

Optimization of Sensor Nodes Deployment for Drive-by Sensing and k-hop Calibration

Hassan Zarrar

This is a digitised version of a dissertation submitted to the University of Bedfordshire.

It is available to view only.

This item is subject to copyright.



Optimization of Sensor Nodes Deployment for Drive-by Sensing and k-hop Calibration

Author Name: **Hassan Zarrar**

A thesis submitted to the University of Bedfordshire, in fulfilment of the requirements for the degree of **PhD**

University Of Bedfordshire

Institute for Research in Applicable Computing (IRAC)

Date of Submission for Examination: **October 2024**

Date of Submission for Repository: **August 2025**

Declaration

I, Hassan Zarrar declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have drawn on or cited the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis or any part of it is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Either none of this work has been published before submission, or parts of this work have been published as indicated in Chapters Background Study, Model Development, Model Implementation, Results and Discussion and Conclusion;

Acknowledgement

I want to express my deepest gratitude to my supervisor, Dr. Vladimir Dyo, whose unwavering support and insightful guidance made this work possible. His expertise, encouragement, and invaluable feedback were instrumental throughout every stage of my research journey, and I am sincerely thankful for his dedication to my academic and professional growth. I also extend my heartfelt thanks to the university's management and staff members, whose assistance and resources created a favorable environment for my research. Their help in providing access to essential materials and administrative support was vital in ensuring the successful completion of this work.

I want to extend my special thanks to my colleagues and fellow researchers for their support, shared knowledge, and the collaborative atmosphere they fostered, all of which greatly enriched my research experience. I am especially grateful to my wife and my family for their continuous encouragement and understanding throughout my research. Their patience, belief in me, and emotional and moral support were invaluable, enabling me to persevere through the challenges of this journey. Your prayers and support sustained me, and for that, I am forever grateful.

Abstract

Air pollution presents significant health related risks, requiring detailed and accurate monitoring of air quality with high spatio-temporal resolution. Drive-by sensing is a popular air quality monitoring technique due to its high spatio-temporal coverage, lower deployment costs, and high-resolution data compared to traditional approaches. In this thesis we address the optimization of mobile or bus-based low-cost sensors deployment to maximize air quality monitoring coverage and reference-grade stations placement to make sure that the low-cost sensors mobile sensors deployed on selected routes are continuously calibrated for reliable air quality measurements. In this thesis, we have proposed an optimized approach to select the bus routes for deployment of bus-based sensors on for maximizing the total unique sensing area and deployment of reference-grade sensors on geolocations, for maximizing the probability of k-hop calibration. A metric-based system developed in this model, using geospatial operations which includes both spatial and temporal joins to quantify the contribution of each bus route and rank them accordingly. We have formulated the optimization problem by considering it as a mixed integer linear program (MILP) for the selection of the best possible subset of bus routes required for bus-based sensors deployment while maintaining the level of sensing and temporal coverage contribution similar to the total unique area. We have adopted greedy heuristic approach to solve this problem and use real-world bus-transit data from two cities, Toronto, Canada, and Manchester, UK, for ease of implementation of our approach. Using bus transit datasets in two different cities gives us an opportunity to analyze our proposed model's reliability and adaptability in two different environments. For performance comparison, the greedy approach outperformed the random approach, where the results showed that to achieve 50% area coverage for Toronto bus transit dataset using greedy approach proposed 39 bus routes at 100m buffer distance while for the random algorithm 64 bus routes at buffer distances of 100m. Moreover, we have developed alternative methods to analyze the performance of our proposed weighted approach where the alternate method where the alternative non-weighted approach use metrics (area and trip frequency) separately. We use this comparative analysis of our model in different environment, algorithms and buffer distances, for highlighting the relevance of the proposed metric-based system. The results of the alternative non-weight approach from the Toronto and Manchester bus-transit dataset indicate that our weighted method performed better in certain aspects in terms of spatio-temporal coverage when compared to the alternative approaches with raw or absolute data of area coverage and trip frequency.

Publications

Table 1: Status and description of published papers in peer-reviewed journals.

Sr.	Published Paper Reference	Description
1.	H. Zarrar and V. Dyo, "Drive-by Air Pollution Sensing Systems: Challenges and FutureDirections," *IEEE Sensor Journal,* vol. 23, no. 19, pp. 23692-23703, 2023.	Reviewing drive-by air quality sensing, which offers dynamic, high-resolution data with low operational costs. The focus is on low-cost sensor technologies, vehicular platforms, deployment strategies, calibration, and the challenges of spatial and temporal coverage.
2.	H. Zarrar, M. Limbu, S. Haxha and V. Dyo, "Mobile Calibration for Bus-based Sensing," *IEEE Sensor Journal,* vol. 25, no. 3, pp. 5576 - 5583, 2024.	Proposed a k-hop calibration approach for joint route selection and reference station placement to maximize sensing area using geographical set operations, evaluated with Manhattan bus data, and compared with a random baseline.
3.	H. Zarrar and V. Dyo, "Bus-Based Sensor Deployment for Intelligent Sensing Coverage and k-Hop Calibration," *IET Smart Cities,* vol. 7, no. 1, p. e70004, 2025.	Optimizing the placement of mobile sensors and reference stations to maximize bus route coverage and maintain k-hop calibratability using MILP and a greedy algorithm, and applied to bus-transit data from Toronto and Manchester.

List of Abbreviations

1. AQI: Air Quality Index
2. DEFRA: Department for Environment, Food and Rural Affairs
3. EPA: Environmental Protection Agency
4. DAQI: Daily Air Quality Index
5. LCS: Low-cost Sensor
6. EC: Electro-Chemical
7. MO: Metal-oxide
8. NDIR: Non-dispersive Infrared
9. PM: Particular Matter
10. CO: Carbon Monoxide
11. NO2: Nitrogen Dioxide
12. O3: Ozone
13. SUV: Sport Utility Vehicle
14. NO: Nitrogen Monoxide
15. UAV: Unmanned Aerial Vehicle
16. ST-ICM: Spatial-Temporal Inference Calibration Model
17. SCAN: Sensor Array Network Calibration
18. V2V: Vehicle to Vehicle
19. V2N: Vehicle to Network
20. V2I: Vehicle to Infrastructure
21. ACO: Ant Colony Optimisation
22. LP: Linear Programming
23. ILP: Integer Linear Programming
24. ARIMA: Auto Regressive Integrate Moving Average
25. PCR: Principal Component Regression
26. GEP: Gene Expression Programming
27. MCLB: Maximizing Coverage with Limited Budget
28. NP: Non-deterministic Polynomial time
29. DP: Dynamic Programming
30. GA: Genetic Algorithm
31. WSN: Wireless Sensor Network
32. DV-Hop: Distance Vector-Hop
33. HOLA: Heuristic and Opportunistic Link Selection Algorithm

34. SA: Simulated Anneal
35. PSO: Particle Swarm Optimization
36. PSO-ILS: Particle Swarm Optimization-Iterated Local Search
37. EA: Evolutionary Algorithms
38. ANN: Artificial Neural Network
39. AQMS: Air Quality Monitoring Station
40. RCGA: Real Coded Genetic Algorithm
41. GTFS: General Transit Feed Specification
42. SD: Standard Deviation
43. ID: Identity or Identification
44. WSM: Weighted Sum Method
45. CFD: Computational Fluid Dynamics
46. GSM: Global System for Mobile communication

Table of Contents

Declaration 2
Acknowledgement 3
Abstract 4
Publications 5
List of Abbreviations 6
List of Tables 11
Tables of Figures 12
Chapter 1: Introduction 13
1.1. Background 13
1.2. Problem Formulation 14
1.3. Proposal and Contribution 17
1.5. Aims and Objectives 20
1.6. Thesis Structure 21
Chapter 2: Background Study 23
2.1. Air Pollutants, Standards, and Sensors 23
2.1.1. Air Pollutants 23
2.1.2. Air Quality Standards 24
2.1.3. Low-Cost Sensor Technology 25
2.1.4. Discussion 30
2.2. Vehicular Platform 31
2.2.1. Public transit 31
2.2.2. Municipal Transport 33
2.2.3. Private Transportation 34
2.2.4. Discussion 35
2.3. Spatio-Temporal Resolution 36
2.4. Mobile Sensor Calibration 37
2.5. Optimisation Problems 40
2.5.1. Linear and Non-linear Programming 40
2.5.2. Combinatorial Optimization 44
2.6. Optimisation Algorithms 44
2.6.1. Exact Algorithms 45
2.6.2. Heuristic Algorithms 48
2.6.3. Meta-Heuristic Algorithms 52

2.6.4. Comparative Analysis of Modelling Approaches 56
Chapter 3: Model Development 58
3.1. Framework 58
3.2. Model Core Elements 61
3.2.1. Route Definition 61
3.2.2. Sensing Coverage 61
3.2.3. Sampling Coverage 62
3.2.4. K-hop Calibration 62
3.3. Weight Determination 63
3.3.1. Metric Selection and Standardization 63
3.3.2. Weights Calculations 64
3.4. Mixed Integer Linear Program Formulation 65
3.4.1. Bus Route Model for Maximal Spatio-Temporal Coverage 65
3.4.2. Minimal Bus Stop Selection Model 66
3.5. Greedy Approximation 67
3.6. Baseline Random Algorithm 69
Chapter 4: Model Implementation 71
4.1. Data Segmentation 71
4.1.1. Shapes (shapes.txt) 73
4.1.2. Route (routes.txt) 73
4.1.3. Trip (trips.txt) 74
4.1.4. Stop (stops.txt) 74
4.2. Weights Calculation 74
4.3. Route Selection 76
4.4. Bus Stop Selection 77
Chapter 5: Results and Discussion 81
5.1. Dataset Description 81
5.1.1. Toronto Dataset 81
5.1.2. Manchester Dataset 82
5.1.3. Buffer Distance and Pollutant Dispersion 82
5.1.4. Calibration and Rendezvous Ranges 83
5.2. Evaluation of Bus Route Model for Maximal Spatio-Temporal Coverage 83
5.2.1. Toronto and Manchester Dataset Results 83
5.2.2. Effect of Z-scores and Weight on Coverage Model 86
5.3. Minimal Bus Stop Selection Model Evaluation 90

5.3.1. Toronto and Manchester Dataset Results 90
5.3.2. Interpretation of Bus Stop Selection Patterns 92
5.4. Discussion 92
Chapter 6: Conclusion 96
6.1. Summary of Our Findings 96
6.2. Contributions 98
6.3. Limitations 99
6.4. Future Work 101
References 102
Appendices 120
Ethics Form 121
Research Data Management Plan 128

List of Tables

Table 1: Publication status of peer-reviewed papers 5
Table 2: Variety of air pollutants. 24
Table 3: Standards for air pollutants. 25
Table 4: Daily Air Quality Index 25
Table 5: Existing Related Survey/Review Papers. 29
Table 6: Comparison of LCS types with pros and cons 30
Table 7: Pre-processed Toronto city bus transit dataset. 82
Table 8: Pre-processed Manchester city bus transit dataset. 82

Tables of Figures

Figure 1: Single-hop calibration and multi-hop calibration. 38
Figure 2: Various platforms in the wireless multi-hop network 39
Figure 3: Branching structures of optimization approaches. 60
Figure 4: Block diagram of the proposed model 60
Figure 5: Greedy and random algorithm routes optimisation for Toronto city 85
Figure 6: Greedy and random algorithm routes optimisation for Manchester city 85
Figure 7: Weighted and absolute route area approach for Toronto city 88
Figure 8: Weighted and absolute route trips approach for Toronto city 88
Figure 9: Weighted and absolute route area approach for Manchester city 89
Figure 10: Weighted and absolute route trips approach for Manchester city.. 89
Figure 11: Bus stops required to k-hop calibrate for Toronto dataset……..……91
Figure 12: Bus stops required to k-hop calibrate for the Manchester dataset……..……92

Chapter 1: Introduction

This chapter presents the overview of the study by detailed discussion of the background and context, followed by the research problem formulation. Next, we address several challenges facing drive-by sensing deployment, and outlining the research aims and objectives. Also included are the posed research questions and concluding with a review of prior similar work and their discussion.

1.1. Background

Air pollution are categorized as harmful substances that are released into the air from various sources, which contaminate the surrounding environment [1]. Air pollutants can be emitted from a wide number of sources, including man-made or artificial sources such as emissions from vehicles, industrial activities (for example, from factories and power plants), and from land usage (for example, from agricultural and urban expansion) [2]. Air pollutants are also released into the air from natural sources, such as volcanic eruptions, forest fires and natural events which involve combustion [2]. Outdoor ambient air pollution is considered as a serious concern for global public health risk, which contributes to premature deaths in both urban and rural regions [3]. Air pollution also adversely impacts the surrounding ecosystem and local biodiversity, affecting plant and animal life [4]. It is estimated that by 2050 around 70% of the global population is expected to be concentrated in urban regions [5]. Moreover, air pollution is not considered an isolated phenomena which is restricted to a specific environment [6]. Therefore, policymakers and agencies have considered several options to reduce the impact of air pollution and developed legal and institutional frameworks. Initiatives have been taken to control air pollution; however, it is important to evaluate the effectiveness of the employed initiatives or policies. To achieve this, it is important to accurately collect or measure air pollution concentration data [7], for analysis, which can be used for several applications, for instance, locating air pollution emission sources, for public awareness campaigns, forecasting of air pollutant levels, ensuring air quality standards are being followed, etc.

Several cities have deployed fixed air quality monitoring stations, where high-grade sensors are utilized; however, they are expensive and therefore are deployed in fewer numbers, alternatively low-cost sensors are deployed in large numbers or a combination of both [8]. Low-cost sensor (LCS) technology has emerged as a popular approach to monitoring air quality because of their features, such as their cost-effectiveness, easy to operate and maintain, providing fast real-time measurements, and commercial availability [6]. By increasing the number of low-cost sensors, high spatio-temporal coverage and real-time air quality data at a lower cost can be achieved compared to deploying fewer reference sensors. There are several Internet of Things (IoT) based applications which have included low-cost sensors and placed in dense sensor networks for monitoring air quality in urban sectors [9]. Fixed air quality monitoring sensors are more commonly

placed in close proximity to emission sources, for instance, commercial, and residential areas due to their high population density and near industrial areas for measuring real-time data. As we mentioned earlier, the drawback of deploying reference-grade sensor is that they require a significant amount of finances for purchase, placement and maintenance, making them less feasible for projects with budget constraints [8]. Moreover, deploying fixed monitoring sensors require infrastructure for their installation, which may require construction, and as they are located in a fixed point, therefore, they cannot measure spatial variations in air pollution levels. Spatial interpolation techniques have been used to address this limitation by utilizing air pollution data from sampled points collected at other fixed monitoring stations to estimate air pollution levels at unsampled points. However, spatial interpolation approach provide more accurate estimation at unsampled points which are in close proximity to the sampled locations compared to unsampled points that are further away [10].

Drive-by sensing has emerged as a popular concept where different types of transport modes which serve as mobile sensing platforms for monitoring the urban environment [11] [12]. Different sensors when mounted on vehicles like buses, taxis, and utility vehicles, can be used for sensing, air pollution [12] [13], road surface conditions [14], noise levels [15], etc., as the vehicles perform their routine operations. Public transport buses are a particularly attractive mobile platform due to their availability, their predictable or fixed trajectories, frequent and reliable operation, and city-wide coverage [16] [17] [18] [19]. There are several projects for monitoring air quality in urban sector, for instance, OpenSense [20], and Mosaic [21], and there are also projects that have introduced models and algorithms to improve measurement coverage and accuracy. There are trade-off between to be considered, when deploying the number of sensors and the resulting spatio-temporal coverage of bus-based sensing networks which is a key research challenge [16] [14]. Ali et al. [14], formulated the original problem of optimal mobile sensor deployment for coverage maximization and proposed an algorithm that was evaluated using the London bus route dataset. Cruz et al. [22], developed a linear optimization model to optimize spatio-temporal coverage and evaluated it using mobility traces from Rio de Janeiro for applications such as waste management, air quality and noise monitoring and ranking buses accordingly. Paliwal et al. [23], proposed leveraging the spatial and temporal correlations in air quality data to select buses that sample a diverse and representative set of locations. Their approach was evaluated using GTFS data from Delhi, India.

1.2. Problem Formulation

Drive-by sensing provides significant spatio-temporal coverage and reduces deployment and overhead costs which are associated with fixed monitoring stations. However, drive-by sensing also presents unique challenges in terms of deployment and costs, which affect scalability of the mobile sensor network.

Furthermore, the accuracy and reliability of the mobile sensor is also effected due to external and internal factors which effect the collected data; for this, mobile sensors require frequent calibration.

Spatio-temporal coverage problem—There are several factors that impact the spatio-temporal coverage drive-by sensing, which depends on the selected mobile platforms. Mobile platform features must be considered, for instance, their trajectory or path, spatial coverage, operations and schedule, availability, and cost of sensor placement. In prior works [11] [24] [17] bus transit are utilized as a mobile platform for data collection, as it offers several advantages, such as, fixed routes, which makes its trajectory predictable. Moreover, bus transit provides high spatial and temporal coverage as it operates in large fleet, covers a wide area in any city and frequently repeats each route daily. In addition to wide coverage and repeated movement on each route, most metropolitan cities operates bus transit. Despite several advantage of utilizing bus transit for drive-by sensing, there are some limitations, for instance, deploying an entire bus fleet with air quality sensors increases maintenance and deployment costs making it infeasible. However, reducing the number of mobile sensors lowers spatial and temporal coverage, which affects the granularity of collected data. Therefore, creating a balance between the cost of deployment and high spatio-temporal coverage is a challenging task, as shown in prior studies [14] [25] [26]. Another challenge is prioritizing bus routes for mobile sensor deployment based on spatial and temporal coverage. these challenges include the number of buses operating each bus route and its coverage, for example, certain bus routes providing high spatial coverage may have fewer buses operating on them, therefore, limiting the temporal coverage. While bus routes with lower spatial coverage which operate a high number of buses may provide more frequent temporal measurements. Therefore, geospatial coverage requirements may not align with the bus schedules, creating gaps in temporal coverage requirements. Finding the balance between maximizing area coverage and frequent data collection is a challenge in deploying bus based sensors, with limited resources.

In-field sensor calibration problem— Utilizing low-cost sensors as an affordable alternative to high-cost reference-grade sensors is increasing. However, low-cost sensors are susceptible to factors, such as internal factors causing temporal drifts due to ageing of sensor components and external factors causing sensor drifts due to environmental conditions and sensor manufacturing tolerances, which can impact on their accuracy and stability [27]. Therefore, periodic calibration of LCS is required to minimize uncertainty and maintain data measurement accuracy. Sensor calibration involves collocating the LCS with a high-grade reference sensor to correct the systematic bias error in the LCS readings [28]. Low-cost sensors may require frequent calibration, for instance, on a weekly [29], or monthly [30] basis to maintain reading accuracy. For example, Vito et al. [31] reported a significant sensor drift after only one month of deployment. Calibrating in-field or post-deployed mobile LCS is a challenging task, as it requires co-locating with fixed reference monitoring stations, which requires suspension of operations of mobile platform. Therefore, a single-hop

approach is utilized to address this problem where mobile sensors are calibrated against a fixed reference grade sensor via a rendezvous phenomenon where both mobile and reference-grade sensors are in close proximity and sensing the same environmental factor. Reference sensors are generally expensive to deploy and maintain compared to mobile low-cost sensors. Reference-grade sensor placement on each route increases the deployment cost, therefore, to address this issue, strategic placement of reference sensors to perform 1-hop calibration of mobile sensors is possible on multiple routes. Alternatively, the multi-hop approach can be used in which freshly calibrated sensor through a single-hop approach, which can then be used to calibrate other mobile sensors when they rendezvous with each other, performing 2-hop, or even 3-hop calibration or more generally known as k-hop calibration. Main benefit of using a multi-hop approach is that fewer reference-grade sensors are required to calibrate multiple mobile sensors. The number of reference-grade stations and their geolocations for placement also requires optimization to reduce the deployment costs.

The research problem discussed and addressed in this thesis includes the optimization of mobile or bus-based low-cost sensors deployment for maximizing air quality monitoring coverage and reference-grade stations placement to ensure that the deployed mobile low-cost sensors are periodically calibrated to ensure sensor measurement reliability. Some research questions have been formulated to address the identified challenges in optimizing drive-by sensing deployment for air pollution monitoring which are as follows:

Q1. How can the bus routes be prioritized to maximize spatio-temporal coverage with limited resource allocation?

- To effectively monitor air quality, a wide coverage of the target area and frequent data sampling are required. However, this is difficult to achieve with limited resources, therefore, optimization of bus routes selection for mobile LCS deployment is required to maximize sensing coverage with budget constraints.

Q2. How can we optimize the number of reference station, and ensure k-hop calibrability of the entire bus-based sensor network with limited number of reference sensors, while considering the prioritized bus routes?

- Calibration of deployed mobile sensors to maintain the accuracy of the sampled data throughout the wireless sensor network. To ensure k-hop calibrability, mobile sensors are required to rendezvous with the deployed reference-grade sensors or recently calibrated mobile sensors. To k-hop calibrate the entire network with a minimum number requires the optimization of reference sensor placement and their geolocations.

Q3. What are the different challenges and opportunities when utilizing current drive-by sensing approaches in a dynamic urban environment, and how can the optimization process address them?

- There are several limitations of using drive-by sensing, such as irregularities in sensing coverage due to dynamic network topology, mobile platform type and behavior, and difficulties in calibrating the post-deployed mobile low-cost sensors. By optimizing mobile and reference sensors deployment can improve sensing coverage and periodic calibration, ensuring reliable and accurate data collection at reduced deployment costs.

1.3. Proposal and Contribution

In this thesis, we have proposed an optimized approach for both the deployment of bus-based sensors on bus routes for maximizing the total unique sensing area and the deployment of reference-grade sensors to maximize the probability of k-hop calibration. The approach first selects an optimized subset of routes and then computes the least possible number of reference stations and locations for each subset of routes selected. In this model, the geolocation of reference stations is restricted to bus stops due to their near-road geolocation, existing infrastructure requiring minimum construction or changes, and available power supply [32] [33] and to limit the impact of apparent wind on the measurements [34], since calibration can be done when the bus is stationary. Similarly to [32], the approach uses an area-based metric rather than the total length of route segments for coverage computation. In our model, a metric-based system is developed which uses geographical set operations, which includes spatial and temporal joins, to quantify the contribution of each bus route and rank them accordingly. We have formulated the optimization problem as a Mixed Integer Linear Program (MILP) for the selection of the best possible subset of bus routes required for bus-based sensors deployment while maintaining the level of sensing and temporal coverage contribution similar to the total unique area. We have adopted greedy heuristic approach to solve this problem and use real-world bus-transit data from two cities, Toronto, CA [35], and Manchester, United Kingdom [35] for ease of implementation of our approach. Both metropolitan cities have different geospatial features such as road network topology for comparison to better understand our proposed approach. The results of our proposed system shows fewer bus routes are required for monitoring, therefore reducing the number of buses required for sensor node deployment. Furthermore, we have also compared the performance of our metric-based system with a random approach and alternative methods. Our research's main contributions include:

1. Comprehensive Literature Review: In this thesis, a comprehensive review of existing literature has been conducted regarding several aspects of drive-by sensing, and air quality monitoring. This includes in-depth analysis of the different types of air pollutants and their concentration standards, the role of low-cost sensor technologies in monitoring urban air quality, evaluation of different types of vehicular or mobile

platforms for drive-by sensing such as public transit, municipal transport, private vehicles, and unmanned aerial vehicles (UAVs). Limitations regarding post-deployment or in-field calibration techniques for mobile sensors in dynamic urban environments and their opportunities have also been included. Moreover, our literature review also includes an analysis of between achieving high spatial and temporal resolution or coverage in mobile sensor networks and their trade-offs. The findings of this comprehensive literature review have been published as a survey paper, highlighting an in-depth discussion and analysis of drive-by sensing and identified key research areas for future research. Moreover, in this thesis we have also included a comprehensive review of various optimization problems and algorithms, which includes exact and approximate approaches to solve different optimization problems.

2. Novel Optimization Framework: We have developed a metric-based system for this framework, which quantifies the contribution of each bus route and prioritizes each route according to their ranks. For this, we have used bus transit data which includes geographical operations, in terms of spatial (route area) and temporal (route trips) joins, which was collected from GTFS files. We have formulated the maximum sensing coverage problem as an MILP, by utilizing the weighted sum method (WSM), which enables our multi-objective optimization problem and converts it in a single-objective mathematical optimization problem. We have normalized the two metrics which includes each route area and its trips, using z-score standardization to convert our dataset metrics at a more comparable scale and for calculating the weights of each metric and each route coverage contribution. In addition to optimizing bus-based sensor deployment on selected bus routes. Moreover, our model calculates the minimum number of reference sensors required for k-hop calibration and identifying the most effective geolocation for achieving successful rendezvous to k-hop calibrate the deployed bus-based sensors.

3. Comparative Analysis: We have employed a greedy heuristic approach to solve the optimization problem, and for performance comparison of our weighted metric-based system we used the results from a random approach. In addition to the random algorithm, we have developed two alternative approach for comparative analysis, where we used absolute or raw route area and trips data separately, without weighted metrics, and we also used greedy approximation to reduce computational complexities and time. Furthermore, we have applied our weighted approach and alternative methods to two cities which have different geospatial features to analyze how it will affect the performance of our model. The inclusion of two alternative approach adjacent to the random approach has provided an in-depth insight into the contribution to our weighted metric-based approach's performance under different or dynamic environments.

Similar Work—The prior work on coverage optimization tends to focus on a selection of subsets of routes for best coverage. Ali et al. [14], investigated the optimal coverage for monitoring road surface condition

using a limited number of bus-based mobile sensors. The authors developed a coverage model by street segmentation, defined as sections of a street between two adjacent intersections. Mixed Integer Linear Program (MILP) formulation was then developed to address the maximum coverage problem using a greedy algorithm to determine the approximate solution. Caminha et al. [11], presented a coverage model to optimize the spatial coverage of a bus-based sensor network. They propose formulating the problem as an MILP, which includes the street section lengths. Cruz et al. [22] proposed a metric to quantify the contribution of individual buses and applied the metric to bus mobility traces from Rio de Janeiro for applications of waste management, air quality, and noise monitoring, ranking buses accordingly. Biondi et al. [24] proposed an approach in which buses are used for deploying sensor nodes to increase the number of urban sites monitored. Sun, et al. [36] propose a super-efficient data envelopment analysis (DEA) to categorize and evaluate existing routes and a comparative analysis to assess bus lines for filtering out candidates for further optimization. It should be noted that this study used a length-based metric, i.e., the total length of selected bus routes for quantifying coverage. In contrast, our metric-based system quantifies the contribution of each bus route based on the area of spatial joins and bus-transit operations data, which includes the number of trips for each route.

More recently, Zarrar et al. [32], investigated the problem of optimal reference station placement for maximizing total unique sensing area for k-hop calibration. The proposed work aims to utilize an area-based metric for coverage computation and constrained reference station placement near bus stops. The approach directly optimizes the reference station placement using a greedy heuristic and also computes coverage as the total unique sensing area of all bus routes serving selected reference stations. In contrast, in this work, we are k-hop calibrating only the subset of bus routes that have been selected for sensing. In other words, the approach first selects the subset of bus routes that provides maximum coverage, and then calculates the minimum number and location of reference stations for the selected subset of bus routes. Moreover, this work develops a metric-based system for quantifying the contribution of each bus route in both spatial and temporal aspects. Moreover, we propose a MILP formulation that includes each bus route area for addressing the spatial coverage and its bus route trips to address the temporal coverage aspect of the optimization problem.

Huang, et al. [37], presented a two-bus selection model in which the spatio-temporal coverage is considered and the temporal gap between consecutive coverage at certain locations, to optimize the bus-based mobile sensing strategies and to improve the performance of air quality monitoring. The authors evaluated the importance of the area based on five criteria that they identified. Dai, et al. [38] propose a multi-line drive-by sensing bus assignment and scheduling problem, which aims to maximize sensing score and minimize operational costs simultaneously, while serving all timetabled trips bound by a set of operational constraints.

The authors have also explored the drive-by sensing power of bus fleets through active scheduling. In contrast, we developed a metric to measure and quantify each bus route sensing contribution and calibration. Yu, et al. [39] formulate the Bus Sensor Deployment Problem (BSDP) and solve this problem by a chemical reaction optimization (CRO) technique to select a subset of bus lines.

Ji et al. [40] introduce the Open Team Orienteering Problem with Repeatable Visits to model vehicle-based mobile sensing for targeted urban monitoring. The study employs an Adaptive Large Neighbourhood Search algorithm to plan coordinated vehicle routes that repeatedly visit key points of interest, ensuring sufficient coverage. Zoysa et al. [41] proposed to leverage public transport system sensor network for monitoring road surface conditions. This approach is particularly useful in developing countries where there are insufficient monitoring and reporting infrastructure which can lead to delays in road maintenance and resurfacing. The proposed bus-based sensor network also utilizes fewer sensors for data collection from different locations. In contrast, our proposed drive-by sensing model focuses on optimizing bus route selection using each route sensing contribution to maximize spatio-temporal coverage with budget constraints.

Ji et al. [42] present a trip-based model for optimization of sensor deployment in which area coverage is defined through spatial grids using urban bus transit systems. The target area is divided into discrete units, and using spatial weights critical areas are prioritized, and by using grid intersection the coverage is determined. The temporal coverage is managed by using temporal intervals which are segmented monitoring periods, and based on the presence of sensors within the grids during specific intervals, the coverage was assessed. To prioritize certain time periods temporal weights are then applied, allowing for decoupling of spatial and temporal structures. The multi-stage solution framework contributes to computational efficiency, and also utilizes linear integrated programming to further reduce computational complexity and ensure efficient sensor distribution. In contrast, our model prioritizes sensor deployment to maximize spatio-temporal coverage by quantifying the contribution of each bus route sensing coverage by utilizing spatial (area) and temporal (trip frequency) metrics and applying their weights. We have formulated our problem as Mixed Integer Linear Programming and greedy algorithm to solve this problem which also reduces computational complexities. Several benefits of using MILP includes solving complex problems with continuous and discrete variable and offers the introduction of budget constraints. Additionally, our approach integrates maximizing spatio-temporal coverage with optimizing reference sensor placement to ensure frequent k-hop calibration of the entire sensor network.

1.5. Aims and Objectives

We aim to achieve an optimized drive-by sensing framework for monitoring air quality by prioritizing bus route selection for bus-based sensor deployment and reference-grade sensor placement, using a metric-

based system to maximize the unique spatio-temporal coverage and ensures that all mobile sensors in the network are k-hop calibratable. This research focuses on the following objectives:

- To conduct a comprehensive literature review on several aspects of drive-by sensing and air quality monitoring which includes in-depth analysis of mobile platforms, sensors technologies, air pollutants, calibration methods and optimization strategies. Moreover, to investigate the challenges of deploying mobile sensors regarding the trade-offs between achieving high spatial and temporal coverage.
- To develop a metric-based framework that quantifies each bus route spatio-temporal contribution using spatial (area-based)and temporal (trip-frequency) data.
- To formulate optimization problems of bus route selection as a Maximum Coverage and reference sensor placement as a Set Cover Problem.
- To implement k-hop calibration model by identifying the minimum number of reference-grade sensors required and their geolocations (Bus stops), to ensure that all mobile network-wide is periodically calibrated.
- To implement the proposed optimization framework using real-world bus transit data from two cities, and to evaluate its performance between two heuristic approach (Greedy and Random) and alternative methods, and present the final findings.

1.6. Thesis Structure

This thesis has been structured systematically, where research gaps and challenges have been addressed. In Chapter 2, we have conducted a comprehensive literature review where we discussed several aspects of drive-by sensing and air quality monitoring, including air pollutants, air quality concentration standards, and low-cost sensor technologies, and their limitations, and calibration methods. In this thesis, a discussion on the impact of deploying low-cost sensors on various vehicular or mobile platforms and their various challenges have been conducted. An analysis of the impact of mobile sensor placement on spatio-temporal resolution has been conducted and their trade-offs have been discussed. Furthermore, several optimization problems and techniques have been discussed and how to address them using different algorithms, such as exact and approximate algorithms. The comprehensive literature review conducted provided the foundation for our optimization model, which aims to address the highlighted challenges and research gaps. In Chapter 3, we have mainly focused on the development of our optimization model framework. We describe the basic concept of our model, where we define different aspects of the development of our metric-based system and then develop the optimization model formulation. The problem formulation includes the selection of

key metrics and their normalization, and then progress to the development of the mathematical formulation of the MILP, supporting the greedy and random algorithms used to solve the formulated problem.

Next, in Chapter 4, we outline our optimization model implementation, which includes detailed process of the data segmentation, key metrics normalization, and their integration into our model. Utilizing greedy approximation, the process of the optimized selection of a subset of bus routes for deployment of bus-based sensors and selection of the right subset of bus stops for sensor placement for k-hop calibration has been described. Additionally, we have also described the implementation of a random algorithm for performance analysis. In Chapter 5, we present the results of our optimization model by applying it to datasets from Toronto and Manchester, and comparing the effectiveness of our greedy approach with the random algorithm and alternative methods. We discuss the impact of our weighted method on the sensing coverage and have described the patterns observed in the selection of the bus routes and bus stops under different scenarios. In Chapter 6, this thesis concludes by summarizing the key research findings, outlining the contributions, and addressing limitations. Moreover, we have discussed possible future research directions, which includes several aspects that can improve our proposed model and promote broader applications. In this chapter we have also discussed the potential contributions of our model to smart city initiatives. Following the conclusion, this thesis paper includes a list of references that cite all the sources and literature used and the appendix section.

Chapter 2: Background Study

In this chapter, we discuss various applications and research gaps related to drive-by sensing for air quality monitoring, problem formulation, and solution techniques. Initially, we will discuss the air quality standards and low-cost sensor technologies used to monitor their concentration levels. In this section, we provide an overview of different types of air pollutants, the sources of emission, and the effects they have on the surrounding environment and public health. Following this, we detail the concentration standards set by organizations and the associated health implications. We evaluate multiple types of low-cost gas sensor technologies and outline their benefits and limitations. We also focus on the various vehicular or mobile platforms which can be candidates for mobile sensing. We provide a comprehensive review of the vehicular platforms available for sensor deployment, including private and public transport, municipal transport, and unmanned aerial vehicles. We discussed their contribution to air quality monitoring, their benefits and the challenges they present. Following this discussion, we also addressed the importance of spatio-temporal resolution in sensor placement for air quality monitoring. We highlight the importance of achieving high spatio-temporal resolution and various challenges when achieving this with mobile sensors, and highlight issues between fixed monitoring stations and mobile sensing platforms. We also discuss the challenges of using low-cost sensors for monitoring air quality after deploying them on mobile platforms and how to address them using mobile sensor calibration techniques. In this chapter, we focus on the importance of periodic calibration of low-cost sensors and the different techniques used in the existing literature for calibrating in-field or post-deployment of mobile sensors. We also mention the challenges facing the maintenance of sensor data accuracy in the context of varying temporal and spatial coverage. Finally, we discuss in detail different types of optimization problems and algorithms to solve the various problems, which include exact and approximate approaches and review existing literature with regard to air quality monitoring and sensor placement. The comprehensive overview of various existing literature provides a better understanding of the benefits, challenges and future research opportunities in mobile air quality monitoring, which has led to the detailed methodologies and findings presented in the following chapters. The content of this chapter is based on the author's published review paper, "Drive-by Air Pollution Sensing Systems: Challenges and Future Directions" [16], and has been adapted to suit the objectives of this thesis.

