of the true value is stable, the predicted value can match it well[225]. In the prediction of AAQI, the degree of match between the predicted value and the true value is similar to that of AQI. When the fluctuation of the true value is relatively stable, it shows a good match, and when the true value fluctuates greatly, there will be a certain difference. The prediction of HAQI is worse than AQI and AAQI, especially when the true value fluctuates greatly, the match with the predicted value will be very low. This is consistent with the value of R^2 (0.4528) in the previous data analysis.

Taken together, the predicted value fits well with the true value when the true value fluctuates more steadily, and the predicted value is significantly higher than the true value when the true value fluctuates too much. The model performs better in predicting more stable pollution. This is similar to the conclusion reached by Ni[226], Lin[227], Cheng[228] et al. in their study on air pollution prediction. The volatility of air pollution directly affects the accuracy of the prediction.

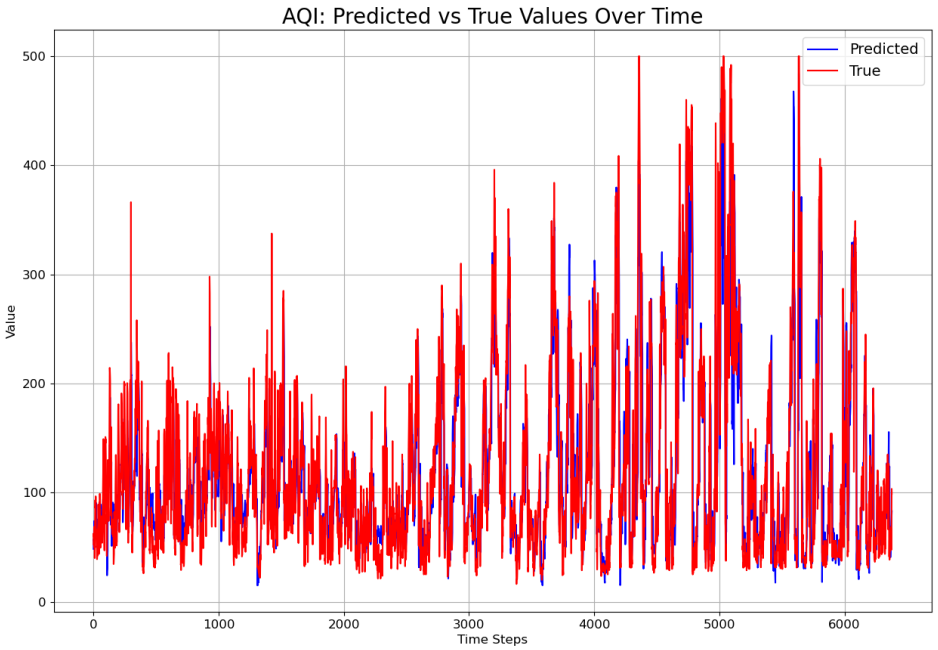


Figure 4-9: Comparison of predicted and actual values of AQI for 6-hour at Aotizhongxin monitoring station.

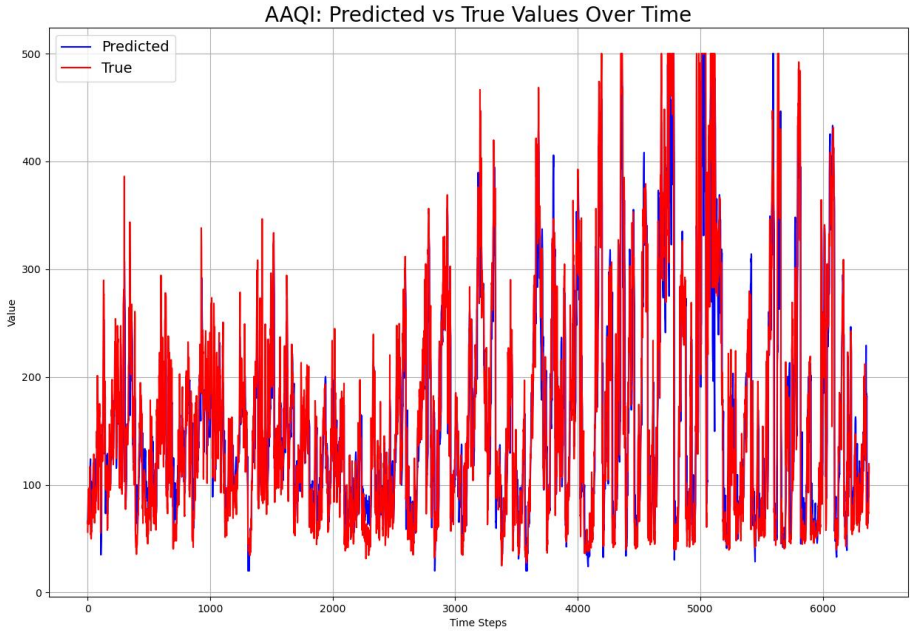


Figure 4-10: Comparison of predicted and actual values of AAQI for 6-hour at Aotizhongxin monitoring station.

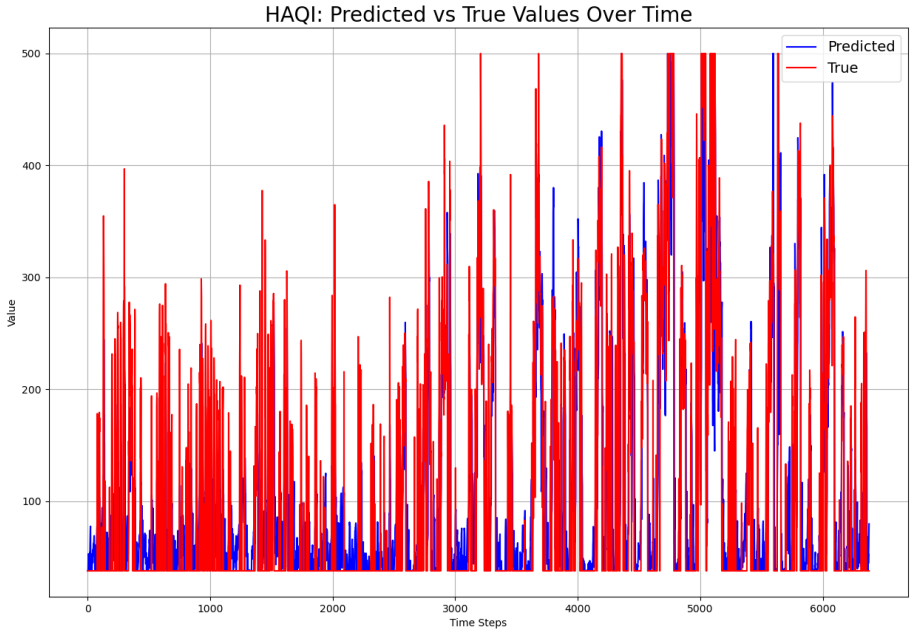


Figure 4-11: Comparison of predicted and actual values of HAQI for 6-hour at Aotizhongxin monitoring station.

4.5 Chapter summary

This chapter proposes a composite model composed of multiple sub-models to achieve high-precision predictions of three air pollution indices (AQI, AAQI and HAQI). This model addresses the limitation of existing air pollution prediction models, which can only predict a single air quality index (AQI) and fail to comprehensively assess air pollution concentration and its associated risks. Furthermore, by integrating multiple neural network models, the proposed approach enhances both prediction accuracy and stability.

Firstly, the correlation between AQI, AAQI and HAQI as well as with each pollutant is analysed, and the results show that the correlation between the three air pollution

indices is strong, above 0.95. The correlation with each pollutant also ranged from 0.4 to 0.85. This lays the foundation for the next prediction. In the prediction of each pollution index, the combined model consists of LSTM, MLP and Transformer, and the model with the best performance is automatically selected. Based on the results it can be seen that LSTM is the best performing model in most cases as the output of the results. In addition, from the results, it can be seen that the model has excellent prediction performance for 1, 3 hours and is able to predict the results of AQI, AAQI and HAQI very well, and the three-evaluation metrics MAE, MSE and R^2 are all at a high level(In terms of averages, both MAE and MSE are less than 1 and R^2 is above 0.5). In the more difficult 6-hour prediction, the prediction performance of AQI and AAQI is also excellent(R^2>0.4), while the prediction of HAQI, which is a more complicated index to calculate, is not satisfactory at some stations, especially R^2, which is poor (less than 0.4), while MAE and MSE are excellent (MAE, MSE less than 1). Overall, the model is successful and the predictions of AQI, AAQI and HAQI for 1, 3 and 6 hours are satisfactory. Since the prediction of multiple pollution indices is proposed for the first time, this study adopts the method of comparing the results with the traditional prediction method (prediction of AQI). The results show that the prediction results of AQI, AAQI and HAQI in this study are better than the prediction results of AQI in the traditional prediction.

Chapter 5. Combined model GCN-LSTM for air pollution prediction

5.1 Introduction

Following the evaluation of the predictive performance for AQI, AAQI, and HAQI in the previous section, this chapter narrows the focus to the prediction of AQI derived from the six major pollutants. As outlined in Chapter 4, the three indices differ substantially in their pollutant-weighting mechanisms. AQI adopts a single-pollutant dominant approach, where the overall index is determined by the maximum sub-index without assigning explicit weights across pollutants. AAQI applies a nonlinear aggregation of sub-indices, allowing all pollutants to contribute to the final score, while HAQI incorporates health-risk–based weighting, linking each pollutant's contribution to its estimated health impacts.

Given these distinctions, AQI is the most suitable index for the present analysis. The objective of this chapter is to assess prediction performance improvements achieved through model enhancements, with each pollutant modelled separately. Consequently, indices that aggregate or weight multiple pollutants—such as AAQI and HAQI—are not required. AQI therefore provides a parsimonious and appropriate basis for examining the effect of the proposed modelling strategies.

Timely and accurate prediction of air pollution is important for government departments to formulate preventive and control measures and for the public to prevent hazards. With the rapid development of artificial intelligence, the prediction of air pollution has been changed from the traditional statistical method to the prediction by neural network model, this has greatly improved the accuracy, speed, and ability to

predict complex conditions in predicting air pollution. This study will further improve the prediction accuracy by improving the method based on neural network prediction.

Changes in air pollution are affected by a variety of parameters, and in addition to meteorological information, changes in spatial location can also have an impact on the air pollution situation. Therefore, this study is based on the LSTM model, which is good at predicting time series data, and adds GCN to extract spatial information, which is used as a parameter to improve the prediction ability. General spatial data requires more process to collect, this study proposes to use the abstract spatial information generated by the similarity of historical data of air pollution instead of the real spatial information, which achieves the step of simplifying the process. The results are then compared with other models in terms of assessment metrics MAE, MSE and R^2 to validate the performance of the model. Overall, the main contributions are as follows:

- Incorporating spatial information as parameters enhances the predictive performance of the model.
- Using GCN to extract the spatial information of each time step, rather than relying on actual spatial data, reduces process complexity, thereby increasing the model's universal applicability.
- The model's performance was verified by combining GCN and LSTM into a composite model and comparing the prediction results for six major pollutants with three prediction models (LSTM, GRU, and LSTM+Attention).

The rest of the chapter is organized as follows. Section 5.2 analyses the correlation between the six pollutants and the pertinent meteorological data, as well as the annual variability of each pollutant. Section 5.3 describes the model parameters, composition, and method of operation, among other things. Section 5.4 presents the results of the experiments and a discussion thereof. A summary of the results of this chapter can be found in section 5.5.

5.2 Analysis of pollution data

The correlation of data plays an important role in predictive modelling, and understanding and exploiting the correlation between data can significantly improve the performance and accuracy of predictive models[229]. The correlation of each pollutant and associated climatic parameters is analysed in Figure 5-1 using heat maps. The lighter the colour, the stronger the correlation. As illustrated in Figure 5-1, $PM_{2.5}$ and PM_{10} have the strongest correlation among all variables, the study by Simran et al.[230] also showed a strong correlation between $PM_{2.5}$ and PM_{10}. Except for O_3, all pollutants show strong correlations with each other.

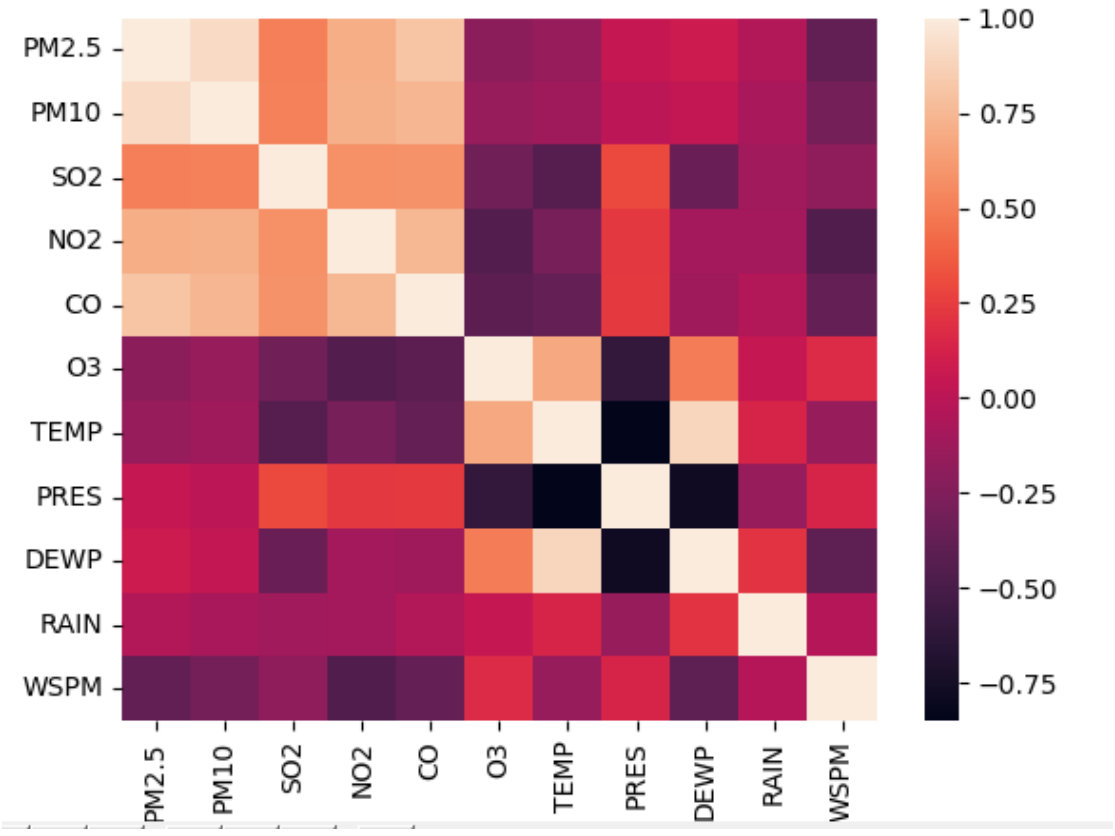
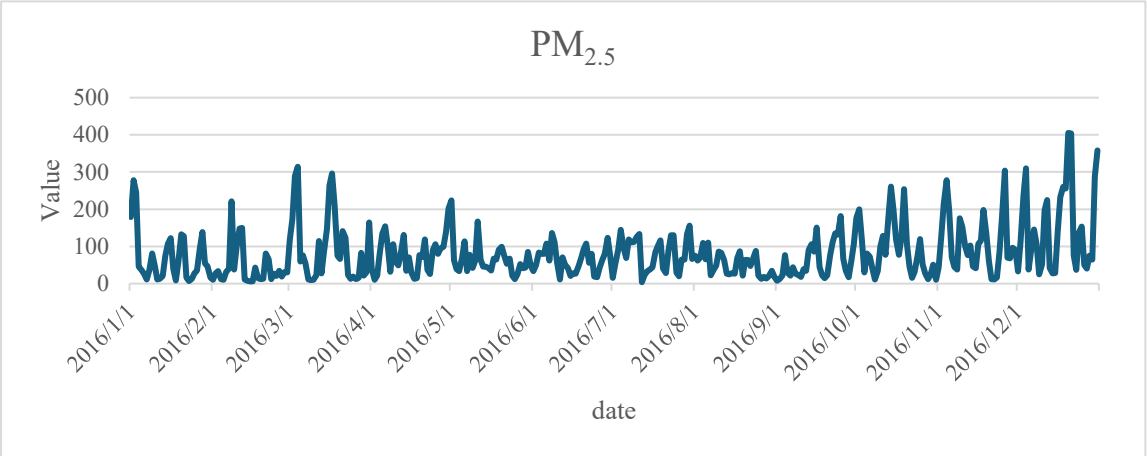


Figure 5-1: Heat map of correlation among each variable.

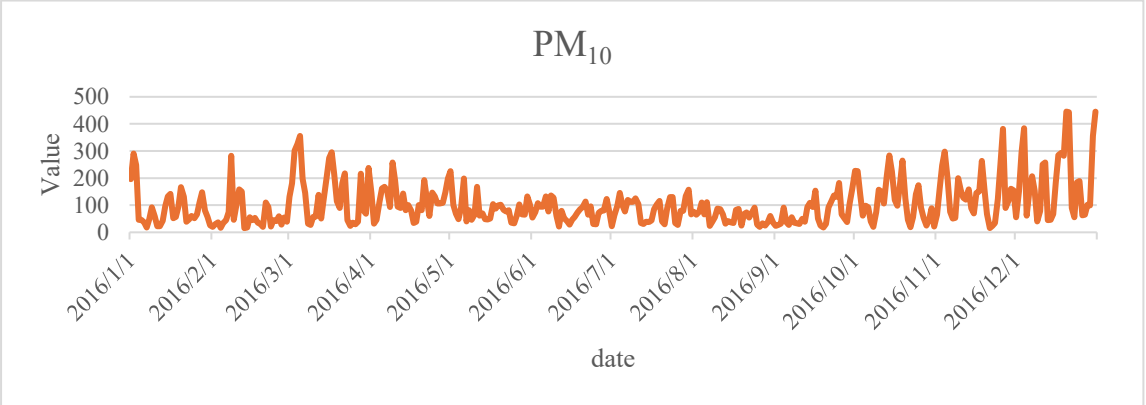
Additionally, there is a noticeable correlation between pollutants and climate conditions, with atmospheric pressure exhibiting a relatively high correlation[231]. O_3 is significantly affected by seasons, leading to a strong correlation with air temperature[232]. As temperature increases, the O_3 pollution index also rises. The observed correlations between pollutants and climate data indicate that the data is predictable.

The annual trends in air pollution also show the correlation between pollutants and the influence of seasonal climate and other factors on pollutant changes[233]. As can be seen from the annual trends for each pollutant in 2016 (Most complete and relatively new data (2016) in the dataset was chosen) in Figure 5-2, all pollutants except O_3 exhibit a positive correlation that is roughly classified into two categories: serious pollution in autumn and winter, and slight pollution in spring and summer[234]. This phenomenon is attributed to Beijing's use of central heating during winter, which involves burning significant amounts of coal and consequently produces elevated levels of pollutants[235]. Particularly, SO_2 and CO showed the most pronounced differences. In contrast, O_3 presents a trend of greater severity during the summer months and less severity during the winter months. The elevated ozone levels observed during the summer are attributed to higher temperatures, which facilitate the catalytic and production processes of O_3[236]. The high correlation among pollutants, excluding O_3, along with the predictable patterns of O_3 fluctuations, create optimal conditions for accurate pollution prediction, thereby ensuring the reliability of pollution forecasting.

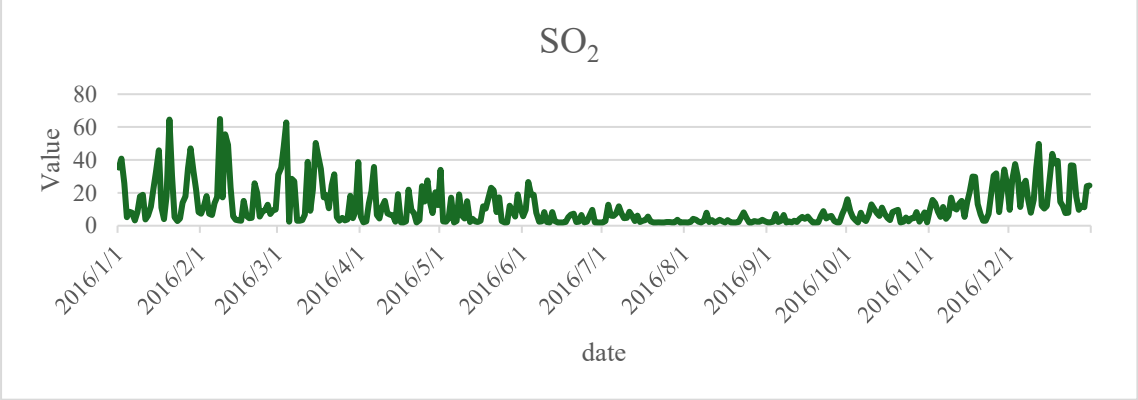
(a)


PM2.5
500
400
300
200
100
0
Value
2016/1/1
2016/2/1
2016/3/1
2016/4/1
2016/5/1
2016/6/1
2016/7/1
2016/8/1
2016/9/1
2016/10/1
2016/11/1
2016/12/1
date

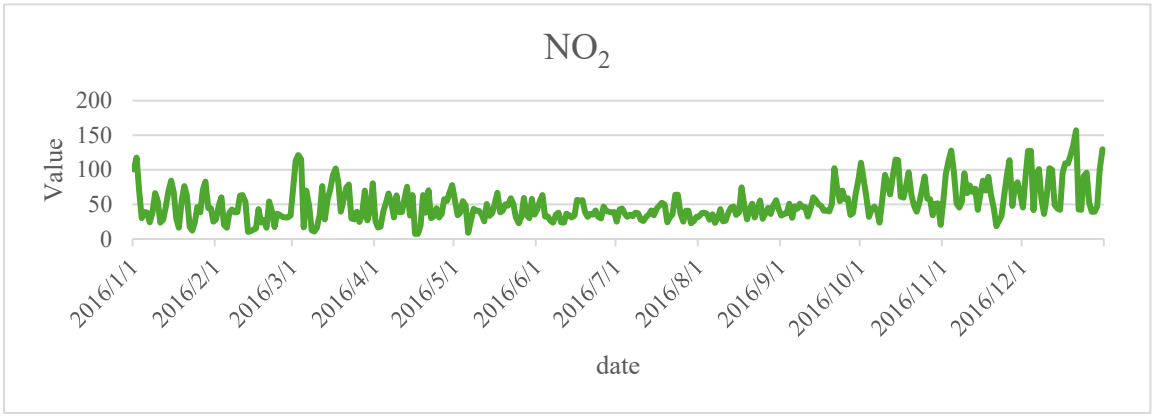
(b)


PM10
500
400
300
200
100
0
Value
2016/1/1
2016/2/1
2016/3/1
2016/4/1
2016/5/1
2016/6/1
2016/7/1
2016/8/1
2016/9/1
2016/10/1
2016/11/1
2016/12/1
date

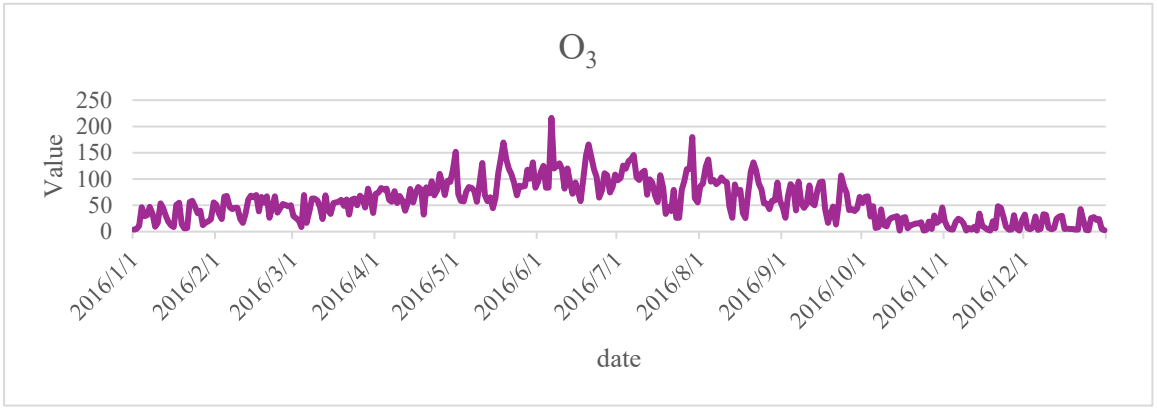
(c)


SO2
80
60
40
20
0
Value
2016/1/1
2016/2/1
2016/3/1
2016/4/1
2016/5/1
2016/6/1
2016/7/1
2016/8/1
2016/9/1
2016/10/1
2016/11/1
2016/12/1
date

(d)



(e)



(f)

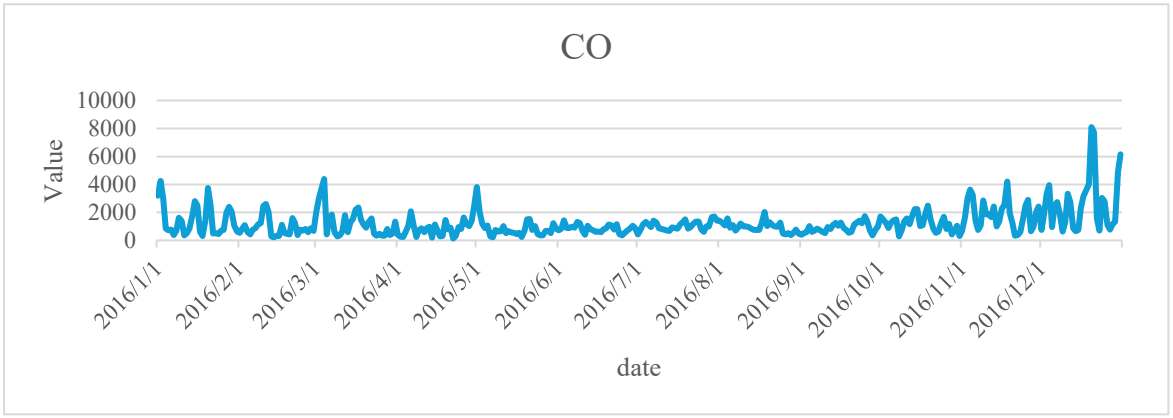


Figure 5-2: Annual change trend of six major pollutants $PM_{2.5}$, PM_{10}, CO, O_3, SO_2, and NO_2 in 2016. The X-axis is the date, and the Y-axis is the AQI value. Since the CO value is too small to show its change, it is enlarged by 1000. (a) $PM_{2.5}$, (b) PM_{10}, (c) SO_2 (d) NO_2 (e) O_3 (f) CO

5.3 Experimental settings

5.3.1 Model operating environment

The experiments were conducted using the Hamilton8 High-Performance Computing system provided by Durham University. This system has 15,616 CPU cores, 36 TB RAM, and 1.9 PB disk space and runs on a Linux operating system. The following software and libraries were used for data pre-processing and modelling: Python (version 3.6) with Matplotlib, NumPy, Pandas, Attention, TensorFlow, Sklearn, Metrics and Tsfresh. These tools provided the necessary computational and analytical processes required for the experiments.

5.3.2 Structure of GCN+LSTM

The structure of the GCN+LSTM model is depicted in Figure 5-3 where the pollution prediction model primarily involves (1) Data Pre-processing; (2) Extraction of Data Spatial Information at Each Time Step Using GCN and Training the Data Using LSTM and (3) Evaluation of Model Performance using MAE, MSE, and R^2.

The air quality dataset used in this study was obtained from UCI machine learning repository, which provides hourly measurements of major pollutants. Before model development, a series of quality control procedures were applied, including the removal of clearly erroneous records (e.g., negative concentrations) and consistency checks across neighbouring stations. Since the raw dataset is recorded at an hourly temporal resolution and contains missing values, the preprocessing workflow begins by aggregating the hourly observations into daily averages. Subsequently, missing values are imputed using the backward-fill (B-fill) method, in which each missing entry is

replaced with the most recent subsequent non-missing observation to ensure temporal continuity [237]. The pre-processed data is imported into the GCN model to extract the spatial structure information (Similarity of historical data) between the data points. The extracted spatial information at each time step is then output to the LSTM model. In the LSTM model, the historical data of seven days is used to predict the data for the next day. The parameters for the model are shown in Table 5-1. The number of epochs is set to 500, with a patience value of 25 for early stopping to prevent overfitting[238].

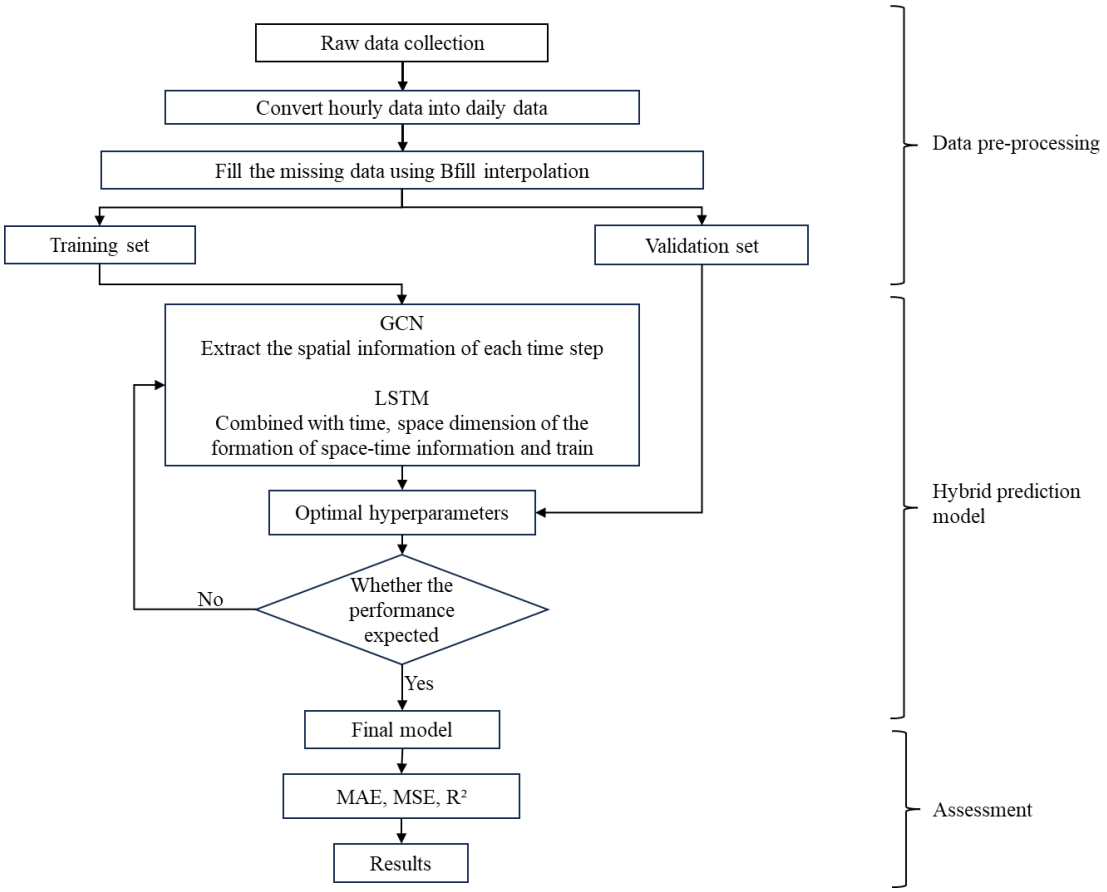


Figure 5-3: The structure of the GCN+LSTM model.

The training data is then exported and compared with the data from the validation set. The comparison between the training set and the validation set facilitates the optimization of model parameters and provides a robust evaluation of the model's fitting performance. MAE, MSE, and R^2 are used as evaluation metrics to obtain the results[206].

