2015-09-18

Real-time Queue Length Estimation on Freeway Off-ramps Using Case Based Reasoning Combined with Kalman Filter

Heshami, Seiran

UNIVERSITY OF CALGARY

Real-time Queue Length Estimation on Freeway Off-ramps Using Case Based Reasoning

Combined with Kalman Filter

by

Seiran Heshami

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN CIVIL ENGINEERING

CALGARY, ALBERTA

SEPTEMBER, 2015

Abstract

Real-time queue length estimation and prediction provides useful information for proactively managing transportation networks. Queue spillback from off-ramps onto main freeway lanes is a serious traffic issue that can be efficiently managed using dynamic queue information. In this thesis, a case-based reasoning algorithm combined with a Kalman filter is developed to provide real-time queue length measurements and predictions on long freeway off-ramps. Estimations are based on occupancy readings from three loop detectors installed on the ramp. The proposed method is examined using a micro-simulation model in a Quadstone Paramics package on an off-ramp with a length of 650 meters. The simulation results demonstrate that the model is capable of estimating and predicting the queue length on long off-ramps in 60 second time intervals. The performance of the algorithm is examined under various demand loading scenarios, estimation time intervals and number of detectors through several experiments.

Acknowledgements

Table of Contents

List of Tables

List of Figures

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
AI	Artificial Intelligence
AITS	Advanced Travelers Information System
API	Application Programming Interface
ANN	Artificial Neural Network
CBR	Case Based Reasoning
CORSIM	CORidor SIMulation
HCM	Highway Capacity Manual
I/O	Input/Output Models
ITS	Intelligent Transportation Systems
KF	Kalman Filtering
KNN	Kth Nearest Neighborhood
LWR	Lighthill and Witham and Richards shockwave model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MOC	Model of Charachteristics
PASSER	Progression Analysis and Signal System Evaluation Routine
PGF	Probability Generating Function
TRANSIT	TRAffic Network StudY Tool
VIVIDS	Video Imaging Vehicles Detection System
VSL	Variable Speed Limit

CHAPTER ONE: **INTRODUCTION**

1.1 Research Background

Urban freeways, one of the most important components of transportation networks, play a crucial role in improving urban transportation capacities. However, population growth and the overuse of personal vehicles in recent years have increased the chance of queues forming and generating bottlenecks on freeways, especially in the merge and divergent points of freeways.

There are several causes of vehicular queue formation in traffic networks. The formation of queues can be due to either recurrent or non-recurrent congestion. In a recurrent congestion queue, the traffic demand exceeds the available capacity of a network component. Examples include queues forming at signalized intersections, metered motorways, unmetered freeway on-ramps and off-ramps, toll booths, and at-grade crossings with other modes. Non-recurrent events, such as the occurrence of collisions, incidents, road construction, special events, and inclement weather conditions, may also result in queue formation.

Information on the spatial-temporal evolution of queue length is one of the most important traffic characteristics that can be utilized in a diverse traffic network analysis and in optimization efforts to improve network performance, congestion mitigation strategies, and fuel consumption and emission reduction measures. Real-time queue length estimation and prediction has extensive applications in modern transportation management methods. This information is vital for several Intelligent Transportation Systems (ITS) applications, such as an Advanced Travelers Information System (ATIS), queue warning, incident management, adaptive signal control, adaptive ramp metering, variable speed limit, and other advanced traffic management systems, which aim to improve the transportation network performance.

A queue on a freeway off-ramp can generate bottlenecks whenever the queue extends beyond the right lane and a spillback occurs onto the freeway. Congestion on freeway off-ramps undermines the enormous investments in the infrastructure, reduces the capacity of the freeway, and increases travel time and potential safety hazards. Two of the main reasons queues form on off-ramps are the operation of a signalized intersection in the downstream urban link and the large demand of traffic exiting the freeway. The mutable nature of traffic flow makes it impossible to design a fixed signal control plan that is simultaneously effective and efficient. In contrast, reactive and proactive signal control strategies are capable of dynamically managing the downstream signalized intersection in an efficient manner. Real-time queue length estimation and prediction can provide useful information to allocate reasonable green time for each phase of the signal to manage the queue more efficiently.

1.2 Research Challenges and Motivations

In the literature, various queue length estimation approaches are discussed. The majority of studies focus on queue estimation in signalized intersections or metered on-ramps. These approaches often target a given congestion level (i.e. under-saturated or saturated conditions), queue types, and particular measuring devices. Previously developed models are mainly unable to measure long queues due to the specific theoretical basis or data requirements.

