Dynamic OD Estimation with Bluetooth Data Using Kalman Filter

Sudeeksha Murari

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Civil Engineering

Montasir M. Abbas, Chair
Antoine G. Hobeika
Linbing Wang

August 10th, 2012
Blacksburg, VA

Keywords: Kalman Filter, Dynamic OD Estimation, QueensOD, Bluetooth Data Collection

ABSTRACT

In this thesis, Kalman Filter based dynamic OD estimation methods were explored. Dynamic OD estimation calls for updating OD flow estimates continuously based on measurements made on field. Often, the measurements available may not directly give the OD flow estimates. Kalman filter is a tool that allows us to first make a prediction of OD flow estimates and then update them based on the measurements that become available. Kalman filter is perfectly suited for online ATIS and ATMS based applications. IT can be used to make a prediction prior to the time when measurements become available, and once they are available, we can update the prediction to obtain an estimate. This estimate can then be used to make a prediction for the following time step. Three Kalman Filter methods were implemented in the work for this thesis. The first two methods (Case 1 and Case 2) were largely based on previously used methods, with modifications made to the prediction step in the Kalman Filter. In the prediction step, we have a model that can relate the estimates of the previous time step to the state variable (OD proportions) of the current time step. This model was modified to use Bluetooth OD counts as a prediction. The Bluetooth OD counts capture the traffic patterns on the network. This information can be used to supplement the measurements (link counts, exit and entry volumes).

ACKNOWLEDGEMENTS

I would sincerely like to thank my advisor, Dr. Abbas for his continuous support, guidance and patience throughout the course of my Masters. He gave me the push I required when I was completely lost and encouraged me when I was disheartened. Working with him has been a wonderful learning experience.

I would also like to thank my committee members Dr. Hobeika and Dr. Wang for their valuable inputs towards the completion of my thesis.

My parents and brother have been an undying source of inspiration, support and love. I take this opportunity to thank them for always being there.

Last but not the least, I would like to thank my lab-mates, flat-mates and all my friends in the United States and India for making the good times special and the bad times less rough.

I would like to dedicate this thesis to my parents.

Table of Contents

1 Introduction 1

1.1 Thesis Objective 2

1.2 Thesis Contribution 2

1.3 Thesis Layout 3

2 Literature Review 4

2.1 OD Estimation Methods: Development of OD estimation techniques 4

3 Comparing Kalman Filter to Synthetic OD Estimation using Bluetooth Data 7

ABSTRACT 8

3.1 Introduction 9

3.2 ATMS and ATIS 10

3.3 Bluetooth data collection 11

3.4 Dynamic OD Estimation 12

3.5 Reston Parkway Arterial 13

3.6 O-D Estimation Methodology 14

3.7 Kalman Filter formulation 15

3.8 QueensOD 20

3.9 Algorithm Structure 20

3.10 Results and Comparison with QueensOD 21

3.11 Conclusions and Further Research 26

3.12 References 26

4 Application of a Modified Kalman Filter with Bluetooth Data for OD Estimation on Reston Parkway 28

ABSTRACT 29

4.1 Introduction 30

4.2 Reston Parkway Arterial 32

4.3 Kalman Filter 33

4.4 Equal Distribution method of OD estimation 38

4.5 Experiment 38

4.6 Results and Conclusions 38

4.7 References 43

5 Conclusions 45

6 References 46

LIST OF FIGURES

Figure 1-1: Summary of Kalman Filter for dynamic OD estimation..1
Figure 3-1: The main goals of ATIS and the ATMS process. 10
Figure 3-2: Bluetooth data collection 11
Figure 3-3: Description of events in one time step. 12
Figure 3-4: Section of Reston parkway arterial considered for the study...............................12
Figure 3-5: Flowchart showing the steps of the experiment. 13
Figure 3-6: Summary of Kalman Filter 14
Figure 3-7: Prediction step of Kalman Filter. 14
Figure 3-8: Measurement update step with example. 15
Figure 3-9: Summary of Kalman Filter steps. 15
Figure 3-10: Structure of MATLAB code written to implement Kalman Filter 20
Figure 3-11: Mean square error for Case 1 with 5% penetration rate plotted for each time step (5mins) in a day. 21
Figure 3-12: The barcharts show the range of error that is seen in each method. The left plot shows results for QueensOD and the right plot is for Kalman Filter Case 1 (5% penetration rate)...21
Figure 3-14: Mean square error for Case 2 with 5% penetration rate plotted for each time step (5mins) in a day alongside QueensOD. 22
Figure 3-14: The barcharts show the range of error that is seen in each method. The left plot shows results for QueensOD and the right plot is for Kalman Filter Case 2 (5% penetration rate)...23
Figure 3-15: Plots of OD flows. Left, Case1 compared with QueensOD, right Case 2 compared with QueensOD...23
Figure 3-16: Plots of OD flows. Left, Case1 compared with QueensOD, right Case 2 compared with QueensOD...24
Figure 3-17: Plot of total square error varying with penetration rates for Case 2. 24
Figure 4-1: Figure showing the difference between Bluetooth detector location...................30
Figure 4-2: Section of Reston parkway arterial considered for the study...............................32
Figure 4-3: Kalman Filter modified for OD estimation for network covered partially by Bluetooth detectors. 33
Figure 4-4: Example of EDM. 37
Figure 4-5: Plot showing the total error varying with the penetration rates. 38
Figure 4-6: Plot of number of vehicles in the network at each time step(5mins) in a day. Kalman Filter is implemented with 5% penetration rate. 39
Figure 4-7: Plot showing the variation of mean square error as time passes in a day. The methods are implemented with 5% penetration rate. 39
Figure 4-8: Plots showing the number of time steps in a day that have error between 0-100, 100-200 and so on. The plot on the left is for Kalman Filter and the plot on the right is EDM. The methods are implemented with 5% penetration rate...40
Figure 4-9: Plot showing OD flow for each time step between OD pair 13...........................40
Figure 4-10: Plot showing OD flow for each time step between OD pair 26.........................41
Figure 4-11: Plot showing OD flow for each time step between OD pair 92.........................41

LIST OF TABLES

Table 1-1: Table showing the state variables and measurements for different cases of study. . 1
Table 3-1: Table showing the state variables and measurements for different cases of study.15

1 Introduction

Origin-Destination (OD) trip information is necessary for various transportation planning activities. The high manpower requirements and expenses to obtain this information has prompted researchers to develop methods to estimate OD matrices using models that utilize link flow or link volume information, which is more readily available. The methods developed can be broadly categorized as parameter calibration methods and matrix estimation methods. Parameter calibration uses linear or non-linear regression analysis to estimate demand assuming gravity flow pattern. These methods require zonal data, which limit their use to places where zonal data is available. Matrix estimation methods need apriori trip table information and traffic counts. This category can be further divided into statistical estimation techniques and mathematical programming methods based on maximum entropy or minimum information and other network equilibrium principles. Statistical methods give future estimates using past information by using methods like Bayesian inference, least squares estimation, etc. The mathematical programming methods assume proportional assignment where certain fractions are used to determine the proportion of trips between an OD pair. When congestion effects become prominent in a network, non-proportional assignment is assumed. In these cases equilibrium principles like that of Wardrop [1] are applicable. Some artificial intelligence (e.g., neural networks) approaches and estimation with partial link counts with bi-level programming were developed as well.

Dynamic OD estimation converts the under specified OD problem into an over specified one by utilizing information from past times to make a prediction for the current time.

Dynamic OD estimation using a technique called Kalman filter allows us to make a prediction of the OD flows in a network at the beginning of a time step. It then allows us to correct the prediction once the measurements (link counts, travel times, Bluetooth OD matrix) become available. This is a continuous process which has been proven efficient by many researchers. The Kalman Filter for dynamic OD estimation can be summarized as shown in Figure 1-1.

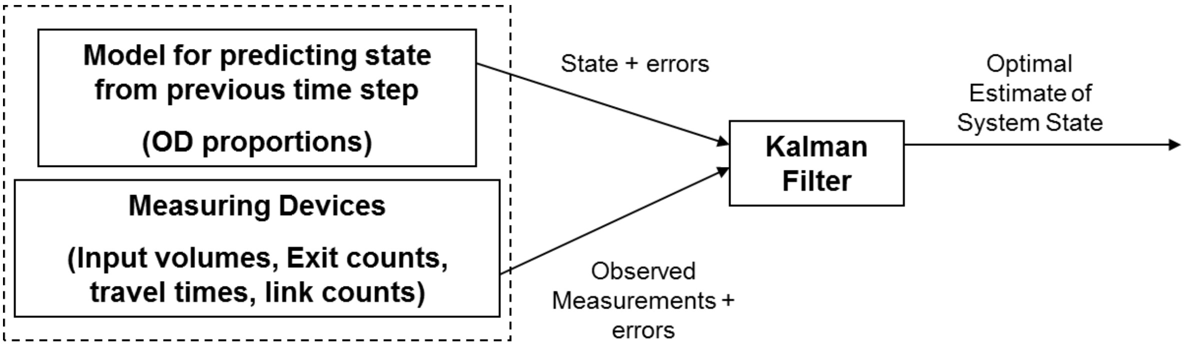


Figure 1-1: Summary of Kalman Filter for dynamic OD estimation.

1.1 Thesis Objective

The main objectives of this thesis are:

- Formulate Kalman Filter method for Dynamic OD estimation using Bluetooth data
- Compare the Kalman filter method with synthetic OD estimation methods (e.g., QueensOD)
- Compare an integrated Kalman filter/ Bluetooth-based OD estimation methods to to simple Bluetooth-based methods

In our work, we have three cases of study that are summarized in Table 1-1. The first 2 cases are discussed in the first paper. Case B is described in the second paper in this document.

Table 1-1: Table showing the state variables and measurements for different cases of study.

Case Number	State Variable	Measurements
Case 1	OD Proportions	Link counts(including entry and exit counts), Travel times from Bluetooth data
Case 2	OD Proportions	Link counts(including entry and exit counts), Travel times from Bluetooth data, Bluetooth OD matrix
Case B	OD Proportions	Link counts(including entry and exit counts), Travel times from Bluetooth data, Bluetooth OD matrix covering a part of the network

1.2 Thesis Contribution

Many dynamic OD estimation methods have been developed in the past using data from various sources like Automatic Vehicle Identification (AVI), Loop detector data, Bluetooth tracking of vehicles across the network, etc. Data from each of these sources have their own advantages and disadvantages.

This thesis focuses on using Bluetooth data collection technique and Loop detector data as the source for data for dynamic OD estimation.

Bluetooth data collection technique tracks vehicles across a roadway network using the MAC addresses of Bluetooth devices in the vehicles. When each vehicle is detected by a Bluetooth antenna installed at various locations on the network, they record the MAC address of the device detected and also the time at which it was detected.

We can extract average travel times for various Origin-destination pairs on the network. Since travel times between two locations on the network do not remain constant throughout the day, we need to have the dynamic travel times to estimate OD matrices correctly.

Bluetooth data collection has one disadvantage. Since the number of vehicles equipped with Bluetooth devices is not very high[2], it is not advisable to use this method for estimating the ODs directly. Barcelo et. al [3] showed that using Bluetooth counting of vehicles across the network and the penetration rate information for OD estimation, resulted in unacceptably high errors .

1.3 Thesis Layout

This thesis document is organized as two papers:

1. Comparing Kalman Filter to Synthetic OD Estimation using Bluetooth Data
2. Application of a Modified Kalman Filter with Bluetooth Data for OD Estimation on Reston Parkway

The first paper works with two Kalman Filter cases and compares their performance with the QueensOD as a benchmark [4].

The second paper works with a modified Kalman Filter and compares it with a simplified node-distribution method [2].

2 Literature Review

2.1 OD Estimation Methods: Development of OD estimation techniques

OD trip matrices are obtained by conducting road surveys or other surveys that involved high manpower requirements. The high cost of this approach limited the use of these techniques. The other techniques used are applicable to smaller studies and required manpower, including limited roadside interviews, flagging techniques (e.g., plate number surveys), aerial photography and "car following". O-D matrix construction can be divided into three groups:

- **direct sample estimation** of the O-D matrix by using home or destination interviews, roadside interviews, flagging techniques, etc. or combination of these surveys. The estimates are usually biased because of "non-response" or systematic measurement errors.
- **model estimation** by applying a set of models that can estimate trips made in a certain period of time.
- **traffic flow data**, where traffic counts are obtained using loop detectors and other recent technologies like Bluetooth detectors, RTMS etc. are continuously used to estimate dynamic OD matrices.

Technological advancements have enabled the utilization of techniques like RTMS, Floating car method, Bluetooth technology, GPS etc. for OD estimation. These methods will be briefed about later in this chapter.

The rest of this chapter summarizes development of OD estimation techniques by briefly describing representative works carried out by researchers.

OD estimation methods have been investigated since the early 70's. Low (1972) [5], Hogberg (1976) [6] and Holm et al. (1976)[7] put forward conventional gravity models requiring easily obtainable data, e.g. population and employment by zone along with traffic counts.

Robillard (1975)[8] estimates OD matrix using observed link volumes by using proportional assignment methods. Symons et al. (1976)[9] combine some concepts of Central Place Theory to produce a gravity type of model for intercity travel. Nguyen (1977)[10] suggests an equilibrium assignment approach but fails to find a unique trip matrix. Van Zuylen and Willumsen (1980)[11] use Information minimization and entropy maximization principles to estimate OD matrix from traffic counts. The models use full information of the counts and they can be used in cases where a gravity model assumption is not justified. Cascetta (1984)[12] put forth the Generalized Least Square estimator of OD matrix using traffic counts using an assignment model. Measurement errors and time variability of link counts are dealt with. These methods either require an apriori matrix or they are based on an assignment method which may not reflect the effects of congestion.

Some of the static OD estimation techniques were designed specifically for urban networks with multiple routes between a given OD pair. Works of Spiess(1990)[13], Florian and Chen (1995)[14], Codina and Barcelo (2004)[15] are examples of such techniques. In our work, we use a linear network, which means that there exists only one path between a given OD pair, thereby eliminating the route choice factor.

Van Aerde et.al. (2003)[4] developed the QUEENSOD software that can estimate the OD matrix without the requirement of flow continuity at a network node. Maximum likelihood estimation is carried out using Sterling's approximation. The objective function has two terms, one is the error term between observed and estimated flows and the second is the likelihood term. Objective is to find a set of OD flows which are most likely for a given set of traffic counts.

Moving on to the dynamic OD estimation methods, we summarize works that are relavant to our study.

With Maher (1983)[16], we see an introduction to dynamic OD estimation techniques. He uses Bayesian inference principle to estimate OD matrix. The prior beliefs are modified by observations to produce posterior beliefs, which are a weighted average of the prior beliefs and the observations. A prior belief, in this case, is a target trip matrix. The prior beliefs help find a unique solution, while the minimal information and maximum entropy methods allow the problem to be solved like an optimization problem to achieve the target matrix with a certain number of constraints. Though this method has limitations like the requirement of an apriori matrix, the Bayesian inference method allows us to make very good use of the obtained traffic counts. The Kalman Filter method used in our work is based on the Bayesian inference.

The work of Cremer and Keller (1987)[17] was a comparative study between a few chosen OD estimation techniques. The basic idea was that traffic flow through a network is a dynamic process with sequences of exit flows that depend on time variable sequences on entrance flows. The methods used to solve the problem included Ordinary least squares estimator with cross-correlation matrices, constrained optimization, simple recursive estimation and Kalman filtering. This comparative study allows us to see the potential of Kalman Filtering technique for dynamic OD estimation.

