Comparative Assessment on Static O-D Synthesis

by

Tina Fang Yu Lin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Recognizing the benefits of data and the information it provides to travel demand is pertinent to network planning and design. Technological advances have led the ability to produce large quantities and types of data and as a result, many origin-destination (O-D) estimation techniques have been developed to accommodate this data. In contrast to the abundant choices on data types, data quantity and estimation procedures, there lacks a common framework to assess these methods. Without consistency in a baseline foundation, the performances of the methodologies can vary greatly based on each individual assumption.

This research addresses the need for techniques to be tested on a common framework by establishing a baseline condition for static O-D estimation through a synthetic Vissim model of the Sioux Falls network as a case study area. The model is used to generate a master dataset, representing the ground-truth, and a subset of the master dataset, emulating the data collected from real world technologies. The subset of data is used as the input for the O-D estimation techniques where the input is varied to evaluate the effects of different levels of coverage/penetration of each data type on estimation results. A total of five estimation techniques developed by Cascetta and Postorino (2001), Castillo et al. (2008b), Parry and Hazelton (2012), Feng et al. (2015) and X. Yang et al. (2017) are tested with three data types (link counts, partial traces, and full traces) and two traffic assignment conditions (all-or-nothing and user equilibrium).

The result of this research highlights the uniqueness of each network situation and highlights the outcomes of each approach. The wealth of data does not directly equal better information for every methodology. The insights that each data type provides each estimation technique reveals different results. The findings of this research demonstrate and supports that an established testbed framework supports and enhances future O-D estimation scenarios as it pertains to general O-D estimation and extensions of existing techniques.

Acknowledgements

Somewhere in the future, I hope to read this page and reflect on the positive experience in research and the tremendously supportive people who made my work possible. To them I am sincerely grateful.

I would like to start by expressing my deepest gratitude to Professor Chris Bachmann, for your invaluable guidance and profound patience through this process. This thesis would not have been possible without your support. Chris is an example of an attentive mentor with a work ethic that is truly inspiring. I am incredibly grateful for our fascinating conversations and your words of encouragement; it was during these conversations that I was able to pull through with confidence. To Professor Bruce Hellinga and Professor Liping Fu, thank you for reviewing my thesis. Your time and attention are greatly appreciated.

Special thanks to my fellow transportation graduate students, especially to Jess, Devin, and Jacob who have kept me motivated throughout my studies. Your exceptional knowledge, moral support, and advice has gotten me through some tough workdays. I appreciate our shared time during study breaks and shared projects, thank you for your friendship.

To my partner, undergrad friends, and childhood friends, thank you for making my life exciting when school was not. Joshua, thank you for making life more memorable, for supporting me in the moments that are the toughest, and simply for making me happy. Mary, thank you for being by my side and always making everything better. Angela, thank you for your support during undergrad, without you those days would have been tougher than they already were.

Finally, thank you to my family (who will likely never read this) for giving me the love and security I needed to focus on my learning. To mom and dad, thank you for being an example of strong and hardworking individuals, you inspire me everyday. I am truly blessed with great people in my life and each of you have played an integral part of my life and this process.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Well-conceived and well-executed transportation management systems are the basis for livable and sustainable cities. The densification of cities facilitates economic prosperity, cultural diversity, and vibrant activities, but is also the basis of problems such as traffic congestion, crime, and pollution. As urbanization intensifies in many places around the world, moving people and goods in an efficient and rapid manner becomes a challenging fundamental task. Poor planning and infrastructure investments can result in travelers spending significant amounts of time in congested road networks. The resulting delays and poor levels of service due to congestion not only negatively impacts the individual traveler but also damages public health, the environment, general mobility, and the economy. In Canada and the United States, the cost of congestion in the largest metropolitan cities cost each country billions of dollars per year (Pishue, 2021; Transport Canada, 2006). It is critical that the correct tools are available for governing bodies such as Infrastructure Canada to make better informed decisions pertaining to the $28.7 billion invested in public transit infrastructure, and the $10.1 billion invested in trade and transportation infrastructure over the next decade (Infrastructure Canada, 2018).

Understanding transportation systems requires reliable and quality input data for model formulation, calibration, and validation. Despite the useful improvements and advances in transportation modelling since the 1950s from one of the first comprehensive transport study released in Chicago (Weiner, 1997), progress has been limited by the quality of traditional data inputs. Due to the evolution of technology, the ubiquitous boom of computing devices has created a wealth of data of traffic measurements for transportation applications. As Smart Cities initiatives are being implemented worldwide, the promise of emerging data require the right tools and approaches for solving the myriad of challenges in urban systems.

1.1.1 Travel Demand Models

Travel demand models (TDM) are widely used to manage existing transportation systems and plan for future infrastructure investments. The models generally fall into two categories: (1) traditional four-step model which originated in the 1950s (Weiner, 1997), and (2) activity-based model which emerged in the late 1970s/early 1980s (McNally & Rindt, 2007).These models rely on quality detailed data as input and commonly include transportation network information, socioeconomic data, and travel activity details. The travel activity information is typically collected by conducting household travel surveys revealing the trip diaries of individuals within a regional level, for one day of travel.

Household travel surveys have highly detailed data useful for transportation modelling, but unfortunately, practitioners have always been confronted with issues such as expensive administering costs, declining response rates, unreported trips, low sampling rates, and rigorous manual data processing (Stopher & Greaves, 2007) The resulting challenges are reflected in the restriction of survey frequency (typically one or two times per decade), and thus only providing a cross-sectional sampling frame of the region, unable to capture the dynamics of continuous human mobility.

1.1.2 Origin-Destination Matrices

TDM results in origin-destination (O-D) matrices, which provide crucial information for traffic management and transportation planning. These matrices estimate flows from every Origin to every Destination, commonly represented by geographic zones, in a specific time-period for a transportation network. As part of the TDM process, O-D matrices are traditionally determined based on survey data, thus having the same limitations previously described. Unfortunately, this can often lead to O-D matrices with empty cells. Alternatively, O-D estimates can be derived directly from traffic observations (e.g., link counts) rather than through a travel demand model relying on survey data.

These O-D matrices are typically separated into two types: static O-D matrices and dynamic O-D matrices. Static O-D matrices are aggregated travel demands over a time interval whereas dynamic O-D matrices include a temporal variation of travel demands within the analysis period.

The scale and application of O-D matrices have evolved overtime due to emerging technologies and the availability of data. Early O-D estimation research emerged through updating sampled matrices from household travel surveys with a set of recent traffic counts. Typically, static methods are considered for longer-term transportation planning and purpose design whereas dynamic methods are considered for short-term strategies such as traffic control.

1.1.3 Emerging Data Sources

Opportunely, emerging and modern data sources have been developed with the establishment of intelligent transportation systems (ITS) for monitoring, understanding, and improving transportation infrastructure. The pervasive computing equipment has advanced electronic sensor technologies, data transmission technologies, and intelligent control technologies into transportation systems (Tsai et al., 2015). Consequently, larger volumes of data can be obtained from diverse sources such as GPS devices, mobile phones, video cameras, active sensors (e.g. license plate recognition), road side sensors (e.g. Bluetooth identification), and smart cards, which can be used to observe point measurements (e.g. link flows, vehicle speed, and occupancy), and point-to-point measurements (e.g. travel time, and trajectories).

The amount of data generated in ITS can reach the Petabyte level (Stopher & Greaves, 2007) and thus provides researchers an opportunity to develop methodologies to further the understanding of travel demand estimation and resolve existing challenges.

1.2 Problem Statement

The O-D estimation process is generally conducted from observed traffic data since the direct observation of an O-D matrix is rare. While demand models can also provide estimates of O-D matrices, the details of daily fluctuations present are not captured in these results. The process of O-D estimation consists of data inputs (e.g., link flows, traces, prior O-D matrix, etc.), an estimation model, and outputs (e.g., O-D matrix estimate, route assignment, etc.). Some of the main challenges of the estimation models consist of the following:

- an underdetermined problem where the number of O-D pairs is generally much larger than the variables that can be directly observed with traffic data, thus leading to non-unique solutions, and

- the reliance on prior data with inconsistent quality.

Many approaches have been proposed in literature with emerging data sources to address the main challenges. Before recognizing the benefits of the data sources, proper methodology and evaluated limitations must be supported with scientific rigor. Practical applications of recent datasets have been widely explored throughout literature, resulting in sophisticated developments in the domain of transportation modelling. Unfortunately, in contrast to the depth of O-D estimation methods, few of these methods are assessed with the criterion of both data variety and amount in application. Therefore, there are substantial opportunities in exploring the data collection process to improve the accuracy of the estimation of O-D matrices.

Selecting relevant and comprehensive indicators for transportation systems is not a trivial task. While the existing body of literature may propose great solutions when given opportunistic datasets, an alternative targeted question should be posed: if the means to collect vast amounts of data exists, what is the best practice to collect said data for each application? The trivial answer to that question would be to replace the data collected through traditional practices with new data collected through ITS, but this does not leverage new technologies nor solve existing shortcomings. It would be particularly helpful to understand the value of different varieties and amounts of data as they emerge, thus giving the proper tools to select correct data collection technologies, and methodologies for model application.

The data fusion technique is often posed as a solution to decrease the uncertainty related to a single data source. Although multi-sourced data is complementary in nature, development in statistical methods to estimate the trade-offs for having less data or improvements for the addition of data is required for specific applications.

With the emerging areas of technology, the focus should be directed towards the procedures on how we obtain data to build accurate models rather than accepting whatever data exists or blindly collecting data. By investigating this area, there are opportunities to answer the following questions:

1. How much data should be collected for a given level of accuracy in O-D estimation? (dataset size)
2. What kind of data best suits this experiment/model? (data source)
3. Should multiple types of data be used? (data fusion)

In addition to the data collection problem itself, there exists a challenge when comparing the vast amounts of algorithms and formulations. Many of these applications conclude to good results with their own data, some of which may have baseline comparison, but overall, there is no consistency between the experimental design from one methodology to the next. Therefore, the best practice to model the different sources of data does not exist. While many approaches use accuracy as the performance evaluation indicator, the baseline reference point is rarely consistent. Without consistent baseline comparisons, the concluding performance can range due to methodological assumptions, data input, network topology, and traffic conditions.

1.3 Research Objective

With the previously identified problems in the research area, the objectives of this work are separated into primary and secondary objectives. This research focuses on static O-D estimation within the context of a medium-sized network with various data collection sources.

The primary objective of this research is to addresses the lack of consistency when comparing static O-D estimation algorithms. To accomplish this objective, an evidenced-based benchmarking framework is developed for evaluating static O-D estimation techniques of various data inputs. This strategy provides a common testbed network where a ground-truth scenario can be established with simulated data. The results provided by this framework gives a common denominator when comparing the results of different algorithms, therefore giving an impartial evaluation.

To develop a framework to test static O-D estimation techniques, a review is conducted to cover the most recent methodologies and understand the general approaches with their respective limitations. Furthermore, the review considers the data collection technologies used to produce the input dataset along with their representation of information and technical capabilities.

The secondary objective of this research is to quantify the value of information with the experimentation of different scenarios and evaluating the impact on performance indicators. To accomplish this objective, sensitivity scenarios are conducted to evaluate the impacts of different types (single source or multi-source) of datasets and database size to the measure of performance.

With the objectives in mind, this research provides a guidance tool for choice and performance of static O-D estimation techniques.

1.4 Research Scope

The application of this research focuses on aggregate O-D demand estimation where each trip between O-D pairs is viewed as the unit to be estimated through observed data in the road network. This thesis concerns the estimation of O-D estimates for operational type (e.g., microsimulation) models, although the estimation techniques tested in this research are flexible with respect to network size and temporal resolution.

This research provides a common framework where static O-D estimation techniques can be evaluated with an established baseline condition. This is accomplished through the implementation of a synthetic model testbed tasked to compare various algorithms and evaluate performance metrics to assess of the value of the input data. It should be noted that this research focuses on O-D estimation techniques which differs from O-D prediction techniques. O-D prediction typically requires the support of real-time control and utilizes data that reflect temporal variations within a network.

The testbed uses the Sioux Falls network as the case study area. However, the aim is not to represent to the actual City of Sioux Falls, South Dakota and remains a fictitious case. The scope of this research is limited to the estimation of unimodal demand of motorized traffic. Methodological developments of newer techniques of static O-D estimation techniques are not within the scope of this thesis.

1.5 Structure of Thesis

The structure of the thesis is organized in chapters and is structured in the following order: This chapter (Chapter 1) introduces the research area and sets the context of the research problem.

Chapter 2 provides an extensive coverage of the existing literature related to this research area. It reviews recent work that investigates methodological approaches to O-D synthesis and defines a list of data sources as input for modelling. First, a brief overview of O-D estimation is provided with a sequential progression from traditional to modern estimation techniques of O-D flows. Subsequently, a comprehensive review of the adaptation of methods from emerging data sources and passively collected datasets is provided. The literature also covers O-D estimation techniques as part of an overall travel demand estimation model. The chapter concludes with the current research gaps and foundation for the remainder of this thesis.

Chapter 3 presents the proposed framework of this thesis. First, the procedure for algorithm selection is covered. A generic formulation of the static O-D estimation problem is then presented as a basis of all proposed methods. The design of the framework allows a common test-ground of all scenarios and approaches. The following sections of this chapter covers the model building and data extraction, O-D estimation approaches, and generation of scenarios.

The last two chapters cover the outputs of this research. Chapter 4 summarizes the results of the chosen algorithms and testing scenarios. Finally, Chapter 5 concludes the thesis with a summary of key findings of this work and provides recommendation for future research.

Chapter 2

Literature Review

The topics reviewed in this literature review focus on the techniques of O-D estimation. As a foundation for TDM, some methodologies are presented as stand-alone procedures while others are presented as part of TDM and specialty cases. Some other foundational considerations differentiate O-D estimation to O-D prediction. O-D estimation often is referred to as the estimation of the demand matrix for a certain network for an existing time period whereas O-D predictions are demand matrices predicted for future time periods. While O-D estimation can occur both in an online and offline application, O-D predictions are only typically conducted in online applications. This literature review focuses on estimation techniques. The definitions and differentiation of O-D estimation and O-D prediction are illustrated in Figure 2.1.

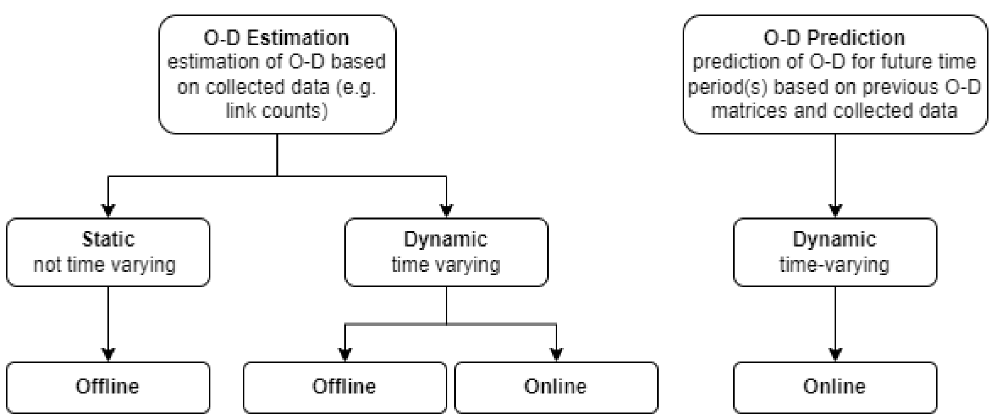

Figure 2.1: O-D Estimation vs O-D Prediction

The focus of this chapter is directed towards the stand-alone procedures, but some special recognition is provided for other cases. To finish, a summary of key findings is presented at the end of this chapter, focusing on the research gap that this work aims to address.

2.1 Origin-Destination Flow Estimation

The O-D flow estimation problem is a long established and well-known area of scientific investigation with a vast body of research covering half-a-century. The classical and widely adopted procedure of O-D estimation relies on updating prior O-D matrices by using traffic counts and survey data (Cascetta et al., 2013; Simonelli et al., 2012). The foundation of these estimation techniques relies on computing link and path travel times from the network model while link flows are not available. Thus, the underdetermined problem results in estimates that are affected by notable errors which requires mitigation by using prior O-D flow estimates created from traffic counts. This classical problem of estimating and updating O-D flows has been studied extensively for static (steady-state) networks through four main solution methodologies: entropy maximization (Van Zuylen & Willumsen, 1980), maximum likelihood (Bell, 1983; Cascetta & Nguyen, 1988; Spiess, 1987), generalized least-squares (GLS) (Cascetta, 1984), and Bayesian inference (Maher, 1983).

Further development led to the extension of O-D estimation for dynamic (time dependent) models, and thus new estimation algorithms have been proposed for within-day and day-to-day dynamics. This problem is commonly represented as a state-space model (Nihan & Davis, 1987), which originated with the contributions by Okutani and Stephanedes (1984) and Cremer and Keller (1987). Subsequently, the dynamic estimation problem is then extended to the within-day framework (Cascetta et al., 1993) and and day-to-day dynamic framework (Zhou & Mahmassani, 2007; Zhou et al., 2003).

The rest of this chapter reviews the development O-D estimation techniques classified by methodological procedure. The literature is then reviewed in a loosely chronological manner as new data collection technologies are introduced with advancements in ITS.

2.1.1 Entropy Maximization

Entropy maximization (EM) or information minimization (IM) was a foundational method developed by Van Zuylen and Willumsen (1980) for O-D estimation. Their paper offers two models based on two theories. The first model starts with calculating the information contained in a trip matrix T_{ij} and aims to minimize information added onto the observed flows. This model is formulated in the following equation:

$$\min_{T_{ij}} \sum_a \sum_{ij} \widehat{T}_{ij} p_{ij}^a \ln \frac{\widehat{T}_{ij} \sum_{ij} \widehat{T}_{ij} p_{ij}^a}{V_a T_{ij}} \tag{2.1}$$

where:

$\widehat{T}_{ij}$ = estimated trip with origin i and destination j
T_{ij} = prior trip with origin i and destination j
V_a = set of traffic counts on link a
p_{ij}^a = fraction of trips travelling on link a with origin i and destination j

The second model tackles the same problem following an entropy maximizing approach. The basis of these models assume routes are known while estimating an O-D matrix best consistent with observed data and prior matrix information. This model is formulated with the following equation:

$$\max_{T_{ij}} \ln \sum_{ij} (T_{ij})! - \sum_{ij} (T_{ij} \ln T_{ij} - T_{ij}) \tag{2.2}$$

The EM principle has been employed on the multiclass O-D estimation problem by (Wong et al., 2005) which can addresses vehicular traffic flow amongst different vehicle types. The model is based on multiclass traffic counts with the estimated matrix output being highly dependent on the reliability of the counts. Furthermore, the EM methodologies have been extended to subnetworks analysis, which addresses the variable nature of travel demand by using subnetwork link flows as constraints (Xie et al., 2011). The approach uses elastic travel demand to consider traffic diversion effects under network changes (e.g. traffic incidents, traffic congestion) and O-D flows are distributed by maximum entropy in the subnetwork while demand function parameters, for the given elastic demand function of each O-D pair, are estimated using least-squares estimation error.

Efforts of the EM approach have also contributed to the estimation of dynamic O-D matrices. Xie and Duthie (2015) have formulated the problem as an excess demand dynamic traffic assignment (DTA) which can be solved using DTA methods and software tools. The basis of this problem employs the synthetic result of individual route cost minimization, traffic count matching error minimization, and O-D demand entropy maximization. It should be noted that the authors suggest a series of experiments to implement and compare solution algorithms for traffic count matching error minimization and O-D demand entropy maximization for a best-fit procedure. Bauer et al. (2018) utilized the principle of EM for quasi-dynamic O-D estimation which is where O-D flows are assumed to be constant across a period and total flows at each origin are assumed to be varying for each sub-period within the reference period. The presented technique is based on continuous count data from detectors and floating car data (FCD) through taxi trajectories without the need of prior O-D matrices. It addresses the non-observability issue with EM and combines a least-squares principle to reproduce link flows. Thus, their research gives the flexibility to produce matrices with or without historical knowledge.

With the increasing ability to collect rich datasets, Ge and Fukuda (2016) leverage passively collected mobile phone traces based on the EM principle to estimate trip flows. The findings of this research are consistent to travel demands in the study area but is constrained to only predicting work-related trips with the dataset. Additionally, Vogt et al. (2019) also leverages passively collected data from FCD, comprised of mobile phone and navigation device traces, in addition to detector counts for O-D matrix estimation through a IM based data fusion technique. This model was evaluated on a simulated testbed network with an established ground truth scenario and shows high goodness of fit (Pearson correlation coefficients of 0.9 or better, root mean square errors of 14.55 or lower, and GEH statistic 1.41 or lower) between the estimated and ground truth matrices with low FCD penetration rates (less than 7%).

EM methodologies are of interest due to its ability to integrate equilibrium-based assignment into O-D estimation. Despite the abilities of this methodology, it has unfortunately shown to be incapable when handling uncertainties in the input data (e.g. traffic counts, prior matrices), thus challenging the confidence of outputs.

2.1.2 Maximum Likelihood

The premise of maximum likelihood estimation (MLE) works by maximizing the probability of model parameters given the observation of the target O-D matrix from survey data and observed link counts. Following the primary works of Bell (1983), Spiess (1987), and Cascetta and Nguyen (1988), the total (joint) likelihood of observing two sets of statistically independent data is given by:

$$L(\widehat{T}, \widehat{V}|T) = L(\widehat{T}|T)L(\widehat{V}|T) \tag{2.3}$$

where:

$\widehat{V}$ = set of traffic count vectors
$\widehat{T}$ = prior O-D matrix
T = estimated O-D matrix

Typically, the traffic counts are hypothesized to follow Poisson distribution or multivariate normal (MVN) distribution. The sample O-D might also be assumed to follow the same distribution (Bera & Rao, 2011) but the statistical distribution would ultimately depend on the sampling strategy (Toledo et al., 2014).

The stochastic nature of flows was considered by Lo et al. (1996) by the introduction of population parameters for the O-D matrix and link choice proportions. Previously, the link choice proportions were assumed to be constant and estimated using traffic assignment techniques. The result of poorly estimated link choice proportions was shown to produce significant errors in O-D estimation. Therefore, this method proposes survey data to be treated as random variables while using the MLE and Bayesian (discussed further in 2.1.4) approach for the O-D matrix estimation of a small network. Hazelton (2000) furthers the comprehension of demand stochasticity by undertaking the parameter identifiability issue and derives the full likelihood function. While the model can consistently estimate O-D matrices based on link data alone, the full likelihood poses an issue of computational burden, thus being too complicated for practical use. Instead of exact likelihood, inference is possible with the MVN approximation or generalized least-squares technique (discussed further in 2.1.3). Taken to a further step, Lo and Chan (2003) incorporated a procedure with a stochastic user equilibrium (UE) principle, to simultaneously estimate the O-D matrix and route choice dispersion parameter for congested networks. The statistical model is separated into two connected steps comprised of the likelihood maximization and the stochastic UE assignment until convergence is obtained for best estimates. With further consideration of stochastic travel demand in O-D estimation, Parry and Hazelton (2012) combines vehicle tracking data with link counts to approximate path choice proportion through likelihood-based techniques. The combination of data results in a statistical model for the portion of vehicles in the network with routing information.

More recent studies explore the advantages of emerging technologies to estimate O-D matrices that are more resilient to data input errors (H. Kim et al., 2018). While the quality of a single data source input can severely affect model accuracy, the combination multiple fields of data result in a more robust model. H. Kim et al. (2018) combined GPS systems, toll collection, and traffic counts for O-D estimation with three objective functions minimizing the gap for the three data types. The combined effort consists of an MLE approach for the navigation system, and two generalized least-squares approaches (discussed further in 2.1.3) for the toll collection and traffic counts.

The spatiotemporal correlations of certain O-D pairs can improve prediction accuracy due to enhanced traffic network knowledge (Xian et al., 2021). A multivariate Poisson log-normal model was formulated by Xian et al. (2021) for the real-time prediction of O-D demand. The problem is formulated with a parameterized covariance matrix that can automatically account for traffic domain knowledge. The parameter estimation is conducted through an expectation maximization algorithm to iteratively obtain the MLE.

2.1.3 Least-Squares

The least-squares approach can produce an estimate of the O-D matrix by solving a system of linear stochastic equations which represent demand modelling errors, flow measurement errors, and traffic assignment errors that are reformulated as an optimization problem. This methodology has the advantage of not requiring a distributional assumption on the dataset, thus producing an unbiased estimation.

The generalized least square (GLS) approach is widely adopted for O-D estimation due to its effectiveness with empirically established variance matrices. This approach first appeared in Cascetta (1984) with the following formulation:

$$\widehat{T} = T + \eta \tag{2.4}$$

$$\widehat{V} = V(T) + \varepsilon \tag{2.5}$$

where:

$\widehat{V}$ = set of traffic count vectors
$\widehat{T}$ = prior O-D matrix
$V(T)$ = flow on a link from the assignment of T
T = estimated O-D matrix
η = sampling error of $\widehat{T}$
ε = sampling error of $\widehat{V}$

Then the GLS estimator based on the formulation above is shown as:

$$\min(\widehat{T} - T)^T Z^{-1}(\widehat{T} - T) + (\widehat{V} - V(T))^T W^{-1}(\widehat{V} - V(T)) \tag{2.6}$$

where:

Z = the variance-covariance matrix of η
W = the variance-covariance matrix of ε

Cascetta et al. (1993) then extended the static problem to offline dynamic O-D estimation. The authors combined traffic count data with other available information (e.g., outdated matrices and surveys). Two estimators are proposed through a GLS framework: a simultaneous estimator to produce joint estimates for a whole set of O-D matrices (one per time slice), and a sequential estimator to produce a sequence of estimates for successive time slices. While the simultaneous estimator is more efficient, the sequential estimator can be adapted for online dynamic estimation. This work was developed further by Ashok and Ben-Akiva (2002) with a state-space model, which is discussed further in 2.1.5. Sherali and Park (2001) proposed a parametric optimization approach using link volumes. The model is formulated as a least-squares problem that seek path flows that compromise between the least path cost and link flow matching. A column generation approach is used to solve a sequence of dynamic shortest path subproblems to provide the O-D estimate. Representative assignment matrices are critical to the O-D estimation process, which seem to cause computational limitations, thus having a model difficult to solve. Zhou and Mahmassani (2006) established an offline dynamic O-D estimation framework from automatic vehicle identification (AVI) data, link counts, and historical demand information using a nonlinear ordinary least-squares (OLS) approach. Point-to-point split-fraction information is obtained from the AVI counts without estimating penetration rates and identification rates. The research is conducted through experimental control using synthetic data. Extending to the real-time O-D estimation, Bierlaire and Crittin (2004) proposed a least-square formulation for large-scale networks and solves the formulation as an iterative algorithm showing to be more efficient than the commonly used Kalman filter (discussed further in 2.1.5).

Path flow estimation (PFE) based techniques estimate O-D flows using path flows as the primary variable. This approach reflects the interdependency between route choice and a final O-D table based on a logit-based path choice model. Sherali et al. (1994). first introduced a UE based linear programming model to reproduce observed link flows using a column generation technique which uses a shortest path sub-problem to generate the paths needed for an optimal solution. Successive to the previous work (Sherali et al., 1994), Sherali et al. (2003) adapts to a sequential linear programming model where complete link volume observation is no longer required and O-D flows can by synthesized based on partial link volume observation. Further adaptation led to Cascetta and Postorino (2001) to assume a stochastic UE based fixed point problem. The problem is then solved using an iterative structure based on the method of successive averages which a GLS based estimator is used to update each iteration. H. Yang et al. (2001) propose an O-D estimation solution for congested networks by presenting a non-linear optimization model to simultaneously estimate the O-D matrix and travel cost coefficient from a logit-based stochastic UE model.

To solve the simultaneous estimation model, a successive quadratic programming method was employed with GLS (or optionally EM) principles to reach a solution. Another attempt was made by A. Chen et al. (2005) to address the O-D estimation problem assuming stochastic UE with the PFE technique. In contrast to previous techniques, this process enforces constraints to link flows and the O-D matrix to exist within a certain range.

