Urban Air Pollution Modelling with Machine Learning using Fixed and Mobile Sensors

Jun Song

A thesis submitted for the degree of Doctor of Philosophy of

Imperial College London

Center for Transport Studies

Department of Civil and Environmental Engineering

Imperial College London, United Kingdom

July 2021

Table of Contents

Affirmation

In this thesis, several datasets are supported by the third institution, as is clearly stated where appropriate.

The work submitted in this thesis solely belongs to the author and has not been previously submitted for any other degree or diploma of any University. Some work has been submitted for journal publication and listed in the reference statement, and these works belong to my individual work under the academic suggestions and support from my supervisor.

Jun Song

July 2021

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

AOD	Aerosol Optical Depth
AOI	Area of Interest
AQ	Air Quality
AQI	Air Quality Index
CMAQ	Community Multiscale Air Quality
CNN	Convolutional Neural Network
CO	Carbon Monoxide
CO_2	Carbon Dioxide
CTM	Chemical Transport Model
Deep AQ	Machine Learning based Air Quality Inference
Deep GP	Deep Gaussian Process
DEM	Digital Elevation Model
DT	Decision Tree
DFeaST-Net	Deep Feature Spatial-Temporal Learning Network
DFeaST-Tree	Deep Feature Spatial-Temporal Learning Tree
DM-LSTM	Deep Multi-output LSTM
EPA	Environmental Protection Agency
ESCAPE	European Study of Cohorts for Air Pollution Effects
FCN	Fully Convolutional Network
GBDT	Gradient Boosting Decent Tree

GEOS-CHEM	Global Atmospheric Chemistry Model
GI	Geographical Interpolation
GIS	Geographic Information System
HOCV	Hold-out Cross Validation
IDW	Inverse Distance Weighted
IoT	Internet of Things
KKF	Kriged Kalman Filtering
KNN	K-Nearest Neighbours
LBS	Location-based Services
LBS-R	Location-based Service-Request
LOOCV	Leave-one-out Cross Validation
LSTM	Long Short-Term Memory
LUR	Land Use Regression
MAUP	Modifiable Areal Unit Problem
MISR	Multiangle Imaging Spectroradiometer
MLR	Multivariate Linear Regression
ML	Machine Learning
MLR	Multivariate linear regression
MSU	Micro Sensing Unit
Multi AP	Multi-pollutant Spatial-Temporal Learning Network
NEIC	National Environment Information Centre
NLR	Nonlinear Regression
NO_x	Nitrogen Oxide

OFCLP	Online Field Calibration Learning Platform
POI	Point of Interest
PM	Particulate matter (addresses both PM_{10} and $PM_{2.5}$)
$PM_{2.5}$	Particulate matter which passes through a size-selective inlet with a 50% efficiency cut-off at 2.5um aerodynamic diameter
PM_{10}	Particulate matter which passes through a size-selective inlet with a 50% efficiency cut-off at 10um aerodynamic diameter
O_3	Ozone
RF	Random Forest
SAVIAH	Small Area Variations in Air Quality and Health
SRS	Satellite Remote Sensing
STK	Spatio-temporal Kriging
ST-MASK	Spatio-temporal Labelling Mask
ST-FILTER	Spatio-temporal convolution Filter
SO_2	Sulphur Dioxide
SVM	Support Vector Machine
WHO	World Health Organization
XGBOOST	Scalable, Portable and Distributed Gradient Boosting

Abstract

Detailed air quality (AQ) information is crucial for sustainable urban management, and many regions in the world have built static AQ monitoring networks to provide AQ information. However, they can only monitor the region-level AQ conditions or sparse point-based air pollutant measurements, but cannot capture the urban dynamics with high-resolution spatio-temporal variations over the region. Without pollution details, citizens will not be able to make fully informed decisions when choosing their everyday outdoor routes or activities, and policy-makers can only make macroscopic regulating decisions on controlling pollution triggering factors and emission sources. An increasing research effort has been paid on mobile and ubiquitous sampling campaigns as they are deemed the more economically and operationally feasible methods to collect urban AQ data with high spatio-temporal resolution.

The current research proposes a *Machine Learning based AQ Inference (Deep AQ)* framework from data-driven perspective, consisting of data pre-processing, feature extraction and transformation, and pixelwise (grid-level) AQ inference. The Deep AQ framework is adaptable to integrate AQ measurements from the fixed monitoring sites (temporally dense but spatially sparse), and mobile low-cost sensors (temporally sparse but spatially dense). While instantaneous pollutant concentration varies in the micro-environment, this research samples representative values in each grid-cell-unit and achieves AQ inference at 1 km × 1 km pixelwise scale. This research explores the predictive power of the Deep AQ framework based on samples from only 40 fixed monitoring sites in Chengdu, China (4,900 km^2, 26 April - 12 June 2019) and collaborative sampling from 28 fixed monitoring sites and 15 low-cost sensors equipped with taxis deployed in Beijing, China (3,025 km^2, 19 June - 16 July 2018).

The proposed Deep AQ framework is capable of producing high-resolution (1 km × 1 km, hourly) pixelwise AQ inference based on multi-source AQ samples (fixed or

mobile) and urban features (land use, population, traffic, and meteorological information, etc.). This research has achieved high-resolution (1 km × 1 km, hourly) AQ inference (Chengdu: less than 1% spatio-temporal coverage; Beijing: less than 5% spatio-temporal coverage) with reasonable and satisfactory accuracy by the proposed methods in urban cases (Chengdu: SMAPE < 20%; Beijing: SMAPE < 15%). Detailed outcomes and main conclusions are provided in this thesis on the aspects of fixed and mobile sensing, spatio-temporal coverage and density, and the relative importance of urban features. Outcomes from this research facilitate to provide a scientific and detailed health impact assessment framework for exposure analysis and inform policy-makers with data driven evidence for sustainable urban management.

Acknowledgments

During my research time at Imperial College London, I have had the great pleasure of meeting many brilliant people, and there are far too many to name here. Thank you all.

I would like to express my respect and sincere gratitude to my supervisor **Dr. Marc E.J. Stettler** over the past time. **Dr. Stettler**, thank you for your systematic academic guidance, constructive advice, criticism, continuous support and encouragement. **Prof. Washington Yotto Ochieng** (Imperial College London), **Prof. Yike Guo** (Imperial College London), **Prof. Chaozhe Jiang** (Southwest China Jiaotong University) and **Prof. Liping Fu** (University of Waterloo), thanks you all for your advice and encouragement. Also, thanks to **Ke Han** for the past. Thank you all again for your strategic suggestions, assistance, and encouragement throughout my research work, which are incredibly valuable to me.

Thanks for the experience of participating in the **U-BLOX** project where **Alghananim Mamon** and I learned machine learning techniques together. Thanks to **Yang Yu**, **Shiming Xu** for their encouragement at the later stage of my Ph.D. study. And thanks to all the CTS and DSI members.

I am grateful to the **China Scholarship Council (CSC)** for awarding me a Ph.D. scholarship. Also, gratitude goes to the Cloud-Guizhou Data Science Research Institute for providing the mobile sensing datasets in Beijing case study. Without these data, this research would not have been possible.

Finally, I would like to thank my family members (**my parents, brother and Jessie Jiang**) for your continuing love, which means a great deal to me more than how I can express.

Chapter 1

Introduction

1.1 Research context and motivation

The dramatic increase in air pollution globally, particularly within developing countries as a result of rapid urbanisation and motorisation, has created obstacles that threaten economic development and public health in serious ways. Worldwide, 6.67 million deaths are attributable to poor air quality and this pollution is partially responsible for 12% of all global deaths (State of Global Air, 2020). In China, $PM_{2.5}$ pollution concentrations remain above the WHO's lowest target (State of Global Air, 2019). The issue of air pollution has become a major problem in developing countries due to the increasing population's activities, traffic emissions, extended industrialisation, insufficient monitoring resources, and various policy issues (Kumar et al., 2013; Mihăiţă et al., 2019). According to the WHO, (1999), the scales of atmospheric processes can be classified into (i) macro-scale (length scale $\geq$ 1000 km); (ii) meso-scale (1 km < length scale < 1000 km); and (iii) micro-scale (length scale $\leq$ 1 km). It is of crucial significance to both public health and environmentally-sustainable urban planning that details concerning air pollution, its causes, and its impact on small-scale city dynamics, are available (Lü et al.,2015). Air pollution varies greatly in micro environments, influenced by the density of pollution transmission sources, population activity, and more (Sloan et al., 2016). Even for people who live in the same area, exposure can vary within a range of more than 30% in one day (Sloan et al., 2016; Dons et al., 2011). Policy-makers usually publish air pollution health risk information to help citizens reduce exposure in highly polluted areas. However, air pollution

---

The partial content in this chapter has been published in the IEEE Internet of Things Journal (J. Song, K. Han and M. Stettler, "Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing",** ***IEEE Internet of Things Journal**, DOI: 10.1109/JIOT.2020.3041047).**

monitoring networks in many regions are still insufficient to provide air pollution details (Adams & Kanaroglou, 2016) and policy-makers can be reluctant to acknowledge the problem or lack the tools to address it. Without the details of local AQ information, citizens cannot make fully-informed decisions about their daily outdoor routes or other pursuits, and policy-makers can only make macroscopic regulating decisions on the control of air pollution triggering factors and emission sources. Over a number of years, air quality details have been supplied by fixed monitoring sites, and this data has been used for a range of aims, including the spreading of information, generation of pollution maps, and policy appraisal. However, the air pollutant measurements collected from fixed monitoring sites are temporally dense but spatially sparse and cannot provide information in regions without AQ sensing at the high cost of building fixed monitoring sites.

Regional air pollution in metropolitan areas is varied along fairly confined spatio-temporal scales. These variations are fueled by various influences, such as weather conditions, how land is used, geographical features, traffic emission, and social behaviour. Mobile and ubiquitous sampling campaigns facilitate the opportunity to cover areas more densely and to collect air quality data in greater detail with finer spatio-temporal granularities. This approach has attracted greater attention of late, representing a cost-effective and practically viable method of surveying the atmospheric environment using sensors that are cheap and easily transportable (Chong & Kumar, 2003; White et al., 2012). In comparison with fixed monitoring sites, mobile air pollution sensing offers the capability to achieve dense, high-resolution dense coverage of the environment relatively cheaply (Kumar et al., 2015; Hasenfratz et al., 2015). It is then possible to transmit these measures of gaseous pollutants (e.g., NO_x, SO_2, O_3) and particulate matter ($PM_{2.5}$ and PM_{10}) wirelessly, sending them to be processed, analysed and applied through a centralized server (US DOE, 2010). Many approaches to collect air quality data through innovative mobile technology have been tried around the world. This has been observed by scholars including Wallace et al. (2009), Wang et al. (2009), Mead et al. (2013), Maag et al. (2018) and Alvear (2018).

Furthermore, the AQ spatio-temporal varying process contains the attributes of diversity, high-dimension, multi-scale, non-stationarity, non-separability, non-linearity, and others (Wang et al., 2009). Diversity refers to the dynamic process of spatio-temporal data in trend, periodicity, and closeness. High-dimension indicates the collection of large amounts of features in both the spatial and temporal domains. Multi-scale indicates different patterns and characteristics in different scales of time granularities and spatial resolutions, which play a critical role in studying gradual patterns of characteristics in generalisation and refinement of spatio-temporal data. The term non-stationarity refers to the statistical characteristics (mean and variance) of spatio-temporal series data that varies with time and location. Non-separability refers to the coherence of the spatio-temporal data and whether the data correlate and influence each other. The term non-linearity means that there are many samples of noise and the spatio-temporal data essentially express a complex dynamic process. Compared with the common time-series data or spatial data, the non-stationarity and non-separability are the most prominent characteristics of pollutant measurements. Non-stationarity and non-separability may be caused by complex influences of meteorology, topography and emissions patterns on AQ conditions (Meiring et al., 1998). Therefore, approaches to modeling AQ conditions must consider the spatio-temporal dependency and construct a non-separable spatio-temporal correlation structure, in that the correlation cannot be separately expressed as a product of a spatial component and a temporal component (Meiring et al., 1998) which models the spatio-temporal correlations simultaneously.

Existing AQ modelling techniques encompass bottom-up (physical/chemical) models and top-down (data-driven) models. The former allows for high model fidelity in relation to air pollution's physical and chemical processes; they rely on expensive computing resources and empirical condition settings, and parameter assumptions that might not be adaptable to all environmental situations (Chen et al., 2018). In contrast to the physical and chemical models, another option is data-driven models which aim to provide air pollution mapping by establishing spatio-temporal correlations with reference to types of land use, geography, meteorological conditions, emissions from vehicles, social behaviours and other explanatory factors (Zheng et al., 2013; Qi et al., 2018).

1.2 Research goal and objectives

This research mainly focuses on particulate matter ($PM_{2.5}$: particulate matter with less than 2.5 um; and PM_{10}: particulate matter with less than 10 um), considering the highly adverse health effects and increased mortality risks of each (Giannadaki et al., 2014; Ho et al., 2018; Madaniyazi et al., 2015) after the public is subjected to long-term exposure.

Achieving AQ inference over spatio-temporal domains means computing the AQ estimations at any location contiguously without gridding resolutions in the study area, and 'maintaining continuous density; however, it computes expensively if learning a continuous function over spatio-temporal (e.g., GPs) to achieve AQ inference. Inspired by the gridding method of Zheng et al., (2013) and motivated by the public health resolution of policies such as the Qianliyan Initiative, (2018), this research makes a choice of 1 km × 1 km as the scaling unit and the research goal is set to achieve high-resolution (1 km × 1 km, hourly) AQ inference in a sizeable urban space (Chengdu, 4,900 km^2; Beijing, 3,025 km^2) supported by multi-source AQ samples (fixed or mobile) and urban features. Figure 1.1 demonstrates the diagram of the end-end pixelwise AQ inference. This research mainly focuses on particulate matter ($PM_{2.5}$ and PM_{10}), considering their highly detrimental health effects and public mortality risks after persistent exposure. In contrast to existing bottom-up approaches, this enquiry explores ingrained spatio-temporal correlations between AQ conditions and feature configurations by leveraging ML-based modelling techniques to construct generalised correlations to model air pollution in a spatio-temporal domain. The research enquiry is therefore data-driven in a way that bottom-up approaches are not. This is achieved via rich and heterogeneous urban datasets, including pollutant measurements from fixed and mobile sensors, and urban features including meteorology data, traffic data, population data, and land use data.

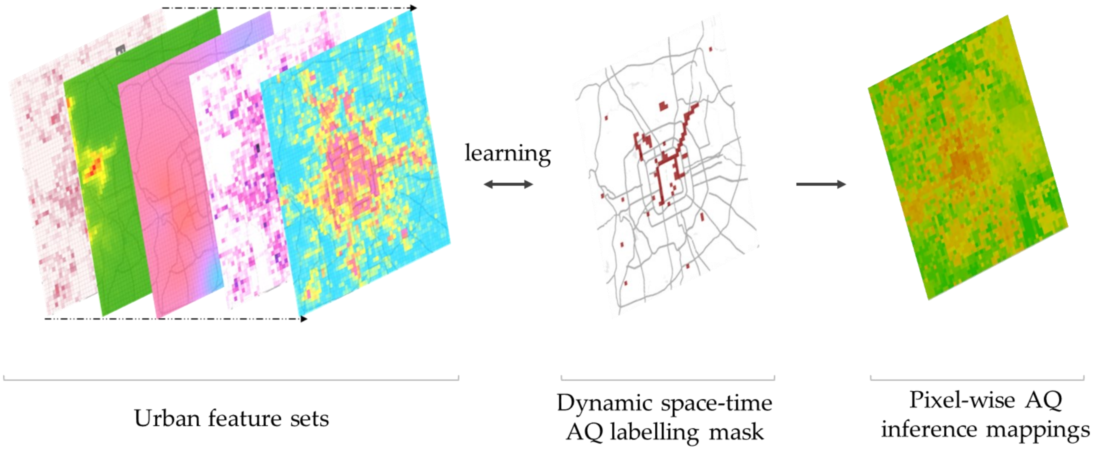


Figure 1.1 Diagram of end-end pixelwise AQ inference

To achieve this, the primary objectives have been formulated as follows, based on studies of Chengdu and Beijing:

1. Conduct a feasibility study through an extensive literature review in Chapter 2 on viable AQ sensing and modeling approaches, in conjunction with ML in common spatio-temporal analytics. This informs the subsequent data-related requirements (such as the quality, quantities, types and granularity of data), suitability, and capacity to choose ML-based algorithms based on practical applications and the granularity exhibited by the output of the models.

2. Collect multi-source AQ samples and urban features for data-driven spatio-temporal analytics in Chapter 3, which informs the measurements from fixed or mobile sensors and urban feature configurations and dynamics. In the Chengdu case study, only fixed monitoring sites are used to explore the Deep AQ framework's predictivity in Chapter 6. In the Beijing case study, both fixed monitoring sites and low-cost mobile sensors are used in the Deep AQ framework to explore the predictivity of this framework in detail in Chapter 7. Representing urban characteristics in detail and in a balanced way is a crucial component of data-driven approaches to reducing overfitting and increasing the model's interpretability. These urban features have been arranged in four categories: geographical and land use, traffic, population, and meteorological variables.

3. Propose spatio-temporal learning and a spatial inference framework (Deep AQ framework) in Chapter 5 to incorporate multi-source AQ samples (fixed or mobile) and urban features to achieve high-resolution (1 km × 1 km, hourly) pixelwise AQ inference. To achieve this objective, urban features are transformed from micro-view, meso-view, and macro-view as input variables for the Deep AQ framework. Encompassing the endogenous features, samples from the fixed monitoring sites surrounding the study area are classed as macro-view features, allowing the model to account for regional transportation of air pollution.

4. Apply the Deep AQ framework to the Chengdu case study using fixed monitoring sites to evaluate the model's performance and applicability in Chapter 6. The model will be compared with existing methodologies based on the HOCV tests, and the LOOCV tests are also applied to evaluate the model's predictivity and to achieve AQ inference at unsampled grid-cells. In addition, the spatial block CV tests are complemented to evaluate the model' performance.

5. Apply the Deep AQ framework to the Beijing case study using both fixed monitoring sites and low-cost mobile sensors to evaluate the model's performance and applicability in Chapter 7. The model will be compared with existing methodologies based on the HOCV tests, the LOOCV tests are also applied to evaluate the AQ inference at unsampled grid-cells. Sensitivity analysis of noise variance and mobile AQ spatio-temporal coverage are respectively carried out to evaluate their influence on the model's predictivity. Besides, spatial block CV tests are performed in order to evaluate the model's performance.

6. Propose a methodological framework dynamically modelling pixelwise cumulative inhalation exposure in Chapter 8. The framework will involve pixelwise AQ inference, population dynamics, inhalation rate, cumulative inhaled $PM_{2.5}$ masses, and multi-scale (i.e., hourly, daily) temporal inhalation exposure analysis.

This thesis focuses on the analysis of the Chengdu and Beijing case studies, given the availability of data discussed in section 3.2. Furthermore, this research achieves high-resolution (1 km × 1 km, per-hour) pixelwise AQ modelling through multi-source AQ sensing (fixed or mobile) and urban features, which provides comprehensive data as a basis for dynamic exposure analysis.

1.3 Challenges

Modelling urban air pollution to achieve high-resolution AQ inference based on fixed or mobile sensing in this thesis mainly met the following challenges.

Modelling urban air pollution to achieve high-resolution AQ inference based on fixed or mobile sensing in this thesis involved the following main challenges.

1. Urban features encompass static and dynamic multi-source features, and the dynamic features are constantly changing in spatio-temporal domains; thus, it remained challenging to decide how to process and extract these urban features and fuse them with an appropriate data structure for model training.

2. Considering the influence of various strides of neighbouring time slices and Euclidean distance of spatial areas, the spatio-temporal structures had to facilitate the potential to model and construct correlations for neighbouring features adaptively; it was therefore challenging to construct the spatio-temporal structures adaptively to build up the correlations between close time slices and neighbouring spatial areas.

3. Most existing studies consider urban endogenous influencing factors, overlooking potential influence of pollution transported from nearby regions outside the study area (e.g., as much as 60% of air pollution was due to regional transport in Beijing in March 2019). By using only endogenous factors for modelling creates a high risk of overfitting, and compromises the interpretability of the model results; therefore, the question of how to utilise the background information (e.g., regional transportation of air pollution) to gain information remains a challenge.

4. Data mining from multiple data views in spatio-temporal analytics plays a key role in achieving high-resolution AQ inference; thus, the fusion of high-

dimensional spatio-temporal features from multiple data views (micro, meso, and macro) remains challenging.

5. Low-cost sensors are sensitive (Chu et al., 2020) and tend to generate higher bias and uncertainties with longer operating time (Bi et al., 2020). As a result, it remained a challenge to know how to design the online-field calibration strategy to rectify deviations and keep the regional sensors well-calibrated (Chu et al., 2020) during deployment.

6. Unlike inferring AQ based purely on fixed monitoring sites where pollutant measurements are continually perceptible for a given period in a specified place, the distribution of mobile measurements is sparse and irregular. As a result, accommodating the irregular dynamic of AQ samples remains a challenge.

7. The AQ samples are collected from different sources (fixed or mobile), and the mobile measurements require calibration using fixed data in order to adjust for the bias relating to ambient temperature and humidity. Due to the variability of multiple sensors, inherent inconsistencies exist between mobile and fixed measurements; therefore, correcting mobile measurements by soft calibration to eliminate discrepancies is also a challenge.

8. There is potential for model overfitting as a result of either limited training data or an unbalanced representation of urban feature configurations and dynamics. It is essential to evaluate datasets which exhaustively cover urban features pertinent to air pollution.; therefore, reducing model overfitting is another challenge.

9. Existing health studies mainly consider daily or long-term exposure and health analysis at regional level or city level; thus, modelling inhalation exposure analysis locally at a high-resolution and with fine-granularity remains a challenge.

1.4 Thesis outline

This thesis comprises nine chapters, as elucidated below.

Chapter 1: Introduction

Chapter 1 identifies the research background, provides an overview introduction of the high-resolution pixelwise AQ inference modelling process and the primary objectives to be achieved, and sets out the research goals, main objectives, and challenges.

Chapter 2: A critical review of relevant literature

Chapter 2 reviews literature about the importance of air quality to public health and introduces viable sensing means to survey atmosphere environment and AQ modelling approaches. The advantages and drawbacks of remote sensing, fixed monitoring sites, and mobile air pollution sensing are respectively introduced. This chapter also points out that mobile air pollution sensing is a cost-effective and practically viable means of surveying details of AQ information with high spatio-temporal resolution. In addition, AQ modelling approaches, including bottom-up models and data-driven models, are illustrated in detail with reference to their advantages and limitations. Furthermore, ML-based spatio-temporal modelling analytics in relation to other research problems (such as traffic prediction) are also illustrated in this chapter.

Chapter 3: Urban case studies

Chapter 3 introduces the Chengdu and Beijing case studies and describes in detail the datasets used. This chapter introduces the AQ sampling mechanisms and urban feature datasets in the Chengdu and Beijing case studies. This chapter also introduces the details of mobile sampling campaigns and pre-processing, including the description of the IoT sensing platform and online field calibration platform for mobile sensing, and also demonstrates mobile air pollution sensing power by matching the daily spatial coverage during the period studied. Furthermore, this chapter also defines the components applied in the framework, including the concept of grid-cell unit, label responses, and AQ inference, and explains the configurations of urban features.

Chapter 4: Review of standard ML algorithms

Chapter 4 reviews the standard LUR and ML-based algorithms used in the Deep AQ framework. An introduction to the concept and algorithmic process of LUR models is provided, and convolution techniques, the principle of bagging or boosting tree-based approaches, fully convolutional network (FCN), and Huber loss function are presented in this chapter. Firstly, the algorithmic process of LUR models is reviewed. Secondly, an illustration is provided of how convolution operates to gather neighbourhood information together, in addition to the related sub-techniques (pooling, padding). Thirdly, the principle of the tree-based approaches and ensemble methods (e.g., bagging and boosting) are reviewed. Fourthly, the principle of FCN and how it works to achieve the pixelwise inference by the end-end learning is introduced. Fifthly, the Huber loss function is explained to improve the robustness of model performance.

Chapter 5: Development of Deep AQ framework

Chapter 5 proposes the Deep AQ framework, which can be applied to achieve high-resolution (1 km × 1 km, hourly) pixelwise AQ inference through multi-source (fixed or mobile) AQ samples and urban features. In this chapter, novel approximate functional correlations among multi-source urban features and AQ samples are proposed, and the relevant transformed or generalised components for urban features are demonstrated in detail. Additionally, this chapter describes in detail how to use the multi-source datasets. Based on the reviews in Chapter 4, the tree-based and network-based algorithms are selected as the fundamental supervised learning approach used in the Deep AQ framework and are independently named DFeaST-Tree and DFeaST-Net. In addition to DFeaST-Net, a novel multi-pollutant spatial-temporal network (Multi-AP) is constructed based on DFeaST-Net to estimate multiple pollutant species simultaneously within one integrated learning network. In this chapter, the process for extracting, transforming, generalisinge, and combining the micro-view, meso-view, and macro-view features are illustrated, and the technical details about the model structure of Deep AQ framework are demonstrated.

Chapter 6: Predicting AQ in Chengdu using fixed monitoring sites

Chapter 6 investigates the Deep AQ framework in Chengdu concerning $PM_{2.5}$, PM_{10} and O_3 using fixed monitoring sites. This chapter illustrates the experimental details and conclusions of the Chengdu case study. Firstly, the Deep AQ framework is compared with other benchmarks through 5-fold cross-validation to test the model's performance against the ground-truth values of all monitoring sites. Secondly, the standard deviation of different evaluation metrics for $PM_{2.5}$, PM_{10} and O_3 are compared, and the results show that the Deep AQ framework's performance is more consistent than other approaches. Thirdly, to investigate the availability of various feature variable sets for the Deep AQ framework, this chapter carries out sensitivity analysis with various subsets in micro-view (L) feature inputs. Fourthly, independent LOOCV tests are applied to evaluate the Deep AQ framework's predictivity at unsampled sub-areas. Fifthly, high-resolution (1 km × 1 km, hourly) pixelwise AQ inference mappings are depicted to show the spatio-temporal heterogeneity. Sixthly, conclusions are made about the Chengdu case study.

Chapter 7: Predicting AQ in Beijing using fixed and mobile sensors

Chapter 7 investigates the Deep AQ framework in Beijing to $PM_{2.5}$ and PM_{10} with collaborative sampling based on fixed and mobile sensors. This chapter illustrates the experimental details and conclusions of the Beijing case study. Firstly, this chapter discusses the details of how to correct mobile measurements to fixed observations. Secondly, this chapter compares the Deep AQ framework with other benchmarks by 5-fold cross-validation to test the model performance against the ground-truth values. Thirdly, high-resolution (1 km × 1 km, hourly) pixelwise AQ inference mappings are shown to depict the spatio-temporal heterogeneity. Fourthly, the sensitivity analysis of mobile AQ samples is carried out to analyse the influence, including the comparison with corrected and uncorrected mobile data, cases with or without mobile data in training or testing datasets. Fifthly, uncertainty analysis of noise variance is carried out to test the robustness of the model in relation to data noises. Sixthly, sensitivity analysis of mobile sensing spatio-temporal coverage is carried on to analyse how mobile sensing's spatio-temporal coverage influences the model's performance. Seventhly, independent LOOCV tests are applied to evaluate the Deep AQ framework's predictivity in unsampled sub-areas. Eighthly, the relative importance of urban features is quantitively analysed to produce insights into their possible relationship with air pollution. Ninthly, this chapter indicates the economic value of saving hardware investments to achieve such spatio-temporal coverage and accuracy. Finally, conclusions are made about the Beijing case study.

Chapter 8: Towards modelling pixelwise cumulative inhalation exposure

Chapter 8 proposes a methodological framework for modelling pixelwise population-weighted cumulative inhaled $PM_{2.5}$ masses. Firstly, relevant health studies about short-term and long-term exposure to $PM_{2.5}$ and health risk are reviewed. Secondly, the methods in existing health studies of estimating cumulative inhaled $PM_{2.5}$ masses are reviewed. Thirdly, the proposed methodological framework incorporates inhalation rate with pixelwise population dynamics and AQ inference to compute the pixelwise population-weighted cumulative inhaled $PM_{2.5}$ masses at fine granularity and high-resolution (1 km × 1 km, hourly). Fourthly, pixelwise population-weighted cumulative inhaled $PM_{2.5}$ masses are estimated based on LBS-R data at different temporal scales to achieve pixelwise mappings, and the spatial heterogeneity is analysed. Fifthly, the hourly and daily exposure assessments are compared based on LBS-R-based and census-based methods.

Chapter 9: Conclusions and future work

Chapter 9 outlines the key aspects and conclusions of each chapter, summarises the details of achieving research objectives and lists the contributions made by this research, its limitations, and future work.

1.5. Research publication

The partial content presented in Chapter 1, 2, 3, 5 and 7 are preprinted in Arxiv (Exploring Urban Air Quality with MAPS: Mobile Air Pollution Sensing, arXiv:1904.12303v1), which has been published in the IEEE Internet of Things Journal (J. Song, K. Han and M. Stettler, 'Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing', *IEEE Internet of Things Journal*, DOI: 10.1109/JIoT.2020.3041047). The partial content in Chapter 3, Chapter 5 and Chapter 6 is under review ('A novel multi-pollutant spatio-temporal learning network for air pollution inference'). The partial content in Chapter 7 and Chapter 8 is under preparation for journal submission (' Towards high-resolution population-related inhalation exposure modelling based on LBS-R data').

Chapter 2

A critical review of relevant literature

2.1 Introduction

Chapter 1 identifies the background, goal, objectives, and challenges of this research. This chapter presents a critical perspective on relevant literature that relates to the research in this thesis. A review of air pollution sensing methods, the bottom-up (deterministic) and data-driven (statistical) models, are provided in this chapter. Controlling air pollution is currently an on-going effort, and the same goal is undertaken by most policy-makers across the world, whereby a large amount of money is invested in preventing air pollution damage (Breathe London, 2018; Qianliyan, 2018). The global market for ubiquitous air pollution detection is fast expanding: its mean growth rate is 14.3% and by 2024 the value of the market is set to reach $530m (Wire, 2019), while China is making efforts to control air pollution with investment opportunities of $3.2 trillion from 2018-2030 (China Daily, 2018). According to the 2012 AQ assessment (MEDE, 2012), urban air pollution is mainly caused by activities involving industry and domestic behaviours.

2.2 Air quality importance to public health

The State of Global Air (2019) approximates that 95% of people globally encounter air pollution, which is deemed by the WHO to be an environmental risk. The WHO also suggests that 3.7 million deaths, almost nine tenths of which were in developing nations, were caused by outdoor air pollution in 2012. The State of Global Air (2019) notes that a one-day-old child's life expectancy is shortened by an average of 20 months by being exposed long-term to air pollution. The health risks linked to air

---

The partial content in this chapter has been published in the IEEE Internet of Things Journal (J. Song, K. Han and M. Stettler, "Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing",** ***IEEE Internet of Things Journal**, DOI: 10.1109/JIOT.2020.3041047).**

pollution include heart and respiratory diseases, lung cancer, and others. If air pollution levels are reduced, the effects of these diseases will be reduced. Outdoor air pollution caused over 4 million premature deaths in 2016, while indoor pollution has been found to pose a serious health risk to over 3 billion people. Across Southern Asia, throughout most of the calendar year, levels of fine particulate matter ($PM_{2.5}$ and PM_{10}) exceed the WHO's limits. China records as many as 1.3 million deaths annually as a result of air pollution, with the mortality and disease involved was estimated to have a monetary value of $1.4 trillion in 2010, a figure which continues to rise. The number of premature deaths in the United Kingdom attributed to air pollution is approximated to be 40,000 people per annum (State of Global Air, 2019). According to health studies (Atkinson et al., 2014; Xu et al., 2018), fine particulate matters are considered the major air pollutant which increases the morbidity and mortality risks. This research selects $PM_{2.5}$ and PM_{10} as pollutants of concern due to their adverse health effects, which are epidemiologically associated with the risk of human health and lung diseases (Du et al.,2010; Qiu et al., 2013).

2.3 Air pollution sensing

In respect of air pollution sensing methods, fixed monitoring sites act as the most widely used methods to monitor air pollution. However, several other methods for obtaining information on AQ conditions have appeared along with technological developments in recent years, including: (i) satellite remote sensing, which has enough spatial coverage but limited temporal resolution; and (ii) mobile sensing using low-cost sensors, with the potential for both good spatial coverage and temporal granularity comparing to fixed sensing, especially with regards to spatial resolution. Figure 2.1 compares the advantages and limitations of these air pollution sensing methods (O'Keeffe et al. 2019). In brief, satellite remote sensing is capable of providing enough spatial coverage (e.g., whole city-level), but their temporal granularity or spatial resolution is limited. Conversely, fixed monitoring sites are capable of continually collecting high-quality AQ observations for long periods (temporal density), but with sparse spatial coverage (e.g., immediate vicinity level). Mobile sensing utilises float vehicles as 'data mules', which offers a low-cost, scalable, and sustainable way to dynamically scan cities and obtain spatio-temporal details of AQ conditions, but low-cost sensors sometimes generate low-quality measurements (Chu et al., 2020).

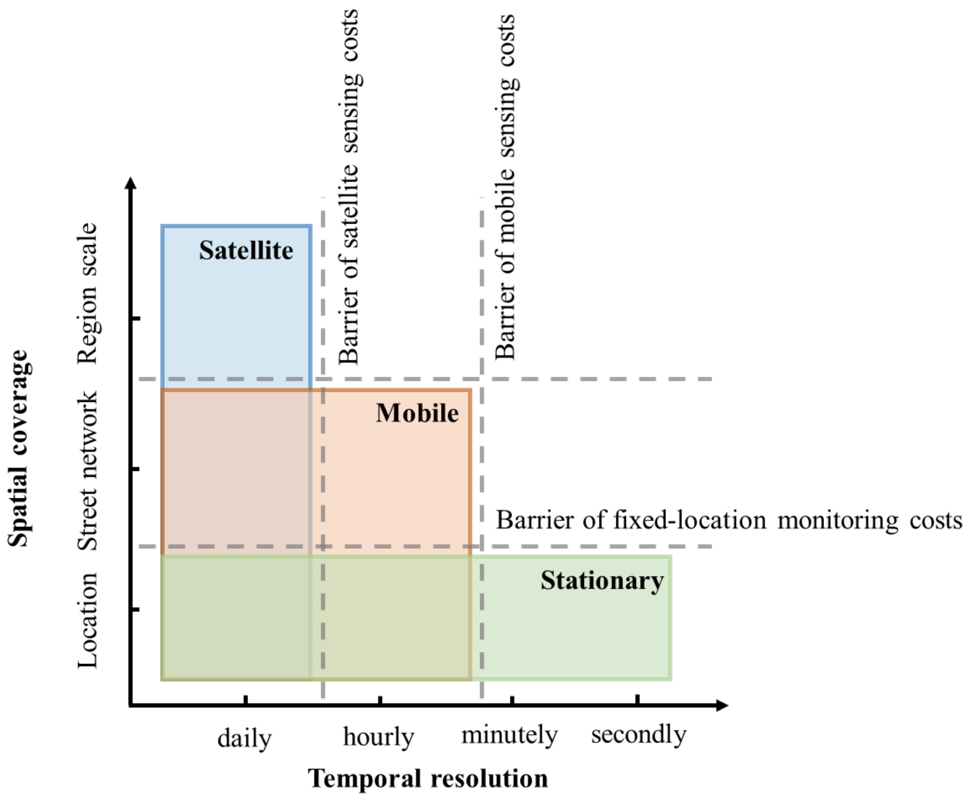


Figure 2.1 Comparison of AQ sensing methods (O'Keeffe et al. 2019).

2.3.1 Satellite remote air pollution sensing

Satellite remote sensing is a method commonly used to obtain information on AQ conditions at a large-scale (Wang & Christopher 2003, Gupta et al. 2006, Lee et al. 2009, Ma et al. 2014, Li et al. 2015). Figure 2.2 shows a diagram of satellite remote sensing. However, satellite remote sensing is inappropriate to monitor the near-surface AQ conditions, as it scans outside the atmosphere's clouds and thus can only provide indirect information. The information also varies significantly depending on location, determined by multiple factors, including local meteorological conditions, land use type, and others (Van et al. 2010, Zheng et al. 2017). In fact, satellite-derived atmospheric aerosol optical depth (AOD) (Van et al., 2006; Van et al., 2010; Wang & Christopher, 2003; Chu et al., 2016) has significantly advanced the understanding of the space-time variations of $PM_{2.5}$ and PM_{10} concentrations at large scales.

Considering the complete spatial coverage of satellite sensing, the AOD retrievals can remedy the gaps in locations where monitors are missing (Chu et al., 2016; Xu et al., 2018). However, using AOD to replace the ground-level AQ conditions at micro-scale directly may be inappropriate due to the complex relationships between AOD and ambient particulate matters, as illustrated by various scholars (Lee et al., 2011, Paciorek et al., 2008, Xu et al., 2018, Song et al., 2019). Furthermore, auxiliary information, including geographical and meteorological data, is exploited to improve modelling performance (Xu et al., 2018).

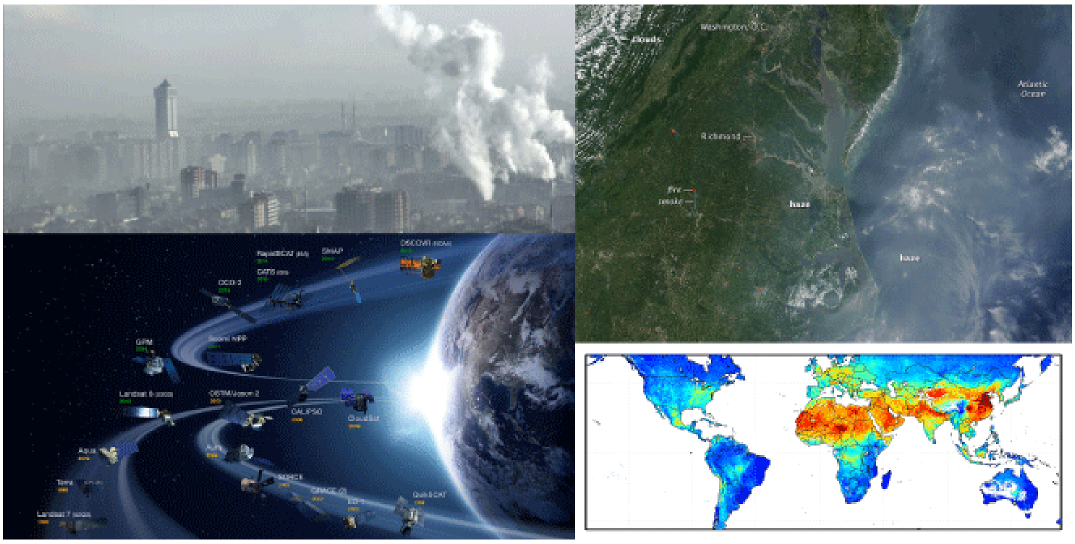


Figure 2.2 Satellite remote sensing (NASA, 2017)

2.3.2 Fixed air pollution sensing

Currently, policy-makers monitor AQ conditions regionally through a sparse monitoring networkm with fixed monitoring sites used to measure pollutants including carbon monoxide (CO), nitrogen oxides (NO_x), sulphur dioxide (SO_2), ozone (O_3) and particulate matter (PM). Some of the fixed monitoring sites in Beijing and Chengdu are shown in Figure 2.3. AQ conditions are often provided along with details of pollutant concentrations or air quality index (AQI) using a scale that the public can easily interpret. However, these AQIs have different definition standards in different regions (Plaia and Ruggieri, 2011), and these standards are sometimes different between countries.

Fixed air pollution sensing can offer high-precision results, but has high maintenance costs (Chong and Kumar, 2003). Often, the fixed monitoring network's spatio-temporal resolution is far too sparse to provide detailed information on AQ conditions. Consequently, there are 40 fixed monitors within the Chengdu study and 28 fixed monitors in the Beijing study, which are unevenly distributed in different directions (Zheng et al., 2013). The fixed monitoring sites are capable of providing regional insight, but they cannot identify details of the micro-environment. Thus, these tools offer the emergent idea of using low-cost sensors which can be deployed densely with cost-effective maintenance.

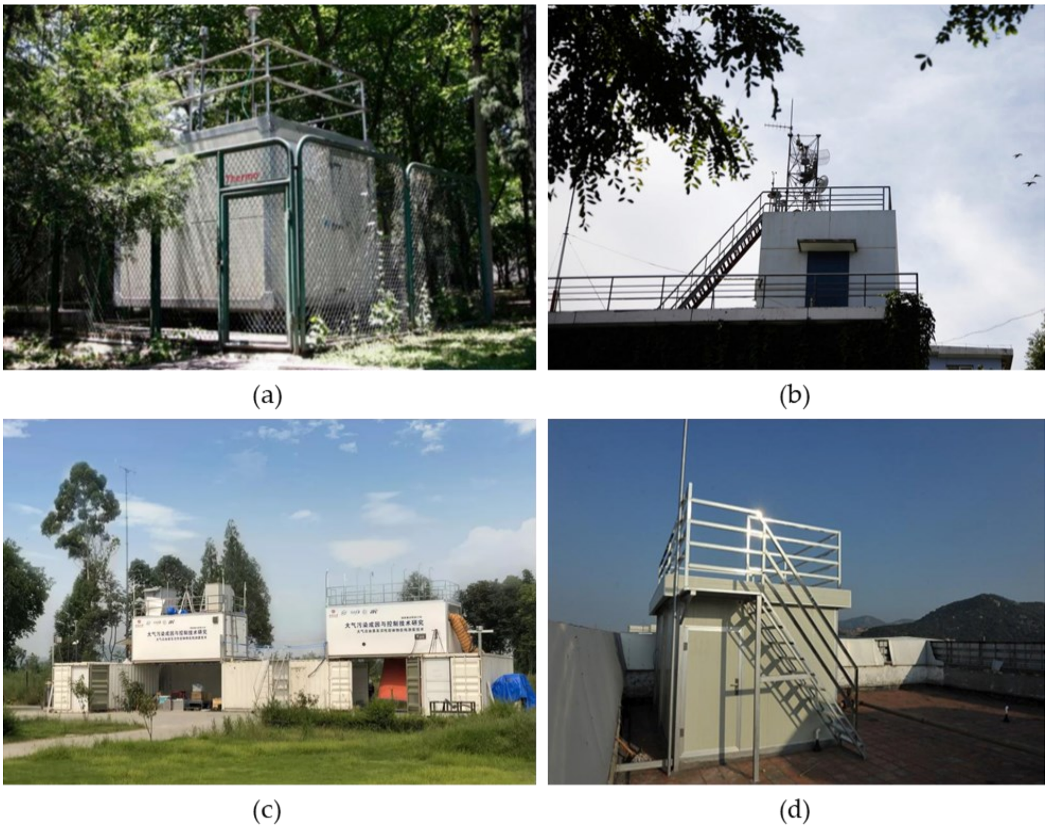


Figure 2.3: Fixed AQ monitoring sites. (a) Located in the Dongcheng area of Beijing; (b) Located in the Nongzhan Guan area of Beijing; (c) Located in the Xinjin area of Chengdu; (d) Locatied in the Longquan area of Chengdu

2.3.3 Mobile air pollution sensing

Mobile air pollution sensing refers to mobile sampling campaigns which have the capability to inexpensively enlarge the existing fixed AQ monitoring networks and generate novel AQ measurements in real time. Figure 2.4 shows a diagram of mobile sampling instruments.

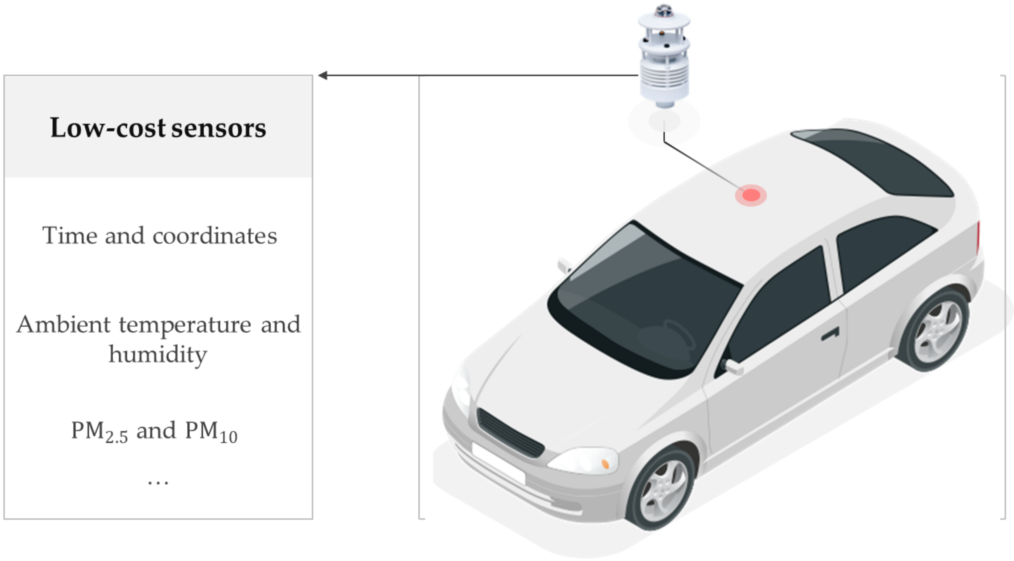


Figure 2.4: Diagram of mobile sampling instruments

Through modern technological advancement and the spread of AQ sensors, there are now additional ways to fine-tune the surveying of the atmosphere, using multiple micro-sensing units (MSUs), which consume less power than distributed mobile sensing nodes, to observe detailed pollutant measurements (Mihăiţă et al.,2019). There is potential for mobile sampling campaigns that deploy low-cost (< $300) sensors to make possible a non-expensive and workable method to bring advantages in respect of real-time pollutant measurements, and the exploration of potential emission sources, and real-time dynamic sensing nodes, make it possible to rapidly acknowledge the air pollution problem and help policy-makers make effective supporting decisions. Due to the high distribution of sensing and the rapid access to real-time pollutant observations that they provide, various worldwide projects are exploring the exploitation of mobile air pollution sensing, including Air Quality Egg, (2016), Citizen Sense, (2016), the Smart Citizen Kit, (2016), Breathe London, (2018), Qianliyan Initiative, (2018) and others. These initiatives collect, process and – in real time – map data centrally. However, these mobile sensing applications still give rise to numerous challenges associated with a large volume of generated sparse data and the low accuracy of their measurements (Doering, 2011; Devarakonda et al., 2013; Mead et al ., 2013; Hasenfratz et al., 2015; Kumar et al., 2015; Marjovi et al., 2015; Suriano et al., 2015; Adams & Kanaroglou, 2016; Esposito et al ., 2016; Apte et al., 2017; Hu et al., 2017; Alvear, 2018; Breathe London, 2018; Chen et al., 2018; Mihăiţă et al., 2019; SM et al., 2019). This means the algorithms play prominent roles. Devarakonda et al. (2013) present a vehicle-based method of mobile sensing used to collect highly covered AQ observations, and the LR is applied to estimate carbon monoxide concentration. SM et al. (2019) reveal the non-linear correlations between pollutant concentrations through using portable air pollutant measuring equipment carried by pedestrians or installed on buses in India. Zappi et al. (2012) and Suriano et al. (2015) deploy mobile low-cost sensors to scan streets and collect the air pollutant measurements for visualisation. Minet et al. (2017) use crowdsensing to measure NO_2 concentration in Montreal and capture the sensitivity of the LUR models in short-term mobile sampling; however, the development of LUR models show poor performance. Marjovi et al. (2015) use mobile measurements in Lausanne, Switzerland. They constructed a probabilistic graphical model to measure lung-deposited surface area at street level, although this inference was confined to parts of the streets with mobile sensing coverage, while the probabilistic method performs poorly when dealing with temporally and spatially sparse mobile data (Leung et al., 2019), hindering its application in sizeable urban locations. The results are confined to street level and have restricted temporal granularity and spatial resolution (daily, weekly, monthly). Hu et al. (2017) use various regression approaches to approximate the CO concentration surface in Sydney - they use both fixed and mobile sensing data, although they do not apply other explanatory parameters. Studies like these have contributed to advancement of data-driven tools, however there remains no systematic modelling framework which is able to produce high-resolution AQ inference mappings (e.g., 1 km × 1 km, per hour) that is derived from irregular sensing and spatio-temporal coverage and which also uses multi-source features to ensure that the model is robust and accurate.

2.4 Air pollution modelling

A large number of studies have attempted numerous approaches to estimate air pollutant concentrations over the past decades. To recover pollution maps, the most straightforward approach is to use interpolation based on sparse AQ monitors. The interpolation approach can also be applied efficiently in large areas with the help of GIS from the monitored air pollutant measurements (Briggs et al., 2005). However, this still represents a relatively undeveloped process. Prizor (1988) divides interpolation approaches into global approaches (e.g., trend surface analysis) and local approaches (e.g., IDW, kriging, spline) in which the local estimates are computed with the nearest data points. Instead, the performances of these interpolation approaches tend to be selected based on the underlying characteristics of space-time data (such as distribution and sampling density) (Briggs et al., 1997). Among these interpolation approaches, Kriging is most popular in air pollution modelling applications (Briggs et al., 2005). However, these approaches are too simplistic with such sparse observations, which require more predictive models for dynamic air pollution modelling.

Scientific figures have studied emissions, chemical interactions, and the transportation of air pollutants using bottom-up approaches, using existing data on emission sources (Beckx et al., 2009a, 2009b). An alternative to bottom-up models is to apply data-driven approaches for AQ inference in order to explore the spatio-temporal correlation and explain it with reference to variables including weather, traffic and urban planning. Table 2.1 shows examples of air pollution modelling. Different models are capable of establishing the correlations between air pollutant measurements and the various selected explanatory factors using statistical and ML-based approaches.

Table 2.1: Examples of air pollution modelling for AQ inference

Approach	Procedure	Method	Authors
Interpolation	Kriging	Using ordinary or universal kriging to interpolate air pollution mappings from monitored data	Liu and Rossini (1996) Nori-Sarma et al., (2020)
	Inverse distance weighting	Using inverse distance to interpolate air pollution mappings from monitored data	Hoek et al. (2001, 2002) Masroor et al., (2020)

Dynamic modeling	Bottom-up model	Using models of dispersion pathways and rates to estimate air pollution	Gibson et al., (2013); Fakinle et al., (2021)
	Data-driven model	Using land use regression models or ML-based approaches to estimate air pollution	Zheng et al., (2013) Wang et al. (2020)

2.4.1 Bottom-up air pollution modelling

For bottom-up models, Chemical Transport Models (CTMs) (Simpson et al., 2012) are widely used to model air pollution variations. These models replicate chemical processes affected by atmospheric conditions, and 'this is the main use of bottom-up models. This process is also suitable for use alongside an inverse method which allows global gas emissions to be estimated (Chen & Prinn, 2006). The Community Multiscale Air Quality (CMAQ) model proposed by Environmental Protection Agency (EPA) is a widely used CTM model to simulate the variations and depositions of airborne gases and particles on the earth's surface (Binkowski & Roselle, 2003). The Gaussian Plume Dispersion Model is the core of most regulatory atmospheric dispersion models (Gibson et al., 2013) where the environmental parameters (e.g., wind conditions) have been used in roadside dispersion modeling which uses data on emissions from vehicles while assuming that pollutants disperse vertically and horizontally (Zhang et al., 2013). Moreover, the Operational Street Pollution Model (Hertel et al., 1991) simulates the conditions of the atmosphere in street canyons by combining the Gaussian Plume Model and a box model which accounts for street-level recirculation by pollutants (Ottosen et al., 2015). Stationarity in a spatio-temporal domain is assumed by models that draw on the Gaussian approach, thereby limiting how applicable it is to either instant or time series emission data. Differential equations and other mathematical methods are used within Computational Fluid Dynamic Models (Chu et al., 2005; Parra et al., 2010), simulating liquid or gaseous flows within parameters dictated by data. Although the method has since been used extensively in software (Glatzel et al., 2008), this tends to involve consuming a large amount of memory and computational time. Castell et al., (2018) use a three-dimensional Eulerian dispersion model known as the EPISODE dispersion model (a 3D) to estimate urban pollutant concentrations (CO, NO, NO_2 and O_3) based on 17 low-cost sensor nodes at kindergartens in Oslo, Norway. These bottom-up models are capable of offering considerable model fidelity regarding the physical and chemical components of air pollution. However, such models are usually computationally cost-intensive while being dependent on applicable assumptions which constrain how easily they can be integrated with other data (Beckx et al., 2009a, 2009b).

2.4.2 Data-driven air pollution modelling

In addition to the bottom-up approach, data driven approaches to research aims to estimate AQ conditions through finding spatio-temporal correlations based on explanatory factors including layout and planning of metropolitan spaces, the movement of vehicles, and meteorological conditions, using statistical approaches. Here, the statistical approaches include numerical and probabilistic approaches. The LUR models (Eeftens et al., 2012; Mavko et al., 2008; Su et al., 2008; Meng et al., 2015) are optimisation-based and offer feasibility to construct the correlations between land use parameters and pollutant measurements (Briggs et al., 1997; Briggs et al., 2005; Slama et al., 2007; Ryan et al., 2007; Ryan et al., 2008; Hoek et al., 2008; Dons et al., 2013; Gulliver & De, 2015; Chu et al., 2016). The LUR models regard ground-based pollutant measurements as the dependent variable, and predicators including meteorological variables, land use information, etc., as the independent variables (Hu, 2009; Kloog et al., 2011; Lee et al., 2011; Kloog et al., 2012; Ma et al., 2014; Chu et al., 2016) which utilise the MLR as a standard surrogate model to achieve AQ inference. The predicator selection also plays an important role (Hoek et al., 2008). Nevertheless, existing LUR models tend to be constructed to estimate long-term AQ conditions in the smallest scale (point-specific) (Hennig et al., 2016). The study of SAVIAH (Small Area Variations in Air Quality and Health) (Briggs et al., 1997) is the first to apply LUR for micro-scale air pollution modelling (Hoek et al., 2008) and uses different buffer sizes for point locations (i.e., monitors) to extract geographical features (e.g., traffic, terrain, etc.) to estimate the AQ variations. Afterwards, the LUR models have been popularised with the developments of GIS in the past decades and are used predominantly in large urban areas within Europe, North America, and China (Gilbert et al., 2005; Ross et al., 2007; Wheeler et al., 2008). Ryan et al. (2007) review predicators adopted in various studies from 1997-2007 and indicate that the primary four types contain the sort of road, traffic level, height above sea level, and land coverage. The ESCAPE project (Eeftens et al., 2012), which built models for 20 European areas, demonstrates a detailed modelling procedure and has become a standard approach. After the successful deployment of the ESCAPE project, other studies began to adopt the ESCAPE lists in many countries, such as Australia (Dirgawati et al., 2015), South Africa (Muttoo et al., 2018) and China (Meng et al., 2015). In fact, there are many different types of LUR studies with extendable predicators, but they mainly involve the following elements: monitoring networks, land use maps, buffer zones, dependent variables and additional independent variables. Table 2.2 summarises the key aspects of typical studies, involving traffic, demographic and land use variables ((Briggs et al., 1997; Briggs et al., 2000; Ross et al., 2007; Su et al., 2008; Mavko et al., 2008; Eeftens et al., 2012; Dons et al., 2013). Most studies explore a series of potential predicators to model AQ conditions.

Table 2.2: Summary of typical LUR studies.

Authors	Study Area	Monitor Sites	Dependent Variables	Independent Variables	Buffer Sizes	R-Square	Significant Variables
Briggs et al. (1997)	Amsterdam (NL) Huddersfield (UK) Pragure (CZ) UK	80 80 37	NO_2	Traffic, land use, topography information	0.3 km	0.59 ~ 0.86	Altitude, sampling height
Briggs et al. (2000)	Huddersfield (UK), Sheffield (UK), Northampton (UK)	20 28 35	NO_2	Road, traffic, land use, and topography information	0.3 km	0.58 ~ 0.76	Road length, intensity, housing, altitude
Ross et al. (2007)	New York City (US)	28-49	$PM_{2.5}$	Traffic, land use, population information and emission data	0.05, 0.1, 0.3, 0.5 and 1 km	0.61 ~ 0.64	Traffic within a buffer of 0.3 or 0.5 km, population density, industrial and green cover, total country primary $PM_{2.5}$ emissions

Su et al. (2008)	Greater Vancouver Regional District, (Canada)	116	NO/NO_2	Road, traffic, meteorology information	3 km	0.53 ~ 0.60	Road, traffic, meteorology (wind speed, wind direction and cloud cover/insolation)
Mavko et al. (2008)	Portland, (US)	77	NO_2	Road, traffic, land use, DEM, and meteorology information	0.05, 0.1, 0.25, 0.3, 0.35, 0.4, 0.5 and 0.75 km.	0.66 ~ 0.81	Major roads within 0.25 km, industrial land use within 0.5 km, minor roads within 0.1 km, Traffic volume on major roads within 0.1 km, parks within 0.1 km
Eeftens et al. (2012)	20 European regions	20/region	$PM_{2.5}$, PM_{10} and PM Coarse	Traffic, population, and land use information	0.025, 0.05, 0.1, 0.3 and 0.5 km	0.35 ~ 0.94	Two and five predictor variables in each area
Dons et al. (2013)	Flanders, (Belgium)	63	Black carbon	Road, traffic, population, and land use information	0.05, 0.1 and 1 km	0.44 ~ 0.77	Total road length; Heavy traffic in a buffer with radius 0.1 km; heavy traffic in a buffer with radius 0.05 km; distance to the nearest road

Furthermore, the advancement and wider use of LUR models is thanks to mobile air pollution sensing. For example, Hankey and Marshall (2015) use more than 85 hours of data gathered from a cycling platform in Minneapolis to construct moderately well-fitting LUR models incorporating particle size, black carbon, and $PM_{2.5}$. Others have used vehicle-based mobile measurements to estimate street-level variations of $PM_{2.5}$ and PM_{10} in downtown Hong Kong (Shi et al., 2016). Meanwhile, Apte et al. (2017) analyse results gathered by a Google Street View vehicle whose AQ sensors sampled each road within a $30\,km^2$ area of Oakland multiple times. Using this data, they simulate the distribution of urban air pollution, giving more spatially precise results (4-5 orders of magnitude) than existing central-site ambient monitoring. LUR models in these aforementioned studies are generally intuitive and interpretable (Wang et al., 2020). Nonetheless, the simplicity shown by LUR models places constraints on how well it can perform when modelling complicated AQ conditions (Wang et al., 2020). Recent studies have exploited ML-based means of overcoming LUR's limitations when trying to capture the non-linear interrelationships between pollutants and predictors (Champendal et al., 2014). Zheng et al. (2013) put forward the suggestion that Beijing's AQI could be inferred using a spatio-temporal classifier, mixed monitoring data and a co-training framework. Brokamp et al. (2017) use the RF with LUR framework to estimate particulate matter elemental components in Cincinnati, Ohio. Qi et al. (2018) assess the potential of using fixed sensing data to estimate Beijing's $PM_{2.5}$ concentration through semi-supervised study. Lim et al. (2019) propose LUR models incorporating ML methods of examining street-level $PM_{2.5}$ concentrations in Seoul to identify the pollution hotspots. Wang et al. (2020) explore the limitations of LUR models in addition to how ML-based models (ANN and XGBOOST) construct correlations between AQ and predicting factors by unveiling the black-box nature of ML with a mobile sampling campaign in Toronto for fine particulate matter and black carbon. Chen et al. (2020) model what happens as a result of exposure over a long period of time to eight components (nickel, copper, potassium, sulfur, iron, vanadium, silicon, and zinc) of $PM_{2.5}$. They use standardised measures of single-year intervals from October 2008 to April 2011 across 19 regions of study, using LR and RF algorithms. In addition to these numerical methods, there is another probabilistic sequence method that includes the Gaussian process (GP) model and hierarchical Bayesian model (HBM). For example, Sahu & Mardia (2005) use a Bayesian Kriging Kalman filter (KKF) spatiotemporal process to simulate the concentration of airborne pollutant particles in New York City. Sahu et al. (2006) simulate fine atmospheric particulate matter data collected in the United States using a Bayesian-based spatiotemporal stratification approach. Calder (2008) uses the Bayesian dynamic process convolution method to interpolate $PM_{2.5}$ and PM_{10} concentration data from various parts of Ohio in the time and space domains. Carslaw & Ropkins (2012) develop a software package named 'OpenAir' with various tools for importing and manipulating data to conduct an extensive analysis of air pollution data, including pollutant sources and impact factors. Bakar (2012) use Bayesian methods to analyse the maximum daily ozone level and developed the first software package named SPTIMER, using spatio-temporal HBMs to fit spatiotemporal air pollutant data. Blangiardo et al. (2013) propose a Bayesian approach to simulate the spatio-temporal variability of AQ conditions. Vitolo et al. (2018) confirm that Bayesian networks are a promising method to identify the dependencies between variables and estimate AQ conditions. Kreuzer & Czado (2019) perform Bayesian inference using the Hamiltonian Monte Carlo method to estimate particulate matter measurements in Beijing. Hamelijnck et al. (2019) propose a multi-resolution and multi-task (MRGP) framework involving shallow GP mixtures to estimate local AQ conditions in London through focused gaussian model and variational threshold extension embodies the superiority of on the background of time and space synthesis and robustness, in a variety of sensing model under different spatio-temporal resolution. Wang et al. (2020) describe a framework by constructing non-stationary and non-separable random fields based on an infinite mixture of convolved stochastic processes, demonstrating the superiority of these random field GPs to separable methods and their improvement comparing to the existing non-separable methods. Ding et al. (2021) propose a flexible HBM which can predict AQ conditions in space and time and outperform the nonlinear regression (NLR) and spatial temporal Kriging (STK) in terms of efficiency, accuracy and robustness.