2.1. Air Pollutants, Standards, and Sensors

2.1.1. Air Pollutants

Air pollutants are high concentrations of materials in the air that exceed the environment's capacity to dilute or decompose in a timely manner and have potentially adverse effects on human health, plant life and climate. Researchers have identified common air pollutants, including oxides of sulphur (SO_x), oxides of nitrogen (NO_x), carbon monoxide (CO), carbon dioxide (CO_2), ozone (O_3), fine particulate matter (e.g.,

PM_{10} and $PM_{2.5}$) and Volatile Organic Compounds [43]. There are several emission sources of air pollution, but they are broadly categorized into two main categories: 1. artificial sources and 2. natural sources. Artificial sources of air pollution include emissions from transportation, industrial activities like factories, power generation plants, and land usage practices like irrigation and urban expansion [2]. Another prominent artificial source of air pollutants is from transportation modes, which consist of those generated through fuel combustion and incomplete combustion in vehicle engines [44]. Natural sources also emit air pollutant particles which affect the environment, such as volcanic eruptions, forest fires and other natural events [2].

Table 2: A variety of air pollutants.

Air pollutants	**Sources**	**Description**	**Health-related issues**
Particulate Matter (PM) [45]	Chemical reactions, industrial sites, fuel combustion, and fires	PM2.5= 2.5 micrometers diameter, PM10= 10 micrometers diameter	Cardiovascular and respiratory diseases
Sulphur dioxide (SO_2) [46] [47]	Combustion of materials or fuel containing sulphur	Colourless has an odour (irritates) [47]	Respiratory-related issues [47]
Nitrogen dioxide (NO_2) [46] [47]	Combustion, road traffic and power generation	Colourless, acidic, highly corrosive, and has an odour	Respiratory infection, asthma, and chronic lung disease
Ozone (O_3) [47]	Combustion, road traffic, bushfires, industrial power generation	Colourless, it is highly reactive and has an odour	Can cause cardiac and respiratory-related problems
Carbon monoxide (CO)	Created by incomplete combustion of fossil fuel, cars and engine exhaust fumes, heating stoves, etc [48]	Colourless, odourless, and non-irritating gas [47]	Reduces oxygen supply to various parts of the human body [47]
Volatile Organic Compounds (VOCs)	Roadside traffic, factories, indoor emission sources, chemical processes, fuel combustion, etc [49]	Different gas examples: 1,3-butadiene, benzene, styrene, etc [49]	Respiratory problems, organ damage, neurological disorders, cardiovascular disease [50]

2.1.2. Air Quality Standards

Air quality indices (AQIs) are developed as indicators in numerical form that provide current air pollution levels, and provide their impact on public health, and provide cautionary guidance for the people [51]. Governments and environmental agencies have established specific concentration limits for each air pollutant. When these limits are exceeded, the air poses a risk to public health and the environment. To illustrate, the air pollutant concentration limits we have provided are air pollutant concentrations from three different agencies, shown in Table 3. The air pollution data shown in Table 3 are reported as an average concentration over different periods or time durations, which include hourly, annually, or during peak seasons. Different agencies have developed their air quality index (AQI) values and levels. For instance, the two agencies we have selected are the Department for Environment, Food and Rural Affairs (DEFRA)

in the United Kingdom. And the U.S. Environmental Protection Agency (EPA), as a comparative, provides air quality-related information for public awareness about air pollution levels in specific areas, as shown in Table 4. To calculate the air quality indices (AQI) levels, air pollutant concentration data is used.

Table 3: Different standards of different air pollutants.

Air Pollutant	**Averaging period**	**Agencies**		
		United States Environmental Protection Agency, Air pollutant concentration [45]	**European Commission, Air pollutant concentration** [52]	**World Health Organization, Air pollutant concentration** [53]
Particulate Matter ($PM_{2.5}$)	24 hours	35 µg/m^3	-	15 µg/m^3
	Annual	12 µg/m^3	25 µg/m^3	5 µg/m^3
Particulate Matter (PM_{10})	24 hours	150 µg/m^3	50 µg/m^3	45 µg/m^3
	Annually	-	40 µg/m^3	15 µg/m^3
Carbon Monoxide (CO)	1 hour	35ppm	-	-
	8 hours	9ppm	10 mg/m^3	-
	24 hours	-	-	4 mg/m^3
Ozone (O_3)	8 hours	0.070 ppm	120 µg/m^3	100 µg/m^3
	Peak Season	-	-	60 µg/m3
Nitrogen dioxide (NO_2)	1 hour	100 ppb	200 µg/m^3	-
	24 hours	-	-	25 µg/m^3
	Annually	53 ppb	40 µg/m^3	10 µg/m^3
Sulphur dioxide (SO_2)	1 hour	-	350 µg/m^3	-
	24 hours	-	125 µg/m^3	40 µg/m^3
	Three months	0.15 µg/m^3	-	-

Table 4: Daily Air Quality Index by Department for Environment, Food & Rural Affairs, UK.

Air Pollution Banding (DEFRA) [54]	**Values**	**Levels of Concern (US EPA)** [55]	**Values**
Low	1-3	Good	0-50
Moderate	4-6	Moderate	51-100
High	7-9	Unhealthy for sensitive groups	101-150
Very High	10	Unhealthy	151-200
		Very Unhealthy	201-300
		Hazardous	301-500

2.1.3. Low-Cost Sensor Technology

Low-cost sensor technologies have emerged as a flexible solution for various monitoring and detection applications, collecting data across fields such as environmental monitoring and healthcare. Their popularity stems from benefits like cost-effectiveness, small size for large-scale deployment, and the ability to gather real-time data. These benefits make them suitable for applications requiring wide coverage and dense data sampling. Researchers have deployed low-cost sensors to monitor air pollution concentrations.

For instance, Li et al. [56] propose an approach in which multiple measurements are integrated. For this data collected for 2,363 AirBox low-cost air quality sensors, 75 fixed monitoring stations and the Terra remote sensing satellite is synergized to estimate the ground-level $PM_{2.5}$ concentrations for Taiwan Main Island during July 14 2018 to Oct 31 in 2018. Despite numerous benefits, low-cost sensors also have certain drawbacks. Compared to more expensive reference-grade sensor devices, low-cost sensors often come with performance trade-offs. Some of the challenges include reduced accuracy and reliability, decreased linearity, and slow response times. Environmental conditions can also affect sensor accuracy, requiring periodic calibration. Therefore, careful consideration is needed when selecting and deploying low-cost sensors to ensure reliable data quality. Below is a summary of different low-cost sensor types used for air pollution monitoring.

A. Electrochemical (EC) sensor

When electrochemical gas sensors come into contact with the target gas, they react with it, thereby producing a measurable potential difference or voltage between two electrodes that corresponds to the concentration of the target gas [57]. They are relatively simple and easy to produce, providing reasonable sensitivity, requiring less power to operate, and environmental conditions like temperature and pressure have less effect on them [58]. Another advantage is that these sensors have a short response time [58], which means that after the sensor is exposed to the target gas, the time required for the electrochemical sensor to move from its baseline signal to a specific percentage of its full response is short [59]. Some studies have indicated that this percentage is typically 90% of the total response, and recovery time, which refers to the time required to return to 90% of the initial baseline signal after the target gas has been removed [60]. EC sensors also exhibit sufficient selectivity [58], meaning they can effectively differentiate the target gas from interfering gas molecules [61]. For measuring air quality monitoring, electrochemical gas sensors are used such as, in work done by Arroyo, et al. [62], where they present a cost-effective IoT system in which a portable device to measure outdoor air pollutants concentration which is based on commercial electrochemical gas and the optical particle matter sensor that is connected to the internet via GSM module to transmitt all real-time data to the cloud platform. Bales, et al. [63] present CitiSense, an air quality monitoring mobile sensor node that utilizes an electrochemical sensor designed to sample air pollutants and environmental parameters. Mead, et al. [64], deisigned portable low-cost electrochemical gas sensing units which can measure CO, CO_2 and NO_2 and calibrated, and then deployed city-wide. Despite these advantages and usage, there are several shortcomings. Cross-sensitivity to other gases may occur, where electrical changes during the detection of the target air pollutant may resemble those of another gas [65]. Ageing affects the sensitivity of electrochemical gas sensors, leading to signal drifts.

B. Metal Oxide (MO) sensors

Metal oxide gas sensors, also known as semiconductor sensors, consist of a surface layer made of one or multiple metal oxides, a heater to heat the membrane and a sensing chip; when the metal oxide interacts with the target gas, conductivity increases, which is then measured by the sensing chip [66]. The small size and low-cost are well-suited for portable and remote sensing applications. Their benefits include long-term stability, extended lifetime, and sufficient sensitivity [67]. As mentioned earlier, these sensors have also been used for air quality monitoring, such as in the work conducted by Bart et al. [68], who deployed low-cost gas-sensitive semiconducting oxide (GSS) sensors to monitor ozone (O_3) with solar power, and automated cell-phone communications and validated with 50 sensors in the Lower Fraser Valley of British Columbia over a three-month period from May to September 2012. Parmar, et al. [69], developed a prototype for an Environmental Air Pollution Monitoring System to measure concentrations of air pollutant gases using low-cost semiconductor gas sensors with Wi-Fi modules. Firdhous, et al. [70], proposed an IoT-based indoor air quality monitoring system with a semiconductor capable of monitoring ozone concentrations, which were installed near a high-volume photocopier. The device was programmed to collect measurements at an interval of five minutes and transmit via Bluetooth device to a gateway node. Metal oxide sensor performance is affected by interference from other gases in its surroundings, and can also be affected by variations in temperature and humidity [71]. Sensitivity towards variations in atmospheric temperature and humidity affects the metal oxide sensor conductivity response to target gases, as it is non-linear, which is a challenge [72]. Certain approaches can be used to address these issues, for instance, applying physical and chemical gas filters to block or delay different non-target gases from reaching the surface of the sensor, preventing them from interfering [73]. Another important issue that can impact the MO sensor's long-term stability is the baseline drift [74]. MO sensors require heating for their activation, and high heating temperatures can also affect their stability [75].

C. Non-Dispersive Infrared (NDIR) sensors

The NDIR sensor works by emitting IR radiation onto the gases contained within the gas chamber, and then identifies the target gas based on its absorption characteristics [76]. The basic components of an NDIR sensor include a gas container or a cell, an IR source, an IR detector, and an optical or light filter. There are several benefits to using NDIR sensors for detecting and identifying gases, for instance, their robustness, high selectivity, and long lifespan [77], making the NDIR sensor measurement more accurate. NDIR have been used for air pollutants monitoring, such as in the work by Mendes et al. [78] who suggested Non-Dispersive Infra-Red (NDIR) based sensors for intensive spatial field monitoring of CO_2 concentrations in the naturally ventilated dairy cow barn. Müller, et al. [79] deployed non-dispersive infrared (NDIR) low-cost sensors to measure CO_2 and assessed them for long-term operation within the Carbosense CO_2 sensor

network in Switzerland. The sensors were calibrated by collocating them with reference instruments in a pressure and climate chamber, as well as under ambient conditions, before deployment. Liu, et al. [80] deployed a NDIR low-cost sensor of the vehicle to observe the on-road CO_2 concentrations during the COVID-19 pandemic in Beijing city. However, there are challenges such as high detection limits, spectral interference, and sensitivity reduction caused by exposure to moisture and non-target gases [81].

D. Optical Particulate-Matter (PM) sensors

Optical sensors operate using the light scattering method, where particles in the sampled air are scattered by laser light, and the scattered light is collected by a photo detector at a specific angle, enabling the sensor to measure the particle size and concentration [82]. Additionally, the optical sensor also includes a fan which directs the airflow, which contains the particles, through the chamber and also contains a set of focusing lenses [83]. Optical sensors are widely used due to their low power usage, affordability, and fast response time [82]. In prior studies, particulate matter (PM) optical sensors have been used to measure PM pollutants, for instance, Li, et al. [56] deployed low-cost PM sensors to measure ground-level $PM_{2.5}$ concentrations. Gao, et al. [21] present Mosaic, a low-cost mobile sensing system to monitor urban air quality. PM sensors were deployed on buses to measure $PM_{2.5}$ air pollutant. Cheng, et al. [84] present implement and evaluate AirCloud, an air quality monitoring system consisting of affordable low-cost PM sensors to obtain $PM_{2.5}$ concentration data. Gao, et al. [85], developed Portable University of Washington Particle (PUWP) monitor, which includes a low-cost PM sensor to measure particle counts. There are several benefits to using optical sensors, however, there are certain drawbacks; for instance, non-target particles (which cause noise) can influence the sensor's accuracy [86]. Additionally, measurement accuracy can also be affected by the reliability of the component used in the sensor, interference from ambient sources and factors that affect the airflow containing the target particles [86]. Moreover, the performance of optical PM sensors can also be affected by high relative humidity conditions [87]. Under these conditions, where relative humidity (RH) levels are high, certain particles that are hygroscopic tend to absorb moisture, increasing their size; therefore, the optical or PM sensors may overestimate particle mass concentrations [88].

We reviewed several existing surveys and studies that examine the performance and capabilities of various low-cost sensors, as summarized in Table 5 below. In addition to evaluating different types of low-cost sensors, we have also conducted a comparative analysis of low-cost sensor attributes in detail, and based on the findings from prior literature review papers, presented in Table 6, see below. Our comprehensive review into the functionality of different types of low-cost sensors highlights their strengths, weaknesses and opportunities. Our analysis outlines key factors that may impact the low-cost sensor performance, including their accuracy, sensitivity, cost effectiveness and limitations under different conditions.

Table 5: Existing related Survey/Review Papers**.**

Ref. Paper	Description
[89]	This study paper reviews and conducts a comparative analysis of commercially available low-cost sensors used for measuring ambient gases. This paper also assesses their potential for large scale coverage and their compliance with the European Air Quality Directive 2008/50/EC.
[90]	This review paper investigates several characteristic of amperometric electrochemical gas sensors which includes their application, development, limitations, challenges, and the contribution in the field for urban air quality monitoring.
[91]	This work evaluates environmental microsensors for monitoring air quality, focusing on improving their performance via machine learning and data modeling, and their ability to meet the European Air Quality Directive standards (2008/50/EC).
[92]	This study paper evaluates the performance of low-cost sensors used for monitoring air pollution and provides scientific guidance to the end-users.
[93]	This study paper reviews and analyzes the low-cost sensor technologies and their machine learning calibration and focus on current research highlighting various challenges and future opportunities.
[94]	This study evaluated the performance of low-cost sensors used for air quality sensors. To conduct this analysis the authors reviewed 112 studies.
[95]	This study provides a brief assessment of low-cost commercial air quality sensor reliability and affordability in comparison to reference measurements.
[20]	This article provides a brief overview of issues related to air pollution modeling, and presents a new dataset of mobile air quality measurements in Zurich. The authors also discuss challenges in interpretating mobile air quality measurements dataset.
[96]	This article provides a comprehensive review of low-cost sensors for air pollution monitoring, LCS calibration techniques, and challenges and limitations in accurate air quality monitoring.
[97]	This literature review examines the applications of low-cost sensing technologies used for monitoring air quality and provides an assessment of their exposure.
[98]	This literature review provides a critical analysis of low-cost sensors used for monitoring air quality, including the potential errors and temporal drift of LCS, and a step-by-step guidance to developing a more reliable and sustainable air quality monitoring setup. Furthermore, this paper also shows opportunities for future improvements.
[99]	This study provides an overview of commercially available chemical sensors that detect volatile organic compounds in both outdoor and indoor environments. This paper discusses sensor designs, their performance parameters and prospects for future improvements.
[100]	This paper studies existing literature on crowd-sourced air quality monitoring and portable sensors. The authors highlight the current progress and future requirements for developing affordable, portable air pollutants monitors, underlining their role in advancing ‘smart cities’.
[101]	In this literature review, the author investigates the working of the conductometric semiconducting metal oxide gas sensors and how different influencing factors affect their sensitivity.
[102]	This report guides the deployment of low-cost air quality sensors, which is based on the AirSensEUR sensor system experience.
[103]	This study paper provides an assessment of metal oxide semiconducting sensors' performance, their structure and operating principles in electronic nose applications. The review also outlines its benefits, limitations, challenges, and implications for odor detection.
[104]	This study explores the use of the Dynamic Neural Networks approach to the stochastic estimation of air pollutant concentrations using chemical multi-sensor devices, demonstrating improved performance in real-world scenarios, particularly during rapid variations in air pollutant concentration.
[60]	This review paper evaluates 1-D nanostructured metal-oxide gas sensors, summarizes their principle mechanisms, their performance, and future research prospects.
[75]	This review paper covers recent developments and advances in the stability of metal oxide semiconductor gas sensors, discussing several factors that influence the sensors, strategies to improve them, important achievements, and future research directions.

2.1.4. Discussion

In dynamic urban environment settings where variations in air quality concentration levels are frequent, the response time of low-cost sensors plays an important role. This is because the temporal resolution of sensor measurements is affected by the LCS response time. This allows for more frequent data points within a specific time interval, leading to more detailed and accurate air quality readings of variation over time. For example, if the mobile sensor measures air quality data every few minutes while being in constant motion, the measured air pollution data may be associated with the location that the mobile sensor passed earlier. Low-cost sensors with quick response times are also affected by environmental factors, such as rapid variations in temperature, humidity, and atmospheric pressure, which cause rapid fluctuations in signals and result in inaccurate readings. Also, sensors with quick response time can require more power, which is an issue for applications where sensors require an independent and limited power supply and may also require more complex calibration methods, causing higher deployment and overhead costs. Additionally, high sensitivity towards other gases and environmental factors may contaminate the collected data, increasing uncertainty. To address these challenges, certain steps are required, such as optimizing sensor network designs, developing more efficient calibration techniques, and deploying physical filters to prevent environmental factors and other gases from interfering with reading. Furthermore, developing advanced data analysis and processing systems to filter out noise from the collected data, while also considering environmental factors, can be applied to improve the quality of low-cost sensor data.

Table 6: Comparisons of different low-cost sensor types, including pros and cons.

LCS Type	**Metal oxide**	**Electro-chemical**	**Non-dispersive infrared radiation absorption**	**Photo-ionization detectors**	**Optical particle counters**
Cost	Low	Low	Moderate	High	High
Accuracy	Low	Good	High	High	High
Sensitivity	Good	Good	Good	Good	Good
Response Time	Long (between 5 and 50 min) [102]	Short (between 30 and 200 s) [102]	Short (between 20 and 120 s) [102]	Fast [102]	Fast (in order of 1 second) [102]
Power Consumption	High	Low	Low	High	High
Selectivity	Poor	Poor	Good	Good	Good
Life Span	Long	Short	Long	Long	Long

2.2. Vehicular Platform

2.2.1. Public transit

Public transit is a mode of mass transport for public use and service that operates primarily in urban areas such as metropolitan cities. Some public transport typically follows fixed or established routes, while others, like taxis, have random routes. Some types of public transport modes provide a wide coverage in urban areas and are commonly available in every major city. In this section, we will discuss three primary types of public transit modes, which include city buses, taxis and trains or trams.

A. Bus Transit

When considering the mobile platform for sensor deployment, bus transit has gained significant popularity, due to its availability in larger numbers, extensive spatial coverage covering most of the urban areas, and reliable and continuous operations [38]. Moreover, bus transit also operates on fixed or set routes and schedules, which are typically available to the public for viewing, making their trajectory predictable [105]. Additionally, city buses also repeat their designated routes several times a day, providing high temporal resolution for data collection for drive-by sensing-related applications. Outfitting an entire bus fleet with sensors for environment sensing, involuntary increases sensing network deployment as well as overhead costs; therefore, considerable efforts have focused on optimizing spatio-temporal coverage using a limited or reduced number of vehicles for sensor deployment or drive-by sensing [11] [14] [23]. Ali, et al. [14] proposed a greedy heuristic approach to select optimal routes for improving sensing coverage by analyzing spatial coverage utilizing the real-bus route dataset of London, UK. Similarly, Caminha et al. [11] analyzed spatial coverage and developed an optimization approach using real-world bus trajectory data from Rio de Janeiro, Brazil. Later, [8], proposed a Modulo system to optimize sensor deployment using a greedy heuristic, taking into consideration aspects like spatiotemporal coverage, budget constraints, and utilized San Francisco and Rome bus datasets. In another study, Paliwal et, al. [23] proposed a regressive facility location approach for drive-by sensing for mapping dense air quality. This proposed approach efficiently selects a subset of vehicles (buses) from the public transport fleet for air quality sampling that incorporates the smoothness in neighbouring locations and autoregressive time correlation for effective spatio-temporal sampling. The authors demonstrated that the proposed problem is submodular, which enables a greedy algorithm with performance guarantees. The system was implemented in the public transport fleet of Delhi, India.

In prior studies, [105] [106], the authors analyzed historical trajectories of buses to select the optimal set of buses to maximize spatio-temporal coverage. The overlapping of bus routes is common, which may lead to redundant data collection, however, this overlapping nature in routes can be utilized for cross-checking

readings from individual buses and mutual sensor calibration [107], for instance, for single-hop and multi-hop calibration. However, the work on the calibration of bus-mounted sensors is currently limited [108]. There are several interesting challenges, including determining the geolocation for deploying fixed reference grade sensor stations based on real bus transit data, which includes routes and schedules. Moreover, calibration performance requires evaluation and considers the trade-offs between maximum calibration opportunities and cost. Another challenge is that the fixed routes and schedules of bus transit restrict their mobility, making them less flexible. This restriction means that drive-by sensing operations are limited to only pre-defined bus routes and schedules. To achieve more flexibility, if required, other vehicular platforms such as taxi cars and unmanned aerial vehicles can be utilized. Further discussion on alternative modes of vehicular platforms for drive-by sensing is in the following sections.

B. Trains and Trams Transport

Similar to bus transit, trains and trams also operate on fixed, pre-defined routes and schedules, and repeat their journey several times daily on the same route. Their pre-defined routes make their trajectory predictable, and repeated journeys provide high temporal resolution similar to bus transit. Moreover, trains and trams operate on dedicated lines or infrastructure, free from traffic and traffic signals at junctions. Therefore, drive-by sensing operations are unaffected by road traffic delays or other related challenges. A well-known project, OpenSense, deployed sensors on trams in Zurich and on buses in Lausanne to monitor air quality in real-time [109]. However, they are not available in every city and do not provide a wide coverage, thereby introducing a spatial bias in air quality monitoring. Their fixed trajectories and limited coverage to specific urban areas restrict data collection to these locations, thereby making it difficult to monitor air quality elsewhere. To achieve greater flexibility, if required, other vehicular platforms, such as taxi cars or unmanned aerial vehicles, can be utilized.

C. Taxis

Similar to bus transit, taxis are also available in most cities, providing wide coverage in urban regions. This is because they are available in large numbers, providing high spatio-temporal resolution. They do not have fixed or pre-defined routes; therefore, they have more flexible mobility. In several prior studies, taxis have been used as sensor platforms, including [110] [111] [112] [113]. Wang, et al. [110] deployed low-cost sensors on a fleet of taxis to measure key air pollutants, CO, NO2 and O3 and utilized an advanced calibration algorithm to reduce sensor data errors and uncertainties. The authors deployed the taxi-based low-cost sensors to study urban air quality over the course of a year and to assess the impact of COVID lockdowns in Nanjing, China. To map key air pollutants, they employed GIS technology to generate high-resolution spatial distribution (50m x 50m), to identify air pollution sources and to analyze traffic-related emission patterns. Other vehicles available can also be used in combination to add value, for instance, in a

study by Yeom [111] to measure real-time air quality and visualize high spatial resolution, where low-cost sensors were deployed on top of taxi roofs and SUVs, and the study took place in Daejeon, Korea.

Taxis do not have pre-defined routes, and their movement or trajectory is random, making them unpredictable. Furthermore, they are more likely to cluster around areas where there is more human activity, such as near shopping malls, airports, train stations, etc., creating a spatial bias [112]. Additionally, taxi mobility is influenced by several other factors such as road traffic conditions, drivers' routing preferences for selecting favourable routes, and passengers' routing preferences, making taxi trajectories random and unpredictable [113]. The unpredictable and spatially biased movement of taxis can lead to inconsistent data collection in some parts of the urban areas. This inconsistency results in sparse data collection and time variations in coverage [114]. This presents a challenge for sufficient spatio-temporal coverage. To address this issue of sparse and time-varying data coverage, Xu et al. [114] developed an adaptive hybrid model-enabled sensing system (HMSS) to optimize the quality of sensing coverage and to provide estimates of air pollution data in detail. The authors collected over 53.5 million air pollution data samples during a 14-day period from 47 self-designed portable sensors, deployed on taxis and at fixed locations, to evaluate the system's performance. The tests were conducted in two cities, Foshan for controlled and in Shenzhen for uncontrolled experiments for performance evaluation.,

2.2.2. Municipal Transport

In addition to buses, municipal transport also operates in cities, such as dump trucks for waste management, delivery vans, and ambulances for patient transport, which are generally used for public services-related tasks other than public transportation. High spatial coverage depends largely on the type of public service vehicle used for drive-by sensing. For example, police patrol cars provide broad coverage, while emergency vehicles such as fire trucks and ambulances offer limited coverage, as they are only used in emergency situations and do not operate during non-emergency situations. However, vehicles such as delivery vans and dump trucks provide sufficient spatial coverage, covering most commercial and residential areas along their routes. Although such vehicles also exhibit biased behaviour since their route trajectory depends on the destination, which can vary occasionally, and the driver's decision-making. Their operations are also intermittent, stopping at certain points for a time period and suspending or terminating operations when the task is completed. In addition to their biased behaviour and intermittent operations, their limited fleet sizes result in providing low spatio-temporal coverage, leaving sparse data coverage. Qin et al. [115] mounted low-cost sensors on 17 postal vans whose sampling routes were relatively random, in Antwerp, Belgium, and proposed a model for mapping fine-grained urban air quality and used sparse NO_2 measurements to validate the model. The sampling campaign typically operates on weekdays and Saturdays from 6:00 to

23:00, and sampling intervals during the day were set to 10 seconds. Sampling operations during night-time, when the postal vans were parked, the sampling intervals were increased to 10 minutes.

2.2.3. Private Transportation

Vehicles that are privately owned and not accessible to the general public for usage are referred to as private transportation. There are several types of private transportation vehicles that can be selected for low-cost sensor deployment in drive-by sensing applications, including cars, bicycles, sub-utility vehicles (SUVs), vans, etc. There are certain benefits to using privately owned vehicles, as the user can decide on operational timings, schedules, and routes at the user's convenience. However, there are drawbacks as well, which are discussed in this section.

A. Dedicated vehicles

Private vehicles which are modified specifically for a particular drive-by sensing application only are referred to as dedicated private vehicles. For instance, vehicles deployed with air pollution measurement sensors operate for the specific role of collecting air pollution data from various locations. An example of this is, researcher equipped two Google Street View cars with fast-response, research-grade instruments to collect nitric oxide (NO) and black carbon measurements in Oakland, California [116] [12] [117]. A similar dedicated system was used in the 'Breathe London' project, where Google Street View cars were equipped with sensors to collect NO, NO_X and CO_2 data at 1-second intervals between September 2018 and October 2019, in London, UK. Chiesa et al [118] proposed an approach where a diffused use of low-cost sensors designed to measure atmospheric gases, particulate matter and other atmospheric physical parameters was deployed on the roofs of vans. The experiment gathered data in Rome, Italy and presented the subsequent processing results. Algorithms were utilized for data fusion from different simulated monitoring systems and for data interpolation for developing a more geographically dense air pollution map, which was then produced and integrated into the CIPCast platform.

B. Personal vehicles

Personal transport vehicles provide flexible mobility as their routes and schedules can be tailored to the user's requirements or modified to accommodate the drive-by sensing application. However, the personal transport vehicles are commonly owned and operated by private individuals. Therefore, the route planning and operational timetable are determined by the owner or operator, which can introduce spatial bias and random movements. In HazeWatch [13], the researchers deployed low-cost sensors on cars for crowd-sourced fine-grained spatial measurements of air pollutants CO, NO_2, and O_3 in Sydney. In a study, Wesseling et al. [119] deployed 500 dust (PM2.5) sensors on bicycles to estimate the PM2.5 levels that the cyclists are exposed to in the province of Utrecht, the Netherlands, in the year 2020. To calibrate and validate the data from mobile sensors, they are collocated at reference stations. Similarly, Gómez-Suárez,

et al. [120] installed low-cost NO_2, O_3 and particulate matter (PM) sensors on 8 bicycles to monitor air quality in Badajoz, Spain, over several days.

C. Unmanned Aerial Vehicles (UAV)

Unlike ground vehicles, which are limited by road networks and traffic, UAVs can travel directly to their destinations or follow arbitrary paths for data collection. One of the benefits of employing UAVs for measuring air pollution or other drive-by sensing applications is that they can collect data from areas that are typically inaccessible via ground due to difficult or no pathway, such as landfills [121]. UAVs are also an ideal choice for measuring at high altitudes, particularly in areas designated as hazardous or dangerous for human safety. Despite these benefits, there are some drawbacks; for example, the measurements for the sensors deployed on UAVs may be influenced by the wind generated by the rotors [122]. Prior studies have addressed this issue [123] by analyzing the UAV's structure. Arroyo, et al. [124] deployed low-cost air quality sensors on a UAV that can measure from on-the-move and stationary positions that are difficult to access. The electronic instrumentation has been optimized to minimize the interference from different devices. The design includes a pneumatic pump and a tube for sampling air to prevent or reduce the influence of the propellers. The sensors deployed on the UAV were calibrated by co-locating with certified reference stations. The results show that the accuracy of the prototype measurements were not affected by the movement of the drone, its downwash effect, or electromagnetic disturbances. The communication signal range between the controller and the UAV can be a challenge, as the UAV may lose control when it goes out of range. To address this issue, communication through satellite and telecommunication links can potentially increase the range of signals.

2.2.4. Discussion

Beyond their mobility, one crucial advantage of using vehicles as sensing platforms is that the sensors can access the vehicle's power system, making them less limited in energy and physical size compared to mobile sensors in other settings [125]. Spatial and temporal coverage are key factors when determining the vehicle platform for sensor deployment. Among the diverse options, bus transit is particularly suitable due to its predictable routes, large fleet size, and repeated operations throughout the day, enabling high spatial and temporal resolution for sensing applications. In contrast, taxis offer flexible and widespread coverage but are prone to spatial bias and unpredictability, often clustering around areas of high human activity. Trains and trams also provide predictable and frequent movement, but their fixed tracks limit spatial coverage to specific corridors. Municipal vehicles such as delivery vans or dump trucks can reach many urban areas, though their route patterns are task-driven and intermittent, which may lead to sparse data collection. Private transport platforms, including dedicated sensing vehicles or bicycles, offer flexible routing but often suffer from user-dependent randomness and limited operational consistency. UAVs, while excellent for

monitoring inaccessible or hazardous areas, are generally unsuitable for continuous near-road sensing due to regulatory, technical, and environmental constraints. Therefore, the choice of a vehicular platform must carefully consider context-specific requirements, balancing spatial and temporal coverage with operational feasibility, cost, and data quality.

2.3. Spatio-Temporal Resolution

Deployment of a drive-by sensing system for air pollution monitoring or other applications requires balancing between spatial and temporal resolution. Stationary sensors measure continuous real-time data, providing high temporal resolution. However, the drawback of the fixed location restricts their spatial coverage. Balancing detailed temporal data and the requirement of achieving broader spatial coverage from the stationary sensors is a challenge in air quality monitoring. To improve spatial resolution, spatial interpolation techniques can be used to address these limitations by applying mathematical models to estimate the air pollutant values in unsampled geo-points. For instance, Joseph et al. [126] used simple spatial interpolation methods to estimate 8-hour ozone concentrations at unsampled locations between monitoring stations in two urban areas. Tong et al. [127] performed an analysis of different kriging interpolation techniques using the air quality index in Wuhan and assessed their effectiveness. Recent studies have analyzed machine learning and neural networks as alternative methodologies to traditional mathematical and statistical models. For instance, Wahid et al. [128] proposed a neural network-based meta-modelling approach to estimate the spatial distribution of ozone concentrations in the Sydney basin, Australia. Pfeiffer et al. [129] present a novel approach based on diffusive sampling measurements and artificial neural network evaluation to calculate the average spatial distribution of air pollutants. For accurate calculation of the spatial distribution of air pollutants values, environmental conditions or factors must also be taken into consideration, as Nicoletta et al. [130] used simple dispersion models to examine the effect of wind direction and intensity on the results of the air quality monitoring of a spatial interpolation technique. Korunoski et al. [131] introduced an Internet of Things-based sensing system architecture which includes air pollution measurements and meteorological parameters for intelligent air pollution visualization and future pollution prediction.

On the other hand, drive-by sensing systems offer high spatial resolution or coverage due to the mobility of the sensing platform, enabling a more detailed spatial distribution map of air pollutants across different geographical locations. Drive-by sensing can be used to identify air pollution hotspots and spatial trends. Several prior studies have deployed low-cost sensors on various types of vehicles, such as buses, to achieve maximum spatial coverage, for instance, Wang et al. [110] deployed low-cost sensors on a fleet of taxis to measure key air pollutants CO, NO2 and O3 and utilized an advanced calibration algorithm to reduce sensor data errors and uncertainties. The authors deployed taxi-based low-cost sensors to study urban air quality

over the course of a year and to assess the impact of COVID lockdowns in Nanning, China. To map key air pollutants, they employed GIS technology to generate high-resolution spatial distribution (50m x 50m), to identify air pollution sources and to analyze traffic-related emission patterns. Yeom [111] to measure real-time air quality and visualize high spatial resolution, where low-cost sensors were deployed on top of taxi roofs and SUVs, and the study took place in Daejeon, Korea. However, the continuous movement of mobile platforms restricts their temporal resolution, as continuous data sampling at a single geo point is challenging. To address this problem, the number of mobile sensors can be increased, which would allow for more frequent data sampling at each location [119], improving the temporal resolution; however, this will increase the deployment and the overhead costs. An alternative approach to this problem would be to combine collected data from mobile sensors with the data of stationary sensors, where the stationary sensors will provide high temporal resolution and mobile sensors will provide high spatial resolution [132].

2.4. Mobile Sensor Calibration

The accuracy and stability of low-cost sensors are affected by various internal and external phenomena, including sensor component age-related temporal drift and extreme variations in meteorological conditions [96]. Due to these factors, periodic calibration of low-cost sensors is necessary, where the systematic bias in sensor readings is determined and corrected [28]. Calibrating low-cost sensors deployed on mobile platforms is challenging due to their dynamic nature, which necessitates the suspension of drive-by sensing operations to facilitate co-location with a reference station for sensor calibration [16]. There are certain calibration approaches that do not require suspension of operation, for instance, a single-hop approach where mobile sensors are calibrated by reference stations by a phenomenon known as rendezvous [27]. For rendezvous to take place, both the mobile sensor and the reference sensor must be spatially and temporally sensing the same air pollutant [133]. Several prior studies have employed a single-hop approach to calibrate mobile sensors using reference stations or other sensors, for instance, [27] [11] [32]. Single-hop calibration of mobile sensors offers a straightforward and direct calibration approach, requiring at least one reference station or sensor to be deployed along the route of the mobile sensor, as illustrated in Figure 1. However, reference stations or sensors are typically expensive to purchase and deploy; therefore, deploying a reference station on every route is not a scalable solution. Selecting an optimal set of reference stations or sensors to make the on-field calibration of mobile sensors is required [16]. Fu, et al. [134] proposed an algorithm to optimally design a reference sensor deployment to ensure that the sensors, network-wide are *k*-hop calibratable if the calibration path is no larger than *k,* and to evaluate the proposed algorithms' performance, real-world data sets were used. In prior studies, work on optimizing the placement of sink nodes has been done, which can be used to optimize reference stations placement for single-hop calibration, for example, Keung, et al. [135] studied the delay-constrained information coverage problem in mobile

sensor networks, and presented a new formulation of this problem which aims to optimize sink nodes placement according to data collection time constraints.