Table 5-1: The parameters of the models.

Parameter	Value
number of records	17532
time interval(day)	1
training set	80%
test set	20%
prediction length	1
history length	7
training epochs	500
early stopping patience	25

For Evaluation, the model will predict the concentrations of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3. To comprehensively validate the performance of the proposed model, this study selected LSTM, GRU, and LSTM-Attention as comparison models, considering three distinct perspectives. First, LSTM, as a widely used time-series model, serves as a baseline to evaluate fundamental performance[23]. Second, GRU, a simplified version of LSTM, offers a comparison to assess the trade-off between model complexity and effectiveness[239]. Finally, LSTM-Attention, an enhanced model with attention mechanisms, is included to examine whether the proposed model can achieve superior results under more complex frameworks[240]. This comparison will help evaluate the performance of the GCN+LSTM model.

5.4 Results and discussion

Table 5-2, Table 5-3, Table 5-4 compare the MSE, MAE, and R^2 indices of the experimental GCN+LSTM model and three popular models: LSTM, GRU, and LSTM+Attention. The optimal metrics for each pollutant are indicated in bold. The data demonstrate that the GCN+LSTM model outperforms the comparison models for each pollutant.

For $PM_{2.5}$ prediction (Table 5-2), the GCN+LSTM model achieves an MSE of 35.43, which is significantly lower than that of the other models[241], [242].

Table 5-2: The comparison of MSE, MAE and R^2 of GCN+LSTM, GRU, LSTM and LSTM+ATTENTION for $PM_{2.5}$ and PM_{10}.

Pollutants	Model	MSE	MAE	R^2
$PM_{2.5}$	GCN+LSTM	**35.4319**	**22.524**	**0.7551**
	GRU	38.3313	24.1133	0.7134
	LSTM	41.0988	26.0669	0.6706
	LSTM+attention	37.0409	23.3606	0.7324
PM_{10}	GCN+LSTM	**44.0816**	**28.4466**	**0.679**
	GRU	45.0773	29.0929	0.6643
	LSTM	52.8607	33.0108	0.5384
	LSTM+attention	44.3411	28.8996	0.6752

The MAE and R^2 values also indicate superior performance, with the GCN+LSTM model showing the best fit. LSTM+Attention as a comparison model, MSE is 1.5 higher than GCN+LSTM, MAE is 0.8 higher, and R^2 is worse by 0.02. The gap is smaller between the hybrid models and larger between GCN+LSTM and the single models GRU and LSTM. This suggests that the hybrid model performs better than the single

models in predicting $PM_{2.5}$ concentrations and the prediction performance of GCN+LSTM model proposed in this study is the best. For PM_{10} (Table 5-2), the GCN+LSTM model again outperforms the other models, with MSE, MAE, and R^2 values of 44.08, 28.45, and 0.68, respectively. The combined model demonstrates superior performance compared to single models, with LSTM showing the least effective outcome. Ban et al.[243] also expressed the same view in his research on composite models. Composite models can give full play to the advantages of each model, improve predictive accuracy and realize the prediction of more complex data.

In predicting SO_2 (Table 5-3), the GCN+LSTM model continues to show superior performance compared to other models. MSE, MAE and R^2 reached 6.3252, 3.6076 and 0.6584 respectively. The LSTM model still demonstrates the least effective outcome, the predictive performance of a single model is often not as good as that of a hybrid model[244]. Moreover, in the prediction of SO_2, the prediction performance of the combined model LSTM+Attention is lower than that of GRU, in particular, the R^2 of GRU is higher than 0.6, while the R^2 of LSTM+Attention is 0.542. The fitting of the predicted and actual values of SO_2 was the least optimal among all pollutants, likely due to the irregular nature of SO_2 emissions[245]. In the prediction of SO_2, GCN+LSTM still shows the best performance, but the single model GRU has the second-best performance in this prediction after GCN+LSTM. This shows that the performance of different models for different pollutants is different. Finally, to predict NO_2 (Table 5-3), the GCN+LSTM model achieves MSE, MAE, and R^2 values of 12.44, 9.16, and 0.81, respectively. These values are all better than those of the comparison models (LSTM: 13.5829, 10.1648, 0.7777; GRU: 13.0708, 9.6554, 0.7941; LSTM+Attention: 12.8088, 9.7122, 0.8023), and the degree of fit is the highest among all pollutants. This indicates that the predicted values are closest to the actual values. On the other hand, the results of GRU and LSTM+Attention are not similar, the MAE of GRU is better than that of LSTM+Attention, while MSE and R^2 are worse than that

of LSTM+Attention by a very small margin. Overall, it is still the combined model that outperforms the single model[16] [20].

Table 5-3: The comparison of MSE, MAE and R^2 of GCN+LSTM, GRU, LSTM and LSTM+ATTENTION for SO_2 and NO_2.

Pollutants	Model	MSE	MAE	R^2
SO_2	GCN+LSTM	**6.3252**	**3.6076**	**0.6584**
	GRU	6.8055	3.8639	0.6046
	LSTM	7.4091	4.2442	0.5313
	LSTM+Attention	7.324	4.1615	0.542
NO_2	GCN+LSTM	**12.4429**	**9.1611**	**0.8134**
	GRU	13.0708	9.6554	0.7941
	LSTM	13.5829	10.1648	0.7777
	LSTM+Attention	12.8088	9.7122	0.8023

Overall, the results indicate that the GCN+LSTM model consistently provides better predictive performance than the single and other combined models across all pollutants[244]. This highlights the effectiveness of integrating spatial information through GCN and temporal information through LSTM for air pollution prediction [246]. Furthermore, the model's performance on CO (Table 5-4) data further supports the superiority of the GCN+LSTM model over the comparison models. MSE, MAE and R^2 are 0.4961, 0.2967 and 0.7935 respectively. Specifically, the GCN+LSTM model outperforms the other models across all metrics while the GRU model demonstrates better results than the LSTM+Attention and LSTM models in terms of MSE, MAE, and R^2. Due to the low concentration of CO in the air, both MSE and MAE values are relatively low. The prediction of O_3 (Table 5-4) also highlights the excellent performance of the GCN+LSTM model, which saw significant improvements in all

three evaluation metrics. MSE, MAE and R^2 are 18.8972, 12.9872 and 0.7862 respectively. The MSE decreased by nearly 2 units, and the R^2 increased by 2%. In turn, the other three comparison models (LSTM, GRU, and LSTM+Attention) showed similar performance without any significant variations across the evaluation metrics. The composite model LSTM+Attention is also worse than GRU in terms of MAE, and the results of these two models are not very different. Adriana[247] predicted ozone concentrations at different times using the MLP model. The results verified the complexity of ozone prediction. The average R^2 value of the prediction results was only 0.6, which was lower than the results of this study. The average MAE value was 13, which was also lower than the results of this study. The research of Ayman[248] and Pan[249] also proved that the prediction results of this study have significant advantages.

Table 5-4: The comparison of MSE, MAE and R^2 of GCN+LSTM, GRU, LSTM and LSTM+ATTENTION for CO and O_3.

Pollutants	Model	MSE	MAE	R^2
CO	GCN+LSTM	**0.4961**	**0.2967**	**0.7935**
	GRU	0.5965	0.321	0.7014
	LSTM	0.6116	0.3393	0.6861
	LSTM+Attention	0.614	0.349	0.6836
O_3	GCN+LSTM	**18.8972**	**12.9872**	**0.7862**
	GRU	20.4123	14.0218	0.7506
	LSTM	20.4306	14.4466	0.7501
	LSTM+Attention	20.2362	14.6011	0.7548

These results show that the combined GCN+LSTM model consistently performs better than single models (LSTM, GRU) and other combined models (LSTM+Attention). The combined model LSTM+Attention and the single model GRU show different prediction

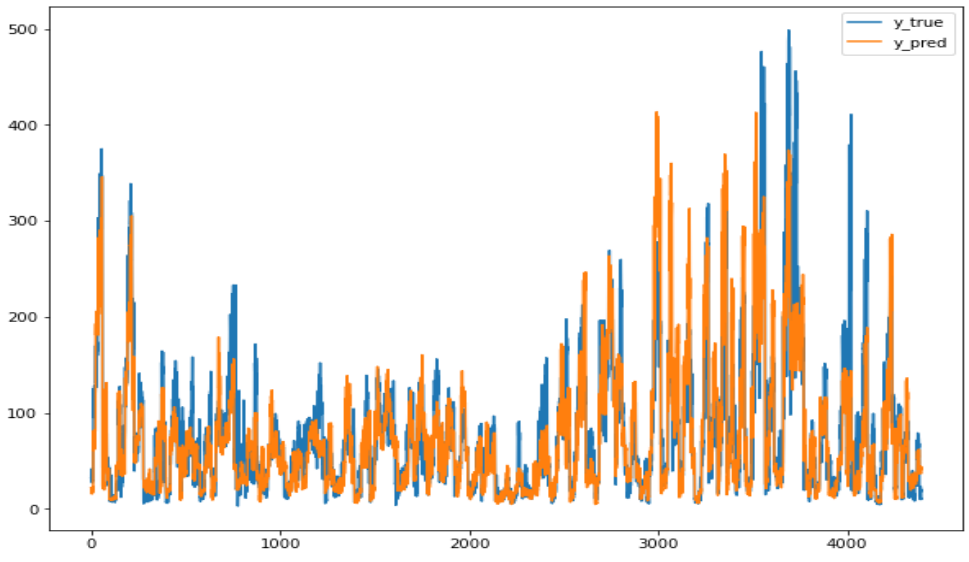
performance for different pollutants. GCN+LSTM shows stable prediction performance for all pollutants. Compared with the prediction of single pollutants in other prediction models[249], [250], the prediction ability of this model is more comprehensive. The integration of GCN for spatial information extraction and LSTM for temporal information processing allows the model to leverage the strengths of both approaches, resulting in improved predictive performance. A similar approach by Qi et al. [27] used the GCN+LSTM model for $PM_{2.5}$ prediction and reported an MAE of 23.10 for 24-hour forecasts. However, Qi used real spatial data as parameters, and the process was more complicated. In contrast, the model in this study achieved a lower MAE of 22.524, demonstrating superior performance by extracting spatial structures more effectively.

The results clearly indicate that incorporating spatial data enhances the accuracy of pollution forecasting. The GCN component efficiently captures spatial relationships, while the LSTM component handles temporal dynamics, leading to more accurate and reliable predictions[25] [26].

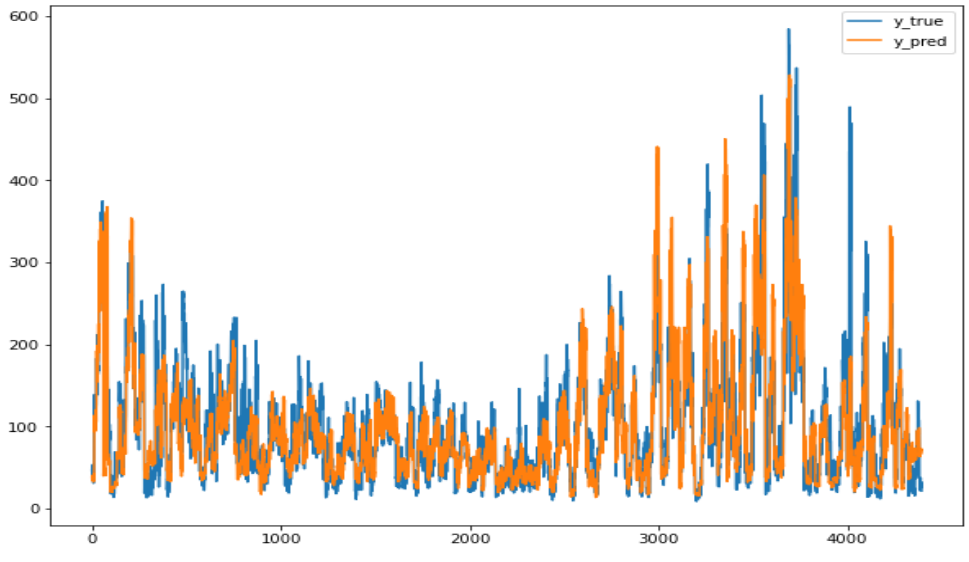
Figure 5-4 shows the fitted graphs of the predicted and actual values for the six pollutants using the GCN+LSTM model. The blue line represents the actual values, while the orange line depicts the predicted values. The fit between the predicted value and the true value can intuitively show the stability of the prediction results[252]. The close alignment between the two lines indicates that the model accurately fits the real values, particularly when pollution concentrations are stable.

The analysis conducted shows that the GCN+LSTM model exhibits a robust ability to predict various pollutants, outperforming other models in terms of MSE, MAE, and R^2 metrics. The combination of spatial and temporal data processing significantly enhances prediction accuracy, making this approach highly effective for air pollution forecasting. This model, which integrates the strengths of GCN and LSTM, significantly improves air pollution prediction accuracy, thereby enhancing

environmental monitoring and informing policymaking. The findings underline the importance of spatial information in improving the precision of pollution forecasts, ultimately contributing to better air quality management and public health outcomes.

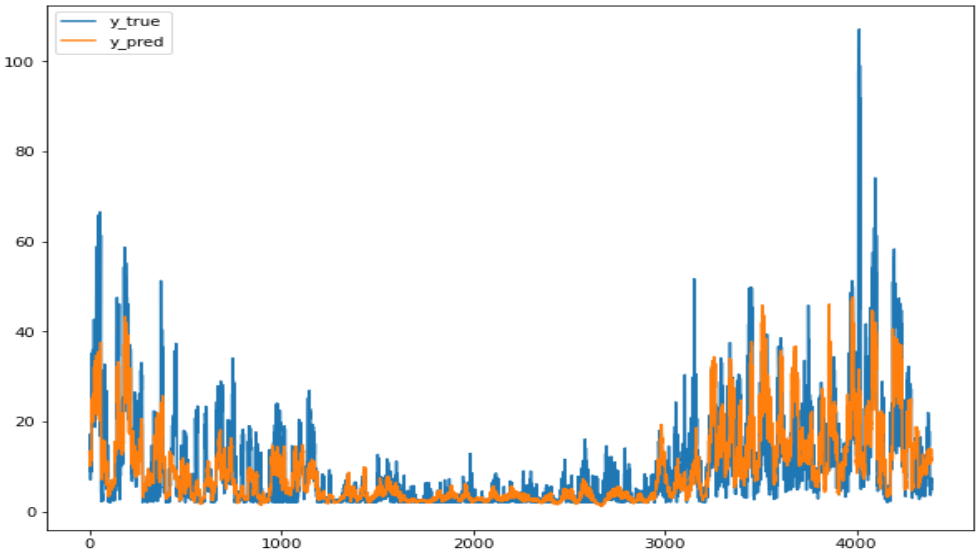


(a)

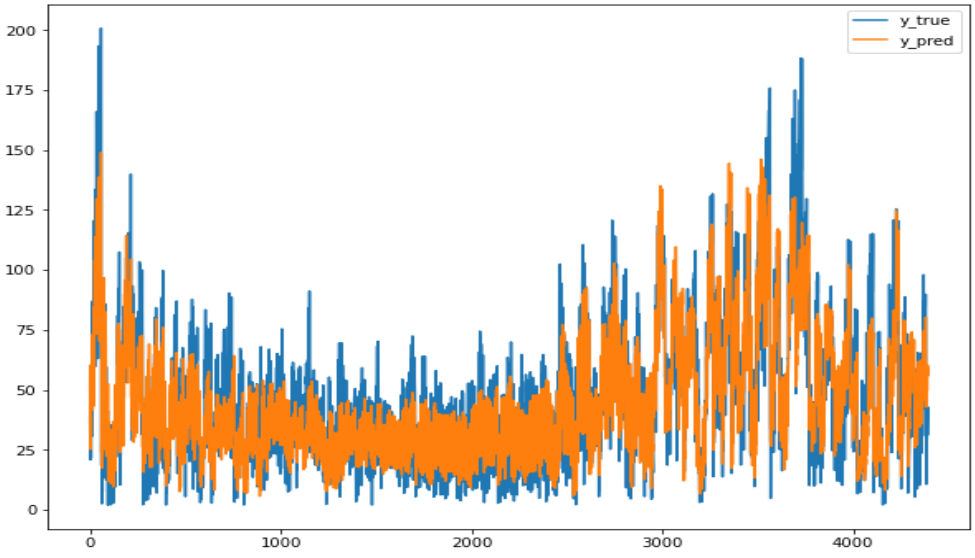


(b)

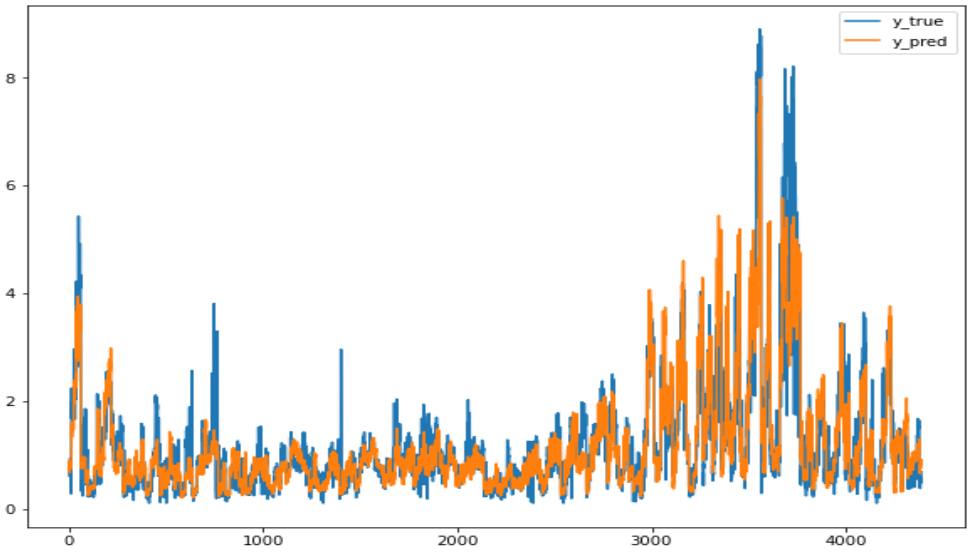
(c)

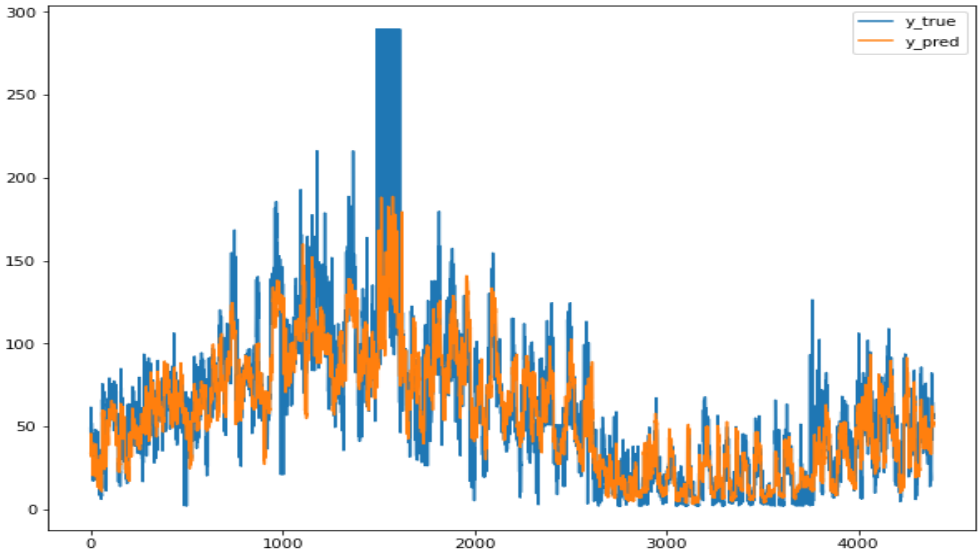

y_true
y_pred
100
80
60
40
20
0
0
1000
2000
3000
4000

(d)


y_true
y_pred
200
175
150
125
100
75
50
25
0
0
1000
2000
3000
4000

(e)


y_true
y_pred
8
6
4
2
0
0
1000
2000
3000
4000



(f)

Figure 5-4: The Comparison of predicted value and real value of GCN+LSTM model. (a), (b) are $PM_{2.5}$ and PM_{10}, (c), (d) are SO_2 and NO_2, (e), (f) are CO and O_3. The blue line is the true value, and the orange line is the predicted value.

The results of this study show that the proposed approach, which extracts spatial information based on historical data correlations, outperforms traditional methods that rely on explicit spatial inputs. Not only does it achieve higher prediction accuracy, but it also eliminates the need for additional procedures to collect spatial data. Consequently, the model exhibits strong transferability and can be readily applied to other geographical locations. Nonetheless, if substantial discrepancies exist among monitoring stations in a new region, some degree of parameter fine-tuning may be necessary to adapt the model to local conditions.

5.5 Chapter summary

Most of prediction models, despite their high accuracy, often neglect spatial information,

resulting in predictions that lack the influence of spatial considerations. Furthermore, existing models that incorporate spatial structures generally assume that climatic and other conditions at nearby stations are similar, which limits their applicability to stations that cannot be directly mapped. This model addresses these limitations by leveraging the similarity of historical data to incorporate spatial information, thereby avoiding the need for direct geospatial calculations of site locations. This approach enhances prediction accuracy while maintaining convenience and simplicity in including spatial information. The annual changes and correlation analysis of various pollutants indicated conductive conditions for the prediction of each pollutant. By comparing the new model against three high-precision models—LSTM, LSTM+Attention, and GRU—results demonstrated this model consistently produced improved prediction results across all pollutants, achieving the best performance metrics.

Chapter 6. The impact of the COVID-19 pandemic on air pollution and human health

6.1 Introduction

The first two chapters of this thesis focus on constructing a modelling framework capable of simultaneously predicting multiple air quality indices (AQI, AAQI, and HAQI) and enhancing AQI prediction for six conventional pollutants ($PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3) by extracting abstract spatial features from historical monitoring data. These chapters centre on methodological innovation, aiming to provide more accurate technical support for real-time air quality warnings and environmental management.

Building on this foundation, Chapter 3 shifts the research perspective from *prediction* to *retrospective assessment* by systematically analysing the evolution of air pollution across 13 cities in the Beijing–Tianjin–Hebei region from 2014 to 2023. The focus moves from how to predict pollution more accurately to understanding how pollution has actually changed over the past decade, what factors drive these changes, and how external shocks—particularly the COVID-19 pandemic—have influenced long-term governance outcomes. This transition from model development to empirical evaluation allows the practical relevance of the prediction framework to be examined and contextualised within real historical scenarios.

This chapter deepens and broadens the overall objectives of the thesis by introducing new analytical dimensions, including shifts in dominant pollutant frequencies, annual and seasonal AQI trends, ArcGIS-based visualisation of inter-city spatial disparities, and health-impact–oriented assessment using AirQ+ to estimate attributable

proportions (AP) of pollution-related diseases such as respiratory conditions. While the chapter 4 and 5 necessarily focused on the six standard pollutants for modelling purposes, the present chapter adopts a more holistic, health-oriented evaluation perspective. By translating changes in pollutant concentrations into quantifiable public health burdens, it aligns more directly with the ultimate goal of air quality management: safeguarding population health[174].

A particularly important contribution of this chapter is the incorporation of a complete pre-, mid-, and post-COVID-19 time series extending through 2023, which remains scarce in existing literature. Most studies examining the pandemic's impact on air quality conclude with data from 2021 or 2022. By extending the analysis into the post-pandemic period, this chapter provides new evidence on whether the temporary emission reductions during COVID-19 resulted in sustained improvements or whether a rebound occurred during economic recovery[137]. Furthermore, by statistically isolating the pandemic as an exogenous shock, the analysis offers a more accurate evaluation of the actual effectiveness of long-standing regional air pollution control policies—such as the Air Pollution Prevention and Control Action Plan—thereby providing a more reliable basis for future policy optimisation[143]. Overall, the main contributions are as follows:

- This study provides a comprehensive analysis of changes in six major pollutants across the Beijing-Tianjin-Hebei region over a ten-year period, offering one of the most extensive temporal and spatial insights into regional air quality trends to date.
- The study establishes a direct link between long-term air quality changes and health-related impacts over a decade, highlighting the broader implications of air pollution control policies on public health.
- By incorporating post-COVID-19 air pollution data, this study rigorously evaluates the pandemic's impact on air quality and demonstrates the tangible improvements

achieved through sustained local pollution control efforts following the pandemic, bridging a critical gap in understanding long-term policy effectiveness.

This chapter not only verifies whether the environmental context underlying the earlier prediction models is stable and whether long-term trends can be reliably captured, but also substantially enhances the practical relevance and scholarly value of the thesis by incorporating health-attribution analysis and the assessment of pandemic-related disruptions. Together, these components form a coherent analytical arc—moving from methodological development to empirical validation, and finally to impact evaluation. In doing so, the thesis establishes a complete framework that fulfils its central objective: advancing air quality prediction capabilities to better support evidence-based decision-making and the protection of public health.

The chapter is organized as follows. Section 6.2 describes the experimental procedure and design. Section 6.3 describes the results of the analysis of the changes in pollution in the Beijing-Tianjin-Hebei region. Section 6.4 describes the results of the analysis of the impacts of air pollution on health and related illnesses. Section 6.5 summarizes the chapter.

6.2 Experimental Procedure and Design

As illustrated in Figure 6-1, this study first collected data on air pollution trends in 13 cities across the Beijing-Tianjin-Hebei region from 2014 to 2023. The air quality dataset was obtained from https://aqicn.org/data-platform/register/, which provides publicly accessible and routinely calibrated measurements of major pollutants. The six main pollutants are $PM_{2.5}$, PM_{10}, CO, O_3, SO_2 and NO_2. The data were recorded on a daily basis. To ensure the consistency and reliability of the dataset, missing observations were imputed using linear interpolation, while anomalous values were identified and

removed during the preprocessing procedure. The relatively complete data covering 13 regions over ten years, after preprocessing, met the requirements for subsequent data analysis. Using Python and Word software, the frequencies of six major air pollutants were analyzed, and frequency variation graphs for these pollutants were generated. Subsequently, quarterly changes in air pollutant concentrations were visualized and analyzed. In Python, NumPy and Pandas are primarily used for data processing, including filtering and statistical analysis, while Matplotlib and the basic graphing functions in Word are utilized for creating the corresponding figures.

ArcGIS was then employed to create change graphs and examine the annual variations in air pollutant levels across the 13 cities. The air pollution concentration in each city is represented using a gradient that transitions from green to red, indicating air quality levels ranging from excellent to poor. A uniform standard is applied to generate maps that illustrate the changes and distribution of air pollution across cities over the years. Finally, the AirQ+ software was used to generate statistical graphs depicting changes in mortality, as well as respiratory and circulatory health impacts, associated with variations in air pollutant levels, which were further analyzed in detail. Using the software, one can calculate the disease burden attributable to air pollution, including premature deaths and disease incidence, by processing air pollution data such as concentrations of $PM_{2.5}$, PM_{10}, NO_2, O_3, and other pollutants.

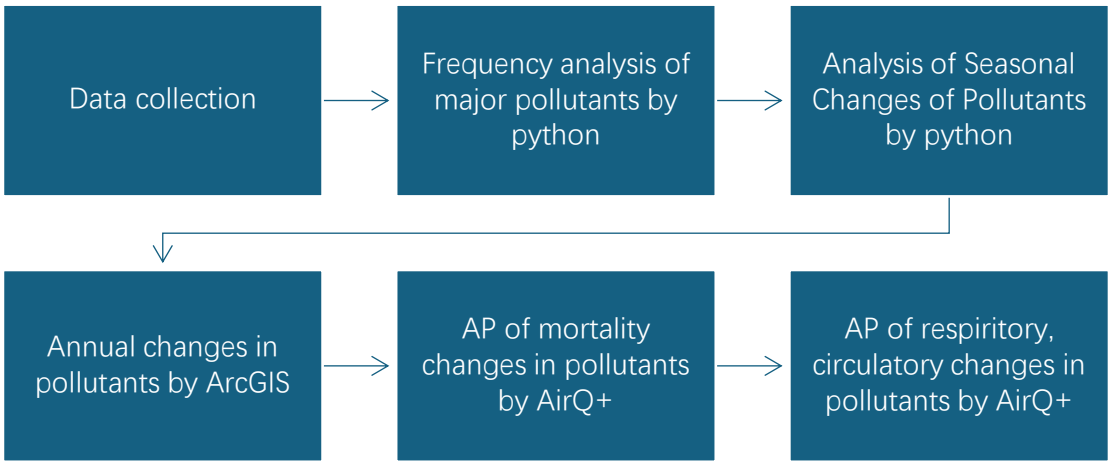


Figure 6-1: Experimental process of changes in air pollution and related health impacts in the Beijing-Tianjin-Hebei region from 2014 to 2023.

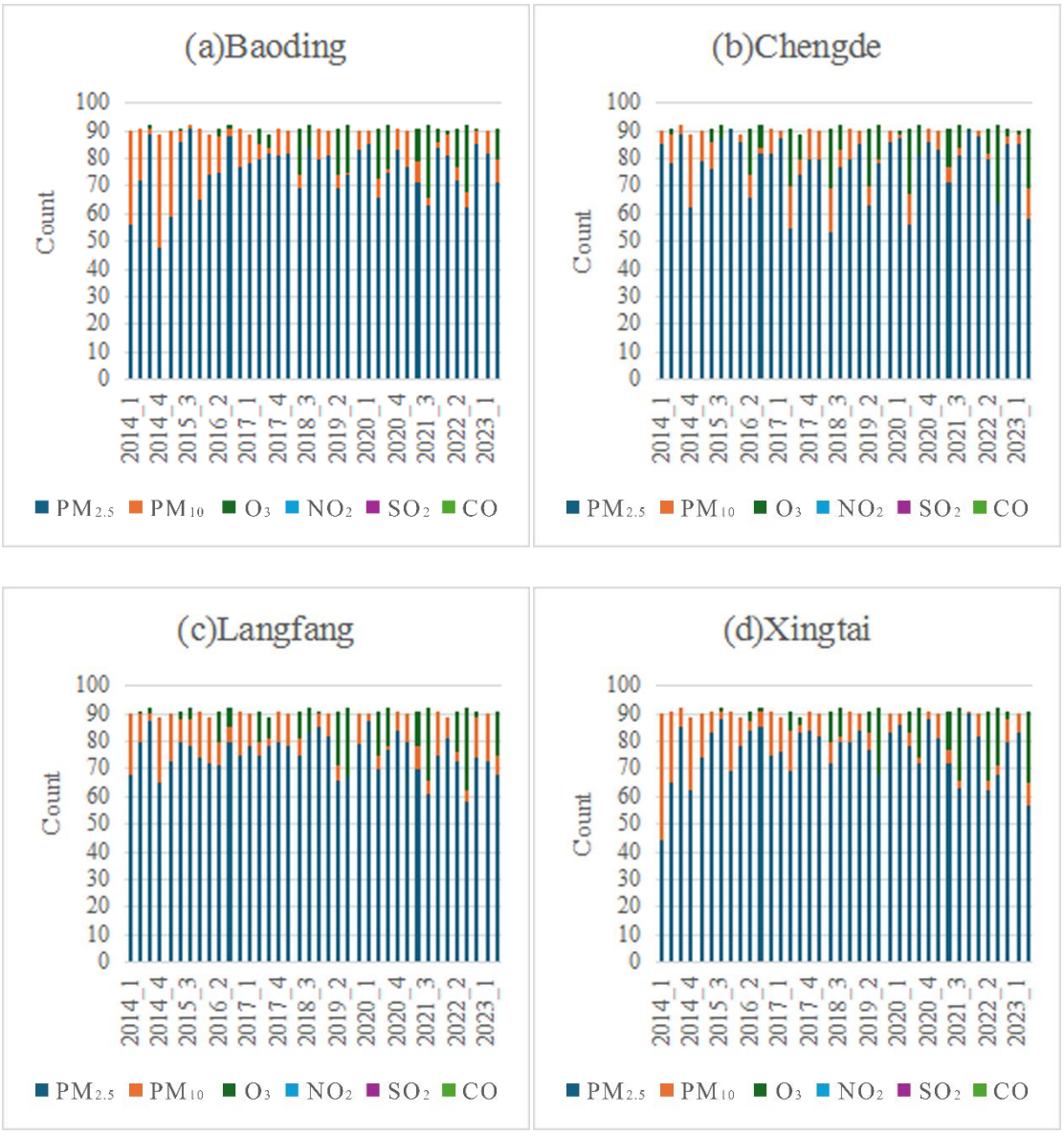
6.3 Results and discussion

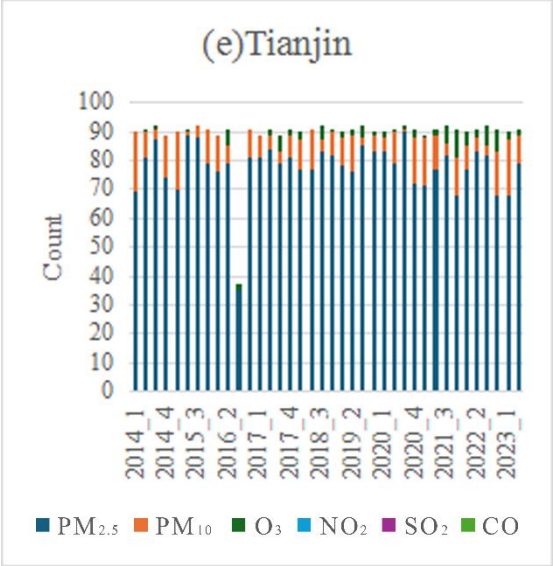
This section analyses the various dimensions of decadal air pollution change in the BTH region. The analysis encompasses the pre- and post-outbreak periods, including the closure and control measures implemented in response to the onset of the pandemic. It aims to provide a comprehensive assessment of the impact of the pandemic on air quality scenarios. Furthermore, the health-related impacts of air pollution will be examined.