A few studies explore techniques to estimate a queue length longer than the distance between the intersection stop line and the advanced detector. However, arrival traffic flow information is required or a constant arrival rate has to be assumed for each signal cycle (Muck, 2002). High resolution traffic signal data might be useful to recognize the relationship between signal phase changes and the long queue behavior, but it is not reliable for congested traffic

conditions in which the break points in the traffic pattern are not recognizable, especially on off-ramps that may be located a relatively long distance from the intersection (Liu et al., 2009).

Further, there are only a few studies in the literature that discuss off-ramp queue length estimation. The geometric characteristics of freeway exit lanes are generally determined based on providing a safe access to the arterial roadways with a tolerable deceleration level, which requires a large radius and, consequently, a long length for the off-ramp. Moreover, a critical traffic situation occurs when a queue spillover disturbs the freeway main stream traffic. Efficient control strategies to mitigate this issue require a queue length estimation approach capable of measuring such long queues. However, there is no study in the literature that focuses on the dynamic evolution of long off-ramp queues.

The specific theoretical basis and data requirements for prevalent queue estimation techniques cannot be directly applied to long off-ramp queue length estimation. One study in the literature focuses on off-ramp queue length estimation (Qian et al., 2012) for uncongested traffic conditions with an off-ramp of 100 meters in length. However, the simplifying assumptions in this approach restrict its application in congested traffic conditions in which the free flow traffic is not observed for a number of consecutive time intervals and it causes significant errors for long queues.

In addition, few studies focus on predicting the queue length in the near future. Anticipating future queue length is an important step to develop proactive, rather than reactive, traffic control and management schemes as traffic monitoring alone cannot improve the safety or efficiency of the transportation system. Kaysi et al. (1993) found that the use of real-time information as a basis for real-time control is significantly inferior to the use of predictive information because of the resultant time lag. In other words, by the time control is formulated

and deployed, traffic conditions may have changed. On the other hand, predictive information allows the traffic control and management system to issue control and rerouting information in a proactive fashion rather than reporting on previously detected conditions that may have passed.

This thesis develops a novel real-time queue length estimation and prediction model for long freeway off-ramps. An integrated case-based reasoning (CBR) and Kalman filtering framework is developed. The use of these two complementary approaches is shown to significantly improve the accuracy of estimating and predicting long queues as compared to the use of each method separately. One of the main advantages of this approach is its ability to estimate and predict queue length in the case of non-recurrent events corresponding to extreme fluctuations in traffic demand.

The developed approach was tested using simulation experiments due to the absence of real observed queue data. A rigorous sensitivity analysis was conducted to evaluate the performance of the approach for various congestion levels, measurement errors, number of detectors, and interval step estimation.

1.3 Research Contributions

This thesis develops an integrated framework for real-time queue length estimation and prediction targeting long freeway off-ramps combining case-based reasoning (CBR) and Kalman filtering framework. CBR, a problem-solving paradigm in artificial intelligence (AI), is capable of forming a relatively large library that includes a priori queue estimates that correspond to both recurrent and non-recurrent traffic congestion such as accidents, special events, and adverse roadway conditions. CBR uses the hypothesis that similar occupancy patterns should have a similar queue length. Although the CBR system goes through a four-

step process (retrieve, reuse, revise, and retain), its main task is to find a similarity measure; thus, it is computationally efficient in finding an appropriate solution. In addition, online learning allows CBR to update its library in case new traffic states are presented, which refreshes the library with recent information and avoids it to become obsolete. Our interest is to estimate and predict the traffic state (i.e. queue length) for current and future situations; therefore, a Kalman Filter (KF) is an excellent candidate for augmenting the CBR method and refining the estimates and predictions of future states. KF processes real-time observations sequentially following an auto-regressive relation and allows additional measurements to be sequentially incorporated. Another important feature of KF is its ability to take into account noisy measurements, which could arise during incidents, to predict future states. Thus, based on a KF, the algorithm refines and predicts the queue.

Although KFs were applied in studies for other queue estimation cases (e.g. on-ramps and at signalized intersection), the developed integrated method improved the technique by incorporating the CBR algorithm, which makes it applicable in a variety of traffic conditions. Moreover, training and implementing the CBR are based on a simple yet robust framework compared to other machine learning algorithms that are discussed in the literature. Revising and updating the library utilizes accessible and affordable loop detector information, which is simple to apply.