2.5 Deep learning in common spatio-temporal analytics

This section introduces the general NN-based techniques used in spatio-temporal analytics, which solve other urban common spatio-temporal problems (e.g., traffic prediction). The spread of urban data over recent years, made possible by information and communication technology and widespread sensors, has facilitated in the swift advancement of computational architectures (Tang et al., 2015) and spatio-temporal data analytics (Birant & Kut, 2007). Academic enquiries have suggested various methods of constructing spatio-temporal correlations both jointly and separately. Deep neural networks like Long-Short-Term Memory (LSTM) make it possible to capture long-term temporal dependencies in conjunction with short-term variations (Zhao et al., 2017). Through the use of convolutional neural networks (CNN), Zhang et al. (2017) predict crowd flow within an urban area by translating information on a region's traffic intensity into a pixel on an image. Such a model is able to be used to predict occurrences in the short term based on information gradually acquired from images that show changes in traffic quantity as time passes. The technique known as precipitation nowcasting is performed by Shi et al. (2015) by combining CNN and LSTM (Conv LSTM), whereby the input values and prediction targets are spatio-temporal sequences. Moreover, Yao et al. (2018) anticipate the spatio-temporal taxi demand using existing data based derived from multi-view spatial-temporal network in order to construct spatial correlations and temporal dependencies at the same time in conjunction with semantic relations. Also, Zhou et al. (2019) present a training model derived from a Deep Multi-output LSTM (DM-LSTM) neural network model in order to predict the AQ within Taipei. However, the training label responses are either available for the entire area at a specific time (as is the case in relation to precipitation and taxi demand) or they are available at numerous separate locations through a non-interrupted time series (such as traffic flow) in all these studies. Not one of the research investigations above consider issues such as AQ inference with both fixed and mobile sensing, which on its own exhibits nonstationary, non-separability, and complex dependency structure of which the training label responses are irregularly distributed.

2.6 Summary

This chapter addresses the first and basic step in research objectives, which is to:

Conduct a feasibility study through an extensive literature review on viable AQ sensing and modelling approaches in conjunction with machine learning in common spatio-temporal analytics.

This chapter reviews literature about the AQ sensing and modelling approaches, highlighting the advantages and disadvantages in data-driven analytics. Besides, this chapter also introduces relevant ML-based approaches for modelling spatio-temporal learning problems. Section 2.1 reviews the critical importance of controlling air pollution and acquiring AQ details to public health and sustainable urban management. Section 2.2 introduces the concept, advantages, and drawbacks of satellite remote sensing, fixed monitoring, and mobile sensing to obtain AQ observations. Section 2.2.1 indicates that the AOD retrievals are not near ground-level, which can be used as supplementary data to obtain macroscopic changes of atmospheric conditions. Section 2.2.2 indicates that the records from fixed monitoring sites are spatially sparse and limited. Although they can produce very accurate measurements, it cost too much to build up too many fixed monitoring sites to achieve high-resolution AQ inference. Section 2.2.3 indicates that mobile sensing method is gradually being used to collect AQ observations with the development of sensor technology, and more studies are exploring to employ mobile sampling campaigns to survey the atmosphere environment in urban environment. It can be concluded that AQ observations from fixed monitoring sites are temporally dense but spatially sparse, and the AQ observations from mobile sampling campaigns are spatially dense but temporally sparse, which means that the combination of fixed and mobile sensing facilitates the potential to achieve high-resolution spatio-temporal coverage to survey atmosphere environment. Section 2.3 introduces the concept, advantages, and limitations of the AQ modelling techniques, including bottom-up models and data-driven models. Section 2.3.1 introduces the physical and chemical principle of widely used bottom-up models and indicates their limitations. Section 2.3.2 lists studies of data-driven models, including LUR models and ML-based approaches. Section 2.4

reviews literature about ML-based approaches for other common spatio-temporal problems (e.g., traffic prediction).

Chapter 3

Urban case studies

3.1 Introduction

Chapter 2 reviews literature relevant to this research, and this chapter introduces urban case studies. In this thesis, two Chinese cities are chosen as urban studies, one is Chengdu and the other is Beijing. Chengdu is a western Chinese city located between 30.05 ~ 31.26N latitude and 102.54 ~ 104.53E longitude, and only fixed monitoring sites are utilised in the Chengdu case study. Furthermore, Beijing is the capital of China which is located between 39.26N ~ 41.03N latitude and 115.25E ~ 117.30E longitude. Both fixed and mobile sensing are used in the Beijing case study and thus, collaborative sampling from fixed monitoring sites and low-cost mobile sensors to achieve urban spatio-temporal coverage. Based on granularity and spatial extent, this is the biggest method of such sampling and it makes available nuanced air quality data in sizeable urban areas that hasn't been present in related enquiries (Zheng et al., 2013; Marjovi et al.,2017; Qi et al., 2018). In addition, as numerous endogenous urban factors and the regional transport of air pollution influence air quality, modelling high-resolution air pollution requires comprehensive urban feature datasets. Furthermore, this research exploits 1 km × 1 km as micro-scale size in this research and the reasons are two-fold: (1) as set out in the Qianliyan Initiative in China, a spatial resolution of 3 × 3 km or 1 × 1 km is as the suggested default sensing resolution for nation-wide adoption. (2) Despite the potentially significant spatial variability of the concentration of air pollutants, spatial (and temporal) resolution of a ML-based model should not be too granular to improve its robustness against random events or data noise. Moreover, it means more sparse distribution of labels if the grids are divided into smaller sections due to limited mobile sensing coverage, adding significant challenges to spatial inference and validation of the results.

The partial content in this chapter has been published in the IEEE Internet of Things Journal (J. Song, K. Han and M. Stettler, "Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing",** ***IEEE Internet of Things Journal**, DOI: 10.1109/JIOT.2020.3041047).**

3.2 Description of Chengdu case study

In the Chengdu case study, a 70 km × 70 km square area was studied, spanning the around-city ring road, depicted in Figure 3.1. The pollutant samples ($PM_{2.5}$, PM_{10} and O_3) are measured from 40 fixed monitoring sites and the meteorological parameters of pressure, temperature, relative humidity, wind speed and direction, and water vapor pressure are kept in mind. The temporal scope is 0:00-23:00 (24 hours a day) for 48 days (2019/04/26 -2019/06/12), with hourly granularity.

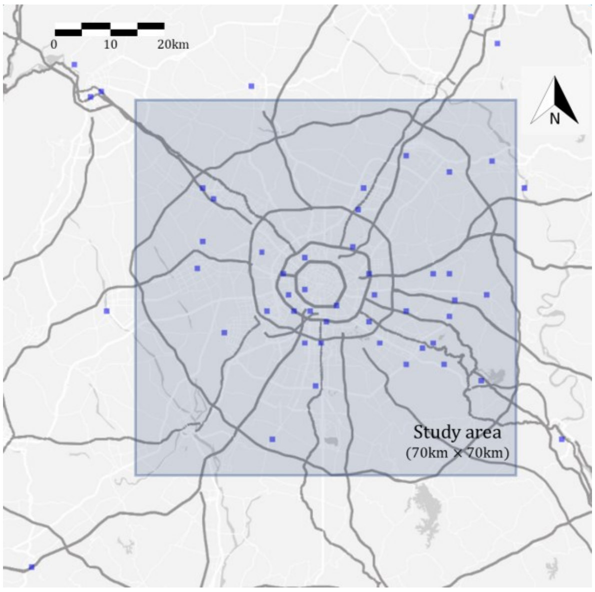


Figure 3.1: The 40 fixed monitoring sites across 70 km × 70 km study area in Chengdu and 9 fixed monitoring sites surrounding the study area.

The grid-based meteorological conditions are provided by China Meteorological Centre, integrating with the satellite-sourced and ground-sourced meteorological monitoring data. The public vitality data are provided by Sina Weibo and traffic dynamics are provided by Gaode map. The land use information is downloaded from the Chengdu Municipal Government Public Data Open Platform. The collection period of the pollutant, traffic, and meteorological data are during 00:00-23:00 (24 hours) from 2019/04/26 to 2019/06/12 (48 days). Table 3.1 lists the AQ sensing data and urban features used in the Chengdu case study, and these datasets are collected from open sources and local governments.

Table 3.1: AQ sensing data and urban features used in the Chengdu case study.

Data Category	Data Type		Spatial Resolution
Geographic & Land use	POIs (21 categories)		1 km × 1 km Grid-level Static (2019)
	AOIs (19 categories)		
	Area of green cover (%)		
	Area of water cover (%)		
	Digital elevation (m)		
Traffic	Traffic network	Number of intersections	1 km × 1 km Grid-level Static (2019)
		Lengths about different typed roads	
	Traffic dynamics	Percentage of congested roads in grid unit	1 km × 1 km Grid-level Hourly (26/04/2019-12/06/2019)
Public vitality	Sina Weibo	Number of users, posts, and comments	1 km × 1 km Grid-level Annual average (2015-2018)
Meteorology	Temperature (°C)		1 km × 1 km Grid-level Hourly (26/04/2019-12/06/2019)
	Pressure (kPa)		
	Water vapor pressure (hPa)		
	Relative humidity (%)		
	Wind direction ([0°,390°])		
	Wind speed (m/s)		
Air pollutant measurements	Fixed monitoring sites	$PM_{2.5}$, PM_{10}, O_3 (ug/m^3)	Point-based Hourly (26/04/2019-12/06/2019)
		Location ([lat, lng])	
*All the datasets used in Chengdu case study are collected from the local government and open sources.			

3.3 Description of Beijing case study

The region of Beijing studied is 55 km × 55 km square and spans the sixth ring road, depicted in Figure 3.2 (a). Its temporal scope is 0:00-23:00 (24 hours a day) over 28 days (2018/06/19 - 2018/07/16), incorporating hourly granularity.

In this Beijing case, as shown in Figure 3.2 (a), 28 fixed monitoring sites are situated in a location 3,025-km^2 in size. These sites provide data every hour on air pollutants to support innovative research and development in air pollution modelling (Zheng et al.,2013; Qi et al.,2018). These mobile observations were measured from 15 probe vehicles that collected data over 28 days, demonstrated in Figure 3.2 (c)-(e). Figure 3.2 (b) depicts two side-by-size fixed monitoring sites in Beijing along with their $PM_{2.5}$ measurements. Although the two time-series exhibit similar trends, considerable differences may exist in the hourly measurements. is the cause of such differences is variations in local air pollution that sparsely distributed fixed monitoring sites cannot adequately observe. The 28 fixed sites inside the area studied generate per-hour concentrations of particulate matter ($PM_{2.5}$ and PM_{10}) in addition to inorganic gaseous pollutants (CO_2 , CO , SO_2 , NO_x), plus meteorological parameters of pressure, humidity, temperature, water vapor pressure and wind speed and direction. These mobile measurements encompass $PM_{2.5}$ and PM_{10} measurements in conjunction with temperature and humidity data, which are sampled from low-cost mobile sensors. The directions travelled by the probe vehicles are mapped in Figure 3.2 (c).

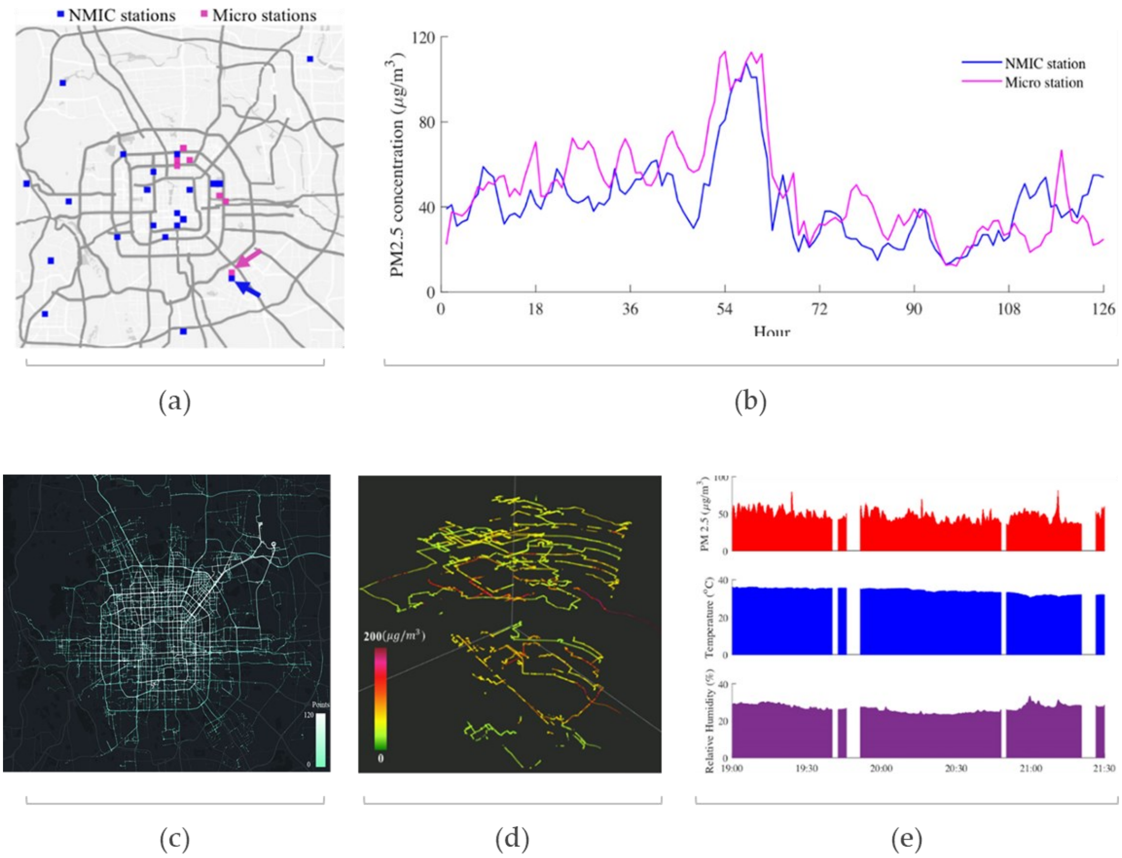


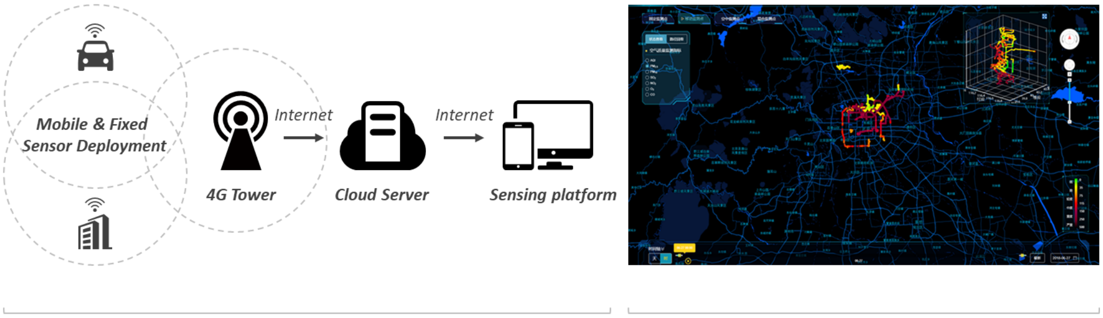
Figure 3.2: Visualisation of spatio-temporal data in the study case of Beijing. (a) The location studied (55 km × 55 km) including 28 fixed stations. (b) $PM_{2.5}$ was measured each hour by 2 adjacent stations, readings being taken during 7:00-24:00 on 2018/06/19-2018/06/26. (c) Mobile sensing coverage at street-level throughout the time studied in Beijing. (d) Spatial-temporal trajectory and $PM_{2.5}$ measurements of a probe vehicle between 2018/06/28 and 2018/07/10. (e) The measurements taken on 2018/06/21 by one of the 15 vehicles, showing $PM_{2.5}$ concentration, relative humidity and temperature at 19:00-21:30.

3.3.1 Scope of the work

This research constructs a high-resolution (1 km × 1 km, hourly) pixelwise air quality inference framework using fixed and mobile air pollution sensing in Beijing. The air pollution sensing IoT (Internet of Things) system is the basis to measure AQ observations and transmit them to the cloud server, and the online field calibration platform for mobile sampling campaigns is deployed to calibrate the mobile measurements generated from the low-cost sensors in real time. Furthermore, the mobile $PM_{2.5}$ and PM_{10} measurements are processed in advance, then geo-meshed and combined as grid-cell units, after which they are adjusted with reference to fixed AQ monitoring measurements; further detail will be given in Chapter 7 in this thesis. Figure 3.6 shows the grid-unit-based spatio-temporal coverage of the fixed and mobile sensing. Figure 3.6 (b)-(c) shows that mobile sensing provides temporally scattered data, but over a larger area. Although this may be counteracted by using a greater number of low-cost sensors, the difficulty remains of fusing heterogeneous fixed and mobile data.

3.3.1.1 IoT sensing platform for mobile sampling campaigns

Using IoT to sense air pollution requires a central network. In such a network, distributed sensing nodes measure AQ samples and transmit them into the cloud server (it should be noted that the IoT system is developed by the third institute and this research reviews the details here to provide a better understanding of the sensing data basis). Figure 3.3 (a) depicts the sensing nodes and their mobile and static configurations. Figure 3.3 (b) shows the GUI of monitoring system. The up-to-date status and trajectory of mobile sampling campaigns and AQ measurements can be seen from each sensing node.



(a) Architecture of IOT sensing platform (b) GUI of sensing platform

Figure 3.3: A centralized air pollution sampling system for (a) Architecture of IoT sensing platform; (b) GUI of sensing platform.

Mobile sampling is conducted via transportation and there are 15 taxies mounted with fast-response measurement devices which collect a large volume of AQ observations in Beijing, including low-cost $PM_{2.5}$ and PM_{10} sensors, temperature sensors, humidity sensors, and GPS receivers. Each probe vehicle equipped with low-cost sensors can be regarded as a mobile sampling node.

The frequency of the entire nodes sample, as defined by the software, is 0.833 Hz, which is the same as 1 clock cycle per 1.2 seconds. The internal optical particle sensor of the Particulate Matter Sensor SPS30 (dimensions: 40.6 × 40.6 × 12.2 mm^3; detectable particle range: 1 to 1000 ug/m^3; lower limit of detection: 0.3 um) from SENSIRION Co., Ltd. (UK), is connected to the air pollution sensing IoT system for mobile sampling campaigns. Appendix II Table I depicts the full specifications of the Particulate Matter Sensor SPS30 and Appendix II Table II depicts the full specifications of the humidity and temperature sensors. The Particulate Matter Sensor SPS30, a low-cost (unit price: $29.73) optical particle sensor, employs laser scattering. Each of the mobile sensing nodes can collect concentrations of particulate matter in local storage, in conjunction with storing the data offline. An onboard 20-channel SIRF-III-based GPS model provides location and time information at a sampling rate of 1 Hz. This IoT sensing platform for mobile sampling campaigns can collect continuous measurements (intervals as low as per 1.2 s) and map them in GIS based on their GPS coordinates. The sensors mentioned in this thesis are set up in a laboratory before being deployed so that that they will reliably yield stable readings. In view of how varied every sensor is, the mobile sensing equipment of every probe vehicle is replaced each week with a different sensing box located in the vicinity of the adjacent fixed sites (it should be noted that the engineering implementation work is supported by the third institute and this research designs the technique route of calibration experiment). To reconcile the different sensing environments for fixed and mobile sensors, the mobile AQ measurements are further corrected against fixed station readings (following the procedure in section 7.2) and this type of correction is performed for every hour.

3.3.1.2 Online field calibration platform for mobile sampling campaigns

To adjust for potential variability of sensors, the calibration steps for low-cost sensors include two processes: laboratory calibration and field calibration (it should be noted that this research designs the technique route of online field calibration experiment and the implementations are supported by the cooperated institute). Laboratory calibration refers to the calibration of sensors, which aims to adjust them to the standard measurement level. Field calibration is to assess and calibrate the low-cost sensor readings to fixed observations and eliminate the possibility of problematic observations in the field (Bi et al., 2020). Mobile sampling boxes (sensors are embedded and processed with laboratory calibration beforehand) are equipped at the same place together with the fixed monitoring sites, and over ten identical mobile sampling boxes are placed at each reference site. Air pollutant, ambient temperature, and humidity measurements are sampled from both fixed monitoring sites and low-cost mobile sensors and are transmitted into the Online Field Calibration Learning Platform (OFCLP) in conjunction with geographic data (DEM, location) by strict collocation with nearest neighbour strategy; thus, the individual calibration coefficients from each monitor are dynamically considered and updated. The calibration learning process runs online and adopts the MLR (Bi et al., 2020) to fit the functional parameters between fixed and mobile observations, involving the ambient temperature, humidity, DEM, and location data as independent variables. Meanwhile, the learning parameters are real-time updated into the edge-computing unit in each mobile sampling box mounted on the vehicles running on the street segments across the study area. Figure 3.4 shows the field calibration for mobile sampling campaigns in the Beijing case study.

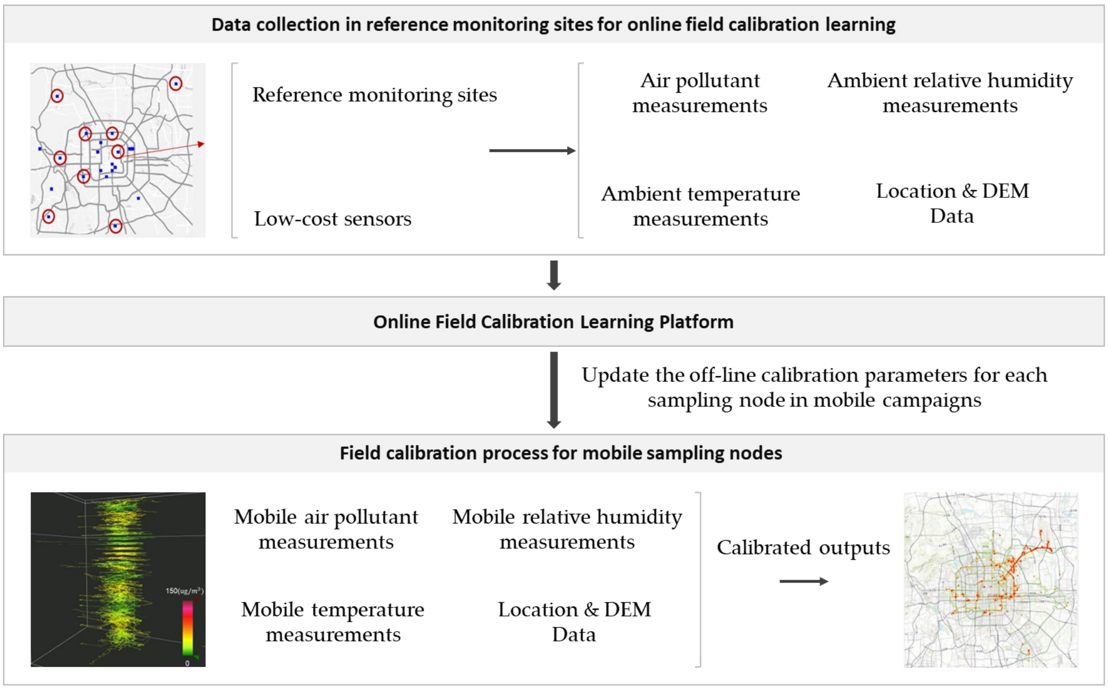


Figure 3.4: Diagram of the field calibration for mobile sampling campaigns

Note: Considering the potential variability of intra-instruments, the mobile sampling box mounted on each probe vehicle is periodically (weekly) replaced by another sampling box placed around the adjacent fixed AQ monitoring sites recently.

3.3.1.3 Spatial extent of mobile sensing

Mobile sensing provides densely spatial extent—in the Beijing case study, just 15 random taxis cover one-third of street segments each day. Figure 3.5 shows the daily spatial coverage of fixed and mobile sensing, which demonstrates that most sub-areas in the study area are dynamically covered and shows that mobile sampling campaigns are capable of making up for the shortcomings (spatially sparse) of fixed AQ monitoring sites. In this research, the sampling routes are chosen following a mixed strategy of pre-determined routing (aiming to maximize spatio-temporal coverage) and opportunistic routing (e.g., informed by real-time traffic conditions or following drivers' personal preferences). However, there still exist unsampled sub-areas every day, which may cause overfitting or underfitting at unsampled grid-cells because of the asymmetrically distributed AQ samples. Clearly, the routing strategy could be further improved through more rigorous planning and optimisation, but this is beyond the research scope here and will be investigated in future work.

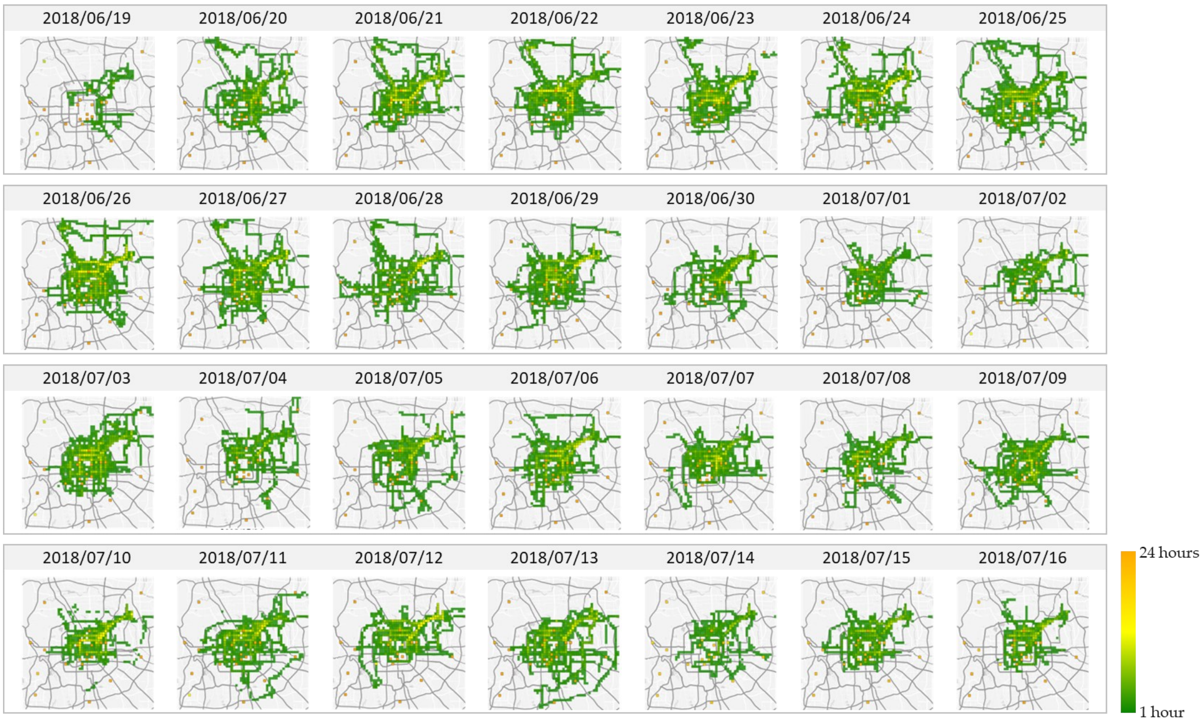


Figure 3.5: Daily spatial coverage by fixed and mobile sensing in Beijing case study between 2018/06/19-2018/7/16.

To get a sense of the spatio-temporal coverage scale, Figure 3.6 (a) shows the amount of labelled grid-cell units per hour during the observing period across the study area in Beijing. The number of hourly spatial sensing covered grid-cells ranges from 50 to 150, of which the density is < 5% of the overall region studied. Moreover, Figure 3.6 (b) shows the labeled times (both fixed and mobile sensing coverage) of spatial grids in the study area, Figure 3.6 (c) shows the labeled times (only mobile sensing coverage) of spatial grids in the study area of Beijing, and Figure 3.6 (d) illustrates the labeled grid-cell units from 8:00-10:00 on 2018/06/24. The apparent spatio-temporal dynamics and sparsity of the air pollution sensing pose a significant challenge to the AQ inference problem. The research proposes to tackle this problem by exploring the spatio-temporal correlations of different grid-cell components combined with their urban features. This will be discussed in forthcoming sections.

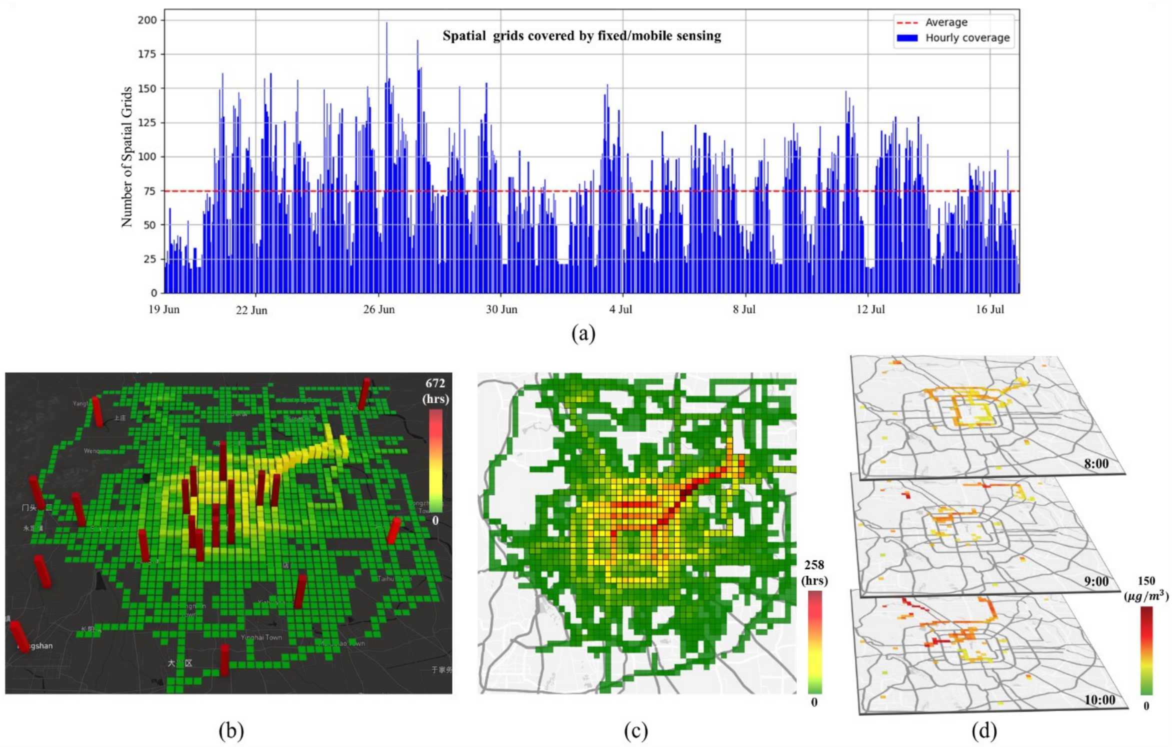


Figure 3.6: Spatio-temporal coverage of fixed and mobile sensing in Beijing. (a): Number of grids covered by fixed/mobile sensing in Beijing on an hourly basis. (b): Number of hours with fixed and mobile coverage in Beijing; (c): Number of hours with mobile coverage in Beijing. (d): Time-variant fixed/mobile sensing coverage in Beijing between 8:00-10:00, 2018/06/24.

Table 3.2 displays the fixed and mobile air pollution sensing and multi-source urban feature datasets used in the Deep AQ framework, including geographic data, land use data, transport data, public vitality data, population data, and meteorology data.

Table 3.2: Fixed and mobile air pollution sensing and multi-source urban feature datasets used in Beijing case study.

<table>
<tr><th>Data Category</th><th colspan="2">Data Type</th><th>Resolution</th></tr>
<tr><td rowspan="5">Geographic & Land Use</td><td colspan="2">Point of Interest (21 categories)</td><td rowspan="5">1 km × 1 km
Static (2018)</td></tr>
<tr><td rowspan="3">Land use categories</td><td>Area of Interest (19 categories)</td></tr>
<tr><td>Area of Green Cover (%)</td></tr>
<tr><td>Area of Water Cover (%)</td></tr>
<tr><td colspan="2">Digital Elevation (m)</td></tr>
<tr><td rowspan="3">Transport Data</td><td rowspan="2">Network</td><td>No. of signalized intersections</td><td rowspan="2">1 km × 1 km
Static (2018)</td></tr>
<tr><td>Total lengths of different typed roads</td></tr>
<tr><td>Traffic Conditions</td><td>Percentages of roads with light/medium/heavy conditions; categorized traffic conditions</td><td>1 km × 1 km
Hourly (2018)</td></tr>
<tr><td>Public Vitality</td><td>Sina Weibo</td><td>Number of users, posts, and comments</td><td>1 km × 1 km
Annual average (2015-18)</td></tr>
<tr><td>Population</td><td>LBS-R</td><td>Dynamic density of Wechat active users</td><td>1 km × 1 km
Hourly (2018)</td></tr>
<tr><td rowspan="6">Meteorology</td><td colspan="2">Temperature (°C)</td><td rowspan="6">1 km × 1 km
Hourly (2018)</td></tr>
<tr><td colspan="2">Pressure (kPa)</td></tr>
<tr><td colspan="2">Water vapor pressure (hPa)</td></tr>
<tr><td colspan="2">Relative humidity (%)</td></tr>
<tr><td colspan="2">Wind direction ([0°,360°])</td></tr>
<tr><td colspan="2">Wind speed (m/s)</td></tr>
<tr><td rowspan="2"></td><td rowspan="2">Fixed monitoring sites</td><td>Particulate matter concentration (μg/m^3)</td><td rowspan="2">Point-based
Hourly (2018)</td></tr>
<tr><td>Location ([lat, lng])</td></tr>
</table>

Air pollutant measurements		Time, location	
	Low-cost mobile sensors	Temperature (°C)	Trajectory-based 2s - 15s (2018)
		Relative humidity (%)	
		Particulate matter concentration (μg/m^3)	
* The fixed and mobile data are supported and provided by the Cloud-Guizhou Data Science Research Institute; the aggregated LBS-R (WeChat) data is provided by Tencent Co., Ltd; the urban feature datasets are collected from local government and open sources.			

3.4 Description of urban features and model component definition

Geographical phenomena vary with changes in measurement scale, and different scales have different representative information values (Ge et al.,2019). All the concentrations of air pollutant species and urban features are defined and depicted as spatio-temporal units shown in the following paragraph.

Definition 3.1. (Grid-unit) Let $\mathcal{S}$ be the set of 1 km × 1 km spatial grid-cell collections in the study area, and $\mathcal{T}$ be the set of 1-hour periods. According to the spatio-temporal resolution requirement, the spatio-temporal domain can be discretized into $M_x \times M_y$ (M_x refers to the number of grid-cells on the x-axis; M_y refers to the number of cells on the y-axis) congruent grid-cells $(x_i \times y_j) \in S$ and time slices $t_k \in \mathcal{T}$, which corresponds to longitude, latitude, and time index, respectively. The grid-cells and time slices are combined to construct the smallest spatio-temporal system unit element $u = (x_i, y_j, t_k) \in S \times \mathcal{T}$. In Beijing and Chengdu case study, the study areas are separately divided into 3,025 and 4,900 1 km × 1 km spatial grids, as shown in Figure 3.7. The research in Beijing has that $|S| = 55 \times 55 = 3,025$, $|\mathcal{T}| = 24$ (hrs/day) × 28 (days) = 672 (hrs) and in Chengdu has that $|S| = 70 \times 70 = 4,900$, $|\mathcal{T}| = 24$ (hrs/day) × 48 (days) = 1,152 (hrs).

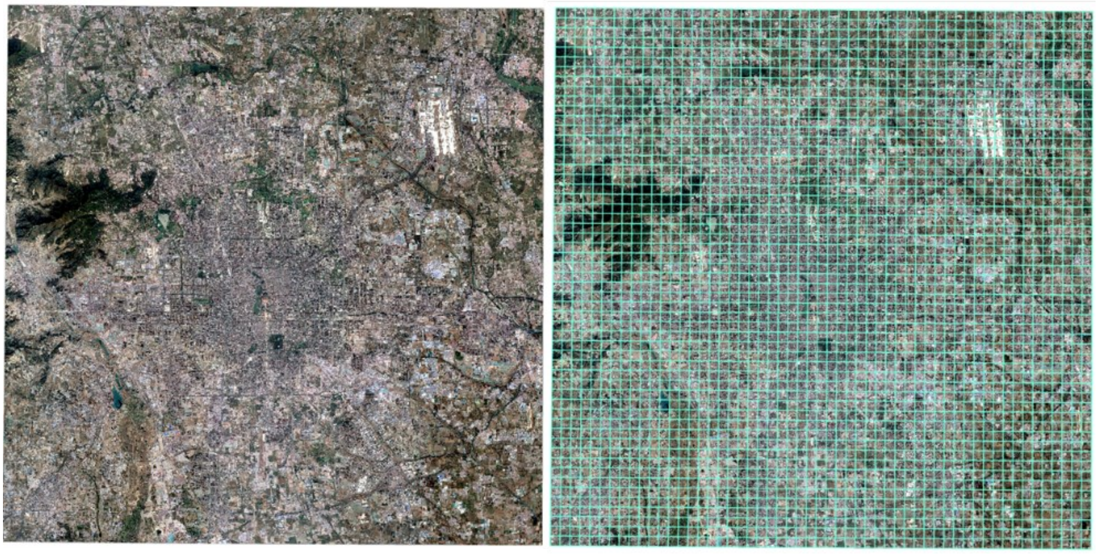


Figure 3.7: Diagram of the geo-meshing to generate pixelwise grid-cells

Definition 3.2. (Feature) Given the $M_x \times M_y$ spatial grids, a static feature is a matrix $S_i \in R^{M_x \times M_y}$ characterising certain aspects within the individual grids (such as traffic infrastructure or POI). One dynamic feature is a tensor $D_i \in R^{M_x \times M_y \times |\mathcal{T}|}$ which describes the location-specific dynamics of a certain quantity (such as relative humidity and traffic congestion). Moreover, these features are processed in advance, before being combined and collected in spatial grids if static or in units if dynamic. Figure 3.8 depicts various examples.

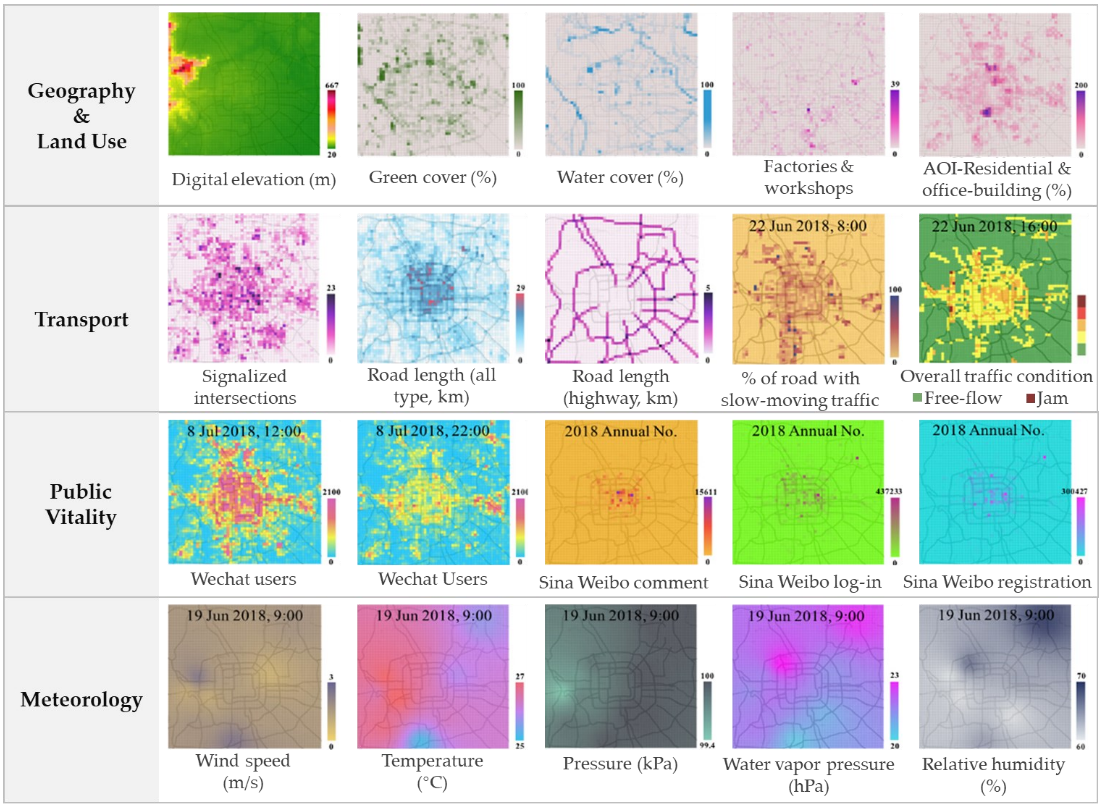


Figure 3.8: Selected (non-exhaustive) urban features in Beijing case study, which are used for model training.

Definition 3.3. (AQ label responses) The label response (Cristianini, 2014) refers to the representative value of the air pollutant measurements which serve as dependent variable by mapping with appropriate units based on their timestamps and coordinates. Note that during one hour, there might be many different measurements from mobile sensors in a cell, but these would only generate one hourly representative value through data preprocessing and statistical aggregation (detailed in Section 5.2.1). Given the subset $\mathrm{u_L} \subset \mathrm{S} \times \mathrm{T}$ containing labeled units, where this research uses $\mathrm{g}^{\mathrm{u}}_{\mathrm{AQ}}$ to denote the label responses in grid-unit $u \in u_L$. The fixed AQ sensing labelled grid-unit can be described as $\mathrm{g}_{\mathrm{AQ}}^{\mathrm{u_{fixed}}}$ and the mobile sensing labelled grid-unit can be described as $\mathrm{g}_{\mathrm{AQ}}^{\mathrm{u_{mobile}}}$.

Definition 3.4. (Air pollutant inference) The objective of AQ inference is to estimate for all grid-cell units based on training the spatio-temporal correlations between pollutant measurements in the labelled grid-units and multi-source urban features.

Geography and land use - this includes how the land cover is used (for example, whether it is a commercial area or factory), in addition to points of interest (POIs). In this research, 21 categories of POIs are considered in Table 3.3 and the numbers of them are used, including restaurants, shopping, schools, hotels, firms, and scenic spots. Estimations are made regarding the number of various sorts of POIs and what percentage of the land they take up (Areas of Interest, AOIs) within a 1 × 1 km area. Additionally, the geographical characteristics of the area are further added to by incorporating the levels of green and water coverage and digital elevation. The POI and AOI collection of the i, j th grid-cell are denoted as $POI^{i,j}$ and $AOI^{i,j}$ respectively, which consist of the number and area of interests with each category in the i, j th grid-cell.

Table 3.3: POI and AOI Categories

NO.	Categories
1	Factory
2	Building
3	Industrial firm
4	Commercial area
5	Transport infrastructure, e.g., charging stations, bus stop
6	Educational institute
7	Financial institute
8	Hotel
9	Beauty spa
10	Scenic spot
11	Restaurant, bar
12	Automobile service
13	Living service
14	Cultural attraction
15	Leisure
16	Hospital
17	Gym
18	Government agency
19	Natural scenery
20	Green Cover
21	Water Cover

Traffic - including static road network structure and dynamic traffic conditions. The details of the road network inside the 1×1 km area includes how many signalised junctions there are and how long the primary, secondary, tertiary, and quaternary roads are. Such factors impact on how much pollution is emitted by traffic, in that static features affect traffic volume and composition, in conjunction with the frequency of stop-and-go driving cycles. The proportions of varied traffic conditions are shown after such data has been gathered from each cell unit based on real-time map navigating services. This information directly indicates traffic levels and as such gives insight into vehicle emissions.

Note: The traffic proxy variable can be represented with the form of one vector with Equation (3.1):

$$\mathrm{V}_{\partial_{\mathrm{trans}}} = f\left(\mathrm{P}_{\mathrm{Trans}_{\mathrm{light}}}, \mathrm{P}_{\mathrm{Trans}_{\mathrm{medium}}}, \mathrm{P}_{\mathrm{Trans}_{\mathrm{heavy}}}\right) \tag{3.1}$$

where $\mathrm{V}_{\partial_{\mathrm{trans}}}$ defines a traffic proxy variable to implicitly represent the potential vehicle emission at the grid-cell unit and $\mathrm{P}_{\mathrm{Trans}_{\mathrm{light}}}, \mathrm{P}_{\mathrm{Trans}_{\mathrm{medium}}}, \mathrm{P}_{\mathrm{Trans}_{\mathrm{heavy}}}$ respectively refers to the road percentages with different traffic statuses (expedite, congested, and blocked) in a grid cell-unit. Here, this research adopts the GBDT approach to construct the correlations among the traffic parameters and the air pollutant concentration of grid-unit to generate the traffic index. Besides the road percentages with different traffic statuses, the ambient temperature, humidity, and position are used as independent predictive variables (described in Table 3.4) to model the traffic index; see more details about sensitivity analysis of these variables in Appendix I.

Table 3.4: Predictors used to model the traffic index.

Predictors	**Description**	**Variable type**
$\mathrm{V}_{\mathrm{pos}}$	Grid-cell position	Discrete (1-3025)
$\mathrm{V}_{\mathrm{temp}}$	Ambient temperature	Continuous
V_{rh}	Relative humidity	Continuous

$P_{Trans_{light}}$	Percentage of the road with expedite traffic status	Continuous
$P_{Trans_{medium}}$	Percentage of the road with congested traffic status	Continuous
$P_{Trans_{heavy}}$	Percentage of the road with blocked traffic status	Continuous

Population and public vitality - this term relates to the level of active social media activities by the population, as facilitated by Location Based Service Request (LBS-R). It should be noted that the data is statistically aggregated after systematic desensitisation which does not involve personal or private data. In the Chengdu case study, only the public vitality data are used to calculate the annual average (avg: 2015 - 2018). In the Beijing case study, the active user data from WeChat (hourly aggregated: 2018/06/19 0:00 - 2018/07/16 23:00) and the statistical public vitality data (avg: 2015 - 2018) are both used. The public vitality data contains statistics on logins, posts, and comments from Sina Weibo, which can be regarded as indirect indicators of the way land is used (such as business districts and rural areas). Active user data from WeChat give instantaneous population density information and is aggregated hourly. The WeChat density maps can show the number of active people, in contrast to actual population density, considering that people in some groups who do not use mobile phones as readily (such as elderly citizens, children and people who are economically disadvantaged) and are therefore partially ignored in this data compared with the community as a whole. However, the WeChat data can reflect the population dynamic patterns and the maps can be utilised to redistribute demographic data combining with the census-based data in the study area according to the methods in Equation (3.2) - (3.4). This research assumes that the inner-urban population mobility does not dramatically affect the total population, which denotes that the population movement across the study area is not considered in this research.

$$POP = \frac{\sum_1^T \sum_1^N pop_i^t}{\mathcal{T}} \tag{3.2}$$

$$w_i^t = \frac{pop_i^t}{POP} \tag{3.3}$$

$$\mathrm{TP}_i^t = \mathrm{TP} \times w_i^t \tag{3.4}$$

where pop_i^t refers to the aggregated number of active WeChat users at the i th grid-cell unit in the t th hour of the $\mathcal{T}$ period, N reflects the total amount of the grid-cells in the area studied, POP reflects the number of average total actual users, w_i^t refers to the distribution weight of the i th grid-cell unit, TP refers to the total number of census-based population and the TP_i^t refers to the redistributed population at the i th grid-cell unit in the t th hour of the $\mathcal{T}$ period.

Meteorology - this encompasses information gathered every sixty minutes on local temperature, humidity, pressure, and wind speed and direction. This information is collected from the ground meteorological station measurements. In the Beijing study, the hourly meteorological measurements from 18 ground-stations were obtained and spatially interpolated into grid-based quantities by IDW method.

Note: The wind can be represented with the form of two vectors reflecting in both x-axis and y-axis with Equation (3.5) and Equation (3.6), and Figure 3.9 shows this vectorising operation:

$$\mathrm{V}_{\mathrm{w}_x} = \mathrm{V}_\mathrm{w} \cos \varphi \tag{3.5}$$

$$\mathrm{V}_{\mathrm{w}_y} = \mathrm{V}_\mathrm{w} \sin \varphi \tag{3.6}$$

where V_w denotes the wind speed and the φ denotes to the angle of wind direction in each grid-cell unit. The weather conditions (including sunny, rainy, cloudy, and more) are transformed into a one-hot encoding vector.

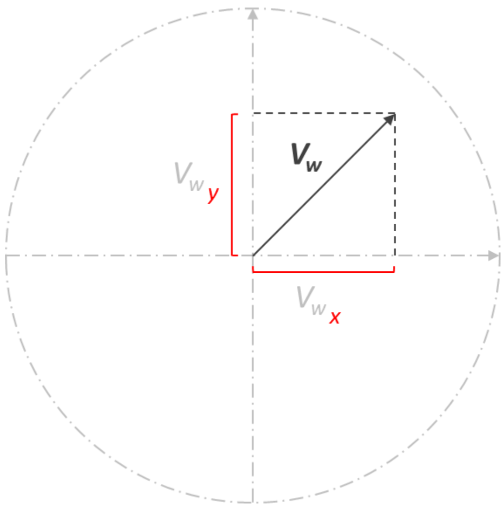


Figure 3.9: Diagram of vectorizing wind speed and wind direction into the horizontal and vertical representation

3.5 Summary

This chapter addresses the second research objective of this thesis, which is to:

Collect multi-source AQ samples and urban features for data-driven analytics.

This chapter describes the Chengdu and Beijing case studies and introduces the urban datasets used in the two case studies in detail. Section 3.1 introduces the AQ sampling mechanisms based on fixed monitoring sites and urban feature datasets in Chengdu case study. Section 3.2 introduces the fixed and mobile air pollution sensing and urban feature datasets used in Beijing case study. Section 3.2.1 introduces the work scope in Beijing, including the description of the IoT platform and online field calibration platform for mobile sampling, and also illustrating the power of mobile air pollution sensing by demonstrating the daily spatial coverage during the study period. Section 3.3 describes the definitions of the model components, including the concept of grid-cell unit, label responses, AQ inference, etc., and also explains the configurations of multi-source urban features.

Chapter 4

Review of standard ML algorithms

4.1 Introduction

This chapter reviews the standard ML algorithms used in the Deep AQ framework. In this thesis, the proposed Deep AQ framework refers to an end-to-end learning framework which is capable of generating pixelwise AQ inference by learning dynamic correlations among multi-source AQ samples and urban features. This research develops the Deep AQ framework based on LUR frameworks combining with ML-based methods employing convolutional techniques to construct the spatio-temporal correlations from meso-view in the Deep AQ framework. Tree-based and network-based surrogate models are selected and achieved in Deep AQ framework. This chapter describes the standard ML algorithms to develop the Deep AQ framework.

4.2 Model development from LUR

LUR is a general model that simulates the spatial difference of air pollutant concentration at the urban scale and is suitable for air pollution information research applications with high spatiotemporal resolution for atmospheric health research, although it requires a lot of data preparation, but can be effectively applied to the simulation of various air pollutant levels due to its intuitive mathematical analysis and strong operability (Hennig et al., 2016). The LUR models provide a series of promising results in the existing studies (Table 2.2) and have five major components: the dependent variables, a monitoring network, land use maps, buffer zones, and other independent variables. Variable selection aims to obtain spatial predictors that best explain variability for AQ modelling within the most parsimonious model (Morley &

Gulliver, 2018). Figure 4.1 shows the flow of the LUR model for AQ modelling, and the calculating process is as follows:

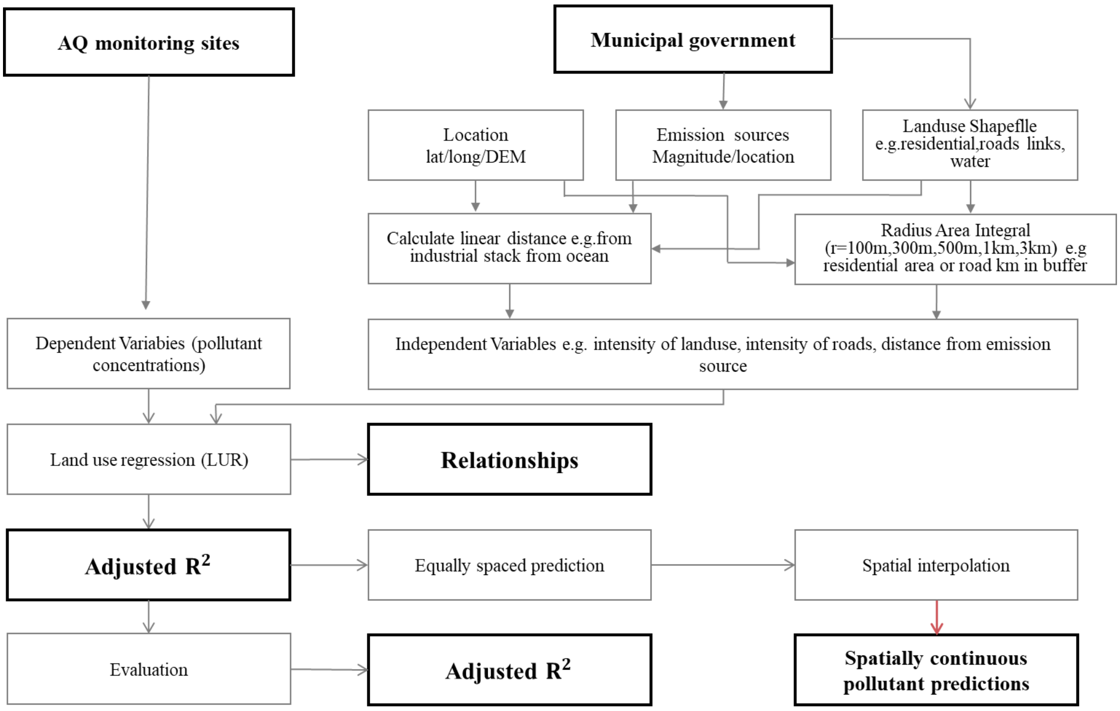


Figure 4.1: Flow of LUR model

1) Field measurements of air pollutant concentration at a set number of locations are collected in the study area, which constitutes the dependent variable.
2) For independent variables, the GIS predictor variables (typical predictors usually involve position, land use information, etc.) are also generated at these points to create the training dataset.
3) Multiple linear regression (MLR) is used to model the functional relationships between the dependent variable and independent variables.
4) Fitted functional relationships are used to achieve equally spaced predictions.
5) Spatial interpolation is used to achieve spatially continuous AQ mappings.
6) R-square is used to evaluate the model performance.

MLR is an analysis method that uses a mathematical model to describe the approximate relationship between the predictor variable and the predicted variable based on the multiple sets of observation value, the mathematical model expression is called the regression equation. If it is assumed that there is a linear relationship between the predictor variable and the predicted variable, and the multiple linear regression can be expressed by the following Equation (4.1):

$$y_i = \beta_0 + \sum_{k=1}^{m} \beta_k \, x_{ik} + \varepsilon_i \tag{4.1}$$

Here y_i is the i-th observed dependent variable, indicating the concentration of air pollutants to be predicted. x_{ik} is the i-th observed independent variable, indicating factors in the regression model that have an impact on the concentration of pollutants. β_k is the weight of the influence coefficient of each influence factor on the target variable y, and each β_k needs to be determined from the observation samples. ε_i is an independent random error term that normal distribution with mean value equals to zero, which represents the possible impact of other random influencing factors that are not involved in the model.

However, LUR models are usually built to predict temporally-stable long-term AQ estimations (Hennig et al., 2016). LUR models in the aforementioned studies are generally intuitive and interpretable (Wang et al., 2020). However, simple statistical nature of LUR models limits their performance when modelling complex AQ conditions (Wang et al., 2020). The Deep AQ framework develops from the LUR framework combining with ML-based approaches to continuously build up correlations between multi-source features and AQ samples to achieve pixelwise AQ inference in sizeable urban areas.

4.3 Convolution techniques in Deep AQ framework

Convolution is a mathematical operation achieved with two functions (f and g) that produces a third function to express how the shape of one is modified by the other. The convolution operator contains the elementwise components and the computing function which can be used for feature extraction. As the depth of the convolution increases, deep high-level features will be extracted, and finally the high-level features will be used for tasks such as classification in the fully connected layer. In the application of neural networks, the input multi-dimensional array is usually called tensor, and the kernel is the parameter of the multi-dimensional array optimized by the learning algorithm. For example, if a two-dimensional array $M(m, n)$ is used as input, the convolution kernel K used is also two-dimensional, and the convolution operation formula is as the following Equation (4.2):

$$S(i,j) = (M * K)(i,j) = \sum_{m}\sum_{n} M(m,n)K(i-m, i-n) \tag{4.2}$$

As aforementioned in Chapter 3, the geo-meshing is adopted to extract urban features and transform them into vector for grid-cell units. In this research, the convolution filter (coined as ST-filter) is used to construct the spatio-temporal correlations in the neighbouring receptive field for each grid-cell unit.

4.3.1 Discrete convolutions

Discrete convolution preserves the notion of ordering which is a linear transformation of neighbouring element values. As there are only a few input grid-units that contribute to the given output grid-unit, it can be defined as 'sparse'. Also, for multiple locations of input, identical weights are applied, so it is reusing parameters. To produce an output, the result can only be passed through a nonlinearity after a biased vector is added. Thus, firstly, a vector is used as the input; then, the vector is multiplied with a matrix. Equation (4.3) provides the definition formula for discrete convolution:

$$(f * g)(n) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(n-\tau) \tag{4.3}$$

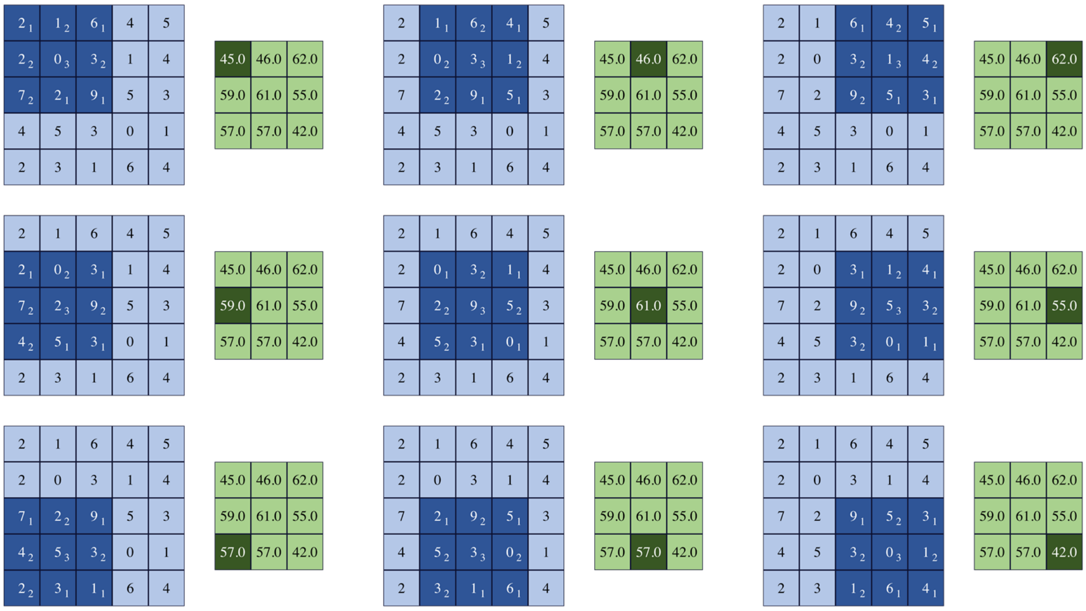


Figure 4.2: Diagram of producing output using discrete convolution

Figure 4.2 provides a diagram of the output production by using a discrete convolution operator. In the case of the grid-cells in light blue colours, they are the ones that contribute to the input feature maps for the given grid-cell that is within the shaded areas in the center. It is indeed rather common that several feature maps may be stacked together. Across all input feature maps, a kernel (the shaded area) of value can be seen shifting around. The duplicated elements of the kernel are calculated at each location, and the sum of the results is used to identify the output in its current location. For example, as shown in Figure 4.2(a), the computing process for the output at this point can be expanded as follows:

$$2 \times 1 + 1 \times 2 + 6 \times 1 + 2 \times 2 + 0 \times 3 + 3 \times 2 + 7 \times 2 + 2 \times 1 + 9 \times 1 = 45$$

Different kernels are applied in the above procedure in a loop in order to produce as many output feature-maps as needed (Figure 4.2). Every single the feature map is firstly equally convolved with a specific kernel. Subsequently, a sum of the generated feature maps will then be computed to generate the output feature map.

From Figure 4.2, it is apparent that the convolution operation involves an instance utilization of a kernel of 3 × 3 size for 2-D convolution. There also appears a possibility to expand it to N-D convolutions. In the case of a 3-D convolution, for instance, across the height, width and depth of the input feature map, there will see the kernel as a cuboid shifting across. Based on the collection of kernels, there is a discrete convolution which is shaped according to the permutation of $(n, m, k_1,..., k_N)$ in which: n means the quantity of output feature maps, m is the quantity of input feature maps, k_j means size of the kernel along axis j. Regarding the output size o_j of a convolutional operator along axis j, pproperties listed below (all of which are along axis j) will have an impact on it: i_j means input size along axis j, k_j means kernel size along axis j, s_j means the stride (i.e. the distance between two consecutive kernel positions) along axis j, and p_j means the zero padding (i.e. number of '0' concatenated) along axis j.

4.3.2 Pooling operations

Pooling refers to the function which uses the overall statistics of the adjacent output in a specific location to substitute the output of the network at the exact location. For example, 'max pooling' can be used to maximise the value in the adjacent area. Commonly used pooling functions include maximum pooling, average pooling, L2 norm, and centre distance weighted pooling. The main functions of the pooling layer are as follows:

(1) Pooling has translation invariance which means that regardless of the pooling function, the input representation can be unchanged when a small amount of translation is made in the input, preserving the position of the feature.

(2) Pooling integrates feedback from all fields, increases the receptive field, and improves the computational efficiency of the network. It also reduces the storage requirements for parameters.

An example of average pooling can be seen in Figure 4.3(a), whereas an example of max pooling is demonstrated in Figure 4.3(b).

