Dependable Monitoring with Low-cost Sensor Networks and Visualisation for Urban Air Quality

by

Hoang Trung Le

Thesis submitted in fulfilment of the requirements for the degree of
Master of Engineering (Research)
under the supervision of Associate Professor Quang Ha

School of Electrical and Data Engineering
Faculty of Engineering and Information Technology
University of Technology Sydney
3rd November, 2024

Certificate of Original Authorship

I, Hoang Trung Le, declare that this thesis is submitted in fulfilment of the requirements for the award of Master of Engineering (Research), in the School of Electrical and Data Engineering of the Faculty of Engineering and IT at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution.

Signature:

Production Note:
Signature removed prior to publication.

Lê Hoàng Trung
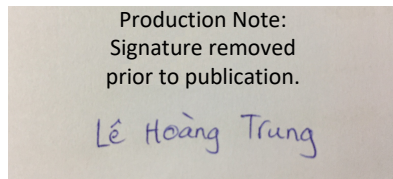
Date: 3rd November, 2024

Abstract

Health risks posed by a polluted atmosphere have significantly escalated in severity. Therefore, the importance of accessibility to reliable air quality monitoring and forecasting has been widely emphasised in recent years. In response, wireless sensor networks are deployed as complementary data sources to regulatory monitoring stations. However, ensuring the reliability and accuracy of air quality sensor data for public dissemination and forecasting models requires robust solutions.

This thesis introduces a fault-tolerant sensing system realised by employing a dependable monitoring configuration for redundancy. Furthermore, the system adopts Dempster-Shafer theory, which effectively handles uncertain and conflicting evidence, for validating sensor readings against reference-grade stations. The application of this theory enables the identification of the most reliable sensor within a network of redundant low-cost sensors, enhancing data reliability prior to training predictive models. The resulting monitoring configuration demonstrates improved robustness and consistency in environmental sensing, particularly under fluctuating air quality conditions. Thereafter, the validated data is streamlined to an ensemble deep learning nowcasting framework, which dynamically selects the best-performing model based on changing environmental conditions. By leveraging the Dempster-Shafer theory for data fusion, the adaptive framework utilises the strengths of multiple models to enhance the robustness of air quality nowcasting.

As for real-world impact, this work demonstrates the operation of a visualisation dashboard, bridging the gap between real-time monitoring and deep learning-based predictions for an urban air quality monitoring network. The proposed system aims to improve early detection and decision-making with a user-friendly interface and comprehensible format, ensuring that communities are better prepared for air quality challenges. All efforts reiterate the primary objectives of this research, which are to enhance public health protection and support sustainable environmental initiatives.

Dedication

Like trees branching out to reach the sun, grounded by strong roots, or rivers flowing to open seas thanks to their tributary streams, who I am and what I have achieved come largely from my family—the place I proudly call home.

Acknowledgement

Hoang Trung Le
3rd November, 2024
Sydney, Australia

Publications and projects

Publications and projects on my own work, resulting in the contributions of this thesis (Chapters 3, 4 and 5)

- **Le, Hoang Trung**; Nguyen, Huynh A.D.; Ha, Quang; Ahmed, Masrur; Barthelemy, Xavier; Azzi, Merched; Duc, Hiep; Jiang, Ningbo; Riley, Matthew; "Dependable Dempster-Shafer Inference Framework for Urban Air Quality Monitoring," *IEEE J. Sensors.* Vol.: 25, No. 14, pp. 27662 -27672, 2025. Doi: 10.1109/JSEN.2025.3577122

- **Trung H. Le**, Huynh A.D. Nguyen, X. Barthelemy, Tien T. Nguyen, Quang P. Ha, N. Jiang, H. Duc, M. Azzi, and M. Riley, "Visualization platform for multi-scale air pollution monitoring and forecast," *2024 IEEE/SICE International Symposium on System Integration.* https://doi.org/10.1109/SII58957.2024.10417539.

- **Trung H. Le**, Huynh A.D. Nguyen, Quang P. Ha, Minh Q. Tran, Masrur Ahmed, Jing Kong, Hiep Duc, Xavier Barthelemy, Ningbo Jiang, Merched Azzi, and Matthew Riley, "Dempster-Shafer ensemble learning framework for air pollution nowcasting," *The 2025 International Conference on Energy, Infrastructure and Environment Research (EIER 2025)*, Hochiminh City Vietnam, 8-10 JAN 2025. https://doi.org/10.1051/e3sconf/202562601003.

Projects

These projects are in collaboration between UTS and the Department of Climate Change, Energy, the Environment and Water of New South Wales:

- PRO22-16023: "Extended deep learning for air pollution forecast enhancement".

- PRO23-18694: "Hierarchical Data Fusion Framework for Air Pollution Nowcasting".

Publications and project in my collaboration not included in the main contributions of this thesis (but mentioned in Chapters 1 and 2)

- Huynh A. D. Nguyen, **Trung H. Le**, Merched Azzi, and Quang P. Ha, "Monitoring and Estimation of Urban Emissions with Low-cost Sensor Networks and Deep Learning," *Ecological Informatics*, Volume 82, September 2024, 102750. https://doi.org/10.1016/j.ecoinf.2024.102750.

- Huynh A.D. Nguyen, **Trung H. Le**, Tien T. Nguyen, Quang P. Ha, Hiep Duc and Merched Azzi, "Particulate Matter Monitoring and Forecast with Integrated Low-cost Sensor Networks and Air-quality Monitoring Stations," *The 2024 International Conference on Energy, Infrastructure and Environment Research (EIER 2024)*, Vungtau Vietnam. 17-18 JAN 2024, pp. 1-6, paper 04001, E3S Web of Conferences 496, (2024). https://doi.org/10.1051/e3sconf/202449604001.

- Huynh A.D. Nguyen, **Trung H. Le**, Quang P. Ha and Merched Azzi, "Deep learning for construction emission monitoring with low-cost sensor network," *2023 Int. Symp. Automation and Robotics in Construction*, pp. 450-457. https://doi.org/10.22260/ISARC2023/0061.

- L.V. Nguyen, **Trung H. Le**, and Q.P. Ha, "Prototypical digital twin of multi-rotor UAV control and trajectory following," *2023 Int. Symp. Automation and Robotics in Construction*, pp. 148-155. https://doi.org/10.22260/ISARC2023/0022.

- Lanh V. Nguyen, I.T. Herrera, **Trung H. Le**, M. D. Phung, R. Aguilera and Q. P. Ha, "Stag hunt game-based approach for cooperative UAVs," *2022 Int. Symp. Automation and Robotics in Construction*, Bogota Columbia, 12-15 JUL 2022, pp. 367-374. https://doi.org/10.22260/ISARC2022/0051.

- I. T. Herrera, L. V. Nguyen, **Trung H. Le**, R. P. Aguilera and Q. Ha, "UAV Target Tracking using Nonlinear Model Predictive Control," *2022 International Conference on Electrical, Computer and Energy Technologies* (ICECET), 2022, pp. 1-7, doi: 10.1109/ICECET55527.2022.9873035.

- Nguyen, Huynh A.D.; **Trung H. Le**; Barthelemy, Xavier; Azzi, Merched; Duc, Hiep; Jiang, Ningbo; Riley, Matthew; Ha, Quang, "Deep-learning based visualization tool for air pollution forecast," *IEEE Software*, in resubmission with minor revision.

Project

This project is in collaboration between UTS and the Department of Climate Change, Energy, the Environment and Water of New South Wales:

- PRO24-19438: "Towards Air Quality Forecast in Adverse Conditions: Data Constraints and Extreme Events."

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivations

Engineering is about solving challenging problems that require knowledge, skilful hands, and ingenuity. The technological solutions advance the world to a time of unprecedented quality of life. However, it can be observed from another perspective that air quality has followed an inverse tendency to societal improvement since the beginning of industrialisation. As a collective, we have disrupted the delicate balance of natural cycles and altered the very composition of the atmosphere. The consequence of our progress is the harmful pollutants released into the atmosphere from human activities. These by-products exacerbate the alarming climate change crisis and are inhaled daily by a staggering 99% of the global population [1]. Urgent calls from the world's most influential health organisation urge governments to embrace intensive actions to mitigate the severity of the current situation, one of which is to monitor air quality exhaustively. With the knowledge of the concentration of substances in the atmosphere, governing bodies can plan proper action and advise the people to take appropriate precautions. This urgent health problem requires significant collaboration and innovative approaches from the global community.

Simultaneously, there is a huge migration of people from war-torn and less-developed areas to bustling metropolises where they can find opportunities to support themselves and their families. The latest estimation states that around 57% of the world's population is now living in cities [2]. Thus, countries are pushing the urbanisation process, reshaping our cities to accommodate this population transition. Inadvertently, the new artificial infrastructure creates distinct environmental dynamics, which are generally referred to as microclimates. These localised phenomena are unpredictable from current forecast systems and can uniquely manifest in each area. Therefore, providing accurate ambient conditions tailored to each community becomes crucial.

Australia, particularly New South Wales (NSW), faces the same global problems. The Department of Climate Change, Energy, the Environment and Water (DCCEEW) of

NSW is at the forefront of tackling the challenges in managing air quality and climate change. They have taken the initiative to integrate atmospheric science with state-of-the-art artificial intelligence (AI) technologies. This approach aims to provide the general public and authorities with insights into environmental conditions across the state, facilitating informative decision-making to enhance the socio-economic and health benefits of the people and mitigate adverse impacts of extreme events. In collaboration with the NSW-DCCEEW scientists, my research endeavours support their missions by delving into the realm of data analysis and visualisation. It encompasses data acquisition from ambient sensing instruments, data assimilation with standard monitoring systems, and an interactive data visualisation platform.

The core objectives of this research emerge from real-world needs identified by DC-CEEW: the utilisation of low-cost wireless sensor networks (LWSN) in street-level monitoring and forecasting of air pollution and the accessibility of environmental data for the citizens of NSW. Notably, there is already a substantial air quality sensor grid distributed across Greater Sydney's urban landscapes, which constantly provides a high-resolution, real-time data stream. The first major milestone seeks to address the reliability issues of that LWSN. Although sensing technologies have made impressive progress, the cost-effective nature of ambient sensors still raises legitimate concerns over data accuracy. The research aims to use the reference-grade readings from state-run air quality monitoring stations (AQMS) as the benchmark to assess the reliability and improve the dependability of low-cost sensors (LCS). Given the positive results from the reliability assessment, invaluable data from the sensors are input to a proposed adaptive ensemble learning model, whose contributing learners are built upon existing deep learning (DL) models fine-tuned for the characteristics of the Sydney Basin area [3, 4, 5]. The ensemble architecture employs a multi-criteria selection process, allowing it to dynamically adjust to volatile inputs and prioritise the most reliable forecasts across predictive models. This adaptive mechanism makes the model well-suited to the diverse conditions encountered by LCSs in urban environments. In the third milestone, the research follows with the development of a demonstrative dashboard hosted on the web environment. This digital canvas is planned to be a user-friendly hub for exploring and accessing important environmental information about the vast state and specific neighbourhoods. The web application is designed to be a gathering point of real-time data from AQMSs and LWSNs. Moreover, it also displays the forecast pollutant concentrations resulting from a diverse source of prediction models. The platform commits to bridging the gap between microclimate and regional scales in environmental monitoring and forecasting. Many stakeholders, ranging from policymakers and researchers to the general public, will gain a holistic insight into convoluted patterns of ambient circulation at different scales, empowering them to make informed decisions and positive changes.

1.2 Contributions

The contributions of the thesis include:

- **Dependable Monitoring of Air Pollution for Low-Cost Sensors**: This thesis proposes a dependable system for air quality monitoring using low-cost sensors. By employing a combination of hardware redundancy and a Dempster-Shafer-based decision-making algorithm, the system ensures reliable and accurate data collection, mitigating the inherent limitations of low-cost sensors in volatile environmental conditions.

- **Ensemble Learning Framework for Air Pollution Nowcasting**: An ensemble learning model is developed, integrating multiple deep learning models to forecast short-term air quality. The Dempster-Shafer-based data fusion method enhances decision-making by selecting the best-performing predictions from the ensemble, offering improved accuracy in predicting air pollution.

- **Visualisation Dashboard for Environmental Data**: A web-based dashboard is designed to provide accessible, real-time visualisations of air quality data. The dashboard serves multiple stakeholders by delivering both monitoring data from low-cost sensors and standard monitoring stations, as well as air quality forecasts resulting from multiple learning-based models, in an intuitive and user-friendly format.

- **Research Prototyping**: This project exemplifies translational research by bridging the gap between academic research on air pollution prediction and real-world applications. The visualisation dashboard converts complex, scientifically validated forecasts into accessible, actionable information. This allows public stakeholders, policymakers, and scientists to make informed decisions about air quality conditions and mitigation strategies, ensuring that scientific insights translate effectively into practical societal benefits.

Chapter 2

Literature review

2.1 Introduction

This chapter reviews state-of-the-art technologies and methodologies that form the foundation of the proposed work. The focus is on the applications of reliable air quality monitoring systems using low-cost ambient sensors. In recent years, the increasing availability of such sensors has presented an opportunity to enhance the spatial and temporal resolution of air quality monitoring systems, thereby providing more granular insights into pollution levels. However, ensuring the reliability and accuracy of measurements from these LCSs is a challenge due to their inherent volatility and susceptibility to environmental factors.

This chapter critically reviews the most recent advancements and applications in the field of low-cost sensor networks for monitoring air pollution and meteorological parameters. A comprehensive overview of the literature will be presented, with particular attention to the effectiveness and challenges of using these sensor networks in diverse environments.

Following this, the chapter will delve into various strategies and methodologies aimed at enhancing the reliability of LWSNs. Since LCSs are prone to measurement drift and inconsistencies, it is essential to explore calibration techniques, data correction algorithms, and other approaches that can improve their accuracy and stability in real-world deployments.

The subsequent section of this chapter will cover the emerging field of nowcasting, which is becoming an increasingly important tool in air quality forecasting. Nowcasting provides short-term predictions of pollution levels using real-time data, enabling quick responses to environmental hazards. This section will review current methodologies and the latest research on nowcasting for air quality, with a focus on integrating it with low-cost sensor networks.

Additionally, the chapter will provide a review of existing air quality monitoring plat-

forms, including those used by government agencies, commercial providers, and scientific research groups. These platforms often incorporate visualisation dashboards that communicate air quality data to the public or specific stakeholders. The review will assess their functionalities, features, and limitations, with a view to identifying potential areas for improvement or innovation.

Finally, while not the primary focus of this thesis, mobile methods of collecting ambient data, such as the use of unmanned aerial vehicles (UAVs), will be briefly discussed. UAVs provide a flexible means of gathering atmospheric data over large or inaccessible areas and can complement static sensor networks. Though this section will be brief, it will highlight the growing interest in such technologies for environmental monitoring.

2.2 Low-cost sensors for air quality monitoring

The Industrial Revolution is fuelling the advances of all engineering and technological fields. Sensor technologies, the means to gather data from the physical world, have witnessed unimaginable progress. Sensing devices have increased in reliability and energy efficiency but decreased in size and price. We can see that they have been extensively deployed all around us, providing a massive source of data for further analysis to better understand the world around us and monitor the status of our invention. That is also true for the application of LCSs in the field of environmental monitoring, where the principal objective is the collection of meteorological data and air quality information, providing essential insights into the convolution of natural dynamics.

Recent reports from the World Meteorological Organization (WMO), the official voice of the United Nations (UN) on the Earth's atmospheric and hydrological states and behaviours, affirmed the promising potential of integrating LCSs in sensing networks as an official branch of environmental monitoring [6]. This effort has led to wide-scale deployments of LWSNs in developing countries, enhancing the accessibility of air quality information to regions and communities lacking the services of conventional monitoring methods. Also, by utilising the information provided by LWSNs, governing bodies and authorities benefit from another layer of insights for effective, data-driven decision-making in urgent circumstances as well as in higher-level policy-making. The consolidation of LWSNs in monitoring is considered by the UN Environment Programme (UNEP) as a way to address two of the UN Sustainable Development Goals (SDG) [7], which are *13 Climate Action* and *15 Life On Land*, illustrated in Fig. 2.1.

LCSs have gained global traction as effective tools for environmental monitoring. These sensors, whether funded through crowd-sourced projects, developed through scientific research, or offered commercially, provide unprecedented insights into urban air quality and pollution exposure at fine spatial and temporal scales.

Research indicates that certain types of LCSs are well-suited for specific applications,

Figure 2.1: United Nations Sustainable Development Goals

underscoring the need for careful selection based on context and sensor characteristics [8]. For instance, studies have shown the potential of air quality sensors to yield high-resolution data that, when properly calibrated and validated, contribute significantly to urban pollution monitoring initiatives [9, 10]. Additional studies have also confirmed the stability and utility of LCSs over long-term deployments, enhancing our understanding of ambient pollution sources and their variability under diverse environmental conditions [11]. However, while LCSs are instrumental in extending air quality networks, they are not intended to replace regulatory AQMS equipped with highly specialised, standardised instruments [12].

Projects that integrate sensor data with professional stations and numerical weather predictions (NWP) have demonstrated improved accuracy in air quality mapping and forecasting. For instance, [13] showed that blending data from both LCSs and regulatory monitors, alongside NWP from the Community Multiscale Air Quality (CMAQ) model, significantly enhances Air Quality Index (AQI) interpolation, enabling more responsive monitoring.

An extensive study reviewed the application of LCSs worldwide, highlighting their efficacy in measuring particulate matter (PM) levels in urban areas and assessing local population exposure to air pollutants [14]. This work substantiated the employment of LWSNs in environmental monitoring, particularly for raising public awareness and fostering community engagement in air quality management efforts.

Several real-world applications illustrate the versatility of LCSs. The KOALA project (Knowing Our Ambient Local Air-quality), led by the Queensland University of Technology, has demonstrated significant impacts in various contexts, such as air quality monitoring during major events and identifying pollution sources in urban centres [15]. KOALA's deployment at the Commonwealth Games showed negligible impact on local air quality and offered a high-resolution dataset for post-event analysis [16]. Furthermore, KOALA sensors have detected patterns in CO emissions in Brisbane's central business district and

validated findings against station data on ship emissions affecting nearby communities [17, 18].

The proliferation of LCSs marks a paradigm shift in environmental observation, complementing the spatial gaps in sparse networks of standard stations to provide real-time, high-resolution insights into urban air quality and meteorological dynamics. This wealth of data offers a promising foundation for integrating learning-based predictive models to acquire more knowledge in spatio-temporal air pollutant circulation, enhancing air pollution estimations for communities. As sensor technologies continue to evolve, LWSNs will play an important role in sustainable environmental management in smart cities and contribute to more resilient, health-focused urban planning that aligns with global climate goals.

2.3 Reliability of low-cost wireless sensor network

The origin of meteorological data observation and collection tools can be traced back thousands of years in early civilisation. Many primitive devices that have been preserved to this day or discovered through excavation are strong evidence of a substantial history of meteorology. In today's increasing need for ambient knowledge, highly precise technologies have been applied in this domain, which provide valuable inputs for our current monitoring and forecasting networks. In NSW, DCCEEW has employed various techniques and sophisticated instruments, namely ultraviolet spectroscopy for ozone, infrared spectrometry for carbon monoxide, Tapered Element Oscillating Microbalance (TEOM) and Beta Attenuation Monitor (BAM) for PM measurements [19]. Figure 2.2 shows the setting of a DCCEEW-operated AQMS with the mission to observe key air pollutants listed in Table A.1. They are scarcely but strategically located across the vast state to capture the most representative ambient pattern of specific regions. Having a method to effectively obtain fine-grained environmental data can dramatically improve the ability to deliver more accurate monitoring and forecasting information to individual communities as well as the entire state. Thus, low-cost air quality sensors come into the picture, offering significant advantages in deployment, scalability and affordability. The enhancement and proliferation of these technologies are driven by the vital demand to acquire a large amount of information to supplement the sparse government-operated monitoring stations and fuel the data-driven forecasting techniques [10].

The advantages of LCSs in environmental monitoring covered in the previous subsection may lead to biased promotion and exaggerated advertisements. Therefore, researchers have raised questions about sensors' reliability in terms of measurement accuracy. Numerous studies have attempted to answer the concerns surrounding the performance and accuracy of air quality LCSs.

PurpleAir sensors [20], our primary objects of interest, have been established in sub-

Figure 2.2: An air quality monitoring station operated by NSW-DCCEEW

stantial monitoring networks around the world, with thousands of nodes across all five continents. They provide crucial environmental insights for governments, meteorology agencies and the general public with extensive data streams. In general, these networks of PurpleAir sensors (PAS) are great platforms for researchers to study and implement innovative ideas for ambient monitoring. Different studies were conducted in populated cities of Greece, as well as California, the United States (US), to evaluate the reliability of the PAS through extensive statistical metrics [21, 22]. The final outcomes strongly suggested that compared to the referenced high-precision instruments, such as BAM and TEOM, this sensor network shows credible stability in the tested temporal periods and ambient volatility. However, LCSs raise challenges related to the calibration and reliability of the collected data. In [23, 24], approaches to derive correction strategies have been proposed to reduce the measurement bias of PASs. Using high-precision devices as references, their methods provide correction factors for dust monitoring that improve the quality of collected data over a large scale, laying the important groundwork for integrating this granular data into the regulatory monitoring system. Inheriting the idea, a framework involving spatial calibration and down-weighting modelling of PurpleAir data was developed in the US. The outcomes indicated a promising method for mitigating noises and reducing bias from the LWSN, creating opportunities for incorporating this rich source of environmental data with the state-run monitoring system and satellite information to predict episodic events and dust hotspots [25].

In an effort to enhance reliable observations, researchers have proposed an extended fractional-order Kalman filter to improve the continuous flow of data from developed wireless sensors used for monitoring air pollution in Australia [26]. For long-term sensor operation, a dependable scheme employing a continuous Markov chain model was then proposed, demonstrating its ability to observe air particles in 72-hour horizons, even under extreme conditions [27]. In Nanjing, China, a dense deployment of wireless sensors

was verified for reliability through statistical analysis and correlation assessments [28]. To detect, identify and remove outliers, the Volterra graph-based method was used in [29] for maintaining the data quality of the low-cost sensor network data in air pollution estimation. These studies have demonstrated robust and feasible methods for sensing networks under a specific configuration of the monitoring system. However, they lack a comprehensive framework that can be applied to various monitoring systems with the capacity for fault tolerance and reliable fusion of data.

2.4 Air pollution monitoring platform

Humans are instinctively perceptive beings, absorbing knowledge more effectively through the senses. Our visual system, one of the most intricate feats of evolution in our neural network, plays a dominant role in our daily activities. We often rely on visual aids to interpret complex results in mathematics and phenomena in the world around us. In environmental science, this necessity for intuitive comprehension has led to the development of platforms that translate atmospheric data into accessible formats. In this context, visualising meteorological observations from multiple sources—alongside forecasting results from various models—is not only beneficial for communication but is also considered essential by guidelines from the WMO [30].

To date, a diverse list of applications has been developed for the visualisation and analysis of environmental data from different sources. The Giovanni dashboard, operated by the National Aeronautics and Space Administration (NASA), collects global ambient data monitored from satellites, then produces exhaustive plots and utilises the feature-rich support from Google Earth to portray data in a 3D context [31]. Employing service from the same geographic information system (GIS) provider, real-time air quality data from governmental monitoring centres has been spatially overlaid on a virtual Beijing city in a web interface [32]. With a more systematic initiative to explore spatio-temporal ambient data for the end users, a dashboard was developed in [33], incorporating intuitive navigation and understanding of the massive air quality data at a nationwide scale. In Melbourne, Australia, the residents are experienced with a user-centric dashboard, namely Context-Aware Visualization of Outdoor Air Pollution with Internet of Things (IoT) Platforms (CAVisAP), in which the functionalities are customised considering the context of individual preferences [34]. Taking a further step, AiR is an augmented reality-based solution hosted on smart mobile devices to promote wide accessibility and user experience in urban areas [35]. These platforms realistically depict monitoring data, mainly at a global scale and haven't garnered significant attention from the wider public yet.

With the recent advancement in computer science, driven by the accelerating computing power, outcomes from weather forecasting methods are significantly improved in reliability and accuracy. Two predominant methods of forecasting, the traditional

NWP model and the learning-based technique, are substantially leveraged by technological progress. Dashboards operated by environmental agencies around the globe are also relaying forecasts from different numerical models to citizens [36, 37]. The model-based methods known as numerical models with complex physical-chemical equations are developed and applied in many countries, such as the CMAQ governed by the US Environment Protection Agency (EPA) [38] or the hybrid model of the Conformal Cubic Atmospheric Model (CCAM) and Chemical Transport Model (CTM) currently used by the Australian government [39]. However, these models rely critically on multiple assumptions of atmospheric states, classical emission inventory, and uncertainties in physical-chemical modelling. These complex modelling inputs of physical-chemical models produce large-size outputs with inadequate accuracy. Thus, this limits the capacity of visualisation to the public.

At a global scale, the World Health Organization (WHO) has put into operation their developed visualisation tool for major air pollutants [40], which ambitiously aims to unify all national air quality standards under the recommended ones scientifically issued by the Organization. This web-based interactive dashboard is the product of the study [41], which stated that the main sources of data rely heavily on the data availability of member countries. It is acknowledged that the project somewhat achieved its objectives, fuelled by great intention to scientifically address the severity of air pollution-related health risks. However, the visualisation platform remains simple due to the lack of data and provides low-resolution air quality monitoring for reference only.