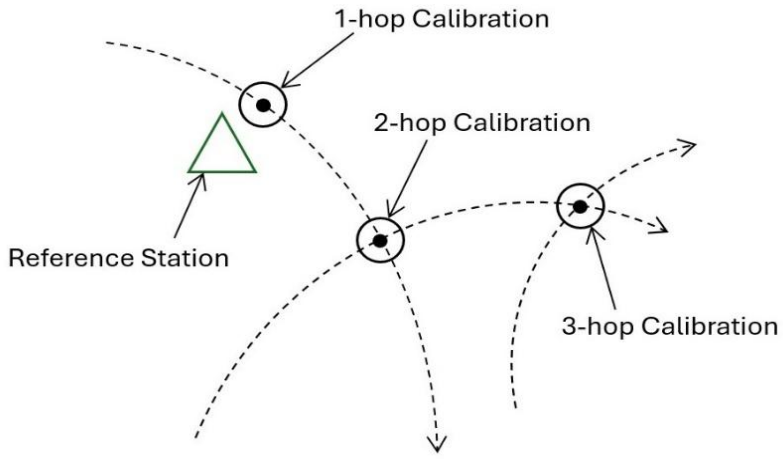


Figure 1: Single-hop and multi-hop calibration methods for mobile sensors, using reference stations (green triangles) and peer-to-peer rendezvous for k-hop calibration.

An alternate approach to the single-hop calibration approach is multi-hop calibration, which involves calibrating using rendezvous where freshly calibrated mobile sensors calibrate other uncalibrated mobile sensors in the wireless sensor network [136]. Results in work done by Xiang, et al. [137] suggest that for ultrafine particle monitors and black carbon monitors, the optimal intervehicle rendezvous calibration distances are 10–80 m and 0–30 m, respectively. Benefits of utilizing multi-hop approach include placement of reference station or sensor on each route operating mobile sensor, which significantly reduces the number of reference grade sensors required.

Markert, et al. [138] used the combination of private proximity tests for calibration with an anonymizing MIX network via sensor rendezvous to introduce a privacy-preserving multi-hop sensor calibration scheme for air quality sensing. Yu, et al. [139] have developed a Gaussian Process Regression-based end-to-end inference calibration solution: Spatial-Temporal Inference Calibration Model (ST-ICM), to tackle reference data scarcity and restricted calibration confront. ST-ICM created calibration grids with different levels of confidence and then used a variance threshold to choose only the high-quality ground drafts. Authors demonstrate the proposed method on a real-world dataset in Nanjing, China and show that compared to the baseline, this method yields a 24% improvement in error calibration. The multi-hop calibration methods provide mobile wireless sensor networks with more flexibility to extend the distance of deployment; however, error accumulation over the hops may occur [140]. To remedy this, Saukh, et al. [136] introduced a novel multi-hop calibration algorithm that minimized error propagation in the wireless network by geometric mean regression. The results show that the proposed system has better performance and is more

efficient than existing calibration algorithms, not only in a multi-hop scenario but it also need fewer number of reference measurements to have the desired estimation. Maag, et al. [140] introduced SCAN: a new application of constrained multi-dimensional linear regression for sensor array calibration in mobile systems to improve measurement accuracy across multiple hops. Maag, et al. [141] proposed an orthogonal approach to, (i) selecting reliable measurements for calibration at every hop, (ii) estimating the data-induced uncertainties at a high level, and inventing an uncertainty-based metric for data filtering. Real-world ozone sensor deployment was used to test the proposed methods. Experimental results show that the approach can be used with both linear and non-linear models, leading to up to 25% improvement in calibration performance on multi-line-of-sight scenarios as compared to other alternate techniques.

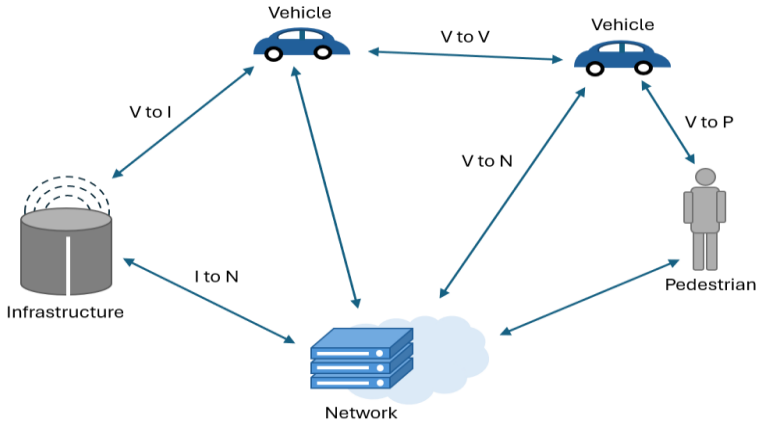


Figure 2: Wireless network platforms for sensor data transmission via Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I), and Vehicle to Network (V2N) communication.

Vehicle-based platforms have dynamic trajectories. which means that the wireless network topology is continuously changing. This could lead to disconnection in communication links between sensors and servers, where collected data is processed and stored [142]. To prevent this and improve the overall reliability of vehicle-based sensor networks by implementing vehicle-to-vehicle (V2V), vehicle-to-network (V2N) and vehicle-to-infrastructure (V2I) architectures. In several prior studies, for instance in [143] [144] [145] [146], various methods have been developed which are based on the use of V2V and V2I communication. These architectures can function as a wireless multi-hop network, where sensory data is transferred from V2V to V2I or V2N, as illustrated in Figure 2. These architectures can also reduce the cost of deployment and maintenance, enable lower power consumption, and allow for smaller data latencies. For multi-hop calibration, the routes of the mobile sensors must intersect or, more preferably, overlap with each other. Similar to the path of mobile sensors, mobile sensors must also intersect or overlap with the reference station's position in order to calibrate them through a single hop. Conclusion – Hence, it is crucial to select the most suitable mobile platform that best meets the required criteria for wireless sensor

deployment. Mobile platforms with pre-defined routes and schedules have a predictable trajectory and behavior which meet the set requirements mentioned earlier. For example, bus transit provides predefined and fixed routes and schedules, which provide higher opportunities for rendezvous with reference stations placed in routes and with other recently calibrated mobile sensors..

2.5. Optimization Problems

Optimization problems aim to find the best possible solution from all possible solutions and are usually applied when the optimal solution is unknown, and they are typically used for maximizing and minimizing [147]. Optimization problems have been applied in many areas such as engineering, economics, and environmental studies, as the problem becomes more complex. Based on the nature of the objective function, constraints and decision variables, we can broadly categorize optimization problems. Optimization problems can be further categorized into exact and heuristic algorithms. Exact Algorithms: As the name suggests, the exact algorithms produce an exact solution for a problem. Approximation (Heuristic)–In this case, the algorithms are heuristic, which means that the solutions generated by them are not exact and may not even lead to an optimal solution. The heuristic approach is more useful in solving more complex and large-scale problems that are more computationally intense, making the process more feasible [148]. It is applicable to most optimization problems, and here we review some of the practically done algorithms, such as greedy algorithms, neural networks, ant colony optimization (ACO), local search, and tabu search. Meta-heuristic algorithms are approximate and usually non-deterministic, and are also used to solve complex optimization problems where a near-optimal solution may be required. Some examples of meta-heuristics, including genetic algorithms, simulated annealing and particle swarm optimization are flexible and have been widely used for different applications, which makes them adaptable to solve a variety of optimization problems. A comprehensive review of optimization and algorithms provides a detailed understanding of these categories and can be helpful in decision-making for the most effective optimization algorithm for a given problem.

2.5.1. Linear and Non-linear Programming

A. Linear programming (LP)

Linear programming (LP) is a mathematical technique for optimizing a linear objective function subject to a set of linear equality and inequality constraints [149]. It is used in a range of applications, such as profit or cost functions, and more by maximizing or minimizing an objective function subject to constraints [150]. Supply chain management, transport logistics, and manufacturing utilize linear programming (LP) for efficient resource allocation to enhance operational efficiency. Taylor [151] used linear programming to solve the global constrained optimization problem with preferences specified for area chairs and affinities.

This optimization (linear program) is performed subject to a set of linear constraints, which essentially define the limitations or requirements of the problem, e.g. availability of resources or maximum capacity available. A typical linear programming problem consists of a linear objective function to be optimized (either maximized or minimized) subject to a set of linear constraints on the decision variables [150].

There are several limitations to this, including the linear assumption, which does not always accurately reflect real-world relationships that tend to be non-linear. In addition, LP assumes that all decision variables have continuous values, which may not be suitable for problems that require discrete or integer solutions. For non-linear relationships or when solutions have integer constraints, alternative methods such as nonlinear programming and integer programming may be more suitable. Linear programming has been widely employed in the optimization of sensor placement, resource allocation and monitoring schedule for air quality monitoring. For instance, in a study [152], a linear programming model is developed to evaluate the economic feasibility of air pollution control measures. The model is applied to the St. Louis Airshed to assess sulfur content restrictions in coal and to recommend efficiency improvements for furnaces. Lamego, et al. [153], describes the application of linear programming in receptor modeling for identifying the sources of air pollution, specifically to determine the contribution of industrial activity to atmospheric aerosol pollution in the Metropolitan Region of Vitoria (MRV), which is a densely populated area in Southeast Brazil.

B. Integer Linear Programming (ILP)

Integer Linear Programming is an optimization technique that uses integer variables for optimizing (maximizing or minimizing) an objective function [154]. Integer programming allows the use of integer constraints, and linear programming uses continuous variable constraints; however, in scenarios where discrete or binary (0 or 1) or mixed-integer decisions are required, linear programming is not suitable [155]. Optimization problems in drive-by sensing have been solved using integer linear programming. For example, Agarwa, et al. [8] present a city-scale drive-by sensing system where an integer linear program was used to formulate the problem of selecting the optimal set of vehicles. Ji, et al. [42] formulated the bus line selection sub-problem as a linear integer program. The authors provide a real-world case study with a radius of 400 km^2 in central Chengdu, so the application of this model can result in solving large-scale area issue successfully. Integer Linear Programming (ILP) is employed in methods for reference sensor placement for both calibrating mobile sensors and coverage. For example, Fu, et al. [134] formulate the reference sensors placement problem as a set cover problem and use an integer linear program to address the problem. The objective is to reduce the number of reference sensors required to ensure that the entire network is k-hop calibrable, and use a real-world dataset to evaluate the performance of the algorithms. Rebai, et al. [156] utilized Integer Linear Programming (ILP) to address the optimization problem of

deploying the minimum number of sensors that are required to provide coverage to the sensing area and still maintain their connectivity with each other. The model has been developed to address the 2-dimensional critical coverage problem. Xu, et al. [157] formulate an integer linear programming problem to deploy sensors at minimal cost while guaranteeing target coverage, and propose a greedy heuristic for solving the problem.

C. Mixed-Integer Linear Programming (MILP)

Mixed-Integer Linear Programming (MILP) are another type of LP that addresses optimization problems in which some variables are constrained integers, and some can be continuous [158]. MILP can be applied to a wide range of applications, such as planning [159], scheduling [160], and route selection [161], etc. Kotb, et al. [162] proposed a new system for smart parking based on mixed integer linear programming that minimizes the monetary cost and maximizes parking resources. Several algorithms like branch-and-cut, branch-and-bound, cutting plane method and heuristic and metaheuristic algorithms like genetic and greedy algorithms, etc., are also used to solve MILP problems. Guo, et al. [161], developed a mixed integer linear programming (MILP) model for a multivehicle routing problem with bus stop locations and passenger assignments. For a numerical example, this study compared the genetic algorithm (GA) and the branch-and-cut algorithm.

MILP has also been applied to solving optimization problems in the planning of vehicle trajectory or path planning. Waen, et al. [163] proposed using mixed integer linear programming to develop a scalable UAV trajectory planning system. The system divides the problem into smaller MILP sub-problems using two sets of heuristics, each sub-problem models a segment of the trajectory. The two heuristics solve the problem in a sequence, the first minimizes the sub-problem difficulty and ensures stability, while the second limits the selection of obstacle amount to be modelled in each sub-problem, ensuring consistency. This model was used in real-world maps of two cities with trajectories ranging over several kilometers to demonstrate its scalability. Trotta, et al. [164] outlined a network design and optimization framework for city-wide Points of Interest (PoI) monitoring by UAVs. This consists of UAV-to-bus recharging and MILP-based route optimization. The UAV scheduler is formulated as a Multicommodity flow problem. Richards, et al. [165] introduces a multi-aircraft trajectory optimization to prevent collision, based on mixed-integer linear programming (MILP). With the linear constraints, they developed an approximate model of aircraft dynamics that can be used for collision avoidance with the MILP approach. Multiple waypoint path-planning can also be included in the formulation, where a set of points are required to be visited by each vehicle in order chosen within the optimization.

MILP has also been applied to sensor deployment optimization in a wireless sensor network. Xiang, et al. [166] proposed a mixed integer linear programming (MILP) based placement algorithm of fixed sensors in a hybrid sensor network for collaborative calibration of mobile sensors. Boubrima, et al. [167] formulated a mixed integer linear programming (MILP) model for the optimal deployment of air quality sensors, reducing costs while preserving network connectivity and data assimilation. To solve this problem efficiently, a heuristic algorithm was employed. Güney, et al. [168] proposed a new MILP models to find the optimal sink location and information flow path of the sensor to sink when sensor locations are given. The first group of models focuses on minimizing routing energy while the second group of models focuses on minimizing the total cost. Caminha et al. [11] proposed a spatial coverage model of a bus-based mobile wireless network leveraging fixed bus trajectories and proposed an MILP formulation to identify the minimum subset of buses that should be equipped with mobile sensing nodes while maintaining the same level of coverage as the complete fleet. They utilized a large-scale mobility dataset from Rio de Janeiro, Brazil, to evaluate their model. After reviewing various studies, Mixed Integer Linear Programming (MILP) proves to be an effective approach for solving complex optimization problems, particularly in the context of air quality management. MILP can handle complex problems and utilize both integer and continuous variables, making it a flexible approach.

D. Non-linear programming

Non-linear programming is an optimization method where the objective function or constraints are non-linear [169]. Prior research has studied nonlinear issues surrounding air quality network management such as network optimization and the cost-based selection of pollution control strategies employing nonlinear models. For example, Carnevale, et al. [170] propose a two-objective optimization approach to reduce the costs of emission reduction and increase air quality through artificial neural networks modelling between emissions pollutants and air quality. Pareto optimal solutions are calculated using Weighted Sum Strategy, and the proposed method has been applied in Northern Italy to a complex domain, including the Milan metropolitan area. Trujillo-Ventura, et al. [171] addressed the problem of optimizing a nonlinear, nonconvex, and nondifferentiable model for an air pollution monitoring network. This involves Simulating pollutant concentrations, applying kriging interpolation and integrating the interpolation errors using an adaptive quadrature. The Hooke and Jeeves Discrete Step optimization algorithm is used to solve this model. Shishegaran, et al. [172] proposed linear and non-linear statistical models for predicting AQI in Tehran. The four models used are Auto Regressive Integrate Moving Average (ARIMA) as a single time series method, Principal Component Regression (PCR) as a hybrid regression model, the hybrid model that combines ARIMA and PCR with a neural network derivative Gene Expression Programming (GEP) and a combination of ARIMA and GEP to predict daily AQI in four seasons over the period 2012–2016.

2.5.2. Combinatorial Optimization

Combinatorial optimization is a type of mathematical model which is used to find optimal solutions from a finite set of discrete alternatives options [173]. Some common combinatorial optimization problems include the Traveling Salesman Problem, which seeks to find the shortest route that visits a set of given cities [174], the Knapsack Problem involves selecting the items with the objective to maximize total value within a weight limit given the weights and values of the items [175], and Graph coloring is the problem of assigning colors to each vertex of a graph such that no adjacent vertices share the same color [176]. Chen, et al. [177] investigated the problem of how to maximize the quality of coverage in a mobile crowd-sensing network for air quality monitoring within a given budget. They refer to the problem as Maximizing Coverage with Limited Budget (MCLB) problem, which they prove is an NP-Complete problem by converting it into a knapsack problem. A greedy algorithm was proposed to solve the MCLB problem. Zarrar, et al. [32] aim to determine the optimal subset of bus stop locations to maximize total sensing area, and formulate the sensor calibration problem as a knapsack problem.

2.6. Optimization Algorithms

Optimization algorithms are fundamental for solving complex problems in many areas, such as engineering, economics, artificial intelligence and logistics. Those algorithms are used to find the optimal solution (or near-optimal solutions) by improving the existing objective function while adhering to constraints. See Figure 3, on page 50, which shows the difference between a near-optimal and an optimal solution. In general, there are two types of optimization techniques based on the manner in which they approach finding the solution: exact algorithms and approximate algorithms. Exact algorithms solve optimization problems by systematically exploring all possible solutions (or a reduced search space using mathematical techniques) to guarantee the discovery of the optimal solution, even when it is not known in advance. Such methods will provide an optimal output, but they may not be feasible for large or complex problems as they demand heavy computation. Branch and Bound, Dynamic Programming, and Simplex methods are well-known exact methods that work best when higher accuracy solutions are required and sufficient computational resources are available. In contrast, the objective of the approximate algorithm is to find near-optimal solutions by sacrificing accuracy for computational efficiency. Approximation approaches are particularly beneficial when solving large-scale problems, as exact approaches are very slow or resource-intensive for such problems. Heuristics and metaheuristics like evolutionary approach, genetic approach, and particle swarm optimization are approximate techniques which provide practical solutions used in real-world scenarios where exact solutions cannot be obtained easily due to time or resource constraints.

2.6.1. Exact Algorithms

Exact algorithms ensure the determination of the optimal solution to an optimization problem. Methods such as branch and bound, dynamic programming, and the simplex method for linear programming (LP) systematically investigate the solution in detail. The computational complexity of applying exact algorithms can effectively solve small to medium-sized optimization problems. However, larger and more complex scheduling problems would require higher computational resources.

A. Branch and bound

The Branch and Bound technique is used for the solution of optimization problems, especially in combinatorial optimization, integer programming and non-linear problems. It begins by segmenting the problem into subproblems called branches and removing sections of the branches based on the bounds on the optimal solutions. This elimination of branches continues until the required optimal solution is required. Branch and bound can provide optimal solutions without requiring an exhaustive iteration through all possible solutions. By removing or pruning branches that do not provide the optimal solution, the algorithm is computationally efficient, requiring less time and resources compared to methods that consider all possible solutions. The branch and bound method is more suitable for solving problems where finding exact solutions is required, such as the knapsack problem, the travelling salesman problem, and integer linear programming. There is significant literature and research available where branch and bound algorithms have been applied to address complex optimization problems and have been used in a wide range of applications, which also includes sensor placement optimization. Zhao, et al. [178] proposed a branch-and-bound algorithm to address the problem of optimizing sensor placement in water networks that combines greedy heuristics with convex relaxation, minimizing the consumption of contaminated water before contamination detection. By repeating sampling of random subsets of contamination events, the sensor placement strategies have improved in comparison to previous studies. The authors in [179], developed a branch-and-bound algorithm to find optimal scheduling to address a three-agent scheduling problem with an overlap effect. The aim is to reduce the total tardiness of jobs from agent 3 while ensuring that $PM_{2.5}$ emissions do not exceed a specified limit from agent 1, so that agent 2 maintenance activities can be accomplished within the set time period. The authors also performed computational experiments to highlight the quality of the solution and speed of the proposed algorithm. Lee, et al. [180] suggested an imitation learning approach to speed up the branch-and-bound algorithm. They introduced an auxiliary pruning policy developed through training a deep neural network with a mixing training strategy and a novel loss function. This method is designed to accelerate the branching process in the branch-and-bound algorithm while managing the trade-offs between optimality and computational complexity for device-to-device communications. Mansourkiaie, et al. [181] proposed an optimization framework to minimize the

probability of collision using cooperative routing in a wireless sensor network. For finding optimal routing, a set of relays and power allocation for cooperative transmission links by using a branch and bound algorithm with reduced search space. To further reduce computational complexity, a near-optimal algorithm was introduced, designed to decouple power allocation from route selection and solve the problem using a branch-and-bound space-reduced method. In comparison with existing schemes, they presented algorithms simulation results show that the collision probability can be significantly reduced.

Branch and Bound success is necessarily dependent on the quality of the bounding functions and on the structure of a problem, although it is very effective when applicable. Inefficient pruning over large-scale problems can lead to computational costs and other complexities. Despite this, it has still remained a widely used exact algorithm in optimization, as it combines the advantages of systematic search and intelligent pruning.

B. Dynamic programming

Dynamic programming is an algorithmic technique that solves complex problems by segmenting the problem into simpler subproblems. Dynamic programming is only used under two conditions: first, the problem has an optimal structure where the optimal solution can be achieved by combining solutions from smaller subproblems. Second, it has overlapping subproblems, where the same subproblems are solved multiple times. It prevents unnecessary calculations by solving the subproblem (memorization) and storing it only once in a sequence (tabulation), thereby providing solutions efficiently and can be used to solve a wide range of complex problems. This, in particular, makes it very efficient for recursive problems, such as the Knapsack Problem, the Fibonacci sequence, and shortest path algorithms like Dijkstra's algorithm. Several prior studies, Kumar, et al. [182] developed a dynamic programming (DP) model to address air pollution control in urban areas. The model determines emission reduction technologies based on technical and policy restrictions. The DP method is applied to a simulated city with three principal sources of pollution, and it aims to minimize the total costs. This study shows that the DP is suitable for the optimum control strategies when considering multiple sources, technologies, and a single pollutant. Ono, et al. [183] formulated the problem as an optimization one and used the dynamic programming technique to determine the optimal location of monitoring posts on a highway, considering the social impact on nearby areas. The objective function is defined as the sum of the products of vehicle emissions and social indicators of adjacent areas, extending the concept of the dosage-population product. A case study on the Osaka-Nishinomiya Line of the Hanshin Superhighway demonstrated the effectiveness of the proposed method in planning the optimal placement of monitoring posts. Anastasiadis, et al. [184] introduced a new network model for pollution control, using game theory and focusing on computational aspects. The problem, modelled on graphs, represents agents as pollution sources and the government as the regulator for setting

pollution limits. They proved the NP-hardness of the problem and applied approximation algorithms on trees and planar graphs, which model pollution in water and air, respectively. For trees, they developed a two-level dynamic.

Dynamic programming has several disadvantages. Dynamic programming uses recursion where it repeatedly solves smaller subproblems, which requires more memory in the call stack, and leads to a stack overflow condition in the runtime. It takes a large amount of memory to store the solutions of each sub-problem and not all solutions stored will be used later during implementation. High memory usage might also lead to degraded performance depending on the dynamic programming algorithm and the programming language.

C. Simplex method

The simplex algorithm was introduced by George B. Dantzig in 1947 (Dantzig, 1990). The simplex method is a systematic process to solve linear programming problems which begins by finding an initial feasible solution at one of the vertex (corner point) of the search space, which is defined by the linear constraints [185]. Iteratively explores the search space, visiting and testing the adjacent vertex, each time finding an improved objective function value (either maximum or minimum), until the optimum solution is found in one of the vertices [185]. The Simplex Method is very effective when solving large-scale linear programming problems across various fields. such as logistics, resource allocation, and financial planning, where decisions must be made within specific constraints. It eliminates the need to evaluate all possible solutions by focusing solely on the most promising paths, which significantly reduces computational effort in practice. The simplex method and its hybrid forms have been studied in prior works and are used in several applications. For example, Zhang, et al. [186] used the Nelder-Mead Simplex Optimization Method (SOM) to improve localization in large-scale wireless sensor networks. Their distributed algorithm minimizes a range of errors between unknown nodes and neighboring anchors. Through an event-driven approach, they reduced error accumulation, achieving significantly lower localization errors compared to maximum likelihood estimation. Ma, et al. [187] investigated gas leakage source identification using optimization methods such as genetic algorithms, simulated annealing, pattern search, and Nelder-Mead simplex, along with hybrid approaches. Genetic algorithm-Pattern search hybrid optimization performed best for source strength and location, while Nelder-Mead simplex hybrids were superior in terms of time and robustness. They also introduced a Gaussian-RBF neural network model, which improved accuracy in location estimation, especially when combined with error correction coefficients.

Wang, et al. [188] proposed a hybrid algorithm combining genetic algorithms (GA) with the simplex method to improve localization accuracy in wireless sensor networks (WSNs). Applied in the third phase

of the weighted DV-Hop algorithm, this approach optimizes unknown node coordinates. Experiments showed significant improvements in both localization accuracy and network coverage, making the GA + simplex method effective for WSN localization. Authors in [189] studied energy consumption in large-area wireless sensor networks with cluster-based routing protocols, focusing on optimizing the number of clusters (k) and transmission range (R). They used the Nelder-Mead Simplex method to achieve joint optimization of (k_{opt},R_{opt}) and reduce energy consumption in both the cluster head and member nodes. Simulations showed that their approach significantly improved energy efficiency in clustering and data transmission, which can be applied to existing routing schemes to enhance network lifespan. There are several journal papers on optimizing sensor placement for air quality monitoring using various optimization methods, but there is limited use of the Simplex Method specifically. However, the method has some drawbacks. While it performs well for many real-world problems, the Simplex Method can suffer from degeneracy, where the algorithm cycles and revisits previously explored vertices without making progress. Additionally, the Simplex Method may not always be the fastest option for very large or complex problems, where interior-point methods or other optimization algorithms might be more efficient. Despite these limitations, the Simplex Method remains one of the most widely applied algorithms for linear programming due to its reliability and effectiveness in practical applications.

2.6.2. Heuristic Algorithms

Heuristic algorithms use rules of thumb, intuition, or simplified models to solve optimization problems that do not guarantee the best solutions but provide near-optimal solutions within a reasonable time frame. However, solutions can be optimally accurate and still considered heuristic until proven to be optimal. These methods are useful where problem-solving involves large-scale datasets, where exact methods are infeasible due to computational complexity. Previous studies have focused on improving various aspects of network and device optimization, primarily within the Internet of Things (IoT) and communication networks. For example, Kaustubh et al. [190] propose HOLA, a heuristic algorithm to reduce and balance energy consumption in Industrial IoT networks. They use a heuristic method to opportunistically offload IoT data to smart devices carried by workers, which then choose the most energy-efficient communication link (e.g., Bluetooth, Wi-Fi, or 3G/4G) based on link quality and energy cost. Hussain, et al. [191] propose an improved heuristic-based algorithm for optimizing Home Energy Management Systems (HEMS). The algorithm schedules home appliances based on demand response, renewable energy availability, and battery conditions. To minimize energy costs and time-based discomfort, they use a multi-objective genetic algorithm (MOGA) with Pareto optimization, allowing consumers to choose the best solution.

However, using a heuristic approach presents several limitations. In considering the heuristic approach for solving optimization problems, it depends on the requirements or criteria of the user, which specific method

or strategy is most appropriate. For example, the objective is to find the optimal solution, determine the number of optimal solutions required, as some problems may yield multiple results, and assess computational efficiency and the time frame required to complete the process. Certain optimization problems require complex mathematical analysis, and employing approximation approaches like heuristic methods can be challenging as the optimal solution is unknown.

A. Greedy Algorithm

The function of a greedy algorithm is based on the principle of local optimal selection; in order to maximize immediate benefits, near-optimal choices are made at each step. The basic concept of this approach is to make a greedy choice, which suggests that it makes a locally optimal choice with the expectation that it might lead to a globally optimal solution [192]. These types of algorithms rely on the optimal substructure property, where the problem can be divided into smaller sub-problems that can be solved independently and efficiently, similar to the knapsack problem and the minimum spanning tree [193]. Greedy algorithms are known for being simple and manageable, and typically require less time and space complexity compared to other exhaustive search methods [194]. The greedy approach is deterministic, meaning that for the same inputs, it produces consistent results, ensuring a reliable and predictable system. However, this approach is more suitable for problems that have a greedy choice property and optimal substructure, as there do not guarantee an optimal solution when applied to a more complex scenario [195], (See Figure 3). The greedy algorithm is a widely used tool due to its ease of implementation and computational efficiency in algorithm design. There are some common uses of greedy algorithms in various applications, including wireless networking [196], task scheduling [197], and optimization problems [198].

Recently, the greedy algorithm has been utilized to solve mobile sensor network and spatial-temporal coverage problems with a budget constraint for drive-by sensing applications. Modifications are made to the standard greedy approach to handle specific constraints, e.g., budget, and collocation requirements. Chen, et al. [199] developed a greedy algorithm to choose participants in order to maximize coverage of sampling points while ensuring that the total rewards for selected participants stay within a limited budget. The results of an extensive simulation were produced to evaluate the validity and efficiency of the algorithm, which demonstrated that the greedy algorithm outperforms an existing method. Ali, et al. [14] present a novel optimization approach based on a greedy heuristic to select optimal mobile sensor deployment to maximize the total coverage along the length of the route to inspect the roads for potholes. Zarrar, et al. [32] formulate the problem and describe an efficient greedy heuristic and propose an optimization approach for a joint placement of reference stations for k-hop calibration and route selection to maximize the total spatial coverage. Hanh, et al. [200] proposed a hybrid algorithm by combining greedy heuristic and spanning tree techniques to determine the minimal number of sensors and relay nodes to

ensure target coverage. In particular, greedy algorithms are well-suited for optimization problems like the maximum coverage problem, and often guarantee strong performance. The effectiveness of the algorithm is measured using the approximation ratio, which suggests the accuracy of the algorithm's solution. For the maximum coverage problem, the greedy algorithm achieves an approximation ratio of $1-1/e$, where e = 2.718 denotes Euler's number [201]. This ratio means that the solution from the greedy algorithm will cover at least $1-1/e \approx 0.63$ of the maximum possible coverage [201]. Agarwa, et al. [8], utilized greedy algorithm to obtain an approximation ratio (1-1/e) solution for the maximum coverage problem, which has been considered as NP-hard.

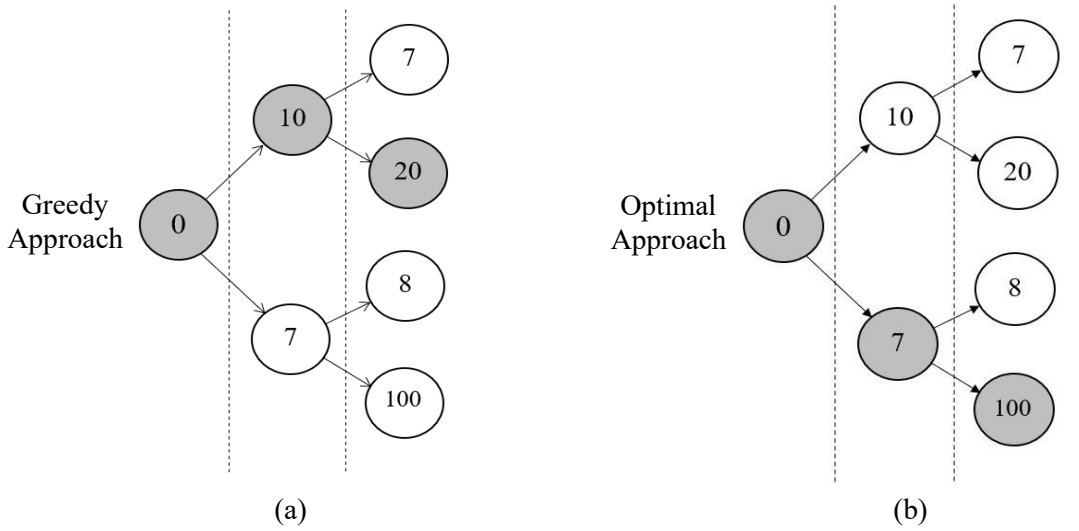


Figure 3: Branching structures comparing optimization approaches: (a) Approximate solution using greedy algorithm, selecting nodes with highest value at each step (total sum = 30), (b) Exact solution considering all outcomes to achieve optimal result (total sum = 107).

B. Simulated annealing (SA)

Simulated annealing (SA) is another metaheuristic optimization algorithm, taking inspiration from the metallurgical process of annealing, where the material is heated and then cooled to increase the size of crystals and decrease the number of defects [202]. It changes a suboptimal solution to a problem iteratively, and relies on logic and rules to seek and find the best possible solution, which is usually near optimal [203]. In practice, simulated annealing does not guarantee global optimal solutions, but usually produces good solutions, and similar to genetic algorithms, repeated process runs can provide high-quality solutions [204]. Simulated Annealing is highly flexible and can be applied to a wide variety of optimization problems, particularly those that involve complex, non-linear [203], and high-dimensional search spaces. It is used is in applications to solve problems such as scheduling problems [205], the travelling salesman problem [206], and resource allocation [207]. Simulated annealing provides a means to escape local optima by allowing

hill-climbing moves (i.e., moves that worsen the objective function value) in hopes of finding a global optimum [208]. The simulated annealing optimization approach has been used in prior studies in optimizing sensor placement for air quality monitoring. Zhang, et al. [209] proposed a weighted DV-Hop method combined with simulated annealing to address the positioning accuracy of the sensors problem. The proposed approach calculates the influence of different known nodes on unknown nodes and assigns weights to the average distances of each known node and at this stage, the average distance between adjacent nodes is closer than the actual average distance. The simulated annealing algorithm is used to find the location of unknown nodes. MATLAB simulation results indicate higher precision of the proposed positioning. Kannan, et al. [210] developed a simulated annealing (SA) based localization algorithm for ad hoc wireless sensor networks to address flip ambiguity problems, which cause location errors. Their proposed method consists of two phases, where in the first phase, an accurate location estimate for nodes is obtained using simulated annealing. In the second phase, the focus is on nodes likely affected by flip ambiguity, which are identified and relocated to correct positions, based on the neighborhood information of nodes. A simulation was done on a sensor network of 200 nodes whose distance measurements were corrupted by Gaussian noise to evaluate the proposed scheme's performance. Results from the proposed novel scheme simulations indicate accurate and consistent location estimates of the nodes and mitigate errors due to flip ambiguity. The proposed algorithm performed better compared to traditional methods such as DV-hop and convex optimization-based semi-definite programming. Zhang, et al. [211] address the high aggregation degree and low coverage rate of sensor nodes when they are randomly deployed. The developed methods are based on a grey wolf algorithm optimized with simulated annealing to address the issue. First, a mathematical model is established for WSN coverage optimization; then, simulated annealing is integrated into the grey wolf algorithm to improve the global optimization ability and convergence rate of the grey wolf algorithm. Results from simulation experiments show that the improved algorithm performs better than the particle swarm optimization algorithm and the standard grey wolf optimization algorithm. The improved algorithm has faster optimization speed, improves network coverage, reduces energy consumption, and prolongs the network life cycle.

However, Simulated annealing has some disadvantages. A major disadvantage of the algorithm is that the convergence process may be very slow when solving a complex system, requiring much more processing time than some conventional algorithms [212]. While this simulated annealing algorithm is good at avoiding local optima, the algorithm does not guarantee a global optimal solution [204]. If the cooling schedule is too fast, the algorithm may converge to a local optimum; however, if the cooling schedule is too slow, the algorithm may take longer to converge. When solving complex problems, the process often requires multiple iterations to obtain a satisfactory solution, which may be computationally intensive.

2.6.3. Meta-Heuristic Algorithms

Metaheuristic algorithms are an advanced form of heuristic methods that combine both random and local search algorithms, allowing them to explore the solution space more thoroughly and repeatedly until a convergence criterion is met, thereby achieving optimal solutions. Common metaheuristic algorithms techniques used in different applications include Genetic algorithms (GA), Particle swarm optimization (PSO), and Simulated annealing (SA), which we will discuss later in the chapter.

A. Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a global algorithm that provides a stochastic approach for solving optimization problems. It was initially developed by Professor Kennedy and Eberhard in 1995 [213] [214]. Similar to other swarm algorithms, PSO is based on the concept of bird flock social behavior, which involves cooperation and information exchange among individuals [215]. Particles evaluate the environment to identify attractive or unattractive features for decision making, then compare their positions with other particles to determine those with better positions, and finally imitate the ones with superior positions [216]. This principle allows particles to adapt, address challenges and effectively solve complex problems. PSO is popular for its simplicity, computational efficiency, and effectiveness in solving complex, non-linear, and high-dimensional optimization problems, making it useful across various fields like engineering, machine learning, and environmental monitoring. The particle swarm optimization approach has been employed in prior studies to optimize sensor placement for air quality monitoring. For example, Gupta, et al. [217], proposed a sensor placement approach for an air quality monitoring system using a bio-inspired particle swarm optimization (PSO) algorithm. The proposed framework finds the optimum location for placing sensors to optimize data coverage while reducing resource use. Extensive experiments and evaluation demonstrate that the PSO-based strategy outperforms conventional methods. The results show improvements in data coverage and accuracy as well as in cost-efficiency. Chelbi, et al. [218] proposed an optimum coverage and connectivity rate with the minimum number of nodes using a combination of particle swarm optimization and iterated local search (PSO-ILS) algorithms. The proposed algorithm is used to identify the optimal positions for placing the minimum number of relay nodes. The proposed algorithm was compared with canonical PSO, differential evolution, and genetic algorithms to evaluate its performance, and the results show an improvement in terms of nodes needed to ensure full coverage and connectivity. Islam, et al. [219] address the challenge of efficient sensor deployment using particle swarm optimization (PSO) in a wireless sensor network. The proposed approach uses PSO due to its simplicity, high quality of solution, first convergence and minimal computational requirements. The aim is to achieve an optimal solution to coverage problems with a minimum number of sensors in WSN. The results shows that the proposed PSO method provides robust and effective solutions for the coverage problem in WSNs and is

considered to give almost optimal solutions. Wang, et al. [220] addressed the hotspots problem in wireless sensor network with fixed sink nodes and developed a particle swarm optimization based on clustering with mobile sink. They perform virtual clustering technique during the routing process utilizing the particle swarm optimization algorithm. The primary parameters to select the cluster head are the residual energy and position of the nodes, and a well-designed control strategy for a mobile sink is employed to collect data from the cluster efficiently. Results from simulations show that the proposed method reduces energy consumption, prolongs network lifetime, and decreases transmission delay compared to other methods.

Despite their (PSO) usage in a wide range of applications, there are limitations that need to be mentioned. Pre-mature convergence can occur in which it may converge to sub-optimal solutions instead of finding the global optimum, especially in high-dimensional spaces [221]. This could be attributed to the loss of diversity in the swarm, which could reduce its ability to effectively explore other search spaces. Therefore, PSO may not always provide the global optimum. During the iterative process, PSO may exhibit a slow convergence rate, which could occur when it is applied to solving high-dimensional and complex problems. This limitation has been addressed by Li, et al. [221] who proposed an adaptive particle swarm optimization algorithm to overcome the limitations. The proposed approach is based on a directed weighted complex network in which the method uses the topology of a small-world network to initialize the particles position and uniformly scatter the particles over the search space. The evolutionary mechanism is used to transform the particle network into a scale-free network when the in-degree obeys a power-law distribution. The simulation results demonstrate that the proposed method effectively avoids premature convergence problems and achieves a faster convergence rate compared to other algorithms.

B. Evolutionary algorithms (EA)

Evolutionary Algorithms (EA) are a type of meta-heuristic, and are also known as stochastic search methods. They are inspired by biological evolutionary concepts to solve optimization problems [222]. Such evolutionary algorithms follow an iterative process based on the development and evolution of a population. The process randomly selects from the population, which then mixes where they interact and compete for dominance until the fittest survives, so that the desired objective is achieved [223]. Some of the works based on EA, such as forecasting air pollution, Espinosa, et al. [224] introduced a spatio-temporal approach using multi-objective evolutionary algorithms to forecast air pollution. They combined multiple non-dominated linear regression models into an ensemble learning model to determine the Pareto front to build multiple forecast models geographically distributed. The proposed approach showed promising results for one week of NO_2 predictions in southern Spain. Prior studies have also been conducted for optimizing sensor placement leverage evolutionary approach, for example, Dinh, et al. [225] addressed the novel issue of finding the monitoring frequencies for maximizing spatial-temporal coverage, reducing monitoring costs

and balancing load across the vehicles and proposed an objective function that consider all three goals. They developed an algorithm to determine the optimal monitoring frequencies by utilizing an evolutionary approach. The proposed approach performance was evaluated, and the results showed that the proposed objective function improved by a factor of 1.33 to 4 compared to other methods. Saukh, et al. [107], Addressed the problem of selecting a subnetwork of a city's public transport network to achieve good area coverage. Low-cost sensors exhibit failures and drift over time; vehicles selected for sensor deployment need to be in each other's vicinity periodically for comparing sensor readings, referred to as checkpoints. Due to the high computational complexity of the route selection problem, both with and without checkpointing support, the authors adapt an evolutionary algorithm solution and evaluate its output based on the tram network of Zurich.

Evolutionary algorithms are very effective in solving complex problems. However, there are several drawbacks. They evaluate several solutions over multiple generations, which can increase run-time, and therefore can require high computational costs [222]. In addition, they may converge to suboptimal solutions or pre-mature convergence, loss of diversity in population, and may require parameter tuning, affecting their performance, and their scalability, especially when solving optimization problems that are complex and large-scale. There are several types of evolutionary algorithms, such as genetic algorithms.

B. Genetic algorithms (GA)

Genetic algorithms (GA) are another type of meta-heuristic algorithm, and they take inspiration from the biological evolution process or natural selection [226]. The GA process produces a new population, iteratively using genetic operators on individuals within the population [227]. The key genetic elements include chromosome representations, selection, crossover, mutation, and fitness function. The process starts by randomly iterating through the chromosomes and evaluating their fitness, modifying them. GA can be used for both constrained and unconstrained optimized problems [228]. Genetic Algorithms have been applied in several studies to enhance air quality monitoring and optimize pollution control measures. Such as, for air pollution prediction, a study in Tehran, Asghari, et al. [229] used a hybrid model combining Genetic Algorithms (GA) with Artificial Neural Networks (ANN) to predict PM10 air pollution levels, using data from local weather stations from 2007 to 2013. The GA-enhanced model (BP-GA) outperformed traditional ANN models, showing higher accuracy in predicting daily PM10 concentrations, especially for shorter time periods. In terms of optimizing monitoring networks, Hao, et al. [230] used the Weather Research and Forecasting–California PUFF (WRF-CALPUFF) model combined with a genetic algorithm (GA) to optimize air quality monitoring in Shijiazhuang, China. Their objective was to maximize coverage while minimizing overlap between monitoring sites. The design results show the coverage efficiency of the urban area for SO_2 to 60.99% and NO_2 to 76.05%, and the redistributing design, including 8 sites, was

proposed, reducing overlap areas by 85.87% compared to the original network. Wang, et al. [231] address the NP-hard problem of designing air quality monitoring stations (AQMS) and propose using a genetic algorithm (GA) to optimize site selection, factoring in environmental, social, and economic aspects. Their method outperforms random site selection and can also be applied to other network designs, like placing wireless base stations efficiently to improve service quality while reducing costs.

For optimizing sensor placement, Kumar, et al. [232] propose a Real Coded Genetic Algorithm (RCGA) to solve the smog-free tower installation problem to find optimal tower locations to minimize residents' exposure to secondhand smog. They introduce a new crossover operator, combining information from two parent solitons to create two new offspring solutions using a random crossover point and two scaling factors to control the amount of information exchanged. RCGA was evaluated on CEC-2017 benchmark problem set and compared to other commonly used operators, and results show that it outperforms other operators and produces high-quality solutions in terms of convergence speed and solution quality. To optimize mobile sensor placement, Anh, et al. [233] addressed the limitations of stationary air quality monitoring stations and proposed using vehicles with deployed sensors to extend the monitoring area. The approach uses a genetic algorithm (GA) to select the optimal vehicles for sensor placement thereby maximizing the area covered. The authors propose two algorithms to calculate the monitored area and a genetic algorithm (GA)-based approach to determine the optimal buses for placing the sensor. Liang, et al. [234] propose a genetic algorithm to optimize the movement of mobile sensor nodes in a wireless sensor network for covering moving objects along a predetermined path. Due to the limited range of sensors, complete coverage requires strategic sensor movements. The genetic algorithm employs tailored crossover and mutation operations to identify an optimal movement pattern, thereby minimizing both the total distance and the maximum distance that sensors need to move. This approach effectively reduces energy consumption compared to the greedy method, thereby extending the operational lifetime of the sensors.

There are several disadvantages to using genetic algorithms (GAs). GA has a slow convergence rate, as it is time-consuming, especially when using complex data and solving large-scale problems, as it relies on random search techniques. Pre-mature convergence where the genetics algorithm converges on a sub-optimal solution, due to loss of population diversity and search space as the solutions' population becomes too similar to each other. GA employs fitness-based selection processes, which may result in over-representation of certain solutions in the population. GA is also computationally expensive when solving large-scale problems. GA performance is also affected by the parameter selection, such as population size, crossover and mutation rate, which require careful parameter tuning and are time-consuming.

2.6.4. Comparative Analysis of Modeling Approaches

This work have been contextualized within the existing literature; five recent studies have been reviewed that address similar challenges in deployment of mobile sensor using public transport. Various optimization techniques have been proposed in these studies to improve sensing coverage in spatial and temporal terms. We summarize each approach, outlining their methods, problem formulations, advantages and disadvantages such as Zarrar, et al. [32] aim to determine the optimal subset of bus stop locations for placing a limited number of reference stations to maximize total sensing area via k-hop calibration. And formulated the sensor calibration problem as a knapsack problem and proposed greedy algorithm to solve the problem by iteratively selects bus stops that offer the highest additional sensing coverage. To handle the linear constraints in its basic form the knapsack problem is suited, however, standard knapsack problem would not be well suited for a scenario in which non-linear cost functions are used. Greedy algorithm is used to reduce computational stress; however, sub-optimal results will be produced. The main focus of this study is bus stops and the sensing area coverage provided by the bus routes which intersect each bus stop. However, the study assesses route sensing coverage collectively by using spatial joins only, instead of the contribution of individual route coverage using spatial and temporal joins. This could lead to suboptimal coverage as some routes that intersect bus stop may provide low coverage. Ji et al. [42] presented a trip-based sensor deployment model that explicitly considers timetabled trips executed by a bus fleet with the aim to deploy a limited number of sensors for ensuring optimal spatio-temporal distribution. The authors formulate the problem as a bi-level optimization model, and use linear integer program where the upper level performs line pre-selection and the lower level assigns sensors to selected trips. However, using heuristic method like greedy approach would make implementation easier and also lower computational expense particularly as the problem size grows. The study focuses on coordinating sensing tasks with scheduled operations, not individual bus route geometries such as area coverage, route lengths or spatial joins which may lead to low sensing coverage in spatial terms.

Caminha et al. [11], proposed a coverage model to optimize spatial coverage of a bus-based sensor network. The authors have formulated the problem as a MILP and incorporating factors, such as street section lengths, to determine the minimum number of buses required to optimize spatial coverage. MILP is an exact algorithm, however, using heuristic method such as greedy approach would make implementation easier and reduces computational cost particularly for large datasets. Additionally, if temporal aspects such as bus schedules from bus operation timetable are included, this could improve sensing frequency and provide better data sampling or temporal coverage. Anh, et al. [233] addresses the problem of maximizing spatial air quality monitoring coverage using mobile sensors mounted on buses. For this, they developed two algorithms, the grid-based approximation approach and exact algorithm for calculating the monitored area

and introduced a genetic algorithm (GA)-based approach to select the optimal number of buses for sensor deployment, and tested on real world bus dataset from Hanoi, Vietnam. The dual method provides more flexibility to handle the size of the dataset and the required level of accuracy. The grid-based method is computationally faster but provides approximate solutions and the exact algorithm uses detailed information on road geometries and bus trajectories, providing more spatial accuracy. However, they require more time to compute. Genetic algorithms can handle complex optimization problems including non-linear, multi-modal landscapes and require large search spaces, without depending on gradient information. However, it requires parameter tuning, such as mutation rate, and population size, which affects their performance such as providing sub-optimal solutions, slow convergence and increased computational cost. Some meta heuristics algorithms can be utilized as they are more structured, require fewer parameters to tune and have faster convergence depending on the problem and its implementation. However, utilizing greedy heuristics as an alternative can provide faster approximate solutions with lower computational cost.

Ali et al. [14] address the problem of optimal deploying a limited number of mobile sensors on public transit vehicles, specifically buses for road surface condition monitoring. The goal is to maximize the total length of road segments inspected for defects such as potholes by strategically placing sensor on buses operating on predetermined routes. The problem is formulated as a linear optimization problem to maximize the number of road segments covered with limited number of vehicles. The paper proposes greedy heuristic algorithm that iteratively selects the route segments that has the most uncovered road length to the total coverage. Spatial features such as street segments, i.e., lengths are used to select the routes, however, it may be useful to include temporal features to improve sampling rate. The advantage of using greedy heuristic includes computational efficiency, and ease of implementation, however, it only provides near-optimal solution. Problem formulation and selection of optimization approach to solve them depends on several factors such as objectives, the type and granularity of data, the requirements for spatio-temporal coverage, available resources and constraints. Heuristic approaches such as greedy heuristic, and meta heuristics such as Genetic Algorithm, Simulated Annealing, Particle Swarm Optimization, etc., provide near-optimal solutions with lower computational cost, they are more scalable, and flexible, particularly when handling non-linear constraints and dynamic scenarios. Alternatively, exact methods such as Mixed-Integer Linear Programming are more suitable for scenarios that require accurate solutions and have well-defined constraints. However, they are time-consuming especially when computing large datasets, therefore, to solve this, algorithms such as greedy heuristic is used to reduce the computational time. But the trade-off is reduced accuracy or provides near-optimal solution.

Chapter 3: Model Development

This chapter provides a brief description of the development of the mathematical model designed to quantify the coverage contribution of each bus route, for optimizing the deployment of bus-based sensors and reference sensor stations for monitoring air quality and mobile sensor calibration. This model aims to maximize spatio-temporal coverage within the set budget constraints and ensure k-hop calibration of the entire deployed mobile sensor network. Section 3.1 presents the framework for addressing sensing coverage maximization problem with resource constraints. Here we discuss the procedure of our dataset preparation, route spatial and temporal features, and using standardization method for scaling key metrics, formulating Mixed Integer Linear Programming and a greedy algorithm for sensor deployment. In Section 3.2, we describe the key elements of the model framework, by defining Sensing Coverage, Sampling Coverage, and k-hop calibration. In Section 3.3 we describe the procedure of quantifying the relative importance of each route feature used in the optimization model. The procedure includes standardization of key metrics into comparable scale and weight determination for the Weight Sum Method. Next, in Section 3.4 we present the development of the Mixed Integer Linear Programming formulation to evaluate the sensing coverage score which represents the contribution of each bus route identifying the routes that provide the most effective coverage and for bus based senor deployment. The model is then extended to formulate the optimization of reference sensor placement which include identifying the minimum number of bus stops required and their geolocations for reference sensor placement to ensure k-hop calibrability of the mobile sensors on selected routes. We have used greedy approximation to solve the optimization problem and in Section 3.5, we describe the application of this approach. We have also described a random approach implementation for comparative analysis to evaluate the performance of our proposed model. The contents included in this chapter is based on the paper published by the author of this thesis, “Bus-Based Sensor Deployment for Intelligent Sensing Coverage and k-Hop Calibration” [235], and has been adjusted to align with the objectives of this thesis.

3.1. Framework

We address the maximizing sensing coverage problem with limited resources, where our optimization model focuses on optimizing the placement of mobile sensors and reference sensor geolocations . We collected and prepared our dataset, which is a General Transit Feed Specification (GTFS) dataset from two cities: Toronto, Canada, and Manchester, UK. These cities have been selected for our experiment due to their different road network designs; therefore, we are able to analyze the adaptability of our proposed model in different environments. The GTFS dataset consists of detailed bus transit information (agency information, routes, stops, trip schedules, stop times, and fares) on public transit data collected. The route features or attributes selected for our optimization model are as follows:

- **Area Coverage** (A_r^d): The spatial extent coverage of each route (r) provides in terms of area, which are adjusted for different buffer distances (d).
- **Trip Frequency** (T_r): The temporal coverage of each route (r) indicating the total number of trips made during a specific timeline.

The two metrics have different scales; therefore, we will standardize them to ensure they are comparable. We are applying z-score normalization to convert them to the same scale. We then calculate the coefficient weights, which determine the relative importance of the two factors (area coverage and trip frequency), in selecting the right set of bus routes for bus-based sensor deployment. We utilize the standardized values of the two factors to calculate the weights and then integrate them into our model formulation. We formulate the maximizing sensing coverage problem using Mixed Integer Linear Programming (MILP) within the budget constraints. We utilize the greedy algorithm to solve the problem, which iteratively selects the best possible set of routes that maximize the weighted coverage within the budget constraint, updating total coverage and cost after each selection. Following the selection of routes, we find the number of bus stops (geolocations) iterating through only the optimized routes selected (i.e. constraints) with the aim to minimize the number of bus stops required. The block diagram illustrates the process of our model's sensor deployment strategy:

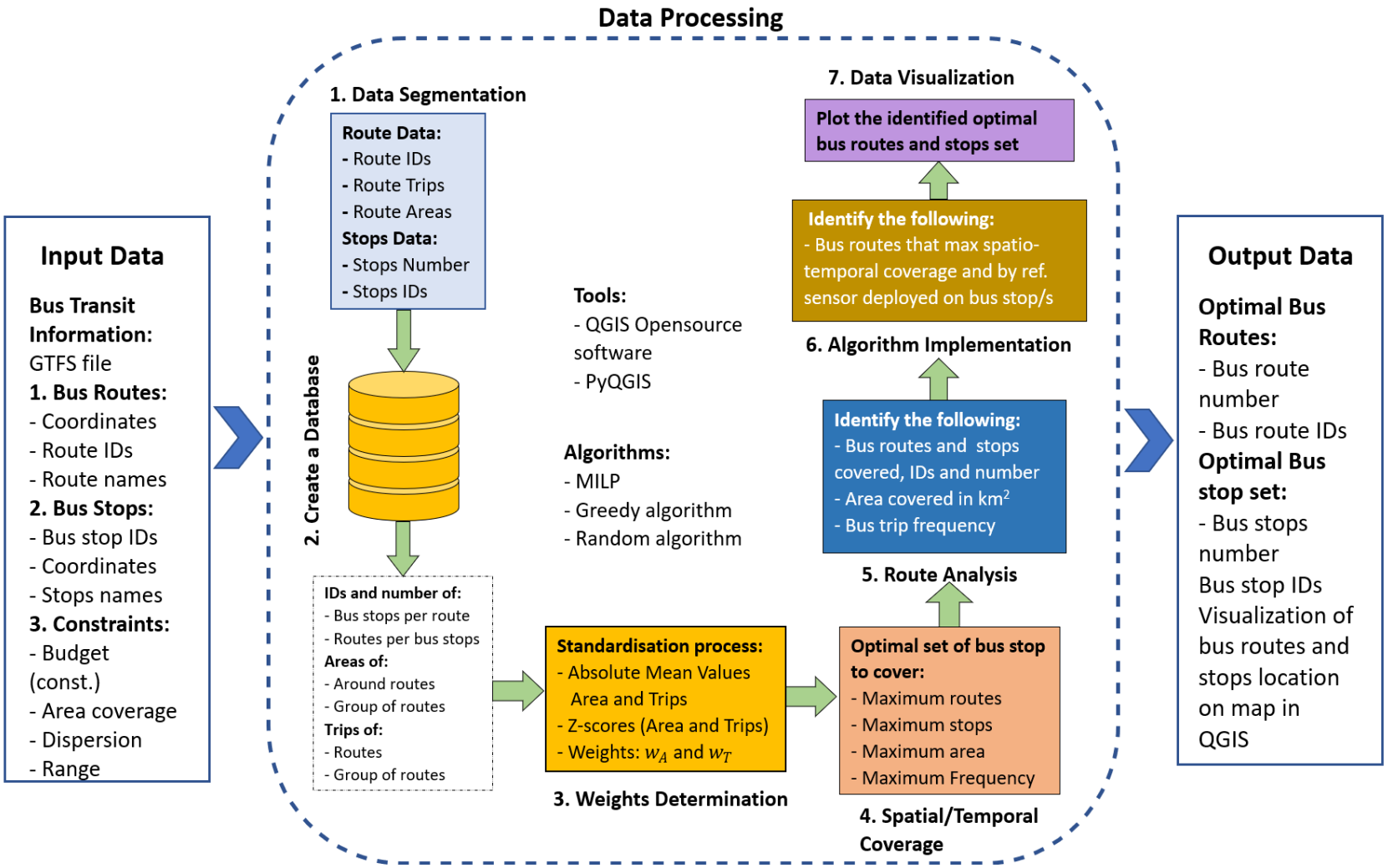


Figure 4: Block diagram of the proposed model.

3.2. Model Core Elements

Bus-transit systems offer high temporal coverage and reliable operations due to their set routes, schedules, and frequent daily movements [236] [38]. Our model aims to strategically place sensor nodes in the transit network to maximize unique sensing area coverage with a minimum number of bus routes selected and a minimum number of reference stations for calibrating the bus-based sensors on the selected routes. The relative importance of the routes depends on the geographical extent of the selected public transit system. The routes were selected to provide maximum geographical coverage, and to avoid redundancy, we considered sections that overlap with each other. These considerations have been included in our optimization model for the strategic placement of air quality sensors. The objective of our model is to balance the selected route's spatial extent with its trip frequency, ensuring that the data collected by mobile sensors deployed on selected routes provides both 1. Maximum area coverage and 2. Optimum sampling rate. We define the sampling rate as the number of trips per route within a set time duration, and a route is defined as any contiguous street segment covered by the selected transit vehicle (bus). The key components of our model are defined in detail in the following sub-sections.

3.2.1. Route Definition

The route network is a key component of public transit operations as it provides the geographical extent of a transit system. In this work, a bus route section refers to any contiguous street segment covered by public transit (bus). Features of a route section have a specific starting and ending point, area, and can overlap with other route sections. A similar definition is found in the work by Zarrar et al. [32]. The spatial coverage is calculated by the extent of the route area, which varies due to different buffering distances, which is the width of the route from the center and the length of the route. The temporal coverage represents the sampling rate along the entire route length, which correlates with the trip counts of transit vehicles on the defined route. Each trip count is defined as one complete rotation of each transit vehicle, which starts at a point and ends at the same starting point on the route. These two key components have been incorporated into our model to find the optimum spatio-temporal coverage. We have utilized an area as a metric to define the spatial extent of each route, as it accommodates two dimensions (length and width). The benefits of selecting an area as a metric are described in detail in Section 5.1.3.

3.2.2. Sensing Coverage

Sensing Coverage refers to the spatial extent or the area of the route/s that is monitored or sensed by the deployed mobile sensors on the route/s. In our context, sensing coverage is determined by the routes chosen for sensor deployment and the buffer distance (d) applied to these routes. As the low-cost sensors are mounted on buses, the sensing coverage is dependent on the range of the selected bus routes and the area it

covers while in motion. The combination of higher distribution density of bus routes and larger sensor detection range enhances the overall coverage. Therefore, the relationship between sensing coverage is defined by the spatial extent of the routes and the detection range of the sensors. The objective is to ensure that the sensors can detect air pollution levels across the maximum possible area. The density and the distribution of the road or route network design determine the sensing coverage (A_r^d). This is also influenced by the range of the sensor detection capabilities, which is represented by the buffered zones, hence the incorporation of area as a core component or metric in our sensing coverage problem.

3.2.3. Sampling Coverage

Sampling coverage refers to the temporal aspect of the data collected by mobile sensors deployed on the route, which correlates with the trip counts of transit vehicles on the defined route. As discussed in Section 2.3, where we mentioned the trade-offs between achieving spatial and temporal resolution, mobile sensors have the ability to achieve high spatial resolution but low temporal resolution, as they cannot measure or sample data at a single point frequently due to their dynamic nature in terms of geo-spatial considerations. To achieve an adequate temporal resolution, it is necessary that the mobile sensor revisits the same geo-spatial point periodically or frequently. In this work, this periodic or frequent revisit is determined by the trip counts (T_r) of the transit vehicle, which is defined by the number of trips or journeys repeated on the same route. The number of trips per route is related to the sampling rate for measuring data continuously overtime. Therefore, the higher the number of repeat journeys or trips, the higher the number of samples collected at the same geo-spatial point, resulting in higher temporal sampling coverage.

3.2.4. K-hop Calibration

Calibration Coverage denotes the extent to which mobile sensors are calibrated using reference stations to maintain data accuracy and reliability. Calibration coverage is ensured by the placement of reference sensor stations (N) at strategic geolocations , which ensures k-hop calibration of the entire mobile sensor network in the designated sensing area. We have taken advantage of the city's infrastructure, which includes bus stops, which are an integral part of the bus transit network. We have selected bus stops as the geolocation for reference sensor placement due to several benefits, including their proximity to roads, existing infrastructure that requires minimal construction or modification, and an available power supply. *K*-hop calibration allows sensors to be calibrated directly and indirectly through intermediate stops within *k*-hops of a reference station ($N_{is} \leq k$). Calibration begins with k=1 representing direct or single-hop calibration of bus based sensors by a reference station, and continues (k=1,2,3….,n), allowing for multi-hop calibration of mobile sensors through freshly calibrated mobile sensors. Multi-hop calibration is only possible if bus routes overlap or intersect allowing two mobile sensors to rendezvous. This strategy ensures calibration

coverage with a minimum number of reference stations required, thereby optimizing resource allocation and reducing costs.

3.3. Weight Determination

The objective is to quantify the contribution of each route in terms of spatial and temporal coverage, utilizing the dataset of the two factors (area coverage and trip frequency). We propose a two-step process for determining the coefficient weights for each factor. In the first step, we standardize the two factors by transforming them into a more comparable scale, and in the second step, we calculate the coefficient weights using the standardized scores.

3.3.1. Metric Selection and Standardization

The metrics have been selected which accurately represent the factors that reflect the core components of drive-by sensing and determine our proposed approach performance:

1. **Area Coverage:** This metric contributes to the spatial coverage of each route, and each route's area is calculated based on the three buffered distances or pollutant dispersion distances.
2. **Trip Frequency:** This metric provides the number of trips made by the buses on each route. Frequent trips provide more data collection opportunities at the same geographic points, thereby improving the temporal resolution.

These metrics represent both the spatial and temporal extent of mobile sensing, ensuring that the selected routes provide maximum sensing coverage and data collection frequency. The two metrics have different dimensions and scales; therefore, we need to process the data points to make them more comparable. We employ the z-score to standardize each route area coverage and trip frequency metric. The z-score of each metric is calculated using the following formula [237]:

1. Z-score standardization for area coverage:

$$Z_A = \frac{A_r^d - \mu_A}{\sigma_A} \qquad (1)$$

2. Z-score standardization for trip frequency:

$$Z_T = \frac{T_r - \mu_T}{\sigma_T} \qquad (2)$$

where:

- A_r^d (sensing coverage) and T_r (trip counts) are the two metric values for a route.
- μ_A, is the mean and σ_A, is the standard deviation (SD) of the entire route area dataset.
- μ_T, is the mean and σ_T, is the standard deviation of trip frequency of the entire route dataset.

The standardization of the metric values allows for their integration into the weight calculation, as the metrics of different scales will not interfere with the calculation. The sum of the metrics z-scores area and trips means will be equal to 0 (area z-score mean + trip z-score mean = 0), and their standard deviation sum will be equal to 1 (area z-score SD + trip z-score SD = 1). This will verify that the process and the calculated values are valid.

3.3.2. Weights Calculations

As mentioned previously, the two metrics selected to determine the weights are the area coverage (A_r^d) and trip frequency (T_r). In the previous step, we also calculated the z-score values of the two primary criteria (Z_A and Z_T), to balance the different scales and dimensions of each metric. We now determine the coefficient weights of the respective factors using the standardized z-scores. The weights are proportional to the contribution each factor provides. To validate the values, the sum of the weights of two metrics ($w_A + w_T$) is equal to 1. The weights calculation steps are as follows:

$$w_i = \frac{x_i}{\sum_{i=1}^{n} x_i} \qquad (3)$$

Where:

- $\boldsymbol{n}$ is the number of metrics,
- $\boldsymbol{w_i}$ is the weight of the $\boldsymbol{i}$th metric,
- $\boldsymbol{x_i}$ is the value of the $\boldsymbol{i}$th value, and
- $\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_i}$ is the sum of all the metric values.

The process involves calculating the total value ($Z_A + Z_T$) and then deriving the weight using the following formulas:

1. Weights for Area:

$$w_A = \frac{Z_A}{total\ value} \qquad (4)$$

2. Weights for Trips:

$$w_T = \frac{Z_T}{total\ value} \qquad (5)$$

By using this method, the weights assigned to area coverage and trip frequency are proportional to their respective contributions.

3.4. Mixed Integer Linear Program Formulation

3.4.1. Bus Route Model for Maximal Spatio-Temporal Coverage

In our model, we define the total coverage as an objective function for each bus route, quantifying their contribution based on their area and trip counts. The total sensing coverage (C) is the weighted sum of each bus route contribution, which includes spatial and temporal contributions. This allows for measuring the sensing coverage of the selected set of routes (R) on which the buses will be deployed with sensors, where each route $r \in R$. The weighting coefficient w_A reflects the normalized importance of the area covered by each route, while w_T represents the weighting coefficient of the normalized importance of the number of trips on each route. The objective is to maximize the total area covered by the selected routes while also considering the number of trips to ensure adequate temporal coverage. The weighted objective function for an optimized route selection is shown below:

Maximize:

$$C = \sum_{r \in R} (w_A . Z_A + w_T . Z_T) . x_r \qquad (6)$$

Subject to:

1. Budget Constraint:

$$\sum_{r \in R} C_r . x_r \leq B \qquad (7)$$

2. Binary Decision Variables:

$$x_r \in \{0,1\} \forall_r \in R \qquad (8)$$

where:

- Z_A, is the calculated z-score of the area.
- Z_T, is the calculated z-score of the trips count.
- C_r, is the cost constraint.
- B is the total budget constraint.
- x_r, is the binary decision variable (1 if route included or 0 if route not included).

This formulation identifies the right subset of bus routes required for bus-based sensors deployment to maximize sensing and sampling coverage, which would maximize the total unique sensing area.

3.4.2. Minimal Bus Stop Selection Model

To ensure that the selected bus routes are properly calibrated, reference sensor stations are placed enroute of the selected set of routes. The placement of these reference sensor stations on every route is infeasible, as reference sensors are expensive to purchase, deploy, and maintain. Therefore, minimizing the number of bus stops is the second objective of our model, where a minimal number of reference sensors are placed ensuring complete *k*-hop calibration coverage. The objective function and associated constraints ensure that this selection maximizes coverage and minimizes the number of stops required for calibration. The bus stop selection objective function is given by:

Minimize:

$$min \sum_{s \in S} y_s \qquad (9)$$

where:

- S is the set of all bus stops.
- y_s is a binary decision variable indicating if bus stop s is selected (1 if selected, 0 otherwise).

Constraints for Bus Stop Selection:

1. K-hop calibration constraint:
   Each selected route r must be within k-hops of the selected bus stop for calibration coverage. The constraint ensures that each selected route has at least one bus stop within k-hops:

   $$\sum_{s \in S} y_s . N_{sr} \leq k \quad \forall_r \in R \qquad (10)$$

where:

- N_{sr} is the number of hops between bus stop s and route r.
- R is the set of selected routes.

2. Binary Decision Variables:
   The decision variable y_s is binary, indicating whether a bus stop is selected or not:

   $$y_s \in \{0,1\} \ \forall_s \in S \qquad (11)$$

This formulation is designed to determine the minimum number of bus stops required to effectively *k*-hop calibrate bus-based sensors in the optimum selected bus routes.

3.5. Greedy Approximation

We consider the problem of selecting an optimized set of bus routes for mobile sensor deployment as the Maximum Coverage Problem, which is NP-hard; therefore, solving this problem requires using an approximation. For our model, we have employed a greedy algorithm with an approximation ratio of 1–1/e ≈ 0.632. The proposed greedy algorithm iteratively selects the best possible set of bus routes that maximizes total coverage restricted by the budget constraints (B), evaluates the contribution of each route for area coverage (Z_A) at a specific buffered distance (d) and the number of trips (Z_T), selecting the route that provides the highest sensing coverage. The decision variables x_r(indicating if the route r is selected) are updated at each step until no further routes can be selected while remaining within the set constraints. Post-selection of the subset of bus routes, the proposed greedy approach continues with the selection of the minimum number of bus stops, which is formulated as a Set Cover Problem required for reference station placement to ensure effective calibration of the selected bus routes. The variables in this process are y_s, which indicates whether a bus stop (s) is selected, and the distance function N_{sr}, which measures the distance between bus stop (s) and route (r). The proposed model is designed to ensure that there are sufficient reference stations available to guarantee k-hop calibrability for the selected bus routes with deployed mobile sensors. While the greedy algorithm does not provide an optimal solution, it provides a good approximation for our NP-hard problems, achieving a provable approximation factor. The greedy approach is typically simple to implement and is efficient in terms of runtime, often running in polynomial time, which is crucial for problems with large datasets. The pseudo-code on the next page outlines the steps of the greedy algorithm for bus route selection in bus-based sensor deployment. Algorithm 1 is the algorithm we have adopted for our proposed framework solution.

Algorithm 1: Greedy algorithm for route selection and bus stop selection.

Data: Toronto and Manchester GTFS Data
Constraints: B; C_r; Z_A; w_A; w_T; Z_T; N_{sr}; k
BusStopSet = total number of bus stops
BusRouteSet = total number of bus routes

Initialization:
SelectedRoutes := ∅;
TotalCoverage := ∅;
TotalCost := ∅;
maxRoute := null;

 # Route Selection
while TotalCost <= B **do**
maxCoverage := 0
for each route r **in** R **do**
if route r not **in** SelectedRoutes **then**
routeCoverage := $w_A * Z_A + w_T * Z_T$
if routeCoverage > maxCoverage and TotalCost + C_r <= B **then**
maxCoverage := routeCoverage
maxRoute := r
if maxRoute is not null **then**
Add maxRoute to SelectedRoutes
Update TotalCoverage += $w_A * Z_A + w_T * Z_T$
Update TotalCost += C_r of maxRoute
else
break

 # Bus Stop Selection
SelectedBusStops := ∅;
CoveredRoutes := ∅;
maxCoverage := 0;
maxStop := null;
for each stop s **in** S **do**
if stop s not **in** SelectedBusStops **then**
Coverage := 0
for each route r **in** SelectedRoutes **do**
if N_{sr} <= k **then**
Coverage += N_{sr}
if Coverage > maxCoverage **then**
maxCoverage := Coverage
maxStop := s
if maxStop is not null **then**
Add maxStop to SelectedBusStops
for each route r **in** SelectedRoutes **do**
if N_{sr} <= k **for** maxStop **then**
remove route r from SelectedRoutes
else
break
return SelectedRoutes, SelectedBusStops

3.6. Baseline Random Algorithm

The random algorithm is used as a baseline to evaluate the performance of the proposed optimization method. This algorithm works by randomly selecting a set of bus routes without taking their coverage metrics or trip frequencies into account. However, for the bus stop selection, we use the routes selected from the greedy approach. This approach provides a comparative benchmark to understand the effectiveness of the greedy algorithm and the associated weighted selection approach. In random approach the algorithm selects random set of routes without replacement, which means that once a route has been selected, the same route cannot be selected again to avoid duplication and prevent the algorithm from going into a loop. The random algorithm process follows the steps below:

1. **Initialization:** Define the budget constraints and initialize variables for routes and stops.
2. **Random Route Selection:** Selects routes randomly until the budget is exhausted.
    - Route Inclusion: For each randomly selected route, check if adding it is within the budget constraints.
    - Budget Update: If the route can be added, update the total cost.
3. **Random Bus Stop Selection:** Using the set of selected routes from the greedy approach to randomly select bus stops to ensure k-hop calibration coverage.
    - Stop Inclusion: For each randomly selected bus stop, check if adding it is within the budget constraints.
    - Coverage Update: Update the list of selected stops for k-hop calibrating all the bus routes selected.

Algorithm 2: Random algorithm for route selection.

Data: Toronto and Manchester GTFS Data
Constraints: B; C_r; Z_A; w_A; w_T; Z_T
BusStopSet = total number of bus stops
BusRouteSet = total number of bus routes

Initialization:
SelectedRoutes := ∅;
TotalCoverage := ∅;
TotalCost := ∅;

while TotalCost <= B **do**
randomRoute := select_random_route()
if randomRoute not in SelectedRoutes and TotalCost + C_r <= B then
Add randomRoute to SelectedRoutes
Update TotalCoverage += w_A* Z_A + w_T* Z_T
Update TotalCost += C_r of randomRoute
else
break
return SelectedRoutes

Algorithm 3: Random algorithm for bus stop selection.