Nie and Zhang (2008) proposed a relaxation strategy by formulating dynamic O-D estimation as a single-level problem using a weighted sum of the upper-level (least-square estimator) and lower level (UE) objectives from a bilevel formulation as the new problem objective. A column generation algorithm is used to solve the proposed variational inequality formulation which is used to determine the DTA endogenously. The framework avoids the bilevel structure while considering congestion effects, but the results are highly dependent on initial traffic data and inconsistent with equilibrium conditions. Alternatively, Lundgren and Peterson (2008) extends the bilevel programming approach to larger scale networks by solving the formulation with a descent gradient method. The proposed approach is a heuristic approach; thus, convergence is highly dependent and not guaranteed. Abdelghany et al. (2015) furthered the research in congested networks by proposing a comparison of two estimation methods with time-dependent link flow observations and time-dependent link density observations. The method adopts a bilevel framework where the upper-level forms a least-squares error minimization between estimated and observed values while the lower-level solves the time-dependent link proportions using a simulation based DTA. The results from this study shows that the density-based estimation method is better suited for congested networks since it captures associated flow breakdown and spillback phenomena. Furthermore, Fujita et al. (2017) have also adopted the bi-level framework to obtain time coefficients for hourly O-D estimation with a goal to reduce the number of parameter estimates and computational costs. While the upper level objective stays the same, the lower level adopts a semi-DTA from static UE with elastic demand, thus accounting for residual demand (i.e. adding residual demand onto the next time period). X. Yang et al. (2017) has proposed an O-D estimation model based on GPS probe vehicles and link flows. From the probe vehicle ratios, a scaled O-D matrix is built as a priori which is then corrected by a GLS framework using link counts. The results from this is then extended with a probe-ratio assignment model to relate links and O-D pairs. The outcome of this model shows a reduction of bias towards observed scenarios in direct scale models and ultimately fill the research gap concluding that permanent heterogeneous data sources such as probe vehicles can be practically used in O-D estimation.

Menon et al. (2015) introduces a methodology for sparse O-D estimation using a least-squares methodology while addressing an automated zoning problem. The traffic zoning problem is based on O-D pairs that best explain observed link flows under a UE assignment, and the O-D estimation problem is then encouraged using a ℓ_1 regularization. Hu et al. (2017) formulates a two-stage sequential model to calibrate dynamic O-D matrices under congested conditions. The commonly seen least-squares type two-norm formulation is more computationally intensive than the proposed one-norm formulation approach due to the linear structure. The first stage of this model estimates an O-D matrix by minimizing the link demand deviation with one-norm formulation and the second stage formulates a time dependent UE traffic assignment model to iteratively adjust the time profile.

Caggiani et al. (2013) formulated the congested O-D estimation problem through a fuzzy-GLS method with fuzzy constraints using a triangular membership function. It is formulated as a fixed-point problem and improves the classic GLS solution by including uncertain available data. Ultimately, by including additional sources of available data, an optimal O-D estimate can be produced in a more cost-effective manner. Talebian and Shafahi (2015) then extends the fuzzy-GLS method to the dynamic O-D estimation problem while considering the uncertainty issue in the input information and dynamic network. Uncertainty is addressed by using fuzzy rules and fuzzy C-Mean clustering within the model and results show robust performance even with low quality historical matrices.

More recently, W. Ma and Qian (2018c) have proposed a dynamic O-D estimation framework for a sequence of multiple consecutive days over several years based on a statistical equilibrium assumption. The problem is solved using a non-negative least-squares model with a clustering technique for dimension reduction based on high-granular multi-sourced data consisting of historical traffic counts and speeds. Furthermore, W. Ma and Qian (2018b) also developed a theoretical framework to estimate the mean and variance/covariance matrix of O-D demands to reduce errors in estimation results. Historical traffic data is treated as one data point, and the data points collected over time are used to estimate the probability distribution of O-D demand. An iterative GLS approach is used to estimate the O-D mean vector and variance/covariance matrix with multi-sourced data and without aggregation. Additionally, Mo et al. (2020) developed a hybrid framework leveraging the use of license plate recognition (LPR) data for dynamic O-D estimation. Using a Bayesian path reconstruction model, the framework does not require other information sources to produce an estimate, therefore eliminating artificial assumptions about traveller preferences. The outputs of the path reconstruction model are incorporated into an OLS technique to estimate the final O-D matrix. The methodology is shown to be most successful when the LPR coverage rate is above 50%. Additionally, when compared to a bilevel framework, this model shows practicality for cities where road networks are particularly difficult to calibrate (e.g., link-impedance functions) since calibration of road networks are not needed for this model.

From the wide array of research into the least-squares approach for O-D estimation, it is shown to be a preferable approach due to absence of needing to derive the likelihood function as well as the its ability to reflect route choice behaviours.

2.1.4 Bayesian Inference

Bayesian inference combines traffic counts with prior subjective information from O-D demand through maximization of logged posterior probability. The posterior probability can be obtained by:

$$p(T|\widehat{V},\widehat{T}) = \frac{p(\widehat{V}|T,\widehat{T})p(T|\widehat{T})}{p(\widehat{V}|\widehat{T})} \propto p(\widehat{V}|T,\widehat{T})p(T|\widehat{T}) \tag{2.7}$$

where:

$\widehat{V}$ = set of traffic count vectors
$\widehat{T}$ = prior O-D matrix
T = estimated O-D matrix

Bayesian inference for O-D estimation was introduced by Maher (1983) using traffic counts and flexibility on the degree of beliefs placed on prior estimates. The prior estimates are modified by more recent observations to produce posterior beliefs where relative weights are applied to both the priori and observation to estimate a final O-D. An MVN distribution is assumed with known covariance for the priori matrix and traffic counts. Hazelton (2001) then defines another O-D estimation procedure through Bayesian inference as three subproblems: a reconstruction problem which estimates the O-D matrix in the observation period, a estimation problem which estimates an expected number of O-D trips, and a prediction problem which estimates future O-D trips. Perrakis et al. (2012) utilized information derived from census studies to predict O-D matrices by applying Bayesian statistics. The framework assumes both a Poisson and negative binomial distribution but found the negative binomial distribution to be more suitable than Poisson due to the overdispersion issue.

Feng et al. (2015) proposes a vehicle trajectory reconstruction for large-scale networks using AVI and loop detector data with a particle theory framework. The framework accounts for five trajectory correction factors: path consistency, travel time consistency, measurability criterion, gravity flow model, and path-link flow matching model. A Bayesian methodology is adopted to approximate the ground truth trajectory, updated by the probabilities of most likely trajectories. This framework is tested under the condition where all links are equipped with loop detectors and does not consider inconsistencies from AVI data, thus limiting the practical application of using AVI data for path reconstruction. Additionally, many artificial behavioral assumptions (e.g., people will prefer the shortest path) are made which may make practical sense but should not be applied universally.

Castillo et al. (2008b) formulated a problem covering both optimal counting locations and O-D matrix updating as a Bayesian model. The O-D estimation problem uses a Wardrop-minimum variance model to estimate O-D pairs and unobserved link flows based on observations. Further discussion of the optimal counting locations and network sensor location problem is summarized in 2.3.1. Furthermore, Castillo et al. (2013) proposed a special conjugate Bayesian method for reconstructing traffic flows from plate scanning based on a shifted-Gamma distribution. Sample flows are reconstructed with a GLS technique with flow definition laws and a Bayesian approach is used to estimate different route O-D pairs, scanned links, or counted link flows.

Lately, this research area has explored the use of mobile phone data, specifically in the format of call detail record (CDR), for O-D estimation purposes. Bachir et al. (2019) proposed an end-to-end framework for the estimation of dynamic O-D matrices by transport mode. In combination with CDRs, other sources of data such as geospatial information, travel surveys, census surveys, and travel card information are used to construct O-D matrices. Transport mode from CDRs are obtained by a two-step semi-supervised learning algorithm involving clustering of areas and Bayesian inference to generate trajectory probabilities. Extending to the day-to-day dynamic problem, Pitombeira-Neto et al. (2020) presents a Bayesian model to estimate day-to-day O-D flows for congested and uncongested networks with traffic link counts. The hierarchical framework is specified into three levels: the first level specifies a dynamic Gaussian model describing the evolution of O-D flows, the second level models the assignment of route flows using route choice probabilities that depend on past user-experience cost and a learning parameter, and the third level models the observed link volumes given route flows.

2.1.5 State-Space Models

The dynamic O-D estimation problem has received much attention since the development of ITS due to the consideration of time-variant demand between each O-D pair which also poses as an essential input for DTA models. Dynamic O-D estimation can also be characterized as non-DTA based models where statistical approaches are used to update the O-D matrix over time. A non-DTA based dynamic model was pioneered by Cremer and Keller (1987) where traffic flows within a network were considered to be a dynamic process.

The stochastic nature of travel demand is considered, and four estimators were proposed including an OLS method involving cross-correlation matrices, constrained optimization, simple recursive formula, and Kalman filtering. Furthermore, both Cremer and Keller (1987) and Nihan and Davis (1987) formulated a recursive algorithm and determined split parameters for input-output relationship of traffic flow.

These previous studies did not consider an assignment model for the DTA framework; thus, it only employ techniques to update O-D matrices overtime without account for the dynamic estimation of link usage. The research area is then furthered by Ashok and Ben-Akiva (2000) where the state variables are redefined as deviations of departure rates from O-D shares instead of destination flows. Ashok and Ben-Akiva (2002) then introduces randomness into the dynamic O-D estimation framework by explicitly estimating a stochastic assignment matrix within the state-space formulation based on travel times and route choice in the road network.

Leveraging the ubiquitous usage of mobile phones, Sohn and Kim (2008) used mobile phone data as traffic probes for dynamic O-D estimation. A GLS and Kalman filter (KF) approach were used for estimation while a simulated environment was used for validation. The effects of market penetration and cell coverage presented as contributing factors in accurate O-D predictions. Previous algorithms were typically tested on a freeway network where alternative paths between O-D pairs are not abundant. Advancing the research area, Barceló et al. (2013) extended the KF approach using Bluetooth data for a small urban network where multiple paths between O-D pairs exist. The formulation accounts for multiple paths according to a dynamic UE model.

The quasi-dynamic framework makes inferences about the evolution of demand between time periods in order to reduce the number of unknowns in the network. Cascetta et al. (2013) proposed a O-D estimation technique for the quasi-dynamic framework based on traffic counts with the assumption that O-D shares remain constant within a time period while demand varies within the same period. In comparison to simultaneous estimates, historical O-D flows show improvement in the performance of KF. Z. Lu et al. (2015) proposes a KF approach using multi-sensor data to better capture the complex characteristics of a dynamic network. The estimation is represented as a dynamic relationship between the O-D structural deviation, random error and historical traffic data. This framework offers an effective way to utilize multi-sensor data with better performance than traditional link-volume based or turning-movement based methods. Furthermore, Carrese et al. (2017) adopts FCD for online and offline dynamic demand estimation within an urban network. The offline problem is formulated as a bilevel optimization problem while the upper level adopts a non-linear state space model that experiments with an extension of the KF approach called Local Ensemble Transformed Kalman Filter (LETKF), eliminating the need for an explicit assignment matrix. The results of the LETKF outperformed common non-linear KF implementations. More recently, J. Liu et al. (2020) leveraged LPR data to estimate dynamic O-D matrices in a three-step approach which shows high prediction accuracy. First a principal component analysis (PCA) algorithm is used to reduce the dimension of the historical O-D and separate the main structure patterns from noisier components. Second, a state-space model is established for significant structure patterns, deviations, and to make a prediction using a Kalman filter. Lastly, a k-Nearest Neighbour (k-NN) based pattern recognition method is used to identify and predict structure patterns and deviations.

2.1.6 Emerging Methodologies

Due to the emerging area of ITS, other modelling techniques have also been used to solve the O-D estimation to better address existing problems or incorporating emerging data sources.

In contrast to the actively collected data used in transportation analysis, the ubiquitous usage of mobile phones, and social media has produced massive amounts of passively collected data. There have been considerable attempts to reconstruct O-D matrices from the passively collected data produced by the traces of mobile devices (Alexander et al., 2015; Bonnel et al., 2018; Caceres et al., 2020; Fekih et al., 2020; Gundlegård et al., 2016; Toole et al., 2015; Wu et al., 2018). An example of this type of data is referred to as call detail records (CDRs), which document the information for a telecommunications transaction (e.g. phone call, text message). Alexander et al. (2015) has inferred individual O-D matrices based on activity inference from triangulated CDRs which were then expanded to the population of the user's home Census Tract. Toole et al. (2015) combined the CDRs with crowdsourced geospatial data, census records, and travel surveys to obtain attributes such as trip purpose and travel mode. Inferred O-D demand are again expanded to the population and compared to O-D estimates from travel survey. In similar methods, the work in O-D estimation from CDRs continued in Rhone-Alpes, France Bonnel et al. (2018) and Lyon, France Fekih et al. (2020) both providing a validation step using household survey data. Gundlegård et al. (2016) addresses the methodology of entering CDRs into traditional travel demand models which include O-D matrix estimation and mobility metrics development. Portable mobile devices equipped with GPS antennas are also a source of continuous information for O-D estimation research (Ge & Fukuda, 2016; Moreira-Matias et al., 2016). An incremental framework to maintin statistics on urban mobility dynamics over a time-evolving O-D matrix was proposed by Moreira-Matias et al. (2016) using highstreams of GPS traces from heterogeneous sources. The framework is able to maintain flexible O-D matrices and discover dense regions of interest for real-time analysis on human mobility. Additionally, social media data has also gained attention for providing additional information in travel demand models. Z. Cheng et al. (2020) explored the utility of social media in combination with household travel surveys to improve O-D matrix quality based on a random forest technique. A comparison is conducted between O-D generated by travel survey and O-D generated with the addition of social media in eight US cities/regions consisting of metropolitan areas, rural areas and tourist areas. The results show that the social media data generally improves travel demand estimates in all areas, with metropolitan areas being most notable.

The bilevel framework is typically used to address the interdependency between the O-D matrix and traffic assignment, mentioned in previous chapters. The upper-level uses a statistical technique (e.g. GLS, MLE) to solve the O-D matrix while the lower level solves the assignment with UE. Frederix et al. (2013) addresses the effects of congestion in dynamic O-D estimation and the limitations of the commonly used bilevel formulation. These models may have biased results by assuming a linear relationship between link flow and O-D flow in a congested network. Additionally, the objective to locate local minima is an incorrect interpretation of information, revealing another source of error. To address the bias, a response function, represented as a first order Taylor expansion, is used to approximate the relationship between O-D flows and link flows and an O-D matrix representing similar congestion patterns to be used as the initial matrix. Furthermore, Gómez et al. (2015) evaluates the tradeoffs between loop detector coverage and penetration of FCD with respect to the accuracies O-D estimates. The methodology is based on a bilevel optimization employing fuzzy logic theory showing promising results that could applied to real-time traffic management.

Genetic algorithms have also been introduced for solving the bilevel programming models. Baek et al. (2004) proposed a multi-vehicle O-D estimation technique using a genetic algorithm from traffic counts. This approach is especially promising for urban networks where a true O-D matrix is not known, however at the expense of longer computational times. Furthermore, Lee et al. (2011) addresses the limitations of inherent non-convexity in a bilevel models by estimating dynamic O-D flows based on Kerner's three phase traffic theory. Using real-time traffic flow, speed, and occupancy, a genetic algorithm is employed to estimate the dynamic O-D flows.

Simultaneous perturbation stochastic approximation (SPSA) is typically used as a solution for the dynamic O-D problem due to its easy implementation and ability to solve general formulations from a wide range of data source. The algorithm, however, has limitations due to noise generated by uncorrelated measurements that grows rapidly as the problem scale increases and as a result, the accuracy and convergence rate suffers (L. Lu et al., 2015). Tympakianaki et al. (2015) proposed a modified SPSA algorithm to overcome some of the aforementioned limitations, improve the stability , and reduce the noise of uncorrelated measurements. The goal of the modified algorithm is to cluster components of the O-D matrix into a small number of "homogeneous" clusters via k-means clustering. The SPSA algorithm then approximates the gradient separately for O-D pairs belonging to a specific cluster which shows to outperform traditional SPSA algorithms in terms of stability, quality, and computation efficiency. Nigro et al. (2018) uses a SPSA coupled with asymmetric design (AD) for gradient computation and polynomial interpolation (PI) of the objective function for linear optimization. The research evaluates the improvement of dynamic O-D estimation problem from using FCD. Ros-Roca et al. (2021) proposes an SPSA technique to extend the conventional bilevel static estimation to dynamic estimation while providing an analysis of the solutions obtained in both convergences to observed traffic data and structural similarity to an a priori O-D matrix.

A seminal contribution by Antoniou et al. (2016) draws the necessity to provide a benchmarking platform for comparing dynamic O-D estimation techniques with a common evaluation setup under standardized conditions. It is built on synthetic testbeds, utilizing a simulated ground truth to test five methodologies consisting of LSQR, SPSA AP-PI, enhanced SPSA AD-PI, SPSA conjugate gradient (CG)-trust region (TR).

To support the dynamic and complex nature of transportation systems, neural networks have been regarded as an opportunistic tool for O-D estimation. Lorenzo and Matteo (2013) proposed a multilayer feedforward neural network (NN) for O-D estimation using link flows. To reduce the dimensionality of the input data and improve significance for NN learning, a PCA is proposed. The trained model shows capability of generating estimation in almost real-time and better performance than commercial O-D estimation programs regarding error and computing time. Sana et al. (2018) used information from Google Aggregated and Anonymized Trips (AAT) and household travel surveys to predict O-D person trips. The relationship between anonymized weights and actual trips is modelled using machine learning techniques consisting of OLS, k-NN, support vector regression, random forest regression, and NN. Ou et al. (2019) proposed an estimator for dynamic O-D flows using machine learning algorithms in a learn-assign-search framework. A convolutional neural network (CNN) is designed as a learner to capture patterns to characterize dynamic mappings between estimated O-D flows and assigned link flows. The trained CNN is then used as the assigner to estimate a regular and real-time O-D sequentially and finally, two genetic algorithms are developed as the searcher. Similarly, Tang et al. (2021) utilizes a three-dimensional CNN to learn dimensional correlation between traffic patterns from AVI observations and O-D flows. Furthermore, Krishnakumari et al. (2020) developed a data-driven method to construct a system of equations utilizing all given data without making more assumptions than needed. The framework requires three ingredients consisting of: a method to predict production and attraction time series, a method to compute N-shortest paths in each O-D pair, and two assumptions on the magnitude of N and on proportions of path flows between O-D pairs. A PCA is introduced for large networks where ingredients may be insufficient to solve the resulting system of equations and an artificial NN employed to model the relationship between supply patterns and productions/attractions. Multi-task deep NNs are presented by Katranji et al. (2020) to estimate individual O-D matrices based on static censuses. The model learns the temporal distribution of displacement from census data and applied on mobility sources. The resulting model yields higher accuracy than single task learning and is portable to other locations which can be especially helpful for cities without household surveys.

Newer research methods have proposed hybrid algorithms (X. Li et al., 2017), other machine learning techniques (W. Ma et al., 2020; Woo et al., 2016; Wu et al., 2018), and primal-dual algorithms (Michau et al., 2017) to address the O-D estimation problem. X. Li et al. (2017) has combined the nonnegative matrix factorization (NMF) algorithm and the Autoregressive (AR) model to estimate matrices in different temporal and spatial scales. The NMF algorithm reveals the characteristics of travel flow and constructs the basis matrix. The AR model then estimates the nonlinear time series coefficient matrix, which was extracted from the NMF algorithm. Finally, the O-D is then predicted based on the estimated coefficient matrix and basis matrix. Woo et al. (2016) developed a framework to predict future O-D demand for real-time series. The methodology searches a large historical data reserve for k-number of dates with similar traffic pattern to the current state. Three strategies are used to implement different feature vectors for k-NN of a single-level O-D demand, multilevel O-D demand, and single-level point demand. Wu et al. (2018) presented a layered computation graph to estimate hierarchical travel demand using multi-source data to better address travel behavior and system observability. The model utilizes data from travel surveys, CDRs, FCD, and traffic sensors to calculate travel demand, path flows, and link flows with a multi-layered hierarchical flow network. This work provides a solution for a single-class static O-D estimation problem which is then extended by W. Ma et al. (2020) to addresses the multi-class dynamic O-D estimation problem. Michau et al. (2017) extends O-D estimation to link-dependent O-D estimation as a optimization problem with traffic counts and probe trajectories. The objective function, consisting of five convex functions each representing a model constraint or property, is minimized and estimated with a proximal primal-dual algorithm.

2.2 O-D Estimation Model Evaluation

In the existing research, statistical measures have been used to quantify the quality of O-D estimates in comparison to a ground truth. Some of the commonly used and notable measures of goodness-of-fit are summarized in Table 2.1.

Table 2.1: O-D Model Evaluation Measures

Measure	**Equation[1]**	**Used by**
Percent error (PE)	$\frac{x_i - y_i}{y_i}$	(Bierlaire & Crittin, 2004), (A. Chen et al., 2005)
Root mean square error (RMSE)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2}$	(Cascetta, 1984), (Nihan & Davis, 1987), (Sherali et al., 1994), (Ashok & Ben-Akiva, 2000), (Hazelton, 2000), (Hazelton, 2001), (Ashok & Ben-Akiva, 2002), (Sherali et al., 2003), (Lo & Chan, 2003), (A. Chen et al., 2005), (Zhou & Mahmassani, 2006), (Nie & Zhang, 2008), (Mínguez et al., 2010), (Xie et al., 2011), (Castillo et al., 2013), (Talebian & Shafahi, 2015), (Tympakianaki et al., 2015), (Z. Lu et al., 2015), (Xie & Duthie, 2015), (Menon et al., 2015), (Ge & Fukuda, 2016), (Moreira-Matias et al., 2016), (Antoniou et al., 2016), (X. Li et al., 2017), (Michau et al., 2017), (Fujita et al., 2017), (X. Yang et al., 2017), (Tesselkin & Khabarov, 2017), (Sana et al., 2018), (Vogt et al., 2019), (Z. Cheng et al., 2020), (Pitombeira-Neto et al., 2020), (Mo et al., 2020), (J. Liu et al., 2020)

Root mean square normalized error (RMSNE) or Root mean square percentage error (RMSPE)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - y_i}{y_i}\right)^2}$	(Cascetta et al., 1993), (Ashok & Ben-Akiva, 2000), (Cascetta & Postorino, 2001), (Ashok & Ben-Akiva, 2002), (Wong et al., 2005), (Barceló et al., 2013), (Talebian & Shafahi, 2015), (Menon et al., 2015), (Antoniou et al., 2016), (X. Yang et al., 2017), (W. Ma & Qian, 2018b), (Bachir et al., 2019)
Mean square error (MSE)	$\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2$	(Cremer & Keller, 1987), (Bierlaire, 2002), (Cascetta et al., 1993), (Lo et al., 1996), (Cascetta & Postorino, 2001), (Cascetta et al., 2013), (Caggiani et al., 2013), (X. Yang et al., 2017), (Wu et al., 2018)
Mean absolute error (MAE)	$\frac{1}{N}\sum_{i=1}^{N}\lvert x_i - y_i\rvert$	(Cascetta, 1984), (Sherali et al., 1994), (Wong & Tong, 1998), (Sherali & Park, 2001), (Sherali et al., 2003), (Menon et al., 2015), (Tympakianaki et al., 2015), (Z. Lu et al., 2015), (Feng et al., 2015), (Moreira-Matias et al., 2016), (Antoniou et al., 2016), (Hu et al., 2017), (Carrese et al., 2017), (Tesselkin & Khabarov, 2017), (X. Li et al., 2017), (Bauer et al., 2018), (Katranji et al., 2020), (J. Liu et al., 2020)

Relative Mean Error (RME) or Mean absolute normalized error (MANE) or Mean absolute percentage error (MAPE)	$\frac{1}{N}\sum_{i=1}^{N}\frac{\lvert x_i - y_i\rvert}{y_i}$	(Cascetta et al., 1993), (Baek et al., 2004), (Sohn & Kim, 2008), (Lee et al., 2011), (Feng et al., 2015), (Menon et al., 2015), (Gómez et al., 2015), (Z. Lu et al., 2015), (Antoniou et al., 2016), (Ge & Fukuda, 2016), (Woo et al., 2016), (X. Li et al., 2017), (Ye & Wen, 2017), (X. Yang et al., 2017), (Carrese et al., 2017), (Krishnakumari et al., 2020), (Mo et al., 2020), (J. Liu et al., 2020)
Mean normalized error (MNE)	$\frac{1}{N}\sum_{i=1}^{N}\frac{x_i - y_i}{y_i}$	(Talebian & Shafahi, 2015), (Antoniou et al., 2016)
GEH Statistic	$\sqrt{\frac{2(x_i - y_i)^2}{x_i + y_i}}$	(Z. Lu et al., 2015), (Antoniou et al., 2016), (Nigro et al., 2018), (Ou et al., 2019), (Vogt et al., 2019)
Total demand scale (TDS)	Procedure detailed by (Bierlaire, 2002)	(Bierlaire, 2002), (A. Chen et al., 2005)
Theil's inequality coefficient (U)	$\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}y_i^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}}$	(Antoniou et al., 2016), (Tympakianaki et al., 2015), (Barceló et al., 2013)

[1] x_i = predicted measure; y_i = observed measure; N = number of measures

Some other additional but less common statistical measures used for O-D estimation include: average percentage error (Lorenzo & Matteo, 2013), weighted absolute percentage error (Sana et al., 2018), scaled RMSE (Parry & Hazelton, 2012), weighted MAPE (Woo et al., 2016), maximum error (X. Li et al., 2017), maximum possible relative error (Ehlert et al., 2006), weighted maximum possible error (Fu et al., 2019), and total demand captured (Nie & Zhang, 2008).

Additionally, correlation coefficients are also of interest to determine the correlation between the estimated results and external data. For example, the Pearson coefficient is used to calculate correlation between mobile phone data input and the estimation results (Bachir et al., 2019; Vogt et al., 2019). Spearman's rho measurement (Menon et al., 2015) and phi coefficient (Sherali et al., 2003) are also used to determine the strength of association between predicted link flow and true link flow. More recently, structural similarity measures evaluation has been of interest in the emerging O-D estimation methodologies (Behara et al., 2020; Caceres et al., 2020; Mungthanya et al., 2019). It can be used as an measure of effectiveness when comparing matrices of different dimensions since the statistical measures cannot be applied (Mungthanya et al., 2019).

2.3 Specialty Cases

To date, much of the research related to O-D estimation has focused on understanding the demand of underlying vehicles, however, the research area has extended to include specialty cases and subproblems such as network sensor location and demand estimation for transit.

2.3.1 Network Sensor Location

The network sensor location problem (NSLP) in research has mostly been regarded as a subproblem of the O-D estimation. An originating study by Lam and Lo (1990) examines how traffic count information affects the model performance of a EM based O-D estimation model. The principle of traffic flow coverage for priority links was proposed in this study. H. Yang and Zhou (1998) addresses the traffic count location problem with determining the optimal number and locations of traffic count points for a given prior O-D distribution pattern with fixed link choice properties. Bianco et al. (2001) proposed an iterative two-stage procedure based on greedy heuristics to determine optimal sensor location for the best mean O-D demand. Ehlert et al. (2006) expands the research area by adjusting the objective to find a second-best location procedure for the scenario where the maximum number of sensors cannot cover all O-D pairs. This problem is then extended by Hadavi and Shafahi (2016) where the trade-off between sensor budget and percentage of covered paths is studied with consideration of two target flows: O-D flows, and path flows. A series four models are considered with respect to the following: minimum number of sensors to ensure uniqueness of path flows, maximum number of observable path flows within budget constraint, minimum number of sensors to ensure uniqueness of O-D flows, and maximum number of observable O-D flows within budget constraint. Additionally, Fei et al. (2013) addresses maximizing the reliability of estimations when solving the NSLP and accounts for uncertain events in a transportation network (e.g. traffic incidents).

Further research led the to the combination of both the traffic count location and the O-D estimation problems. Zhou and List (2010) proposed a methodology for locating a limited set of traffic count location and AVI readers within a given network. The methodology aims to maximize expected information for the subsequent O-D estimation problem. Additionally, Mínguez et al. (2010) extends the NSLP to plate scanning devices given a prior O-D distribution pattern and budget constraints.

More recently, the NSLP has been extended to compressed sensing framework by Ye and Wen (2017) The basis of compressed sensing can recover signal vector directly from under sampled measurements and slightly promotes reconstruction accuracy. Additionally, Fu et al. (2019) optimizes traffic count locations and combines the O-D estimation problem with stochastic link proportions. The study optimize considers the covariance effects which is important in O-D estimation.

2.3.2 Public Transit

The O-D estimation problem has also been extended to estimate and update passenger trip matrices within transit networks using a EM framework (Nguyen et al., 1988; Wong & Tong, 1998) With the implementation of automatic fare collection (AFC) systems in transit agencies worldwide, smart card data has become a large-scale source for investigating passenger pattern movement.