Nihan and Davis (1987)[18] proposed a Kalman filter based technique for OD estimation. They assumed a network with detectors present at all origins and exits. Their state variables were OD proportions and their measurements were traffic counts obtained from detectors. Their problem is very similar to the one we are trying to solve in our work. The only limitation of the work of Nihan and Davis was that they did not consider travel times, which makes their work applicable only in cases where the time interval of the Kalman filter is comparable to the maximum travel time on the network.

Similar methods were proposed by Bell(1990)[19], Van Der Zijpp and Hamerslag (1994)[20]. They are applicable only when travel times are not important in the study, which also makes them ineffective in capturing effects of congestion on a network. Chang and Wu (1994)[21] improved these methods by including the travel time in their estimation, but the relationship between their state variables (OD proportions) and their measurements (observed traffic counts) was non-linear, which calls for the application of extended Kalman filter.

With advances in technology, newer means of obtaining traffic data became available. These include GPS, Bluetooth detectors, Floating Car Data, Remote Traffic Microwave Sensors etc. The works summarized below demonstrate the use of such technology in dynamic OD estimation.

Castillo et al. (2008)[22] demonstrated that one can effectively extract OD information and path flow estimates from a wide network coverage of automated registration plate scanners. Mobile traffic sensors like onboard GPS devices and GPS navigation systems can track the path of equipped vehicles through the network. They worked with an approximate penetration rate of 5%. If the technology becomes widely available, the OD estimation will just be enumeration.

Hui et.al. (2010)[23] try to have a practical approach towards estimating time-varying OD demands incorporating both floating car data (FCD) and remote traffic microwave sensors (RTMS) data. The first stage is to obtain the static OD demands based on RTMS data for the entire modeling period to ensure that the total modeled demands match the total observed demands. The second stage is to manipulate static OD demands so that dynamic OD demands can be computed based on time-varying splitting rates extracted from FCD and RTMS data. The methodology was tested on the Western 3rd Ring-Road corridor network in Beijing. It was concluded that the method proved to be accurate and practical.

Barcelo et. al.(2010)[3] used Bluetooth detectors for tracking vehicles across a network. They use those counts in a Kalman Filter technique to obtain dynamic OD flows. Their work used OD proportions as state variables and Bluetooth detector counts as measurements. The results were good for high penetration rates (% of Bluetooth equipped vehicles on the road network), which is also a limitation since high penetration rates are generally not observed. Our work uses a similar technique with several modifications to make better use of the data from the Bluetooth detectors.

Parry and Hazelton (2012)[24] utilized link counts and sporadic routing data for their proposed methodology. It is practical to assume that some vehicles on the network have routing information readily available. In addition to that, link count data is available. The work proposes a statistical model to estimate OD matrix for the entire network. The penetration rate is known and maximum likelihood based inference for the normal models can be used.

Gharat (2011)[2] used Bluetooth data and a fixed distribution technique to obtain OD flows. The Bluetooth OD counts were simply divided between the nearest exits and entries. This method might work well in networks where the traffic flows are relatively uniform.

It can be inferred from the literature review that there have been considerable efforts made to establish a method to estimate an OD matrix. In this thesis, we will attempt to make improvements to an existing Kalman Filter method and compare it with works of Van Aerde et. al (2003)[4] and Gharat (2011)[2].

3 Comparing Kalman Filter to Synthetic OD Estimation using Bluetooth Data

By

Sudeeksha Murari
Graduate Research Assistant
Charles Via Department of Civil and Environmental Engineering
301-D, Patton Hall
Virginia Polytechnic Institute and State University
Blacksburg, VA, 24061
Phone: (540) 998-1228
E-mail: murari@vt.edu

Montasir M. Abbas, Ph.D., P.E.
Associate Professor
Charles Via Department of Civil and Environmental Engineering
301-A, Patton Hall
Virginia Polytechnic Institute and State University
Blacksburg, VA, 24061
Phone: (540) 231-9002
Fax: (540) 231-7532
E-mail: abbas@vt.edu

Submission date: August 1st, 2012

Word count: words (3,807 words for text and 3750 words for 15 Figures)

Prepared for the
Transportation Research Board 2013

ABSTRACT

Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS) utilize real-time information to apply measures improve the transportation system performance. Two key inputs for ATMS and ATIS are dynamic travel times and dynamic OD matrices. Bluetooth devices detection technology has been increasingly used to track vehicle movements on the network. This possibility naturally raises the question of whether this information can be used to improve the dynamic estimation of OD matrices. Previous research efforts rely entirely on the Bluetooth OD counts for estimation, which is why they require high penetration rates. In our study, we use Bluetooth data to supplement loop detector data while estimating dynamic OD matrices using Kalman filter. We use OD proportions as state variables and travel times, link counts, Bluetooth OD matrix and input and exit volumes as measurements. A simulation experiment is conducted in VISSIM and is designed such that the traffic network emulates the observed traffic patterns. Two case studies are performed for comparison. One uses Bluetooth OD matrices as input for estimation while the other does not. The Bluetooth ODs used in the Kalman filter estimation was found to improve the OD flow estimates. The developed methods were compared with synthetic OD estimation software (QueensOD) and were found to be more effective in obtaining dynamic OD flow estimates.

Keywords: Bluetooth data, OD Estimation, Kalman Filter, QueensOD

3.1 Introduction

Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS) utilize real-time information to apply measures improve the transportation system performance. For the proper functioning of these systems, we need to be able to accurately assess the traffic conditions and forecast how they might change in the near future (i.e., short term forecasting). Two key inputs for ATMS and ATIS are dynamic travel times and dynamic Origin-Destination (OD) matrices.

Measurements form inductive loop detectors have been used in the past for static OD estimation. The problem is underspecified since there exist more OD pairs than the number of link counts on the network. This problem is not seen in dynamic OD estimation methods, since they use time series traffic counts, where the number of equations is more than the number of OD pairs. New techniques for data collection include Automatic Vehicle Identification, License Plate Recognition, and Detection of Bluetooth Devices onboard vehicles. These technologies open up new ways of tracking vehicles on a road network.

The detection of Bluetooth devices in vehicles allows us to use their unique MAC addresses to track their movements on the network. This possibility naturally raises the question of whether this information can be used to estimate dynamic OD matrices.

Researchers have proposed many methods to perform dynamic OD estimation. Some of the important works are summarized below.

The literature can be broadly divided into two parts, one where travel times were not considered for OD estimation and the other where travel time effects were considered.

Travel time effects not considered:

Cremer and Keller (1987)[1] used a dynamic approach to the OD problem. The idea was that traffic flow through a network is a dynamic process with sequences of exit flows that depend on time variable sequences on entrance flows. They used to solve the problem included Ordinary least squares estimator with cross-correlation matrices, constrained optimization, simple recursive estimation and Kalman filtering. The potential of Kalman filtering technique was realized in this work.

Nihan and Davis (1987)[2] proposed using the Recursive Prediction Error estimator which can be interpreted as a recursive least-squares algorithm, a Kalman Filter. Recursive estimators allow tracking time-varying OD patterns. The disadvantage of this method is that it requires detectors to be present at each entry and exit.

Van der Zijpp & Hamerslag (1994)[3] used time varying OD proportions as state variables and used main section counts, exit ramp counts on a freeway to estimate dynamic OD matrix. The travel times between OD pairs were assumed fixed. This is ineffective in capturing congestion effects.

Ashok and Ben Akiva (1995)[4] used OD flow deviates from historical data as their state variables and measurements used were volumes on links and the average speeds on links. This method requires an Apriori matrix which is not always available.

Travel time effects considered:

Advances in technology allowed use of Bluetooth technology, GPS, Floating Car Data etc. to estimate OD flows. The following works could use travel time information and information about traffic patterns to estimate OD matrix.

Barcelp et. al. (2010)[5] developed Kalman Filtering methods that can use OD proportions, OD flows and deviates of OD flows from historical data as state variables and measurements were obtained from Bluetooth data collection technique(Travel times, exit volumes, main section counts). The main drawback of this method is that it requires a very high penetration rate for good performance.

Castillo et al. (2008)[6] demonstrated that one can effectively extract OD information and path flow estimates from a wide network coverage of automated registration plate scanners. Mobile traffic sensors like onboard GPS devices and GPS navigation systems can track the path of equipped vehicles through the network. They worked with an approximate penetration rate of 5%. If the technology becomes widely available, the OD estimation will just be enumeration.

Hui et.al. (2010)[7] had a practical approach towards estimating time-varying OD demands using both floating car data (FCD) and remote traffic microwave sensors (RTMS) data. The first stage was to obtain the static OD demands based on RTMS data for the entire modeling period, so as to ensure that the total modeled demands match the total observed demands. The second stage is to manipulate static OD demands so that dynamic OD demands can be computed based on time-varying splitting rates extracted from FCD and RTMS data.

In this paper we propose a method of performing dynamic OD estimation using Kalman filter. We use OD proportions as state variables and travel times, link counts, Bluetooth OD matrix and input and exit volumes as measurements. Travel times and Bluetooth OD matrix are obtained by the Bluetooth data collection technique described in Section 4.2.

3.2 ATMS and ATIS

Advanced Traveler Information System (ATIS) strives to provide information about the expected travel time that they will experience while traversing a roadway segment. Advanced Traffic Management System (ATMS) estimates current traffic state on a roadway segment and forecasts how it evolves in short term. Travel time forecasting and Dynamic OD estimation are the two key components of ATIS and ATMS. Figure 3-1 shows the main goals of ATIS and shows the flow of events for ATMS. A simplified ATMS procedure would be to collect information, process it, and dispatch relevant information to the users. This procedure helps achieve the goals of ATIS.

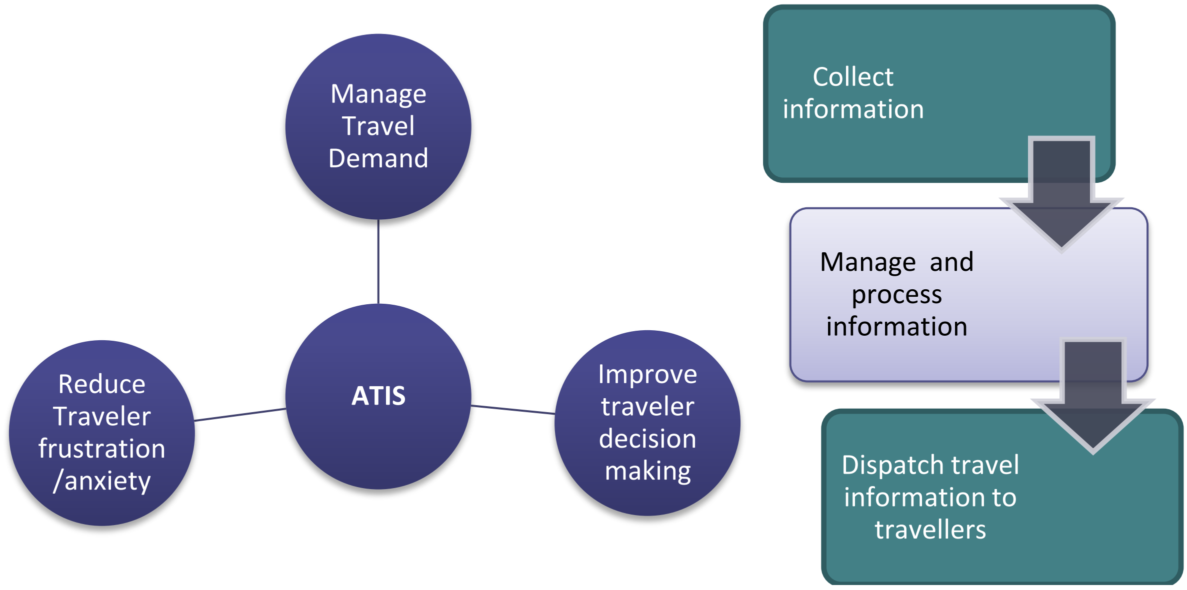


Figure 3-1: The main goals of ATIS and the ATMS process.

Along with the measurements obtained from traditional technologies like inductive loop detectors, new technologies like Automatic Vehicle Location, License Plate Recognition, detection of Bluetooth devices on vehicles, etc., can be used for dynamic OD estimation. In this paper, we explore the application of Bluetooth data collection technology for estimation of time-dependent OD matrices. The next section describes the Bluetooth data collection technology.

3.3 Bluetooth data collection

The Bluetooth data collection technology uses equipment on roadsides to detect Bluetooth devices within its coverage radius. Bluetooth is a standard protocol (IEEE 802.15.1) for information exchange between devices using the 2.4GHz radio frequency. When a Bluetooth device is detected by the Bluetooth Antenna, it captures a hexadecimal code from the Bluetooth device. This contains the MAC address of the device which is detected. Each Bluetooth device has a unique MAC address. The Bluetooth antenna also logs the time at which it captured a particular MAC address.

Figure 3-2 shows the scheme of Bluetooth data collection. The MAC addresses of detected vehicles can be used to identify them as and when they cross Bluetooth detectors located across the network. When a Bluetooth detector detects a device, it logs the time of detection, which enables us to obtain the travel time between any two Bluetooth detectors that a vehicle has crossed. Since vehicle movements are tracked across the network, their entry and exit locations can be identified. Thus, we can develop an OD matrix, which will be referred to as the Bluetooth OD matrix and the OD flows will be referred to as the Bluetooth OD counts or Bluetooth OD flows.

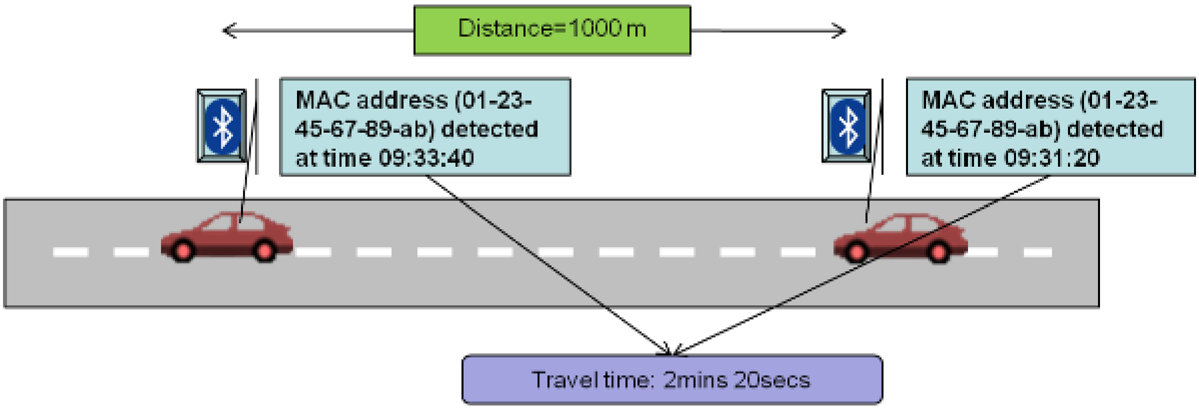


Figure 3-2: Bluetooth data collection

While using the Bluetooth data collection technique, we need to be mindful of the fact that not all vehicles on the road are equipped with Bluetooth devices. The ratio of the number of Bluetooth equipped vehicles to the total number of vehicles on the roadway is known as the Penetration Rate (***PR***). The formula for obtaining penetration rate is shown in Equation (3-i)

(3-i)

An added advantage of the Bluetooth data collection technique is that it doesn't invade the privacy of the drivers. The MAC addresses that are captured by the Bluetooth detectors do not have any personal information about the owner of the device, which ensures that the privacy of the drivers is respected.