Data: Toronto and Manchester GTFS Data
Constraints: N_{sr}; k
BusStopSet = total number of bus stops
BusRouteSet = total number of bus routes

Initialization:
SelectedRoutes := ∅;
CoveredRoutes := ∅;
maxCoverage := 0;
maxStop := null;

for each stop s in S **do**
if stop s not **in** SelectedBusStops **then**
Coverage := 0
for each route r **in** SelectedRoutes **do**
if N_{sr}<= k **then**
Coverage += N_{sr}
if Coverage > maxCoverage **then**
maxCoverage := Coverage
maxStop := s
if maxStop is not null **then**
Add maxStop to SelectedBusStops
for each route r **in** SelectedRoutes **do**
if N_{sr}<= k **for** maxStop t**hen**
remove route r **from** SelectedRoutes
else
break
return SelectedRoutes

Chapter 4: Model Implementation

In this section, we detail the steps taken towards implementing our optimization framework. We segment our data by geographical area (by bus route and bus stop) to allow us to calculate coverage of the route area and frequency of trips. Then we discuss the standardization step with z-score technique to standardize the data and quantify weights of each metric. Using a greedy algorithm that prioritizes coverage within the budget, is described in route selection. Lastly, it addresses the issue of bus stop selection for reference stations to ensure sensor calibration is effective but also that a minimum number of bus stops are used. The chapter provides a step-by-step description into the approach used in improving the quality of the urban air monitoring. The content of this chapter is based on the author's published paper, "Bus-Based Sensor Deployment for Intelligent Sensing Coverage and k-Hop Calibration" [235], and has been adapted to suit the objectives of this thesis.

4.1. Data Segmentation

Here, we focus on systematically process and compile the collected data into useful data for detailed analysis and for optimization strategies. This analysis is based on data sources which includes bus transit information related to the selected metrics, such as the area covered by each route and the frequency of bus trips on each route. The current analysis of the efficiency and coverage of the bus network is based on this segmented data, which aims to improve mobile and reference sensor deployment for air quality monitoring as well as additional intelligent tasks.

Steps in Data Segmentation:

1. **Geographical Segmentation:**

   We divide the geographical information into smaller, and more manageable datasets which enables us to perform more accurate analysis. We first define the study area by specifying geographic parameters or coordinates.

   - **Input Coordinates:** A list of coordinates (latitude and longitude) provides the geographical target area mapping. These coordinates define the parameters of the zone which is also divided into small sections to obtain a more detailed overview. This helps us identify the various section of the study area but also provides us with geolocation of bus stops as well as the bus routes and their extents.
   - **Buffer Zones:** In order to improve our analysis, we have created buffer zones around each bus route at three different distances. Setting different buffer distances allows for the creation of both immediate and extended areas influenced by the bus routes. The three distances we have selected are 25m, 50m and 100m, each providing a different area coverage of each route.

2. **Metrics Extraction:**

   In the previous step we divided the geographical components or characteristics of the given dataset into manageable components. We separate the key metrics from the earlier segmented data set for our evaluation of the overall potential or contribution of each bus route sensing coverage in terms of both spatial and temporal dimensions.

   - **Area Coverage Calculation:** As we have mentioned before, we have buffered the bus routes at three different distances (25m, 50m, and 100m). The buffer distances are from the center of the route and provides the spatial extent of the sensing coverage. We will calculate three different area coverages of each route which represent the different setting for the creation of both immediate and extended areas influenced by the bus routes.
   - **Trip Frequency Calculation:** Another key metrics which contributes to the importance of sensing coverage is the trip frequency at each route. This is the number of trips buses on each route take during a specific period of time. The GTFS dataset provides this data which is in the trips text file. The time frame on how often the buses complete their trips on each route can be in daily or weekly. This metric is important due to its contribution in sensing coverage in temporal context, the higher the number of trips of buses the higher the sampling rate at each point of the route.

3. **Data Structuring:**

   The extracted data (metrics) organized is in a specific format so that the information can be organized, processed, stored, and retrieved quickly and effectively. They are a means by which we can ensure that the data is ready for use for optimization and decision-making processes.

   - **Organize Data:** The data that has been extracted and calculated is first organized systematically, which is related to the bus transit data. This includes route data, their IDs, areas, trips counts and also bus stops data, their unique ID and the route each bus stops intersects. This structured format simplifies the dataset removing redundant data like their long and short names as their unique IDs provide enough and easy to use imitation.
   - **Store Data:** When the data has been organized, it is compiled and stored in designated files form where it is easily and efficiently retrieved and used. The file for geospatial data allows simple processing and visualization of the data compiled. Every bus route receives a unique identifier together with its relevant metrics, making it easy to access for use in subsequent optimization tools, including Tabu Search or other heuristic methods.

We downloaded the required data from GTFS files we obtained from an open source portal [35]. GTFS establishes a universal method for the organization of data related to various kinds of public transportation services. The GTFS files provided rich details about the cites bus routes and bus stops, featuring their

special IDs, names and positions, as well as routes paths and details on their trips. For our model we will use the text files which are relevant to our analysis. All text files contain data columns that show different public transit features, where some are important and some are optional. We have only included the text files and columns in them that contain data relevant to this work. Columns with additional data but irrelevant are ignored or removed which reduces processing time and management resources.

4.1.1. Shapes (shapes.txt)

The shape files provide information and geospatial information like coordinates for the routes or paths that the transit vehicle type will travel on. The relevant columns are as follows:

- shape_id: This column contains unique ID or identifier for each route shape along the road network and every point for the shape contains the same value.
- shape_pt_lat: This column contains the latitude coordinates of every point of the shape associated with the unique shape ID.
- shape_pt_lon: This column contains the longitude coordinates of every point of the shape associated with the unique shape ID.
- shape_pt_`sequence: An integer number zero that defines point ordering in a shape. Each shape should always contain unique values therefore not repeating again within the same shape.

4.1.2. Route (routes.txt)

The route files provide information regarding group of trips which provide the travellers with displayed service information.

- route_id: This column contains unique ID or identifier for each route. The route ID must be specified.
- agency_id: This column provides information related to the transit operating agency of the routes.
- route_short_name: The column provides a short version of the route name which is less descriptive than the route long name. The short name must be specified.
- route_long_name: This column provides a more descriptive name of the route compared to the route short name. the long route name must be specified.
- route_desc: This column provide a description of the route information which is useful and should not be duplicate of route short name and route long name.
- route_type: This column specifies what type of transportation operates the routes for example Trams, bus, subway, ferry, etc.

4.1.3. Trip (trips.txt)

The trips files contain information regarding scheduling and routing of each route trip, within a specific time.

- route_id: This column identifies route ID, which is associated with the trip, corresponding in the route.txt file.
- service_id: This column provides information regarding schedules of each trip which are a set of dates about the services availability on each route.
- trip_id : This column contains a unique ID or identifier for each trip. The trip ID must be specified.
- shape_id: This column contains the description of the particular geospatial shape of the route for the corresponding trip.

4.1.4. Stop (stops.txt)

The stops file contains information regarding the points or stops, their schedules and locations which are available for public access.

- stop_id: This column contains a unique ID or identifier for each bus stop. The stop ID mut be specified.
- stop_name: The name given to the stop, commonly found in schedules and maps to assist passengers in identifying their stops.
- stop_lat: This column contains the latitude coordinates of every point of the stop location associated with the unique stop ID.
- stop_lon: This column contains the longitude coordinates of every point of the stop location associated with the unique stop ID.

4.2. Weights Calculation

In step 2, we will use standardization process to covert the different variables of the two metric to common scale. this process allows us to compare the two metrics and implement them in our model. We will also determine or calculate the weights of each metric utilizing the standardized variables. The weights will be assigned to each metrics or criterion which will help us to achieve accurate decision making process for optimizing sensor deployment.

Steps for metric scale standardization and weights calculations:

1. **Standardization:**

   We will first use the standardization process for transforming the two different metrics or criteria into a common scale. This process allows us to compare the two metrics and implement them in our model.

We will use z-score standardization process for this conversion where we will use their mean and standard deviation values.

- **Z-score Calculation:** We apply the z-score standardization process to both metrics of area coverage and trip frequency separately. These two metrics values are very different variables, as they have different scales, units, measures and also different dimensions. We have mentioned the z-score formulation in Section 3.3.1, to verify the results of the converted values are valid, the Mean of the z-scores of each metric should be equal to 0, and the standard deviation of each metric z-scores should be equal to 1. We will utilize z-score values in calculating the weights we will assign to each metric value.

2. **Weight Calculation:**

As mentioned earlier we will use the standardized z-score values to calculate the weights that we will assign to each metric value of each route which will help us in determining the relative importance of each route sensing coverage and help us in decision making regarding bus route and stop selection for sensor placement.

- **Weight Derivation from z-scores:** We use the calculated z-scores to calculate the weights, because the z-score ensures that no single metric variable disproportionally influences the weights. This process is important to prevent bias in our model and in our calculations and when assigning the weights to each metric. This disproportional effect occurs because the two metrics in question have different scales, units, measures and different dimensions also. For example, if a specific route provides extensive area coverage but has a low trip frequency, weight adjustment ensures that both factors are considered, even if they operate on different scales of influence.

3. **Weight Integration:**

We will now integrate the calculated weights into our proposed model. We will assign the weights to each metric or criteria which will help us to achieve an accurate decision making process for optimizing sensor deployment. The weight represent the relative importance or the contribution of each metric proportionally. The integration of weights help us in quantifying the sensing area coverage importance of each bus route which will help us in identifying optimizing the sensor deployment.

- **Incorporation into Optimization Modes:** The assigned weights are applied to our formulation which combines the two metrics into a single score representing area coverage of each route. To validate the weights calculation, we sum the weights of the two metrics of each route which is equal to 1 (W1+W2=1). In our proposed approach we utilize weighted sum method (WSM) in our formulation providing a linear combination of the multiple criteria. The WSM allows easy integration of different metrics, computationally efficient and flexible in handling qualitative and quantitative criteria making it suitable for various optimization and decision making methods.

4.3. Route Selection

We have used the weights of the two metrics to quantify the sensing coverage contribution into a single score. We will utilize the greedy approximation to efficiently select the subset of bus routes that provides the optimum sensor coverage while remaining within budget constraints. The process iteratively selects the best possible solution in order to optimize the bus-based sensor deployment on the selected routes. This process repeats under different budget constraints that represents different scenarios to find the best possible solution in varying budget constraints. Further description of the process is given below:

Optimize bus route subset selection process using Greedy approach:

1. **Initialization:**
    - Set up the initial conditions for the preferred number of bus routes by declaring the variables which include the quantified sensing coverage scores, budget constraints and route related dataset like IDs.
2. **Evaluate Routes:**
    - Evaluate each bus route in terms of network coverage considering spatial and temporal dimensions.
    - This contribution is determined via weighted metrics such as the land coverage for each route or frequency how often buses run time.
3. **Iterative Selection:**
    - Resets the budget constraints each time the process repeats to select the best possible set of bus routes that provide optimum sensing coverage at every iteration.
    - As bus routes are numerically ordered with every selection, the combined sensing coverage of the route selected will be optimized starting with most important routes.
4. **Maximizing Coverage:**
    - The system continues to add more routes to increase spatial extent and sampling frequency, to ensure mobile sensor placement is optimized.
5. **Finalizing the Selection:**
    - This process continues to select the best possible set of bus routes restricting itself to budget constraints until all the set of routes in the study has been selected and numerically listed according to their importance.
    - This ensures that the final outcome includes a set of bus routes.

4.4. Bus Stop Selection

Following the bus route selection the process includes the selection of bus stops for reference sensors placement to guarantee k-hop calibrability of the mobile sensors which are located along the selected bus routes. The importance of this stage is maintaining the deployed mobile sensor calibration for collecting reliable and accurate air quality data. As the deployed bus-based sensors intersect the bus stops deployed with reference sensors, which serve as meeting points or rendezvous points for frequent single hop calibration. The aim of this stage is to calculate the minimum number of reference sensors while guaranteeing that all selected bus routes have sufficient calibration coverage or opportunities.

Input:

1. **SelectedRoutes:** The subset of bus routes after route optimization.
2. **BusStops:** Dataset that contain bus stop coordinates, IDs, and the list of intersecting routes.
3. **StopRouteIntersecting:** Intersecting bus stops and bus routes stored in a pre-processed dictionary.
4. ***k*:** Maximum number of hops allowed for calibration.

Process for Bus Stop Selection for Reference Station Placement:

1. **Selection Process:**
    - Extract the set of variable required to initialize the process. Set of variable contain the subset of bus routes selected at varying budget constraints, bus stop dataset include bus stop location (with coordinates), and IDs, and bus stops intersecting each bus route.
    - Selection of bus stop also consists of budget constraints; the objective of the process is to minimize the number of bus stops for reference sensor placement to ensure calibration of the bus routes selected in step 3 of route selection process only.

    Method Description:

    - **Filtering:** Filter only those bus stops intersecting the selected bus routes (using StopRouteIntersecting) from the bus stop dataset).
    - **Data Preparation:** Setup a temporary list for bus routes intersecting the bus stops during iteration to ensure efficient access to save processing time.
    - **Initialization:** Create two empty sets or lists that will store the following:
        - Selected Bus stops.
        - Already covered Routes.
    - The bus stops selection is updated during each iteration and setting up the temporary list ensure efficient access to data.

1. $\boldsymbol{S}$**:** Set of bus stop candidates.
2. $\boldsymbol{R}$**:** Set of bus routes selected in the route optimization process in Step 3.
3. $\boldsymbol{N_{rs}}$**:** Binary relation.

   Binary relation = 1 if bus stop s∈S, can calibrate route r∈R (within k-hop range), otherwise 0

2. Calculate Coverage Overlap:

- We calculate the coverage for every bus stop intersecting bus routes will provide if the bus stop is selected for reference sensor placement. Bus stops intersecting more bus routes have high coverage and if the bus stops are selected, they can calibrate more bus-based sensors on the routes they are deployed on.
- The goal is to select the minimum number of bus stops for reference sensor stations and their geolocations that provide the highest calibration opportunities for the mobile sensors on the selected bus routes in step 3 of the route selection process.
- Bus routes also intersect other bus routes therefore they are also included in this calculations for multi hop calibration, and determine the minimum number of bus stops required to calibrate the bus based sensors over multiple hops.
- This allows a single reference sensor station deployed on a bus stop to increase its reach over hops.

Method Description:

- **For each bus stop**:
  - Acquire the list of routes from StopRouteIntersecting and find the bus routes associated with the bus stop.
  - Compare the list with the current intersected bus routes set.
  - Count the number of routes intersected by each bus stop can be calibrated.
- **Coverage Score:** The number of bus routes a bus stop can calibrate becomes its coverage score.
- **Goal:** Find the set of bus stops that can maximize coverage.
- **Multi-hop Case:** Account for the bus stops that indirectly connect to other bus routes (i.e., routes intersecting other routes which do not intersect the bus stop) in multi-hop (k > 1) calibration through rendezvous with freshly calibrated mobile sensor.
- For bus stop s covering more routes are expressed as:
  - For each bus stop $s \in S$, define:

$$s = \sum_{r \in R} N_{rs}$$

3. **Select Stops Based on Maximum Coverage:**
    - Bus stops intersect bus routes, therefore, if the bus stop is selected with high coverage will calibrate more bus-based sensors. The bus stop selected will be the one that intersects with the highest number of routes with the most calibration possibilities at one place.
    - The proposed algorithm will select the bus stops with the highest coverage for single hop and multi hop calibration.

    Method Description:

    - **Coverage Evaluation:** Lookup StopRouteIntersecting data for each candidate bus stop and find the number of bus routes it intersects.
    - **Candidate Selection:** Select the candidate bus stop that intersect with the maximum number of bus routes.
    - **Update State:** Add the selected bus stop (s) to the set of Bus Stops Selected (S). Add the new bus routes (r) to the set of Selected Routes (R).
    - At each iteration, for each bus stop $s \in S$, compute:
      Select $s \in S$ such that

$$s = argmax_{s \in S}\, Coverage\ (s)$$

4. **Iterate until all Routes are Covered:**
    - After every iteration to select the best possible set of bus stops that maximizes calibration opportunities, the budget constraint is reset.
    - Moreover, the system also provides the geolocations for each bus stop selected using the bus stop dataset (coordinates).

    Method Description:

    - **Repeat Evaluation:** Repeats evaluation of all remaining bus stops to identify the next best candidate that can calibrate the remaining uncalibrated routes.
    - **Updates Set:** After each selection, updates the set of:
        - Bus stops Selected S
        - Routes calibrated R

- **Stopping Condition:** The iteration stops when every selected subset of bus routes has a candidate bus stop for reference sensor placement that can directly calibrate them or via k-hop proximity.

$$\sum_{s \in S} y_s \,.\, N_{rs} \geq 1 \qquad \forall_r \in SelectedRoutes$$

This ensures that each selected route r is covered by at least one bus stop s such that the distance relation N_{rs} satisfies the k-hop calibration condition.

5. **Minimize the Number of Bus Stops**:
    - The main objective of this step is to find the least number of bus stops required for reference stations installation for calibrating the bus-based sensors. This step is also for calibrating the bus-based sensors that do not intersect directly with the selected bus stop achieving multi hop calibration. This step can be expressed as:

$$minimize \sum_{s \in S} y_s$$

6. **Final Placement of Reference Stations:**
    - In this final step the geo location of the selected bus stops are given where the reference sensors will be placed.

Method Description:

- **Objective:** Following the selection of the best possible set of bus stops, their geolocation or coordinates are shown where the reference sensors will be placed.
- **Output:** The finalized list of reference sensor placements is deployed to bus stops $s \in S$ where $y_s = 1$.

$$Final\ Set\ of\ Stops = \{s \in S \mid y_s = 1\}$$

Chapter 5: Results and Discussion

This chapter provides a detailed discussion and results of this study and experiment. We begin with collection, and description of the bus-transit dataset from the GTFS format and processed data which includes segmentation and produce more useful dataset to help in solving the formulated problem. We have used a bus transit dataset from two cities of Toronto, CA and Manchester, UK. We test the proposed optimization model and discuss and investigate how using z-scores standardization values and assigned weights affect the sensing coverage. We have also conducted comparative analysis of alternative methods to solving the optimization coverage problem. Our analysis of implementing our optimization coverage method of two cities with different road networks has also been discussed and their results presented. We have provided the results of our optimization approach of selecting the right subset of bus routes for achieving maximum sensing coverage and minimum number of bus stops required to k-hop calibrate the selected bus routes. We also discuss how the increasing number of k-hops affects the number of bus stops for reference sensors placement. Finally, we present our findings and list the contributions and potential applications in urban air quality monitoring. The contents included in this chapter is based on the paper published by the author of this thesis, "Bus-Based Sensor Deployment for Intelligent Sensing Coverage and k-Hop Calibration" [235], and has been adjusted to align with the objectives of this thesis.

5.1. Dataset Description

The proposed approach was tested based on the bus transit datasets of Toronto, CA [35] and Manchester, UK [35]. We relied on General Transit Feed Specification (GTFS) files, which are in the format of structure that public transport data should be provided in, as well. GTFS files also carried in-depth know-how relating to the bus routes and bus stops which even included their unique identifications (IDs), their names, geolocation coordinates, trajectories of bus routes and their trips information.

5.1.1. Toronto Dataset

We have pre-processed 188 bus routes and 9,464 bus stops for the Toronto dataset which have unique IDs and we have calculated the time period of the number of trips for each routes on a weekly basis. The pre-processed dataset of Toronto city bus transit is provided in Table 7. We have calculated the values in the table from the GTFS data which provides important geo-spatial information which we have used in the optimization model. We applied different buffer distances to the routes in spatial terms, to calculate three different geo-spatial measurements. The three buffered distances we have employed on each bus route are 25m, 50m, and 100m.

Table 7: Pre-processed Toronto city bus transit dataset.

Metrics	Values 25m buffer	Values 50m buffer	Values 100m buffer
Mean Route Area	672711.06 m^2	1323514.03 m^2	2595680.88 m^2
Median Route Area	582669.81 m^2	1146834.25 m^2	2244301.15 m^2
Max Route Area	2753012.46 m^2	5210444.85 m^2	9732316.15 m^2
Min Route Area	124141.73 m^2	243590.95 m^2	464329.93 m^2
Total Route Area	67475973.36 m^2	128420674.8 m^2	237130165.2 m^2

5.1.2. Manchester Dataset

We have also pre-processed 659 bus routes and 13,780 bus stops for the Manchester dataset which have unique IDs and we have calculated the time period of the number of trips for each routes on a weekly basis. The pre-processed dataset of Manchester city bus transit is presented in Table 8. We have used the same method to calculate geo-spatial data as the Toronto dataset.

Table 8: Pre-processed Manchester city bus transit dataset.

Metrics	Values 25m buffer	Values 50m buffer	Values 100m buffer
Mean Route Area	631014.92 m^2	1254472.26 m^2	2496575.26 m^2
Median Route Area	574590.41 m^2	1142592.53 m^2	2282959.28 m^2
Max Route Area	2920782.33 m^2	5844375.25 m^2	11699992.90 m^2
Min Route Area	34214.65 m^2	71230.06 m^2	153663.21 m^2
Total Route Area	153249842.5 m^2	260923824.5 m^2	448628671.4 m^2

5.1.3. Buffer Distance and Pollutant Dispersion

We employed the buffer distances of various distances (25m, 50m and 100m) for bus routes as the possible sensing ranges or distances for mobile sensors in our study. Computational fluid dynamics (CFD) models have been frequently used to predict pollution concentrations in the near vicinity of roadways in air pollution dispersion studies. These models consider factors such as vehicular emissions, meteorological conditions, morphologies of cities to predict the dispersion of pollutants. A study [238] showed that when traffic-related air pollutants decay in concentration with distance from the road by different decay rates at near-road range (i.e. 20 –100 meters) as compared to further ranges (i.e. 100 –300 meters). The author suggests that near-road zones are mainly affected by primary direct traffic emissions and the zones that were further away were affected by other factors. Based on this study, we have used variable buffer

distances concerning bus routes (25m, 50m and 100m) as a representative of sensing ranges or distances as they are applicable for mobile sensors in this work.

The distances we are interested in are between 20m and 100m as it better represents near to road traffic emissions. Distances further away may impact the air pollution measurements which could be influenced by factors such as urban infrastructure or meteorological factors. Hence, we are assuming that the maximum distance for air pollutant dispersion is 100m. The area of each route calculated will be based on the pollutant dispersion distances or range. This approach of using multiple buffer distances helps to analyze how these different ranges affect the strategy for sensor nodes placement. In our model, the 25m buffer range allows us to evaluate immediate roadside air pollution levels, and the buffer distance of 50m provides a balanced approach in which we cover a wider area. For maximum air pollutant dispersion range with a buffer distance of 100m, which provides extensive sensing coverage.

5.1.4. Calibration and Rendezvous Ranges

Reference sensors or calibrated mobile sensors need to be within rendezvous range for 1-hop or multi-hop calibration. Saukh, et al. [136], showed that the maximum spatial distance between sensors should be 50m and a maximum time of 5 minutes for successful rendezvous to take place. Results in another work done by Xiang, et al. [137], suggest that the ideal rendezvous calibration distances for ultrafine particle monitors and black carbon monitors, are 10–80 m and 0–30 m, respectively. Based on these studies we have concluded that the bus stops which are selected for geolocation for reference sensor placement are within the selected buffer distances or air pollutant dispersion range for successful k-hop calibration.

5.2. Evaluation of Bus Route Model for Maximal Spatio-Temporal Coverage

In both Toronto and Manchester, the sensing coverage optimization results of the greedy algorithm as well as random algorithm are significantly different in their coverage efficiency. The greedy algorithm, where we iteratively choose bus routes which cover as many areas as possible while satisfying the budget (B) and area coverage constraints, performs considerably better. The selection of route is also optimized with the respective weights for area (w_A) and trips (w_T).

5.2.1. Toronto and Manchester Dataset Results

The graphs in Figure. 5 shows the total route area coverage in square meters as a function of selected routes comparing the proposed greedy algorithm with the baseline random algorithm at different buffer distances (in meters). The trend can be clearly seen in the figure where total route area coverage increases as we add more routes to the baseline random algorithm. By comparison, our proposed greedy algorithm consistently outperforms the baseline random algorithm for the number of selected routes in terms of route area

coverage. We have implemented greedy algorithm on Toronto dataset and the results show that to achieve 50% of the total area coverage the greedy approach proposed that only 42 bus routes at 25m buffer and 50m buffer distances, and proposed 39 bus routes at 100m buffer distance. To achieve the total area coverage of 70% the greedy algorithm proposed 76 bus routes at 25m buffer distance, 74 bus routes at 50m buffer distance and at 100m buffer distance 73 bus routes are proposed. The random baseline algorithm in contrast is less efficient as to achieve the same 50% of the total area coverage the random approach proposed 66 bus routes at 25m, and 64 bus routes at buffer distances of 50m and 100m. The results have also demonstrated that for the greedy approach, as the buffering distance increases, the number of bus routes selected decreases to cover the same percentage of area which is due to assigning weight assigned to each route based on their importance in terms of area and trips.

In Figure 6, where we show Manchester bus transit dataset for the total area coverage in which we use the same parameters as the Toronto bus-transit dataset shown in Figure 3. In Figure 4, we have compared the proposed greedy approach and the random baseline algorithm trendlines. As observed with the Toronto results, the greedy algorithm consistently shows improvements in the correlation between the number of selected routes and area coverage. In contrast to our proposed greedy approach, the results from random baseline algorithm shows relatively less efficient trend as less area coverage with same number of bus routes as greedy approach. When implementing greedy algorithm for the Manchester dataset, the algorithm proposed 142 bus routes at 25m buffer distance, 136 bus routes at 50m buffer distance and 105 bus routes at 100m buffer distance in order to achieve 50% area coverage. The random baseline algorithm in contrast is less efficient as to achieve the same 50% of the total area coverage the random approach proposed 191 bus routes at 25m, and 154 bus routes at buffer distance of 50m and at 100m buffer distance the proposed bus routes are 140. The graph in Figure 4 shows a similar increase in area coverage trendline as the selected number of bus routes increases. The comparison shows that the performance of our proposed greedy approach is better than the baseline algorithm when applied to both cities dataset for optimizing sensing coverage.

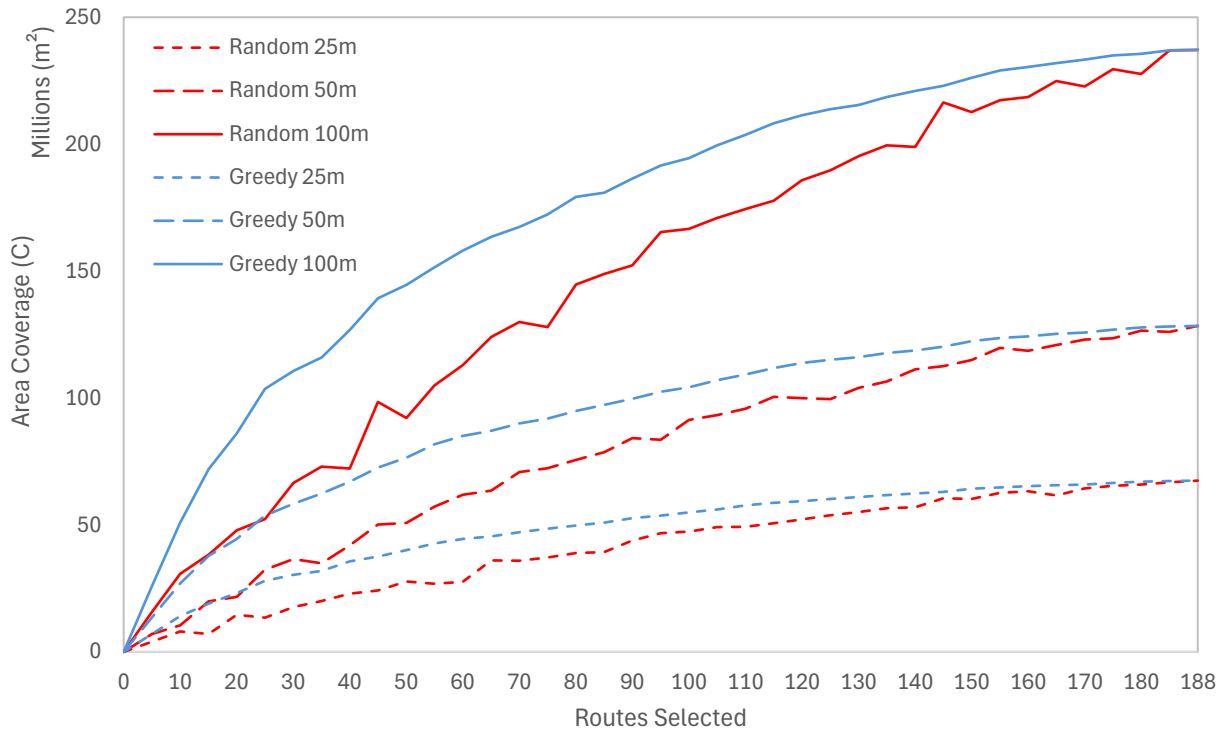


Figure 5: Greedy (in blue) and random (in red) algorithm routes optimize for Toronto city for 25m, 50m and 100m buffered width of bus routes.

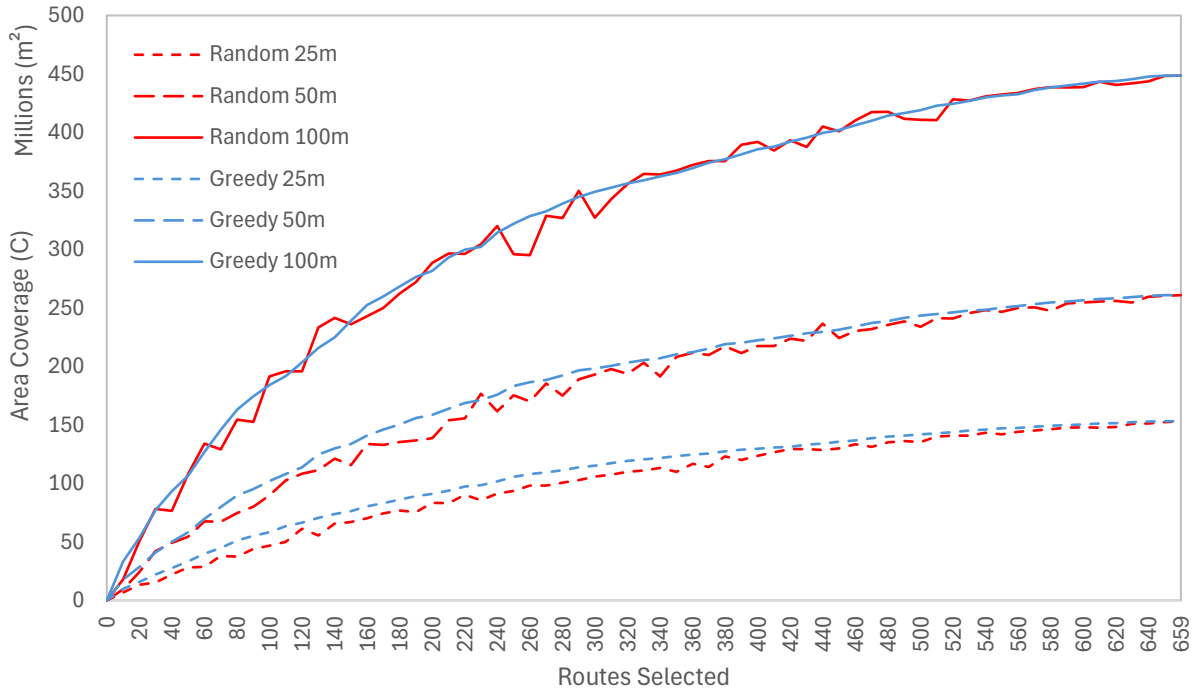


Figure 6: Greedy (in blue) and random (in red) algorithm routes optimization for Manchester city for 25m, 50m and 100m buffered width of bus routes.

5.2.2. Effect of Z-scores and Weight on Coverage Model

The assigned weights are calculated based on two key metrics 1. Route area coverage, and 2. Trips per route or trip frequency. The z-score standardization is used so that no single metric variable will disproportionally influence the calculation of the assigned weights, which represents the relative importance or the contribution of each metric and help us in quantify the sensing area coverage importance of each bus route.

A. Weight Calculation:

Weights have been derived from two standardized metrics or criteria:

- Area Coverage: This criteria provides the spatial extent of which the sensing coverage is possible for each route. We have buffered the bus routes at three different distances (25m, 50m, and 100m). The buffer distances are from the center of the route which represent the different settings for the criterion of both immediate and extended areas influenced by the bus routes.
- Trip Frequency: This particular metric provides the temporal aspect of the sensing coverage. Which means that the number of trips buses on each route take during a specific period of time. This metric is important due to its contribution in sensing coverage in temporal context. Where the higher the number of trips of buses the more the sampling rate at each point of the bus route.

By including these metrics we can investigate the impact of weight calculation method, in this experiment we have compared the results of our proposed approach against alternative approaches whereby each route area and trips are used separately as raw data (or absolute route area and trips respectively) without the standardization process or calculating and assigning weights. We are using an alternative approach for our comparative analysis for a better understanding of the performance of our optimization model which includes coefficient weights while the alternative approaches do not.

B. Alternative Method Comparison:

We conducted a comprehensive analysis by plotting the total route area coverage as a function of the number of selected routes for both the weighted approach and the alternative method. In this comparative analysis we are comparing the performance of the weighted and non-weight or single-metric method, therefore, we are using the greedy algorithm to calculate the results in both methods. The main differences in coverage efficiency between these two approaches were examined to understand the impact of our weight calculations. See Figures 6 and 7. This comparison is conducted to isolate the individual contributions of spatial (area coverage) and temporal (trip frequency) factors, and to validate whether combining them through calculated weights leads to more balanced and effective sensor deployment. In the single-metric approach, raw, unstandardized values of area coverage and trip frequency were used independently–on

optimization performance without transformation. Standardization process is used when combining multiple metrics or criterion (like in our weighted approach) to ensure they are on the same scale with proportional contribution. However, in the single-metric approach, only one variable is used at a time, with no risk of scale disproportion between features and will not improve the analysis. Two alternative methods, objective functions:

Maximize:

$$\sum_{r \in R} x_r \,.\, A_r \qquad (11)$$

$$\sum_{r \in R} x_r \,.\, T_r \qquad (12)$$

where:

- x_r is a binary decision variable indicating if route r is selected (1) or not (0).
- A_r is the raw area covered by each route r.
- T_r is the raw trips covered by each route r.

Constraints:

1. **Budget Constraint:**

$$\sum_{r \in R} C_r . x_r \leq B \qquad (13)$$

- C_r is the cost of selecting route r.
- B is the total budget available for route selection.