A comprehensive monitoring dashboard for the entire world is an arduous yet encouraging attempt to help us understand air quality better. It is certain that, at this moment, much effort is spent on developing improved platforms. The foundation of that hope is based on the inspiring work done at the continental scale. The Copernicus Atmosphere Monitoring Service (CAMS) within the Earth observation programme of Europe is an exemplary and pioneering project in monitoring and forecasting numerous parameters and conditions in different environments [42]. The objectives of this project are to ensure the provision of quality-controlled, reliable atmospheric, land and marine observations from satellites and in-situ measuring instruments on Earth along with forecasts to a wide audience and help in understanding and combating climate change [43, 44]. Notably, outcomes of the project, which is an operational dashboard [45], have served hundreds of millions of users worldwide [46]. The dashboard employs a simplistic yet intuitive interface, engaging the majority of users in the exploration for monitoring and forecasting data through the convenience of a slide show of time-series contour maps of Europe. It also supports a seamless transfer of data for other communities, such as scientists, developers and interested individuals. Alongside, the European Air Quality Index Dashboard, a project of the European Environment Agency, is specifically dedicated to conveying air quality monitoring information from routine monitoring stations from all

territories of EU member states [47]. The dashboard allows users to view real-time as well as the previous 48-hour air quality data. Clicking on a marker, users can explore details of the station, including the percentage of air quality-categorised day in a year and the concentration of main contributing factors to the air quality, which are the measured values of $PM_{2.5}$, PM_{10}, NO_2, O_3 and SO_2. Moreover, the time slider of the dashboard enables users to explore air quality forecasts in the next 24 hours, providing additional insights for the public or authorities to be proactive in pre-planning schedules, activities and events. The simple graphical user interface (GUI) and interactive functions of the dashboard, along with its mobile app version, signify the aim of delivering a more specific type of environmental information to a broader range of audiences, setting it apart from a more scientific-oriented CAMS implemented for Europe.

Across the Atlantic Ocean, several governmental agencies of the US are pioneering in the missions of monitoring Earth's climate, meteorology, hydrology and other natural conditions. NASA is one of the global-leading experts in this field. Their development of the Worldview dashboard [48], a web-based application for visualising rich and diverse resources of the agency in remote sensing, is based on NASA's dense Earth-observing satellite network. The dashboard promotes interactive exploration of high-definition, diverse satellite imagery layers captured at a global scale and updated at hourly frequency. This near-real-time access supports time-critical actions concerning natural hazards and calamities, including storm tracking, wildfire management, air quality monitoring, and flood surveillance. Understanding the spatio-temporal characteristics of geoscientific data, the dashboard introduces functionalities such as animation and dual comparison to users, allowing them to leverage visual descriptions of the changes in Earth's natural conditions through the lens of orbital satellites. The tool democratises access to NASA's Earth observations, facilitating critical analysis and important discoveries in science through the sharing of data, informing timely insights for decision-making in natural hazards, encouraging the next generations of scientists and activities, and providing scientifically accurate information to the general public and media as an outreach attempt.

The problems of climate change and air pollution are globally indiscriminate in whom they affect. Everyone on the planet is considered a stakeholder. Therefore, the private sector is taking numerous proactive actions, joining with the tireless efforts of governmental agencies in combating climate change and raising environmental awareness among the public. This spirit of unity is best reflected in the visualisation dashboard of IQAir [49], a Swiss company specialising in air quality monitoring and physical technologies such as environmental sensors and air cleaning products. IQAir dashboard is considered one of the most comprehensive platforms for global environmental monitoring, featuring the display of over 80,000 sources from monitoring stations of governmental agencies to crowdsourced LCSs alongside the integration of satellite imagery [50]. Therefore, this integrated platform creates a multiscale, high-resolution map of environmental conditions.

Users have access to publicly available, real-time weather and air quality information with eye-catching graphics and engaging functionalities, enabling them to exhaustively explore meteorological conditions and air quality status in their areas of interest. With a similar objective in environmental monitoring, PurpleAir develops their own dashboard [20] displaying a dense monitoring network of PASs worldwide. The dashboard is dedicated to exhaustive means of visualising all parameters within the measuring capability of PASs. These commercial platforms have been shown to leverage the environmental awareness of a wide range of audiences by capturing their attention and delivering easily digestible information that meets common needs without diving too deep into other layers of environmental monitoring. On the other hand, the absence of the forecasting component in these dashboards can be explained by the fact that these companies merely displayed observed values from their sensors and third-party sources.

Recently, the big data observed from AQMSs or sensor networks leverages data-driven methods such as machine learning (ML) or DL models for improving the accuracy of air pollution forecasts, which are visualised together with real-time monitoring information on multiple platforms. For example, a web interface with the integration of IoT-enabled data and a DL model is developed in North Macedonia for urban monitoring and predictions [51]. Similar studies have been reported in Korea [52] and Taiwan [53] exploiting big data approaches in air quality visualisation.

Nowadays, smartphones have become the one inseparable electronic device of almost everyone, especially since the percentage is extremely high among people living in cities around the world. Air quality monitoring and forecasting platforms have been specifically designed for mobile phones as a way to proliferate environmental awareness and improve the accessibility of vital air quality information to as many people as possible. Acknowledging this trend, the studies [54] resulted in a mobile application called IoT-Mobair. The research employed a LWSN that measured gases such as CO, CO_2, and CH_4, which are major factors contributing to air pollution. The collected data from the sensors were sent to a cloud server where the proposed algorithm from the research group performed calculations to derive the AQI at the locations of respective deployed sensors. The processed data was then sent to the mobile application for display. However, the focal point of the research did not lie in the forecast of AQI level, which is equally important to the monitoring of air quality.

The public greatly benefits from perceptive tools such as those covered above in gaining necessary air quality information for making informative decisions in their daily activities and short-term plans. However, we must not forget that they are not the only beneficiaries of such helpful systems. The people working in the field of weather forecasting, where their titles can be climatologist, meteorologist, atmospheric scientist, data scientist, GIS analyst and so on, can considerably enhance their working performance with visualisation interfaces for the extensive environmental datasets.

The current landscape of air quality visualisation tools reported in the literature either concentrates on regulatory monitoring stations operated by governmental agencies or low-cost air quality sensor networks. The majority of established platforms targets the delivery of pollution and meteorological event forecasts resulting from complex numerical models. Others are transitioning to the latest advances in environmental forecasting, namely big data and DL-based techniques. A noticeable gap in comprehensive tools integrating information from state-run monitoring stations and decentralised networks of ambient sensors, as well as intelligent forecasting models, has been observed in this domain. Our work is then directed to address this absence of a unified platform capable of merging multiple sources of environmental observations and incorporating predictive frameworks. Various stakeholders can gain a holistic understanding of air quality situations from this public visualisation platform, which is beneficial for informed decision-making and the biophysical and socioeconomic situations of the interested region.

2.5 Nowcasting

Nowcasting is an integral aspect of meteorological forecasting. The part "now" in the word does not take the literal meaning of the exact moment of time in the presence. Rather, in the vast field of forecasting, the concept of nowcasting is extended a bit further in the future compared to its common meaning in everyday conversation. Specifically, the definition of nowcasting given by WMO is essentially "forecasting with local detail, by any method, over a period from the present to 6 hours ahead, including a detailed description of the present weather" [30]. From this point onward, every mention of nowcasting in this document should explicitly follow the standardised interpretation of WMO.

While other types of forecasting attempt to resolve the meteorological status for the next 24, 48, 72 hours or even longer time range, the aim of nowcasting is to unveil the environmental conditions from the current up to 6 hours in the future. The classification of nowcasting into a unique category emphasises the significance and specific role that it serves in the big picture of forecasting. With the knowledge of 6 hours in advance, nowcasting addresses the need for timely, location-specific predictions of high-impact weather that is of utmost importance to a diverse range of activities. It can be used for any human activities such as public outdoor activities, entertainment and sporting events, construction, transportation and the list goes on. For each of those activities, the nowcasting results can be crucial decision-making information for the authorised bodies. They can issue appropriate cautions and warnings directly to the public or relevant third parties in accordance with the short-term forecasts.

According to WMO, nowcasting is applied to weather that develops on the mesoscale and local scales and over very short time periods. The common meteorological phenomena predicted by nowcasting methods can be listed as thunderstorms, tornadoes, hail, heavy

precipitation that causes flash floods and types of winter precipitation, severe wind and visibility/fog. It is due to the small-scale nature of these events that prevents explicit prediction of their occurrence several hours ahead of time. That is stated for the contextual understanding of nowcasting in the field of meteorology. However, we adopt the concept of nowcasting in this work by keeping true to the original definition and broadening its applications. We attempt to apply nowcasting techniques to unveil the short-term status of certain air pollutants in a localised area.

In a global sense, the adverse effect of climate change is being felt in every corner of the Earth. The frequency of natural calamities and smaller in size hazards has significantly increased. The UN has announced that they will act as a trailblazer in actions that aim to safeguard humankind from catastrophic climate events through the Early Warnings for All initiative. This involves an attempt to deliver a global Multi-Hazard Early Warning System (MHEWS), especially aiming to address a lack of such a system in developing countries and the vulnerability of their citizens. A comprehensive four-pillar structure has been constructed to ensure the delivery of the Early Warning for All initiative, following the UN SDG of *11 Sustainable Cities and Communities* and *13 Climate Action* as shown in Fig. 2.1. It is acknowledged that the current revolution in computing and AI have complemented each other to propel the advancement in forecasting in general. Where substantial NWP models benefit from the computing power and the theories of complex learning-based models can be applied. Alongside forecasting, nowcasting is within one of the major milestones, "Enhancing forecasting capabilities", in Pillar 2: Detection, observation, monitoring, analysis, and forecasting of the Initiative [55].

2.6 Air pollution monitoring with UAV

The paradigm shifts from conventional research ways to innovative approaches have been iterated and discussed throughout this literature review section. An exponential pace of technological advancement is among the significant reasons that propel this positive trend in academia and professional practices. From the conversation about LWSN complementing the spatial scarcity of AQMSs, researchers transform stationary monitoring nodes into high-mobility devices. Specifically, various studies have proven multi-rotor copters as a capable host for air pollution sensing equipment.

The main challenge that UAVs offer viable solutions is the examination of air pollution distribution and dispersion in various settings. A meticulous analysis of gaseous concentration at high and low altitudes in different metropolitan areas and surrounding suburbs of Romania was presented in [56]. Distinctive patterns were unveiled and assessed to generate heatmaps of the pollutants, which were only possible with the flexibility of their proposed hexacopter platform. In addition, traffic, being an integral infrastructure, is often the leading contributor to pollution emissions in urban areas. The excessive ex-

haust fumes from the high volume of vehicles create a unique environmental zone along the transportation grid. To study the distribution of various types of PMs and other meteorological parameters, [57] employed a customised UAV to collect the data in vertical and horizontal profiles. These works and many others further establish the capabilities of UAVs to be effective platforms for environmental and meteorological research.

The versatility of drones allows the integration of a diverse range of peripheral sensing technologies onto the platforms. Recent advancements in pollution detection have pushed forward the implementation of vision-based methodologies to estimate the AQI. Using innovative approaches in computer vision and learning-based frameworks, researchers are able to obtain reliable ambient parameters from the images captured by optical instruments onboard UAVs. For instance, the study in [58] acquired panoramic image-based data from an aerial platform for haze estimation in a wide urban area. An AQI inference model proposed by the researchers with optimisation strategies in UAV altitudinal and positional placement during the data acquisition stage claimed high recognition accuracy of haze while requiring only a small dataset for pretraining. This framework, named AQ360, confirms the promising capabilities of drones in environmental monitoring missions with the collaboration of technologies such as computer vision and machine learning. Utilising the high maneuverability of UAVs, a group of researchers devised effective algorithms to deploy the monitoring drone in urban settings and map the fine-grained AQI there in real time [59]. The paper discussed the potential outcomes of the hybrid AQI prediction framework, which combines a physical particle dispersion model used in meteorology and a neural network scheme. This system offered a higher accuracy than existing models at that time, emphasising its ability to produce meter-level granularity in AQI mapping within a short time period. On top of the proposed nowcasting method, the researchers introduced an adaptive algorithm for air pollution monitoring based on the idea of selective monitoring for energy and trajectory optimisation. Simulated 2-D and 3-D scenarios of parks and build-up zones are put to the test, and all results indicate the robustness and precision of the overall system. However, special remarks were made that their system required fine-tuning for different circumstances, demanding tradeoffs to be compromised between the accuracy of AQI forecasts and the power consumption of the aerial vehicle. In an attempt to extend their previous efforts, the research group presented a more solid architecture for the aerial AQI monitoring system, specifically aimed at being an efficient IoT solution for the current transition toward smart cities [60]. Moving on from simulated environments, the upgraded system was practically experimented within the boundary of university campuses, a miniature version of the proposed smart city, reaffirming the system's plausibility in the real world. Moreover, an additional presentation layer was introduced in the overall architecture, which provides a GUI containing necessary AQI information to the general public and smart city managers.

In contrast to pre-planned flight approaches that are reviewed above, recent literature

in UAV-based air quality monitoring explores dynamic navigation and active guidance methods, propelling environmental surveillance to new heights. A notable example is the employment of a deep reinforcement learning technique, the Deep Q-Network, which enables UAVs to autonomously track signatures of pollution hotspots in polluted zones [61]. A multitude of test cases was carried out, each with a different number of extreme hotspots, to validate the adaptability of the proposed frameworks to environmental volatility. With the same goal to rapidly identify the most polluted area in an interested region, [62] advanced the idea of merging the bio-inspired chemotaxis movement with the adaptive spiral mobility pattern derived from Particle Swarm Optimisation (PSO). The final outcomes of both studies and many others going along similar paths yielded swift tracking and coverage time as well as high precision in the identification of polluted locations. These approaches leverage the power of artificial intelligence in guidance and extreme event recognition, enhancing the efficiency and responsiveness of UAVs in capturing air quality data. Furthermore, a growing number of researchers are delving into swarm robotics applied for UAVs in environmental monitoring. One example of this adoption was presented in [63], where an architecture for a federated learning scheme was designed for AQI forecasting. Claiming to be a pioneer in this field, the overall system aggregates data collected from both the ground-based air quality sensors and images taken from the airborne drone squadron. Then, implementing a distributed computing process in IoT, a hybrid LSTM model was utilised for the spatio-temporal prediction of AQI. Through these transformative studies, environmental monitoring and forecasting can benefit from state-of-the-art technologies, ranging from IoT, cooperative UAV navigation and various forms of artificial intelligence. The ambient surveillance system of this innovative era does not have to rely solely on scarcely positioned AQMS of governmental agencies since UAVs are demonstrating their promising competence by solving coverage, accuracy and real-time problems in monitoring-related missions. These active methodologies showcase the evolving landscape of UAV applications, highlighting the integration of cutting-edge technologies for more responsive and adaptive environmental surveillance.

Digital twin, an innovative concept strengthened by the Industrial Revolution 4.0, is gradually securing a prominent position in UAV research. Virtual replicas of drones, capable of simulating the dynamics and performance of their physical counterparts, accelerate the rigorous testing phase and deployment to real-world missions [64]. It is made possible by also generating simulated surroundings that mimic the actual external working conditions of the aerial vehicles, such as temperature, wind and obstacles. In the context of environmental monitoring, UAV digital twin holds significant importance due to enhancements in testing abilities and analysis in any virtual scenarios. Researchers can pre-plan the environmental monitoring missions in computerised areas of interest, swiftly validating different strategies in terms of coordination, safety and energy efficiency. Moreover, complex phenomena such as air pollution dispersion patterns and weather variations can

be embedded in the virtual world. This enormous advantage effortlessly aids in devising appropriate testing procedures for extreme and episodic events and predicting expected outcomes in reality. Therefore, digital twin facilitates effective field tests and safe UAV operation. In previous studies [65, 66, 67], actual environments with a defined number of meteorological parameters are virtualised in digital twin platforms to prove the plausibility and evaluate the performance and benefits of this technology. Surveying the literature, there remains a lack of proposed digital twins specialised for environmental monitoring missions, due in part to the computational demands of simulating ambient characteristics and the challenges associated with ensuring sensor accuracy in virtual environments. While this thesis does not focus on these challenges as primary research objectives, the identified gaps and the potential of UAV digital twins are recognised as promising directions for future investigation—particularly for extending reliable monitoring frameworks to mobile platforms such as drones.

2.7 Conclusion

This chapter has reviewed recent advancements in low-cost sensors for air quality monitoring, which attempt to provide fine-grained, real-time pollution data to the public and implement microclimatic management strategies within smart cities. Given the inherent variability and uncertainty in the measurements of low-cost sensors, various approaches for enhancing their performance and reliability have been put under survey.

The chapter has also explored current methodologies for nowcasting in general, an essential tool for predicting short-term meteorological parameters, pollution levels, and extreme events. Going through the literature, promising areas for improving the overall predictive accuracy in the presence of data uncertainty will be applied in the proposed framework for air pollution nowcasting using data from professional monitoring stations with volatile measurements from low-cost sensors.

Moreover, a review of existing platforms for visualising environmental data highlighted key features and limitations of commercial, governmental, and research-oriented platforms. Identifying the strengths and shortcomings of those applications serves as a foundation to develop the proposed platform for the local EPA, which seeks to provide a comprehensive and user-friendly tool for presenting air quality monitoring and forecasting data to a diverse group of stakeholders.

Lastly, the chapter briefly discusses mobile data collection methods using UAVs, presenting solutions to reach areas that are difficult to monitor using stationary sensors. This section paves the way to implement the proposed reliable monitoring in onboard instruments of UAVs in future research.

In summary, this chapter has provided a thorough examination of the relevant literature in low-cost sensor networks, reliability enhancement techniques, nowcasting method-

ologies, and data visualisation platforms. These reviews offer an up-to-date perspective of current innovations in air quality monitoring to better inform the development process of the scientific proposals in this thesis.

Chapter 3

Dependable monitoring for air quality

3.1 Introduction

The severity of health risks associated with air pollution exposure remains an increasing concern with a high rate of hospitalisation [68]. Recognising the problem, governments and citizens have emphasised the importance of reliable access to air quality monitoring and forecasting. To this end, LWSNs are being extensively deployed as a complementary data source to state-run AQMSs. To achieve reliability in using LWSN for air quality monitoring, it is essential to ensure information availability and estimation accuracy in the face of adverse conditions in ambient environments [27]. Along with urbanisation and the emergence of smart cities, this issue has attracted a number of studies aiming to enhance the performance of sensor networks as well as data quality using IoT-enabled sensors in the environmental monitoring domain.

The concept of dependability in engineering is critical for ensuring the consistent and trustworthy operation of systems, especially in applications like ambient monitoring, where data quality directly impacts decision-making. Dependability is often defined by six key characteristics: availability, reliability, resilience, maintainability, integration, and safety. Availability is the readiness of a system to provide service when required, while reliability ensures that the system performs its intended functions accurately over time. Resilience pertains to a system's capacity to recover quickly from faults or unexpected conditions, and maintainability emphasizes ease of system repair and upkeep. Integration ensures that various components operate cohesively, allowing efficient data fusion and processing from multiple sources. Finally, safety guarantees that system operations do not pose risks to users or the environment. In ambient monitoring, these attributes are essential for LWSNs deployed in dynamic urban environments, where equipment may face challenges from both internal imperfections and external volatility. By designing with

dependability in mind, ambient monitoring systems can improve data accuracy, continuity, and overall reliability, supporting the effective use of LWSNs alongside regulatory AQMSs.

Owing to the cost-effectiveness advantage, monitoring systems using LWSN benefit from dense deployments with colocated settings to enhance their dependability, including system availability and reliability against uncertainties, device imperfectnesses and environmental volatility [27]. As such, numerous techniques have been developed to enhance the integrity of data collected from LWSN with other sources, such as learning-based models using neural networks, Kalman filtering, Bayesian inference and Markov chain [69, 70, 71, 72]. For data fusion and reliability enhancement, the Dempster-Shafer evidence theory (DSET), a generalisation of the Bayesian method, appears promising for its adaptability and ability to handle conflicting evidence hidden within substantial amounts of data while requiring no presumed probabilities [73]. DSET has been proven to be a robust method in monitoring systems such as the localisation of sensors [74] and fault detection [75].

This chapter proposes a novel dependable Dempster-Shafer inference (DDSI) framework designed for colocated low-cost wireless sensors to monitor air quality in cities. The approach is built on DSET mathematical background to provide an effective mechanism for enhancing the dependability of the monitoring system. Additionally, this framework incorporates common faulty features from LWSN to infer basic probabilities which, upon ranking, maintain the continuity of air quality data based on the most reliable sensing information at the monitoring location. Extensive experiments are conducted in various settings to enrich the knowledge of data uncertainty and potential faults of the colocated LWSN in practical operations. The proposed framework will then be implemented on real-world monitoring systems across suburbs in Sydney, Australia.

The work presented in the paper *Dempster-Shafer ensemble learning framework for air pollution nowcasting* is elaborated in this chapter as follows. Section 3.2 presents the development of the proposed DDSI framework following a brief background on the evidence theory. Section 3.3 describes the design of the colocated LWSN incorporating the proposed DDSI algorithm for evaluation and verification of its system performance in monitoring air quality under laboratory conditions. Section 3.4 presents the real-world implementation of the DDSI LWSN monitoring system along with its on-field results and discussion. Section 3.5 concludes the chapter.

3.2 Dependable monitoring system using Dempster-Shafer evidence theory

In environmental monitoring, large-scale deployment of low-cost air quality sensors becomes promising for gathering high-resolution data [28]. However, these sensors are sus-

ceptible to environmental volatility, imperfectnesses and events that may affect their functions, potentially leading to false readings. As a result, the likelihood of issues increases with the number of devices deployed. Identifying these problems is essential to avoid inaccurate observations, erroneous data for further analysis, and hence, inappropriate decision-making.

This section presents our reliable sensing framework using a dependable configuration to maintain the highest quality of the sensing information from the colocated sensors available as a result of Dempster-Shafer (DS) inference.

3.2.1 Frame of discernment

Our development is based on DSET, which is centred on the frame of discernment (FoD), a finite set encompassing all hypothesised states of the monitoring system. Here, the FoD is used for making decisions [76] for the dependable configuration of all accountable states (hypotheses) of the air quality sensors. We consider four types of hypotheses, including (i) normal operation, (ii) physical faults of sensors, (iii) data anomalies, and (iv) other undefined states. These conditions are selected based on long-term operation in our real-world LWSN implementation under volatile environments [27, 26]. The FoD can then be represented by vector Φ with elements $\phi_k, k = 0, 1, \ldots, K$, where K is the number of predefined faults or imperfect states, and ϕ_k is the probability associated with them. In particular, ϕ_0 denotes the sensor's reliable operation, and ϕ_K is the probability for all uncertain or abnormal conditions influencing the standby sensor.

For computational purposes, the hypothesis FoD Φ is projected to a reference matrix $R = [r_{kj}], j = 1, 2, \ldots M$, where M is the number of interested ambient parameters (e.g., temperature, relative humidity, fine particles) collected by the LWSN.

3.2.2 Basic probability assignment

The inference using DSET is quantified with the basic probability assignment (BPA) or mass function. This function assigns a degree of belief to each subset in the space containing all subsets of Φ, or the power set 2^Φ expressed as,

$$2^\Phi = \{\varnothing, \{\phi_0\}, \{\phi_1\}, \ldots, \{\phi_K\}, \{\phi_0, \phi_1\}, \ldots, \Phi\}. \tag{3.1}$$

The mass value or probability associated with a certain hypothesis in the defined FoD should be in the range between 0 and 1. Therefore, mathematically, the mass function, denoted as m, is portrayed as [77],

$$m : 2^\Phi \rightarrow [0, 1],$$

where it satisfies

$$\begin{cases} m\left(\emptyset\right) = 0, \\ \sum_{\phi \subseteq \Phi} m\left(\phi\right) = 1. \end{cases} \tag{3.2}$$

For LWSN reliability assessment, the mass function numerically interprets the level of confidence in the examined hypotheses as well as operational imperfectnesses of sensors based on the collected data. For their sampling, the sampling matrix $S = [s_{ij}, i = 1, \ldots, L]$ is defined, where L is the number of timestamped readings from the sensors considered for the reliability assessment. The following steps describe the procedure to acquire mass values for all sensors.

To derive BPAs complying with the conditions (3.2) from the dataset obtained from the LWSN, we propose to assess the dissimilarity between sensor data and a reference source from a nearby AQMS considered as their benchmark. This evaluation ensures the alignment of data from reliable sensors and serves as a means of generating BPAs in DSET. Firstly, we employ an absolute error between measured and reference values as a distance metric to quantify the evidence differences for a predefined fault. The expression for this distance is:

$$d_{ijk} = |s_{ij} - r_{kj}|. \tag{3.3}$$

The probability associated with a hypothesised state k can be derived as,

$$q_{ijk} = \frac{d_{ijk}^{-1}}{\sum_{k=0}^{K-1} d_{ijk}^{-1}}, \tag{3.4}$$

which forms the probability matrices $P_k = [q_{ijk}]$.

For data fusion of available sources, the Shannon entropy

$$\varepsilon_{ij} = -\sum_{k=0}^{K-1} q_{ijk} \ln\left(q_{ijk}\right), \tag{3.5}$$

is well-known for obtaining the amount of information in light of the information theory [78]. This is particularly promising for enhancing the dependability of the monitoring system since the colocated air quality sensors considered here are a multi-sensor source available for data fusion with data from AQMS, numerical models and also meteorological parameters by using the evidence measures [79]. The technique is applied here to support the detection of possible sensor abnormalities based on the data collected from LWSN to maintain the flow of reliable monitoring information under the proposed DDSI scheme.

In the next step, the Shannon entropy is integrated as the discount factor for the calibration of the probability distribution to take into account the inherent volatility of low-cost air quality sensors (3.4) [80]. The discount matrix is defined as $\Psi = [\psi_{ij}]$, where its elements are obtained after a normalisation:

$$\psi_{ij} = 1 - \frac{\varepsilon_{ij}}{\max \varepsilon_{ij}}. \tag{3.6}$$

Then, with the discount factor, the BPA is adjusted by

$$m_{ijk} = \psi_{ij} \times q_{ijk}. \tag{3.7}$$

Finally, the probability assignment for all uncertainties m_{ijK} is obtained as,

$$m_{ijK} = 1 - \sum_{k=0}^{K-1} m_{ijk}. \tag{3.8}$$

3.2.3 Dempster's rule of combination

Dempster's rule of combination in DSET is pivotal in extracting insights from data gathered from various sources. The combination rule is hence employed in fusing data obtained from sensors in LWSN, or other estimation and monitoring sources. In the proposed DDSI framework, we use the combination rule to derive a confidence level with uncertainty probability for each hypothesis in the FoD.