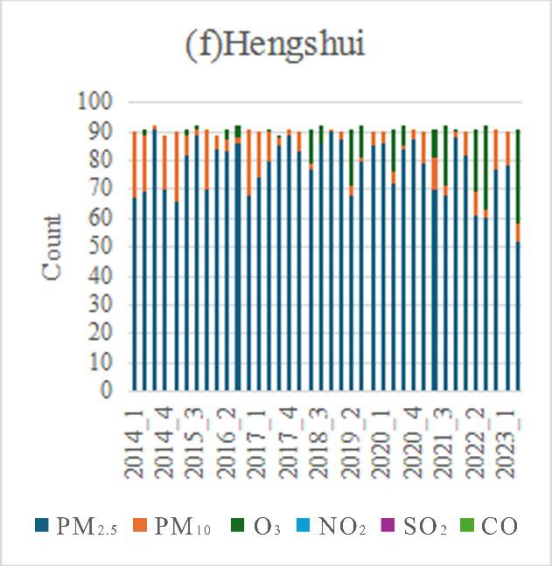
6.3.1 Frequency of primary pollutants

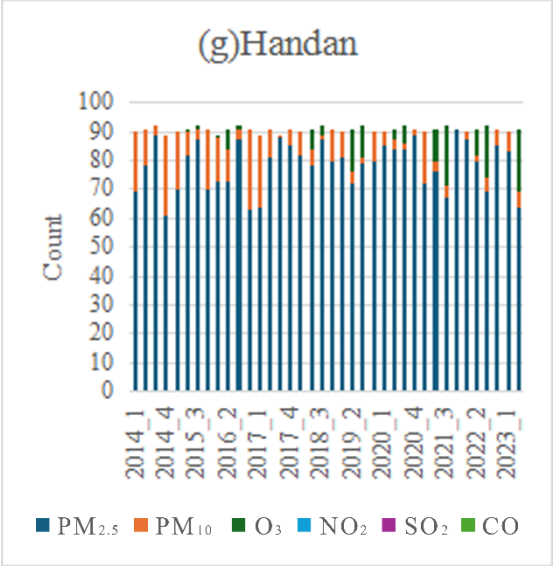
The study encompasses six air pollutants with a significant impact on human health. In

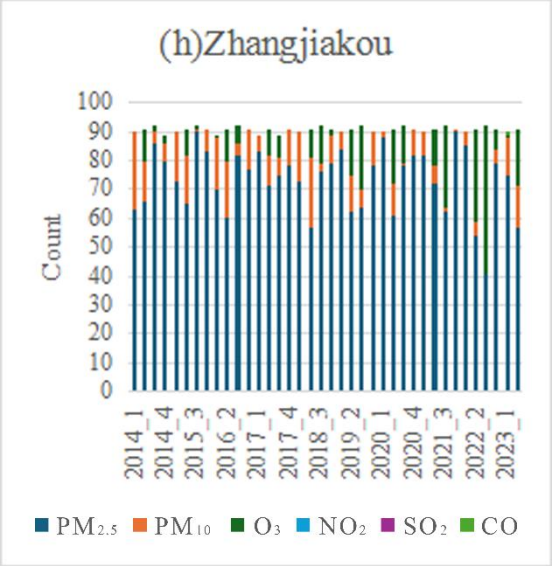
accordance with the WHO and Chinese local regulations [214], the pollutant with the highest individual pollution index is designated as the primary pollutant, and the primary pollutant directly determines the AQI value[253]. Figure 6-2 shows a series of stacked bar charts, delineating the frequency of primary pollutants in the atmosphere of 13 cities within the BTH region between the years 2014 to 2023. The figures illustrate the changes in primary air pollutants over the ten-year period of the local government's air pollution control initiative.

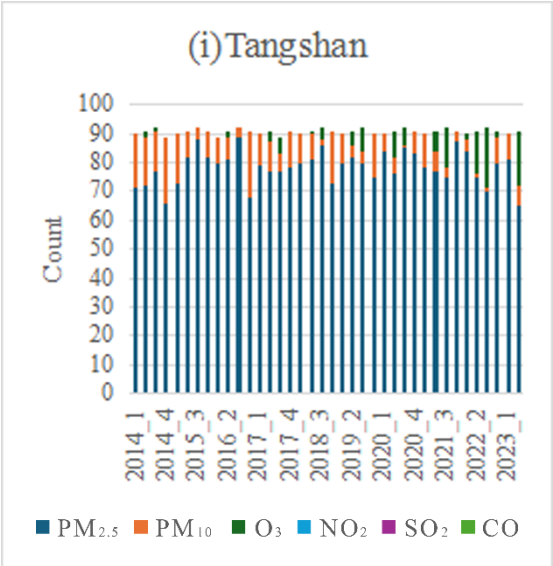


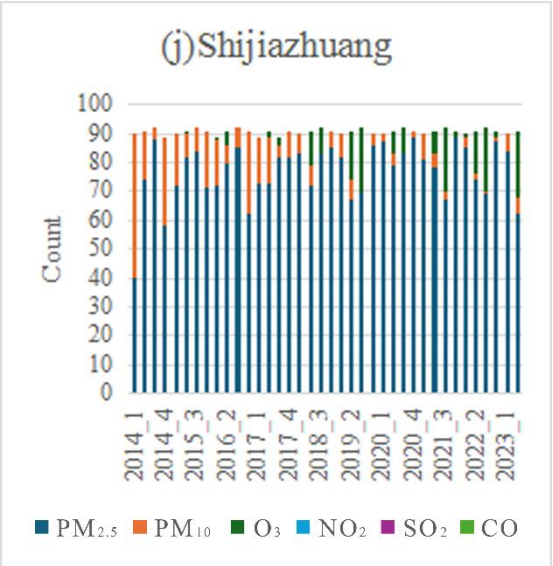

(e)Tianjin
Count
100
90
80
70
60
50
40
30
20
10
0
2014_1
2014_4
2015_3
2016_2
2017_1
2017_4
2018_3
2019_2
2020_1
2020_4
2021_3
2022_2
2023_1
PM2.5
PM10
O3
NO2
SO2
CO


(f)Hengshui
Count
100
90
80
70
60
50
40
30
20
10
0
2014_1
2014_4
2015_3
2016_2
2017_1
2017_4
2018_3
2019_2
2020_1
2020_4
2021_3
2022_2
2023_1
PM2.5
PM10
O3
NO2
SO2
CO


(g)Handan
Count
100
90
80
70
60
50
40
30
20
10
0
2014_1
2014_4
2015_3
2016_2
2017_1
2017_4
2018_3
2019_2
2020_1
2020_4
2021_3
2022_2
2023_1
PM2.5
PM10
O3
NO2
SO2
CO


(h)Zhangjiakou
Count
100
90
80
70
60
50
40
30
20
10
0
2014_1
2014_4
2015_3
2016_2
2017_1
2017_4
2018_3
2019_2
2020_1
2020_4
2021_3
2022_2
2023_1
PM2.5
PM10
O3
NO2
SO2
CO


(i)Tangshan
Count
100
90
80
70
60
50
40
30
20
10
0
2014_1
2014_4
2015_3
2016_2
2017_1
2017_4
2018_3
2019_2
2020_1
2020_4
2021_3
2022_2
2023_1
PM2.5
PM10
O3
NO2
SO2
CO


(j)Shijiazhuang
Count
100
90
80
70
60
50
40
30
20
10
0
2014_1
2014_4
2015_3
2016_2
2017_1
2017_4
2018_3
2019_2
2020_1
2020_4
2021_3
2022_2
2023_1
PM2.5
PM10
O3
NO2
SO2
CO

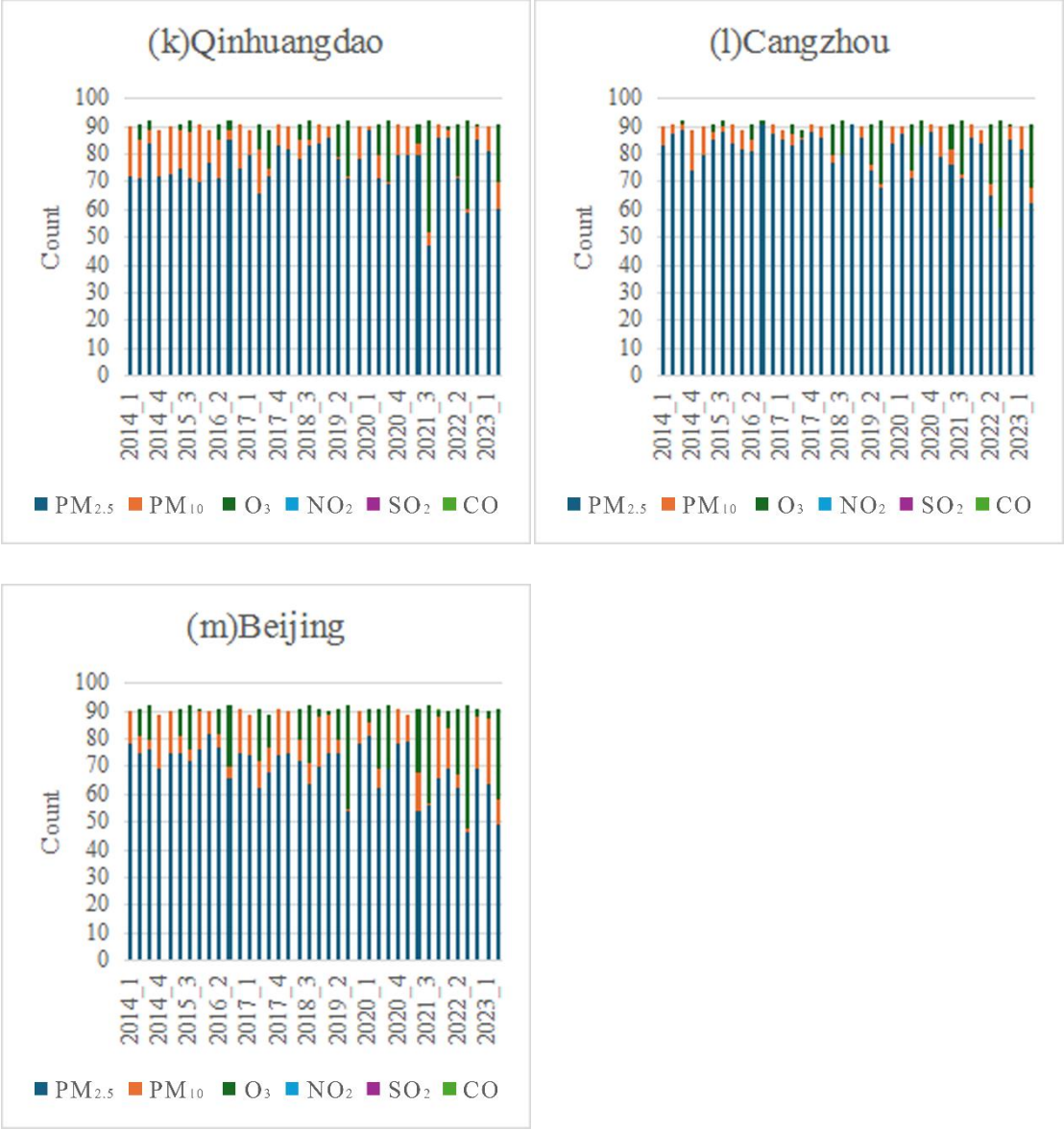


Figure 6-2(a)-(m): Frequency (Occurrences) of each pollutant as a primary pollutant per quarter from 2014 to 2023 for 13 cities in the BTH region.

Overall, $PM_{2.5}$ has the highest frequency as a primary pollutant in every quarter and in every city, and thus dominates the pollution impact, occurring more frequently than all other pollutants combined. This aligns with Xiao's study[254], which concludes that $PM_{2.5}$ has the highest average proportion among air pollutants. Therefore, $PM_{2.5}$ is the most critical pollutant in pollution control. PM_{10} is the second most frequent pollutant in each quarter, and is also higher than all other pollutants overall. In contrast, O_3 is

slightly less frequent than PM_{10} overall, and occurs mainly in the 2nd and 3rd quarters (April to September), as O_3 is more affected by the air temperature[255]. The remaining pollutants, namely NO_2, SO_2 and CO exhibit an almost negligible frequency, with the exception of Zhangjiakou, where CO is identified as the primary pollutant during the initial quarter of 2023. This is because the production of these three gaseous pollutants is frequently accompanied by the production of particulate matter, which is more harmful when produced with greater frequency. In the trend of single pollutants, $PM_{2.5}$ is undoubtedly the primary pollutant. However, the frequency of PM_{10} and O_3 as the primary pollutants is not stable. At times, PM_{10} has a high frequency, while at sometimes, O_3 is much higher than PM_{10}.

From a temporal perspective, the frequency of PM_{10} as a primary pollutant demonstrates a decreasing trend in most cities. This may be attributed to the fact that the reduction in PM_{10} is more pronounced with the implementation of air pollution control measures[10]. The aforementioned trend is particularly evident in the cities of Shijiazhuang, Xingtai, Baoding, and Handan, where the frequency of PM_{10} as a primary pollutant has exhibited a discernible decline since 2014. In contrast, ozone exhibited a markedly different trend, with a relatively limited occurrence as a primary pollutant between 2014 and 2017. In Beijing, however, it was used as a primary pollutant with a frequency comparable to that of PM_{10}. In 2018 and subsequent years, the frequency of O_3 as a primary pollutant increased rapidly, reaching a level similar to or even higher than that of PM_{10}. This is attributed to the fact that O_3 is primarily influenced by temperature, with its concentration exhibiting minimal fluctuations over the ten-year period[91]. Therefore, following an extended period of intervention, the impact of pollution control in the BTH region became discernible from 2018 onwards. This resulted in the concentration of the more stable O_3 becoming the primary pollutant, exhibiting a higher frequency. Although the frequency of ozone, as the primary pollutant in Beijing did not show a significant change compared to other cities, it demonstrates a gradual increase. This is attributable to the fact that Beijing has fewer

heavy polluting industries, resulting in relatively superior overall air quality compared to other cities that are in the initial stages of implementing air pollution control measures[256]. As pollution levels gradually improve, the overall air quality is also expected to improve.

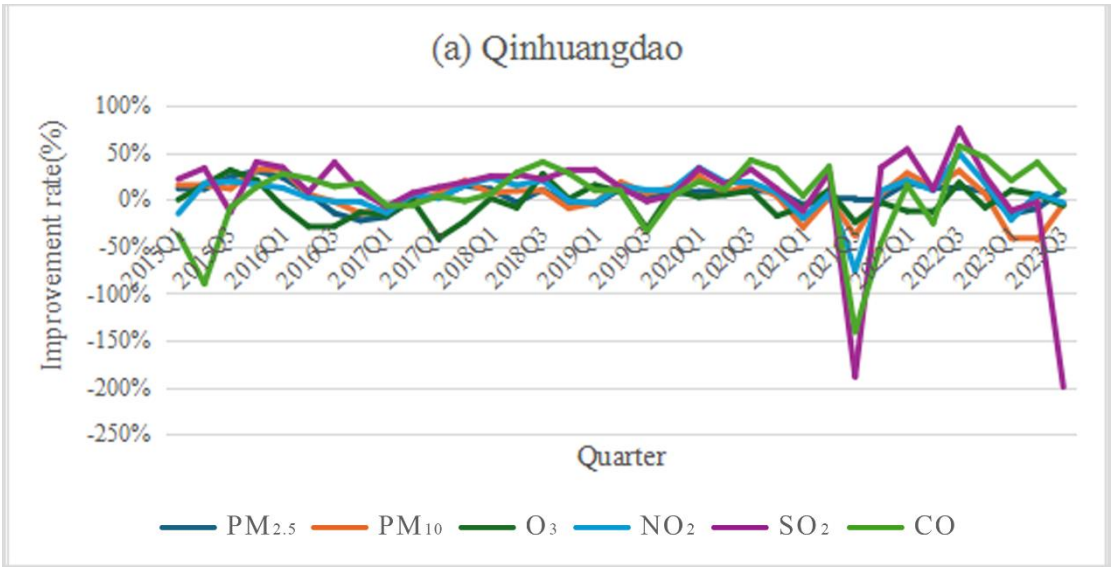
During the period of the global pandemic caused by COVID-19 from the beginning of 2020 to the end of 2022, there was an unparalleled improvement in air quality[31]. This was accompanied by a reduction in industrial and domestic production, which resulted in a significant overall decline in the frequency of $PM_{2.5}$ and PM_{10} as the primary pollutants. Conversely, there was a notable increase in the frequency of O_3 as the primary pollutant during this period. Notably, Beijing and Zhangjiakou, which have historically demonstrated relatively superior air quality within the BTH region, exhibited the most rapid surge in the prevalence of O_3 as a primary pollutant during the pandemic. Ozone becomes the main pollutant because other pollutants are greatly affected by human activities, while ozone is mainly influenced by weather conditions. Higher temperatures lead to higher ozone concentrations[257]. Furthermore, even in regions such as Handan, Qinhuangdao, Hengshui, Cangzhou, Shijiazhuang, Xingtai, Langfang, Baoding, and Chengde, which previously exhibited poor air quality and where O_3 was not a significant pollutant, the frequency of O_3 as a primary pollutant increased considerably during this period. Ozone as the primary pollutant usually means that the concentrations of other pollutants are at low levels and the overall air quality is good[258].

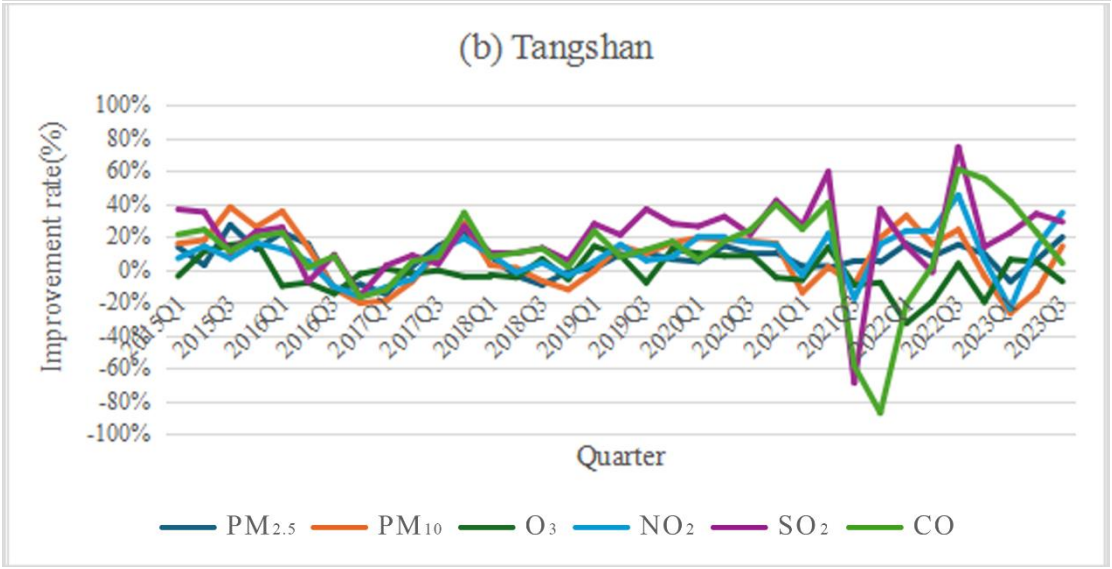
Following the lifting of the pandemic control measures in 2023, the resumption of production and the return to normalcy of everyday life will inevitably result in an increase in emissions of air pollutants in comparison to the levels observed during the pandemic period. The data in Figure 6-2 indicate a decrease in the frequency of O_3 as the primary pollutant and an increase in the frequency of $PM_{2.5}$ and PM_{10} as the primary pollutants in the cities of Baoding, Chengde, Langfang, Xingtai, Zhangjiakou, Qinhuangdao, and Cangzhou[190]. This suggests that the resumption of industrial and

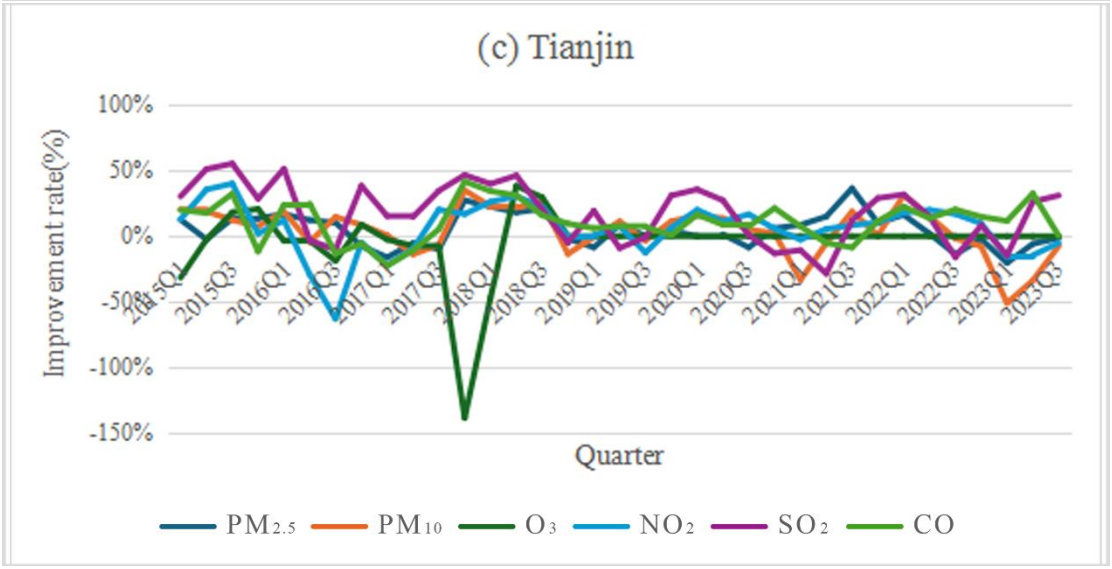
economic activity, coupled with a return to normal levels of activity and population growth, has resulted in a notable increase in air pollution. However, in six cities, the frequency of O_3 as a primary pollutant either increased or remained relatively unchanged, particularly in comparison to data collected prior to the pandemic. Furthermore, the frequency of O_3 as a primary pollutant increased in nearly all cities. This indicates that the impact of the COVID-19 pandemic on air quality was considerable, with the effects of the pandemic on air pollution in the BTH region still discernible during the latter stages of the pandemic, and overall air quality continuing to improve[54].

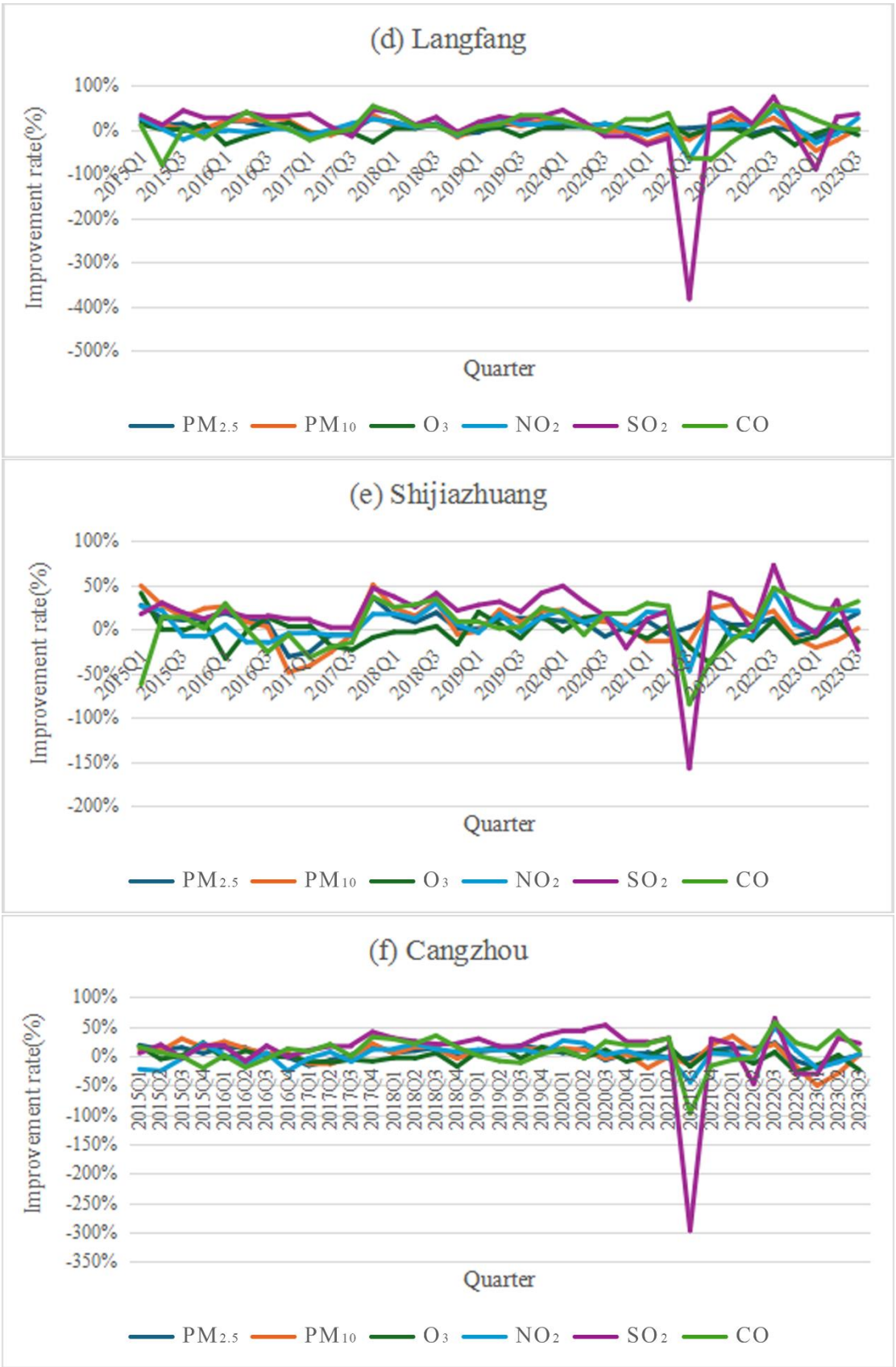
6.3.2 Quarterly fluctuations in air pollutants

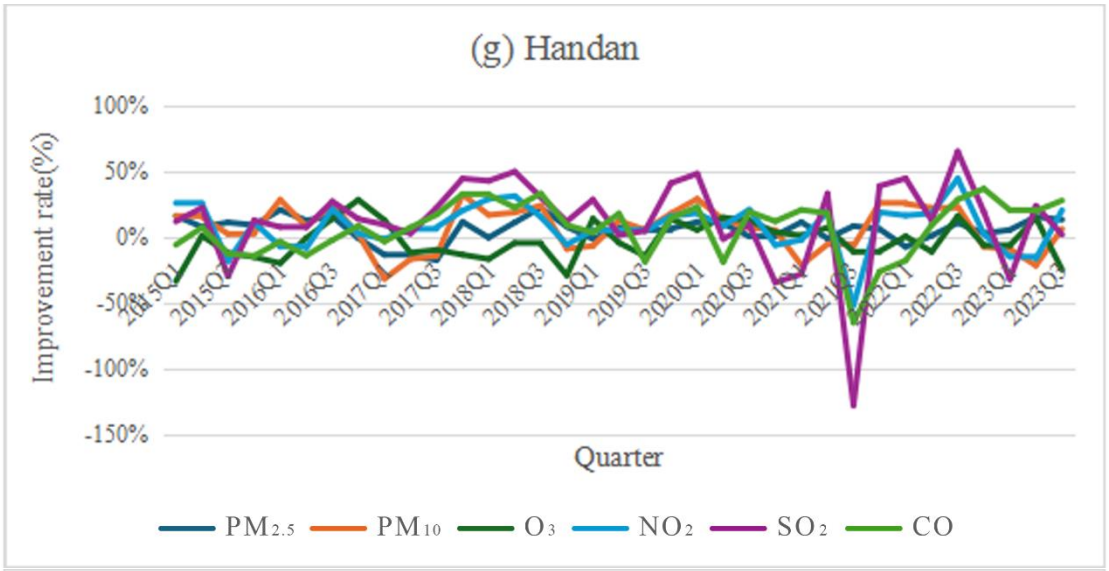
The annual fluctuations in air pollutants can serve as a direct reflection of the impact of air treatment on the enhancement of air quality. However, as changes in air pollution are seasonal in nature, a quarterly division of annual changes in air pollution can provide a more detailed illustration of the changes in air quality[128]. Figure 6-3 illustrates the quarterly changes in the six major pollutants for the 13 municipalities in the BTH region from the beginning of 2014 to 2023. This time period encompasses the initial period of elevated air pollution at the outset of the treatment, the subsequent period of annual improvement following the implementation of the treatment, and the period during the COVID-19 pandemic and the subsequent period following the complete lifting of the pandemic. This provides a comprehensive overview of the enhancement of air quality and the impact of the COVID-19 pandemic.