Substantial work has been done to estimate station-to-station O-D matrices from smart card data for metropolitan cities across the world and has shown promising results for the replacement of passenger O-D surveys which are costly and time consuming (Barry et al., 2002; Munizaga & Palma, 2012; Zhao et al., 2007). Due to the high comprehensive spatiotemporal coverage of smartcard data, it has become an important element of transit management and planning.

2.4 Summary

Contrasting to the richness of methodologies in the literature, only select authors have evaluated their estimation methods in a structured framework to evaluate performances in an array of scenarios. Of the reviewed literature from the past two decades, select authors have tested their methodologies against different estimator/predictor, data type and/or data amount. A summary of these methodologies is provided in Table 2.2.

Although Table 2.2 has shown how some researchers conducted sensitivity analyses on different properties to evaluate select impacts of their methodology, very few have established a benchmarking procedure. The framework of benchmarking dynamic O-D estimation was first introduced by Djukic et al. (2011). This research was furthered explored by Antoniou et al. (2016) where a comparison of five estimators and four data inputs is conducted on a single testbed. With the implementation of a benchmarking framework, only then can the true strengths and limitations of the estimation methods be shown with quantitative results. Furthermore, of the summarized research in Table 2.2, there has yet to be an investigation of varying all properties (i.e., estimator, data type, and data amount) under a benchmarking framework. Additionally, many authors also indicate that comparisons of their proposed methodology is required in future research (Abdelghany et al., 2015; Mamei et al., 2019; Mitra et al., 2020; Xie & Duthie, 2015; X. Yang et al., 2017) This work bridges the gap in research to cover the effects of testing the performance when adjusting all three properties.

Table 2.2: Properties of O-D Estimation/Prediction

Research	**Estimator**	**Data Type**	**Data Amount**
(Tang et al., 2021)	+		
(Ros-Roca et al., 2021)	+		
(Rostami Nasab & Shafahi, 2020)		+	+
(W. Ma et al., 2020)		+	+
(Mo et al., 2020)			+
(Sana et al., 2018)	+		
(W. Ma & Qian, 2018b)	+		
(Wu et al., 2018)		+	
(Rao et al., 2018)			+
(Nigro et al., 2018)		+	+
(X. Li et al., 2017)	+		+
(Carrese et al., 2017)		+	+
(Fujita et al., 2017)			+
(X. Yang et al., 2017)			+
(Antoniou et al., 2016)	+	+	
(Moreira-Matias et al., 2016)	+		
(Gómez et al., 2015)		+	+
(Tympakianaki et al., 2015)	+		
(Feng et al., 2015)			+

(Barceló et al., 2013)		+
(Cascetta et al., 2013)		+
(Caggiani et al., 2013)	+	+
(Sun & Feng, 2011)		+
(Djukic et al., 2011)		+
(Lou & Yin, 2010)	+	
(Mussone et al., 2010)		+
(Sohn & Kim, 2008)	+	+
(Zhang et al., 2008)	+	+
(Tsekeris & Stathopoulos, 2008)	+	
(Park et al., 2008)	+	
(Gan et al., 2005)		+
(Chootinan et al., 2005)		+
(Baek et al., 2004)	+	
(Bierlaire & Crittin, 2004)	+	
(Sherali et al., 2003)	+	
(Nie & Lee, 2002)	+	
(H. Kim et al., 2001)	+	+

Chapter 3

Methodology

As shown in the literature review, existing research covers a wide range of methodologies and data types to solve the O-D estimation problem. However, these models cannot be fairly compared to one another for the evaluation of effectiveness, nor the value of information added with additional data quantities or sources. The methodology in this research, which is categorized into three overarching procedures and illustrated in Figure 3.1, addresses this challenge by utilizing a simulation-based approach to establish a ground-truth scenario.

The first procedure covers the model building portion and data extraction (3.1), which documents the setup of the testbed environment with a Vissim network to replicate a consistent experimental design. For every scenario selected, a ground-truth benchmark is established with *true* O-D demand. Simultaneously, a master dataset is generated with a selected subset of data that emulate the types of data extracted from real world technologies. Additionally, by using a simulation approach, this research allows room for experimentation by varying the levels of coverage/penetration of each data type to be input into the O-D estimation algorithm.

The second procedure reviews the scenario selection criteria (3.3). To start this section, the selection procedure of the existing O-D estimation algorithms covered in Chapter 2 are filtered. Then, the detailed methodology of each estimation algorithm is provided in this section.

Lastly, the third procedure outlines the comparison of scenarios (3.4).To start this section, a list of evaluated scenarios is provided. Finally, the selected measures of effectiveness are provided for evaluation.

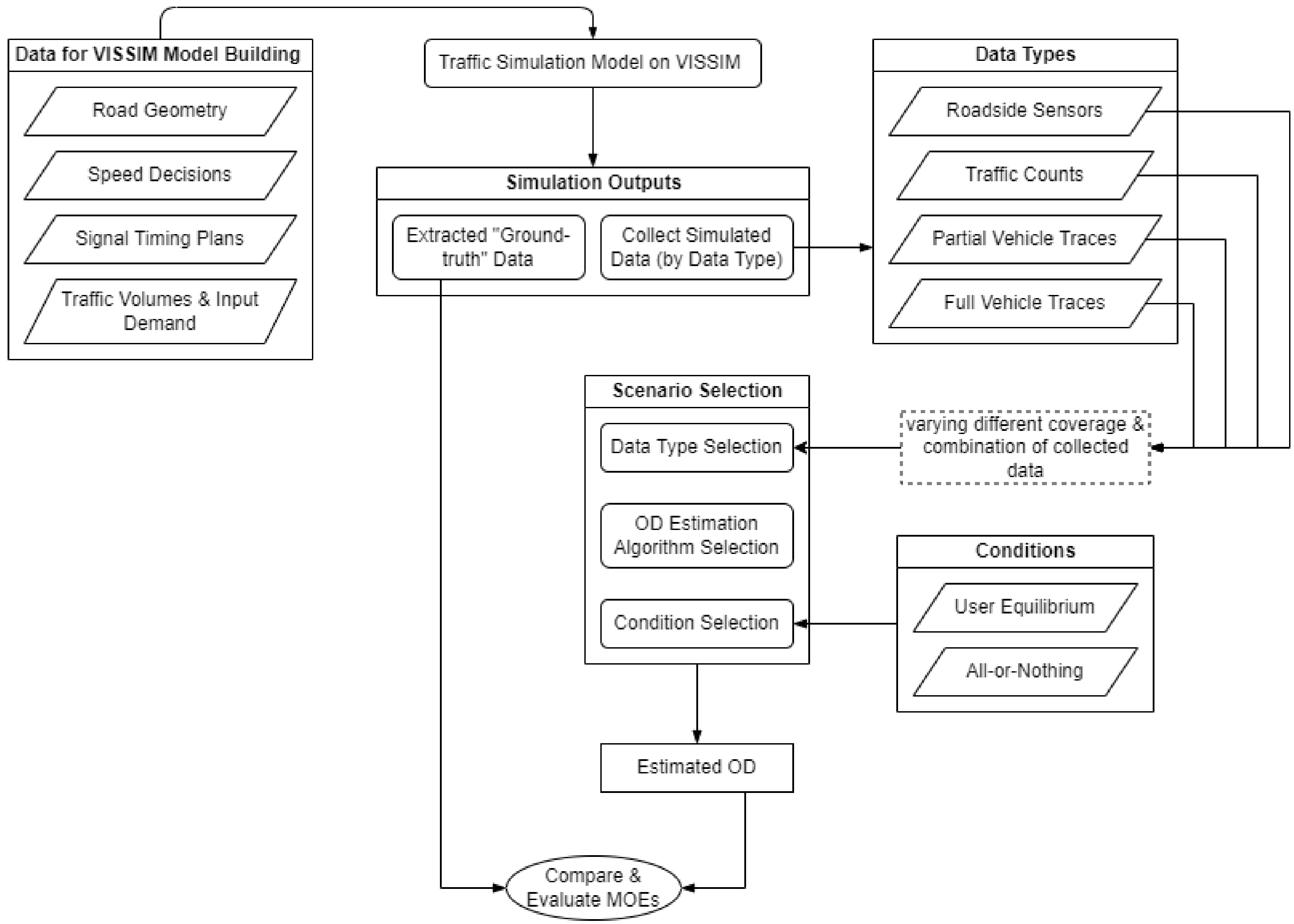

Figure 3.1: Research Framework

3.1 Simulation Model

The simulation model for this research was created in PTV Vissim 2020. A traffic microsimulation method is chosen due to its ability to emulate the dynamics of a network with a given demand. The details of the data input for this model as well as the data extraction are described in this section.

The Sioux Falls network was illustrated in Figure 3.2. The network consists of 24 nodes, 552 O-D pairs and 187 km of road. All nodes act both as origins and destinations in this study. The Sioux Falls location is selected as the test network due to its notoriety within the transport research community. However, it is noted that this research does not intend to replicate the real City of Sioux Falls, SD and remains a fictitious case. The nodal and link references are taken from Transportation Networks for Research Core Team (n.d.), which is an adaptation from LeBlanc et al. (1975).

3.1.1 Data Input

Prior to using the Vissim model for data generation purposes, the Sioux Falls network was first defined with a set of input model characteristics. This process aims to replicate the real network with a reasonable accuracy to produce an accepted level of realism in the data generation process. The subsequent subsections provide details of the data sources used for the creation of this model.

3.1.1.1 Road Geometry

The road geometry of the Sioux Falls network is constructed based off two map sources: OpenStreetMap (OSM) for link location placement and Google Maps for lane configuration. Vissim has built-in map providers, such as OSM, which was used for the link placements on to the Sioux Falls road network. The high visual resolution of Google satellite maps was used to determine the lane configurations of each link. Additionally, using the lane configurations provided by Google satellite maps provide real-world physical characteristics of the Sioux Falls network, thus covering many road function classifications such as freeways, arterials, and collectors.

Corresponding with the nodes and links from LeBlanc et al. (1975), a selected number of intersections from the real-world network are used for this model. These selected intersections are illustrated in Figure 3.3.

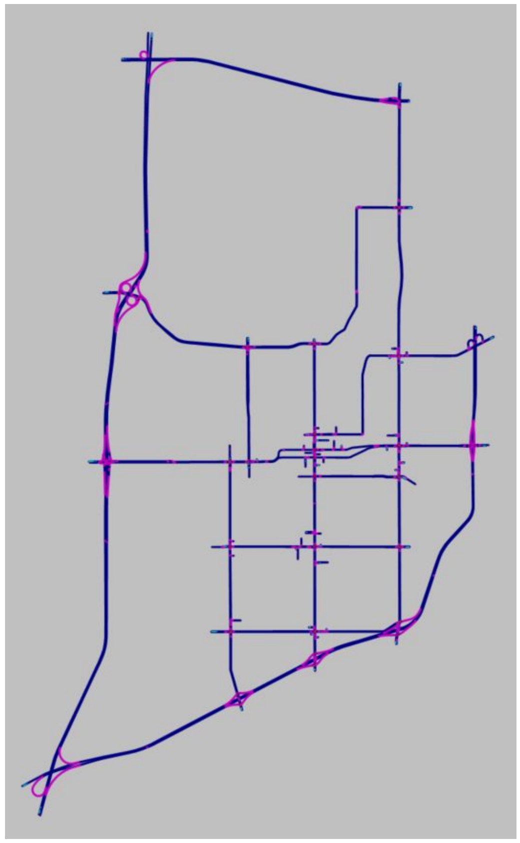

Figure 3.2: Sioux Falls Vissim Network

3.1.1.2 Traffic Volume and Demand

The traffic volumes and demand set for this model are adapted from the full-day demand set by Transportation Networks for Research Core Team (n.d.). The simulation period of this research is conducted for a 1-hour interval after a warm-up period of 15-minutes. Therefore, the demand is adapted as a proportion of the reference demand for all-or-nothing (AON) and user equilibrium (UE) condition. The proportion utilized for this research divides the full-day demand by 24-hours, then multiplying by 2 to create a congested network (i.e., demand exceeding the simulation hour).

Due to the stochastic nature of the Vissim model, each simulation slightly varies the input demand thus, creating a different ground-truth O-D demand for each condition. The procedure to extract the true O-D demand is discussed in Section 3.2.2.

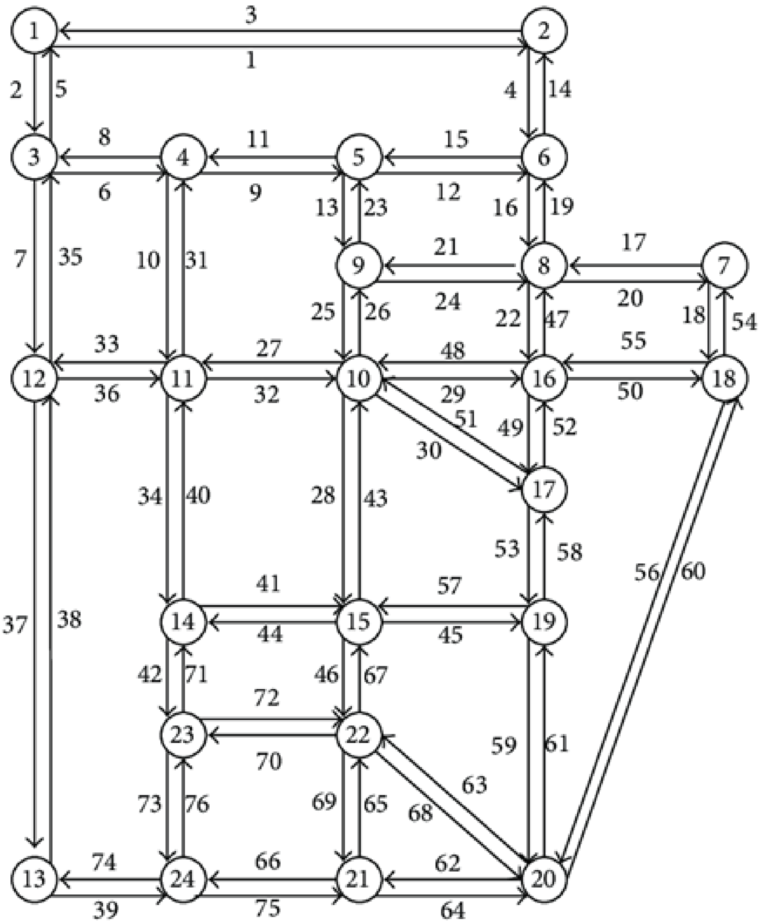

Figure 3.3: Sioux Falls Network – Selected Nodes (H. Liu et al., 2015)

3.1.1.3 Signal Plans

The signal placements of the road network also follow the real-world scenario that is shown in both OSM and Google Maps. The creation of signal timing plans is conducted in PTV Vistro through the built-in signal optimization options. With the established demands from the previous section, the O-D demands are inputted into Vistro. The built network from Vissim is exported into Vistro for the creation of signal timing plans. The final signal plans of the network consist of actuated and semi-actuated signals that are exported from Vistro as ring barrier controllers and directly inputted into Vissim.

3.1.1.4 Paths

A procedure is required to filter the enumerated paths into a set of efficient (or reasonable) paths. Following the procedure outlined by Dial (1971), and to avoid biases of some O-D estimation algorithms with a path enumeration procedure (e.g., logit), a path is assumed to be efficient if it does not backtrack. A visualization of the efficient paths considered is illustrated in Figure 3.4. The resulting network considers a total of 2069 paths.

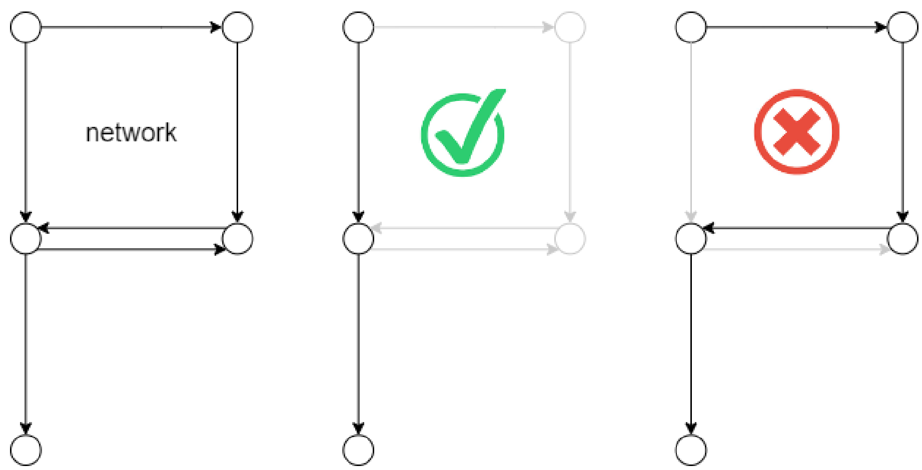

Figure 3.4: Path Consideration Example

3.1.1.5 Conditions

Two assignment conditions related to the path selection are considered in this research: AON assignment and UE assignment, conducted by an incremental assignment algorithm. The AON assignment loads trips between O-D pairs onto the path with shortest travel time. The detailed procedure of incremental assignment is provided in Appendix A.

The two selected conditions are chosen as a basis to evaluate the performance of the O-D estimation techniques where the assignment methodologies are as different as possible. AON concentrates all demand onto the single shortest path, whereas UE diffuses demand to the extent that all used paths have equal travel times. By testing these conditions, the results could illustrate strengths or weaknesses of the estimation techniques in specific scenarios. Since real-world traffic assignment may not reflect an AON or a UE approach (maybe somewhere in between), a strong estimation performance of both conditions may reflect its ability to estimate in all conditions. Lastly, since some estimation techniques rely on an underlying UE assumption as a part of its methodology, the test of an AON assignment may reveal certain limitations of its assumptions.

3.2 Data Generation

The data generation process aims to collect data from the simulation model to emulate real-world data collection technologies. With some of the built-in evaluation outputs of Vissim, many real-world technologies can be reasonably reproduced. Table 3.1 lists the real-world data sources that are grouped with each Vissim evaluation output and the specific formatted data that is extracted as inputs for O-D estimation algorithms.

Table 3.1: Vissim Data Generation

Real World Source	**Vissim Output**	**Extracted Data**
Road Sensors (e.g. Bluetooth detectors, toll plazas)	Data Collection Points	Time Stamp, Speed
Traffic Counts (e.g. loop detectors, cameras)	Data Collection Points	Flow, Time Stamp
Partial Vehicle Traces (e.g. license plate recognition, automatic vehicle identification)	Vehicle Records	Path reconstruction, travel time
Full Vehicle Traces (e.g. probe vehicles, GPS traces)	Vehicle Records	O-D Flows, Path Reconstruction, Travel Time

Given the microscopic nature of the Vissim modelling software, many parameters can be adjusted to closely emulate the details of specific networks (e.g., driving behaviours). However, some of these parameters are not critical to the scope of this research, thus these parameters are set as default. The parameters that are adjusted for this research are listed as follows:

1. **Data collection interval:** this parameter varies the frequency at which data is collected (interval ranges from 1s to 3600s depending on the data required)

2. **Simulation Period:** Selected to be 4500s (allowing a 900s warm-up period)

3.2.1 Vehicle Records

Vissim can extract full trajectories with the *Vehicle Record* evaluation output. For each vehicle within the simulation period, a record is produced at a selected interval. The resolution of the time interval is selected for every 1-time step (i.e., the resolution at which the simulation is updated). The format of the *Vehicle Record* evaluation output is illustrated in Table 3.2.

Table 3.2: Example of Vehicle Record Attributes

Simulation Second	**Vehicle ID**	**Link ID**
2450	3002	6

Therefore, for each simulated scenario, the ground-truth O-D matrix can be extracted by matching all *Link ID* to origins and destinations for every unique *Vehicle ID*.

In addition to the ground truth O-D matrix, *Vehicle Records* are also used for partial and full traces. The partial traces are representative of data collection methods that are typically fixed in location and can identify unique vehicles. Thus, the raw vehicle records are aggregated on the link level, and the levels of penetration with respect to data amount ranges from 0 to 76 links (e.g., 50% of partial traces means 38 of 76 links collect the *Simulation Second* and *Vehicle ID* of all vehicles).

In contrast to partial traces, full traces are representative of a full path reconstruction for an individual *Vehicle ID*. Therefore, the raw vehicle records are aggregated on the *Vehicle ID* level, and the level of penetration with respect to data amount occurs on the total demand level (e.g., 5% of full traces means that 1,040 of 20,818 vehicles collect *Simulation Second* and *Link ID* at every time step).

3.2.2 Data Collection Points

Data Collection Points can be defined on specific location(s) of a link within the Vissim network that would collect similar measurements to road sensors and traffic count devices. The record for *Data Collection Points* is produced at a selected 3600s interval (i.e., the simulation hour) where data for a specific link is aggregated as an average for speed or a sum for link counts. The format of the *Data Collection Points* evaluation output is illustrated in Table 3.3.

Table 3.3: Example of Data Collection Point Attributes

Time Interval [s]	Link ID	Vehicle Count	Average Point Speed [km/hr]
900-4500	22	172	57.04

3.3 Scenario Selection

The scenario selection procedure covers the data selection, algorithm selection and conditions selection of the framework. To start, the algorithm selection procedure is presented in 3.3.1, where existing methodologies are filtered based on a set of scope requirements for this research. Prior to the presentation of modelling steps of chosen algorithms, a generic formulation of the O-D estimation problem is first in presented in Section 3.3.2 as a basis for all methods. Subsequently, the procedures of the selected algorithms are presented in Section 3.3.2.1 to 3.3.2.5.

3.3.1 Algorithm Selection

This section covers the selection procedure of the existing O-D estimation techniques covered in the literature review. The following list provides the exclusion procedure of the algorithm selection:

- Focus is on static O-D estimation techniques, thus dynamic and quasi-dynamic estimation algorithms are excluded.
- The algorithm is to be tested on a network with many O-D pairs. This eliminates algorithms that have not yet been shown to estimate demands successfully on larger and more representative networks.
  - As a basis, the smallest acceptable network consists of 20 O-D pairs, 20 nodes and 20 links.
- Network chosen cannot be single corridor network (i.e., no corridors/single major freeway or arterial, and no toy networks).
- Method cannot require vastly large amounts of data, multi-day data or datasets that cannot be reproduced (e.g., significant historical data).

- Techniques established from 2000 or earlier are excluded. While many fundamental procedures were established at or before this time, since then, other methodologies have claimed to provide additional benefit to the fundamental formulations.
- Estimation is for general O-D matrix (i.e., estimates O-D from data points to matrix) (non-specific types, e.g., sub systems, internal-internal , external-internal, etc.).
- Methodology provides a procedure for end-to-end estimation, thus eliminating procedures to improve a specific portion of the O-D estimation framework.

A detailed overview of the selection matrix is summarized in Appendix B for all static O-D estimation techniques reviewed previously and considered in this research.

3.3.2 Generic Formulation of the O-D Estimation Problem

To start, a generic formulation of an example network, is represented by a directed graph $G(N, L)$ consisting of node set N and link set L. Let t_{ij} be the O-D demand between origin i and destination j, and t be the vector with components t_{ij}. For every given link L, the flow f is the sum of the path flows using that link:

$$f = AF \tag{3.1}$$

where:

$$\begin{aligned} &f && = \text{link-flow vector} \\ &F && = \text{path-flow vector} \\ &A = a_{lk} = \begin{cases} 1, & \text{if } l \in k \\ 0, & \text{otherwise} \end{cases} && = \text{link-path incidence matrix} \end{aligned}$$

Furthermore, some algorithms require an average generalized cost associated with each link on the network. For consistency's sake, this research utilizes is the Bureau of Public Roads (BPR) cost function as it is the most used and explicitly selected as the function of choice in Castillo et al. (2008b). Thus, for every given link L, the cost c is

$$c_l(f_l) = T_l + \alpha_l \left(\frac{f_l}{k_l}\right)^{\beta_l} \tag{3.2}$$

where:

$c_l(f_l)$ = cost of link l
T_l = free flow travel time on link l
k_l = capacity on link l
α_l, β_l = link-specific constants

Consistent with typical values, $\beta = 4$ is held constant while other link properties (i.e., t_l, k_l, and α_l) are determined through the Vissim network. These properties are summarized in Appendix C. The resulting cost of path K is equal to the sum of link costs forming said path:

$$g = A^T c \tag{3.3}$$

where:

g = path-cost vector
c = link-cost vector

Finally, for the algorithms that require a target O-D demand as a source of information, a consistent target matrix is created by extracting 5% of all vehicles (non-homogeneous) and expanded to mimic an O-D matrix obtained from a household travel survey. The initial matrices used for both the UE and AON networks are provided in Appendix D.

3.3.2.1 Algorithm 1 – Fixed Point Approach to Estimating O-D Matrices Using Traffic Counts on Congested Networks

The first methodology by Cascetta and Postorino (2001) utilizes link counts as a data source for O-D demand estimation. From the generalized formulation, this algorithm utilizes a path enumeration approach for the estimation of the final demand vector. Thus, let p_{kij} be the fraction of demand of t_{ij} using path k. Thus, the actual path flow F can be expressed as:

$$F = tP \tag{3.4}$$

where:

$P = p_{kij}$ = path choice fraction matrix

By substituting Equation 3.4 into 3.1, the link flow can be expressed as the following:

$$f = tAP = H(c)t \tag{3.5}$$

where:

H = assignment matrix

The path-choice mode is obtained using a logit stochastic UE (SUE) model where first, path performance is calculated based on Equation 3.3 and the path probability matrix is

$$P = p_{kij} = \frac{\exp(-g_{kij}/\theta)}{\sum_{q \in k} \exp(-g_{qij}/\theta)} \tag{3.6}$$

where:

θ = dispersion parameter

The value of θ affects diversion probabilities and as the value of θ increases, the level of uncertainty increases, conversely as θ decreases, it corresponds to a more accurate view of route costs and the probability of trip using the shortest path increases (Dial, 1971). Consistent with the author (Cascetta & Postorino, 2001), a value of $\theta = 1.4$ is utilized where the parameter would be considered a middle value and does not to represent any extremes.

With the path probability matrix, the estimated assignment matrix can be found by the following:

$$H^* = AP \tag{3.7}$$

Furthermore, the flow f^* by assigning the actual demand t to the network can then be written as

$$f^* = APt = H^*t \tag{3.8}$$

The estimated assignment matrix H^* is generally different from the actual assignment matrix H due to approximations underlying the assignment model. Thus, the relationship between H^* and H is represented as

$$H = H^*(c) + E^{SIM} \tag{3.9}$$

where:

E^{SIM} = matrix of unknown assignment errors

By substituting Equation 3.9 into Equation 3.5, the link-flow vector can be represented as

$$f = Ht = (H^*(c) + E^{SIM})t = H^*(c)t + E^{SIM}t = H^*(c)t + \varepsilon^{SIM}t \tag{3.10}$$

where:

ε^{SIM} = vector of assignment error related to each link-flow

The premise of equation 3.10 illustrates that even if the real demand vector t were to be assigned to the network by the assignment model by the assignment matrix $H^*(c)$, the resulting flow vector would be different than the real one due to the error term ε^{SIM}. Additionally, the counted flow vector (i.e., measured on the path), $\bar{f}$ can also differ from the actual flows due to several reasons (e.g., measurement errors). Therefore, the discrepancy is represented as

$$\bar{f} = f + \varepsilon^{MEAS} \tag{3.11}$$

where:

ε^{MEAS} = vector of measure error related to each link-flow

By combining Equations 3.10 and 3.11 to express counted link flow from unknown O-D demand

$$\bar{f} = H^*(c)t + \varepsilon \tag{3.12}$$

where:

$\varepsilon = \varepsilon^{SIM} + \varepsilon^{MEAS}$ = both assignment and measurement errors

Moving forward, a fixed-point formulation is utilized for the estimation problem alongside information provided by link counts $\bar{f}$ and a sample-based (5% in this research) estimate of O-D demand $\bar{t}$. The sample-based O-D estimate is synthesized as a target and can be expressed as

$$\bar{t} = t + \eta \tag{3.13}$$

where:

η = vector of errors of target demand $\bar{t}$

In the situation of a congested network, the assignment map is nonlinear for demand t and link costs depend on link counts that are not known and thus having to be estimated.