3.4 Dynamic OD Estimation

The Bluetooth data collection technique described in Section 3.2 naturally raises the question: Is it possible to estimate OD matrices by Bluetooth data collection technique?

Barcelo et. al (2010)[5] performed a study which concluded that a simple counting of vehicles detected by the Bluetooth detectors to generate an OD matrix can lead to unacceptably high errors when the penetration rates are low.

However, we can extract a wealth of information from the Bluetooth data collection technique, such as:

- Travel times between Origins and Destinations
- Bluetooth (initial) OD matrix

Along with the Bluetooth data, we have data from the loop detectors on the road network which provide us with link counts.

For the purpose of this study we divide a day (24hrs) into 288 time steps of 5 min duration. We intend to perform OD estimation for each time step, i.e. the OD matrix is updated every 5 minutes. Figure 3-3 summarizes the tasks carried out in every time step.

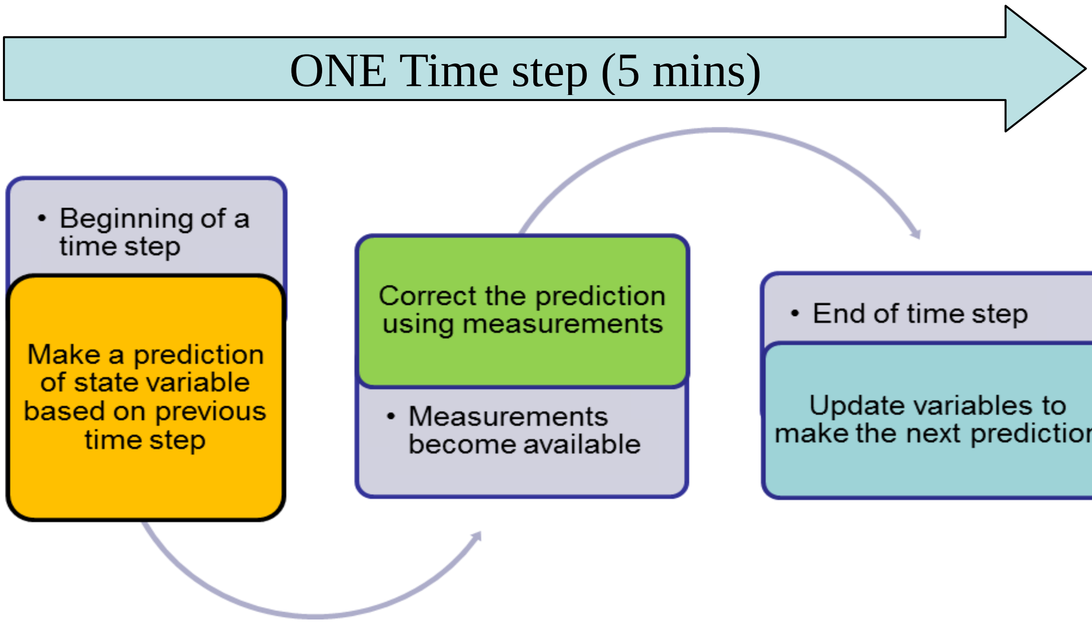


Figure 3-3: Description of events in one time step.

3.5 Reston Parkway Arterial

We considered a section of Reston Parkway for our study. Reston Parkway is located in the Fairfax County in the state of Virginia, USA. The considered network has a total of 5 intersections as shown in Figure 3-4. The speed limit for the main arterial is 45 mph, and it ranges between 15 mph and 45 mph for the side streets.

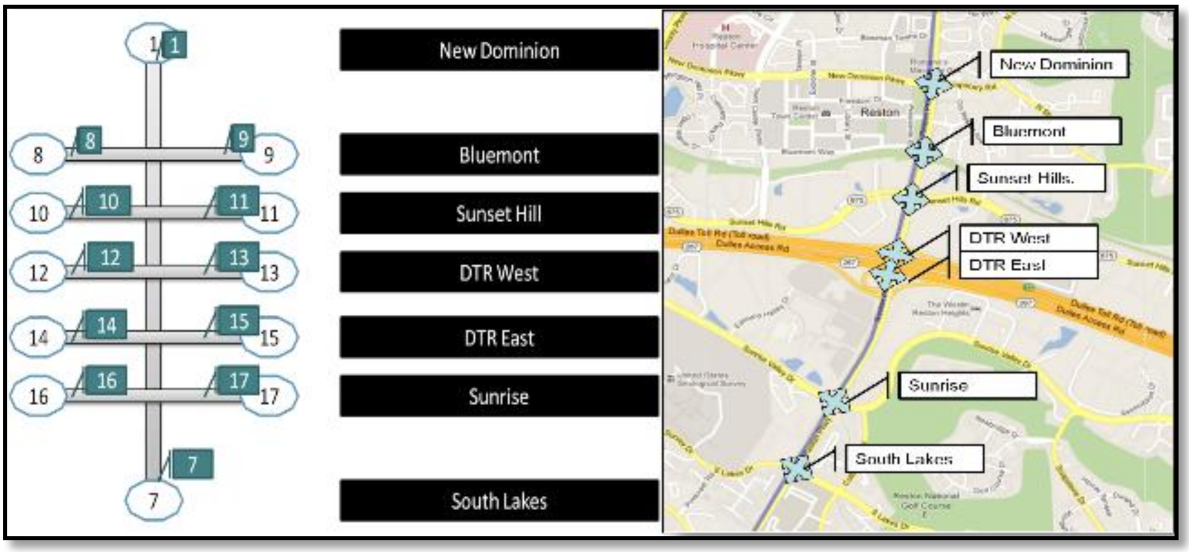


Figure 3-4: Section of Reston parkway arterial considered for the study.

Figure 3-4 (left) shows the schematic of Reston Parkway considered for the study. There are 10 origins in the network and 10 destinations. Nodes 12 and 15 are destinations only; nodes 13 and 14 are origins only; the rest of the nodes are both origins and destinations. The blue boxes indicate the location of the Bluetooth detectors with the Bluetooth detector numbers in the boxes and the white ovals contain entry and exit numbers. There is no route choice since there is only one possible route between any given OD pair. Bluetooth detectors are assumed to be placed at every origin and destination in this hypothetical case.

3.6 O-D Estimation Methodology

For the purpose of the Kalman filter verification study, we consider the network shown in Figure 3-4. This network is assumed to have Bluetooth detectors at every entry and exit. The loop detector data is used for obtaining the number of vehicles entering the network at every entry, the number of vehicles leaving at every exit, the number of vehicles on each link on the network at any given point of time.

The simulation experiment is conducted for a duration of 24hrs. Each time step is 5 mins long and the OD matrix is updated every 5mins. The Simulation is carried out in VISSIM. The experiment is designed such that the traffic network emulates the observed traffic patterns. The simulation experiments allowed us to treat a randomly identified vehicle as a vehicle equipped with Bluetooth device and then track its movements across the network.

A flowchart presented in Figure 3-5 shows the steps followed in conducting the simulation experiment and analyzing the results. Once the simulation was completed, we could extract the actual OD patterns on the network using data collection points on the network. The number of vehicles treated as Bluetooth equipped vehicles was determined by the assumed penetration rate. The travel times of these vehicles are computed for each time step. Along with the travel time information, we can also obtain an OD matrix for the Bluetooth equipped vehicles, which will be referred to as the Bluetooth OD matrix in this paper.

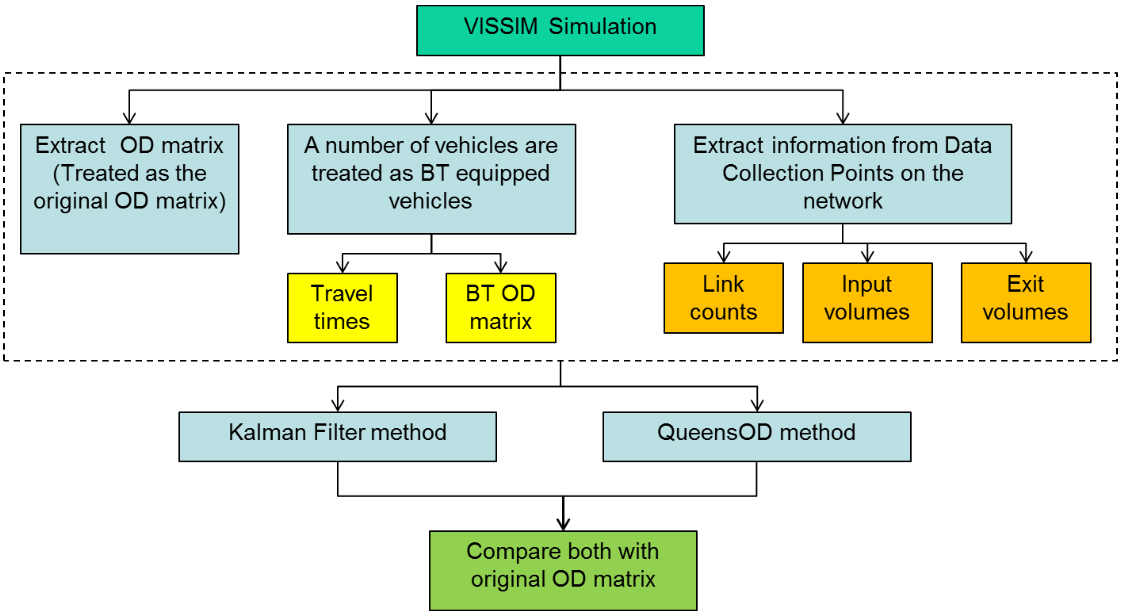


Figure 3-5: Flowchart showing the steps of the experiment.

The Bluetooth OD matrices are created from VISSIM data such that it emulates the true OD patterns. We assume that the penetration rate all across the network is uniform and doesn't change throughout the day.

3.7 Kalman Filter formulation

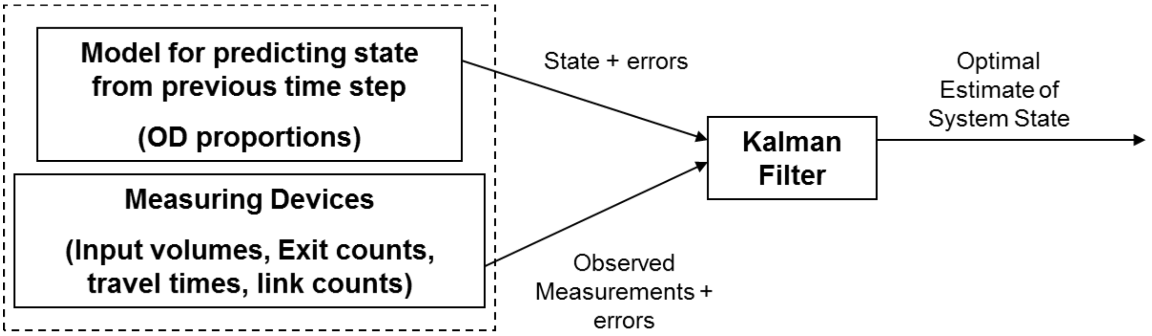


Figure 3-6: Summary of Kalman Filter

The Figure 3-6 gives a brief outline of the Kalman Filter. Kalman Filter is a tool that helps us correct a set of predictions made based on a model. We use weights to decide how much the predictions need to be modified based on the obtained measurements. In our case, we are dealing with OD proportions as the state variables. Our predictions are based on the OD proportions from the previous time step. The measurements are the Bluetooth OD counts that we obtain at each time step. Since Bluetooth OD counts cannot be used directly when the penetration rates are low (Barcelo et. al.(2010)[5]), we use Kalman filter method to incorporate those measurements to improve our predictions and therefore obtain a more accurate estimate.

The Kalman Filter formulation consists of two main steps:

- Prediction step

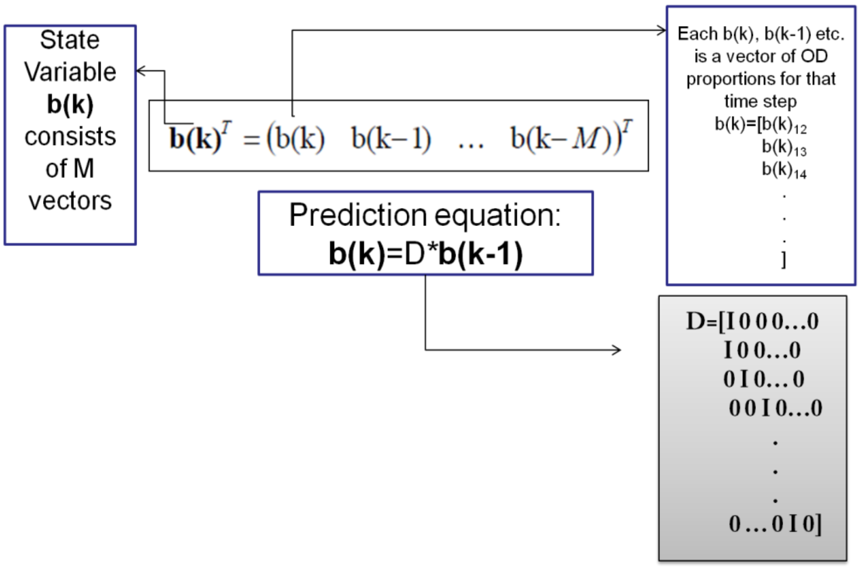


Figure 3-7: Prediction step of Kalman Filter.

The prediction step shown in Figure 3-7 is the first step in Kalman Filter. Here ***b(k)*** is the state variable which is itself a set of OD proportions belonging to time step *k, k-1…k-M*. *M* is the maximum number of time steps required by a vehicle to traverse the network. Each ***b(k)*** is assumed to be linearly related to its previous time step. *D* is the transition matrix that relates ***b(k)*** and ***b(k-1)***.

- Measurement update step

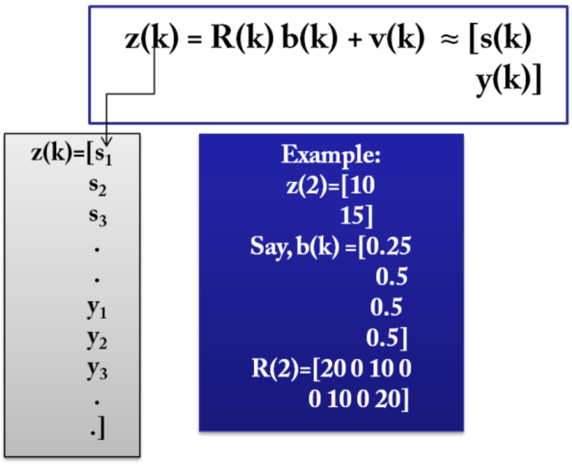


Figure 3-8: Measurement update step with example.

In the measurement update step, described in Figure 3-8 shows the relationship between measurements *z(k)* and state variable *b(k)*. *R(k)* is a matrix constructed with input volumes such that, when multiplied with the OD proportions, i.e. *b(k)*, the resulting matrix is *z(k)* which contains exit volumes *s1, s2* etc. and link counts *y1, y2* etc.

Values in *z(k)* are treated as measurements. When there is a difference between the measurements in *z(k)* and the predicted exit volumes and link counts given by *R(k)*b(k), v(k)* is the residual.