Regarding the pooling window along axis j, the properties listed below (all of which are along axis j) will affect its output size o_j: i_j means the input size along axis j, k_j means the size of pooling window along axis j, *and* s_j means the stride (i.e. the distance between the pooling window's two consecutive positions) along axis j.

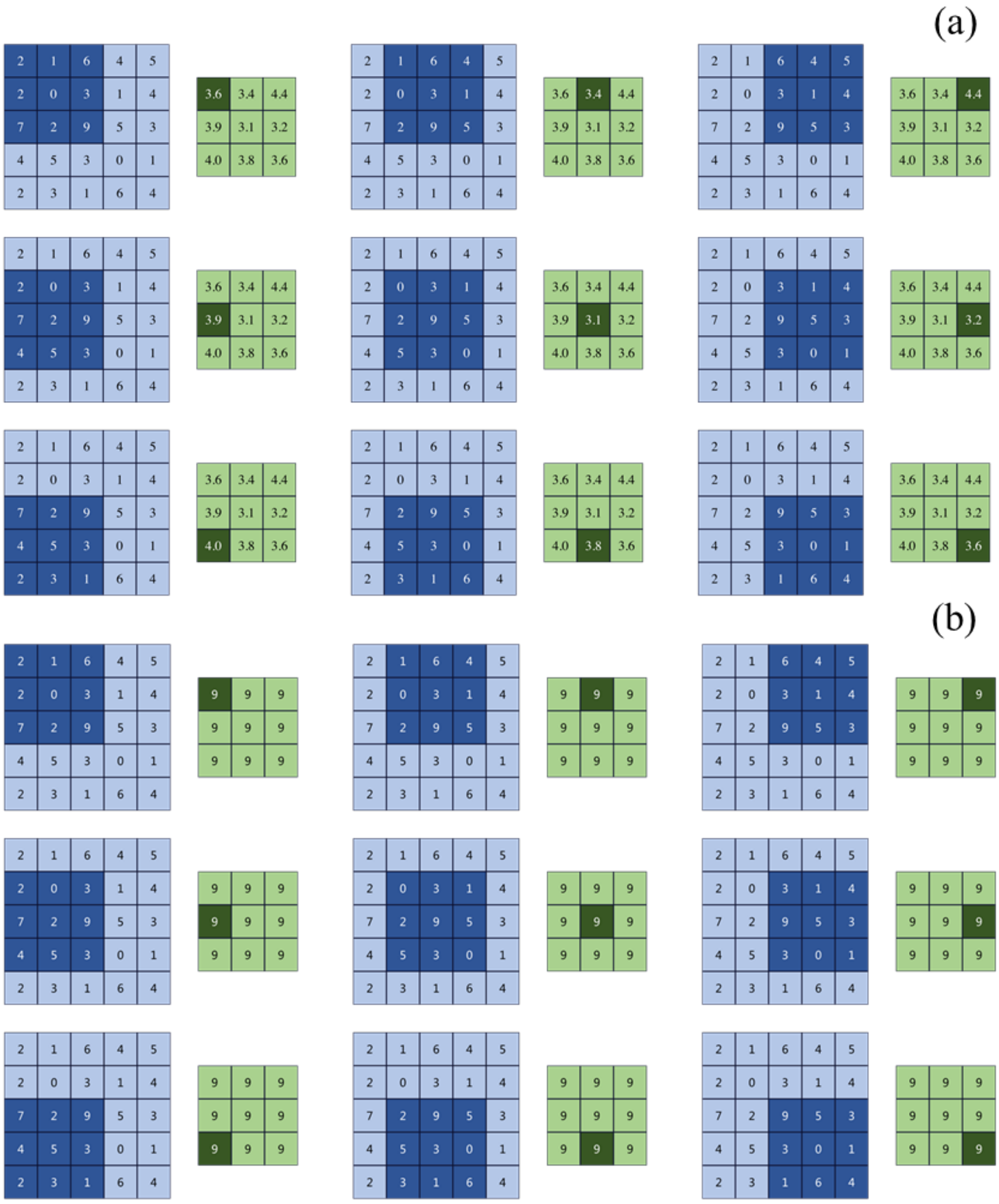


Figure 4.3 (a) shows a 1 × 1 stride with a 5 × 5 input where the output values of a 3 × 3 average pooling operation is calculated. Figure 4.3 (b) shows a 1 × 1 stride with a 5 × 5 input where the output values of a 3 × 3 max pooling is calculated.

4.3.3 Padding

This research aims to construct the neighbourhood correlations for each grid-unit, which means that the input size and the convolved output size should be identical. For those grid-units on the study area's border (Figure 4.4), the dotted area is the blank position which has no truth value but should be filled with the padding value. This research employs half padding to fix the blank position, that promises the same output size as the input size. Figure 4.4 is an example of half padding and unit strides.

It can be considered a desirable property when the input size is identical to the output size (i.e., $o_j = i_j$):

Relationship: For any i_j and for k_j odd $\left(k_j = 2n + 1, n \in \mathrm{N}\right)$, $s_j = 1$ and $p_j = \left\lfloor \frac{k_j}{2} \right\rfloor = n$, $o_j = i_j + 2\left\lfloor \frac{k_j}{2} \right\rfloor$ - (k - 1) $= i_j + 2n - 2n = i_j$.

The above relationship is often called the 'half padding' or the 'same padding'. A demonstration of $i_j = 5$, $k_j = 3$ and (therefore) $p_j = 1$ can be seen in Figure 4.4.

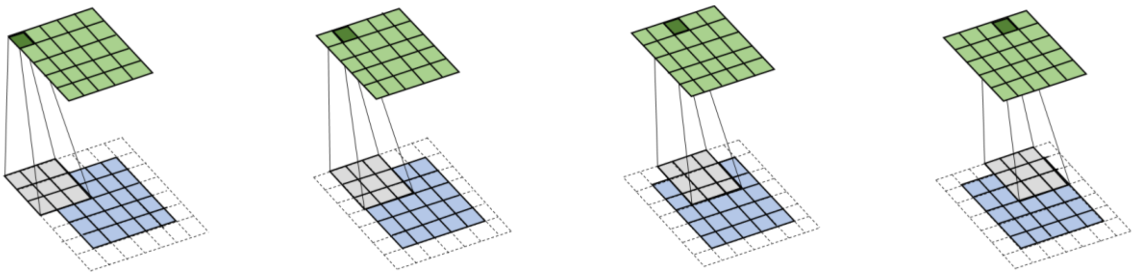


Figure 4.4: An example of half padding and unit strides (i.e., i = 5, k = 3, s = 1 and p = 1) on a 5 × 5 input with a convolved a 3 × 3 kernel.

4.4 Machine learning approaches in Deep AQ framework

4.4.1 Optimisation problem in Deep AQ framework

This research proposes a spatio-temporal mask (ST-Mask) to label the grid-unit and aggregates representative values from pollutant measurements as label responses which are dynamically covered by fixed or mobile sensing, where V_{input} (feature collections) for each labelled grid-unit is identical to the model input features. This research holds an assumption that the model input variables consist of ρ-dimensional input features ($X_i = (x_i^1, x_i^2, \ldots, x_i^\rho)$) and one-dimensional output label responses. The sample collection for model training can be defined as $Z = \{(X_1,y_1), (X_2,y_2),\ldots,(X_N,y_N)\}$(in which N refers to training samples). For the given dataset $\{X_i,y_i\}_1^N$, the training objective is to construct an appropriate fitting function $f(x)$ to generate and describe the correlations between input feature variables $\{X_i\}_1^N$ and label response $\{y_i\}_1^N$ to estimate the results for unsampled grid-units.

The loss (L) process for estimating y is expressed in Equation (4.4) as follows:

$$L(f) = \sum_{i=1}^{N} L(y_i, f(X_i)) \tag{4.4}$$

4.4.2 Tree-based algorithm

Decision trees divide samples into smaller and more homogenous subsets based on the feature variables. The recursive division of datasets means the subsets on lower levels have very few instances. In some datasets, the instances at a specific leaf node are as low as 2-5 instances (Mitchell,1997).

A decision tree partitions the instance space recursively. The node with no incoming edges is called root node (Maimon & Rokach, 2008) and all other nodes have exactly one incoming edge. A node with outgoing edges is coined to as an 'internal' or 'test' node. All terminal nodes are called 'leaves' (also known as decision nodes). Each internal node in a decision tree partitions the instance space into two or more sub-regions and classify the instances into corresponding sub-regions according to the attribute values of input feature variables. In the simplest and most frequent case, each partition test considers a single attribute where the instance space is divided according to the attribute value. Each leaf is assigned to a range representing the most appropriate target value. Alternatively, the leaf may hold a probability vector (affinity vector) indicating the probability of having a specific value based on the target attribute. Two or more branches may grow from each interior node (i.e., not a leaf). Each node corresponds with a specific attribute, and the branches are partitioned according to the value range of particular attribute. Instances are divided and classified by navigating them from the tree's root to the terminal leaves, according to the partition tests along the path. Specifically, the process is concluded as the following steps:

1) Starting with the root of a tree;
2) Considering the features corresponding to the root;
3) Defining the features corresponding to which branches or sub-regions;
4) Considering the node in which the given branch appears;

5) Repeating same operations for this node, etc., until reaching a leaf;

Note that a decision tree is capable of dealing with both categorical and numeric attributes. Each node is labelled with the attribute it tests, and its branches are classified with the corresponding instances.

The goal is to minimize $L(y, f(x))$ for $f(x)$, where here $f(x)$ refers to a sum of trees. And the minimizing procedure is a numerical optimisation as written in Equation (4.5):

$$\hat{f} = arg\min_{f} L(y, f(x)) \tag{4.5}$$

where the parameters $f \in R^N$ represent the approximating function $f(x_i)$ at each data point x_i in N training samples:

$$f = T(f(x_1), f(x_2), \ldots, f(x_N)) \tag{4.6}$$

Algorithm 4.1: Decision Tree

Input: Training sample collection Z^* of size N in the instance space
Attributes set $X_i = (x_i^1, x_i^2, \ldots, x_i^\rho)$
Begin do:
Generate a node;
if all samples in Z^* belong to the same category C **then**
Mark the node as a leaf-like node of C **return**
end if
if $X_i = \emptyset$ **or** samples in Z^* have the same value in X_i **then**
Mark the node as a leaf node, and mark its category as the category with most samples in Z^*; **return**
end if
Choose the optimal attribute x_i^ρ in X_i to partition;
for each attribute x_i^ρ in X_i **do**
Generate a branch for the node; make $\widetilde{\mathbb{Z}}^{*m}$ denotes the bootstrap from the $\mathbb{Z}^*$ that Z^* takes the value x_i^ρ in X_i;
if $Z^* = \emptyset$ **then**
Mark the branch node as a leaf node, and mark its category with the most samples in Z^*; **return**
else
Take $(Z^*, X_i \backslash \{x_i^\rho\})$ as the branch node
end if

end for

Output: A decision tree with the node as the root node.

4.4.2.1 The principle of the tree-based algorithm

Growing regression trees involves generating a binary tree recursively by choosing the best appropriate feature variables and split points. The tree structure is able to be adaptive according to the attributes of the dataset. Therefore, it can deal with categorical and numerical variables simultaneously. However, when the structure of a decision tree (DT) is too complicated, the problems of overfitting or being trapped into local minimum points may appear (Maimon & Rokach, 2008).

Partitioning the samples in feature space into multiple sub-regions is usually the first step of a DT-based algorithm. The next step is to fit the corresponding simple model in each sub-region. Figure 4.5(a) shows an example of a 2-D feature space partition. In this figure, V_1,V_2 refer to the 2-D feature spaces, and $\{s_1, s_2, s_3, s_4, s_5\}$ refer to the splitting node. Figure 4.5 (b) demonstrates the splitting process. Firstly, all the samples split at $\widetilde{V}_1 = s_1$, then the samples in subregion $\widetilde{V}_1 \leqslant s_1$ split at $\widetilde{V}_2 = s_2$, and the samples in subregion $\widetilde{V}_1 > s_1$ split at $\widetilde{V}_3 = s_3$. Finally, the samples in subregion $\widetilde{V}_1 > s_3$ split at $\widetilde{V}_2 = s_4$. The splitting process partitions all the samples into five sub-regions $\{R_1, R_2, R_3, R_4, R_5\}$, as shown in Figure 4.5 (b).

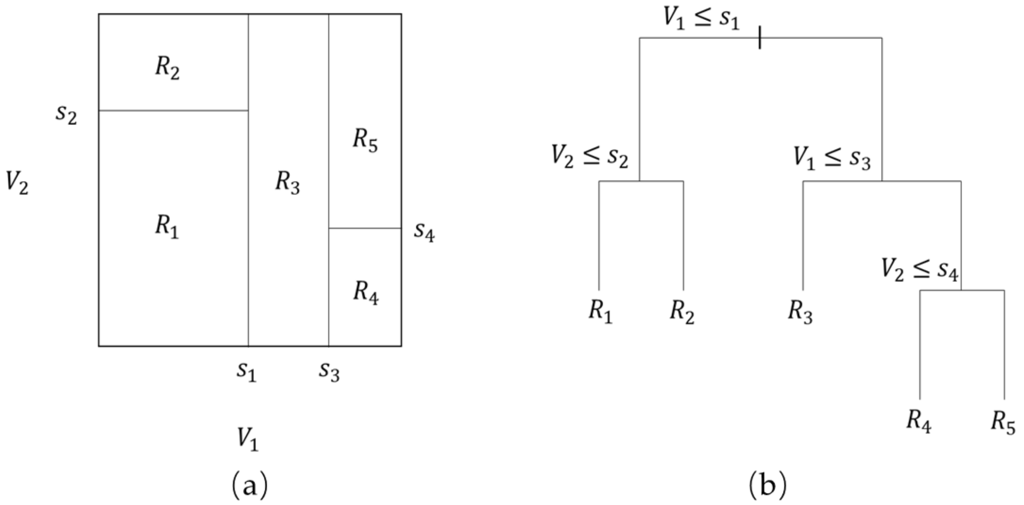


Figure 4.5: Diagram of the principle of a DT. (a) An example of the recursive partition for instances from the 2-D feature space. (b) A simple tree indicating the splitting

process of (a) which derives the partitioning rules. The following rules is derived from the tree shown in Figure 4.5 (b):

$$\mathrm{R}_1(\mathrm{V}) = \mathrm{I}(\mathrm{V}_1 \leqslant \mathrm{s}_1) \cdot \mathrm{I}(\mathrm{V}_2 \leqslant \mathrm{s}_2)$$

$$\mathrm{R}_2(\mathrm{V}) = \mathrm{I}(\mathrm{V}_1 \leqslant \mathrm{s}_1) \cdot \mathrm{I}(\mathrm{V}_2 > \mathrm{s}_2)$$

$$\mathrm{R}_3(\mathrm{V}) = \mathrm{I}(\mathrm{V}_1 > \mathrm{s}_1) \cdot \mathrm{I}(\mathrm{V}_1 \leqslant \mathrm{s}_3)$$

$$\mathrm{R}_4(\mathrm{V}) = \mathrm{I}(\mathrm{V}_1 > \mathrm{s}_1) \cdot \mathrm{I}(\mathrm{V}_1 > \mathrm{s}_3) \cdot \mathrm{I}(\mathrm{V}_2 \leqslant \mathrm{s}_4)$$

$$\mathrm{R}_5(\mathrm{V}) = \mathrm{I}(\mathrm{V}_1 > \mathrm{s}_1) \cdot \mathrm{I}(\mathrm{V}_1 > \mathrm{s}_3) \cdot \mathrm{I}(\mathrm{V}_2 > \mathrm{s}_4)$$

The tree-based algorithm is capable of flexibly choosing splitting feature variables and nodes. Here, it is assumed that all the data samples can be partitioned into the M disjoint regions recursively, which are expressed as R_1, R_2,..., R_M, and the functional rule can be modelled by $\mathfrak{R}_\mathrm{m}$, written as Equation (4.7):

$$X_m \in R_m \Rightarrow f_m(X_m) = \mathfrak{R}_m \tag{4.7}$$

where X_m refers to the instance space in the m region.

Thus, a DT is described using Equation (4.8) as follows:

$$\mathcal{T}_m(x) \ = \ \mathfrak{R}_m I(X_m \in \ R_m) \tag{4.8}$$

For classifying tress, the accuracy and Gini impurity (which measures the diversity in a set) are used to select the best feature to split the data. For the regression trees, the loss errors (e.g., absolute errors and squared errors) were used to select the best feature to split the data. The best binary partition in terms of the minimum sum of squares is impossible to identify computationally. As Maimon and Rokach (2008) observed, it is very commonly a greedy algorithm is adopted. By defining a dividing feature variable v at the split node s, the pair of half-planes is expressed as Equation (4.9):

$$R_-(v,s) \ = \ \{X \,|X_v \leq \ s\} \text{ and } R_+(v,s) \ = \ \{X \,|X_v > s\} \tag{4.9}$$

Then the dividing feature variable v and spilt node s are continuously sought to meet the requirements in Equation (4.10):

$$\min_{v,s}\left[\min_{R_-}\sum_{x_i\in R_-(v,s)}(y_i-R_-)^2+\min_{R_+}\sum_{x_i\in R_+(v,s)}(y_i-R_+)^2\right] \tag{4.10}$$

The feasible best pair (v, s) can be determined by scanning all the instance samples to search for the best split node s for each dividing feature variable v. For each best split, the data samples are partitioned into two sub-regions, and the splitting is conducted repeatedly on all resulting sub-regions.

4.4.2.2 Bagging and boosting

Bagging and boosting are techniques for ensemble, where a strong leaner will be created by combining a group of weak learners so as to outperform any single learner (Maimon & Rokach, 2008). By generating extra data in the training stage, the techniques of Bagging and boosting will get N learners. Then, by replacement from the original set and the use of random re-sampling, new training data sets can be generated. Due to the re-sampling process with replacement, there may be duplicated observations within the new training data sets. All elements have equal possibility to appear in a new data set in bagging. In contrast, as boosting actually will weigh the observations, some observations may therefore appear more frequently in the new data sets than the others. Figure 4.6 shows how additional data is generated in the training stage of bagging and boosting.

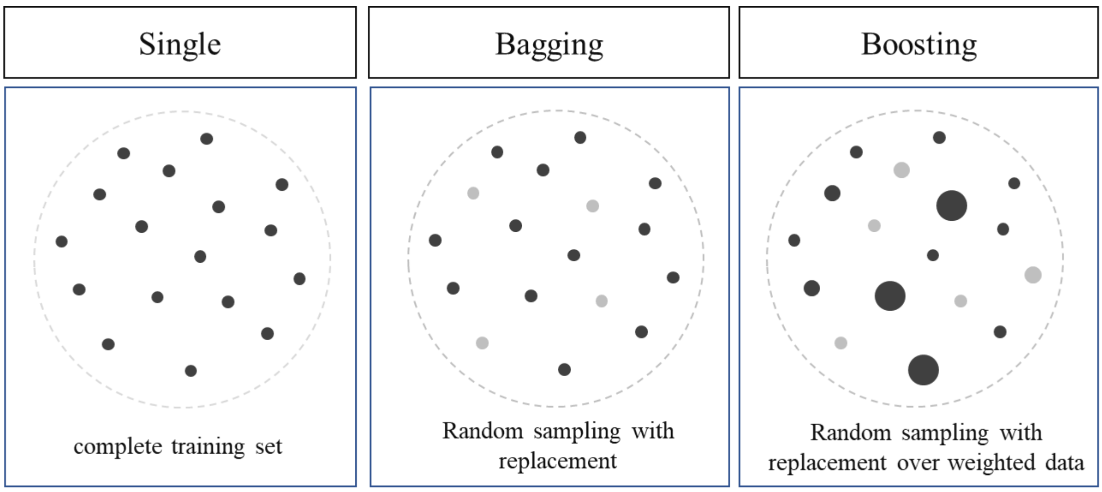


Figure 4.6: Generation of additional data in the training stage of bagging and boosting

While the training stage in Bagging is parallelly carrying on (i.e., building each model in a separate manner), new learner in boosting is created sequentially (Maimon & Rokach, 2008). In boosting algorithms ensure every learner is trained on data, while each learner's success will be taken into account by the subsequent learner. When all the training steps are done, the redistribution of sample weights will take place, then the weights of misclassified data will be increased to emphasize. Following this sense, these data will be emphasised during the training of the subsequent learners. Figure 4.7 shows the divergence of bagging and boosting.

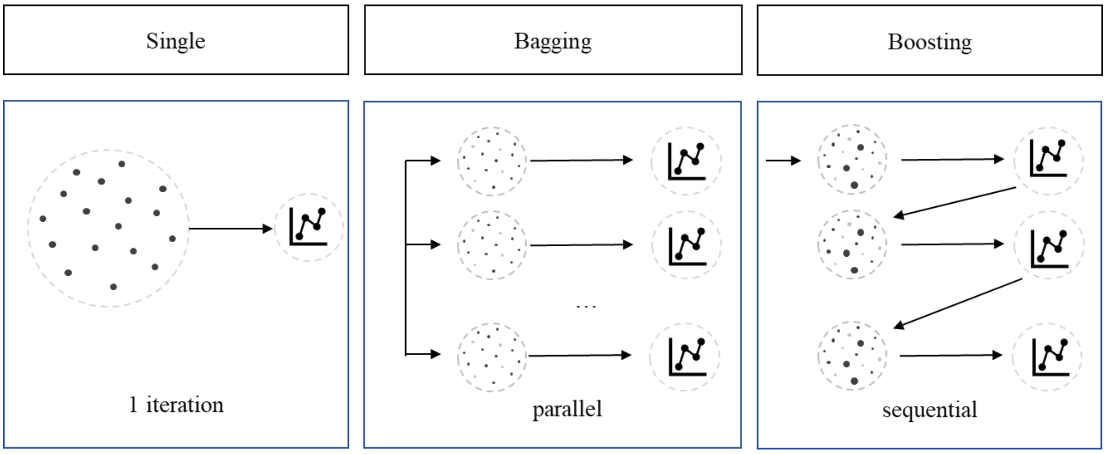


Figure 4.7: Divergence of bagging and boosting

Through using the average results of the N learners (or majority vote), the result in Bagging can be obtained. In comparison, in Boosting, a second set of weights will be assigned in order to take the weighted average for their estimates for the N learners. Each learner is allocated its weight by the algorithm in the Boosting training stage. The more satisfactory and reasonable the learner's estimated result on the training data, the higher weight it will be assigned. Thus, errors of each learner will be recorded in order to inform the training of a new learner. Figure 4.8 shows the training procedure of bagging and boosting.

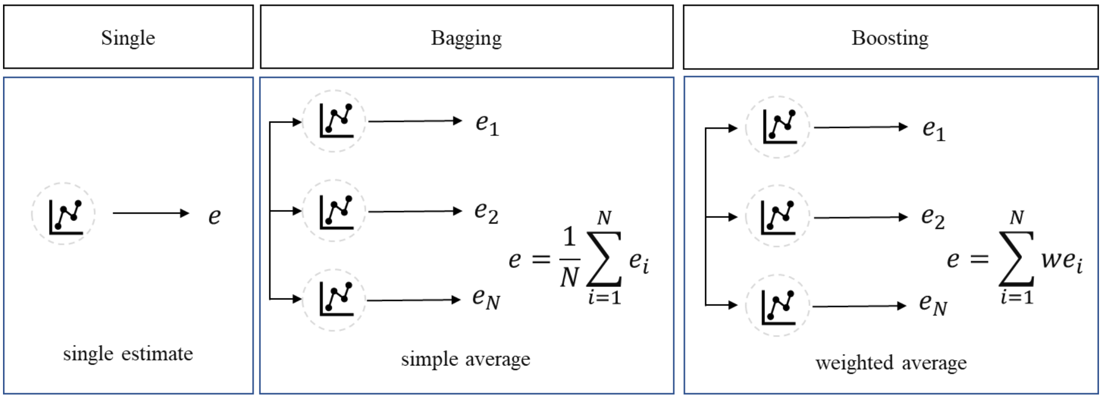


Figure 4.8: Training procedure of bagging and boosting

4.4.2.3 Bagging trees

Bagging is powerful in reducing high variance which is especially suitable for tree-structured approaches considering their ability to construct complex correlations in data learning (Rokach & Maimon, 2008). Bagging trees can bootstrap versions of training data set by fitting the same DT as many times as required, while averagely compute the estimation results. Considering the notoriously noisy characteristics of trees, the tree-based approaches can significantly benefit from the bagging. However, all trees distribute identically when they grow in bagging, and the expectation of the M such trees' average result is identical to that of every one of the trees. This implies that the bias of bagging trees equals to that of individual trees, thus only improves through variance reduction.

Through using the subsets of samples, Bagging can compute the average estimation values to eventually reduce their variance (Rokach & Maimon, 2008). For each sample subset $\tilde{Z}^{*m}, m = 1, 2, \ldots, M$, the model produces the estimation as written in Equation (4.11):

$$\hat{f}_{Bag}(X) = \frac{1}{M}\sum_{m=1}^{M} \mathcal{T}_m(X) \tag{4.7}$$

Equation (4.11) derives the correct bagging estimate in terms of a Monte Carlo estimate, which approaches it as $M \rightarrow \infty$.

Only under the situation when the latter builds up with nonlinear functions will the bagged estimate will be different from the initial estimate. There are usually different features in each bootstrap tree compared to the original tree; the terminal nodes may also be different in number. The average estimation based on the M trees is the bagged estimate.

It is worthwhile to note that there will be a loss on any simple structure in the model when bagging a model. In this case, a bagged tree can no longer be regarded as a tree.

Random forests refer to a substantial modification of bagging (Rokach & Maimon, 2008), which builds many de-corrected DTs and generates a final estimate by averaging them. Equation (4.12) shows the average variance Var when (i) the feature variables are distributed in an identical manner, and (ii) the feature variables are not independent from positive pairwise correlation ρ:

$$Var = \rho\sigma^2 + \frac{1-\rho}{M}\sigma^2 \tag{4.12}$$

As M increases, the second term gradually disappears, but the first term still remains. As demonstrated in Algorithm 4.2, by lowering the correlations among the DTs, the function of random forest can eliminate the variance in bagging. Moreover, by using random selection of the input features, it can be achieved in growing trees.

Algorithm 4.2: Random Forest
Input: Training sample collection Z^* of size N in instance space **Begin do:** For $m = 1$ to M: a) Sub-select a bootstrap $\widetilde{\mathbb{Z}}^{*m}$ from the $\mathbb{Z}^*$ b) Grow a tree $\mathcal{T}_m(x)$ on the sub-selected dataset through repeatedly rerunning the three steps as listed before on each terminal node until reaching the minimum node size of n_{min}. I. From the p feature space, sub-selecting the m features randomly; II. Within the k, selecting the best feature/split-point; and III. Dividing the node into two sub-nodes. **Output:** Overall ensemble estimate $$\hat{f}^{M}_{RF}(X) = \frac{1}{M}\sum_{m=1}^{M}\mathcal{T}_m(X)$$

Especially: Before each split, when growing a tree on a sub-selected dataset and identifying the candidates for splitting, typical values for k (within the randomly selected $k \leqslant p$ of the input features) are $\sqrt{p}$.

4.4.2.4 Boosting trees

Unlike bagging, boosting is initially proposed as a committee for weak tree-based learners to evolve and cast weighted votes (Maimon & Rokach, 2008). The motivation of boosting, which perspective is similar to bagging, is to form a powerful committee through grouping together the outputs of multiple learners from the DT (ibid). There are two purposes of Boosting: firstly, it aims to constantly modify the versions of data through applying learner of DT in sequential order and secondly to generate a sequence of weak DT learners $\mathcal{T}_m(x), m = 1, 2, \ldots, \mathrm{M}$. Figure 4.9 demonstrates the diagram of the DT learners' training on the weighted versions of dataset modifications.

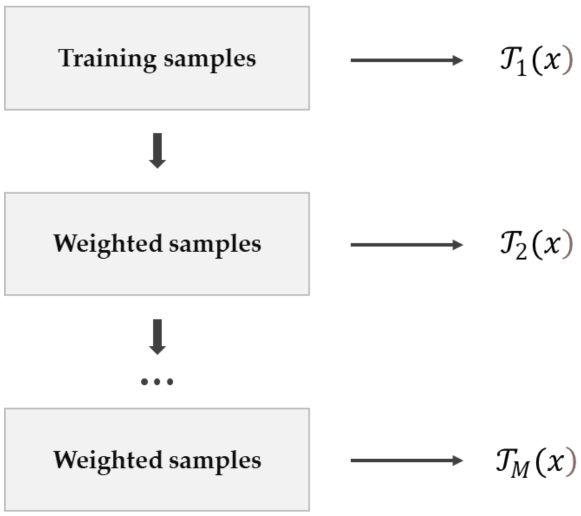


Figure 4. 9: Diagram of learners training on the weighted versions of dataset modifications.

The estimates from all weak tree-structured learners are combined through a weighted majority vote written in Equation (4.13) to produce the final estimation:

$$\hat{f}^{M}_{Boost}(X) = \sum_{m=1}^{M} \gamma_m \mathcal{T}_m(X) \tag{4.13}$$

Here, it calculates γ_1, $\gamma_2,\ldots,\gamma_M$, while the contribution of every weak tree-structured learner is evaluated and weighted by the boosting algorithm.

The modifications of the data throughout the Boosting involve the procedure of applying weights w_1, w_2, ..., w_N to every single sample of training. To begin with, to train the model on the data in a usual way, $w_i = 1/N$ will be set as the initial weight of all. The observation weights for each subsequent iteration (i.e. $m = 2,\ldots,M$) will be adjusted individually, while the algorithm will be repeatedly applied to the weighted samples.

Gradient Boosting Decision Tree (GBDT) is boosting trees with base learners consisted of various DTs. No DT in a GBDT is independent; the emphasis of any misclassified samples recorded from former learners will be increased by the newly added DT. The residual from previous DT will be, as shown in Figure 4.10, taken into consideration as an input for the subsequent DT learner. The subsequent DT leaner is expected to lower the residual and create a negative gradient direction in each iteration in order to reduce the loss. Finally, based on the sum of all DT learners' results, the output result can be obtained estimated (Liang et al., 2020).

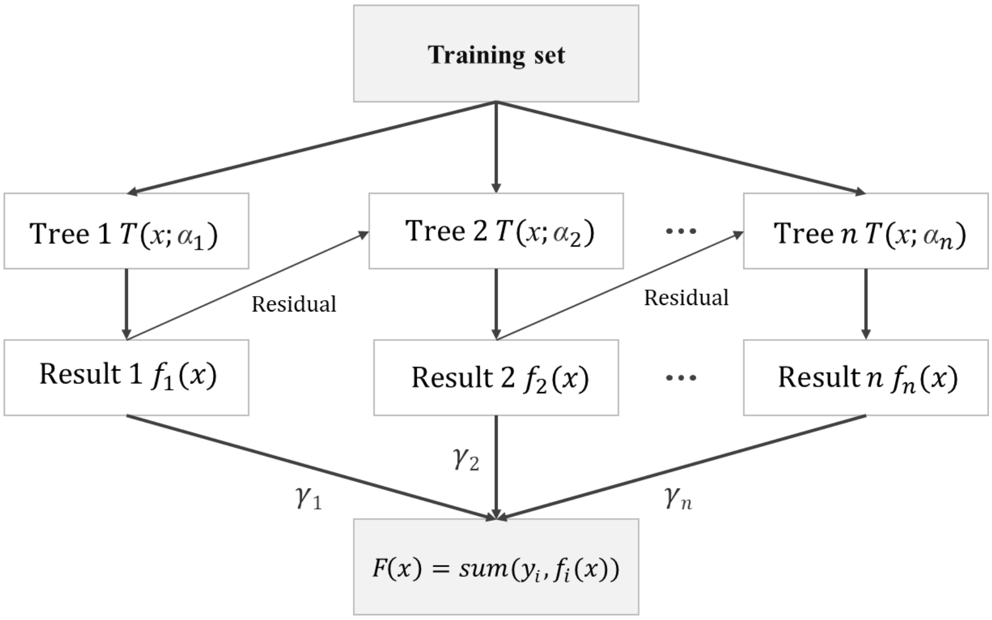


Figure 4.10: Diagram of GBDT algorithm

Let $\{X_i, y_i\}_{i=1}^{N}$ indicate the data samples, where $X_i = \left(x_i^1, x_i^2, \ldots, x_i^\rho\right)$ represent the feature variables, and y_i denotes the label response. GBDT can be calculated through the steps listed below:

Step 1: Obtaining the the initial constant value γ through Equation (4.14):

$$F_0(x) = arg\,\min_{\gamma} \sum_{i=1}^{N} L(y_i, \gamma) \quad (4.14)$$

where $L(y_i, \gamma)$ is the loss function.

Step 2: Equation (4.15) is for the gradient direction of the residual:

$$\hat{y}_i = -\left[\frac{\partial L\big(y_i, F(X_i)\big)}{\partial F(X_i)}\right]_{f(x)=f_{m-1}(x)} \quad (4.15)$$

where the number of iterations is indicated by m, and $m = 1,2, \dots, \mathrm{M}$.

Step 3: By fitting data samples, the initial model $T(X_i; \alpha_m)$ can be obtained, and based on the loss function (as shown in Equation (4.16), the parameter α_m can be calculated:

$$\alpha_m = arg\,\min_{\alpha,\beta} \sum_{i=1}^{m} L(\hat{y_i} - \beta T(X_i; \alpha)) \quad (4.16)$$

Step 4: The current model weight is represented in Equation (4.17) by minimizing the loss function:

$$\gamma_m = arg\,\min_{\gamma} \sum_{i=1}^{m} L\big(y_i, F_{m-1}(x) + \gamma T(X_i; \alpha_m)\big) \quad (4.17)$$

Step 5: The model is updated as Equation (4.18):

$$F_m(x) = F_{m-1}(X_i) + \gamma_m T(X_i; \alpha_m) \quad (4.18)$$

This loop is to be carried on until the specified times of iterations or the convergence conditions are satisfied. In Algorithm 4.3, the computing process of GBDT is illustrated.

Algorithm 4.3: GBDT Algorithm
Initialize $F_0(x) = arg\,\min_{\gamma} \sum_{i=1}^{N} L(y_i, \gamma)$ For $m = 1$ to M: a) For $i = 1,2,\dots,\mathrm{N}$ calculate $r_i^m = -\left[\frac{\partial L\big(y_i, F(X_i)\big)}{\partial F(X_i)}\right]_{f(x)=f_{m-1}(x)}$

b) By giving the terminal regions $\mathrm{R_m}, m = 1,2,\ldots,\mathrm{M}$, grow the regression tree to the targets r_i^m

c) For $m = 1,2,\ldots,\mathrm{M}$ calculate

$$\gamma_m = arg\,\min_{\gamma} \sum_{i=1}^{m} L\big(y_i, F_{m-1}(X_i) + \gamma T(X_i;\alpha_m)\big)(X_i \in R_m)$$

d) Update

$$F_m(x) = F_{m-1}(X_i) + \gamma_m T(X_i;\alpha_m)(X_i \in R_m)$$

Output: $\hat{F}(x) = F_m(x)$

Specifically, according to Chen et al., (2016), XGBoost is a scalable and parallel algorithm based on GBDT, which can produce state-of-the-art results by solving various ML challenges. It can compute more quickly and efficiently than directly using GBDT, meaning they are more suitable for implementing the real-time processing and analysis system.

4.4.2.5 Relative importance of feature

DT algorithms can sort the relative importance of features, and the relative importance of each feature attribute can be regarded as a kind of quantifying approach for measuring contributions of features and their influence on estimating label responses. Each feature variable's relative importance is a very effective metric for measuring the contribution of each input feature variable. It can be defined as a measure of a feature's effectiveness in fitting the functional relationships between features and label responses (Mitchell, 1997), impacting each feature's impact on the AQ label responses.

As Breiman et al. (1984) stated, for any single DT $\mathcal{T}$:

$$L_v^2(\mathcal{T}) = \sum\nolimits_{s=1}^{J-1} \hat{L}_s^2\, I(v(s) = v) \tag{4.19}$$

as the feature variable v 's measure of relevance. The whole relative importance $L_v^2(\mathcal{T})$ for variable v in the tree $\mathcal{T}$ can be computed and defined by the sum of the $J-1$ the tree's internal nodes. The threshold sample chosen from the input instance space is the one that can achieve the maximum estimated improvement $\hat{\mathcal{L}}_s^2$ in the loss of squared error risk over that of the constant fit over the entire region. Thus, the feature variable $\mathcal{v}$'s squared relative importance equals the total squared improvements within all internal nodes.

This relative importance measure can be easily summarised using the additive tree expansions. Due to the stabilizing effect of averaging (which is shown in Equation (4.20)), the overall relative importance of feature variable $\mathcal{v}$ can be computed by averaging over the whole additive trees with more reliable output comparing with the single tree.

$$\bar{\mathcal{L}}_{\mathcal{v}}^2 = \frac{1}{M}\sum_{m=1}^{M}\mathcal{L}_{v}^2(\mathcal{T}_m) \tag{4.20}$$

The computing measure of relative importance refers to the squared relevance meaning that the actual relevance lies in their own square roots. It seems appropriate to allocate and scale the relative importance values into [0, 100] based on the consideration that these measures are relative.

Note that some feature variable will be completely ignored in boosting, but will not be ignored in bagging. Thus, the possibility for any feature variable to be included in a bagging tree will be increased thanks to the process of selecting candidate split feature, comparing to boosting where there does not exist such a process.

4.4.3 Neural network-based algorithm

Inspired by image denoising methods, this research adopts a neural network-based surrogate model to recover the AQ maps.

4.4.3.1 Introduction to neural networks

It is a commonly known fact that neural networks are to imitate the human brain and its learning process: a human brain can process input to generate relevant output, and that is what a neutral network tries to achieve. The three actions of receiving input (information), processing the input, and generating output are named as input, hidden and output respectively within the three layers of a neutral network. A training process of a neutral network can be accomplished in two steps, as detailed below:

- Forward propagation

Images are taken in in the form of numbers into the input layer. In other words, the intensity of pixels of an image are represented numerically before passing on to the next layer: the hidden layer. In the hidden layer, a number of mathematical operations will be applied by the neurons on these given values. There are a number of randomly assigned parameter values to aid the performance of these mathematical operations. The results of these operations will then be sent to the output layer to generate the final prediction.

- Backward propagation

After an output has been obtained, the output will be compared with the actual value. The values of the parameters will be updated based on two factors: 1) the final output, and 2) the distance (i.e. the error) between the final output and the actual value. Last but not least is the forward propagation process which generates new outputs by using the updated parameter values repeatedly.

4.4.3.2 Convolutional Neural Network (CNN) architecture

Similar to a standard multilayer neural network, a convolutional neural network (CNN) consists of convolutional layers which has a subsequent, fully-connected layer. The CNN is designed in conjunction with the strengths of the input image's 2D structure. With local connections and tied weights together with the pooling operations process to produce translated invariant features, CNN can be achieved. Below shows the two parts of a convolutional neural network:

- **Convolution layers**: these layers are for extracting features from the input
- **Fully connected (dense) layers**: these layers utilise the data from convolution layer to generate output

Figure 4.11: Workflow of CNN.

In the CNN, the convolution layers constitute of a group of parallel feature maps. These feature maps are formed of a set of convolutional kernels (i.e. feature detectors) which are sliding over the pixelwise input and reflecting the element-wise dots in the feature maps. Every layer of data in CNN consists of an array which is 3-D with a size $of\ h \times w \times d$ in which spatial input dimensions are represented by h and w, while the feature dimension is represented by d. This sliding operation is handled by a set of kernel filters $K = (k_1, k_2, \dots, k_n)$ which slide over the $h \times w \times d$ dimensions in each layer. Pixelwise input is in the first layer, where the pixel size is $h \times w$, and the feature dimension is d (Long et al, 2015). Besides, stride padding is the critical measure in terms of the convolution layers' spatial size control. Feature maps are then generated while being stacked together, resulting in the convolution layers. Convolution, pooling, and activation functions are the three essential components of convolutional networks. All three of them perform on local input regions and rely on relative spatial positions considering translation invariance characteristics. The data vector at location (m, n) in a specific layer can be expressed as $V_{m,n}$, and $O_{m,n}$ for the following layer with transformation output of $V_{m,n}$, and Equation (4.21) can be used to describe the functional relationships:

$$O_{m,n} = \otimes \{ f_{k_1 s}\left(\{V_{m,n}\}\right), \dots, f_{k_n s}\left(\{V_{m,n}\}\right)\} \tag{4.21}$$

where $K = (k_1, k_2, \dots, k_n)$ refers to various kernel sizes, s refers to the stride, and f_{ks} represents the type of the layer, for instance, whether it is a matrix multiplication for convolution or average pooling, a spatial max for max pooling, or an elementwise nonlinearity for an activation function (e.g., ReLU), and so on for other types of layers. The concatenate operation (⊗) combines the convolved results produced by different sizes of receptive fields. Algorithm 4. 4 demonstrates the process of CNN algorithms.

Algorithm 4.5: CNN Algorithm
Input: Matrix $H_1 \times W_1 \times D_1$ **Parameter:** Number of filters K; Size of filter F; Stride S; Padding P **Begin do:** a) Preprocessing i. Mean removal, pull the center of the samples back to the origin of coordinator. ii. Normalized, reduce the interference caused by the difference in the value range of each dimension. b) Structure of CNN calculation i. Convolutional layer. Each convolution filter generates $F \times F \times D_1$ weight and

K bias.

ii. Linear rectifier layer. The ReLU function, as the neuron activation function, defines the non-linear output result of the neuron after linear transformation.

iii. Pooling layer. Feature down-sampling, extract salient features to ensure that information can be abstracted faster.

Fully connected layer. Map the distributed features to the labeled sample space.

Output: Feature matrix $H_2 \times W_2 \times D_2$, where $H_2 = \frac{H_1-F+2P}{S} + 1,\ W_2 = \frac{W_1-F+2P}{S} + 1,$ $D_2 = K$

Since the complex relationships cannot be captured by considering the linear transformation alone, there was an additional component (a.k.a. the activation function) in the architecture of the network to add nonlinearity to the data. The filter matrix, which is randomly initialized during the forward propagation process, is adopted as the operating parameters in the convolution layer. The gradient of the parameter has to be identified in order to update the filter matrix $-\partial\ell/\partial f$. Figure 4. 12 shows a diagram of the computation process for backward propagation in CNN:

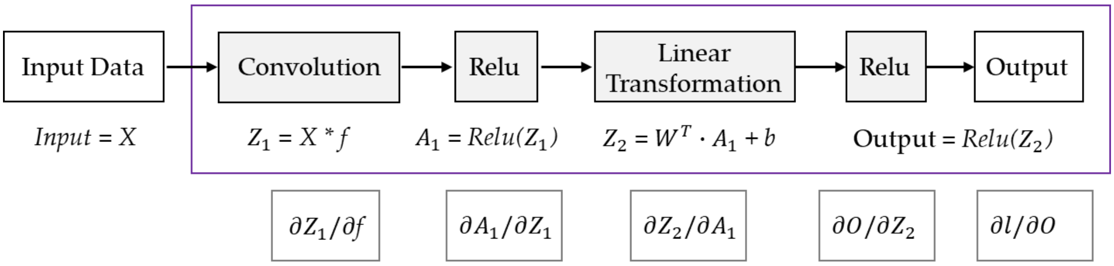


Figure 4.13: Diagram of computing process for backward propagation in CNN.

4.4.3.3 Fully Convolutional Network (FCN) architecture

Pixelwise AQ inference refers to the process of fitting relationships between feature variables and AQ label responses towards pixelwise grid-cell units. Furthermore, the translation of feature fitting upon features and label responses should slide and operate on each grid-unit, which means that traditional CNNs are not suitable for the requirement of this inference learning structure, but an end-end fully convolutional network (FCN) (Long et al, 2015) should be adaptable to this learning structure.

Assuming $\text{grid}(i,j)$ represents the location of the grid-cell. Writing $y_{i,j}$ for the ground truth value, $\hat{y}_{i,j}$ for the network estimation value at $\text{grid}(i,j)$, and $O_{i,j}$ for the value output from the previous layer. Let ϕ be the set of grid-units labelled by ST-Mask and grid$(i, j) \in \phi$. If the loss function $\ell(\phi)$ is a sum over the spatial dimensions of the final layer, its gradient will be the sum over the gradients of each of its spatial components in ϕ:

$$\ell(\phi) = \sum_{i,j} \ell'(y_{i,j}, \hat{\gamma}_{i,j}) \,, (i, j) \in \phi \tag{4.22}$$

During network training, there is a need to backpropagate the gradient of loss $\ell(\phi)$ through this transformation:

$$\frac{\partial \ell(\phi)}{\partial \hat{\gamma}_{i,j}} = \begin{cases} \frac{\ell(\phi)}{\partial_\phi} \cdot \frac{\partial \ell'(y_{i,j}, \hat{\gamma}_{i,j})}{\partial \hat{\gamma}_{i,j}} & if \;\; (i, j) \in \phi \\ 0 & \text{otherwise} \end{cases} \tag{4.23}$$

Then use the standard chain rule to backpropagate the gradients of the previous layer.

$$\frac{\partial \ell(\phi)}{\partial \mathcal{O}_{i,j}} = \frac{\partial \ell(\phi)}{\partial \hat{\gamma}_{i,j}} \cdot \frac{\partial \hat{\gamma}_{i,j}}{\partial \mathcal{O}_{i,j}} \tag{4.24}$$

Finally, update the trainable parameters of the network using the gradients.

Figure 4.11 shows the diagram of the end-end pixelwise inference using FCN. As shown in Figure 4.10, the input is a multi-dimensional array with N channels, and the FCN allows the use of operation filters with any shape. The FCN has multi-layers, and for the m_{th} layer, this research applies two filters with kernel sizes of $3 \times 3 \times N_m$ and $5 \times 5 \times N_m$. Their results are combined using the concatenate operation, and the channel dimension is doubled to $2N_m$ to generate the channels in the next hidden layer. Combining hidden layers with filters of various sizes facilitates the model's capability to predict from different receptive fields. Subsequently, the ReLU function (Eckle & Schmidt-Hieber, 2018) is used as the activation function. Finally, for an output layer of size $h * w$ with one channel, the component for predicting at every grid-unit is a $1 \times 1 \times 1$ small kernel to produce pixelwise inference results.

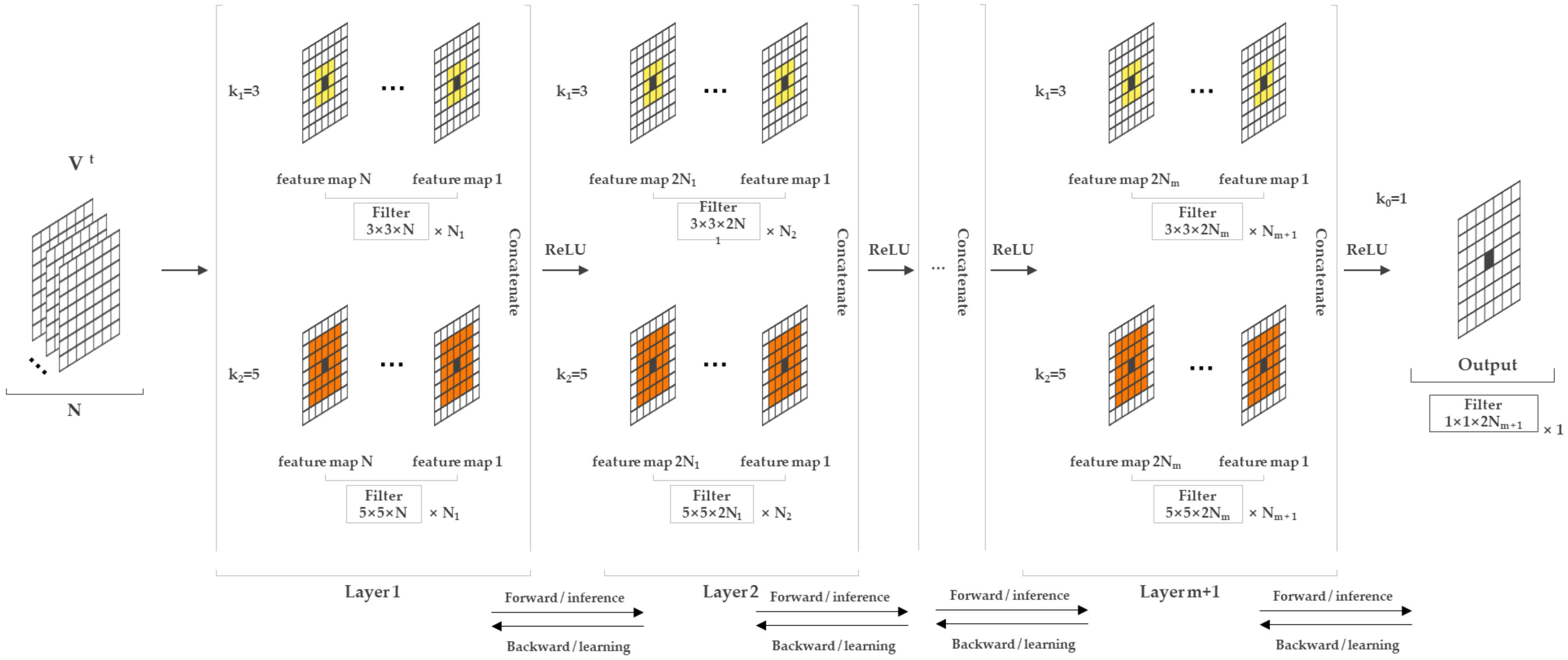


Figure 4.14: Diagram of end-end pixelwise inference by FCN.

4.4.4 Robust loss functions

The squared-error loss $L(y, f(x)) = (y - f(x))^2$ and absolute loss $L(y, f(x)) = |y - f(x)|$ are most used loss functions in many regression problems. During the training process, much more emphasis is placed on the observations in the training samples $|y_i - f(x_i)|$ due to squared-error loss. Thus, as Friedman (2001) concluded, it is inferior in terms of robustness and its performance in for long-tailed error distributions, especially for grossly outliers is unsatisfactory. In this thesis, the Huber loss criterion (Cavazza, 2016) is adopted to improve the robustness. The Huber loss function can be described using Equation (4.25) as follows:

$$L(y, f(x)) = \begin{cases} |y - f(x)|^2 & for |y - f(x)| \leq \delta \\ 2\delta|y - f(x)| - \delta^2 & otherwise \end{cases} \tag{4.25}$$

δ is an adaptive constant, when $\delta \rightarrow 0$, the Huber loss is more trended to absolute loss; when $\delta \rightarrow \infty$, the Huber loss is more trended to squared-error loss;

The Huber loss function (ibid), as described above, is strongly convex in a uniform neighbourhood of its minimum (i.e. $y - f(x) = 0$) at the boundary of which there is a differentiable extension to an affine function at points $y - f(x) = -\delta$ and $y - f(x) = \delta$ for the Huber loss function. Thanks to these properties, the combination of the sensitivity of the mean-unbiased, minimum-variance estimator of the mean (using the quadratic loss function) and the robustness of the median-unbiased estimator (using the absolute value function) is enabled. Figure 4.12 shows the standard deviation variations of the Huber loss function with different values of δ.

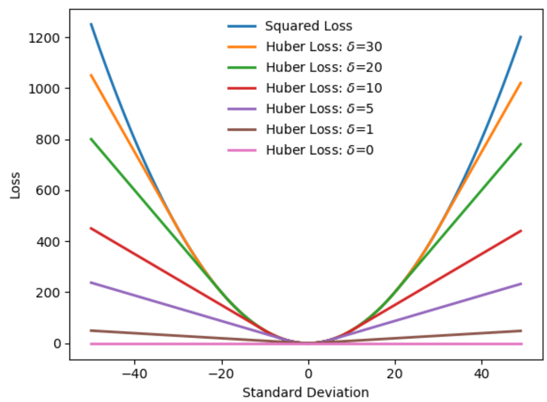


Figure 4.15: Standard deviation variations of Huber loss function with different values of δ.

4.5 Summary

This chapter reviews the standard ML algorithms for the third research objective of this thesis, which is to:

Propose a spatio-temporal learning and inference framework called Deep AQ framework to incorporate multi-source AQ samples and urban features to achieve high-resolution pixelwise AQ inference mappings.

This chapter mainly introduces the standard ML algorithms used in the Deep AQ framework. Section 4.1 reviews the standard ML algorithms for Deep AQ framework. Section 4.2 reviews the flow of LUR model. Section 4.3 introduces the concept of convolution approaches, including convolution, pooling, and paddling techniques, which are widely used in image recognition to extract and pad information in the neighbouring receptive field. Section 4.4 introduces the principle of decision tree-based and network-based ML-based models and efficient improvement by bagging or boosting approaches. Section 4.4.1 restates the optimisation regression problem. Section 4.4.2 introduces the tree-based approaches and the bagging or boosting trees, including RF and GBDT. Section 4.4.3 introduces the principle and diagram of end-end pixelwise inference by FCN. Section 4.4.4 presents the Huber loss function and compares standard deviation variations of Huber loss function with different values of δ

Chapter 5

Development of Deep AQ framework

5.1 Introduction

Chapter 4 reviews the standard ML-based methods for the development of Deep AQ framework. This chapter proposes a ML-based spatio-temporal learning framework (Deep AQ) for AQ inference based on multi-source urban features, including geographical, land use, traffic, population, and meteorological data. Data from monitoring sites from outside of the study area are employed to construct feature from a macro-view in addition to the said endogenous features. This enables the Deep AQ framework to also consider the regional spread of air pollution. The Deep AQ framework offers flexibility in accommodating sparse and irregular distribution of AQ measurements in spatio-temporal domains and is capable of integrating fixed and mobile AQ samples adaptively. There are a number of difficulties for AQ inference based on multi-source AQ samples that are either fixed or mobile and urban features. These difficulties include heterogeneous datasets that are substantially different in terms of spatio-temporal resolution, numerical scale, and veracity. Thus, transforming urban features with proper spatio-temporal data structures to integrate endogenous factors contributing to local air pollution (e.g., traffic, meteorology, land use data) but also external pollution influence (e.g., regional transport of air pollution) is essential.

5.2 Overall technical route

The Deep AQ framework for high-resolution AQ inference has been proposed, executed, and evaluated in this thesis, which is detailed in Figure 5.1.

The partial content in this chapter has been published in the IEEE Internet of Things Journal (J. Song, K. Han and M. Stettler, "Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing",** ***IEEE Internet of Things Journal**, DOI: 10.1109/JIOT.2020.3041047).**

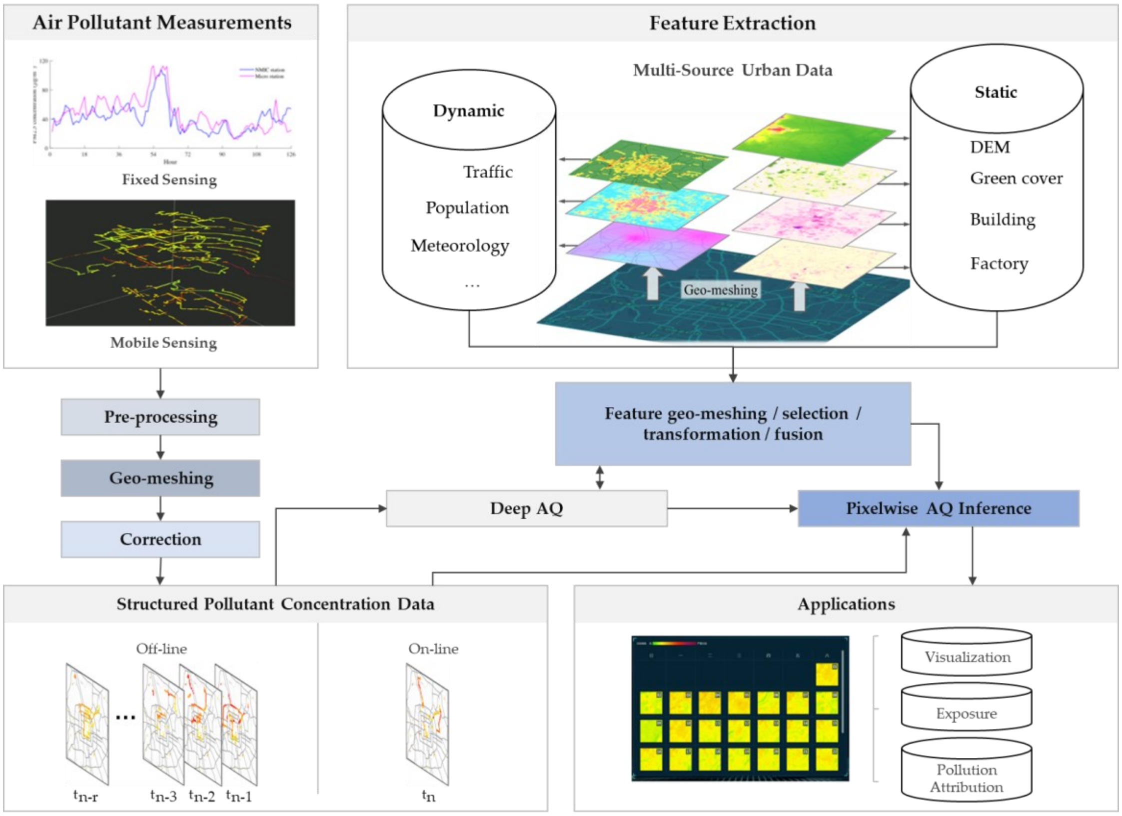


Figure 5.1: Methodological process with Deep AQ framework

The methodological process with Deep AQ framework mainly consists of five components: pre-processing of AQ measurements, geo-meshing for the representative value of grid-cell units to generate label responses, calibration/correction of mobile AQ measurements, geo-meshing/selection/transformation/fusion of feature maps, and pixelwise AQ inference.

5.2.1 Pre-processing of mobile trajectories and measurements

As the data quality of the mobile trajectories and measurements plays a crucial role in this research, a set of pre-processing steps for the collection of mobile sensing data are executed. Figure 5.2 shows the said pre-processing steps for mobile sampling campaigns.

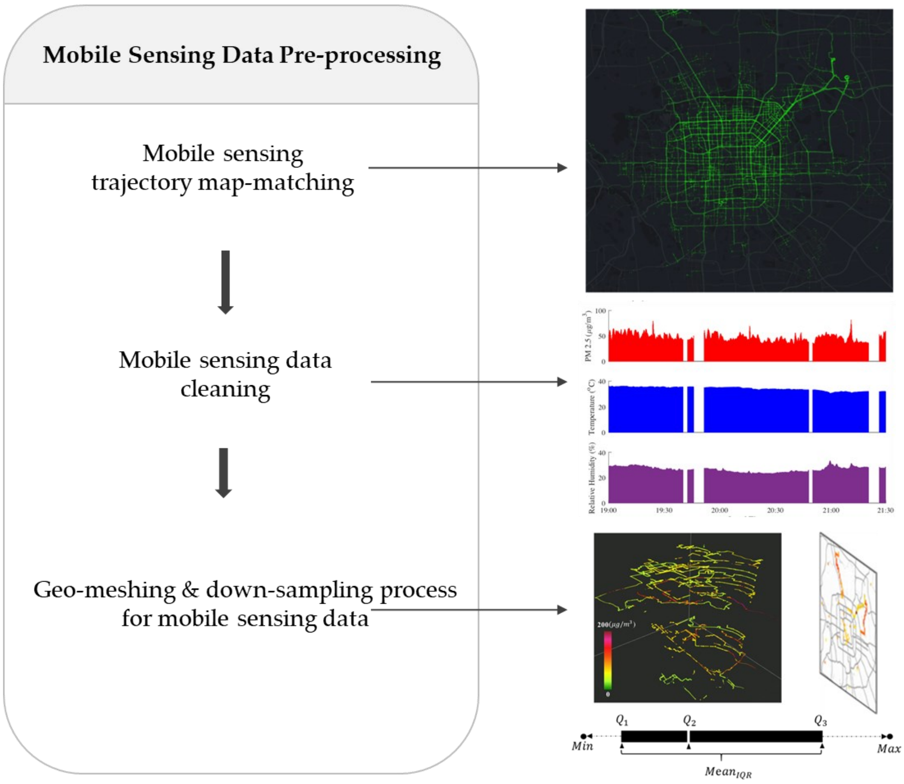


Figure 5.2: Pre-processing for mobile air pollution sensing

1) *Mobile sensing trajectory map-matching,* which matches the GPS trajectories into the corresponding road segments on the map.

This step uses a geometric refinement to remove errors by considering the GPS sensors' random shifts. Figure 5.3 shows a diagram of the map-matching for trajectory. When a trajectory Γ is under the constraints of a road-network G, the mobile sampling points $\{p_1, p_1, \ldots, p_n\}, n \in \mathbb{N}$ can be aligned to the corresponding road segments $\{r_1, r_1, \ldots, r_n\}, n \in \mathbb{N}$ in the road network G by map-matching (Ji et al., 2016). Γ is then transferred to $\Gamma_{matched}$ which is the representation of the road-network constrained trajectory.

Figure 5.3: Diagram of map-matching for trajectory

2) *Mobile sensing data cleaning*, which removes the outliers of GPS trajectories and AQ measurements.

   This step cleans the raw trajectories and the wireless collection of raw AQ readings and other relevant data including time, coordinates, temperature and relative humidity from low-cost mobile sensors. Data that are beyond the normal range will be cleaned after this step. This research considers samples as anomalous data based on 4 criteria: 1) samples with temperature over 40 ℃ or below 0 ℃, 2) samples with pressure over 1500 kPa, 3) samples with humidity over 100%, and 4) samples where the wind direction is over 360° or below 0°. Besides, this research regards the air pollutant concentration as outliers when the value is greater than 500 (ug/m^3). For these anomalous data, this research proposes to fill them with the linear interpolation approach. Due to the random errors of mobile sensors, some GPS points or AQ measurements may be missing or appear as outliers which cause noises in the application.