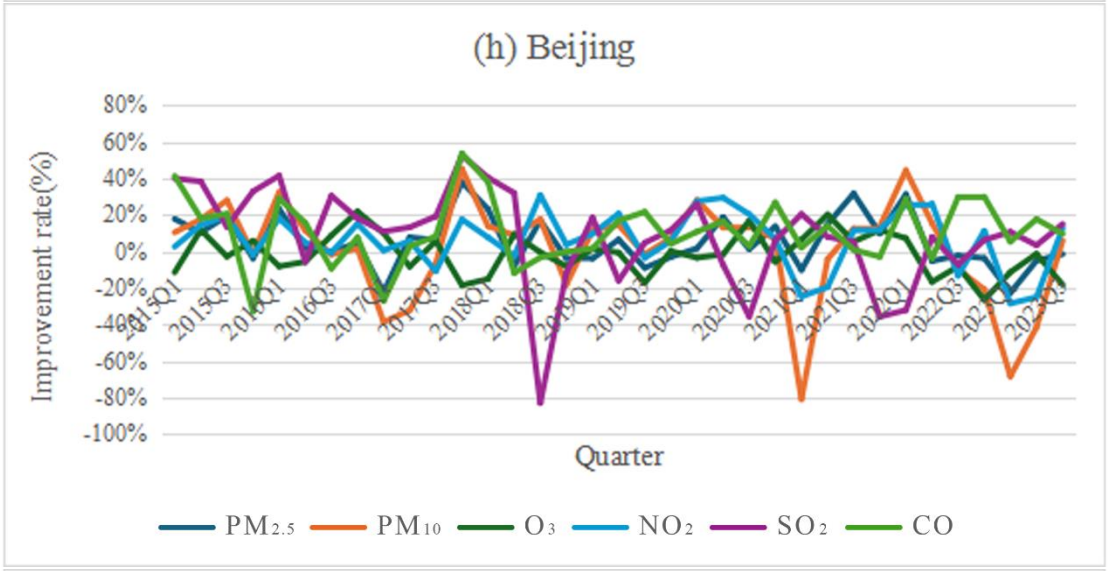

(a) Qinhuangdao
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
-200%
-250%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO

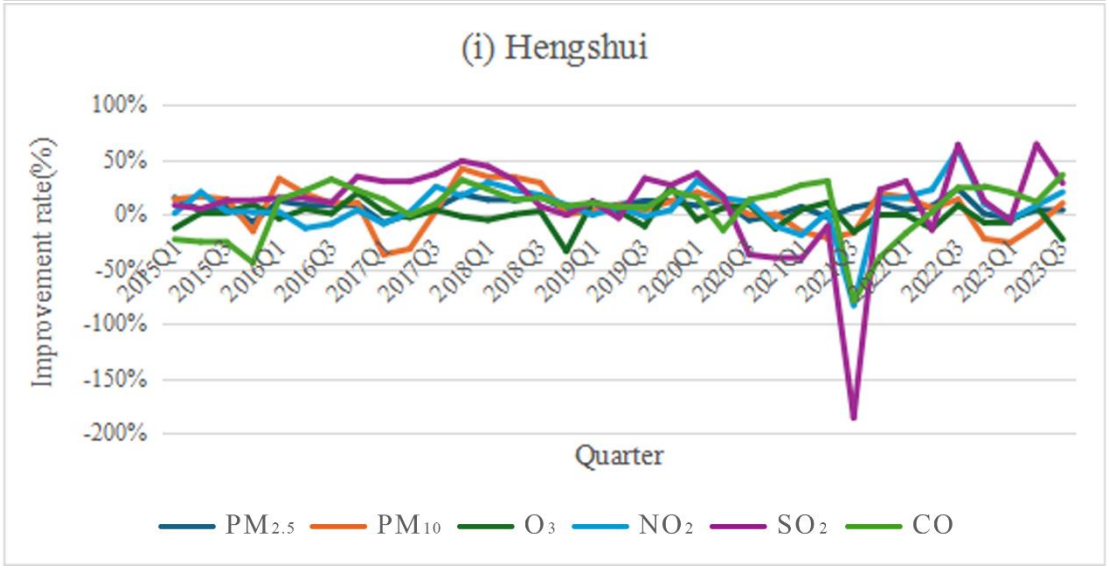

(b) Tangshan
Improvement rate(%)
100%
80%
60%
40%
20%
0%
-20%
-40%
-60%
-80%
-100%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO

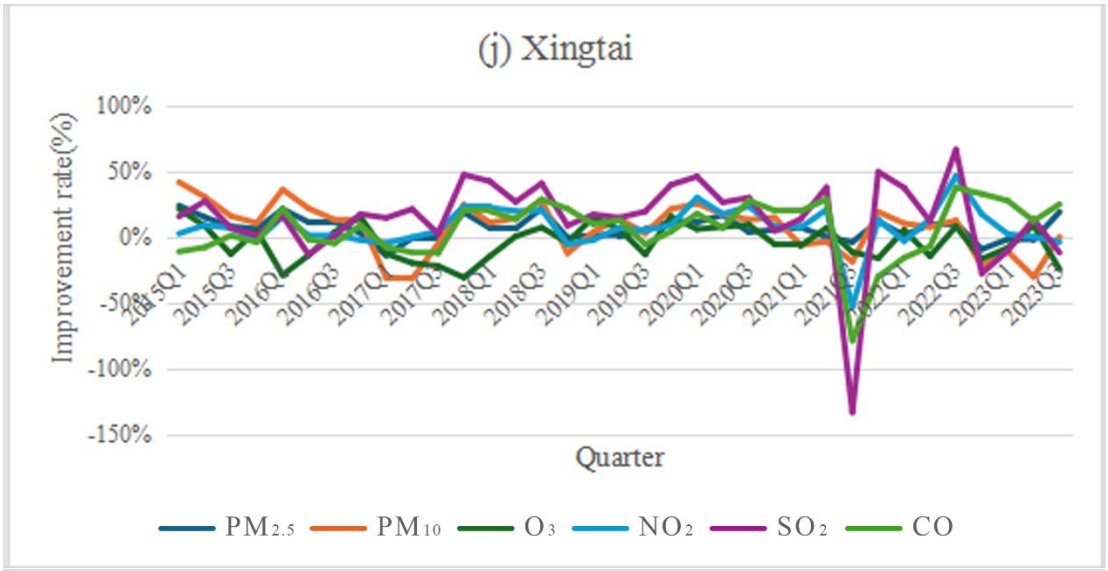

(c) Tianjin
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO

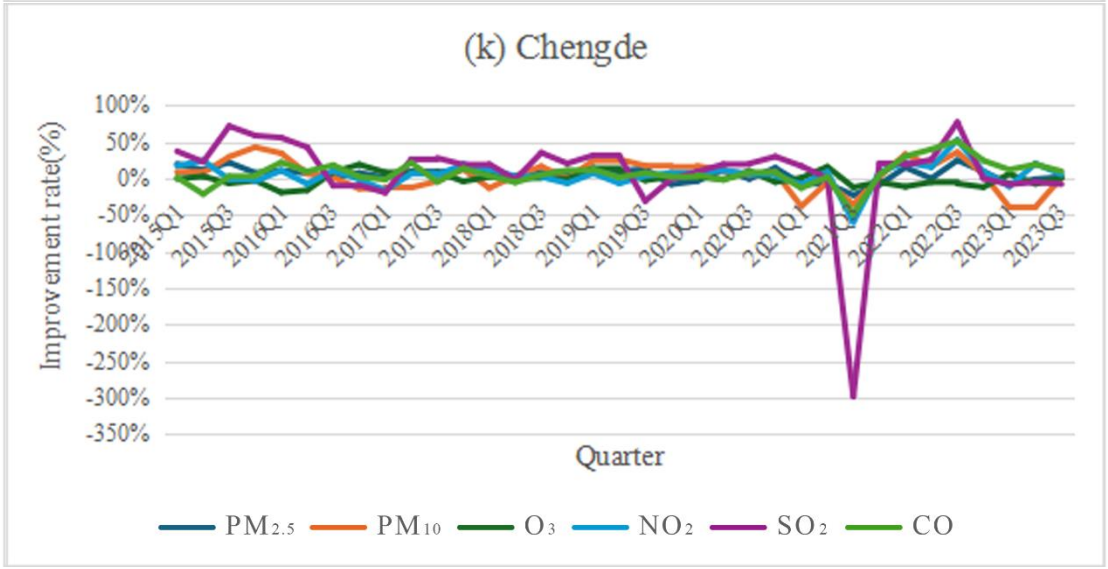

(d) Langfang
Improvement rate(%)
100%
0%
-100%
-200%
-300%
-400%
-500%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO
(e) Shijiazhuang
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
-200%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO
(f) Cangzhou
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
-200%
-250%
-300%
-350%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO

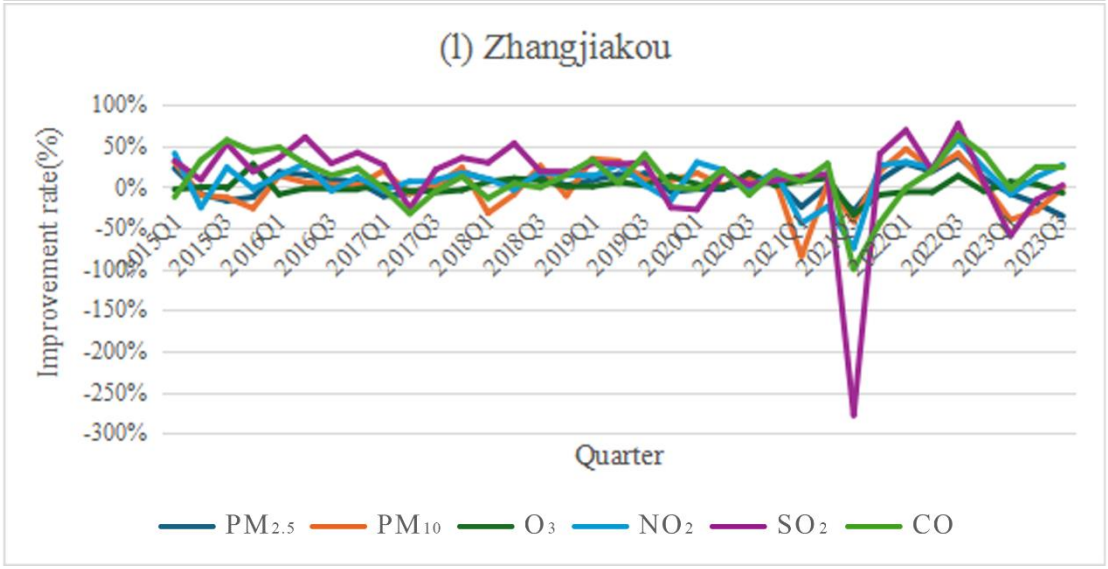

(g) Handan
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO


(h) Beijing
Improvement rate(%)
80%
60%
40%
20%
0%
-20%
-40%
-60%
-80%
-100%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO


(i) Hengshui
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
-200%
Quarter
PM2.5
PM10
O3
NO2
SO2
CO


(j) Xingtai
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
2015Q1
2015Q3
2016Q1
2016Q3
2017Q1
2017Q3
2018Q1
2018Q3
2019Q1
2019Q3
2020Q1
2020Q3
2021Q1
2021Q3
2022Q1
2022Q3
2023Q1
2023Q3
Quarter
PM2.5
PM10
O3
NO2
SO2
CO


(k) Chengde
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
-200%
-250%
-300%
-350%
2015Q1
2015Q3
2016Q1
2016Q3
2017Q1
2017Q3
2018Q1
2018Q3
2019Q1
2019Q3
2020Q1
2020Q3
2021Q1
2021Q3
2022Q1
2022Q3
2023Q1
2023Q3
Quarter
PM2.5
PM10
O3
NO2
SO2
CO


(l) Zhangjiakou
Improvement rate(%)
100%
50%
0%
-50%
-100%
-150%
-200%
-250%
-300%
2015Q1
2015Q3
2016Q1
2016Q3
2017Q1
2017Q3
2018Q1
2018Q3
2019Q1
2019Q3
2020Q1
2020Q3
2021Q1
2021Q3
2022Q1
2022Q3
2023Q1
2023Q3
Quarter
PM2.5
PM10
O3
NO2
SO2
CO

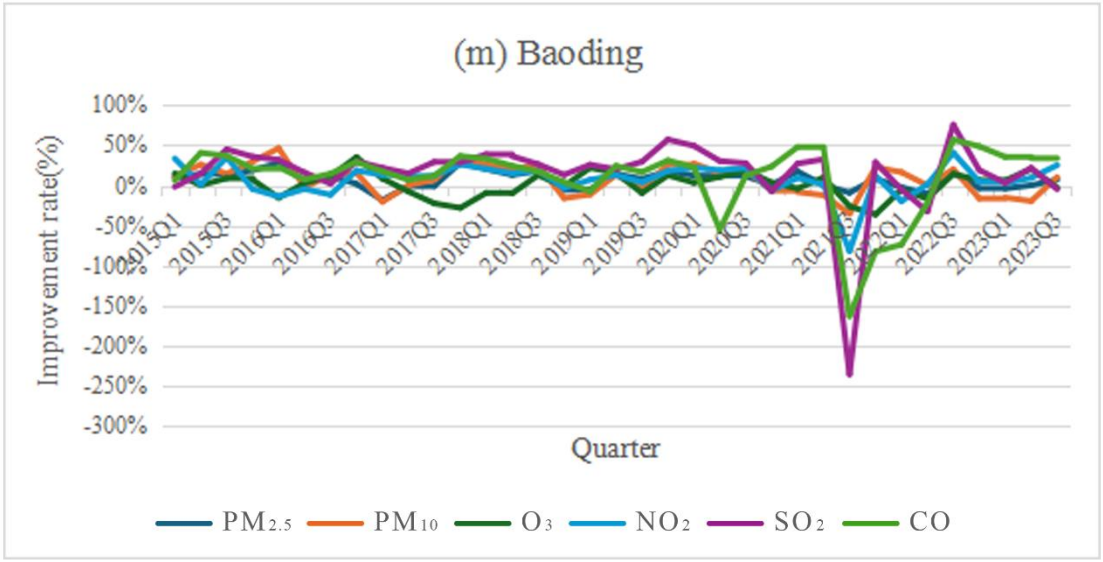


Figure 6-3(a) –(m): Quarterly year-on-year improvement rate (%) in air pollutants from 2014 to 2023 for 13 cities in the BTH region. The x-axis is time, and the y-axis is the rate of air quality improvement in each year over the same quarter of the previous year.

Due to the seasonal fluctuations in air pollution levels in the BTH region, it can be observed that the pollution levels are relatively high during the first and fourth quarters, and relatively low during the second and third quarters[70]. Therefore, the pollution trend illustrated in Figure 6-3 demonstrates an improvement in each quarter of each year in comparison to the same quarter of the previous year. This provides a more rigorous representation of the improvement in air quality. When the percentage change in the graph is above 0%, this indicates an improvement in air quality in comparison to the previous year. Overall, the air quality in each city exhibits a fluctuating pattern; however, it demonstrates a consistent trend of improvement over the years. This suggests that the level of air pollution is declining on an annual basis as a result of policy intervention. Moreover, the enhancement in air quality has been more consistent, oscillating between 0% and 50% on an annual basis. The improvement in air quality is relatively stable, with an average annual increase of between 0% and 50%. Some cities, including Zhangjiakou have seen the annual changes of CO, SO_2, and Chengde's SO_2, Beijing's CO and Shijiazhuang's PM_{10} and Tianjin's SO_2, have experienced an

improvement of over 50% in certain periods. Overall, there has been an improvement in air quality; however, there have been fluctuations. For example, according to the data in Figure 6-3 between the fourth quarter of 2016 and the first quarter of 2017, ten cities (Tianjin, Langfang, Shijiazhuang, Cangzhou, Handan, Beijing, Hengshui, Xingtai, Chengde, and Zhangjiakou) exhibited a decline in air quality compared to the same quarter of the previous year[190]. Of the cities in question, Shijiazhuang experienced the most pronounced decline, with $PM_{2.5}$ levels dropping by 30.92% and PM_{10} by 48.61% in the fourth quarter of 2016, and $PM_{2.5}$ and PM_{10} levels falling by 25.83% and 41.44%, respectively, in the first quarter of 2017. The fluctuations in air quality are influenced by a multitude of factors, both anthropogenic (industrial emissions, transportation) and natural (wind speed, temperature, precipitation)[259]. Consequently, the observed overall improvement in air quality is to be expected, given the impact of natural factors on these changes.

In addition to the typical fluctuations observed in air quality, the greatest variation was noted during the period of the COVID-19 outbreak. In response to the COVID-19 pandemic, the BTH region implemented stringent measures to restrict human activity and mobility, resulting in a notable reduction in anthropogenic emissions of air pollutants[131]. The enhancement in air quality for the six principal pollutants exhibited an increase of over 10% for all cities in 2020. On average, no more than two pollutants exhibited a decline in air quality per city, with all declines occurring at a rate of less than 5%. In comparison to preceding years, there has been a slight increase in the degree of improvement in air quality, accompanied by a notable reduction in the extent of deterioration for each pollutant[190]. The reduction in anthropogenic activities due to the impact of the COVID-19 pandemic had a significant effect on the improvement of air quality. However, in 2021, there was a sudden and dramatic decline in air quality. This was due to the Chinese government's strict control of the spread of the COVID-19 pandemic which has resulted in the vast majority of the country being cleared of COVID-19, and the restoration of the most productive human activities[260].

Therefore, the air pollution situation in 2021 appears to be significantly worse than the air quality in 2020, when the entire country was subject to severe restrictions on production and daily life. It is particularly noteworthy that the rise in SO_2 levels, which are strongly linked to industrial and vehicular emissions has been considerable in a number of urban areas. In the cities of Langfang, Cangzhou, Chengde, Zhangjiakou, and Baoding, there has been a marked increase in the levels of SO_2 and CO with rises of over 200% and 100% respectively. This trend has also been observed in the cities of Qinhuangdao, Zhangjiakou, and Baoding[190]. Furthermore, other pollutants demonstrated considerable annual increases in comparison to the data recorded in 2020. The presented data substantiate the assertion that the impact of human activity on air quality is considerable, particularly in comparison to the relatively static situation observed in 2020[259]. Moreover, in 2022 COVID-19 underwent a mutation that markedly increased its transmission rate compared to the previous variant, while concurrently reducing its virulence. Consequently, the local control policy was modified from an overall containment strategy to a regional containment approach. As illustrated in Figure 6-3, the data indicate a notable enhancement in air quality in 2022 in comparison to the preceding year, 2021. In particular, the pollutants that exhibited a notable decline in air quality in 2021 demonstrated a significant increase in 2022. Overall, the improvement in air quality for each pollutant is more pronounced than in previous years, with numerous pollutants exhibiting an improvement of over 50%. This provides further evidence that a reduction in human production and activity has a considerable effect on air quality.

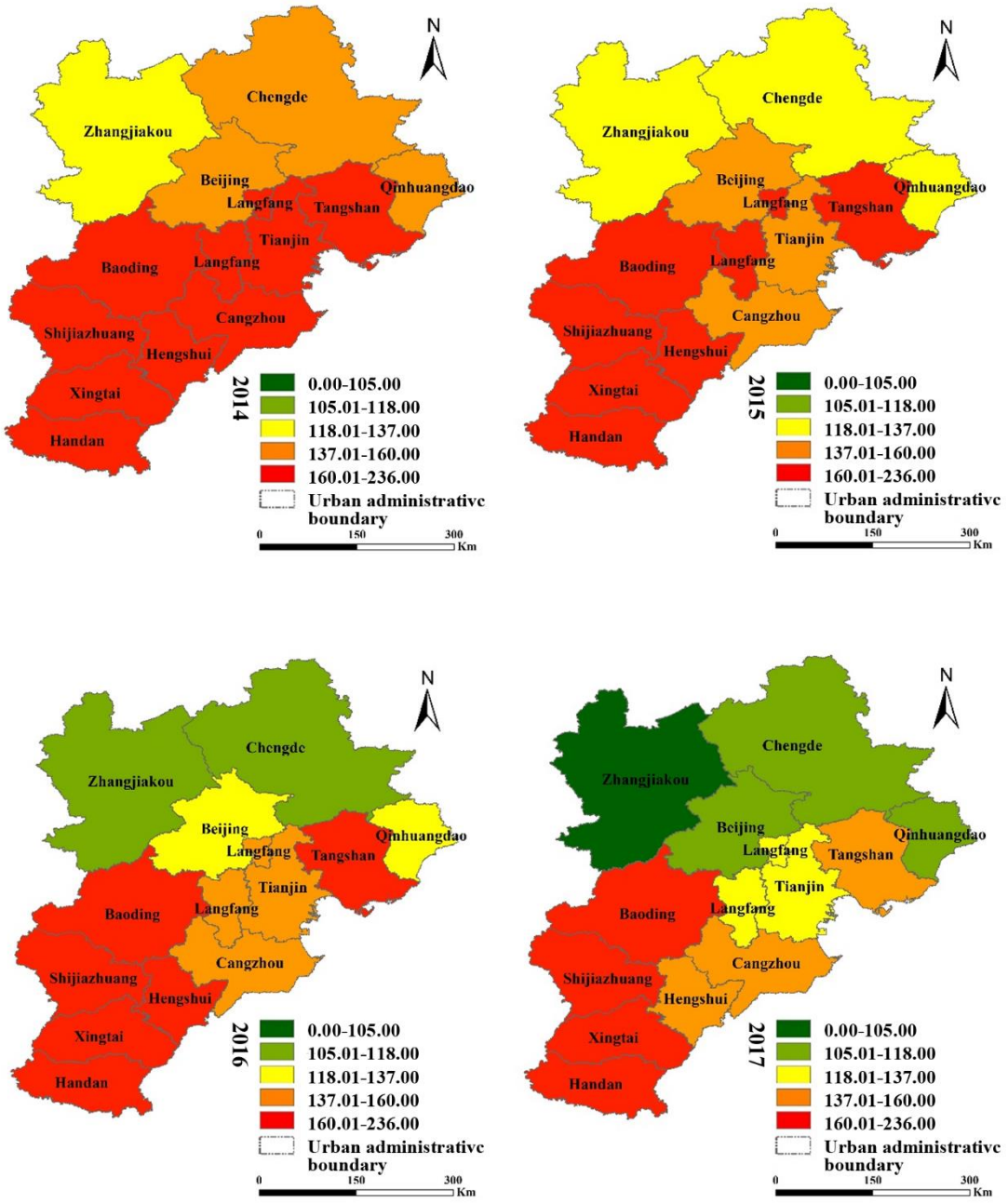
The air quality data from 2023, the year of full relaxation of the COVID-19 in China, is crucial for a comprehensive evaluation of the consequences of limiting human activity on air quality. The data from 2023 will reflect not only the impact of COVID-19 on air quality, but also the improvement of the region’s air quality following the removal of the impacts. Figure 6-3 illustrates that the majority of cities exhibited minimal fluctuations in the intensity of each pollutant, accompanied by slight declines

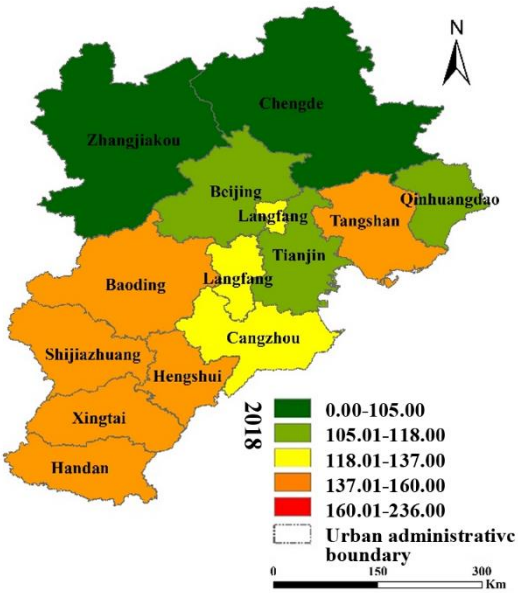
in air quality and, in some instances, modest enhancements. With the exception of Beijing and Cangzhou, where the PM_{10} pollution index increased by over 50%, the pollution indices of the remaining cities demonstrate a relatively stable trend. The change in air pollution in 2023 demonstrates that the impact of the pandemic on air quality improvement is significant, yet the implementation of air pollution control measures in the BTH region has led to a sustained improvement in air quality. Even in the absence of the pandemic's influence on air quality the region continues to experience a gradual enhancement in air quality.

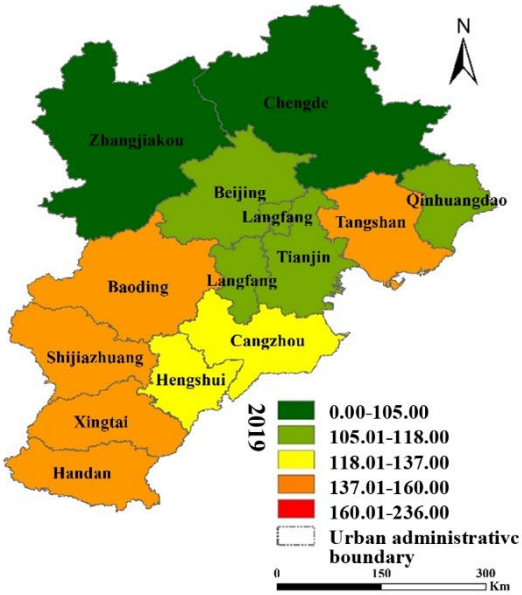
6.3.3 Annual changes in air quality index

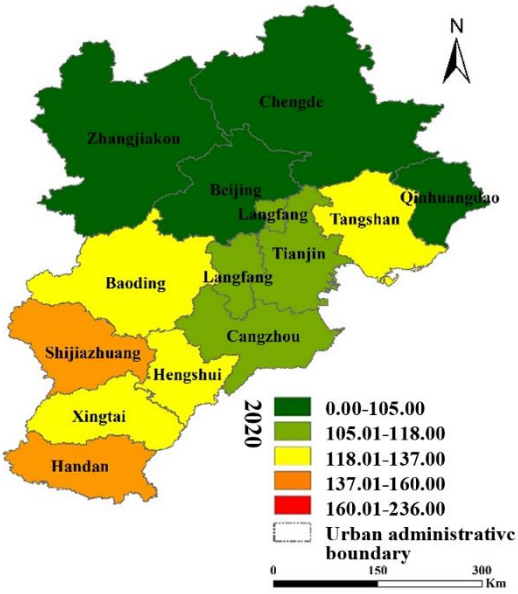
The distribution of air pollutants is typically random, exhibiting considerable variation in the AQIs observed across different cities within the same region. The regional distribution of air pollution is in the BTH region, as demonstrated in Figure 6-4, which depicts the geographical distribution and fluctuations in AQI from 2014 to 2023 for 13 cities within the BTH region. The colour spectrum, from green to red represents the quality of the air, with green indicating optimal conditions and red indicating poor air quality. The air quality is superior in the northern regions of the BTH region when compared to the southern regions. Zhangjiakou, located in the north-western region, has demonstrated the best air quality, while Baoding, Shijiazhuang, Xingtai and Handan, situated in the south-western region, have consistently exhibited the poorest air quality within the region. The cities of Chengde and Qinhuangdao in the north-eastern region demonstrated a more pronounced improvement (From orange to green), whereas the cities of Langfang, Cangzhou and Hengshui in the south-central region, and Tangshan in the south-western region exhibited a comparatively lesser degree of improvement than the aforementioned cities. The air quality in the central municipalities of Beijing and Tianjin exhibited a trajectory of gradual improvement, situated between the north

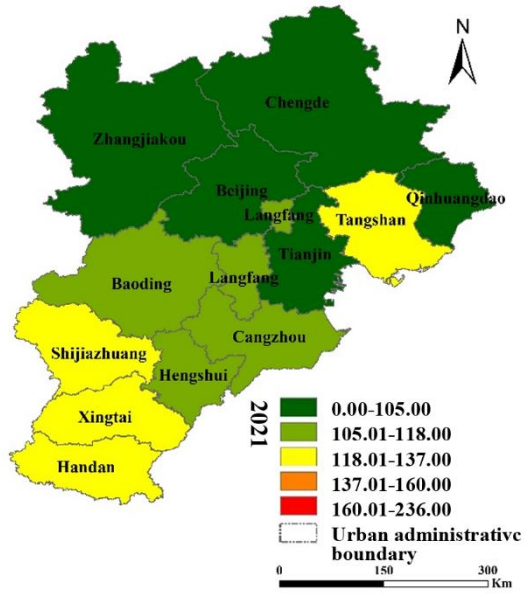
and the south, and attained the air quality standards, this may be related to the geo-location, human involvement and industrial setting etc[10].




N
Chengde
Zhangjiakou
Beijing
Qinhuangdao
Langfang
Tangshan
Tianjin
Baoding
Langfang
Cangzhou
Shijiazhuang
Hengshui
Xingtai
Handan
2018
0.00-105.00
105.01-118.00
118.01-137.00
137.01-160.00
160.01-236.00
Urban administrativc boundary
0
150
300
Km


N
Chengde
Zhangjiakou
Beijing
Qinhuangdao
Langfang
Tangshan
Tianjin
Baoding
Langfang
Cangzhou
Shijiazhuang
Hengshui
Xingtai
Handan
2019
0.00-105.00
105.01-118.00
118.01-137.00
137.01-160.00
160.01-236.00
Urban administrativc boundary
0
150
300
Km


N
Chengde
Zhangjiakou
Beijing
Qinhuangdao
Langfang
Tangshan
Tianjin
Baoding
Langfang
Cangzhou
Shijiazhuang
Hengshui
Xingtai
Handan
2020
0.00-105.00
105.01-118.00
118.01-137.00
137.01-160.00
160.01-236.00
Urban administrativc boundary
0
150
300
Km


N
Chengde
Zhangjiakou
Beijing
Qinhuangdao
Langfang
Tangshan
Tianjin
Baoding
Langfang
Cangzhou
Shijiazhuang
Hengshui
Xingtai
Handan
2021
0.00-105.00
105.01-118.00
118.01-137.00
137.01-160.00
160.01-236.00
Urban administrativc boundary
0
150
300
Km

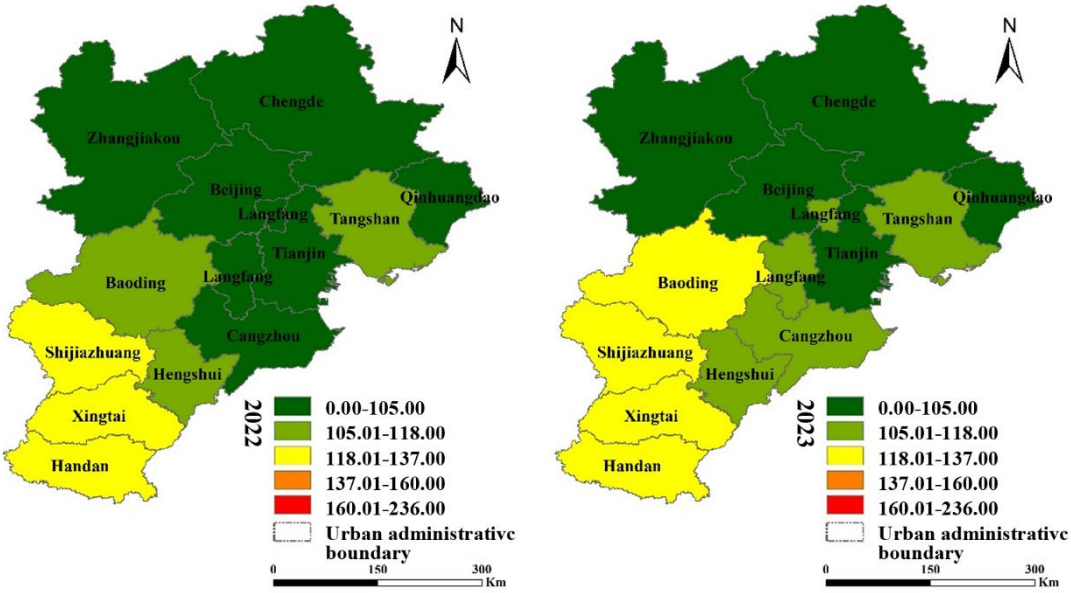


Figure 6-4: Spatial distribution of the annual changes in AQI from 2014 to 2023 for 13 cities in the BTH region. Colours range from green (low AQI) to red (high AQI).