The link flows are estimated as a result from the assignment of O-D demand t^*. The circular problem can obtain an estimated link flow $f(t^*)$ and its resulting costs $c(f(t^*))$. Thus, the equilibrium assignment model can is expressed as the following fixed-point problem.

$$f^* = AP[c(f^*)]t \tag{3.14}$$

Finally, the estimation problem is solved by Cascetta and Postorino (2001) through the following formulation

$$t = \underset{x \geq 0}{\operatorname{argmin}} (x - \bar{t})^T V^{-1} (x - \bar{t}) + (H^*(c(f(t^*)))x - f)^T W^{-1} (H^*(c(f(t^*)))x - f) \tag{3.15}$$

where:

V = variance-covariance matrix of error term η
W = variance-covariance matrix of error term ε

The final solution algorithm follows the procedure outlined by the author through functional iteration, which is described in the following:

Functional Iteration Algorithm

Step 0: Initialization $k = 0$
Initial demand vector $t^0 = \bar{t}$ explicit evaluation of the assignment matrix H^0 by using SUE assignment model with $c = c(t^0)$

Step 1: Update $k = k + 1$

Step 2: Solve the problem:

$$t = \operatorname{argmin}_{x \geq 0} (x^k - \bar{t})^T V^{-1} (x^k - \bar{t}) + (H^{k-1} x^k - \bar{f})^T W^{-1} (H^{k-1} x^k - \bar{f})$$

Step 3: Solve the SUE assignment problem for demand vector t^k, compute $c(f^k)$ and evaluate H^k:

$$f^k = f(t^k) \qquad c^k = c(f^k) \qquad H^k = H(c^k)$$

Step 4: Termination test on maximum percentage difference between the elements of t:

$$\max \frac{t_{ij}^k - t_{ij}^{k-1}}{t_{ij}^k} < \varepsilon$$

If the test is satisfied the algorithm ends, otherwise return to *Step 1*

When the termination test is satisfied, the final estimated matrix is obtained from *Step 2*. Further details of this algorithm are provided by its authors, Cascetta and Postorino (2001).

3.3.2.2 Algorithm 2 – Traffic Estimation and Optimal Counting Location Without Path Enumeration Using Bayesian Networks

The second methodology by Castillo et al. (2008b) utilizes link counts as a data source for O-D demand estimation. The approach of this methodology utilizes a Wardrop-minimum variance (WMV) method for assignment alongside a Bayesian network approach for estimation. For background, a brief summary of the Bayesian network proposed in (Castillo et al., 2008a) is provided, without which the following summary of the algorithm is difficult to understand.

To start, a Bayesian network model (Castillo et al., 2008b) is defined as a pair (G, P) where G, is a directed acrylic graph defined on a set of nodes X (the random variables), $P = p(x_1|\pi_1), \ldots, (x_m|\pi_m)$ is a set of m conditional probability densities (CPD) for each variable and Π_i is the set of parents of node x_i in G. The set P defines associated joint probability density of all nodes, which is provided in the following

$$p(x) = \prod_{m=1}^{n} p(x_m|\pi_m) \tag{3.16}$$

The Gaussian Bayesian network is defined with the joint probability distribution (JPD) associated with its variable X having a normal distribution $N(\mu, \Sigma)$. The author proposes a Gaussian Bayesian network as a basis for O-D estimation based on the following assumptions:

1. Link flows satisfy following relations

$$f = \beta t + \varepsilon \tag{3.17}$$

where:

$\beta = \beta_{lij}$ = matrix of flow proportions with origin i and destination j using link l
ε = link measurement errors

2. A correlation structure exists amongst the T random variables. The following assumption is made to represent these correlations and find the associated variance-covariance matrix

$$t = KU + \eta \tag{3.18}$$

where:

U	= random positive variable that measures the total level of flow
$K = k_{ij}$	= column matrix for O-D pair ij with respect to the total traffic flow

It is noted that the initial K value is derived from the same prior matrix $\bar{t}$, utilized in Algorithm 1. The resulting t vector is a multivariate normal $N(\mu_t, \Sigma_t)$ random variable with mean μ_t and variance-covariance matrix Σ_t. The author assumes that O-D and link flows satisfy the following system of equations.

$$E[t] = E[U]K \tag{3.19}$$

$$E[f] = E[U]\beta K + E[\varepsilon] \tag{3.20}$$

$$D_\eta = \text{Diag}(\eta_1, \eta_2, ..., \eta_r) \tag{3.21}$$

$$D_\varepsilon = \text{Diag}(\varepsilon_1, \varepsilon_2, ..., \varepsilon_r) \tag{3.22}$$

$$\sum_{tt} = \sigma_U^2 K K^T + D_\eta \tag{3.23}$$

$$\sum_{tf} = \sum_{tt} \beta^T \tag{3.24}$$

$$\sum_{ft} = \sum_{tf} \tag{3.25}$$

$$\sum_{ff} = \beta \sum_{tt} \beta^T + D_\varepsilon \tag{3.26}$$

The following Equations (3.27 - 3.30) update the mean and covariance when some variables are observed (further details can be seen in Castillo et al. (1997) As variables become observed, the other variables are conditioned based on the observation, thus, the expected values and covariances of remaining variables change.

$$\mu_{Y|Z=z} = \mu_Y + \sum_{YZ} \sum_{ZZ}^{-1} (z - \mu_Z) \tag{3.27}$$

$$\sum_{Y|Z=z} = \sum_{YY} - \sum YZ \sum_{ZZ}^{-1} \sum_{ZY} \tag{3.28}$$

$$\sum_{Z|Z=z} = 0 \tag{3.29}$$

Y = set of unobserved variables
Z = set of observed variables

The following equation fixes the mean value of the observed variable to actual observed value.

$$E[Z|Z] = z \tag{3.30}$$

Finally, the following equation uses the mean as the prediction of O-D pair assigned to different paths.

$$E[t] = E[Y|Z = z]_{t \subseteq Z} \tag{3.31}$$

Moving forward to the WMV equilibrium model and recall the traffic network $G(N, L)$ from the generic O-D formulation, which estimates link flows associated with each O-D pair. The following is the WMV optimization problem, which is a mixture of Wardrop equilibrium and minimum variance problem, and is defined as

$$\min_v Z = \sum_{l \in A} \int_0^{c_l(\sum_{ij} f_{ij})} c_l(f) dv + \frac{\lambda}{m} \sum_{l \in A} \sum_{ij} (f_{lij} - \mu)^2 \tag{3.32}$$

The values of c_l are obtained from Equation 3.2 in the generic O-D formulation. Further details of the WMV equilibrium model are provided by its authors Castillo et al. (2008b).

Traffic Estimation Bi-Level Algorithm Without Path Enumeration

Input: Data required consists of the following $E[U]$, K, c_l, α_l, q_l,γ_l, $\forall \in A$, and the observed link flows.

Step 0: Initialize the O-D flow to the initial guess for $E[t]$:

$$t_0 = E[t] = E[U]K$$

Step 1: Solve the master problem:

$$\min_v Z = \sum_{l \in A} \int_0^{c_l(\sum_{ij} f_{ij})} c_l(f)dv + \frac{\lambda}{m} \sum_{l \in A} \sum_{ij} (f_{lij} - \mu)^2$$

Step 2: Calculate the β matrix:

$$\beta_{lij} = \frac{v_{lij}}{t_{ij}}$$

Step 3: Update the O-D and link flow predictions using the Bayesian network. The new O-D pair and link flows are predicted with Equations 3.23 - 3.29.

Step 4: Convergence checking

$$error = (t_0 - E[t])^T (t_0 - E[t])$$

If the error is less than the tolerance, stop and return the values of t and f. Otherwise, let $t_0 = t$ and continue with *Step 1*

When the termination test is satisfied, the final estimated matrix is obtained from *Step 3*. Finally, additional details and assumptions are provided by its authors Castillo et al. (2008b).

3.3.2.3 Algorithm 3 – Estimation of Origin-Destination Matrices from Link Counts and Sporadic Routing Data

The third methodology by Parry and Hazelton (2012) estimates O-D matrices using a formal likelihood based approach from link counts and sporadic routing data, otherwise known as partial traces in this research. Furthering the general notation provided in Section 3.3.2, some additional variables and notations are defined.

Let K index the set of routes established in Section 3.1.1.4, with r denoting the total number of routes and K_{ij} representing the subset of routes with origin i and destination j.

By definition, $E[F^*] = F$, where the F^* denotes the random vector of realized route flows during the observational window. The distribution of F^* employs the Poisson model $F_k^* \sim Pois(F_k)$ and is independent of $F_{ij}^* : i \neq j$ for all $k \in K$. A Gaussian approximation is employed if traffic flows are not too small.

$$F^* \; N(F, \Lambda) \tag{3.33}$$

where:

$\Lambda = \text{diag(F)}$

The first data source, link counts are defined and related to route flows by Equation 3.1. Based on a normal approximation, the distribution of link counts is given by the following.

$$f = N(AF, A\Lambda A^T) \tag{3.34}$$

It is assumed that the covariance matrix $A\Lambda A^T$ is invertible.

The second data source, partial traces, provides routing information for a sample of vehicles. It is assumed that each vehicle on path k has a probability of p_k of being tracked, and the corresponding vector of probabilities is defined as $p = (p_1, \ldots, p_r)^T$.The observed path flows $F_{trk}^* = (F_{trk,1}^*, \ldots, F_{trk,r}^*)^T$ are based on tracked vehicles alone and applying normal approximation, the unconditional distribution of observed route flows are as follows

$$F_{trk}^* = N(PF, P\Lambda) \tag{3.35}$$

where:

$P = \text{diag(p)}$

The link counts f and partial traces F_{trk}^* are not statistically independent since they are observed contemporaneously. However, the two data sources can be decomposed to link counts as

$$f = f_{trk} + f_{not} \tag{3.36}$$

where:

$f_{trk} = AF_{trk}^*$ = the contribution of tracked vehicles to link counts
f_{not} = the contribution of non-tracked vehicles to link counts

Under standard assumptions F^*_{trk} and f_{not} are not independent. The distribution of F^*_{trk} is given in Equation 3.38 and the distribution of f_{not} is given in the following

$$f_{not} = N(AQF, AQ\Lambda A^T) \tag{3.37}$$

where:

$Q = I - P$ with I being the identity matrix

The links between mean vectors and covariance matrices have been defined in Equations 3.35 and 3.37. This methodology assumes that the covariance matrix of F is fixed (i.e., $Var(F^*) = \Sigma = diag(\sigma)$ and not dependent on F to simplify a complex log-likelihood (Hazelton, 2003, 2010). Thus, the simplified model is as follows

$$F^*_{trk} = N(PF, P\Sigma) \tag{3.38}$$

$$f_{not} = N(AQF, AQ\Sigma A^T) \tag{3.39}$$

Moving onto the likelihood theory of this method, the following likelihood functions are all in the following form given the independence of F^*_{trk} and f_{not}.

$$\mathcal{L}(F, p) = g(f_{not}|F, p)g(F^*_{trk}|F, p) \tag{3.40}$$

where:

$g =$ a probability mass/density function

The vector p in Equation 3.40 is a parameter that will be estimated in tandem with F. In principle p can be estimated alongside F using likelihood methods, however, there would be identifiability problems. The details of these issues are outlined in Parry and Hazelton (2012). Thus, a structure is assumed on p resulting in a simpler estimation problem. The assumption is that the penetration of technology is homogeneous, thus $p = p_0 1$ for a scalar p_0. The details of derivation and justification of assumptions can be found in Parry and Hazelton (2012) and the final estimator can be found by

$$p_0 = \frac{1^T AF^*_{trk}}{1^T(AF^*_{trk} + f_{not}} \tag{3.41}$$

Continuing with the likelihood theory, two models are presented. The chosen models tested for this research are the simplified models due to similarities in the performance accuracies and the computational advantages. The model utilizes link counts and partial traces, which constructs the following log-likelihood.

$$\begin{aligned} l(F) = &-\frac{1}{2}\log(|AQ\Sigma A^T|) - \frac{1}{2}(f_{not} - AQF)^T(A\Sigma QA^T)^{-1}(f_{not} - AQF) \\ &- \frac{1}{2}\log(|P\Sigma|) - \frac{1}{2}(F^*_{trk} - PF)^T(P\Sigma)^{-1}(F^*_{trk} - PF) \end{aligned} \tag{3.42}$$

Finally, further details and assumptions of the likelihood theory are outlined by its authors Parry and Hazelton (2012).

3.3.2.4 Algorithm 4 – Vehicle Trajectory Reconstruction using Automatic Vehicle Identification and Traffic Count Data

The fourth methodology by Feng et al. (2015) focuses on O-D estimation through full trajectory reconstruction based on partial vehicle traces and spot speed data utilizing a Kalman filter approach. This approach updates the state-space probability curve through four correction factors that approximate the true state-space probability curve of vehicles. It is noted that the original methodology consists of five correction factors, however, the measurability criterion is not considered in the research since the basis of this correction factor is to correct data collection errors. Since the data collection process is simulated, the error detection would not be necessary. The particle set consists of all potential trajectories between updates of the vehicle. To start, the vehicle trajectory reconstruction is presented as

---

Vehicle Trajectory Reconstruction

Step 1: Initialize the particle set:

Set $x^1, x^2, ..., x^I$ to represent the initial particle set and $P(x^1), P(x^2), \ldots, P(x^I)$ to represent the previous particle probability. I denotes the number of particles with an initial probability set to $1/I$. All possible paths are acquired from procedure outlined in Section 3.1.1.4.

Step 2: Importance sampling:

The probability distribution of particles is assumed to obey the density function that is derived from previous importance sampling based on four correction factors. Starting with the path consistency sampling, the initial particle weight follows a uniform distribution. The path consistency probability density function is derived from network topology and assigned an AON distribution. Path consistency is expressed through a weight-update equation presented in the following.

$$W_1^i = W_0^i \times \frac{P(y|\hat{x}_1^i)P(\hat{x}_1^i|x_0^i)}{q(\hat{x}_1^i|x_0^i, y)} \tag{3.43}$$

where:

W_1^i	=	non-normalized weight of possible trajectory i after path consistency updating
W_0^i	=	initial prior weight of possible trajectory i
y	=	objective data collected
$\hat{x}_1^i$	=	state space of possible trajectory i
x_0^i	=	particle for which the initial possible trajectory is i
$P(y\|\hat{x}_1^i)$	=	probability of possible trajectory i based on path consistency
$P(\hat{x}_1^i\|x_0^i)$	=	transformation probability from the previous data to path consistency (in the static case, the state transition probability is regarded as a constant)
$q(\hat{x}_1^i\|x_0^i, y)$	=	previous probability density function, in which the possible trajectory i is the true trajectory under a path consistency addition

The particle gathering equation is expressed as

$$N_1^i = \frac{W_1^i}{\sum_{i=1}^{N} W_1^i} \times N \tag{3.44}$$

where:

N_1^i = particle gathering number of possible trajectories i after the path consistency computations

Three correction factors remain: travel time consistency, gravity flow model and path-link flow matching. The remaining four resampling processes are conducted prior to *Step 3*. The details of the resampling process are provided in detail further on in this Section.

Step 3: Output the 'true' trajectory:

Based on particle filter theory, the posterior probability of all possible trajectories is represented as

$$P(X|y) \approx \hat{P}(X|y) = \frac{1}{N}\sum_{i=1}^{N}\delta(X^i)d(X) \tag{3.45}$$

where:

$P(X|y)$ = posterior probability density function after the updated probability

With the expectation function represented as

$$E(g(X)) = \int g(X)P(X|y)dX \approx \frac{1}{N}\sum_{i=1}^{N}N^i g(X) \tag{3.46}$$

where:

$g(X)$ = integral form of $P(X \mid y)$

The final result of vehicle reconstruction is obtained from the largest weight of the posterior state spatial curve.

The first correction factor is the path consistency analysis, which adopts the notion that all traces from the data must belong to a path in the set of paths. The updating process relies on an AON distribution where $W_0^i \sim \frac{1}{N}$ represents the path consistency probability density function. All possible paths follow a uniform distribution. Thus, the probability update formula is represented as

$$\begin{cases} P_{path}^i = 1 & \text{if } N \in N_{path} \\ P_{path}^i = 0 & \text{if } N \notin N_{path} \end{cases} \tag{3.47}$$

where:

N_{path} = node set of a complete trajectory
P_{path}^i = updated probability based on path consistency analysis

Moving on to the second correction factor, travel time consistency adopts the notion that a vehicle will select a path in which the average travel time is similar to the collected travel time between two data collection facilities. The methodology is adopted from R. Li et al. (2006), modified by Feng et al. (2015) for this algorithm, and estimates travel time from spot speeds, as shown in the following equation.

$$\tilde{t}(n_1, n_2) = \frac{2(x_{n1}, x_{n2})}{\tilde{v}(x_{n1}, \Delta t) + \tilde{v}(x_{n2}, \Delta t)} \tag{3.48}$$

where:

$\tilde{t}(n_1, n_2)$	= average travel time
n_1	= beginning of the link
n_2	= end of the link
(x_{n1}, x_{n2})	= length of the link
$\tilde{v}(x_{n1}, \Delta t)$	= average spot speed at the beginning of the calculation period
$\tilde{v}(x_{n2}, \Delta t)$	= average spot speed at the end of the calculation period

The following equation compares the true travel time between two data collection facilities and the average travel time.

$$P^i_{time} = \begin{cases} 0.01 & \text{if } \tilde{t}_i > 2t' \\ 2 - \frac{\tilde{t}_i(DC_1, DC_2)}{t'(DC_1, DC_2)} & \text{if } t' < \tilde{t} < 2t' \\ \frac{\tilde{t}_i(DC_1, DC_2)}{t'(DC_1, DC_2)} & \text{if } t' > \tilde{t} \end{cases} \tag{3.49}$$

where:

$\tilde{t}_i(DC_1, DC_2)$	= average travel time of possible path i between two data collection facilities
$t'(DC_1, DC_2)$	= true travel time between two data collection facilities
P^i_{time}	= updated probability based on travel time consistency

At this trajectory correction factor, the initial probability is $W_1^{(i)}$, which was calculated with the path consistency analysis correction factor, shown in Equation 3.43. Additionally, $P(y|\hat{x}_2^{(i)})$ is calculated through Equation 3.49 and $q(\hat{x}_2^{(i)}|x_1^{(i)}, y)$ is the probability density function of travel time consistency (regarded as a cost density distribution function). Finally, the probability of possible trajectory is updated with Equation 3.46.

The next trajectory correction factor is the gravity flow model which is based on the classic gravity model for trip distribution. The basis of this correction factor assumes that a vehicle always arrives from an adjacent zone with a higher volume. At this trajectory correction factor, the initial probability is $W_2^{(i)}$, calculated at the previous trajectory correction factor, travel time consistency. Additionally, $P(y|\hat{x}_3^{(i)})$ is calculated through Equation

3.50 and $q(\hat{x}_3^{(i)}|x_2^{(i)}, y_{t+\Delta t})$ can be defined as a constant probability density in the following equation due to the volume reflecting the assignment relationship of the O-D.

$$P_{flow}^i = \tilde{G} = \frac{F \times F_{car}}{(f(x))^2} \tag{3.50}$$

where:

$\tilde{G}$	= gravitational value
F	= volume at the entrance/exit of the zone
$F_{car} = 1$	= vehicle unit
$f(x)$	= road impedance function and is replaced by detected travel time
P_{flow}^i	= updated probability based on gravity flow model

Finally, the last correction factor is the path-link flow matching model which is based off the link capacity reliability in graph theory. At this trajectory correction factor, the initial probability is $W_3^{(i)}$, calculated at the previous trajectory correction factor, gravity flow model. Additionally, $P(y|\hat{x}_4^{(i)})$ is calculated through Equation 3.51 and $q(\hat{x}_4^{(i)}|x_3^{(i)}, y_{t+\Delta t})$ can be defined as a constant probability density in the following equation due to an existing assignment relationship in path and link flows.

$$P_{path-flow}^i = \frac{\min(F_{N1-N2}^i)}{\sum_{i=1}^{N} \min(F_{N1-N2}^i)} \tag{3.51}$$

where:

F_{N1-N2}^i	= observed volume between node N_1 and N_2
$P_{path-flow}^i$	= updated probability based on path-link flow matching model

Finally, additional details and assumptions are provided by its authors Feng et al. (2015).

3.3.2.5 Algorithm 5 – Origin-Destination Estimation Using Probe Vehicle Trajectory and Link Counts

The fifth and final methodology by X. Yang et al. (2017) estimates O-D matrices through full vehicle trajectories and link counts with a GLS approach. Following the generic formulation, some additional definitions and notations are required. It is noted that the original formulation of this procedure can accommodate data collected in S equal time intervals within an analysis period. For consistency of testing, only one time interval is considered (i.e., $S = 1$) in this research with the following formulation reflecting the single interval.

Let l denote the links installed with sensors where l is a subset of L. Thus, $\bar{f}_l$, represents the observed link count of all vehicles at link location l. The formulation of this methodology assumes two types of vehicles in the network: probe vehicles and regular vehicles. Let $\bar{z}_{ij}$ denote the probe vehicle O-D flow with origin i and destination j for each O-D pair.

The proportion of probe vehicles in the total vehicle population is the O-D probe vehicle penetration ratio, represented as γ_{ij}. The observed number of probe vehicles passing a sensor location is represented as $\bar{h}_l$. Finally, the probe link flow to corresponding link traffic flow is represented as $\bar{\theta}_l$.

This methodology introduces two types of models for O-D estimation: scaled probe O-D matrix as prior matrix (SPP) method, and probe ratios assignment (PRA) model.

Starting with the SPP method, it consists of two steps:

1. The prior matrix $\bar{t}_{ij}$ is estimated by scaling up by the probe O-D flows $\bar{z}_{ij}$ with corresponding O-D probe penetration ratios γ_{ij} estimated by averaging the link probe ratios across the network
2. Solving for O-D flows x_{ij} with a GLS formulation

The direct scaling method utilized by the author was previously proposed by Van Aerde et al. (1993). The estimated O-D probe ratios is computed with the following

$$\gamma = \frac{\sum_{l \in L} \bar{h}_l}{\sum_{l \in L} \bar{f}_l} \tag{3.52}$$

where:

$\bar{h}_l$ = observed link probe flows
$\bar{f}_l$ = observed link flows
γ = common value of γ_{ij}

The average ratio of the total number of probe vehicles to the total number of vehicles observed across the entire network is represented in Equation 3.53. The calculation implies O-D probe ratios are homogeneous amongst all O-D pairs.

$$\bar{t}_{ij} = \frac{\bar{z}_{ij}}{\gamma} \tag{3.53}$$

where:

$\bar{z}_{ij}$ = observed probe O-D flow

Furthermore, the SPP model utilizes the following formulation to adjust the prior O-D matrix

$$\min_{x_{ij}} \sum_{ij} \frac{(x_{ij} - \bar{t}_{ij})^2}{\omega_{ij}^2} + \sum_{l} \frac{\left(f_l - \bar{f}_l\right)^2}{q_l^2} \tag{3.54}$$

$$f_l = \sum_{ij} a_{ij,l} x_{ij} \tag{3.55}$$

where:

x_{ij} = unknown O-D flow
f_l = estimated link flow
ω_{ij}^2 = variance of x_{ij}
q_l^2 = variance of f_l
$a_{ij,l}$ = flow assignment fraction

Equation 3.55 represents the assignment procedure which follows the author's procedure of computing flow assignment fractions directly from probe vehicle trajectories. The computation of $a_{ij,l}$ is first discussed by defining $\bar{z}_{ij,l}$, which is the observed number of probe vehicles traveling between O-D pair with origin i and destination j passed through sensor l. Subsequently, the flow assignment fraction of probe vehicles is expressed as

$$a_{ij,l} = \frac{\bar{z}_{ij,l}}{\bar{z}_{ij}} \tag{3.56}$$

In Equation 3.56, the denominator represents the total number of probe vehicles between O-D pairs ij. Additional assumptions and details are provided by its authors X. Yang et al. (2017).

The PRA model considers the correlation between the O-D probe ratio and the observed link probe ratios, and is expressed in the following equation.

$$\theta_l = P(\gamma_l) = \sum_{t} \sum_{ij} \rho_{ij,l} \gamma_{ij} \tag{3.57}$$

where:

P = assignment matrix of probe vehicle ratios
$\rho_{ij,l}$ = probe ratio assignment fraction

The value of $\rho_{ij,l}$ represents the assignment of probe vehicles ratio between origin i and destination and j which passes through the probe ratio link l. Extending the SPP formulation to include the observed link probe ratios $\bar{\theta}_l$, the PRA model is represented as

$$\min_{x_{ij}} \sum_{ij} \frac{(x_{ij} - \bar{t}_{ij})^2}{\omega_{ij}^2} + \sum_{l} \frac{\left(f_l - \bar{f}_l\right)^2}{q_l^2} + \sum_{l} \frac{\left(\theta_l - \bar{\theta}_l\right)^2}{v_l^2} \tag{3.58}$$

$$f_l = \sum_{ij} a_{ij,l} x_{ij} \tag{3.59}$$

$$\theta_l = \sum_{ij} \rho_{ij,l} \left(\frac{\bar{z}_{ij}}{x_{ij}} \right) \tag{3.60}$$

where:

v_l^2 = variance of θ_l

The computation of $a_{ij,l}$ follow Equation 3.57. The probe ratio assignment is approximated using the observed number of probe vehicles passing through sensor l and the fraction is computed with the following equation.

$$\rho_{ij,l} \approx \frac{\bar{z}_{ij}}{f_l} \tag{3.61}$$

Finally, additional details and assumptions are provided by its authors X. Yang et al. (2017).

3.4 Output Comparison

The final procedure of the framework compares the outcome of the estimated O-D matrix to the ground truth O-D matrix with a MOE. The range of data coverage for partial and full traces follows the same coverage shown in Algorithm 5 by X. Yang et al. (2017). This coverage is a middle ground between the other algorithms which range from 18% to 100%.

Table 3.4: Tested Scenarios

Algorithm (Conditions)	**Data Type**	**Data Amount**
1 (UE, AON)	Link counts	10-100% link count coverage
2 (UE, AON)	Link counts	10-100% link count coverage
3 (UE, AON)	Link counts and partial traces	10-100% link count coverage, 5-30% partial trace coverage
4 (UE, AON)	Partial traces	5-30% partial trace coverage
5 (UE, AON)	Full traces	5-30% full trace coverage

Finally, the chosen MOE is RMSE (Equation 3.62) due to its popularity amongst the authors considered in this research and in the literature review.

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(t_{ij} - \hat{t}_{ij})} \tag{3.62}$$

where:

t_{ij} = estimated value
$\hat{t}_{ij}$ = ground truth value
T = number of estimates

Chapter 4

Results

4.1 Algorithm Results

To evaluate the capabilities of each estimator in this research, they are applied, and results are compared on the basis of RMSE. To start, the findings of each algorithm are presented with an established set of scenarios for consideration. Each scenario consists of 100 runs and the results are then illustrated in box charts where the first quartile, third quartile and the mean RMSE values are shown.

4.1.1 Algorithm 1 Results

Algorithm 1 by Cascetta and Postorino (2001) utilizes link counts as a data source and estimates O-D matrices through a fixed-point formulation. Two conditions, UE and AON, are tested with 10% to 100% link count coverage, in increments of 10%. Thus, a total of 20 scenarios are conducted for this algorithm. The UE and AON results are illustrated in Figure 4.1 and Figure 4.2, respectively.