Dempster's rule of combination is used to combine in the same field of discernment Φ the evidence that supports the combination of independent sources of data, e.g., from LWSN or AQMS. For two BPAs with mass functions of m_1 and m_2 associated with subsets B and C of Φ, this combination uses the orthogonal sum to aggregate them, resulting in the joint m_{12} as calculated by

$$m_{12}(B \cap C) = m_1(B) \oplus m_2(C) = \frac{\sum m_1(B) \, m_2(C)}{1 - K_f}, \tag{3.9}$$

if $B \cap C \neq \emptyset$, and

$$m_{12}(\emptyset) = 0,$$

where

$$K_f = \sum_{B \cap C = \emptyset} m_1(B) \, m_2(C). \tag{3.10}$$

Here, the coefficient K_f represents the degree of conflict between pair-wise BPAs. This conflict factor is calculated by summing the products of two BPAs for all sets where their intersection is empty [81].

3.2.4 DDSI framework for LWSN Reliable Monitoring

The development described above is now encapsulated in Algorithm 1, which represents a
step-by-step procedure for implementing our dependable DSET-based monitoring system
using colocated low-cost sensors.

Algorithm 1 Dependable Dempster-Shafer Inference

1: **function** QUANTIFY_RELIABILITY($sampling\_matrix$, $reference\_matrix$)
2:     Compute distances (Eq. 3.3)
3:     Compute probabilities (Eq. 3.4)
4:     Compute discounted BPA (Eq. 3.6)
5:     **function** DISCOUNTED_MASS($discount\_matrix$, $probability\_matrix$)
6:         Replicate $discount\_matrix$ to match the dimension of $probability\_matrix$
7:         Compute the discounted mass (Eq. 3.7)
8:         Consider all hypothesised states of sensors to compute the final $discounted\_mass$
        (Eq. 3.8)
9:     **end function**
10:     **function** COMBINED_MASS($discounted\_mass$)
11:       **for** $j = 1 : K$ **do**
12:         Compute conflict coefficient K_f (Eq. 3.10)
13:         Compute intermediate mass $inter\_mass[j]$ from pair-wise faults (Eq. 3.9)
14:       **end for**
15:       $combined\_mass \leftarrow inter\_mass[K]$
16:     **end function**
17:     **function** SAMPLING_COMBINED_MASS($discounted\_mass$)
18:       **for** $i = 1 : L$ **do**
19:         $combined\_mass[i] \leftarrow$ COMBINED_MASS ($discounted\_mass[i]$)
20:       **end for**
21:       $lwsn\_states\_prob \leftarrow$ COMBINED_MASS ($combined\_mass$)
22:     **end function**
23:     **return** $lwsn\_states\_prob$
24: **end function**

The Dempster-Shafer inference here plays an integral role in deducing the probabilities
of sensors' states, known as combined mass functions, by utilising data from the colocated
sensors. These outputs are then passed into the decision-making layer of the proposed
DDSI framework for the colocated LWSN. The overall DDSI-LWSN system for reliable
monitoring of air quality is shown in the diagram of Fig. 3.1. Therein, the probabilities
of health states of all sensors are ranked to select the most reliable sensor in the network.
Following the probability ranking, this sensor can take the role of the duty sensor to
provide air quality data to maintain monitoring performance. The probability ranking and
data switching take place in the decision-making layer in the pipeline and are described
in Algorithm 2, where the LWSN output always yields the most accurate data provided
by the duty sensor and the remaining sensors of the mote stay in the standby status.
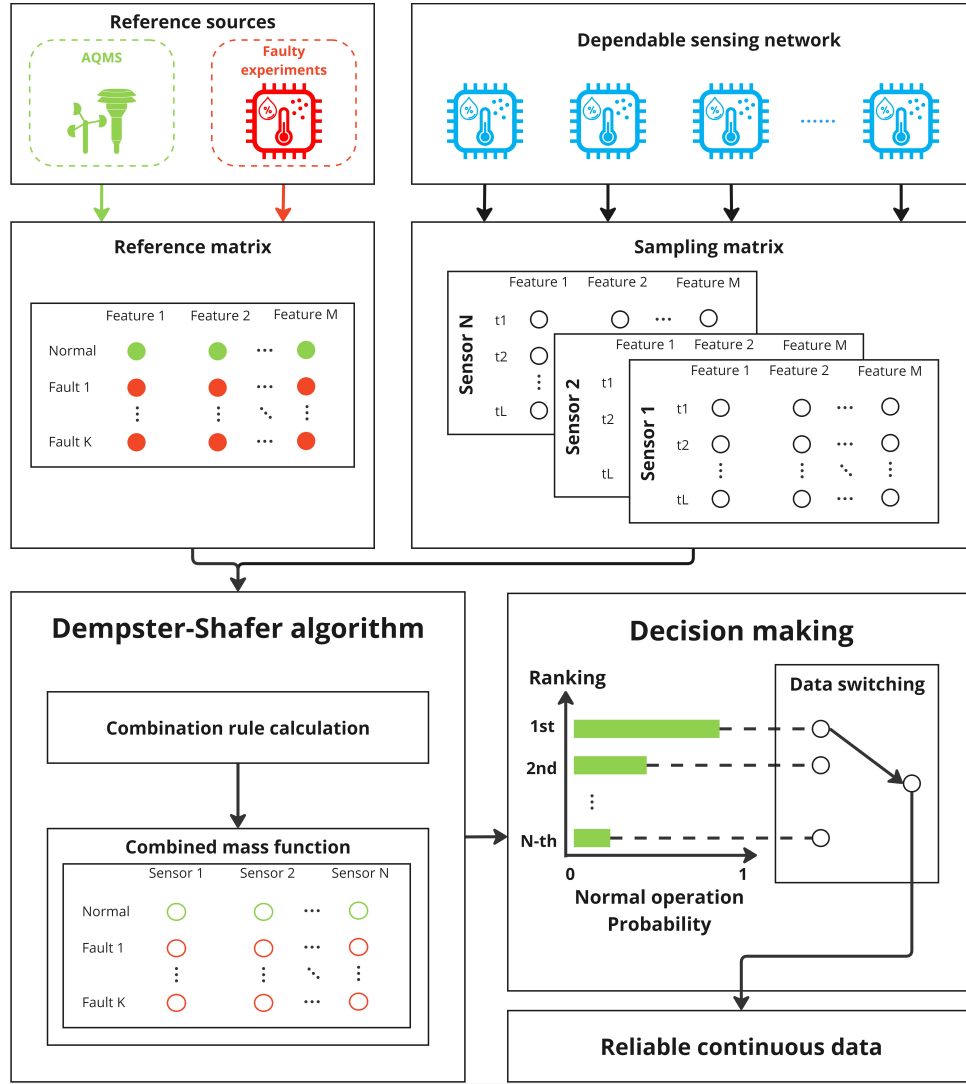
Figure 3.1: Proposed DDSI-LWSN architecture for reliable monitoring.

Algorithm 2 Dependable monitoring with DS-inferred LWSN

1: **function** SENSOR_SCHEDULING($lwsn\_states\_prob$)
2:    **for** each sensor n, states $states\_prob$ in $lwsn\_states\_prob$ **do**
3:       **if** $max(states\_prob) == q^n_{normal}$ **then**
4:          $normal\_probs[n] \leftarrow q^n_{normal}$
5:       **else** Raise flag for faulty operation
6:       **end if**
7:    **end for**
8:    **if** $selected\_sensor = \varnothing$ **then**
9:       Raise flag for dependable system corruption
10:    **else**
11:       $reliable\_sensor \leftarrow max(normal\_probs)$
12:    **end if**
13:    **return** reliable sensor and sampling data $S^{reliable\_sensor}$
14: **end function**

3.3 Identification of faulty scenarios for LWSN

To establish the FoD with all hypothesised states of the LWSN incorporating the DDSI algorithm, we conduct experiments with four colocated air quality sensors in laboratory conditions.

3.3.1 Experimental setup

In this work, PASs are selected among the popular types of LCSs for air pollution monitoring due to their high reliability, ease of deployment, maintenance with standard calibration and weather resilience. Additionally, large-scale PAS networks are currently deployed in the EPAs of several countries, including the US [24] and Australia [82]. Technical specifications of the PAS can be found from the equipment manufacturer [83] or in [82]. Here, external factors affecting the monitoring accuracy of the PurpleAir LWSN may include voltage variations, overheating, abnormal aerosol concentrations, communication issues, and other uncertainties. We have conducted a number of laboratory experiments and real-world tests to assess these factors, as elaborated in Table 3.1.

Table 3.1: Faulty scenarios in reliability assessment of dependable air quality monitoring system

ID	Case	Test Setup	Experimental Rationale and Description
1	Normal operation		All sensors in the dependable monitoring system are working normally.
2	Voltage fluctuation	Voltage supply is changed abruptly.	Inaccurate measurements can originate from an unstable power supply to the sensors. Given that the operated range of voltage supplied for sensors is 5V [83], we selected four lower voltage values—3.5V, 3.8V, 4.0V, and 4.2V—to observe how voltage fluctuations affect the collected data.
3	Overheating	Temperature is increased with a heater.	The air quality sensors are susceptible to a variety of volatile meteorological conditions, for example, an extreme rise of temperature, such as an urban heat island in cities [84].
4	Humidity	A humidifier is located close to the sensor to create an increase in aerosol.	Using laser counters for measuring PM levels, the air quality sensors may yield inaccurate measurements in an adverse impact of aerosol emissions, as recognised in [85].
5	Communication issue	Disconnection of WiFi signals.	Communication loss may happen that affects a wireless sensor, causing its missing information [27].
6	Other uncertainties		An air quality sensor may experience unexpected/undefined fault(s) during its operation due to some impecfectnesses or uncertainties[23].

The setup is shown in Fig. 3.2, comprising four colocated PASs. The dependable quad-sensor system is secured on a 3D-printed holder to collect data necessary for normal operation reference matrix under conditions of voltage fluctuation, overheating temperature, abnormal humidity, communication problems, and other uncertain imperfectnesses.

They were conducted in a controlled laboratory environment, ensuring that any influencing factors different from those in consideration were negligible. The general procedure for the subsequent tests is described in the following:

1. Data from all sensors in the dependable configuration during their normal working state were collected and averaged to obtain reference matrix R;

2. A sensor from the LWSN was applied with specific faults according to the test case. The faulty hypothesis was then formed by averaging the measured parameters over a certain period of time;

3. The sensor in consideration was then subject to the predefined faults, and the resulting data were recorded for the sampling matrix, S.

Figure 3.2: Colocated sensors

3.3.2 Test results and Discussion

The laboratory tests were conducted to identify the feasibility of the proposed framework of dependable DS inference in diagnosing the health of dependable air quality monitoring systems. Throughout the sampling period of 10-minute frequency, recorded data from the quad-sensor system were analysed by the DDSI algorithm for detecting predefined faults. Figure 3.3 presents the DS-inferred probability of sensor faults from the tests to demonstrate that the framework correctly flagged the faults influencing the sensor in various scenarios.

For the low-voltage test, the framework was able to distinguish sudden drops and rapid recoveries of voltage supplied to the sensor, as shown in Fig. 3.3(a). In the abnormal heat test, the resultant probability assignment is depicted in Fig. 3.3(b), showing an increase in the fault probability. DDSI recognised the temperature rise introduced to the sensor and deduced it as a changing dominance between uncertain and heat-related faults. However, the probability of normal operation was computed to be notably low, apparently rendering the temperature change in this case has less effect on the quad-sensor system.
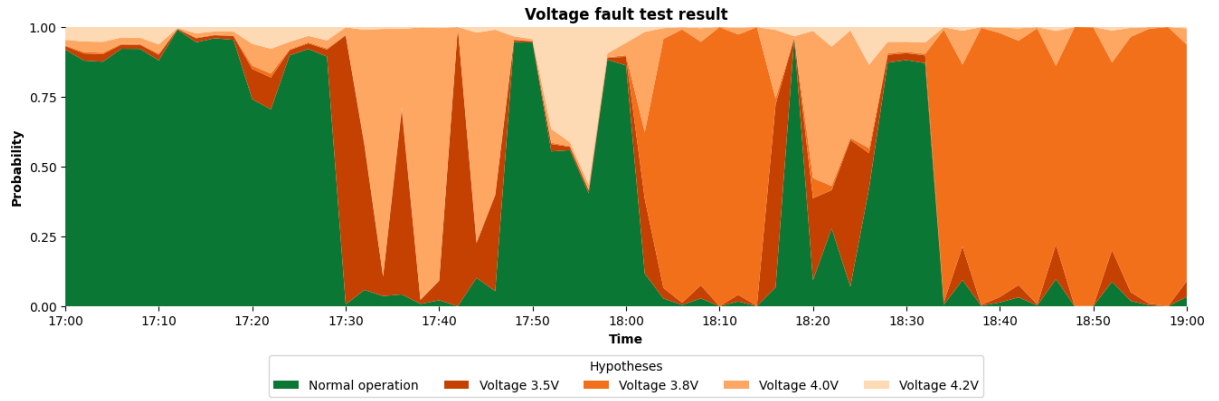
(a) Voltage variations
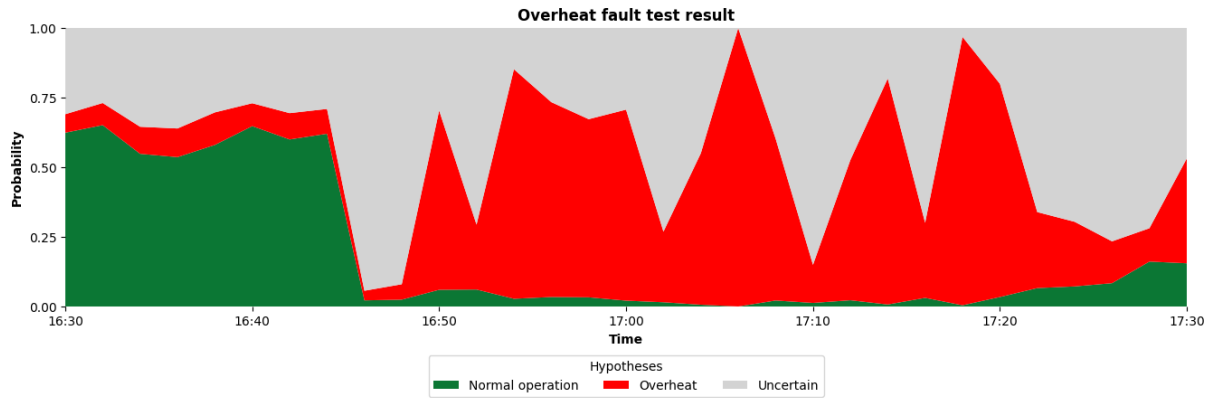
(b) Overheat
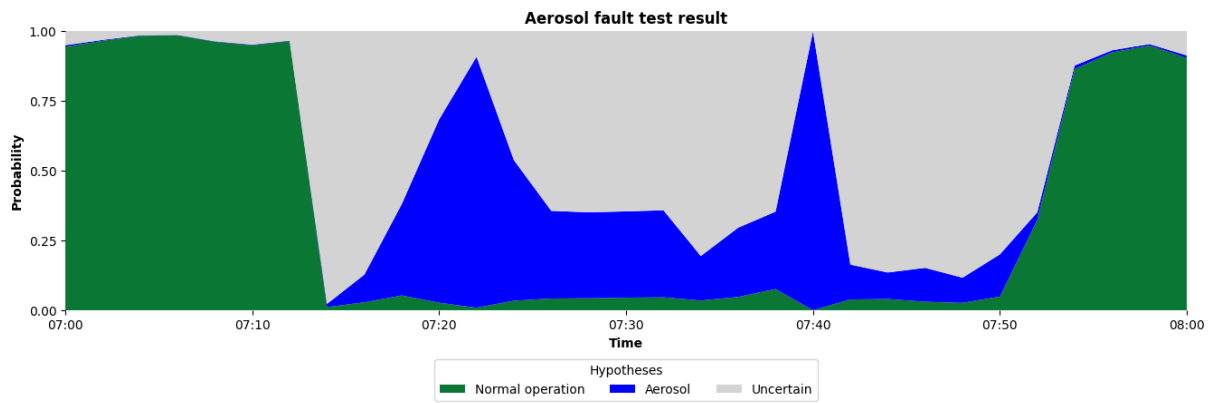
(c) Abnormal aerosol

Figure 3.3: Experimental results for different faulty cases

The humidity increase detection illustrated in Fig. 3.3(c) shows a trend that resembles the temperature rise test, where a sharp decline in normal operation probability signifies the beginning of the excessive injection of PM concentration due to highly concentrated aerosol level. The results confirmed that the DDSI could suggest the sensor was experiencing a humidity fault. Overall, the framework proved that it could differentiate normal operation from either abrupt or prolonged periods of faulty conditions, which effectively signified a higher probability of sensor malfunction for the exclusion of the corrupted data and switching to the most reliable sensor available in the colocated sensor network.

These experimental setup and results provide valuable insights into the framework performance as a basis for moving forward to real-world implementation for urban monitoring of air quality, which is becoming an important problem given the emergence of smart cities.

3.4 Reliable monitoring of urban air quality with colocated low-cost sensor system

This section presents the application of the DDSI framework to monitor real-world air quality in a suburb with a LWSN using PASs.

3.4.1 Real-world system description

To evaluate the real-world implementation of the DDSI algorithm, colocated PASs are installed to form a LWSN near an AQMS in Lidcombe, a suburb in the state of NSW, Australia, located at coordinates -33.88143°S, 151.04676°E. This reference station is currently being operated and maintained by the local authority. An array of PASs is arranged around the AQMS at Lidcombe, as shown in Fig. 3.4, where the sensors are depicted as in its inlet. Here, a triple-sensor mote with three colocated PurpleAirs are referred to as S1, S2, and S3.

Data from the air quality monitoring station and colocated sensors are available from the EPA's respective cloud servers [86]. The datasets of all concerned sources, spanning from March 1, 2021, to December 31, 2022, were collected at an hourly frequency for the reliability assessment. Four common air quality parameters, namely humidity, temperature, $PM_{2.5}$ and PM_{10} were chosen for the evaluation of the developed DDSI-LWSN monitoring system for urban air quality.

3.4.2 Low-cost sensor calibration

The accuracy of the measurement of outdoor PM concentrations depends on the performance of onboard laser counters in PASs [87]. As such, it is essential to conduct a proper

Figure 3.4: PASs and the referent AQMS colocated at Lidcombe, Australia

data calibration prior to their utilisation. To this end, a multi-segment data correction method has been developed by the United States Environmental Protection Agency (US EPA) for PASs operating in outdoor settings. This method could improve the precision of sensor recordings by aligning them with reference-grade data from colocated monitoring stations. The proven efficacy of the correction method drives its implementation and adoption by the Australian EPA [88], as well as the NSW authority.

The calibration of $PM_{2.5}$ readings of each PAS was based on its raw measurement data, taking into account the relative humidity readings. Built on the establishment, the US EPA correction on the PurpleAir dataset ensures that the input data to the DDSI framework matches the state's requirement for air quality monitoring accuracy.

Table 3.2: US EPA set of correction equations for raw PurpleAir $PM_{2.5}$ measurements [89]

$PM_{2.5}$ range	Correction equation
$0 \leq x < 30$	$0.524x - 0.0862 \cdot RH + 5.75$
$30 \leq x < 50$	$\left(0.786\left(\dfrac{x}{20} - \dfrac{3}{2}\right) + 0.524x\left(1 - \left(\dfrac{x}{20} - \dfrac{3}{2}\right)\right)\right) - 0.0862 \cdot RH + 5.75$
$50 \leq x < 210$	$0.786x - 0.0862 \cdot RH + 5.75$
$210 \leq x < 260$	$\left(0.69\left(\dfrac{x}{50} - \dfrac{21}{5}\right) + 0.786x\left(1 - \left(\dfrac{x}{50} - \dfrac{21}{5}\right)\right)\right) - 0.0862 \cdot RH\left(1 - \left(\dfrac{x}{50} - \dfrac{21}{5}\right)\right) + 2.966\left(\dfrac{x}{50} - \dfrac{21}{5}\right)$ $+ 5.75\left(1 - \left(\dfrac{x}{50} - \dfrac{21}{5}\right)\right) + 2.966\left(\dfrac{x}{50} - \dfrac{21}{5}\right) + 8.84 \times 10^{-4}x^2\left(\dfrac{x}{50} - \dfrac{21}{5}\right)$
$x \geq 260$	$2.966 + 0.69x + 8.84 \times 10^{-4}x^2$

3.4.3 Urban air quality monitoring implementation

The proposed DDSI framework was implemented in the colocated triple-sensor mote network to improve the accuracy and reliability of urban air quality monitoring. The factors influencing the sensing performance commonly considered include overheating, excessive aerosol, loss of communication and other imperfectnesses. Given that the supply power for these sensors is generally stable in city conditions, we excluded the cases of insufficient voltage in our analysis. After applying US EPA calibration, the corrected PurpleAir data, along with AQMS data, were appropriately assigned to the reference and sampling matrices for the reliability assessment process in Algorithm 1. Here, the operational states of three colocated sensors at Lidcombe are reflected as shown in Fig. 3.5.
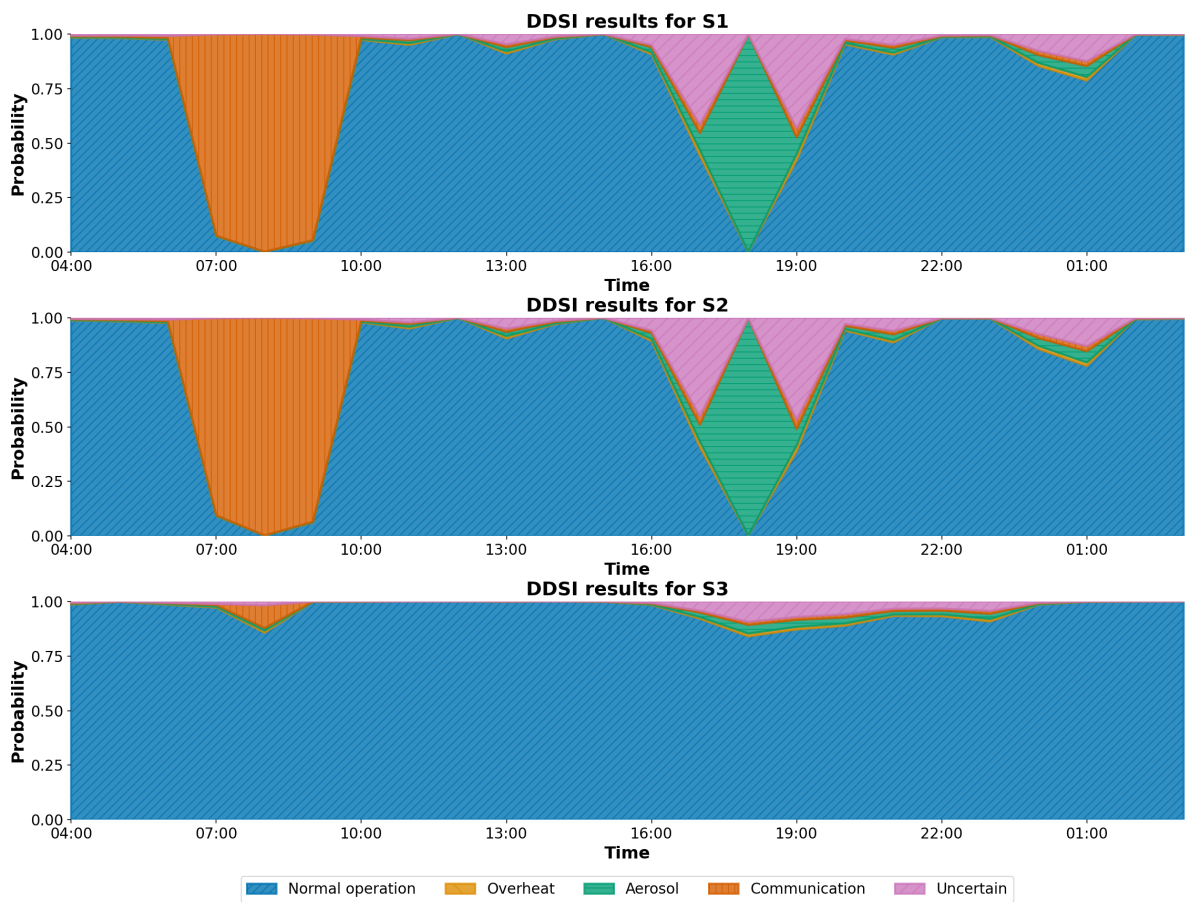
Figure 3.5: DDSI performance with the colocated sensors at Lidcombe

As a general trend, Sensor S1 and Sensor S2 presented relatively similar readings, while Sensor S3 returned different measurements. DDSI made two significant observations. The first one was where DDSI indicated a high probability of communication fault coming from S1 and S2, whereas S3 was less susceptible to that error and maintained a closer alignment to AQMS readings. In this case, the measurements of S1 and S2 provided values of all parameters quite lower than those collected by the station. Eventually, the dependable sensors could manage to recover and presented strong similarity to the station data.

The second drastic change in the result was that DDSI recognised a strong likelihood of excessive aerosol levels influencing S1 and S2. On the other hand, S3 was still identified to be working properly, unaffected by exogenous abnormalities. Another remark from this result was a consistently low probability of overheating fault among all sensors.