3) *Geo-meshing and down-sampling process for mobile sensing data,* which group and match the mobile sensing data into corresponding grid-cells based on trajectories. The label response (Cristianini, 2014) refers to the representative value of the air pollutant measurements which serve as dependent variable by mapping with appropriate units based on their timestamps and coordinates. It should be noted that there might be many different measurements from mobile sensors in a cell

every hour, but these would only generate one hourly representative value through data pre-processing and statistical aggregation. It is necessary to calculate the representative AQ value as label response in each grid-unit as the adjustment for eliminating the potential temporal variability, considering AQ differs largely by distances however short it is due to the uneven distribution of emission sources, dilution, and physical-chemical transformations (Herwehe et al., 2004; Apte et al., 2017), it's necessary to calculate the representative AQ value as label response in each grid-unit as the adjustment for eliminating the potential temporal variability. Say the spatio-temporal domain as a discrete rectangle area $\mathrm{M_x} \times \mathrm{M_y}$ within a discrete time duration $\mathcal{T}$, the pollution map ϕ represents a three-dimensional matrix of size $\mathrm{M_x} \times \mathrm{M_y} \times \mathcal{T}$, and $\phi\ [i, j, k]$ represents the pollution concentration at spatio-temporal coordinate $[i, j, k]$ (corresponding to the x-axis, the y-axis of space, and t-axis of time). This step blinds the trajectory points with the index of each grid-unit (1 km × 1 km) with the form of $\phi\ [i, j, k]$ before down-sampling to generate AQ label responses by computing the mean value ($\mathrm{Mean_{Q1,Q3}}$) between the $[\mathrm{Q_1}, \mathrm{Q_3}]$ quartile values (in the study of Statistics, quartile refers to each of the three points where a set of numbers is divided into four equal parts) in a grid-unit as shown in Figure 5.4. Measurements generated in each the mobile sensors in each grid-cell can be different even if they are collected within the same hour, but these would only generate one aggregating hourly value.

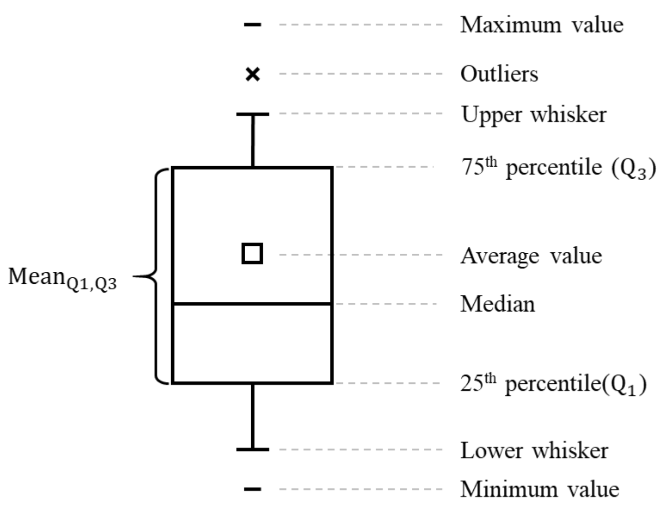


Figure 5.4: Diagram of computing the mean value ($Mean_{Q1,Q3}$) between the $[Q_1, Q_3]$ quartile values

Figure 5.5 shows an example of distribution visualization of mobile AQ measurements in a grid-cell unit. For arithmetic average method, the extreme outliers will sometimes significantly influence the whole average results; Thus, to eliminate the influence of extreme outliers as much as possible, this research exploits the quartile approach to filter out the outliers to eliminate the potential noise interference and make it more representative for each grid-unit to generate AQ label response with time interval.

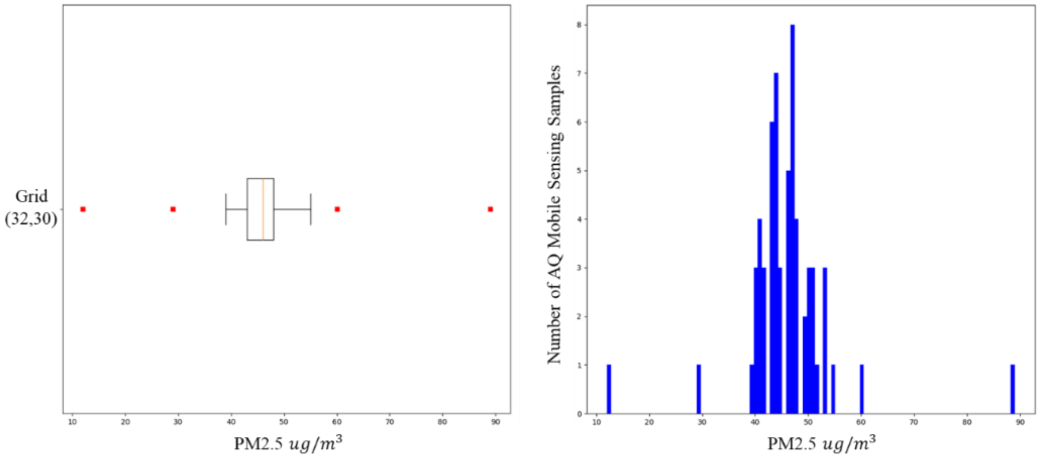


Figure 5.5: Mobile measurements ($PM_{2.5}$) in grid (32,30) during 11:30:00 – 12:29:59 in 2018/07/02

5.2.2 AQ labelling matrix for grid-cell units

Urban air pollution sensing IoT system consists of multi-array of fixed and mobile AQ sensors to continuously collect pollutant measurements, and the AQ samples collected from the sensing-covered grid-cell units make up the labelled observation dataset (Ma et al., 2020). During the data collection period, the samples are labelled after their timestamps and coordinates. Therefore, this research reformulates these data to construct a labelling matrix ϕ_{label} which size is the same as the pollution map ϕ (Ma et al., 2020). However, AQ measuring values are assigned to the entries with observations, whereas the value of zero is assigned to each entry without observations. Also, this research holds an assumption that all the values of AQ sensing are above zero, which is to prevent the possibility that the unobserved entries with zeros may lead to ambiguity (Ma et al., 2020). In this case, an objective will be drawn up to infer at unsampled positions to generate the complete air pollution map, given partial observations.

5.2.3 Urban feature selection, generalization, and fusion

In this research, selecting and combining urban features are a key part because the AQ conditions are strongly correlated to multi-feature factors, regardless of whether it is direct or indirect urban features. For example, static and dynamic raw urban feature maps may include features such as land use data, traffic dynamics. These features are grouped into four main categories in the Beijing case study, and the defined macro-view features are used to consider the potential influence of regional transport of air pollution. These explicit urban features allow data-driven algorithms to construct their spatio-temporal correlations, which is crucial to achieve pixelwise AQ inference.

Feature fusion for AQ inference combines geo-meshed AQ label responses and urban features to learn their coherence, relevance, importance, and interdependencies in the spatio-temporal domain. Figure 5.6 shows the geo-meshing for urban features (e.g., water cover area, green cover area, etc.). Figure 5.6(a) shows the mixed distribution of water cover area and road structures, which can be further extracted with only water cover area distribution as demonstrated in Figure 5.6(b). Figure 5.6(c) showcases the distribution of attribution value about water cover area, and the colour bar represents the value range about the water cover area in each grid-unit. Figure 5.6(d) shows the mixed distribution of green cover area and road structures and 5.6(e) demonstrates when only green cover area distribution is extracted. Figure 5.6(f) shows the distribution of attribution value about the green cover area, and the colour bar represents the value range about the green cover area in each grid-unit. Figure 5.6(g)-(i) respectively shows the distribution of attribution value about the POI of building, traffic light, and factory, and the colour bar represents the value range about the POIs in each grid-unit.

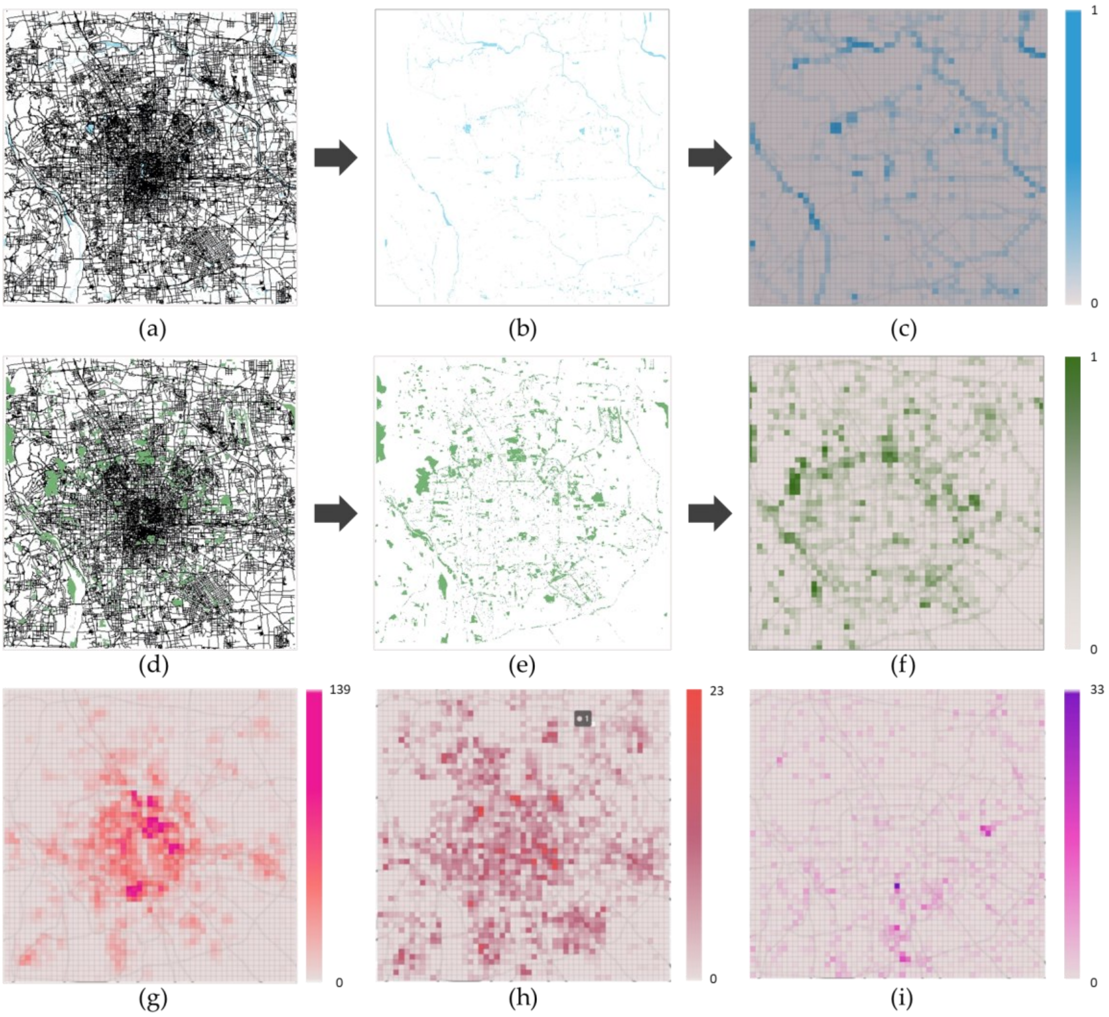


Figure 5.6: Geo-meshing process for urban features (e.g., AOI of water, green cover and POI of building, traffic light, and factory)

These urban features are geo-meshed corresponding to grid-cell units, and the geo-meshed grid-cell unit-based feature maps are processed to create the multi-dimensional features for grid-cell units. All the raw features in each grid-cell unit are then concatenated into local feature collection as the input vector which is defined as V_{micro}. Considering the spatio-temporal correlations, the meso-view features are generated by compressing neighbouring feature sets as the input vector which is defined as V_{meso}. The hourly pollutant measurements from fixed monitoring sites in different directions surrounding the study area are used to generate macro-view feature sets as the input vector which is defined as V_{macro}. Details of the micro-view, meso-view, and macro-view generalized features are discussed in the current chapter, and Figure 5.7 shows the process of geo-meshing, transformation, and fusion for feature maps to form generalized features from micro-view, meso-view, and macro-view.

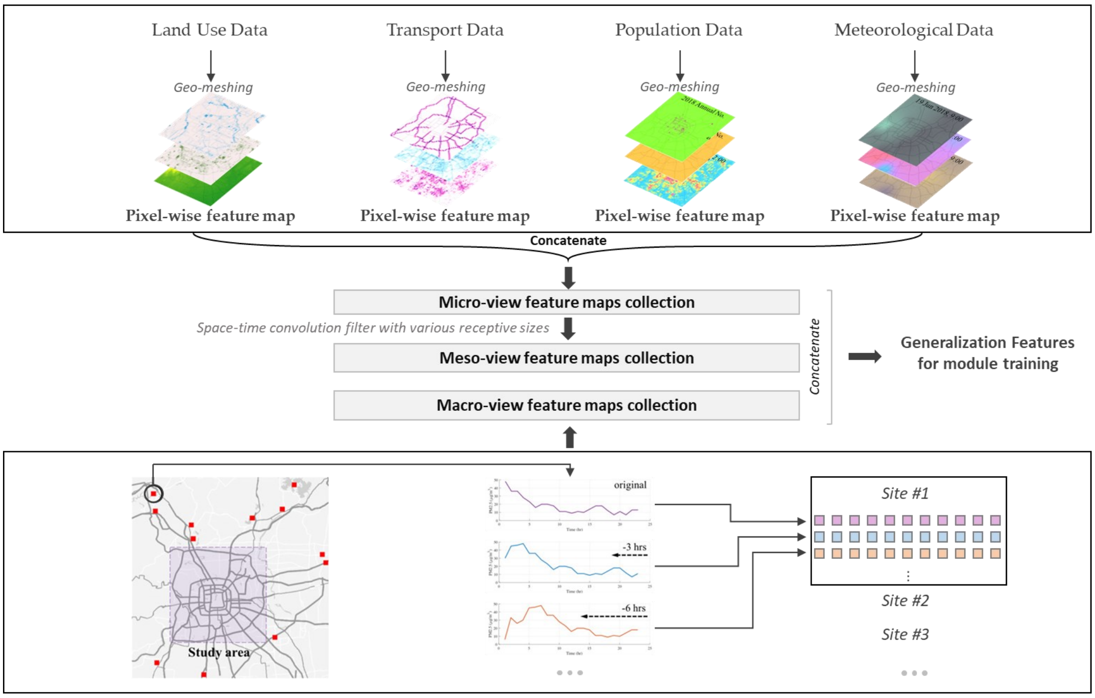


Figure 5.7: Diagram of geo-meshing, transformation, and fusion for generalized feature maps for training

5.2.4 Tree-based spatio-temporal learning framework (DFeaST-Tree)

Figure 5.9 illustrates DFeaST-Tree, which refers to the tree-based Deep AQ framework. Three viewed types of generalized features are considered: micro-view, meso-view, and macro-view.

- **Micro view (Local features)**

Local features are a mix of geographic, land use, transport, public vitality, and meteorological information all of which are defined for each grid-cell in the case of static feature, and grid-time tuple (s, t) in the case of dynamic feature (Chengdu: 52 static and 9 dynamic local features; Beijing: 52 static and 10 dynamic local features).

- **Meso view (Neighbouring correlations)**

According to first law of geography by Tobler's (1970), everything is interrelated, and the closer things are the stronger relation they share. Neighbouring features are compressed for each given unit by including features of adjacent units in space and time. The research is therefore able to establish these urban features' spatio-temporal correlations. Following the convolution technique (Nielsen, 2015) the extraction of neighbouring features is carried out, and this research draws a clear line between static and dynamic features. Both static and dynamic features are multi-dimensional, and the convolution filter operates on all dimensional feature maps.

The N_s static features are treated as N_s input images of size $M_x \times M_y$, forming an input volume $V \in \mathbb{R}^{M_x \times M_y \times N_s}$. This research chooses two sets of filters by sensitivity analysis (details in Appendix VI) on both the number of sets and sizes of the receptive field and the filters of $\{A^1, \ldots, A^{L_s}\} \subset \mathbb{R}^{3\times3\times N_s}$ and $\{B^1, \ldots, B^{L_s}\} \subset \mathbb{R}^{5\times5\times N_s}$ are applied to V with stride 1 and padding 0 ($L_s = N_s$ in the thesis which is the dimensional number of static features). Specifically, the k -th channel A_k^i (or B_k^i) of A^i (or B^i) is a mean filter with a random multiplicative weight:

$$A_k^i = w_k^i \times \frac{1}{9} \times 1(3,3), B_k^i = \omega_k^i \times \frac{1}{25} \times 1(5,5), 1 \le i \le L_s, 1 \le k \le N_s$$

where w_k^i and ω_k^i are i.i.d following the standard normal distribution. And two feature maps are generated by convolution with the two sets of filters

$$V_A = \oplus (V_A^1, \ldots, V_A^{L_s}), \quad V_B = \oplus (V_B^1, \ldots, V_B^{L_s}) \tag{5.1}$$

where

$$V_A^i = f(V * A^i),\ V_B^i = f(V * B^i)\ \ 1 \le i \le L_s \tag{5.2}$$

Here $*$ is the convolution operator; $\oplus$ is the concatenation operator. The mean filters perform arithmetic average within their receptive fields, while the random weights are used to sample different combinations of the input features. Finally, the output is $W = \oplus (V_A, V_B) \in \mathbb{R}^{M_x \times M_y \times 2L_s}$; see Figure 5.9 (top).

The N_d dynamic features are convolved similarly. For every time $t \in \mathcal{T}$, the input volume V_t consists of the dynamic features of the present and previous time steps to account for the temporal dependencies. Figure 5.8 demonstrates the spatio-temporal transformation and fusion for neighbouring features with convolution operators.

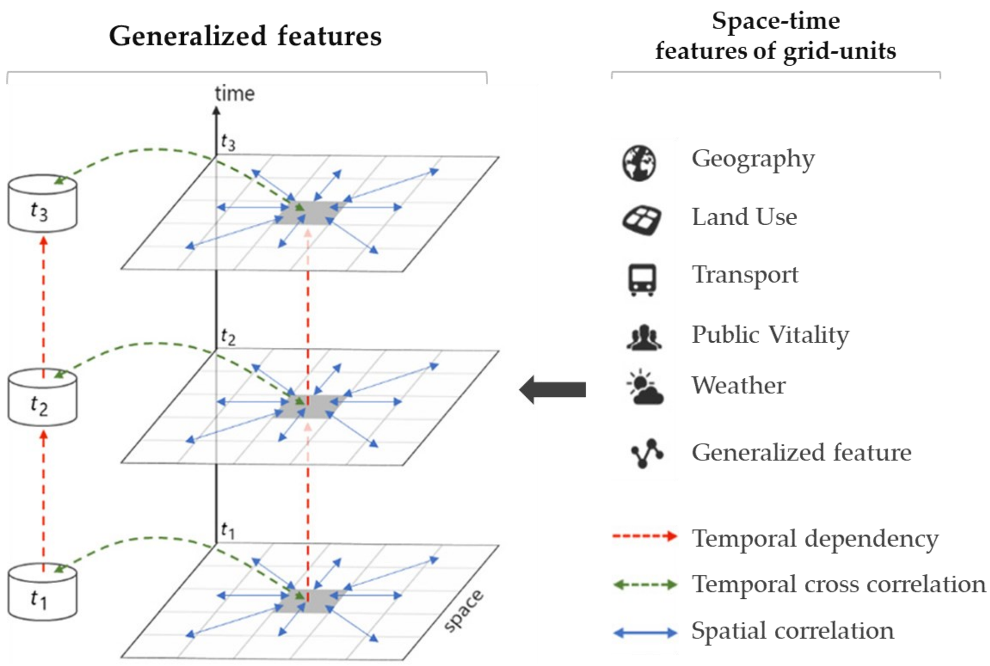


Figure 5.8: Spatio-temporal transformation and fusion for neighbouring features with convolution operators

This research then applies three sets of different sizes of filters $\{C^1,\ldots,C^{L_d}\} \subset \mathbb{R}^{1\times1\times2N_d}, \{D^1,\ldots,D^{L_d}\} \subset \mathbb{R}^{3\times3\times2N_d}, \{E^1,\ldots,E^{L_d}\} \subset \mathbb{R}^{5\times5\times2N_d}$, whose structures are similar to those of A^i and B^i and obtain:

$$C_k^i = v_k^i \times 1(1{,}1), D_k^i = v_k^i \times \frac{1}{9} \times 1(3{,}3), E_k^i = \mu_k^i \times \frac{1}{25} \times 1(5{,}5) \tag{5.3}$$

$$V_C^t = \oplus\, (V_C^{t,1},\ldots,V_C^{t,L_d}), V_D^t = \oplus\, (V_D^{t,1},\ldots,V_D^{t,L_d}), V_E^t = \oplus\, (V_E^{t,1},\ldots,V_E^{t,L_d}) \tag{5.4}$$

$$V_C^{t,i} = f(V^t * C^i), V_D^{t,i} = f(V^t * D^i), V_E^{t,i} = f(V^t * E^i) \tag{5.5}$$

where $1 \le i \le L_d, 1 \le k \le 2N_d$, and the output for time t is $W^t = \oplus\, (V_C^t, V_D^t, V_E^t) \in \mathbb{R}^{M_x \times M_y \times 3L_d}$; see the parts about static and dynamic neighbouring features in Figure 5.9 (middle).

- **Macro View (Background information)**

In order to also consider regional transport of pollutants without excessive need to expand the study area, this research employs additional features by including the AQ data collected from fixed monitoring stations outside of the selected area. The premise of the research is that there does exist the probability that these data will be temporally cross-correlated (that is, with possible time lags) with the background pollutant concentration. The the inclusion of the macro-view features, it can ensure that the results of AQ inference are more independent from locally defined features so as to lower the risk of model overfitting.

This research collects pollutant concentration data from N_G monitoring sites outside the study area, as shown in Figure 5.9 (bottom). By analyzing these time series with time shifts depending on the distance of the stations and average wind speed, the research may indirectly account for pollutants' transport dynamics. Specifically,$\forall 1 \leq i \leq N_G$, let $\{y_i(t), t \in \mathbb{Z}\}$ be the hourly time series provided by the i-th monitoring site. For a given time $t \in \mathcal{T}$, the research defines the macro-view features for grid-cell units $u \in \mathcal{S} \times \{t\}$ as $\{y_i(t-\theta): 1 \leq i \leq N_G, \theta \in \Theta$, where Θ is the set of backward time shifts (Θ = {0, 1, 3, 5, 7} (hours) in this thesis). Alternatively, the macro-view features can be viewed as a set of $M_x \times M_y$ images:

$\{M_{i,\theta}^t: 1 \leq i \leq N_G, \theta \in \Theta, t \in \mathcal{T}\} \subset \mathbb{R}^{M_x \times M_y}$ where $M_{i,\theta}^t = y_i(t-\theta) \times 1(M_x, M_y)$; here, 1 (M_x, M_y) denotes a $M_x \times M_y$ matrix of one's; see Figure 5.9 (bottom).

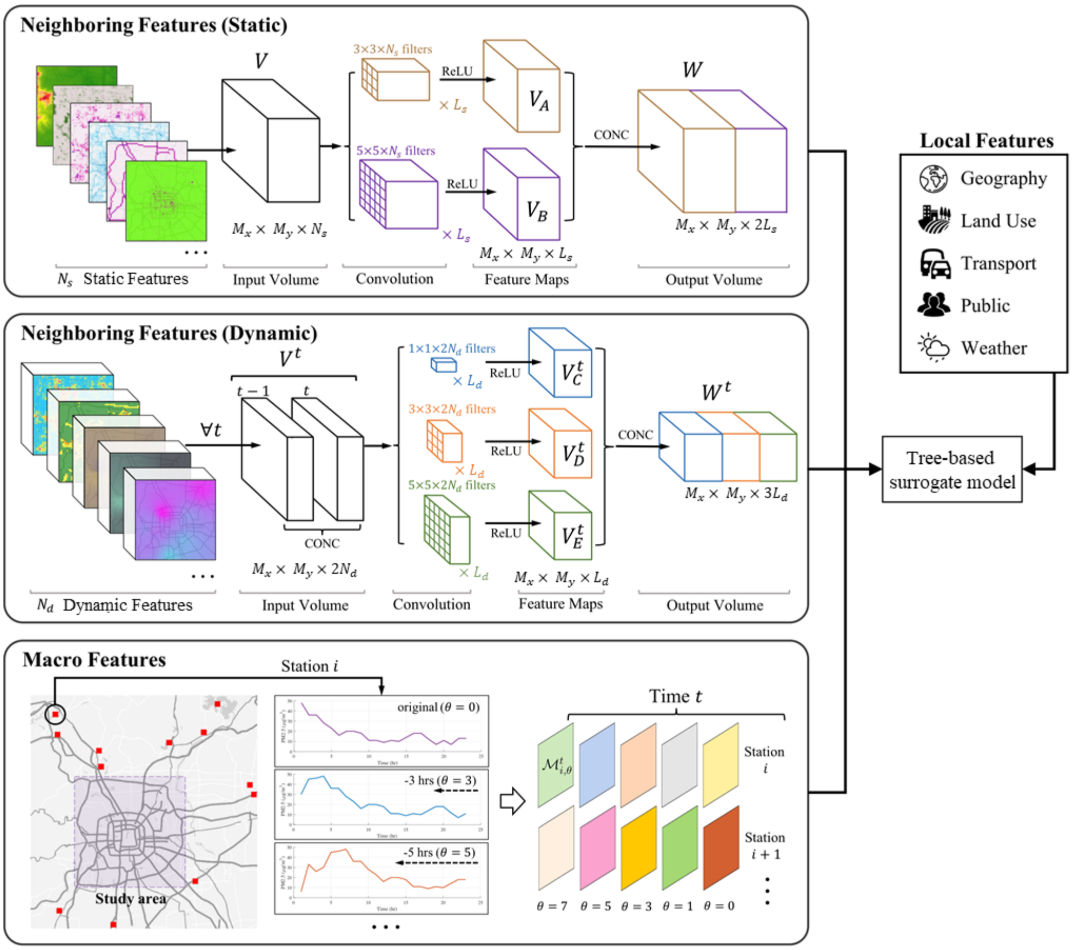


Figure 5.9: Feature extraction and learning framework for DFeaST-Tree

Compare to the problems of many other spatio-temporal machine learning, training label responses in this research are distributed based on the probe vehicles' trajectories, and hence they are more irregular in terms of space and time. In this respect, the DFeaST-Tree creates a label mask for screening out output elements in $\{W^t: t \in \mathcal{T}\}$ that are invalid according to the training label responses to collect the training samples from labelled grid-cell units at each time slice before achieving AQ inference based on the bagged or boosted model such as Random Forest (RF), Gradient Boosting Decision Trees (GBDT). Algorithm 5.1 details the learning structure of such DFeaST-Tree.

Algorithm 5.1: Training procedure for DFeaST-Tree
Input: Static features $\{\mathcal{S}_1, \dots, \mathcal{S}_{N_s}\} \subset \mathbb{R}^{M_x \times M_y}$ Dynamic features $\{D_1^t: t \in \mathcal{T}\}, \dots, \{D_{N_d}^t: t \in \mathcal{T}\} \subset \mathbb{R}^{M_x \times M_y}$ Macro features $\{M_{i,\theta}^t: 1 \le i \le N_G, \theta \in \Theta, t \in \mathcal{T}\} \subset \mathbb{R}^{M_x \times M_y}$ **Output**: Trained DFeaST-Tree for AQ inference

begin

Concatenate static features to form the input volume: $V = \bigoplus(\mathcal{S}_1, \ldots, \mathcal{S}_{N_s})$;

For $1 \le i \le L_s, 1 \le k \le N_s$, generate i.i.d. weights $\omega_k^i, \omega_k^i \sim N(0,1)$

$A_k^i \leftarrow \omega_k^i \times \frac{1}{9} \times 1(3,3)$, $B_k^i \leftarrow \omega_k^i \times \frac{1}{25} \times 1(5,5)$;

$W = \bigoplus(V_A, V_B)$ where V_A, V_B are given by (5.1)- (5.2)

end

for $\forall t \in \mathcal{T}$ **do**

Form the input volume $V^t = \bigoplus(D_1^t, \ldots, D_{N_d}^t, D_1^{t-1}, \ldots, D_{N_d}^{t-1})$;

For $1 \le i \le L_d, 1 \le k \le N_d$,generate i.i.d. weights $\upsilon_k^i, \upsilon_k^i, \mu_k^i \sim N(0,1)$

$C_k^i \leftarrow \upsilon_k^i \times 1(1,1), D_k^i \leftarrow \upsilon_k^i \times \frac{1}{9} \times 1(3,3), E_k^i = \mu_k^i \times \frac{1}{25} \times 1(5,5)$;

$W^t = \bigoplus(V_C^t, V_D^t, V_E^t)$ where V_C^t, V_D^t, V_E^t are given by (5.3) -(5.4)

end

Train the tree-based surrogate model using input features $W, \{W^t : t \in \mathcal{T}\}$ and $\{M_{i,\theta}^t : 1 \le i \le N_G, \theta \in \Theta, t \in \mathcal{T}\}$;

5.2.5 Network-based spatio-temporal learning framework (DFeaST-Net)

In addition to DFeaST-Tree, the network-based model DFeaST-Net is also proposed and applied so that the weights of filters weights can be updated iteratively and automatically through back-propagation in Fully Convolutional Network (FCN). Figure 5.10 illustrates the said process.

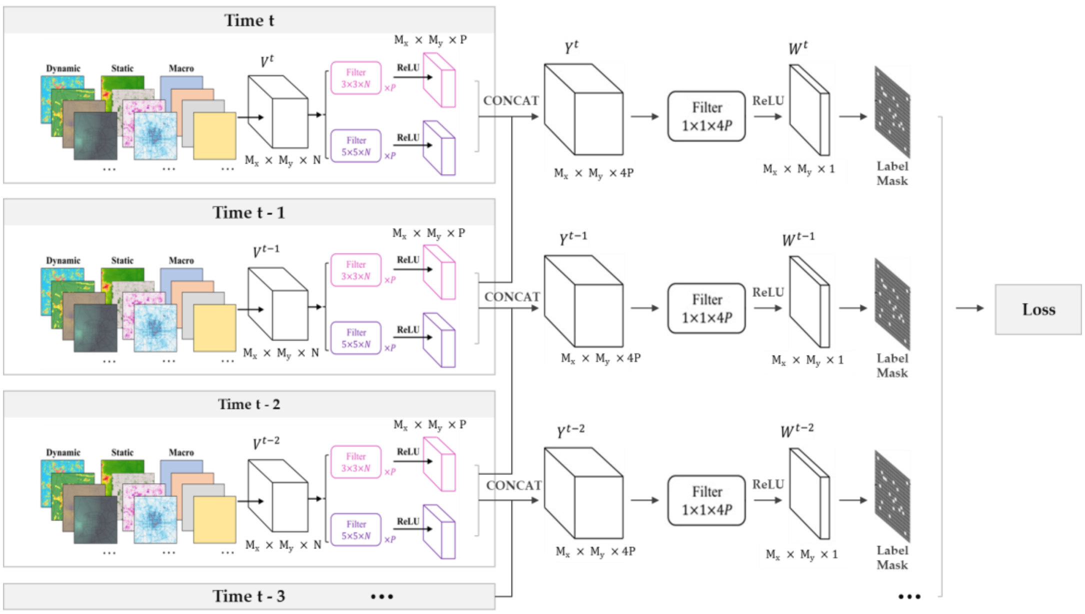


Figure 5.10: Feature extraction and learning framework for DFeaST-Net

There is a total of N ($N = N_s + N_d + N_G$) input feature variables. Given any $t \in \mathcal{T}$, the FCN first convolves around an input volume. $V^t \in \mathbb{R}^{M_x \times M_y \times N}$ consists of micro-view (static and dynamic) and macro-view features correlates with the time t. With a learnable parameter σ (all time steps use the same parameters), two sets of filters $\{F^1,\ldots,F^P\} \subset \mathbb{R}^{3\times3\times N}$ and$\{G^1,\ldots,G^P\} \subset \mathbb{R}^{5\times5\times N}$ are applied. This is a step where two feature maps V_F^t, $V_G^t \in \mathbb{R}^{M_x \times M_y \times P}$ are computed through the rectifier. These two feature maps correspond to different fields of the filters respectively. In order to take temporal dependencies into account, the feature maps from the previous time step $t-1$ is further incorporated in the current research:

$$Y^t = \oplus\,(V_F^t, V_G^t, V_F^{t-1}, V_G^{t-1}) \in \mathbb{R}^{M_x \times M_y \times 4P} \tag{5.6}$$

An additional convolution layer a $1 \times 1 \times 4P$ filter is applied in this research with learnable parameters λ (parameters are listed in Appendix VI Table I). This is followed by the yield of the final output map $W^t \in \mathbb{R}^{M_x \times M_y}$ for time t through rectifier. To address the dynamic irregular distribution problem, the DFeaST-Net also constructs a label mask to filter out output elements that are invalid in $\{W^t : t \in \mathcal{T}\}$ according to the training label responses. Whereas the set of grid-cell units with training label responses are represented by u_L, y_i and $\widetilde{y_i}$ represent the ground-truth and model inference values respectively. Algorithm 5.2 details the training procedure for DFeaST-Net which is to minimize the aforementioned objective function until a termination criterion (Chandramouli et al., 2007).

Algorithm 5.2 Training procedure for DFeaST-Net

Input: Static features $\{\mathcal{S}_1,\ldots,\mathcal{S}_{N_s}\} \subset \mathbb{R}^{M_x \times M_y}$
Dynamic features $\{D_1^t : t \in \mathcal{T}\},\ldots,\{D_{N_d}^t : t \in \mathcal{T}\} \subset \mathbb{R}^{M_x \times M_y}$
Macro features $\{M_{i,\theta}^t : 1 \le i \le N_G, \theta \in \Theta, t \in \mathcal{T}\} \subset \mathbb{R}^{M_x \times M_y}$
Output: Trained *DFeaST-Net* for AQ inference
for $t \in \mathcal{T}$ **do**
Concatenate features to form the input volume for time *t*:
$V^t = \oplus\,(\mathcal{S}_1,\ldots,\mathcal{S}_{N_s}, D_1^t,\ldots,D_{N_d}^t,\ldots,M_1^t,\ldots,M_{N_G}^t)$
End
Normalize each feature of the input
Initialize an FCN and pretrain the parameters σ, λ
FCN = $\oplus$ (Input layer, Middle layers, output layer)
Fine Tune the parameters of FCN

Train the FCN and optimize σ and λ by minimizing the Huber loss function through backpropagation

until Termination criterion is met;

5.3 Network-based multi-pollutant spatio-temporal learning framework (Multi-AP)

The Deep AQ (Multi-AP) achieves AQ inference based on DFeaST-Net for multiple pollutants simultaneously within one integrated learning network, which encompasses both the shared network layers to construct the correlations among all pollutants and the individual network layer for each pollutant's inference independently. This will enable the balancing of commonalities and differences among these AQ inference tasks. Figure 5.11 shows the Multi-AP learning network's technical framework, and Algorithm 5.3 describes the Multi-AP learning network's training procedures.

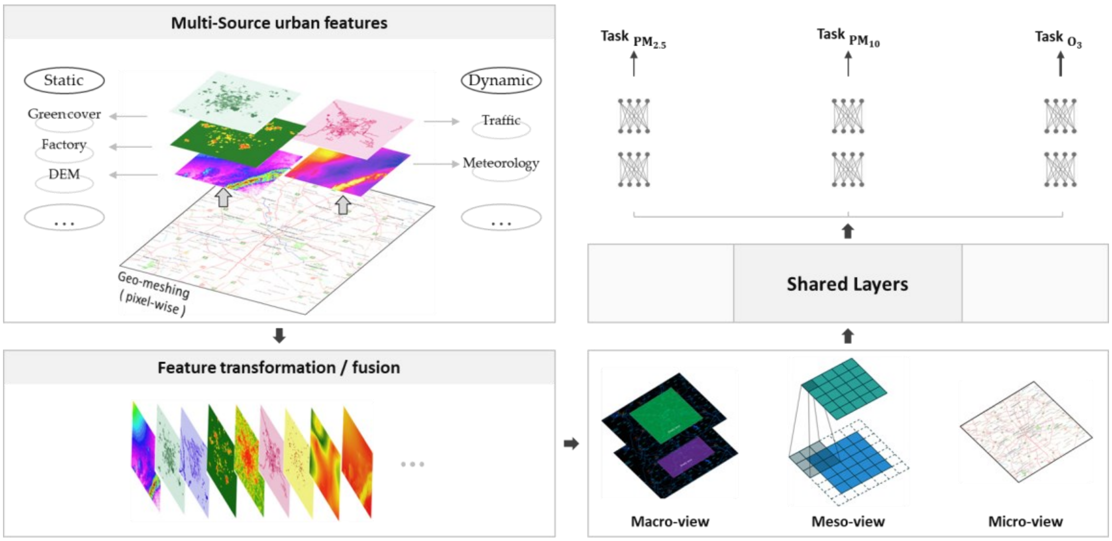


Figure 5.11: Technical framework of Multi-AP learning network

The final training process is trained jointly for multiple AQ inference tasks, with an aim to minimize the following objective function until a termination criterion (Chandramouli et al., 2007):

$$\mathrm{Loss(AQ)} = \mathrm{Loss}(\sigma_1, \lambda_1)_{\mathrm{AQ}_1} + \cdots + \mathrm{Loss}(\sigma_Z, \lambda_Z)_{\mathrm{AQ}_Z} \quad (5.7)$$

$\mathrm{AQ_i}$ refers to the i_{th} air pollutant species and the total number of pollutant species is assumed to be Z.

Algorithm 5.3 describes the training procedures of the Multi-AP learning network.

Algorithm 5.3: Training procedure of Multi-AP learning network

Input: Static features $\{\mathcal{S}_1, \ldots, \mathcal{S}_{N_s}\} \subset \mathbb{R}^{M_x \times M_y}$

Dynamic features $\{D_1^t: t \in \mathcal{T}\}, \ldots, \{D_{N_d}^t: t \in \mathcal{T}\} \subset \mathbb{R}^{M_x \times M_y}$

Macro features $\{M_{i,\theta}^t: 1 \le i \le N_G, \theta \in \Theta, t \in \mathcal{T}\} \subset \mathbb{R}^{M_x \times M_y}$

Output: Trained Multi-AP for AQ inference with multi-pollutant simultaneously

for t ∈ $\mathcal{T}$ **do**

Concatenate features to form the input volume for time *t*:

$V^t = \oplus (\mathcal{S}_1, \ldots, \mathcal{S}_{N_s}, D_1^t, \ldots, D_{N_d}^t, \ldots, M_1^t, \ldots, M_{N_G}^t)$

End

Normalize each feature of the input

Initialize an FCN and **pretrain** the parameters σ_z, λ_z $(1 \le z \le Z)$

FCN = $\oplus$ (Input layer, Middle layers, output layer)

Fine Tune the parameters of FCN

Train the FCN and optimize σ_z and λ_z to get the final minimization Loss (AQ) function (5. 7) for all tasks jointly through the backpropagation.

Until Termination criterion is met to get minimized Loss(AQ);

5.4 Summary

This chapter describes the thesis's third research objective, which is to:

Propose a spatio-temporal learning and inference framework called Deep AQ framework by incorporating the multi-source AQ samples and urban features to achieve high-resolution pixelwise AQ inference.

This chapter introduces two types of Deep AQ frameworks: 1) the tree-based model called DFeaST-Tree, and 2) the network-based model called DFeaST-Net. Section 5.1 introduces the technical route of the methodological framework. Section 5.2.1 describes the details of pre-processing for mobile trajectories and measurements, including trajectory map-matching, data cleaning, geo-meshing, and down-sampling to generate representative values (label responses) corresponding to the specific grid-cell units. Section 5.2.2 illustrates the AQ labelling matrix for grid-cell units. Section 5.2.3 introduces the urban feature selection, generalization, and fusion processes. Section 5.2.4 proposes the tree-based model DFeaST-Tree, which transforms the generalized features from micro-view (local features), meso-view (neighbouring spatio-temporal correlations), and macro-view (regional transport of air pollution). Algorithm 5.1 describes the training procedure of DFeaST-Tree and demonstrates the transformation and generalization process for urban features, the generating process for dynamic label responses, and the training procedure developed from bagging or boosting decision tree models (e.g., GBDT). Section 5.2.5 proposes the network-based spatio-temporal learning framework DFeaST-Net, which also transforms and generalizes features from micro, meso, and macro views. Algorithm 5.2 describes the training procedure of DFeaST-Net and demonstrates the feature transformation and generalization process, the process to generate dynamic label responses, and the model training process. It should be noted that the meso-view features are generated by basic convolution operations (average and max pooling) with no learnable backward updating mechanism but generated by the convolution operator (static weight of the kernel) in DFeaST-Tree, which has a learnable backward updating mechanism in DFeaST-Net. Section 5.3 proposes a network-based multi-pollutant spatio-temporal learning framework which is able to build multiple AQ inference tasks in one-integrated network based on DFeaST-Net for multiple air pollutant species parallelly and simultaneously. The two options for DFeaST-Tree and DFeaST-Net are respectively developed from different principled model structures (tree-based and network-based), which are exploited to compare their commonalities and differences in the proposed methodological framework to practice in actual uses (to be detailed in Chapter 6 for Chengdu case study and in Chapter 7 for Beijing case study).

Chapter 6

Predicting AQ in Chengdu using fixed monitoring sites

6.1 Introduction

Chapter 5 discusses the Deep AQ framework incorporating multi-source AQ samples and urban features to achieve high-resolution (1 km × 1 km) pixelwise AQ inference for pollution map-recovery and evolution. Traditionally, estimating AQ conditions is mainly achieved upon fixed monitoring sites (Wang et al., 2020). Various inquiries have examined and contrasted the performance of LUR models, and ML-based approaches at various spatial-temporal scales using fixed AQ monitoring data (Brokamp et al., 2017; Zheng et al., 2013; Qi et al., 2018). Brokamp et al. (2017) use the RF method to estimate elemental components of particulate matter within Cincinnati, Ohio. Zheng et al. (2013) and Qi et al. (2018) consider ML-based techniques in LUR models (Zheng et al.: co-training framework; Qi et al.: semi-supervised learning) to estimate AQ conditions using fixed monitoring sites in Beijing. However, these studies are restricted to limited spatio-temporal resolution (street level; daily) and have not constructed a systematic high-resolution (1 km × 1 km, hourly) AQ inference framework with reference to limited, sparse AQ measurements and urban features (e.g., traffic, population, land use data, etc.).

In this chapter, the Deep AQ framework is applied in Chengdu to estimate pixelwise pollutant concentrations based on fixed AQ monitoring data and multi-source urban features, including land use information, traffic data, and meteorological conditions. Experimental conclusions about model performance and robustness, relative pollution attribution and appraisal are made in this chapter to provide evidence-based support for sustainable urban management.

Chengdu is a western Chinese city located between 30.05 ~ 31.26N latitude and 102.54 ~104.53E longitude. There are 40 fixed monitoring sites in the study area of Chengdu (4,900 km^2, 26 April - 12 June 2019), where the concentrations of air pollutant species ($PM_{2.5}$, PM_{10} and O_3) are collected from fixed monitoring sites to train the model to estimate hourly pollutant concentration in 4,900 1 km × 1 km grids.

The objectives in this chapter are highlighted as follows:

1) Apply Deep AQ framework using fixed AQ data combining with multi-source urban features (e.g., population, traffic, land use information) and external background information (e.g., regional transport of air pollution) to achieve pixelwise (1 km × 1 km, hourly) AQ inference within the location studied (4,900 km^2).
2) Evaluate the Deep AQ framework's performance by comparing it with other benchmarks based on HOCV tests.
3) Investigate the performance improvement with meso-view and macro-view features in the Deep AQ framework.
4) Validate the Deep AQ framework's robustness and stability by comparing the standard deviation on the different validation datasets.
5) Investigate the variability of the Deep AQ framework's performance to the availability of various feature predicators.
6) Evaluate the Deep AQ framework's predictivity to achieve AQ inference at fully unsampled grids by deploying independent LOOCV tests.
7) Evaluate the Deep AQ framework's predictivity to achieve AQ inference at independent block by deploying the spatial block CV tests.
8) Analyse the pollution maps recovered by AQ inference to provide evidence-based support for sustainable urban management.

6.2 Materials and methods

Modelling urban air pollution with the Deep AQ framework involves heterogeneous urban features, which differs in spatio-temporal resolution, mathematical scale, and accuracy. Thus, it's critical to transform characteristics using suitable spatio-temporal structures so as to integrate both endogenous (e.g., population, traffic, land use information) and external background information (e.g., regional transportation of air pollution). The Deep AQ framework's technical details are described in Chapter 5, and three types of feature views are transformed in the Deep AQ framework: micro-view, meso-view, and macro-view.

For the Chengdu case study in this chapter, the region studied is 70 × 70 km squared, and it covers the around-city ring road, depicted Figure 3.1 and the pollutant samples ($PM_{2.5}$, PM_{10} and O_3) are measured from 40 fixed monitoring sites and using meteorological parameters. The timespan is 0:00-23:00 (24 hours a day) over 48 days (2019/04/26 - 2019/06/12), with granularity per hour. Moreover, the collection period of the pollutant, traffic, and meteorological data is during 00:00-23:00 (24 hours) from 2019/04/26 to 2019/06/12 (48 days). Table 3.1 lists the AQ sensing data and urban features used in the Chengdu case study, and these datasets are supported by local governments.

To compare how this suggested model performs, the common or state-of-the-art approaches in existing studies (Liu & Rossini, 1996; Hoek et al., 2001, 2002; Cheng et al., 2014; Qi et al., 2018; Ma et al., 2020; Wang et al., 2020) are used. These approaches include: Arithmetic Average (AA), Spatial Interpolation (SI), Land Use Regression (LUR), K-Nearest Neighbours (KNN), Auto Encoder (AE), Gaussian Process Regression (GPR), Deep Air Learning (DAL) and Support Vector Regression (SVR), . And the aforementioned Deep AQ framework produces the results with variations of micro-view (L), meso-view (N), and macro-view (G) features. The network-based Deep AQ frameworks refer to DFeaST-Net and Multi-AP. The tree-based Deep AQ frameworks refers to DFeaST-Tree (DT), DFeaST-Tree (RF), and DFeaST-Tree (GBDT).

The following approaches are compared alongside the proposed benchmarks and the parameter settings are listed in Appendix V Table I.

- **Arithmetic Average (AA):** The arithmetic average approach is applied for every $t \in \mathcal{T}$ to estimate the mean values of air pollution concentration in fixed monitoring points in the study area for each grid-unit in time slices. It disregards all relevant features and estimates the arithmetic average value of all training data points for every test data point, which refers to the simplest baseline that other approaches should outperform.

- **Spatial Interpolation (SI):** SI is applied to estimate pollutant concentration per $t \in \mathcal{T}$ with reference to 2 popular approaches: Kriging and Inverse Distance Weighting (IDW). Hoek et al. (2001, 2002) and Liu & Rossini (1996) respectively use IDW and Kriging to interpolate and generate AQ mappings from monitored data. The aforementioned geostatistical interpolation approaches can be used to estimate the pollutant concentration, but these SI approaches are not capable of expressing variability in view of the local-scale land cover (Gulliver et al., 2011). In contrast, in relation to the Kriging method, the ordinary Kriging is adopted here and selects the linear function as the variogram model.

- **Land Use Regression (LUR):** LUR is done by constructing multiple linear regression models describing the relationships between AQ observations at a sample of monitors and relevant land use variables are integrated to compute which can account for the spatial dependency, using GIS, for zones of influence around each site (Details can be seen in section 4.2) which can account for spatial dependence. Although LUR model is generally intuitive, interpretable (Wang et al., 2020), its basicness confines how well it performs in dealing with complicated AQ conditions (Wang et al., 2020).

- **K-Nearest Neighbours (KNN):** Nearest neighbour techniques have their basis in the labels of the K-nearest patterns in data space (Kramer, 2013). Nearest neighbour approaches perform well with sizeable datasets and low dimensions (Kramer, 2013). KNN, a regression approach, employs the k closest training points to make AQ inferences.

- **Support Vector Regression (SVR):** This method seeks to minimize errors associated with generality, rather than minimising the observed training error

(Smola & Schölkopf, 2004). Its basis is in the computation of a linear regression function within a high-dimensional space in which a nonlinear function is used to map input data (Smola & Schölkopf, 2004; Satapathy & Rath, 2014). SVR instigates nonlinear regression with kernels which are derived from similar principles of support vector machine classifier.

- **Gaussian Process Regression (GPR):** The Air Cloud team models the air pollution field as a GP and the gaussian kernel is adopted which is a kind of RBF using the adjustable parameter σ to tune the performance of the kernel illustrated in Equation (6.1).

$$k(\gamma_i, \hat{\gamma}_i) = exp(-\frac{\left\|\gamma_i - \hat{\gamma}_i\right\|^2}{2\sigma^2}) \tag{6.1}$$

If overestimated σ, the exponential will lose the non-linear power at high-dimensional space and perform linearly. If underestimated σ, it will be highly sensitive to noise and lack regularization. Thus, the grid search method (Wang & Li, 2005) is adopted to find the appropriate σ. Samples are used to inform probability distribution before being applied to estimate unobserved phenomena (Cheng et al., 2014). This approach is regarded as performing well in achieving AQ inference using mobile sampling. Furthermore, deep gaussian process (Deep GPs) are considered as competitive models and will be explored in the future work which are a deep belief network derived from GP mapping where data is modelled as the output of a multivariate GP and a different GP governs the input (Damianou & Lawrence, 2012).

- **Auto Encoder (AE):** Auto Encoder framework is proposed for AQ inference by Ma et al. (2020), consisting of 2 models at the inference stage, namely Encoder and Decoder. An autoencoder framework is used to apply the partially-observed pollution map into a low-dimensional variable under the autoencoder framework, before the variable is decoded in order to assemble a complete map of the pollution situation.

- **Deep Air Learning (DAL):** DAL (Qi et al., 2018) achieves AQ inference and predictions by lying in embedding feature selection and semi-supervised learning in different deep learning network layers of the deep learning network. DAL is an enhanced and advanced version of the U-Air modelling framework which are all based on semi-supervised learning; however, DAL improves the performance by selecting features at the input layer of the network; the optimisation and good performance of this is simple to achieve in order to find the features that are most applicable to AQ variations.

The following evaluation metrics (RMSE, SMAPE and R^2 stand for root mean square error, symmetric mean average percentage error, and R^2). These metrics are for use in evaluating accuracy and goodness of fit.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\gamma_i - \hat{\gamma}_i)^2} \qquad \mathrm{SMAPE} = \frac{2}{N}\sum_{i=1}^{N}\left|\frac{\gamma_i - \hat{\gamma}_i}{\gamma_i + \hat{\gamma}_i}\right| \qquad \mathrm{R}^2 = 1 - \frac{\sum|\gamma_i - \hat{\gamma}_i|^2}{\sum|\gamma_i - \overline{\gamma_i}|^2}$$

In this study, three kinds of validation tests are applied: one is HOCV tests on all the monitoring sites which randomly samples the validation datasets on the whole datasets; the other two are the LOOCV tests on each independent site (Wang et al., 2012) and the spatial block CV tests on each block (Valavi et al., 2018).

Validation data are referred to as ground-truth values to evaluate the model performance (Jason, 2018). HOCV tests are primarily used in ML-based modelling and can approximate the performance of a ML-based model when faced with unseen data (Jason, 2018). HOCV tests involve randomly splitting the observations into k groups (known as folds) whose sizes are around the same. The first is treated as a validation set, and the method is fit on the remaining $k - 1$ folds. The multi-fold cross-validation is employed to get mean value at the evaluation step, quality-checking the model's predictions and fixing an arbitrary validation set (Jason, 2018). Here this research evaluates all the approaches by 5-fold cross-validation to test the model performance against the ground-truth values based on the whole fixed monitoring sites and accounts for the standard deviation across these 5-fold validations for $PM_{2.5}$, PM_{10} and O_3 to compare the stability and variability of the benchmarks.

To investigate the model's predictivity to achieve AQ inference at unsampled sensing sub-areas, this study splits the independent validation test subset to see how the model performs against the measurements from each independent monitoring site. The independent validation tests are carried on with the LOOCV tests (Wang et al., 2012): the datasets at $N - 1$ sites are employed to train the model, which is subsequently employed for AQ inference at the separated monitoring site. After repeating this process$N$ times, the metrics in terms of RMSE, SMAPE and R^2 are used to examine how well the model performed (Meng et al., 2015), and details are introduced in section 6.4. The aforementioned K-fold HOCV and LOOCV tests are two popular methods which have been widely adopted to validate the model performance in AQ epidemiological studies, such as Amini et al., (2012), Yanosky et al., (2014), Meng et al., (2015), Frauke et al., (2016), Xu et al., (2017), Ke et al., (2017), Zhong et al., (2018), Maddix & MD (2020), Wang et al., (2020) and more. The AQ varying process is a complex and fuzzy process and the AQ data is dependent in spatio-temporal (hence not independent) and in reality, it is also non-stationary (hence not identical). Both K-fold CV and LOOCV are appropriate for independent and identically distributed (i.i.d) data, while the AQ process and its data is not appropriate to either. Although the AQ data are spatio-temporal and non-stationary, it doesn't mean that the K-fold CV and LOOCV are not appropriate for the evaluation of the modelling effect. The CV results in this thesis are all averagely-based on multi rounds which can offset the potential biases of the evaluations to some extent; thus, they are acceptable for the generalisation goal at hand. In fact, the purpose of the model is to learn the spatio-temporal correlations with the randomly and unevenly distributed AQ observations and make spatial inference; thus, the model should offer the capability to train the model structures based on the randomly and unevenly distributed AQ labelling grid-units and finally make AQ predictions at the unlabelled grid-units. For the process of spatial inference, it is beneficial to incorporate more relevant labelled grid-units no matter whether they are explicit or implicit relevant to the unlabelled grid-units which offers more gain information for the model training to estimate at the unsampled grid-units. To make the outcomes more convinced, the spatial block CV tests are also conducted to evaluate the model performance where the data are split into blocks (e.g., rectangles) to treat together (Valavi et al., 2018).

6.3 Results and discussions

In the Chengdu case study, a total of 42,584 grid-units are labelled (sampling with fixed monitoring), of which 20% (8,525 units) are sampled in the validation test subset. For each 1-hour period, this research randomly samples 20% of the labelled grid-units as a validation dataset, i.e., 8 of the 40 fixed stations.

6.3.1 Performance of benchmarks

In this chapter, the capacity of the aforementioned benchmarks to make accurate predictions is compared. Table 6.1 shows the model results with mean accuracy based on 5-fold CV and compares various benchmarks and the Deep AQ frameworks where the results clearly show the superior performance of the Deep AQ framework (Multi-AP; L + N + M) over other benchmarks and the standard errors across different folds are listed in Table 6.5. The average arithmetic method is the simplistic baseline to compare and is widely used as a summary statistic. Compared with the arithmetic average method, the IDW and Kriging methods improve the interpretability of the methods accounting for the location factors (Inspired by the first law of geography from Tobler's (1970)). LUR shows better R-square than the aforementioned results but worse RMSE and SMAPE which is because the LUR cannot construct the non-linear relationships among land use information and pollutants. KNN or SVR performs worse than the aforementioned approaches, which indicates they are not suitable surrogate models. The performance of GPR, AE and DAL models from recent studies significantly increase compared with the aforementioned methods, and AE is more consistent with DAL considering they are both network-based models although constructed with different network techniques. Furthermore, the variations of Deep AQ frameworks all outperform the aforementioned models. For the variations of Deep AQ frameworks, the Multi-AP learning network and DFeaST-Tree facilitate the potential to provide reasonable results, and the Multi-AP performs best in this case study. For the tree-based Deep AQ frameworks, both the GBDT-based and RF-based models outperform the DT-based models, indicating that boosting and bagging offer the capability to improve the model performance and the GBDT-based model achieves the best which is slightly better than the RF-based model.

Table 6.1: The performance of various approaches and feature inputs (Views refers to the perspectives of feature transformation, including L, N, and M)

Methods	Views	$PM_{2.5}$			PM_{10}			O_3		
		RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Mean	/	9.63	25.44	0.467	16.42	24.06	0.565	25.80	31.63	0.692
IDW	/	8.30	22.17	0.566	15.67	21.52	0.635	19.35	23.72	0.731
Ordinary Kriging	/	9.26	24.17	0.583	17.33	22.32	0.614	18.21	22.48	0.744
LUR	L	13.98	42.72	0.61	27.28	44.28	0.653	32.95	41.56	0.767
KNN	L	13.88	37.58	0.668	25.39	35.37	0.686	33.73	39.65	0.582
SVR	L	14.47	43.09	0.604	28.74	45.30	0.669	35.13	40.86	0.547
GPR	L	10.81	23.44	0.726	16.90	24.31	0.734	20.49	26.91	0.811
AE	L	8.98	21.89	0.732	14.31	19.97	0.781	18.81	24.17	0.839
DAL	L	9.18	22.01	0.738	14.89	20.15	0.775	18.92	24.24	0.833
DFeaST-Net	L	9.05	22.43	0.733	14.85	20.09	0.781	18.89	24.22	0.839
	L+N	8.76	21.86	0.749	14.52	19.42	0.803	17.86	22.51	0.846
	L+M	8.47	21.42	0.757	14.58	18.94	0.807	17.51	22.18	0.851
	L+N+M	7.63	19.93	0.791	12.89	16.85	0.835	15.29	20.08	0.887
DFeaST-Net (Multi-AP)	L	8.93	21.97	0.745	14.35	19.83	0.794	17.97	23.13	0.861
	L+N	8.36	20.93	0.755	13.92	18.91	0.821	17.01	22.25	0.885
	L+M	7.59	20.37	0.772	13.08	17.77	0.829	16.32	21.89	0.892
	L+N+M	6.51	18.97	0.814	10.54	15.78	0.849	14.17	19.92	0.903
DFeaST-Tree (DT)	L	9.64	22.81	0.717	16.39	21.88	0.691	20.42	25.56	0.805
	L+N	9.02	22.39	0.733	15.62	21.22	0.712	19.53	23.98	0.821
	L+M	8.91	21.87	0.749	14.79	20.76	0.745	18.01	22.03	0.837
	L+N+M	8.05	20.99	0.768	13.73	18.69	0.771	17.17	21.95	0.873
DFeaST-Tree (RF)	L	7.89	20.45	0.770	11.78	17.06	0.809	17.60	23.27	0.865
	L+N	7.78	20.24	0.773	11.47	16.97	0.814	16.91	22.36	0.872
	L+M	7.72	20.21	0.779	11.12	16.71	0.815	16.84	21.78	0.878
	L+N+M	7.45	19.71	0.788	10.99	16.62	0.823	14.13	19.85	0.889
DFeaST-Tree (GBDT)	L	7.80	20.32	0.772	11.56	16.88	0.817	17.19	23.03	0.874
	L+N	7.73	20.22	0.786	11.25	16.49	0.822	16.81	22.44	0.885
	L+M	7.68	20.17	0.801	10.81	16.19	0.833	16.72	21.54	0.895
	L+N+M	7.26	19.43	0.812	10.59	15.82	0.847	14.15	19.81	0.901

Furthermore, to exhibit the applicability and satisfactory performance of the proposed model, this research quantifies the improvement by comparing the Deep AQ (Multi-AP, L + N + M) model's results with the worst results in other benchmarks across all metrics about each pollutant species. Table 6.2 shows the relevant margins.

Table 6.2: The performance improvement of the Deep AQ (Multi-AP, L + N + M) model compared with the worst results in other benchmarks across all metrics.

Methods	$PM_{2.5}$			PM_{10}			O_3		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Deep AQ (Multi-AP)	-7.96	-24.1	+0.347	-18.2	-29.52	+0.284	-20.96	-21.64	+0.356

6.3.2 Performance improvement with meso and macro features

To investigate the performance improvement with macro-view features in the Deep AQ framework, this research conducts comparison experiments of 'L + N' and 'L + N + M', which clearly shows that the macro-view features significantly improve the model's performance, which Table 6.3 lists.

Table 6.3: The performance improvement of the Deep AQ framework using 'L + N + M' instead of 'L + N'.

Methods	$PM_{2.5}$			PM_{10}			O_3		
	RMSE (ug/ m^3)	SMAPE (%)	R^2	RMSE (ug/ m^3)	SMAPE (%)	R^2	RMSE (ug/ m^3)	SMAPE (%)	R^2
Deep AQ (Multi-AP)	-1.85	-1.96	0.059	-3.38	-3.13	0.028	-2.84	-2.33	0.018
Deep AQ (GBDT)	-0.47	-0.79	0.026	-0.66	-0.67	0.025	-2.66	-2.63	0.016

To investigate the effect of meso-view features on the proposed models, this research carries out comparison experiments of 'L + N' and 'L', which clearly shows that the meso-view features significantly increase the model performance as listed in Table 6.4, suggesting that it is crucial to construct the spatio-temporal correlations among neighbouring grid-units.

Table 6.4: The performance improvement of the Deep AQ framework using 'L + N' instead of 'L'.

Methods	$PM_{2.5}$			PM_{10}			O_3		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Deep AQ (Multi-AP)	-0.57	-1.04	0.01	-0.43	-0.92	0.027	-0.96	-0.88	0.024
Deep AQ (GBDT)	-0.07	-0.1	0.014	-0.31	-0.39	0.005	-0.38	-0.59	0.011

Figure 6.1(a)-(c) show scatter plots of the validation datasets for $PM_{2.5}$, PM_{10} and O_3 species independently (5-folds combined, over 42,584 points) produced by Deep AQ (Multi-AP, L + N + M) model. The scatter plots ($PM_{2.5}$: $R^2 = 0.814$; PM_{10}: $R^2 = 0.849$; O_3: $R^2 = 0.903$) suggest that Deep AQ (Multi-AP, L + N + M) model facilitates the potential to produce excellent goodness of fit.