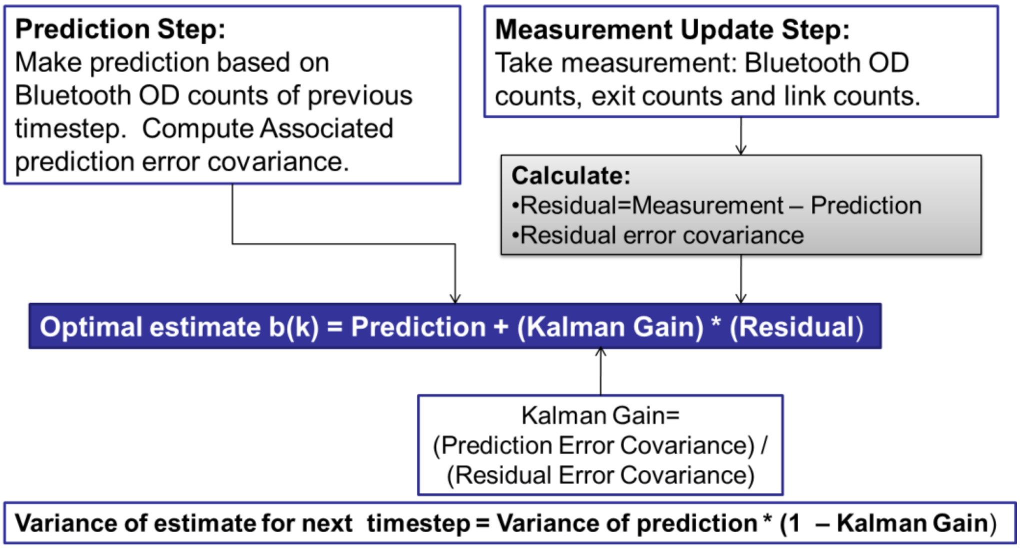


Figure 3-9: Summary of Kalman Filter steps.

Figure 3-9 describes the various calculations involved in the Kalman Filter.

In this paper, we have two cases of study, summarized in Table 3-1.

Table 3-1: Table showing the state variables and measurements for different cases of study.

Case Number	State Variable	Measurement
Case 1	OD Proportions	Link counts(including entry and exit counts), Travel times from Bluetooth data
Case 2	OD Proportions	Link counts(including entry and exit counts), Travel times from Bluetooth data, Bluetooth OD matrix

The Case 2 differs from Case 1 in the prediction step of the Kalman filter. In Case 2, the Bluetooth OD proportions from the previous time step is used as the prediction for the current time step for which we are estimating the OD matrix.

State variable

State variable for the Kalman filter in our study is a vector of OD proportions. Each element in this vector is the proportion of flow from an origin **i**, exiting at a destination **j** in time step **k,** represented by *b(k)* in Equation (3-ii)

(3-ii)

(3-iii)

$\boldsymbol{b_k}$ is a vector of b(k), b(k-1)… to b(k-m), where, m is the maximum number of time steps taken by any vehicle to traverse the network.

Prediction step

CASE 1

A prediction of the state variable is made for a given time step k based on estimates of the state variables from the previous time steps. The equations (3-iv), (3-v), (3-vi) and (3-vii) are used in the prediction step.

(3-iv)

(3-v)

(3-vi)

(3-vii)

Where,

w is white noise with an expected value of zero

is the predicted state variable

is the estimate of the state variable from the previous time step

is the prediction error covariance matrix

is the prediction error covariance matrix for the previous time step *m* is the maximum number of time steps any vehicle can take to traverse the network

n is the total number of time steps of 5min duration in a day

W is a matrix given by and is of size (m+1)*n n*n

D is a transition matrix of size (m+1)*n n*n given by

I is an identity matrix of size n n given by

CASE 2

Case 2 differs from Case 1 in the prediction step. We make use of the Bluetooth OD data collected using Bluetooth detectors to improve our prediction in Case 2. The Bluetooth OD counts collected at time step ***k-1*** are used to calculate the prediction of OD proportions for time step ***k*.** The prediction equation (3-vi) changes to (3-viii). By using this equation, we are providing more information to the Kalman Filter about how the OD patterns are evolving.

(3-viii)

is the matrix with Bluetooth OD counts from time step ***k-1***.

This is an improvement over the method used by Barcelo et. al.(2010)[5] who used the Bluetooth data collection technique and Kalman Filter to estimate OD proportions from the collected data. The improved method implemented in Case 2 makes use of travel time information as well as the Bluetooth OD counts to obtain a better estimate of OD proportions.

Another advantage of this method is that it performs well even with very low penetration rates. The results section of this paper illustrates this fact.

Measurement update step

Measurement vector z_k is given by Equation (3-ix).

(3-ix)

Where,

s(k) is a vector of exit volumes and *y(k)* is a vector of link counts for time step k.

The residual ε_k is computed using Equation (3-x).

(3-x)

Where,

is the residual

is the measurement matrix

is a transition matrix of dimension (m+1)*n (m+1)*n constructed such that it satisfies Equation (3-xi).

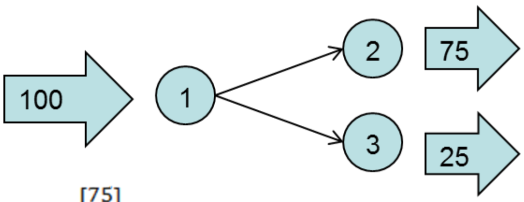
(3-xi)

(3-xii)

Where, r is assumed to be white noise with an expected value of zero. If our prediction b_k is accurate, r will be zero.

consists of input volumes, which when multiplied with the corresponding OD proportions in b_k and summed up, give exit volumes and link counts.

Example:



$$z_2 = \begin{bmatrix} 75 \\ 25 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 0.75 \\ 0.25 \end{bmatrix} \qquad R_2 = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$$

Kalman gain is computed using the following equation:

(3-xiii)

Where,

G is the Kalman gain

(3-xiv)

R is the measurement covariance matrix.

The predicted state variable is then updated using the following equation:

(3-xv)

At the step where the prediction is updated in the code written in MATLAB, a MATLAB function called ‘lsqlin’ is used to solve for by minimizing the difference between the left hand side and the right hand side of equation (3-xv) and also applying the constraints that require OD proportions summed over a common origin should total to 1. This method ensures that there is no overestimation of vehicles in the network. We also define a lower-bound value of zero for all OD proportions in the ‘lsqlin’ solver to avoid negative estimates.

The prediction error covariance is updated using Equation (3-xvi).

(3-xvi)

3.8 QueensOD

The results of the experiment conducted are compared with results from QueensOD software developed by Van Aerde Associates. QueensOD is a model developed by Van Aerde et.al. (2003)[8] that can estimate an OD matrix from link counts, without the requirement of flow continuity at network nodes. Maximum Likelihood OD estimation is carried out using Sterling’s approximation. The objective function has two terms, one is the error term between observed and estimated flows and the second is the likelihood term. The objective is to find a matrix with maximum likelihood of the values of OD flows for a given set of link counts while minimizing the error.

QueensOD method was used as a benchmark for comparison because it allows us to clearly see the effects of incorporating travel time information and some amount of traffic flow information. QueensOD belongs to a class of OD estimation methods that used neither of these.

3.9 Algorithm Structure

The implementation of the Kalman Filter is done in MATLAB. The basic structure of the code is as shown in Figure 3-10. The data extracted from VISSIM simulation is compiled into excel files. The Excel files contain entry volumes, exit volumes, link counts, travel times and an OD matrix that is treated as the original matrix. The QueensOD software was used to obtain OD matrices using link counts, against which Case 1 and Case 2 are compared.

The code was written in MATLAB because it can handle large matrices that we encounter while implementing Kalman Filter. MATLAB also allows us to repeat the procedure many times, which lets us compute the OD matrix for an entire day with 5min time steps.

The next section presents the results of Kalman Filter method compared with QueensOD.

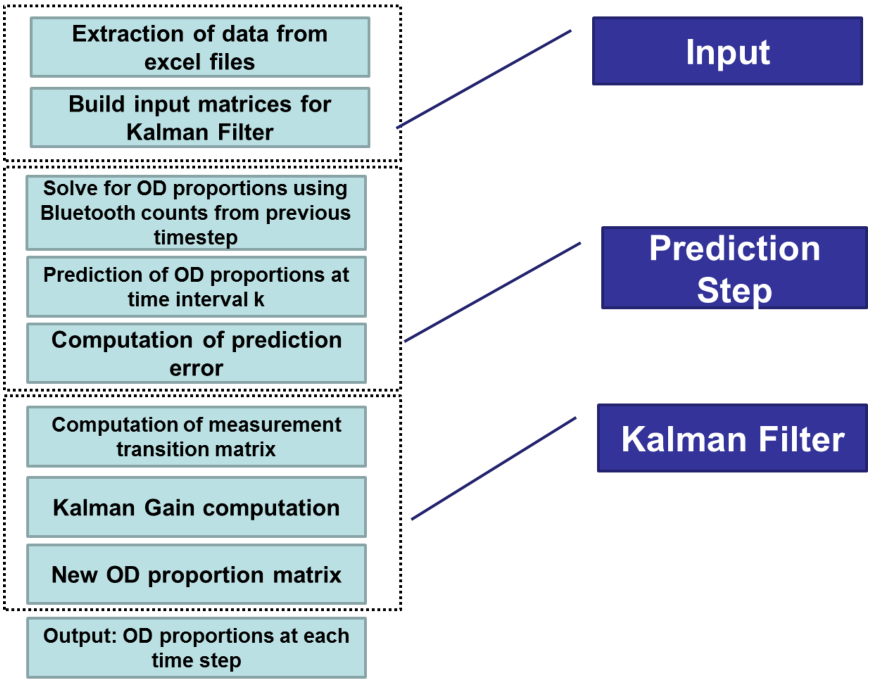


Figure 3-10: Structure of MATLAB code written to implement Kalman Filter

3.10 Results and Comparison with QueensOD

The Kalman Filter cases were compared with QueensOD. The Cases were then compared with each other to assess their performance.

The results are organized as follows:

- Case1 with penetration rate 5%
    - Mean square error for each time step summed over all OD pairs
    - Frequency of error showing the number of time steps that have mean square errors falling in a given range
- Case 2 with penetration rate 5%
    - Mean square error for each time step summed over all OD pairs
    - Frequency of error showing the number of time steps that have mean square errors falling in a given range
- Plots of OD flows for Case 1 and Case 2 for penetration rate 5%
- Plot of total error for Case 2 with varying penetration rates

Results for Case 1 compared with QueensOD (penetration rate 5%)

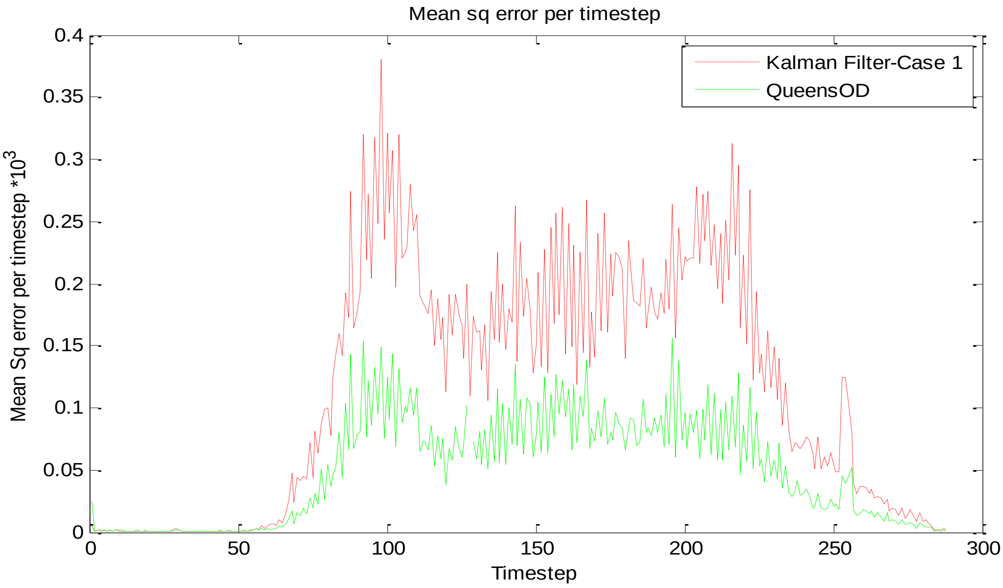


Figure 3-11: Mean square error for Case 1 with 5% penetration rate plotted for each time step (5mins) in a day.

The mean square error is calculated by taking the average of square errors across all OD pairs in the network for a given time step. The mean square error plot (Figure 3-11) for Case 1 shows that the Kalman Filter method is performing reasonably well with acceptable errors. The QueensOD plot has lesser error as compared to the Kalman Filter in Case 1.

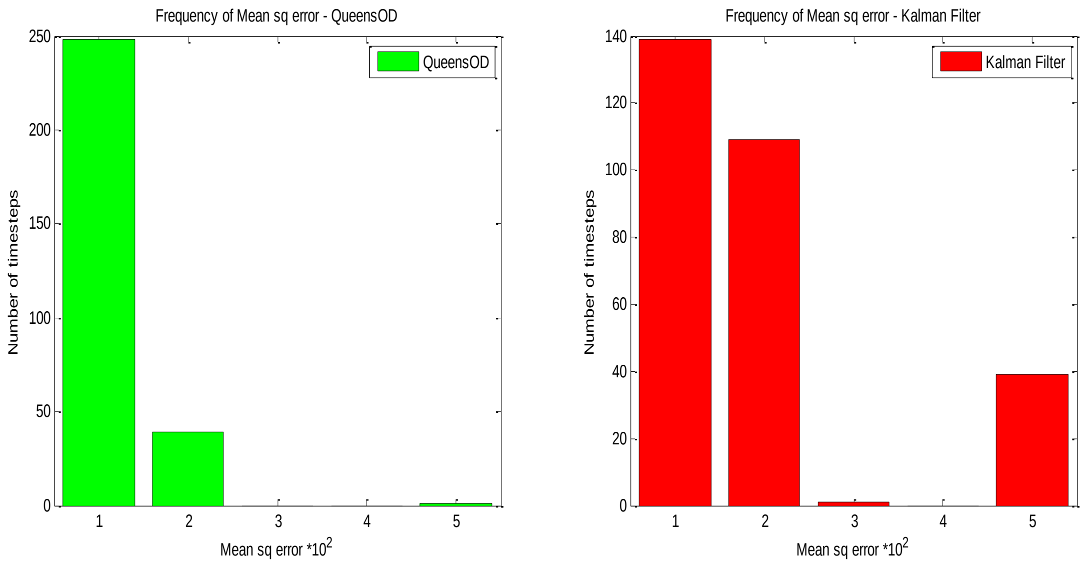


Figure 3-12: The barcharts show the range of error that is seen in each method. The left plot shows results for QueensOD and the right plot is for Kalman Filter Case 1 (5% penetration rate).

The frequency of error plots (Figure 3-12) tell us that the Queens OD has fewer time steps with mean square errors greater than 100 and the Kalman Filter Case 1 has more points with errors greater than 100.