In both scenarios, RMSE decreases as the percentage of links sampled increases. However, there are notable differences in the performance when comparing Figure 4.1 and Figure 4.2 in terms of RMSE values and the spread of upper and lower extremes. While the UE scenario shows a steady decrease in mean RMSE value as the percentage of link sampled increases, the AON scenario shows stagnation of mean RMSE values from 50% to 90% link sample coverage. For link coverage at 10% and 20%, the upper extreme of the whiskers on the plot are also significantly higher in the AON scenario. Additionally, all

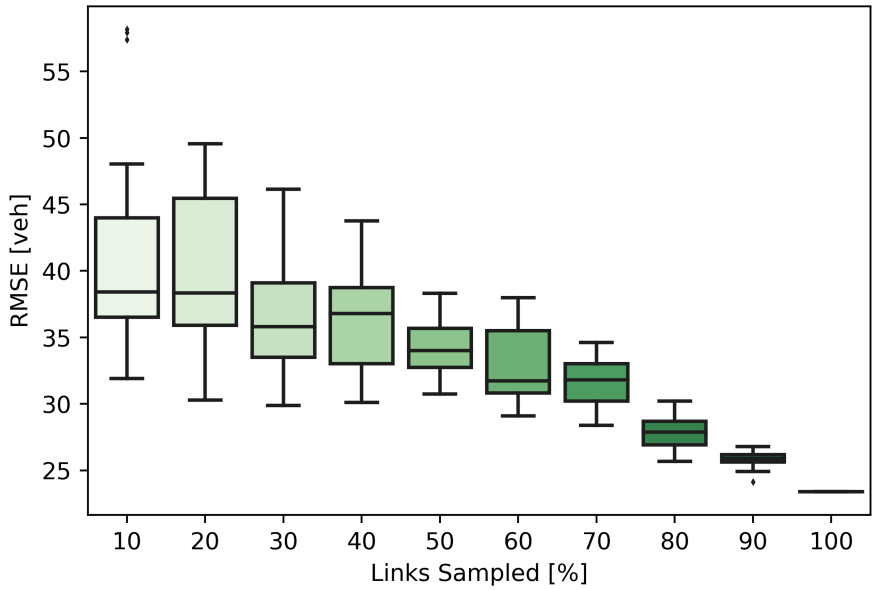

Figure 4.1: Algorithm 1 - UE Results

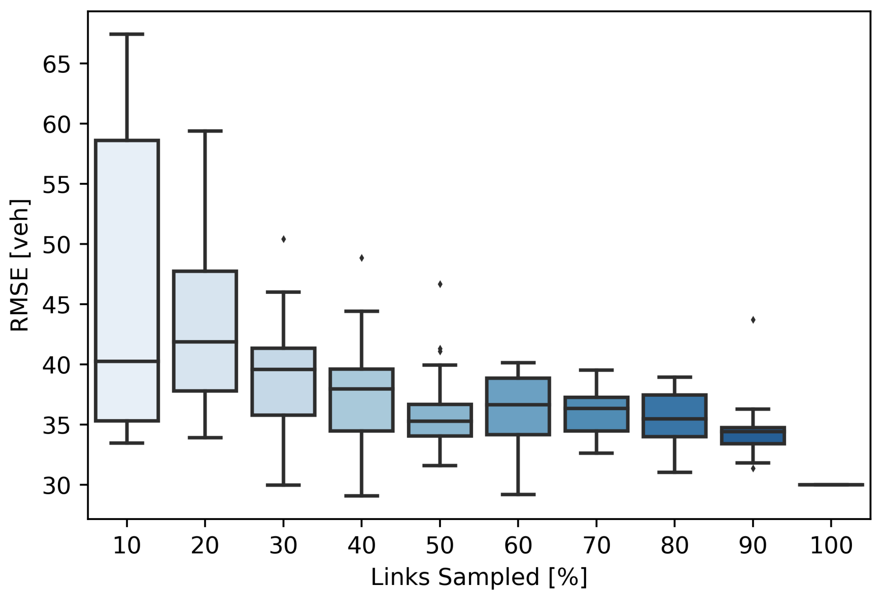

Figure 4.2: Algorithm 1 - AON Results

mean RMSE values for all link coverage percentages are higher in the AON scenario when compared to the UE scenario.

4.1.2 Algorithm 2 Results

Algorithm 2 by Castillo et al. (2008b) utilizes link counts as a data source and estimates O-D matrices through WMV methodology alongside a Bayesian network approach. Two conditions, UE and AON, are tested with 10% to 100% link count coverage (10% increments). Thus, a total of 20 scenarios are conducted for this algorithm. The UE and AON results are illustrated in Figure 4.3 and Figure 4.4, respectively.

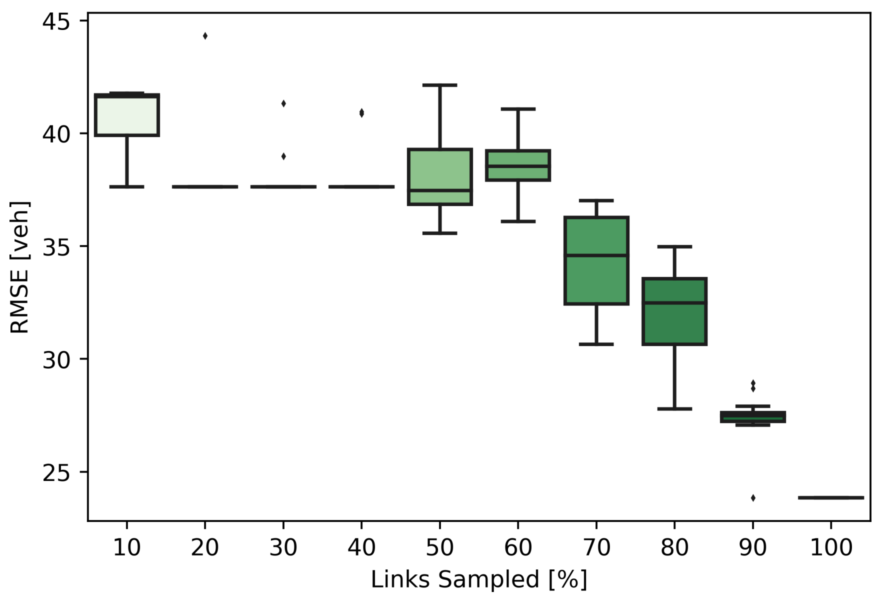

Figure 4.3: Algorithm 2 - UE Results

In both scenarios, RMSE decreases as the percentage of links sampled increases. Figure 4.3 and Figure 4.4 reveal similar results between the UE and AON scenarios. However, the UE scenario for link coverage at 90% and 100% indicate a slightly better estimate. The RMSE remains stagnant with link coverage below 60%. Improvement in estimation results begin when the percentage of links sampled increases above 60%, which can indicate low confidence in the estimator with link count coverage below 60%.

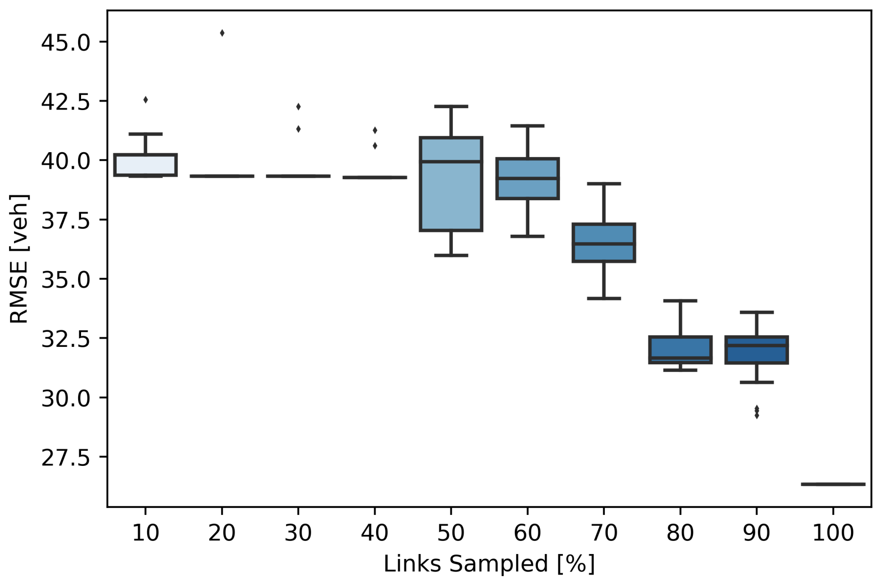

Figure 4.4: Algorithm 2 - AON Results

4.1.3 Algorithm 3 Results

Algorithm 3 by Parry and Hazelton (2012) utilizes link counts and partial traces to estimate O-D matrices using a formal likelihood approach. Two conditions, UE and AON are tested with 10% to 100% link count coverage in 10% increments and 5% to 30% partial traces coverage in 5% increments. Thus, a total of 120 scenarios (i.e., 10 link coverage scenarios x 6 partial trace scenarios × and 2 assignment conditions) are conducted for this algorithm. The UE and AON results formulated with both link counts and partial traces are illustrated in Figure 4.5 and Figure 4.6, respectively.

In the UE and AON scenarios with link count and partial traces data, RMSE generally decreases as the percentage of links sampled increases. Figure 4.5 and Figure 4.6 reveal nearly identical results when compared to each other. The RMSE mostly remains stagnant when the percentage of links sampled is below 60%. Improvement in estimation results begin when the coverage is above 60%, which can indicate low confidence in the estimator with link count coverage below 60%. The most significant improvement in estimation happens at 100% link sampled coverage for all partial trace coverage.

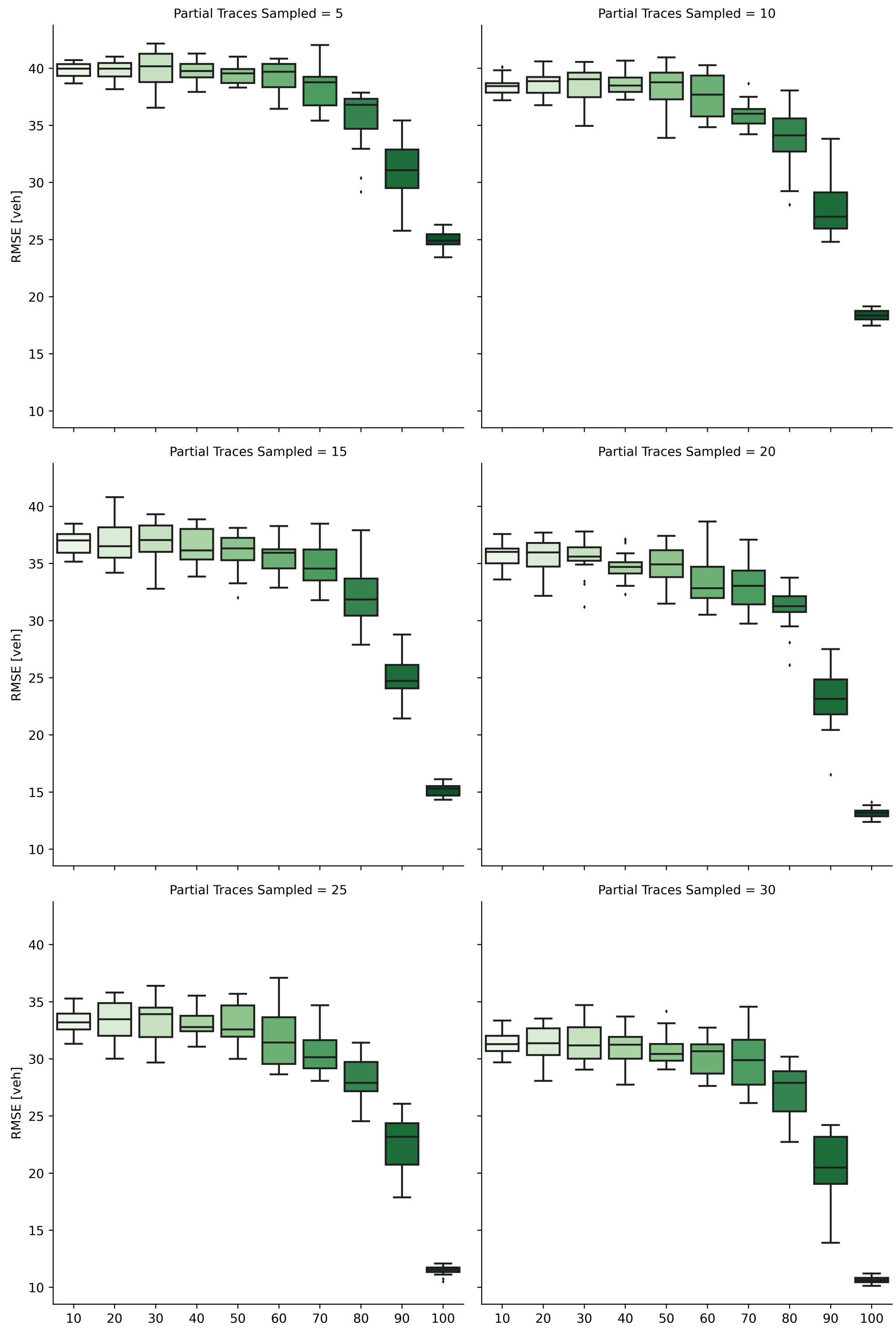

Figure 4.5: Algorithm 3 - UE Results

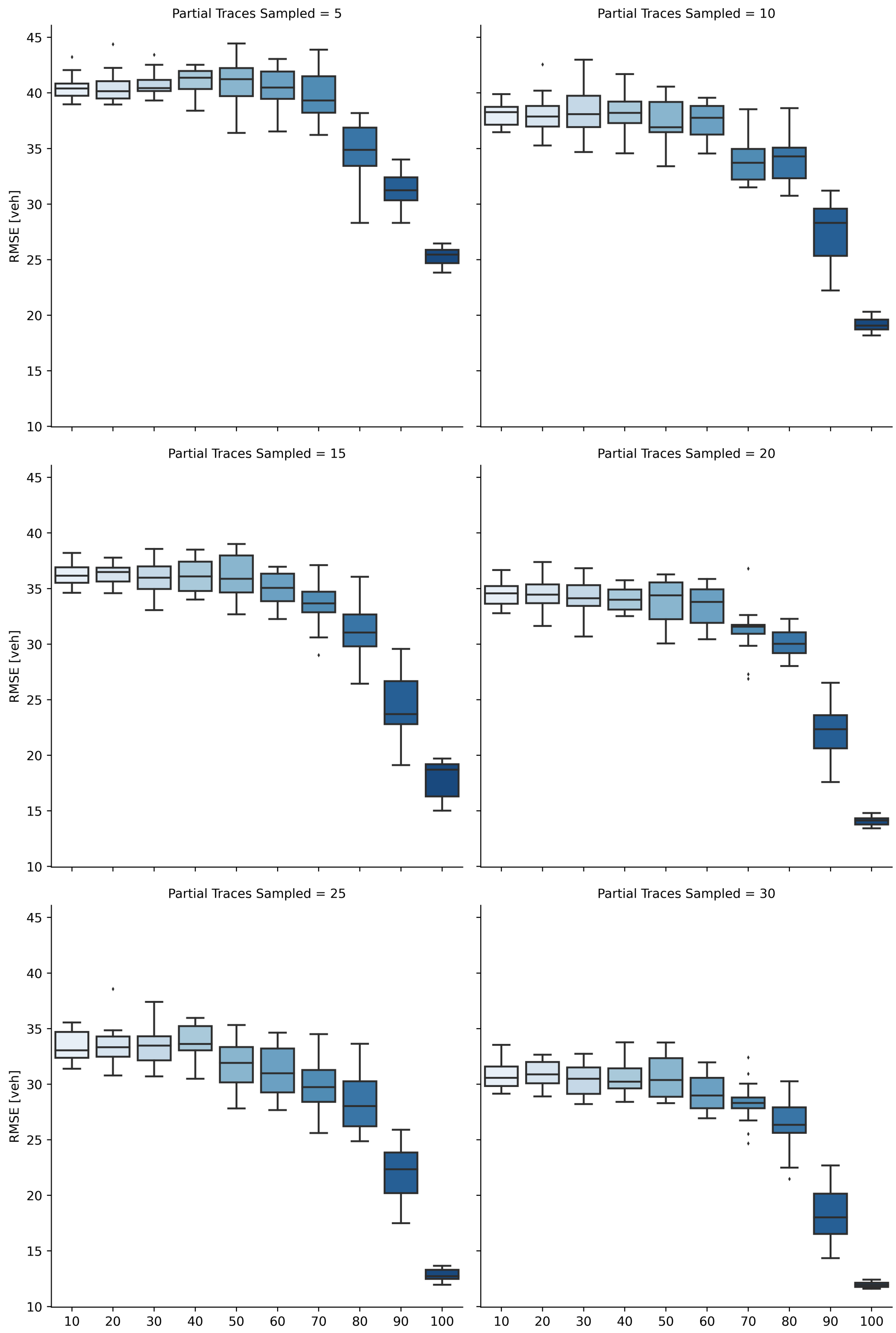

Figure 4.6: Algorithm 3 - AON Results

4.1.4 Algorithm 4 Results

Algorithm 4 by Feng et al. (2015) utilizes partial traces and spot speed data to estimate O-D matrices through a particle filter approach. Two conditions, UE and AON are tested with 10% to 30% partial traces coverage (5% increments). This results in a total of 12 tested scenarios. Unlike the other four algorithms, link count data is not utilized in this method. The UE and AON results are illustrated in Figure 4.7 and Figure 4.8, respectively.

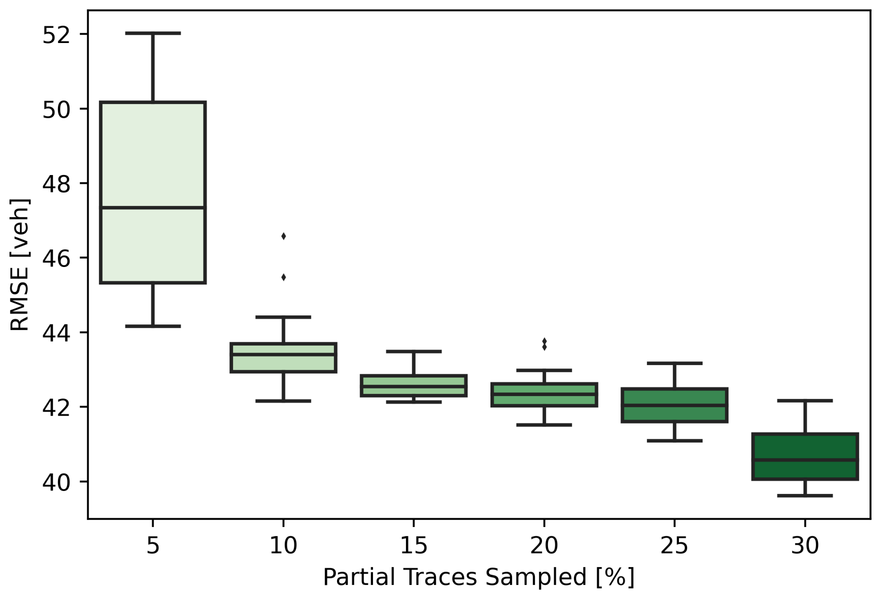

Figure 4.7: Algorithm 4 - UE Results

In both scenarios, RMSE decreases as the percentage of partial traces sampled increases. Figure 4.7 and Figure 4.8 reveal similar results between the UE and AON scenarios with the most drastic improvement occurring at 10% coverage. However, the UE scenarios indicate a slight spread of the upper and lower quartile.

4.1.5 Algorithm 5 Results

Algorithm 5 by X. Yang et al. (2017) utilizes link counts and full traces to estimate O-D matrices using a GLS formulation. There are two models, SPP and PRA, that are proposed by the author of this methodology. Two conditions, UE and AON are tested with 10% to

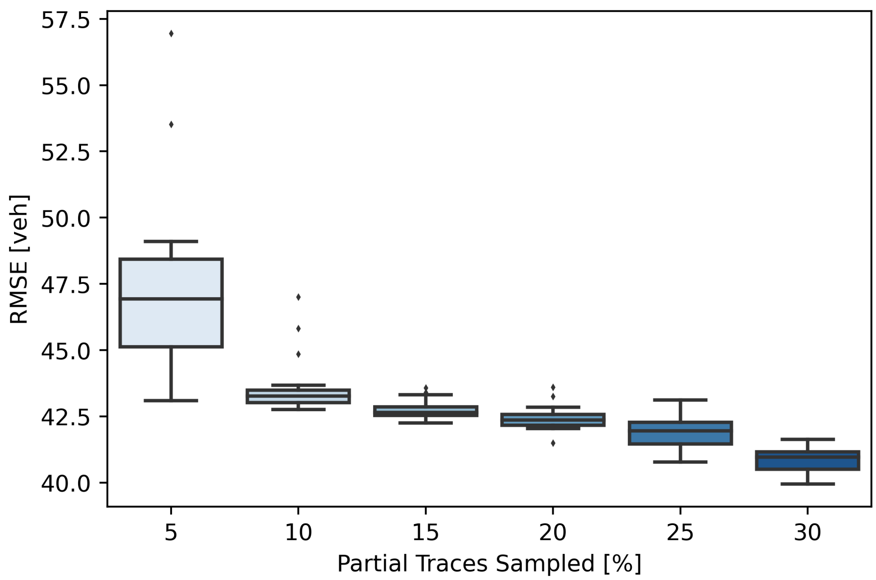

Figure 4.8: Algorithm 4 - AON Results

100% link count coverage and 5% to 30% full traces coverage. This results in a total of 240 tested scenarios, where 120 scenarios are tested with the SPP model and 120 are tested with the PRA model.

First, the UE and AON results for the SPP model are illustrated in Figure 4.9 and Figure 4.10, respectively. Furthermore, the UE and AON results for the PRA model are illustrated in Figure 4.11 and Figure 4.12, respectively.

In the UE and AON scenarios for the SPP model, RMSE generally decreases as the percentage of links sampled increases. Figure 4.9 and Figure 4.10 reveal nearly identical results when compared to each other. The RMSE mostly remains stagnant when the percentage of links sampled is above 60% when 5% of full traces are sampled, 70% when 100% of full traces are sampled and 80% when 15% to 30% of full traces are sampled. The most significant improvement in estimation happens prior to the stagnation and is most noticeable when link sampled from 10% to 50%. This indicates that increasing link coverage prior to stagnation is valuable to improving estimation. Once stagnation occurs, there are diminishing returns to the estimated O-D demand.

In the UE and AON scenarios for the PRA model, RMSE generally decreases as the percentage of links sampled increases. Figure 4.11 and Figure 4.12 reveal nearly identical results when compared to each other. The RMSE mostly remains stagnant when the percentage of links sampled is above 50% when 5% of full traces are sampled, 70% when 100% of full traces are sampled and 80% when 15% to 30% of full traces are sampled. Similar to the SPP model, the most significant improvement in estimation happens prior to the stagnation and is most noticeable when link sampled from 10% to 50%. Once again, this indicates that increasing link coverage prior to stagnation is valuable to improving estimation but there are diminishing returns to the estimated O-D demand once stagnation occurs.

When comparing the results of the SPP model (i.e., Figure 4.9 and Figure 4.10) and PRA model (i.e., Figure 4.11 and Figure 4.12), results are nearly identical. At high (25% and 30%) full traces coverage and high (90% and 100%) link sample coverage, the PRA model has a slightly better estimation.

4.2 Summary of Algorithms

To close, an illustration of the performance of all algorithms is summarized for the AON and UE scenarios in Figure 4.12 and Figure 4.14, respectively. For brevity, the algorithms that have partial or full traces as data source only illustrate the lowest trace coverage (5%) and highest trace coverage (30%).

For both AON and UE scenarios, algorithms that utilize only link counts or link counts and partial traces have a concave decreasing relationship with respect to increasing collected sample data. In contrast, algorithms that utilize link counts and full traces have a convex decreasing relationship with respect to increasing collected sample data. For link counts and partial traces, this indicates that the percentage of collected data is pertinent to the confidence of the estimator. Full traces, on the other hand, indicate that any amount of sampled data improves estimator performance. However, at certain coverage of link samples, there is diminishing value in increasing link data collection (e.g., generally after 50% for 5% full traces, and 80% for 30% traces).

For most cases, low percentages of sampled links reveal similar performance resulting between a RMSE value of 37 and 45 in the tested network. The exception to this is Algorithm 3 with 30% of partial traces with a RMSE value closer to 30. In addition, most cases of 100% of sampled links revel performances resulting in a RMSE value between 24 and 34. Again, Algorithm 3 with 30% of partial traces is an exception to this with a RMSE value closer to 10.

Usually, only link counts cannot obtain the same quality of O-D estimates when compared to the ones obtained through full traces, even at low full trace percentages (i.e., 5%).