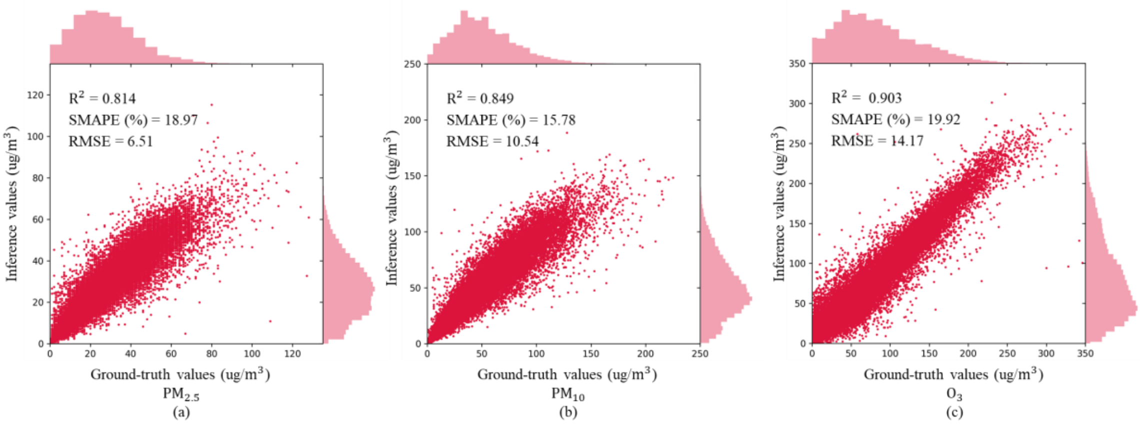


Figure 6.1: Scatter plots of testing points (5 folds) (Deep AQ (Multi-AP, L + N + M)) for (a) $PM_{2.5}$; (b) PM_{10}; (c) O_3.

6.3.3 Stability of the Deep AQ framework

To validate the stability of the Deep AQ framework on different validation datasets, this research carries out stability analysis to compare the proposed model with other benchmarks on the aspect of the standard deviation with different evaluation metrics for $PM_{2.5}$, PM_{10} and O_3. The results show that the Deep AQ model (Multi-AP, L + N + M) is more consistent than other listed benchmarks in Table 6.5.

Table 6.5: Standard deviation of the evaluation metrics across the 5-fold validation for $PM_{2.5}$, PM_{10} and O_3.

Methods	Standard deviation of metrics								
	$PM_{2.5}$			PM_{10}			O_3		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Mean	2.681	3.242	0.051	1.932	4.672	0.076	1.475	5.631	0.091
Kriging	1.152	1.258	0.042	1.267	2.325	0.041	1.113	2.176	0.026
IDW	1.096	0.976	0.039	1.215	1.962	0.033	1.086	1.972	0.021
LUR	1.271	2.461	0.036	1.418	3.984	0.052	1.275	4.013	0.075
KNN	1.346	2.752	0.045	1.526	4.537	0.072	1.294	4.556	0.078
SVR	1.375	2.954	0.051	1.589	5.178	0.087	1.356	4.829	0.085
GPR	1.015	1.043	0.032	1.084	1.625	0.026	1.021	1.417	0.019
AE	0.941	0.863	0.029	0.987	1.356	0.019	0.892	1.249	0.012
DAL	0.982	0.915	0.034	1.042	1.435	0.021	0.908	1.185	0.014
Deep AQ (Multi-AP)	0.562	0.645	0.026	0.793	1.175	0.015	0.642	0.921	0.008

To claim the significance of the results, this research performs a paired t-test between the 5 folds for the Deep AQ (Multi-AP) model. The results, shown in Table 6.6, demonstrate that there are no significant differences between the results. The t-value refers to a ratio to describe the difference within the groups where the larger t-value means more difference and the smaller one means more similarity. Besides, the p-value refers to a ratio to describe the probability that the results from the samples occurring by chance.

Table 6.6: Claiming the significance of results using the paired t-tests.

Methods	$PM_{2.5}$		PM_{10}		O_3	
	t-value	p-value	t-value	p-value	t-value	p-value
Deep AQ (Multi-AP)	3.22	0.001	3.75	0.004	2.84	0.002

To evaluate and exhibit the model stability, this research plots the scatter of Deep AQ (Multi-AP, L + N + M) inference results and ground-truth values of each fold train-validation dataset in (a1-5), (b1-5), (c1-5) for $PM_{2.5}$, PM_{10} and O_3as shown in Figure6.2 (a)-(c). (a1) - (a5) respectively shows the $PM_{2.5}$ results from fold1 to fold5. (b1) - (b5) respectively shows the PM_{10} results from fold1 to fold5. (c1) - (c5) respectively shows the O_3 results from fold1 to fold5. For each group of (a), (b), and (c), it exhibits similar distribution of results and fitting correlation, and the goodness of fit correlation test of each fold all suggests that the Deep AQ framework is capable of producing stable results.

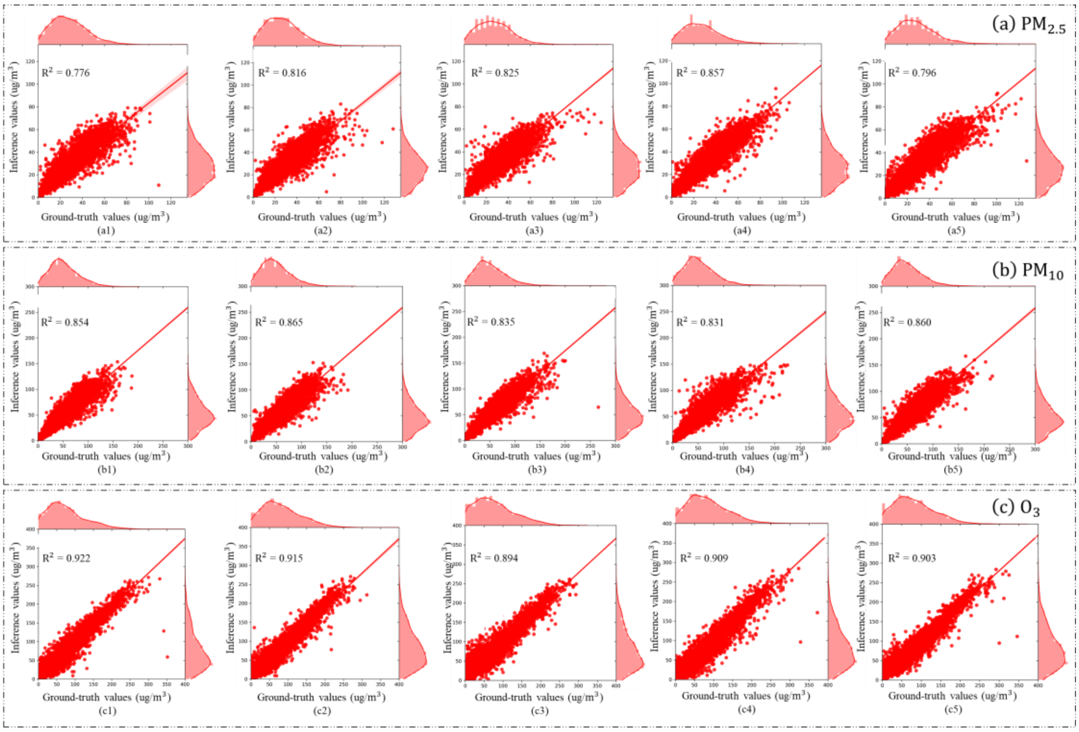


Figure 6.2: Scatter plot of model estimates (Deep AQ (Multi-AP, L + N + M)) and ground-truth values of each fold in (a1-5), (b1-5), (c1-5) for $PM_{2.5}$, PM_{10} and O_3

6.3.4 Performance variability of the Deep AQ framework with various combinations of potential predictors

To investigate the performance variability of Deep AQ framework to the availability to features, this study evaluates the model performance with various feature sets in micro-view (L) inputs and the analysis is similar to that of Analitis et al. (2020). The model performance metrics for these tests are shown in Table 6.6, and it can be concluded that the M + LU + Geo + T + P input data contribute to the best model performance. Note that the M + LU + Geo + T + P input data comprises the local features (micro-view) used elsewhere in this chapter, and that the performance metrics shown in Table 6.7 are identical to those shown for Deep AQ (Multi-AP; L) in Table 6.1.

Table 6.7: The performance of the Deep AQ (Multi-AP) with different input feature variable set from the micro-view.

	Air pollutant species

Variables	$PM_{2.5}$			PM_{10}			O_3		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
M	13.31	24.15	0.65	19.12	22.46	0.71	22.29	25.49	0.79
M + LU + Geo	10.25	23.46	0.69	17.05	21.97	0.73	19.95	25.12	0.81
M + LU + Geo + T	9.08	22.86	0.73	15.28	20.65	0.77	18.74	24.44	0.85
M + LU + Geo + T + P	8.93	21.97	0.75	14.35	19.83	0.79	17.97	23.13	0.86

* M - Meteorological features; LU – Land use features (POIs & AOIs); Geo – Geographical features (Location & DEM); T – Traffic data; P – Public vitality

6.4 Independent LOOCV tests

This study sets 40 rounds of LOOCV tests and deliberately excludes measurements from a monitoring site in each round to set as a testing dataset, achieve AQ inference from other monitoring sites (training dataset), and repeat this procedure until all predictions were done. The actual concentrations reported by the fixed monitoring sites are then used as the ground truth to measure the model performance. Each grid-cell with a fixed monitoring site is tested in this way every hour. Table 6.7 shows the performance range (exceeding outliers) of the Deep AQ (Multi-AP, L + N + M) model showing the predictivity for AQ inference at unsampled sub-areas as shown in Table 6.8.

Table 6.8: The performance of the Deep AQ (L + N + M) in 40 LOOCV tests

Methods	$PM_{2.5}$			PM_{10}			O_3		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2

Deep AQ (Multi-AP)	[5.6,7.5]	[17.4,24.2]	[0.73,0.81]	[9.8,11.4]	[15.8,27.9]	[0.64,0.85]	[13.4,15.1]	[13.9,25.7]	[0.84,0.91]
Deep AQ (GBDT)	[5.8,7.5]	[19.1,23.3]	[0.69,0.79]	[10.2,11.1]	[14.2,20.4]	[0.8,0.86]	[13.1,15.5]	[19.3,24.5]	[0.82.0.93]

Similar land use characteristics of grids have similar geographical variation patterns (Huang, 1992), which means that if the land-use patterns in one place are similar to another place in different locations, then the social activities in two places are analogous, which may appear similar air pollution emission and formation patterns. This research employs the agglomerative clustering approach (Kamvar, 2002) to categorize grid-cells into 7 clusters. Grid-cells in each cluster are classified by similar characteristics on the aspect of land use information, as depicted within Figure 6.3 (a). And Figure 6.3 (b) demonstrates the cluster of the grid-cells at which located fixed monitoring sites according to the land use characteristics, and Table 6.9 describes the characteristic aspects of the 7 clusters.

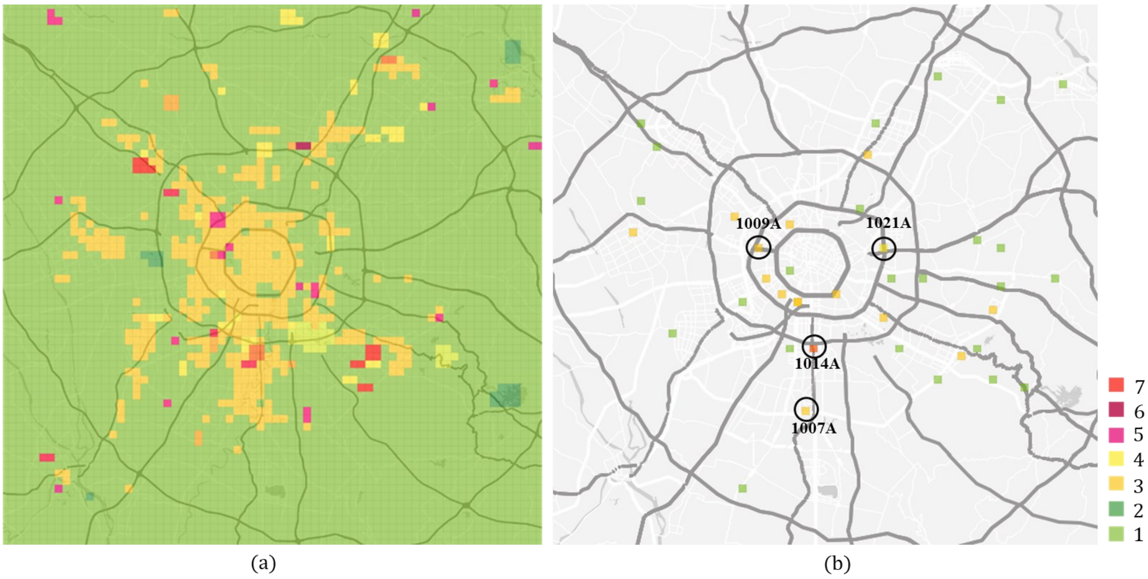


Figure 6.3: Categorized grids within 7 clusters for (a) distribution of clustered grid-cells, (b) examples of representative monitors used for LOOCV tests at clusters

Table 6.9: Characteristic aspects of the clusters.

Cluster	**Characteristic (sequence of the main covered interests)**	**Number of monitors contained in the cluster**

1	Ring roads, commercial areas, buildings, transport infrastructure	24
2	Scenic spots or water areas	0
3	Urban roads, ring roads, intersections, business areas	14
4	Educational, sport areas	1
5	Industries,	0
6	Governments, leisure areas	0
7	Transport infrastructure (gas stations, parking places)	1

Figure 6.4 (a)-(b) shows that LOOCV tests at cluster 7 (e.g., 1014A) and cluster 4 (e.g., 1021A) for $PM_{2.5}$, PM_{10} and O_3 behave poorly, considering no other fixed monitoring site data from the same clustered grids in the training dataset.

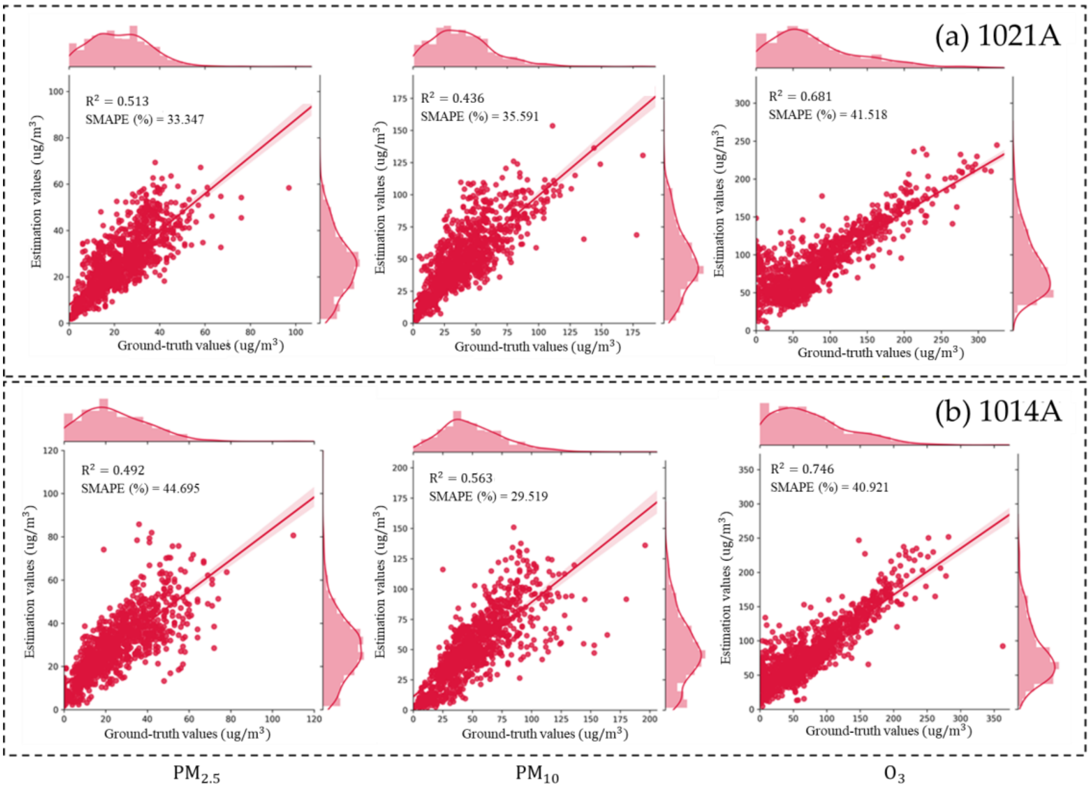


Figure 6.4: Results of LOOCV test on two monitors in (a) Cluster 4 (1021A) and (b) Cluster 7 (1014A) for $PM_{2.5}$, PM_{10} and O_3 (left to right).

Figure 6.5 (a)-(b) shows that LOOCV tests in sub-areas (e.g., 1007A) with rare monitors do not perform as well as (e.g., 1009A) in sub-areas with more monitors for $PM_{2.5}$, PM_{10} and O_3. Thus, the model performs well at the grid-cells which have similar other grid-cells belonging to the same cluster, and the model performs better at the grid-cells where more monitors are located nearby compared to those where monitors are rarely placed in nearby sub-areas.

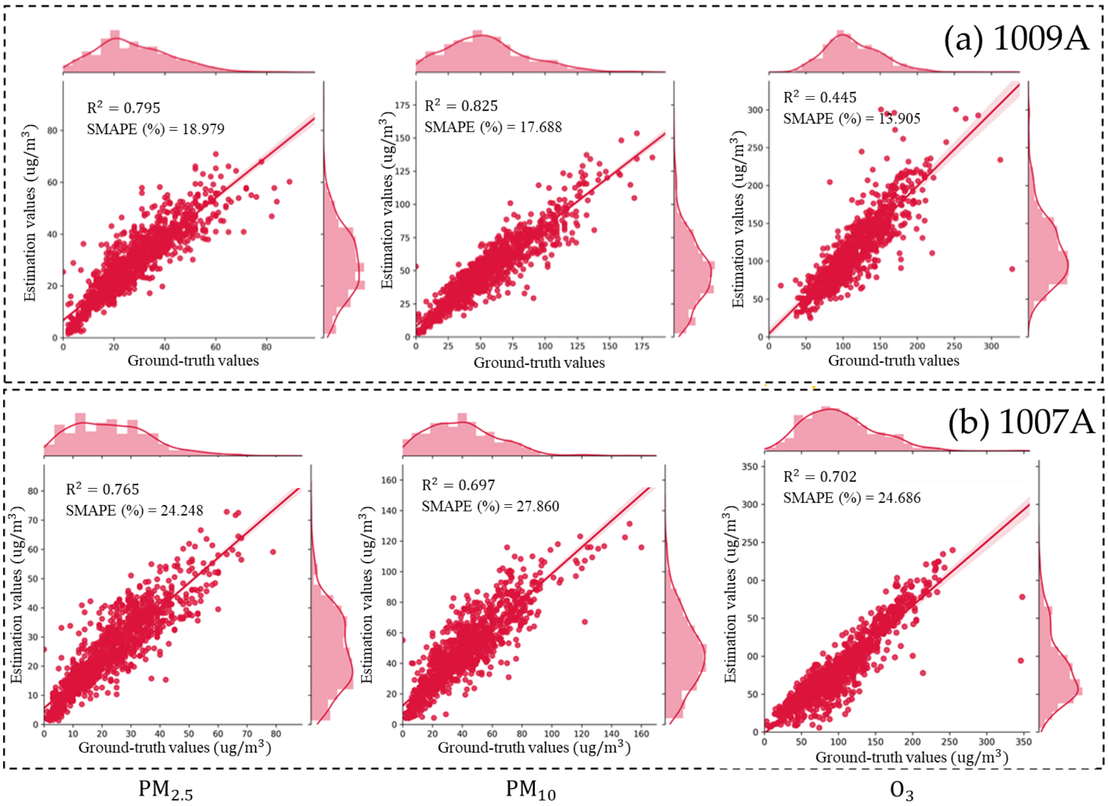


Figure 6.5: Results of LOOCV test on two monitors in Cluster 3 (a) (1009A) and (b) (1007A) for $PM_{2.5}$, PM_{10} and O_3 (left to right).

6.5 Spatial block CV tests

This study sets 5 spatial blocks as shown in Figure 6.6 and makes it possible to allocate blocks to folds in either a random or systematic way.

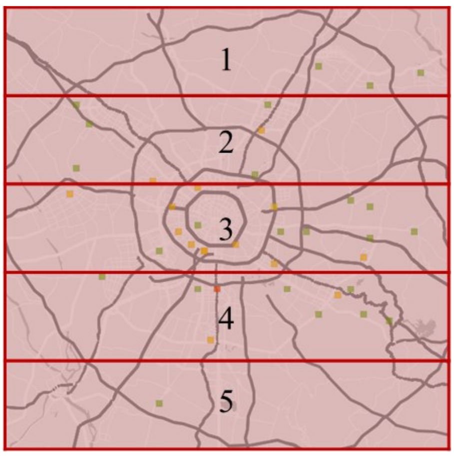


Figure 6.6: Spatial blocks outlined in red. Fold numbers are shown within blocks are fold numbers and illustrate the allocation of blocks to folds.

There are 5 rounds of spatial block CV tests and deliberately excludes measurements from the block in each round sequentially to set as the testing dataset, achieving AQ inference by training on samples from other blocks and repeat this procedure until all the predictions were made. It can be found that the results in Table 6.10 are wholly better than that in Table 6.8 which show that the block CV tests of Deep AQ show more stable performance than LOOCV tests, which can further explain the consistent conclusions of the accuracy and robustness of the model.

Table 6.10: The performance of the Deep AQ (L + N + M) in the 5 block CV tests.

Methods	$PM_{2.5}$			PM_{10}			O_3		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Deep AQ (Multi-AP)	[6.3,7.1]	[18.5,19.3]	[0.77,0.86]	[9.7,10.8]	[15.2,16.1]	[0.82,0.85]	[13.7,14.4]	[19.5,21.2]	[0.89,0.93]
Deep AQ (GBDT)	[6.9,7.4]	[19.2,19.6]	[0.78,0.85]	[10.1,10.7]	[15.5,16.3]	[0.81,0.86]	[14.1,14.9]	[19.6,20.9]	[0.89.0.92]

6.6 Analysis of air pollution maps produced by the Deep AQ framework

The AQ inference mappings of air pollutant species depict the spatio-temporal heterogeneity produced by the Deep AQ framework. Figure 6.6-6.8 presents the pixelwise mappings of $PM_{2.5}$, PM_{10} and O_3 concentrations individually about Chengdu on 13 May 2019 from 0:00-23:00 as demonstrations of the high-resolution inference results of air pollutant species produced by the Deep AQ framework. The trend of concentration of each pollutant can be affirmed by the official data which gives hourly statistical average concentration values, depicted at the top of each figure. The Deep AQ framework offers the capability to produce high-resolution (1 km × 1 km) AQ inference mappings which show spatio-temporal heterogeneity and variations. The hourly mappings of air pollutant species in Figure 6.6 and Figure 6.7 show similar spatio-temporal distributions and variations, which both belong to particulate matters. The $PM_{2.5}$ and PM_{10} concentrations both slowly rise after the unexpected decline between 00:00-04:00, most probably a result of the continuous influence of the north eastern sandstorm passing through Chengdu during the period in that night (Tianqihoubao, 2019), which offers a likely explanation about the spatio-temporal variation details presented by the pixelwise AQ mappings.

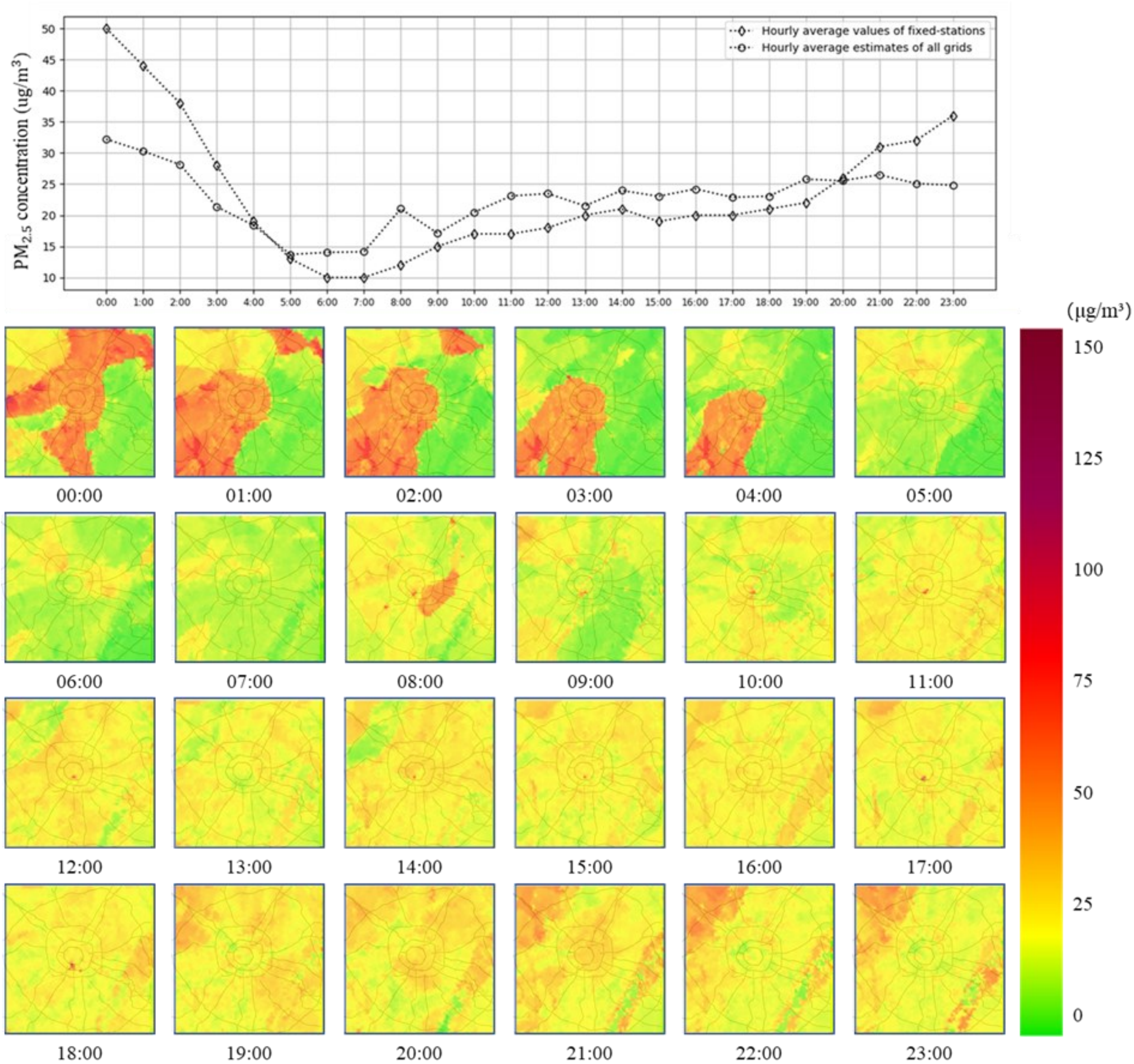


Figure 6.7: $PM_{2.5}$: Top subplot compares the hourly average values of monitors and estimates of all grid-cells. The bottom sub-plot shows inference mappings produced by Deep AQ (Multi-AP, L + N + M) model between 0:00-23:00 on 13 May 2019.

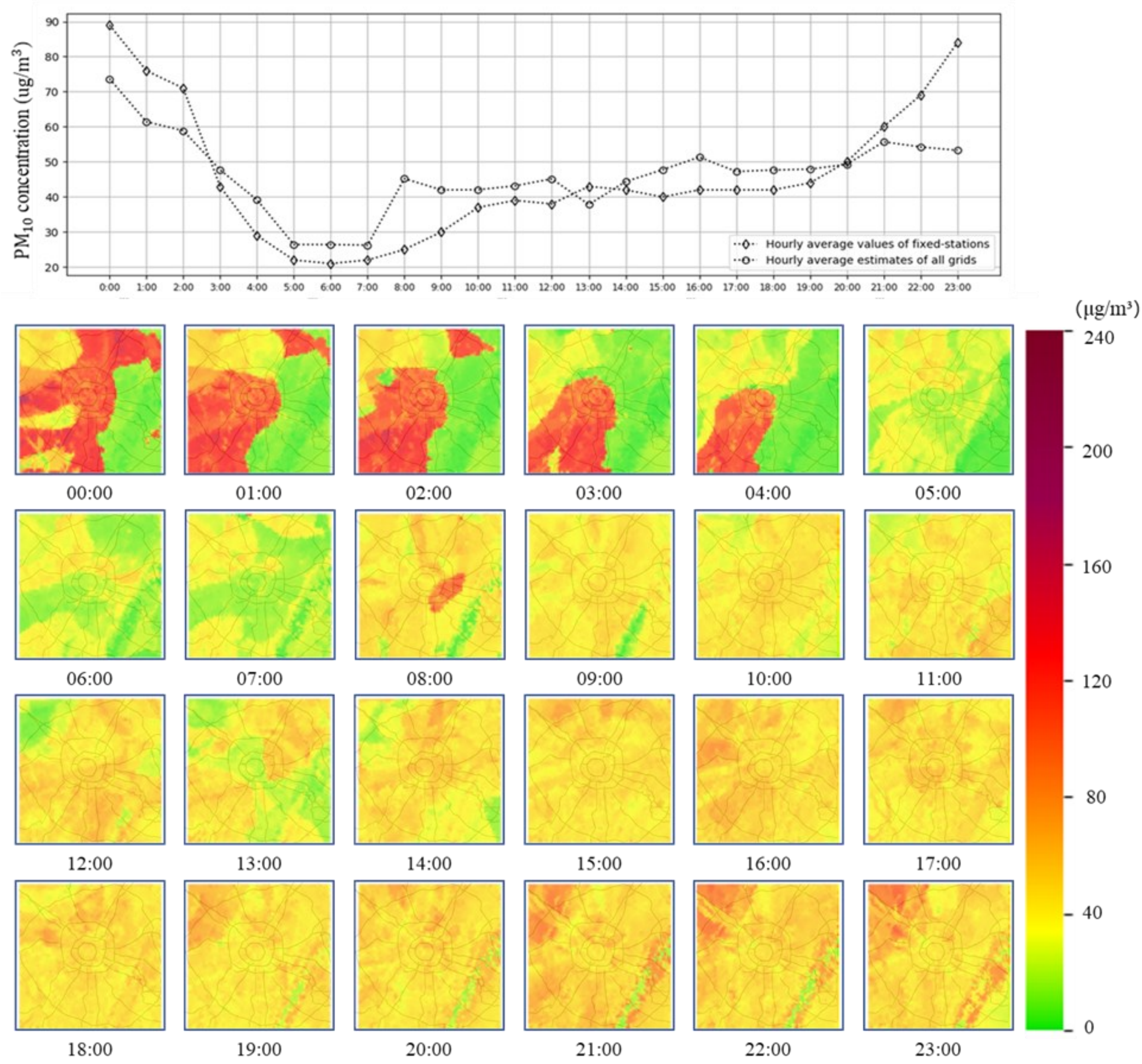


Figure 6.8: PM_{10}: Top subplot compares the hourly average values of monitors and estimates of all grid-cells. The bottom sub-plot shows inference mappings produced by Deep AQ (Multi-AP, L + N + M) model between 0:00-23:00 on 13 May 2019.

However, Figure 6.8 shows that the variation of O_3 does not show the similar spatio-temporal patterns like $PM_{2.5}$ and PM_{10} which means that O_3 is not influenced by the sandstorm and can be considered that O_3 does not belong to particulate matters, which differs from $PM_{2.5}$ and PM_{10} in physical and chemical properties. However, it can be found that the concentration of O_3 rises gradually after 9:00, which is likely caused by the sunlight in the day-time (De, 1989). However, Figure 6.6-6.8 shows that there still exist anomalous values which is considered causing by the limited and sparse AQ sensing spatial coverage.

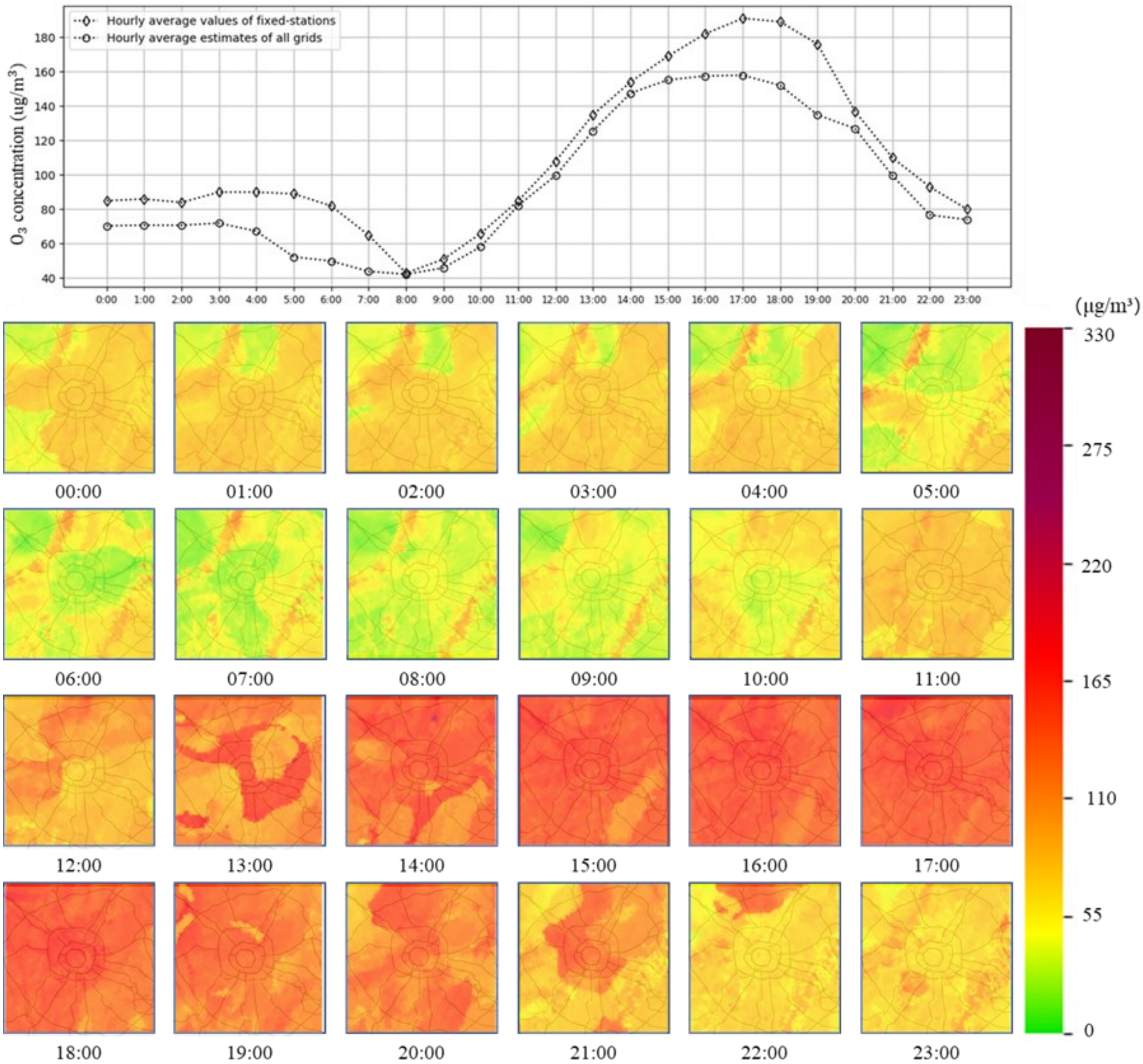


Figure 6.9: O_3: Top subplot compares the hourly average values of monitors and estimates of all grid-cells. The bottom sub-plot shows inference mappings produced by Deep AQ (Multi-AP, L + N + M) model between 0:00-23:00 on 13 May 2019.

The AQ inference mappings produced by the Deep AQ framework offer the capability to express spatio-temporal heterogeneity. However, it can be seen from the aforementioned inference mappings that there still depicts some anomalous values with extremely high or low values comparing to the values in their surrounding places, which is likely to be caused by the overfitting problems considering that the amount of AQ monitored data is rarely minor and unbalanced to such sizeable urban area.

6.7 Conclusions

This chapter reviews the theory basis for the fourth research objective of this thesis, which is to:

Apply the Deep AQ framework to the Chengdu case study based on fixed monitoring sites to evaluate the model performance and applicability.

This chapter mainly investigates the Deep AQ framework to the Chengdu case study with samples from fixed monitoring sites. In this chapter, the Deep AQ framework is investigated based on fixed monitored AQ data in conjunction with relevant urban features, using which to create fine-granular, pixelwise AQ mappings of sizeable urban spaces. The steps of the process include identification of factors that affect emissions, and dispersion of air pollution, including the information related to the regional transportation of air pollution. The Deep AQ framework suits the case study of Chengdu reasonably well (listed in Table 6.1), which outperforms other listed benchmarks. Its outcomes demonstrate how effective it can be to use micro-view, meso-view, and macro-view features.

The conclusions are made as follows:

1. This study achieves high-resolution ($1\,\text{km} \times 1\,\text{km}$, hourly) pixelwise AQ inference in the study area (4,900 km^2) with reasonable performance with Deep AQ framework on fixed AQ data combining with multi-source urban features (e.g., population, traffic, land use information, regional transport of air pollution, etc.).

2. Macro-view features improve the model performance by reducing overfitting, in that inferences made about air quality are influenced to a lesser extent by local factors and provide quantitative evidence for regional transportation of air pollution.

3. Transforming features with meso-view is capable of building up the neighbouring spatio-temporal correlations and improve the model performance.

4. The Deep AQ framework is capable of reaching reasonable accuracy with less than 1% of urban AQ sensing coverage based on the HOCV, LOOCV, and spatial block CV tests.

5. The standard deviation of model performance on different validation datasets is compared, and the goodness-of-fits of the scatter plots are estimated, which all suggests that the Deep AQ framework is capable of producing stable and robust results.

6. Performance variability of the Deep AQ framework with various potential predictors are compared, which shows that the combination of all feature sets outperforms other combinations of reduced feature sets.

7. The AQ inference mappings produced by the Deep AQ framework can express the spatio-temporal heterogeneity although there still exist anomalous values that are likely caused by overfitting problems because of the limited and unbalanced data.

The AQ inference mappings produced by the Deep AQ framework are capable of expressing the spatio-temporal heterogeneity. However, it can be seen from the aforementioned AQ inference mappings that there still depicts some anomalous values with extreme high or low values comparing to the values in their surrounding places which is likely to be caused by the overfitting problems considering that the amount of AQ monitored data are rarely small and unbalanced to such large urban area. Chapter 7 will investigate the Deep AQ framework with reference to data from fixed and mobile sensing in the Beijing case study.

Chapter 7

Predicting AQ in Beijing using fixed and mobile sensors

7.1 Introduction

Chapter 6 investigates the Deep AQ framework to the Chengdu case study based on fixed monitoring sites. In this chapter, it will be investigated using fixed and mobile sensors to achieve high-resolution (1 km × 1 km) pixelwise AQ inference for pollution map-recovery and evolution in the Beijing case study. With the development of low-cost sensors and data mining techniques, an increasing number of studies have developed models based on mobile measurements (Hankey & Marshall, 2015; Shi et al., 2016; Apte et al., 2017; Lim et al., 2019; Wang et al., 2020). Hankey and Marshall (2015) use more than 85 hours of data gathered from a cycling platform in Minneapolis to construct moderately well-fitting LUR models incorporating particle size, black carbon, and $PM_{2.5}$. Others have used vehicle-based mobile measurements to estimate street-level variations of $PM_{2.5}$ and PM_{10} in downtown Hong Kong (Shi et al., 2016). Meanwhile, Apte et al. (2017) analyse results gathered by a Google Street View vehicle whose AQ sensors sampled each road within a 30km^2 area of Oakland multiple times. Using this data, they simulate the distribution of urban air pollution, giving more spatially precise results (4-5 orders of magnitude) than existing central-site ambient monitoring. Lim et al. (2019) propose LUR models incorporating ML methods of examining street-level $PM_{2.5}$ concentrations in Seoul to identify the pollution hotspots. Wang et al. (2020) explore the limitations of LUR models in addition to how ML-based models (ANN and XGBOOST) construct correlations between AQ and predicting

---

The partial content in this chapter has been published in the IEEE Internet of Things Journal (J. Song, K. Han and M. Stettler, "Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing",** ***IEEE Internet of Things Journal**, DOI: 10.1109/JIOT.2020.3041047).**

factors by unveiling the black-box nature of ML with a mobile sampling campaign in Toronto for fine particulate matter and black carbon. However, such inference results are restricted to the street level, with limited temporal granularity and spatial resolution (daily, weekly, monthly). These studies have advanced data-driven tools for achieving AQ inference based on ubiquitous sensing. However, there is still a lack of a systematic framework that is capable of producing high-resolution AQ inference (e.g., 1 km × 1 km, hourly) to recover the pollution map based on irregular AQ observations with limited spatio-temporal coverage while ensuring the robustness and accuracy of the model by leveraging multi-source urban features.

Beijing is the capital of China which locates approximately between 39.26N ~ 41.03N latitudes and 115.25E ~ 117.30E longitudes. 28 fixed monitoring sites are set in Beijing and they release hourly average concentration of air pollutant species such as $PM_{2.5}$, PM_{10}, O_3, CO, NO_2, and SO_2 (AQ data are accessed from official sources). This project deploys 15 taxies mounted low-cost sensors ($PM_{2.5}$ and PM_{10}) for mobile sampling campaigns to survey the atmosphere environment across the city. Similar to the fixed monitors, these low-cost mobile sensors collect the AQ measurements but with a different recording frequency, i.e., every 10s along their routes. The study period is from 19 June 2018 to 16 July 2018, and there are 650,814 records measured from low-cost mobile sensors during this period. The number of records vary among individual sensors because of the difference in routes chosen by the taxis t which the sensors are attached. Table 7.1 lists the statistics of the measured values from both fixed and mobile sources.

Table 7.1: Statistics of measured values from both fixed and mobile sources.

Variable	Unit	Range	Mean	St. Dev.
$PM_{2.5}$	ug/m^3	[2,413]	69.89	46.97
PM_{10}	ug/m^3	[2,864]	77.74	51.2
Temp	°C	[19.21,55.29]	32.59	5.29
Relative humidity	%	[0.61,97.88]	46.98	20.43

Pressure	hPa	[963.27,1008.49]	997.86	3.94
Water Vapor	hPa	[8.85,34.44]	23.00	5.45
Precipitation	mm	[0,31.77]	0.21	1.33
Wind Speed	m/s	[0,6.94]	1.68	0.81

*Unit, range, mean, and standard deviation values from 19/06/2018-16/07/2018

The objectives in this chapter are highlighted as follows:

1) Apply the Deep AQ framework using fixed and mobile AQ sensing combining with multi-source urban features (e.g., population, traffic, land use information) and external background information (e.g., regional transport of air pollution) to achieve pixelwise (1 km × 1 km, hourly) AQ inference in the selected study area in Beijing (3,500 km^2).
2) Design the correction experiments for the field calibration after the laboratory calibration and online field calibration, and correct the mobile sensing data to reduce discrepancies and gaps with fixed monitored data.
3) Evaluate the performance of the Deep AQ framework by comparing with other benchmarks based on HOCV tests.
4) Evaluate the performance improvement with meso-view and macro-view features in the Deep AQ framework.
5) Depict the heterogeneity of AQ inference and analyse the pollution maps to provide evidence-based support to inform sustainable urban management.
6) Analyse the performance of the Deep AQ framework with corrected or uncorrected mobile data and the mobile data's influence on model performance.
7) Analyse sensitivity with data noise variance to validate the robustness of the model.
8) Investigate the variability of the model performance to the availability to various spatio-temporal coverage by mobile sampling campaigns.

9) Evaluate the Deep AQ framework's predictivity to achieve AQ inference at fully unsampled grid-cells by conducting independent LOOCV tests.

10) Evaluate the Deep AQ framework's predictivity to achieve AQ inference at independent block by deploying the spatial block CV tests.

11) Analyse feature interpretation about potential pollution contributors to achieve relative attribution of pollution in Beijing to provide evidence-based support for sustainable urban management.

7.2 Correction of mobile sensing data

The mobile sensors used in this study are first calibrated in a lab environment before deployment to ensure providing reliable readings (this step is finished by the cooperated research institute). Here the correction of mobile sensing data refers to the re-calibration process applied after adopting the online field calibration strategy by OFCLP mentioned in section 3.3.1.2. Considering the variability of each low-cost mobile sensor, the mobile sampling box equipped on each probe vehicle is periodically (weekly) replaced by another sampling box placed at the adjacent reference monitoring site recently. The raw mobile data include timestamps, location (coordinates), $PM_{2.5}$ and PM_{10} concentration, temperature, and humidity. These time-series data are denoised using wavelet transform (Jing & Lin, 2012) to reduce temporal variations and eliminate extreme outliers. The resulting spatio-temporal data are then aggregated into spatial grid-cells and time intervals via arithmetic average with quartile method. The mobile and fixed sensors have distinct sensing environments; that is, the mobile sensors are deployed on the street level while the latter usually on higher grounds or buildings (3~5m). In the correction process, the between-instrument variability is not accounted for during the re-calibration processes, as measurements from different mobile sensors are merged because both procedures require adequate sample sets. Moreover, meteorological parameters, such as temperature and humidity, are considered in the multi-variate regression model. To reconcile the mobile data, this research further corrects the mobile label responses using fixed observations as the reference instruments (Saukh et al., 2017, Maag et al., 2017) in the overlapped grid-cell units and it is necessary to ensure that both (fixed and mobile) share the time-space from time to time. This is achieved by using multivariate regression involving the location (grid s), time (hourly interval t), temperature ($temp$) and relative humidity (RH). Specifically, the multivariate linear regression is expressed as Equation 7.1:

$$y_f = \beta_0 + \beta_1 \cdot y_m + \beta_2 \cdot s + \beta_3 \cdot t + \beta_4 \cdot temp + \beta_5 \cdot RH \tag{7.1}$$

where y_f and y_m denote fixed and mobile AQ label responses, which are collected from the same grid s at the same time interval t. In other words, only those mobile label responses sharing the same location (grid-cell) with fixed monitors can be used for the multi-variate regression. The total number of valid mobile label responses used in the regression is 1,348 for $PM_{2.5}$ and 1,472 for PM_{10}. The regression result is visualized in Figure 7.1, with a goodness-of-fit coefficient R^2 = 0.70 ($PM_{2.5}$) and R^2 = 0.74 (PM_{10}).

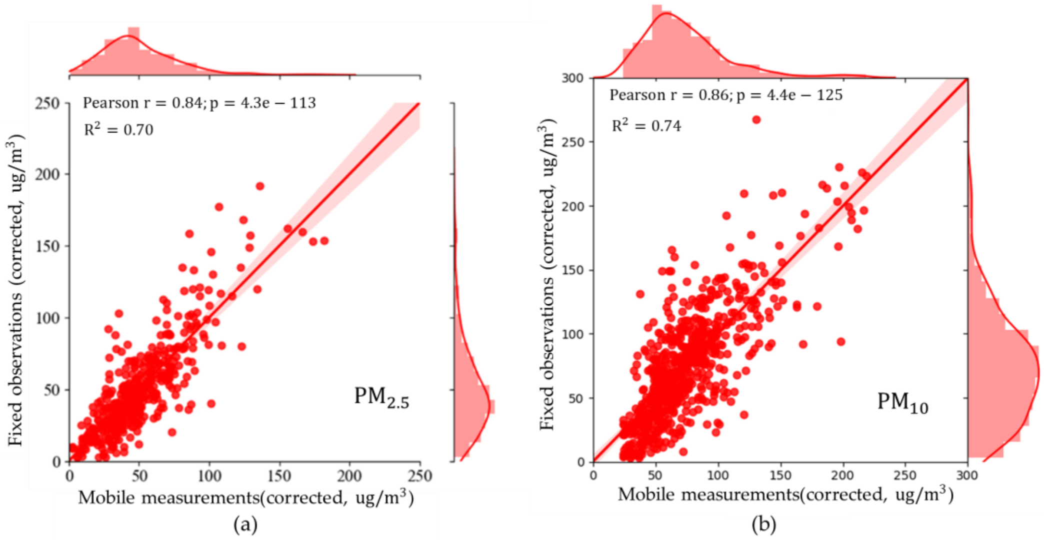


Figure 7.1: Correction of mobile label responses: scatter plots of fixed and corrected mobile label responses.

Let β_i^* ($i = 0, \ldots, 5$) be the coefficients obtained from the multi-variate linear regression, then for any mobile data $\hat{y}_m$ not sharing the same grid-cell unit with any fixed monitors, its value is corrected as Equation 7.2:

$$\bar{y}_m = \beta_0^* + \beta_1^* \cdot \hat{y}_m + \beta_2^* \cdot s + \beta_3^* \cdot t + \beta_4^* \cdot temp + \beta_5^* \cdot RH \tag{7.2}$$

where $s, t, temp, RH$ are associated with the label response $\hat{y}_m$.

7.3 Results and discussions

In the Beijing case study, a total of 50,086 grid-cell units are either labelled as fixed or mobile sensing, of which 20% (10,017 units) are used for the independent validation dataset. For each 1-hour period, the research evenly samples 20% of the labelled spatial grids as a validation dataset. The same evaluation metrics (RMSE, SMAPE, and R^2) are applied to evaluate the model performance.

7.3.1 Performance of benchmarks

This chapter employs the same benchmarks in section 6.2 to evaluate the model performance in terms of their predictivity based on samples from both fixed and mobile sensing. The model settings are listed in Appendix V Table I. Table 7.2 compares the results from these benchmarks combining with micro-view, meso-view, and macro-view generalized features. For model inference in this chapter, all the grid-unit-based mobile label responses are corrected.

Table 7.2: The performance of various benchmarks to achieve AQ inference.

Methods	Views	Air pollutant species					
		$PM_{2.5}$			PM_{10}		
		RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Mean	/	30.43	26.52	0.538	30.12	22.99	0.240
IDW	/	20.09	23.34	0.803	26.99	19.96	0.661
Ordinary Kriging	/	20.17	27.21	0.769	28.14	20.57	0.647
LUR	L	33.93	31.31	0.485	30.09	26.14	0.562
SVR	L	40.40	41.16	0.269	32.56	26.57	0.483
KNN	L	24.18	26.28	0.738	29.14	23.82	0.575
GPR	L	19.83	24.54	0.811	25.77	18.98	0.703
AE	L	17.19	20.47	0.840	23.51	17.95	0.712
DAL	L	15.61	18.33	0.861	21.14	16.07	0.746
Deep AQ	L	16.53	17.85	0.879	18.92	14.89	0.801
	L+N	16.26	17.67	0.883	18.64	14.77	0.806

(Net)	L+M	15.65	17.05	0.891	18.18	14.42	0.816
	L+N+M	15.54	16.99	0.893	17.76	14.24	0.822
Deep AQ (Multi AP)	L	15.89	16.94	0.896	18.31	14.57	0.809
	L+N	15.31	16.72	0.901	17.69	14.42	0.813
	L+M	15.23	16.69	0.902	17.44	14.36	0.818
	L+N+M	15.01	16.58	0.907	17.03	14.18	0.824
Deep AQ (DT)	L	21.37	21.55	0.795	25.64	17.40	0.700
	L+N	21.33	21.38	0.794	24.87	17.06	0.718
	L+M	20.63	20.93	0.810	25.21	17.14	0.710
	L+N+M	20.65	20.85	0.809	24.51	16.93	0.726
Deep AQ (RF)	L	14.99	16.61	0.899	19.99	14.21	0.810
	L+N	14.83	16.53	0.902	19.66	14.03	0.823
	L+M	14.55	16.05	0.905	19.02	14.07	0.834
	L+N+M	14.38	15.89	0.907	18.74	13.75	0.852
Deep AQ (GBDT)	L	14.49	16.48	0.897	20.27	14.10	0.813
	L+N	14.46	15.88	0.902	19.98	13.92	0.825
	L+M	13.92	15.52	0.905	17.37	14.01	0.841
	L+N+M	13.24	14.72	0.909	16.84	13.66	0.863

*L = local features (micro-view); N = neighbouring correlations (meso-view); M = background air pollution information (macro-view).

Table 7.2 shows the result which are based on the average of 5-fold-cross validation, from which the superior performance of the Deep AQ framework over other benchmarks is evident. Besides, there also demonstrates a significant improvement on the accuracy of the macro-view features regardless of the applied surrogate models through comparison of 'L' with 'L + M', or 'L + N' with 'L + N + M'. Furthermore, performance of collaboration with meso-view features 'L + N' (or 'L + N + M') is superior to micro-view features 'L' (or 'L + M'), which implies the importance to construct the spatio-temporal correlations of various urban features among neighbouring grid-cell units and this research achieves it by using convolution techniques. For the tree-based model, both the GBDT-based and RF-based models outperform the DT-based model, indicating that boosting and bagging offer the capability to improve the model's performance. Besides, the GBDT-based model can achieve the best performance, and performs slightly better than RF-based model. Figure 7.2 shows that the Deep AQ (GBDT, L + N + M) model outperforms other benchmarks.

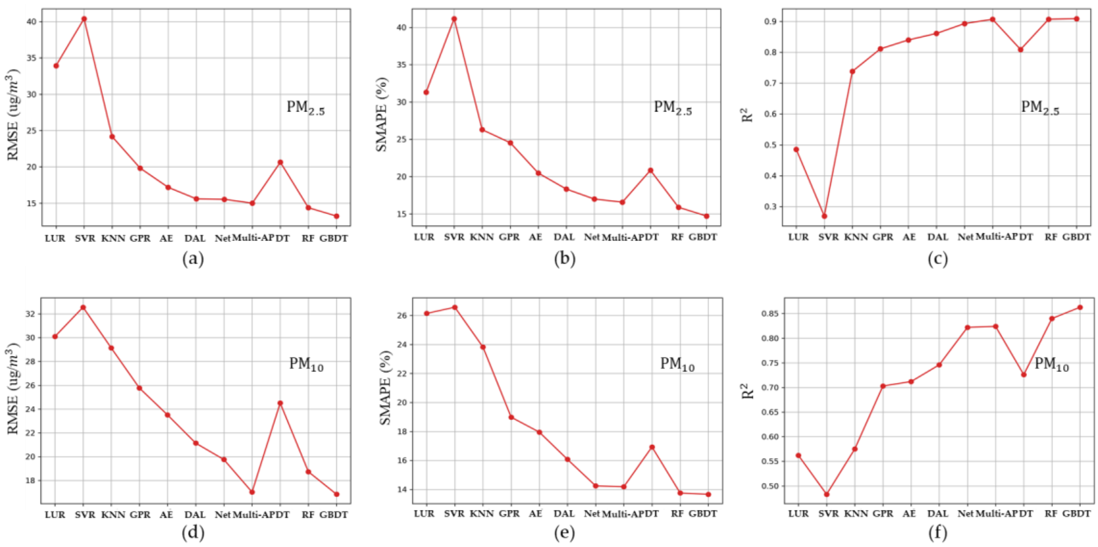


Figure 7.2: The performance comparison of approaches for $PM_{2.5}$ (a)RMSE, (b)SMAPE, (c) R^2 and PM_{10} (d)RMSE, (e)SMAPE, (f) R^2

Figure 7.3 shows the performance of the Deep AQ framework with multi feature views and the 'L + N + M' combined feature input outperforms than 'L', 'L + N' and 'L + M'.

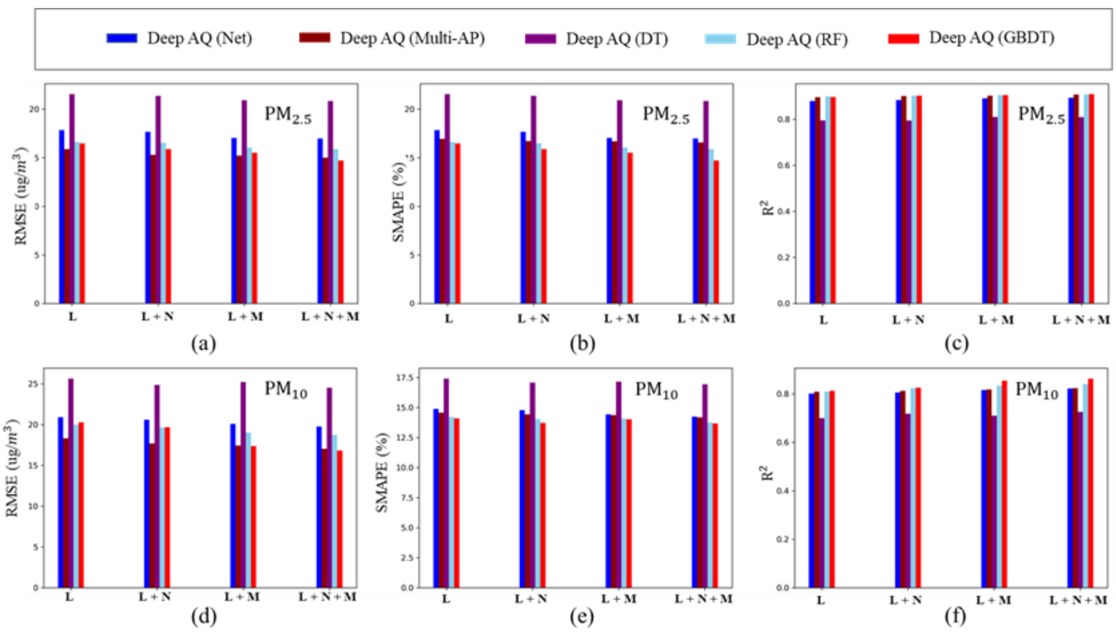


Figure 7.3: The performance of the Deep AQ framework with multi feature views for $PM_{2.5}$ (a)RMSE, (b)SMAPE, (c) R^2 and PM_{10} (d)RMSE, (e)SMAPE, (f) R^2

The importance of the macro-view and meso-view features are also apparent when using 'L + N + M' instead of 'L' which shows performance improvement of the Deep AQ framework as listed in Table 7.3.

Table 7.3: The performance improvement using 'L + N + M' instead of 'L' features for the Deep AQ framework

Methods	$PM_{2.5}$			PM_{10}		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Deep AQ (Multi-AP)	-0.88	-0.36	0.011	-1.28	-0.39	0.015
Deep AQ (GBDT)	-1.25	-1.76	0.012	-3.43	-0.44	0.05

7.3.2 Stability of the Deep AQ framework

To investigate the stability of the Deep AQ framework with different validation datasets, this research performs stability analysis to compare the proposed model with other benchmarks in the aspect of standard deviation with different evaluation metrics for $PM_{2.5}$ and PM_{10}. The results show that the Deep AQ framework (GBDT, L + N + M) is more consistent than other benchmarks listed in Table 7.4.

Table 7.4: Standard deviation of the evaluation metrics across the 5-fold validation for $PM_{2.5}$ and PM_{10}.

Methods	Standard deviation of metrics					
	$PM_{2.5}$			PM_{10}		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Mean	3.153	2.819	0.056	2.485	3.237	0.069
Kriging	1.227	1.843	0.048	1.357	2.816	0.053

IDW	1.438	1.225	0.051	1.172	2.614	0.047
LUR	1.782	2.335	0.049	1.752	2.831	0.052
KNN	1.775	2.653	0.069	1.975	3.682	0.065
SVR	2.482	2.142	0.063	2.917	3.954	0.074
GPR	0.824	0.761	0.037	0.822	1.374	0.031
AE	0.853	0.798	0.042	0.834	1.471	0.026
DAL	0.871	0.818	0.034	1.128	1.144	0.023
Deep AQ (GBDT)	0.415	0.517	0.019	0.539	0.752	0.017

The evaluation metrics's standard deviation across the 5-fold validation suggests that the Deep AQ framework is capable of producing stable and robust results.

To claim the significance of the results, this research performs a paired t-test between the 5 folds for the Deep AQ (Multi-AP) model. The results, shown in Table 7.5, demonstrate that there are no significant differences between the results. The t-value refers to a ratio to describe the difference within the groups where the larger t-value means more difference and the smaller one means more similarity. Besides, the p-value refers to a ratio to describe the probability that the results from the samples occurring by chance.

Table 7.5: Claiming the significance of results using the paired t-test.

Methods	$PM_{2.5}$		PM_{10}	
	t-value	p-value	t-value	p-value
Deep AQ (GBDT)	2.674	0.003	2.519	0.005

7.4 Analysis of $PM_{2.5}$ pollution maps produced by the Deep AQ framework

Figure 7.4 presents the pixelwise AQ inference mappings about $PM_{2.5}$ on 19 June 2018 between the hours of 00:00 to 23:00 in Beijing produced by the Deep AQ (GBDT, L + N + M) model. The hourly $PM_{2.5}$ concentration shows that concentration rises after 02:00, which is likely caused by the dust brought along by the east wind from the east area outside of Beijing (Appendix IV Figure I shows the wind direction and speed) from midnight to morning (03:00-09:00) (Tianqihoubao, 2018). Appendix IV Figure II shows the hourly variations of humidity (relative) on 19 June 2018, which provide possible reasons for explaining the dissipation of $PM_{2.5}$ after 08:00 and the accumulation of $PM_{2.5}$ between 16:00-23:00. The dissipation of $PM_{2.5}$ beginning at 8:00 is partially due to the humidity decreasing throughout 08:00-15:00. Moreover, the concentrations of $PM_{2.5}$ increase as humidity increases throughout 16:00-23:00, which echoes prior research evidence that high humidity tends to facilitate the formation of particulate matters (Feng et al., 2014), or that the low-cost sensors are cross-sensitive (Talampas & Low, 2012, Mei et al., 2020). This study also conducts an experiment to compare the AQ inference mapping with or without mobile samples. Appendix IV Figure III shows the mappings without mobile samples which depict similar spatio-temporal distribution and variation patterns comparing with Figure 7.4. Thus, the comparing results demonstrate that the increasing concentrations of $PM_{2.5}$ are mainly caused by the increasing humidity but not influenced too much by the cross-sensitivity of low-cost sensors. Nevertheless, the AQ inference mapping shown in Figure 7.4 is a complex manifestation of a combination of different static and dynamic features, as well as higher-order features, the above is only tentative interpretation using aforementioned features for the ease of conceptualization. The hourly average values published on official channels (as shown on top of Figure 7.4) has confirmed such a trend.