A real-time queue estimation method should be an accurate model that can be applied in a variety of traffic conditions, including recurrent and non-recurrent congestions. Further, any real-time estimation model requires data to be provided continuously in relatively short time intervals. Thus, affordability and maintainability of the data gathering devices have a considerable effect on the overall cost of the approach in real world applications. The main

objectives of this research, considering the discussed challenges and motivations, are the following:

1) Developing a real-time queue estimation/prediction approach that is capable of producing accurate estimations of long queues on freeway off-ramps for various traffic conditions and ramp lengths;

2) Integrating an AI algorithm with an advanced estimation/prediction method to improve the accuracy of estimations;

3) Examining the performance of the developed approach utilizing a Quadstone Paramics microsimulation model in the absence of real observed data.

In this thesis, an integrated CBR and KF framework is developed to estimate and predict the queue length of freeway off-ramps. CBR is capable of forming a relatively large library that includes queue estimates that correspond to both recurrent and non-recurrent traffic congestion such as accidents, special events, and adverse roadway conditions.

This research makes the following contributions to the literature:

1) Estimating and predicting the queue length on long freeway off-ramps in the case of recurrent congestions and non-recurrent events corresponding to extreme fluctuations in traffic demand;

2) Utilizing the CBR method to provide indirect measurements of queue length based on a library of cases;

3) Updating and correcting CBR measurements dynamically by integrating a KF process to improve the accuracy of the model;

4) Updating the library of events dynamically to extend the applicability of the model as well as to reduce the computational iterations and improve the performance of the model.

5) Predicting the queue length over a short time period; and

6) Conducting a sensitivity analysis to examine the performance of the developed approach for various levels of measurement errors, demand fluctuations, estimation intervals, and number of detectors;

1.4 Organization of Thesis

This thesis is organized in five chapters as follows. Chapter Two contains a review of the literature on different queue estimation methods with a specific emphasis on real-time queue estimation approaches. In Chapter Three, the proposed integrated CBR and Kalman filtering approach is presented. A brief description of CBR and the KF algorithm is also provided in this chapter. The performance of the developed method is examined in Chapter Four using a microsimulation network. Several sensitivity analyses are also conducted and discussed in this chapter to evaluate the effects of different parameters on the accuracy of the model. Chapter Five summarizes the major results and conclusions of the research and also recommends some directions for future studies based on this thesis.

CHAPTER TWO: **LITERATURE REVIEW**

Congestion on urban streets, freeways, and merges and diverges has a negative impact on the performance of a transportation network, the reliability of a public transit system, and it also affects other road users. Vehicular queues, as a measure of congestion, is the main reason for delays and has been a controversial issue for transportation planners during recent decades. Recognizing and detecting queue characteristics, especially queue length, have been investigated in several studies to develop accurate and reliable queue estimation and management methodologies. In this section, a review of various data collection methods and their application in diverse queue length estimation models for different traffic network elements, including signalized intersections, urban networks and freeways, and metered and un-metered ramps, are discussed.

2.1 Review of traffic data collection methods

Various data types and data collection methods reviewed in this section are utilized in the literature to provide information on traffic characteristics and for queue length estimation.

Video cameras are one of the most widely used traffic monitoring and detection devices. Cameras detect multiple lanes and areas based on the position and the required performance and resolution. Speed measurements, traffic count data and vehicle types and classification can be provided using video cameras. However, there are limitations in using these devices especially for real-time data collection purposes. Image processing technologies are required to extract numerical traffic data from recorded videos. In addition, the accuracy and reliability of collected data affected by visibility issues including inclement weather conditions, transition of days and nights and the location of camera(Cheek et al., 2008).

Travelling vehicles on the network can be detected using floating data from probe vehicle which provide real-time information about the speed, location, and travel time. Various technologies are utilized to develop probe vehicle systems including Global Positioning Systems (GPS), Automatic Vehicle Location (AVL), Automatic Vehicle Identification (AVI), and cellular probes (Qui, 2007). Another advanced technology that is moving into deployment is Connected Vehicles Initiative which combines GPS navigation, advanced vehicle sensors including Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I), and Infrastructure to Vehicle (I2V) communications (Venkatanarayana et al. 2011). These systems are expected to provide more accurate real-time traffic information. However, the main limitations are privacy issues (Herrera et al., 2009) and the effect of probe vehicles penetration rate on the accuracy of the data (Comert and Cetin, 2008).