The concept of dependability illustrated above can be clearly explained in the decision-making layer of our proposed system as shown in its architecture of Fig. 3.1, wherein the normal operation probability of each dependable sensor returned by DDSI was ranked in descending order with the highest selected as the standby sensor considered to be the most reliable at a specific sampling. This is where the sensor switching schedule is devised, as presented in Fig. 3.6(a) for a day when DDSI detected abnormal readings of $PM_{2.5}$ concentrations from S1, S2 and switched to S3. With the proposed DDSI algorithm, although S3 largely outperformed its counterparts, there were instances when the other sensors provided more reliable measurements. This can be observed during 3-4 pm, 8-9 pm and 10-12 pm on the day.
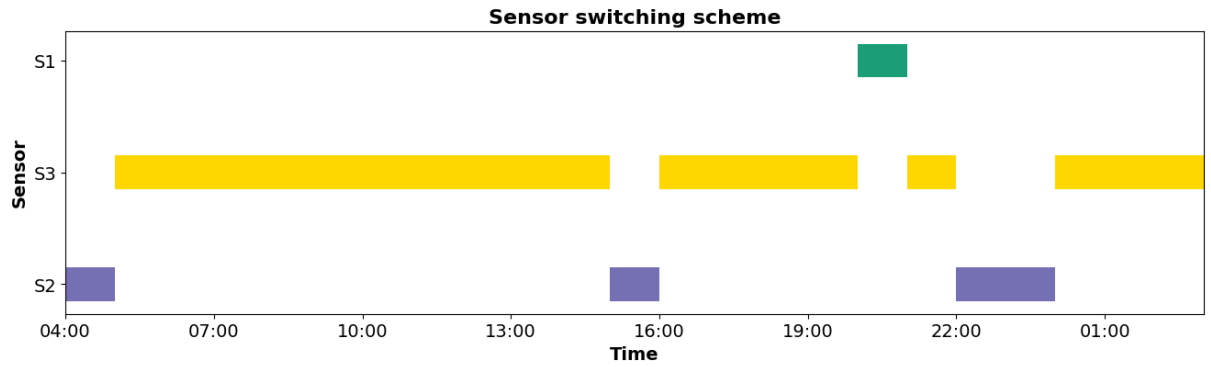
(a) Reliable sensor selection schedule based on DDSI results
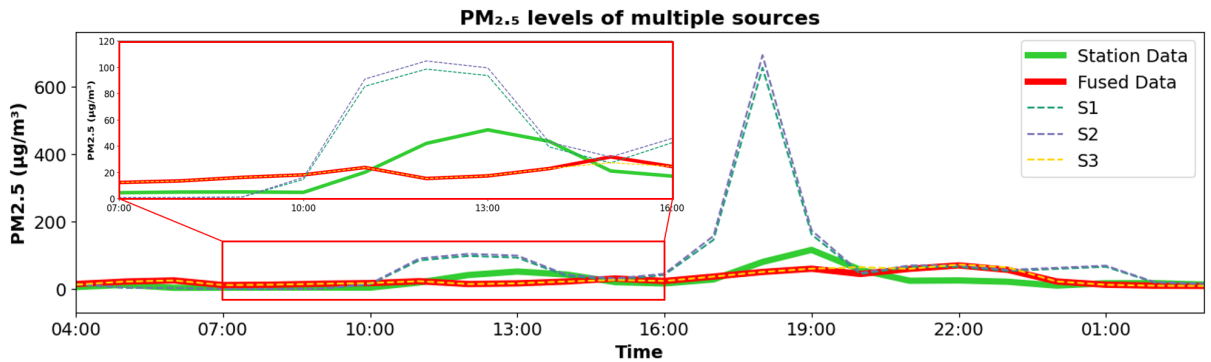
(b) $PM_{2.5}$ concentration comparison between station, fused, and individual sensor data

Figure 3.6: Measured and fused data from DDSI-LWSN compared to station data

3.4.4 DDSI-LWSN monitoring results

To assess the accuracy and viability of the DDSI-based LWSN for urban monitoring of air quality, we compare the results obtained from the implementation described above

with those obtained from AQMS, using widely recognised metrics such as the Root Mean Squared Error (RMSE) and coefficient of determination (R^2), which are formulated below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}, \tag{3.11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{3.12}$$

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{3.13}$$

where N is the size of the considered datasets, y_i and $\hat{y}_i$ are, respectively, the output of the proposed dependable LWSN using colocated PurpleAirs and Lidcombe AQMS data.

As a result of the DDSI algorithm, the output of the proposed framework is a continuous data stream constructed from the most reliable sensor in the dependable configuration at specific time intervals. In particular, Fig. 3.6(b) illustrates the monitoring process for $PM_{2.5}$ inside the LWSN with three sensors in separation compared to the results obtained from the DDSI fusion and the benchmark from AQMS. In accordance with the moment when a humidity state was detected by the DS inference, as shown in Fig. 3.5, the $PM_{2.5}$ level from S1 and S2 exhibited discernible spikes. Furthermore, the zoom-in image shown in the inlet of Fig. 3.6(b) shows the concentration of $PM_{2.5}$ estimated by the DDSI framework follows the most reliable sensor (here S3) and is very close to the station data.

Notably, under normal conditions, each sensor may perform better in estimating certain parameters, for example, humidity, PM_{10} with S2, temperature with S3 and $PM_{2.5}$ with S1. Overall, Table 3.3 presents the statistical results obtained from running DDSI at the surveyed site. The evaluation metrics are obtained from the data of Lidcombe station and individual sensors as well as their fusion. Alongside, the improvements of the fused data compared with individual sensors are included in the table. This verifies the merits of the DDSI-LWSN monitoring system and advocates further for its application in complement to the limited number of state-run stations for reliable monitoring of air quality.

3.4.5 Discussion

The developed DDSI framework is straightforward and computationally inexpensive, taking into consideration the data-intensiveness of near real-time data processing of air quality sensors. The proposed approach amalgamates the capability of DS-inferred fault detection and probability-based dependable switching with the availability of the data-rich colocated LCSs to provide the best estimation of air quality by fusing measurements of

Table 3.3: Statistical comparison of colocated AQMS data against fused and individual sensor data and the improvement of fused data over each sensor in the dependable scheme

Meteorological parameter	Humidity				Temperature			
	RMSE (%)	Improved RMSE (%)	R^2	Improved R^2 (%)	RMSE (°C)	Improved RMSE (%)	R^2	Improved R^2 (%)
Fused data	**2.773**	N/A	**0.918**	N/A	**0.581**	N/A	**0.906**	N/A
S1	3.259	14.9	0.886	3.6	1.277	54.5	0.547	65.7
S2	3.037	8.664	0.900	1.8	1.453	60.0	0.414	119.0
S3	4.678	40.719	0.764	20.0	0.677	14.2	0.873	3.8
Air pollutant	$PM_{2.5}$				PM_{10}			
	RMSE ($\mu g/m^3$)	Improved RMSE (%)	R^2	Improved R^2 (%)	RMSE ($\mu g/m^3$)	Improved RMSE (%)	R^2	Improved R^2 (%)
Fused data	**1.159**	N/A	**0.892**	N/A	**2.110**	N/A	**0.721**	N/A
S1	1.238	6.381	0.878	1.595	3.417	38.237	0.268	169.0
S2	1.566	25.990	0.805	10.807	3.361	37.215	0.292	147.2
S3	3.428	66.190	0.067	1231.3	3.728	43.392	0.128	461.0

various ambient parameters. The monitoring system can also be extended to incorporate the fusion of multiple sources. Here, the DDSI switching mechanism plays a pivotal role in the framework, enabling a dynamic selection of the most reliable sensor at a given timestamp.

Our statistical analysis, summarised in Table 3.3, evaluates the framework's performance based on the fused data resulting from the DS-inferred algorithms. With respect to the metrics RMSE and R^2, the measurement quality of each sensor may not be consistent for a particular parameter of air quality. As such, a data fusion technique is expected to retain the best responses of each sensor to achieve a good alignment between the sensors and station data. Here, the proposed DDSI framework for LWSN has significantly improved the monitoring performance of the system with all air quality parameters considered, namely humidity, temperature, and PM. In all cases, our fused data approach outperforms individual sensors with relatively low RMSE compared to the station observations across various meteorological parameters and air pollutants in monitoring. Notably, with our monitoring system for estimating the $PM_{2.5}$ level, the R^2 metric could reach 0.894 in relation to AQMS, quite consistent with the correlation of 0.88 of colocated monitoring devices with accurate methods adopted by the US National Institute of Standards and Technology as reported in [90].

The results obtained indicate the promising application of the proposed approach for data fusion of multiple sources to improve accuracy and reliability in air quality monitoring. Our future work will be devoted to implementing the DSET and improving the discount methods to better handle environmental unpredictability and sensor variability.

3.5 Conclusion

In this chapter, we have presented the development and implementation of a reliability-enhanced inference framework for dependable low-cost air quality sensor systems based on DSET. Along with the proliferation of LWSNs, the approach offers a viable solution to tackle the problem of air quality monitoring in city suburbs. Test results in laboratory settings indicated that the proposed framework could differentiate between normal operation and predefined faulty scenarios. The monitoring system was then applied in real-world conditions with PASs colocated at a state-run monitoring station. Here, the dependable Dempster-Shafer inference monitoring system was developed and deployed in a suburb to confirm its ability to improve its accuracy and reliability compared to benchmark data collected by the station. By fusing data from multiple sensing units, the dependable monitoring system could discern their healthy states and erroneous measurements, ensuring data integrity with a continuous, reliable stream of air quality information as the final outcome. The development is promising for data fusion of multiple sources, with the proposed framework being available for a dense network of colocated LCSs to monitor urban air quality in conjunction with existing stations.

Chapter 4

Data Fusion with Ensemble Learning Model

4.1   Introduction

DL has emerged as a powerful approach to significantly improve forecast accuracy for air quality estimation. Several models have been developed, demonstrating their own merits in some scenarios and for certain pollutants. In nowcasting, the prediction of air pollution over a small time period essentially demands accurate and reliable estimates, especially in the event cases. From there, selecting the most suitable model to achieve the required forecast performance remains challenging. This chapter presents an ensemble framework based on DSET for data fusion to identify the most accurate and reliable forecasts of air pollution obtained from multiple deep neural network models. Our framework is evaluated against three popular ML methods, namely, LightGBM (Light Gradient-Boosting Machine), Random Forest (RF), and XGBoost (Extreme Gradient Boosting). Experiments are conducted on two horizons: 6-hour and 12-hour predictions using real-world air quality data collected from state-run monitoring stations and LWSNs.

One in three people worldwide lacks access to any form of multi-hazard early warning system, despite its proven ability to substantially reduce disaster-related fatalities by threefold [55, 91]. Nowcasting, which provides weather predictions up to six hours ahead, is one of the integral components in such systems, facilitating timely detection of natural hazards, as highlighted by the UN' Early Warning for All initiative [55]. In the context of smart cities, nowcasting adds significance to the microclimate management scheme, where it can offer local insights for public information dissemination and, in cases of extreme events, community-specific dispatch of responsive measures.

Due to the stochastic and volatile nature of environmental conditions, achieving accurate nowcasting results remains challenging, which requires multiple models for different spatial and temporal scenarios [5]. Therefore, researchers worldwide have been exploring

various approaches, especially ensemble models, to predict weather events effectively [92]. For instance, by leveraging radar data from the Romanian National Meteorological Administration, [93] developed NowDeepN, an ensemble of deep neural networks capable of accurately nowcasting heavy precipitation and hail. Furthermore, data from multiple remote sensing sources collected by the National Oceanic and Atmospheric Administration and the China Meteorological Administration contributed to the development of Nowcast-Net [94], integrating the physics-informed layer with neural models. In Spain, researchers successfully addressed the issue of fog-related accidents by implementing an ensemble DL framework. This framework demonstrated higher performance in nowcasting fog events compared to individual models [95]. Nowcasting in urban air quality estimation plays an important role in microclimate management in cities, though its applications are relatively limited.

The real-time demand of nowcasting requires straightforward and robust methods that can efficiently handle uncertainty in environmental conditions [96]. DSET has proven to be an effective data fusion approach, where it fused predictions from an ensemble of ML models to enhance precipitation classification and rainfall estimation in Algeria [97]. In another research, DSET was employed to fuse flood susceptibility forecasts from RF and Support Vector Machine models, improving the ultimate accuracy in the face of a multitude of conflicting flood conditioning factors [98]. The ground for adopting DSET in nowcasting is evident in previous works through its notable ability to assimilate environmental parameters and forecasts. Given the heterogeneity of urban environments, where localised ambient conditions may affect the performance of certain models, there is a clear need for a DS-based model selection at a decision level of an ensemble nowcasting framework.

We propose a dynamic model selection based on DS data fusion scheme for an ensemble DL nowcasting framework designed for air quality in urban areas. The ensemble framework is intended to select the most accurate nowcast among member models. Through implementation on real-world air quality datasets and extensive statistical analysis, the proposed framework demonstrates its feasibility and effectiveness in urban air quality nowcasting. The contributions of this work to the literature include (i) ensemble DL nowcasting framework on real-world datasets collected by multi-scale air quality monitoring network based on the substantial works [4, 5, 3], and (ii) the development of DS-based algorithm for multi-metrics fusion for dynamic selection of best-performing nowcasting results originally from the paper *Dempster-Shafer ensemble learning framework for air pollution nowcasting*.

The remainder of the chapter is organised as follows. Section 4.2.1 introduces the DL models integrated into the ensemble nowcasting framework and the development of the DS-based mechanism for the dynamic selection of best-performing predictions with multi-metrics criteria. Following that, we present the nowcasting results of the proposed

framework on the urban air quality monitoring network in Section 4.2.2 and discuss its overall performance. The chapter is concluded in Section 4.3.

4.2 Dempster-Shafer ensemble learning framework for air pollution nowcasting

4.2.1 Dempster-Shafer-based ensemble learning framework

Deep learning models

An ensemble model combines multiple ML or DL models to improve predictive performance. Each model, or "learner," captures distinct patterns in the data, making the ensemble more capable of handling different scenarios [99]. This approach ensures robustness and reliability by leveraging the strengths of individual models to adapt to varying temporal dynamics. This adaptability is essential for accurate air quality nowcasting in urbanised areas, where many localised environmental factors, such as traffic flows, household emissions from energy usage and interactions between changing weather conditions, govern the shift in pollutant levels.

In the proposed framework, each DL model chosen as a member learner exhibits its own strengths in predicting certain pollutant patterns from the input data. The following learners, resulting from a series of collaborative projects between UTS and NSW-DCCEEW [4, 5, 3], are incorporated into the ensemble architecture and their configurations are tabulated in Table 4.1:

- **1D-CNN (Convolutional Neural Network)**: A robust convolutional neural network for capturing temporal patterns.

- **Artificial Neural Network (ANN)**: A general network can model all types of data including time series.

- **Long Short-Term Memory (LSTM)**: A special recurrent model captures long-term patterns of time series.

- **Gated Recurrent Unit (GRU)**: A type of lightweight LSTM with less number of internal gates.

- **Bidirectional LSTM (BiLSTM)**: A type of LSTM network with learning capacity in two directions.

- **Convolutional LSTM (CNN-LSTM)**: A hybrid model with robustness of spatial-temporal learning.

Table 4.1: Member learners of ensemble learning framework

Member learners	Configurations
1D-CNN	Conv1D(64) - MaxPooling - Dense(64)
ANN	Dense(100, 2 layers)
LSTM	LSTM(128) - Dense(64)
GRU	GRU(128) - Dense(64)
BiLSTM	BiLSTM(128) - Dense(64)
CNN-LSTM	Conv1D(64) - LSTM(128) - Dense(64)

The selection, after tuning, of hyperparameters for each DL model in the ensemble learning framework is presented in Table 4.2. These hyperparameters are shared across all member learners, ensuring consistency in the training and prediction processes while preserving each model's particular suitability in handling temporal distributions.

Table 4.2: Hyperparameters for ensemble learning framework

Hyperparameters	Values and types
Input layer (historical data)	12
Output layer (prediction horizon)	6 - 12
Epoch	50
Batch size	512
Learning rate	0.001
Patience (Early stopping)	5
Loss function	Mean squared error
Optimiser	Adam

DS framework for ensemble learning

Multiple strategies for constructing decision levels in ensemble learning have recently been established [100]. Given the advantages and disadvantages of each nowcasting model, and in the face of air pollution volatility in urban conditions, for instance, PM originating from anthropogenic sources, leveraging the forecast from all models may still encounter inaccuracy in the final estimation. As such, an ensemble model that employs a dynamic model selection mechanism to utilise the capabilities of each learner offers adaptability in real time for various environmental conditions and thus, can improve forecast accuracy [101]. For this, the Dempster-Shafer-based ensemble learning framework (DSEL) is designed to enhance overall predictive performance by dynamically selecting the most reliable forecast from the diversity of DL learners to meet the nowcasting requirements.

The architecture of DSEL, depicted in Fig. 4.1, demonstrates the flow of information from individual nowcasts produced by member learners to the DS selection process for determining the best-performing model. Going in-depth, key components of this architecture include the calculation of performance indices between each learner's predictions

and real observations from monitoring instruments. The DS algorithm, which is defined in 4.2.1, receives those metrics and then computes the similarity level for each model. Finally, the predictions that align most closely with real-world measurements are made the concluding nowcasting results of DSEL.
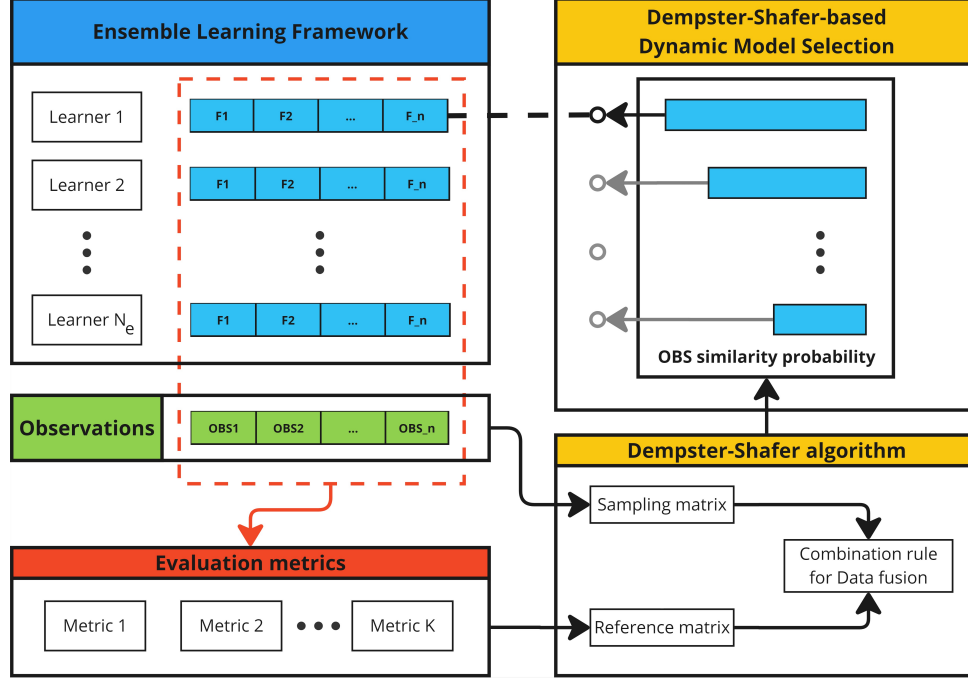
Figure 4.1: DSEL framework for nowcasting: the architecture

Dynamic model selection mechanism

In this section, we elaborate on the model selection approach based on DSET at the decision-making unit of the ensemble learning framework. The FoD in DSET lays the foundation for the ensemble model selection mechanism. FoD declares a finite set of all possible hypotheses of the concerned problem, which are the individual learners here. We denote the FoD as the set

$$H = \{p_j, j = 1, \ldots, N_e\}, \tag{4.1}$$

where N_e is the total number of nowcasting estimators, and p_j denotes the probability of model j achieving predictions similar to real observations.

For evaluation, we consider the performance-indicative metrics, where each provides different insights into the relation of prediction results with real observations. They are listed in the following,

- The RMSE: A common metric for measuring the average magnitude of the prediction error, taking into account both the magnitude and variability of errors.

- Pearson's correlation coefficient r: Indicates the strength and direction of the linear relationship between predictions and observations.

- The Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{4.2}$$

where n is the prediction horizon, y_i and $\hat{y}_i$ are, respectively, the real observations from monitoring instruments and the ensemble nowcasting results. MAE provides the average absolute difference between predictions and observations.

The metrics RMSE, MAE, and correlation coefficient r are popular in DL time-series forecasting. However, the selection of a best-performing learner based on a single metric may overlook other indications, leading to some bias. To identify the best-fit prediction from an individual learner, DSET is tasked with the fusion of multiple evaluation metrics to reach a balance by using a multi-criteria selection process. Here, the metrics are considered sources of evidence for the development of the BPA in DSET. Quantifying the degree of support for the similarity between forecast values and observations, a BPA for each member learner is assigned based on the evidence provided by the model's performance indices. It is essential to note that the mass, or probability, associated with a certain hypothesis in the defined FoD must be in the range between 0 and 1. Therefore, mathematically, the mass function, denoted as m, is portrayed as

$$m : 2^H \to [0, 1],$$

where it must satisfy

$$\begin{cases} m\left(\varnothing\right) = 0, \\ \sum_{p \subseteq H} m\left(p\right) = 1. \end{cases} \tag{4.3}$$

For computation, we first formulate a reference matrix that represents the FoD and BPAs. The reference matrix

$$R = [r_{jk}], \ k = 1, 2, \ldots K, \tag{4.4}$$

where K is the number of evaluation metrics used in the multi-criteria analysis. Since we are concerned with the resemblance of forecasts to observational values while considering multiple metrics, the sampling vector $S = [s_k]$ is derived from the assumption of exact alignment of forecasts to observations where the evaluation statistics are ideal. In other words, the sampling vector contains the optimal values of the metrics where they indicate the best performance.

In compliance with the conditions (4.3), we took the first step in deriving BPAs by the quantification of the similarity between nowcasting values and the ideal scenario of

exact fit with measured values from monitoring devices, which is expressed as

$$d_{jk} = |s_{jk} - r_{jk}|. \tag{4.5}$$

The multi-metrics-fused similarity probability associated with an individual learner j is calculated as

$$q_{jk} = \frac{d_{jk}^{-1}}{\sum_{k=1}^{K} d_{jk}^{-1}}, \tag{4.6}$$

which establish the probability matrix $Q = [q_{jk}]$.

The statistical values of the evaluation metrics are transformed into the degree of forecasts aligning with actual observations in the matrix Q. However, the entries in each column only represent the supporting degree from one metric. Here, Dempster's rule of combination in DSET is imperative in combining evidence from multiple metrics to derive a balanced and inclusive confidence level of a learner closely matching observations. If evidence from different metrics is represented by a mass function m, then given any two mass functions m_1 and m_2, Dempster's rule aggregates them into a joint belief m_{12} by using the orthogonal sum ($\oplus$) as expressed in 3.9, 3.10.

The pseudocode presented in Algorithm 3 summarises the DS-based model selection mechanism and describes the step-by-step procedure for dynamically selecting the best model at each nowcasting interval. By incorporating DSET in the decision-making layer of DSEL, the proposed ensemble learning framework gains the necessary flexibility to continuously varying input data to improve the accuracy required for nowcasting.

4.2.2 DS ensemble learning implementation for urban air quality nowcasting

Multi-scale air quality monitoring network

Rapid urban growth in smart cities necessitates microclimate management to address challenges posed by climate change and safeguard public health. Multi-scale environmental monitoring combining reference-grade AQMS and low-cost air quality sensors is becoming more and more popular, e.g., in Europe [102] and Australia [82]. To this end, our proposed DSEL framework has been implemented to perform along such networks in urban settings within the city of Sydney, NSW, Australia. Our sites of interest, illustrated in 3.4, include the monitoring station located in Liverpool suburb at the co-ordinates -33.93132°S, 150.90727°E and PASs located in Lidcombe suburb at coordinates -33.88143°S, 151.04676°E. These data sources are under the management of the local authority, enabling a quality-assured, continuous and real-time information stream.