An analysis of the data reveals that in 2014, the air quality in the BTH area was consistently poor. Cities situated in the centre and south of the region were classified as red, indicating that the AQI was at a markedly elevated level. Figure 6-4 indicates that even in the northern cities, where air quality is relatively good, only Zhangjiakou was classified as yellow, while Beijing, Chengde and Qinhuangdao were all designated as orange. By 2015, there had been a slight improvement in air quality, with Chengde and Qinhuangdao added to the list of cities in yellow and Tianjin and Cangzhou added to the list of cities in orange. The remaining cities remained in the red category. In 2016, further improvements were observed, with Zhangjiakou and Chengde in the north reaching light green and Beijing and Qinhuangdao raising their AQIs to yellow. Additionally, the city of Langfang was assigned an orange level classification. By 2017, for the first time, the city of Zhangjiakou exhibited air quality at the green level, while Beijing and Qinhuangdao also demonstrated improvement, reaching the light green level. The cities of Langfang and Tianjin were elevated from the orange category to the

yellow category. Additionally, Hengshui, situated in the southern region, was elevated from the red level to the orange level for the first time. By the time 2018 had elapsed, there had been a significant improvement in air quality in the BTH region, eliminating all cities designated as having poor air quality. Furthermore, the AQIs of the southern region's cities with the poorest air quality had all improved to orange levels. Both Zhangjiakou and Chengde, situated in the northern part of the region, have attained the green level. By 2019, air quality had improved to a lesser extent than in 2018, with Langfang moving from yellow to light green. The year 2020 was characterised by the outbreak of the COVID-19, which led to a drastic reduction in human activity in the BTH region in order to control the spread of the pandemic[137]. This resulted in a notable improvement in air quality, although it had a significant impact on the lives of people in the region. In comparison to the preceding year, 2019, the air quality in the four northern cities of Zhangjiakou, Chengde, Qinhuangdao and Beijing all reached the green level. Additionally, Tangshan achieved its first yellow rating, while Cangzhou, situated in the southern region, reached light green. Furthermore, the heavily polluted cities of Baoding and Xingtai, which had previously been rated orange, improved to yellow. By 2021, there was a significant improvement in air quality, with five cities (Zhangjiakou, Chengde, Beijing, Tianjin, Qinhuangdao) reaching the green level. In 2020, the yellow level was attained by Baoding and Hengshui, which were subsequently upgraded to the light green level. Similarly, the orange-coloured cities of Shijiazhuang and Handan also reached the light green level. Consequently, all air quality in the entire BTH region was upgraded to yellow and above. In 2022, Tangshan was upgraded to the light green level for the first time. This indicates that the restrictions imposed by the global pandemic have had a significant impact on air quality. The impact of COVID-19 on air pollution changes has been widely studied. For example, Uzair et al. [144] investigated air pollution variations during the lockdown period and the temporary lifting of restrictions. However, in these studies, the lifting of lockdown measures was temporary, making it difficult to account for long-term changes and

seasonal effects. In this study, we incorporate air pollution data from 2023, after the complete removal of lockdown measures in the region. In 2023, following the relaxation of restrictions imposed by the pandemic, economic activity and social life resumed at pre-pandemic levels. This resulted in a slight improvement in air quality across the region, with Baoding moving from the light green level to yellow and Langfang and Cangzhou shifting from green to light green. This suggests that when human activity resumes at pre-pandemic levels, air quality declines slightly and is more susceptible to the effects of COVID-19. Furthermore, following the removal of the impact of the pandemic from the data set for the period 2020 to 2022, a comparison of the air quality situation in the BTH region with that of 2019 reveals a significant improvement. Additionally, the number of cities with air quality reaching the standard is on the rise. The air quality of the cities in the northern region has reached the green level. The key areas for air pollution control are Baoding, Shijiazhuang, Xingtai and Handan in the south-west, where there is still scope for improvement in air quality.

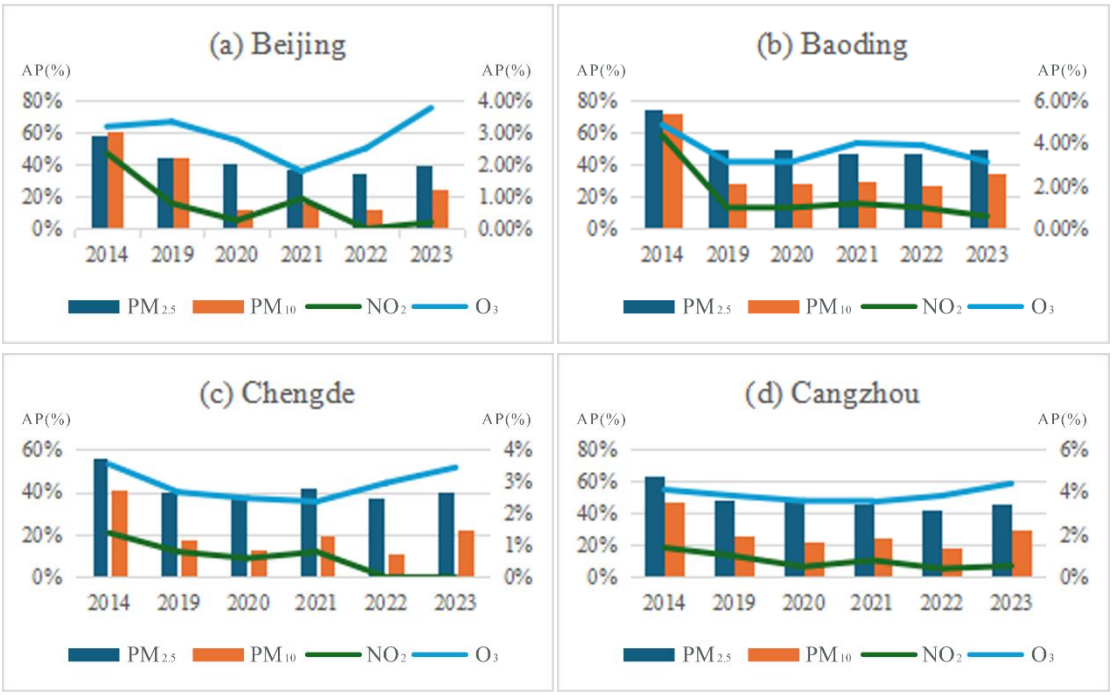
6.4 Proportion of health effects attributable to air pollutants

The most significant risk associated with air pollutants is their detrimental impact on human health. There is a discernible correlation between prolonged or acute exposure to air pollution and an increased risk of mortality and morbidity[261]. In order to evaluate the potential risks of air pollution on human health, the WHO has developed AirQ+, a software tool that can be utilized in all regions to assess the relationship between air pollutants and attributable mortality and morbidity[216]. The software will be employed in this study to evaluate the long-term attributable proportions of pollutants on total mortality, respiratory and circulatory outcomes between 2014 and 2023 in 13 cities within the BTH region. This will facilitate the connection between changes in air pollution and associated health risks, enabling a more intuitive response

to improvements in air quality through the implementation of air pollution control measures. Furthermore, it will support the ongoing assessment of the impact of air pollution on health hazards, both during the pandemic and beyond.

6.4.1 Attributable proportion of total mortality

Figure 6-5 illustrates the shift in the proportion of attribution of the four major pollutants, $PM_{2.5}$, PM_{10}, NO_2, O_3, to total mortality from 2014 to 2023 for the 13 cities in the BTH region. As illustrated in the figure, the most significant pollutants are particulate matter ($PM_{2.5}$ and PM_{10}), while gaseous pollutants (NO_2 and O_3) exert a comparatively lesser influence.



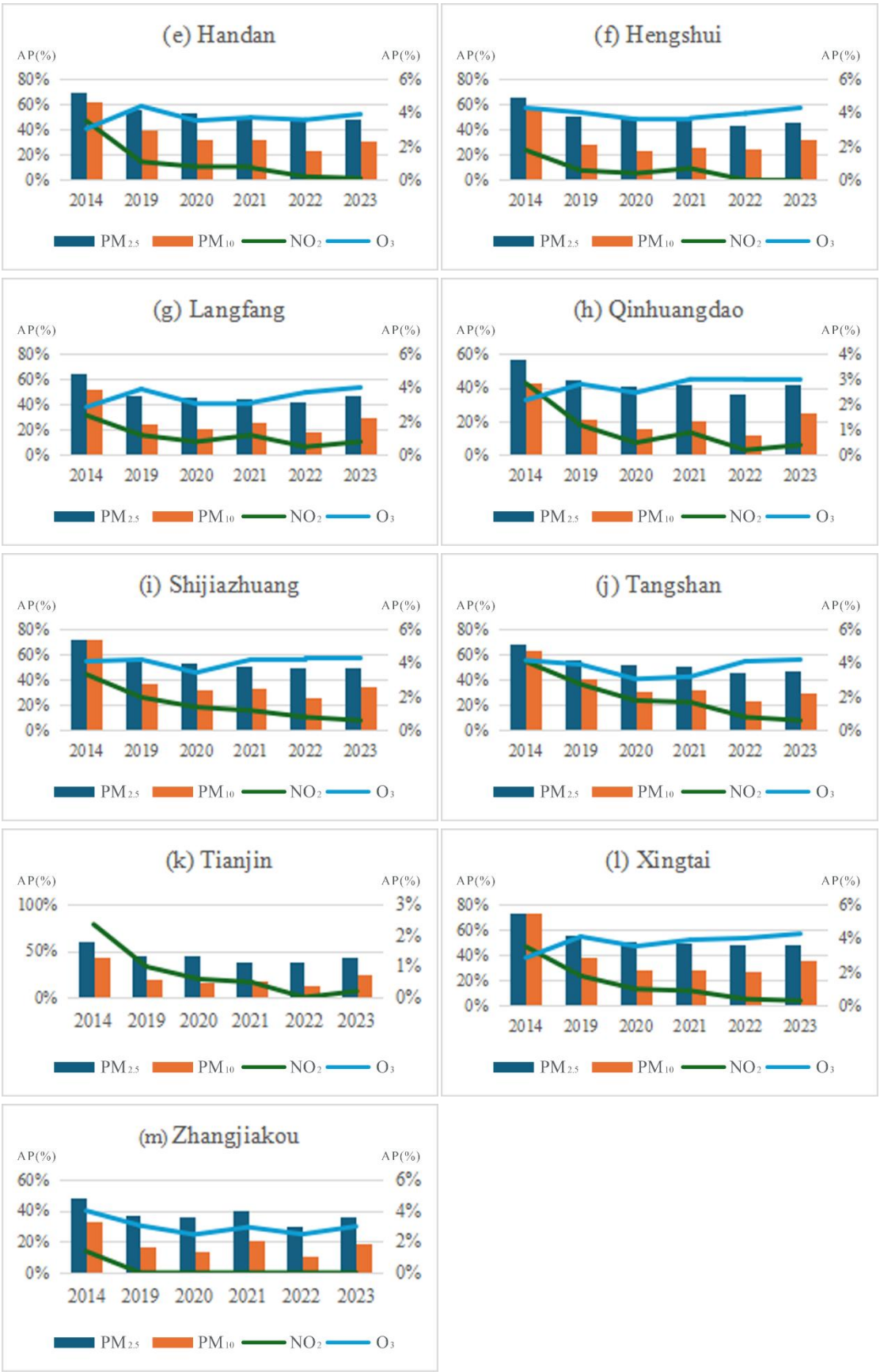


Figure 6-5 (a-m): Annual total mortality attributable proportion (%) for $PM_{2.5}$, PM_{10},

NO_2 and O_3 in the BTH region in 2014 and from 2019 to 2023.

The analysis commences with an examination of the attributable proportions (AP) of particulate matter. $PM_{2.5}$ is the most harmful pollutant, exhibiting the highest levels in every year across the 12 cities with the exception of Beijing. Moreover, Figure 6-5 indicates that Beijing's PM_{10} has also declined to a greater extent than $PM_{2.5}$ following years of treatment. From 2020 onwards, the AP of PM_{10} begins to be significantly lower than AP of $PM_{2.5}$. In the initial phase of air pollution control, namely 2014, the AP of $PM_{2.5}$ was markedly elevated, exceeding 60% in eight cities: Baoding, Cangzhou, Handan, Hengshui, Langfang, Shijiazhuang, Tangshan and Xingtai. In 2014, long-term exposure to air pollution was identified as an extremely hazardous factor for human health. Following years of intervention, the attributable proportion exhibited a marked decline in each city, with an average reduction of over 15% observed in 2019. The city of Zhangjiakou exhibited the lowest decline, at 11.43%. This may be attributed to the relatively favourable overall air quality in the region, where the attributable proportion of $PM_{2.5}$ was as low as 36.98% in 2019, comparable to other cities after 2020. The city with the most significant reduction in attributable proportion is Baoding, with a decrease of up to 25%. This is due to the fact that Baoding is a city with a greater concentration of factories that contribute to air pollution, and its attributable proportion was also the highest in the BTH region in 2014, at 74.98%[185], [256].

In addition, the reduction in the AP for $PM_{2.5}$ between 2019 and 2020 is less pronounced. The most substantial decline is observed in Xingtai, at a rate of 4.31%, while Tangshan also exhibits a decline exceeding 4%, at 4.25%. The decline in the AP for $PM_{2.5}$ is less than 4% in other cities, with Tianjin and Baoding exhibiting a mere 0.09% and 0% decline, respectively. The data demonstrate that the impact of the COVID-19 pandemic on the AP of $PM_{2.5}$ is relatively limited. In contrast, the change in the AP for $PM_{2.5}$ from 2020 to 2021 exhibits somewhat disparate trends across different cities. The majority

of cities continue to demonstrate a gradual decline in the AP, with three exceptions: Chengde (3.92%), Qinhuangdao (0.45%), and Zhangjiakou (3.58%), which exhibit an increase. This slight increase is a normal fluctuation and is related to the relatively good air quality in these cities. This may be attributable to the shorter period of closure and control for these cities in 2021. In 2022, the rapid spread of the mutant strain of the COVID-19 resulted in a significantly shorter period of containment compared to 2021 across all regions. This is also reflected in the change in the AP of $PM_{2.5}$ from 2021 to 2022. In comparison to the preceding year, the AP of $PM_{2.5}$ in 2022 is observed to decline in all cities except Baoding, where it is noted to increase by 0.2%. It is further noted that seven cities experienced a decline of over 3%. The cities of Qinhuangdao and Tangshan exhibited decreases of 5.15%, while Zhangjiakou demonstrated the most substantial decline, at 9.75%. The AP for $PM_{2.5}$ in Zhangjiakou in 2022 is the lowest of the 13 cities in a calendar year, at 30.08%. The data from 2022 indicates that the implementation of stringent restrictions on human activities has a considerable impact on the improvement of air quality and the reduction of the adverse effects of air pollution on human health[16]. The year 2023 is of particular significance, marking the complete lifting of COVID-19 restrictions in China. This has led to a gradual return to normalcy in terms of both production and quality of life. This phenomenon has resulted in an increase in the AP of $PM_{2.5}$ in 11 cities when compared to the data from 2022. The declines observed in Handan and Shijiazhuang were relatively modest, at 1.96% and 0.08%, respectively. In the case of the cities with elevated APs, Beijing exhibited an increase of 5.42%, Cangzhou 4.29%, Langfang 4.46%, Qinhuangdao 5.38%, Tianjin 4.59%, and Zhangjiakou demonstrated the most significant elevation at 6.42%. The data from the latter part of the closure period provides further evidence of the impact of the pandemic on air pollution. The closure had a considerable impact on the improvement of air quality, with all cities exhibiting a notable enhancement in air quality as a result of the closure prompted by the advent of the pandemic[147]. Furthermore, a comparison of the pre- and post-COVID-19 periods allows for the identification of

changes in air quality following the removal of the impact of COVID-19. This may be indicative of an improvement in air quality management in the BTH region following the year 2019. The annual mean concentration of $PM_{2.5}$ in 2023 demonstrates a downward trajectory in all 13 cities when compared to the 2019 data. Cities exhibiting a lesser degree of decrease include Baoding, Chengde, Langfang and Zhangjiakou, with a reduction of 0.59%, 0.23%, 0.61% and 0.48%, respectively. Notable decreases were observed in Beijing, Handan, Hengshui, Shijiazhuang, Tangshan and Xingtai, with reductions of 4.88%, 8.24%, 5.56%, 5.18%, 8.90% and 6.62%, respectively. The data demonstrate that the impact of the pandemic on air quality improvement is considerable. However, following the exclusion of the effects of the pandemic, it is evident that the implementation of air pollution control measures has resulted in a notable enhancement of air quality, with six cities exhibiting a particularly noteworthy improvement.

The AP for PM_{10} exhibited a considerably more pronounced decline than that observed for $PM_{2.5}$ in urban areas. In particular, the average decrease in 2019 compared to 2014 was 26.02%. Four cities exhibited a decline in excess of 30%, namely Baoding, Hengshui, Shijiazhuang and Xingtai. The city of Baoding, which exhibited the most pronounced decline, demonstrated a reduction of 44.02%. In the initial year of the global pandemic, 2020, there was a reduction in the AP for PM_{10} across all cities in comparison to 2019. However, the extent of this decline exhibited considerable variation. The city of Beijing exhibited the most substantial decline, at 32.64%. With the exception of Xingtai, which exhibited a decrease of 9.40%, the remaining cities demonstrated a reduction of approximately 3%. In the second year affected by the global pandemic of 2021, the closure time (2020 was the first year of COVID-19. At the beginning of the year, there was a 4-month centralized lockdown. In 2021, the overall situation was stable, with a lockdown period of 1 to 2 months.[262]) is reduced in comparison to the previous year, resulting in higher APs for PM_{10} in 2021 compared to 2020 for all 13 cities. Nevertheless, the extent of the increase is relatively modest, with an average magnitude of approximately 3%. The cities with the highest elevations

are Chengde, Langfang, and Zhangjiakou, with respective increases of 5.91%, 5.39%, and 6.81%. In 2022, the cities' closure controls became more stringent, resulting in a notable decline in AP across the 12 cities, with an average reduction exceeding 7%. The city of Zhangjiakou, which experienced the most significant decline, saw a reduction of 10.26%. Moreover, in 2023, following the complete lifting of the closure control, the AP of all cities exhibited an upward trajectory, with five cities demonstrating an increase of over 10%. The data demonstrate that the COVID-19 pandemic has a substantial impact on PM_{10} levels, with a notable surge in AP observed following the lifting of sealing control measures. Furthermore, the data for 2023, following the lifting of the closure, demonstrated a slight increase in AP in eight cities and a decrease in AP in only five cities, in comparison to the data for 2019, before the advent of the pandemic. This suggests that the impact of PM_{10} management strategies is no longer significant after 2019, with indications of a potential rebound[185]. However, the AP for PM_{10} is already at approximately 30% of its value in 2023, with even Zhangjiakou's AP dropping to 19.4%. Overall, the local government's measures to combat air quality are effective; however, the rate of improvement will likely slow down or fluctuate as air quality gradually improves[185].

The APs for the gaseous pollutants NO_2 and O_3 are considerably lower than those for particulate matter, with none of the APs exceeding 5% in any of the previous years[216]. In 2014, the APs of NO_2 exceeded the 3% threshold in five cities. The following cities have been identified as having exceeded the APs for NO_2: Baoding, Handan, Shijiazhuang, Tangshan and Xingtai. In 2019, following years of air pollution control measures, only six cities – Handan, Langfang, Qinhuangdao, Shijiazhuang, Tangshan and Xingtai – exhibited APs above 1%. The highest concentration was recorded at 2.73% in Tangshan. During the subsequent period of the global pandemic caused by the COVID-19, the annual mean concentration of NO_2 in urban areas declined markedly, with only four cities, namely Baoding, Langfang, Shijiazhuang and Tangshan, exhibiting concentrations above the 1% threshold. In 2021, both Baoding and Langfang

exceeded 1%, which can be attributed to the shorter closure and control periods in that year, coupled with a return to more normal production and activity levels. The cities of Shijiazhuang and Tangshan have experienced a greater duration of concentrations above 1%, yet these values are exhibiting a gradual decline on an annual basis[190]. Among the cities with the most significant improvements in air quality, Beijing and Tianjin reached 0 AP in 2022 and rebounded to 0.20% in 2023, with NO_2 already at a non-hazardous level. Furthermore, Chengde and Hengshui reached zero in both 2022 and 2023, while Zhangjiakou reached zero from 2019 to 2023, achieving a level that is no longer hazardous. In 2023, only five cities exhibited an AP exceeding 0.5%, with all remaining values below 0.8%. Furthermore, the BTH region demonstrated minimal health impacts from NO_2 in 2023. O_3 is a more specific indicator, exhibiting less sensitivity to air pollution control measures and displaying a pronounced correlation with temperature[236]. As temperature rises, ozone concentrations tend to increase. As illustrated in Figure 6-5, the AP of O_3 exhibits a downward trend, with all values below 5%, which has a negligible impact on human health. With regard to the evolution of individual cities, six cities, namely Baoding, Cangzhou, Chengde, Hengshui, Tangshan and Zhangjiakou, exhibited a decline in the AP of O_3 between 2014 and 2019. The BTH region displays a pattern of alternating increases and decreases in the concentration of O_3, with half of the cities showing an upward trend and the other half a downward one. However, Tianjin lacks the requisite data on the concentration of O_3, and thus has been excluded from this analysis. Furthermore, during the period of the global pandemic caused by COVID-19, the AP of O_3 in all cities demonstrated a downward trajectory. However, most cities exhibited an increasing or fluctuating trend during the 2021 and 2022 periods. In 2023, following the global pandemic of COVID-19, only the city of Baoding and the city of Qinhuangdao exhibited a decline in the annual mean value of O_3's AP, while the remaining ten cities demonstrated a slight increase. The AP of O_3 in eight cities, namely Beijing, Cangzhou, Handan, Langfang, Qinhuangdao, Shijiazhuang, Tangshan, and Xingtai, all exhibited an increase (0.58%, 0.29%, 0.86%, 1.16%, 0.77%,

0.24%, 0.05%, 1.42%) in 2023 compared to 2014[190].

The aforementioned analyses demonstrate a general decline in AP for particulate matter, with PM_{10} exhibiting the most pronounced reduction. Moreover, both $PM_{2.5}$ and PM_{10} were impacted by COVID-19, with their respective APs exhibiting reduced levels during the period of the pandemic. Following the relaxation of restrictions associated with the COVID-19 pandemic, there has been a resurgence in several locations. However, when compared with the data from 2019 prior to the emergence of the virus, the overall trend is still downward. The air quality management strategies employed in the BTH region have been instrumental in reducing the AP of particulate matter. The APs for gaseous pollutants are at minimal levels and do not present a significant risk to human health. Notably, NO_2 exhibits the most pronounced decline, with the AP reaching zero in numerous cities and other cities also reducing to a minimal level of approximately 1%. O_3 is a particularly noteworthy pollutant due to its sensitivity to various environmental factors, including temperature. While the overall decrease in O_3 levels is not significant, there are instances where its AP has increased, particularly in some cities. However, its AP remains at a relatively low level, indicating that it does not pose a significant health risk. Furthermore, the impact of the pandemic on NO_2 was minimal, with only five cities exhibiting a modest rebound following the onset of the pandemic. The impact of the COVID-19 pandemic on O_3 levels has been relatively minor. While there has been a slight decline in O_3 levels across all cities during the period under review, this decrease has not reached a statistically significant level. In addition, the observed rebound in O_3 levels in 2023 after the pandemic is minimal.

In general, the larger APs are primarily associated with $PM_{2.5}$, and despite years of intervention, APs in cities other than Zhangjiakou still exceed 40%, which should be the primary focus of treatment. The overall decline in APs for PM_{10} is notable; however, seven cities still exhibit APs above 30%. The APs for gaseous pollutants are all low and have minimal impact on human health. In addition, compared with other studies[31], [262], this study supplemented the changes in AP in the BTH region affected by

COVID-19 by adding the changes in AP after COVID-19 control was lifted.

6.4.2 Attributable proportion of respiratory

In addition to evaluating the AP of each pollutant on total mortality, AirQ+ also assesses the AP of specific diseases[216]. Figure 6-6 illustrates the results of the AP of respiratory. Four pollutants were identified as exerting an influence on the AP of respiratory: $PM_{2.5}$, PM_{10}, NO_2 and O_3. The figure illustrates the annual fluctuations in the four pollutants' impact on respiratory AP between 2014 and 2023. As illustrated in Figure 6-6, particulate matter exhibits a markedly elevated AP, whereas gaseous pollutants display a comparatively lower AP.

In 2014, the level of air pollution in the BTH region was particularly concerning, as evidenced by the data. The average AP of $PM_{2.5}$ on respiratory reached 71.61%, with some regions, such as Baoding and Xingtai, exceeding 80%. Following years of intervention, by 2019, all cities had achieved a notable reduction in AP, with an average reduction of 15%. The city of Baoding, which had the highest AP in 2014, demonstrated a reduction from 82.01% to 61.08%, representing a 20% decrease. The APs in Chengde and Zhangjiakou were reduced to below 50% by 2019. These findings suggest that the implementation of air pollution control measures has been effective in reducing the AP of $PM_{2.5}$ on respiratory. In the subsequent years, 2020 to 2022, the implementation of control measures to combat the spread of COVID-19 in the BTH area led to a notable decline in the average AP for all three years. In 2020, only Handan and Shijiazhuang exhibited APs above 60%, while four cities (Beijing, Chengde, Qinhuangdao, and Zhangjiakou) demonstrated APs below 50%. In 2021, with the exception of Zhangjiakou, Chengde, and Qinhuangdao, where the AP increased slightly, the remaining ten cities all exhibited a decrease, with the AP of all cities falling below 60% and five cities below 50%. By 2022, only Baoding's AP had increased by 0.2% in

comparison to 2021. The remaining 12 cities exhibited a decrease to varying degrees, resulting in an average AP below 50% for the year, at 49.03%. The number of cities with an AP below 50% reached seven, with Zhangjiakou exhibiting the lowest AP of all cities in all years, at 35.80%. This demonstrates that the control measures implemented during the COVID-19 pandemic had a favourable and substantial impact on the improvement of the AP of $PM_{2.5}$ on respiratory outcomes. Following the lifting of all restrictions on the COVID-19 in the BTH area in 2023, the AP demonstrated a resurgence. The mean value increased by 3% in comparison to the preceding year, with only Handan and Shijiazhuang exhibiting a slight decrease, while the remaining 12 cities demonstrated an increase. Notable increases were observed in Beijing (46.69%, up from 40.68% in 2022), Zhangjiakou (43.01%, up from 35.80%), and Tianjin (49.65%, up from 44.62%). Furthermore, the number of cities with APs below 50% also decreased, reaching a total of five. Furthermore, when compared with the figures from 2019, which were recorded prior to the onset of the pandemic, it is evident that the current levels remain lower overall[216]. The mean value was 4% lower, and all cities exhibited a varying degree of decline in the AP. This demonstrates that, when the impact of the closure caused by the pandemic is excluded, the AP of the 13 cities in the BTH region continues to demonstrate a downward trend following 2019[137]. This illustrates the significant effect of air control.

The overall effect of PM_{10} on AP of the respiratory system is greater than that of $PM_{2.5}$, which is somewhat different from the effect of these two particles on total mortality. The respiratory system is more affected by PM_{10}[263]. In 2014, the AP of PM_{10} on respiratory had a markedly elevated impact, with a mean value of 85.56% observed in 13 cities within the BTH region. In excess of 90% of the total number of cases were recorded in six cities, with the highest incidence occurring in Xingtai, where 97.76% of cases were reported. Following a lengthy period of intervention, the AP in the cities exhibited a marked improvement in 2019, with the mean value declining from 85.56% to 59.55%, representing a reduction of 26.01%. The AP in Beijing, Chengde, Tianjin

and Zhangjiakou fell below 50%. The data indicates a notable decline in the AP during the period of the global pandemic caused by COVID-19, with a 9% reduction in 2020 compared to the previous year, a 5% increase in 2021 compared to the preceding year, and a further 11% reduction in 2022 compared to the previous year. In 2020, the AP of Beijing, Chengde, Qinhuangdao, Tianjin, and Zhangjiakou fell below 40%, representing a notable improvement. In 2021, however, no city had an AP below 40%, which may be attributed to the fact that there were fewer closures in 2021[264]. In 2022, the number of closures increased once more, with the APs of Chengde and Zhangjiakou exceeding 30% (29.62% and 27.19%, respectively). Furthermore, the average AP of 13 cities also declined to 45.07%. Following the lifting of COVID-19 restrictions in 2023, the AP saw a notable recovery, with an average increase of 18%. The most severe case is that of Xingtai, which saw an increase to 92.21%, a figure that is cause for serious concern. Furthermore, in comparison to 2019, AP has increased by 4%. Upon the removal of the impact of the COVID-19 pandemic, the impact of PM_{10} on AP of respiratory disease demonstrated an upward trajectory in recent years. Consequently, the control of PM_{10} must be intensified. Moreover, this suggests that respiratory systems are more vulnerable to PM_{10} than $PM_{2.5}$, and therefore more susceptible to the effects of PM_{10}[265].

The AP response to gaseous pollutants is significantly less than that observed with particulate matter. The highest AP for NO_2 was 6.43% in 2014, with an average of 3.93% across the 13 cities. Following several years of treatment, the mean value decreased significantly to 1.76% in 2019. Nine cities exhibited concentrations below the 2% threshold. With the exception of Tangshan, which exhibited a higher value in 2014, there was a 2% decline, yet the city still exhibited the highest concentration in 2019 at 4.05%. Conversely, Zhangjiakou achieved the desired AP of 0 for NO_2 in 2019. This is consistent with Wu's study[266] that Zhangjiakou's air quality is the best in the BTH region. Following the period of the COVID-19 (2020-2022), the overall AP declines again. In 2020, only two cities had an AP higher than 2%, with the mean AP decreasing

to 1.09%. There was a slight increase in 2021, to 1.23%, but this was minimal, and only Tangshan's AP exceeded 2%. In 2022, the AP for NO_2 was even lower, with the mean dropping to 0.49%. The highest city, Baoding, had an AP of only 1.47%. Furthermore, Beijing, Chengde, Hengshui, Tianjin and Zhangjiakou have their APs reduced to 0, which is optimal for the health risk of NO_2. In 2023, following the lifting of restrictions on the spread of COVID-19, the number of cities with an AP of 0 decreased to three. However, the mean value remained at 0.49%, and the overall AP was not affected by the increase in human activity. In particular, Zhangjiakou is the most optimal location for NO_2 levels, where the AP has been 0 since 2019. When the changes over the years are combined, it can be seen that the AP for NO_2 has experienced a significant decrease in 2019 and 2022, and that the impact of COVID-19 has been a contributing factor in this decrease in AP. In addition, the data subsequent to the advent of COVID-19 demonstrates that the impact of NO_2 management in the BTH region is sustained. Once the influence of the pandemic is excluded, the AP continues to decline at a gradual pace and has reached a more optimal state.