2. **Binary Decision Variables:**

$$x_r \in \{0,1\} \qquad (14)$$

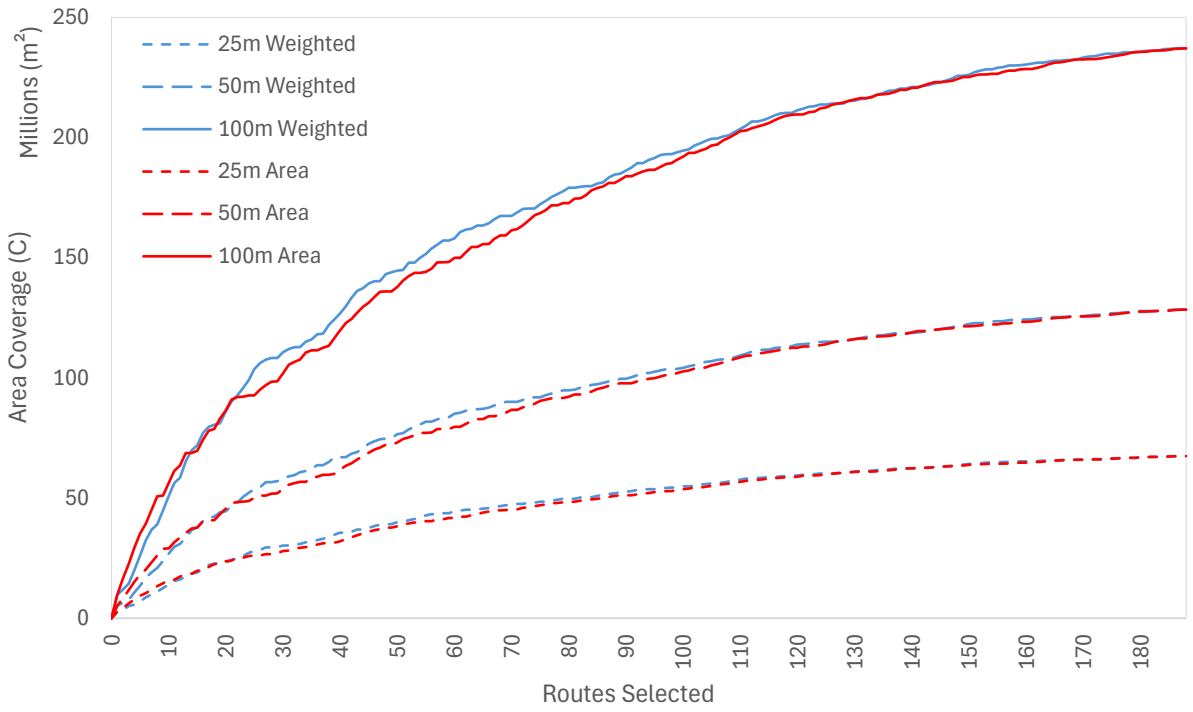


Figure 7: Comparison of weighted (blue) and absolute route area (red) approaches using greedy method for Toronto city at 25m, 50m, and 100m buffer widths.

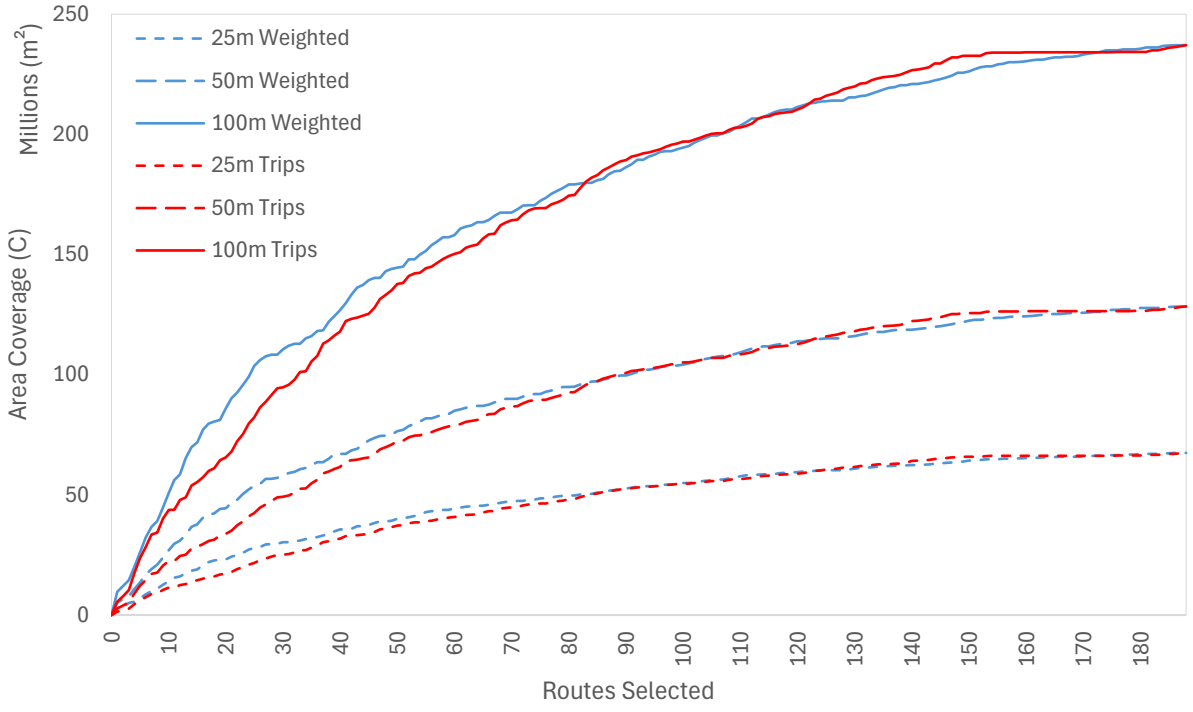


Figure 8: Comparison of weighted (blue) and absolute route trips (red) approaches using greedy method for Toronto city at 25m, 50m, and 100m buffer widths.

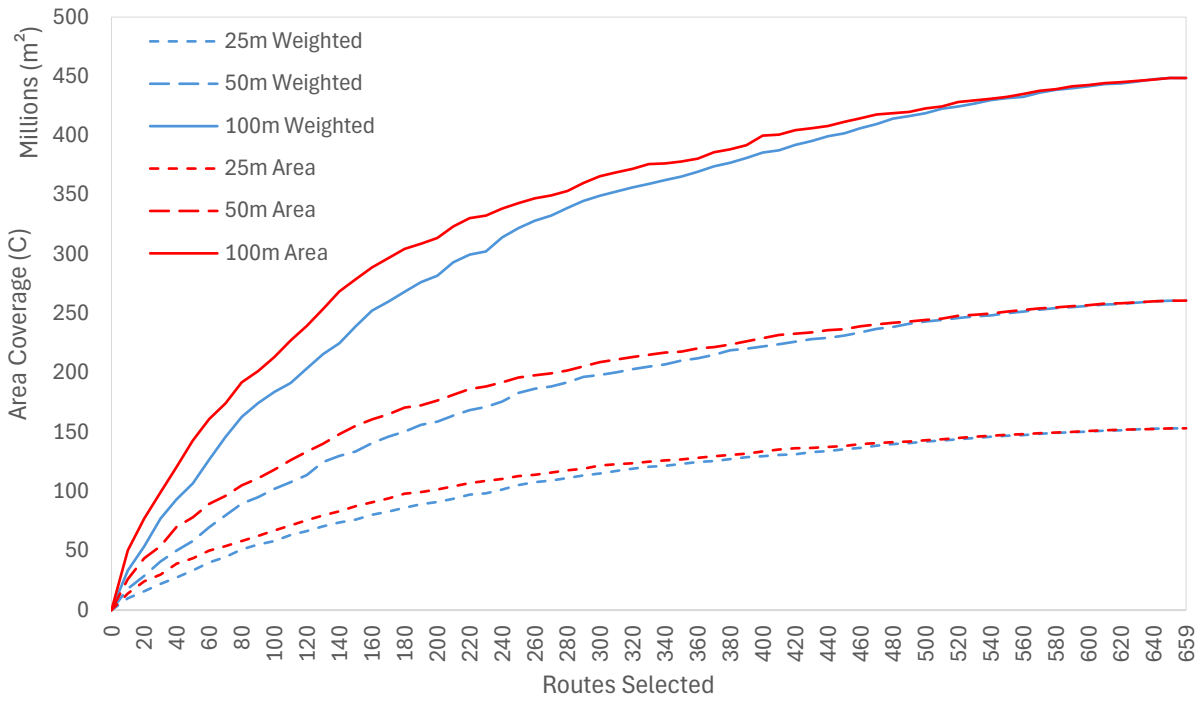


Figure 9: Comparison of weighted (blue) and absolute route area (red) approaches using greedy method for Manchester city at 25m, 50m, and 100m buffer widths.

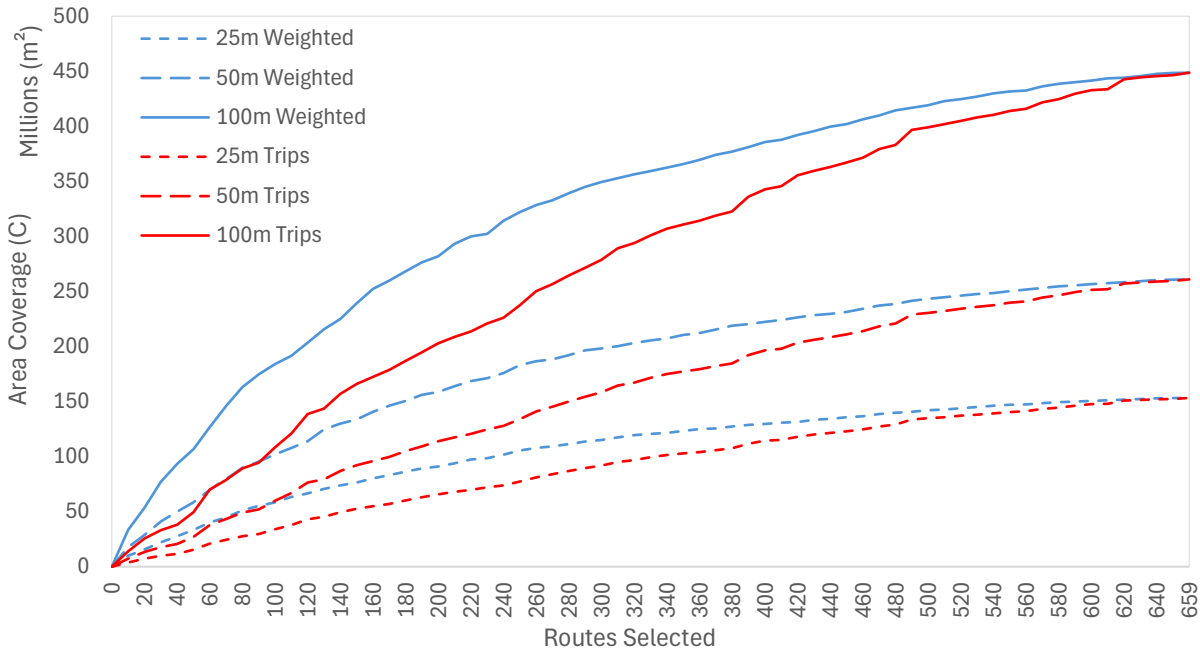


Figure 10: Comparison of weighted (blue) and absolute route trips (red) approaches using greedy method for Manchester city at 25m, 50m, and 100m buffer widths.

5.3. Minimal Bus Stop Selection Model Evaluation

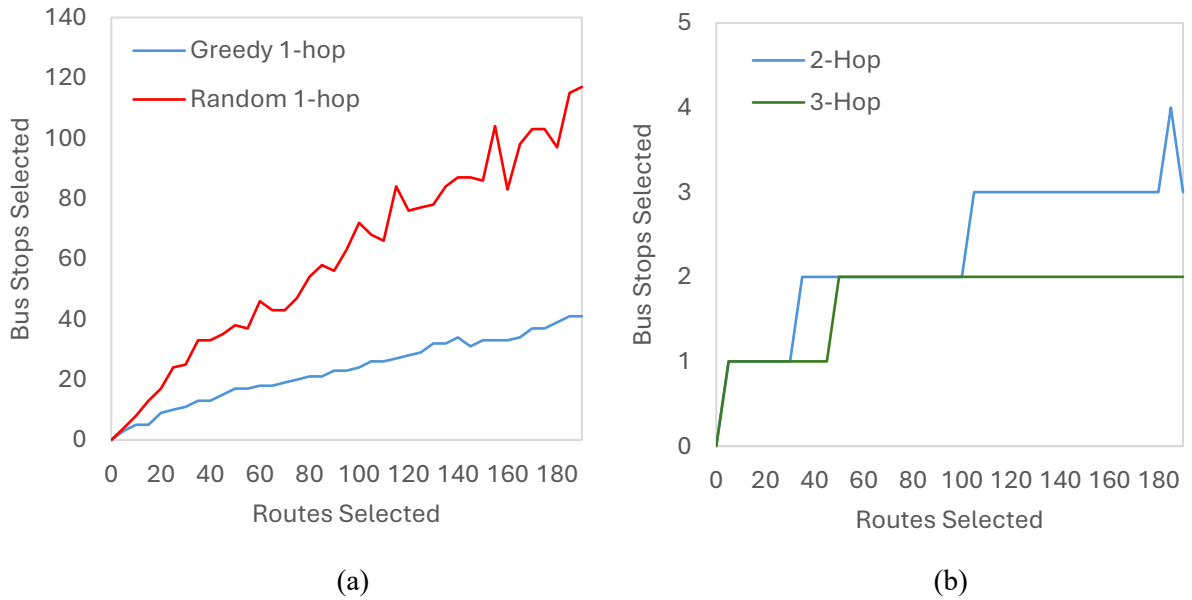
The results of the reference station geolocation optimization using the greedy algorithm and the random algorithm for both Toronto and Manchester show clear differences in coverage efficiency. The optimization of bus stop placements for reference stations ensured k-hop calibration of the selected bus routes, using variables such as the number of bus stops required (N_{sr}). The greedy algorithm was able to effectively reduce the total number of bus stops while ensuring that all the bus-based sensors on the selected bus routes were k-hop calibrated.

5.3.1. Toronto and Manchester Dataset Results

For this section we have considered single hop and multi hop sensor calibration in which we have three criteria. Our first experiment will only consider 1-hop calibration of mobile sensors while for the second and third experiment which are multi hop, we will consider 2-hop and 3-hop calibration. We have implemented greedy algorithm on Toronto dataset and the results show that to 1-hop calibrate 10 and 85 bus routes subset, only 5 and 21 bus stops are proposed respectively, for reference sensor placement. While only 21 bus stops have been proposed for reference stations placement to 1-hop calibrate the entire set of 188 bus routes. The result for multi hop approach show further efficiency, to successfully calibrate bus-based sensors deployed on all 188 bus routes using 2-hop approach, the greedy algorithm proposed only 3 bus stops for reference sensor placement, while for 3-hop approach only 2 bus stops were proposed.

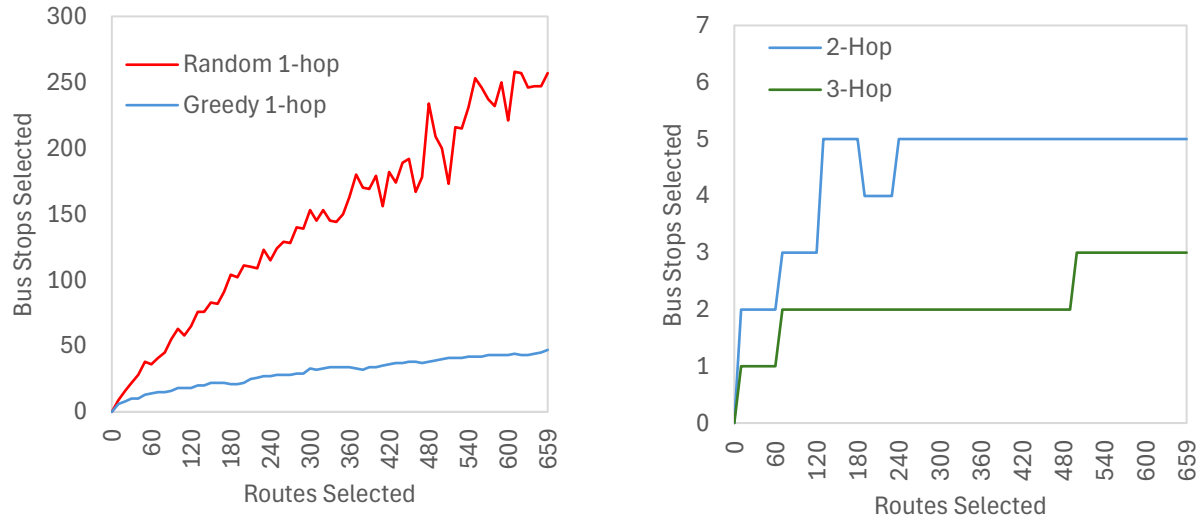
The results from using the Manchester dataset show a similar trendline, where the implemented greedy algorithm proposed 10 bus stops for reference sensor placement to 1-hop calibrate bus-based sensors on 30 selected bus routes, and proposed 18 and 25 bus stops for reference sensor deployment to 1-hop calibrate bus-based sensors operating on 110 and 210 selected bus routes, respectively. While only 47 bus stops have been proposed for reference stations placement to 1-hop calibrate the entire set of 659 bus routes. The results for multi hop approach show further efficiency, to successfully calibrate bus-based sensors deployed on all 659 bus routes using 2-hop approach, the greedy algorithm proposed only 5 bus stops for reference sensor placement, while for 3-hop approach only 3 bus stops were proposed. The random algorithm is less effective in both datasets, necessitating more bus stops for reference station deployment to achieve the same level of calibration as the greedy algorithm. The findings the results indicate that the greedy algorithm consistently outperforms the random algorithm in minimizing the number of bus stops required for effective calibration across both datasets. The efficiency of our proposed greedy algorithm is apparent in selecting the minimum number of bus stops for reference sensor deployment and achieving maximum area coverage for 1-hop calibration of bus-based sensors deployed on the selected bus routes. The k-hop calibration results demonstrate that increasing the number of hops reduces the number of required bus stops as illustrated in Figures 9 and 10.

In summary, the evaluation of the minimal bus stop selection model underscores the effectiveness of the greedy algorithm in optimizing bus stop selection for sensor calibration, achieving significant improvements over the random algorithm. Our approach minimizes the number of bus stops which serve as a geolocation for reference sensor placement, for k-hop calibrating the bus-based sensors deployed on the selected bus routes, therefore reducing the resource allocation and lowering the overall sensor deployment and overhead costs.



(a) (b)

Figure 11: Reference station placement for Toronto dataset: (a) comparison between greedy and random algorithm for 1-hop, (b) comparison between 2-hop and 3-hop calibration using greedy algorithms.



(a) (b)

Figure 12: Reference station placement for Manchester dataset: (a) comparison between greedy and random algorithm for 1-hop, (b) comparison between 2-hop and 3-hop calibration using greedy algorithms.

5.3.2. Interpretation of Bus Stop Selection Patterns

The selection of bus stops for reference station deployment revealed consistent patterns in the number of stops selected. For instance, the number of bus stops proposed by greedy approach were 17 bus stops for 1-hop calibration of 48 to 56 subset of bus routes, for the Toronto bus-transit dataset. The reason for this situation occurs when more and more bus routes overlap each other, the marginal benefit of selecting higher number of bus stops for reference sensor deployment reduces as the current number of selected bus stops is enough to k-hop calibrate additional bus routes. A similar pattern can be seen in other parts of the bus stop selection trendline within the Toronto dataset as well as in the Manchester dataset. This is also seen in the multi hop approach trendlines as seen in Figures 9(b) and 10(b), therefore, requiring less reference stations, thus reducing resource allocation. During the process of selecting bus stops for reference station deployment, fluctuations in the number of bus stops selected were observed in both 1-hop and multi-hop scenarios. For example, the proposed number of bus stops required for reference sensor deployment to 1-hop calibrate bus-based sensors deployed on 58 selected bus routes are 18, however in order to 1-hop calibrate the bus-based sensors deployed on 58 selected bus routes the proposed number of bus stops required are 17. The reduction in the bus stop number required for reference sensor placement can be attributed to the varying bus route locations. The selection process of the number and geolocation of bus stops for reference sensor placement to k-hop calibrate completely depends on the selection of the bus routes and their geolocation. For instance, it is not necessary that all the new set of bus routes that are selected have the same geolocation as the set of bus routes that were selected previously. The decrease in the number of bus stops as the number of routes increases might also be due to the threshold effect, where adding new routes does not immediately require additional bus stops due to existing coverage. This reason could be due to the redundancy of a number of bus stops for reference station placement as they no longer provide sufficient coverage for k-hop calibration.

In summary, the findings highlighted the performance of the greedy algorithm in identifying key bus stops that maximize calibration opportunities with minimal resources. The fluctuations in the selected number of bus stops for reference stations placement were observed in both 1-hop and multi-hop scenarios and consistent patterns in the bus stops selection highlights the importance of selecting bus routes that overlap and provide marginal coverage which benefits in optimizing reference sensor placement.

5.4. Discussion

The evaluation of the results from the Toronto and Manchester bus-transit datasets show that our proposed weighted method performed better in certain domains in terms of spatio-temporal coverage when compared to the alternative methods where we used raw or absolute data of area coverage and trip frequency. This improvement was particularly seen in the initial stages of the experiment and the trendlines remain similar

even when the three different buffering distances are applied. In the initial stages of the experiment, our proposed weighted approach showed a logarithmic improvement in sensing area coverage because our model considers the contribution of area coverage and trip frequency proportionally of each route. At the start of route selection, several routes that provide high coverage are available. In the beginning of the bus route selection process the routes that provided high spatio-temporal coverage are always available. Our weighted method helps in identifying the routes that have high importance and selecting them, which leads to a rapid increase in spatio-temporal coverage. However, as the number of selected routes increases, the rate of improvement decreases. The bus routes that are less likely to be selected in the initial stages typically have a lower contribution to the overall spatio-temporal coverage. At this stage, both the weighted method and the alternative method start selecting routes that are less critical, and the advantage of using weight decreases because the remaining routes have relatively uniform impact in terms of raw area coverage. Around the selection of 80 routes, the performance gap between the two methods narrows. This is because our approach has already selected most of the routes which contribute the most, and the remaining routes offer similar contributions to the total unique coverage. In the initial stages of the experiment which involved the Toronto bus transit dataset, the results of the alternative methods showed less improvements than our proposed approach. However, as the number of bus routes increased for both methods, the alternative methods' performance improved. Since the remaining routes have lower but relatively uniform coverage, the difference in the methods' effectiveness reduces. The evaluation of the results from our experiment showed that our proposed weighted approach performed better in the beginning. However, as the experiment progressed, the alternative methods showed improvement in their performance as the number of bus routes selected increased.

In addition, the results from the random approach showed significantly lower performance in the Toronto dataset compared to Manchester as shown in Figures 5 and 6 in Section 5.2.1. The difference in the results could be due to the geographical features of the two cities' transit networks. For instance, the number of bus routes and bus stop of Toronto dataset in 188 and 10,869, which reduces the chances of selecting the routes that provide high coverage when randomly selected. In comparison, Manchester city has a significantly larger network, with 659 bus routes and 36,365 bus stops, and has a higher degree of route overlapping one another. This increases the chances of randomly selecting the bus routes that provide higher area coverage. This results in a random approach having better performance which can appear closer to the performance of the greedy approach. The different road network topology is another contributing factor where Toronto seems to have a more grid-like road layout, with more distinct non-overlapping routes, which increases the variation in random route selection. In comparison, the Manchester road network layout is more radial and irregular, which increases the chances of selecting routes with overlaps providing relevant area coverage.

In summary, the overall results for using the two cities dataset showed that our proposed weighted method performed better in certain domains, such as the early stages of route selection, where it achieved higher unique area coverage with fewer routes selected, and where the integration of spatial and temporal metrics led to more efficient sensing. The results of the experiment conducted with Manchester bus transit dataset showed that the alternate method which uses raw or absolute route area measurement performed better than our proposed weighted method, in the later stages of route selection, where most high-coverage routes had already been selected and the remaining routes contributed similarly to overall coverage. However, the trendline of our proposed approach remains closely aligned with the trendline of the alternate method throughout the experiment until the gap begins to close as the number of selected routes approaches the total number of bus routes available. This suggests that although the alternative method shows slightly better results in the Manchester dataset, our weighted method still performs similarly overall. As more routes are selected, the difference between the methods becomes smaller because the remaining routes contribute less to total coverage. As seen in the trendlines that our proposed approach and the alternate method remains closely aligned throughout the experiment until the gap closes, it shows that our method is still effective. This results could also be attributed to the inclusion of trips frequency metric into our proposed method, as sampling rate also needs to be considered for air quality monitoring. Additionally, when comparing the inclusion of raw trips data, the results show that in the Toronto dataset, the trendline for raw trips data is similar to that of the raw area data. However, in the Manchester dataset, the trendline for raw trips data is notably below both our weighted approach and the method using raw area data. The results of integrating trip frequency metric into the alternate method for optimizing bus routes selection showed lower performance compared to the results of when routes area metric is used with our proposed weighted method.

Additionally, the results of the bus stop selection process further highlight the efficiency of the weighted method, particularly in minimizing the number of bus stops required for effective sensor calibration. The optimization of bus stop placements for reference stations ensured k-hop calibration of the selected bus routes, showing that the greedy algorithm effectively minimized the total number of selected bus stops. In the Toronto dataset, the greedy algorithm required only 5 reference stations to calibrate bus-based sensors on 10 routes, and 41 reference stations for all 188 bus routes. In comparison, for a successful calibration of bus-based sensors deployed on all 188 bus routes using 2-hop approach, the greedy algorithm proposed only 3 bus stops for reference sensor placement, while for 3-hop approach only 2 bus stops were proposed. The results from using the Manchester dataset show a similar trendline, where the implemented greedy algorithm proposed 10 bus stops for reference sensor placement to 1-hop calibrate bus-based sensors on 30 selected bus routes, and proposed 25 bus stops for reference sensor deployment for 210 subset of bus routes, while only 47 bus stops have been proposed for reference stations placement for the entire set of 659 bus

routes. The result for multi hop approach show further efficiency, to successfully calibrate bus-based sensors deployed on all 659 bus routes using 2-hop approach, the greedy algorithm proposed only 5 bus stops for reference sensor placement, while for 3-hop approach only 3 bus stops were proposed. The pattern of bus stop selection reveals efficiency gains with the greedy algorithm, as it consistently reduced the number of bus stops compared to the random approach. Additionally, fluctuations in the number of bus stops such as requiring fewer bus stops as the number of routes increases demonstrates how route overlaps and marginal coverage benefits affect stop selection.

As more routes are selected, the rate of improvement slows, and the difference between the weighted and alternative methods narrows as the remaining routes offer more uniform contributions to total coverage. The results for our proposed weighted method and the alternate methods showed variation in their performance depending on the data the alternate methods were utilizing. The alternate method using raw route area as a metric outperformed our weighted approach, specifically in cases where spatial coverage was the primary objective, such as in the Manchester dataset under conditions that emphasized total route area without accounting for trip frequency. However, our approach outperformed the alternate method using trip frequency as a metric. The results for minimal bus stop selection for reference stations using our proposed greedy approach outperformed the baseline random approach. The fluctuations in the number of bus stops for reference stations placement were observed in both 1-hop and multi-hop scenarios and consistent patterns in the bus stops selection highlights the importance of selecting bus routes that overlap and provide marginal coverage which benefits in optimizing reference sensor placement. Overall, the study highlights the strength of the weighted method in both route and bus stop selection, ensuring balanced spatial and temporal coverage while minimizing resource usage. While exact algorithms can provide optimal solutions, they often involve complex formulations and may prune decisions early based on local information. In our context, where spatial and temporal interactions significantly affect coverage, such pruning risks discarding routes or stops that could contribute meaningfully in later steps. Moreover, the greedy approach, by contrast, incrementally builds the solution based on maximum marginal gain and is less computationally intensive making it more suitable for large datasets.

Chapter 6: Conclusion

In this thesis, an optimization model has been proposed for bus-based sensors deployment on a subset of bus routes for maximizing the total unique sensing area and optimizing reference-grade sensors placement to maximize the probability of k-hop calibration. In our model we have developed a metric-based system by quantifying the contribution of each route and rank them accordingly by utilizing geographical set operations which includes spatial (route area-based) and temporal (bus trips frequency) joins. We have formulated the optimization problem as a Mixed Integer Linear Program (MILP) to select the best possible subset of bus routes for deployment of bus-based sensors and to maintain the same level of sensing and temporal coverage contribution as the total unique area. We have utilized a greedy heuristic approach to solve this problem and for performance evaluation we compare the results to random approach and alternative methods. We use real-world bus-transit data from two cities, Toronto, CA, and Manchester, UK. The contents included in this chapter is based on the paper published by the author of this thesis, "Bus-Based Sensor Deployment for Intelligent Sensing Coverage and k-Hop Calibration" [235], and has been adjusted to align with the objectives of this thesis.

6.1. Summary of Findings

For our model we have used a bus transit dataset of two cities, Toronto, CA and Manchester, UK. We have also conducted comparative analysis of alternative methods for solving the optimization coverage problem by using raw or absolute data instead of standardized dataset. We have also analyzed the implementation of our proposed optimization coverage approach to the two cities with different road networks or topologies and discussed the corresponding outcomes. The results includes the selection of the optimized subset of bus routes for mobile sensor deployment to achieve maximum sensing coverage and identifying the minimum number of reference sensors for placement on bus stops required to k-hop calibrate the mobile sensors on selected bus routes. Our proposed greedy approach consistently outperforms in comparison to the random approach in the selection of the number of routes and in terms of route area coverage. When we implement greedy algorithm on Toronto dataset, the results show that 50% of the total area coverage is achieved with a selection of only 42 bus routes at buffer distances of 25m and 50m, and selection of only 39 bus routes at buffer distance of 100m.

However, the random approach, in comparison, is less efficient as to achieve the same total area coverage of 50%, random approach selected 66 bus routes at a buffer distance of 25m, and selected 64 bus routes at 50m and 100m buffer distances. The results also demonstrated that as the buffering distance increases when using the greedy approach, the selected number of subset of bus routes decreases to achieve the same percentage of sensing area coverage. The results of applying greedy and random algorithm for comparison

using the Manchester bus transit dataset were not so dissimilar from the results Toronto dataset. To achieve 50% area coverage, the greedy approach selected 122 bus routes at 25m buffer distance, 116 bus routes at 50m buffer distance and 108 bus routes at 100m buffer distance. In comparison, the random approach is less efficient as to achieve the same 50% of the total area coverage, 191 bus routes were selected at 25m buffer distance, 154 bus routes were selected at buffer distance of 50m and at 100m buffer distance the bus routes selected are 140. The results of reference sensor optimization and their geolocation using the greedy and the random approach for both cities of Toronto and Manchester show clear differences in calibration coverage efficiency.

For our experiment we have selected reference sensors based on the number of hops, which have increased gradually until the number of reference sensors required for calibrating the entire network becomes constant. For example, we started with 1-hop or single hop calibration of mobile sensors while for the second and third experiment which are multi hop, we have 2-hop and 3-hop calibration. To calibrate the entire set of 659 bus routes of Manchester city, with 1-hop our proposed approach selected only 47 bus stops for reference stations placement, while for 2-hop approach the greedy algorithm selected only 5 bus stops for reference sensor placement. And for a 3-hop approach only 3 bus stops were selected for reference sensors, and as the number of hops increased the selected bus stops remained only 3, indicating that 3 reference sensors should be enough to multi hop calibrate the entire mobile network. In comparison, the random approach is less effective in both cites datasets, requiring more bus stops for reference station deployment to achieve calibration of the entire mobile network as a greedy approach.

The evaluation of the results from the bus-transit dataset of both Toronto and Manchester cities show that our proposed weighted method's performance is better in certain domains, such as coverage optimization with initial route selection, where with fewer routes a higher unique area coverage was achieved, by utilizing the combined use of spatial and temporal metrics for improved efficiency. In the experiment conducted with the alternate method which uses raw or absolute route area measurement using the bus transit dataset of Manchester city, the results showed that it performed better than our proposed weighted method. As the route selection progressed and most of the routes providing high coverage had already been selected and the remaining routes offered limited additional sensing coverage. However, the trendlines of our proposed weighted approach and the alternative approach remains closely aligned with each other throughout the experiment, and as the number of selected routes approaches the total number of bus routes selected where the gap between the trendline begins to close.

As the outcome of both methods suggest that while the alternative method in which absolute area as a metric is used shows slight improvement in the Manchester dataset. However, the effectiveness of our weighted method still remains comparable to the area based alternate method, particularly as route selection

increases. The results can be attributed to the trip frequency metric integrated into our weighted approach, which is important as air quality monitoring must also consider data sampling rate. Additionally, we have also analyzed the results of the alternate method where we included absolute trip data into the Toronto dataset, which showed that the trendline of both absolute trip data and area data closely resemble each other. In contrast, the trendline for the raw trip data is significantly lower than both our weighted approach and the alternate method based on raw area data, in the Manchester dataset. Additionally, temporal (trip frequency metric) data is integrated into the alternative method for optimizing bus route selection, resulting in a lower performance compared to using the spatial (area-based metric) data with our proposed weighted model.

6.2. Contributions

1. Comprehensive Literature Review: We have conducted a comprehensive review of existing literature regarding multiple aspects of drive-by sensing and air quality monitoring which includes the analysis the various air pollutants and their current concentration standards, the role of low-cost sensor technology for urban air quality monitoring, evaluation of multiple types of vehicular platforms for drive-by sensing which includes public transit, municipal transport, private vehicles, and unmanned aerial vehicles (UAVs). And limitation post-deployment or in-field calibration techniques of mobile sensors in dynamic urban environments and their opportunities. Moreover, our review also included the comparative analysis between achieving high spatial and temporal resolution or coverage in mobile sensor networks and their trade-offs. The findings of this comprehensive literature review were published as a survey paper, which provides an in-depth understanding of drive-by sensing as well as identifying the key research areas for further work. Moreover, we have included a comprehensive review of different types of optimization problems and algorithms to solve the various problems discussed which includes exact and approximate approach.

2. A Novel Optimization Framework: This framework we develop a metric-based system which quantifies each bus route contribution and prioritize each route accordingly. For this we consider bus transit geographical operations, in terms spatial (route area) and temporal (route trips) joins data collected from GTFS files. We formulate the max coverage problem as a MILP where we use weighted sum method (WSM) allowing our multi-objective optimization problem to be projected as a single-objective mathematical optimization problem. We normalize the two metrics, each route area and trips, using z-score standardization to make data points from our dataset metrics at a more comparable scale and to calculate each route coverage contribution and calculate the weights of each metrics. In addition to optimizing bus-based sensor deployment, we have calculated the minimum number of reference sensors and their

geolocation required for co-location to k-hop calibrate the deployed bus-based sensors on selected bus routes.

3. Comparative Analysis: Initially we utilized a greedy heuristic approach to solve the max coverage optimization problem, we used baseline random algorithm results to compare the performance of our weighted metric-based system. In addition to the random algorithm, we developed two alternative methods for comparative analysis where we used absolute or raw route area and trips data separately, without weighted metrics where we have also used greedy approximation to reduce computational complexities and time. Furthermore, we have applied our approach and alternative methods to two cities with two different road network layout to analyze the effect of all techniques mentioned earlier. The inclusion of the two alternative methods to a random baseline approach has given us an in-depth insight to the contribution to our weighted metric-based approach performance under different or dynamic environments.

Published research papers in journals (see Table 1 for summary):

1. Drive-by Air Pollution Sensing Systems: Challenges and Future Directions [16].

2. Mobile Calibration for Bus-Based Urban Sensing [32].

3. Bus-Based Sensor Deployment for Intelligent Sensing Coverage and k-Hop Calibration [235].

6.3. Limitations

Given that our proposed method has achieved satisfactory results, certain limitations should be mentioned. The principle function of greedy approximation is based on locally optimal selection to maximize or minimize local benefits, therefore, at each step near-optimal choices are made to achieve this. Therefore, this approach is more suitable for problems with a greedy selection and optimal substructure requirements, and may not be effective for more complex problems and have complex dependencies between variables, where locally optimal choices do not lead to a globally optimal solution. However, greedy algorithm is a popular tool due to its easy implementation and requires less processing time to find solution. The effectiveness of the greedy algorithm is measured by its approximation ratio, which provides how accurate the algorithm's solution is. Bus transit has fixed and predictable trajectories, which is an advantage, however they are not flexible and will follow the designated routes and the bus route also overlap with one another. The overlapping of bus routes is common, which may lead to redundant data collection, which cannot be avoided due to their fixed route plans. However, this overlapping nature in routes is utilized for rendezvous between calibrated bus-based sensors and uncalibrated bus-based sensors for multi-hop sensor calibration therefore avoiding redundant mobile sensor data collection is difficult. Therefore, the trade-off

between more opportunities for multi-hop calibration of bus-based sensors and reducing redundant sensor data collection is a challenge.