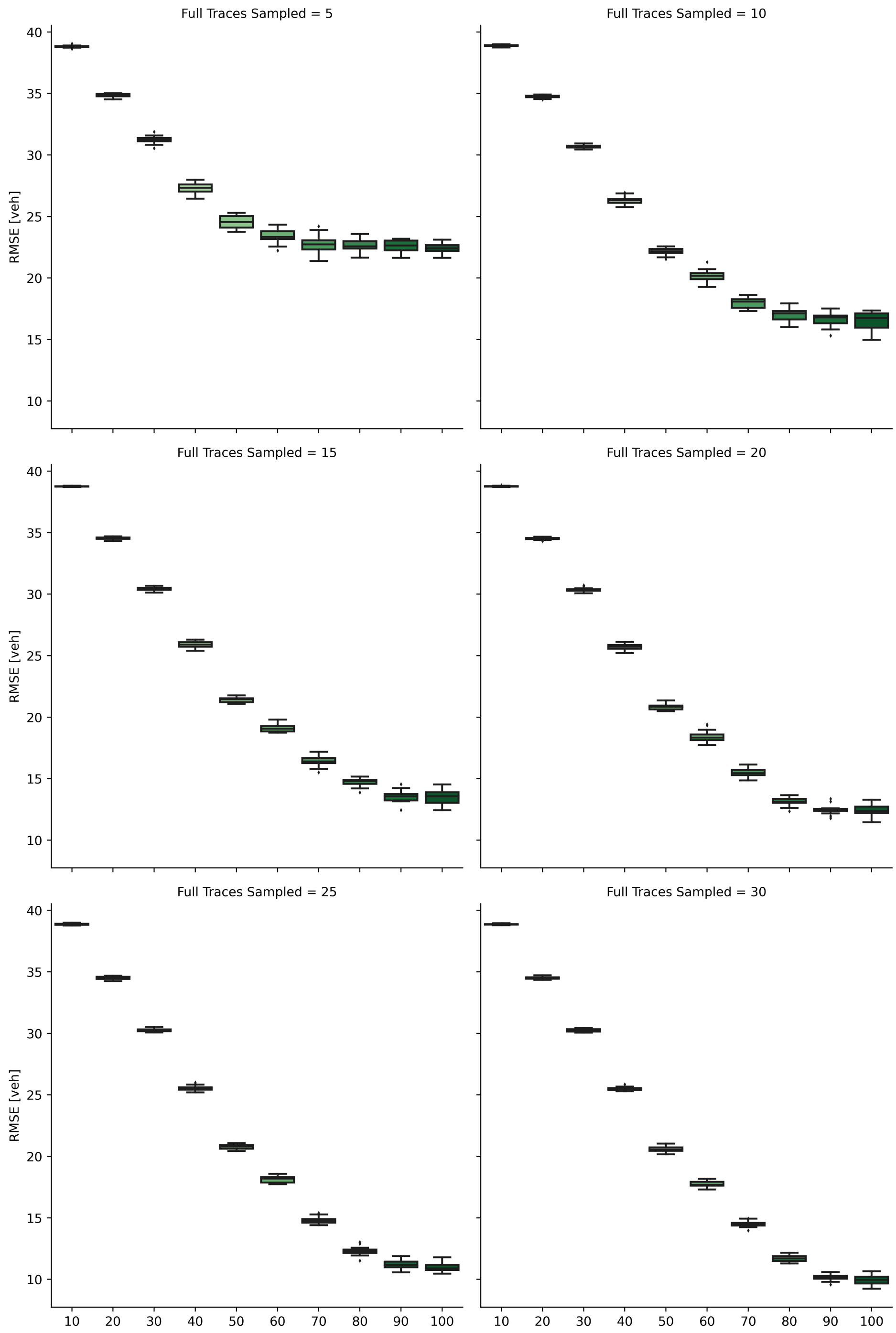

Figure 4.9: Algorithm 5 (SPP) - UE Results

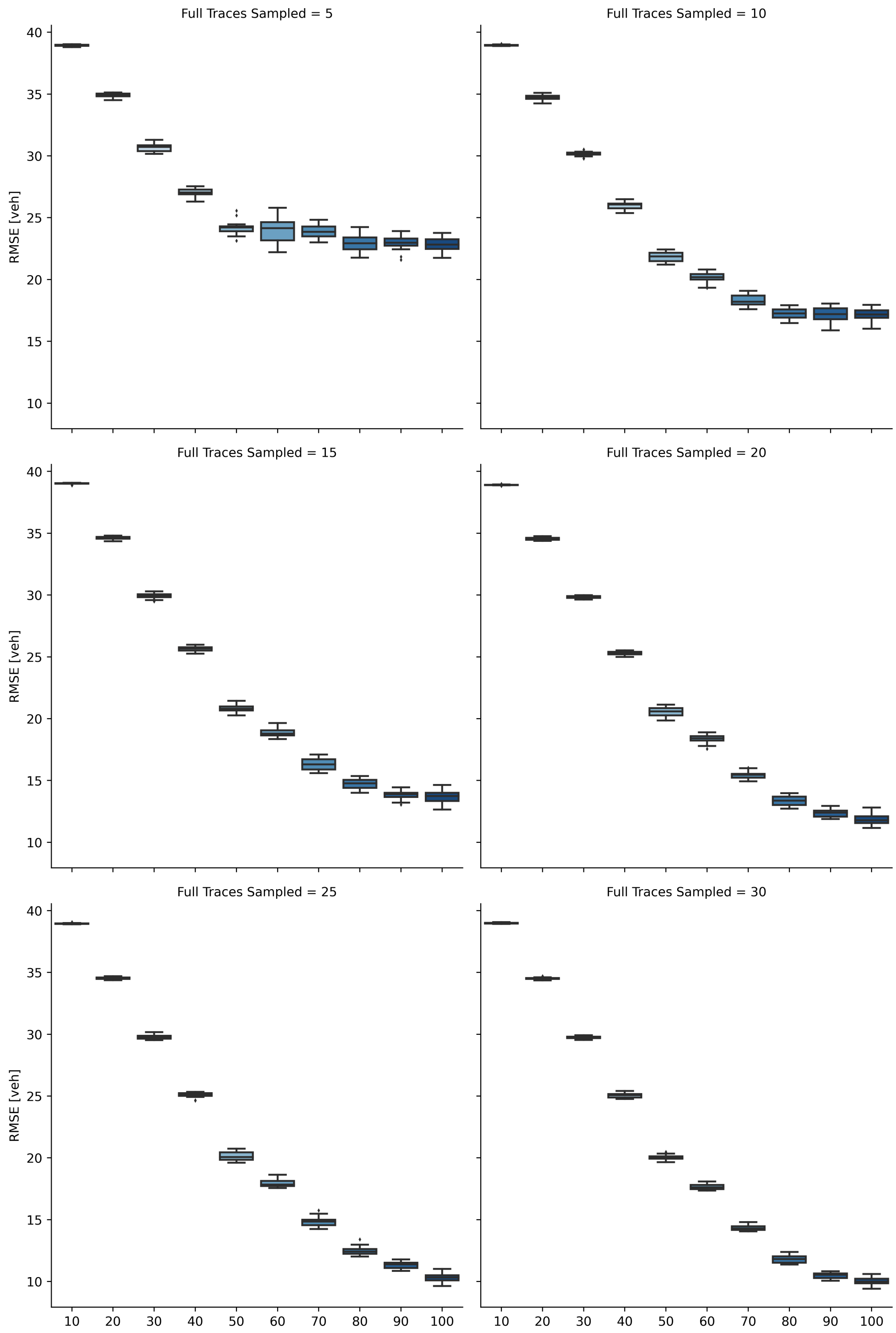

Figure 4.10: Algorithm 5 (SPP) - AON Results

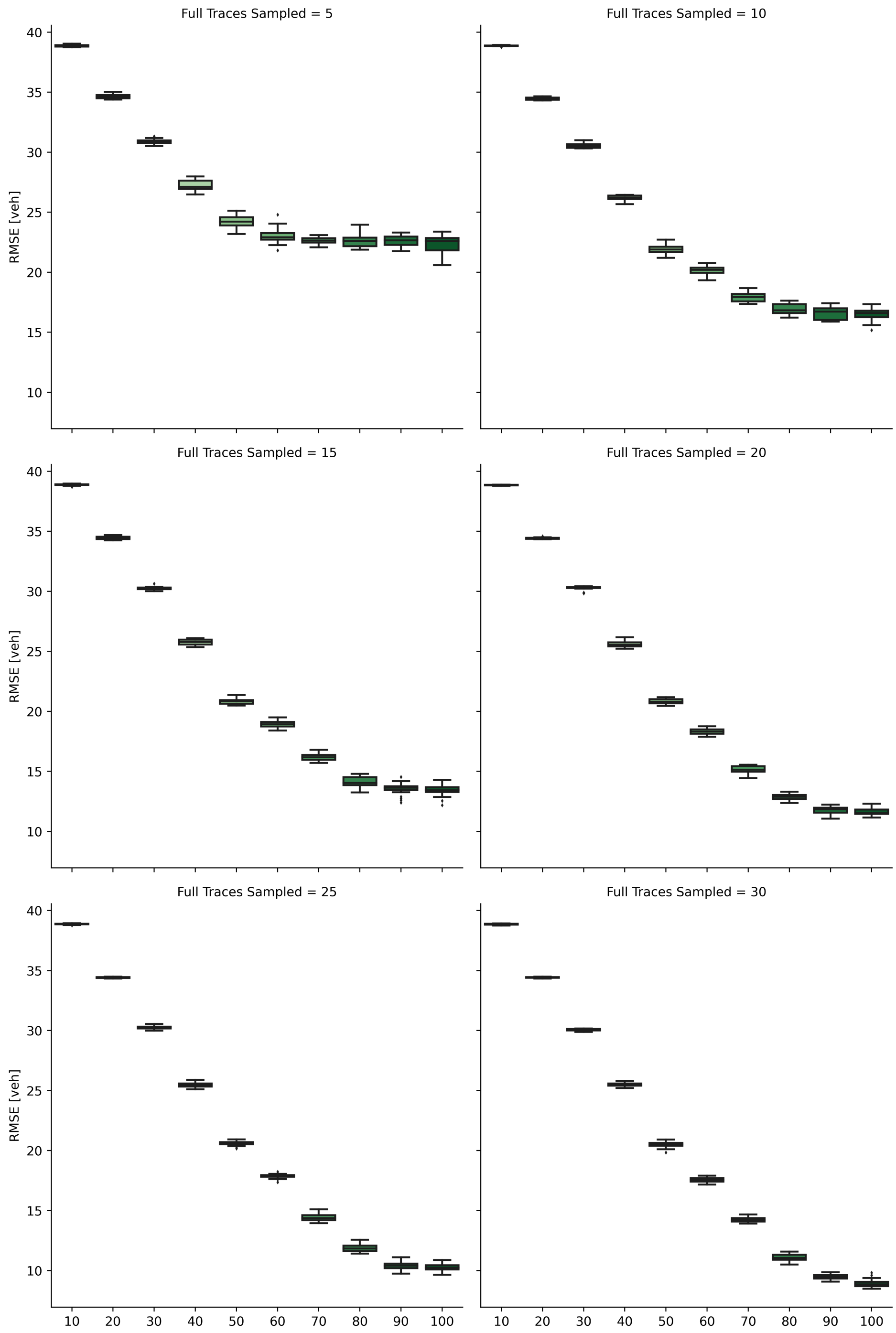

Figure 4.11: Algorithm 5 (PRA) - AON Results

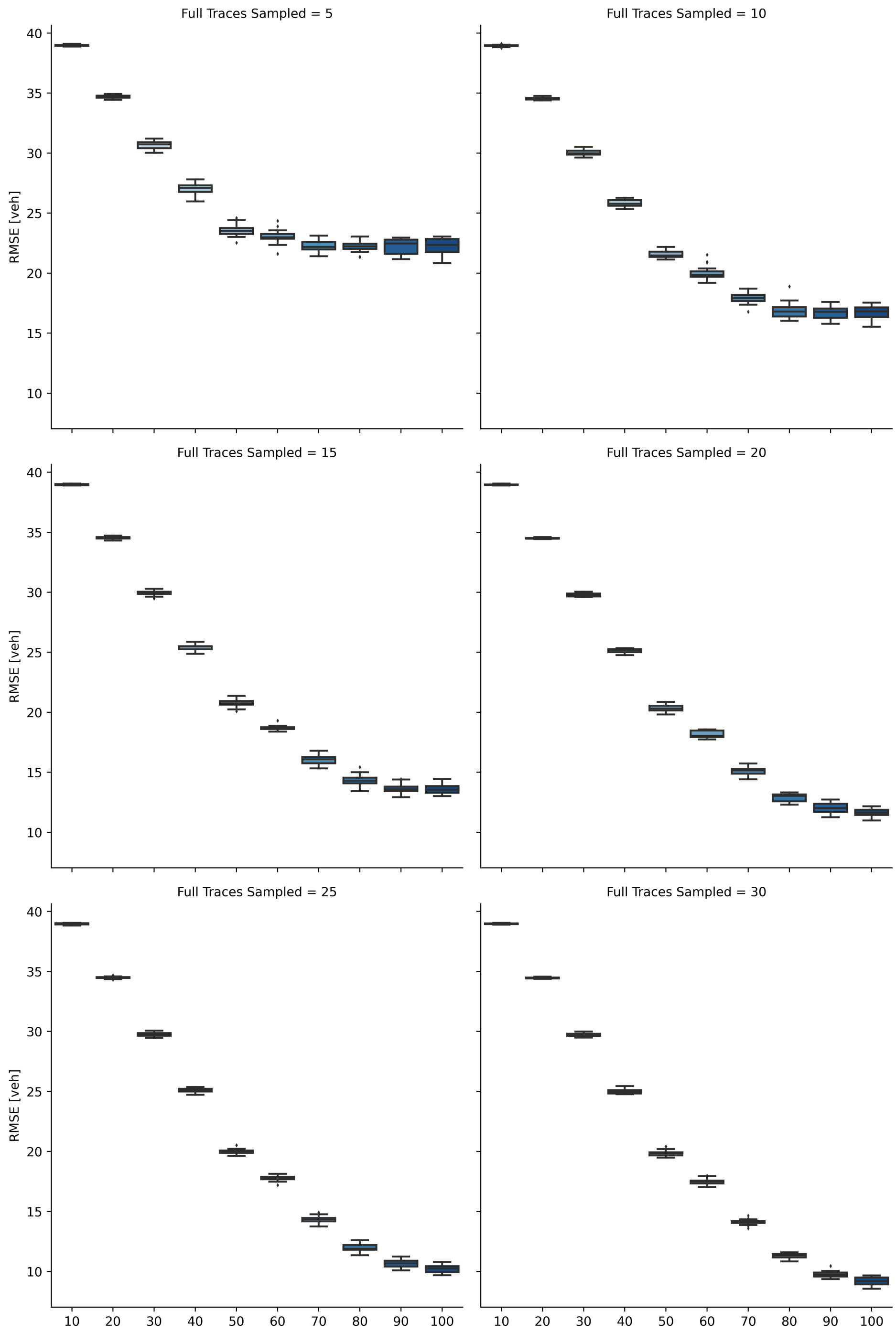

Figure 4.12: Algorithm 5 (PRA) - AON Results

The exception to this is at 100% links sampled where the UE scenario indicates similar performances of Algorithm 1, 2 and 3 (with 5% of partial traces). In some cases, link counts can compete with low percentages (5%) of full traces.

While 100% link coverage with 30% of partial traces reveal the best performance, obtaining full trace data or high levels of link count data may not feasible. Thus, the following data coverage are shown to be most valuable in these specific situations:

- when link coverage is below 30%, it is more valuable to collect 30% of partial traces rather than increasing link coverage or full trace coverage
- when link coverage at 100% with 30% of partial traces reveal a similar performance to link coverage at 100% of 30% of full traces.

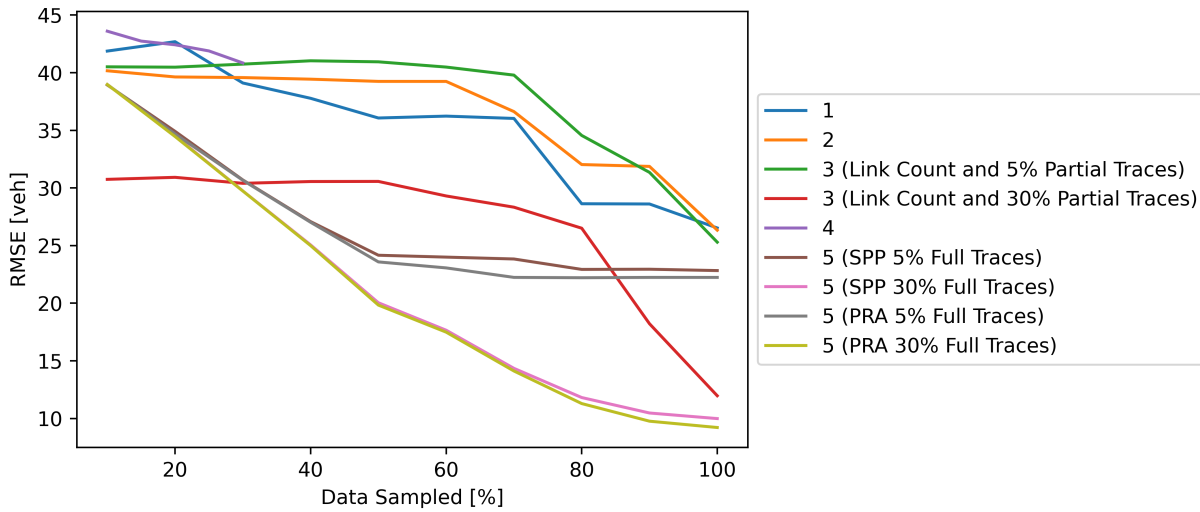

Figure 4.13: Performance Summary (AON)

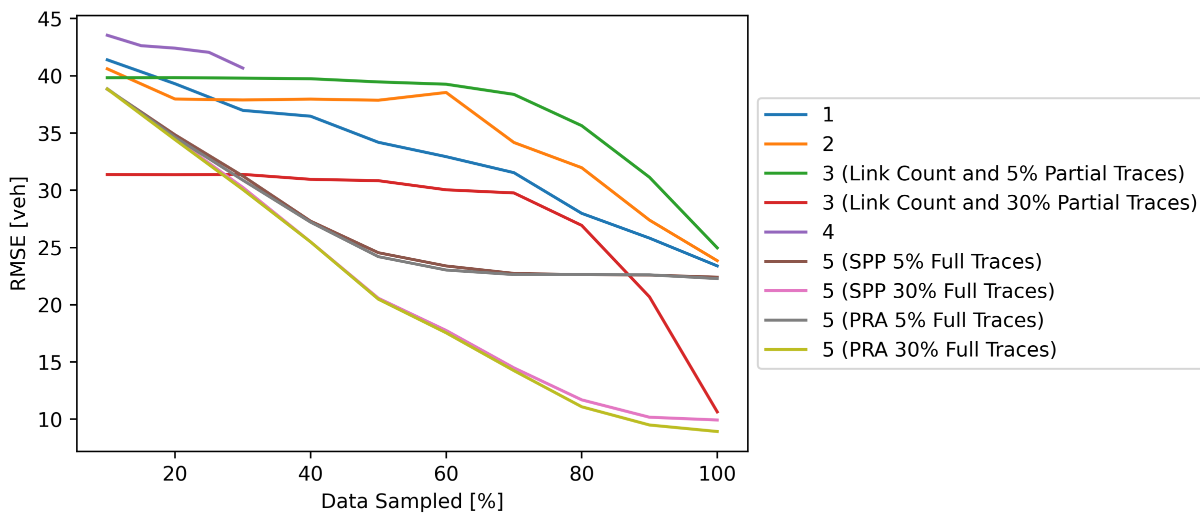

Figure 4.14: Performance Summary (UE)

Chapter 5

Discussion and Conclusions

This thesis presents a framework to prepare and test static O-D estimation algorithms on a common testbed with an established baseline condition. The framework is implemented in a simulation environment consisting of a Vissim model of the Sioux Falls network as a case study area. The network characteristics of the model environment consists of 24 nodes, 552 O-D pairs and 2,069 paths. A total of five algorithms were tested with four data types and two traffic assignment conditions. The performances of the methodologies were repeated 100 times for each set of scenarios (412 total scenarios) for a total of 41,200 O-D matrices estimated amongst the five algorithms.

This research does not conclude to a "best" approach for O-D estimation; rather it provides support in testing the effects of data type, data amount, and a variety of scenarios and conditions. To close, limitations and future recommendations are provided at the end of this chapter based on the findings of this research. Some future recommendations are provided through the lens of the challenges met within this research.

5.1 Major Findings

A brief overview of some general findings of the algorithms are presented. Subsequently, specific details of each algorithm are reviewed in the following subsections.

5.1.1 Algorithm 1

While the results for this algorithm are promising within the testbed framework, there are specific details utilized in the framework that are the basis for its success. External knowledge is pertinent to the success of this methodology since the relationship between link counts and O-D estimates hold an indeterminate relationship. Since paths are enumerated in the methodology, the link costs and spread parameters are key inputs for an accurate assignment matrix and thus, also a reliable O-D estimate. This approach assumes an equilibrium-based assignment. Although this assumption may be favourable to reflect traveler route choice behaviour under congested conditions, equilibrium may not be reached in some network cases. The results of this method reflect this limitation as a better estimate is shown in the UE scenario in comparison to the AON scenario. It is noted that the calibration of spread parameters can be adjusted on a case-by-case basis to better reflect each network and scenario. However, the performance would still be dependent on the accuracies of these parameters.

Since the framework utilized in this research is a synthetic testbed, the BPR cost function is easily identified through Vissim, which then leads to an extremely accurate path cost. Consistent with the findings by Cascetta and Postorino (2001), the utilization of a fixed-point approach within this method shows to be capable of O-D estimation within a congested scenario. Finally, with this external knowledge, the estimated assignment matrix is especially close to the actual assignment matrix. In addition to path enumeration, the prior O-D matrix also holds imperative information to a successful O-D estimate. Since the extracted O-D matrix found is theoretically representing a prior O-D obtained by a household travel survey, the prior O-D is represented as the same dataset as a 5% full trace dataset in this research. Unsurprisingly, since the prior O-D is of high quality, a lower RMSE result is already established even without any implementation of link count data. Limitations of low-quality prior O-Ds are not reflected in this research.

5.1.2 Algorithm 2

The authors of Algorithm 2, Castillo et al. (2008b), have also tested their methodology on the Sioux Falls network, although their results are not reported. Their findings are said to show high precision in O-D estimates even at low link counts. Some of the findings in this research are consistent with Castillo et al. (2008b), however, results only show confidence in the methodology at higher link count coverage (60%+). It is noted that Castillo et al. (2008b) provided a two-step framework in their paper: the first step covering an optimal subset of links to be observed, followed by the second step where O-D estimates are conducted utilizing the optimal links identified in the first step. The first step utilizes a prior O-D matrix as a basis for finding a set of optimal link counts to collect. This approach also relies on a high-quality prior O-D for necessary knowledge in the O-D estimation step. Since the scope of this research is limited to the O-D estimation procedure, the first step for optimal link counts is excluded. The difference in results indicates that a specific optimal link set is important to the success of estimating O-D flows with low link coverage.

Consistent with the results on a testbed network (4 O-D pairs, 13 nodes, and 19 links) by Castillo et al. (2008b), the quality of results improve as the number of observed links increases. Although this methodology only utilizes prior O-D flows and observed link counts as input data, the Bayesian network contains two sources of information: the joint normal distribution of links and O-D flows, and the observed link flows. Therefore, with a prior O-D, an estimate can be provided without additional link count data.

5.1.3 Algorithm 3

Objective functions in this method are similar to the GLS estimation of O-D matrices where the routing data acts like the target matrix. Unlike algorithms requiring a prior O-D, this algorithm's performance is not reliant on a high-quality target. It is important to note that utilizing both the partial traces and link counts provide important information in this estimation method and is crucial to its success. While one might think that the information that partial traces hold has precedence when compared to link count data, the model results reflect that even at high partial trace coverage (30%), the best performance only exists with the presence of high link count coverage. In fact, at high partial trace coverage and low link sampling rates, results indicate low confidence in the estimator's performance.

Since this algorithm assumes a restricted set of routes, potential biases of the route set may exist. While Parry and Hazelton (2012) restricts O-D pairs to have no more than six or seven routes, this research does not limit routes to six or seven O-D pairs. However, establishing limited routes reveals similar limitations to restriction established by Parry and Hazelton (2012). Establishing an unrestricted set of routes would not be plausible for this estimation method as Parry and Hazelton (2012) state, networks with tens of thousands of routes, or greater, would present significant computational burden.

Lastly, the penetration rates of the partial traces are critical factor to the success of this estimator. The assumption of homogeneity of traces amongst routes, presented Equation 3.41, is a limiting factor since constant penetration rate cannot be guaranteed in data collection. Similar to the results presented by Parry and Hazelton (2012), this research produced O-D estimates that are not greatly affected by the assumption of homogeneous penetration rates. Although the variations of the probability of tracking vehicles exists in our testbed, this methodology shows quality O-D estimates based on the results.

5.1.4 Algorithm 4

The methodology of Algorithm 4 only utilizes partial trace as a data source for O-D estimation. Hence, the information that the partial traces provide is pertinent to the success of this method.

The limitations of this procedure at low partial traces coverage are susceptible to errors potentially due to a lack of scalability within the methodology. Extremely confident results are reported by Feng et al. (2015) when the trace coverage is greater than 60% since a 60% sample rate covers all vehicles within their network. The large amount of information obtained at high partial trace coverage is directly related to the confidence in performance of this algorithm. Since this research shows tests with coverage from 5% to 30%, a confident estimate has not been established. Additional effort in increasing information through different avenues are required for estimating at lower partial trace coverage for this estimator.

5.1.5 Algorithm 5

Unsurprisingly, the information of full traces proved to show the best O-D estimates. The success of this algorithm shows full traces to have information for quality estimates without external knowledge. Like Algorithm 3, link count data provide valuable important information in this estimation method based on model results. The inclusion of link data notably increases the quality of the estimates, especially since the best performances are with a presence of high link count coverage. In fact, at high partial trace coverage and low link sampling rates, results indicate low confidence in the estimator's performance.

Within this testbed, the penetration of full traces is conducted heterogeneously. X. Yang et al. (2017) reported findings show a better performance in the PRA model for heterogeneous penetration rates. Although not as significant as the results provided by X. Yang et al. (2017), the findings of this research also support that the PRA model performs slightly better with the heterogeneous penetration rates of this testbed. The slight improvement in performance could be explained by how the variation of the heterogeneity of the penetration in this research testbed is not as significant as the ones by X. Yang et al. (2017).

The original network utilizes a range of full traces from 5% to 30% with a fixed constant link count coverage of 51%. Their findings reveal unreliable estimates at the 5% trace coverage. Further testing in this research reveals exceeding the 50% link count coverage would not significantly improve the estimates at low full trace coverage (5%).

5.1.6 Prior Knowledge

Due to the observability issue in static O-D estimation based solely on link counts, some algorithms in this study rely on a prior O-D matrix. While each algorithm depends on prior knowledge differently, the quality of this external knowledge reveals differences in performance amongst algorithms. Though it is rationale to think that more prior data is better, it is not equivalent to better estimation quality since some priors create a less accurate estimation due to variable reasons (e.g., outdated estimates).

Additionally, the algorithms that require a path-incidence matrix as part of their methodology risk eliminating routes that carry flow. While this research does not consider paths that backtrack, some specific networks could contain a reasonable amount of backtracking as plausible routes. By utilizing a constrained set of routes, these algorithms introduce some slight biases into their estimates. Furthermore, for larger and more complex networks, establishing a restricted set of routes poses its own challenges. Alternatively, an entire set of unrestricted routes can be established through search algorithms, but many estimation methods may have capacity limitations to estimate O-D flows on an unrestricted set of paths.

5.1.7 Data Type and Data Amount (Sample Size)

Results of this research provide insight into the value of information gained from each data type. When utilizing only link count data, critical results occur at coverages of 80% or higher. These algorithms (i.e., Algorithm 1 and Algorithm 2) reveal the most confidence and substantial improvement at this coverage. At coverages below 80%, the results of all algorithms utilizing only link counts do not show substantial additional value.

The addition of partial traces to link count data provides increasing value of information when the coverage of partial traces is 15% or higher. At 15% or lower, the addition of more partial traces shows no substantial value when compared to only having link counts.

When adding full traces to link count data, all full trace coverage rates provide additional value, especially when link count coverage is between 20% and 90%. At low full trace coverages (5% and 10%), the most impact from link counts is had by increasing coverage up to 50%; additional link coverage beyond 50% does not show substantial value. However, as full trace coverage increases above 10%, diminishing returns begin to occur when link count coverage surpass 80%. Thus, there is added value of information when increasing both full traces and link count coverage (up until 80% in these experiments).

At high sample sizes, the best performing algorithms in the tested conditions are as follows:

- high link coverage and high partial trace coverage (Algorithm 3), and
- high link coverage and medium-to-high full trace coverage (Algorithm 5).

At these large sample sizes, the RMSE value decreases by more than 50% when compared to other scenarios. While these high sample sizes show the best performances, the data collection required by these scenarios may not be feasible in practice.

Therefore, for medium data coverage, the best performing options with similar results are as follows:

- medium link coverage and medium partial trace coverage (Algorithm 3), and
- medium link coverage and low full trace coverage (Algorithm 5).

At lower data coverages, the implementation of a full trace (also in a low percentage) appears to have the best value of information. Thus, Algorithm 5 would likely perform the best with a low sample size. This is also apparent in the results of Algorithm 5, where the plots (Figure 4.9 - Figure 4.12) illustrate a convex shape of the results of this algorithm, demonstrating the steepest slopes of the function occur at the lower data sampled.

5.2 Limitations and Future Research

Based on the research framework, there are some limitations as it pertains to real world applications. The findings of this research are confined to the scenarios that are tested. Technological, measurement, and human errors are not uncommon in data collection and would affect the overall performances of O-D estimation algorithms. This research represents a simplification of reality where modifications and simplifications make it possible to model in a simulated environment. Many additions and enhancements can be conducted to better represent scenarios unique to specific models

The presented testbed framework and demonstrated results supports and enhances the testing of future general O-D estimation methods and extensions of existing algorithms. Additionally, with the common testbed, the value of data type and data amount can be tested as new emerging technologies are being utilized. The ability to gain insight to the value of data for suitable applications support allocating resources in evidence-based decision making.

References

Abdelghany, K., Hassan, A., Alnawaiseh, A., & Hashemi, H. (2015). Flow-based and density-based time-dependent demand estimation for congested urban transportation networks. *Transportation Research Record*, *2498*, 27–36. doi: 10.3141/2498-04

Alexander, L., Jiang, S., Murga, M., & González, M. (2015). Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, *58*, 240–250. doi: 10.1016/j.trc.2015.02.018

Antoniou, C., Barceló, J., Breen, M., Bullejos, M., Casas, J., Cipriani, E., ... van Lint, H. (2016). Towards a generic benchmarking platform for origin-destination flows estimation/updating algorithms: Design, demonstration and validation. *Transportation Research Part C: Emerging Technologies*, *66*, 79–98. doi: 10.1016/j.trc.2015.08.009

Ashok, K., & Ben-Akiva, M. (2000). Alternative approaches for real-time estimation and prediction of time-dependent Origin-Destination flows. *Transportation Science*, *34*(1), 21–36. doi: 10.1287/trsc.34.1.21.12282

Ashok, K., & Ben-Akiva, M. (2002). Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows. *Transportation Science*, *36*(2), 184–198. doi: 10.1287/trsc.36.2.184.563

Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M., & Puchinger, J. (2019). Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, *101*, 254–275. doi: 10.1016/j.trc.2019.02.013

Baek, S., Kim, H., & Lim, Y. (2004). Multiple-vehicle origin-destination matrix estimation from traffic counts using genetic algorithm. *Journal of Transportation Engineering*, *130*(3), 339–347. doi: 10.1061/(ASCE)0733-947X(2004)130:3(339)

Baek, S., Lim, Y., Rhee, S., & Choi, K. (2010). Method for estimating population OD matrix based on probe vehicles. *KSCE Journal of Civil Engineering*, *14*(2), 231–235. doi: 10.1007/s12205-010-0231-4

Barceló, J., Montero, L., Bullejos, M., Serch, O., & Carmona, C. (2013). A kalman filter approach for exploiting bluetooth traffic data when estimating time-dependent od matrices. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *17*(2), 123–141. doi: 10.1080/15472450.2013.764793

Barry, J., Newhouser, R., Rahbee, A., & Sayeda, S. (2002). Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record*, *1817*, 183–187. doi: 10.3141/1817-24

Bauer, D., Richter, G., Asamer, J., Heilmann, B., Lenz, G., & Kölbl, R. (2018). Quasi-Dynamic Estimation of OD Flows from Traffic Counts Without Prior OD Matrix. *IEEE Transactions on Intelligent Transportation Systems*, *19*(6), 2025–2034. doi: 10.1109/TITS.2017.2741528

Behara, K., Bhaskar, A., & Chung, E. (2020). A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies*, *111*, 513–530. doi: 10.1016/j.trc.2020.01.005

Behara, K., Bhaskar, A., & Chung, E. (2021). Single-level approach to estimate origin-destination matrix: exploiting turning proportions and partial OD flows. *Transportation Letters*. doi: 10.1080/19427867.2021.1932182

Bell, M. (1983). The Estimation of an Origin-Destination Matrix from Traffic Counts. *Transportation Science*, *17*(2), 198–217. doi: 10.1287/trsc.17.2.198

Bera, S., & Rao, K. (2011). Estimation of origin-destination matrix from traffic counts: The state of the art. *European Transport - Trasporti Europei*(49), 3–23.