We retrieved the data of Liverpool AQMS through the publicly available online portal

---

Algorithm 3 Dynamic Dempster-Shafer-based model selection for ensemble learning

---

1: **function** STATISTICS($ensemble\_nowcast$, obs)
2:    **for** $k = 1 : K$ **do**
3:       Compute evaluation metrics $stats_k$ (Eq. 3.11 4.2, 3.13)
4:       Compute intermediate mass $inter\_mass[j]$ from pair-wise hypotheses (Eq. 3.9)
5:    **end for**
6: **end function**
7: **function** MATRIX_FORMATION($ensemble\_nowcast$, obs)
8:    **for** $j = 1 : N_e$ **do**
9:       $stats \leftarrow$ STATISTICS($ensemble\_nowcast_j$, obs)
10:      $ref\_mat[j] \leftarrow stats$
11:   **end for**
12:   $samp\_mat$ contains ideal statistics
13: **end function**
14: **function** COMBINED_MASS($discounted\_mass$)
15:   **for** $j = 1 : N_e$ **do**
16:      Compute conflict coefficient K_c (Eq. 3.10)
17:      Compute intermediate mass $inter\_mass[j]$ from pair-wise faults (Eq. 3.9)
18:   **end for**
19:   $combined\_mass \leftarrow inter\_mass[N_e]$
20: **end function**
21: **function** DSET_ALGORITHM($ref\_mat, samp\_mat$)
22:   Compute distances (Eq. 4.5)
23:   Compute probabilities (Eq. 4.6)
24:   $similarity\_prob \leftarrow$ COMBINED_MASS($ensemble\_nowcast_j$, obs)
25: **end function**
26: **function** MODEL_SELECTION($similarity\_prob$)
27:   **if** $max(similarity\_prob) == similarity\_prob_j$ **then**
28:      $best\_model \leftarrow j$
29:   **end if**
30:   **return** best-fit nowcast $ensemble\_nowcast_{best\_model}$
31: **end function**
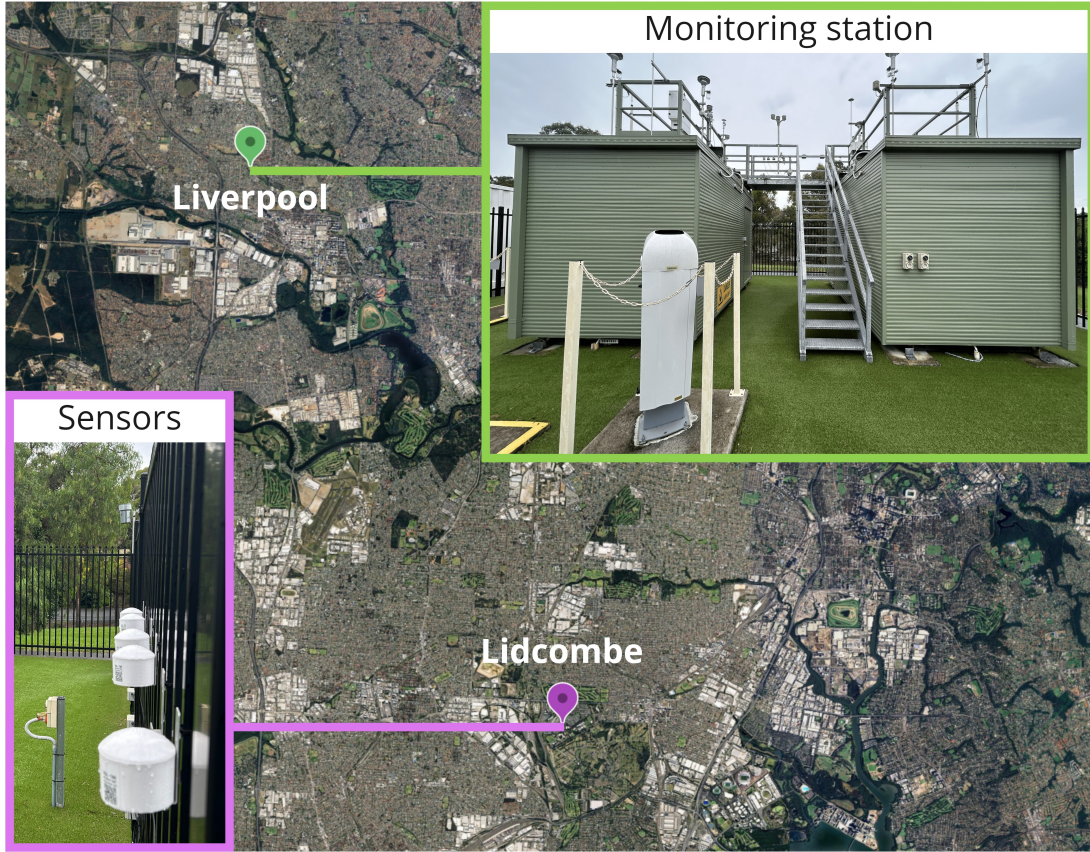
Figure 4.2: Surveyed urban air quality monitoring instruments in Sydney

of the local environmental agency. This dataset spans the period from January 1, 2018 to September 30, 2023. As for the PAS in Lidcombe, we collected the data from March 1, 2021 to June 30, 2022. The focal point of both datasets is hourly $PM_{2.5}$ concentrations, the primary target in this study, serving for the training and evaluation of the proposed DSEL framework against other ensemble models for cross-comparison purposes. The training and nowcasting of $PM_{2.5}$ for all ensemble models were performed on an Interactive High Performance Computing (iHPC) server equipped with the NVIDIA A2 GPU to ensure uniformity in the learning process and support rigorous evaluation afterwards.

Air pollution nowcasting with DS ensemble learning: Results and Evaluation

To demonstrate the application of our proposed framework on real-world datasets, we present the final results of DSEL in the form of 6-hour nowcasting and 12-hour very short-term forecasting for the reference-grade station in Liverpool and PAS in Lidcombe. Both horizons provide necessary information for quick responses and early warning in local communities [30]. The performance of our proposed framework is compared with different ensemble ML methods LightGBM, RF regression and XGBoost. These models were chosen due to their ability to handle complex, high-dimensional data and capture both linear and nonlinear patterns, making them potential for air quality estimation.

- LightGBM: A gradient-boosting model known for its speed and accuracy, particularly on high-dimensional datasets. LightGBM's leaf-wise tree growth enhances its ability to capture complex patterns, essential for fluctuating air quality data.

- RF regression: An ensemble of decision trees that reduces variance by averaging predictions, improving stability. RF is effective at modelling intricate relationships in air quality features and performs well with missing data.

- XGBoost: This model is optimised for predictive power by applying regularisation to reduce overfitting, making it well-suited for the irregular trends and outliers common in air pollution data collected by LCSs.

Each of these ensemble methods was trained on the exact datasets as DSEL, enabling a comprehensive comparison with the DSEL model.

Figure 4.3 presents the time-series prediction from DSEL and other ensembles plotted alongside real observations from professional-grade and cost-effective instruments. Specifically, Fig. 4.3a and Fig. 4.3c show a general trend of transitioning from a *Fair* to a *Good* level of $PM_{2.5}$ concentration as categorised by the local authority [103]. While all ensemble models considered here are able to recognise the constant inclination of observations, DSEL presents the ability to trace the magnitude of its target when the others show significant biases in magnitude. The decreasing tendency of $PM_{2.5}$ in the 12-hour forecast accompanies fluctuations in the form of sudden spikes and valleys. Despite the challenges posed, the 12-hour DSEL forecasts can adapt well to such abrupt changes to match observational values at Liverpool station, as depicted in Fig. 4.3c.

The measurements from low-cost sensors are expected to have sharper edges than their counterparts collected by regional monitoring stations due to inherent differences in hardware and deployment sites. The PAS readings shown in Fig. 4.3b illustrate the rapid fluctuation of $PM_{2.5}$ values. It is observed that DSEL predictions can identify and follow the general trend of the PAS measurements in spite of its high instability, whereas predictions from other ensemble models struggle to capture the extrema and instead converge toward the overall mean value. Figure 4.3d portrays a different scenario with a steep and jagged decline of the $PM_{2.5}$ ground-truth. In this case, it is visible that DSEL forecasts can follow the general downward trend from PAS, especially in the middle of the prediction window. At the two ends, the forecast shows some biases compared to observations, likely due to extreme shifts in $PM_{2.5}$ levels. The underperformance of ML ensemble models is exhibited in cases of considerably high concentrations of PM throughout the test datasets. However, at relatively lower concentrations, the distinction is not as significant and at a comparable level to DSEL. To comprehensively evaluate the performance, we tabulated the statistics in Table 4.3 and 12-hour forecasting in Table 4.4.

Through statistical analysis, the results for 6-hour horizon produced by DSEL have significantly lower RMSE and MAE compared to other ensemble nowcasting models,
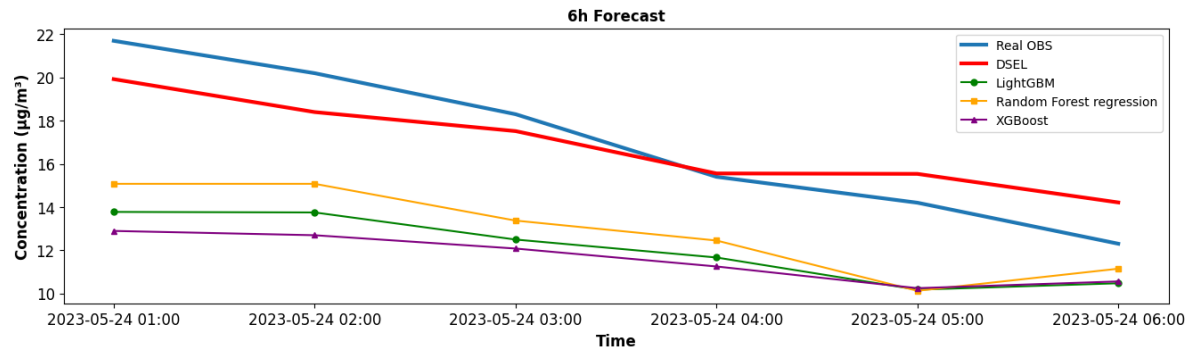
(a) Nowcasting results for AQMS
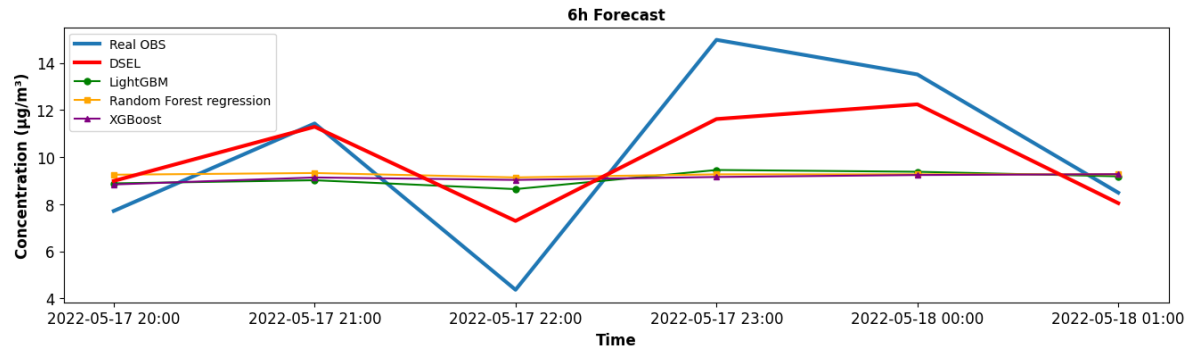
(b) Nowcasting results for PAS

(c) 12-hour forecasts for AQMS

(d) 12-hour forecasts for PAS

Figure 4.3: Comparison of real observations from monitoring instruments against the proposed DSEL, LightGBM, RF regression and XGBoost.

Table 4.3: Performance comparison between different ensemble models on nowcasting of $PM_{2.5}$

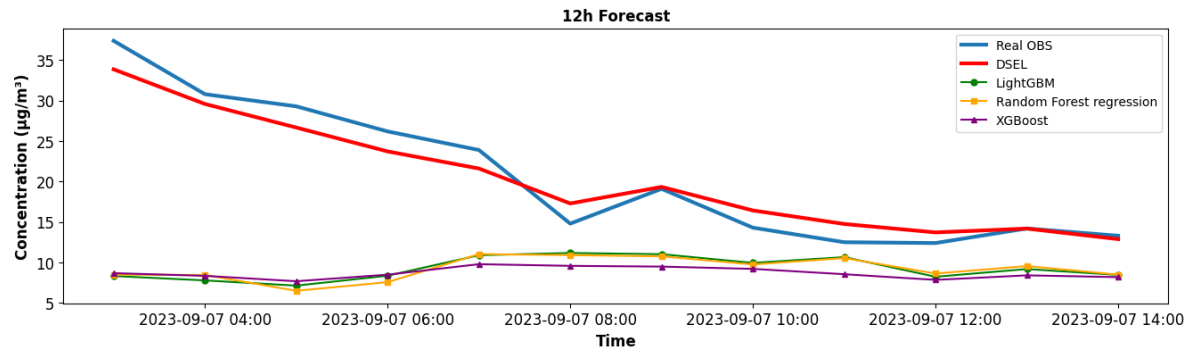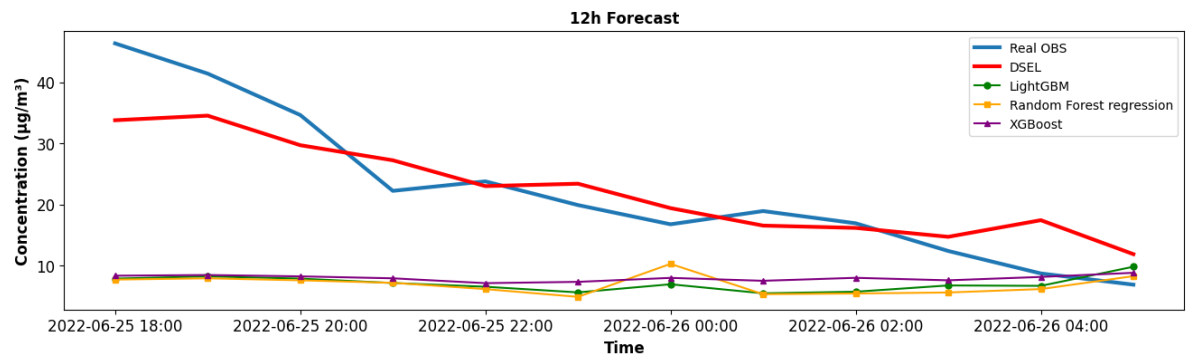| Prediction horizon (hr)	Data source	Model	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	Pearson's r
6	AQMS	DSEL	**1.442**	**1.294**	**0.990**
		LightGBM	4.084	3.662	0.817
		RF regression	4.207	3.605	0.799
		XGBoost	4.312	3.872	0.769
	PAS	DSEL	**1.972**	**1.570**	**0.935**
		LightGBM	3.503	3.037	0.916
		RF regression	3.671	3.193	0.706
		XGBoost	3.677	3.162	0.486

meaning the nowcast from DSEL closely matches the magnitude of real observations, with RMSE ranging from 1.44 to 1.97 $\mu g/m^3$ and MAE between 1.29 to 1.57 $\mu g/m^3$ for the AQMS and PAS respectively. The correlation coefficient also indicates the outperformance resulting from the proposed DSEL, which is tabulated as 0.99 for AQMS and 0.93 for PAS. Within the 12-hour prediction horizon, the statistics of AQMS forecasts from different ensemble models are quite comparable. Nevertheless, it is acknowledged that DSEL forecasts offer better accuracy. In contrast, there is a clear distinction between the performance of DSEL in the 12-hour forecast of PAS data compared to other ensemble models. The RMSE and MAE of the 12-hour forecast on PAS data are 1.94 and 1.63 $\mu g/m^3$, within close proximity to their counterparts in the nowcasting, while maintaining a high correlation of 0.92. These results emphasise DSEL's high predictive accuracy even in a longer prediction window, especially in dealing with erratic data patterns of low-cost ambient sensors.

Table 4.4: Performance comparison between different ensemble models on very short-term forecasts of $PM_{2.5}$

Prediction horizon (hr)	Data source	Model	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	Pearson's r
12	AQMS	DSEL	**3.778**	**3.030**	**0.960**
		LightGBM	3.908	3.362	0.921
		RF regression	4.702	4.064	0.958
		XGBoost	4.055	3.525	0.934
	PAS	DSEL	**1.943**	**1.637**	**0.924**
		LightGBM	2.974	2.337	0.840
		RF regression	2.845	2.202	0.839
		XGBoost	3.185	2.630	0.917

4.3 Conclusion

The ability to predict the status of particulate matter concentrations in urban areas for short time horizons is pivotal for supporting early warning systems in a metropolis. This chapter presents a DSEL framework for the multi-criteria model selection mechanism to achieve the best forecast performance. The predictions obtained from real-world datasets suggested that the proposed DSEL is capable of tracking the overall propensity observed by monitoring instruments and outperforms other ensemble ML models, particularly in the presence of high concentrations of the pollutant. The adaptability and accuracy of the framework consolidate its suitability in volatile environments. For future work, we plan to extend it to accommodate multivariate nowcasting and broaden the outputs to other key pollutants that may cause critical impacts on the environment and human health. The ultimate aim is to improve the computational efficiency and scalability of DSEL across the multi-scale monitoring network to provide fine-grained air quality nowcasting to the community level and comprehensively facilitate the informative decision-making process and timely responses from the stakeholders.

Chapter 5

Visualisation platform for multi-scale air quality monitoring and forecast

5.1   Introduction

Building on the premises established in the preceding chapters, this chapter introduces the development and implementation of a visualisation platform to deliver presentable outputs of the proposed frameworks in a manner accessible to stakeholders. While ensuring reliable monitoring data from LWSN and enhancing the accuracy of air pollution forecasting through ensemble learning techniques are essential for resilient environmental management infrastructures, making the data intelligible and easily accessible to a wide range of users is equally influential. Without an intuitive interface, the practical value of these data—particularly in time-sensitive decision-making—may be diminished.

Globally, air pollution remains a severe public health issue, with the majority of the world population constantly breathing air that surpasses the recommended threshold [1]. The problem is exacerbated when the general public is not equipped with comprehensible resources. Given the known health impacts of air pollution, clear and timely communication is of great concern for governmental agencies, the private sector, and the public. Monitoring and forecasting tools that provide interpretable outputs can support decisions across health advisories, policy planning, and individual behavioural changes.

From an operational standpoint, visualisation systems play a vital role in climate studies management, serving as an interface between complex datasets and end users. As highlighted by WMO, real-time nowcasting necessitates rapidly updated, high-resolution observations that must be presented within integrated platforms capable of handling heterogeneous data streams [30]. The statement remains valid when being extended beyond nowcasting to a broader context of environmental monitoring and forecasting. Whether the data is derived from ground-based instruments, remote sensing technologies or NWPs, such visualisation systems allow experts to assess environmental conditions and forecasts

through a unified interface. However, specific challenges arise in implementing such platforms—particularly when integrating real-time data from both low-cost and regulatory monitoring systems, aligning asynchronous data formats and update frequencies, and visualising predictions from learning-based models in ways that are both scientifically valid and publicly interpretable.

The primary objective of our development is to provide a visualisation tool that operates on the Science Data and Compute (SDC) facility managed by the NSW-DCCEEW. This tool is capable of seamlessly accessing forecasts resulting from multiple predictive models, encompassing both numerical models utilised by the Australian government and extensive DL algorithms. Hosted in a high-performance computing environment, the back end of the tool is designed as an intermediate interface with the database of DL forecasts, providing the most current updates on learning-based forecasting information to all stakeholders and users [82]. Moreover, the server side is tasked with dynamically collecting real-time observations from PurpleAir LWSN and regulatory monitoring stations maintained by the local EPA.

All data processing, such as retrieving, aggregating, and processing forecast and monitoring data, occurs in the back end, leveraging the high-performance server to ensure efficient handling of large datasets. This architecture is designed to facilitate a cohesive system that supports both monitoring and forecasting needs by offloading heavy computational tasks from the front end, preventing any potential performance bottlenecks on user devices. By relieving the front end of resource-intensive processing, the system allows the user-facing interface to focus solely on rendering a smooth, responsive experience. A robust pipeline between the back end and front end ensures seamless data transfer and interaction, guaranteeing that the front end remains lightweight and optimised for user interaction while still delivering real-time, complex environmental data visualisations.

The front-end design of the dashboard emphasises user-friendliness and interactivity, enabling a wide range of end users to advantageously and meaningfully explore air pollution data across the state. Users are offered the flexibility to customise various options such as the regions of interest, time scopes of forecast, air pollutant types, and DL models used for visualisation. Through graphical representations and the degree of freedom, the dashboard shows engaging experiences that leverage it from a tool for illustrating data repositories to a platform for exploration and decision-making. This integrated solution aims to provide stakeholders with a holistic yet explicit perspective on the spatial-temporal disperses of air pollutants with accuracy and reliability. Such insights into the ambient conditions can contribute to raising environmental awareness, providing input to microclimate management, enhancing the ability to make informed decisions, analysing meteorological patterns, supporting policy formulation, and facilitating proactive climate responsiveness.

The chapter is organised as follows. After the introduction, Section 5.2 describes the

environmental dataset, including multi-scale monitoring data from [20, 38, 39, 104] and learning-based predictions from [3, 4, 5] and the submitted *Dempster-Shafer ensemble learning framework for air pollution nowcasting* paper. The architecture of the visualisation platform along with the functionalities and merits of our dashboard and plans for further development are elaborated in Sections 5.3 and 5.4 respectively, which is the original contribution in the published paper [82]. Finally, we draw the conclusion in Section 5.5.

5.2 Multi-scale monitoring and prediction

5.2.1 State-run air quality stations

The AQMS network is operated and maintained by the NSW DCCEEW. Their purpose is to provide timely and accurate environmental characteristics on a vast scale to all residents. Thus, these state-run stations are strategically positioned to capture the most representative ambient conditions of specific regions [105]. Each station hosts sophisticated scientific instruments for gathering air-related information continuously, as shown in Fig. 3.4. Accordingly, the concentration levels of key air pollutants, along with auxiliary parameters, are recorded. Table 5.1 shows the main air pollutants observed by DCCEEW-operated AQMS. The collected data provide inputs to models for estimating regional and global air circulation patterns. The knowledge significantly helps assess the impact of long-range transport of pollutants.

Table 5.1: Main observed pollutants by NSW-DCCEEW [106]

Air pollutants	Symbols	Units	Standard thresholds (daily)
Ozone	O3	pphm	6.5
Nitrogen dioxide	NO2	pphm	8
Carbon monoxide	CO	ppm	9.0
Sulfur dioxide	SO2	pphm	10
Particulate matter	$PM_{2.5}$	$\mu g/m^3$	25
	PM_{10}		50
Visibility	Bsp	$10^{-4}m^{-1}$	3

5.2.2 Low-cost air quality sensor network

Recent trends in sustainability and smart cities are characterised by deploying a thorough monitoring network over the metropolitan and sprawling suburbs. Many studies have proved that microclimate conditions are formed differently in various parts of an area, which are not adequately reflected due to the sparsity of stations [107, 108, 109]. Therefore, localised monitoring is promising to be used along with station observations

in forecasting the dynamic dispersion of air pollution in urban areas, filling in the gaps between stations.

Central districts of Sydney and surrounding suburbs have a substantial grid of PASs [20]. This network comprises devices commissioned by DCCEEW and is open for further extension. They are low-cost, off-the-shelf sensors designed to measure ambient conditions and air pollutant levels, as displayed in the bottom left of Fig. 4.2. All models of PASs are equipped with dual laser counters, ambient condition sensors and IoT modules as given in Table 5.2.

Table 5.2: Specifications of PurpleAir sensors [83, 110]

PurpleAir sensor	Classic	Flex
Dual Laser Particle Counters	PMS-5003	PMS-6003
Pressure, Temperature, Humidity & Gas Sensor	BME280	BME688
Built-in SD card data logger	Available	
WiFi module	Available	

The placement of the sensors complements the sparsely located AQMS to increase the spatial continuity in the input data. A finer resolution of inputs is then sent to the data-driven model to increase its accuracy.

5.2.3 Predictive air quality data from multiple sources

In NSW, the government regularly makes predictions of air pollution through complex physical-chemical-based models, including CTM and CMAQ in fusion with meteorological modelling CCAM and WRF (Weather Research and Forecasting model) to formulate hybrid models, namely CCAM-CTM, WRF-Chem and WRF-CMAQ [39, 111]. Each model has a different domain configuration, but they produce grids of predictions up to 72-hour future values to forecast the dispersion of air pollution over multiple scales from 3×3 km to 80×80 km [39]. These models provide users with different options for forecast information. Besides, we used future values from CCAM-CTM as an input to our recently developed deep learning model using LSTM-BNN to improve the forecast accuracy of air pollutant patterns [4].

The prediction originates from our recently developed hybrid LSTM-BNN model [4]. Notably, this model displays substantial improvements in forecast accuracy for air pollutants, such as ozone (O_3) and $PM_{2.5}$, compared with the government-managed model (i.e., CCAM-CTM). Here, the LSTM network is the central core for learning patterns of air pollution data with robustness from multiple memory-controlled mechanisms (i.e., input gate, forget gate, hidden states) inside each LSTM layer. To quantify the forecast uncertainty, dropout layers are inserted between LSTM layers to reduce overfitting and effectively approximate the Bayesian inference of neural network [112]. The configuration

of our model integrated into the dashboard is shown in Table 5.3. In our experiment, input and output layers share the same values of time steps that produce good performances of periodic patterns (e.g., diurnal features). We selected 12, 24, 48 and 72 values to match with time steps of predictions from the CCAM-CTM model as these future estimations enhance pattern recognition with our developed model [39, 4]. The Adaptive learning rate (Adam) optimisation and Rectified Linear Unit (ReLU) activation function are selected here given high nonlinearity in the air pollution data [113]. Besides, the output layer is attached with a linear activation function for generating forecast regression values [114].

Table 5.3: DL model configuration [4]

Configurations	Selected values	Notes
Input/Output layers	12, 24, 48, 72	No. of time steps
Hidden LSTM layers	3 layers	128 nodes per layer
Dropout layers	3 layers	Dropping rate: 0.2 - 0.4
Optimiser	Adam	Adaptive learning rate
Activation functions	ReLU function	Used in LSTM layers
	Linear function	Used in other layers

In this work, we also experimented with our model for local forecasts in a fusion of data from stations and a developed sensor network for monitoring air quality at an interested location of a suburb. The results displayed outperform other DL forecasting models, encompassing the GRU, CNN, as well as the hybrid CNN-LSTM model [3]. These proofs assure the reliable performance of our DL forecasting model in a real-world operation for the developed dashboard.

5.3 Visualisation dashboard framework

5.3.1 Client side

In the web application, the GUI is the principal visualisation for user interaction. The client interface of this web-based dashboard prioritises fast responses and intuitive experiences, focusing on delivering clear and actionable air quality insights.

Here, core front-end web development technologies are employed, including HTML (Hyper-Text Markup Language), CSS (Cascading Style Sheets) and JavaScript, to ensure cross-platform compatibility and accessibility. Additionally, the dashboard applies Shadcn UI for customisable UI components, built on Tailwind CSS, offering flexibility in designing user interfaces that are both visually appealing and responsive. Tailwind's utility-first approach accelerates the development process, while Shadcn UI provides a streamlined and lightweight way to create reusable, consistent interface elements.

Utilising React.js, one of the most popular JavaScript frameworks, allowed us to leverage its component-based approach and extensive functionalities, promoting modularisa-

tion, reusability and scalability. This choice simplifies continuing feature extensions and future maintenance as evolving demands for new functionalities are incorporated into the dashboard. Given the need to spatially illustrate time-series ambient information across the state of NSW, the tool features the integration of web-mapping and data visualisation libraries. Leaflet.js, a trusted GIS framework by various meteorological bureaus, is integrated into our platform to facilitate user-friendly interactions within our geographical area of interest. In parallel, the Chart.js library is complementarily run with Leaflet.js, ensuring comprehensive visualisation in both spatial and temporal aspects of air quality observations and DL forecasts. Furthermore, Zustand is utilised for efficient and minimalistic state management, making the handling of global states across the application both lightweight and scalable. The client side is structured with a well-organised folder hierarchy, as shown in Fig. 5.1. Each folder serves a specific role, ensuring a clean separation of functionalities and facilitating smooth collaboration and future development. Specifically, the main directories contain:

- assets: houses static files such as CSS styling, images, and icons.