While higher k values can reduce the number of reference stations needed, they introduce several challenges. For $k > 1$, calibration depends on mobile sensors encountering each other within a specific distance and time threshold to relay calibration information. This requires overlapping bus routes and synchronized schedules to enable effective rendezvous events. As the value of k increases, the probability of rendezvous events decreases, due to route scheduling mismatches and sparse or irregular transit networks. In several scenarios the length of overlapping spatial joins may be too short, providing insufficient time for the calibration to occur. This spatial limitation, combined with the uncertainty of synchronized bus arrivals, makes successful rendezvous less probable as k-hop increases. Missed or brief encounters cause reduced rendezvous opportunities, creating gaps in the calibration chain, which increases the probability of error accumulation over hops reducing data quality.

Our proposed framework for optimizing bus-based air quality sensing can be affected by several critical factors that influence sensor deployment—namely, k-hop calibration strategy, sensor sensing range, and variation in low-cost sensor (LCS) types and their associated costs. As shown in Table 6, the sensor response time vary depending on sensor type which could affect the temporal aspect of coverage. Sensors with fast response time can collect data more frequently and in scenario where sensors with slow response time are utilized will require the trips frequency of bus based sensors on the routes to be higher. LCS are also commercially available with varying costs depending on sensor type and manufacturer, and different LCS measure different or specific types of air pollutants. Depending on the deployment scenario and requirements, different types of LCS can be required which will change or increase the deployment cost. Sensor selectivity can affect the reliability of the measured data as multiple substances will be present in the environment. Low-cost sensors with low selectivity will require modifications to prevent other air pollutants from interfering with the measurements which will also increase the sensor network cost. Although our model uses MILP formulation, which include cost variables, however, the varying costs cannot be determined due to the factors mentioned earlier, we have considered the uniform cost of LCS. This removes or reduces the effect of cost on our framework and focuses on route contribution for deployment optimization. Temporal coverage is also included in our formulation, however, depending on the deployment scenario different LCS types may be used, which have different response times that can effect data measurement frequency. Similar to cost we have also considered uniform data measurement frequency to remove the effect of response time of the LCS on temporal resolution.

6.4. Future Work

There are several areas where this research model can be extended, where bus transit can be used as mobile platforms, due to its various advantages. The areas may include monitoring meteorological factors (e.g., temperature, humidity, etc.), noise level, traffic flow. Several smart city initiatives, for instance urban planning, resource management, and public safety can also follow this research model to provide data. However, bus-transit as a mobile platform lacks flexibility, as buses must follow designated routes, and these routes often overlap with one another, providing redundant coverage. To achieve more flexibility (if required), other vehicular platforms such as taxi cars, unmanned aerial vehicles, etc., can be utilized. Further discussion on alternative modes of vehicular platforms for drive-by sensing is provided in section 2.2. Further research can be done to achieve more balance between the spatial and temporal trade-offs, sensing redundant data and achieving higher multi-hop calibration opportunities. The greedy algorithm is popular due to several benefits; however, it does not guarantee an optimal solution, especially when applied to more complex problems. Therefore, to further optimize the model more complex algorithms can be utilized such as bound and box, hill climb and genetic algorithm. We have also used buffered distances to simulate air pollutant dispersion around the routes area, however, incorporating more environmental conditions such as meteorological conditions, traffic conditions can be incorporated in the model to better understand and simulate air pollutant dispersion. In this thesis work, we incorporated two metrics—route area coverage and bus transit trip count—derived from the GTFS dataset; however, additional metrics available within the GTFS or extracted through further processing can be introduced to improve and expand the model.

References

[1] A. C. Kemp, B. P. Horton, J. P. Donnelly, M. E. Mann, M. Vermeer and S. Rahmstorf, "Climate related sea-level variations over the past two millennia," *Proceedings of the National Academy of Sciences,* vol. 108, no. 27, pp. 11017-11022, 2011.

[2] L. Bai, J. Wang, X. Ma and H. Lu, "Air Pollution Forecasts: An Overview," *International Journal of Environmental Research and Public Health,* vol. 15, no. 4, 2018.

[3] "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019," vol. 396, pp. 1204-1222, 2020.

[4] P. Sicard, E. Agathokleous, S. C. Anenberg, A. D. Marco, E. Paoletti and V. Calatayud, "Trends in urban air pollution over the last two decades: A global perspective," *Science of The Total Environment,* vol. 858, 2023.

[5] United Nations, "The World's cities in 2018," *The World's Cities in,* p. 34, 2018.

[6] T. Feng, Y. Sun, Y. Shi, J. Ma, C. Feng and Z. Chen, "Air pollution control policies and impacts: A review," *Renewable and Sustainable Energy Reviews,* vol. 191, 2024.

[7] M. Pavani and P. Rao, "Urban air pollution monitoring using wireless sensor networks: A comprehensive review," *International Journal of Communication Networks and Information Security (IJCNIS),* vol. 9, no. 3, pp. 439-449, 2017.

[8] D. Agarwa, S. Iyengar, M. Swaminathan, E. Sharma, A. Raj and A. Hatwar, "Modulo: Drive-by Sensing at City-scale on the Cheap," in *COMPASS '20: Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*, 2020.

[9] P. Han, H. Mei, D. Liu, N. Zeng, X. Tang, Y. Wang and Y. Pan, "Calibrations of Low-Cost Air Pollution Monitoring Sensors for CO, NO2, O3, and SO2," *Sensors,* vol. 21, no. 1, 2021.

[10] L. Li, X. Zhang, J. B. Holt, J. Tian and R. Piltner, "Spatiotemporal Interpolation Methods for Air Pollution Exposure," in *Proceedings of the Ninth Symposium on Abstraction, Reformulation and Approximation*, 2011.

[11] P. H. C. Caminha, R. d. S. Couto, L. H. M. K. Costa, A. Fladenmuller and M. D. d. Amorim, "On The Coverage of Bus-Based Mobile Sensing," *Sensors (Basel),* vol. 18, no. 6, 2018.

[12] J. S. Apte, K. P. Messier, S. Gani, M. Brauer, T. W. Kirchstetter, M. M. Lunden, J. D. Marshall, C. J. Portier, R. C. Vermeulen and S. P. Hamburg, "High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data," *Environmental Science & Technology,* vol. 51, no. 12, 2017.

[13] V. Sivaraman, J. Carrapetta, K. Hu and B. G. Luxan, "HazeWatch: A participatory sensor system for monitoring air pollution in Sydney," in *38th Annual IEEE Conference on Local Computer Networks - Workshops*, Sydney, NSW, Australia, 2014.

[14] J. Ali and V. Dyo, "Coverage and Mobile Sensor Placement for Vehicles on Predetermined Routes: A Greedy Heuristic Approach," in *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) - WINSYS*, Madrid, Spain, 2017.

[15] R. M. A.-P. Hernandez-Jayo, F. Alías and I. Angulo, "Design of a Mobile Low-Cost Sensor Network Using Urban Buses for Real-Time Ubiquitous Noise Monitoring," *Sensors,* vol. 17, no. 1, 2017.

[16] H. Zarrar and V. Dyo, "Drive-by Air Pollution Sensing Systems: Challenges and FutureDirections," *IEEE Sensor Journal,* vol. 23, no. 19, pp. 23692-23703, 2023.

[17] S. Kaivonen and E. C.-H. Ngai, "Real-time air pollution monitoring with sensors on city bus," *Digital Communications and Networks,* 2020.

[18] Y. M. Idir, O. Orfila, V. Judalet, B. Sagot and P. Chatellier, "Mapping Urban Air Quality from Mobile Sensors Using Spatio-Temporal Geostatistics," *Sensors,* vol. 21, no. 14, 2021.

[19] P. Wei, P. Brimblecombe, F. Yang, A. Anand, Y. Xing, L. Sun, Y. Sun, Mengyuan and C. Z. Ning, "Determination of Local Traffic Emission and Non-Local Background Source Contribution to On-Road Air Pollution Using Fixed-Route Mobile Air Sensor Network," *Environmental Pollution,* vol. 290, 2021.

[20] J. J. Li, B. Faltings, O. Saukh, D. Hasenfratz and J. Beutel, "Sensing the Air We Breathe — The OpenSense Zurich Dataset," in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[21] Y. Gao, W. Dong, K. Guo, X. Liu, Y. Chen, X. Liu, J. Bu and C. Chen, "Mosaic: A low-cost mobile sensing system for urban air quality monitoring," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016.

[22] P. Cruz, R. S. Couto, L. H. M. K. Costa, A. Fladenmuller and M. D. d. Amorim, "Per-Vehicle Coverage in a Bus-Based General-Purpose Sensor Network," *IEEE Wireless Communications Letters,* vol. 9, no. 7, pp. 1019 - 1022, 2020.

[23] C. Paliwal and P. Biyani, "Dense Air Quality Maps Using Regressive Facility Location Based Drive By Sensing," *arXiv:2201.09739,* 2022.

[24] S. M. Biondi, V. Catania, S. Monteleone and C. Polito, "Bus as a sensor: A mobile sensor nodes network for the air quality monitoring," in *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Rome, Italy, 2017.

[25] R. Choudhary and A. Agarwa, "Optimization of Bus Routes to Maximize," in *6th Conference of the Transportation Research Group of India (CTRG-2021)*, 2021.

[26] L. H. Anh, C. M. Hai, N. P. Truong, T.-H. Nguyen and P. L. Nguyen, "Optimal Deployment of Vehicular Mobile Air Quality Monitoring Systems," in *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, 2020.

[27] A. Arfire, A. Marjovi and A. Martinoli, "Model-based rendezvous calibration of mobile sensor networks for monitoring air quality," *IEEE SENSORS,* pp. 1-4, 2015.

[28] Z. Idrees and L. Zheng, "Low cost air pollution monitoring systems: A review of protocols and enabling technologies," *Journal of Industrial Information Integration,* vol. 17, 2020.

[29] M. Kamionka, P. Breuil and C. Pijolat, "Calibration of a multivariate gas sensing device for atmospheric pollution measurement," *Sensors and Actuators B: Chemical,* vol. 118, no. 1-2, pp. 323-327, 2006.

[30] M. Mueller, J. Meyer and C. Hueglin, "Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich," *Atmospheric Measurement Techniques,* vol. 10, no. 10, 2017.

[31] S. D. Vito, E. Massera, M. Piga, L. Martinotto and G. D. Francia, "On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario," *Sensors and Actuators B: Chemical,* vol. 129, no. 2, pp. 750-757, 2008.

[32] H. Zarrar, M. Limbu, S. Haxha and V. Dyo, "Mobile Calibration for Bus-based Sensing," *IEEE Sensor Journal,* vol. 25, no. 3, pp. 5576 - 5583, 2024.

[33] Y.-C. Chen, P.-Y. Chen and C.-Y. Wen, "Distributed bus information management for mobile weather monitoring," in *2017 15th International Conference on ITS Telecommunications (ITST)*, Warsaw, Poland, 2017.

[34] P. Castello, C. Muscas, P. A. Pegoraro, D. Sitzia and S. Sulis, "Influence of Apparent Wind on Particulate Matter Monitoring through Low-Cost Sensors," in *2024 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0 & IoT)*, Firenze, Italy, 2024.

[35] "OpenMobilityData," [Online]. Available: https://transitfeeds.com/.

[36] D. J. Sun, S. Chen, C. Zhang and S. Shen, "A bus route evaluation model based on GIS and super-efficient data envelopment analysis," *Transportation Planning and Technology,* vol. 39, no. 4, pp. 407-423, 2016.

[37] M. Huang, X. Li, M. Yang and X. Kuai, "Intelligent coverage and cost-effective monitoring: Bus-based mobile sensing for city air quality," *Computers, Environment and Urban Systems,* vol. 108, 2024.

[38] Z. Dai and K. Han, "Exploring the drive-by sensing power of bus fleet through active scheduling," *Transportation Research Part E: Logistics and Transportation Review,* vol. 171, p. 103029, 2023.

[39] J. J. Yu, V. O. Li and A. Y. Lam, "Sensor deployment for air pollution monitoring using public transportation system," in *2012 IEEE Congress on Evolutionary Computation*, Brisbane, QLD, Australia, 2012.

[40] W. Ji, K. Han and Q. Ge, "Route planning of mobile sensing fleets for repeatable environmental monitoring tasks," *Computers, Environment and Urban Systems,* vol. 119, p. 102285, 2025.

[41] K. D. Zoysa, C. Keppitiyagama, G. P. Seneviratne and W. W. A. T. Shihan, "A public transport system based sensor network for road surface condition monitoring," in *NSDR '07: Proceedings of the 2007 workshop on Networked systems for developing regions*, 2007.

[42] W. Ji, K. Han and T. Liu, "Trip-based mobile sensor deployment for drive-by sensing with bus fleets," *Transportation Research Part C: Emerging Technologies,* vol. 157, p. 104404, 2023.

[43] L. Curtis, W. Rea, P. Smith-Willis, E. Fenyves and Y. Pan, "Adverse health effects of outdoor air pollutants," *Environment International,* vol. 32, 2006.

[44] S. Vardoulakis, B. E. Fisher, K. Pericleous and N. Gonzalez-Flesca, "Modelling air quality in street canyons: a review," *Atmospheric Environment,* vol. 37, 2003.

[45] United States Environmental Protection Agency, "National Ambient Air Quality Standards (NAAQS)," [Online]. Available: https://www.epa.gov/criteria-air-pollutants. [Accessed 01 November 2022].

[46] World Health Organization, Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide, World Health Organization, 2006.

[47] Queensland Government, "Queensland Government," [Online]. Available: https://www.qld.gov.au/environment/management/monitoring/air/air-pollution/pollutants. [Accessed 7 11 2022].

[48] J. Raub, "Health effects of exposure to ambient carbon monoxide," *Chemosphere - Global Change Science,* vol. 1, 1999.

[49] Y. M. Kim, S. Harrad and R. M. Harrison, "Concentrations and Sources of VOCs in Urban Domestic and Public Microenvironments," *Environmental Science & Technology,* vol. 35, 2001.

[50] United States Environmental Protection Agency (EPA), "Volatile Organic Compounds' Impact on Indoor Air Quality," [Online]. Available: https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality. [Accessed 10 11 2022].

[51] M. S. Jassim and G. Coskuner, "Assessment of spatial variations of particulate matter (PM10 and PM2.5) in Bahrain identified by air quality index (AQI)," *Arabian Journal of Geosciences,* vol. 10, 2017.

[52] European Commission (EC), "EU air quality standards," [Online]. Available: https://environment.ec.europa.eu/topics/air/air-quality_en. [Accessed 11 November 2022].

[53] World Health Organization (WHO), "What are the WHO Air quality guidelines?," [Online]. Available: https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines. [Accessed 11 November 2022].

[54] Department for Environment Food & Rural Affairs (Defra), UK, "Daily Air Quality Index (DAQI)," [Online]. Available: https://uk-air.defra.gov.uk/air-pollution/daqi. [Accessed 11 November 2022].

[55] U.S. Environmental Protection Agency, "Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI)," 2018.

[56] J. Li, H. Zhang, C.-Y. Chao, C.-H. Chien, C.-Y. Wu, C. H. Luo, L.-J. Chen and P. Biswas, "Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level PM2.5," *Atmospheric Environment,* vol. 223, p. 117293, 2020.

[57] N. Stradiotto, H. Yamanaka and M. V. Zanoni, "Electrochemical sensors: a powerful tool in analytical chemistry," *Journal of the Brazilian Chemical Society,* vol. 14, no. 2, 2003.

[58] A. Nasri, M. Pétrissans, V. Fierro and A. Celzard, "Gas sensing based on organic composite materials: Review of sensor types, progresses and challenges," *Materials Science in Semiconductor Processing,* vol. 128, 2021.

[59] M. Tonezzer and L. V. Duy, "Gas Sensors," *Encyclopedia of Sensors and Biosensors,* vol. 3, pp. 185-208, 2023.

[60] M. M. Arafat, B. Dinan, S. A. Akbar and A. S. M. A. Haseeb, "Gas Sensors Based on One Dimensional Nanostructured Metal-Oxides: A Review," *Sensors,* vol. 12, no. 6, 2012.

[61] M. Wusiman and F. Taghipour, "Methods and mechanisms of gas sensor selectivity," *Critical Reviews in Solid State and Materials Sciences,* vol. 47, no. 3, pp. 416-435, 2022.

[62] P. Arroyo, J. Gómez-Suárez, J. I. Suárez and J. Lozano, "Low-Cost Air Quality Measurement System Based on Electrochemical and PM Sensors with Cloud Connection," *Sensors,* vol. 21, no. 18, 2021.

[63] E. Bales, N. Nikzad, N. Quick, C. Ziftci, K. Patrick and W. Griswold, "Citisense: Mobile air quality sensing for individuals and communities Design and deployment of the Citisense mobile air-quality system," in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, San Diego, CA, USA, 2012.

[64] M. Mead, O. Popoola, G. Stewart, P. Landshoff and M. Calleja, "The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks," *Atmospheric Environment,* vol. 70, pp. 186-203, 2013.

[65] S. S. M. Alabdullah, A. Z. M. AL-Bassam and N. Asaad, "Electrochemical sensors and its applications," *International Journal of Research in Engineering and Innovation (IJREI),* vol. 5, no. 5, pp. 258-262, 2017.

[66] A. M. Collier-Oxandale, J. Thorson, H. Halliday, J. Milford and M. Hannigan, "Understanding the ability of low-cost MOx sensors to quantify ambient VOCs," *Atmospheric Measurement Techniques,* vol. 12, no. 3, 2019.

[67] N. Kularatna and B. H. Sudantha, "An Environmental Air Pollution Monitoring System Based on the IEEE 1451 Standard for Low Cost Requirements," *IEEE Sensors Journal,* vol. 8, 2008.

[68] M. Bart, D. E. Williams, B. Ainslie, I. McKendry, J. Salmond, S. K. Grange, M. Alavi-Shoshtari, D. Steyn and G. S. Henshaw, "High Density Ozone Monitoring Using Gas Sensitive Semi-

Conductor Sensors in the Lower Fraser Valley, British Columbia," *Environmental Science & Technology,* vol. 48, no. 7, 2014.

[69] G. Parmar, S. Lakhani and M. K. Chattopadhyay, "An IoT based low cost air pollution monitoring system," in *2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)*, Bhopal, India, 2017.

[70] M. Firdhous, B. Sudantha and P. Karunaratne, "IoT enabled proactive indoor air quality monitoring system for sustainable health management," in *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*, Chennai, India, 2017.

[71] AnanyaDey, "Semiconductor metal oxide gas sensors: A review," *Materials Science and Engineering: B,* vol. 229, 2018.

[72] N. Masson, R. Piedrahita and M. Hannigan, "Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring," *Sensors and Actuators B: Chemical,* vol. 208, 2015.

[73] M. Fleischer and H. Meixner, "Selectivity in high-temperature operated semiconductor gas-sensors," *Sensors and Actuators B: Chemical,* vol. 52, 1998.

[74] C. Zhang, Y. Luo, J. Xu and M. Debliquy, "Room temperature conductive type metal oxide semiconductor gas sensors for NO2 detection," *Sensors and Actuators A: Physical,* vol. 289, pp. 118-133, 2019.

[75] H. Chai, Z. Zheng, K. Liu, J. Xu, K. Wu, Y. Luo, H. Liao, M. Debliquy and C. Zhang, "Stability of Metal Oxide Semiconductor Gas Sensors: A Review," *IEEE Sensors Journal,* vol. 22, no. 6, 2022.

[76] R. K. Jha, "Non-Dispersive Infrared Gas Sensing Technology: A Review," *IEEE Sensors Journal,* vol. 22, 2022.

[77] A. Sklorz, A. Schafer and W. Lang, "Merging ethylene NDIR gas sensors with preconcentrator-devices for sensitivity enhancement," *Sensors and Actuators B: Chemical,* vol. 170, 2012.

[78] L. B. Mendes, N. W. M. Ogink, N. Edouard, H. J. C. V. Dooren, I. D. F. F. Tinôco and J. Mosquera, "NDIR Gas Sensor for Spatial Monitoring of Carbon Dioxide Concentrations in Naturally Ventilated Livestock Buildings," *Sensors,* vol. 15, no. 5, pp. 11239-11257, 2015.

[79] M. Müller, P. Graf, J. Meyer, A. Pentina, D. Brunner, F. Perez-Cruz, C. Hüglin and L. Emmenegger, "Integration and calibration of non-dispersive infrared (NDIR) CO2 low-cost sensors and their operation in a sensor network covering Switzerland," *Atmospheric Measurement Techniques,* vol. 13, no. 7, pp. 3815-3834, 2020.

[80] D. Liu, W. Sun, N. Zeng, P. Han, B. Yao, Z. Liu, P. Wang, K. Zheng, H. Mei and Q. Cai, "Observed decreases in on-road CO2 concentrations in Beijing during COVID-19 restrictions," *Atmospheric Chemistry and Physics,* vol. 21, no. 6, p. 4599–4614, 2021.

[81] T.-V. Dinh, I.-Y. Choi, Y.-S. Son and Jo-ChunKim, "A review on non-dispersive infrared gas sensors: Improvement of sensor detection limit and interference correction," *Sensors and Actuators B: Chemical,* vol. 231, 2016.

[82] X. Su, L. Sutarlie and X. J. Loh, "Sensors and Analytical Technologies for Air Quality: Particulate Matters and Bioaerosols," *Chemistry – An Asian Journal,* vol. 15, no. 24, 2020.

[83] M. Badura, P. Batog, A. Drzeniecka-Osiadacz and P. Modzel, "Optical particulate matter sensors in PM2.5 measurements in atmospheric air," in *10th Conference on Interdisciplinary Problems in Environmental Protection and Engineering EKO-DOK 2018*, 2018.

[84] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia and X. Jiang, "AirCloud: a cloud-based air-quality monitoring system for everyone," in *SenSys '14: Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, 2014.

[85] M. Gao, J. Cao and E. Seto, "A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China," *Environmental Pollution,* vol. 199, pp. 56-65, 2015.

[86] H. Cho and Y. Baek, "Practical Particulate Matter Sensing and Accurate Calibration System Using Low-Cost Commercial Sensors," *Sensors,* vol. 21, 2021.

[87] Y. Wang, J. Li, H. Jing, Q. Zhang, J. Jiang and P. Biswas, "Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement," *Aerosol Science and Technology,* vol. 49, no. 11, 2015.

[88] L. R. Crilley, M. Shaw, R. Pound, L. J. Kramer, R. Price, S. Young, A. C. Lewis and F. D. Pope, "Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring," *Atmospheric Measurement Techniques,* vol. 11, no. 2, 2018.

[89] M. Aleixandre and M. Gerboles, "Review of small commercial sensors for indicative monitoring of ambient gas," *Chemical Engineering Transaction,* vol. 30, pp. 169-174, Sep 2012.

[90] R. Baron and J. Saffell, "Amperometric Gas Sensors as a Low Cost Emerging Technology Platform for Air Quality Monitoring Applications: A Review," *ACS Sensors,* vol. 2, no. 11, p. 1553–1566, 2017.

[91] C. Borrego, M. C. J. Ginja, C. Ribeiro, K. Karatzas, T. Sioumis, N. Katsifarakis, K. Konstantinidis, S. D. Vito, E. Esposito, M. Salvato, P. Smith, N. André, P. Gérard, L. Francis, N. Castell, P. Schneider, M. Viana, M. C. Minguillón, W. Reimringer, R. P. Otjes, O. v. Sicard, R. Pohle, B. Elen, D. Suriano, V. Pfister, M. Prato and S. Dipi, "Assessment of air quality microsensors versus reference methods: The EuNetAir Joint Exercise – Part II," *Atmospheric Environment,* vol. 193, pp. 127-142, 2018.

[92] A. C.Rai, P. Kumar, F. Pilla, A. N.Skouloudis, S. D. Sabatino, C. Ratti, A. Yasar and D. Rickerby, "End-user perspective of low-cost sensors for outdoor air pollution monitoring," *Science of the Total Environment,* vol. 607–608, pp. 691-705, 2017.

[93] F. Concas, J. Mineraud, E. Lagerspetz, S. Varjonen, X. Liu, K. Puolamäki, P. Nurmi and S. Tarkoma, "Low-Cost Outdoor Air Quality Monitoring and Sensor Calibration: A Survey and Critical Analysis," *ACM Transactions on Sensor Networks,* vol. 17, 2021.

[94] Y. Kang, L. Aye, T. D. Ngo and J. Zhou, "Performance evaluation of low-cost air quality sensors: A review," *Science of The Total Environment,* vol. 818, 2022.

[95] F. Karagulian, M. Barbiere, A. Kotsev, L. Spinelle, M. Gerboles, F. Lagler, N. Redon, S. Crunaire and A. Borowiak, "Review of the Performance of Low-Cost Sensors for Air Quality Monitoring," *Atmosphere,* vol. 10, no. 9, 2019.

[96] B. Maag, Z. Zhou and L. Thiele, "A Survey on Sensor Calibration in Air Pollution Monitoring Deployments," *IEEE Internet of Things Journal,* vol. 5, no. 6, 2018.

[97] 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. Hagler, R. Jayaratne, P. Kumar, A. K. Lau, P. K. Louie, M. Mazaheri, Z. Ning, N. Motta, M. M. Rahman, M. Md 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?," *Science of The Total Environment,* vol. 116, pp. 286-299, 2018.

[98] M. V. Narayana, D. Jalihal and S. M. S. Nagendra, "Establishing A Sustainable Low-Cost Air Quality Monitoring Setup: A Survey of the State-of-the-Art," *Sensors,* vol. 22, no. 1, 2022.

[99] B. Szulczynski and J. Gebicki, "Currently Commercially Available Chemical Sensors Employed for Detection of Volatile Organic Compounds in Outdoor and Indoor Air," *Environments,* vol. 4, no. 1, 2017.

[100] J. E. Thompson, "Crowd-sourced air quality studies: A review of the literature & portable sensors," *Trends in Environmental Analytical Chemistry,* vol. 11, pp. 23-34, 2016.

[101] C. Wang, L. Yin, L. Zhang and D. X. a. R. Gao, "Metal oxide gas sensors: Sensitivity and influencing factors," *Sensors,* vol. 10, no. 3, 2010.

[102] S. G. M. European Commission and Joint Research Centre and Yatkin, A. Borowiak and M. Signorini, "Guidance on low-cost air quality sensor deployment for non-experts based on the AirSensEUR experience," Publications Office of the European Union, 2022.

[103] Ö. Örnek and B. Karlık, "An overview of metal oxide semiconducting sensors in electronic nose applications," in *Proceedings of the 3rd International Symposium on Sustainable Development*, 2012.

[104] E. Esposito, S. D. Vito, M. Salvato, V. Bright, R. Jones and O. Popoola, "Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems," *Sensors and Actuators B: Chemical,* vol. 231, pp. 701-713, 2016.

[105] C. Wang, C. Li and X. Li, "Maximizing spatial–temporal coverage in mobile crowd-sensing based on public transports with predictable trajectory," *International Journal of Distributed Sensor Networks,* vol. 14, 2018.

[106] N. H. Tonekaboni, L. Ramaswamy, D. Mishra, O. Setayeshfar and S. Omidvar, "Spatio-Temporal Coverage Enhancement in Drive-By Sensing Through Utility-Aware Mobile Agent Selection," in *Internet of Things - ICIOT 2020*, Cham, 2020.

[107] O. Saukh, D. Hasenfratz, A. Noori, T. Ulrich and L. Thiele, "Route selection for mobile sensors with checkpointing constraints," in *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, 2012.

[108] T. Xi, W. Wang, E. C.-H. Ngai and X. Liu, "Spatio-Temporal Aware Collaborative Mobile Sensing with Online Multi-Hop Calibration," in *Mobihoc '18: Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2018.

[109] B. Predić, Z. Yan, J. Eberle, D. Stojanovic and K. Aberer, "ExposureSense: Integrating daily activities with air quality using mobile participatory sensing," in *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2013.

[110] S. Wang, Y. Ma, Z. Wang, L. Wang, X. Chi, A. Ding, M. Yao, Y. Li, Q. Li, M. Wu, L. Zhang, Y. Xiao and Y. Zhang, "Mobile monitoring of urban air quality at high spatial resolution by low-cost sensors: impacts of COVID-19 pandemic lockdown," *Atmospheric Chemistry and Physics,* vol. 21, no. 9, 2021.

[111] K. Yeom, "Development of urban air monitoring with high spatial resolution using mobile vehicle sensors," *Environmental Monitoring and Assessment,* vol. 193, 2021.

[112] K. P. O'Keeffe, A. Anjomshoaa, S. H. Strogatz, P. Santi and C. Ratti, "Quantifying the sensing power of vehicle fleets," in *Proceedings of the National Academy of Sciences of the United States of America*, 2019.

[113] X. Chen, S. Xu, J. Han, H. Fu, X. Pi, C. Joe-Wong, Y. Li, L. Zhang, H. Y. Noh and P. Zhang, "PAS: Prediction-Based Actuation System for City-Scale Ridesharing Vehicular Mobile Crowdsensing," *IEEE Internet of Things Journal,* vol. 7, 2020.

[114] X. Chen, S. Xu, X. Liu, X. Xu, H. Y. Noh, L. Zhang and P. Zhang, "Adaptive Hybrid Model-Enabled Sensing System (HMSS) for Mobile Fine-Grained Air Pollution Estimation," *IEEE Transactions on Mobile Computing,* vol. 21, 2022.

[115] X. Qin, T. H. Do, J. Hofman, E. R. Bonet, V. P. L. Manna, N. Deligiannis and W. Philips, "Fine-Grained Urban Air Quality Mapping from Sparse Mobile Air Pollution Measurements and Dense Traffic Density," *Remote Sensing,* vol. 14, no. 11, 2022.

[116] K. P. Messier, S. E. Chambliss, S. Gani, R. Alvarez, M. Brauer, J. J. Choi, S. P. Hamburg, J. Kerckhoffs, B. LaFranchi, M. M. Lunden, J. D. Marshall, C. J. Portier, A. Roy, A. A. Szpiro, R. C. H. Verme and J. S. A. , "Mapping air pollution with google street view cars: efficient approaches

with mobile monitoring and land use regression," *Environmental Sciences & Technologies,* vol. 52, pp. 12563-12572, 2018.

[117] S. E. Alexeeff, A. Roy, J. Shan, X. Liu, K. Messier, J. S. Apte, C. Portier, S. Sidney and S. K. V. D. Eeden, "High-resolution mapping of traffic related air pollution with Google street view cars and incidence of cardiovascular events within neighborhoods in Oakland, CA," *Environmental Health,* vol. 17, 2018.

[118] S. Chiesa, A. D. Pietro, M. Pollino and S. Taraglio, "Urban Air Pollutant Monitoring through a Low-Cost Mobile Device Connected to a Smart Road," *International Journal of Geo-Information,* vol. 11, no. 2, 2022.

[119] J. Wesseling, W. Hendricx, H. d. Ruiter, S. v. Ratingen, D. Drukker, M. Huitema, C. Schouwenaar, G. Janssen, S. v. Aken, J. W. Smeenk, A. Hof and E. Tielemans, "Assessment of PM2.5 Exposure during Cycle Trips in The Netherlands Using Low-Cost Sensors," *International Journal of Environmental Research and Public Health,* vol. 18, 2021.

[120] J. Gómez-Suárez, P. Arroyo, R. Alfonso, J. I. Suárez, E. Pinilla-Gil and J. Lozano, "A Novel Bike-Mounted Sensing Device with Cloud Connectivity for Dynamic Air-Quality Monitoring by Urban Cyclists," *Sensors,* vol. 22, no. 3, 2022.

[121] R. Sharma and R. Arya, "UAV based long range environment monitoring system with Industry 5.0 perspectives for smart city infrastructure," *Computers & Industrial Engineering,* vol. 168, 2022.

[122] G. M. Bolla, M. Casagrande, A. Comazzetto, R. D. Moro, M. Destro, E. Fantin, G. Colombatti, A. Aboudan and E. C. Lorenzini, "ARIA: Air Pollutants Monitoring Using UAVs," in *2018 5th IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, Rome, Italy, 2018.

[123] H. Patrick, C. Balistreri, P. Pontelandolfo and G. Triscone, "Development of an unmanned aerial vehicle UAV for air quality measurement in urban areas," in *32nd AIAA Applied Aerodynamics Conference*, 2014.

[124] P. Arroyo, J. Gómez-Suárez, J. L. Herrero and J. Lozano, "Electrochemical gas sensing module combined with Unmanned Aerial Vehicles for air quality monitoring," *Sensors and Actuators B: Chemical,* vol. 364, 2022.

[125] V. Dyo and C. Mascolo, "Efficient Node Discovery in Mobile Wireless Sensor Networks," in *Distributed Computing in Sensor Systems*, Berlin, Heidelberg, 2008.

[126] J. Joseph, H. O. Sharif, T. Sunil and H. Alamgir, "Application of validation data for assessing spatial interpolation methods for 8-h ozone or other sparsely monitored constituents," *Environmental Pollution,* vol. 178, pp. 411-418, 2013.

[127] Y. Tong, Y. Yu, X. Hu and L. He, "Performance analysis of different kriging interpolation methods based on air quality index in Wuhan," in *2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP)*, 2015.

[128] H. Wahid, Q. Ha, H. Duc and M. Azzi, "Neural network-based meta-modelling approach for estimating spatial distribution of air pollutant levels," *Applied Soft Computing,* vol. 13, no. 10, 2013.

[129] H. Pfeiffer, G. Baumbach, L. Sarachaga-Ruiz, S. Kleanthous, O. Poulida and E. Beyaz, "Neural modelling of the spatial distribution of air pollutants," *Atmospheric Environment,* vol. 43, no. 23, 2009.

[130] L. Nicoletta, S. Daniele, G. Aristide, B. Diego and P. Massimo, "Spatial Interpolation Techniques For innovative Air Quality Monitoring Systems," *Chemical Engineering Transactions,* vol. 86, 2021.

[131] 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*, Novi Sad, Serbia, 2019.

[132] O. Cetinkaya, B. Zaghari, F. M. J. Bulot and W. Damaj, "Distributed Sensing with Low-Cost Mobile Sensor Toward a Sustainable IoT," *IEEE Internet of Things Magazine,* vol. 4, no. 3, pp. 96-102, 2021.

[133] O. Saukh, D. Hasenfratz, C. Walser and L. Thiele, "On rendezvous in mobile sensing networks," *Real-World Wireless Sensor Networks,* 2013.

[134] K. Fu, W. Ren and W. Dong, "Multihop calibration for mobile sensing: K-hop Calibratability and reference sensor deployment," in *Conference: IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017.

[135] G. Y. Keung, Q. Zhang and B. Li, "The delay-constrained information coverage problem in mobile sensor networks: single hop case," *Wireless Networks,* vol. 16, 2010.

[136] O. Saukh, D. Hasenfratz and L. Thiele, "Reducing multi-hop calibration errors in large-scale mobile sensor networks," in *IPSN '15: Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 2015.

[137] J. Xiang, E. Austin, T. Gould, T. Larson, M. Yost, J. Shirai, Y. Liu, S. Yun and E. Seto, "Using Vehicles' Rendezvous for In Situ Calibration of Instruments in Fleet Vehicle-Based Air Pollution Mobile Monitoring," *Environmental Science & Technology,* 2020.

[138] J.-F. Markert, M. Budde, G. Schindler, M. Klug and M. Beigl, "Private Rendezvous-based Calibration of Low-Cost Sensors for Participatory Environmental Sensing," in *Urb-IoT '16: Proceedings of the Second International Conference on IoT in Urban Space*, 2016.

[139] C. Yu, J. Luo, R. Shi, X. Liu, F. Dang and X. Chen, "ST-ICM: Spatial-Temporal Inference Calibration Model for Low Cost Fine-grained Mobile Sensing," in *MobiCom '22: Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, 2022.

[140] B. Maag, Z. Zhou, O. Saukh and L. Thiele, "SCAN: Multi-hop calibration for mobile sensor arrays," in *Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies*, 2017.