Bianco, L., Confessore, G., & Reverberi, P. (2001). A network based model for traffic sensor location with implications on O/D matrix estimates. *Transportation Science*, *35*(1), 50–60. doi: 10.1287/trsc.35.1.50.10140

Bierlaire, M. (2002). The total demand scale: A new measure of quality for static and dynamic origin-destination trip tables. *Transportation Research Part B: Methodological*, *36*(9), 837–850. doi: 10.1016/S0191-2615(01)00036-4

Bierlaire, M., & Crittin, F. (2004). An efficient algorithm for real-time estimation and prediction of dynamic OD tables. *Operations Research*, *52*(1), 116–127. doi: 10.1287/opre.1030.0071

Bonnel, P., Fekih, M., & Smoreda, Z. (2018). Origin-Destination estimation using mobile network probe data. In *Transportation research procedia* (Vol. 32, pp. 69–81). doi: 10.1016/j.trpro.2018.10.013

Caceres, N., Romero, L., & Benitez, F. (2020). Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys. *Transportmetrica A: Transport Science*, *16*(3), 574–601. doi: 10.1080/23249935.2020.1720857

Caggiani, L., Ottomanelli, M., & Sassanelli, D. (2013). A fixed point approach to origin-destination matrices estimation using uncertain data and fuzzy programming on congested networks. *Transportation Research Part C: Emerging Technologies*, *28*, 130–141. doi: 10.1016/j.trc.2010.12.005

Carrese, S., Cipriani, E., Mannini, L., & Nigro, M. (2017). Dynamic demand estimation and prediction for traffic urban networks adopting new data sources. *Transportation Research Part C: Emerging Technologies*, *81*, 83–98. doi: 10.1016/j.trc.2017.05.013

Carvalho, L. (2014). A bayesian statistical approach for inference on static origin-destination matrices in transportation studies. *Technometrics*, *56*(2), 225–237. doi: 10.1080/00401706.2013.826144

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. *Transportation Research Part B*, *18*(4-5), 289–299. doi: 10.1016/0191-2615(84)90012-2

Cascetta, E., Inaudi, D., & Marquis, G. (1993). Dynamic estimators of origin-destination matrices using traffic counts. *Transportation Science*, *27*(4), 363–373. doi: 10.1287/trsc.27.4.363

Cascetta, E., & Nguyen, S. (1988). A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B*, *22*(6), 437–455. doi: 10.1016/0191-2615(88)90024-0

Cascetta, E., Papola, A., Marzano, V., Simonelli, F., & Vitiello, I. (2013). Quasi-dynamic estimation of o-d flows from traffic counts: Formulation, statistical validation and performance analysis on real data. *Transportation Research Part B: Methodological*, *55*, 171–187. doi: 10.1016/j.trb.2013.06.007

Cascetta, E., & Postorino, M. (2001). Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks. *Transportation Science*, *35*(2), 134–147. doi: 10.1287/trsc.35.2.134.10138

Castillo, E., Gutierrez, J. M., & Hadi, A. S. (1997). *Expert Systems and Probabilistic Network Models*. Springer New York. doi: https://doi.org/10.1007/978-1-4612-2270-5

Castillo, E., Jiménez, P., Menéndez, J., & Nogal, M. (2013). A Bayesian method for estimating traffic flows based on plate scanning. *Transportation*, *40*(1), 173–201. doi: 10.1007/s11116-012-9443-4

Castillo, E., Menéndez, J., & Sánchez-Cambronero, S. (2008b). Traffic estimation and optimal counting location without path enumeration using bayesian networks. *Computer-Aided Civil and Infrastructure Engineering*, *23*(3), 189–207. doi: 10.1111/j.1467-8667.2008.00526.x

Castillo, E., Menéndez, J. M., & Sánchez-Cambronero, S. (2008a). Predicting traffic flow using Bayesian networks. *Transportation Research Part B: Methodological*, *42*(5), 482–509. doi: https://doi.org/10.1016/j.trb.2007.10.003

Chen, A., Chootinan, P., & Recker, W. (2005). Examining the quality of synthetic origin-destination trip table estimated by path flow estimator. *Journal of Transportation Engineering*, *131*(7), 506–513. doi: 10.1061/(ASCE)0733-947X(2005)131:7(506)

Chen, A., Chootinan, P., & Recker, W. (2009). Norm approximation method for handling traffic count inconsistencies in path flow estimator. *Transportation Research Part B: Methodological*, *43*(8-9), 852–872. doi: 10.1016/j.trb.2009.02.007

Chen, D., Sun, J., & Li, K. (2011). Dynamic OD estimation under automated vehicle identification environment. *Key Engineering Materials*, *467-469*, 835–840. doi: 10.4028/www.scientific.net/KEM.467-469.835

Cheng, L., Zhu, S., Chu, Z., & Cheng, J. (2014). A Bayesian network model for origin-destination matrices estimation using prior and some observed link flows. *Discrete Dynamics in Nature and Society*, *2014*. doi: 10.1155/2014/192470

Cheng, Z., Jian, S., Rashidi, T., Maghrebi, M., & Waller, S. (2020). Integrating Household Travel Survey and Social Media Data to Improve the Quality of OD Matrix: A Comparative Case Study. *IEEE Transactions on Intelligent Transportation Systems*, *21*(6), 2628–2636. doi: 10.1109/TITS.2019.2958673

Cho, H.-J., Jou, Y.-J., & Lan, C.-L. (2009). Time dependent origin-destination estimation from traffic count without prior information. *Networks and Spatial Economics*, *9*(2), 145–170. doi: 10.1007/s11067-008-9082-7

Choi, K., Jayakrishnan, R., Kim, H., Yang, I., & Lee, J. (2009). Dynamic origin-destination estimation using dynamic traffic simulation model in an urban arterial corridor. *Transportation Research Record*(2133), 133–141. doi: 10.3141/2133-14

Chootinan, P., Chen, A., & Recker, W. (2005). Improved path flow estimator for origin-destination trip tables. *Transportation Research Record*(1923), 9–17. doi: 10.3141/1923-02

Cremer, M., & Keller, H. (1987). A new class of dynamic methods for the identification of origin-destination flows. *Transportation Research Part B*, *21*(2), 117–132. doi: 10.1016/0191-2615(87)90011-7

Dial, R. B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, *5*(2), 83–111. doi: https://doi.org/10.1016/0041-1647(71)90012-8

Diao, Y., Juan, Z.-C., & Ni, A.-N. (2012). Dynamic OD matrix estimation model of urban network. *Shanghai Jiaotong Daxue Xuebao/Journal of Shanghai Jiaotong University*, *46*(3), 436–440.

Dixon, M., & Rilett, L. (2005). Population origin-destination estimation using automatic vehicle identification and volume data. *Journal of Transportation Engineering*, *131*(2), 75–82. doi: 10.1061/(ASCE)0733-947X(2005)131:2(75)

Djukic, T., Van Lint, J. W. C., & Hoogendoorn, S. P. (2011). Efficient methodology for benchmarking dynamic origin-destination demand estimation methods. *Transportation Research Record*(2263), 35–44. doi: 10.3141/2263-05

Ehlert, A., Bell, M., & Grosso, S. (2006). The optimisation of traffic count locations in road networks. *Transportation Research Part B: Methodological*, *40*(6), 460–479. doi: 10.1016/j.trb.2005.06.001

Etemadnia, H., & Abdelghany, K. (2009). Distributed approach for estimation of dynamic origin-destination demand. *Transportation Research Record*(2105), 127–134. doi: 10.3141/2105-16

Fei, X., Mahmassani, H., & Murray-Tuite, P. (2013). Vehicular network sensor placement optimization under uncertainty. *Transportation Research Part C: Emerging Technologies*, *29*, 14–31. doi: 10.1016/j.trc.2013.01.004

Fekih, M., Bellemans, T., Smoreda, Z., Bonnel, P., Furno, A., & Galland, S. (2020). A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of Lyon region (France). *Transportation*. doi: 10.1007/s11116-020-10108-w

Feng, Y., Sun, J., & Chen, P. (2015). Vehicle trajectory reconstruction using automatic vehicle identification and traffic count data. *Journal of Advanced Transportation*, *49*(2), 174–194. doi: 10.1002/atr.1260

Frederix, R., Viti, F., Himpe, W. W. E., & Tampère, C. M. J. (2014). Dynamic origin-destination matrix estimation on large-scale congested networks using a hierarchical decomposition scheme. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *18*(1), 51–66. doi: 10.1080/15472450.2013.773249

Frederix, R., Viti, F., & Tampère, C. (2013). Dynamic origin-destination estimation in congested networks: Theoretical findings and implications in practice. *Transportmetrica A: Transport Science*, *9*(6), 494–513. doi: 10.1080/18128602.2011.619587

Fu, H., Lam, W., Shao, H., Xu, X., Lo, H., Chen, B., … Sumalee, A. (2019). Optimization of traffic count locations for estimation of travel demands with covariance between origin-destination flows. *Transportation Research Part C: Emerging Technologies*, *108*, 49–73. doi: 10.1016/j.trc.2019.09.004

Fujita, M., Yamada, S., & Murakami, S. (2017). Time coefficient estimation for hourly origin-destination demand from observed link flow based on semidynamic traffic assignment. *Journal of Advanced Transportation*, *2017*. doi: 10.1155/2017/6495861

Gan, L., Yang, H., & Wong, S. C. (2005). Traffic counting location and error bound in origin-destination matrix estimation problems. *Journal of Transportation Engineering*, *131*(7), 524–534. doi: 10.1061/(ASCE)0733-947X(2005)131:7(524)

Ge, Q., & Fukuda, D. (2016). Updating origin-destination matrices with aggregated data of GPS traces. *Transportation Research Part C: Emerging Technologies*, *69*, 291–312. doi: 10.1016/j.trc.2016.06.002

Gómez, P., Menéndez, M., & Mérida-Casermeiro, E. (2015). Evaluation of trade-offs between two data sources for the accurate estimation of origin-destination matrices. *Transportmetrica B*, *3*(3), 225–245. doi: 10.1080/21680566.2015.1025892

Gundlegård, D., Rydergren, C., Breyer, N., & Rajna, B. (2016). Travel demand estimation and network assignment based on cellular network data. *Computer Communications*, *95*, 29–42. doi: 10.1016/j.comcom.2016.04.015

Hadavi, M., & Shafahi, Y. (2016). Vehicle identification sensor models for origin-destination estimation. *Transportation Research Part B: Methodological*, *89*, 82–106. doi: 10.1016/j.trb.2016.03.011

Hazelton, M. (2000). Estimation of origin-destination matrices from link flows on uncongested networks. *Transportation Research Part B: Methodological*, *34*(7), 549–566. doi: 10.1016/S0191-2615(99)00037-5

Hazelton, M. (2001). Inference for origin-destination matrices: Estimation, prediction and reconstruction. *Transportation Research Part B: Methodological*, *35*(7), 667–676. doi: 10.1016/S0191-2615(00)00009-6

Hazelton, M. (2003). Some comments on origin–destination matrix estimation. *Transportation Research Part A: Policy and Practice*, *37*(10), 811–822. doi: https://doi.org/10.1016/S0965-8564(03)00044-2

Hazelton, M. (2010). Bayesian inference for network-based models with a linear inverse structure. *Transportation Research Part B: Methodological*, *44*(5), 674–685. doi: 10.1016/j.trb.2010.01.006

Hu, X., Chiu, Y.-C., Villalobos, J., & Nava, E. (2017). A Sequential Decomposition Framework and Method for Calibrating Dynamic Origin-Destination Demand in a Congested Network. *IEEE Transactions on Intelligent Transportation Systems*, *18*(10), 2790–2797. doi: 10.1109/TITS.2017.2661751

Infrastructure Canada. (2018). *Investing in Canada.* Transport.

Katranji, M., Kraiem, S., Moalic, L., Sanmarty, G., Khodabandelou, G., Caminada, A., & Hadj Selem, F. (2020). Deep multi-task learning for individuals origin–destination matrices estimation from census data. *Data Mining and Knowledge Discovery*, *34*(1), 201–230. doi: 10.1007/s10618-019-00662-y

Kattan, L., & Abdulhai, B. (2012). Sensitivity analysis of an evolutionary-based time-dependent origin/destination estimation framework. *IEEE Transactions on Intelligent Transportation Systems*, *13*(3), 1442–1453. doi: 10.1109/TITS.2012.2189437

Kim, E., Mahmassani, H. S., Ale-Ahmad, H., & Ostojic, M. (2019). Day-to-Day Learning Framework for Online Origin–Destination Demand Estimation and Network State Prediction. *Transportation Research Record*. doi: 10.1177/0361198119852075

Kim, H., Baek, S., & Lim, Y. (2001). Origin-destination matrices estimated with a genetic algorithm from link traffic counts. *Transportation Research Record*(1771), 156–163. doi: 10.3141/1771-20

Kim, H., & Jayakrishnan, R. (2010). The estimation of a time-dependent OD trip table with vehicle trajectory samples. *Transportation Planning and Technology*, *33*(8), 747–768. doi: 10.1080/03081060.2010.536629

Kim, H., Nam, D., Suh, W., & Cheon, S. (2018). Origin-destination trip table estimation based on subarea network OD flow and vehicle trajectory data. *Transportation Planning and Technology*, *41*(3), 265–285. doi: 10.1080/03081060.2018.1435437

Krishnakumari, P., van Lint, H., Djukic, T., & Cats, O. (2020). A data driven method for OD matrix estimation. *Transportation Research Part C: Emerging Technologies*, *113*, 38–56. doi: 10.1016/j.trc.2019.05.014

Kwon, J., & Varaiya, P. (2005). Real-time estimation of origin-destination matrices with partial trajectories from electronic toll collection tag data. *Transportation Research Record*(1923), 119–126. doi: 10.3141/1923-13

Laharotte, P.-A., Billot, R., Come, E., Oukhellou, L., Nantes, A., & El Faouzi, N.-E. (2015). Spatiotemporal analysis of bluetooth data: Application to a large urban network. *IEEE Transactions on Intelligent Transportation Systems*, *16*(3), 1439–1448. doi: 10.1109/TITS.2014.2367165

Lam, W., & Lo, H. (1990). Accuracy of O-D estimates from traffic counts. *Traffic Engineering & Control*, *31*(6), 358–367.

LeBlanc, L., Morlok, E., & Pierskalla, W. (1975). An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research*, *9*(5), 309–318. doi: 10.1016/0041-1647(75)90030-1

Lee, S., Heydecker, B., Kim, Y., & Shon, E.-Y. (2011). Dynamic OD estimation using three phase traffic flow theory. *Journal of Advanced Transportation*, *45*(2), 143–158. doi: 10.1002/atr.117

Li, R., Rose, G., & Sarvi, M. (2006). Evaluation of Speed-Based Travel Time Estimation Models. *Journal of Transportation Engineering*, *132*(7), 540–547. doi: 10.1061/(ASCE)0733-947X(2006)132:7(540)

Li, X., Kurths, J., Gao, C., Zhang, J., Wang, Z., & Zhang, Z. (2017). A Hybrid Algorithm for Estimating Origin-Destination Flows. *IEEE Access*, *6*, 677–687. doi: 10.1109/ACCESS.2017.2774449

Li, Y., Jiang, Z., Wang, H., & Lei, Z. (2016). A Novel Method for Estimation of Dynamic OD Flow. In *Procedia engineering* (Vol. 137, pp. 94–102). doi: 10.1016/j.proeng.2016.01.238

Lin, P.-W., & Chang, G.-L. (2006). Modeling measurement errors and missing initial values in freeway dynamic origin-destination estimation systems. *Transportation Research Part C: Emerging Technologies*, *14*(6), 384–402. doi: 10.1016/j.trc.2006.10.002

Liu, H., Wang, D., & Yue, H. (2015). Global optimization for transport network expansion and signal setting. *Mathematical Problems in Engineering*, *2015*. doi: 10.1155/2015/385713

Liu, J., Zheng, F., Van Zuylen, H., & Li, J. (2020). A dynamic OD prediction approach for urban networks based on automatic number plate recognition data. In *Transportation research procedia* (Vol. 47, pp. 601–608). doi: 10.1016/j.trpro.2020.03.137

Lo, H. P., & Chan, C.-P. (2003). Simultaneous estimation of an origin-destination matrix and link choice proportions using traffic counts. *Transportation Research Part A: Policy and Practice*, *37*(9), 771–788. doi: 10.1016/S0965-8564(03)00048-X

Lo, H. P., Zhang, N., & Lam, W. H. (1996). Estimation of an origin-destination matrix with random link choice proportions: A statistical approach. *Transportation Research Part B: Methodological*, *30*(4 PART B), 309–324. doi: 10.1016/0191-2615(95)00036-4

Lorenzo, M., & Matteo, M. (2013). OD matrices network estimation from link counts by neural networks. *Journal of Transportation Systems Engineering and Information Technology*, *13*(4), 84–92. doi: 10.1016/S1570-6672(13)60117-8

Lou, Y., & Yin, Y. (2010). A decomposition scheme for estimating dynamic origin-destination flows on actuation-controlled signalized arterials. *Transportation Research Part C: Emerging Technologies*, *18*(5), 643–655. doi: 10.1016/j.trc.2009.06.005

Lu, L., Xu, Y., Antoniou, C., & Ben-Akiva, M. (2015). An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models. *Transportation Research Part C: Emerging Technologies*, *51*, 149–166. doi: 10.1016/j.trc.2014.11.006

Lu, Z., Rao, W., Wu, Y.-J., Guo, L., & Xia, J. (2015). A Kalman filter approach to dynamic OD flow estimation for urban road networks using multi-sensor data. *Journal of Advanced Transportation*, *49*(2), 210–227. doi: 10.1002/atr.1292

Lundgren, J., & Peterson, A. (2008). A heuristic for the bilevel origin-destination-matrix estimation problem. *Transportation Research Part B: Methodological*, *42*(4), 339–354. doi: 10.1016/j.trb.2007.09.005

Ma, W., Pi, X., & Qian, S. (2020). Estimating multi-class dynamic origin-destination demand through a forward-backward algorithm on computational graphs. *Transportation Research Part C: Emerging Technologies*, *119*. doi: 10.1016/j.trc.2020.102747

Ma, W., & Qian, Z. (2018a). Estimating multi-year 24/7 origin-destination demand using high-granular multi-source traffic data. *Transportation Research Part C: Emerging Technologies*, *96*, 96–121. doi: 10.1016/j.trc.2018.09.002

Ma, W., & Qian, Z. (2018b). A Generalized Single-Level Formulation for Origin–Destination Estimation under Stochastic User Equilibrium. *Transportation Research Record*, *2672*(48), 58–68. doi: 10.1177/0361198118782041

Ma, W., & Qian, Z. (2018c). Statistical inference of probabilistic origin-destination demand using day-to-day traffic data. *Transportation Research Part C: Emerging Technologies*, *88*, 227–256. doi: 10.1016/j.trc.2017.12.015

Ma, Y., Kuik, R., & Van Zuylen, H. J. (2013). Day-to-day origin–destination tuple estimation and prediction with hierarchical Bayesian networks using multiple data sources. *Transportation Research Record*, *2343*, 51–61. doi: 10.3141/2343-07

Maher, M. (1983). Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach. *Transportation Research Part B*, *17*(6), 435–447. doi: 10.1016/0191-2615(83)90030-9

Mamei, M., Bicocchi, N., Lippi, M., Mariani, S., & Zambonelli, F. (2019). Evaluating origin–destination matrices obtained from CDR data. *Sensors (Switzerland)*, *19*(20). doi: 10.3390/s19204470

McNally, M. G., & Rindt, C. R. (2007). The Activity-Based Approach. In D. A. Hensher & K. J. Button (Eds.), *Handbook of transport modelling* (Vol. 1, pp. 55–73). Emerald Group Publishing Limited. doi: 10.1108/9780857245670-004

Menon, A., Cai, C., Wang, W., Wen, T., & Chen, F. (2015). Fine-grained OD estimation with automated zoning and sparsity regularisation. *Transportation Research Part B: Methodological*, *80*, 150–172. doi: 10.1016/j.trb.2015.07.003

Michau, G., Pustelnik, N., Borgnat, P., Abry, P., Nantes, A., Bhaskar, A., & Chung, E. (2017). A Primal-Dual Algorithm for Link Dependent Origin Destination Matrix Estimation. *IEEE Transactions on Signal and Information Processing over Networks*, *3*(1), 104–113. doi: 10.1109/TSIPN.2016.2623094

Mínguez, R., Sánchez-Cambronero, S., Castillo, E., & Jiménez, P. (2010). Optimal traffic plate scanning location for OD trip matrix and route estimation in road networks. *Transportation Research Part B: Methodological*, *44*(2), 282–298. doi: 10.1016/j.trb.2009.07.008

Mitra, A., Attanasi, A., Meschini, L., & Gentile, G. (2020). Methodology for O-D matrix estimation using the revealed paths of floating car data on large-scale networks. *IET Intelligent Transport Systems*, *14*(12), 1704–1711. doi: 10.1049/iet-its.2019.0684

Mo, B., Li, R., & Dai, J. (2020). Estimating dynamic origin–destination demand: A hybrid framework using license plate recognition data. *Computer-Aided Civil and Infrastructure Engineering*, *35*(7), 734–752. doi: 10.1111/mice.12526

Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., & Damas, L. (2016). Time-evolving O-D matrix estimation using high-speed GPS data streams. *Expert Systems with Applications*, *44*, 275–288. doi: 10.1016/j.eswa.2015.08.048

Mungthanya, W., Phithakkitnukoon, S., Demissie, M., Kattan, L., Veloso, M., Bento, C., & Ratti, C. (2019). Constructing Time-Dependent Origin-Destination Matrices with Adaptive Zoning Scheme and Measuring Their Similarities with Taxi Trajectory Data. *IEEE Access*, *7*, 77726–77737. doi: 10.1109/ACCESS.2019.2922210

Munizaga, M., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, *24*, 9–18. doi: 10.1016/j.trc.2012.01.007

Mussone, L., Grant-Muller, S., & Chen, H.-B. (2010). A neural network approach to motorway OD matrix estimation from loop counts. *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/ Journal of Transportation Systems Engineering and Information Technology*, *10*(1), 88–98. doi: 10.1016/s1570-6672(09)60026-x

Nguyen, S., Morello, E., & Pallottino, S. (1988). Discrete time dynamic estimation model for passenger origin/destination matrices on transit networks. *Transportation Research Part B*, *22*(4), 251–260. doi: 10.1016/0191-2615(88)90002-1

Nie, Y., & Lee, D.-H. (2002). Uncoupled method for equilibrium-based linear path flow estimator for origin-destination trip matrices. *Transportation Research Record*(1783), 72–74. doi: 10.3141/1783-10

Nie, Y., & Zhang, H. (2008). A variational inequality formulation for inferring dynamic origin-destination travel demands. *Transportation Research Part B: Methodological*, *42*(7-8), 635–662. doi: 10.1016/j.trb.2008.01.001

Nigro, M., Cipriani, E., & del Giudice, A. (2018). Exploiting floating car data for time-dependent Origin–Destination matrices estimation. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *22*(2), 159–174. doi: 10.1080/15472450.2017.1421462

Nihan, N., & Davis, G. (1987). Recursive estimation of origin-destination matrices from input/output counts. *Transportation Research Part B*, *21*(2), 149–163. doi: 10.1016/0191-2615(87)90013-0

Okutani, I., & Stephanedes, Y. (1984). Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B*, *18*(1), 1–11. doi: 10.1016/0191-2615(84)90002-X

Osorio, C. (2019). Dynamic origin-destination matrix calibration for large-scale network simulators. *Transportation Research Part C: Emerging Technologies*, *98*, 186–206. doi: 10.1016/j.trc.2018.09.023

Ou, J., Lu, J., Xia, J., An, C., & Lu, Z. (2019). Learn, Assign, and Search: Real-Time Estimation of Dynamic Origin-Destination Flows Using Machine Learning Algorithms. *IEEE Access*, *7*, 26967–26983. doi: 10.1109/ACCESS.2019.2901289

Ouyang, L. Q., & Huang, D. (2018). An activity-based model for the estimation of origin-destination matrix on road networks. *Tehnicki Vjesnik*, *25*(5), 1530–1537. doi: 10.17559/TV-20180527042248

Park, E. S., Rilett, L. R., & Spiegelman, C. H. (2008). A Markov Chain Monte Carlo-based origin destination matrix estimator that is robust to imperfect intelligent transportation systems data. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *12*(3), 139–155. doi: 10.1080/15472450802262364

Parry, K., & Hazelton, M. (2012). Estimation of origin-destination matrices from link counts and sporadic routing data. *Transportation Research Part B: Methodological*, *46*(1), 175–188. doi: 10.1016/j.trb.2011.09.009

Perrakis, K., Karlis, D., Cools, M., Janssens, D., Vanhoof, K., & Wets, G. (2012). A Bayesian approach for modeling origin-destination matrices. *Transportation Research Part A: Policy and Practice*, *46*(1), 200–212. doi: 10.1016/j.tra.2011.06.005

Pishue, B. (2021). *2021 INRIX Global Traffic Scorecard* (No. December). Retrieved from `https://www2.inrix.com/e/171932/021-INRIX-Scorecard-Report-pdf/5p2p8h/907921818?h=5jLvVtgvvn4LWcenQuJzWJRQUs8OaNtR3hACL2Ld7us`

Pitombeira-Neto, A., Loureiro, C., & Carvalho, L. (2020). A Dynamic Hierarchical Bayesian Model for the Estimation of day-to-day Origin-destination Flows in Transportation Networks. *Networks and Spatial Economics*, *20*(2), 499–527. doi: 10.1007/s11067-019-09490-5

Pitombeira-Neto, A., & Loureiro, C. F. G. (2016). A dynamic linear model for the estimation of time-varying origin–destination matrices from link counts. *Journal of Advanced Transportation*, *50*(8), 2116–2129. doi: 10.1002/atr.1449

Rao, W., Wu, Y.-J., Xia, J., Ou, J., & Kluger, R. (2018). Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data. *Transportation Research Part C: Emerging Technologies*, *95*, 29–46. doi: 10.1016/j.trc.2018.07.002

Ros-Roca, X., Montero, L., & Barceló, J. (2021). Investigating the quality of Spiess-like and SPSA approaches for dynamic OD matrix estimation. *Transportmetrica A: Transport Science*, *17*(3), 235–257. doi: 10.1080/23249935.2020.1722282

Rostami Nasab, M., & Shafahi, Y. (2020). Estimation of origin–destination matrices using link counts and partial path data. *Transportation*, *47*(6), 2923–2950. doi: 10.1007/s11116-019-09999-1

Sana, B., Castiglione, J., Cooper, D., & Tischler, D. (2018). *Using Google's Passive Data and Machine Learning for Origin-Destination Demand Estimation* (Vol. 2672) (No. 46). doi: 10.1177/0361198118798298

Shafei, S., Saberi, M., Zockaie, A., & Sarvi, M. (2017). Sensitivity-based linear approximation method to estimate time-dependent origin-destination demand in congested networks. *Transportation Research Record*, *2669*, 72–79. doi: 10.3141/2669-08

Shafiei, S., Saberi, M., & Sarvi, M. (2016). Application of an exact gradient method to estimate dynamic origin-destination demand for melbourne network. In *Ieee conference on intelligent transportation systems, proceedings, itsc* (pp. 1945–1950). doi: 10.1109/ITSC.2016.7795870

Sheffi, Y. (1984). *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods.*

Sherali, H., Narayanan, A., & Sivanandan, R. (2003). Estimation of origin-destination trip-tables based on a partial set of traffic link volumes. *Transportation Research*

Part B: Methodological, *37*(9), 815–836. doi: 10.1016/S0191-2615(02)00073-5

Sherali, H., & Park, T. (2001). Estimation of dynamic origin-destination trip tables for a general network. *Transportation Research Part B: Methodological*, *35*(3), 217–235. doi: 10.1016/S0191-2615(99)00048-X

Sherali, H., Sivanandan, R., & Hobeika, A. (1994). A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes. *Transportation Research Part B*, *28*(3), 213–233. doi: 10.1016/0191-2615(94)90008-6

Simonelli, F., Marzano, V., Papola, A., & Vitiello, I. (2012). A network sensor location procedure accounting for o-d matrix estimate variability. *Transportation Research Part B: Methodological*, *46*(10), 1624–1638. doi: 10.1016/j.trb.2012.08.007

Sohn, K., & Kim, D. (2008). Dynamic origin-destination flow estimation using cellular communication system. *IEEE Transactions on Vehicular Technology*, *57*(5), 2703–2713. doi: 10.1109/TVT.2007.912336

Spiess, H. (1987). A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B*, *21*(5), 395–412. doi: 10.1016/0191-2615(87)90037-3

Stathopoulos, A., & Tsekeris, T. (2005). Methodology for validating dynamic Origin-Destination matrix estimation models with implications for Advanced Traveler Information Systems. *Transportation Planning and Technology*, *28*(2), 93–112. doi: 10.1080/03081060500053368

Stopher, P., & Greaves, S. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, *41*(5), 367–381. doi: https://doi.org/10.1016/j.tra.2006.09.005

Sun, J., & Feng, Y. (2011). A Novel OD Estimation Method Based on Automatic Vehicle Identification Data. *Communications in Computer and Information Science*, *135*(PART 2), 461–470. doi: 10.1007/978-3-642-18134-4{_}74

Talebian, A., & Shafahi, Y. (2015). The treatment of uncertainty in the dynamic origin–destination estimation problem using a fuzzy approach. *Transportation Planning and Technology*, *38*(7), 795–815. doi: 10.1080/03081060.2015.1059124

Tang, K., Cao, Y., Chen, C., Yao, J., Tan, C., & Sun, J. (2021). Dynamic origin-destination flow estimation using automatic vehicle identification data: A 3D convolutional neural network approach. *Computer-Aided Civil and Infrastructure Engineering*, *36*(1), 30–46. doi: 10.1111/mice.12559

Tesselkin, A., & Khabarov, V. (2017). Estimation of Origin-Destination Matrices Based on Markov Chains. In *Procedia engineering* (Vol. 178, pp. 107–116). doi: 10.1016/j.proeng.2017.01.071

Toledo, T., Kolechkina, T., Wagner, P., Ciuffo, B., Azevedo, C., Marzano, V., & Flötteröd, G. (2014). *Network model calibration studies.* doi: 10.1201/b17440

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*. doi: 10.1016/j.trc.2015.04.022

Tornero, R., Martínez, J., & Castelló, J. (2012). Computing real-time dynamic origin/destination matrices from vehicle-to-infrastructure messages using a multi-agent system. *Advances in Intelligent and Soft Computing*, *156 AISC*, 147–154. doi: 10.1007/978-3-642-28762-6{_}18

Transport Canada. (2006). *The Cost of Urban Congestion in Canada* (No. March 22). Environmental Affairs.