Point detection technologies including loop detectors, microwave radar, and infrared are widely used in traffic monitoring. These types of sensors are able to measure traffic flow, time occupancy, and possibly speed. Loop detector is one the most commonly used sensors which is widely used in detecting vehicle presence in actuated traffic signals. However loop detectors cannot directly measure space mean speed and additional assumptions and computations are required to determine queue length from loop detector occupancy readings (Sreedevi, 2007). However, these devices are still widely used because of their availability and possibility of collecting real-time traffic information.

2.2 Queue Length Estimation at Signalized Intersections

Information on delays and queue lengths is fundamental for any responsive or adaptive traffic signal control system. The two main analytical queue estimation techniques that are prevalent for signalized intersections are the following:

1) Input-Output technique (I/O)

2) Shockwave-based approach

Several data-driven queue estimation techniques have also been developed over the past few years, which are distinguished by their unique direction of queue estimation. These techniques include the following:

3) Probabilistic probe vehicle-based positioning methods

4) Kalman Filtering and Marcovian Chains-Based Models

5) Direct video-based queue estimation methods

6) Neural Network and Neuro-Fuzzy methods for queue prediction

A brief review of the above methods and their utilization in prior studies are discussed in the following subsections.

2.2.1 Input-Output Techniques

The fundamental theory of the I/O model is mainly based on the fundamental vehicles' conservation equation that estimates the queue length by calculating the difference between the total number of arrivals and the total number of departures at a signalized intersection (Anusha et al., 2013). Arrivals and discharge flow profiles produced by stop bar and advanced detectors combined with signal cycle information are used to determine the queue length and the performance measure (i.e. delay) at the intersection (Sharma et al., 2007).

The basic idea of input-output flow has been utilized and improved in several queue estimation studies for more than 50 years (Newell, 1965; Gazis, 1974; Catling, 1977;

Stephanopoulos ad Michatopoulos, 1978; Akcelik, 1999; Sharma et al., 2007). From a graphical point of view, Lawson et al. (1996) developed the spatial-temporal evolution of the queue for three different cases: i) constant departure rate from the bottleneck, ii) one change in the departure rate, and iii) an under-saturated signalized intersection. Newell's basic triangular flow-density relationship was assumed to describe the flow characteristics that provided a precise theoretical definition of a vehicular queue. However, the applicability of this method is limited in real world implementations, especially for real-time estimation purposes, due to several simplifying assumptions including:

- Under-saturated traffic conditions

- A known arrival flow profile

- Instantaneous change in speed from free flow speed to queued speed

- Zero speed in the queue

In several subsequent studies, the I/O method was improved by changing the source of data and by integrating other mathematical models into the method. Detector actuation data and signal timing information were utilized to develop the queue polygon using the flow conservation equation (Sharma et al. 2007). The mean error of this method was reported as 1.3 vehicles in each cycle. However, this method could only be applied in low volume traffic situations.

The majority of traffic modeling software produced queue length estimations based on the principles of the I/O method and its specific definition of queue length. Mystkowski and Khan (1999) compared the queue length estimated by SIGNAL 94, SYNCHRO3, TRANSIT-7F, PASSERII-90, and CORSIM with observed field data and reported reasonable results 50% to 85% of the time. A classification framework also developed by Viloria et al. (2000) in which a

conversion factor was determined to translate the estimations of each model to the equivalent HCM 2000 outputs. Such categorization and comparison may provide useful information to choose the appropriate modeling program. However, the results are not reliable for over saturated traffic conditions, and their application is limited to static queue estimation approaches (Viloria et al. 2000).

2.2.2 Shockwave-Based Approaches

Integrating the theory of fluid dynamics with transportation problems was initially proposed by Lighthill and Witham (1955) and was further developed by Richards (1956), who introduced shockwaves to define highway traffic flow dynamics (LWR shockwave theory). The shockwave theory in transportation considers the movement of a group of vehicles as a continuously distributed flow with a certain density (Dazhi et al., 2007). The shockwave is defined as a discontinuity, which propagates over time; the wave velocity is determined by equation (2.1):

$$u_w = \frac{\Delta q}{\Delta k} = \frac{q_2 - q_1}{k_2 - k_1} = \frac{k_2 u_2 - k_1 u_1}{k_2 - k_1} \qquad (2.1)$$