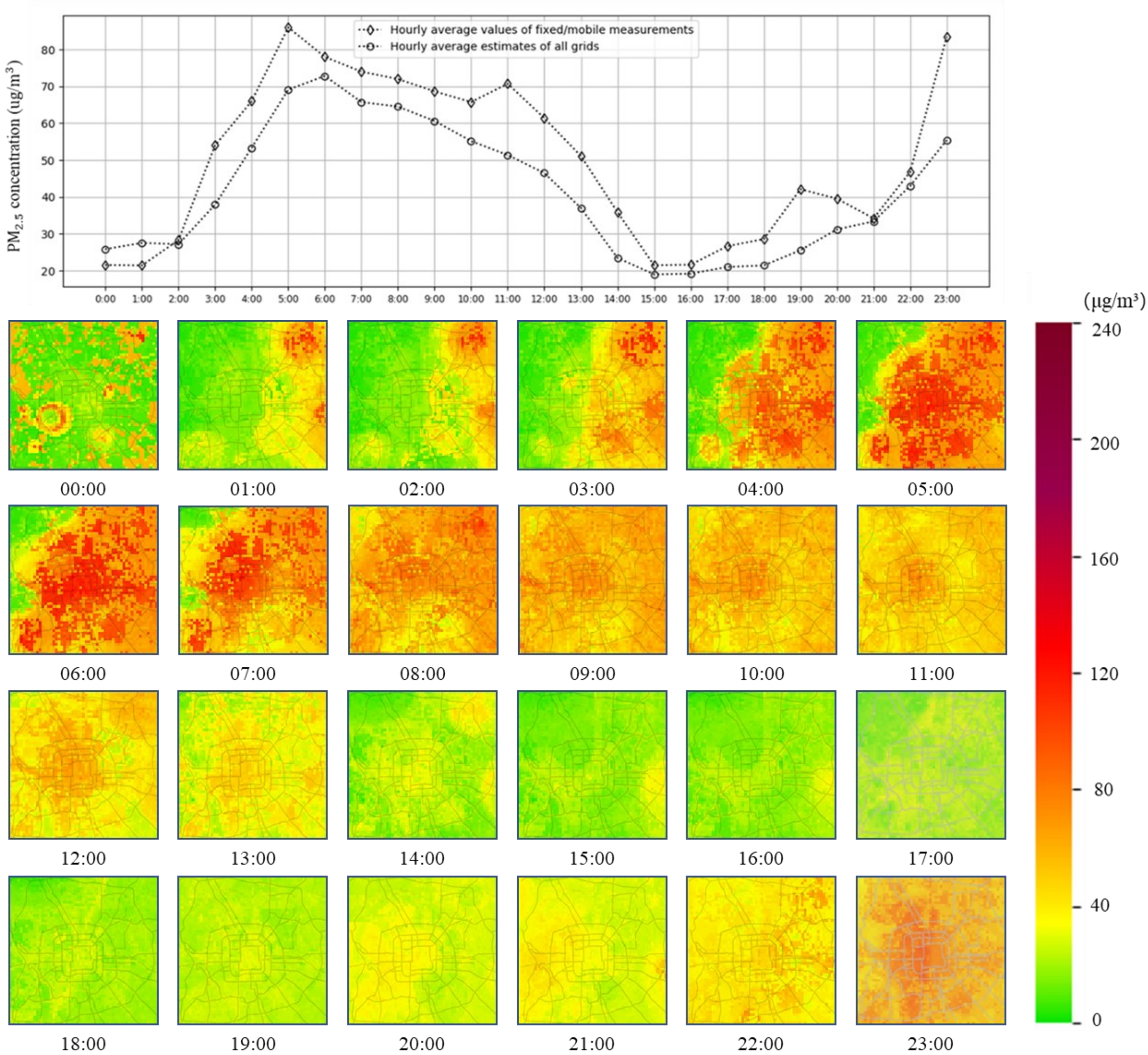


Figure 7.4: $PM_{2.5}$: Top subplot compares the fixed AQ monitoring network's hourly average values and estimates of all grids. The bottom subplot shows inference mappings produced by the Deep AQ (GBDT, L + N + M) model between 0:00-23:00 on 19 June 2018.

Figure 7.5 presents the pixelwise AQ inference mappings about $PM_{2.5}$ on 30 June 2018 between 00:00 to 23:00 in Beijing produced by the Deep AQ (GBDT, L + N + M) model. On this day, the $PM_{2.5}$ concentrations are generally lower than other days, and the hourly air pollution mappings show relatively higher concentrations between 05:00-19:00 than other periods. The time windows of 07:00-9:00, 10:00-14:00, and 16:00-18:00 are the rush hours in the morning, noon, and evening, during all of which it can be observed that the inner urban areas suffered from much higher concentrations than outer urban areas during these hours which is likely caused by traffic activities. Nevertheless, the inference mappings shown in Figure 7.5 are a complex manifestation of a combination of different static and dynamic features, as well as higher-order features. Moreover, the above is only a tentative interpretation using just one single feature for ease of conceptualization. Based on the hourly average values published on official channels as shown on the top subplot in Figure 7.5, this trend is confirmed.

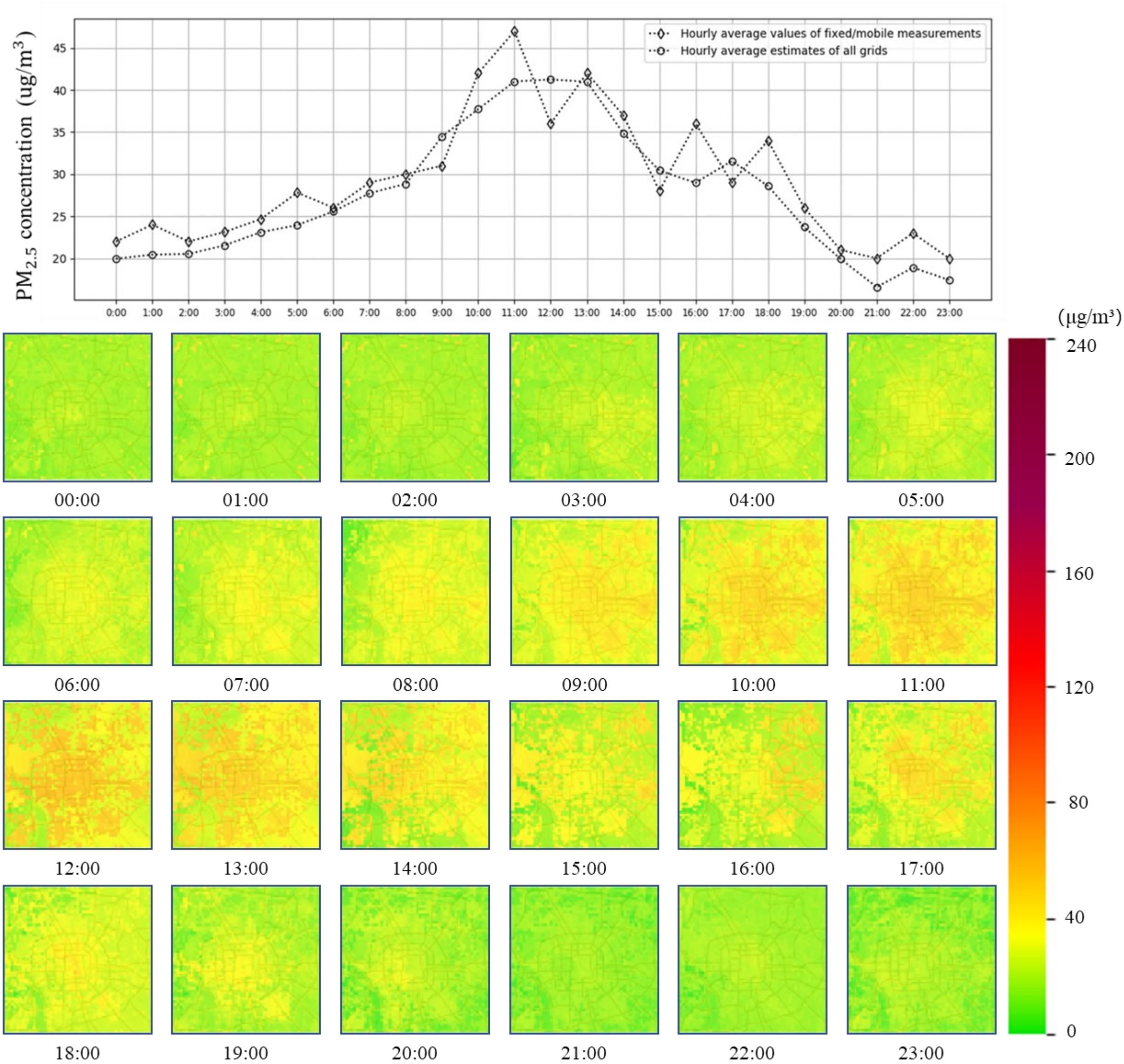


Figure 7.5: $PM_{2.5}$: Top subplot compares the hourly average values of fixed AQ monitoring network and average estimates of all grid-cells. The bottom subplot shows inference mappings produced by the Deep AQ (GBDT, L + N + M) model between 0:00-23:00 on 30 June 2018.

Figure 7.6 presents the pixelwise $PM_{2.5}$ inference mappings in Beijing on 5 July 2018 from 0:00 - 23:00 produced by the Deep AQ (GBDT, L + N + M) model. On this day, the hourly pollution maps show relatively higher between 0:00-16:00 than other time periods. Besides, it can be observed that the city road suffered from much higher $PM_{2.5}$ concentrations in most time slices. Appendix IV Figure IV shows the hourly variations of humidity(relative) on 5 July 2018. A relatively high and steady concentrations can be identified in the hourly pollution maps for most of the day as shown in Figure 7.5, and it then sees a sudden drop between the three-hour period between 16:00 to 19:00. This was probably the result of the on-going thundershower and Appendix IV Figure IV provides a reasonable explanation of the sudden change (the elbow of humidity time-series). Moreover, the dissipation of $PM_{2.5}$ beginning at 16:00 is partially due to continuous thundershower around the same time (Tianqihoubao,2018). Nevertheless, the inference result shown in Figure 7.6 is a complex manifestation of a combination of different static and dynamic features, as well as higher-order features. Moreover, the above is only a tentative interpretation using just one single feature for ease of conceptualization. Based on the hourly average values published on official channels as shown on the top of Figure 7.6, such variation trend can be confirmed.

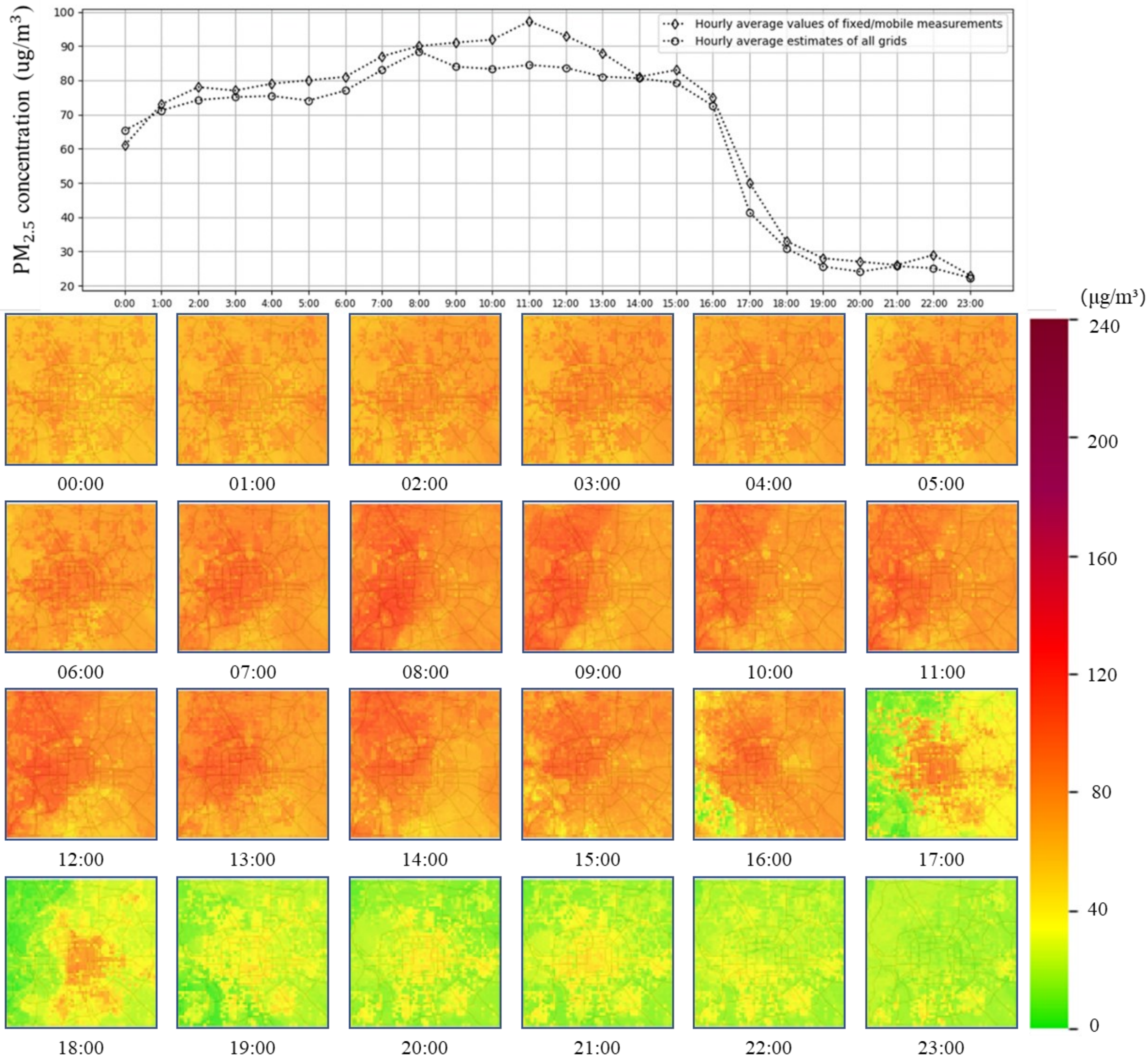


Figure 7.6: $PM_{2.5}$: Top subplot compares the hourly average values of fixed AQ monitoring network and average estimates of all grid-cells. The bottom subplot shows inference mappings produced by the Deep AQ (GBDT, L + N + M) model between 0:00 - 23:00 on 5 July 2018.

Figure 7.7 presents the pixelwise $PM_{2.5}$ inference mappings in Beijing on 6 July 2018 from 00:00-23:00 produced by the Deep AQ (GBDT, L + N + M) model. The pollution concentration between 00:00-04:00 is relatively lower than other periods influenced by the thundershower and hail, and the pollution concentration accumulates after 05:00 when the traffic activities begin to increase. Considering the day is Friday, traffic patterns are unusual, where many people go out for their weekly social activities. It can be clearly observed that the pollution concentration continuously increases between 06:00-23:00. Furthermore, the city road experiences much higher $PM_{2.5}$ concentrations than other urban land use areas in most time slices which is also likely caused by traffic activities, especially between 16:00-23:00 after most citizens finish their day-time work and seek leisure activities. Besides, most people drive home or pick others up after their social activities after 20:00, so the increasing traffic activities are likely resulting in the formulation of higher dust. Nevertheless, the inference result shown in Figure 7.7 is a complex manifestation of a combination of different static and dynamic features, as well as higher-order features. Moreover, the above is only a tentative interpretation using the aforementioned features for ease of conceptualization. Based on the hourly average values published on official channels, as shown on the top subplot in Figure 7.7, the trend of variation is confirmed.

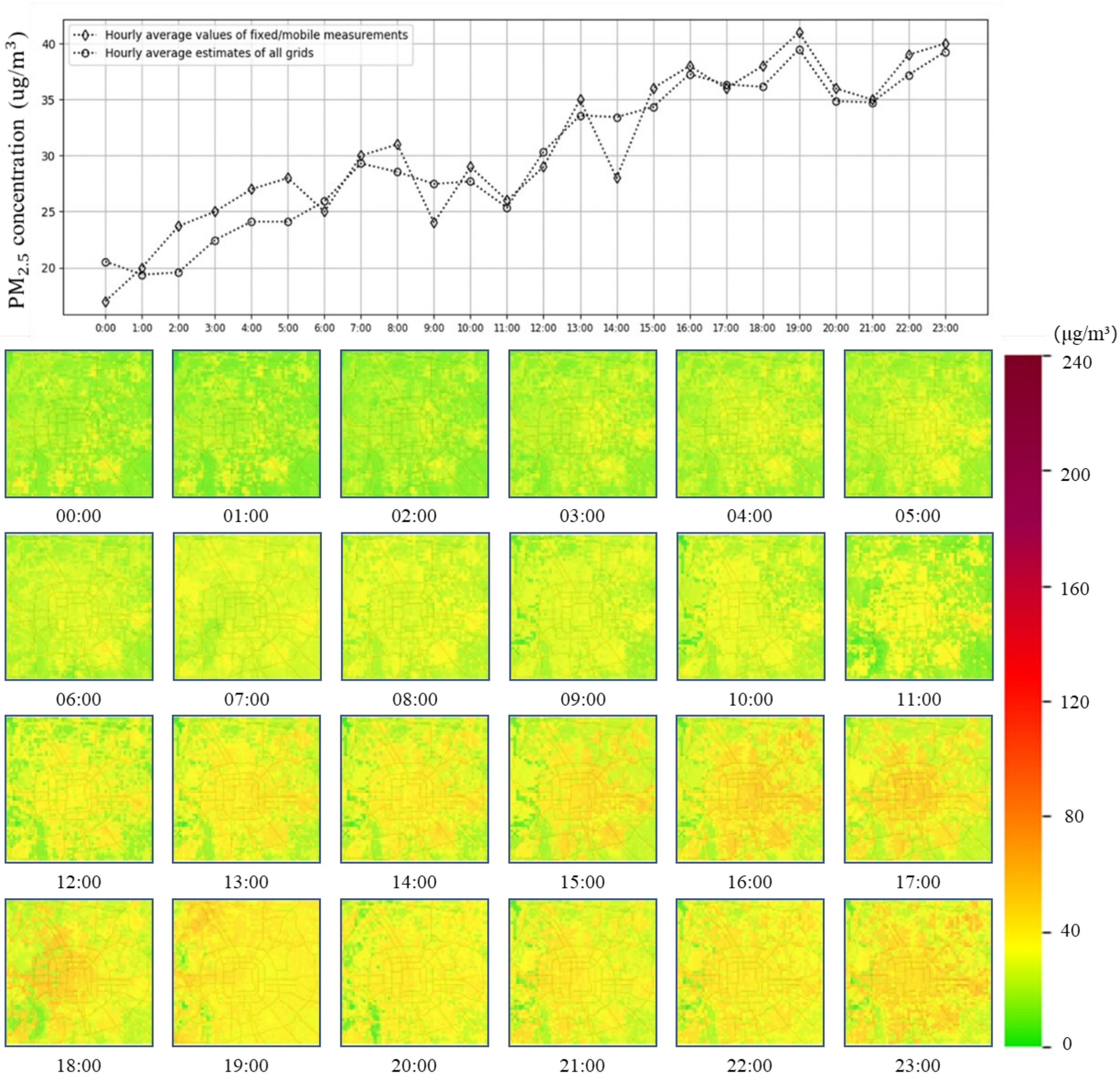


Figure 7.7: $PM_{2.5}$: Top subplot compares the hourly average values of fixed AQ monitoring network and average estimates of all grid-cells. The bottom subplot shows inference mappings produced by the Deep AQ (GBDT, L + N + M) model between 0:00-23:00 on 6 July 2018.

7.5 Impact of mobile AQ samples

Due to different surveying environments, there can be discernible discrepancies in the measurements of air pollutant collected from fixed monitoring stations and mobile sensors. There are two ways to properly address this problem: either to correct the mobile AQ label responses following section 7.2 or to distinguish fixed and mobile AQ label responses by using a categorical feature. A majority of the fixed stations are positioned in high grounds or secluded areas while also serving as reference instruments for the correction of mobile label responses. Therefore, there is a tendency that the correction process will scale down the mobile values rather frequently, which is confirmed in Figure 7.8. Table 7.4 compares the performance of the Deep AQ framework with corrected and uncorrected mobile AQ label responses.

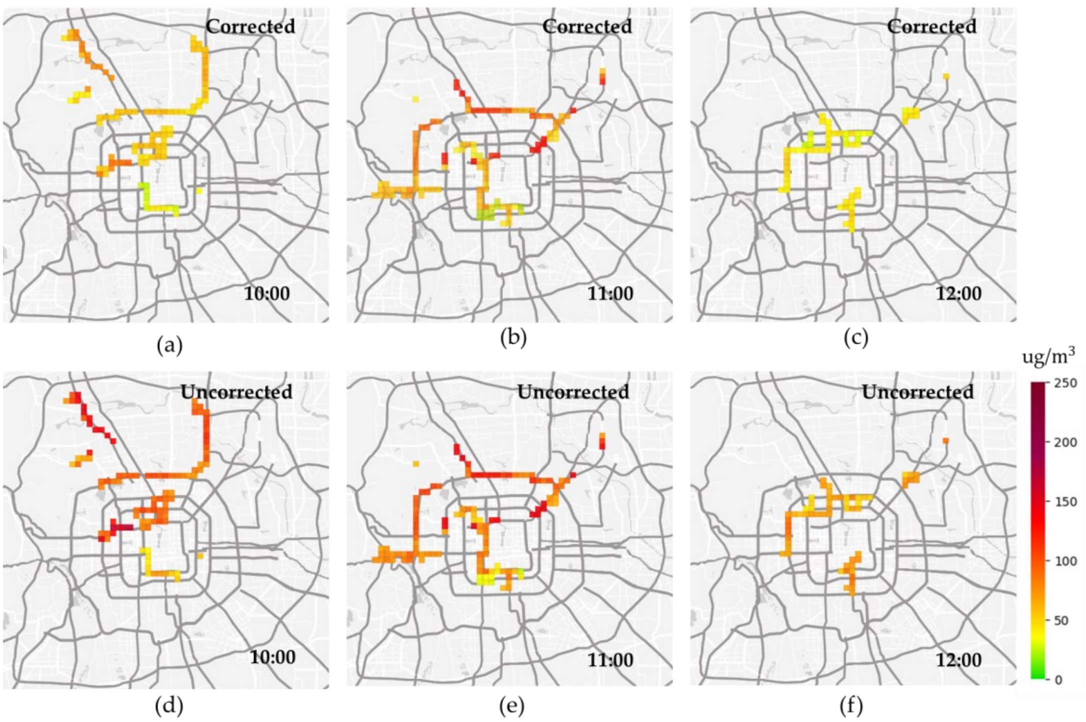


Figure 7.8: Time variant sensing coverage by corrected for (a)-(c) and uncorrected for (d)-(f) mobile AQ label responses between 10:00-12:00 on 20 June 2018.

Furthermore, due to their collection environment (including heavily polluted urban streets), exceptionally high values may be recorded in the uncorrected mobile AQ label responses. However, it is not uncommon that these extreme values may be ignored by the AQ inference. DFeaST-Tree actually possesses robust mechanisms to prevent overfitting, but it is at the cost of higher RMSE and SMAPE with lower goodness of fit (R^2) when differentiating fixed mobile AQ label responses from the uncorrected ones (as shown in Table 7.6).

Table 7.6: The performance of the Deep AQ (GBDT, L + N + M) model based on uncorrected or corrected mobile AQ label responses.

Air pollutant species	Mobile AQ samples	RMSE (ug/m^3)	SMAPE (%)	R^2
$PM_{2.5}$	Uncorrected	14.28	15.25	0.846
	Corrected	13.24	14.72	0.909
PM_{10}	Uncorrected	17.02	14.19	0.854
	Corrected	16.84	13.66	0.863

Logically speaking, one way to improve the model performance of mobile AQ label responses is to lower their discrepancies through the reconciliation of fixed and mobile AQ label responses. However, findings of this research suggests that the representativeness of the corrected mobile AQ label responses are no longer the case for the ground-level concentration of air pollutant species. Instead, they are artifacts created specifically to close the gap between fixed AQ label responses. For a further investigation of the impacts of the interaction between fixed and (corrected) mobile data, this research summarizes the performances with mobile training and testing data (as shown in Table 7.5). Cases (i)-(ii) are to testify the internal consistencies of fixed and mobile data, while (iii)-(iv) are to evaluate the level of integration of fixed and mobile data. As demonstrated in Table 7.7, if the test dataset is entirely made of fixed sensing data (i, iii) or mobile data (ii, iv), the margin results from (i) with (iii) and (ii) with (iv) show either fixed or mobile data improve the model performance and the mobile data improve more compared to fixed data. Nevertheless, by limiting the test set so that it only consists of fixed sensing data, the Deep AQ framework is further tested whereas (i) fixed sensing data; or (iii) fixed and mobile sensing data make up the entire training set. For case (i), similar to Chapter 6, it is an extension to spatial inference based on merely fixed data (e.g., Zheng et al., 2013; Qi et al., 2018); for case (iii), it supplements the training set with mobile data. By including mobile data, the SMAPE (%) sees a 2% reduction for $PM_{2.5}$ and a 3% reduction for PM_{10}. This is the model can construct a better spatio-temporal correlations and lower overfitting thanks to a spatially diversified survey of urban features and pollutant concentrations. In addition, the outputs are usually of better quality when using collaborative samples which are a mix of fixed and mobile sensing (iii, iv) compared to using single samples (i, ii). By comparing (i) and (v), it is apparent that it is sufficient to use fixed sensing data to infer data of the same kind. However, there can be significant misinterpretation on the margin of mobile data, which serves as proof of the incapability of a fixed-location AQ monitoring in capturing complex urban atmospheric environment due to their spatio-temporal sparsity or surveying environment such as in high grounds or secluded areas. This also implies that there may be substantial noises captured in mobile sensing data induced by their complex surveying environments. In the case of case (vi), given in Table 7.2, its performance metrics falls between (iii) and (iv). This result is logical given the fact that the general performance on fixed and mobile test datasets is the weighted average of (iii) and (iv).

Table 7.7: The performance of the Deep AQ (GBDT, L + N + M) model

Case	Training Set	Test Set	Air pollutant species					
			$PM_{2.5}$			PM_{10}		
			RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
(i)	F	F	15.12	17.76	0.867	21.08	16.87	0.853

(ii)	M	M	12.45	14.60	0.900	14.19	12.19	0.892
(iii)	F + M	F	14.41	15.77	0.898	17.18	13.84	0.859
(iv)	F + M	M	12.13	13.66	0.914	13.97	12.11	0.895
(v)	F	F + M	42.23	46.17	0.419	54.81	47.61	0.453
(vi)	F + M	F + M	13.24	14.72	0.909	16.84	13.66	0.863

*F refers to fixed sensing data; M refers to mobile sensing data

It can be seen that the R^2 is often high combining with mobile ones in training data which is considered that the mobile data are more consistent in sampling campaigns.

7.6 Independent LOOCV tests

This study sets 28 rounds of LOOCV tests and deliberately excludes measurements from a monitoring site in each round to set as the testing dataset, achieving AQ inference by training on other monitoring sites and mobile sensors (training dataset). This procedure is repeated until all the predictions were made. The actual concentrations reported by the monitoring sites are then used as the ground truth values to measure the model performance. Each grid-cell with a fixed monitoring site is tested in this way every hour. Table 7. 8 shows the performance range (exceeding outliers) of the Deep AQ (GBDT, L + N + M) model which shows the model predictivity to infer at unsampled sub-areas.

Table 7.8: The performance of the Deep AQ (L + N + M) in the 28 LOOCV tests.

Methods	$PM_{2.5}$			PM_{10}		
	RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Deep AQ (Multi-AP)	[14.5,17.2]	[17.3,18.9]	[0.84,0.89]	[18.7,20.2]	[14.9,18.2]	[0.77,0.84]

Deep AQ (GBDT)	[12.1,15.7]	[14.2,18.2]	[0.85,0.89]	[17.3,19.6]	[14.3,16.9]	[0.81,0.86]

7.7 Spatial block CV tests

This study sets 5 spatial blocks as shown in Figure 7.9 and makes it possible to allocate blocks to folds in either a random or systematic way.

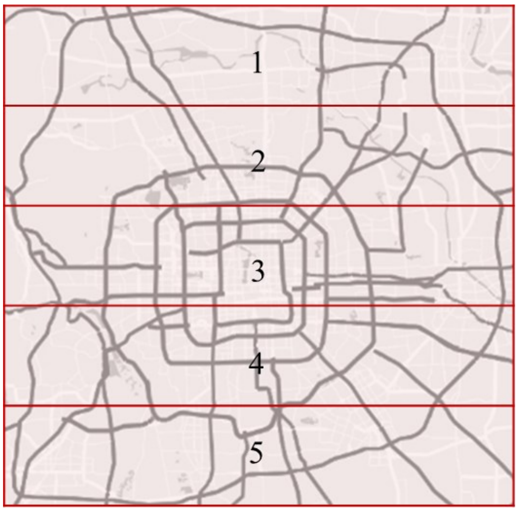


Figure 7.9: Spatial blocks outlined in red. Fold numbers are shown within blocks are fold numbers and illustrate the allocation of blocks to folds.

There are 5 rounds of spatial block CV tests and deliberately excludes measurements from the block in each round sequentially to set as the testing dataset, achieving AQ inference by training on samples from other blocks and repeat this procedure until all the predictions were made. Table 7.9 shows the performance of the Deep AQ in the 5 spatial block CV tests. It can be found that the results in Table 7.9 are wholly better than that in Table 7.8 which show that the block CV tests of Deep AQ show more stable performance than LOOCV tests, which can further explain the consistent conclusions of the accuracy and robustness of the model.

Table 7.9: The performance of the Deep AQ (L + N + M) in the 5 block CV tests.

Methods	$PM_{2.5}$			PM_{10}		
	RMSE	SMAPE (%)	R^2	RMSE	SMAPE (%)	R^2

	(ug/m^3)			(ug/m^3)		
Deep AQ (Multi-AP)	[14.8,15.1]	[16.4,16.7]	[0.89,0.91]	[16.8,17.3]	[14.1,14.3]	[0.81,0.85]
Deep AQ (GBDT)	[13.1,13.4]	[14.5,14.8]	[0.89,0.92]	[16.7,17.1]	[13.5,13.8]	[0.82,0.87]

7.8 Robustness of the model with data noise variance

Mobile sampling campaigns generate a large volume of AQ measurements from low-cost sensors, which contains unexpected measurement noise. The impact of the noise is an important issue to decide the appropriate model structure to be less affected by the measurement noise. To quantify the influence of data noise, sensitivity analysis is conducted by adding various percentages of noises to evaluate the model's robustness to outliers. As shown in Figure 7.10 (a)-(c) and Figure7. 10 (d)-(f) respectively, the results of experiments with various percentages of training noises σ, to compare the model performance for $PM_{2.5}$ and PM_{10} in metrics including RMSE, SMAPE and R^2. In the uncertainty analysis of data noise variance, the noise data are randomly generated according to the setting percentages of the whole training dataset, and the percentages of noise variance are set as $\sigma \in [1\%, 2\%, 3\%, 4\%, 5\%]$ where the data noise are gradually added. For either tree-structured or net-structured Deep AQ framework, the average errors enlarge across all metrics with the adding data noises.

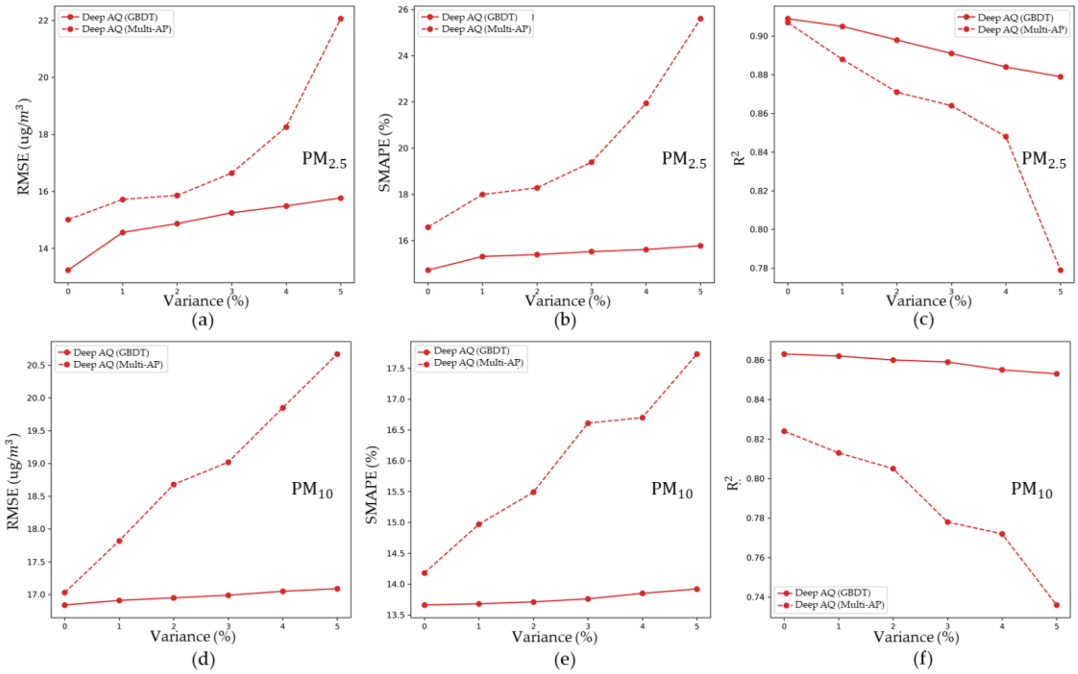


Figure 7.10: Performance of the Deep AQ framework (L + N + M) with data noise variance for $PM_{2.5}$ (a)RMSE, (b)SMAPE, (c) R^2 and PM_{10} (d)RMSE, (e)SMAPE, (f) R^2

As shown in Figure 7. 11, the Deep AQ (GBDT) model reduces the error accumulation compared to the Deep AQ (Multi-AP) model either for $PM_{2.5}$ or PM_{10} . The accumulated error of results from the Deep AQ (Multi-AP) model is much higher than that of the Deep AQ (GBDT) model, especially when under the circumstance where there are 5% extra data noises (as shown in Figure 7). 11 and Appendix III Table I lists the accumulated errors compared to no-adding extra noises. Comparing with the Deep AQ (GBDT) model, the Deep AQ (Multi-AP) model is significantly influenced by the amount of adding noises and performs worse than the Deep AQ (GBDT) model, which shows that the tree-structured are more robust to noises and more applicable.

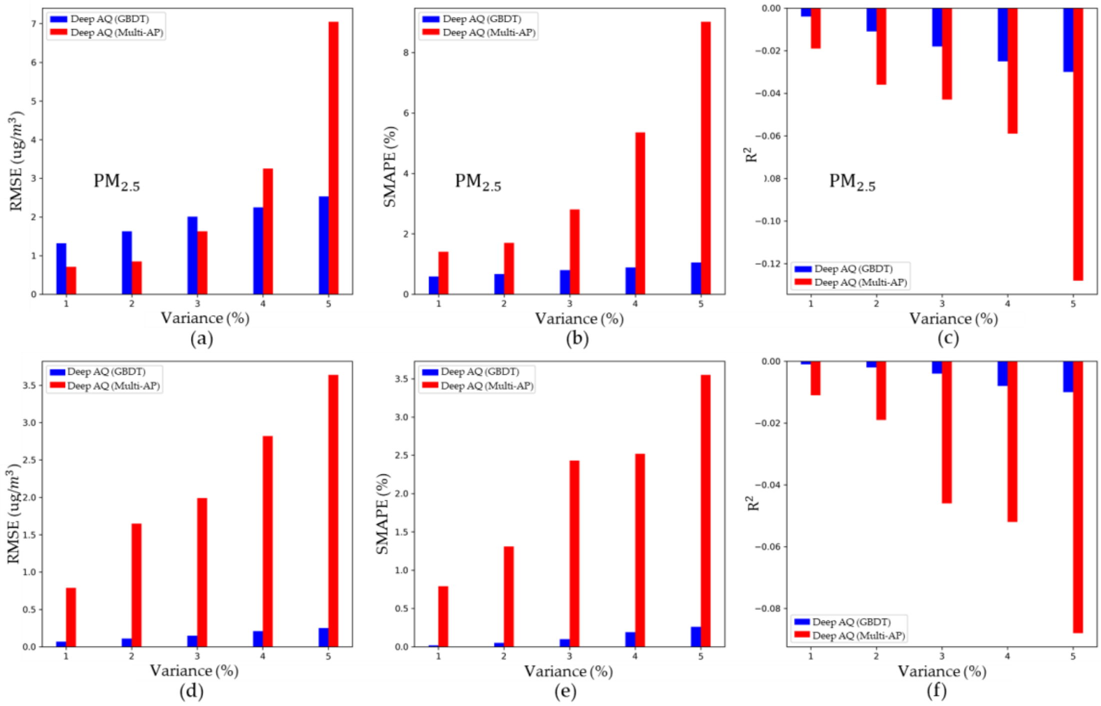


Figure 7.11: Error accumulation of the Deep AQ framework with noise variance for $PM_{2.5}$ (a)RMSE, (b)SMAPE, (c) R^2 and PM_{10} (d)RMSE, (e)SMAPE, (f) R^2

As demonstrated in Figure 7.12 (a) and Figure 7.12 (c) respectively, the points in the two residual plots about $PM_{2.5}$ and PM_{10} from the Deep AQ (GBDT, L + N + M) model spread across the horizontal axis in a random manner. This means that the model and the data are a decent fit. However, Figure 7.12 (b) and Figure 7.12 (d) show that small bias still exists in model fitting as the residuals' sum and mean are not equal to zero.

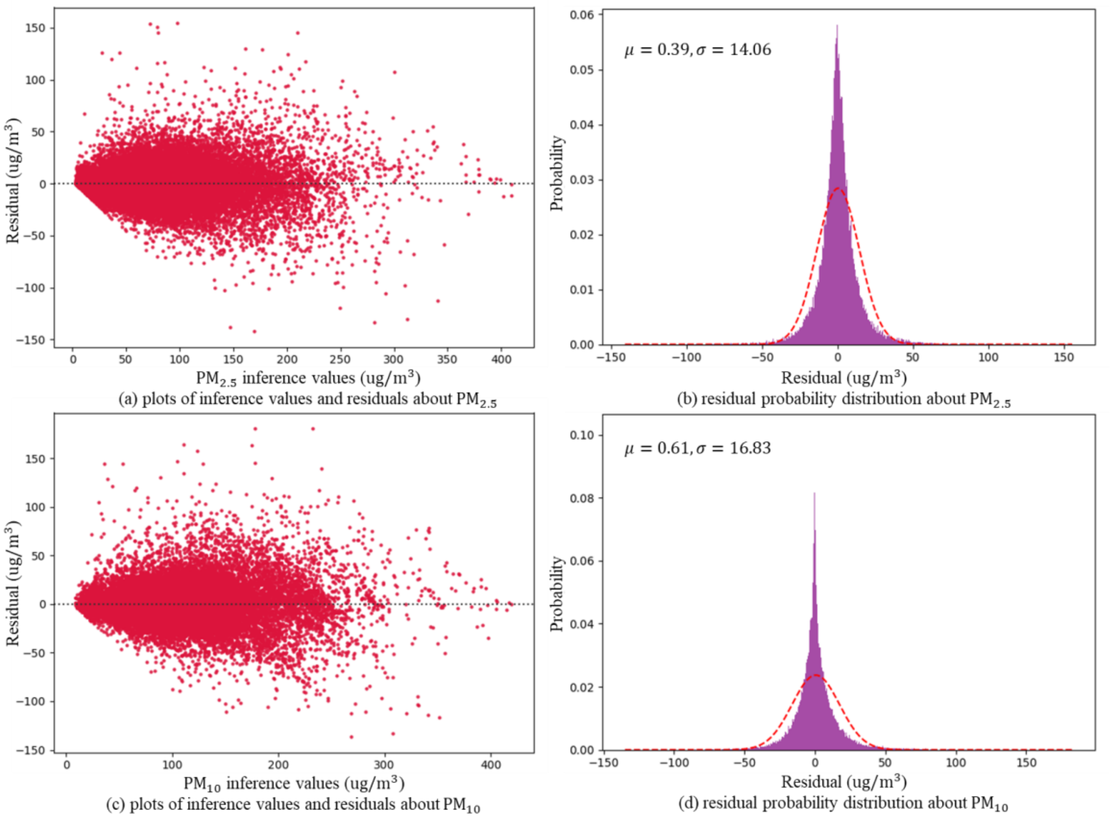


Figure 7.12: (a) plots of inference values and residuals about $PM_{2.5}$; (b) probability distribution of residuals about $PM_{2.5}$; (c) plots of inference values and residuals about PM_{10}; (d) probability distribution of residuals about PM_{10}.

7.9 Impact of mobile sensing spatio-temporal coverage

Within mobile sensing, a main practical aspect is that spatio-temporal sensing coverage is required to achieve AQ inference. In other words, it means the relationship between the accuracy of the proposed model and the quantity of training data the mobile sensing can provide. Taking the Deep AQ (GBDT, L + N + M) model as an example, the research tests its performance on an independent validation dataset of 10,017 units which is a combination of samples collected from fixed and mobile sensing. Starting with only samples from fixed sensing, samples from mobile sensing are then added one by one into the training dataset. The model performance in different metrics is shown in Figure 7.12.

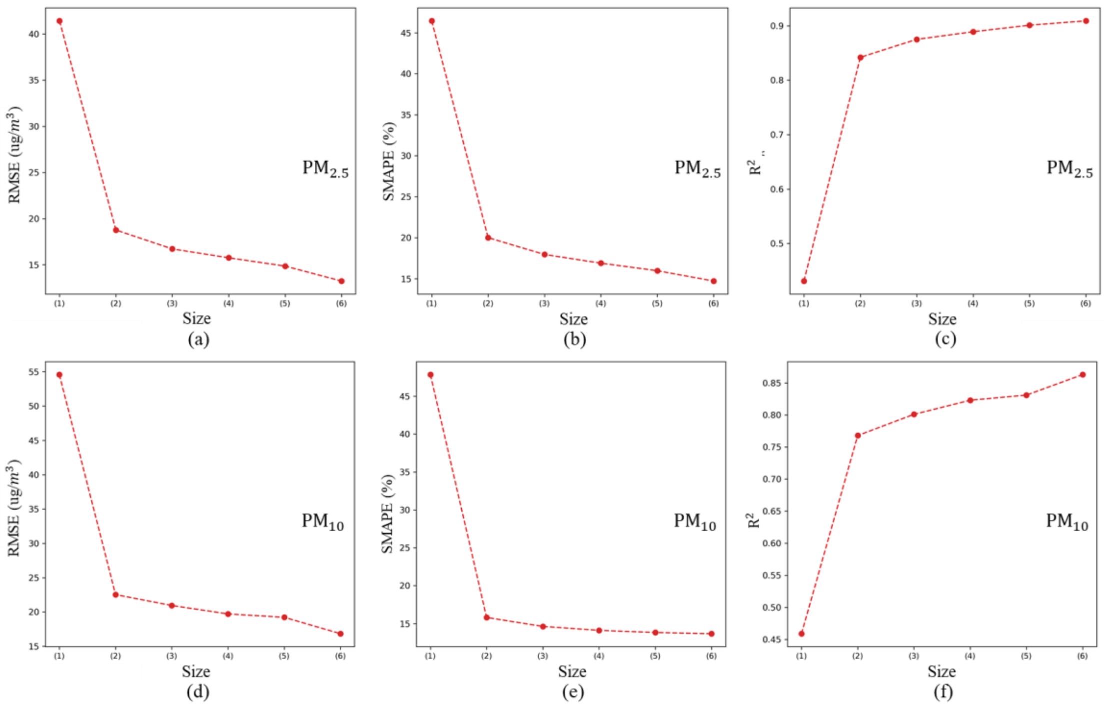


Figure 7.13: Trends of performance metrics with training datasets of different sizes. (1): Fixed data only; (2): Fixed data + 20% mobile data; (3): Fixed data + 40% mobile data; (4): Fixed data + 60% mobile data; (5): Fixed data + 80% mobile data; (6): Fixed data + 100% mobile data.

As demonstrated in Figure 7.12, the accuracy variation of the Deep AQ (GBDT, L + N + M) model in the same validation datasets for $PM_{2.5}$ and PM_{10} with various sizes of the training datasets. In both situations, an identical trend is found: when there are only fixed samples in the training dataset, the errors are high; as soon as the first pieces of mobile samples, despite being a small number, are added into the training dataset, the 'elbow' point appears. When more and more mobile samples are brought into the training dataset, the accuracy of the models increases accordingly. Having said that, with the continuous increment of mobile samples, the margin of improvement sees a gradual decrease. This implies that it does not take a large quantity of mobile samples for the proposed model to achieve its satisfactory AQ inference. As reported in Table 7.2, it is sufficient to achieve such point with the spatial coverage below 5%.

In the next experiment, by analysing model performance on grid-cells that vary in terms of temporal coverage, this study conducts sensitivity analysis to test the model robustness. The analysis is based on the observation as shown in Figure 3.6 that, there is a significant difference in the temporal coverage of grid-cells. Thus, examining the model performance's variability at these grid-cells is essential. The RMSE and SMAPE of the Deep AQ framework at grid-cells whose temporal coverage is under y hrs are shown in Table 7. 10. It appears that as y increases, the performance improves slightly accordingly. This entails that the more the labelling samples, the better fit for the spatial grid-cell units. Nonetheless, the difference of RMSE is lower than 1.39 μg/m^3 and the difference of SMAPE is 1.5% about $PM_{2.5}$; the difference of RMSE is lower than 1.85 μg/m^3 and the difference of SMAPE is 1.29% about PM_{10}. These results demonstrate how the model handles grid-cells with different temporal coverage with robustness.

Table 7.10: The performance of the Deep AQ framework on grids with ≤ y hrs of temporal coverage.

Air pollutant species	y	10	20	30	60	90	120	150	180	210
/	No. of grids	983	1170	1270	1438	1509	1552	1588	1601	1613
$PM_{2.5}$	RMSE (μg/m^3)	15.73	15.43	14.91	14.54	14.46	14.43	14.40	14.35	14.34
	SMAPE (%)	16.48	16.07	15.84	15.37	15.20	15.15	15.09	15.01	14.98
PM_{10}	RMSE (μg/m^3)	18.87	18.41	18.26	18.04	17.83	17.59	17.33	17.21	17.02
	SMAPE (%)	16.04	15.66	15.45	15.19	15.04	14.96	14.92	14.87	14.75

The scatter plots of the validation testing points (five folds combined, over 50,086 points) shown in Figure 7.13 are generated by the Deep AQ (GBDT, L + N + M) model for both $PM_{2.5}$ and PM_{10}. The R^2 suggest that the proposed model is capable of yielding effective results.

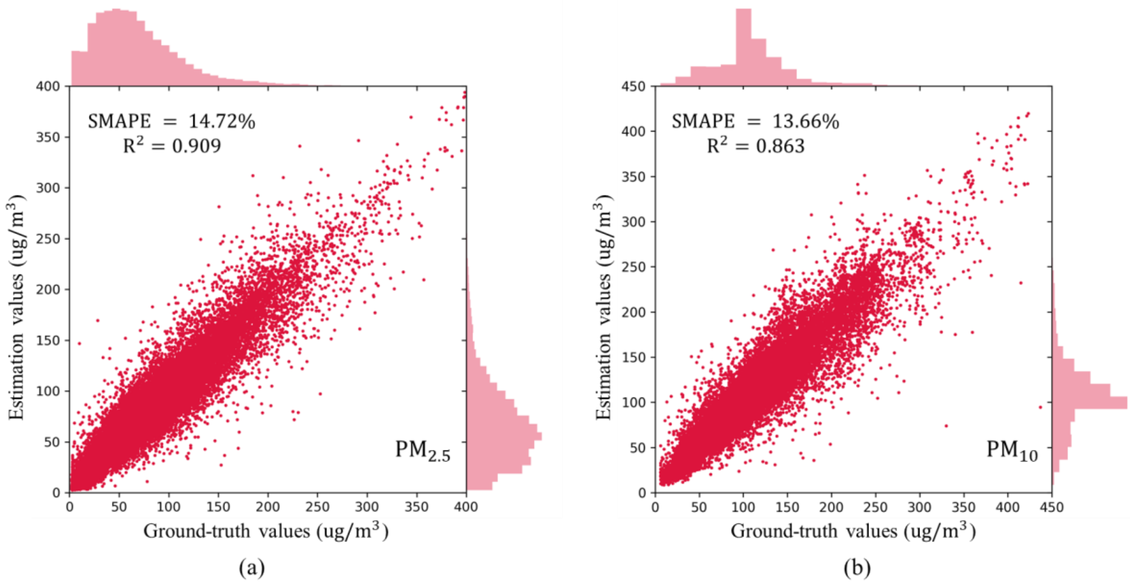


Figure 7.14: Scatter plots of validation testing points in all five folds for (a) $PM_{2.5}$; (b) PM_{10}.

7.10 Feature analysis and interpretation of the AQ inference results

The tree-structured approach is capable of computing weights/importance of input feature variables by investigating each feature's impact on the model output qualitatively (Wang et al., 2020), which is one of its many advantages. The average importance amongst all the decision trees within the model is also referred as the importance of a feature variable. Meanwhile, the importance in a single tree can be found by calculating the total of reductions in the loss function at all the split points for which the feature is responsible. That is to say, by providing a better fit between the observations and outputs of the model, the importance of a feature variable can measure its contribution to loss reduction. This research aims to provide insights regarding potential causes and relevant factors of particulate matter concentrations by investigate the weights of features generated by the Deep AQ (GBDT, L + N + M) model to analyse their relative importance. Through offering a better fit between model outputs and observations, the weight of an input feature can evaluate its contribution to the loss reduction.

Figure 7.14 (a) demonstrates the relative importance of main feature categories according to the Deep AQ framework results. The top four features are meteorology, POI and AOI, macro-feature (i.e., domestic transport of air pollution), and traffic conditions respectively. Accumulated research evidence, such as Zheng et al. (2013), has been suggesting that, in explaining particulate matter, meteorological parameters, traffic, and POI play Important roles. Indeed, the main causes of particulate matter formation have long been identified as local restaurants, auto services, residential buildings and the likes which are represented by POI, AOI. Meanwhile, traffic is found to be the main source of fugitive dust and secondary reactions. However, compared with some local features, authors such as Zheng et al., (2013) do not take into consideration the pollutants' regional transport and their weights. As demonstrated in Figure 7.14, it is basically a matter of the meteorological parameters that is accountable for the particulate matter concentrations in summer time. Indeed, regarding particulate matter concentrations, wind speed, pressure, water vapor pressure, temperature, and humidity are all identified as highly relevant. These observations are logical given that with stronger winds, pollutants are spread more quickly, and the correlation between the higher pressure and higher wind speed in summer. It has also been found that high humidity contributes to a more favourable environment for the formation of particulate matter. In their study, Qi et al., (2018) argue that pressure and humidity are irrelevant features for particulate matter concentrations. Their different observation can be due to the fact that their study took place in winter, meaning the changes in pressure and humidity are relatively smaller. In Figure 7.14 (c)-(d), macro-view features and the particulate matter's regional transport are illustrated. The regional transport of particulate matter can be traced through the relative weights of individual fixed monitoring sites outside the study area. Two sites in the Northwest, namely Dingling and Changping, recorded the heaviest weights; another two cites from the South and Southwest, namely Yufa and Liuli River, have the second highest recorded weights. The recorders are collected between 19 June to 16 July 2018, showing the two major directions towards which the particulate matter transport. Figure 7.14, which is the wind rose diagram, offers valuable data that demonstrate the particulate matter's physical transport. This finding also partially echoes the findings of Feng et al., (2014) and Qi et al. (2018), where the main source of industrial pollutants that affects Beijing is found to have come from Hebei Province, which is to the South of Beijing.

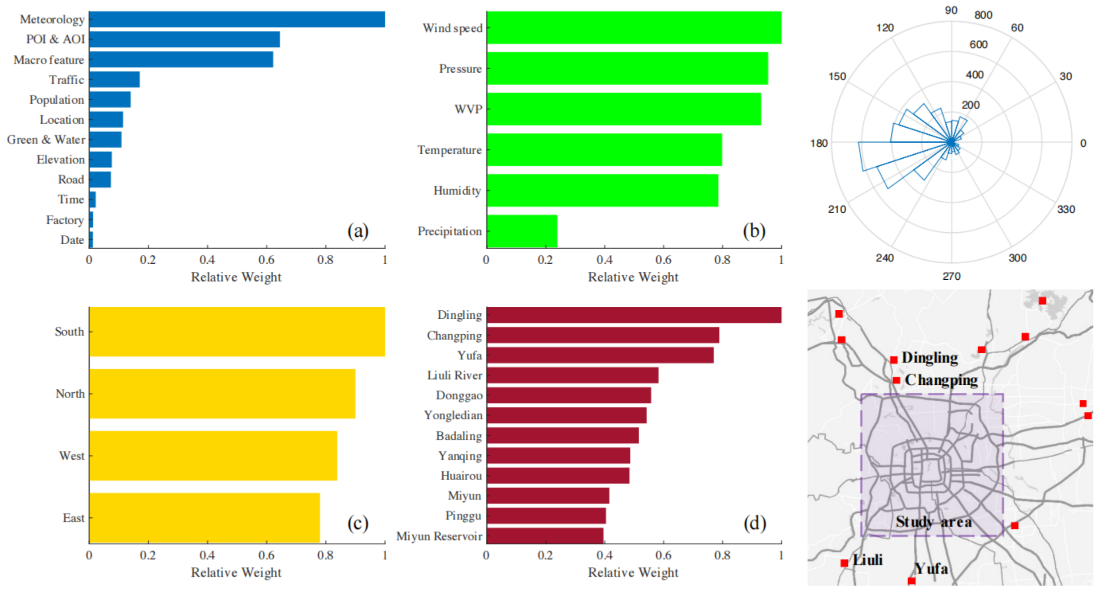


Figure 7.15: Feature analysis. (a)-(d) demonstrate the statistical relative weights of the features by groups, (c)-(d) indicate the regional transport of $PM_{2.5}$ and PM_{10} where (c) is obtained by aggregating relevant fixed stations from (d) into four main directions.

7.11 Potential impact of Deep AQ framework in Beijing case study

August 2018 marked the start of the Chinese government's Qianliyan Initiative. The aim behind the initiative is to collect data from the Jing-Jin-Ji Area (comprising Beijing, Tianjin and Hebei) and 28 nearby cities on the conditions of the atmosphere, with a specific focus on $PM_{2.5}$. As part of the initiative, this entire region was divided for the purposes of the research into 36,793 grids, 3 km × 3 km in size. The spatial resolution of such grids varied slightly, in that they could be as small as 1 km × 1 km in the region's largest cities such as Beijing. By adopting and applying mobile sensing infrastructure based on the Deep AQ framework, the amount of money required to be invested in a vast air quality monitoring project like this one would be dramatically curbed, provided at least one fixed monitoring site per spatial grid-cell were used. Consider the Beijing case study; it required a minimum of 3,025 × 300k (RMB/station) of investment in order to fully cover the 3,025 grid-cells. This amounts to over 900m RMB, and this excludes other costs such as yearly maintenance and administrative costs. In fact, 300k RMB represents a rather optimistic estimate to secure a good-quality

AQ monitoring micro-station. On the flipside, the Deep AQ framework achieves 85% inference accuracy (SMAPE < 15%) with 15 probe vehicles and 28 fixed monitoring sites. This costs less than one tenth of the aforementioned 900m RMB investment. Notwithstanding any extra fees associated with obtaining the urban data necessary to operate the Deep AQ framework, this saves a significant amount. Furthermore, the Deep AQ framework's benefit exceeds mere monetary savings. It also offers a greater amount of data and insight, as this research demonstrates. Thus, the Deep AQ framework is a scientifically and financially viable tool for sustainable urban management with a potential for global adaptation.

7.12 Conclusions

This chapter reviews the theoretical basis for the fifth of this thesis's objectives, namely to:

Apply the Deep AQ framework to Beijing case study with samples from fixed monitoring sites and low-cost mobile sensors to evaluate the model's performance and applicability.

In this chapter, the Deep AQ framework was used to examine the Beijing case study, thereby investigating the high-resolution AQ predictions created using mobile and fixed air quality data. Although it has become far easier, and far more popular, to use ubiquitous air quality sensing with cheap, moveable sensors, this study was the first of its kind to combine mobile and fixed air quality data in conjunction with urban features as a means of producing fine-granular, pixelwise AQ mapping for sizeable urban spaces. This exercise involves pinpointing the factors that affect the emission, dispersion, transportation and dissipation of air pollution by examining urban data in a range of areas, such as using macro-view features to represent how pollutants are transported within a locality. The application of the Deep AQ framework to this case study demonstrated an acceptable level of accuracy, and it surpassed other standards in this respect. In addition, the results of this study demonstrate how effective it can be to use micro-view, meso-view, and macro-view features in Deep AQ frameworks. Moreover, by interpreting the results, insights can be garnered regarding the origins and behaviour of particulate matter in Beijing.

The following conclusions are made in this chapter:

1. This study achieves high-resolution (1 km × 1 km, hourly) pixelwise air quality predictions within an area of3,500 km^2, with satisfactory performance by the Deep AQ framework using fixed and mobile AQ sensors combined with multi-source urban features (e.g., population, traffic, land use information, regional transport of air pollution).
2. Macro-view features improve the model's performance by reducing overfitting, which assists the model in becoming less dependent on locally defined features and provides quantifiable evidence of the regional transportation of air pollution.
3. Transforming features with meso-view is capable of modelling the neighbouring spatio-temporal correlations and improving the model's performance.
4. The Deep AQ framework is capable of reaching satisfactory accuracy with less than 5% of urban AQ sensing coverage based on the HOCV, LOOCV, and spatial block CV tests.
5. The Deep AQ framework's AQ inference mappings can demonstrate spatio-temporal heterogeneity, although there still exist anomalous values that are likely caused by overfitting problems because of the limited and unbalanced data.
6. Correcting the mobile AQ label responses using fixed label responses as a reference improves the accuracy and goodness-of-fit of the model in terms of RMSE, SMAPE and R^2. However, by correcting the data, it cannot thereafter be said to be representative of the ground-level concentration of pollutants.
7. Uncorrected mobile data – unlike fixed data – are mixed with more astonishing noises that give rise to a greater number of inaccuracies. Such noises tend to be synonymous with high levels of pollutant concentration.
8. It is possible to generate AQ predictions with greater accuracy by utilising fixed and mobile data for model training, in contrast to utilising fixed observations on their own.
9. Sensitivity analysis of the model with data noise variance shows that the tree-based Deep AQ frameworks are more robust to network-based Deep AQ frameworks.

10. The addition of mobile information to the training dataset increases how accurate the model inference is, although this results in a decrease in the improvement margin.
11. In respect of the Beijing case study, from 19th June - 16th July 2018, most relevant to $PM_{2.5}$ and PM_{10} concentration were weather, POI & AOI, macro-view characteristics (such as localised transportation of air pollution), and traffic. The main routes of localised transportation are, in the north-west, Dingling and Changping, and in the south-west, Yufa and the Liuli River.

Chapter 8

Towards modelling pixelwise cumulative inhalation exposure

8.1 Introduction

Chapter 7 investigates the Deep AQ framework to the Beijing case study with fixed and mobile AQ sensing. This chapter proposes a methodological framework towards modelling pixelwise population-related cumulative inhalation exposure based on AQ inference from the Deep AQ framework and population dynamics in Beijing case study. $PM_{2.5}$ is a major component of air pollution which is formed of a mixture of particles. Accumulated evidence usually suggests that it causes different levels of adverse health effects according to the levels of exposure (Pinault et al., 2017). In other words, $PM_{2.5}$ is a public health hazard (Huang & Zhang, 2013) which deserves more academic attention. Thus, $PM_{2.5}$ is selected as the pollutant of concern in this chapter. Many studies have associated long-term exposure to $PM_{2.5}$ with natural-cause mortality (Cao et al., 2011; Eeftens et al., 2012; Beelen et al., 2014). For instance, research evidence suggests that there is a significantly higher risk of lung cancer and for people who experience long-term exposure to $PM_{2.5}$, even with low concentration Specifically, lung cancer and adenocarcinoma risk will raise by 18% and 55%, respectively every the 5 ug/m^3 increases in $PM_{2.5}$ concentration (Nielsen et al., 2013). A correlation of lung cancer mortality with 10 ug/m^3 changing of $PM_{2.5}$ concentrations was identified in the study of Pope et al. (2002). In the study by Harrison et al. (2004), it was found that health risk increased by 8% using $PM_{2.5}$ concentration data from 1979 to 1983, by 13% using $PM_{2.5}$ samples collected from 1999 to 2000. According to prior research, exposure to $PM_{2.5}$ can lead to various long-term or short-term adverse health impacts (Zanobetti et al., 2014), such as all-cause mortality (Anenberg et al., 2010), cardiopulmonary mortality (Boldo et al., 2006), respiratory disease (Hao et al., 2015), cardiovascular disease, and lung cancer (Pope et al., 2011), and brain damages (Holgate et al., 2016). Other diseases that can be caused by long-term exposure to $PM_{2.5}$ include cardiovascular and respiratory diseases (Pope and Dockery, 2006). It cannot be ruled out there is possibility where the endpoints of health can occur under seemingly less life-threatening disease, such as respiratory diseases (Analitis et al., 2006; Hoek et al., 2013; Katsouyanni et al., 1997).