Results for case 2 compared with QueensOD (Penetration rate 5%)

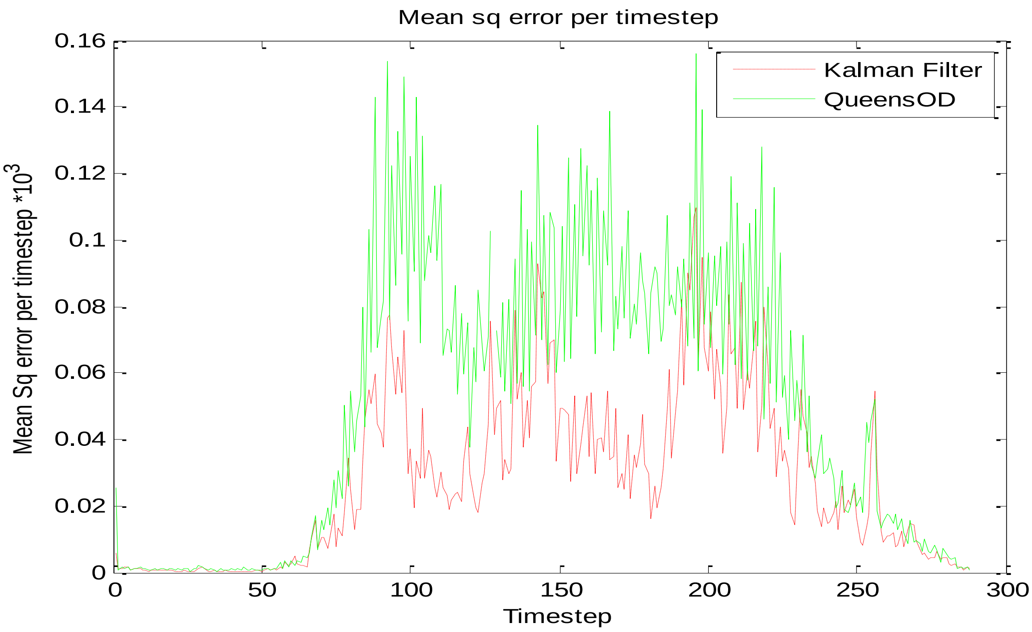


Figure 3-14: Mean square error for Case 2 with 5% penetration rate plotted for each time step (5mins) in a day alongside QueensOD.

The Mean square error plot (Figure 3-14) for Case 2 suggests that the error reduces greatly in Case 2 as compared to Case 1(Figure 3-11). We can conclude that using Bluetooth OD counts in the prediction step of the Kalman Filter helps it get a better estimate of OD counts.

The frequency of error plots (Figure 3-14) tell us that the Kalman Filter Case 2 has fewer time steps with mean square errors greater than 100 and the QueensOD has more points with errors greater than 100.

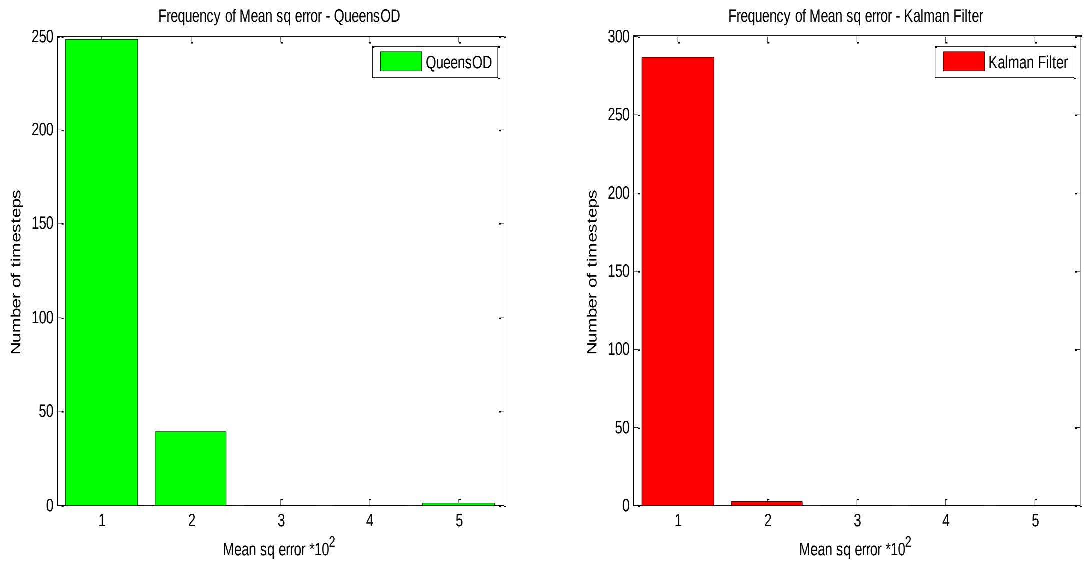


Figure 3-14: The barcharts show the range of error that is seen in each method. The left plot shows results for QueensOD and the right plot is for Kalman Filter Case 2 (5% penetration rate).

OD pair 3 (5% penetration rate)

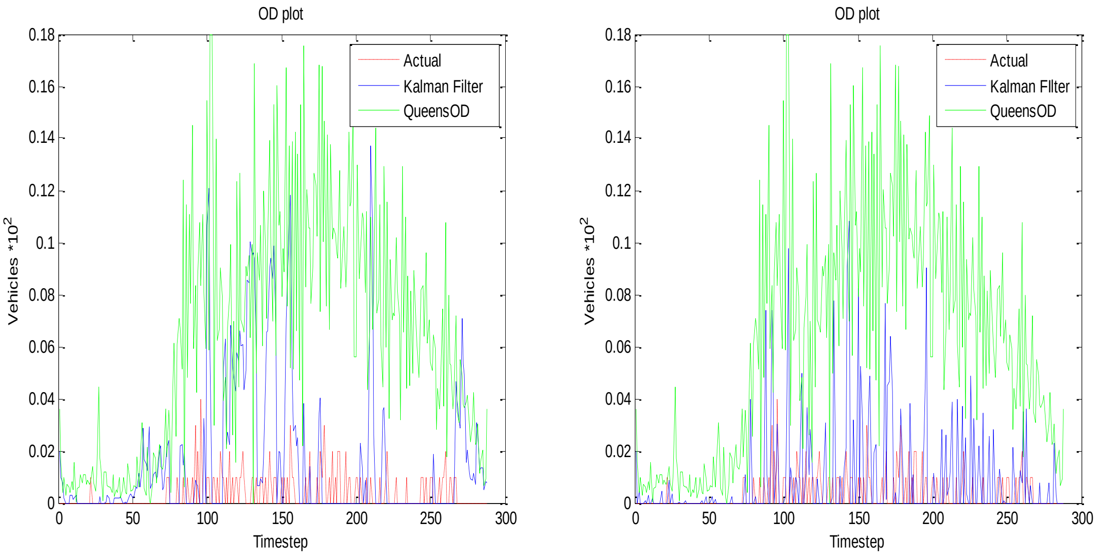


Figure 3-15: Plots of OD flows. Left, Case1 compared with QueensOD, right Case 2 compared with QueensOD

OD pair 21 (5% penetration rate)

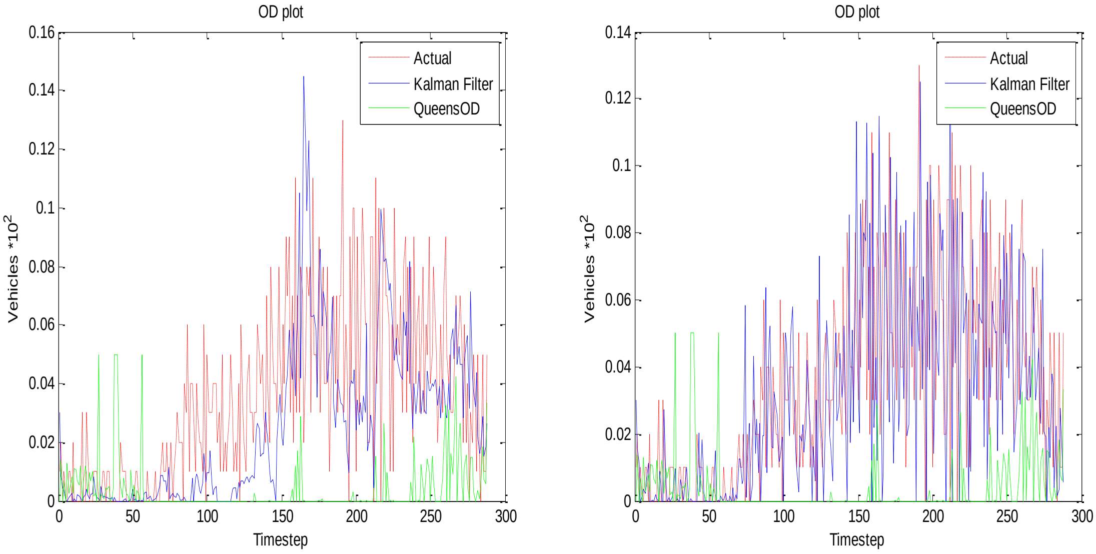


Figure 3-16: Plots of OD flows. Left, Case1 compared with QueensOD, right Case 2 compared with QueensOD

The OD plots shown in Figure 3-15 and Figure 3-16 further reaffirm the fact that Case 2 performs better than Case 1. The left plots in each figure are from Case 1 where the Kalman filter and QueensOD both do poorly in capturing the OD flow pattern through the day. The right plots in the figures show that Kalman Filter does a better job of capturing the OD flow pattern than QueensOD. This further supports our conclusion that using Bluetooth OD counts with the Kalman Filter helps it perform better than without it.

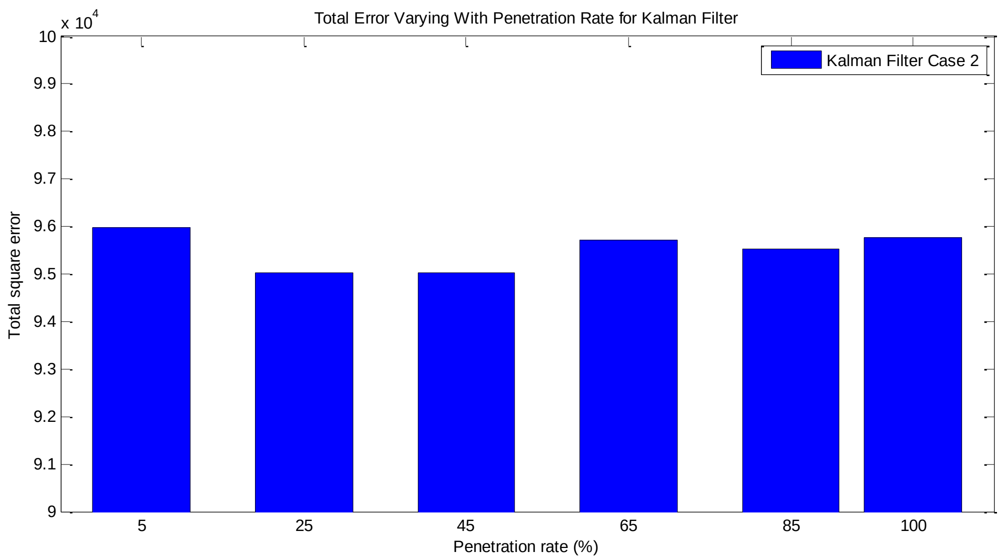


Figure 3-17: Plot of total square error varying with penetration rates for Case 2.

Figure 3-17 shows that there is not much variation of total error with the penetration rates. This is because the Bluetooth OD matrix that we chose has uniform penetration rates all across the network, which means that the Bluetooth OD matrix was a good representation of the actual OD matrix. Thus, penetration rates did not significantly affect the total error. In case of non-uniform penetration rates, we could expect the total error to decrease as the penetration rates increase. This aspect needs to be researched further.

3.11 Conclusions and Further Research

We studied two cases of Kalman Filter in our study. The first one uses Bluetooth data to obtain travel times and uses entry volumes, exit volumes and link counts to obtain estimates of OD flows. The second case uses the Bluetooth OD counts as the predicted OD flows instead of using the values from the previous time step as the predicted OD flows. This helps capture the traffic patterns much more accurately.

Case 2 clearly performs better than Case 1. Case 2 also performs better than QueensOD. The Bluetooth OD counts used as prediction in the Kalman filter improves the estimates of the Kalman filter. The previous research efforts rely on the Bluetooth counts for OD estimation (Barcelo et. al. 2010[5]), which is why they require high penetration rates. In our study, we use Bluetooth data to supplement loop detector data while estimating dynamic OD matrices. When we worked with a penetration rate of 5%, the performance of the method was acceptable. The developed method is thus proven to work well with low penetration rates.

Case 1 does not use the Bluetooth counts directly in OD estimation. It uses the entry volumes, exit volumes and link counts along with travel time obtained using Bluetooth data to estimate OD flows. Case 2 has an advantage over case 1 because the Bluetooth OD counts are used as predictions. It can be seen clearly in the results that the Case 2 captures variations in OD flows much better than Case 1.

In the Measurement Update step, we incorporate a constraint that ensures that the OD proportions associated with a particular origin sum to 1. This keeps the method from over estimating vehicles in the network. It is one of the reasons why OD proportions as state variables is a better choice than OD flows as state variables.

Future research should focus on networks where the Bluetooth detectors are not present at every entry and exit.

3.12 References

1. Cremer, M. and H. Keller, *A New Class of Dynamic Methods for the Identification of Origin-Destination Flows.* Transportation Research, 1987. **21B(2)**: p. 117-132.
2. Nihan, N.L. and G.A. Davis, *Recursive estimation of origin-destination matrices from input/output counts.* Transportation Research Part B: Methodological, 1987. **21**(2): p. 149-163.
3. Zijpp, N.J.V.d. and R. Hamerslag, *Improved Kalman Filtering Approach For Estimating Origin-Destination Matrices For Freeway Corridors.* Transportation Research Record, 1994. **1443**: p. 54-64.
4. Ashok, K., & Ben-Akiva, M. E. (1995). Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin–Destination Flows. *Transportation Science, 34*(1), 21-36.

5. Barcelo, J., Lidin, M., Laura, M., & Carlos, C. (2010). Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring. *Transportation Research Record: Journal of the Transportation Research Board*, *2175/2010*, 19-27.
6. Castillo, E., J.M. Menéndez, and S. Sánchez-Cambronero, *Traffic Estimation and Optimal Counting Location Without Path Enumeration Using Bayesian Networks.* Computer-Aided Civil and Infrastructure Engineering, 2008. **23**(3): p. 189-207.
7. Hui, Z., et al., *Estimation of Time-Varying OD Demands Incorporating FCD and RTMS Data.* Journal of Transportation Systems Engineering and Information Technology, 2010.
8. Aerde, M.V., H. Rakha, and H. Paramahamsan, *Estimation of Origin-Destination Matrices: Relationship Between Practical and Theoretical Considerations.* Transportation Research Record: Journal of the Transportation Research Board, 2003. **1831/2003**: p. 122-130.