Where, $q_1, k_1,$ and u_1 are the flow, density, and velocity of the upstream region, respectively, and $q_2, k_2,$ and u_2 are the flow, density and velocity of the downstream region, respectively. The graphical illustration of a traffic shockwave based on the fundamental flow-density diagram is represented in Figure 2.1. $v_1, v_2,$ and v_3 define the shockwave velocity between any two points on the curve.
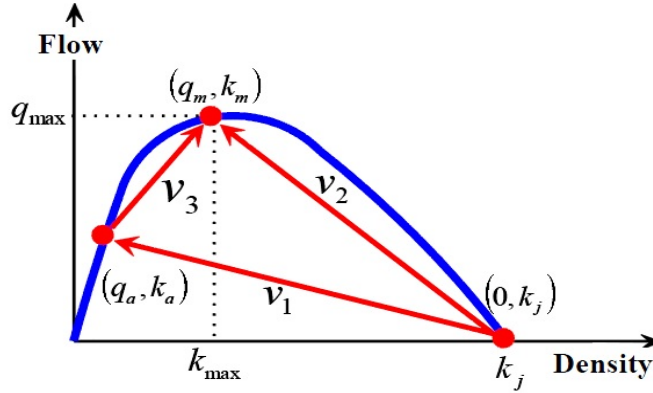
Figure 2. 1: Shockwave speed on the fundamental diagram (Dazhi et al., 2007)

This theory was later utilized to describe different traffic characteristics, especially dynamic queue length estimations. Yi et al. (2001) developed a mathematical framework based on the continuum principle and the method of characteristics (MOC). In this approach, two parameters, α and β, describe the flow dynamics to improve the general speed-density relationship. This method resulted in a large variation of errors depending on the selected speed-density model, frequency of updating parameters, and the data utilized for density derivation.

Utilizing the high resolution detector data, Liu et al. (2009) improved the application of the LWR theory in dynamic queue length estimation. This approach was further developed using a fusion of probe vehicle and detector data that increased the accuracy of real-time queue length estimations (Badillo et al. 2012). The accuracy of shockwave-based approaches was reported relatively high for long queues. However, these methods are not applicable for oversaturated traffic conditions where the breakpoints in shockwave and arrival flow are not identifiable.

Ramezani and Geroliminis (2015) improved the application of the LWR theory by developing a queue profile estimation method that did not require signal cycle data; therefore, it can be applied in actuated signal setting situations. Another improvement was that the proposed

model did not require the assumption of a known arrival flow profile and, as a result, the model was more consistent with real world conditions. Probe vehicle information was categorized into stopped and moving vehicles and was utilized to develop a flow-density curve and queue profile. The model was also able to approximate the signal setting; however, the accuracy of the results in both queue profile and signal setting estimation was dependent on the probe vehicles' penetration rate. Figure 2.2 illustrates an example of queue profile identification based on this proposed method.
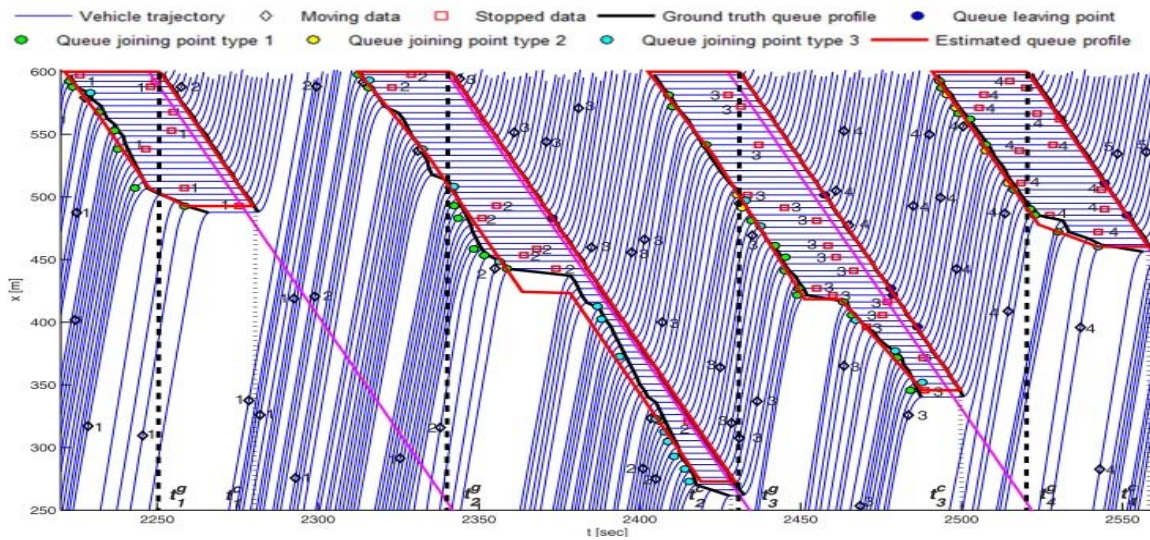
Figure 2.2: Shockwaves-based queue profile estimation (source: Ramezani and Geroliminis, 2015)