O_3 is a distinctive pollutant, exhibiting a distinctive pattern of change compared to other pollutants[255]. Fluctuations are more stable from year to year. The AP of O_3 in the BTH area is below 10%, indicating a relatively low level of hazard. From 2014 to 2019, the average value of AP increased from 6.36% to 7.31%. During this period, half of the cities exhibited an increase, with Handan showing a particularly notable rise from 1.06% to 8.53%. It is noteworthy that Tianjin will no longer include the concentration of O_3 in its measurements from 2019 to 2023. This suggests that the treatment of air pollution may have a diminished impact on O_3. However, during the period of the COVID-19, the mean value of AP exhibited a decline. In 2020, the AP of 12 cities exhibited a reduction in comparison to 2019, with the mean value also demonstrating a decrease to 6.03%. In 2021, only three cities (Beijing, Cangzhou, Chengde) exhibited a continued decline, while the remaining cities demonstrated a slight increase. In 2022, compared to the preceding year, the AP decreased in only four cities (Baoding, Handan,

Qinhuangdao, Zhangjiakou), while the average value rose to 6.93%. In 2023, following the relaxation of restrictions associated with the COVID-19 pandemic, the mean value of AP increased to 7.42%, representing the highest recorded figure across all years. Additionally, the AP of ten cities demonstrated an upward trajectory. The results demonstrate that the correlation between O_3 and other pollutants is relatively weak, with other pollutants exhibiting a notable decline following years of treatment, whereas O_3 displays a more stable fluctuation. The control of COVID-19 has been observed to have a positive effect on the reduction of O_3, and the reduction of human activities has been shown to result in a reduction in the AP of O_3[259].

The results demonstrate a declining trend in the AP of particulate pollutants on the respiratory system over time, with a reduction in AP during the period of COVID-19 and an increase following the pandemic compared to the period of the pandemic itself. However, when the impact of the pandemic is removed, the AP trend remains downward. This shows that the region's efforts to control air pollution continue to be effective, and the risk posed by particulate pollutants to the respiratory system is decreasing annually. Among the gaseous pollutants, the trend of the AP of NO_2 on respiratory is similar to that of particulate pollutants $PM_{2.5}$ and PM_{10}. The impact of the pandemic on NO_2 is comparable to that observed for the particulate pollutants, with a notable decline during the pandemic and a subsequent, albeit slight, increase following the relaxation of restrictions. However, the AP of NO_2 continues to exhibit a gradual decline compared to the period preceding the pandemic. The situation with O_3 is distinctive. The stringent restrictions on human activities resulted in a notable reduction in the AP of O_3. However, following the deregulation, the AP of O_3 did not decline. Nevertheless, the overall low AP of gaseous pollutants exerts a negligible effect on respiratory function in comparison to particulate pollutants. This study complements the relevant research in the literature on the changes in AP after COVID-19, and provides a more comprehensive and long-term data change analysis for air pollution research in the BTH region[17], [31], [55].

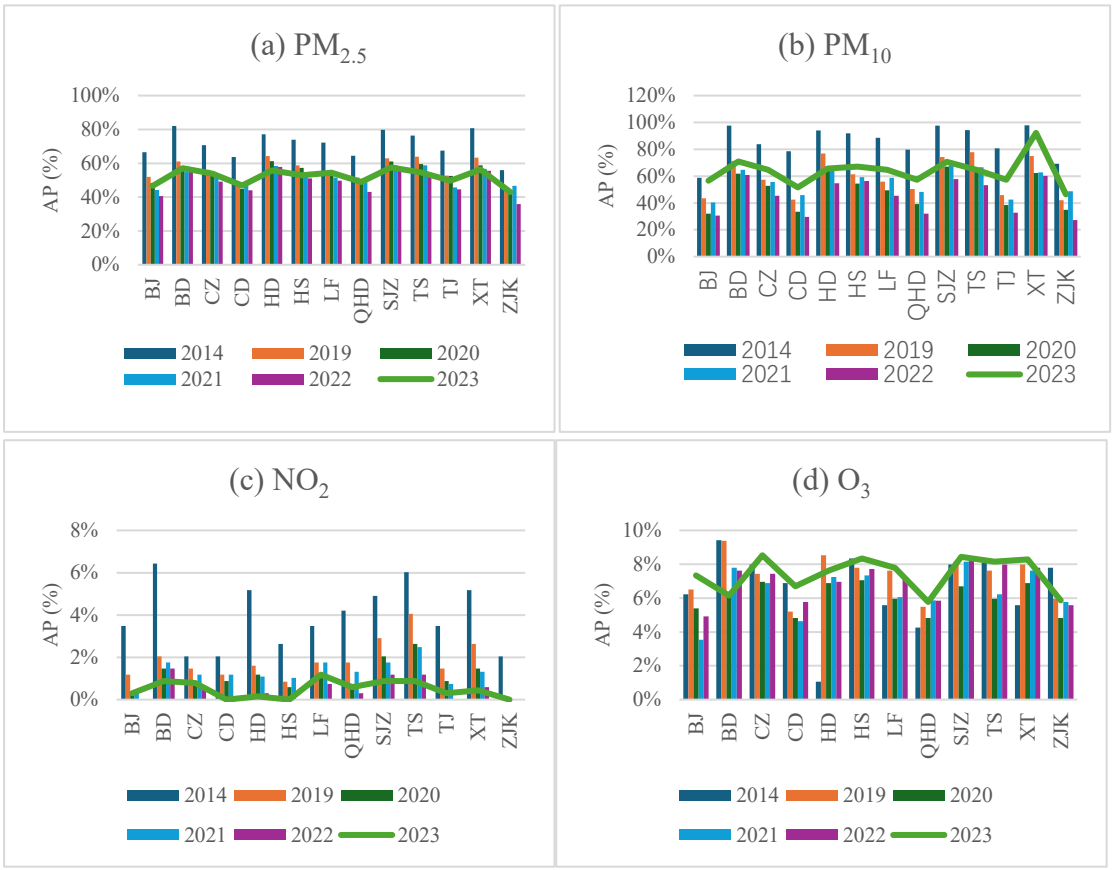


Figure 6-6: Annual changes in Attributable Proportion of respiratory in the BTH region, 2014, 20219-2023. Where (a) is $PM_{2.5}$, (b) is PM_{10}, (c) is NO_2, and (d) is O_3.

6.4.3 Attributable proportion of circulatory

AirQ+ also assesses the AP of $PM_{2.5}$ to the circulatory system[216]. Figure 6-7 illustrates the change in the AP of $PM_{2.5}$ to the circulatory system in the BTH region between the years 2014 and 2019, and subsequently to 2023.

In 2014, the mean value of the AP of $PM_{2.5}$ in the circulatory system was as high as 74.72%, with the AP of Baoding, Handan, Shijiazhuang, and Xingtai all exceeding 80%. The respective values were 84.72%, 80.16%, 82.68%, and 83.56%, which has been identified as a significant risk factor for circulatory. Following years of treatment, the mean AP value for the BTH region in 2019 decreased significantly to 59.21%,

representing a notable decline. Moreover, the highest AP also decreased from a value exceeding 80% to 66.57% in Xingtai. Furthermore, the AP in Zhangjiakou decreased to 46.54%. Overall, the control of air pollution in the region has resulted in a notable reduction in the impact of pollutants on circulatory. From 2020 to 2022, during the COVID-19 period, the AP on circulatory shows a decreasing trend year by year, with figures of 59.21%, 57.05% and 55.65%, respectively. The largest decrease was observed in 2020, with the AP decreasing in 12 cities and only Tangshan exhibiting a slight increase. The data for 2021 indicates a slight increase in the cases of Chengde, Qinhuangdao, and Zhangjiakou, while the overall trend shows a decrease in the number of 12 cities. In 2022, the control of COVID-19 was more stringent than that in 2021. Consequently, in 2022, only Baoding exhibited a slight increase, while the other 12 cities demonstrated a downward trend. In 2023, following the relaxation of restrictions on the spread of COVID-19, the AP of circulatory increased slightly to 55.34%, representing the only instance in which this value exceeded that observed in 2022. Overall, the AP in 2023 is only higher than that in 2022, which suggests that the control of air pollution has significantly reduced the hazard to circulatory. The control of COVID-19 played a beneficial role in reducing the risk to circulatory. However, the exclusion of COVID-19 still demonstrated a declining effect on the hazard to circulatory on an annual basis[267]. The continued control of air pollution in the region resulted in a yearly decrease in the AP for the circulatory system.

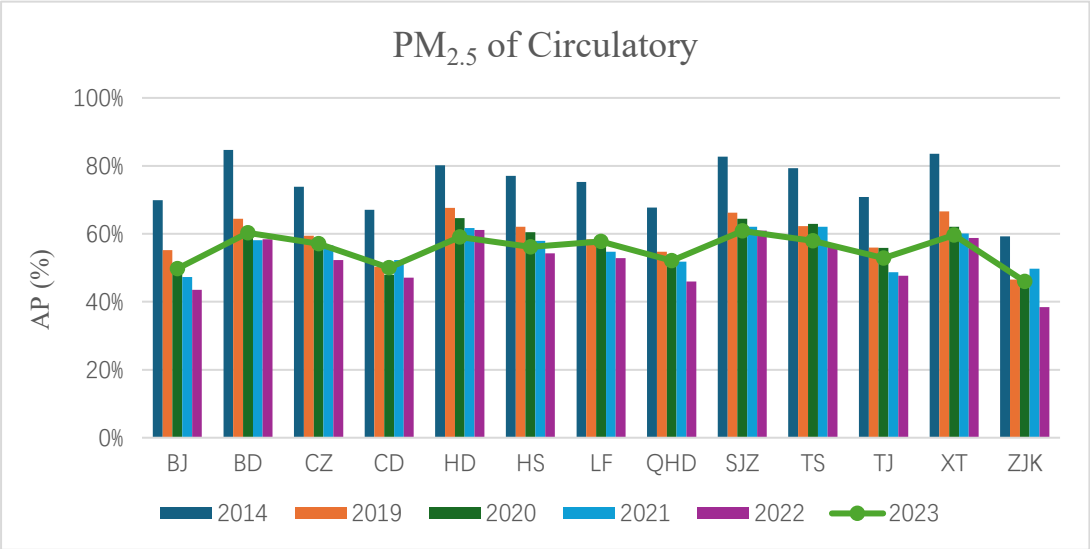


Figure 6-7: Annual changes in the attributable proportion of $PM_{2.5}$ to circulatory diseases in the 13 cities of the Beijing-Tianjin-Hebei region, 2014, 2019-2023.

6.5 Chapter summary

Existing studies on changes in air pollution in the BTH region focus on changes in air pollution before or during the COVID-19 pandemic in 2020. They cannot comprehensively assess the impact of COVID-19 on air pollution control and the improvement of air pollution after removing the impact of COVID-19. Or they focus on individual cities, and the data only includes changes in air quality, without a comprehensive analysis of the association with related health risks.

This study addresses the gap in existing research by incorporating air pollution changes in the BTH region not only during the COVID-19 period but also after it. Additionally, it tackles the lack of analysis on the evolving health risks associated with air pollution, making the study more comprehensive. Therefore, a thorough analysis of air quality improvements and related health risk reductions in the Beijing-Tianjin-Hebei (BTH) region from 2014 to 2023—especially covering air quality changes during and after the

COVID-19 period—holds significant value for policymakers and researchers in guiding future air pollution control efforts.

To address these issues, this study first analyzes the frequency of each pollutant to understand the variation patterns of major pollutants in 13 cities of the Beijing-Tianjin-Hebei region over the past decade. Subsequently, the quarterly fluctuations of pollutants in the 13 cities are examined. The results demonstrate a discernible improvement in air quality across a range of urban centers, with a notable decline in the concentration of pollutants, particularly between 2020 and 2022. During this period, the advent of the COVID-19 led to a substantial reduction in the levels of air pollutants in all monitored areas. Furthermore, following the conclusion of the COVID-19 in 2023, the concentration of pollutants in all cities exhibited a range of increases. However, when compared with 2019, the overall trend remained downward. Following the removal of the impact of COVID-19, the control of air pollution has demonstrated a positive influence. Next, this chapter presents and analyses the annual changes in AQI from 2014 to 2023 in the BTH region in a map format. The results demonstrate the geographical distribution of pollutant concentrations in BTH region, the improvement of pollution in different regions, and the changes in different regions before and after COVID-19. These findings provide meaningful references for the regional management of air pollution.

In addition to analysing the spatiotemporal variations in air pollution, this study also examines air pollution-related health risks to provide a more comprehensive assessment. Specifically, it analyzes the annual fluctuations in AP due to long-term exposure to various pollutants, including total mortality, respiratory diseases, and circulatory diseases. Among them, the AP of the total mortality rate of $PM_{2.5}$ dropped from 64.02% in 2014 to 44.87% in 2023, and the AP of the total mortality rate of PM_{10} dropped from 44.87% in 2014 to 28.95% in 2023.The results demonstrated that the pollutants were markedly detrimental to human health in each region in 2014, the inaugural year of air pollution control, and exhibited a gradual decline over subsequent years. Between 2020

and 2022, the AP of each pollutant declined to very low levels in all cities, with varying degrees of decline. In 2023, following the relaxation deregulation of COVID-19, the health effects of each pollutant rebounded slightly, but still showed a decreasing trend compared to the situation before COVID-19. The region's air pollution control measures continue to play a role in reducing the health impacts of pollutants after the removal of the pandemic. This trend was similarly observed in the period preceding and following the pandemic.

In general, this study incorporated data on changes in air pollution and related health risks after the outbreak, which will help to more comprehensively assess the impact of the epidemic on air pollutants, as well as comprehensively evaluate the long-term effects of air pollution prevention and control in the region after excluding the impact of COVID-19.

Chapter 7. Conclusion and future work

This thesis proposed the prediction methods of multiple air pollution indices based on air pollution data in Beijing city, air pollution prediction models that add spatial information as parameters to improve the prediction performance, and the trend of air pollution in Beijing-Tianjin-Hebei region after treatment. In this chapter, important findings and contributions are summarised and relevant recommendations for future work are discussed.

7.1 Main findings

The main conclusions are summarized as follows:

1. Simultaneous Prediction of Multiple Air Pollution Indices

- Developed a model to predict AQI, AAQI, and HAQI simultaneously, providing a more comprehensive evaluation than using AQI alone.
- Strong correlations exist among AQI, AAQI and HAQI, indicating the complementary value of the additional indices.

2. Multi-Model Prediction Approach

- Proposed a combined prediction framework using LSTM, MLP, and Transformer networks, with automatic selection of the best-performing model.
- The approach improves prediction accuracy for 1-hour, 3-hour, and 6-hour forecasts as evidenced by MAE, MSE, and R^2 metrics.

3. Incorporation of Spatial Information (GCN+LSTM)

- Introduced spatial information extracted from historical data correlations to enhance day-ahead AQI prediction.
- This approach simplifies data collection while achieving superior predictive performance compared to conventional models.

4. Long-Term Air Quality Trends (2014–2023)

- Sustained air pollution control measures have significantly reduced $PM_{2.5}$ and PM_{10} concentrations.
- COVID-19 lockdowns temporarily enhanced air quality, while post-pandemic rebounds were minor, confirming the effectiveness of regional management policies.

5. Spatial Disparities in Pollution Improvement

- Northern cities experienced more pronounced air quality improvements than southern cities, highlighting the need for targeted interventions.

6. Health Impacts of Air Pollution

- Attributable proportions of major pollutants to total mortality and respiratory/circulatory diseases have declined over time.
- The study confirms that long-term policy interventions effectively reduce both air pollution and associated health risks.

Overall, this thesis provides a comprehensive framework for air pollution prediction and impact assessment, demonstrating both methodological innovation and practical relevance for regional air quality management.

7.2 Future work

A high level of prediction has been achieved for multiple air pollution indices. In future research, the prediction performance can be improved by enhancing the prediction model and incorporating pertinent parameters, particularly over extended periods. In addition, the proposed model incorporating spatial information enhances the prediction of air pollution. In subsequent research, the integration of additional parameters, such as industrial production and traffic conditions, may be considered to further improve the model's predictive performance or extend its application to a broader range of areas for evaluation. In future studies, it would be beneficial to consider incorporating air quality changes over a longer time period following the onset of pandemic in the BTH region. This would allow for a more comprehensively analysis of the impact of COVID-19 on air quality and the effectiveness of air pollution control measures in the region after accounting for the effects of the pandemic.

7.2.1 Multi-indices prediction

This study introduces, for the first time, a novel methodology for the simultaneous prediction of multiple air pollution indices using multiple neural network models. Experimental results demonstrate that this approach achieves high predictive performance and effectively forecasts multiple air pollution indices. However, the model's predictive accuracy for the next 6-hour period remains limited. Future research could focus on optimizing the model to enhance its performance and exploring more advanced machine learning or neural network architectures to enable longer-term predictions. Additionally, extending the application of the model to a broader range of cities could improve its generalizability and practical applicability.

7.2.2 Air pollution prediction

The GCN+LSTM model proposed in this study enhances air pollution prediction by efficiently integrating spatial information as a parameter, simplifying the data collection process while improving predictive performance. Future research could explore incorporating additional air pollution-related parameters, such as industrial production levels, energy consumption patterns, and real-time traffic conditions. Leveraging advanced neural network architectures could further streamline the integration of these parameters, enhancing both efficiency and predictive accuracy. Moreover, extending the model's application to a broader range of cities or regions, particularly those with distinct air pollution patterns (e.g., coastal cities), would improve its generalizability and adaptability to diverse environmental contexts.

7.2.3 Air pollution changes

This study systematically analyzed air pollution trends in the Beijing-Tianjin-Hebei region over the past decade, starting from 2014, following the implementation of air pollution control measures. It also examined the interplay between air pollution changes and health-related risk factors during the COVID-19 pandemic and six months post-pandemic. Future research could extend the analysis to cover longer-term pollution trends following the pandemic, providing deeper insights into sustained changes in air quality. Additionally, comparative studies between the Beijing-Tianjin-Hebei region and other urban agglomerations, both within China and globally, could help identify similarities and differences in air quality responses to pandemic-related restrictions and subsequent economic recovery, offering valuable perspectives for broader

environmental and public health strategies.

Appendix A

Respiratory

Beijing total				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	58.73%	60.29%	2.35%	3.18%
2019	44.71%	45.12%	0.79%	3.33%
2020	41.34%	12.48%	0.26%	2.75%
2021	37.61%	16.34%	0.94%	1.79%
2022	34.41%	11.83%	0	2.51%
2023	39.83%	25.04%	0.20%	3.76%
Baoding				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	74.98%	72.37%	4.36%	4.85%
2019	49.98%	28.35%	0.99%	3.13%
2020	49.98%	28.35%	0.99%	3.13%
2021	47.41%	30.29%	1.18%	4%
2022	47.61%	27.79%	0.99%	3.90%
2023	49.39%	34.79%	0.59%	3.13%
Cangzhou				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	62.80%	46.61%	1.38%	4.09%
2019	48.61%	25.48%	0.99%	3.81%
2020	47.00%	22.81%	0.49%	3.57%
2021	45.97%	24.45%	0.79%	3.52%
2022	42.10%	18.93%	0.40%	3.81%
2023	46.39%	30.29%	0.53%	4.38%
Chengde				

	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	55.94%	41.34%	1.38%	3.52%
2019	40.29%	17.48%	0.79%	2.65%
2020	38.18%	13.17%	0.59%	2.46%
2021	42.10%	19.09%	0.79%	2.36%
2022	37.47%	11.45%	0.00%	2.94%
2023	40.06%	22.20%	0.00%	3.42%
Handan				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	69.67%	62.34%	3.50%	3.04%
2019	56.45%	39.82%	1.08%	4.38%
2020	53.50%	31.91%	0.79%	3.52%
2021	50.74%	32.31%	0.73%	3.71%
2022	50.17%	24.01%	0.20%	3.57%
2023	48.21%	30.83%	0.10%	3.90%
Hengshui				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	66.21%	58.13%	1.77%	4.28%
2019	51.12%	28.07%	0.57%	4.00%
2020	49.59%	23.86%	0.40%	3.61%
2021	47.21%	26.64%	0.69%	3.67%
2022	43.85%	24.90%	0.00%	3.95%
2023	45.56%	31.91%	0.00%	4.28%
Langfang				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	64.34%	52.72%	2.35%	2.84%
2019	47.61%	24.60%	1.18%	3.90%
2020	45.56%	20.97%	0.79%	3.04%

2021	44.28%	26.36%	1.18%	3.09%
2022	42.54%	18.93%	0.49%	3.71%
2023	47.00%	30.29%	0.79%	4.00%
Qinhuangdao				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	56.61%	42.48%	2.83%	2.17%
2019	44.28%	21.59%	1.18%	2.80%
2020	41.20%	15.85%	0.49%	2.46%
2021	41.65%	20.35%	0.89%	2.99%
2022	36.50%	12.48%	0.20%	2.98%
2023	41.88%	25.48%	0.40%	2.94%
Shijiazhuang				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	72.55%	72.16%	3.31%	4.09%
2019	55.08%	37.54%	1.96%	4.19%
2020	53.32%	31.91%	1.38%	3.42%
2021	51.12%	32.97%	1.18%	4.19%
2022	49.98%	25.78%	0.79%	4.28%
2023	49.90%	34.53%	0.59%	4.33%
Tangshan				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	68.72%	62.78%	4.07%	4.14%
2019	56.11%	40.65%	2.73%	3.90%
2020	51.86%	31.38%	1.77%	3.04%
2021	51.12%	31.64%	1.67%	3.18%
2022	45.97%	23.11%	0.79%	4.09%
2023	47.21%	30.29%	0.59%	4.19%
Tianjin				

	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	59.67%	43.37%	2.35%	
2019	45.35%	19.09%	0.99%	
2020	45.26%	15.52%	0.59%	
2021	38.89%	17.48%	0.49%	
2022	37.95%	12.83%	0.00%	
2023	42.54%	25.48%	0.20%	
Xingtai				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	73.59%	73.12%	3.50%	2.84%
2019	55.43%	38.03%	1.77%	4.09%
2020	51.12%	28.63%	0.99%	3.52%
2021	49.20%	28.91%	0.89%	3.90%
2022	48.01%	27.22%	0.40%	4.00%
2023	48.81%	35.80%	0.30%	4.26%
Zhangjiakou				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	48.41%	33.50%	1.38%	4.00%
2019	36.98%	17.16%	0.00%	3.04%
2020	36.25%	13.85%	0.00%	2.46%
2021	39.83%	20.66%	0.00%	2.94%
2022	30.08%	10.40%	0.00%	2.48%
2023	36.50%	19.40%	0.00%	2.99%

Appendix B

Circulatory

Beijing total				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	58.73%	60.29%	2.35%	3.18%
2019	44.71%	45.12%	0.79%	3.33%
2020	41.34%	12.48%	0.26%	2.75%
2021	37.61%	16.34%	0.94%	1.79%
2022	34.41%	11.83%	0	2.51%
2023	39.83%	25.04%	0.20%	3.76%
Baoding				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	74.98%	72.37%	4.36%	4.85%
2019	49.98%	28.35%	0.99%	3.13%
2020	49.98%	28.35%	0.99%	3.13%
2021	47.41%	30.29%	1.18%	4%
2022	47.61%	27.79%	0.99%	3.90%
2023	49.39%	34.79%	0.59%	3.13%
Cangzhou				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	62.80%	46.61%	1.38%	4.09%
2019	48.61%	25.48%	0.99%	3.81%
2020	47.00%	22.81%	0.49%	3.57%
2021	45.97%	24.45%	0.79%	3.52%
2022	42.10%	18.93%	0.40%	3.81%
2023	46.39%	30.29%	0.53%	4.38%

Chengde				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	55.94%	41.34%	1.38%	3.52%
2019	40.29%	17.48%	0.79%	2.65%
2020	38.18%	13.17%	0.59%	2.46%
2021	42.10%	19.09%	0.79%	2.36%
2022	37.47%	11.45%	0.00%	2.94%
2023	40.06%	22.20%	0.00%	3.42%
Handan				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	69.67%	62.34%	3.50%	3.04%
2019	56.45%	39.82%	1.08%	4.38%
2020	53.50%	31.91%	0.79%	3.52%
2021	50.74%	32.31%	0.73%	3.71%
2022	50.17%	24.01%	0.20%	3.57%
2023	48.21%	30.83%	0.10%	3.90%
Hengshui				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	66.21%	58.13%	1.77%	4.28%
2019	51.12%	28.07%	0.57%	4.00%
2020	49.59%	23.86%	0.40%	3.61%
2021	47.21%	26.64%	0.69%	3.67%
2022	43.85%	24.90%	0.00%	3.95%
2023	45.56%	31.91%	0.00%	4.28%
Langfang				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	64.34%	52.72%	2.35%	2.84%
2019	47.61%	24.60%	1.18%	3.90%

2020	45.56%	20.97%	0.79%	3.04%
2021	44.28%	26.36%	1.18%	3.09%
2022	42.54%	18.93%	0.49%	3.71%
2023	47.00%	30.29%	0.79%	4.00%
Qinhuangdao				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	56.61%	42.48%	2.83%	2.17%
2019	44.28%	21.59%	1.18%	2.80%
2020	41.20%	15.85%	0.49%	2.46%
2021	41.65%	20.35%	0.89%	2.99%
2022	36.50%	12.48%	0.20%	2.98%
2023	41.88%	25.48%	0.40%	2.94%
Shijiazhuang				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	72.55%	72.16%	3.31%	4.09%
2019	55.08%	37.54%	1.96%	4.19%
2020	53.32%	31.91%	1.38%	3.42%
2021	51.12%	32.97%	1.18%	4.19%
2022	49.98%	25.78%	0.79%	4.28%
2023	49.90%	34.53%	0.59%	4.33%
Tangshan				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	68.72%	62.78%	4.07%	4.14%
2019	56.11%	40.65%	2.73%	3.90%
2020	51.86%	31.38%	1.77%	3.04%
2021	51.12%	31.64%	1.67%	3.18%
2022	45.97%	23.11%	0.79%	4.09%
2023	47.21%	30.29%	0.59%	4.19%

Tianjin				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	59.67%	43.37%	2.35%	
2019	45.35%	19.09%	0.99%	
2020	45.26%	15.52%	0.59%	
2021	38.89%	17.48%	0.49%	
2022	37.95%	12.83%	0.00%	
2023	42.54%	25.48%	0.20%	
Xingtai				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	73.59%	73.12%	3.50%	2.84%
2019	55.43%	38.03%	1.77%	4.09%
2020	51.12%	28.63%	0.99%	3.52%
2021	49.20%	28.91%	0.89%	3.90%
2022	48.01%	27.22%	0.40%	4.00%
2023	48.81%	35.80%	0.30%	4.26%
Zhangjiakou				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	48.41%	33.50%	1.38%	4.00%
2019	36.98%	17.16%	0.00%	3.04%
2020	36.25%	13.85%	0.00%	2.46%
2021	39.83%	20.66%	0.00%	2.94%
2022	30.08%	10.40%	0.00%	2.48%
2023	36.50%	19.40%	0.00%	2.99%

Appendix C

Total mortality

Beijing total				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	58.73%	60.29%	2.35%	3.18%
2019	44.71%	45.12%	0.79%	3.33%
2020	41.34%	12.48%	0.26%	2.75%
2021	37.61%	16.34%	0.94%	1.79%
2022	34.41%	11.83%	0	2.51%
2023	39.83%	25.04%	0.20%	3.76%
Baoding				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	74.98%	72.37%	4.36%	4.85%
2019	49.98%	28.35%	0.99%	3.13%
2020	49.98%	28.35%	0.99%	3.13%
2021	47.41%	30.29%	1.18%	4%
2022	47.61%	27.79%	0.99%	3.90%
2023	49.39%	34.79%	0.59%	3.13%
Cangzhou				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	62.80%	46.61%	1.38%	4.09%

2019	48.61%	25.48%	0.99%	3.81%
2020	47.00%	22.81%	0.49%	3.57%
2021	45.97%	24.45%	0.79%	3.52%
2022	42.10%	18.93%	0.40%	3.81%
2023	46.39%	30.29%	0.53%	4.38%
Chengde				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	55.94%	41.34%	1.38%	3.52%
2019	40.29%	17.48%	0.79%	2.65%
2020	38.18%	13.17%	0.59%	2.46%
2021	42.10%	19.09%	0.79%	2.36%
2022	37.47%	11.45%	0.00%	2.94%
2023	40.06%	22.20%	0.00%	3.42%
Handan				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	69.67%	62.34%	3.50%	3.04%
2019	56.45%	39.82%	1.08%	4.38%
2020	53.50%	31.91%	0.79%	3.52%
2021	50.74%	32.31%	0.73%	3.71%
2022	50.17%	24.01%	0.20%	3.57%
2023	48.21%	30.83%	0.10%	3.90%

Hengshui				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	66.21%	58.13%	1.77%	4.28%
2019	51.12%	28.07%	0.57%	4.00%
2020	49.59%	23.86%	0.40%	3.61%
2021	47.21%	26.64%	0.69%	3.67%
2022	43.85%	24.90%	0.00%	3.95%
2023	45.56%	31.91%	0.00%	4.28%
Langfang				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	64.34%	52.72%	2.35%	2.84%
2019	47.61%	24.60%	1.18%	3.90%
2020	45.56%	20.97%	0.79%	3.04%
2021	44.28%	26.36%	1.18%	3.09%
2022	42.54%	18.93%	0.49%	3.71%
2023	47.00%	30.29%	0.79%	4.00%
Qinhuangdao				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	56.61%	42.48%	2.83%	2.17%
2019	44.28%	21.59%	1.18%	2.80%
2020	41.20%	15.85%	0.49%	2.46%

2021	41.65%	20.35%	0.89%	2.99%
2022	36.50%	12.48%	0.20%	2.98%
2023	41.88%	25.48%	0.40%	2.94%
Shijiazhuang				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	72.55%	72.16%	3.31%	4.09%
2019	55.08%	37.54%	1.96%	4.19%
2020	53.32%	31.91%	1.38%	3.42%
2021	51.12%	32.97%	1.18%	4.19%
2022	49.98%	25.78%	0.79%	4.28%
2023	49.90%	34.53%	0.59%	4.33%
Tangshan				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	68.72%	62.78%	4.07%	4.14%
2019	56.11%	40.65%	2.73%	3.90%
2020	51.86%	31.38%	1.77%	3.04%
2021	51.12%	31.64%	1.67%	3.18%
2022	45.97%	23.11%	0.79%	4.09%
2023	47.21%	30.29%	0.59%	4.19%
Tianjin				
	$PM_{2.5}$	PM_{10}	NO_2	O_3

2014	59.67%	43.37%	2.35%	
2019	45.35%	19.09%	0.99%	
2020	45.26%	15.52%	0.59%	
2021	38.89%	17.48%	0.49%	
2022	37.95%	12.83%	0.00%	
2023	42.54%	25.48%	0.20%	
Xingtai				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	73.59%	73.12%	3.50%	2.84%
2019	55.43%	38.03%	1.77%	4.09%
2020	51.12%	28.63%	0.99%	3.52%
2021	49.20%	28.91%	0.89%	3.90%
2022	48.01%	27.22%	0.40%	4.00%
2023	48.81%	35.80%	0.30%	4.26%
Zhangjiakou				
	$PM_{2.5}$	PM_{10}	NO_2	O_3
2014	48.41%	33.50%	1.38%	4.00%
2019	36.98%	17.16%	0.00%	3.04%
2020	36.25%	13.85%	0.00%	2.46%
2021	39.83%	20.66%	0.00%	2.94%
2022	30.08%	10.40%	0.00%	2.48%

2023	36.50%	19.40%	0.00%	2.99%

Appendix D

Yearly average index for ArcGIS

Year	Baoding	Beijing	Cangzhou	Chengde	Handan	Hengshui	Langfang	Qinhuangdao	Shijiazhuang	Tangshan	Tianjin	Xingtai	Zhangjiakou
2014	235.3759	158.7294	176.72	146.438	209.7347	195.4734	184.5137	148.5458	222.8973	206.4082	162.4725	227.8821	120.332
2015	202.4397	141.6952	155.7984	124.4246	185.0658	185.5835	166.4614	119.8643	185.2	177.601	146.7831	194.2344	120.9841
2016	176.8156	130.4314	146.0761	116.2082	162.2092	167.5024	140.2726	118.0567	184.3548	168.0545	139.9224	170.1551	106.141
2017	170.3532	117.3436	145.5769	107.4602	172.4331	159.8996	136.4761	117.5116	175.4795	155.6137	133.9478	166.5878	104.7539
2018	149.835	107.854	136.6209	101.75	154.758	141.343	127.695	112.75	156.2967	156.846	112.920	152.203	96.86538

	2	4			2	4	1			2	3	3	
2019	137.3941	109.7797	121.302	96.607761	147.85666	129.0672	117.6087	108.4369	142.9086	146.8148	111.6916	143.97761	88.44683
2020	124.96191	99.8110106	116.182	91.313533	137.82667	124.16144	112.09331	98.94009	137.2462	132.7323	111.56007	129.4666	86.19616
2021	117.36681	89.2664648	112.9753	101.64566	128.08852	116.8225	108.294	99.92033	129.1291	128.5659	92.767374	122.66648	94.81319
2022	117.89566	81.8434341	102.3462	88.835166	126.08879	107.4505	103.2308	86.5	124.6429	112.9093	90.2033	118.98835	71.46703
2023	123.866	95.6155	114.215	95.8855	119.59	111.6131	115.87	100.8291	124.575	116.6533	103.4523	121.7688	87.27

Bibliography

[1] WHO, ‘Air pollution’. Accessed: Oct. 19, 2023. [Online]. Available: https://www.who.int/health-topics/air-pollution

[2] ‘Air Pollution and Your Health’. Accessed: Oct. 19, 2023. [Online]. Available: https://www.niehs.nih.gov/health/topics/agents/air-pollution/index.cfm

[3] ‘9 out of 10 people worldwide breathe polluted air, but more countries are taking action’. Accessed: Feb. 01, 2024. [Online]. Available: https://www.who.int/zh/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action

[4] ‘Air Pollution - Definition, Causes, Effects And Control’, BYJUS. Accessed: Feb. 01, 2024. [Online]. Available: https://byjus.com/biology/air-pollution-control/

[5] ‘Ambient (outdoor) air pollution’. Accessed: Oct. 19, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

[6] ‘How is air pollution caused? What are its effects on our health and the environment? | ClientEarth’. Accessed: Oct. 19, 2023. [Online]. Available: https://www.clientearth.org/latest/latest-updates/stories/how-is-air-pollution-caused/

[7] ‘United Nations EcoNOmic Commission for Europe’. Accessed: Oct. 19, 2023. [Online]. Available: https://unece.org/air-pollution-and-ecoNOmic-develoPMent

[8] ‘EcoNOmy and air pollution - Clean Air Fund’. Accessed: Oct. 19, 2023. [Online]. Available: https://www.cleanairfund.org/theme/ecoNOmics/

[9] R. Heinecke, ‘Why is air pollution a bigger topic in developing countries than in developed nations?’, Breeze TechNOlogies. Accessed: Oct. 19, 2023. [Online]. Available: https://www.breeze-techNOlogies.de/blog/why-is-air-pollution-bigger-topic-in-developing-nations-than-in-developed-nations/

[10] S. Cai, Y. Wang, B. Zhao, S. Wang, X. Chang, and J. Hao, ‘The impact of the “Air Pollution Prevention and Control Action Plan” on PM2.5 concentrations in Jing-Jin-Ji region during 2012–2020’, *Sci. Total Environ.*, vol. 580, pp. 197–209, Feb. 2017, doi: 10.1016/j.scitotenv.2016.11.188.