4 Application of a Modified Kalman Filter with Bluetooth Data for OD Estimation on Reston Parkway

By

Sudeeksha Murari
Graduate Student
Charles Via Department of Civil and Environmental Engineering
301-D, Patton Hall
Virginia Polytechnic Institute and State University
Blacksburg, VA, 24061
Phone: (540) 998-1228
E-mail: murari@vt.edu

Montasir M. Abbas, Ph.D., P.E.
Associate Professor
Charles Via Department of Civil and Environmental Engineering
301-A, Patton Hall
Virginia Polytechnic Institute and State University
Blacksburg, VA, 24061
Phone: (540) 231-9002
Fax: (540) 231-7532
E-mail: abbas@vt.edu

Keywords: Bluetooth data, Origin Destination, Kalman Filter, VISSIM

Submission date: August 1st, 2012

Word count: 6605 words (3355 words for text and 3250 for 13 Figures)
Prepared for the
Transportation Research Board 2013

ABSTRACT

Several problems are usually encountered while trying to estimate dynamic OD flows for a given roadway network. Travel times of vehicles, congestion effects and variation of flows with time of day are few examples. Dynamic OD estimation methods have seen considerable improvements over the past decade. Some of this improvement can be attributed to the fact that the technological developments have allowed us to measure more parameters for OD estimation. Travel times can be measured using Bluetooth data collection technology. It is infeasible to place Bluetooth detectors at every entry and exit of the network under study. In this paper, we present an improvement to a Kalman Filter OD estimation method proposed by Murari and Abbas (Under Review) and apply it to Reston Parkway network in Virginia. The purpose of the study is to be able to estimate OD flows for networks when only a portion of the network has travel time information and traffic pattern information available. The developed method is compared with a synthetic OD estimation method developed by Gharat (2011)[1]. The results show that the proposed method performs reasonably well even for low penetration rates.

Keywords: *Bluetooth data, Origin Destination, Kalman Filter, VISSIM*

4.1 Introduction

Dynamic OD estimation methods have been developed over the years to provide accurate OD matrix estimates based on available data. Earlier on data for OD estimation was limited to loop detector counts or video detector counts at intersections. With the advent of newer technologies, other measurements have become available to estimate OD flows.

OD flows are dependent on link flows, travel times, presence of congestion, entry volumes, exit volumes and the distribution pattern. Link flows are generally obtained from loop detectors or video detectors installed at intersections. Similarly, we can obtain entry volumes and exit volumes. Travel times, congestion effects and distribution patterns are not easily discernible. In the following section we review a few relevant efforts that make use of recent technological advances for OD estimation.

Bluetooth Data Collection:

Bluetooth data collection technique involves tracking of vehicles across a network by tracking their MAC address. The Bluetooth detectors located at different points on the network pick up the MAC addresses of the vehicles that pass them. Two detectors that pick up the same MAC address form a Bluetooth OD pair. Counting those vehicles which travel between Bluetooth detectors allows us to estimate a Bluetooth OD matrix.

Barcelp et. al. (2010)[2] developed Kalman Filter based methods that can use OD proportions, OD flows and deviates of OD flows from historical data as state variables and measurements were obtained from Bluetooth data collection technique(Travel times, exit volumes, main section counts). The main drawback of this method is that it requires a very high penetration rate for good performance.

Gharat (2011)[1] used an OD estimation method where the Bluetooth counts of vehicles were used to develop an OD matrix using a simple distribution of vehicles counted at any given Bluetooth detector.

Murari & Abbas (Under Review) used a modified Kalman Filter to estimate OD matrices from Bluetooth data. The location of the Bluetooth detectors was such that the whole roadway network under analysis was covered. In this paper we deal with a network where the Bluetooth detectors do not cover the whole network.

GPS Data:

GPS based data collection involves tracking of vehicles that have an onboard GPS device. The penetration rates available are low, but the prospect of estimating OD matrices is promising.

Castillo et al. (2008)[3] demonstrated that one can effectively extract OD information and path flow estimates from a wide network coverage of automated registration plate scanners. Mobile traffic sensors like onboard GPS devices and GPS navigation systems can track the path of equipped vehicles through the network. They worked with an approximate penetration rate of 5%. If the technology becomes widely available, the OD estimation will just be enumeration.

Floating Car Data and RTMS:

Hui et. al. (2010)[4] had a practical approach towards estimating time-varying OD demands incorporating both floating car data (FCD) and remote traffic microwave sensors (RTMS) data. The first stage was to obtain the static OD demands based on RTMS data for the entire modeling period to ensure that the total modeled demands match the total observed demands. The second stage was manipulating static OD demands so that dynamic OD demands could be computed based on time-varying splitting rates extracted from FCD and RTMS data. This is a slightly more expensive method and is less feasible than the other listed methods, but is more reliable.

Cell Phones:

Friedrich et. al. (2011) [5] presented a study that tracked cell phone trajectories of cell phones in a study area. Based on these trajectories, the origins and destinations of the vehicles were obtained. Link counts were then used to obtain OD counts with the recorded cell phone trajectory patterns. The biggest disadvantage of this method is that it invades the privacy of the users. The cell phones are associated with personal information of the owner, unlike the Bluetooth devices that are associated with MAC addresses that do not contain any personal information about the user.

AVI and Probe Vehicles:

Kwon and Varaiya (2006) used an Automatic Vehicle Identification (AVI) technique and applied a statistical method to obtain OD matrix estimate. They used the Electronic Toll Collection tags to obtain partial trajectories of vehicles to as a source of data to obtain dynamic OD matrices. This method is applicable only to a study area with routes that have tolls.

Nanthawichit et.al. (2007) used Probe vehicle data for dynamic OD estimation. A Kalman filter method was developed to integrate Probe vehicle data for OD estimation. The method performed well with the probe vehicle data than without it. Travel times were then obtained by using the speeds. Travel time forecasting using the predictions of the developed method performed well compared to other travel time forecasting methods. Congestion effects might affect the travel time forecasting using this method. This is an advantage Bluetooth data has over probe vehicle data. The Bluetooth technique measures travel times rather than estimating them.

Since vehicle movements can be tracked on the network using Bluetooth technology, we can obtain counts of vehicles travelling between any two Bluetooth detectors. These counts are referred to as the Bluetooth OD counts or the Bluetooth OD matrices. It is to be noted that the Bluetooth OD counts correspond to vehicles travelling between Bluetooth detectors and not the actual entries and exits in the network.

In this paper we propose a method of performing dynamic OD estimation using Kalman filter. We use OD proportions as state variables and travel times, link counts, Bluetooth OD matrix and input and exit volumes as measurements. The developed method builds on the Kalman Filter method proposed by Murari and Abbas (Under Review). The improvements made to their model are to deal with networks where the Bluetooth detectors are not located at every entry and exit.

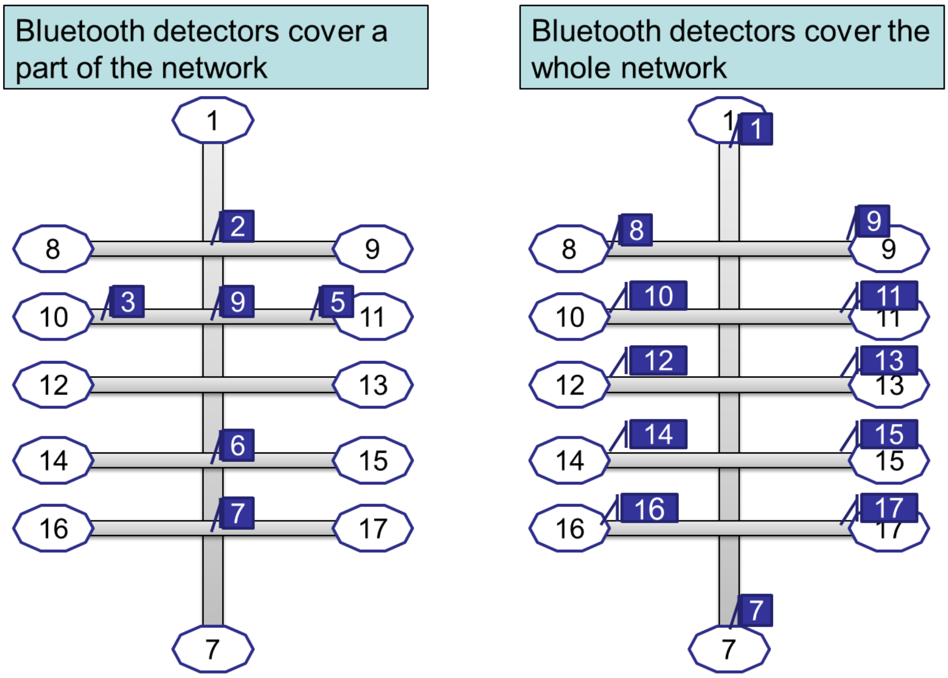


Figure 4-1: Figure showing the difference between Bluetooth detector location.

The Figure 4-1 shows the difference in location of Bluetooth detectors on a roadway network, the blue markers are the Bluetooth detectors and the white ovals represent entry-exit locations. The network on the left shows the case where every entry and exit does not have a detector. The network on the right has Bluetooth detectors at every exit and entry on the network. The network with partial coverage poses an additional difficulty where it becomes harder to determine as to how much flow between an OD pair contributes to the Bluetooth OD count between 2 Bluetooth detectors. For example, if we consider the network on the left in Figure 4-2, the flows between Bluetooth OD pair 2-7, constitutes flows between OD pairs 1-7, 1-16, 1-17, 8-7, 8-16, 8-17, 9-7, 9-16 and 9-17. It is hard to tell which one of those contributes more or which contributes less to the Bluetooth OD count 2-7.

Another factor that affects the Bluetooth data collection method is that not all vehicles on the road will have Bluetooth devices on them. Some might have it turned off which means that those vehicles cannot be detected. The ratio of the number of Bluetooth equipped vehicles to the total number of vehicles on the roadway is known as the Penetration Rate (*PR*). The formula for obtaining penetration rate is shown in Equation (4-i).

(4-i)

4.2 Reston Parkway Arterial

We considered a section of Reston Parkway, VA extending from New Dominion Pkwy to South Lakes Dr. for our study. Reston Parkway is located in the Fairfax County in the state of Virginia, USA. The network shown in Figure 4-2 has a total of 5 intersections. The speed limit for the main arterial is 45 mph, and it ranges between 15 mph and 45 mph for the side streets. These speed limits are used to compute travel times between the Bluetooth detectors and the nearest entry and exit nodes. It will be described in further detail when we discuss the OD estimation methodology.

There are 10 origins in the network and 10 destinations. Nodes 12 and 15 are destinations only; nodes 13 and 14 are origins only; the rest of the nodes are both origins and destinations. The blue boxes indicate the location of the Bluetooth detectors with the Bluetooth detector numbers in the boxes, there are a total of 6 detectors. The white ovals contain entry and exit numbers.

We simulated traffic flows on the network using VISSIM for 24hours. The OD flows were recorded for reference. These are treated as the actual OD flows. For the Kalman filter, we use link counts assumed to come from loop detectors. The data collection points are placed in VISSIM network such that they emulate field conditions. The location of Bluetooth detectors is as shown in Figure 4-2. A given percentage of vehicles (based on the given penetration rate) are treated as those equipped with Bluetooth devices. These vehicles are assumed to be detected by the Bluetooth detectors.

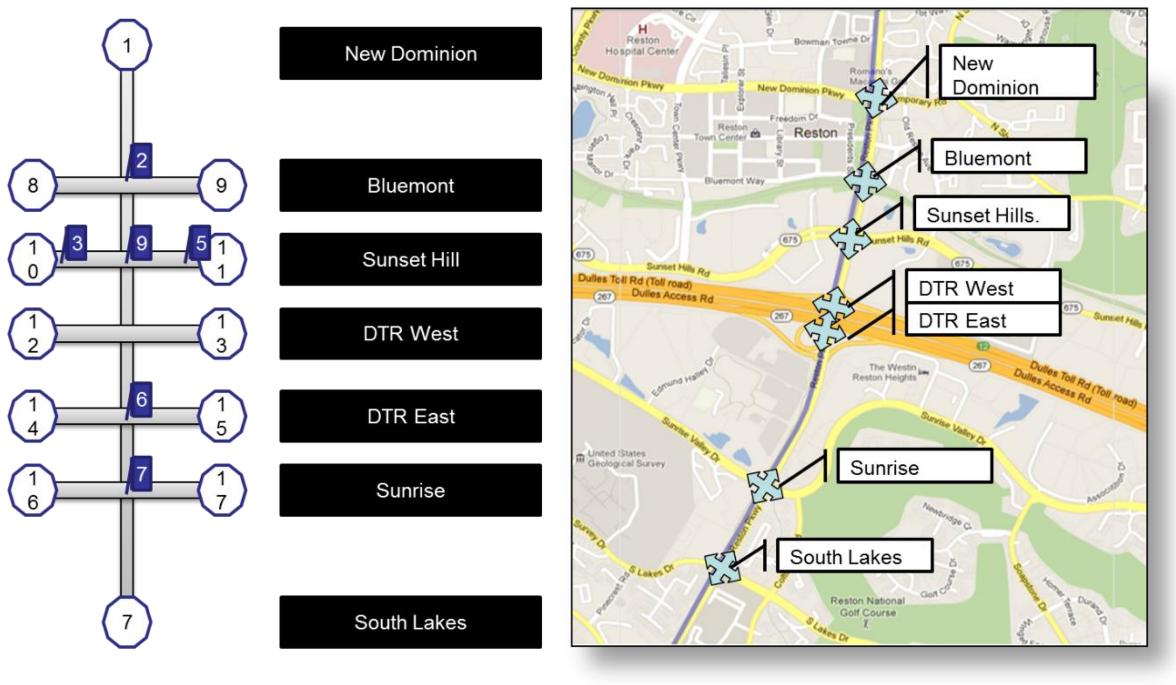


Figure 4-2: Section of Reston parkway arterial considered for the study.

Important points to note about the experiment:

- There is no route choice since there is only one possible route between an OD pair.
- The simulation experiments allowed us to treat a randomly identified vehicle as a vehicle equipped with Bluetooth device and then track its movements across the network.
- Bluetooth detectors are not present at every entry and exit of the network. The Bluetooth OD counts and travel times are available for the part of the network that is covered by the Bluetooth detectors.

The 24hrs study period was divided into 288 time steps of 5mins duration each. In each time step, we update our estimate of the OD matrix based on the available measurements. At the beginning of the time step, we make a prediction of what our OD proportions are, based on Bluetooth OD proportions from the previous time step. When we have measurements available, we update our prediction by checking if they are more reliable than the measurements or not. This is decided by a ratio called the Kalman Gain. After updating the predictions, we compute the error covariance, which basically indicates how close our predictions were to the measured values.

4.3 Kalman Filter

Barcelo et. al(2010)[6] performed a study which concluded that a simple counting of vehicles detected by the Bluetooth detectors to generate an OD matrix can lead to unacceptably high errors when the penetration rates are low. This calls for a method that can make use of the Bluetooth data in OD estimation.

Along with the Bluetooth data which provides us with travel times and some information about the traffic patterns (Bluetooth OD counts), we have data from the loop detectors on the road network which provide us with link counts, entry and exit counts. We need a tool that will allow us to use all the available information to come up with OD flow estimates. One such tool is the Kalman Filter.