2.2.3 Probabilistic Probe Vehicle-Based Positioning Methods

The stochastic behavior of traffic flow motivated a number of studies to explore probabilistic-based queuing models. The queue evolution at both fixed and actuated signal controls was formulated as a randomly distributed process that was required to assume a certain probability distribution for arrivals and departures. The results of this study demonstrated the

model's consistency with microsimulation models and it's applicability in traffic planning and design problems (Viti and Vanzuylen, 2009).

The probe vehicle technology is an application of ITS, which typically refers to a vehicle moving on the transportation network with its location and speed information transferred in real time to a control center (Qui, 2007). The positions and speed of the probe vehicles were used by Comert and Cetin (2008), (2007) in the case of non-saturated conditions to provide information on queue length, assuming a known marginal probability distribution of the queue length. The location of the last probe vehicle is the only required information in this approach. They also developed analytical models to evaluate the impact of the percentage of probe vehicles in a traffic stream on the accuracy of estimations. The assumptions of steady-state conditions and known arrival rate limit the application of the model; also, the estimation accuracy is highly dependent on the number of probe vehicles (Comert and Cetin, 2007, 2008), (Cetin, 2013).

2.2.4 Kalman Filtering and Marcovian Chains-Based Models

A Kalman filter (KF) is named after Radulf E. Kalman, who developed a recursive algorithm to solve a discrete-data linear filtering problem in 1960. A KF is a predictor-corrector type estimator that minimizes the estimated error covariance based on some predetermined conditions. A KF is a very robust tool capable of controlling noisy input measurement data to produce less noisy output. The KF process begins by providing an a priori estimate of the state using state equation, which represents the relationship between states in two consecutive time steps. The initial estimate is updated to an a posteriori estimate of the state using the measurement; the next time step utilizes the a posteriori estimate to predict the state in the next time interval. A more precise description of a KF is presented in Chapter Three. A Marcov chain is also a set of random variables $\varphi = \{\varphi_n : n \in T\}$ where T is a countable time set. In this

stochastic model, the estimation of parameters in the next time interval is dependent only on the current state as the immediate memory (Meyn and Tweedie, 2012).

Pecherkova et. al (2008) investigated the application of the state and parameter estimation methods to provide online estimates of queue length for three different traffic regions, including ring road, peripheral road, and city center. The results showed that for smooth traffic conditions with sufficient detectors on main roads, a KF is a suitable estimation technique. In contrast, if there are some immeasurable roads in the network with non-homogenous traffic flow, nonlinear filters (i.e. an extended KF) are advantageous (Pecherkova et. al, 2008).

Applying data fusion techniques, Friedrich et al. (2003) found that combining the Kimber-Hollis and Marcov chains methods produced an acceptable estimation of the average queue length in signalized intersections with a relatively low degree of saturation. They improved the results for more saturated traffic conditions by utilizing the quasi measurement from the Meuck module in a KF (Friedrich et al., 2003). The combination of a KF with other estimation approaches, and using different sources of data, is utilized in the queue estimation of other network elements, such as street networks and on-ramps; this approach is discussed later in this chapter.

2.2.5 Direct Video Based Methods

Video imaging vehicle detection systems (VIVDS) are methods that provide a direct measurement of queue length. Cheek et al. (2008) proposed a linear regression method combined with a KF to produce a more accurate estimation of queue length using VIVIDS technology. The reported accuracy of the model was 86 percent, but it was only reliable for measuring queues with a length shorter than 400 ft. Distance, visibility issues, geometric characteristics of the network, and weather conditions affected the recognition of detection lines, illustrated in Figure

2.3, and, consequently, affected the reliability of observed data. Furthermore, the high cost of infrastructure, maintenance, and data processing constrained the application of this method in real-time operation (Cheek et al., 2008).

Figure 2.3. Typical VIVID set up for queue detection (source: Cheek et al., 2008)

Harris detector cameras were utilized by Albiol and Mossi (2009) to obtain static and moving corners of vehicles and road markings to calculate the queue length (Figure 2.4). The main advantage of this method is the simple configuration and calculation process. However, the accuracy of the model suffers from the limitations and concerns mentioned for VIVIDs system (Albiol and Mossi, 2009).