- configurations: holds configuration files for managing settings such as environment variables.

- components: contains various sections of the dashboard, the banner, the footer, auxiliary areas and the User Interface (UI) folder with definitions and styling of graphic elements.

- stores: manage state variables across components.

- services: handle Application Programming Interface (API) calls using Axios, enabling communication between the client and server.

- utils: includes helper functions used across the project.

- others: cover miscellaneous package configuration files.

To ensure the interface remains user-centric, its design is developed in consultation with atmospheric scientists from DCCEEW and non-expert representatives. The goal is to reduce complexity while providing large amounts of useful information in an accessible manner. To this end, the incorporation of DL models is restricted to back-end operations of the dashboard to abstract all heavy data processing.

5.3.2 Server side

The server is the brain of a web application, responsible for orchestrating data flow between various components. In the air quality forecast dashboard, the server acts as an
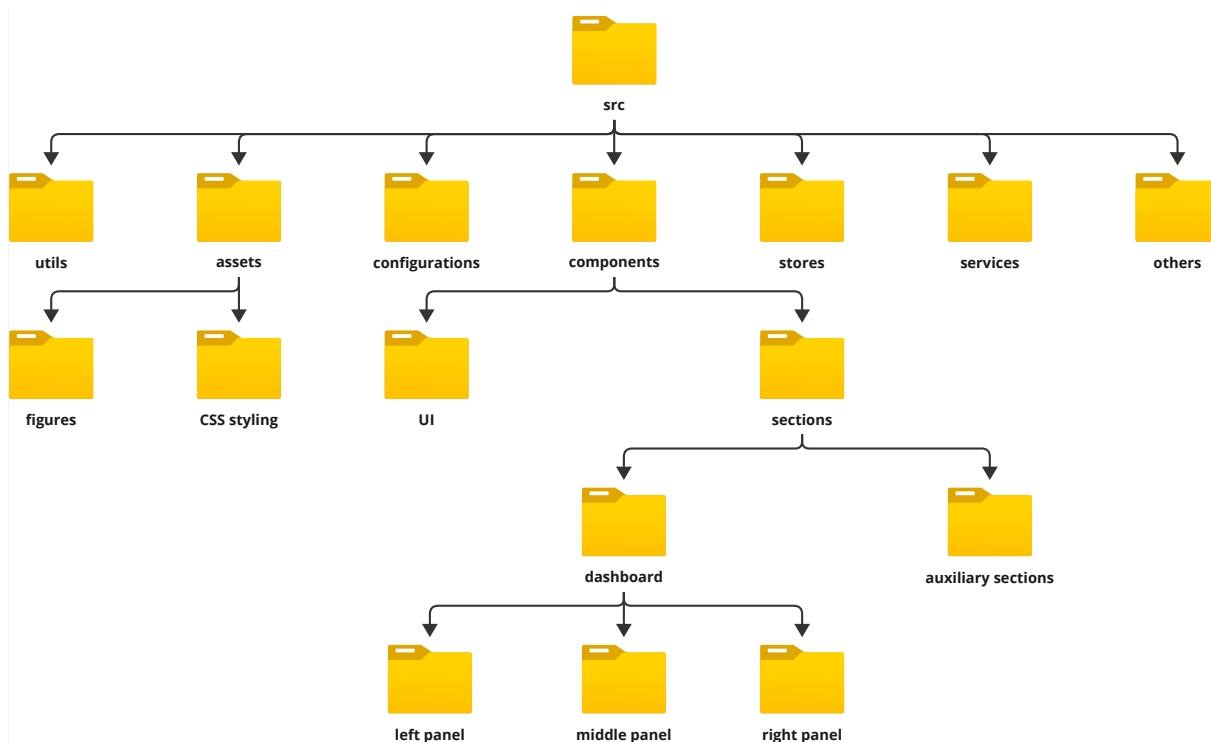
Figure 5.1: Client-side hierarchical development structure

intermediary, handling the retrieval and processing of DL forecasts from the SDC facility and transporting them to the client-side interface.

To ensure cross-platform development within our integrated framework, we employ Node.js, a JavaScript runtime environment, as a skeletal foundation to build the back-end structure. Here, Node.js offers scalability, high performance, and efficiency in handling a multitude of asynchronous requests. Built on top of Node.js is the Express.js web application framework. This widely used tool is lightweight, unopinionated, and middleware-compatible, facilitating a rapid and robust development process. In our application, Express.js also provides a variety of Hypertext Transfer Protocol (HTTP) utility methods essential for RESTful API development and imperative for implementing a standardised data transmission protocol between the client and the server. To facilitate secure data exchanges, CORS (Cross-Origin Resource Sharing) is configured on the server, ensuring that the client can safely request resources from different domains, which is critical when dealing with data streams from multiple sources, including government databases and real-time sensors. Additionally, near real-time updates of environmental monitoring data are scheduled using cron jobs. These scheduled tasks allow the server to periodically fetch the latest air quality observations from both regulatory agencies and PASs, ensuring that the data presented to users are timely and accurate.

The folder structure of the server side, illustrated in Fig. 5.2, is designed to maintain a clean and modular approach:

- configurations: contains configuration files, such as environment variables, which

control server settings and behaviour across different environments (e.g., development, production)

- data: keeps the DL forecast files and other relevant datasets from the SDC facility for efficient access to model outputs and real-time data streams.

- routes: manages data routing logic and API endpoints that facilitate interaction between the client and server. The routing system processes incoming API requests and returns the appropriate data, making it a critical component of the back end.

- src: houses the `server.js` file, which serves as the entry point for the application. This file initialises the Express.js server, handles API requests from the client, and ensures that the server is running correctly.
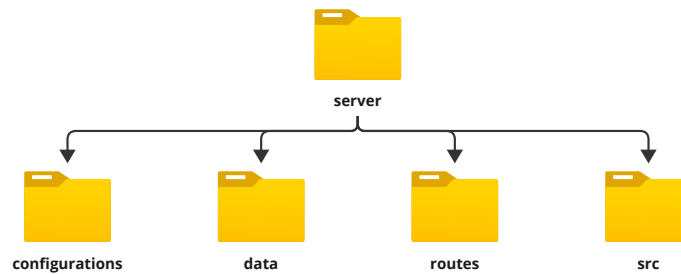
Figure 5.2: Server-side structure

The synergy between Node.js, Express.js, and supporting tools like CORS and node-cron presented here a suitable server-side design for a real-time and data-intensive application in our DL-based air quality forecast dashboard.

5.3.3 System integration

The integration between the client and server forms the complete structure of the visualisation tool. This section elaborates on the interactions between the front end, back end and data sources, ensuring seamless visualisation of real-time monitoring data and DL-based forecasts. Figure 5.3 demonstrates the high-level architecture of our implementation and the communication between key components.

The client side of the tool is a fusion of front-end technologies, which primarily handles data querying, performing logic under engaging functionalities and rendering visualisation. Meanwhile, the server handles requests and routes pertinent tasks in order to respond with desired information. Communication between them adheres to typical client-server protocol, where the front end sends HTTP requests and receives responses in Javascript Object Notation (JSON) format, facilitated by the isomorphic, promise-based Axios.js library.

Figure 5.3: Dashboard architecture.

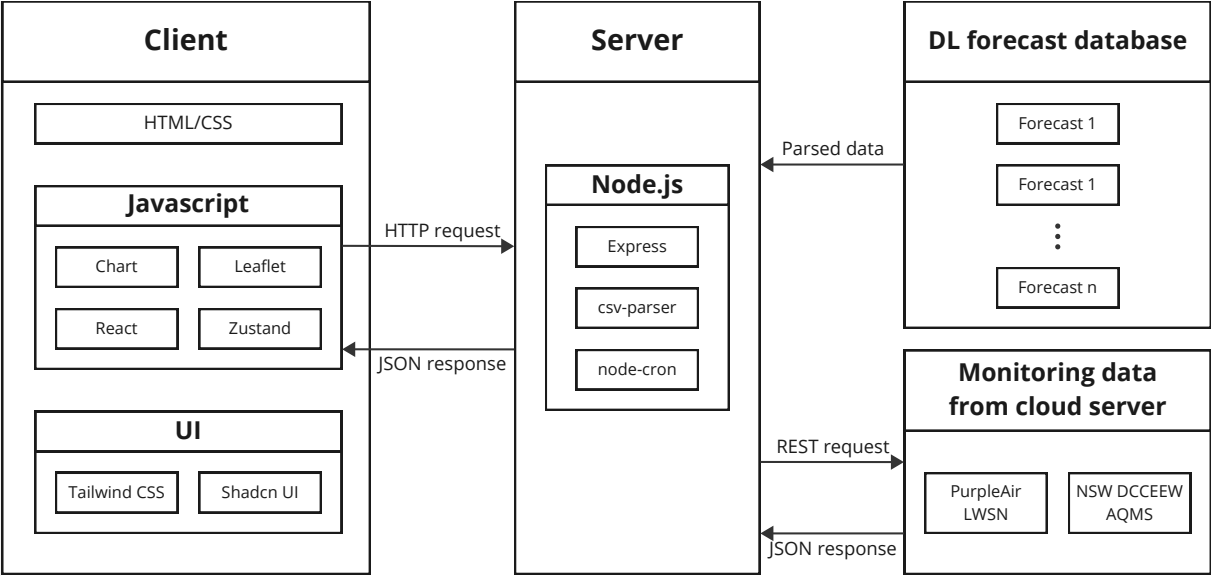Taking advantage of the SDC facility, the back end empowers Node.js and Express.js to retrieve monitoring data, manage the forecasting results, and preprocess them accordingly. As mentioned, all forecasts were performed offline with respect to the dashboard. After going through the QA/QC (quality assurance/quality control) steps of DCCEEW to meet their strict requirements, the approved forecasts were redirected to a storage location within their server as comma-separated values (CSV) files. Due to specific characteristics of environmental data and the operation of NSW-DCCEEW, they do not employ conventional database format (i.e. SQL,...) for the forecasts but rather a dedicated folder in their server, which is the `data` folder listed as a component of the back-end structure in Fig. 5.2. The forecast files are parsed using the csv-parser library, formatting the output appropriately in compliance with the client side.

On the other hand, the real-time monitoring data is fetched from online portals, including the PurpleAir cloud server and DCCEEW database, through RESTful API calls using privately owned keys and public access, respectively. Additionally, cron jobs are implemented using node-cron to schedule periodic retrievals of updated air quality data from external sources, ensuring that users receive up-to-date information in near real-time. This mechanism allows the dashboard to balance real-time responsiveness with efficiency, given the environmental data's continuously evolving nature. All technologies used in the development of the dashboard are tabulated in Table 5.4.

The distributed client-server architecture adopted in our dashboard design divides specialised assignments for the front end and back end, therefore assuring that the system can handle both computationally demanding jobs while maintaining the efficient delivery of real-time data and satisfied experiences with exploring forecasting information.

Table 5.4: Technologies for dashboard development.

Library	Description	Source
Axios.js	Library for HTTP requests	https://axios-http.com
Chart.js	Charting library	https://www.chartjs.org
csv-parser	CSV parsing package	https://www.npmjs.com/package/csv-parser
Express.js	Back-end framework	https://expressjs.com
Leaflet.js	Web mapping library	https://leafletjs.com
Node.js	JavaScript runtime environment	https://nodejs.org
node-cron	Task scheduler	https://www.npmjs.com/package/node-cron
React.js	Front-end framework	https://react.dev
Shadcn UI	Customisable UI component library	https://ui.shadcn.com
Tailwind CSS	Utility-first CSS framework	https://tailwindcss.com
Zustand	State management library	https://github.com/pmndrs/zustand

5.4 Visualisation of air quality monitoring and forecasting

5.4.1 Dashboard overview

The visualisation tool is designed with an intuitive, user-friendly interface that ensures ease of access to air quality monitoring and forecasting data. The dashboard is divided into distinct sections, as annotated in Figure 5.4, each serving a specific purpose to streamline navigation and interaction. Below is a breakdown of the primary components:

- Banner of the dashboard (Label 1): The topmost section contains a succinct description of the tool's purpose.

- Dashboard title (Label 2): This section includes the official logo of NSW-DCCEEW and the title, which also indicates its key service.

- Selection panel and map keys (Label 3): The panel on the left-hand side houses various controls for interacting with the dashboard, including the selection for either *Monitoring* or *Forecasting* mode, drop-down menus for customisable selection and buttons for manipulating different map layers. Map keys for symbols displayed on the map are also included for users' interpretation.

- Interactive map with plots (Label 4): This central feature is an interactive map powered by Leaflet.js, displaying real-time air quality data from various monitoring stations across NSW. Users can click on individual markers to bring up detailed information, including real-time sensor readings, historical data, or forecasts, depending on the selected mode. The map maintains the focus to NSW. On the top left of the map is the utilities to adjust the zoom level of the map and to reset the zoom level to default. The bottom left is a scale and the bottom right is the attributions to Leaflet.

- Playback time slider (Label 5): The playback slider is used to visualise air quality trends over time. Users can slide through different timestamps to see forecast predictions, providing a dynamic temporal exploration of air quality data.

- Air quality ranking panel (Label 6): This section displays a ranked list of stations based on current air quality metrics. It provides users with a quick view of the worst and best air quality in real time across NSW.

- Air quality category legend (Label 7): The legend defines the colour-coding scheme used in the map and ranking panel based on official air pollutant categorised by NSW-DCCEEW (see A.1).

- Auxiliary section (Label 8): Users can find additional resources, including a statistics panel that offers numerical summaries and comparisons of air quality forecasting data in terms of performance metrics, as well as activity guidelines for users for different concentration levels of air pollutants (see B.1).

- Footer (Label 9): The footer contains relevant links and additional information about the tool.
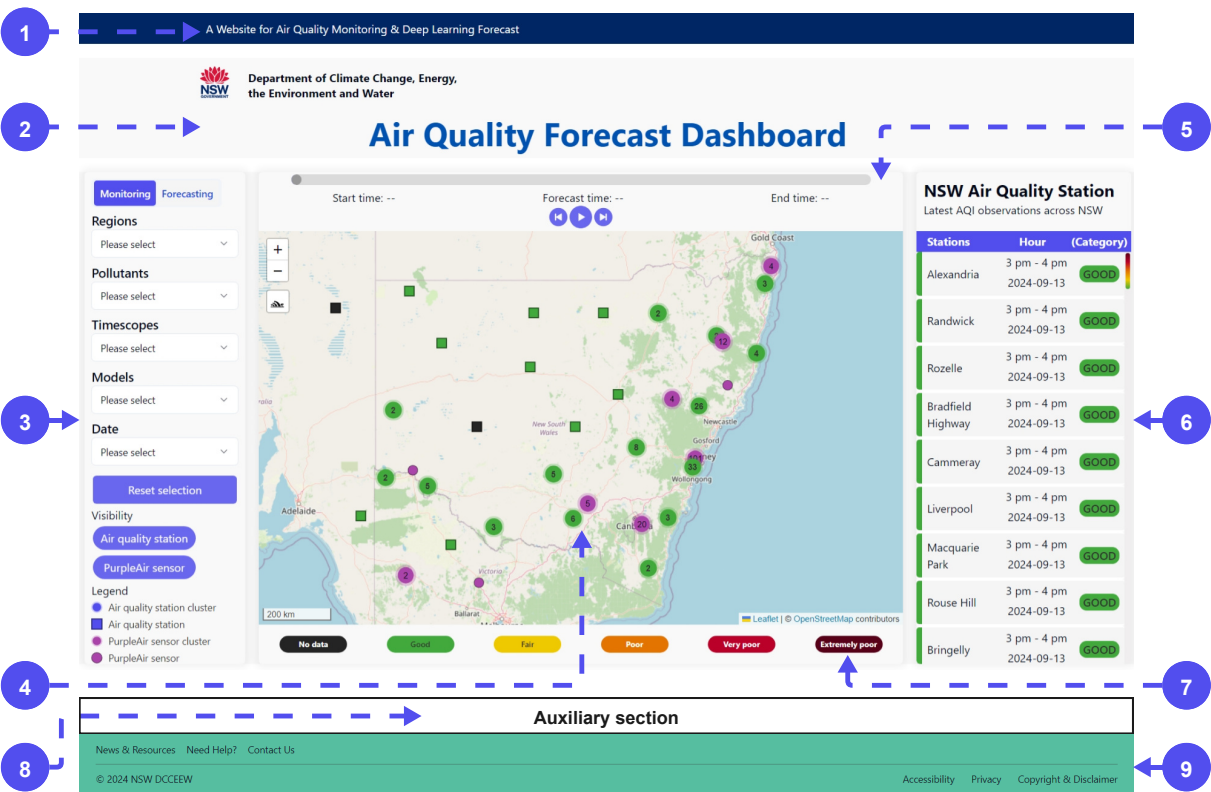

Figure 5.4: Interface of the visualisation tool with main sections.

By following this structure, the dashboard ensures that users can seamlessly interact with the air quality data, whether for real-time monitoring or detailed forecasting. The

combination of map visualisations, real-time plots, and accessible selection tools provides an engaging and informative user experience.

5.4.2 Air quality monitoring

Our proposed dashboard integrates environmental data from multiple sources. A database is established to get environmental information from data sharing services of DCCEEW [115]. This is pivotal groundwork for benchmarking and evaluating the reliability of LCSs. At the local scale, we retrieve real-time data from the cloud storage of the sensors. This is achieved with API calls, a standardised process for data exchange between various platforms, and as such, enabling the coordination between the dashboard and external data providers.

Regulatory monitoring station

The default display of the dashboard presents the *Monitoring* mode for the multi-scale air quality monitoring network across NSW, as depicted in Fig. 5.4. A total of 117 reference-grade air quality monitoring stations are reported in this network, strategically located across diverse geographical regions, capturing a range of air quality conditions in coastal, desert, urban, and suburban environments.

The map visualisation initially highlights the most recent AQI status for each AQMS. As indicated by the map key located in the left-hand panel, each station is represented by a square marker with a black outline, and their background colour reflects the AQI level recorded at each site. Stations that fall within a radius of 70 kilometres from one another are grouped into clusters when the map is viewed at higher zoom levels. These clusters are represented by circles with opaque outlines, where the background colour corresponds to the station within the cluster recording the highest AQI. The number of AQMS within each cluster is displayed in the centre of the circle. Users can navigate the map to explore AQI values across different regions or focus on specific locations by zooming and panning. Hovering over a station marker displays the station's name.

On the dashboard's right-hand panel, AQI observations from all stations are ranked in descending order and displayed in tabular format. This table includes station names, the time of data retrieval, and AQI categories, allowing users to quickly identify potential pollution hotspots across the state. A colour-coded legend in the map key helps users interpret AQI categories of the air pollutants used in our dashboard more easily (see Table 5.5).

LWSN monitoring

While the default map focuses on regulatory AQMS, the network of PurpleAir LWSN, which operates within the state's boundaries, is initially hidden. Users can reveal these

sensors by toggling the *PurpleAir sensor* visibility control. PASs are depicted as circles with a purple background and a black outline. Clustering for these LCSs follows the same logic as for the AQMS network, helping to reduce map clutter when zoomed out.

PASs provide real-time ambient air quality data, offering insights into localised pollution levels in specific suburbs of NSW. When a user selects a PAS marker, detailed environmental data is displayed in a sidebar, as shown in Fig. 5.5. The sidebar header presents the name and geographic coordinates of the sensor's deployment site, with a background colour corresponding to the measured $PM_{2.5}$ concentration. The body of the sidebar includes measurements of $PM_{1.0}$, $PM_{2.5}$ and $PM_{10.0}$, as well as temperature and humidity recordings. Data is retrieved via the PurpleAir cloud database through a licensed API key, ensuring the dashboard remains up-to-date with the latest air quality readings.
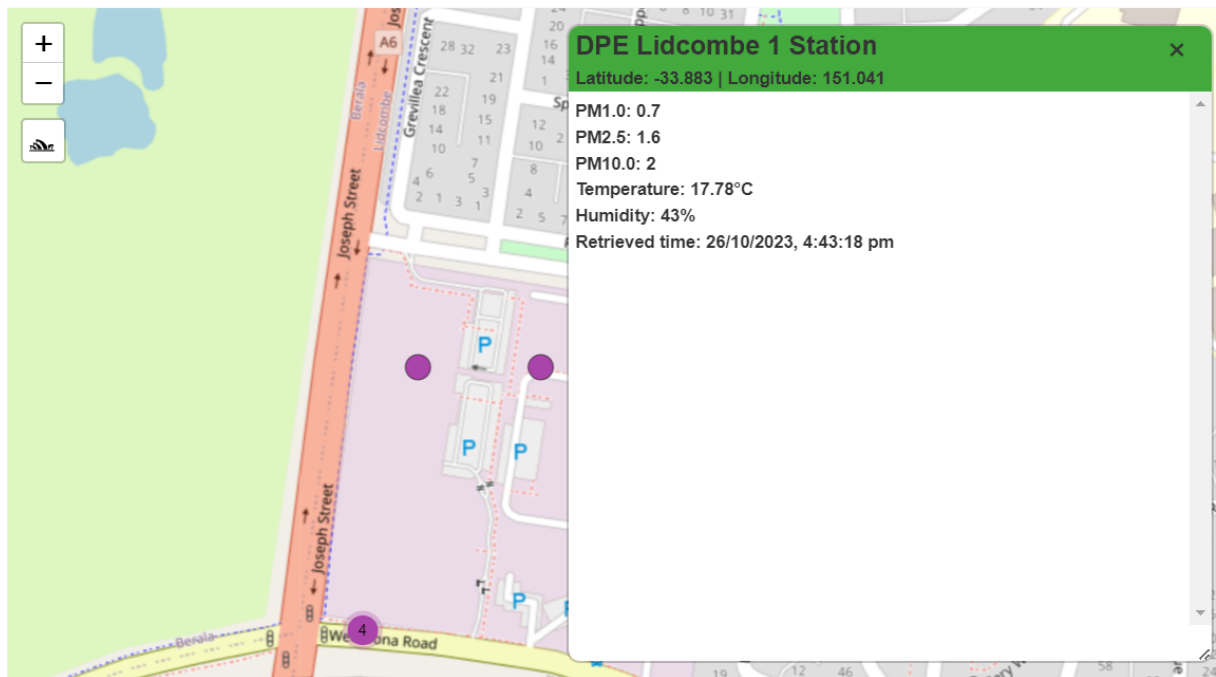
Figure 5.5: Real-time information from a PAS.

5.4.3 Air quality forecasting

Selection instructions for air pollutant forecast

The left panel of the dashboard is designed to intuitively engage end-users in overall dashboard control. Five selections in the form of drop-down lists are prompted in the GUI. The options follow the approach of our DL forecasting framework. As portrayed in Fig. 5.6, users can choose to view a forecast made from a combination of five parameters, i.e., region, pollutant, time scope, model and date.

Diverse climatic zones within NSW require our predictive model to be tailored for each region. The focal point of the model is to examine the complex topography of the Sydney Basin from 246 km to 384 km easting and from 6207 km to 6305 km northing. By offering a list of recognisable geographic areas, the first selection enables users to explore forecasts across the Basin. Some regions are characterised by a high-density population or heavy industrialisation, where different microclimatic disperses and excessive emissions originate. Knowledge of air quality in those areas is in high demand by the general public and state authorities. The forecasting process can employ a range of models, including DL-based ones for forecasting and ensemble learning-based framework for nowcasting, each adapted to the uniqueness of a region. Depending on the region selection, a model automatically reveals itself to end-users in the drop-down list. In parallel with spatial selection, the temporal aspect is one of the key factors of the dashboard. The dashboard has access to air pollutant forecasts for various time scopes, which users can select to view 3, 6, 12, 24, 48 or 72 hours in advance. In our study, the air pollutants of concern are O_3 and $PM_{2.5}$. They are known as the causes of severe respiratory illnesses. Having insights into their concentrations in the future enables appropriate responsive measures from all stakeholders, e.g., the government, to disseminate precautions and the public to adjust outdoor exposure activities. The last drop-down list offers the start date that a forecast is produced, which is beneficial for forecasters to make comparisons between temporally different outputs from the same forecasting model. A forecast file stored in the database that matches the chosen combination is loaded into the front end. Dynamic changes are applied all over the GUI to render requested information to users visually. With a simple click on the reset button, the dashboard removes the selected combination, and users can renew the dashboard exploration by repeating the steps above. The available forecast file that matches the chosen combination is loaded into the interactive map from our database.

Figure 5.6: Selection panel

Exploring visualisations on the interactive map

The dashboard incorporates an interactive map embedded with user-friendly functionalities. After users confirm their selection, the map zooms into the chosen region. Only stations located inside the geographical boundary remain visible.

Upon selecting a station marker, the dashboard renders a sidebar to the right of

the map, where the forecast for that station is plotted. The sidebar header displays the name and coordinates of the station, providing additional location information to users. Moreover, its background is colour-coded to match the air pollutant monitored or predicted for that station. This header decoration quickly aids users in assessing ambient status.