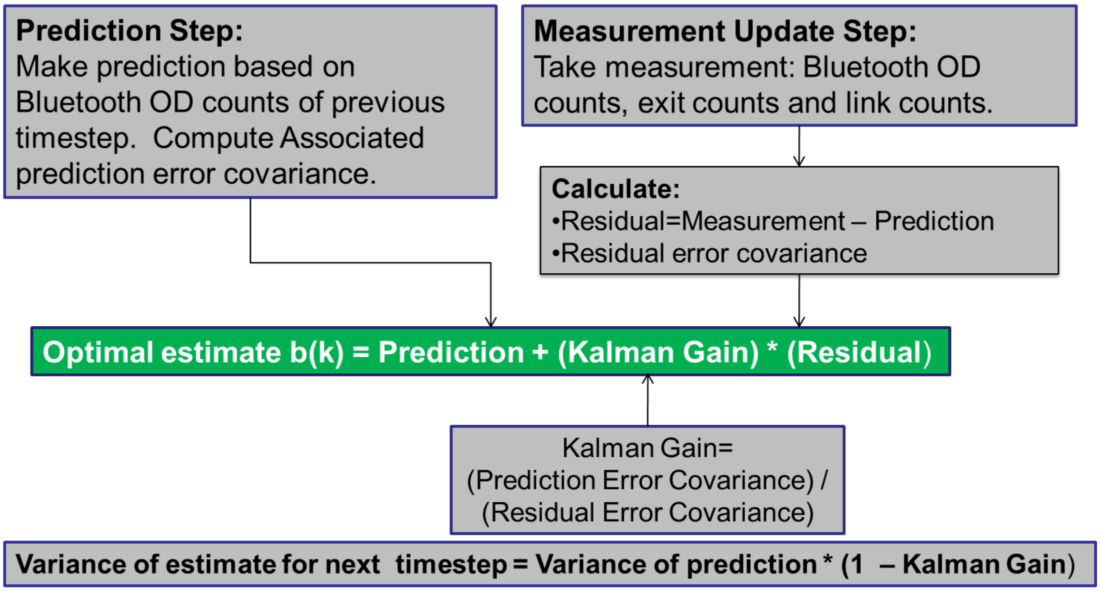


Figure 4-3: Kalman Filter modified for OD estimation for network covered partially by Bluetooth detectors.

In the study conducted by Murari and Abbas (Under Review), they use Bluetooth OD counts as predictions. This becomes clearer as we proceed to explain the formulation of the Kalman Filter.

The Kalman Filter formulation consists of two main steps which are summarized in Figure 4-3:

- Prediction step
- Measurement update step

State variable

We use the same state variables used by Murari and Abbas (Under Review) as shown below. We modify the prediction step to cater to the fact that the location of Bluetooth detectors is not the same as their study.

State variable for the Kalman filter in our study is a vector of OD proportions. Each element in this vector is the proportion of flow from an origin **i**, exiting at a destination **j** in time step **k,** represented by b(k) in Equation (4-ii).

(4-ii)

(4-iii)

$\boldsymbol{b_k}$ is a vector of $b(k)$, $b(k\text{-}1)$… to $b(k\text{-}m)$ as shown in Equation (4-iii), where, m is the maximum number of time steps taken by any vehicle to traverse the network.

Prediction step

The prediction step in the Kalman filter tries to relate the state variables for the current time step and those from the previous time step. In our study we are trying to relate the OD proportions for time step *k* with Bluetooth OD proportions from time step *k-1*. The changes made to the prediction step as compared to the method used by Murari and Abbas (Under Review) constitutes the main contribution of this paper.

A prediction of the state variable is made for a given time step *k* based on Bluetooth OD proportions from the previous time steps. The following equations are used in the prediction step.

(4-iv)

(4-v)

(4-vi)

Where,

w is white noise with an expected value of zero

is the predicted state variable

is the estimate of the state variable from the previous time step

D is a transition matrix of size (m+1)*n n*n given by

n is the total number of time steps in a day (24hrs*60 mins/5 mins = 288)

I is an identity matrix of size n n given by

Equation (4-iv) shows the relationship between Bluetooth OD proportions from the previous time step and the OD proportions of current time step. This is our model for prediction. We are saying that the OD proportions for this time step will be almost the same as the Bluetooth OD proportion from the previous time step.

The Bluetooth detectors cover only part of the network, so the OD matrix developed using the counts from those detectors will give us a partial OD matrix. This matrix is used to compute a set of OD counts for each time step by using a solver in MATLAB. The ‘lsqlin’ solver solves for a set of OD counts that have values for each entry and exit of the network instead of values just for the locations where Bluetooth detectors are present. While solving for the OD flows between each OD pair, we try to minimize the error between OD flow between Bluetooth detectors (not located at the network entries and exits) and the sum of predicted OD flows (going from network entries to exits) passing a given pair of Bluetooth detectors. This solving helps us come up with a prediction (Equation (4-vi)) for the Kalman filter at each time step.

Equation (4-vi) shows the prediction equation. The relationship between Bluetooth OD proportions and the state variable for the current time step are related using a transition matrix *D*. is a variable containing the predicted OD proportions. We solve for by using information from the Bluetooth OD counts available and also looking at OD flows from the previous time step. This step is necessary because we need to have as many Bluetooth OD proportions as the number of OD pairs in the network. Since we do not have Bluetooth detectors present at every entry and exit, we try to solve for a set of OD proportions with the given Bluetooth OD counts. For solving, we try to minimize the error between the estimated OD flows for the previous time step and the predicted set of values.

The prediction error covariance is computed as shown below. This variable helps us compute the Kalman Gain, which ultimately decides how much faith we can put on our predictions.

(4-vii)

Where,

is the prediction error covariance matrix and is the Prediction error covariance for the previous time step.

W is a matrix given by and *w* is white noise with an expected value of zero.

Measurement update step

Measurement update step is when we decide how much we need to modify the predictions, to obtain the updated OD estimates. This step is critical to capturing the variations in traffic flows during the course of the day. It is also effective in capturing the congestion effects.

Measurement vector z_k is given by Equation (4-viii).

(4-viii)

Where,

s(k) is a vector of exit volumes and *y(k)* is a vector of link counts

The residual ε_k is computed using Equation (4-ix).

(4-ix)

Where,

is the residual

is the measurement matrix

is a transition matrix of dimension (m+1)*n (m+1)*n constructed such that :

(4-x)

(4-xi)

Where, *r* is assumed to be white noise with an expected value of zero.

consists of input volumes, which when multiplied with the corresponding OD proportions in b_k and summed up, give exit volumes and link counts.

Kalman gain is computed using Equation (5-xii). It can be defined as the ratio between the prediction error covariance and the residual error covariance.

(4-xii)

Where,

G is the Kalman gain

(4-xiii)

R is the residual error covariance matrix.

The predicted state variable is updated using the following equation:

(4-xiv)

At the step where the prediction is updated in the code written in MATLAB, a MATLAB function called 'lsqlin' is used to solve for by minimizing the difference between the left hand side and the right hand side of equation (4-xiv) and also applying the constraints that require OD proportions corresponding to a given origin should sum to 1. This method ensures that there is no overestimation of vehicles in the network. We also define a lower-bound value of zero for all OD proportions in the 'lsqlin' solver to avoid negative estimates.

The prediction error covariance is updated using Equation (4-xv)

(4-xv)

4.4 Equal Distribution method of OD estimation

The method used for OD estimation by Gharat (2011)[1] will be referred to as *Equal Distribution Method (EDM)* in this paper. The method is a simple distribution of Bluetooth OD counts. The vehicles arriving at a Bluetooth detector location are equally distributed among the nearest exits and similarly, the vehicles departing from a Bluetooth detector location are distributed equally among the nearest entries. The penetration rates are used to bring the Bluetooth counts up to the actual volumes. A small example is shown in Figure 4-4.

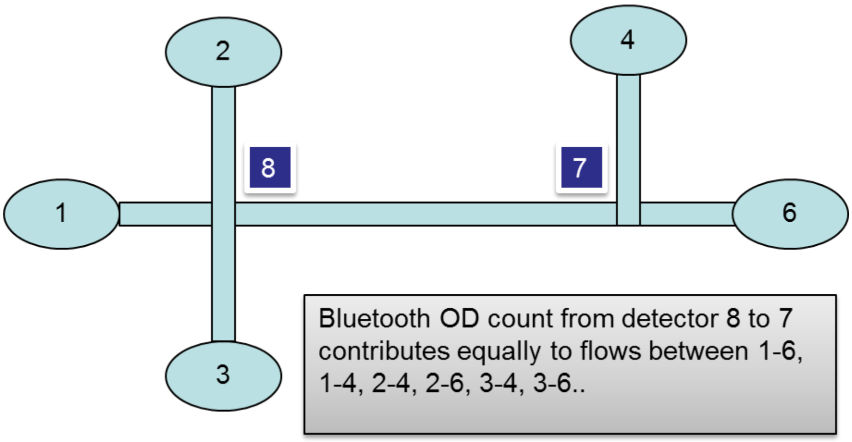


Figure 4-4: Example of EDM.

The main reason for choosing EDM for comparison with our developed method is that it deals with networks similar to ours. The network used in EDM is also covered partially by the Bluetooth detectors. This gives us a good benchmark to compare our method.

4.5 Experiment

24hr simulation was run in VISSIM and the data was organized into input volumes, exit volumes, link counts, travel times, Bluetooth OD matrix and actual OD flows. The actual OD flows are used only at the very end for comparison. EDM and the developed method were given the same inputs of Bluetooth counts.

MATLAB was used for the implementation of the developed method. 'lsqlin' solver was the most important tool used in the code. It was used twice in every iteration. The first is in the prediction step where we try to obtain a prediction using Bluetooth OD flows from previous time step, and the second is when we are modifying our predictions based on the Kalman Gain and the Residual.

The results section elaborates the comparisons carried out between the implemented methods.

4.6 Results and Conclusions

The results are presented as follows:

- Total error vs penetration rates for EDM and Kalman Filter
- Number of vehicles in the network at each time step (5% penetration rate)
- Mean square error for each time step (5% penetration rate)

- OD Flow patterns estimated by each method (5% penetration rate)

Total error is computed by taking the sum of all square errors between the estimated and actual OD flows and then summing them up across all time steps in a day.

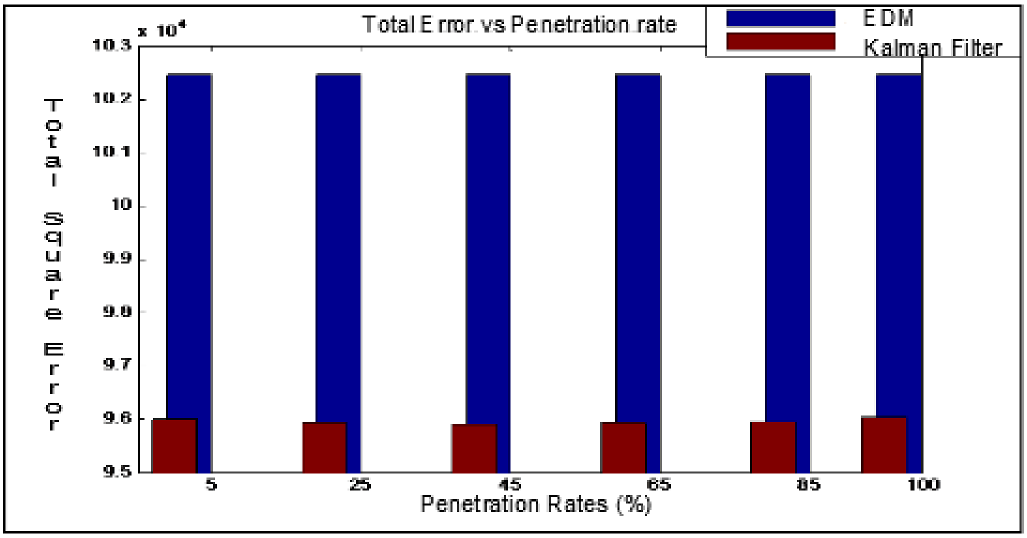


Figure 4-5: Plot showing the total error varying with the penetration rates.

Figure 4-5 shows that the total error doesn't vary much with EDM when penetration rates are varied. This is because the distribution of the Bluetooth OD counts remains the same throughout in EDM irrespective of the penetration rates. If the distribution patterns change with varying penetration rates, the error values also change. Kalman Filter method is seen to have lower total errors as compared to EDM.

The penetration rates are uniform across the network, so the Bluetooth OD counts are a good representation of the actual OD patterns on the network. This is the reason why the total error does not vary much with penetration rate for Kalman Filter.

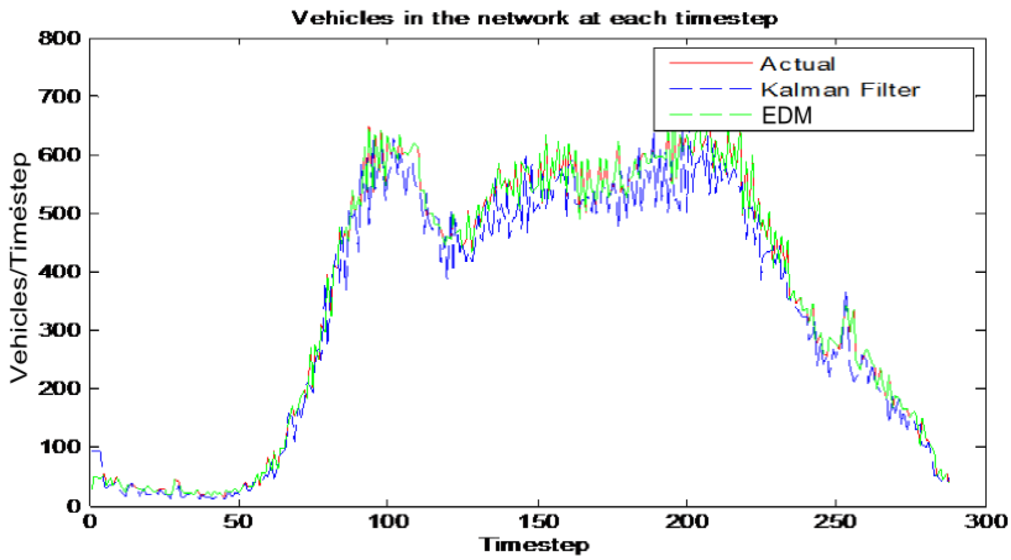


Figure 4-6: Plot of number of vehicles in the network at each time step(5mins) in a day. Kalman Filter is implemented with 5% penetration rate.

The Figure 4-6 shows the number of vehicles that are in the network as time passes in a day. It is to be noted that higher errors are observed during the peak hours when the volumes are higher. The plot is to verify that the vehicles are not over or under estimated. Kalman Filter can perform better if we can impose very strict constraints while estimating OD proportions. If the OD proportions corresponding to a given origin add up to 1, there will be no over or under estimation. Because of the limitations in the software used, we cannot achieve that, but we come very close.

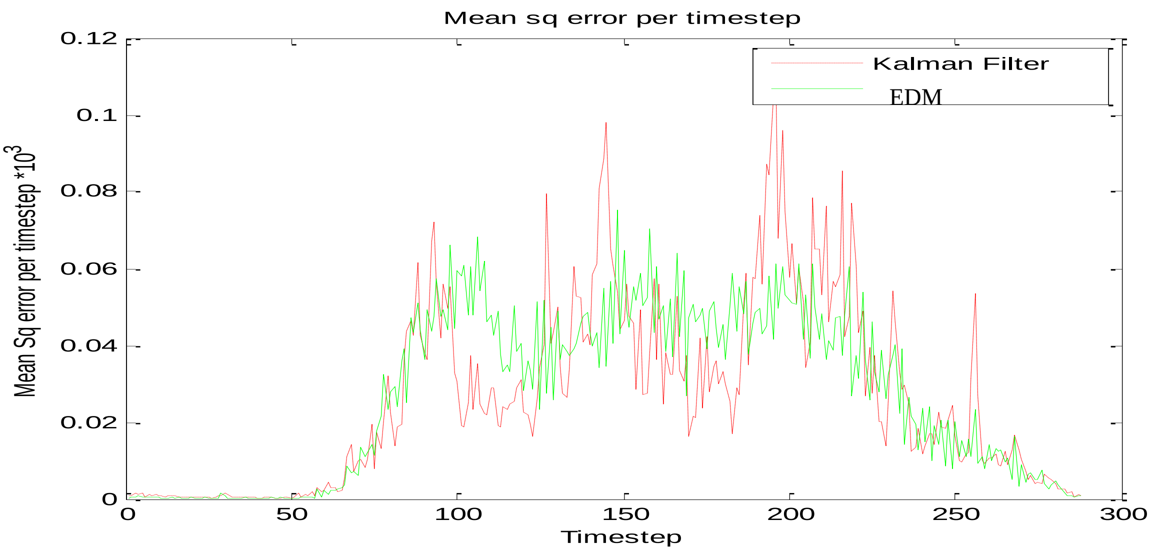


Figure 4-7: Plot showing the variation of mean square error as time passes in a day. The methods are implemented with 5% penetration rate.