(a)

(b)

Figure 2.4. a) Static corners in red, b) Queue start (source: Alboil and Mossi, 2009)

2.2.6 Neural Network-Based and Neuro Fuzzy Methods

During the last decade, artificial intelligence (AI) techniques, especially neural networks, have been used in different traffic management and modeling problems. Some examples include macroscopic modeling of traffic flow and behavior (Zhang et al. 1997, Ledoux, 1997), adaptive

traffic control systems (Saraf et al., 1996), and congestion recognition and prediction (Dougherty et al., 1993, Gilmore and Abe, 1995).

In neural network models, the neurons or nodes are organized into three categories: input, hidden, and output layers. Neurons in the input layer aggregate data from external sources that are processed in hidden layers through a pre-set mathematical process and transmitted to the outside of the network by the neurons in the output layers. The neural network is trained to optimize the connections' weights, which are the knowledge of the network. The optimization is a random process that attempts to minimize the error by comparing the outputs with the training data (Mucsi et al., 2011). The architecture of a basic neural network is illustrated in Figure 2.5.
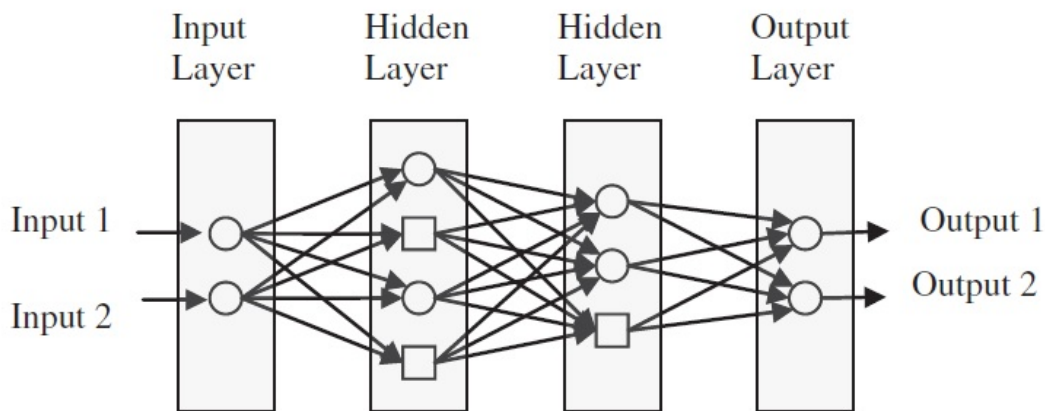
Figure 2.5. Architecture of a basic neural network (source: Mucsi et al., 2011)

Artificial Neural Network (ANN) algorithm was utilized in the prediction of queue length at signalized intersections by Chang and Su (1995). They developed the model by first identifying several interacted parameters such as signal control state, previous queue length, downstream arrivals, and the number of feed-in links. Different scenarios were designed and examined through several neural network models according to the type of information. The most

appropriate ANN model was selected based on the accuracy of the result for three second time intervals; the accuracy of the most appropriate model was 90% for the predicted queue length (Chang and Su, 1995).

Mucsi et al. (2011) developed a fuzzy logic-based neural network model to estimate the number of vehicles in a detection zone. Occupancy times and the corresponding number of vehicles were collected from three detectors in a simulated network; this information was used to train the neural network. The average accuracy of the method was less than two vehicles per lane, and it was sensitive to the length of the detection zone and the detector set-up.

2.3 Queue Length Estimation in Urban Networks and Freeways

The queue estimation approaches described in the previous section were also applied to estimate the queue behavior on street networks. Wirasinghe (1978) utilized the theory of shockwaves to determine the delays caused by traffic storage in a traffic incident on urban networks. Geroliminis and Skabardonis (2011) developed a shockwave-based approach using detector data to identify the spillovers in urban networks with signalized intersections. Identifying spillover involved detecting the queue discharge rates that are lower than the saturation flow. They also investigated the impacts of spillovers on the efficiency of traffic network operation (Geroliminis and Skabardonis, 2011). The proposed model was a monitoring approach used to identify the location of the congestion and the time taken to attract attention to active bottlenecks. However, the results may not be directly usable for designing the control strategies that require more precise information.