(a) 6-hour nowcast for AQMS

(b) 12-hour forecast for PAS

(c) 24-hour forecasts for AQMS

(d) 48-hour forecasts for AQMS
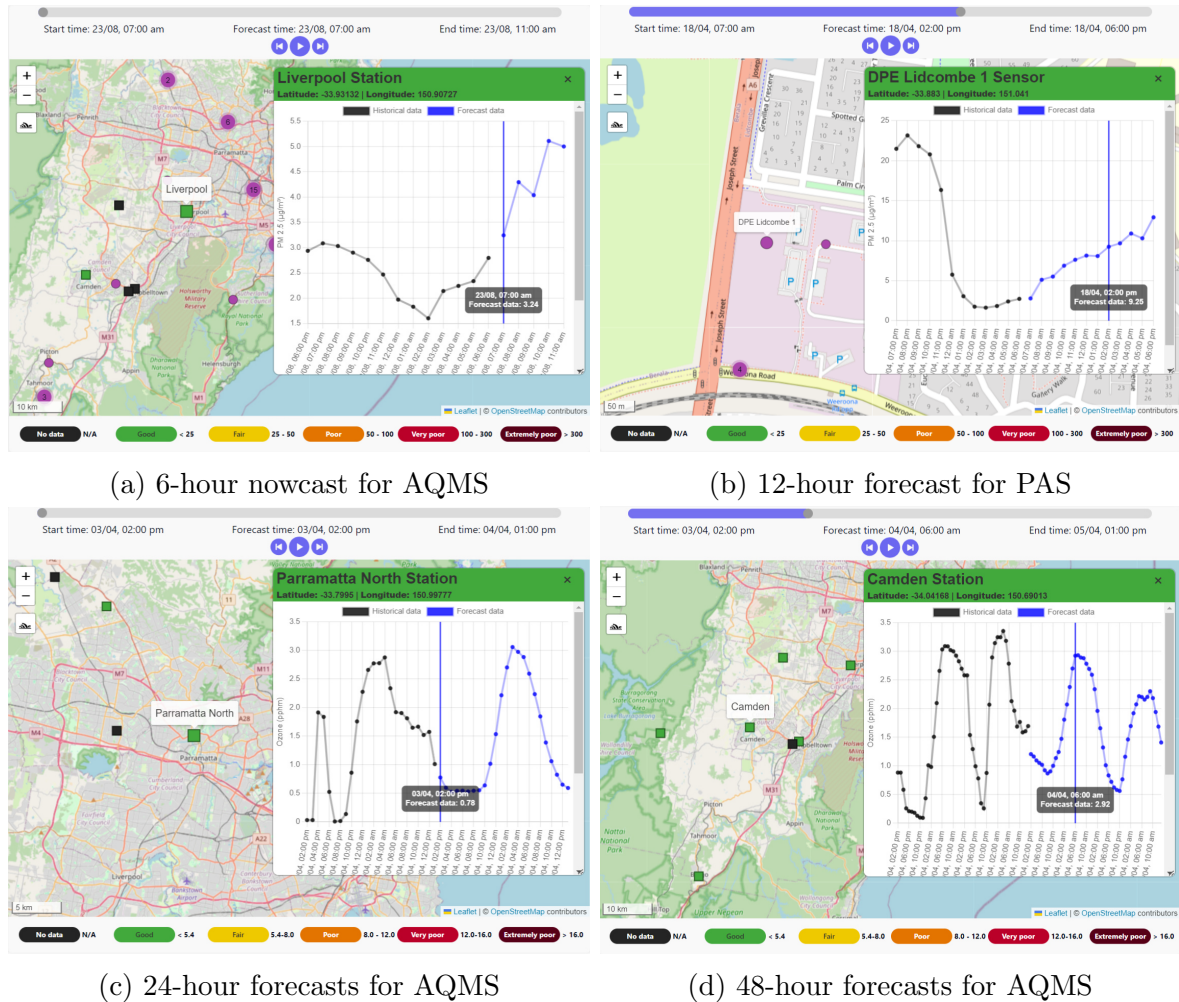
Figure 5.7: Visualisations of air pollution prediction in multi-scale monitoring network.

The sidebar primarily features the graphical format of the selected time-series forecast. The timeline of the graph spans from the historical inputs of air pollutant values into the DL model to the selected prediction time. The vertical axis marks the magnitude of air pollutant concentrations. Adhering to conventions in meteorology, we present the distinction between past and predicted data in appropriate colours. Figure 5.7 illustrates a series of DL-based forecasts as examples for the visualisation of different selection combinations from users. Also, the performance metrics for the forecast of the selected station are listed below the forecast graph.

To enhance user experience, the time slider above the map supports automatic transition for viewing the time series data. Specifically, a line indicator in the graph will step

through each timestamp in the forecast. The legend below the map indicates the pollutant categorisation complying with NSW-DCCEEW standards. Categories of O_3 and $PM_{2.5}$ concentrations and their respective colour codes are shown in Table 5.5.

Table 5.5: Air pollutant categorisation by NSW Government [103].

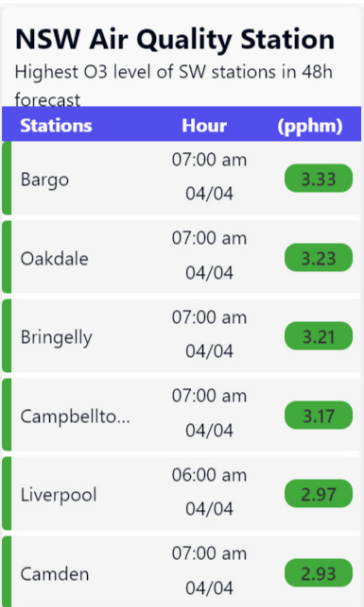| Category	colour code	Ozone (pphm)	$PM_{2.5}$ $(\mu g/m^3)$
Good	#42a93c (■)	< 5.4	< 25
Fair	#eec900 (■)	5.4 - 8.0	25 - 50
Poor	#e47400 (■)	8.0 - 12.0	50 - 100
Very poor	#ba0029 (■)	12.0 - 16.0	100 - 300
Extremely poor	#590019 (■)	> 16.0	> 300

Supporting panels

Figure 5.8: Ranking panel in forecasting mode.

The ranking panel, shown in Fig. 5.8, updates dynamically to reflect user selections, displaying stations in descending order of forecasted air pollutant concentrations. This helps users quickly identify suburbs within their selected region that are expected to have the highest pollution levels in the prediction horizon.

To assess model performance, a variety of evaluation metrics—including MAE, RMSE, Pearson's correlation coefficient, R^2, Mean of Forecast (MEAN_FC), Mean of Observations (MEAN_OBS), Standard Deviation of Forecast (STD_FC), Standard Deviation of Observations (STD_OBS), and Mean Bias Error (MBE)—are calculated and displayed in the dashboard. For example, the evaluation metrics of an LSTM-BNN-based forecast for all six stations within the Sydney Southwest region are tabulated below the map as in Fig. 5.9. They are significant tools, providing weather forecasters a quick assessment of the forecast's performance and offering scientists data for in-depth analysis. By publishing these numbers to all users, the dashboard emphasises research integrity and upholds responsibility in learning-based application practices.

Furthermore, the dashboard relays the activity guidelines for sensitive groups and everyone else, respective to a particular air quality category (see Table B.1), as officially announced in [116]. This feature facilitates each individual in making well-informed decisions about their outdoor schedules. Thus, it reiterates the commitment to promoting public health and, more generally, to climate responsiveness at both global and local scales.

Statistics

Station	MAE	RMSE	PEARSON_R	R_SQUARE	MEAN_FC	MEAN_OBS	STD_FC	STD_OBS	MBE
BARGO	0.6663	0.7936	0.7329	-0.4450	1.8785	1.8968	1.0497	0.7000	1.3828
BRINGELLY	0.5766	0.7304	0.8031	0.4005	1.5216	1.4270	1.1325	0.9771	1.4895
CAMDEN	0.6686	0.8156	0.6942	-0.0437	1.0005	1.4845	0.7077	0.8369	0.5991
CAMPBELLTOWN_WEST	0.6090	0.7700	0.7859	0.3233	1.4009	1.3306	1.1276	0.9742	1.6638
LIVERPOOL	0.5970	0.7588	0.7789	0.3179	1.3504	1.2190	1.1008	0.9632	1.6339
OAKDALE	0.3110	0.3790	0.6219	-0.9534	2.3478	2.4470	0.4296	0.2920	0.5737

Figure 5.9: Statistics panel.

5.4.4 Further Development

For further development, there are possibilities for scaling up the dashboard. Heat maps can be generated from the predicted air pollutant concentrations and overlaid over NSW regions. Animation features of the time slider then render temporal changes of heat maps on demand. The inclusion of methods for reliability assessment of LCSs and evaluation criteria is another milestone in moving forward. Near real-time data collected from neighbouring stations will serve as reference points in the validation algorithm. By reliably increasing the spatial resolution with LCS data, the accuracy of the forecasting model can be enhanced. The visualisation tool can also be extended to deliver more scientific information, specialised standards, and metrics in climatology.

5.5 Conclusion

In this chapter, we have developed a visualisation tool for multi-scale air quality monitoring and DL-based air pollutant forecasts. It is designed to be user-friendly by utilising widely adopted web technologies. At the global scale, the complex forecasting process for state-run stations is abstracted to an intuitive selection flow. We also integrated the learning models along with the LWSN into the developed platform as an integrated system. The interactive map and graphical data representation enable users to comprehend ambient forecasts. At the local scale, real-time information streaming from the LWSN provides fine-grained ambient updates. With a seamless transition between global and local contexts, a wide range of end users can gain a holistic view and new perspective on the environment across the state of NSW. Supported by scientifically validated monitoring methods and learning-based forecasting models, this platform is promising for applications in microclimatic studies and the management of smart cities.

Chapter 6

Conclusion

6.1 Thesis summary

The work conducted in this thesis endeavours to enhance the reliability of air quality monitoring and improve the accuracy of predicting air quality in urban settings with the end goal of promoting the accessibility to environmental data for multiple stakeholders.

Localised environmental dynamics exhibit divergent patterns and progression compared to mesoscale or synoptic meteorological phenomena. The proliferation of IoT technologies has made it feasible and affordable to deploy dense networks of LCS to suburban areas, accumulating necessary data to further enrich insights about microscale meteorology. However, the quality of data collected by these instruments is often questioned due to inherent limitations in their hardware and the volatile working conditions.

This thesis proposes a dependable air quality monitoring system using cost-efficient sensors, where the quality of output data is ensured through both hardware redundancy and a DS-based decision-making algorithm. Controlled laboratory experiments were first set up to test the feasibility of the proposed methodology. After positive results, the DDSI framework was applied on a real-world monitoring site using a tri-sensor system, benchmarked against a nearby standard monitoring station. The output of this practical implementation is a continuous stream of monitoring data, where observational inaccuracies caused by internal faults or non-ambient factors are mitigated by the reliability assessment logic assigning the most reliable sensing mote as an on-duty sensor.

Following the augmentation of monitoring data quality, the thesis takes a further step by routing the localised, quality-enhanced observations to short-term air quality forecasts, or nowcasting. The intrinsic stochastic nature of local air pollutant dispersion is noticeable in those datasets. To address this issue, an ensemble learning model consisting of multiple DL models was employed to utilise the unique capabilities of each learner in capturing specific PM patterns in urban environments. At the decision-making layer of the proposed DSEL is a DS-based fusion, which attempts to balance the indication of

multiple performance metrics between the forecasts and observations to reach an optimal forecast.

Then, a web-based dashboard is developed as a result of the collaboration between researchers at UTS and scientists at NSW-DCCEEW. This tool employs a client-server architecture to effectively distribute specialised tasks between two sides. The back end handles the retrieval and preprocessing of real-time monitoring data from various sources and DL-based air pollutant forecasts, while the front end concentrates on ensuring smooth, interactive engagement with a wide range of pertinent users and rendering comprehensible visualisations based on their requests.

Furthermore, my research endeavours extend to formulating a digital twin package for UAV simulation. This platform can serve as a sturdy groundwork to conduct virtual, preliminary tests of UAVs equipped with instruments measuring ambient conditions, a highly dynamic approach to real-time environmental data collection and monitoring.

6.2 Conclusions

The proposed DDSI framework enhances the reliability of LCS data by adopting dependable strategies through colocation deployment and a reliability assessment algorithm to determine the on-duty sensor. The promising results from laboratory tests were strong precursors for the high statistical alignment between the output from the DDSI framework and benchmarked reference-grade station data. The positive outcomes from real-world implementation encourage the propagation of this method to other locations. This two-layer approach to data collection using low-cost sensing systems underscores the importance of high-quality data for further analysis. However, it should be noted that the current validation is limited to specific sites in Sydney, and further testing in diverse urban environments is necessary to confirm broader applicability.

The development of the DSEL demonstrates how reliable monitoring data from dependable sensing systems, as well as standard regulatory data, can be used to generate short-term air quality forecasts. The DSEL model's ability to outperform conventional ensemble models through its multi-criteria decision-making process highlights the importance of diversifying approaches in environmental forecasting.

In the face of the unpredictable and volatile nature of urban environments, this thesis reaffirms the notion that no single learning-based model can be universally applied to air quality forecasting. The ensemble approach, which leverages the combined strengths of multiple learners, proves to be the most effective method. This thesis emphasises the value of diversification in learning-based methods, especially in the context of air quality forecasting, which is critical for addressing issues of sustainability, resilience, and climate preparedness in urban environments.

The DSEL was developed to learn the features of the quality-enhanced data from the

dependable sensing system, as well as standard regulatory data, to produce short-term predictions of air pollution. With the DS-based data fusion scheme for a multi-criteria decision-making process, DSEL can outperform conventional ensemble models and output the best-performing forecast. Despite the promising results, the generalisability of the framework to other regions and pollutant types remains an important area for future research.

The visualisation dashboard developed in this thesis aims to align with the principles of translational research, which seeks to convert scientific findings into actual applications that directly benefit humans. The visualisation dashboard is intended as a medium for conveying the outcomes from this thesis to a wide range of audiences. The web-based platform delivers up-to-date spatio-temporal ambient observations from both LWSNs and state-run AQMSs, while its forecasting mode swiftly visualises the time-series forecasts in an engaging and intuitive manner and delivers associated health advice for the general public and provides performance indexes of the forecasts for further scientific evaluation. This ensures the intention of developing the visualisation tool as a medium for conveying the meaningful outcomes from this thesis to a broad range of audiences, raising public environmental awareness and advocating research integrity and practical applicability. The current deployment is internal to DCCEEW, so beta testing and stakeholder feedback would further validate its impact and pave the way for public access.

In conclusion, this thesis presents a comprehensive pipeline, starting from the collection of reliable air quality data through a dependable monitoring system, moving to ensemble learning-based forecasting of air pollution with a multi-metrics fusion method for the selection of best-performing predictions, and culminating in the comprehensible visualisation of environmental data on a user-friendly dashboard.

6.3 Future works

The work presented in this thesis lays strong building blocks for future research. There are several avenues for possible improvement and extension that can be pursued to further advance the original objectives of this research.

- **Fusion of multiple data sources:** utilise diverse environmental data sources, ranging from satellite imagery of remote sensing, NWPs from models such as CMAQ, for a comprehensive reliability assessment of LWSN, and to achieve multi-scale air quality forecasting.

- **Enhancement of nowcasting performance:** concentrate on augmenting the performance and accuracy of ensemble learning framework DSEL by boosting individual learners and explore more advanced optimisation techniques and ensemble methods.

- **Broaden the scope of DSEL:** extend the framework to accommodate multivariate input sources and multiple key pollutants as output, implement the framework across the entire LWSN to holistically evaluate its plausibility and provide a scalable solution for air quality nowcasting in urban settings.

- **Visualisation tool improvements:** the GUI of the dashboard can benefit from a more intuitive and user-friendly appearance, while new functionalities can be introduced for real-time diagnosis and historical data analysis of low-cost ambient sensors' operational states.

- **Development of mobile monitoring platforms:** create adverse environmental scenarios in drone-based digital twin platforms to experiment and implement navigation strategies and control algorithms specifically for mobile monitoring, assessment and responses to event cases.

Bibliography

[1] World Health Organization (WHO), "Ambient (outdoor) air pollution," https:// www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-h ealth, accessed: Aug. 12, 2023.

[2] United Nations Human Settlements Programme (UN-Habitat), "World Cities Report 2022: Envisaging the Future of Cities," https://unhabitat.org/wcr, accessed: Oct. 8, 2023.

[3] H. A. D. Nguyen, T. H. Le, Q. P. Ha, and M. Azzi, "Deep learning for construction emission monitoring with low-cost sensor network," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 40. IAARC Publications, 2023, pp. 450–457. doi: 10.22260/ISARC2023/0061 .

[4] H. A. D. Nguyen, Q. P. Ha, H. Duc, M. Azzi, N. Jiang, X. Barthelemy, and M. Riley, "Long short-term memory Bayesian neural network for air pollution forecast," *IEEE Access*, 2023. doi: 10.1109/ACCESS.2023.3265725 .

[5] H. A. D. Nguyen, T. H. Le, M. Azzi, and Q. P. Ha, "Monitoring and estimation of urban emissions with low-cost sensor networks and deep learning," *Ecological Informatics*, vol. 82, p. 102750, 2024. doi: 10.1016/j.ecoinf.2024.102750 .

[6] World Meteorological Organization (WMO), "Integrating Low-Cost Sensor Systems and Networks to Enhance Air Quality Applications," https://wmo.int/publicatio n-series/integrating-low-cost-sensor-systems-and-networks-enhance-air-quality-a pplications, acceesed: Aug. 9, 2024.

[7] United Nations Environment Programme, "Low-cost sensors can improve air quality monitoring and people's health," https://www.unep.org/news-and-stories/press-r elease/low-cost-sensors-can-improve-air-quality-monitoring-and-peoples, acceesed: Aug. 9, 2024.

[8] R. Jayaratne, X. Liu, K.-H. Ahn, A. Asumadu-Sakyi, G. Fisher, J. Gao, A. Mabon, M. Mazaheri, B. Mullins, M. Nyaku, Z. Ristovski, Y. Scorgie, P. Thai, M. Dunbabin, and L. Morawska, "Low-cost PM2.5 Sensors: An Assessment of their Suitability for

Various Applications," *Aerosol and Air Quality Research*, vol. 20, no. 3, pp. 520–532, 2020. doi: 10.4209/aaqr.2018.10.0390 .

[9] A. L. Clements, W. G. Griswold, A. Rs, J. E. Johnston, M. M. Herting, J. Thorson, A. Collier-Oxandale, and M. Hannigan, "Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary)," *Sensors*, vol. 17, no. 11, p. 2478, 2017. doi: 10.3390/s17112478 .

[10] L. Morawska, P. K. Thai, X. Liu, A. Asumadu-Sakyi, G. Ayoko, A. Bartonova, A. Bedini, F. Chai, B. Christensen, M. Dunbabin, J. Gao, G. S. W. Hagler, R. Jayaratne, P. Kumar, A. K. H. Lau, P. K. K. Louie, M. Mazaheri, Z. Ning, N. Motta, B. Mullins, M. M. Rahman, Z. Ristovski, M. Shafiei, D. Tjondronegoro, D. Westerdahl, and R. Williams, "Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone?" *Environment International*, vol. 116, pp. 286–299, 2018. doi: 10.1016/j.envint.2018.04.018 .

[11] X. Liu, R. Jayaratne, P. Thai, T. Kuhn, I. Zing, B. Christensen, R. Lamont, M. Dunbabin, S. Zhu, J. Gao, D. Wainwright, D. Neale, R. Kan, J. Kirkwood, and L. Morawska, "Low-Cost Sensors as an Alternative for Long-Term Air Quality Monitoring," *Environmental Research*, vol. 185, p. 109438, 2020. doi: 10.1016/j.envres.2020.109438 .

[12] M. Van Poppel, P. Schneider, J. Peters, S. Yatkin, M. Gerboles, C. Matheeussen, A. Bartonova, S. Davila, M. Signorini, M. Vogt, F. R. Dauge, J. S. Skaar, and R. Haugen, "SensEURCity: A Multi-City Air Quality Dataset Collected for 2020/2021 Using Open Low-Cost Sensor Systems," *Scientific data*, vol. 10, no. 1, p. 322, 2023. doi: 10.1038/s41597-023-02135-w (online) .

[13] N. Schulte, X. Li, J. K. Ghosh, P. M. Fine, and S. A. Epstein, "Responsive high-resolution air quality index mapping using model, regulatory monitor, and sensor data in real-time," *Environmental Research Letters*, vol. 15, no. 10, p. 1040a7, 2020. doi: 10.1088/1748-9326/abb62b .

[14] A. U. Raysoni, S. D. Pinakana, E. Mendez, D. Wladyka, K. Sepielak, and O. Temby, "A Review of Literature on the Usage of Low-Cost Sensors to Measure Particulate Matter," *Earth*, vol. 4, no. 1, pp. 168–186, 2023. doi: 10.3390/earth4010009 .

[15] Queensland University of Technology, "Establishing Advanced Networks for Air Quality Sensing and Analysis," International Laboratory for Air Quality & Health, https://research.qut.edu.au/ilaqh/projects/koala-sensors, Accessed: Oct. 31, 2024.

[16] T. Kuhn, R. Jayaratne, P. K. Thai, B. Christensen, X. Liu, M. Dunbabin, R. Lamont, I. Zing, D. Wainwright, C. Witte, D. Neale, and L. Morawska, "Air quality during and after the Commonwealth Games 2018 in Australia: Multiple benefits of monitoring," *Journal of Aerosol Science*, vol. 152, p. 105707, 2021. doi: 10.1016/j.jaerosci.2020.105707 .

[17] R. Jayaratne, T. Kuhn, B. Christensen, X. Liu, I. Zing, R. Lamont, M. Dunbabin, J. Maddox, G. Fisher, and L. Morawska, "Using a Network of Low-cost Particle Sensors to Assess the Impact of Ship Emissions on a Residential Community," *Aerosol and Air Quality Research*, vol. 20, no. 12, pp. 2754–2764, 2020. doi: 10.4209/aaqr.2020.06.0280 .

[18] R. Jayaratne, P. Thai, B. Christensen, X. Liu, I. Zing, R. Lamont, M. Dunbabin, L. Dawkins, L. Bertrand, and L. Morawska, "The effect of cold-start emissions on the diurnal variation of carbon monoxide concentration in a city centre," *Atmospheric Environment*, vol. 245, p. 118035, 2021. doi: 10.1016/j.atmosenv.2020.118035 .

[19] New South Wales Department of Planning and Environment (NSW-DPE), "How and why we monitor air pollution," https://www.environment.nsw.gov.au/topics/air/air-quality-basics/sampling-air-pollution, accessed: Aug. 12, 2023.

[20] PurpleAir, "Real-time Air Quality Map," https://map.purpleair.com, accessed: Aug. 9, 2024.

[21] I. Stavroulas, G. Grivas, P. Michalopoulos, E. Liakakou, A. Bougiatioti, P. Kalkavouras, K. M. Fameli, N. Hatzianastassiou, N. Mihalopoulos, and E. Gerasopoulos, "Field Evaluation of Low-Cost PM Sensors (Purple Air PA-II) Under Variable Urban Air Quality Conditions, in Greece," *Atmosphere*, vol. 11, no. 9, p. 926, 2020. doi: 10.3390/atmos11090926 .

[22] Z. Farooqui, J. Biswas, and J. Saha, "Long-Term Assessment of PurpleAir Low-Cost Sensor for PM2.5 in California, USA," *Pollutants*, vol. 3, no. 4, pp. 477–493, 2023. doi: 10.3390/pollutants3040033 .

[23] J. Tryner, C. L'Orange, J. Mehaffy, D. Miller-Lionberg, J. C. Hofstetter, A. Wilson, and J. Volckens, "Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers," *Atmospheric Environment*, vol. 220, p. 117067, 2020. doi: 10.1016/j.atmosenv.2019.117067 .

[24] K. Barkjohn, B. Gantt, and A. Clements, "Development and application of a United States wide correction for PM2.5 data collected with the PurpleAir sensor," *Atmos. Meas. Tech*, vol. 2020, pp. 1–34, 2020. doi: 10.5194/amt-14-4617-2021 .

[25] J. Bi, A. Wildani, H. H. Chang, and Y. Liu, "Incorporating Low-Cost Sensor Measurements into High-Resolution PM2. 5 Modeling at a Large Spatial Scale," *Environmental Science & Technology*, vol. 54, no. 4, pp. 2152–2162, 2020. doi: 10.1021/acs.est.9b06046 .

[26] S. Metia, H. A. D. Nguyen, and Q. P. Ha, "IoT-Enabled Wireless Sensor Networks for Air Pollution Monitoring with Extended Fractional-Order Kalman Filtering," *Sensors*, vol. 21, no. 6, p. 5313, 2021. doi: 10.3390/s21165313 .

[27] H. A. D. Nguyen and Q. P. Ha, "Wireless Sensor Network Dependable Monitoring for Urban Air Quality," *IEEE Access*, vol. 10, pp. 40 051–40 062, 2022. doi: 10.1109/ACCESS.2022.3166904 .

[28] M. A. Zaidan, Y. Xie, N. H. Motlagh, B.-S. Wang, W. Nie, P. Nurmi, S. Tarkoma, T. Petaja, A. Ding, and M. Kulmala, "Dense air quality sensor networks: Validation, analysis, and benefits," *IEEE Sensors Journal*, vol. 22, pp. 23 507–23 520, 2022. doi: 10.1109/JSEN.2022.3216071 .

[29] P. Ferrer-Cid, J. Barceló-Ordinas, and J. García-Vidal, "Volterra graph-based outlier detection for air pollution sensor networks," *IEEE Transactions on Network Science and Engineering*, vol. 9, pp. 2759–2771, 2022. doi: 10.1109/tnse.2022.3169220 .

[30] Y. Wang, E. Coning, A. Harou, W. Jacobs, P. Joe, L. Nikitina, R. Roberts, J. Wang, J. Wilson, A. Atencia, B. Bica, B. Brown, S. Goodmann, A. Kann, P. W. Li, I. Monterio, F. Schmid, A. Seed, and J. Sun, *Guidelines for Nowcasting Techniques*. WMO, 11 2017.

[31] A. I. Prados, G. Leptoukh, C. Lynnes, J. Johnson, H. Rui, A. Chen, and R. B. Husar, "Access, Visualization, and Interoperability of Air Quality Remote Sensing Data Sets via the Giovanni Online Tool," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 3, no. 3, pp. 359–370, 2010. doi: 10.1109/JSTARS.2010.2047940 .

[32] P. Chen, "Visualization of real-time monitoring datagraphic of urban environmental quality," *Eurasip Journal on Image and Video Processing*, vol. 2019, no. 1, pp. 1–9, 2019. doi: 10.1186/s13640-019-0443-6 .

[33] W. Lu, T. Ai, X. Zhang, and Y. He, "An interactive web mapping visualization of urban air quality monitoring data of China," *Atmosphere*, vol. 8, no. 8, p. 148, 2017. doi: 10.3390/atmos8080148 .