Figure 4-7 shows the variation of mean square error with time of the day. The plot around the 100th and 175th time steps represents peak hours. Kalman filter does well in the peak hours as compared to EDM.

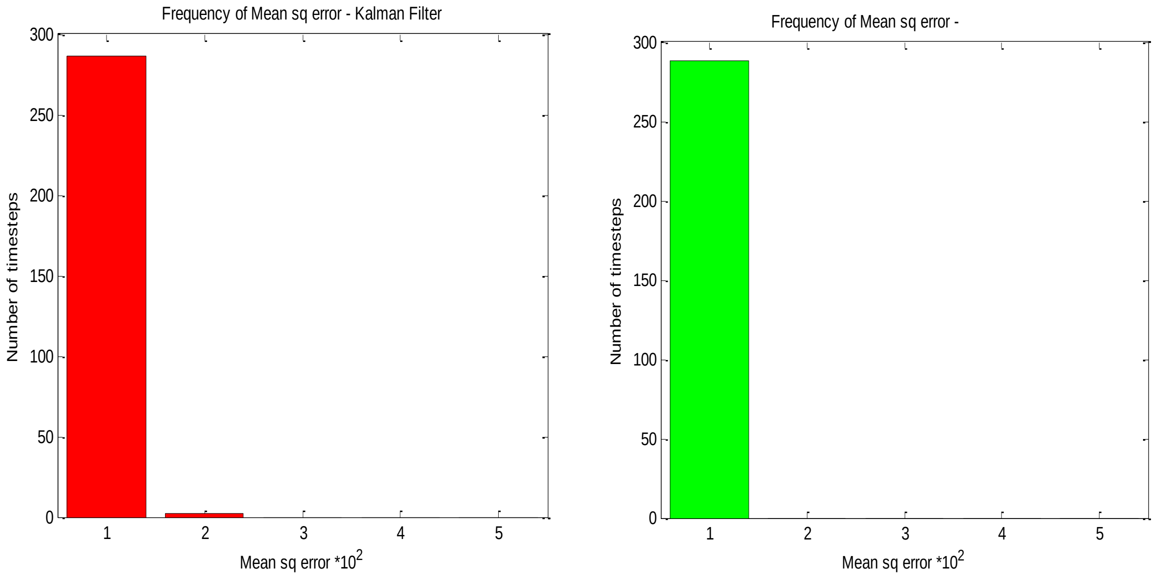


Figure 4-8: Plots showing the number of time steps in a day that have error between 0-100, 100-200 and so on. The plot on the left is for Kalman Filter and the plot on the right is EDM. The methods are implemented with 5% penetration rate.

Figure 4-8 shows plots showing the number of time steps in a day with low errors. Both Kalman Filter and EDM have errors between 0 and 100, which is acceptable.

OD Pair 13

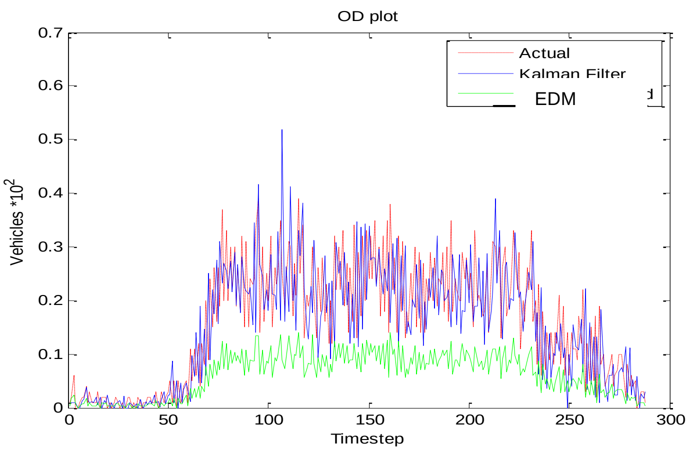


Figure 4-9: Plot showing OD flow for each time step between OD pair 13

OD Pair 26

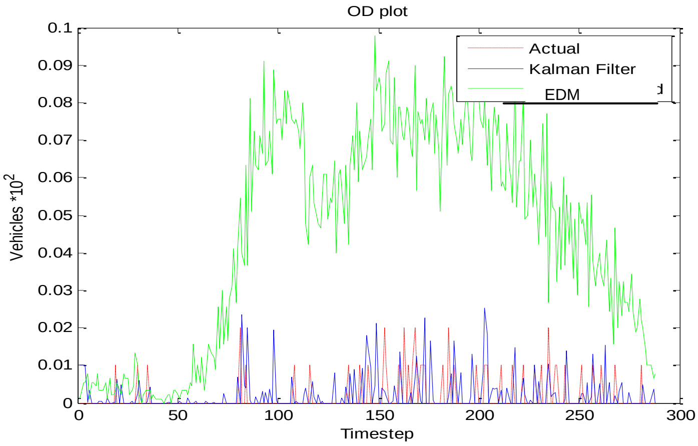


Figure 4-10: Plot showing OD flow for each time step between OD pair 26

OD Pair 92

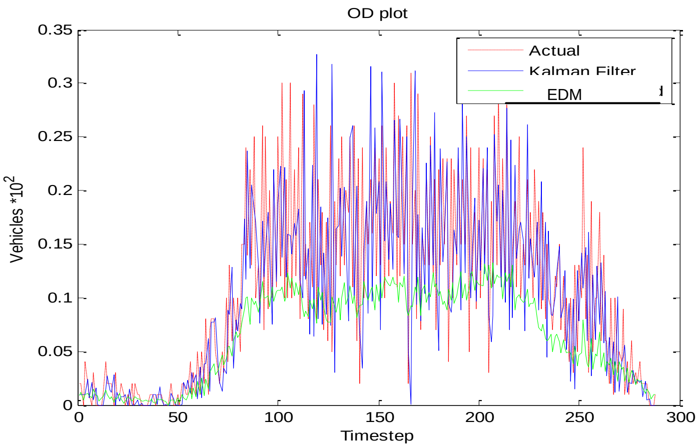


Figure 4-11: Plot showing OD flow for each time step between OD pair 92

Figure 4-9, Figure 4-10 and Figure 4-11 show OD flow patterns varying with time of the day. EDM fails to capture the variations in OD flows. Kalman Filter does a better job of capturing the OD patterns.

Overall we can conclude that the Kalman Filter method implemented to use Bluetooth data for a part of the network, performs reasonably well. It captures the OD patterns better than the method used by Gharat (2011) [1].

Future research can further examine the effect of non-uniform penetration rates in the Kalman Filter method. Using different predictions may improve estimates, which should be researched.

4.7 References

1. Gharat, A., *Bluetooth Based Dynamic Critical Volume Estimation on Signalized Arterials*, in Civil and Environmental Engineering,2011, Virginia Polytechnic and State University: Blacksburg, VA. p. 62.
2. Barcelo, J., Lidin, M., Laura, M., & Carlos, C. (2010). *Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring.* Transportation Research Record: Journal of the Transportation Research Board, 2175/2010, 19-27.
3. Castillo, E., J.M. Menéndez, and S. Sánchez-Cambronero, T*raffic Estimation and Optimal Counting Location Without Path Enumeration Using Bayesian Networks.* Computer-Aided Civil and Infrastructure Engineering, 2008. 23(3): p. 189-207.
4. Hui, Z., et al., *Estimation of Time-Varying OD Demands Incorporating FCD and RTMS Data.* Journal of Transportation Systems Engineering and Information Technology, 2010.
5. Friedrich M. et.al. *Generating Origin-Destination Matrices from Mobile Phone Trajectories*, Transportation Research Record, Vol. 2196/2010, 2011, 93-101
6. Kwon J., and Varaiya P. *Real-Time Estimation of Origin-Destination Matrices with Partial Trajectories from Electronic Toll Collection Tag Data*. Transportation Research Record, Vol. 1923/2005, 2006, 119-126
7. Nanthawichit C., Nakatsuji T., and Suzuki H. *Application of Probe-Vehicle Data for Real-Time Traffic-State Estimation and Short-Term Travel-Time Prediction on a Freeway,* Transportation Research Recoed, Vol. 1855/2003, 2007, 49-59
8. Kalman, R. E. *A New Approach to Linear Filtering and Prediction Problems*. Journal of Basic Engineering, Transactions of the ASME, Vol. 82, No. 1, 1960, pp. 33-45.

5 Conclusions

In this thesis, Kalman Filter based dynamic OD estimation methods were explored. Dynamic OD estimation calls for updating OD flow estimates continuously based on measurements made on field. Often, the measurements available may not directly give the OD flow estimates. Kalman filter is a tool that allows us to first make a prediction of OD flow estimates and then update them based on the measurements that become available.

Kalman filter is perfectly suited for online ATIS and ATMS based applications. IT can be used to make a prediction prior to the time when measurements become available, and once they are available, we can update the prediction to obtain an estimate. This estimate can then be used to make a prediction for the following time step.

Three Kalman Filter methods were implemented in the work for this thesis. The first two methods (Case 1 and Case 2) were largely based on previously used methods, with modifications made to the prediction step in the Kalman Filter. In the prediction step, we have a model that can relate the estimates of the previous time step to the state variable (OD proportions) of the current time step. This model was modified to use Bluetooth OD counts as a prediction. The Bluetooth OD counts capture the traffic patterns on the network. This information can be used to supplement the measurements (link counts, exit and entry volumes).

The first two methods presented in Paper 1 were compared with QueensOD. Case 2 performed much better than Case 1 and QueensOD. It successfully captured the traffic patterns. We can conclude that the inclusion of Bluetooth OD counts in the prediction step is an effective modification. Case 2 performed well with low penetration rates as well. This is an advantage over methods used by Barcelo et. al (2010) since the performance of their method reduced with decreasing penetration rates.

The third method was a modification to Case 2. An additional step was added to the prediction step to help deal with networks that are partially covered by Bluetooth detectors. This additional step solved for a set of predicted flows based on Bluetooth OD counts from the previous time step and the OD flow estimates from the previous time step. This method was compared with Equal Distribution Method used by Gharat and Abbas(2011).

Overall, we can conclude that the inclusion of Bluetooth OD counts in the prediction step of Kalman Filter for dynamic OD estimation enables us to capture the traffic patterns better, therefore obtaining better estimates of OD flows.

Future research should focus on the effect of penetration rates on the developed methods. Another aspect that can be studied is the effect of location of Bluetooth detectors on the dynamic OD estimates.

6 References

1. Wardrop, J.G., *Some theoretical aspects of road traffic research.* Proc. Inst. Civil Eng., 1952. **1**: p. 325–378.
2. Gharat, A., *Bluetooth Based Dynamic Critical Volume Estimation on Signalized Arterials*, in *Civil and Environmental Engineering*2011, Virginia Polytechnic and State University: Blacksburg, VA. p. 62.
3. Barcelo, J., et al., *Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring.* Transportation Research Record: Journal of the Transportation Research Board, 2010. **2175/2010**: p. 19-27.
4. Aerde, M.V., H. Rakha, and H. Paramahamsan, *Estimation of Origin-Destination Matrices: Relationship Between Practical and Theoretical Considerations.* Transportation Research Record: Journal of the Transportation Research Board, 2003. **1831/2003**: p. 122-130.
5. Low, D., *A new approach to transportation systems modelling.* Traffic Quart., 1972. **26(1972)**: p. 391-404.
6. Hogberg, P., *Estimation of parameters in models for traffic prediction: A non-linear regression approach.* Transportation Research 1976. **10**(4): p. 263-265.
7. Holm, J., *Calibrating traffic models on traffic census results only.* Traffic Engng and Control 1976. **17 (1976)**: p. 137–140.
8. Robillard, P., *Estimating an O-D matrix from observed link volume.* Transpn Research 1975. **9 (1975)**: p. 123–128.
9. Symons, J., *A model of inter-city motor travel estimated by link volumes.* ARRB Proc., 1976. **8**: p. 53-65.
10. Nguyen, S., *Estimating an O-D Matrix from Network Data: A Network Equilibrium Approach.* Centre de recherche sur les Transports, Universite de Montreal, 1977. **87**.
11. Zuylen, H.J.v., *The Most Likely Trip Matrix Estimated From Traffic Counts*. Transportation Research Part B: Methodological, 1980. **14B**(3): p. 218-293.
12. Cascetta, E., *Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator.* Transportation Research Part B: Methodological, 1984. **18**(4-5): p. 289-299.
13. H Spiess, *A Gradient Approach for the 0-D Matrix Adjustment Problem.* CRT, Université de Montréal, Montréal, 1990(Publication No. 693).
14. Florian, M. and Y. Chen, *A coordinate descent method for bilevel O-D matrix estimation problems.* Centre de Recherche sur les Transports, 1995. Technical Report CRT-**807**.
15. Codina, E. and J. Barcelo, *Adjustment of OD trip matrices from observed volumes: An algorithmic approach based on conjugate directions.* European Journal of Oper. Research, 2004. **155**(3): p. 535-557.
16. Maher, M.J., *Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach.* Transportation Research Part B: Methodological, 1982. **17**(6): p. 435-447.
17. Cremer, M. and H. Keller, *A New Class of Dynamic Methods for the Identification of Origin-Destination Flows.* Transportation Research, 1987. **21B(2)**: p. 117-132.
18. Nihan, N.L. and G.A. Davis, *Recursive estimation of origin-destination matrices from input/output counts.* Transportation Research Part B: Methodological, 1987. **21**(2): p. 149-163.
19. Bell, M.G.H., *The estimation of an origin-destination matrix from traffic counts.* Transportation Science, 1983. **17(2)**: p. 198-217.

20. Zijpp, N.J.V.d. and R. Hamerslag, *Improved Kalman Filtering Approach For Estimating Origin-Destination Matrices For Freeway Corridors.* Transportation Research Record, 1994. **1443**: p. 54-64.
21. Chang, G.L. and J. Wu, *Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors.* Transportation Research Part B: Methodological, 1994. **28**(2): p. 141-160.
22. Castillo, E., J.M. Menéndez, and S. Sánchez-Cambronero, *Traffic Estimation and Optimal Counting Location Without Path Enumeration Using Bayesian Networks.* Computer-Aided Civil and Infrastructure Engineering, 2008. **23**(3): p. 189-207.
23. Hui, Z., et al., *Estimation of Time-Varying OD Demands Incorporating FCD and RTMS Data.* Journal of Transportation Systems Engineering and Information Technology, 2010.
24. Parry, K. and M.L. Hazelton, *Estimation of origin–destination matrices from link counts and sporadic routing data.* Transportation Research Part B: Methodological, 2012. **46**(1): p. 175-188.