Much research effort (such as that by Katsouyanni et al., 1997, Harrison et al., 2004, Analitis et al., 2006, Boldo et al., 2006, Anenberg et al., 2010, Pope et al., 2011, Hoek et al., 2013, Zanobetti et al., 2014, Hao et al., 2015, Holgate et al., 2016) has been dedicated to the exploration of the adverse impacts of $PM_{2.5}$ on public health. However, all studies have their own practical limitations. For these aforementioned studies in particular, their limitations would be down to the presumption that population distribution and AQ information are constant at region-level or city-level, and the mobility level of people in areas characterized by roughly equal AQ level over one day or another defined period (Ho et al.,2015). These studies cannot dynamically assess public health risk considering the population dynamics and are usually limited by AQ inference performance in spatio-temporal domain. If the AQ maps cannot promise effective inference and reflect enough spatio-temporal heterogeneity, it may produce the potential biases in dynamic exposure analysis. The proposed Deep AQ framework offers the capability to produce pixelwise AQ inference mappings, and Chapter 7 achieves high-resolution (1 km × 1 km, hourly) inference using static and mobile AQ sensing data which provides comprehensive data for dynamic exposure analysis. Compared to previous studies, this chapter provides the pixelwise estimation of population-related cumulative inhaled $PM_{2.5}$ masses at high-resolution and fine-granularity (1 km × 1 km, hourly). Furthermore, the proposed methodological framework provides the potential to achieve multi-scale (hourly, daily, weekly, and monthly) temporal analysis for the cumulative mass of inhaled $PM_{2.5}$. The estimations at various temporal granularities are computed with the average results of the minimum-period (hourly) estimations. Figure 8.1 shows the diagram of dynamically modelling pixelwise population-related cumulatively inhaled mass of $PM_{2.5}$.

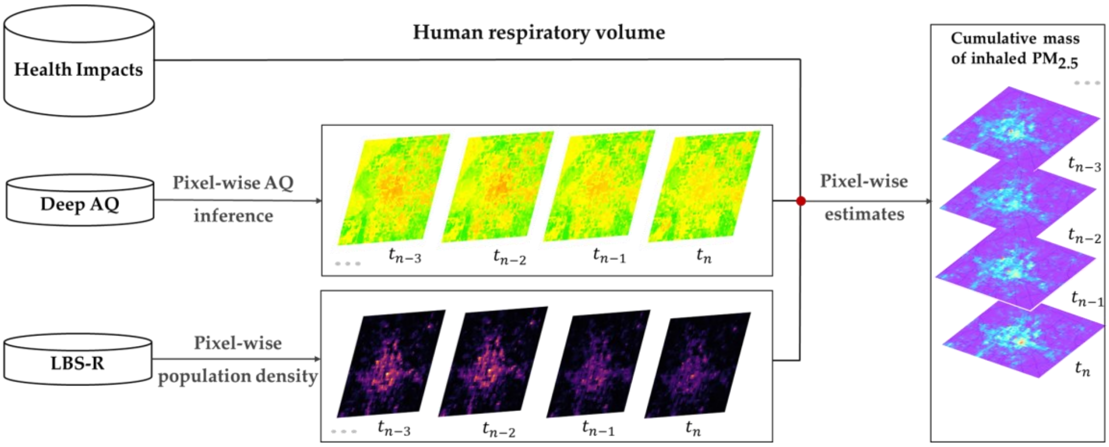


Figure 8.1: Diagram of modelling pixelwise cumulative inhalation exposure.

The objectives of this chapter are highlighted as follows:

1. Propose a methodological framework towards dynamically modelling pixelwise population-related cumulatively inhaled mass of $PM_{2.5}$.
2. Compare and quantify the differences of estimating pixelwise population-related cumulatively inhaled $PM_{2.5}$ mass with reference to LBS-R-based and census-based methods.

8.2 Materials and methods

To study the health risks and life expectancies, it is important to estimate the cumulative mass of inhaled $PM_{2.5}$ (Gamble, 1998; Gavett & Koren, 2001). High-resolution pixelwise AQ inference mappings and population dynamics are essential to constitute a prerequisite to achieve population exposure analysis at high-resolution. The proposed methodological framework in this chapter mainly consists of five components: pixelwise AQ inference from the Deep AQ frameworks, population dynamics, inhalation rate, cumulative inhaled $PM_{2.5}$ masses, and multi-scale (hourly, daily) temporal exposure analysis. This methodological framework provides support to help policy-makers to prioritise measures that limit $PM_{2.5}$ pollution through application of health and economic policy instruments.

8.2.1 Pixelwise AQ mappings and population dynamics

The high-resolution pixelwise AQ inference mappings can be derived by the Deep AQ framework in the Beijing study case during 19 June and 16 July 2018. The dynamic population density can be collected based on LBS-R data from WeChat and every piece of information pertaining to user' identity and other private details had been deleted before usage. The LBS-R data are deemed as instantaneous population density and the collecting period is hourly based. In fact, the density maps using information from WeChat only show active users and not actual population density. Furthermore, the data largely ignores partial groups of people in the society, including people of lower economic status, elderly people and children, due to their less frequent mobile phone usage in these groups compared with the whole community. Therefore, this research regrades the LBS-R data as indicators to quantify the population dynamic patterns and redistribute the demographic data in the study area described with Equation (3.2) and (3.3) based on the census-based data. The census-based population data (approximate 10.5 million people) in study area of Beijing is obtained from the statistics bureau, and this research assumes that the inner-urban population mobility does not dramatically affect the total census-based population, which denotes that the inflow and outflow of population movements of the study area have not been considered in this research. Figure 8.2 shows the pixelwise population density maps in the study area (2018/06/19), where the density illustrates particular population size within every pixelwise grid-unit, using a scale of black to yellow-white (lowest to highest) in order to do so.

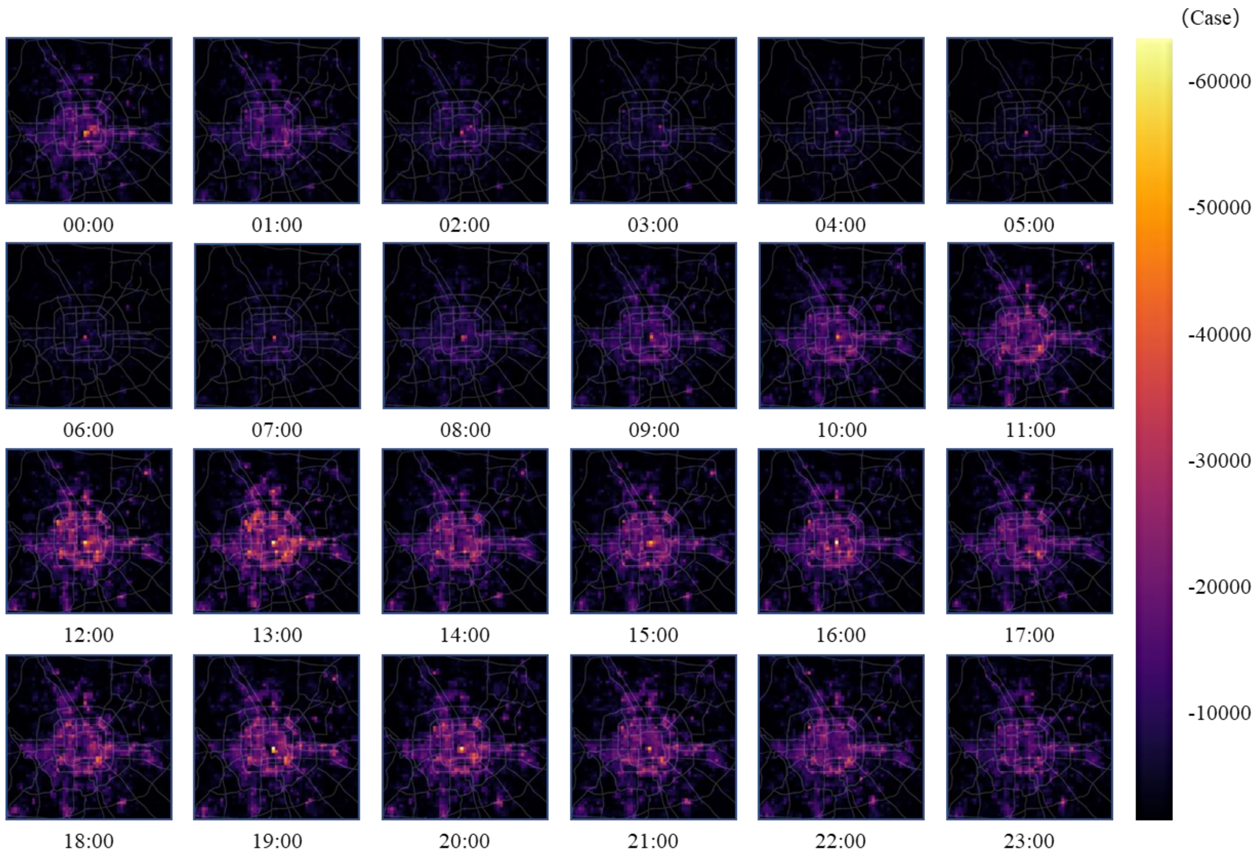


Figure 8.2: Pixelwise mappings of population density in Beijing (2018/06/19).

8.2.2 Estimation of cumulative inhaled $PM_{2.5}$ masses (Chen et al., 2018)

The adverse health effects are mainly caused by the inhalation exposure and it's important to estimate the mass of $PM_{2.5}$ that is inhaled in total, which constitutes one among the most crucial preconditions for modelling the correlation between exposure to $PM_{2.5}$ and public health. To estimate the cumulative inhaled $PM_{2.5}$ masses at a specific area, the inhalation rate for different groups of people, population density and the spatio-temporal variations of $PM_{2.5}$ are fused to achieve a better estimate:

$$Inh_{PM2.5} = \sum_{i=1}^{N} \sum_{t=1}^{\mathcal{T}} \begin{matrix} p_i(t) \cdot h_i \cdot d_i(t) \cdot pm(t) \ + \\ p_i(t) \cdot h_i \cdot (1 - d_i(t)) \cdot pm(t) \ \cdot \alpha \end{matrix} \tag{8.1}$$

where $Inh_{PM_{2.5}}$ denotes the cumulatively inhaled mass of $PM_{2.5}$ (unit: ug) at this specific area, and h_i denotes the inhalation rate (unit: m^3/h) for the i_{th} age group, N represents the overall amount of age groups. t denotes the time (hours in this study), $pm(t)$ and $p_i(t)$ denote the $PM_{2.5}$ concentration (unit: ug/m^3) and population (unit:

case) at time t in specific area, $\mathcal{T}$ is the target temporal period, $d_k(t)$ denotes the percentage of the population outdoors at time t, and α denotes the outdoor-indoor ratio of $PM_{2.5}$ concentration.

Nonetheless, certain limitations of Equation (8.1) exist due to their dependency on multiple variables. For example, the ratio of outdoor to indoor exposure to $PM_{2.5}$ is impacted by various details, including geographic location, building structures, etc. Besides, the inhalation rate varies across age, sex, and other factors affecting the inhaled value (Marty et al., 2002). Hence, the ideal model expressed by Equation (8.1) can be simplified as Equation (8.2) by neglecting the factors mentioned above. Thus, the estimations of cumulative inhaled $PM_{2.5}$ masses can be directly obtained as follows:

$$Inh'_{PM2.5} = \sum_{t=1}^{\mathcal{T}} p(t) \cdot h \cdot pm(t) \tag{8.2}$$

where $\mathrm{Inh}'_{\mathrm{PM}_{2.5}}$ denotes the cumulatively inhaled mass of $PM_{2.5}$ from the simplified model, p(t) represents population data at time t and h reflects the empirical inhaled volume of $PM_{2.5}$. Adams conducted an experiment (Adams, 1993) measuring the hourly average air inhalation volume based on 200 individuals (Chen et al., 2018) and the results show that the average hourly ratio of inhalation volume ranged from 0.42 m^3/h ~ 0.63 m^3/h for sitting, from 1.20 m^3/h ~ 1.44 m^3/h for walking, and from 3.10 m^3/h ~ 3.48 m^3/h for running. This study assumes that an individual inhales, on average, 0.625 m^3/h of $PM_{2.5}$ (Marty et al., 2002).

8.2.3 Comparison of exposure assessments based on the LBS-R-based and census-based methods

As a means of investigating whether incorporating population dynamics makes a difference to exposure analysis, this research quantitatively compared the LBS-R-based and census-based estimations of cumulatively inhaled mass of $PM_{2.5}$ on varying temporal scales (hourly and daily). For LBS-R-based methods, the pixelwise cumulatively inhaled mass of $PM_{2.5}$ is estimated with reference to pixelwise AQ inference and population dynamics, and all the pixelwise estimations are summed up together to compute the inhaled mass of $PM_{2.5}$ for the whole region, as shown in Figure 8.3.

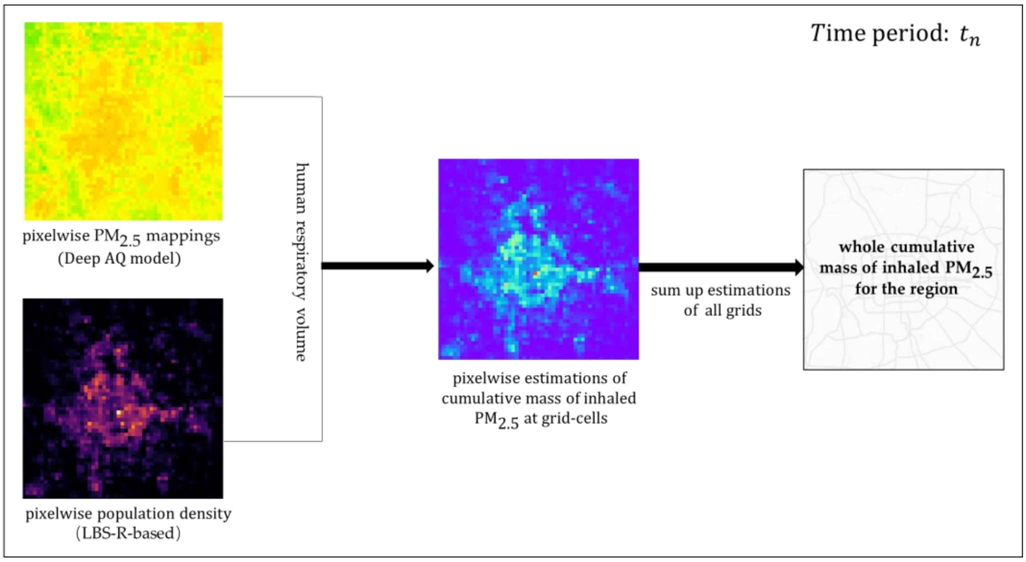


Figure 8.3: Diagram of computing the inhaled mass of $PM_{2.5}$ for the whole region at the time period t_n based on LBS-R-based methods.

For census-based methods, the whole cumulatively inhaled mass of $PM_{2.5}$ for a region are directly estimated using the average AQ observations and total statistical population number, as shown in Figure 8.4.

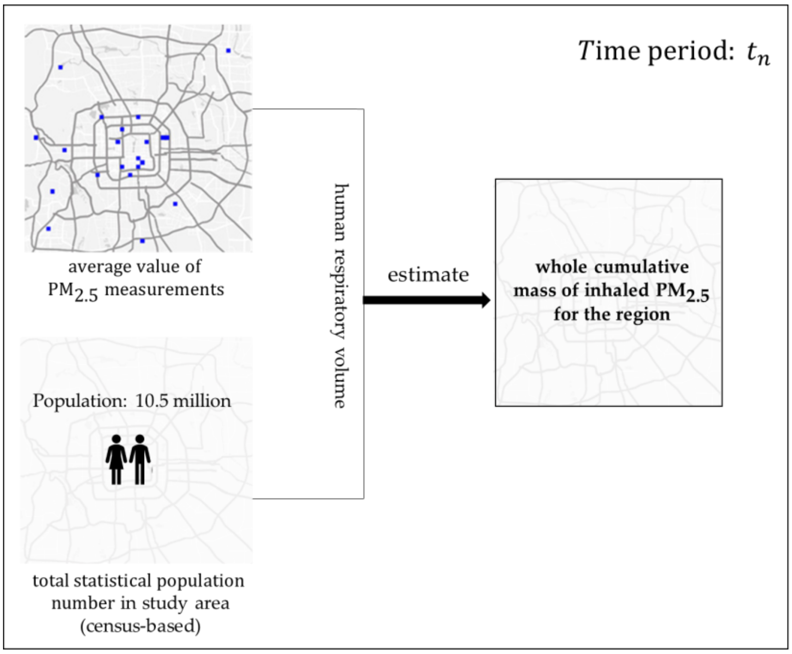


Figure 8.4: Diagram of computing the cumulatively inhaled mass of $PM_{2.5}$ for the whole region at the time period t_n based on census-based methods.

To quantify the deviation of results from LBS-R-based and census-based methods, the Mean Absolute Percent Error (MAPE: described as Equation (8.3)) is used to evaluate the bias.

$$\text{MAPE} = \frac{y_{cens} - y_{lbs}}{y_{lbs}} \tag{8.3}$$

where y_{lbs} refers to the whole cumulatively inhaled mass of $PM_{2.5}$ for the region with reference to the LBS-R-based method and y_{cens} refers to the whole cumulatively inhaled masses with reference to the census-based method for the region.

8.3 Results and discussions

This chapter constructs the functional relationships between exposure and cumulatively inhaled mass of $PM_{2.5}$ quantitatively with Equation (8.2) in the proposed methodological framework to achieve hourly pixelwise estimates and provide the potential to conduct temporal exposure analysis in multi-scales (i.e., hourly, daily).

8.3.1 Cumulative inhaled $PM_{2.5}$ masses

The LBS-R-based technique demonstrates the gridded cumulatively inhaled mass of $PM_{2.5}$ and improves the spatial resolution of basic grid-cells from region-level to pixelwise level. Figure 8.5 shows the hourly pixelwise mappings of cumulatively inhaled $PM_{2.5}$ mass with reference to the LBS-R population density in 2018/06/19, which depicts substantial spatial heterogeneity in each hour. It can be found that fewer inhabitants are impacted by $PM_{2.5}$ pollution during 2:00-4:00 and it's considered that the active users are not too much because most people are in deep sleep during this period. What's more, the hourly estimations provide a comprehensive data basis and offer the potential to compute the daily and monthly results.

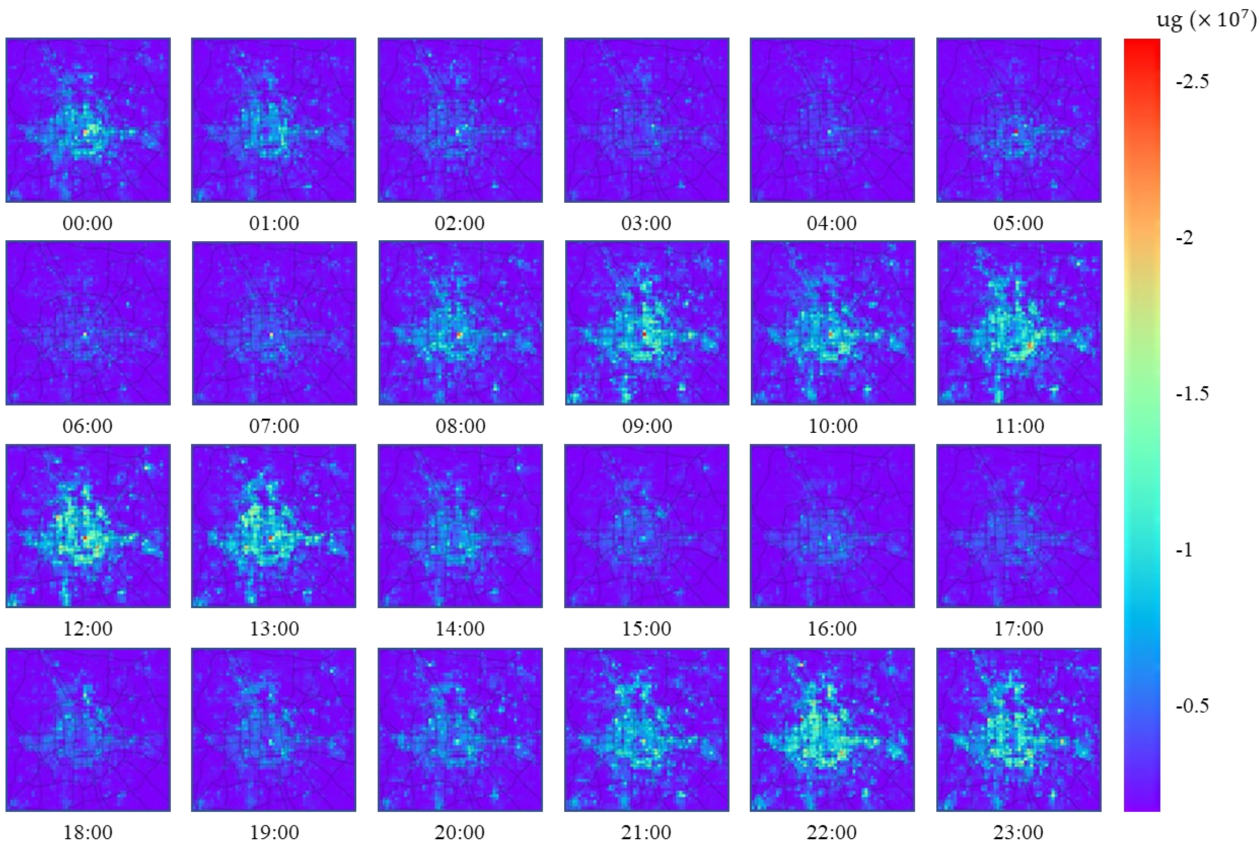


Figure 8.5: Hourly pixelwise mappings of cumulatively inhaled mass of $PM_{2.5}$ with reference to LBS-R population density (2018/06/19).

To better understand the distribution of the grid-cell's cumulative inhalation, this research further aggregated the pixelwise estimations at each hour from 2018/06/19-2018/07/16. Figure 8.6 shows the distribution of cumulatively inhaled quantities of $PM_{2.5}$ at the proportional node in all grid-cells with the sorted ascending order of their corresponding inhalation values in each hour (2018/06/19-2018/07/16). The results distinguish diurnal and daily variations, thereby proving that it is necessary to consider spatio-temporal variability of AQ mappings and population dynamics into exposure analysis and provide detailed evidence for urban health management.

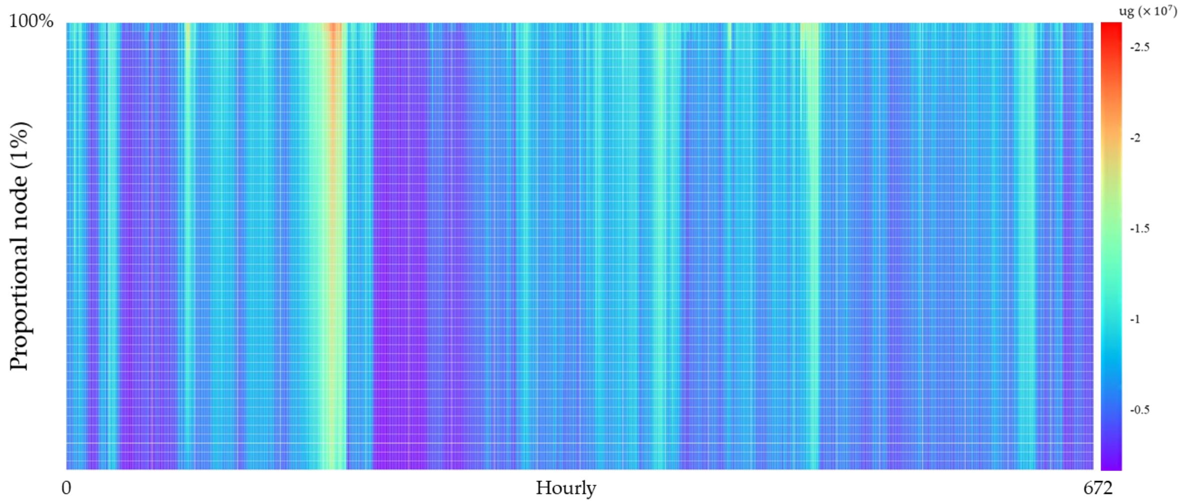


Figure 8.6: Distribution of cumulatively inhaled mass of $PM_{2.5}$ at the proportional node of grids with the sorted ascending order of their corresponding inhalation values in each hour (2018/06/19 0:00-2018/07/16 23:00)

For each grid-cell, the hourly estimates of cumulative masses of inhaled $PM_{2.5}$ in a day are summed up to compute the daily estimate. Figure 8.7 shows the daily pixelwise mappings of cumulative inhaled $PM_{2.5}$ masses from 2018/06/19 to 2018/07/16 based on the LBS-R population density, which depicts daily substantial spatial heterogeneity, and it can be found that a greater number of inhabitants are impacted by $PM_{2.5}$ pollution in central and southern Beijing (conclusions are consistent with Yin et al., 2017), and cumulative masses are directly proportional to size of the population. The distribution patterns look similar in spatial domain as population dynamics in a city have similar patterns; However, the cumulative inhalations of each day are temporally different which is considered mainly caused by different AQ conditions.

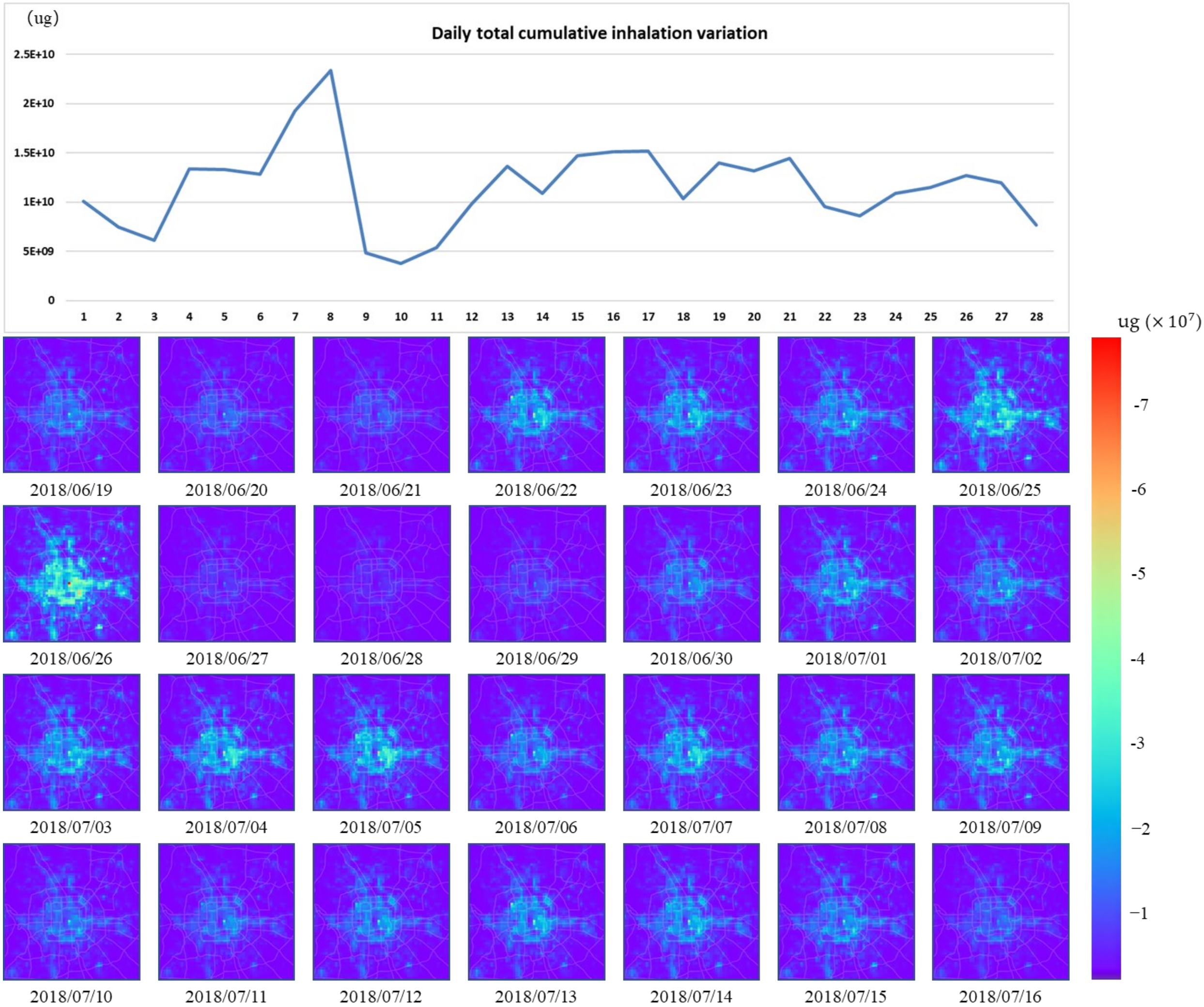


Figure 8.7: Daily pixelwise mappings of cumulative inhaled mass of $PM_{2.5}$ based on the LBS-R population density (2018/06/19-2018/07/16).

8.3.2 Temporal analysis of exposure and cumulatively inhaled $PM_{2.5}$ masses

Moreover, the LBS-R-based and census-based methods can yield different estimations of the total cumulatively inhaled $PM_{2.5}$ mass for a whole region and the deviation can be computed by Equation (8.3) as a scale to quantify the differences between the two methods across different temporal scales (i.e., hourly, daily). The results in Figure 8.8 show the maximum deviation can reach to over 1000% (even 2500% at an hour) at hourly temporal scale and over 150% at daily temporal scale. Meanwhile, the range of deviations at hourly-granularity (-1 ~ 25 × 100%) is significantly larger than that of daily-granularity (-0.6 ~ 1.8 × 100%), which means that the estimations of inhalation exposure between LBS-R-based and census-based methods make a more significant difference at a finer temporal granularity. However, the margins of difference decrease with coarser temporal granularity (hourly → daily).

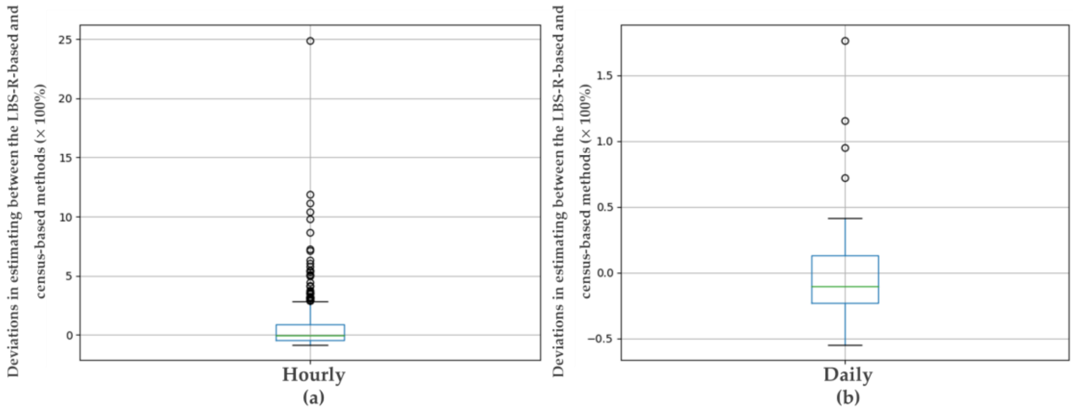


Figure 8.8: Deviations in estimating the hourly (a) and daily (b) cumulatively inhaled mass of $PM_{2.5}$ for the whole region between the LBS-R-based and census-based methods

This chapter considers well on the spatio-temporal variability of both high-resolution population dynamics and pixelwise AQ inference, thus improving assessments of the impact of pollution inhalation, and the relative advantage is likely a result of the considerations below.

1. Spatio-temporal variation in high-resolution population dynamics and pixelwise AQ inference are incorporated in exposure analysis. Given that the pollutant concentrations are constantly changing, populations also move constantly across spatial-temporal scales (Park & Kwan, 2017), and both dynamic characteristics and their interactions at high-resolution spatio-temporal scales should be well involved. However, existing research has consistently utilised census-based information to assume those who do not move within regions are generally exposed to the same AQ conditions at all times, which creates considerable bias when actual exposure is analysed.

2. The proposed methodological framework incorporated inhalation rate, pixelwise AQ inference, and population dynamics to estimate population-weighted cumulatively inhaled mass of $PM_{2.5}$. This is intended to make a contribution to

modelling relationships between exposures and health risks quantitatively. In contrast with a census-based approach, the LBS-R-based method is capable of yielding fine-granular and high-resolution estimations of population inhalation of $PM_{2.5}$. The method can estimate cumulatively inhaled $PM_{2.5}$ mass at a given location (grid-level) and time (hourly) in sizeable urban spaces based on the spatio-temporal mutability of population dynamics and the concentration of pollutants.

8.4 Conclusions

This chapter reviews the theory basis for the sixth research objective in this thesis, which is to:

Propose a methodological framework towards dynamically modelling population-weighted cumulative inhaled $PM_{2.5}$ masses.

This chapter proposes a methodological framework to estimate population-weighted cumulatively inhaled $PM_{2.5}$ at high-resolution (1 km × 1 km, hourly), which provides the potential to achieve multi-scale (i.e., hourly, daily) temporal analysis. The methodological framework involves pixelwise AQ inference from the Deep AQ frameworks, pixelwise population dynamics, inhalation rate, cumulative inhaled $PM_{2.5}$ masses and multi-scale (i.e., hourly, daily) temporal exposure analysis. This methodological framework provides support to help policy-makers to prioritise the reduction of $PM_{2.5}$ pollution through health and economic policy instruments. In this chapter, indoor and outdoor $PM_{2.5}$ concentrations are not distinguished, while the inhalation rate is assumed constant ignoring age, sex, and other factors; thus, these limitations make the results conservative in this chapter. It can be said that, because these approximations of total inhalation volume are the highest since this study, it therefore does not consider outdoor settings or commuter transportation. However, it is possible that cumulative inhalation volume may be higher, as this study adopts one consistent value that represents an adult's low inhalation rate and does not consider physical activity, size, gender or other variables. Therefore, the results could be improved on the aspects of acquiring more detailed information about population including the percentages of indoor and outdoor activities, moving trajectories of inhabitants and etc.

The following conclusions are made in this chapter:

1. Differing from previous health studies at region-level or city-level, this chapter constructs a novel methodological framework towards dynamically modelling population-weighted cumulatively inhaled mass of $PM_{2.5}$ at high-resolution (1 km × 1 km, hourly).
2. In contrast to census-based methods, this chapter introduces LBS-R-based population dynamics and pixelwise AQ inference into the inhalation exposure analysis and the considerable differences are figured out, and in doing so, it provides a more advanced method of assessing the data at different temporal scales to support for sustainable health management and relevant policy-driven actions.

Chapter 9

Conclusions and Future work

This thesis proposes a ML-based pixelwise AQ inference framework called Deep AQ which offers the capability to generate high-resolution (1 km × 1 km, hourly) AQ inference using multi-source (fixed or mobile) AQ samples and urban features. This has been applied and evaluated through two case studies, one in Chengdu involving sampling from fixed monitoring sites, and the other in Beijing involving sampling from both fixed and mobile sensors. In the Chengdu case study (4,900 km^2, 26 April - 12 June 2019), the Deep AQ framework is applied and evaluated where air pollutant ($PM_{2.5}$, PM_{10} and O_3) measurements from 40 fixed monitoring sites are used to train the model to estimate pollutants' concentrations in 4,900 grid-cells (1 km^2). The Deep AQ frameworks are applied to find out the correlations between urban features and air pollution sensing data, which estimates pixelwise (grid-level) concentrations of air pollutant species. Furthermore, the proposed ML-based model allows the achievement of high-resolution (1 km × 1 km, hourly) AQ inference based on pollutant measurements covering less than 1% of the grid-cells with reasonable accuracy, under 20% SMAPE compared to other approaches. In the Beijing case study (3,025 km^2, 19th June to 16th July 2018), the research achieves high-resolution (1 km × 1 km; hourly) AQ inference, with satisfactory accuracy under 15% SMAPE using mobile and fixed air pollution sensing which covers less than 5% of the grid-cells. Furthermore, this research leverages urban features in order to uncover insights into the potential causes of pollution, an endeavour that supports evidence-based urban management which is environmentally sustainable. This research may be especially suitable for vehicle-based endeavors to cover locations without dense AQ monitoring networks, especially in developing countries. In this chapter, the focus will be on the ways in which research objectives have been confronted, before discussing the broader impacts of related outcomes and conclusions.

9.1. Summary of conclusions

This chapter discusses 6 key objectives, as identified in section 1.2 of this thesis, based on an examination of existing literature in this field.

1. Conduct a feasibility study through extensive literature review on viable AQ sensing and modelling approaches, in conjunction with machine learning in common spatio-temporal analytics.

Fixed monitoring sites are widely used as the primary method of collecting AQ observations for decision making. They offer the opportunity to observe continually for long time periods (temporally dense), but have sparse spatial coverage (e.g., immediate vicinity level). Conversely, satellite remote sensing is capable of providing enough spatial coverage (e.g., whole city-level), yet with restricted temporal granularity, confined by the time interval that reflects the frequency of satellites passing over the study area. What' s more, mobile sensing refers to float vehicles using the term 'data mules', which presents a low-cost, easily expandable and environmentally-sustainable way to scan cities to reveal detailed pollution dynamics. For air pollution modelling, policy-makers and researchers have been continually seeking appropriate approaches to estimate concentrations of pollutants at various scales by modeling correlations between the pollutant measurements and a wide range of parameters. In addition to the interpolation approaches, existing modelling approaches encompass the bottom-up (deterministic) and data-driven (statistical) models (Chen et al., 2018). With knowledge of different emission sources, bottom-up models have been widely used by environmental scientists to study how pollutants are emitted and physically transported, and the chemical interactions of these pollutants, by making sets of empirical and theoretical assumptions. These models can be used for AQ inference and predetermination which are continually used and improved by simulating the processes of pollution diffusion, migration, and conversion. In addition to bottom-up models, another approach is using data-driven approaches to correlate data with explanatory factors including the land use, traffic, meteorological information and more. LUR is a typical data-driven approach that utilises linear regression as a standard surrogate model to estimate concentrations of air pollutants which regards AQ measurements as dependent variables and explanatory factors as independent variables. In addition to the typical LUR model, ML-based techniques have been used in recent studies. Furthermore, this research also reviews common spatio-temporal analytics in relation to other research problems, either separately or jointly.

The wider conclusions presented in Chapter 2 are as follows:

a) Mobile air pollution sensing constitutes a low-cost, expandable and environmentally-sustainable means of surveying the atmosphere environment.

b) Bottom-up models provide high model fidelity regarding the physical and chemical processes involved in air pollution. However, such models are usually computationally expensive while relying on assumptions which constrain how feasible it is to integrate them into high frequency and ubiquitous air pollution data.

c) A typical LUR model tends to be characterised by low temporal resolutions (such as daily or weekly), limiting its capacity to uncover nuanced insight into pollution dynamics.

d) Machine learning techniques facilitate the capability to enhance the development of spatio-temporal data analytics.

2. Collect multi-source AQ samples and urban features for spatio-temporal analytics.

In the Chengdu case study, only fixed monitoring sites are used to explore the predictivity of the Deep AQ framework. In the Beijing case study, collaborative sampling from fixed monitoring sites and low-cost mobile sensors is used to achieve urban coverage, offering thorough air quality data within sizeable urban spaces. For mobile AQ samples, measurements are collected from IoT sensing systems in real-time and calibrated by OFCLP, where the distributed low-cost sensors measure AQ observations, calibrate them and transmit into the cloud server. For urban features, nuanced and balanced representations of local characteristics are crucial in order to allow data-driven approaches to minimise overfitting and boost the ease with which the model can be interpreted. All urban characteristics can be divided into 4 categories: geographic and land use category, traffic category, population category, and meteorology category. All the AQ samples and urban characteristics are defined and illustrated with reference to the spatio-temporal grid-cell units.

The wider conclusions presented in Chapter 3 are as follows:

a) The IoT sensing system is the research basis for collecting AQ samples and transmitting them into the cloud server; the online field calibration platform for mobile sampling campaigns is set up to calibrate the mobile measurements generated from the low-cost sensors in real time.

b) The extent of mobile sensing power is surprisingly great – for Beijing, only 15 randomly placed taxis give coverage of one-third of the city's roads each day, offering the capability to make up the shortcomings (spatial sparsity) of fixed monitoring sites. However, mobile samples are irregularly and sparsely distributed, and there exist locations bereft of sensing coverage each day, which may produce overfitting estimates at unsampled grid-cells because of the asymmetrically distributed AQ samples.

c) Urban AQ conditions are influenced by numerous endogenous factors and regional transportation of air pollution. As a result, comprehensive, balanced representations of urban characteristics are crucial within data-driven approaches in order that they can mitigate overfitting and boost models' ease of interpretation.

d) The apparent spatio-temporal dynamics and sparsity of the air pollution sensing pose a significant challenge and the AQ observations should be defined and represented in terms of the spatio-temporal units by exploring the spatio-temporal correlations of different grid-cell units in conjunction with their urban features.

3. Propose a spatio-temporal learning and inference framework called Deep AQ framework to incorporate multi-source (fixed or mobile) AQ samples and urban features to achieve high-resolution pixelwise AQ inference.

A ML-based spatio-temporal learning and inference framework called the Deep AQ framework is proposed for AQ inference using mobile and fixed air pollution sensing.

The proposed model contains multi-source and heterogenous sets of data which vary tremendously in respect of their granularity, numerical scale, and accuracy. As a result, one critical task is transforming urban characteristics using appropriate spatio-temporal structures, combining endogenous variables pertaining to the localised atmospheric conditions (such as traffic and meteorology) and background information indicators (e.g., regional transport of air pollution). The pollutant concentrations of a given grid-cell at a specific time are spatially and temporally related to their neighbours in the spatio-temporal domain. The model is adaptable to accommodate irregularly and sparsely distributed AQ samples from fixed or mobile sensing.

The wider conclusions presented in Chapter 5 are as follows:

a) A systematic high-resolution air pollution modelling framework (Deep AQ) is proposed and developed based on multi-source air pollution sensing (fixed or mobile) and urban features. The proposed framework is adaptable to accommodate irregularly and sparsely distributed AQ samples from either fixed or mobile sensing.

b) Dynamic convolution filters are applied in the Deep AQ framework to effectively transform neighbouring spatio-temporal information for each grid-cell unit with various receptive fields and convert them into meso-view generalised features.

c) Pollutant measurements from monitoring sites surrounding the study area are coined as macro-view features, which improves the accuracy of model inference through limiting overfitting over endogenous characteristics while offering quantifiable evidence of the sources of pollution.

d) DFeaST-Tree is a tree-based learning structure capturing spatio-temporal correlations to construct relationships between AQ data and the micro-view, meso-view, and macro-view features as aforementioned.

e) DFeaST-Net is a network-based learning structure capturing spatio-temporal correlations to construct relationships between AQ data and the micro-view, meso-view, and macro-view features as aforementioned. Compared to DFeaST-Net, the DFeaST-Tree performs more stable and robust operations than DFeaST-Net.

f) Multi-AP is a network-based learning structure that captures spatio-temporal correlations within one integrated learning network, containing both the shared network layers to construct the correlations among multiple pollutants and the individual network layer for each pollutant independently. Multi-AP is capable of estimating concentrations of multiple air pollutant species simultaneously by accounting for potential correlations among these pollutant species.

4. Apply the Deep AQ framework to the Chengdu case study using fixed monitoring sites to evaluate the model performance and applicability.

The Deep AQ framework is applied in the case of Chengdu with reasonable accuracy (SMAPE < 20%) using only fixed monitoring sites, outperforming other listed standards. Additionally, these results indicate how effective it can be to use micro-view, meso-view and macro-view features. The standard deviation of the different evaluation metrics for $PM_{2.5}$, PM_{10} and O_3 are compared and the results show that the Deep AQ framework's performance on different validation datasets is more consistent than other benchmarks. To investigate the sensitivity of the Deep AQ framework to the availability of various feature predictors, different subsets in micro-view (L) inputs are evaluated and compared, which show that the combination of all predicators performs best. Furthermore, independent LOOCV tests are also applied to evaluate the Deep AQ framework's predictivity to achieve AQ inference at unsampled sub-areas. The AQ inference mappings are shown and analysed to depict spatio-temporal heterogeneity.

The wider findings in Chapter 6 are concluded as follows:

a) This study achieves high-resolution ($1\,\text{km} \times 1\,\text{km}$, hourly) pixelwise AQ inference in the study area (4,900 km^2) with reasonable performance, with the Deep AQ framework using fixed monitoring sites in conjunction with multi-source urban features (e.g., population, traffic, land use information, regional transport of air pollution, etc.).

b) Macro-view features improve the model performance by reducing overfitting to assist the model in becoming less dependent on locally defined features and to provide quantifiable evidence of the regional transport of air pollution.

c) Transforming features with meso-view is capable of building up neighbouring spatio-temporal correlations and improving the model's performance.

d) The Deep AQ framework is capable of reaching reasonable accuracy with less than 1% urban AQ sensing coverage based on HOCV, LOOCV, and spatial block CV tests.

e) The standard deviation on different validation datasets is compared, and the goodness-of-fit of all scatter plots suggests that the Deep AQ framework is capable of producing stable and robust results.

f) Performance variability of the Deep AQ framework with various potential predictors is compared, which shows that the combination of all predicators outperforms other combinations of different predicators.

g) The AQ inference mappings produced by the Deep AQ framework can express the spatio-temporal heterogeneity, although there still exist anomalous values that are likely caused by overfitting problems affected by the limited and unbalanced data.

5. Apply the Deep AQ framework to the Beijing case study using fixed monitoring sites and low-cost mobile sensors to evaluate the model performance and applicability.

The Deep AQ framework is applied in Beijing with an acceptable level of accuracy (SMAPE < 15%) based on fixed and mobile sensors, which outperforms other listed standards. Furthermore, these results indicate the efficacy of utilising micro-view, meso-view and macro-view features. The AQ inference mappings are shown and analysed to depict spatio-temporal heterogeneity. In this case study, sensitivity analysis of noise variance is conducted to test the model's robustness towards data noises, and a sensitivity analysis of spatio-temporal coverage is also conducted to quantify how the spatio-temporal coverage influences the model's performance.

Moreover, through interpreting these results, insights into the causes and features of Beijing's particulate pollution can be extracted.

The wider findings in Chapter 7 are as follows:

a) This study achieves high-resolution (1 km × 1 km, hourly) pixelwise air quality prediction within a 3,500 km^2 area, with satisfactory performance by the Deep AQ framework using fixed and mobile AQ data in conjunction with multi-source urban features (e.g., population, traffic, land use information, regional transport of air pollution, etc.).
b) Macro-view features improve the model's performance by reducing overfitting to assist the model less dependent on locally defined features and to provide quantifiable evidence of regional transportation of air pollution.
c) Transforming features with meso-view is capable of constructing neighbouring spatio-temporal correlations and improving the model's performance.
d) The Deep AQ framework can reach a satisfactory accuracy based on HOCV, LOOCV, and spatial block CV tests with less than 5% of urban AQ sensing coverage.
e) The AQ inference mappings produced by the Deep AQ framework can depict the spatio-temporal heterogeneity, although there still exist anomalous values that are likely caused by overfitting problems affecting by the limited and unbalanced data.
f) Correcting the mobile AQ label responses using fixed observations as reference improves the accuracy of model inference and goodness-of-fit in terms of RMSE, SMAPE and R^2. However, by correcting the data, it cannot thereafter be said to be representative of the ground-level concentration of pollutants.
g) Uncorrected mobile data – unlike fixed data – are mixed with more astonishing noises that give rise to a greater number of inaccuracies. Such noises tend to be synonymous with high levels of pollutant concentration.
h) It is possible to generate AQ predictions with greater accuracy by utilising fixed and mobile data for model training, in contrast to utilising fixed observations on their own.

i) Sensitivity analysis of the Deep AQ frameworks with data noise variance shows that the tree-structured Deep AQ frameworks are more robust than the network-structured Deep AQ frameworks.

j) The addition of mobile information to the training dataset increases how accurate the model inference is, although this results in a decrease in the improvement margin.

k) In respect of the Beijing case study, from 19th June - 16th July 2018, most relevant to $PM_{2.5}$ and PM_{10} concentration were weather, POI & AOI, macro-view characteristics (such as localised transportation of air pollution), and traffic. The main routes of localised transportation are, in the north-west, Dingling and Changping, and in the south-west, Yufa and the Liuli River.

6. Propose a methodological framework towards dynamically modelling cumulative inhalation exposure.

This chapter proposes a methodological framework that moves towards dynamically modelling population-weighted cumulative inhaled $PM_{2.5}$ masses at fine-granularity and high-resolution (1 km × 1 km, hourly), which provides the potential to achieve multi-scale (i.e., hourly, daily) temporal exposure analysis. The methodological framework involves pixelwise AQ inference, population dynamics, inhalation rate, cumulative inhaled $PM_{2.5}$ masses and multi-scale (i.e., hourly, daily) temporal exposure analysis. This methodological framework provides support to help policy-makers in making reduction of $PM_{2.5}$ pollution a priority through health and economic policy instruments. In this chapter, the contrasting situation in relation to indoor and outdoor $PM_{2.5}$ concentrations, ratios of different groups (e.g., sex, age, gender) and differences of activities are not distinguished. These limitations limit the results in this chapter and indicate that detailed research to eliminate the limitations should be a focus of future work.

The following conclusions are made in Chapter 8:

a) In contrast to previous health studies at region-level or city-level, this chapter constructs a novel methodological framework that aims to dynamically model population-weighted cumulative inhaled $PM_{2.5}$ masses at high-resolution (1 km × 1 km, hourly).

b) Compared to the census-based method, this chapter introduces the LBS-R-based population dynamics and pixelwise AQ inference into the study's exposure analysis. This provides a better means of assessment at different temporal scales to support sustainable health management and relevant policy-driven actions.

This thesis proposes the Deep AQ framework to produce high-resolution pixelwise AQ inference based on multi-source AQ samples (fixed or mobile) and urban features. Three types of Deep AQ frameworks are proposed to model spatio-temporal correlations to achieve AQ inference. The Deep AQ framework is applied to and evaluated in Chengdu and Beijing to explore the potential of the model. For dynamic exposure modelling, this research takes the variations of pixelwise pollutant concentrations and LBS-R population density into consideration, which provided data as a basis for more reliable exposure analysis.

9.2 Research contributions

This research is one of the systematic studies trying to apply ML-based methods for high-resolution AQ modelling in the most optimum way, and the proposed data-driven models in this thesis have the potential for real-time operations which offer the capability to automatically construct spatio-temporal relationships between multi-source AQ samples and urban features, which can allow researchers to easily aggregate results for regional investigations.

The research contributions in this thesis are outlined as the following:

1. This research advances the frontiers of research into modelling urban air pollution with fixed and mobile AQ sensing supported by multi-source urban features.
2. A systematic high-resolution AQ inference framework (Deep AQ) is proposed and developed based on multi-source AQ samples (fixed or mobile) and urban features. Tree-based and network-based learning structures of Deep AQ frameworks are proposed to perform the pixelwise AQ inference.

3. The strategy of the online field calibration is designed to calibrate the mobile data to reduce discrepancies caused by the low-cost sensors.
4. The proposed Deep AQ framework is capable of accommodating irregularly and sparsely distributed AQ samples from either fixed or mobile sensing.
5. To further deal with ingrained contradictions between mobile and fixed label responses, a correction approach is proposed to correct mobile data based on coordinates, timestamps, temperature, and humidity.
6. Among existing literature of data-driven AQ modelling, it is the first study to utilise measurements from monitoring sites outside the study area as macro-view features to take localised transportation of air pollution into consideration, thereby improving the accuracy of the model inference by reducing overfitting over endogenous urban features while, additionally, offering quantifiable evidence that can be used to determine the sources of pollution.
7. This research constructs spatio-temporal filters based on convolution methods to effectively extract and transform neighbouring spatio-temporal information for each grid-cell unit with various receptive fields and convert them into meso-view generalised features.
8. This research converts the transformed and generalised features from micro, meso, and macro views into suitable variables as model training predicator variables to develop the Deep AQ framework.
9. This research proposes a mechanism to quantify the contributors of relevant urban features considering both the endogenous factors and regional transportation of air pollution.
10. The proposed Deep AQ framework achieves satisfactory performance and offers the potential to invest in ninety percent less hardware than at present, in contrast to ubiquitous sensing which is predominantly dependent on fixed monitoring sites.
11. This research achieves high-resolution (1 km × 1 km, hourly) AQ inference with reasonable accuracy (SMAPE< 20%) in Chengdu (4,900 km^2, 26 April - 12 June 2019) using only 40 monitoring sites. To our knowledge, this research is among the first studies to offer high-resolution pixelwise AQ inference to recover the pollution maps for a large metropolitan area using only fixed monitoring sites.

12. This research achieves high-resolution (1 km × 1 km, hourly) AQ inference with satisfactory accuracy (SMAPE < 15%) in Beijing (3,500 km^2, 19 June - 16 July 2018) using 28 monitoring sites and 15 low-cost mobile sensors. To our knowledge, this research is among the first studies to offer high-resolution pixelwise AQ inference to recover the pollution maps spanning sizeable urban areas using both mobile and fixed sensors.
13. Throughout the Chengdu case study, this research provides in-depth discussions on the selection of predicators, stability and robustness of model, and analysis of pollution maps. An intuitive analysis of the AQ predictions illustrate how well the proposed model suits its intended purpose.
14. Throughout the Beijing case study, this research provides a deep exploration of how to reconcile mobile and fixed data, the impact of mobile sensing coverage, and a significance analysis of urban characteristics. By interpreting the results of the AQ predictions intuitively, the case study also demonstrates the proposed model's applicability to its intended purpose.
15. This research is one of the first systematic studies of fixed and mobile air pollution sensing, and the proposed Deep AQ framework could generally be compared with the LUR model using in other areas and would further contribute to general application in more cities.

9.3 Limitations

Despite the important advantages of the research for high-resolution AQ inference at granular scale, various questions and challenges still remain unanswered. Important aspects under investigation are related to the regulated use of mobile sensors, the data constraints, and more.

1. Mobile sampling measurements are still limited and sparse compared to the sizeable study area, and the sensing coverage is unbalanced in different directions, which causes fitting problems in some locations with unbalanced spatio-temporal sensing coverage.
2. Convolution techniques tackling spatio-temporal data are still limited to small temporal neighbourhood slices; more temporal information from longer

neighbourhood slices is worthy of being incorporated as the gained features to train more advanced deep learning models.

3. The probabilistic ML-based methods such as Deep GPs are considered as very competitive methods of estimating AQ conditions; however, they are not studied and explored in this research, which was limited by the study period.
4. Density maps of LBS-R data can only present active users rather than real population density, and the data tends to exclude several groups of people in society, including children, the elderly, and the poor, considering the lower use of mobile phones in these groups.
5. To derive population dynamics using the LBS-R data, the census data was redistributed. It is assumed that each administrative region (i.e., study area) has a constant total population, as assuming that the overall population inflows and outflows will offset from each other and will not lead to a significant change in the total population in a short time period within the study area. As a result, the effects caused by inter-city migrations and movements are not considered in this study.
6. The Chengdu and Beijing urban case studies in this thesis were both conducted in summer and the experimental outcomes about the application and evaluation of the Deep AQ framework have still not been investigated in other seasons, especially in winter, which limits the transferability of the presented conclusions.
7. The simplified model may provide biased assessment results compared to the actual cumulative inhalation exposure, a result of neglecting the outdoor to indoor ratio of pollutant concentrations and assuming that the respiratory volume is not affected by different age groups, gender, physical activities, and other factors.

9.4 Future work

By constructing this thesis, it has been possible to identify 3 key areas on which to focus future research. The areas of focus are as follows: firstly, limitations of the current dataset; secondly, how transferrable the results are; and thirdly, the model's extension.

1. Data constraints

Within this thesis, the available datasets for AQ modelling are clearly capable of expansion to encompass a greater number of explanatory variables. Such variables might include emission inventory, urban OD flow dynamics, and the amount of power consumed by factories and buildings, in order to represent more accurate spatio-temporal variations between each source of emissions. In addition to ground-based measurements, the AOD data is introduced in Chapter 2, which facilitates the obtaining of AQ conditions in wider spatial areas; however, the AOD data is not utilised in this project. Indeed, remote sensing data has been utilised for low-resolution and large-scale AQ sensing. However, such data do not meet the required spatio-temporal resolution in this thesis, and can only be used as a supplementary dataset to validate the aggregation trends (such as city-wide average) of the results. As data acquisition and model adjustment require much more work beyond the scope of this research, this thesis will leave this to future study. In Chapter 3, the mobile sensing power is exhibited; however, the sampling routes of mobile campaigns are not pre-determined in a rigorous manner, which causes unbalanced spatio-temporal coverage in different directions and sub-areas. Consequently, the number of low-cost sensors and routes of probe vehicles should be enlarged and designed to achieve more effective spatio-temporal coverage and obtain more balanced spatio-temporal AQ sensing coverage. In Chapter 5, convolution techniques are proposed to manage the spatio-temporal data, but they are still limited to small neighbourhood segments and the features containing temporal proximity, periodicity, and trend are not combined and trained; more advanced deep learning structures are not used in this research, considering that the length of the study period is not great enough to train the appropriate model well. Thus, collecting data over a longer time-series (e.g., 3 months) to develop more advanced deep learning structures, including the combinations of FCN+LSTM / ResNets, should be important future work.

2. Transferability of results

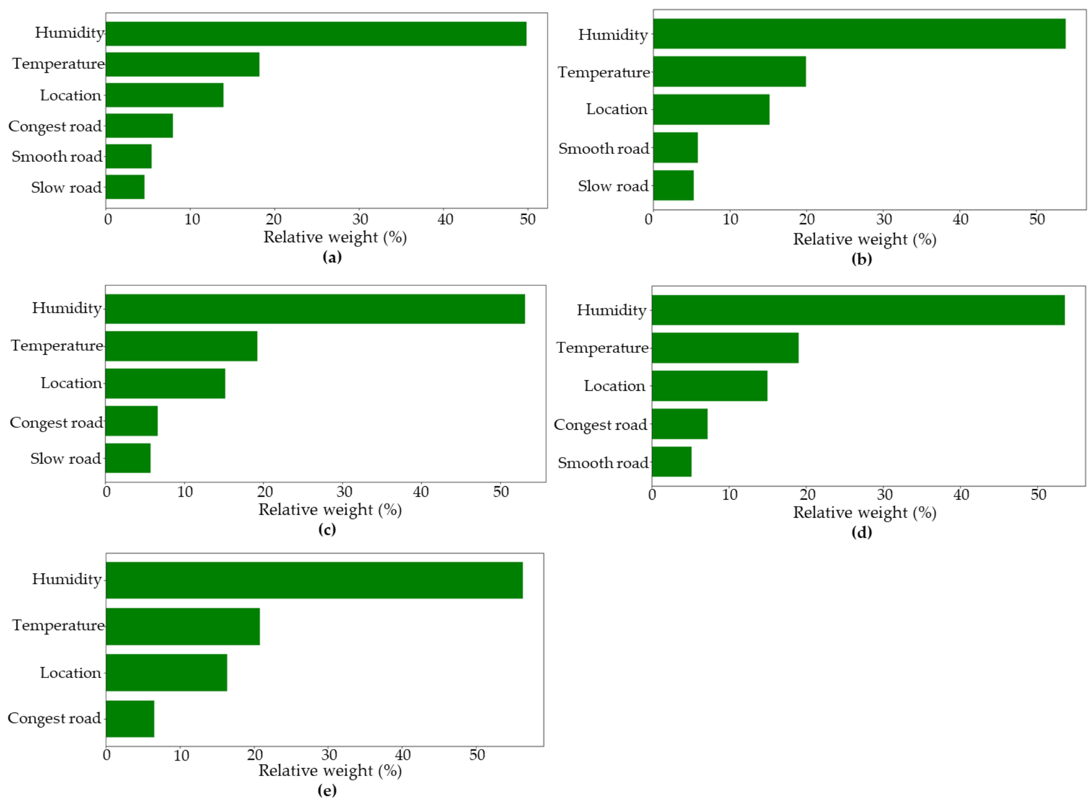
The inquiry that this thesis presents has expanded on extensive previous academic inquiry by developing on the available datasets, model innovation, and systematic case studies. The mobile sampling offers the potential to provide dynamic AQ sensing spatio-temporal coverage to reduce the unsampled locations, and the model offers the capability to achieve high-resolution (1 km × 1 km, hourly) AQ inference based on multi-source AQ samples (fixed or mobile) and urban features. In addition to the air pollutant species in this project, more air pollutant species such as SO_2, NO_x will be analysed in future work. However, the case studies in this project were conducted in summer, and it should be noted that the seasonal factors must be considered to extend how transferrable the results given in this thesis are. As a result, it ought to be considered how transferrable this research and its related results are. In order to explore such a question, this analysis ought to be broadened by incorporating other sampling campaigns for more case studies in various seasons. If these results are transferrable, this thesis's findings might be expanded in order to model air pollution adaptively.

3. Extension for model

The Deep AQ framework is capable of achieving high-resolution (1 km × 1 km, hourly) pixelwise AQ inference using multi-source AQ sensing (fixed or mobile) and urban features, as discussed in this thesis. However, the probabilistic advanced methods (e.g., Deep GPs) are not explored deeply and will instead be studied in subsequent academic studies. Besides, this research develops the online incremental learning system based on the Deep AQ framework, exploring high-resolution pixelwise AQ time-series prediction and an incident hot-spot identification model. High-resolution AQ inference enables the publication of AQ information to allow citizens to take action to reduce its health impacts and identify air pollution hot-spots to make route changes when driving in order to benefit their health (Mihăiţă et al., 2019). In the long-term, this would lead to a reconfiguration of cities based on human health prioritisation.

Appendix I

This appendix presents the sensitivity analysis of traffic parameters mentioned in section 3.4. To analyse the potential correlations between traffic parameters and AQ conditions, this research uses temperature, humidity, and traffic parameters as independent variables and uses the concentrations of particulate matter as the dependent variable. The GBDT is exploited as a surrogate model to construct functional relationships among independent variables and dependent label response. Appendix I Figure I (a)-(e) show that the percentage of congested roads in the grid-cell unit is the parameter of highest relative importance among traffic parameters, and the percentage of slow roads (blocked) in the grid-cell unit is the lowest. Appendix I Table I shows that the selection of the percentage of congested roads as the only traffic parameter offers the potential to yield the best model performance compared with other traffic parameters. Consequently, this research uses the percentage of congested roads to represent traffic conditions and constitutes this as the only proxy input parameter variable for the model training.