[11] K. Yu, ‘Fuzzy Algorithm and Air Quality Ranking—Taking 18 Urban Cities in China as an Example’, *Water. Air. Soil Pollut.*, vol. 234, NO. 11, p. 704, NOv. 2023, doi: 10.1007/s11270-023-06710-0.

[12] A. L. Association, ‘Air Quality Index’. Accessed: NOv. 21, 2023. [Online]. Available: https://www.lung.org/clean-air/outdoors/air-quality-index

[13] P. K. Swamee and A. Tyagi, ‘Formation of an Air Pollution Index’, *J. Air Waste Manag. Assoc.*, vol. 49, NO. 1, pp. 88–91, Jan. 1999, doi: 10.1080/10473289.1999.10463776.

[14] E. K. Cairncross, J. John, and M. Zunckel, ‘A NOvel air pollution index based on the relative risk of daily mortality associated with short-term exposure to common air pollutants’, *Atmos. Environ.*, vol. 41, NO. 38, pp. 8442–8454, Dec. 2007, doi: 10.1016/j.atmosenv.2007.07.003.

[15] J. Hu, Q. Ying, Y. Wang, and H. Zhang, ‘Characterizing multi-pollutant air pollution in China: Comparison of three air quality indices’, *Environ. Int.*, vol. 84, pp. 17–25, NOv. 2015, doi: 10.1016/j.envint.2015.06.014.

[16] H. Luo *et al.*, ‘Air pollution characteristics and human health risks in key cities of NOrthwest China’, *J. Environ. Manage.*, vol. 269, p. 110791, Sept. 2020, doi: 10.1016/j.jenvman.2020.110791.

[17] ‘Air pollution characteristics and health risks in Henan Province, China - ScienceDirect’. Accessed: NOv. 21, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013935117307430

[18] C. Xu, Z. Zhang, G. Ling, G. Wang, and M. Wang, ‘Air pollutant spatiotemporal evolution characteristics and effects on human health in NOrth China’, *Chemosphere*, vol. 294, p. 133814, May 2022, doi: 10.1016/j.chemosphere.2022.133814.

[19] H. Tularam *et al.*, ‘A hybrid air pollution / land use regression model for predicting air pollution concentrations in Durban, South Africa’, *Environ. Pollut.*, vol. 274, p. 116513, Apr. 2021, doi: 10.1016/j.envpol.2021.116513.

[20] H. Liu, G. Yan, Z. Duan, and C. Chen, ‘Intelligent modeling strategies for forecasting air quality time series: A review’, *Appl. Soft Comput.*, vol. 102, p. 106957, Apr. 2021, doi: 10.1016/j.asoc.2020.106957.

[21] W. C. Leong, R. O. Kelani, and Z. Ahmad, ‘Prediction of air pollution index (API) using support vector machine (SVM)’, *J. Environ. Chem. Eng.*, vol. 8, NO. 3, p. 103208, June 2020, doi: 10.1016/j.jece.2019.103208.

[22] H. Sun *et al.*, ‘Assessing the potential of random forest method for estimating solar radiation using air pollution index’, *Energy Convers. Manag.*, vol. 119, pp. 121–129, July 2016, doi: 10.1016/j.enconman.2016.04.051.

[23] S. Tsokov, M. Lazarova, and A. Aleksieva-Petrova, ‘A Hybrid Spatiotemporal Deep Model Based on CNN and LSTM for Air Pollution Prediction’, *Sustainability*, vol. 14, NO. 9, Art. NO. 9, Jan. 2022, doi: 10.3390/su14095104.

[24] Y.-T. Tsai, Y.-R. Zeng, and Y.-S. Chang, ‘Air Pollution Forecasting Using RNN with LSTM’, in *2018 IEEE 16th Intl Conf on Dependable, AutoNOmic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and TechNOlogy Congress(DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2018, pp. 1074–1079. doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00178.

[25] Li L., Gong J., and Zhou J., ‘Spatial Interpolation of Fine Particulate Matter Concentrations Using the Shortest Wind-Field Path Distance’, *PLOS ONE*, vol. 9, NO. 5, p. e96111, May 2014, doi: 10.1371/journal.pone.0096111.

[26] H. Zhou, F. Zhang, Z. Du, and R. Liu, ‘Forecasting PM2.5 using hybrid graph

convolution-based model considering dynamic wind-field to offer the benefit of spatial interpretability', *Environ. Pollut.*, vol. 273, p. 116473, Mar. 2021, doi: 10.1016/j.envpol.2021.116473.

[27] Y. Qi, Q. Li, H. Karimian, and D. Liu, 'A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory', *Sci. Total Environ.*, vol. 664, pp. 1–10, May 2019, doi: 10.1016/j.scitotenv.2019.01.333.

[28] 'Sustainability | Free Full-Text | Urban Air Pollution, Urban Heat Island and Human Health: A Review of the Literature'. Accessed: NOv. 21, 2023. [Online]. Available: https://www.mdpi.com/2071-1050/14/15/9234

[29] 'The impact of environmental accountability on air pollution: A public attention perspective - ScienceDirect'. Accessed: NOv. 21, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421521005991

[30] T. Feng, H. Du, Z. Lin, X. Chen, Z. Chen, and Q. Tu, 'Green recovery or pollution rebound? Evidence from air pollution of China in the post-COVID-19 era', *J. Environ. Manage.*, vol. 324, p. 116360, Dec. 2022, doi: 10.1016/j.jenvman.2022.116360.

[31] 'Air pollution and its associated health risks before and after COVID-19 in Shaanxi Province, China - ScienceDirect'. Accessed: NOv. 21, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749123000921

[32] H. Liu, F. Yue, and Z. Xie, 'Quantify the role of anthropogenic emission and meteorology on air pollution using machine learning approach: A case study of PM2.5 during the COVID-19 outbreak in Hubei Province, China', *Environ. Pollut.*, vol. 300, p. 118932, May 2022, doi: 10.1016/j.envpol.2022.118932.

[33] C. Gao, F. Zhang, D. Fang, Q. Wang, and M. Liu, 'Spatial characteristics of change trends of air pollutants in Chinese urban areas during 2016–2020: The impact of air pollution controls and the COVID-19 pandemic', *Atmospheric Res.*, vol. 283, p. 106539, Mar. 2023, doi: 10.1016/j.atmosres.2022.106539.

[34] T. V. Vu *et al.*, 'Assessing the contributions of outdoor and indoor sources to air quality in London homes of the SCAMP cohort', *Build. Environ.*, vol. 222, p. 109359, Aug. 2022, doi: 10.1016/j.buildenv.2022.109359.

[35] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, 'Environmental and Health Impacts of Air Pollution: A Review', *Front. Public Health*, vol. 8, 2020, Accessed: Feb. 09, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpubh.2020.00014

[36] P. M. Mannucci and M. Franchini, 'Health Effects of Ambient Air Pollution in Developing Countries', *Int. J. Environ. Res. Public. Health*, vol. 14, NO. 9, Art. NO. 9, Sept. 2017, doi: 10.3390/ijerph14091048.

[37] 'Health consequences of air pollution'. Accessed: Aug. 14, 2024. [Online]. Available: https://www.who.int/news/item/25-06-2024-what-are-health-consequences-of-air-pollution-on-populations

[38] M. Greenstone and R. Hanna, 'Environmental Regulations, Air and Water Pollution, and Infant Mortality in India', *Am. Econ. Rev.*, vol. 104, NO. 10, pp. 3038–3072, Oct. 2014, doi: 10.1257/aer.104.10.3038.

[39] C. Wen *et al.*, 'A NOvel spatiotemporal convolutional long short-term neural network for air pollution prediction', *Sci. Total Environ.*, vol. 654, pp. 1091–1099, Mar. 2019, doi: 10.1016/j.scitotenv.2018.11.086.

[40] 'zhihu'. Accessed: Mar. 17, 2024. [Online]. Available: https://daily.zhihu.com/story/2651979

[41] P. Thangavel, D. Park, and Y.-C. Lee, 'Recent Insights into Particulate Matter (PM2.5)-Mediated Toxicity in Humans: An Overview', *Int. J. Environ. Res. Public. Health*, vol. 19, NO. 12, p. 7511, June 2022, doi: 10.3390/ijerph19127511.

[42] 'Nitrogen dioxide', Ministry for the Environment. Accessed: Mar. 20, 2024. [Online]. Available: https://environment.govt.nz/facts-and-science/air/air-pollutants/nitrogen-dioxide-effects-health/

[43] F. and R. A. (Defra) webmaster@defra gsi gov uk Department for Environment, 'Causes of air pollution- Defra, UK'. Accessed: Mar. 20, 2024. [Online]. Available: https://uk-air.defra.gov.uk/air-pollution/causes

[44] 'Ozone (O3)', GOV.UK. Accessed: Mar. 20, 2024. [Online]. Available: https://www.gov.uk/government/statistics/air-quality-statistics/concentrations-of-ozone

[45] 'DeveloPMent of an aggregate Air Quality Index for an urban Mediterranean agglomeration: Relation to potential health effects - ScienceDirect'. Accessed: Mar. 26, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160412007000153

[46] O. US EPA, 'Air Data Basic Information'. Accessed: Mar. 24, 2024. [Online]. Available: https://www.epa.gov/outdoor-air-quality-data/air-data-basic-information

[47] O. US EPA, 'EPA Defines Air Pollution Danger Levels'. Accessed: Mar. 24, 2024. [Online]. Available: https://www.epa.gov/archive/epa/aboutepa/epa-defines-air-pollution-danger-levels.html

[48] 'AQI Basics | AirNOw.gov'. Accessed: Mar. 24, 2024. [Online]. Available: https://www.airNOw.gov/aqi/aqi-basics

[49] 'Guide for Particle Pollution | AirNOw.gov'. Accessed: Mar. 24, 2024. [Online]. Available: https://www.airNOw.gov/publications/activity-guides/air-quality-activity-guide-for-particle-pollution/

[50] 'Technical Regulation on Ambient Air Quality Index (on trial)'. Accessed: Mar. 24, 2024. [Online]. Available: https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/jcffbz/201203/t20120302_224166.shtml

[51] 'DeveloPMent of an aggregate Air Quality Index for an urban Mediterranean agglomeration: Relation to potential health effects'. Accessed: Mar. 25, 2024.

[Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160412007000153
[52] ‘Air pollution characteristics and human health risks in key cities of NOrthwest China’. Accessed: Mar. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301479720307222#bib73
[53] ‘Characterizing multi-pollutant air pollution in China: Comparison of three air quality indices’. Accessed: Mar. 26, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160412015300052#bb0135
[54] Y. Ma *et al.*, ‘Air pollution and its associated health risks before and after COVID-19 in Shaanxi Province, China’, *Environ. Pollut.*, vol. 320, p. 121090, Mar. 2023, doi: 10.1016/j.envpol.2023.121090.
[55] M. Mao, H. Sun, and X. Zhang, ‘Air Pollution Characteristics and Health Risks in the Yangtze River EcoNOmic Belt, China during Winter’, *Int. J. Environ. Res. Public. Health*, vol. 17, NO. 24, Art. NO. 24, Jan. 2020, doi: 10.3390/ijerph17249172.
[56] F. Shen, X. Ge, J. Hu, D. Nie, L. Tian, and M. Chen, ‘Air pollution characteristics and health risks in Henan Province, China’, *Environ. Res.*, vol. 156, pp. 625–634, July 2017, doi: 10.1016/j.envres.2017.04.026.
[57] H. Janjani, M. S. Hassanvand, H. Kashani, and M. Yunesian, ‘Characterizing Multiple Air Pollutant Indices Based on Their Effects on the Mortality in Tehran, Iran during 2012–2017’, *Sustain. Cities Soc.*, vol. 59, p. 102222, Aug. 2020, doi: 10.1016/j.scs.2020.102222.
[58] ‘A systematic survey of air quality prediction based on deep learning’. Accessed: Mar. 28, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110016824002485
[59] ‘Health impacts’. Accessed: Apr. 21, 2024. [Online]. Available: https://www.who.int/teams/environment-climate-change-and-health/air-quality-energy-and-health/health-impacts
[60] Y. Yu, H. Li, S. Sun, and Y. Li, ‘PM2.5 concentration forecasting through a NOvel multi-scale ensemble learning approach considering intercity synergy’, *Sustain. Cities Soc.*, vol. 85, p. 104049, Oct. 2022, doi: 10.1016/j.scs.2022.104049.
[61] Q. Guo *et al.*, ‘Applications of artificial intelligence in the field of air pollution: A bibliometric analysis’, *Front. Public Health*, vol. 10, Sept. 2022, doi: 10.3389/fpubh.2022.933665.
[62] ‘Frontiers | Applications of artificial intelligence in the field of air pollution: A bibliometric analysis’. Accessed: Apr. 21, 2024. [Online]. Available: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.933665/full
[63] ‘Air Pollution Forecasts: An Overview - PMC’. Accessed: Apr. 21, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/PMc/articles/PMC5923822/
[64] K. Liao, X. Huang, H. Dang, Y. Ren, S. Zuo, and C. Duan, ‘Statistical

Approaches for Forecasting Primary Air Pollutants: A Review', *Atmosphere*, vol. 12, NO. 6, Art. NO. 6, June 2021, doi: 10.3390/atmos12060686.

[65] 'ARIMA forecasting of ambient air pollutants (O3, NO, NO2 and CO) | Stochastic Environmental Research and Risk Assessment'. Accessed: Apr. 22, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00477-009-0361-8

[66] 'Air Pollution PM2.5 Data Analysis in Los Angeles Long Beach with Seasonal ARIMA Model | IEEE Conference Publication | IEEE Xplore'. Accessed: Apr. 22, 2024. [Online]. Available: https://ieeexplore-ieee-org.ezphost.dur.ac.uk/abstract/document/5367074

[67] E. Aladağ, 'Forecasting of particulate matter with a hybrid ARIMA model based on wavelet transformation and seasonal adjustment', *Urban Clim.*, vol. 39, p. 100930, Sept. 2021, doi: 10.1016/j.uclim.2021.100930.

[68] 'Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition - ScienceDirect'. Accessed: Apr. 23, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844022035277

[69] 'Predictions and mitigation strategies of PM2.5 concentration in the Yangtze River Delta of China based on a NOvel NOnlinear seasonal grey model - ScienceDirect'. Accessed: Apr. 23, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749121001925#bib45

[70] 'Quarterly PM2.5 prediction using a NOvel seasonal grey model and its further application in health effects and ecoNOmic loss assessment: evidences from Shanghai and Tianjin, China | Natural Hazards'. Accessed: Apr. 23, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11069-021-04614-y

[71] Domańska , 'Application of fuzzy time series models for forecasting pollution concentrations', *Expert Syst. Appl.*, vol. 39, NO. 9, pp. 7673–7679, July 2012, doi: 10.1016/j.eswa.2012.01.023.

[72] A. Y. Watson, R. R. Bates, and D. Kennedy, 'Mathematical Modeling of the Effect of Emission Sources on Atmospheric Pollutant Concentrations', in *Air Pollution, the Automobile, and Public Health*, National Academies Press (US), 1988. Accessed: Apr. 23, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK218138/

[73] O. US EPA, 'CMAQ: The Community Multiscale Air Quality Modeling System'. Accessed: Apr. 23, 2024. [Online]. Available: https://www.epa.gov/cmaq

[74] 'CMAS: Community Modeling and Analysis System'. Accessed: Apr. 23, 2024. [Online]. Available: https://www.cmascenter.org/cmaq/

[75] 'Modeling intercontinental air pollution transport over the trans-Pacific region in 2001 using the Community Multiscale Air Quality modeling system - Wang - 2009 - Journal of Geophysical Research: Atmospheres - Wiley Online Library'. Accessed: Apr. 23, 2024. [Online]. Available:

https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2008JD010807
[76] ‘Understanding of regional air pollution over China using CMAQ, part I performance evaluation and seasonal variation - ScienceDirect’. Accessed: Apr. 23, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S135223101000261X#bib22
[77] J. Hu, J. Chen, Q. Ying, and H. Zhang, ‘One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system’, *Atmospheric Chem. Phys.*, vol. 16, NO. 16, pp. 10333–10350, Aug. 2016, doi: 10.5194/acp-16-10333-2016.
[78] R. Mathur *et al.*, ‘Extending the Community Multiscale Air Quality (CMAQ) modeling system to hemispheric scales: overview of process considerations and initial applications’, *Atmospheric Chem. Phys.*, vol. 17, NO. 20, pp. 12449–12474, Oct. 2017, doi: 10.5194/acp-17-12449-2017.
[79] ‘Investigating the regional difference of aerosol feedback effects over South Korea using the WRF-CMAQ two-way coupled modeling system - ScienceDirect’. Accessed: Apr. 23, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231019306077
[80] ‘Comparison of Conventional Statistical Methods with Machine Learning in Medicine: DiagNOsis, Drug DeveloPMent, and Treatment - PMC’. Accessed: Apr. 23, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/PMc/articles/PMC7560135/
[81] ‘Air pollution prediction by using an artificial neural network model | Clean TechNOlogies and Environmental Policy’. Accessed: Apr. 23, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10098-019-01709-w
[82] ‘Statistical modelling and prediction of atmospheric pollution by particulate material: two NOnparametric approaches - Silva - 2001 - Environmetrics - Wiley Online Library’. Accessed: Apr. 23, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/1099-095X(200103)12:2%3C147::AID-ENV451%3E3.0.CO;2-3
[83] J. Tobin, ‘Artificial intelligence: DeveloPMent, risks and regulation’, July 2023, Accessed: Apr. 24, 2024. [Online]. Available: https://lordslibrary.parliament.uk/artificial-intelligence-develoPMent-risks-and-regulation/
[84] ‘What Is Artificial Intelligence (AI)? | Built In’. Accessed: Apr. 24, 2024. [Online]. Available: https://builtin.com/artificial-intelligence
[85] ‘What is Artificial Intelligence (AI)? | IBM’. Accessed: Apr. 24, 2024. [Online]. Available: https://www.ibm.com/topics/artificial-intelligence
[86] ‘Air quality forecasting with artificial intelligence techniques: A scientometric and content analysis - ScienceDirect’. Accessed: Apr. 24, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364815222000354
[87] I. H. Sarker, ‘Machine Learning: Algorithms, Real-World Applications and

Research Directions', *SN Comput. Sci.*, vol. 2, NO. 3, p. 160, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[88] 'Early History of Support Vector Machines | SpringerLink'. Accessed: Apr. 25, 2024. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-41136-6_3

[89] N. N. Maltare and S. Vahora, 'Air Quality Index prediction using machine learning for Ahmedabad city', *Digit. Chem. Eng.*, vol. 7, p. 100093, June 2023, doi: 10.1016/j.dche.2023.100093.

[90] 'Meteorological pattern analysis assisted daily PM2.5 grades prediction using SVM optimized by PSO algorithm - ScienceDirect'. Accessed: Apr. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1309104218306469#bib26

[91] A.-L. Balogun and A. Tella, 'Modelling and investigating the impacts of climatic variables on ozone concentration in Malaysia using correlation analysis with random forest, decision tree regression, linear regression, and support vector regression', *Chemosphere*, vol. 299, p. 134250, July 2022, doi: 10.1016/j.chemosphere.2022.134250.

[92] 'Machine Learning vs Neural Networks: What is the Difference?', upGrad blog. Accessed: Apr. 25, 2024. [Online]. Available: https://www.upgrad.com/blog/machine-learning-vs-neural-networks/

[93] 'AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference?' Accessed: Apr. 25, 2024. [Online]. Available: https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/

[94] 'Why Deep Learning over Traditional Machine Learning? | by Sambit Mahapatra | Towards Data Science'. Accessed: Apr. 25, 2024. [Online]. Available: https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063

[95] 'Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in NOrthern California - ScienceDirect'. Accessed: Apr. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969712013617

[96] P. Jiang, Q. Dong, and P. Li, 'A NOvel hybrid strategy for PM2.5 concentration analysis and prediction', *J. Environ. Manage.*, vol. 196, pp. 443–457, 2017, doi: 10.1016/j.jenvman.2017.03.046.

[97] 'A review of artificial neural network models for ambient air pollution prediction - ScienceDirect'. Accessed: Apr. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364815218306352

[98] F. Biancofiore *et al.*, 'Recursive neural network model for analysis and forecast of PM10 and PM2.5', *Atmospheric Pollut. Res.*, vol. 8, NO. 4, pp. 652–659, July 2017, doi: 10.1016/j.apr.2016.12.014.

[99] S. Agarwal *et al.*, 'Air quality forecasting using artificial neural networks with

real time dynamic error correction in highly polluted regions', *Sci. Total Environ.*, vol. 735, p. 139454, Sept. 2020, doi: 10.1016/j.scitotenv.2020.139454.

[100] 'Spatial estimation of urban air pollution with the use of artificial neural network models - ScienceDirect'. Accessed: May 02, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231018305119

[101] P. Perez and J. Reyes, 'An integrated neural network model for PM10 forecasting', *Atmos. Environ.*, vol. 40, NO. 16, pp. 2845–2851, May 2006, doi: 10.1016/j.atmosenv.2006.01.010.

[102] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', *Neural Comput.*, vol. 9, NO. 8, pp. 1735–1780, NOv. 1997, doi: 10.1162/neco.1997.9.8.1735.

[103] 'Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts - ScienceDirect'. Accessed: May 03, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652618332694

[104] 'A NOvel optimal-hybrid model for daily air quality index prediction considering air pollutant factors - ScienceDirect'. Accessed: May 03, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969719323290#bb0215

[105] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin, 'An LSTM-based aggregated model for air pollution forecasting', *Atmospheric Pollut. Res.*, vol. 11, NO. 8, pp. 1451–1463, Aug. 2020, doi: 10.1016/j.apr.2020.05.015.

[106] 'Air Pollutant Concentration Prediction Based on GRU Method - IOPscience'. Accessed: May 03, 2024. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1168/3/032058/meta

[107] 'NO2 pollutant concentration forecasting for air quality monitoring by using an optimised deep learning bidirectional GRU model | International Journal of Computational Science and Engineering'. Accessed: May 03, 2024. [Online]. Available: https://www.inderscienceonline.com/doi/abs/10.1504/IJCSE.2021.113652

[108] D. kharkar, 'Unravelling the Power of XGBoost: Boosting Performance with Extreme Gradient Boosting', Medium. Accessed: May 03, 2024. [Online]. Available: https://medium.com/@dishantkharkar9/unravelling-the-power-of-xgboost-boosting-performance-with-extreme-gradient-boosting-302e1c00e555

[109] B. Pan, 'Application of XGBoost algorithm in hourly PM2.5 concentration prediction', *IOP Conf. Ser. Earth Environ. Sci.*, vol. 113, NO. 1, p. 012127, Feb. 2018, doi: 10.1088/1755-1315/113/1/012127.

[110] 'Application of XGBoost algorithm in the optimization of pollutant concentration - ScienceDirect'. Accessed: May 03, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169809522002241

[111] J. Zhang and S. Li, 'Air quality index forecast in Beijing based on CNN-LSTM multi-model', *Chemosphere*, vol. 308, p. 136180, Dec. 2022, doi:

10.1016/j.chemosphere.2022.136180.
[112] ‘Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions - ScienceDirect’. Accessed: May 03, 2024. [Online]. Available: https://www-sciencedirect-com.ezphost.dur.ac.uk/science/article/pii/S1309104215300301
[113] ‘DeveloPMent of artificial intelligence based NO2 forecasting models at Taj Mahal, Agra - ScienceDirect’. Accessed: May 04, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1309104215302567
[114] P. Pakrooh and E. Pishbahar, ‘Forecasting Air Pollution Concentrations in Iran, Using a Hybrid Model’, *Pollution*, vol. 5, NO. 4, pp. 739–747, Oct. 2019, doi: 10.22059/poll.2019.274827.572.
[115] A. Russo and A. O. Soares, ‘Hybrid Model for Urban Air Pollution Forecasting: A Stochastic Spatio-Temporal Approach’, *Math. Geosci.*, vol. 46, NO. 1, pp. 75–93, Jan. 2014, doi: 10.1007/s11004-013-9483-0.
[116] Y. Xu, H. Liu, and Z. Duan, ‘A NOvel hybrid model for multi-step daily AQI forecasting driven by air pollution big data’, *Air Qual. Atmosphere Health*, vol. 13, NO. 2, pp. 197–207, Feb. 2020, doi: 10.1007/s11869-020-00795-w.
[117] Y. Gu, B. Li, and Q. Meng, ‘Hybrid interpretable predictive machine learning model for air pollution prediction’, *Neurocomputing*, vol. 468, pp. 123–136, Jan. 2022, doi: 10.1016/j.neucom.2021.09.051.
[118] S. Zhu, X. Lian, H. Liu, J. Hu, Y. Wang, and J. Che, ‘Daily air quality index forecasting with hybrid models: A case in China’, *Environ. Pollut.*, vol. 231, pp. 1232–1244, Dec. 2017, doi: 10.1016/j.envpol.2017.08.069.
[119] ‘A NOvel hybrid model for hourly PM2.5 prediction considering air pollution factors, meteorological parameters and GNSS-ZTD - ScienceDirect’. Accessed: May 05, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364815223001664
[120] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, ‘Gradient-based learning applied to document recognition’, *Proc. IEEE*, vol. 86, NO. 11, pp. 2278–2324, NOv. 1998, doi: 10.1109/5.726791.
[121] ‘Finite State Automata and Simple Recurrent Networks | Neural Computation | MIT Press’. Accessed: May 06, 2024. [Online]. Available: https://direct.mit.edu/neco/article-abstract/1/3/372/5483/Finite-State-Automata-and-Simple-Recurrent
[122] T. Li, M. Hua, and X. Wu, ‘A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5)’, *IEEE Access*, vol. 8, pp. 26933–26940, 2020, doi: 10.1109/ACCESS.2020.2971348.
[123] ‘A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction | Air Quality, Atmosphere & Health’. Accessed: May 06, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11869-018-0585-1
[124] ‘Sensors | Free Full-Text | A Deep CNN-LSTM Model for Particulate Matter

(PM2.5) Forecasting in Smart Cities'. Accessed: May 06, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/18/7/2220

[125] Y. Chen, X. Chen, A. Xu, Q. Sun, and X. Peng, 'A hybrid CNN-Transformer model for ozone concentration prediction', *Air Qual. Atmosphere Health*, vol. 15, NO. 9, pp. 1533–1546, Sept. 2022, doi: 10.1007/s11869-022-01197-w.

[126] 'Algorithms | Free Full-Text | PM2.5 Concentration Prediction Based on CNN-BiLSTM and Attention Mechanism'. Accessed: May 06, 2024. [Online]. Available: https://www.mdpi.com/1999-4893/14/7/208

[127] Y.-S. Chang, S. Abimannan, H.-T. Chiao, C.-Y. Lin, and Y.-P. Huang, 'An ensemble learning based hybrid model and framework for air pollution forecasting', *Environ. Sci. Pollut. Res.*, vol. 27, NO. 30, pp. 38155–38168, Oct. 2020, doi: 10.1007/s11356-020-09855-1.

[128] 'Air pollution diffusion simulation and seasonal spatial risk analysis for industrial areas - ScienceDirect'. Accessed: May 06, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013935120315929

[129] J. Wang and G. Song, 'A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction', *Neurocomputing*, vol. 314, pp. 198–206, NOv. 2018, doi: 10.1016/j.neucom.2018.06.049.

[130] 'Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction - ScienceDirect'. Accessed: May 06, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045653518324639

[131] L. Saha, A. Kumar, S. Kumar, J. Korstad, S. Srivastava, and K. Bauddh, 'The impact of the COVID-19 lockdown on global air quality: A review', *Environ. Sustain.*, vol. 5, NO. 1, pp. 5–23, 2022, doi: 10.1007/s42398-021-00213-6.

[132] 'COVID-19 outbreak: Migration, effects on society, global environment and prevention'. Accessed: Apr. 02, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969720323998

[133] 'International Tourist Numbers Could Fall 60-80% in 2020, UNWTO Reports | UN Tourism'. Accessed: Apr. 03, 2024. [Online]. Available: https://www.unwto.org/news/covid-19-international-tourist-numbers-could-fall-60-80-in-2020

[134] M. Drewes, F. Daumann, and F. Follert, 'Exploring the sports ecoNOmic impact of COVID-19 on professional soccer', in *COVID-19 and the Soccer World*, Routledge, 2022.