[34] M. Nurgazy, A. Zaslavsky, P. P. Jayaraman, S. Kubler, K. Mitra, and S. Saguna, "CAVisAP: Context-aware visualization of outdoor air pollution with IoT plat-

forms," in *2019 International Conference on High Performance Computing & Simulation (HPCS)*. IEEE, 2019, pp. 84–91. doi: 10.1109/HPCS48598.2019.9188167 .

[35] N. S. Mathews, S. Chimalakonda, and S. Jain, "Air: An augmented reality application for visualizing air pollution," in *2021 IEEE Visualization Conference (VIS)*. IEEE, 2021, pp. 146–150. doi: 10.1109/VIS49827.2021.9623287 .

[36] Umweltbundesamt, "For our environment," https://www.umweltbundesamt.de/en/data/air/air-data/air-quality, accessed: Aug. 12, 2023.

[37] Department for Environment, Food & Rural Affairs, "Air pollution forecast," https://uk-air.defra.gov.uk, accessed: Aug. 12, 2023.

[38] United States Environmental Protection Agency, "Community Multiscale Air Quality Modeling System," https://www.epa.gov/cmaq/cmaq-models-0, accessed Aug. 12, 2023.

[39] H. N. Duc, T. Trieu, Y. Scorgie, M. Cope, and M. Thatcher, "Air quality modelling of the Sydney region using CCAM-CTM," *Air Quality and Climate Change*, vol. 51, no. 1, pp. 29–33, 2017.

[40] WHO. National Air Quality Standards. https://worldhealthorg.shinyapps.io/AirQualityStandards. National Air Quality Standards. Acceesed: Aug. 1, 2024.

[41] M. Kutlar Joss, M. Eeftens, E. Gintowt, R. Kappeler, and N. Künzli, "Time to Harmonize National Ambient Air Quality Standards," *International Journal of Public Health*, vol. 62, no. 4, pp. 453–462, 2017. doi: 10.1007/s00038-017-0952-y .

[42] Copernicus Atmosphere Monitoring Service (CAMS), "Atmosphere Monitoring Service," https://atmosphere.copernicus.eu/, accessed: Aug. 18, 2024.

[43] C. Buontempo, S. N. Burgess, D. Dee, B. Pinty, J.-N. Thépaut, M. Rixen, S. Almond, D. Armstrong, A. Brookshaw, A. L. Alos, B. Bell, C. Bergeron, C. Cagnazzo, E. Comyn-Platt, E. Damasio-Da-Costa, A. Guillory, H. Hersbach, A. Horányi, J. Nicolas, A. Obregon, E. P. Ramos, B. Raoult, J. Muñoz-Sabater, A. Simmons, C. Soci, M. Suttie, F. Vamborg, J. Varndell, S. Vermoote, X. Yang, and J. G. Marcilla, "The Copernicus Climate Change Service: Climate Science in Action," *Bulletin of the American Meteorological Society*, vol. 103, no. 12, pp. E2669–E2687, 2022. doi: 10.1175/BAMS-D-21-0315.1 .

[44] J.-N. Thépaut, D. Dee, R. Engelen, and B. Pinty, "The Copernicus Programme and its Climate Change Service," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 1591–1593. doi: 10.1109/IGARSS.2018.8518067 .

[45] CAMS, "European air quality hourly forecast of regulated air pollutants," https://atmosphere.copernicus.eu/charts/packages/cams_air_quality/products/europe-air-quality-forecast-regulated, accessed: Aug. 17, 2024.

[46] V.-H. Peuch, R. Engelen, M. Rixen, D. Dee, J. Flemming, M. Suttie, M. Ades, A. Agustí-Panareda, C. Ananasso, E. Andersson, D. Armstrong, J. Barré, N. Bousserez, J. J. Dominguez, S. Garrigues, A. Inness, L. Jones, Z. Kipling, J. Letertre-Danczak, M. Parrington, M. Razinger, R. Ribas, S. Vermoote, X. Yang, A. Simmons, J. G. Marcilla, and J.-N. Thépaut, "The Copernicus Atmosphere Monitoring Service: From Research to Operations," *Bulletin of the American Meteorological Society*, vol. 103, no. 12, pp. E2650–E2668, 2022. doi: 10.1175/BAMS-D-21-0314.1 .

[47] European Environment Agency, "European Air Quality Index," https://www.eea.europa.eu/themes/air/air-quality-index, accessed: Aug. 17, 2024.

[48] National Aeronautics and Space Administration, "Worldview," https://worldview.earthdata.nasa.gov/, accessed: Aug. 30, 2024.

[49] IQAir, "Explore your Air Quality," https://www.iqair.com/, accessed: Aug. 28, 2024.

[50] ——, "IQAir AirVisual Platform," https://www.iqair.com/air-quality-monitors/airvisual-platform, accessed: Aug. 28, 2024.

[51] M. Korunoski, B. R. Stojkoska, and K. Trivodaliev, "Internet of Things Solution for Intelligent Air Pollution Prediction and Visualization," in *IEEE EUROCON 2019 -18th International Conference on Smart Technologies*. IEEE, 2019, pp. 1–6. doi: 10.1109/EUROCON.2019.8861609 .

[52] S. K. Shah, Z. Tariq, J. Lee, and Y. Lee, "Real-time machine learning for air quality and environmental noise detection," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 3506–3515. doi: 10.1109/BigData50022.2020.9377939 .

[53] Y.-R. Zeng, Y. S. Chang, and Y. H. Fang, "Data Visualization for Air Quality Analysis on Bigdata Platform," in *2019 International Conference on System Science and Engineering (ICSSE)*, 2019, pp. 313–317. doi: 10.1109/ICSSE.2019.8823437 .

[54] S. Dhingra, R. B. Madda, A. H. Gandomi, R. Patan, and M. Daneshmand, "Internet of Things Mobile–Air Pollution Monitoring System (IoT-Mobair)," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5577–5584, 2019. doi: 10.1109/JIOT.2019.2903821 .

[55] WMO, "Early warnings for all," https://library.wmo.int/idurl/4/58209, accessed: Sep. 19, 2024.

[56] A. Cozma, A.-C. Firculescu, D. Tudose, and L. Ruse, "Autonomous Multi-Rotor Aerial Platform for Air Pollution Monitoring," *Sensors*, vol. 22, no. 3, p. 860, 2022. doi: 10.3390/s22030860 .

[57] A. Samad, D. Alvarez Florez, I. Chourdakis, and U. Vogt, "Concept of using an unmanned aerial vehicle (UAV) for 3D investigation of air quality in the atmosphere—example of measurements near a roadside," *Atmosphere*, vol. 13, no. 5, p. 663, 2022. doi: 10.3390/atmos13050663 .

[58] J. Gao, Z. Hu, K. Bian, X. Mao, and L. Song, "AQ360: UAV-Aided Air Quality Monitoring by 360-Degree Aerial Panoramic Images in Urban Areas," *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 428–442, 2020. doi: 10.1109/JIOT.2020.3004582 .

[59] Y. Yang, Z. Zheng, K. Bian, L. Song, and Z. Han, "Real-Time Profiling of Fine-Grained Air Quality Index Distribution Using UAV Sensing," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 186–198, 2017. doi: 10.1109/JIOT.2017.2777820 .

[60] Z. Hu, Z. Bai, Y. Yang, Z. Zheng, K. Bian, and L. Song, "UAV Aided Aerial-Ground IoT for Air Quality Sensing in Smart City: Architecture, Technologies, and Implementation," *IEEE Network*, vol. 33, no. 2, pp. 14–22, 2019. doi: 10.1109/MNET.2019.1800214 .

[61] A. F. Mohammed, S. M. Sultan, S. Cho, and J.-Y. Pyun, "Powering UAV with Deep Q-Network for Air Quality Tracking," *Sensors*, vol. 22, no. 16, p. 6118, 2022. doi: 10.3390/s22166118 .

[62] O. Alvear, C. T. Calafate, N. R. Zema, E. Natalizio, E. Hernández-Orallo, J.-C. Cano, and P. Manzoni, "A Discretized Approach to Air Pollution Monitoring Using UAV-based Sensing," *Mobile Networks and Applications*, vol. 23, pp. 1693–1702, 2018. doi: 10.1007/s11036-018-1065-4 .

[63] Y. Liu, J. Nie, X. Li, S. H. Ahmed, W. Y. B. Lim, and C. Miao, "Federated Learning in the Sky: Aerial-Ground Air Quality Sensing Framework With UAV Swarms," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9827–9837, 2020. doi: 10.1109/JIOT.2020.3021006 .

[64] L. Lei, G. Shen, L. Zhang, and Z. Li, "Toward Intelligent Cooperation of UAV Swarms: When Machine Learning Meets Digital Twin," *IEEE Network*, vol. 35, no. 1, pp. 386–392, 2021. doi: 10.1109/MNET.011.2000388 .

[65] L. Nguyen, T. Le, and Q. Ha, "Prototypical digital twin of multi-rotor UAV control and trajectory following," in *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*. International Association for Automation and Robotics in Construction (IAARC), 2023. doi: 10.22260/ISARC2023/0022 .

[66] Y. Yang, W. Meng, and S. Zhu, "A Digital Twin Simulation Platform for Multi-rotor UAV," in *2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*. IEEE, 2020, pp. 591–596. doi: 10.1109/ICCSS52145.2020.9336872 .

[67] N. Grigoropoulos and S. Lalis, "Simulation and Digital Twin Support for Managed Drone Applications," in *2020 IEEE/ACM 24th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*. IEEE, 2020, pp. 1–8. doi: 10.1109/DS-RT50469.2020.9213676 .

[68] N. A. Manan, A. Aizuddin, and R. Hod, "Effect of Air Pollution and Hospital Admission: A Systematic Review," *Annals of Global Health*, vol. 84, pp. 670 – 678, 2018. doi: 10.29024/aogh.2376 .

[69] Q. P. Ha, S. Metia, and M. D. Phung, "Sensing Data Fusion for Enhanced Indoor Air Quality Monitoring," *IEEE Sensors Journal*, vol. 20, no. 8, pp. 4430–4441, 2020. doi: 10.1109/JSEN.2020.2964396 .

[70] K. Zhang, K. Yang, S. Li, D. Jing, and H.-B. Chen, "ANN-Based Outlier Detection for Wireless Sensor Networks in Smart Buildings," *IEEE Access*, vol. 7, pp. 95 987–95 997, 2019. doi: 10.1109/ACCESS.2019.2929550 .

[71] S. Metia, Q. Ha, H. Duc, and Y. Scorgie, "Urban air pollution estimation using unscented Kalman filtered inverse modeling with scaled monitoring data," *Sustainable Cities and Society*, vol. 54, p. 101970, 2020. doi: https://doi.org/10.1016/j.scs.2019.101970 .

[72] Q. Jiang, D.-c. Gao, L. Zhong, S. Guo, and A. Xiao, "Quantitative sensitivity and reliability analysis of sensor networks for well kick detection based on dynamic Bayesian networks and Markov chain," *Journal of Loss Prevention in the Process Industries*, vol. 66, p. 104180, 2020. doi: 10.1016/j.jlp.2020.104180 .

[73] G. Jesus, A. Casimiro, and A. Oliveira, "A survey on data quality for dependable monitoring in wireless sensor networks," *Sensors*, vol. 17, no. 9, p. 2010, 2017. doi: 10.3390/s17092010 .

[74] C. Elkin, R. Kumarasiri, D. B. Rawat, and V. Devabhaktuni, "Localization in wireless sensor networks: A Dempster-Shafer evidence theoretical approach," *Ad Hoc Networks*, vol. 54, pp. 30–41, 2017. doi: 10.1016/j.adhoc.2016.09.020 .

[75] N. Ghosh, R. Paul, S. Maity, K. Maity, and S. Saha, "Fault Matters: Sensor Data Fusion for Detection of Faults using Dempster-Shafer Theory of Evidence in IoT-Based Applications," *Expert Syst. Appl.*, vol. 162, p. 113887, 2019. doi: 10.1016/j.eswa.2020.113887 .

[76] Y.-W. Du and J.-J. Zhong, "Generalized combination rule for evidential reasoning approach and Dempster–Shafer theory of evidence," *Information Sciences*, vol. 547, pp. 1201–1232, 2021. doi: 10.1016/j.ins.2020.07.072 .

[77] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A Framework for Decision Fusion in Image Forensics Based on Dempster–Shafer Theory of Evidence," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 593–607, 2013. doi: 10.1109/TIFS.2013.2248727 .

[78] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE mobile computing and communications review*, vol. 5, no. 1, pp. 3–55, 2001. doi: 10.1002/j.1538-7305.1948.tb01338.x .

[79] F. Xiao, "Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy," *Information Fusion*, vol. 46, pp. 23–32, 2019. doi: 10.1016/j.inffus.2018.04.003 .

[80] H. Zhang and Y. Deng, "Engine Fault Diagnosis Based on Sensor Data Fusion Considering Information Quality and Evidence Theory," *Advances in Mechanical Engineering*, vol. 10, no. 11, p. 1687814018809184, 2018. doi: 10.1177/1687814018809184 .

[81] K. Sentz and S. Ferson, "Combination of Evidence in Dempster-Shafer Theory," *Sandia report SAND2002-0835, Sandia National Laboratories*, 2002. doi: 10.2172/800792 .

[82] T. H. Le, H. A. D. Nguyen, X. Barthelemy, T. T. Nguyen, Q. P. Ha, N. Jiang, H. Duc, M. Azzi, and M. Riley, "Visualization platform for multi-scale air pollution monitoring and forecast," in *2024 IEEE/SICE International Symposium on System Integration (SII)*, 2024, pp. 01–06. doi: 10.1109/SII58957.2024.10417539 .

[83] PurpleAir, "PurpleAir Flex Air Quality Monitor," https://www2.purpleair.com/products/purpleair-flex, accessed: Oct. 19, 2024.

[84] C. Heaviside, S. Vardoulakis, and X.-M. Cai, "Attribution of mortality to the urban heat island during heatwaves in the West Midlands, UK," *Environmental Health*, vol. 15, no. 1, p. S27, 2016. doi: 10.1186/s12940-016-0100-9 .

[85] S. Park, S. Lee, M. Yeo, and D. Rim, "Field and laboratory evaluation of PurpleAir low-cost aerosol sensors in monitoring indoor airborne particles," *Building and Environment*, vol. 234, p. 110127, 2023. doi: 10.1016/j.buildenv.2023.110127 .

[86] NSW-DPE, "Search for and download air quality data," https://www.dpie.nsw.gov.au/air-quality/air-quality-data-services/data-download-facility, accessed: Jul. 21, 2024.

[87] K. K. Barkjohn, A. L. Holder, S. G. Frederick, and A. L. Clements, "Correction and Accuracy of PurpleAir PM2.5 Measurements for Extreme Wildfire Smoke," *Sensors*, vol. 22, no. 24, p. 9669, 2022. doi: 10.3390/s22249669 .

[88] D. L. Robinson, N. Goodman, and S. Vardoulakis, "Five Years of Accurate PM2.5 Measurements Demonstrate the Value of Low-Cost PurpleAir Monitors in Areas Affected by Woodsmoke," *International Journal of Environmental Research and Public Health*, vol. 20, no. 23, p. 7127, 2023. doi: 10.3390/ijerph20237127 .

[89] K. K. Barkjohn, A. L. Holder, A. L. Clements, S. G. Frederick, and R. Evans, "Sensor data cleaning and correction: Application on the AirNow Fire and Smoke Map," in *American Association for Aerosol Research Conference*, Albuquerque, NM, October 18-22 2021.

[90] T. Becnel, K. Tingey, J. Whitaker, T. Sayahi, K. Lê, P. Goffin, A. Butterfield, K. Kelly, and P.-E. Gaillardon, "A Distributed Low-Cost Pollution Monitoring Platform," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 738–10 748, 2019. doi: 10.1109/JIOT.2019.2941374 .

[91] C. Demir and S. Kim, "Early Warnings Save Lives and Support Long-Term Sustainability," https://www.undp.org/eurasia/blog/early-warnings-save-lives, United Nations Development Programme, accessed: Sep. 19, 2024.

[92] Q. Liao, M. Zhu, L. Wu, X. Pan, X. Tang, and Z. Wang, "Deep Learning for Air Quality Forecasts: A Review," *Current Pollution Reports*, vol. 6, no. 4, pp. 399–409, 2020. doi: 10.1007/s40726-020-00159-z .

[93] G. Czibula, A. Mihai, and E. Mihuleţ, "NowDeepN: An Ensemble of Deep Learning Models for Weather Nowcasting Based on Radar Products' Values Prediction," *Applied Sciences*, vol. 11, no. 1, p. 125, 2021. doi: 10.3390/app11010125 .

[94] Y. Zhang, M. Long, K. Chen, L. Xing, R. Jin, M. I. Jordan, and J. Wang, "Skilful nowcasting of extreme precipitation with NowcastNet," *Nature*, vol. 619, no. 7970, pp. 526–532, 2023. doi: 10.1038/s41586-023-06184-4 .

[95] C. Peláez-Rodríguez, J. Pérez-Aracil, A. de Lopez-Diz, C. Casanova-Mateo, D. Fister, S. Jiménez-Fernández, and S. Salcedo-Sanz, "Deep learning ensembles for accurate fog-related low-visibility events forecasting," *Neurocomputing*, vol. 549, p. 126435, 2023. doi: 10.1016/j.neucom.2023.126435 .

[96] B. Kang, P. Zhang, Z. Gao, G. Chhipi-Shrestha, K. Hewage, and R. Sadiq, "Environmental assessment under uncertainty using Dempster–Shafer theory and Z-numbers," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 5, pp. 2041–2060, 2020. doi: 10.1007/s12652-019-01228-y .

[97] F. Belmahdi, M. Lazri, F. Ouallouche, K. Labadi, R. Absi, and S. Ameur, "Application of Dempster-Shafer theory for optimization of precipitation classification and estimation results from remote sensing data using machine learning," *Remote Sensing Applications: Society and Environment*, vol. 29, p. 100906, 2023. doi: 10.1016/j.rsase.2022.100906 .

[98] T. Gudiyangada Nachappa, S. Tavakkoli Piralilou, K. Gholamnia, O. Ghorbanzadeh, O. Rahmati, and T. Blaschke, "Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory," *Journal of Hydrology*, vol. 590, p. 125275, 2020. doi: 10.1016/j.jhydrol.2020.125275 .

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

[100] Y. Yang, H. Lv, and N. Chen, "A Survey on ensemble learning under the era of deep learning," *Artificial Intelligence Review*, vol. 56, no. 6, pp. 5545–5589, 2023. doi: 10.1007/s10462-022-10283-5 .

[101] J. F. de Oliveira, E. G. Silva, and P. S. de Mattos Neto, "A Hybrid System Based on Dynamic Selection for Time Series Forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3251–3263, 2022. doi: 10.1109/TNNLS.2021.3051384 .

[102] M. Penza, D. Suriano, V. Pfister, M. Prato, and G. Cassano, "Urban Air Quality Monitoring with Networked Low-Cost Sensor-Systems," *Proceedings*, vol. 1, no. 4, p. 573, 2017. doi: 10.3390/proceedings1040573 .

[103] NSW Environment and Heritage, "About the air quality categories," https://www.environment.nsw.gov.au/topics/air/understanding-air-quality-data/air-quality-categories, accessed: Sep. 27, 2024.

[104] NSW Government, "Air quality monitoring network," https://www2.environment.nsw.gov.au/topics/air/monitoring-air-quality, accessed: Oct. 19, 2024.

[105] M. Riley, J. Kirkwood, N. Jiang, G. Ross, and Y. Scorgie, "Air quality monitoring in NSW: From long term trend monitoring to integrated urban services," *Air Quality and Climate Change*, vol. 54, no. 1, pp. 44–51, 2020.

[106] NSW-DPE, "Standards and goals for measuring air pollution," https://www.environment.nsw.gov.au/topics/air/understanding-air-quality-data/standards-and-goals, accessed: Aug. 12, 2023.

[107] T. K. Lim, N. H. Wong, M. Ignatius, M. Martin, H.-M. H. Juang, J. Lou, and R. L. K. Tiong, "Singapore: an integrated multi-scale urban microclimate model for urban planning in Singapore," *Urban Climate Science for Planning Healthy Cities*, pp. 189–217, 2021. doi: 10.1007/978-3-030-87598-5$_9$.

[108] S. Croce and S. Tondini, "Fixed and Mobile Low-Cost Sensing Approaches for Microclimate Monitoring in Urban Areas: A Preliminary Study in the City of Bolzano (Italy)," *Smart Cities*, vol. 5, no. 1, pp. 54–70, 2022. doi: 10.3390/smartcities5010004 .

[109] I. Kousis, I. Pigliautile, and A. L. Pisello, "Intra-urban microclimate investigation in urban heat island through a novel mobile monitoring system," *Scientific Reports*, vol. 11, no. 1, p. 9732, 2021. doi: 10.1038/s41598-021-88344-y .

[110] PurpleAir, "PurpleAir Classic Air Quality Monitor," https://www2.purpleair.com/products/purpleair-pa-ii, accessed: Oct. 19, 2024.

[111] H. N. Duc, M. Azzi, M. Riley, and K. Monk, "Air Quality Modelling of Natural and Man-made events in New South Wales Using WRF-Chem and WRF-CMAQ," *Air Quality & Climate Change*, vol. 57, no. 2, 2023.

[112] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *Proc. of the 33rd Int. conf. on Machine Learning.* PMLR, 2016, pp. 1050–1059. doi: 10.48550/arXiv.1506.02142 .

[113] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv*, 2014. doi: 10.48550/arXiv.1412.6980 .

[114] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," *arXiv*, 2021. doi: 10.48550/arXiv.2106.11342 .

[115] NSW-DPE, "Air quality data services," https://www.dpie.nsw.gov.au/air-quality/air-quality-data-services, accessed: Aug. 12, 2023.

[116] NSW Government, "Health advice," https://www.airquality.nsw.gov.au/health-a dvice, accessed: Oct. 18, 2024.

Appendix A

Air quality index categorisation

The AQI used in this thesis abides by the defined standards of NSW-DCCEEW as below.

Table A.1: Air quality categories standardised by NSW-DCCEEW

Air pollutant	Averaging period	Units	Air quality categories (AQC)				
			GOOD	FAIR	POOR	VERY POOR	EXTREMELY POOR
Ozone	1-hour	pphm	< 6.7	6.7–10.0	10.0–15.0	15.0–20.0	20.0 and above
O_3	8-hour rolling	pphm	< 5	5–6.5	6.5–9.75	9.75-13	13.0 and above
Nitrogen dioxide NO_2	1-hour	pphm	< 8	8–12	12–18	18–24	24 and above
Visibility Neph	1-hour	bsp	< 1.5	1.5–3.0	3.0–6.0	6.0–18.0	18.0 and above
Carbon monoxide CO	8-hour rolling	ppm	< 6.0	6.0–9.0	9.0–13.5	13.5–18.0	18.0 and above
Sulphur dioxide SO_2	1-hour	pphm	< 13.3	13.3–20.0	20.0–30.0	30.0–40.0	40.0 and above
Particulate matter <10 µm PM_{10}	1-hour	µg/m3	< 50	50–100	100–200	200–600	600 and above
	24-hour	µg/m3	< 33.5	33.5-50	50-75	75-100	100 and above
Particulate matter <2.5 µm $PM_{2.5}$	1-hour	$\mu g/m^3$	< 25	25–50	50–100	100–300	300 and above
	24-hour	$\mu g/m^3$	< 16.75	16.75-25	25-37.5	37.5-50	50 and above

Appendix B

Health advice to the citizens of New South Wales

The NSW-DCCEEW (formerly known as the NSW Department of Planning and Environment) and NSW Health devised an activity guide for the public with an attention to sensitive groups on appropriate course of actions to take following the AQI categorised with indicative colour codes in A.1.

Table B.1: Activity guide from NSW Government for recommended actions to protect citizens' health.

Air quality category	General health advice and recommended actions	
	Sensitive group including: • people with a heart or lung condition, including asthma • people over the age of 65 • infants and children • pregnant women	**Everyone else**
Good	• No change needed to your normal outdoor activities.	• No change needed to your normal outdoor activities.
		Continued on next page

Air quality category	General health advice and recommended actions	
	Sensitive group including	**Everyone else**
Fair	• Reduce outdoor physical activity if you develop symptoms such as cough or shortness of breath. • Consider closing windows and doors until outdoor air quality improves. • Follow the treatment plan recommended by your doctor. • If concerned, call HealthDirect on 1800 022 222 or see your doctor.	• No change needed to your normal outdoor activities.
Poor	• Avoid outdoor physical activity if you develop symptoms such as cough or shortness of breath. • Close windows and doors until outdoor air quality improves. • Follow the treatment plan recommended by your doctor. • In case of symptoms, call HealthDirect or see your doctor. • In a health emergency, call 000 for an ambulance.	• Reduce outdoor physical activity if symptoms arise.
		Continued on next page

Air quality category	General health advice and recommended actions	
	Sensitive group including	**Everyone else**
Very poor	• Stay indoors as much as possible with windows and doors closed. • If your home's air quality is uncomfortable, consider going to a place with cleaner air. • Actively monitor symptoms and follow your doctor's treatment plan. • If concerned, call HealthDirect or see your doctor. • In a health emergency, call 000 for an ambulance.	• Avoid outdoor physical activity if you develop symptoms. • Close windows and doors until outdoor air quality improves. • Call HealthDirect if concerned or see your doctor. • In a health emergency, call 000 for an ambulance.
Extremely poor	• Stay indoors with windows and doors closed. • Reduce indoor activity if needed. • If the air at home is uncomfortable, consider going to a place with cleaner air. • Actively monitor symptoms and follow your doctor's treatment plan. • If concerned, call HealthDirect or see your doctor. • In a health emergency, call 000 for an ambulance.	• Stay indoors with windows and doors closed. • Consider going to a place with cleaner air if it's safe to do so. • Call HealthDirect or see your doctor if concerned. • In a health emergency, call 000 for an ambulance.