Appendix I Figure I: (a) Relative importance of humidity, temperature, location, traffic parameters (percentage of congestion road, slow road, smooth road in the grid-cell unit); (b) Relative importance of humidity, temperature, location, traffic parameters (percentage of slow road, smooth road in the grid-cell unit); (c) Relative importance of humidity, temperature, location, traffic parameters (percentage of congestion road, slow road in the grid-cell unit); (d) Relative importance of humidity, temperature, location, traffic parameters (percentage of congestion road, smooth road in the grid-cell unit); (e) Relative importance of humidity, temperature, location, traffic parameters (percentage of congestion road in the grid-cell unit);

Appendix I Table I: The performance of different traffic variable combinations.

Traffic variables	**SMAPE (%)**	**R^2**
CG+SL+SM	20.62	0.75
CG+SL	20.37	0.756
CG+SM	20.58	0.751
SL+SM	20.53	0.748
CG	20.07	0.759

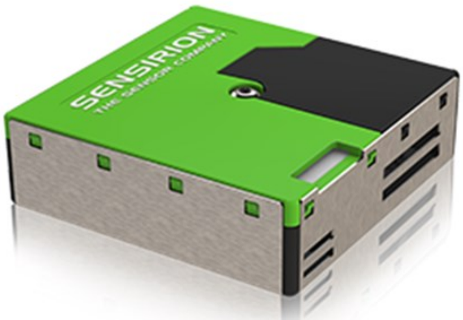
*CG= Percentage of the congested road in the grid-cell unit; SL=Percentage of the slow road in the grid-cell unit; SM=Percentage of the smooth road in the grid-cell unit;

Appendix II

This appendix presents the specification details of particulate matter, temperature, and humidity sensors.

Appendix II Table I: Specifications of particulate matter sensor (Sensirion, 2018)

Particulate Matter Sensor Specifications	
Mass concentration accuracy [1]	$\pm$10 ug/m^3 @ 0 to 100 ug/m^3 $\pm$10 % @ 100 to 1000 ug/m^3
Mass concentration range	1 to 1000 ug/m^3
Mass concentration resolution	1 ug/m^3
Particle detection size range [2]	Mass concentration: $PM_{1.0}$, $PM_{2.5}$, PM_4 and PM_{10} Number concentration: $PM_{0.5}$, $PM_{1.0}$, $PM_{2.5}$, PM_4 and PM_{10}
Lower limit of detection	0.3 um
Minimum sampling interval	1 sec (continuous mode)
Lifetime	> 8 years' operating continuously 24h/day
Dimensions	40.6×40.6×12.2 mm^3
Operating temperature range	-10 to +60 °C
Storage temperature range	-40 to +70 °C
Electrical Specifications	
Interface	UART, I^2C
Supply voltage	4.5 – 5.5 V
Average supply current@ 1 Hz measurement rate	< 60 mA

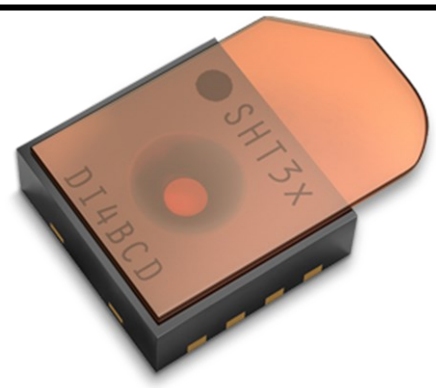


[1] Specified for $PM_{2.5}$ at 25 °C using potassium chloride salt particles and the TSI DustTrak™ DRX Aerosol Monitor 8533 as a reference.

[2] PM_x defines particles with a size smaller than 'x' micrometres (e.g., $PM_{2.5}$ = particles smaller than 2.5 um).

Appendix II Table II: Specifications of temperature and humidity sensors (Sensirion, 2018).

Temperature & Humidity Sensor Specifications	
Temperature operating range	-40 to +125° C (-40 to +257° F)
RH operating range	0 - 100% RH
Energy consumption	4.8μW (at 2.4 V, low repeatability, 1 measurement/s)
RH response time	8 sec (tau63%)
Dimensions	2.5×2.5× 0.9 mm^3
Packaging	Tape & Reel, sealed in ESD bag
Packaging sizes	2’500 & 10’000 pieces per reel
Electrical Specifications	
Interface	Voltage Out, I^2C
Supply voltage	2.15 to 5.5 V



Appendix III

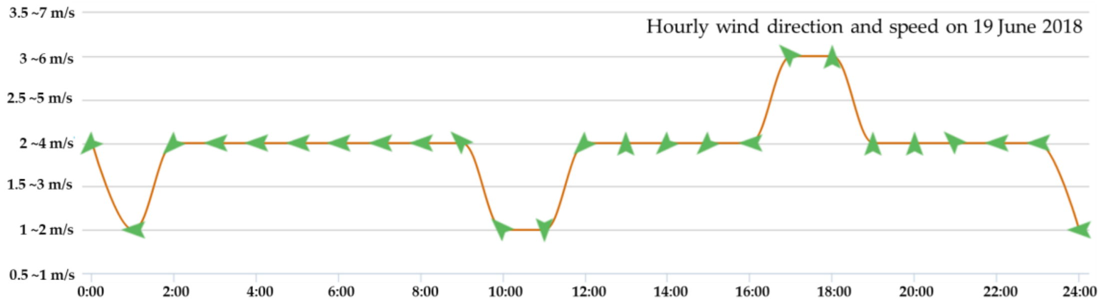
This appendix presents the results used in section 7.7.

Appendix III Table I: Accumulated errors of practicing the Deep AQ frameworks with various percentages of noises compared to the situation with no-adding noises

Methods	Noise variance	$PM_{2.5}$			PM_{10}		
		RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Deep AQ (Multi-AP)	1%	0.71	1.41	-0.019	0.79	0.79	-0.011
	2%	0.85	1.7	-0.036	1.65	1.31	-0.019
	3%	1.63	2.81	-0.043	1.99	2.43	-0.046
	4%	3.25	5.36	-0.059	2.82	2.52	-0.052
	5%	7.05	9.02	-0.128	3.64	3.55	-0.088
Deep AQ (GBDT)	1%	1.32	0.59	-0.004	0.07	0.02	-0.001
	2%	1.63	0.67	-0.011	0.11	0.05	-0.002
	3%	2.01	0.8	-0.018	0.15	0.1	-0.004
	4%	2.25	0.89	-0.025	0.21	0.19	-0.008
	5%	2.53	1.05	-0.03	0.25	0.26	-0.01

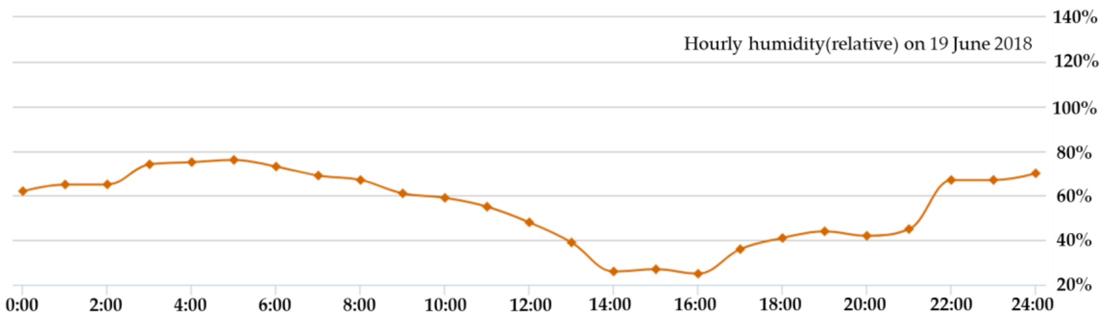
Appendix IV

This appendix presents some reasonable evidence for the depth discussion about the AQ inference in section 7.4. Appendix IV Figure I shows the temporal variations of wind speed and direction at each hour on 19 June 2018, which provides the possible meteorological reasons for explaining the dynamics of AQ distribution.



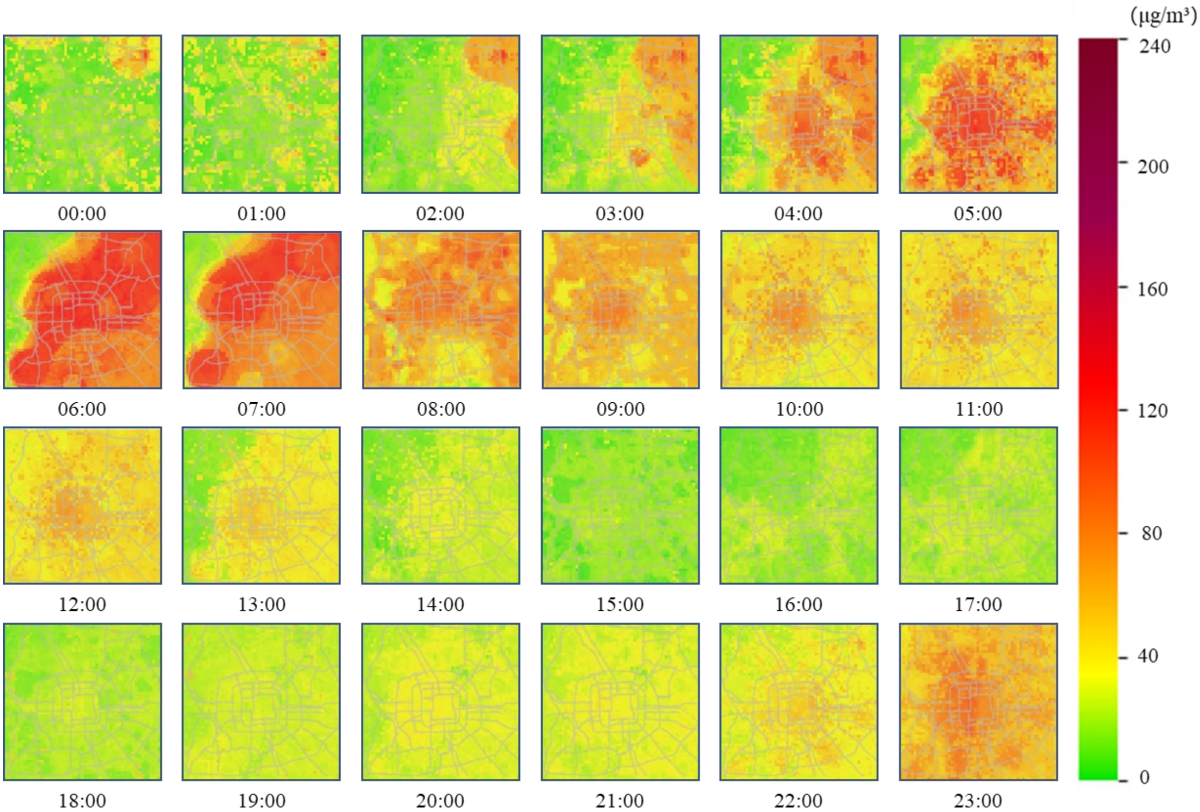
Appendix IV Figure I: Temporal variations of wind speed and direction at each hour on 19 June 2018.

Appendix IV Figure II shows the temporal variations of humidity(relative) at each hour on 19 June 2018, which provides the possible reasons for explaining the dissipation of $PM_{2.5}$ after 8:00 and the accumulation of $PM_{2.5}$ throughout 16:00-23:00.



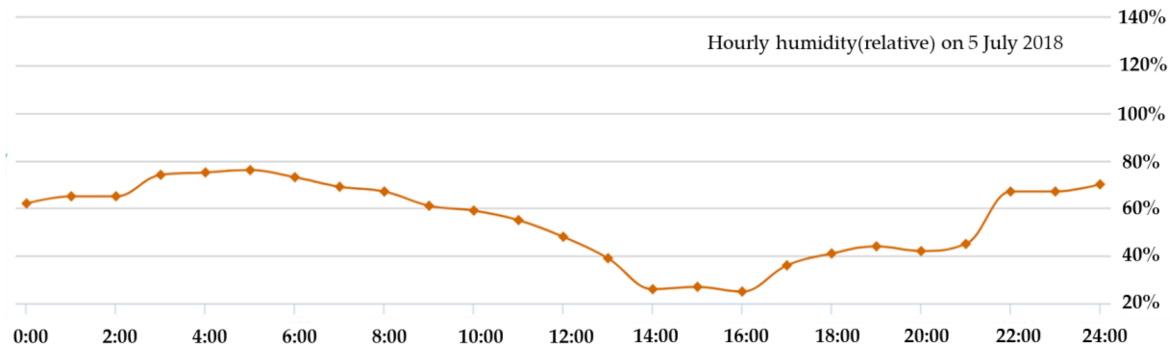
Appendix IV Figure II: Temporal variations of humidity(relative) at each hour on 19 June 2018.

Appendix IV Figure III shows the AQ inference mappings without mobile samples produced by the Deep AQ (GBDT, L + N + M) model between 00:00-23:00 on 19 June 2018. This provides a comparison with Figure 7.4 to prove that the increasing concentrations of $PM_{2.5}$ are mainly caused by the increased humidity and not influenced too much by the cross-sensitivity of low-cost sensors. In other words, if the low-cost sensors are influenced too much, Figure 7.4 should make a significant difference from Appendix IV Figure III.



Appendix IV Figure III: AQ inference mappings without mobile samples produced by the Deep AQ (GBDT, L + N + M) model between 0:00-23:00 on 19 June 2018.

Appendix IV Figure IV shows the temporal variations of relative humidity at each hour on 5 July 2018. In the hourly pollution maps in Figure 7.5, it is apparent that during the majority part of the day, there is a relatively high and steady concentrations. Then, there is a sudden dip between the three-hour timeframe of 16:00-19:00, which was likely due to the on-going thundershower. Appendix IV Figure IV provides reasonable explanation of the sudden change (the elbow of humidity time-series).



Appendix IV Figure IV: Temporal variations of humidity (relative) at each hour on 5 July 2018.

Appendix V

This appendix presents the sensitivity analysis of different sizes and numbers of ST-filters to evaluate the model's performance in both Chengdu and Beijing case study. Appendix V Table I shows the performance of different combinations for the Deep AQ (GBDT; L + N + M). The results show the best combination is to combine ST-filters with 3 × 3 and 5 × 5 receptive fields.

Appendix V Table I: The performance of the Deep AQ (GBDT; L + N + M) with various sizes and number of ST-filters

Case study	Combinations of ST-filters	$PM_{2.5}$			PM_{10}		
		RMSE (ug/m^3)	SMAPE (%)	R^2	RMSE (ug/m^3)	SMAPE (%)	R^2
Chengdu	3×3	7.47	20.75	0.807	10.48	16.39	0.838
	5×5	7.41	20.01	0.809	10.51	16.12	0.839
	7×7	7.34	19.67	0.805	10.49	15.91	0.837
	3 × 3 + 5 × 5	7.26	19.43	0.812	10.59	15.82	0.847
	3 × 3 + 5 × 5 + 7 × 7	7.26	19.45	0.812	10.60	15.82	0.848
Beijing	3×3	13.61	15.08	0.901	17.21	14.56	0.841
	5×5	13.44	14.94	0.903	17.03	14.02	0.853
	7×7	13.35	14.85	0.904	16.92	13.89	0.856
	3 × 3 + 5 × 5	13.24	14.72	0.909	16.84	13.66	0.863
	3 × 3 + 5 × 5 + 7 × 7	13.25	14.72	0.910	16.83	13.65	0.864

Appendix VI

This appendix presents the parameter settings and lists them in Appendix VI Table I.

Appendix VI Table I: The parameter settings in benchmarks

Case study	model	parameters
Chengdu	Ordinary kriging	Params = {‘variogram_model: ‘linear, ‘verbose’: False, ‘nlags’:5, ‘weight’: False, ‘anisotropy_scaling’:1.5, ‘anisotropy_angle’:0.0}
	KNN	Params = {‘n_neighbors’: 1500, ‘leaf_size’: 50, ‘metric’: ‘ Huber’, ‘p:2, ‘n_jobs’:50}
	SVR	Params = {‘kernel’: ‘rbf’, ‘degree’: 2, ‘gamma’: ‘auto deprecated’, ‘tol’:1e-5, ‘C’:1.5, ‘epsilon’:0.08, ‘shrinking’: True, ‘cache_size’:300, ‘max_iter’: -1}
	GPR	Params = {‘kernel: ‘rbf’, ‘theta0’:0.1, ‘thetaL’:0.004, ‘thetaU’:1, ‘nugget’:0.007}
	AE	Params = {‘gpu’:1, ‘lr’:0.0001, ‘max epoch’:10000, ‘batch size’:32’, ‘loss’: ‘Huber’} (10 layers are constructed and ReLU is used as activation function)
	DAL	Params = {‘gpu’:1, ‘lr’:0.0001, ‘max epoch’:10000, ‘batch size’:32’, ‘loss’: ‘Huber’} (10 layers are constructed and ReLU is used as activation function)
	DFeaST-Tree (RF)	Params = {‘n estimators’: 5000, ‘max depth’: 40, ‘min sample split’: 72, ‘learning rate’: 0.01, ‘presort’: ‘auto’, ‘loss’: ‘Huber’}

	DFeaST-Tree (GBDT)	Params = {'n estimators': 5000, 'max depth': 40, 'min sample split': 72, 'learning rate': 0.01, 'presort': 'auto', 'loss': 'Huber'}
	DFeaST-Net	Params = {'gpu':1, 'lr':0.0001, 'max epoch':10000, 'batch size':32', 'loss': 'Huber'} (10 layers are constructed and ReLU is used as activation function)
	DFeaST-Net (Multi-AP)	Params = {'gpu':1, 'lr':0.0001, 'max epoch':10000, 'batch size':32', 'loss': 'Huber'} (10 layers are constructed and ReLU is used as activation function)
Beijing	Ordinary kriging	Params = {'variogram_model: 'linear, 'verbose': False, 'nlags':3, 'weight': False, 'anisotropy_scaling':2.0, 'anisotropy_angle':0.0}
	KNN	Params = {'n_neighbors': 2000, 'leaf_size': 40, 'metric': ' Huber', 'p':1, 'n_jobs':50}
	SVR	Params = {'kernel': 'rbf', 'degree': 4, 'gamma': 'auto deprecated', 'tol':1e-1, 'C':1, 'epsilon':0.08, 'shrinking': True, 'cache_size':200, 'max_iter': -1}
	GPR	Params = {'kernel: 'rbf', 'theta0':0.17, 'thetaL':0.003, 'thetaU':1, 'nugget':0.02}
	AE	Params = {'gpu':1, 'lr':0.0001, 'max epoch':10000, 'batch size':32', 'loss': 'Huber'} (10 layers are constructed and ReLU is used as activation function)
	DAL	Params = {'gpu':1, 'lr':0.0001, 'max epoch':10000, 'batch size':32', 'loss': 'Huber'} (10 layers are constructed and ReLU is used as activation function)

	DFeaST-Tree (RF)	Params = {‘n estimators’: 5000, ‘max depth’: 40, ‘min sample split’: 72, ‘learning rate’: 0.01, ‘presort’: ‘auto’, ‘loss’: ‘Huber’}
	DFeaST-Tree (GBDT)	Params = {‘n estimators’: 5000, ‘max depth’: 40, ‘min sample split’: 72, ‘learning rate’: 0.01, ‘presort’: ‘auto’, ‘loss’: ‘Huber’}
	DFeaST-Net	Params = {‘gpu’:1, ‘lr’:0.0001, ‘max epoch’:10000, ‘batch size’:32’, ‘loss’: ‘Huber’} (10 layers are constructed and ReLU is used as activation function)
	DFeaST-Net (Multi-AP)	Params = {‘gpu’:1, ‘lr’:0.0001, ‘max epoch’:10000, ‘batch size’:32’, ‘loss’: ‘Huber’} (10 layers are constructed and ReLU is used as activation function)

References

Adams, M. D., & Kanaroglou, P. S. (2016). Mapping real-time air pollution health risk for environmental management: Combining mobile and stationary air pollution monitoring with neural network models. *Journal of Environmental Management*, *168*, 133–141.

Adams, W. C. (1993). Measurement of Breathing Rate and Volume in Routinely Performed Daily Activities: Final Report; Contract NO. A033-205; University of California: Davis, CA, USA.

Air Quality Egg. 2016. Retrieved from http://airqualityegg.com [Accessed: 5 January 2018].

Alvear, Ó. P. A. (2018). *Mobile sensing architecture for air pollution monitoring* [PhD Thesis]. Universitat Politècnica de València.

Amini, H., Shahri, S., Naddafi, K., Nabizadeh, R., & Yunesian, M. (2012). Estimating long-term chronic exposure to PM_{10} employing a land use regression model. *24th International Society for Environmental Epidemiology (ISEE) conference: Environmental Health across Land, Air and Sea*.

Analitis, A. , Barratt, B. , Green, D. , Beddows, A. , & Katsouyanni, K., 2020. Prediction of $PM_{2.5}$ concentrations at the locations of monitoring sites measuring PM_{10} and $N0_X$, using generalized additive models and machine learning methods: a case study in london. *Atmospheric Environment*, 240, 117757.

Analitis, A., Katsouyanni, K., Dimakopoulou, K., Samoli, E., Nikoloulopoulos, A. K., Petasakis, Y., Touloumi, G., Schwartz, J., Anderson, H. R., & Cambra, K. (2006). Short-term effects of ambient particles on cardiovascular and respiratory mortality. *Epidemiology*. 17, 230-233.

Anenberg, S. C., Horowitz, L. W., Tong, D. Q., & West, J. J. (2010). An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling. *Environmental Health Perspectives*, *118*(9), 1189–1195.

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C. J., Vermeulen, R. C., & Hamburg, S. P. (2017). High-resolution air pollution mapping with google street view cars: Exploiting big data. *Environmental Science & Technology,* 51(12), 6999-7008.

Atkinson, R. W., Kang, S., Anderson, H. R., Mills, I. C., & Walton, H. A. (2014). Epidemiological time series studies of PM2. 5 and daily mortality and hospital admissions: A systematic review and meta-analysis. *Thorax*, *69*(7), 660–665.

Bakar, K.S. (2012). Bayesian Analysis of Daily Maximum Ozone Levels. *PhD Thesis, University of Southampton*, Southampton, United Kingdom.

Beckx, C., Panis, L. I., Van De Vel, K., Arentze, T., Lefebvre, W., Janssens, D., & Wets, G. (2009). The contribution of activity-based transport models to air quality modelling: A validation of the albatross–aurora model chain. *Science of the Total Environment,* 407(12), 3814-3822.

Beckx, C., Panis, L. I., Vankerkom, J., Janssens, D., Wets, G., & Arentze, T. (2009). An integrated activity-based modelling framework to assess vehicle emissions: Approach and application. *Environment and Planning B: Planning and Design,* 36(6), 1086-1102.

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z. J., Weinmayr, G., Hoffmann, B., Wolf, K., Samoli, E., Fischer, P., & Nieuwenhuijsen, M. (2014). Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. *The Lancet*, *383*(9919), 785–795.

Bi, J., Wildani, A., Chang, H. H., & Liu, Y. (2020). Incorporating low-cost sensor measurements into high-resolution $PM_{2.5}$ modeling at a large spatial scale. *Environmental Science & Technology*, *54*(4), 2152–2162.

Binkowski, F. S., & Roselle, S. J. (2003). Model-3 community Multiscale air quality (CMAQ) model aerosol component 1. model description. *Journal of Geophysical Research: Atmospheres*, 108(D6).

Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering,* 60(1), 208-221.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, 7, 39–55.

Boldo, E., Medina, S., Le Tertre, A., Hurley, F., Mücke, H.-G., Ballester, F., & Aguilera, I. (2006). Apheis: health impact assessment of long-term exposure to $PM_{2.5}$ in 23 European cities. *European Journal of Epidemiology,* 21(6), 449-458.

Breathe London. 2018. Retrieved from https://www.breathelondon.org/

Breen, M. S., Long, T. C., Schultz, B. D., Crooks, J., Breen, M., Langstaff, J. E., Isaacs, K. K., Tan, Y.-M., Williams, R. W., & Cao, Y. (2014). GPS-based microenvironment tracker (micro trac) model to estimate time–location of individuals for air pollution exposure assessments: Model evaluation in central north Carolina. *Journal of Exposure Science & Environmental Epidemiology,* 24(4), 412-420.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*, CRC press.

Briggs, D. (2005). The role of GIS: Coping with space and time in air pollution exposure assessment. *Journal of Toxicology and Environmental Health, Part A,* 68(13-14), 1243-1261.

Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pryl, K., Van Reeuwijk, H., Smallbone, K., & Van Der Veen, A. (1997). Mapping urban air pollution using GIS: A regression-based approach. *International Journal of Geographical Information Science,* 11(7), 699-718.

Briggs, D. J., De, H. C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., & Smallbone, K. (2000). A regression-based method for mapping traffic-related air pollution: Application and testing in four contrasting urban environments. *Science of Total Environment*. 253:151–167.

Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., & Ryan, P. (2017). Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmospheric Environment*, 151, 1-11.

Calder, C. A. (2008). A dynamic process convolution approach to modeling ambient particulate matter concentrations. *Environmetrics: The official journal of the International Environmetrics Society*, 19(1), 39-48.

Cao, J., Yang, C., Li, J., Chen, R., Chen, B., Gu, D., & Kan, H. (2011). Association between long-term exposure to outdoor air pollution and mortality in China: a cohort study. *Journal of Hazardous Materials*,186, 1594-1600.

Carslaw, D. C. , & Ropkins, K. . (2012). Openair — an R package for air quality data analysis. *Environmental Modelling & Software*, 27, 52-61.

Castell, N., Schneider, P., Grossberndt, S., Fredriksen, Mirjam. F., Sousa-Santos, G., Vogt, M., & Bartonova, A. (2018). Localized real-time information on outdoor air quality at Kindergartens in Oslo, Norway using low-cost sensor nodes. *Environmental Research,* 165, 410-419.

Cavazza, J., & Murino, V. (2016). Active regression with adaptive Huber loss.

Champendal, A., Kanevski, M., & Huguenot, P.-E. (2014). Air pollution mapping using nonlinear land use regression models. In *International Conference on Computational Science and Its Applications* (pp. 682-690). Springer.

Chandramouli, V., Lingireddy, S., & Brion, G. M. (2007). Robust training termination criterion for back-propagation anns applicable to small data sets. *Journal of Computing in Civil Engineering, 21*(1), 39-46.

Chen, B., Song, Y., Jiang, T., Chen, Z., Huang, B., & Xu, B. (2018). Real-time estimation of population exposure to $PM_{2.5}$ using mobile-and station-based big data. *International Journal of Environmental Research and Public Health*, *15*(4), 573.

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Weinmayr, G., Bauwelinck, M., van Donkelaar, A., & Hvidtfeldt, U. A. (2020). Development of europe-wide models for particle elemental composition using supervised linear regression and random forest. *Environmental Science & Technology*, 54(24).

Chen, T., Tong, H., & Benesty, M. (2016). Xgboost: extreme gradient boosting.

Chen, X., Xu, X., Liu, X., Pan, S., He, J., Noh, H. Y., Zhang, L., & Zhang, P. (2018). PGA: Physics guided and adaptive approach for mobile fine-grained air pollution estimation. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 1321–1330.

Chen, Y.H., & Prinn, R. G. (2006). Estimation of atmospheric methane emissions between 1996 and 2001 using a three-dimensional global chemical transport model. *Journal of Geophysical Research: Atmospheres,* 111(D10).

Cheng, Y., Li, X., Li, Z., Jiang, S., Li, Y., Jia, J., & Jiang, X. (2014). AirCloud: a cloud-based air-quality monitoring system for everyone. *In Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems. ACM*, 251–265.

China Daily Report. 2018. Retrieved from http://www.ecns.cn/business/2018/04-18/299535.shtml

Chong, C.-Y., & Kumar, S. P. (2003). Sensor networks: Evolution, opportunities, and challenges. *Proceedings of the IEEE*, *91*(8), 1247–1256.

Chu, A. K. M., Kwok, R. C. W., & Yu, K. N. (2005). Study of pollution dispersion in urban areas using computational fluid dynamics (CFD) and geographic information system (GIS). *Environmental Modelling & Software,* 20(3), 273-277.

Chu, H. J., Ali, M. Z., & He, Y. C. (2020). Spatial calibration and $PM_{2.5}$ mapping of low-cost air quality sensors. *Scientific Reports*, *10* (1), 1–11.

Chu, Y., Liu, Y., Li, X., Liu, Z., Lu, H., Lu, Y., Mao, Z., Chen, X., Li, N., & Ren, M. (2016). A review on predicting ground $PM_{2.5}$ concentration using satellite aerosol optical depth. *Atmosphere,* 7(10), 129.

Citizen Sense. (2016). Retrieved from https://citizensense.net/about/people/jennifer-gabrys/.

Cristianini, N. (2014). Label (Labeled Data, Response, Dependent Variable). Dictionary of Bioinformatics and Computational Biology. *John Wiley & Sons, Ltd*.

Damianou, A. C. , & Lawrence, N. D. (2012). Deep gaussian processes. *Computer Science*, 207-215.

De L. F. (1989). Contributions of traffic emissions to mesoscale oxides of nitrogen and ozone concentrations. *Atmospheric Environment,* 23.

Denby et al. 2010. Guidance on the use of models for the European Air Quality Directive. ETC/ACC report. Retrieved from : http://fairmode.ew.eea.europa.eu/fol429189/forumsguidance/model_guidance_document_v6_2.pdf

Devarakonda, S., Sevusu, P., Liu, H., Liu, R., Iftode, L., & Nath, B. (2013). Real-time air quality monitoring through mobile sensing in metropolitan areas. Proceedings of the 2nd ACM SIGKDD international workshop on urban computing (pp. 1-8). ACM.

Ding, W. , Leung, Y. , Zhang, J. , & Fung, T. (2021). A hierarchical bayesian model for the analysis of spatio-temporal air pollutant concentrations and an application to air pollution analysis in Northern China. *Stochastic Environmental Research and Risk Assessment*, 1-35.

Dirgawati, M., Barnes, R., Wheeler, A. J., Arnold, A.-L., McCaul, K. A., Stuart, A. L., Blake, D., Hinwood, A., Yeap, B. B., & Heyworth, J. S. (2015). Development of land use regression models for predicting exposure to NO_2 and NO_x in metropolitan perth, Western Australia. *Environmental Modelling & Software*, 74, 258-267.

Doering, M. 2011. High-resolution large-scale air pollution monitoring: Approaches and challenges. In *Proceedings of the 3rd ACM international workshop on MobiArch* (pp. 5-10). ACM.

Dons, E., Panis, L. I., Van Poppel, M., Theunis, J., Willems, H., Torfs, R., & Wets, G. (2011). Impact of time–activity patterns on personal exposure to black carbon. *Atmospheric Environment*, *45*(21), 3594–3602.

Dons, E., Van Poppel, M., Kochan, B., Wets, G., & Panis, L. I. (2013). Modeling temporal and spatial variability of traffic-related air pollution: hourly land use regression models for black carbon. *Atmospheric Environment*, 74(aug.), 237-246.

Du, X., Kong, Q., Ge, W., Zhang, S., & Fu, L. (2010). Characterization of personal exposure concentration of fine particles for adults and children exposed to high ambient concentrations in Beijing, China. *Journal of Environmental Sciences*, *22*(11), 1757–1764.

Eckle, K., & Schmidt-Hieber, J. (2019). A comparison of deep networks with relu activation function and linear spline-type methods. *Neural Networks*.

Eeftens, M., Tsai, M. Y., Ampe, C., Anwander, B., Beelen, R., Bellander, T., Cesaroni, G., Cirach, M., Cyrys, J., & de Hoogh, K. (2012). Spatial variation of $PM_{2.5}$, PM_{10}, $PM_{2.5}$ absorbance and PM coarse concentrations between and within 20

European study areas and the relationship with NO_2 – Results of the ESCAPE project. *Atmospheric Environment*, 62, 303-317.

Esposito, E., De Vito, S., Salvato, M., Bright, V., Jones, R. L., & Popoola, O. (2016). Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. *Sensors and Actuators B: Chemical,* 231, 701-713.

Feng, X., Li, Q., Zhu, Y., Wang, J., Liang, H., & Xu, R. (2014). Formation and dominant factors of haze pollution over Beijing and its peripheral areas in winter. *Atmospheric Pollution Research*, 5(3), 528-538.

Frauke, H., Dorothea, S., Lilian, T., F Kateryna, Susanne, M., & Karl-Heinz, J., et al. (2016). Comparison of land-use regression modelling with dispersion and chemistry transport modelling to assign air pollution concentrations within the Ruhr area. *Atmosphere*, 7(3), 48.

Fakinle, B. S., Odekanle, E. L., Okunlola, O. S., Oke, D. O., Aremu, C. O., & Sonibare, J. A., et al. (2021). Spatial dispersion modeling of air emissions from a farm using a gaussian model. *Environmental Quality Management*.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.

Gamble, J. F. (1998). $PM_{2.5}$ and mortality in long-term prospective cohort studies: cause-effect or statistical associations. *Environmental Health Perspectives*, 106(9), 535-549.

Gavett, S. H., & Koren, H. S. (2001). The role of particulate matter in exacerbation of atopic asthma. *International Archives of Allergy & Immunology*, 124(1-3), 109-112.

Ge, Y., Jin, Y., Stein, A., Chen, Y., Wang, J., Wang, J., Cheng, Q., Bai, H., Liu, M., & Atkinson, P. M. (2019). Principles and methods of scaling geospatial Earth science data. *Earth-Science Reviews*, 197, 102897.

Giannadaki, D., Pozzer, A., & Lelieveld, J. (2013). Modeled global effects of airborne desert dust on air quality and premature mortality. *Atmospheric Chemistry and Physics*, 13(9).

Gibson, M. D., Kundu, S., & Satish, M. (2013). Dispersion model evaluation of $PM_{2.5}$,

NO_x and SO_2 from point and major line sources in nova scotia, canada using aermod gaussian plume air dispersion model. *Atmospheric Pollution Research,* 4(2), 157-167.

Gilbert, N. L., Goldberg, M. S., Beckerman, B., Brook, J. R., & Jerrett, M. (2005). Assessing spatial variability of ambient nitrogen dioxide in montréal, canada, with a land-use regression model. *Journal of the Air & Waste Management Association*, 55(8), 1059-1063.

Glatzel, T., Litterst, C., Cupelli, C., Lindemann, T., Moosmann, C., Niekrawietz, R., Streule, W., Zengerle, R., & Koltay, P. (2008). Computational fluid dynamics (CFD) software tools for microfluidic applications – a case study. *Computers & Fluids,* 37(3), 218-235.

Gulliver, J., & Briggs, D. J. (2004). Personal exposure to particulate air pollution in transport microenvironments. *Atmospheric Environment*, *38*(1), 1–8.

Gulliver, J., & de Hoogh, K. (2015). Environmental exposure assessment: Modelling air pollution concentrations. In R. DETELS (Ed.). *Oxford textbook of global public health* (pp. 1-21). Oxford University Press.

Gulliver, John, de Hoogh, K., Fecht, D., Vienneau, D., & Briggs, D. (2011). Comparative assessment of GIS-based approaches and metrics for estimating long-term exposures to air pollution. *Atmospheric Environment,* 45(39), 7072-7080.

Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y. C., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, *40*(30), 5880–5892.

Hankey, S., & Marshall, J. D., 2015. Land use regression models of on-road particulate air pollution (particle number, black carbon, $PM_{2.5}$, particle size) using mobile monitoring. *Environmental Science and Technology*, 49(15), 9194-202.

Hao, Y., Balluz, L., Strosnider, H., Wen, X. J., Li, C., & Qualters, J. R. (2015). Ozone, fine particulate matter, and chronic lower respiratory disease mortality in the United States. *American Journal of Respiratory and Critical Care Medicine,* 192(3), 337-341.

Harrison, R. M., Smith, D. J. T., & Kibble, A. J. (2004). What is responsible for the carcinogenicity of $PM_{2.5}$? *Occupational and Environmental Medicine*, *61*(10), 799–805.

Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Arn, T., Beutel, J., & Thiele, L. (2015). Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing*, *16*, 268–285.

Hamelijnck. O., Damoulas. T., Wang. K. & Girolami. M. (2019). Multi-resolution Multi-task Gaussian Processes. *In Advances in Neural Information Processing Systems* (NeurIPS'19).

Hennig, F., Sugiri, D., Tzivian, L., Fuks, K., Moebus, S., Jöckel, K. H., Vienneau, D., Kuhlbusch, T. A., De Hoogh, K., & Memmesheimer, M. (2016). Comparison of land-use regression modelling with dispersion and chemistry transport modelling to assign air pollution concentrations within the ruhr area. *Atmosphere*, 7(3), 48.

Herwehe, J A., J.K S. Ching, and J. SWALL.(2004). Stochatstic description of subgrid pollutant variability in CMAQ. Presented at 2004 Models-3 Conference, Chapel Hill, NC, October 18 - 20.

Ho, H. C., Knudby, A., & Huang, W. (2015). A spatial framework to map heat health risks at multiple scales. *International Journal of Environmental Research and Public Health,* 12(12), 16110-16123.

Ho, H. C., Wong, M. S., Yang, L., Shi, W., Yang, J., Bilal, M., & Chan, T.C. (2018). Spatio-temporal influence of temperature, air quality, and urban environment on cause-specific mortality during hazy days. *Environmental International,* 112, 10-22.

Hoek, G., Beelen, R., De Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment,* 42(33), 7561-7578.

Hoek, G., Brunekreef, B., Goldbohm, S., Fischer, P., & van den Brandt, P. A. (2002). Association between mortality and indicators of traffic-related air pollution in the Netherlands: A cohort study. *Lancet* 360:1203–1209.

Hoek, G., Fischer, P., Van Den Brandt, P., Goldbohm, S., & Brunekreef, B. (2001).

Estimation of long-term average exposure to outdoor air pollution for a cohort study on mortality. *Journal of Exposure Analysis & Environmental Epidemiology*. 11:459–469.

Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., & Kaufman, J. D. (2013). Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environmental Health*,12, 43-57.

Holgate, S., Grigg, J., Agius, R., Ashton, J. R., Cullinan, P., Exley, K., Fishwick, D., Fuller, G., Gokani, N., & Griffiths, C. (2016). Every breath we take: the lifelong impact of air pollution, *Report of a working party. R. Coll. Physicians*. 1-94

http://www.gov.cn/xinwen/2018-08/26/content_5316749.htm (Accessed on 23 February 2020).

Hu, K., Rahman, A., Bhrugubanda, H., & Sivaraman, V. (2017). Hazeest: Machine learning based metropolitan air pollution inference from fixed and mobile sensors. *IEEE Sensors Journal,* 17(11), 3517-3525.

Hu, Z. (2009). Spatial analysis of modis aerosol optical depth, $PM_{2.5}$, and chronic coronary heart disease. *International Journal of Health Geographics,* 8(1), 27.

Huang G.H. (1992). A stepwise cluster analysis method for predicting air quality in an urban environment. *Atmospheric Environment. Part B.* Urban Atmosphere.

Huang, D., Zhang, S., 2013. Health benefit evaluation for $PM_{2.5}$ pollution control in Beijing-Tianjin-Hebei region of China. *China Environmental Science*, 33, 166-174.

Jason B. (2018). A Gentle Introduction to k-fold Cross-Validation. Retrieved from https://machinelearningmastery.com/k-fold-cross-validation/

Ji, Y., Liu, H., Liu, X., Ding, Y., & Luo, W. (2016). A comparison of road-network-constrained trajectory compression approaches. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)* (pp. 256-263). IEEE.

Jing, J., & Lin, J. (2012). A Signal Denoise Algorithm Based on Wavelet Transform. *International Conference on Industrial Control & Electronics Engineering*. IEEE.

Kamvar, S. D. 2002. Interpreting and extending classical agglomerative clustering algorithms using a model-based approach. *Proceedings of ICML*, 2002.

Katsouyanni, K., Touloumi, G., Spix, C., Schwartz, J., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., & Bacharova, L. (1997). Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. *British Medical Journal*. 314 (1658).

Ke, H., Rahman, A., Bhrugubanda, H., and Sivaraman, V. 2017. HAZEST: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors. *IEEE Sensors Journal,* 17 (11)**,** 3517-3525.

Kloog, I., Koutrakis, P., Coull, B. A., Lee, H. J., & Schwartz, J. (2011). Assessing temporally and spatially resolved $PM_{2.5}$ exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment,* 45(35), 6267-6275.

Kloog, I., Nordio, F., Coull, B. A., & Schwartz, J. (2012). Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatio temporal $PM_{2.5}$ exposures in the mid-atlantic states. *Environmental Science & Technology,* 46(21), 11913-11921.

Kramer, O. (2013). K-nearest neighbours. *Intelligent Systems Reference Library*, 41(2), 13-23.

Kreuzer, A., Valle, L. D., & Czado, C. (2019). A Bayesian non-linear state space copula model to predict air pollution in Beijing.

Krupnick, A. J. (2008). Challenges to managing air pollution. *Journal of Toxicology and Environmental Health, Part A*, *71*(1), 13–23.

Kumar, P., Jain, S., Gurjar, B. R., Sharma, P., Khare, M., Morawska, L., & Britter, R. (2013). New directions: Can a ‘blue sky’ return to Indian megacities? *Atmospheric Environment*, *71*, 198–201.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., & Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environment International*, *75*, 199–205.

Lee, H. J., Liu, Y., Coull, B. A., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. *Atmospheric*

Chemistry and Physics, *11*(15), 7991–8002.

Lee, K. H., Li, Z., Kim, Y. J., & Kokhanovsky, A. (2009). Atmospheric aerosol monitoring from satellite observations: A history of three decades. *Atmospheric and Biological Environmental Monitoring*, 13–38.

Leung, Y., Zhou, Y., Lam, K. Y., Fung, T., & H Jung. (2019). Integration of air pollution data collected by mobile sensors and ground-based stations to derive a spatiotemporal air pollution profile of a city. *International Journal of Geographical Information Science*, 33(11), 1-23.

Li, J., Carlson, B. E., & Lacis, A. A. (2015). How well do satellite AOD observations represent the spatial and temporal variability of $PM_{2.5}$ concentration for the United States? *Atmospheric Environment*, *102*, 260–273.

Liang, W., Luo, S., Zhao, G., & Wu, H. (2020). Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms.

Lim, C. C., Kim, H., Vilcassim, M. R., Thurston, G. D., Gordon, T., Chen, L.-C., Lee, K., Heimbinder, M., & Kim, S.Y. (2019). Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in seoul, South Korea. *Environment International*, 131, 105022.

Liu, L.J. S., & Rossini, A. J. (1996). Use of kriging models to predict 12-hour mean ozone concentrations in Metropolitan Toronto—A pilot study. *Environmental International*. 22:677–692.

Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V., & Sarnat, J. A. (2004). Mapping annual mean ground-level $PM_{2.5}$ concentrations using multiangle imaging spectroradiometer aerosol optical thickness over the contiguous United States. *Journal of Geophysical Research: Atmospheres,* 109, D22206.

Lü, J., Liang, L., Feng, Y., Li, R., & Liu, Y. (2015). Air pollution exposure and physical activity in China: Current knowledge, public health implications, and future research needs. *International Journal of Environmental Research and Public Health*, *12*(11), 14887–14897.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640-651.

Ma, R., Liu, N., Xu, X., Wang, Y., Noh, H. Y., Zhang, P., & Zhang, L. (2020). Fine-grained air pollution inference with mobile sensing systems: a weather-related deep autoencoder model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*.

Ma, Z., Hu, X., Huang, L., Bi, J., & Liu, Y. (2014). Estimating ground-level PM2. 5 in China using satellite remote sensing. *Environmental Science & Technology*, *48*(13), 7436–7444.

Maag, B., Zhou, Z., & Thiele, L. (2018). W-air: Enabling personal air pollution monitoring on wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(1), 1–25.

Maag, B., Zhou, Z., Saukh, O., & Thiele, L. (2017). SCAN: Multi-hop calibration for mobile sensor arrays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, No. 19.

Madaniyazi, L., Nagashima, T., Guo, Y., Yu, W., & Tong, S. (2015). Projecting fine particulate matter-related mortality in East China. *Environment Science & Technology*. 49, 11141-11150

Maddix, M., & MD Adams. (2020). Effects of spatial sampling density and spatial extent on linear land use regression modelling NO_2 estimates in an automobile-oriented city. *Atmospheric Environment*, 238, 117735.

Maimon, O. Z., & Rokach, L. (2014). *Data mining with decision trees: Theory and applications* (Vol. 81). World scientific.

Marjovi, A., Arfire, A., & Martinoli, A. (2015). High resolution air pollution maps in urban environments using mobile sensor networks. In *2015 International Conference on Distributed Computing in Sensor Systems* (pp. 11-20). IEEE.

Marjovi, A., Arfire, A., & Martinoli, A. (2017). Extending urban air quality maps beyond the coverage of a mobile sensor network: Data sources, approaches, and performance evaluation. *In Proceedings of the International Conference on Embedded Wireless Systems and Networks*. ACM.

Masroor, K. , Fanaei, F. , Yousefi, S. , Raeesi, M. , & Hadei, M. (2020). Spatial modelling of $PM_{2.5}$ concentrations in tehran using kriging and inverse distance weighting (idw) methods. *Journal of Air Pollution and Health*, 5(2), 89-96.

Mavko, M. E., Tang, B., & George, L. A. (2008). A sub-neighbourhood scale land use regression model for predicting NO_2. *Science of the Total Environment*, 398(1-3), 68-75.

Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J. J., McLeod, M. W., Hodgson, T. F., & Dicks, J. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, *70*, 186–203.

MEDE 2012. *Bilan de la qualit ́e de l'air en France en 2012* (Technical Report). Paris: Minist ́ ere de l' ́ ecologie et du d ́ eveloppement durable, Direction G ́ en ́ erale de l'Energie et du Climate.

Mei, H., Han, P., Wang, Y., Zeng, N., Liu, D., Cai, Q., Deng, Z., Wang, Y., Pan, Y., & Tang, X. (2020). Field evaluation of low-cost particulate matter sensors in beijing. *Sensors*, 20(16), 4381.

Meiring, W., Sampson, P. D., & Guttorp, P. (1998). Spatio-temporal estimation of grid-cell hourly ozone levels for assessment of a deterministic model. *Environmental and Ecological Statistics*, 5(3), 197-222.

Meng, X., Chen, L., Cai, J., Zou, B., Wu, C.F., Fu, Q., Zhang, Y., Liu, Y., & Kan, H. (2015). A land use regression model for estimating the NO_2 concentration in Shanghai, China. *Environmental Research,* 137, 308-315.

Mihăiţă, A. S., Dupont, L., Chery, O., Camargo, M., & Cai, C. (2019). Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling. *Journal of Cleaner Production*, *221*, 398–418.

Minet, L., Gehr, R., & Hatzopoulou, M. (2017). Capturing the sensitivity of land-use regression models to short-term mobile monitoring campaigns using air pollution micro-sensors. *Environmental Pollution,* 230, 280-290.

Mitchell, T. M. (1997). *Machine learning*, McGraw-Hill.

Morley, D. W., & Gulliver, J. (2018). A land use regression variable generation, modelling and prediction tool for air pollution exposure assessment.

Environmental Modelling & Software, 105, 17-23.

Muttoo, S., Ramsay, L., Brunekreef, B., Beelen, R., Meliefste, K., & Naidoo, R. N. (2018). Land use regression modelling estimating nitrogen oxides exposure in industrial South Durban, South Africa. *Science of the Total Environment,* 610, 1439-1447.

NASA. 2017. Retrieved from: https://arset.gsfc.nasa.gov/airquality/webinars/introduction-satellite-remote-sensing-air-quality-applications.

Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 25). Determination press San Francisco, CA.

Nielsen, O. R., Andersen, Z. J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M. J., & Brunekreef, B. (2013). 'Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE),' *The Lancet Oncology*, vol. 14, no. 9, pp. 799–805, 2013.

Nori-Sarma, A. , Thimmulappa, R. K. , Venkataramana, G. V. , Fauzie, A. K. , & Bell, M. L. (2020). Low-cost NO_2 monitoring and predictions of urban exposure using universal kriging and land-use regression modelling in mysore, india. *Atmospheric Environment*, 226, 117395.

O'Keeffe, K. P., Anjomshoaa, A., Strogatz, S. H., Santi, P., & Ratti, C. (2019). Quantifying the sensing power of crowd-sourced vehicle fleets. *PNAS,*116(26) 12752-12757.

Ottosen, T.-B., Kakosimos, K. E., Johansson, C., Hertel, O., Brandt, J., Skov, H., Berkowicz, R., Ellermann, T., Jensen, S. S., & Ketzel, M. (2015). Analysis of the impact of inhomogeneous emissions in the operational street pollution model (OSPM). *Geoscientific Model Development,* 8(10), 3231-3245.

Paciorek, C. J., Liu, Y., Moreno-Macias, H., & Kondragunta, S. (2008). Spatio-temporal associations between goes aerosol optical depth retrievals and ground-level $PM_{2.5}$. *Environmental science & technology,* 42(15), 5800-5806.

Park, Y. M., & Kwan, M.-P. (2017). Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health & Place*, 43, 85-94.

Parra, M. A., Santiago, J. L., Martín, F., Martilli, A., & Santamaría, J. M. (2010). A methodology to urban air quality assessment during large time periods of winter using computational fluid dynamic models. *Atmospheric Environment,* 44(17), 2089-2097.

Pizor, P. J. (1987). Principles of geographical information systems for land resources assessment. *Soil Science*, *144*(4), 306

Plaia, A., & Ruggieri, M. (2011). Air quality indices: a review. *Reviews in Environmental Science and Bio/Technology*, 10, 165–179.

Pope III, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association*, 56(6), 709-742.

Pope III, C. A., Burnett, R. T., Turner, M. C., Cohen, A., Krewski, D., Jerrett, M., Gapstur, S. M., & Thun, M. J. (2011). Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure-response relationships. *Environment Health Perspectives*. 119, 1616-1621.

Qi, Z., Wang, T., Song, G., Hu, W., Li, X., & Zhang, Z. (2018). Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. *IEEE Transactions on Knowledge and Data Engineering,* 30(12), 2285-2297.

Qianliyan Initiative, 2018. Ministry of Ecological Environment, People's Republic of China.Retrieved from

Qiu, H., Yu, I. T., Wang, X., Tian, L., Tse, L. A., & Wong, T. W. (2013). Differential effects of fine and coarse particles on daily emergency cardiovascular hospitalizations in Hong Kong. *Atmospheric Environment*, *64*, 296–302.

Ross, Z., Jerrett, M., Ito, K., Tempalski, B., & Thurston, G. D. (2007). A land use regression for predicting fine particulate matter concentrations in the new york city region. *Atmospheric Environment*, 41(11), 2255-2269.

Ryan, P. H., & LeMasters, G. K. (2007). A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhalation toxicology*, 19(sup1), 127-133.

Ryan, P. H., LeMasters, G. K., Levin, L., Burkle, J., Biswas, P., Hu, S., Grinshpun, S., & Reponen, T. (2008). A land-use regression model for estimating micro environmental diesel exposure given multiple addresses from birth through childhood. *Science of the Total Environment*, 404(1), 139-147.

Sahu, S. K.; Mardia, K. V. (2005) A Bayesian kriged Kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society*, 54, 223–244.

Sahu, S. K., Gelfand, A. E., & Holland, D. M. (2006). Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1), 61-86.

Saukh, O., Hasenfratz, D., Walser, C., & Thiele, L. (2014). On rendezvous in mobile sensing networks. *Lecture Notes in Electrical Engineering*, 281, 29-42.

Satapathy, S. M., & Rath, S. K. (2014). Use Case Point Approach Based Software Effort Estimation using Various Support Vector Regression Kernel Methods. *India Software Engineering Conference* (pp.1-10). ACM.

Sensirion. (2018). Retrieved from https://www.sensirion.com/en/environmental-sensors/particulate-matter-sensors-pm25/particulate-matter-sensor-sps30/

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.K., & Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* (pp. 802-810). ACM.

Shi, Y., Lau, K. K.L., & Ng, E. (2016). Developing street-level $PM_{2.5}$ and PM_{10} land use regression models in high-density Hongkong with urban morphological factors. *Environmental Science and Technology*, 8178-8187.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., & Jonson, J. E. (2012). The EMEP MSC-W chemical transport model-technical description. *Atmospheric Chemistry and Physics*, 12(16), 7825-7865.

Slama, R., Morgenstern, V., Cyrys, J., Zutavern, A., Herbarth, O., Wichmann, H.-E., Heinrich, J., & Group, L. S. (2007). Traffic-related atmospheric pollutants levels during pregnancy and offspring's term birth weight: A study relying on a land-use regression exposure model. *Environmental Health Perspectives,* 115(9), 1283-1292.

Sloan, C. D., Philipp, T. J., Bradshaw, R. K., Chronister, S., Barber, W. B., & Johnston, J. D. (2016). Applications of GPS-tracked personal and fixed-location $PM_{2.5}$ continuous exposure monitoring. *Journal of the Air & Waste Management Association*, *66*(1), 53–65.

SM, S. N., Yasa, P. R., Narayana, M. V., Khadirnaikar, S., & Rani, P. (2019). Mobile monitoring of air pollution using low cost sensors to visualize spatio-temporal variation of pollutants at urban hotspots. *Sustainable Cities and Society*, 44, 520-535.

Smart Citizen Kit. 2016. Retrieved from https://smartcitizen.me/

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3).

Song, Y., Huang, B., He, Q., Chen, B., Wei, J., & Mahmood, R. (2019). Dynamic assessment of $PM_{2.5}$ exposure and health risk using remote sensing and geo-spatial big data. *Environmental Pollution,* 253, 288-296.

Song, J. , & Han, K. (2019). Exploring Urban Air Quality with MAPS: Mobile Air Pollution Sensing. arXiv:1904.12303v1

Song, J. , & Han, K. and Stettler, M. (2021). 'Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing', *IEEE Internet of Things Journal*, DOI: 10.1109/JIoT.2020.3041047

State of Global Air 2019: Air pollution a significant risk factor worldwide | Health Effects Institute. Retrieved from https://www.healtheffects.org/announcements/state-global-air-2019-air-pollution-significant-risk-factor-worldwide

State of Global Air 2020: Air Pollution's Impact on Health: A Global Snapshot | Health Effects Institute. Retrieved from https://www.stateofglobalair.org/sites/default/files/documents/2020-10/soga-global-profile-factsheet.pdf

Stieb, D. M., Burnett, R. T., Smith-Doiron, M., Brion, O., Shin, H. H., & Economou, V. (2008). A new multipollutant, no-threshold air quality health index based on short-term associations observed in daily time-series analyses. *Journal of the Air & Waste Management Association*, 58, 435–450.

Su, J. G., Brauer, M., Ainslie, B., Steyn, D., Larson, T., & Buzzelli, M. (2008). An innovative land use regression model incorporating meteorology for exposure analysis. *Science of the Total Environment*, 390(2-3), 520-529.

Suriano, D., Prato, M., Pfister, V., Cassano, G., Camporeale, G., Dipinto, S., & Penza, M. (2015). Stationary and low-cost mobile gas sensor-systems for air quality monitoring applications. *Fourth Scientific Meeting EuNetAir*, Linkoping University, Linkoping, Sweden.

Talampas, M. C. R., & Low, K. (2012). Maximum likelihood estimation of ground truth for air quality monitoring using vehicular sensor networks. *TENCON 2012 IEEE Region 10 Conference*, 1–6.

Tang, B., Chen, Z., Hefferman, G., Wei, T., He, H., & Yang, Q. (2015). A hierarchical distributed fog computing architecture for big data analysis in smart cities, *Proceedings of the Ase Big data & Social Informatics 2015* (pp. 1-6).

Tianqihoubao, 2018. Retrieved from http://www.tianqihoubao.com/lishi/beijing.html.

Tianqihoubao, 2019. Retrieved from http://www.tianqihoubao.com/lishi/chengdu.html.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*. 46:234–240.

Valavi, R. , Elith, J. , Lahoz-Monfort, J. J. , & Guillera-Arroita, G. (2018). BlockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*.

Van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level $PM_{2.5}$ using aerosol optical depth determined from satellite remote sensing. *Journal of*

Geophysical Research: Atmospheres, 111(D21).

Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, *118*(6), 847–855.

Vitolo, C., Scutari, M., Ghalaieny, M., Tucker, A., & Russell, A. (2018). Modelling air pollution, climate and health data using Bayesian networks: a case study of the English regions. *Earth & Space Science*, 5(12).

Wallace, J., Corr, D., Deluca, P., Kanaroglou, P., & McCarry, B. (2009). Mobile monitoring of air pollution in cities: The case of Hamilton, Ontario, Canada. *Journal of Environmental Monitoring*, *11*(5), 998–1003.

Wang, A., Xu, J., Tu, R., Saleh, M., & Hatzopoulou, M. (2020). Potential of machine learning for prediction of traffic related air pollution. *Transportation Research Part D: Transport and Environment*, 88, 102599.

Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and $PM_{2.5}$ mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21), Article 21.

Wang, X. L., & Li, Z. B. (2005). Identifying the parameters of the kernel function in support vector machines based on the grid-search method. *Journal of Ocean University of Qingdao*.

Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., & Brunekreef, B. (2012). Systematic evaluation of land use regression models for NO_2. Environmental Science & Technology, 46(8), p.4481-4489.

Wang, M., Zhu, T., Zheng, J., Zhang, R. Y., Zhang, S. Q., Xie, X. X., Han, Y. Q., & Li, Y. (2009). Use of a mobile laboratory to evaluate changes in on-road air pollutants during the Beijing 2008 Summer Olympics. *Atmospheric Chemistry and Physics*, *9*(21), 8247–8263.

Wang, K., Hamelijnck, O., Damoulas, T., & Steel, M. (2020). Nonstationary Nonseparable Random Fields. In Proceedings of the 37th International Conference on Machine Learning (ICML), Online

Wang, L., Bi, J., Meng, X., Geng, G., & Y Liu. (2020). Satellite-based assessment of the long-term efficacy of $PM_{2.5}$ pollution control policies across the Taiwan strait. *Remote Sensing of Environment*, 251, 112067.

Wheeler, A. J., Smith-Doiron, M., Xu, X., Gilbert, N. L., & Brook, J. R. (2008). Intra-urban variability of air pollution in windsor, ontario—measurement and modeling for human exposure assessment. *Environmental Research*, 106(1), 7-16.

White, R. M., Paprotny, I., Doering, F., Cascio, W. E., Solomon, P. A., & Gundel, L. A. (2012). Sensors and 'apps' for community-based atmospheric monitoring. *EM: Air and Waste Management Association's Magazine for Environmental Managers*, 5, 36-40.

WHO. 1999. Monitoring ambient air quality for health impact assessment. Copenhagen: WHO Regional Office for Europe.

Wire, T. E. 2019. *Global Air Quality Sensor Market Report 2019-2024*. Retrieved from https://www.theexpresswire.com/pressrelease/Air-Quality-Sensor-Market-2019-Size-by-Product-Types-End-users-Regional-Outlook-Growth-Potential-Price-Trends-and-Forecast-to-2024-Absolute-Reports_10485953.

Xu, W., EA Riley, Austin, E., Sasakura, M., Schaal, L., & Gould, T. R., et al. (2017). Use of mobile and passive badge air monitoring data for NOx and ozone air pollution spatial exposure prediction models. *Journal of Exposure ence & Environmental Epidemiology*.

Xu, Y., Ho, H. C., Wong, M. S., Deng, C., Shi, Y., Chan, T.-C., & Knudby, A. (2018). Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level $PM_{2.5}$. *Environmental Pollution*, *242*, 1417–1426.

Yanosky, J. D., Paciorek, C. J., & F Laden. (2014). Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors. *Environmental Health*, 13(1), 63.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., & Li, Z. (2018). Deep multi-view spatio-temporal network for taxi demand prediction. *In 32th AAAI Conference on Artificial Intelligence*.

Yin, H., Pizzol, M., & Xu, L. (2017). External costs of $PM_{2.5}$ pollution in beijing, china: uncertainty analysis of multiple health impacts and costs. *Environmental Pollution,* 226, 356-369.

Zanobetti, A., Dominici, F., Wang, Y., & Schwartz, J. D. (2014). A national case-crossover analysis of the short-term effect of $PM_{2.5}$ on hospitalizations and mortality in subjects with diabetes and neurological disorders. *Environmental Health*, *13*(1), 1–11.

Zappi, P., Bales, E., Park, J. H., Griswold, W., & Rosing, T. Š. (2012). The citisense air quality monitoring mobile sensor node. In *Proceedings of the 11th ACM/IEEE conference on information processing in sensor networks, Beijing, China*.

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. *In the 31th AAAI Conference on Artificial Intelligence*.

Zhang, Lihui, Yin, Y., & Chen, S. (2013). Robust signal timing optimization with environmental concerns. *Transportation Research Part C: Emerging Technologies,* 29, 55-71.

Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems,* 11(2), 68-75.

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., & Qiu, Y. (2017). Analysis of influential factors for the relationship between $PM_{2.5}$ and AOD in Beijing. *Atmospheric Chemistry and Physics*, *17*(21), 13473–13489.

Zheng, Y., Liu, F., & Hsieh, H.-P. (2013). U-air: When urban air quality inference meets big data. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1436–1444.

Zhou, Y., Chang, F.J., Chang, L.C., Kao, I.F., & Wang, Y.S. (2019). Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *Journal of Cleaner Production,* 209, 134-145.

Zhong, P. R., Huang, S. C., Zhu, Y. H., Yang, L. I., & Lu, M. A. (2018). Application of land use regression model in simulation of spatial distribution of ambient NO_2

concentration. *Journal of Environment and Health*.

Zuurbier, M., Hoek, G., Oldenwening, M., Lenters, V., Meliefste, K., Van Den Hazel, P., & Brunekreef, B. (2010). Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental Health Perspectives*, *118*(6), 783–789.

Zuurbier, M., Hoek, G., Oldenwening, M., Meliefste, K., Van, D. H., P., & Brunekreef, B. (2011). Respiratory effects of commuters' exposure to air pollution in traffic. *Epidemiology*, 219–227.