[141] B. Maag, Z. Zhou and L. Thiele, "Enhancing Multi-hop Sensor Calibration with Uncertainty Estimates," in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2019.

[142] L. Zhao, Z. Yin, K. Yu, X. Tang, L. Xu, Z. Guo and P. Nehra, "A Fuzzy Logic Based Intelligent Multi-Attribute Routing Scheme for Two-layered SDVNs," *IEEE Transactions on Network and Service Management,* 2022.

[143] X. Fan, T. Cui, C. Cao, Q. Chen and K. S. Kwak, "Minimum-cost offloading for collaborative task execution of MEC-assisted platooning," *Sensors,* vol. 19, no. 4, 2019.

[144] K. Zhang, Y. Mao, S. Leng, Y. He and Y. ZHANG, "Mobile-edge computing for vehicular networks: A promising network paradigm with predictive off-loading," *IEEE Vehicular Technology Magazine,* vol. 12, no. 2, 2017.

[145] Z. Ning, X. Wang, J. J. P. C. Rodrigues and F. Xia, "Joint computation offloading power allocation and channel assignment for 5G-enabled traffic management systems," *IEEE Transactions on Industrial Informatics,* vol. 15, no. 5, 2019.

[146] S. J. Muhammad, S. Zhang and V. Dyo, "Network coding for reliable safety message communication in vehicular Ad-Hoc networks: a review," in *2015 Fourth International Conference on Future Generation Communication Technology (FGCT)*, Luton, UK, 2015.

[147] S. Desale, A. Rasool, S. Andhale and P. Rane, "Heuristic and Meta-Heuristic Algorithms and Their Relevance to the Real World: A Survey," *International Journal Of Computer Engineering In Research Trends,* vol. 2, no. 5, pp. 296-304, 2015.

[148] A. M. Al-Shaery, M. O. Khozium, N. S. Farooqi, S. S. Alshehri and M. A. M. B. Al-Kawa, "Problem Solving in Crowd Management Using Heuristic Approach," *IEEE Access,* vol. 10, pp. 25422-25434, 2022.

[149] H. R. Y. Sun, "Chapter 3 - Building energy flexibility: modeling and optimization," in *Building Energy Flexibility and Demand Management*, Academic Press, 2023, pp. 41-62.

[150] R. Bridgelall, Optimization Problems in Transportation and Logistics: A Practical Guide, Basel, Switzerland: MDPI, 2024.

[151] C. J. Taylor, "On the Optimal Assignment of Conference Papers," University of Pennsylvania Department of Computer and Information Science Technical Report, 2008.

[152] R. E. Kohn, "Application of Linear Programming to a Controversy on Air Pollution Control," *Management Science,* vol. 17, no. 10, pp. B-609, 1971.

[153] M. M. Lamego, P. A. {. Souza}, E. {. P. Ferreira} and V. K. Garg, "Linear programming applied to the identification of air pollution sources: A case study," *IFAC Proceedings Volumes,* vol. 32, no. 2, pp. 5800-5805, 1999.

[154] J. S. Arora, "Chapter 15 - Discrete Variable Optimum Design Concepts and Methods," in *Introduction to Optimum Design*, 4 ed., 2017, pp. 683-706.

[155] P. C. H and S. Kenneth, Combinatorial optimization: algorithms and complexity, Courier Corporation, 1998.

[156] M. Rebai, M. L. berre, H. Snoussi, F. Hnaien and L. Khoukhi, "Sensor deployment optimization methods to achieve both coverage and connectivity in wireless sensor networks," *Computers & Operations Research,* vol. 59, pp. 11-21, 2015.

[157] X. Xu and S. Sahni, "Approximation Algorithms for Sensor Deployment," *IEEE Transactions on Computers,* vol. 56, no. 12, pp. 1681-1695, 2007.

[158] J. Zhang, C. Liu, X. Li, H.-L. Zhen, M. Yuan, Y. Li and J. Yan, "A survey for solving mixed integer programming via machine learning," *Neurocomputing,* vol. 519, pp. 205-217, 2023.

[159] T. Wu, K. Akartunalı, J. Song and L. Shi, "Mixed integer programming in production planning with backlogging and setup carryover: modeling and algorithms," *Discrete Event Dynamic Systems,* vol. 23, p. 211–239, 2013.

[160] A. B. Keha, K. Khowala and J. W. Fowler, "Mixed integer programming formulations for single machine scheduling problems," *Computers & Industrial Engineering,* vol. 56, no. 1, pp. 357-367, 2009.

[161] R. Guo, W. G. weig and W. Zhang, "Route Design Problem of Customized Buses: Mixed Integer Programming Model and Case Study," *Journal of Transportation Engineering, Part A: Systems,* vol. 144, no. 11, 2018.

[162] A. O. Kotb, Y.-C. Shen, X. Zhu and Y. Huang, "iParker—A New Smart Car-Parking System Based on Dynamic Resource Allocation and Pricing," *IEEE Transactions on Intelligent Transportation Systems,* vol. 17, no. 9, pp. 2637-2647, 2016.

[163] J. D. Waen, H. T. Dinh, M. H. C. Torres and T. Holvoet, "Scalable multirotor UAV trajectory planning using mixed integer linear programming," in *2017 European Conference on Mobile Robots (ECMR)*, Paris, France, 2017.

[164] A. Trotta, F. D. Andreagiovanni, M. D. Felice, E. Natalizio and K. R. Chowdhury, "When UAVs Ride A Bus: Towards Energy-efficient City-scale Video Surveillance," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Honolulu, HI, USA, 2018.

[165] A. Richards and J. How, "Aircraft trajectory planning with collision avoidance using mixed integer linear programming," in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, Anchorage, AK, USA, 2002.

[166] Y. Xiang, L. S. Bai, R. Piedrahita, R. P. Dick, Q. Lv, M. Hannigan and L. Shang, "Collaborative calibration and sensor placement for mobile sensor networks," in *IPSN '12: Proceedings of the 11th international conference on Information Processing in Sensor Networks*, 2012.

[167] A. Boubrima, W. Bechkit, H. Rivano and L. Soulhac, "Leveraging the Potential of WSN for an Efficient Correction of Air Pollution Fine-Grained Simulations," in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, Hangzhou, China, 2018.

[168] E. Güney, N. Aras, İ. K. Altınel and C. Ersoy, "Efficient integer programming formulations for optimum sink location and routing in heterogeneous wireless sensor networks," *Computer Networks,* vol. 54, no. 11, pp. 1805-1822, 2010.

[169] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, Nonlinear Programming: Theory and Algorithms, 3 ed., John wiley & sons, 2006.

[170] C. Carnevale, E. Pisoni and M. Volta, "A multi-objective nonlinear optimization approach to designing effective air quality control policies," *Automatica,* vol. 44, no. 6, pp. 1632-1641, 2008.

[171] A. Trujillo-Ventura and H. Ellis, "Nonlinear Optimization Of Air Pollution Monitoring Networks: Algorithmic Considerations And Computational Results," *Engineering Optimization,* vol. 19, no. 4, pp. 287-308, 2007.

[172] A. Shishegaran, M. Saeedi, A. Kumar and H. Ghiasinejad, "Prediction of air quality in Tehran by developing the nonlinear ensemble model," *Journal of Cleaner Production,* vol. 259, 2020.

[173] M. R. Ramzan, N. Nawaz, A. Ahmed, M. Naeem, M. Iqbal and A. Anpalagan, "Multi-objective optimization for spectrum sharing in cognitive radio networks: A review," *Pervasive and Mobile Computing,* vol. 41, pp. 106-131, 2017.

[174] N. Mazyavkina, S. Sviridov, S. Ivanov and E. Burnaev, "Reinforcement Learning for Combinatorial Optimization: A Survey," *Computers & Operations Research,* vol. 134, 2021.

[175] P. A. Samuel, L. Ochei and C. B. Marcus, "Optimal Resource Allocation in Exam Scheduling Using the Knapsack Model," *The International Journal of Engineering and Science (IJES),* vol. 13, no. 5, pp. 302-321, 2024.

[176] E. Malaguti, M. Monaci and P. Toth, "A Metaheuristic Approach for the Vertex Coloring Problem," *INFORMS Journal on Computing,* vol. 20, no. 2, pp. 169-331, 2008.

[177] J. Chen and J. Yang, "Maximizing Coverage Quality with Budget Constrained in Mobile Crowd-Sensing Network for Environmental Monitoring Applications," *Sensors,* vol. 19, no. 10, 2019.

[178] Y. Zhao, R. Schwartz, E. Salomons, A. Ostfeld and H. V. Poor, "New formulation and optimization methods for water sensor placement," *Environmental Modelling & Software,* vol. 76, pp. 128-136, 2016.

[179] J.-Y. Wang, "A Branch-and-Bound Algorithm for Minimizing the Total Tardiness of a Three-Agent Scheduling Problem Considering the Overlap Effect and Environmental Protection," *IEEE Access,* vol. 7, pp. 5106-5123, 2018.

[180] M. Lee, G. Yu and G. Y. Li, "Learning to Branch: Accelerating Resource Allocation in Wireless Networks," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 1, pp. 958-970, 2019.

[181] F. Mansourkiaie and M. H. Ahmed, "Optimal and Near-Optimal Cooperative Routing and Power Allocation for Collision Minimization in Wireless Sensor Networks," *IEEE Sensors Journal,* vol. 16, no. 5, pp. 1398-1411, 2016.

[182] G. S. Kumar, A. Sharma, K. Shukla and A. K. Nema, "Dynamic Programming-Based Decision-Making Model for Selecting Optimal Air Pollution Control Technologies for an Urban Setting," in *Smart Cities–Opportunities and Challenges*, vol. 58, Springer Singapore, 2020, p. 709–729.

[183] T. Ono, K. Ito and S. Tsujio, "A method for locating monitoring posts for air pollution on a highway," *International Journal of Environmental Studies,,* vol. 15, no. 2, pp. 123-132, 1980.

[184] E. Anastasiadis, X. Deng, P. Krysta, M. Li, H. Qiao and J. Zhang, "Network Pollution Games," *Algorithmica,* vol. 81, p. 124–166, 2019.

[185] Ş. Y. Balaman, "Modeling and Optimization Approaches in Design and Management of Biomass-Based Production Chains," in *Decision-Making for Biomass-Based Production Chains*, 2019, pp. 185-236.

[186] S. Zhang and G. Li, "Simplex Optimization Localization Algorithm for Wireless Sensor Networks," in *2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing*, Wuhan, China, 2010.

[187] D. Ma, J. Deng and Z. Zhang, "Comparison and improvements of optimization methods for gas emission source identification," *Atmospheric Environment,* vol. 81, pp. 188-198, 2013.

[188] F. Wang, C. Wang, Z. Wang and X.-y. Zhang, "A Hybrid Algorithm of GA + Simplex Method in the WSN Localization," *International Journal of Distributed Sensor Networks,* vol. 11, no. 7, 2015.

[189] S. H. Kang, "Energy Optimization in Cluster-Based Routing Protocols for Large-Area Wireless Sensor Networks," *Symmetry,* vol. 11, no. 1, 2019.

[190] K. Dhondge, R. Shorey and J. Tew, "HOLA: Heuristic and opportunistic link selection algorithm for energy efficiency in Industrial Internet of Things (IIoT) systems," in *2016 8th International Conference on Communication Systems and Networks (COMSNETS)*, Bangalore, India, 2016.

[191] B. Hussain, Q. U. Hasan, N. Javaid, M. Guizani, A. Almogren and A. Alamri, "An Innovative Heuristic Algorithm for IoT-Enabled Smart Homes for Developing Countries," *IEEE Access,* vol. 6, pp. 15550-15575, 2018.

[192] A. Vince, "A framework for the greedy algorithm," *Discrete Applied Mathematics,* vol. 121, no. 1-3, pp. 247-260, 2002.

[193] S. Pirkwieser, G. R. Raidl and J. Puchinger, "Combining Lagrangian Decomposition with an Evolutionary Algorithm for the Knapsack Constrained Maximum Spanning Tree Problem," in *Evolutionary Computation in Combinatorial Optimization*, Berlin, Heidelberg, 2007.

[194] S. Khuller, B. Raghavachari and N. E. Young, "Greedy Methods," in *Handbook of Approximation Algorithms and Metaheuristics*, 2nd ed., vol. 1, 2018, p. 15.

[195] B. I. Simmons, C. Hoeppke and W. J. Sutherland, "Beware greedy algorithms," *Journal of Animal Ecology,* vol. 88, no. 5, pp. 804-807, 2019.

[196] A. Capone, G. Carello, I. Filippini, S. Gualandi and F. Malucelli, "Routing, scheduling and channel assignment in Wireless Mesh Networks: Optimization models and algorithms," *Ad Hoc Networks,* vol. 8, no. 6, pp. 545-563, 2010.

[197] M. Hojati, "A greedy heuristic for shift minimization personnel task scheduling problem," *Computers & Operations Research,* vol. 100, pp. 66-76, 2018.

[198] D. Sashka and I. Russell, "Models of Greedy Algorithms for Graph Problems," *Algorithmica,* vol. 54, p. 269–317, 2009.

[199] J. Chen and J. Yang, "Maximizing Coverage Quality with Budget Constrained in Mobile Crowd-Sensing Network for Environmental Monitoring Applications," *Sensors,* vol. 19, no. 10, 2019.

[200] N. T. Hanh, H. T. T. Binh, N. V. Son and P. N. Lan, "Minimal Node Placement for Ensuring Target Coverage With Network Connectivity and Fault Tolerance Constraints in Wireless Sensor Networks," in *IEEE Congress on Evolutionary Computation (CEC)*, Wellington, New Zealand, 2019.

[201] R.-H. Li and J. X. Yu, "Triangle minimization in large networks," *Knowledge and Information Systems,* vol. 45, p. 617–643, 2015.

[202] A. R. Jordehi, "Optimisation of demand response in electric power systems, a review," *Renewable and Sustainable Energy Reviews,* vol. 103, pp. 308-319, 2019.

[203] P. Bettinger, K. Boston and J. Sessions, Simulated Annealing, Academic Press, 2008, pp. 3255-3261.

[204] A. Leach, "4.05 - Ligand-Based Approaches: Core Molecular Modeling," *Comprehensive Medicinal Chemistry II,* vol. 4, pp. 87-118, 2007.

[205] M. Kolonko, "Some new results on simulated annealing applied to the job shop scheduling problem," *European Journal of Operational Research,* vol. 113, no. 1, pp. 123-136, 1999.

[206] S.-h. Zhan, J. Lin, Z.-j. Zhang and Y.-w. Zhong, "List-Based Simulated Annealing Algorithm for Traveling Salesman Problem," *Computaional Intelligence and Neuroscience,* 2016.

[207] J. C. J. H. Aerts and G. B. M. Heuvelink, "Using simulated annealing for resource allocation," *International Journal of Geographical Information Science,* vol. 16, no. 6, pp. 571-587, 2010.

[208] A. G. Nikolaev and S. H. Jacobson, "Simulated Annealing," *Handbook of Metaheuristics. International Series in Operations Research & Management Science,* vol. 146, p. 1–39, 2010.

[209] G. Zhang and L. Zhang, "WSN Location Algorithm Based on Efficient Simulated Annealing Weighted DV-Hop," in *2019 4th International Conference on Communication and Information Systems (ICCIS)*, Wuhan, China, 2019.

[210] A. A. Kannan, G. Mao and B. Vucetic, "Simulated Annealing based Wireless Sensor Network Localization," *Journal of Computers,* vol. 1, no. 2, 2006.

[211] Y. Zhang, L. Cao, Y. Yue, Y. Cai and B. Hang, "A Novel Coverage Optimization Strategy Based on Grey Wolf Algorithm Optimized by Simulated Annealing for Wireless Sensor Networks," *Computational Intelligence and Neuroscience,* 2021.

[212] J. Leite and B. Topping, "Parallel simulated annealing for structural optimization," *Computers & Structures,* vol. 73, no. 1-5, pp. 545-564, 1999.

[213] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1995.

[214] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, 1995.

[215] X. Zhu and H. W. , “A new inertia weight control strategy for Particle Swarm Optimization,” in *AIP Conference Proceedings*, 2018.

[216] H. Belyadi and A. Haghighat, "Evolutionary optimization," in *Machine Learning Guide for Oil and Gas Using Python*, Gulf Professional Publishing, 2021, pp. 419-454.

[217] D. Gupta, S. Thakur, A. Gode, K. Jagtap and V. S. Gutte, "Optimizing Sensor Placement for Air Quality Monitoring System Using Particle Swarm Optimization," in *2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech)*, Banur, India, 2023.

[218] S. Chelbi, H. Dhahri and R. Bouaziz, "Node placement optimization using particle swarm optimization and iterated local search algorithm in wireless sensor networks," *International Journal of Communication Systems,* vol. 34, no. 9, 2021.

[219] S. M. M. Islam, M. A. R. Reza and M. A. Kiber, "Wireless Sensor Network using Particle Swarm Optimization," in *Proc. of Int. Conf. on Advances in Control System and Electricals Engineering*, 2013.

[220] J. Wang, Y. Cao, B. Li, H.-j. Kim and S. Lee, "Particle swarm optimization based clustering algorithm with mobile sink for WSNs," *Future Generation Computer Systems,* vol. 76, pp. 452-457, 2017.

[221] M. Li, W. Du and F. Nian, "An Adaptive Particle Swarm Optimization Algorithm Based on Directed Weighted Complex Network," *Mathematical problems in engineering,* no. 1, p. 434972, 2014.

[222] M. S. Saddique, A. R. Bhatti, S. S. Haroon, M. K. Sattar, S. Amin and S. S. u. Haq, "Solution to optimal reactive power dispatch in transmission system using meta-heuristic techniques—Status and technological review," *Electric Power Systems Research,* vol. 178, 2020.

[223] K. Cherifi, "An overview on recent machine learning techniques for Port Hamiltonian systems," *Physica D: Nonlinear Phenomena,* vol. 411, 2020.

[224] R. Espinosa, F. Jiménez and J. Palma, "Multi-objective evolutionary spatio-temporal forecasting of air pollution," *Future Generation Computer Systems,* vol. 136, pp. 15-33, 2022.

[225] T. A. N. Dinh, A. D. Nguyen, T. T. Nguyen, T. H. Nguyen and P. L. Nguyen, "Spatial-temporal Coverage Maximization in Vehicle-based Mobile Crowdsensing for Air Quality Monitoring," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, Austin, TX, USA, 2022.

[226] Z. Michalewicz and M. Schoenauer, "Evolutionary Algorithms for Constrained Parameter Optimization Problems," *Evolutionary Computation,* vol. 4, no. 1, pp. 1-32, 1996.

[227] S. Katoch, S. S. Chauhan and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications,* vol. 80, p. 8091–8126, 2020.

[228] N. Srinidhi, S. D. Kumar and K. Venugopal, "Network optimizations in the Internet of Things: A review," *Engineering Science and Technology, an International Journal,* vol. 22, no. 1, pp. 1-21, 2019.

[229] M. Asghari and H. Nematzadeh, "Predicting air pollution in Tehran: Genetic algorithm and back propagation neural network," *Journal of AI and Data Mining,* vol. 4, no. 1, pp. 49-54, 2016.

[230] Y. Hao and S. Xie, "Optimal redistribution of an urban air quality monitoring network using atmospheric dispersion model and genetic algorithm," *Atmospheric Environment,* vol. 177, pp. 222-233, 2018.

[231] K. Wang, H. Zhao, D. Yanana, L. Taoa, H. Lujianb and S. Fengjuanb, "Optimization of air pollutant monitoring stations with constraints using genetic algorithm," *Journal of High Speed Networks,* vol. 21, no. 2, pp. 141-153, 2015.

[232] Y. Kumar, K. Deep and A. K. Nagar, "Installation of Smog-Free Towers Using Novel Real Coded Genetic Algorithm," *IEEE Access,* vol. 12, pp. 15056-15074, 2024.

[233] L. H. Anh, C. M. Hai, N. P. Truong, T.-H. Nguyen and P. L. Nguyen, "Optimal Deployment of Vehicular Mobile Air Quality Monitoring Systems," in *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, Ho Chi Minh City, Vietnam, 2020.

[234] C.-K. Liang and Y.-H. Lin, "A coverage optimization strategy for mobile wireless sensor networks based on genetic algorithm," in *2018 IEEE International Conference on Applied System Invention (ICASI)*, Chiba, Japan, 2018.

[235] H. Zarrar and V. Dyo, "Bus-Based Sensor Deployment for Intelligent Sensing Coverage and k-Hop Calibration," *IET Smart Cities,* vol. 7, no. 1, p. e70004, 2025.

[236] C. Wang, C. Li, C. Qin, W. Wang and X. Li, "Maximizing spatial–temporal coverage in mobile crowd-sensing based on public transports with predictable trajectory," *International Journal of Distributed Sensor Networks,* 2018.

[237] D. A. Anggor and W. Supriyanti, "Improving Accuracy by applying Z-Score Normalization in Linear Regression and Polynomial Regression Model for Real Estate Data," *International Journal of Emerging Trends in Engineering Research,* vol. 7, no. 11, pp. 549-555, 2019.

[238] S. V. Liu, F.-l. Chen and J. Xue, "A meta-analysis of selected near-road air pollutants based on concentration decay rates," *Heliyon,* vol. 5, no. 8, 2019.

Appendices

Ethics Form

UNIVERSITY OF BEDFORDSHIRE

Research Ethics Scrutiny (Postgraduate Research Students)

When completing this form please ensure that you read and comply with the following.
Pay particular attention to the Data Collection under Covid-19 section on the next page.

Researchers must demonstrate clear understanding of an engagement with the following:

1. *Integrity* - The research has been carried out in a rigorous and professional manner and due credit has been attributed to all parties involved.
2. *Plagiarism* - Proper acknowledgement has been given to the authorship of data and ideas.
3. *Conflicts of Interest* - All financial and professional conflicts of interest have been properly identified and declared.
4. *Data Handling* - The research draws upon effective record keeping, proper storage of date in line with confidentiality, statute and University policy.
5. *Ethical Procedures* - Proper consideration has been given to all ethical issues and appropriate approval sought and received from all relevant stakeholders. In addition the research should conform to professional codes of conduct where appropriate.
6. *Supervision* - Effective management and supervision of staff and student for whom the researcher(s) is/are responsible
7. *Health and Safety*- Proper training on health and safety issues has been received and completed by all involved parties. Health and safety issues have been identified and appropriate assessment and action have been undertaken.

The **Research Institutes** are responsible for ensuring that all researchers abide by the above. It is anticipated that ethical approval will be granted by each Research Institute. Each Research Institute will give guidance and approval on ethical procedures and ensure they conform to the requirements of relevant professional bodies. As such Research Institutes are required to provide the University Research Ethics Committee with details of their procedures for ensuring adherence to relevant ethical requirements. This applies to any research whether it be, or not, likely to raise ethical issues. Research proposals involving vulnerable groups; sensitive topics; groups requiring gatekeeper permission; deception or without full informed consent; use of personal/confidential information; subjects in stress, anxiety, humiliation or intrusive interventions must be referred to the University Research Ethics Committee.

Research projects involving participants in the NHS will be submitted through the NHS National Research Ethics Service (NRES). The University Research Ethics Committee will normally accept the judgement of NRES (it will never approve a proposal that has been rejected by NRES), however NRES approval will need to be verified before research can commence and the nature of the research will need to be verified.

Where work is conducted in collaboration with other institutions ethical approval by the University and the collaborating partner(s) will be required.

The **University Research Ethics Committee** is a sub-committee of the Academic Board and is chaired by a member of the Vice Chancellor's Executive Group, appointed by the Vice-Chancellor and includes members external to the University

Research Misconduct: Allegations of Research Misconduct against staff or post graduate (non-taught) research students should be made to the Director of Research Development.

November 2020

Data collection during Covid-19:

At the start of the lockdown, many students applied for extensions or interruptions due to being unable to collect data. This was seen as the most feasible option under the circumstances. Now that the outlook indicates that restrictions will continue (and probably get more serious), extensions and interruptions are no longer the only feasible solution to data collection issues.

The general principles of data collection during Covid-19 (based on the guidelines issued by UK Research and Innovation and UK Research Integrity Office) are as follows:

1. If the project requires access to human participants or data collection in public places the student and their supervisory team will need to consider alternatives to in-person data collection. Online surveys/questionnaires, online interviews or focus groups, using secondary data (social media, blogs, publications), or existing datasets should be a default position.
2. If the methods of data collection have changed the Methodology chapter/section of the project needs to be updated to explain the changes to the nature of data gathering, sample size, and access to the sample. An updated ethics approval should be sought from the Institute's REC.
3. For **online data collection**, extra attention should be paid to the contents of informed consent letters, including explanations of the platforms used for data collection, how the recordings will be stored, and how the researchers will ensure the participants' privacy is not intruded upon (e.g. opting for voice recording only, where feasible). An end user agreement for the platform to be used should be made available to the participants.
4. It is also advised that the researchers over-recruit by a small number to offset the risk of insufficient numbers of participants turning up.
5. The default platform should be Blackboard Collaborate, where sessions can be audio or video recorded. Since the University has control over this platform, the recordings are secured under Data Protection Law/GDPR. The recordings can then be accessed through a secure link. After the recording has been transcribed it should be deleted from the platform. Secure storage of data is expected as per normal practices: either on the University's OneDrive, an encrypted data stick, or on a password-protected hard drive of the University's laptop. All existing guidelines for secure data storage continue to apply.
6. The **use of labs** is allowed under the conditions outlined by individual specialist areas. These have specific requirements as detailed within their local risk assessment documents, in addition to the general University guidelines on accessing campus facilities.
7. If **in-person data collection** is absolutely necessary the student and supervisors should use the risk assessment procedures set up in their School/Faculty. If data collection is carried out overseas the student must adhere to the relevant travel guidelines and local guidelines on social distancing, etc.

November 2020

UNIVERSITY OF BEDFORDSHIRE

Research Ethics Scrutiny (Annex to RS1 form)

SECTION A To be completed by the candidate

Registration No: **1612118**

Candidate: **Hassan Zarrar**

Degree of: **MPhil/Ph.D**

Research Institute: **Institute for Research in Applicable Computing (IRAC)**

Research Topic: **Optimization of Sensor Nodes Deployment for Intelligent Drive-by Sensing and k-hop Calibration**

External Funding: **Yes**

Data Protection Impact Assessment (DPIA) Number: **Not Applicable**

NOTE: If you are required to receive a DPIA approval please complete the Data Protection Impact Assessment Form (located in the 'Research Data Management' area on BREO – RGS Community).

The candidate is required to summarise in the box below the ethical issues involved in the research proposal and how they will be addressed. In any proposal involving human participants the following should be provided:

- clear explanation of whether data collection will be done in person or online,
- if in-person data collection is the only feasible option, a clear explanation of the reasons for this and risk assessment of the process,
- how informed consent will be obtained,
- how confidentiality and anonymity will be observed,
- how the nature of the research, its purpose and the means of dissemination of the outcomes will be communicated to participants,
- how personal data will be stored and secured (including audio or video files of interviews, focus groups, etc.),
- if participants are being placed under any form of stress (physical or mental) identify what steps are being taken to minimise risk

If protocols are being used that have already received University Research Ethics Committee (UREC) ethical approval then please specify. Roles of any collaborating institutions should be clearly identified. Reference should be made to the appropriate professional body code of practice.

No human participants will be involved in this research.

November 2020

Answer the following questions by deleting as appropriate:

1. Does the study require in-person data collection only?

No

If YES: Has risk assessment been carried out and measures taken to ensure the participants', researcher's and others' safety?

Explain in detail which risk assessment protocols were followed (e.g. UoB School's/Department's, other organisations') and what measures are taken to ensure the participants', researcher's and others' safety.

2. Does the study involve vulnerable participants or those unable to give informed consent (e.g. children, people with learning disabilities, your own students)?

No

If **YES**: Have/will Researchers be DBS checked?

3. Will the study require permission of a gatekeeper for access to participants (e.g. schools, self-help groups, residential homes)?

No

4. Will it be necessary for participants to be involved without consent (e.g. covert observation in non-public places)?

No

5. Will the study involve sensitive topics (e.g. sexual activity, substance abuse)?

No

6. Will blood or tissue samples be taken from participants?

No

If YES: Has risk assessment been carried out and measures taken to ensure the participants', researcher's and others' safety?

Explain in detail which risk assessment protocols were followed (e.g. UoB School's/Department's, other organisations') and what measures are taken to ensure the participants', researcher's and others' safety.

7. Will the research involve intrusive interventions (e.g. drugs, hypnosis, physical exercise)?

No

8. Will financial or other inducements be offered to participants (except reasonable expenses)?

No

9. Will the research investigate any aspect of illegal activity?

No

10. Will participants be stressed beyond what is normal for them?

No

11. Will the study involve participants from the NHS (e.g. patients) or participants who fall under the requirements of the Mental Capacity Act 2005?

No

*Please note if the answer to Question 10 is yes then the proposal should be submitted through **NHS research ethics approval procedures** to the appropriate **NRES**. The UREC should be informed of the outcome.

If you have answered yes to **Question 1** then a clear justification for the need for in-person data collection must be provided.

If you have answered yes to **Question 2** then a clear justification for the research must be provided.

If you have answered yes to any other of the above questions or if you consider that there are other significant ethical issues then details should be included in your summary above.

Checklist of documents which should be included:

Project proposal (with details of methodology) & source of funding	✓
Risk assessment documents from UoB and other organisations as relevant – FOR IN-PERSON DATA COLLECTION or USE OF LABS	
Documentation seeking informed consent (if appropriate)	
Information sheet for participants (if appropriate)	
Questionnaire (if appropriate)	

(Tick as appropriate)

Applicant declaration

- I understand that I cannot collect any data until the application referred to in this form has been approved by all relevant parties.
- I understand that if in-person data collection is the ONLY viable option I must carry out all required risk assessment exercises and adhere to the University's and other applicable guidelines on interaction during Covid-19.
- I agree to carry out the research in the manner specified and comply with the statement of ethical requirements on page 1 of this form. If I make any changes to the approved method, I will seek further ethical approval for any changes.

Signature of Applicant:  Date: **01/010/2024**

Signature of Director of Studies: V. Dyo Date: 2/10/2024

This form together with a copy of the research proposal should be submitted to the Research Institute Director for consideration by the Research Institute Ethics Committee/Panel

Note you cannot commence collection of research data until this form has been approved

SECTION B To be completed by the Research Institute Ethics Committee:

Comments: Looks fine with me

Approved Dr X. Feng

Signature Chair of Research Institute Ethics Committee:

Xiaohua Feng (electronics submission) Date: 2/10/2024

This form should then be filed on the student's record

If in the judgement of the committee there are significant ethical issues for which there is not agreed practice then further ethical consideration is required before approval can be given and the proposal with the committees comments should be forwarded to the secretary of the UREC for consideration.

There are significant ethical issues which require further guidance

Signature Chair of Research Institute Ethics Committee:

Date:

This form together with the recommendation and a copy of the research proposal should then be submitted to the University Research Ethics Committee.

Research Data Management Plan

1. Overview

1.1 Project title
Optimization of Sensor Nodes Deployment for Drive-by Sensing and k-hop Calibration
1.2 Student name and Research Institute
Hassan Zarrar, Institute for Research in Applicable Computing (IRAC)
1.3 Supervisor(s)
Note: The Director of Studies is the Data Steward
*Dr. Vladimir Dyo (1st Supervisor), Dr. Vitaly Schetinin (2nd Supervisor), Dr. Dayou Li **
1.4 Project description: aim, objectives, main research methods used (2-3 sentences will suffice)
The aim of this project is to optimize air quality sensor placements for bus-based mobile sensors and static sensors which are reference grade sensors. The objective is to maximize the sensing coverage and calibrating the mobile sensors within limited budget. We have employed weighted sum method (WSM) for multi-criteria decision-making, z-score standardization to normalize and compare metrics, and optimization technique like greedy approach to improve the algorithms efficiency.

2. Compliance

When you submit your DMP you are confirming that you have read and understood all of the legislative, policy and contractual requirements that apply to your project.

2.1 University policy requirements
All data handling has been adopted following the research policies:
University policy or guidance
University of Bedfordshire GDPR Guidance for Researchers (BREO sign-in required)
University of Bedfordshire Research Ethics Policy
University of Bedfordshire Internet-Mediated Research Guidelines (BREO sign-in required)
University of Bedfordshire Intellectual Property Policy
University of Bedfordshire Compliance with the Concordat to Support Research Integrity

2.2 Legal requirements	
'Not Applicable'	
UK Legislation or framework	
2.3 Contractual requirements	
'Not Applicable'	
Name of funder	**Data policy URL**
N/A	N/A

3. Gathering data – see 'Assessment of Existing Data' and 'Information on New Data' in the RDM Checklist. Cross-reference with your Data Protection Impact Assessment.

3.1 Description of the data

3.1.1 Types of data – existing and/or new

- *Types of secondary data used*
  ***Answer:** General Transit Feed Speciation (GTFS) dataset is used for this research experiment which is open source and is publicly available online listed in my Reference.*
- *Types of data generated in the project (primary data set)*
  ***Answer:** Dataset used for this research experiment is open source and is publicly available online listed in my Reference.*
- *Software code (if relevant)*
  This project involves developing a software for which we have provided pseudo codes.

3.1.2 Format and scale of the data

- *Types of files you will be creating (e.g. Excel, Word, etc.)*
  Formats created were, Excel, CSV and JSON.
- *Types of files for long-term archiving: N/A*

3.2 Data collection methods

- *How will you collect your data?*
  ***Answer**: All data used in my research is public or open source downloaded from website that are listed in my Reference list.*
- *Will you collect data from other sources? If so, include citations of these sources*
  No.

3.3 Development of original software
• *Describe scripts, libraries, plug-ins, software tools, applications, etc. that you will develop: what programming language will be used, how you will address dependencies in your code, etc.* ***Answer:*** *All scripts, libraries, plug-ins, software tools, applications used in this research are open source. We used QGIS which is a geographic information system software that is free and open source to develop the software and analyze the results. QGIS uses Python scripting language.* • *Is the software you're developing for this project only or will other versions be released later?* ***Answer:*** *The software developed is only used for this project, but other versions can be produced.*

4. Working with data – see 'Quality Assurance of Data' and 'Backup and Security of Data' in the RDM Checklist. Cross-reference with your Data Protection Impact Assessment.

4.1 Short- and medium-term data storage arrangements
Note: No personal or sensitive data is used or created in this project. No human subjects were involved in this work.
• *Where will the data be stored during the project (short- and medium-term)?* ***Answer:*** *For easy access the data is stored locally (computer) and external hard drive.* • *How will you create backups?* ***Answer:*** *Automatic backups and manual backups on external hard drive.* • *How will you ensure all files are secure?* ***Answer:*** *The data will have restricted access and will be protected via password.*
4.2 Control of access to data and sharing with collaborators (beyond your supervisory team)
• *Not applicable as the data used is publicly available.*
4.3 File organisation and version control
• *Not applicable as the data used is publicly available and relevant information is available in the data gathering sections of the thesis.*
4.4 Documentation that will accompany the data
• *Not applicable as the data used is publicly available.*

5. Archiving data – see 'Preparation of Data for Sharing and Archiving' in the RDM Checklist

5.1 Selection of data to be retained and deleted at the end of the project
• *Not applicable as data used is available publicly.*
5.2 Data preservation strategy and retention period
• *Data will be stored locally (computer) and on external hard drive.*
5.3 Maintenance of original software
• Are you going to archive a binary executable, installer or package file to accompany your data? No. • Will you submit your code to a language-specific network? No. • Will you share your code through a source code repository? No, will provide source code when asked. • Will you continue to develop your code after the project is over? If so, who will have rights to edit the code? No.

6. Sharing data – see 'Preparation of Data for Sharing and Archiving' in the RDM Checklist

6.1 Justification for any restrictions on data sharing
• *There are no restrictions*
6.2 Arrangements for data sharing
• *No plans to using data archive or repository.*

7. Implementation – see 'Responsibilities' in the RDM Checklist

7.1 Review of the Data Management Plan
• *Not applicable.*
7.2 Special resources required for the project
• *No.*

7.3 Further training needs
• No.