[135] H. Rashid, E. Haworth, S. Shafi, Z. A. Memish, and R. Booy, 'Pandemic influenza: mass gatherings and mass infection', *Lancet Infect. Dis.*, vol. 8, NO. 9, pp. 526–527, Sept. 2008, doi: 10.1016/S1473-3099(08)70186-5.

[136] Humans I. W. G. on the E. of C. R., 'Sources of air pollutants', in *Outdoor air pollution*, International Agency for Research on Cancer, 2016. Accessed: Apr. 05, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK368029/

[137] V. Zander S, 'COVID-19 lockdowns cause global air pollution declines'.

Accessed: Apr. 05, 2024. [Online]. Available: https://www.pnas.org/doi/full/10.1073/pnas.2006853117
[138] 'Airborne Nitrogen Dioxide Plummets Over China'. Accessed: Apr. 05, 2024. [Online]. Available: https://www.earthobservatory.nasa.gov/images/146362/airborne-nitrogen-dioxide-plummets-over-china
[139] M. J. Bechle, D. B. Millet, and J. D. Marshall, 'Remote sensing of exposure to NO2: Satellite versus ground-based measurement in a large urban area', *Atmos. Environ.*, vol. 69, pp. 345–353, Apr. 2013, doi: 10.1016/j.atmosenv.2012.11.046.
[140] 'Changes in U.S. air pollution during the COVID-19 pandemic - ScienceDirect'. Accessed: Apr. 05, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969720333842#bb0005
[141] Á. Briz-Redón, C. Belenguer-Sapiña, and Á. SerraNO-Aroca, 'Changes in air pollution during COVID-19 lockdown in Spain: A multi-city study', *J. Environ. Sci.*, vol. 101, pp. 16–26, Mar. 2021, doi: 10.1016/j.jes.2020.07.029.
[142] 'Characterization of summertime single aerosol particles in Chengdu (China): Interannual evolution and impact of COVID-19 lockdown - ScienceDirect'. Accessed: Apr. 08, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969723063921
[143] 'Significant concurrent decrease in PM2.5 and NO2 concentrations in China during COVID-19 epidemic - ScienceDirect'. Accessed: Apr. 09, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1001074220302886
[144] 'Assessing the change of ambient air quality patterns in Jiangsu Province of China pre-to post-COVID-19 - ScienceDirect'. Accessed: Apr. 09, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045653521030411
[145] P. Jha, J. S. Rajput, and A. K. Agarwal, 'Assessment of Air Status Pre-Lockdown, Lockdown and Post-Lockdown in Dewas, M.P. (India)', *Assess. Air Status Pre-Lockdown Lockdown Post-Lockdown Dewas MP India*, vol. 8, NO. 12, p. 7, Jan. 2024, doi: 10.5281/zeNOdo.10450792.
[146] 'Air quality changes during the COVID-19 pandemic guided by robust virus-spreading data in Italy | Air Quality, Atmosphere & Health'. Accessed: Apr. 12, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11869-023-01495-x
[147] S. Wang *et al.*, 'Modeling assessment of air pollution control measures and COVID-19 pandemic on air quality improvements over Greater Bay Area of China', *Sci. Total Environ.*, vol. 926, p. 171951, May 2024, doi: 10.1016/j.scitotenv.2024.171951.
[148] 'Drivers for the poor air quality conditions in NOrth China Plain during the COVID-19 outbreak - ScienceDirect'. Accessed: Apr. 13, 2024. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S1352231020308359
[149] P. Rafaj *et al.*, 'Outlook for clean air in the context of sustainable develoPMent goals', *Glob. Environ. Change*, vol. 53, pp. 1–11, NOv. 2018, doi: 10.1016/j.gloenvcha.2018.08.008.
[150] M. Ghanbari Ghozikali *et al.*, 'Evaluation of Chronic Obstructive Pulmonary Disease (COPD) attributed to atmospheric O3, NO2, and SO2 using Air Q Model (2011–2012 year)', *Environ. Res.*, vol. 144, pp. 99–105, Jan. 2016, doi: 10.1016/j.envres.2015.10.030.
[151] E. Fattore *et al.*, 'Human health risk in relation to air quality in two municipalities in an industrialized area of NOrthern Italy', *Environ. Res.*, vol. 111, NO. 8, pp. 1321–1327, NOv. 2011, doi: 10.1016/j.envres.2011.06.012.
[152] 'Ambient air pollution and daily hospital admissions for cardiovascular diseases in Arak, Iran - PMC'. Accessed: Apr. 14, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/PMc/articles/PMC5677328/
[153] 'A review on analysis methods, source identification, and cancer risk evaluation of atmospheric polycyclic aromatic hydrocarbons - ScienceDirect'. Accessed: Apr. 14, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969721028126
[154] S. P. Nana-Sinkam and C. A. Powell, 'Molecular Biology of Lung Cancer: DiagNOsis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines', *Chest*, vol. 143, NO. 5, Supplement, pp. e30S-e39S, May 2013, doi: 10.1378/chest.12-2346.
[155] 'Environmental determinants of cardiovascular disease: lessons learned from air pollution | Nature Reviews Cardiology'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.nature.com/articles/s41569-020-0371-2
[156] 'Environmental factors in cardiovascular disease | Nature Reviews Cardiology'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.nature.com/articles/nrcardio.2015.152
[157] 'Guidance to Reduce the Cardiovascular Burden of Ambient Air Pollutants: A Policy Statement From the American Heart Association | Circulation'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.ahajournals.org/doi/full/10.1161/CIR.0000000000000930
[158] R. Liang, B. Zhang, X. Zhao, Y. Ruan, H. Lian, and Z. Fan, 'Effect of exposure to PM2.5 on blood pressure: a systematic review and meta-analysis', *J. Hypertens.*, vol. 32, NO. 11, p. 2130, NOv. 2014, doi: 10.1097/HJH.0000000000000342.
[159] 'Geographic variations in the blood pressure responses to short-term fine particulate matter exposure in China - ScienceDirect'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969720313541
[160] 'Effects of Ambient Air Pollution on Blood Pressure Among Children and Adolescents: A Systematic Review and Meta-Analysis | Journal of the American

Heart Association'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.ahajournals.org/doi/full/10.1161/JAHA.120.017734
[161] 'Relations between air pollution and vascular develoPMent in 5-year old children: a cross-sectional study in the Netherlands | Environmental Health'. Accessed: Apr. 17, 2024. [Online]. Available: https://link.springer.com/article/10.1186/s12940-019-0487-1
[162] 'Association of long-term exposure to ambient air pollutants with blood lipids in Chinese adults: The China Multi-Ethnic Cohort study - ScienceDirect'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013935121004680
[163] 'Particulate Matter Air Pollution and Atherosclerosis | Current Atherosclerosis Reports'. Accessed: Apr. 17, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11883-010-0122-7
[164] 'Long-Term PM2.5 Exposure and Risks of Ischemic Heart Disease and Stroke Events: Review and Meta-Analysis | Journal of the American Heart Association'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.ahajournals.org/doi/full/10.1161/JAHA.120.016890
[165] 'Environmental Tobacco Smoke Exposure and Risk of Stroke in Never Smokers: An Updated Review with Meta-Analysis - ScienceDirect'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1052305716303408
[166] I. Bilić, G. Džamonja, I. Lušić, M. Matijaca, and K. Čaljkušić, 'Risk Factors and Outcome Differences between Ischemic and Hemorrhagic Stroke', *Acta Clin. Croat.*, vol. 48, NO. 4, pp. 399–403, Dec. 2009.
[167] A. Bhatnagar, 'Cardiovascular Effects of Particulate Air Pollution', *Annu. Rev. Med.*, vol. 73, NO. Volume 73, 2022, pp. 393–406, Jan. 2022, doi: 10.1146/annurev-med-042220-011549.
[168] Y. Liu *et al.*, 'Short-Term Exposure to Ambient Air Pollution and Asthma Mortality', *Am. J. Respir. Crit. Care Med.*, vol. 200, NO. 1, pp. 24–32, July 2019, doi: 10.1164/rccm.201810-1823OC.
[169] '7 million premature deaths annually linked to air pollution'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.who.int/news/item/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution
[170] 'Modifiers of the effect of short-term variation in PM2.5 on mortality in Beijing, China - ScienceDirect'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013935119308631
[171] 'Short-term associations between size-fractionated particulate air pollution and COPD mortality in Shanghai, China - ScienceDirect'. Accessed: Apr. 17, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749119328209
[172] 'Long-Term Exposure to Ozone and Cause-Specific Mortality Risk in the United States | American Journal of Respiratory and Critical Care Medicine'. Accessed:

Apr. 17, 2024. [Online]. Available: https://www.atsjournals.org/doi/full/10.1164/rccm.201806-1161OC

[173] 'Air pollution and COPD: GOLD 2023 committee report | European Respiratory Society'. Accessed: Apr. 17, 2024. [Online]. Available: https://erj.ersjournals.com/content/61/5/2202469.abstract

[174] 'WHO tool for behavioural insights on COVID-19'. Accessed: Apr. 18, 2024. [Online]. Available: https://www.who.int/europe/tools-and-toolkits/airq---software-tool-for-health-risk-assessment-of-air-pollution

[175] W. Wu, M. Zhang, and Y. Ding, 'Exploring the effect of ecoNOmic and environment factors on PM2.5 concentration: A case study of the Beijing-Tianjin-Hebei region', *J. Environ. Manage.*, vol. 268, p. 110703, Aug. 2020, doi: 10.1016/j.jenvman.2020.110703.

[176] 'Health and ecoNOmic benefits of clean air policies in China: A case study for Beijing-Tianjin-Hebei region - ScienceDirect'. Accessed: Apr. 13, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749121011076?via%3Dihub

[177] X. Yang *et al.*, 'Two-Dimensional Silicon Fingerprints Reveal Dramatic Variations in the Sources of Particulate Matter in Beijing during 2013–2017', *Environ. Sci. TechNOl.*, vol. 54, NO. 12, pp. 7126–7135, June 2020, doi: 10.1021/acs.est.0c00984.

[178] 'Spatiotemporal variations of PM2.5 pollution and its dynamic relationships with meteorological conditions in Beijing-Tianjin-Hebei region - ScienceDirect'. Accessed: Apr. 13, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004565352201133X#bib78

[179] J. Xu, Y. Dong, L. Xie, and S. Chen, 'The pollution haven strikes back?–Evidence from air quality daily variation in the Jing-Jin-Ji region of China', *Environ. Sci. Policy*, vol. 138, pp. 105–121, Dec. 2022, doi: 10.1016/j.envsci.2022.09.014.

[180] Vanessa Zheng, 'Location of Beijing, Beijing' Location map'. Accessed: May 20, 2023. [Online]. Available: https://www.mybeijingchina.com/travel-guide/general-info/location.htm

[181] X. Zhang, M. Lin, Z. Wang, and F. Jin, 'The impact of energy-intensive industries on air quality in China's industrial agglomerations', *J. Geogr. Sci.*, vol. 31, NO. 4, pp. 584–602, Apr. 2021, doi: 10.1007/s11442-021-1860-x.

[182] Y. Tang and X. Meng, 'From Concentration to Decentralization: The Spatial DeveloPMent of Beijing and the Beijing–Tianjin–Hebei Capital Region', in *Chinese Urban Planning and Construction: From Historical Wisdom to Modern Miracles*, L. Bian, Y. Tang, and Z. Shen, Eds, in Strategies for Sustainability. , Cham: Springer International Publishing, 2021, pp. 89–112. doi: 10.1007/978-3-030-65562-4_5.

[183] Z. Dong, S. Wang, J. Xing, X. Chang, D. Ding, and H. Zheng, 'Regional transport

in Beijing-Tianjin-Hebei region and its changes during 2014–2017: The impacts of meteorology and emission reduction', *Sci. Total Environ.*, vol. 737, p. 139792, Oct. 2020, doi: 10.1016/j.scitotenv.2020.139792.

[184] Y. Wang, H. Liu, G. Mao, J. Zuo, and J. Ma, 'Inter-regional and sectoral linkage analysis of air pollution in Beijing–Tianjin–Hebei (Jing-Jin-Ji) urban agglomeration of China', *J. Clean. Prod.*, vol. 165, pp. 1436–1444, NOv. 2017, doi: 10.1016/j.jclepro.2017.07.210.

[185] Y. Song, Z. Li, T. Yang, and Q. Xia, 'Does the expansion of the joint prevention and control area improve the air quality?—Evidence from China's Jing-Jin-Ji region and surrounding areas', *Sci. Total Environ.*, vol. 706, p. 136034, Mar. 2020, doi: 10.1016/j.scitotenv.2019.136034.

[186] 'Jing-Jin-Ji: China's new mega-region'. Accessed: June 27, 2024. [Online]. Available: https://www.worldfinance.com/infrastructure-investment/jing-jin-ji-chinas-new-mega-region

[187] 'Detecting the causality influence of individual meteorological factors on local PM2.5 concentration in the Jing-Jin-Ji region | Scientific Reports'. Accessed: June 27, 2024. [Online]. Available: https://www.nature.com/articles/srep40735

[188] UCI, 'UCI Machine Learning Repository: Beijing Multi-Site Air-Quality Data Data Set'. Accessed: May 20, 2023. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data

[189] M. Moazeni, M. R. Maracy, B. Dehdashti, and A. Ebrahimi, 'Spatiotemporal analysis of COVID-19, air pollution, climate, and meteorological conditions in a metropolitan region of Iran', *Environ. Sci. Pollut. Res.*, vol. 29, NO. 17, pp. 24911–24924, Apr. 2022, doi: 10.1007/s11356-021-17535-x.

[190] T. W. A. Q. I. project, 'Air Quality Historical Data Platform: Institution & University Registration', aqicn.org. Accessed: June 27, 2024. [Online]. Available: https://aqicn.org/data-platform/register/

[191] 'Air Pollutants in Metropolises of Eastern Coastal China - PMC'. Accessed: May 27, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/PMc/articles/PMC9691249/

[192] Y. Yu, X. Si, C. Hu, and J. Zhang, 'A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures', *Neural Comput.*, vol. 31, NO. 7, pp. 1235–1270, July 2019, doi: 10.1162/neco_a_01199.

[193] O. Calzone, 'An Intuitive Explanation of LSTM', Medium. Accessed: May 29, 2024. [Online]. Available: https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c

[194] B. Das, Ö. O. Dursun, and S. Toraman, 'Prediction of air pollutants for air quality using deep learning methods in a metropolitan city', *Urban Clim.*, vol. 46, p. 101291, Dec. 2022, doi: 10.1016/j.uclim.2022.101291.

[195] 'MLP and MLR models for instantaneous thermal efficiency prediction of solar still under hyper-arid environment - ScienceDirect'. Accessed: June 01, 2024. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S0168169916300035
[196] W. Turner, 'Build a model using — Multilayer Perceptron', Medium. Accessed: June 01, 2024. [Online]. Available: https://medium.com/@willturnerau/build-a-model-using-multilayer-perceptron-791a62502387
[197] S. Cristina, 'The Transformer Model', MachineLearningMastery.com. Accessed: June 01, 2024. [Online]. Available: https://machinelearningmastery.com/the-transformer-model/
[198] K. Han *et al.*, 'A Survey on Vision Transformer', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, NO. 1, pp. 87–110, Jan. 2023, doi: 10.1109/TPAMI.2022.3152247.
[199] N. Engel, V. Belagiannis, and K. Dietmayer, 'Point Transformer', *IEEE Access*, vol. 9, pp. 134826–134840, 2021, doi: 10.1109/ACCESS.2021.3116304.
[200] T. N. Kipf and M. Welling, 'Semi-Supervised Classification with Graph Convolutional Networks', Feb. 22, 2017, *arXiv*: arXiv:1609.02907. doi: 10.48550/arXiv.1609.02907.
[201] Thomas Kipf, 'Graph Convolutional Networks'. Accessed: May 29, 2024. [Online]. Available: https://tkipf.github.io/graph-convolutional-networks/
[202] K. Cho *et al.*, 'Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation', 2014, *arXiv*. doi: 10.48550/ARXIV.1406.1078.
[203] Anishnama, 'Understanding Gated Recurrent Unit (GRU) in Deep Learning', Medium. Accessed: June 23, 2024. [Online]. Available: https://medium.com/@anishnama20/understanding-gated-recurrent-unit-gru-in-deep-learning-2e54923f3e2
[204] D. Bahdanau, K. Cho, and Y. Bengio, 'Neural Machine Translation by Jointly Learning to Align and Translate', 2014, *arXiv*. doi: 10.48550/ARXIV.1409.0473.
[205] M.-T. Luong, H. Pham, and C. D. Manning, 'Effective Approaches to Attention-based Neural Machine Translation', 2015, *arXiv*. doi: 10.48550/ARXIV.1508.04025.
[206] Doreswamy, H. K s, Y. Km, and I. Gad, 'Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models', *Procedia Comput. Sci.*, vol. 171, pp. 2057–2066, Jan. 2020, doi: 10.1016/j.procs.2020.04.221.
[207] W. Mao, W. Wang, L. Jiao, S. Zhao, and A. Liu, 'Modeling air quality prediction using a deep learning approach: Method optimization and evaluation', *Sustain. Cities Soc.*, vol. 65, p. 102567, Feb. 2021, doi: 10.1016/j.scs.2020.102567.
[208] Y.-C. Liang, Y. Maimury, A. H.-L. Chen, and J. R. C. Juarez, 'Machine Learning-Based Prediction of Air Quality', *Appl. Sci.*, vol. 10, NO. 24, Art. NO. 24, Jan. 2020, doi: 10.3390/app10249151.
[209] N. H. Van, P. Van Thanh, D. N. Tran, and D.-T. Tran, 'A new model of air quality prediction using lightweight machine learning', *Int. J. Environ. Sci. TechNOl.*, vol. 20, NO. 3, pp. 2983–2994, Mar. 2023, doi: 10.1007/s13762-022-04185-w.
[210] S. Turney, 'Coefficient of Determination (R^2) | Calculation & Interpretation',

Scribbr. Accessed: Aug. 19, 2024. [Online]. Available: https://www.scribbr.com/statistics/coefficient-of-determination/

[211] Aaron Nicolson and Kuldip K. Paliwal, 'Deep learning for minimum mean-square error approaches to speech enhancement', *Speech Commun.*, vol. 111, pp. 44–55, Aug. 2019, doi: 10.1016/j.specom.2019.06.002.

[212] Davide Chicco, Matthijs J. Warrens, and Giuseppe Jurman, 'The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation', *PeerJ Comput. Sci.*, vol. 7, p. e623, July 2021, doi: 10.7717/peerj-cs.623.

[213] 'Descriptive Statistics: Definition, Overview, Types, and Examples', Investopedia. Accessed: June 29, 2024. [Online]. Available: https://www.investopedia.com/terms/d/descriptive_statistics.asp

[214] 'Air Quality Standards'. Accessed: July 01, 2024. [Online]. Available: https://www.who.int/tools/air-quality-standards

[215] R. Python, 'Python Statistics Fundamentals: How to Describe Your Data – Real Python'. Accessed: June 30, 2024. [Online]. Available: https://realpython.com/python-statistics/

[216] 'AirQ+: software tool for health risk assessment of air pollution'. Accessed: June 30, 2024. [Online]. Available: https://www.who.int/europe/tools-and-toolkits/airq---software-tool-for-health-risk-assessment-of-air-pollution

[217] K. R. D. Aniceto *et al.*, 'Seasonal Mapping and Air Quality Evaluation of Total Suspended Particulate Concentration Using ArcGIS-Based Spatial Analysis in Metro Manila, Philippines', in *2021 IEEE 13th International Conference on HumaNOid, NaNOtechNOlogy, Information TechNOlogy, Communication and Control, Environment, and Management (HNICEM)*, NOv. 2021, pp. 1–6. doi: 10.1109/HNICEM54116.2021.9732031.

[218] P. OrellaNO, J. ReyNOso, N. Quaranta, A. Bardach, and A. Ciapponi, 'Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and *meta*-analysis', *Environ. Int.*, vol. 142, p. 105876, Sept. 2020, doi: 10.1016/j.envint.2020.105876.

[219] M. Zamani Joharestani, C. Cao, X. Ni, B. Bashir, and S. Talebiesfandarani, 'PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data', *Atmosphere*, vol. 10, NO. 7, Art. NO. 7, July 2019, doi: 10.3390/atmos10070373.

[220] 'Full article: A hybrid integrated deep learning model for predicting various air pollutants'. Accessed: Mar. 15, 2025. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/15481603.2021.1988429#abstract

[221] 'The application of strategy based on LSTM for the short-term prediction of PM2.5 in city - ScienceDirect'. Accessed: Mar. 17, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969723065191

[222] ‘A new method for interpolation of missing air quality data at monitor stations - ScienceDirect’. Accessed: Mar. 17, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160412022004652
[223] H. Wen, Y. Dang, and L. Li, ‘Short-Term PM2.5 Concentration Prediction by Combining GNSS and Meteorological Factors’, *IEEE Access*, vol. 8, pp. 115202–115216, 2020, doi: 10.1109/ACCESS.2020.3003580.
[224] ‘A spatial correlation prediction model of urban PM2.5 concentration based on deconvolution and LSTM’, *Neurocomputing*, vol. 544, p. 126280, Aug. 2023, doi: 10.1016/j.neucom.2023.126280.
[225] ‘Forecasting hourly PM2.5 concentration with an optimized LSTM model’, *Atmos. Environ.*, vol. 315, p. 120161, Dec. 2023, doi: 10.1016/j.atmosenv.2023.120161.
[226] X. Y. Ni, H. Huang, and W. P. Du, ‘Relevance analysis and short-term prediction of PM 2.5 concentrations in Beijing based on multi-source data’, *Atmos. Environ.*, vol. 150, pp. 146–161, Feb. 2017, doi: 10.1016/j.atmosenv.2016.11.054.
[227] M.-D. Lin, P.-Y. Liu, C.-W. Huang, and Y.-H. Lin, ‘The application of strategy based on LSTM for the short-term prediction of PM2.5 in city’, *Sci. Total Environ.*, vol. 906, p. 167892, Jan. 2024, doi: 10.1016/j.scitotenv.2023.167892.
[228] Y. Cheng, H. Zhang, Z. Liu, L. Chen, and P. Wang, ‘Hybrid algorithm for short-term forecasting of PM2.5 in China’, *Atmos. Environ.*, vol. 200, pp. 264–279, Mar. 2019, doi: 10.1016/j.atmosenv.2018.12.025.
[229] S. Oji and H. Adamu, ‘Correlation between air pollutants concentration and meteorological factors on seasonal air quality variation’, *J. Air Pollut. Health*, May 2020, doi: 10.18502/japh.v5i1.2856.
[230] ‘Characterising temporal variability of PM2.5/PM10 ratio and its correlation with meteorological variables at a sub-urban site in the Taj City - ScienceDirect’. Accessed: Mar. 15, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212095523003577
[231] O. US EPA, ‘Climate Change Impacts on Air Quality’. Accessed: Mar. 15, 2025. [Online]. Available: https://www.epa.gov/climateimpacts/climate-change-impacts-air-quality
[232] ‘On the interpretation of seasonal variations of stratospheric ozone - ScienceDirect’. Accessed: Mar. 15, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0032063389901438
[233] ‘Seasonal Variability of Air Pollutants and Their Relationships to Meteorological Parameters in an Urban Environment - Shelton - 2022 - Advances in Meteorology - Wiley Online Library’. Accessed: Mar. 15, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2022/5628911
[234] S. Shelton *et al.*, ‘Seasonal Variability of Air Pollutants and Their Relationships to Meteorological Parameters in an Urban Environment’, doi: 10.1155/2022/5628911.
[235] Xiao Q., Ma Z., Li S., and Liu Y., ‘The Impact of Winter Heating on Air Pollution

in China', *PLOS ONE*, vol. 10, NO. 1, p. e0117311, Jan. 2015, doi: 10.1371/journal.pone.0117311.

[236] Paul McAleavey, 'Hot summer weather exacerbating ozone pollution — European Environment Agency'. Accessed: Apr. 27, 2024. [Online]. Available: https://www.eea.europa.eu/highlights/hot-summer-weather-exacerbating-ozone-pollution

[237] T. J. Brown, M. Wilkerson, B. Blanton, and S. Bhattacharya, 'Pragmatic and ML Approaches to Backfilling Missing Data Within Time Series Datasets', in *2024 International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2024, pp. 1865–1871. doi: 10.1109/ICMLA61862.2024.00287.

[238] B. M. Hussein and S. M. Shareef, 'An Empirical Study on the Correlation between Early Stopping Patience and Epochs in Deep Learning', *ITM Web Conf.*, vol. 64, p. 01003, 2024, doi: 10.1051/itmconf/20246401003.

[239] 'Air Pollutant Concentration Prediction Based on GRU Method - IOPscience'. Accessed: Mar. 15, 2025. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1168/3/032058/meta

[240] 'Air pollution forecasting based on attention-based LSTM neural network and ensemble learning - Liu - 2020 - Expert Systems - Wiley Online Library'. Accessed: Mar. 15, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12511

[241] A. Dutta and W. Jinsart, 'Air Pollution in Indian Cities and Comparison of MLR, ANN and CART Models for Predicting PM10 Concentrations in Guwahati, India', *Asian J. Atmospheric Environ.*, vol. 15, NO. 1, p. 2020131, Mar. 2021, doi: 10.5572/ajae.2020.131.

[242] Qi Zhang, 'Deep-AIR: A Hybrid CNN-LSTM Framework for Fine-Grained Air Pollution Estimation and Forecast in Metropolitan Cities | IEEE Journals & Magazine | IEEE Xplore'. Accessed: May 29, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9780279

[243] 'PM2.5 Prediction Based on the CEEMDAN Algorithm and a Machine Learning Hybrid Model'. Accessed: Mar. 15, 2025. [Online]. Available: https://www.mdpi.com/2071-1050/14/23/16128

[244] Y. Gu, B. Li, and Q. Meng, 'Hybrid interpretable predictive machine learning model for air pollution prediction', *Neurocomputing*, vol. 468, pp. 123–136, Jan. 2022, doi: 10.1016/j.neucom.2021.09.051.

[245] 'Analysis of SO2 Pollution Changes of Beijing-Tianjin-Hebei Region over China Based on OMI Observations from 2006 to 2017 - Wang - 2018 - Advances in Meteorology - Wiley Online Library'. Accessed: Mar. 16, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2018/8746068

[246] Z. Wu, M. Huang, A. Zhao, and Z. lan, 'Traffic prediction based on GCN-LSTM model', *J. Phys. Conf. Ser.*, vol. 1972, NO. 1, p. 012107, July 2021, doi: 10.1088/1742-6596/1972/1/012107.

[247] 'Hourly ozone prediction for a 24-h horizon using neural networks', *Environ.*

Model. Softw., vol. 23, NO. 12, pp. 1407–1421, Dec. 2008, doi: 10.1016/j.envsoft.2008.04.004.

[248] A. Yafouz, A. N. Ahmed, N. Zaini, and A. El-Shafie, 'Ozone Concentration Forecasting Based on Artificial Intelligence Techniques: A Systematic Review', *Water. Air. Soil Pollut.*, vol. 232, NO. 2, pp. 1–29, Feb. 2021, doi: 10.1007/s11270-021-04989-5.

[249] Q. Pan, F. Harrou, and Y. Sun, 'A comparison of machine learning methods for ozone pollution prediction', *J. Big Data*, vol. 10, NO. 1, Art. NO. 1, Dec. 2023, doi: 10.1186/s40537-023-00748-x.

[250] 'Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China', *Sci. Total Environ.*, vol. 699, p. 133561, Jan. 2020, doi: 10.1016/j.scitotenv.2019.07.367.

[251] 'Graph Convolutional Network - an overview | ScienceDirect Topics'. Accessed: Mar. 16, 2025. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/graph-convolutional-network

[252] 'Air pollutant prediction based on ARIMA-WOA-LSTM model', *Atmospheric Pollut. Res.*, vol. 14, NO. 6, p. 101761, June 2023, doi: 10.1016/j.apr.2023.101761.

[253] J. Tian, C. Fang, J. Qiu, and J. Wang, 'Analysis of Pollution Characteristics and Influencing Factors of Main Pollutants in the Atmosphere of Shenyang City', *Atmosphere*, vol. 11, NO. 7, Art. NO. 7, July 2020, doi: 10.3390/atmos11070766.

[254] K. Xiao, Y. Wang, G. Wu, B. Fu, and Y. Zhu, 'Spatiotemporal Characteristics of Air Pollutants (PM10, PM2.5, SO2, NO2, O3, and CO) in the Inland Basin City of Chengdu, Southwest China', *Atmosphere*, vol. 9, NO. 2, Art. NO. 2, Feb. 2018, doi: 10.3390/atmos9020074.

[255] B. J. Bloomer, J. W. Stehr, C. A. Piety, R. J. Salawitch, and R. R. Dickerson, 'Observed relationships of ozone air pollution with temperature and emissions', *Geophys. Res. Lett.*, vol. 36, NO. 9, 2009, doi: 10.1029/2009GL037308.

[256] 'A high-resolution air pollutants emission inventory in 2013 for the Beijing-Tianjin-Hebei region, China - ScienceDirect'. Accessed: Mar. 11, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231017306386

[257] K. Li, D. J. Jacob, L. Shen, X. Lu, I. De Smedt, and H. Liao, 'Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences', *Atmospheric Chem. Phys.*, vol. 20, NO. 19, pp. 11423–11433, Oct. 2020, doi: 10.5194/acp-20-11423-2020.

[258] J. (Jim) Zhang, Y. Wei, and Z. Fang, 'Ozone Pollution: A Major Health Hazard Worldwide', *Front. ImmuNOl.*, vol. 10, Oct. 2019, doi: 10.3389/fimmu.2019.02518.

[259] 'Impact of environmental pollution from human activities on water, air quality and climate change - ScienceDirect'. Accessed: Mar. 12, 2025. [Online].

Available: https://www.sciencedirect.com/science/article/pii/S2950509724000571

[260] ‘Frontiers | China’s COVID-19 Control Strategy and Its Impact on the Global Pandemic’. Accessed: Mar. 12, 2025. [Online]. Available: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.857003/full

[261] S. Fallahizadeh, M. Kermani, A. Esrafili, Z. Asadgol, and M. Gholami, ‘The effects of meteorological parameters on PM10: Health impacts assessment using AirQ+ model and prediction by an artificial neural network (ANN)’, *Urban Clim.*, vol. 38, p. 100905, July 2021, doi: 10.1016/j.uclim.2021.100905.

[262] ‘Covid-19: China orders millions in Hebei province around Beijing into lockdown | Hong Kong Free Press HKFP’. Accessed: Dec. 02, 2024. [Online]. Available: https://hongkongfp.com/2022/08/30/covid-19-china-orders-millions-in-hebei-province-around-beijing-into-lockdown/

[263] ‘Biological effect of PM10 on airway epithelium-focus on obstructive lung diseases’, *Clin. ImmuNOl.*, vol. 227, p. 108754, June 2021, doi: 10.1016/j.clim.2021.108754.

[264] ‘Does lockdown reduce air pollution? Evidence from 44 cities in NOrthern China - ScienceDirect’. Accessed: Apr. 08, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969720325699

[265] ‘The impact of PM2.5 on the human respiratory system - PMC’. Accessed: Mar. 25, 2025. [Online]. Available: https://PMc.ncbi.nlm.nih.gov/articles/PMC4740125/

[266] ‘Forecasting Air Quality Indicators for 33 Cities in China - Wu - 2020 - CLEAN – Soil, Air, Water - Wiley Online Library’. Accessed: Mar. 12, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/clen.201900097

[267] ‘Interaction Effects of Air Pollution and Climatic Factors on Circulatory and Respiratory Mortality in Xi’an, China between 2014 and 2016’. Accessed: Mar. 12, 2025. [Online]. Available: https://www.mdpi.com/1660-4601/17/23/9027