Transportation Networks for Research Core Team. (n.d.). *Transportation Networks for Research.* Retrieved from `https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls`

Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big Data*, *2*(1), 21. doi: 10.1186/s40537-015-0030-3

Tsekeris, T., & Stathopoulos, A. (2008). Treating uncertain demand information in origin-destination matrix estimation with traffic counts. *Journal of Transportation Engineering*, *134*(8), 327–337. doi: 10.1061/(ASCE)0733-947X(2008)134:8(327)

Tympakianaki, A., Koutsopoulos, H., & Jenelius, E. (2015). c-SPSA: Cluster-wise simultaneous perturbation stochastic approximation algorithm and its application to dynamic origin-destination matrix estimation. *Transportation Research Part C: Emerging Technologies*, *55*, 231–245. doi: 10.1016/j.trc.2015.01.016

Van Aerde, M., Hellinga, B., & Yu, L. (1993). Vehicle Probes As Real-time ATMS Sources Of Dynamic OD And Travel Time Data.

Van Aerde, M., Rakha, H., & Paramahamsan, H. (2003). Estimation of Origin-Destination Matrices: Relationship Between Practical and Theoretical Considerations. *Transportation Research Record*(1831), 122–130.

Van Zuylen, H., & Willumsen, L. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, *14*(3), 281–293. doi: 10.1016/0191-2615(80)90008-9

Vogt, S., Fourati, W., Schendzielorz, T., & Friedrich, B. (2019). Estimation of origin-destination matrices by fusing detector data and Floating Car Data. In *Transportation research procedia* (Vol. 37, pp. 473–480). doi: 10.1016/j.trpro.2018.12.216

Wang, H., & Zhang, X. (2011). Estimation of origin-destination matrix with tolling data. *Applied Mechanics and Materials*, *50-51*, 239–244. doi: 10.4028/www.scientific.net/AMM.50-51.239

Weiner, E. (1997). *Urban transportation planning in the United States : an historical overview.*

Wong, S., & Tong, C. (1998). Estimation of time-dependent origin-destination matrices for transit networks. *Transportation Research Part B: Methodological*, *32*(1), 35–48. doi: 10.1016/S0191-2615(97)00011-8

Wong, S., Tong, C., Wong, K., Lam, W., Lo, H. H., Yang, H., & Lo, H. H. (2005). Estimation of multiclass origin-destination matrices from traffic counts. *Journal of Urban Planning and Development*, *131*(1), 19–29. doi: 10.1061/(ASCE)0733-9488(2005)131:1(19)

Woo, S., Tak, S., & Yeo, H. (2016). Data-driven prediction methodology of origin-destination demand in large network for real-time service. *Transportation Research Record*, *2567*, 47–56. doi: 10.3141/2567-06

Wu, X., Guo, J., Xian, K., & Zhou, X. (2018). Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph. *Transportation Research Part C: Emerging Technologies*, *96*, 321–346. doi: 10.1016/j.trc.2018.09.021

Xian, X., Ye, H., Wang, X., & Liu, K. (2021). Spatiotemporal Modeling and Real-Time Prediction of Origin-Destination Traffic Demand. *Technometrics*, *63*(1), 77–89. doi: 10.1080/00401706.2019.1704887

Xie, C., & Duthie, J. (2015). An Excess-Demand Dynamic Traffic Assignment Approach for Inferring Origin-Destination Trip Matrices. *Networks and Spatial Economics*, *15*(4), 947–979. doi: 10.1007/s11067-014-9277-z

Xie, C., Kockelman, K., & Waller, S. (2011). A maximum entropy-least squares estimator for elastic origin-destination trip matrix estimation. *Transportation Research Part B: Methodological*, *45* ‘(9), 1465–1482. doi: 10.1016/j.trb.2011.05.018

Yang, H., Meng, Q., & Bell, M. (2001). Simultaneous estimation of the origin-destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium. *Transportation Science*, *35*(2), 107–123. doi: 10.1287/trsc.35.2.107.10133

Yang, H., & Zhou, J. (1998). Optimal traffic counting locations for origin-destination matrix estimation. *Transportation Research Part B: Methodological*, *32*(2), 109–126. doi: 10.1016/s0191-2615(97)00016-7

Yang, X., Lu, Y., & Hao, W. (2017). Origin-destination estimation using probe vehicle trajectory and link counts. *Journal of Advanced Transportation*, *2017*. doi: 10.1155/2017/4341532

Ye, P., & Wen, D. (2017). Optimal Traffic Sensor Location for Origin-Destination Estimation Using a Compressed Sensing Framework. *IEEE Transactions on Intelligent Transportation Systems*, *18*(7), 1857–1866. doi: 10.1109/TITS.2016.2614828

Zhang, H., Nie, Y., & Qian, Z. (2008). Estimating time-dependent freeway origin-Destination demands with different data coverage: Sensitivity analysis. *Transportation Research Record*(2047), 91–99. doi: 10.3141/2047-11

Zhao, J., Rahbee, A., & Wilson, N. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering*, *22*(5), 376–387. doi: 10.1111/j.1467-8667.2007.00494.x

Zhou, X., & List, G. (2010). An information-theoretic sensor location model for traffic origin-destination demand estimation applications. *Transportation Science*, *44*(2), 254–273. doi: 10.1287/trsc.1100.0319

Zhou, X., Liu, Z., Zhao, X. X., & Guo, J. H. (2014). Framework for dynamic OD matrix estimation based on multi-source traffic data fusion. *Applied Mechanics and Materials*, *505-506*, 1153–1156. doi: 10.4028/www.scientific.net/AMM.505-506.1153

Zhou, X., & Mahmassani, H. (2006). Dynamic origin-destination demand estimation using automatic vehicle identification data. *IEEE Transactions on Intelligent Transportation Systems*, *7*(1), 105–114. doi: 10.1109/TITS.2006.869629

Zhou, X., & Mahmassani, H. (2007). A structural state space model for real-time traffic origin-destination demand estimation and prediction in a day-to-day learning framework. *Transportation Research Part B: Methodological*, *41*(8), 823–840. doi: 10.1016/j.trb.2007.02.004

Zhou, X., Qin, X., & Mahmassani, H. (2003). Dynamic Origin-Destination Demand Estimation with Multiday Link Traffic Counts for Planning Applications. *Transportation Research Record*(1831), 30–38. doi: 10.3141/1831-04

APPENDICES

Appendix A

Incremental Assignment

The incremental assignment procedure is a heuristic method for obtaining a UE solution based on Sheffi (1984) and outlined below.

Step 0: Preliminaries

- Divide each O-D entry into N equal portion:

$$q_{ij}^n = \frac{q_{ij}}{N}$$

where:

q_{ij}^n = incremental n of the total O-D demand between i and j
q_{ij} = total O-D demand between i and j
N = number of increments

- Set increment to $n = 1$
- Set all link flows to zero:

$$x_a^0 = 0,\ \forall\ a$$

where:

x_a^0 = flow on link a for increment 0

Step 1: Update

- Set

$$t_a^n = t_a(x_a^{n-1}),\ \forall\ a$$

where:

t_a^n	= travel time on link a for increment n
$t_a(x_a^{n-1})$	= travel time on link a after $n-1$ increments of the O-D matrix have been assigned

Step 2: Incremental loading

- Perform AON assignment based on t_a^n, but using q_{ij}^n trip rates for each O-D pair
- These yields flow pattern w_a^n

Step 3: Flow summation

- Set

$$x_a^n = x_a^{n-1} + w_a^n,\ \forall\ a$$

Step 4: Stopping rule

- If $n = N$, stop (the current set of link flows is the solution)
- Otherwise, $n = n + 1$ and go to Step 1

Appendix B

Criteria Selection Matrix

Research	Static/Dynamic	Remove?/Reason
(Fekih et al., 2020)	Static	Yes/Data cannot be replicated
(Behara et al., 2021)	Static	Yes/Testbed network
(Tang et al., 2021)	Dynamic	Prediction
(Ros-Roca et al., 2021)	Dynamic	Yes/ Dynamic estimation
(Xian et al., 2021)	Dynamic	Prediction
(Mitra et al., 2020)	Static	new approach
(Rostami Nasab & Shafahi, 2020)	Static	new approach
(W. Ma et al., 2020)	Dynamic	Yes/ Dynamic estimation
(Mo et al., 2020)	Dynamic	Yes/ Dynamic estimation
(Z. Cheng et al., 2020)	Static	Yes/ Data cannot be replicated
(Pitombeira-Neto et al., 2020)	Dynamic	Yes/ Dynamic estimation
(Krishnakumari et al., 2020)	Static	new approach
(J. Liu et al., 2020)	Dynamic	Yes/ Dynamic estimation
(Mamei et al., 2019)	Dynamic	Yes/ Dynamic estimation
(Bachir et al., 2019)	Dynamic	Yes/ Dynamic estimation
(E. Kim et al., 2019)	Dynamic	Yes/ Dynamic estimation
(Ou et al., 2019)	Dynamic	Yes/ Dynamic estimation
(Vogt et al., 2019)	Static	Yes/ Testbed network
(Osorio, 2019)	Dynamic	Yes/ Dynamic estimation
(Sana et al., 2018)	Static	Yes/ Data cannot be replicated
(W. Ma & Qian, 2018b)	Static	new approach
(Wu et al., 2018)	Dynamic	Yes/ Dynamic estimation

(W. Ma & Qian, 2018a)	Dynamic	Yes/ Dynamic estimation
(Ouyang & Huang, 2018)	Static	Yes/ Testbed network
(Bauer et al., 2018)	Dynamic	Yes/ Dynamic estimation
(H. Kim et al., 2018)	Static	Yes/ Testbed network
(Nigro et al., 2018)	Dynamic	Yes/ Dynamic estimation
(Bonnel et al., 2018)	Static	Yes/ Data cannot be replicated
(X. Li et al., 2017)	Dynamic	Prediction
(Hu et al., 2017)	Dynamic	Yes/ Dynamic estimation
(Carrese et al., 2017)	Dynamic	Yes/ Dynamic estimation
(Michau et al., 2017)	Static	new approach
(Shafei et al., 2017)	Dynamic	Yes/ Dynamic estimation
(Fujita et al., 2017)	Dynamic	Yes/ Dynamic estimation
(Tesselkin & Khabarov, 2017)	Static	Yes/ Weak performance in strongly-connected networks
(X. Yang et al., 2017)	Static	No/Chosen
(Shafiei et al., 2016)	Dynamic	Yes/ Dynamic estimation
(Pitombeira-Neto & Loureiro, 2016)	Dynamic	Yes/ Dynamic estimation
(Gundlegård et al., 2016)	Static	Yes/ Data cannot be replicated
(Ge & Fukuda, 2016)	Static	Yes/ Data cannot be replicated
(Moreira-Matias et al., 2016)	Dynamic	Yes/ Dynamic estimation
(Woo et al., 2016)	Dynamic	Yes/ Dynamic estimation
(Y. Li et al., 2016)	Dynamic	Yes/ Dynamic estimation
(Xie & Duthie, 2015)	Dynamic	Yes/ Dynamic estimation
(Talebian & Shafahi, 2015)	Dynamic	Yes/ Dynamic estimation
(Gómez et al., 2015)	Static	Yes/ Dynamic estimation
(Alexander et al., 2015)	Static	Yes/ Data cannot be replicated
(Toole et al., 2015)	Static	Yes/ Data cannot be replicated
(Laharotte et al., 2015)	Static	Yes/ Data cannot be replicated
(Tympakianaki et al., 2015)	Dynamic	Yes/ Dynamic estimation
(Feng et al., 2015)	Static	No/ Chosen
(Z. Lu et al., 2015)	Dynamic	Yes/ Dynamic estimation
(Abdelghany et al., 2015)	Dynamic	Yes/ Dynamic estimation
(Carvalho, 2014)	Static	Yes/ Testbed network
(Zhou et al., 2014)	Dynamic	Yes/ Dynamic estimation
(Frederix et al., 2014)	Dynamic	Yes/ Dynamic estimation
(L. Cheng et al., 2014)	Static	Yes/ Testbed network
(Barceló et al., 2013)	Dynamic	Yes/ Dynamic estimation

(Y. Ma et al., 2013)	Dynamic	Yes/ Dynamic estimation
(Lorenzo & Matteo, 2013)	Static	Yes/ Data cannot be replicated
(Cascetta et al., 2013)	Dynamic	Yes/ Dynamic estimation
(Caggiani et al., 2013)	Static	Yes/ Testbed network
(Kattan & Abdulhai, 2012)	Dynamic	Yes/ Dynamic estimation
(Tornero et al., 2012)	Dynamic	Yes/ Dynamic estimation
(Diao et al., 2012)	Dynamic	Yes/ Dynamic estimation
(Parry & Hazelton, 2012)	Static	No/ Chosen
(Perrakis et al., 2012)	Static	Yes/ Data cannot be replicated
(Sun & Feng, 2011)	Static	Yes/ Testbed network
(Lee et al., 2011)	Dynamic	Yes/ Dynamic estimation
(Wang & Zhang, 2011)	Static	Yes/ Testbed network
(Xie et al., 2011)	Static	Yes/ Not general O-D estimation
(D. Chen et al., 2011)	Dynamic	Yes/ Dynamic estimation
(H. Kim & Jayakrishnan, 2010)	Dynamic	Yes/ Dynamic estimation
(Baek et al., 2010)	Static	Yes/ Non-testbed network did not establish ground truth
(Lou & Yin, 2010)	Dynamic	Yes/ Dynamic estimation
(Mussone et al., 2010)	Static	Yes/ Not general O-D estimation
(Etemadnia & Abdelghany, 2009)	Dynamic	Yes/ Dynamic estimation
(Choi et al., 2009)	Dynamic	Yes/ Dynamic estimation
(A. Chen et al., 2009)	Static	Yes/ Not end-to-end estimation
(Cho et al., 2009)	Dynamic	Yes/ Dynamic estimation
(Sohn & Kim, 2008)	Dynamic	Yes/ Dynamic estimation
(Zhang et al., 2008)	Dynamic	Yes/ Dynamic estimation
(Tsekeris & Stathopoulos, 2008)	Dynamic	Yes/ Dynamic estimation
(Park et al., 2008)	Static	Yes/ Testbed network
(Castillo et al., 2008b)	Static	No/Chosen
(Nie & Zhang, 2008)	Dynamic	Yes/ Dynamic estimation
(Zhou & Mahmassani, 2006)	Dynamic	Yes/ Dynamic estimation
(Lin & Chang, 2006)	Dynamic	Yes/ Dynamic estimation
(Gan et al., 2005)	Static	Yes/ Testbed network
(Stathopoulos & Tsekeris, 2005)	Dynamic	Yes/ Dynamic estimation
(Wong et al., 2005)	Static	Yes/ Not general O-D estimation

(Dixon & Rilett, 2005)	Static	Yes/ Testbed network
(Chootinan et al., 2005)	Static	Yes/ Not end-to-end estimation
(Kwon & Varaiya, 2005)	Dynamic	Yes/ Dynamic estimation
(Bierlaire & Crittin, 2004)	Dynamic	Yes/ Dynamic estimation
(Van Aerde et al., 2003)	Static	Yes/ Testbed network
(Zhou et al., 2003)	Dynamic	Yes/ Dynamic estimation
(Sherali et al., 2003)	Static	Yes/ Testbed network
(Nie & Lee, 2002)	Static	Yes/ Not end-to-end estimation
(Ashok & Ben-Akiva, 2002)	Dynamic	Yes/ Dynamic estimation
(H. Kim et al., 2001)	Static	Yes/ Testbed network
(Hazelton, 2001)	Static	Yes/ Testbed network
(Cascetta & Postorino, 2001)	Static	No/ Chosen
(H. Yang et al., 2001)	Static	Yes/ Testbed network
(Sherali & Park, 2001)	Dynamic	Yes/ Dynamic estimation
(Hazelton, 2000)	Static	Yes/ Testbed network

Appendix C

Link Properties

Link	t	α	K
1	6	0.9	25.9002
2	4	0.6	23.4035
3	6	0.9	25.9002
4	5	0.75	4.9582
5	4	0.6	23.4035
6	4	0.6	17.1105
7	4	0.6	23.4035
8	4	0.6	17.1105
9	2	0.3	17.7828
10	6	0.9	4.9088
11	2	0.3	17.7828
12	4	0.6	4.948
13	5	0.75	10
14	5	0.75	4.9582
15	4	0.6	4.948
16	2	0.3	4.8986
17	3	0.45	7.8418
18	2	0.3	23.4035
19	2	0.3	4.8986
20	3	0.45	7.8418
21	10	1.5	5.0502
22	5	0.75	5.0458
23	5	0.75	10
24	10	1.5	5.0502
25	3	0.45	13.9158
26	3	0.45	13.9158
27	5	0.75	10
28	6	0.9	13.512
29	5	0.75	5.1335
30	8	1.2	4.9935
31	6	0.9	4.9088
32	5	0.75	10
33	6	0.9	4.9088
34	4	0.6	4.8765
35	4	0.6	23.4035
36	6	0.9	4.9088
37	3	0.45	25.9002
38	3	0.45	25.9002

Link	t	α	K
39	4	0.6	5.0913
40	4	0.6	4.8765
41	5	0.75	5.1275
42	4	0.6	4.9248
43	6	0.9	13.512
44	5	0.75	5.1275
45	4	0.6	15.6508
46	4	0.6	10.315
47	5	0.75	5.0458
48	5	0.75	5.1335
49	2	0.3	5.2299
50	3	0.45	19.6799
51	8	1.2	4.9935
52	2	0.3	5.2299
53	2	0.3	4.824
54	2	0.3	23.4035
55	3	0.45	19.6799
56	4	0.6	23.4035
57	4	0.6	15.6508
58	2	0.3	4.824
59	4	0.6	5.0026
60	4	0.6	23.4035
61	4	0.6	5.0026
62	6	0.9	5.0599
63	5	0.75	5.0757
64	6	0.9	5.0599
65	2	0.3	5.2299
66	3	0.45	4.8854
67	4	0.6	10.315
68	5	0.75	5.0757
69	2	0.3	5.2299
70	4	0.6	5
71	4	0.6	4.9248
72	4	0.6	5
73	2	0.3	5.0785
74	4	0.6	5.0913
75	3	0.45	4.8854
76	2	0.3	5.0785

Appendix D

Initial Matrices

UE Matrix (Origin × Destination)

$$
\begin{bmatrix}
0 & 20 & 20 & 40 & 0 & 0 & 20 & 0 & 20 & 80 & 40 & 60 & 20 & 20 & 20 & 20 & 40 & 20 & 40 & 20 & 20 & 20 & 0 & 20 \\
20 & 0 & 20 & 0 & 0 & 20 & 60 & 0 & 20 & 0 & 20 & 0 & 40 & 0 & 0 & 0 & 20 & 20 & 0 & 40 & 0 & 20 & 20 & 40 \\
40 & 0 & 0 & 20 & 20 & 0 & 100 & 120 & 0 & 20 & 40 & 0 & 0 & 60 & 20 & 20 & 20 & 20 & 0 & 0 & 20 & 0 & 40 & 40 \\
60 & 0 & 0 & 0 & 40 & 20 & 20 & 0 & 40 & 100 & 160 & 0 & 40 & 80 & 60 & 40 & 20 & 20 & 20 & 40 & 20 & 0 & 20 & 40 \\
0 & 0 & 0 & 40 & 0 & 0 & 100 & 120 & 0 & 20 & 60 & 0 & 20 & 40 & 20 & 20 & 0 & 0 & 20 & 20 & 40 & 20 & 0 & 60 \\
0 & 0 & 60 & 40 & 40 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 40 & 60 & 20 & 80 & 80 & 0 & 20 & 0 & 20 & 20 & 20 & 40 \\
80 & 0 & 0 & 0 & 60 & 20 & 0 & 80 & 0 & 60 & 60 & 40 & 60 & 80 & 60 & 40 & 60 & 40 & 20 & 0 & 40 & 20 & 40 & 0 \\
80 & 20 & 20 & 0 & 20 & 40 & 80 & 0 & 40 & 100 & 60 & 60 & 100 & 40 & 20 & 80 & 60 & 0 & 40 & 60 & 0 & 60 & 40 & 0 \\
40 & 0 & 20 & 0 & 0 & 40 & 100 & 40 & 0 & 80 & 40 & 20 & 60 & 0 & 0 & 40 & 40 & 40 & 20 & 20 & 20 & 40 & 0 & 40 \\
20 & 0 & 20 & 40 & 20 & 100 & 60 & 100 & 140 & 0 & 120 & 100 & 60 & 20 & 120 & 220 & 100 & 40 & 40 & 100 & 100 & 60 & 60 & 40 \\
40 & 20 & 20 & 20 & 20 & 20 & 40 & 120 & 20 & 200 & 0 & 80 & 60 & 20 & 20 & 160 & 60 & 20 & 40 & 20 & 80 & 80 & 80 & 80 \\
0 & 0 & 0 & 40 & 0 & 20 & 20 & 20 & 40 & 60 & 40 & 0 & 0 & 40 & 80 & 40 & 20 & 20 & 80 & 0 & 0 & 100 & 40 & 60 \\
20 & 0 & 0 & 60 & 0 & 40 & 40 & 40 & 20 & 120 & 20 & 40 & 0 & 40 & 40 & 20 & 20 & 0 & 0 & 40 & 20 & 20 & 20 & 120 \\
0 & 0 & 0 & 20 & 20 & 20 & 0 & 40 & 40 & 20 & 40 & 20 & 20 & 0 & 60 & 20 & 40 & 0 & 20 & 100 & 40 & 60 & 120 & 0 \\
0 & 40 & 60 & 20 & 0 & 60 & 60 & 40 & 60 & 140 & 60 & 20 & 20 & 40 & 0 & 40 & 60 & 20 & 60 & 20 & 80 & 40 & 80 & 0 \\
0 & 20 & 60 & 0 & 40 & 100 & 120 & 20 & 0 & 160 & 20 & 80 & 80 & 20 & 60 & 0 & 100 & 60 & 80 & 60 & 40 & 20 & 20 & 20 \\
0 & 0 & 20 & 20 & 20 & 60 & 20 & 40 & 60 & 80 & 20 & 20 & 20 & 60 & 40 & 100 & 0 & 40 & 80 & 140 & 20 & 60 & 20 & 60 \\
40 & 40 & 20 & 20 & 20 & 20 & 20 & 20 & 40 & 60 & 40 & 100 & 0 & 80 & 0 & 40 & 0 & 0 & 0 & 40 & 0 & 80 & 20 & 0 \\
60 & 0 & 60 & 0 & 60 & 20 & 40 & 40 & 40 & 60 & 40 & 20 & 100 & 60 & 40 & 80 & 60 & 20 & 0 & 0 & 20 & 60 & 60 & 20 \\
0 & 40 & 20 & 40 & 80 & 20 & 60 & 20 & 40 & 80 & 60 & 0 & 0 & 60 & 40 & 80 & 100 & 0 & 80 & 0 & 20 & 60 & 60 & 20 \\
40 & 0 & 60 & 40 & 20 & 60 & 0 & 40 & 100 & 120 & 60 & 20 & 60 & 0 & 20 & 20 & 0 & 0 & 20 & 40 & 0 & 20 & 20 & 20 \\
20 & 20 & 20 & 20 & 0 & 20 & 0 & 40 & 60 & 0 & 0 & 100 & 80 & 60 & 40 & 120 & 0 & 20 & 20 & 80 & 40 & 0 & 60 & 40 \\
0 & 0 & 100 & 20 & 20 & 60 & 100 & 40 & 40 & 60 & 60 & 80 & 80 & 20 & 40 & 100 & 80 & 60 & 20 & 20 & 60 & 120 & 0 & 20 \\
40 & 60 & 60 & 100 & 80 & 0 & 0 & 20 & 0 & 60 & 40 & 20 & 100 & 100 & 20 & 20 & 40 & 0 & 20 & 0 & 20 & 60 & 20 & 0
\end{bmatrix}
$$

AON Matrix (Origin × Destination)

0	0	0	60	0	20	40	40	80	40	0	0	20	40	0	0	40	0	20	20	0	80	20	0
20	0	40	0	0	0	0	0	40	20	20	0	0	0	20	40	20	20	0	40	20	20	0	20
20	0	0	0	0	0	60	0	20	0	60	40	0	0	20	20	0	20	40	20	20	40	20	0
40	0	0	0	0	0	0	100	20	20	100	20	20	20	60	20	60	20	0	100	0	20	20	60
0	20	0	0	0	80	100	20	40	80	0	20	20	40	20	0	40	0	60	20	40	0	20	20
0	20	0	60	0	0	20	40	0	20	0	0	40	0	0	40	60	40	0	40	20	20	20	20
80	0	20	20	20	0	0	20	120	60	20	20	20	0	40	20	40	60	20	60	20	40	20	0
20	20	0	40	20	40	40	0	60	120	60	20	20	20	80	120	100	0	20	40	100	0	20	0
40	0	0	20	120	0	80	60	0	140	40	40	80	40	20	0	20	40	0	20	40	40	20	40
40	80	20	60	20	20	60	80	100	0	120	40	120	100	80	140	100	40	40	140	40	60	40	40
40	20	20	80	20	0	80	80	180	140	0	40	20	100	40	100	60	60	40	80	40	20	40	120
0	0	0	40	20	20	20	100	60	100	40	0	60	40	80	100	20	120	80	0	0	20	40	40
0	0	0	0	0	0	60	60	0	60	80	80	0	40	40	0	20	20	0	20	40	80	60	60
40	0	0	60	0	20	20	20	60	240	80	60	60	0	40	60	0	60	0	120	60	40	80	80
80	20	40	40	40	60	0	20	40	40	80	20	20	100	0	0	40	40	20	100	0	20	20	0
20	40	80	40	20	20	40	80	80	60	140	60	40	20	60	0	100	0	60	80	20	100	20	40
20	0	40	60	20	0	20	20	40	100	0	60	0	40	80	140	0	20	40	40	60	40	20	60
20	60	40	0	40	0	40	20	40	40	60	60	0	40	20	20	40	0	0	0	0	0	40	0
20	20	80	40	40	0	0	60	20	60	0	100	0	20	80	60	80	0	0	80	40	180	60	20
40	60	0	40	40	0	20	20	40	140	40	20	0	60	60	60	20	0	80	0	40	80	80	0
40	20	0	60	100	20	0	20	20	60	40	40	100	0	40	0	0	0	40	20	0	40	60	20
60	40	0	20	20	80	40	60	80	60	20	40	60	60	120	40	20	0	20	80	40	0	60	80
0	40	100	60	60	40	80	40	60	60	40	80	100	40	40	60	40	60	20	40	40	80	0	20
20	40	0	20	0	40	0	40	40	20	40	60	60	20	20	20	60	0	0	0	40	100	20	0