Considering the stochastic nature of traffic phenomena, Lu and Yang (2014) proposed a mathematical framework that formulated various events including queue formation and

dissipation, platoon dispersion, queue spillover, and blockage using probability generating functions (PGF). Comparing the results with classical analytical models, they proposed a framework based on a more realistic description of traffic flows. However, the complicated mathematical process does not consider partial lane blockages or through vehicles caused by drivers' lane changing, which affect the flow dynamics (Lu and Yang, 2014).

The relationship between queue length and roll time occupancy on an urban street network was developed by Ma et.al (2012) using S-type logistic models. Roll time occupancy is defined as the percentage of time in a time interval in which a detector is occupied. This algorithm was tested under twelve different simulation scenarios. Fluctuation of traffic flow, lane changing, and number of detectors was shown to affect the accuracy of this method. Moreover, the statistical analysis of this method restricts its application in real-time estimations (Ma et.al, 2012).

Freeways are also subjected to queue formation due to recurrent or non-recurrent congestions. Two different scenarios that may result in queue formation on freeways are active bottlenecks and variable speed limit control (VSL). Based on shockwave analysis, a discrete time-space model was developed to estimate queue length for both scenarios on freeways. The accuracy of the method was reported to be acceptable in both bottleneck and VSL cases through macroscopic simulated models. However, the reliability of this method is still questionable for the case of very long queues (Cao et al., 2015).

2.4 Queue Length Estimation in Metered and Un-metered Links

The observations of freeway bottlenecks indicate that queue spillover from on-ramps and off-ramps onto the urban links and freeway mandatory lanes is one of the main reasons

bottlenecks occur (Cassidy, 2001). Real-time information about traffic conditions, especially queue length on ramps, is crucial for designing proactive control strategies to improve on-ramp and off-ramp performance. Some of the estimation approaches discussed in section 2.1 were examined in the literature to provide queue length estimations on ramps.

2.4.1 On-ramps queues

Wu et al. (2008) compared the estimations produced by three different models including KF, linear occupancy, and Highway Capacity Manual (HCM) back of queue. The results indicated that KF and linear occupancy methods are applicable for on-ramp queue length estimation, but the HCM back of queue method is not reliable for real world applications.

A queue length estimation method proposed by Vigos et al (2006, 2008) utilized three traditional loop detectors to measure flow and time occupancy and employed a KF to calculate real-time estimates of queue length in signalized links. Loop detector design and structure is based on activating an electronic pulse when vehicle passes over the detector. The duration of produced digital signal is directly proportional to the vehicle's length and adversely proportional to the speed. As shown by Vigos et al. (2006):

$$t_j = \frac{L_j}{y_j} \tag{2.2}$$

Where,

t_j is the duration of digital pulse,

y_j is the vehicle's speed, and

L_j is the vehicle's length plus detector length.

Digital signal is transformed to a series of 1's and 0's according to a sufficiently short sampling time. The time occupancy $O_t(k) \in [0,1]$ for each time period is calculated based on equation (2.3) that relates average vehicular speed and length to occupancy readings.

$$O_t(k) = \sum_{j}^{N(k)} \frac{t_j}{T} = \sum_{j}^{N(k)} \frac{L_j}{y_j T} \qquad (2.3)$$

Where,

k is the discrete time period $(0,1,2,\ldots)$,

$O_t(k)$ is the time occupancy in the time period k,

T is the duration of time period, and

$N(k)$ is the number of passed vehicles in time period $[kT, (k+1)T]$.

Space occupancy $O_s(k) \in [0,1]$ is the portion of lane length which is covered by vehicles. Vigos et al. (2008) showed that if the traffic condition changes moderately, the space occupancy can be determined based on equations (2.4) and (2.5).

$$O_s(x) = \begin{cases} 1 & \text{if } x \text{ is covered by a vehicle} \\ 0 & \text{else} \end{cases} \qquad (2.4)$$

$$O_s = \frac{1}{\Delta} \int_0^\Delta O_s(x) dx \qquad (2.5)$$

Where,

Δ is the length of the link.

Equation (2.5) produces accurate measurements of space occupancy in uncongested traffic conditions and with occasionally stopped vehicles based on short time intervals (e.g. 20s).

Recently, Bahuleyan and Kattan (2015) improved the detector-based estimation by developing a hybrid prediction model to provide a real-time estimation of queue length on non-signalized freeway exit-ramps. The proposed model utilized two different supervised machine learning algorithms combined with a KF to predict the queue length dynamically. The results showed more than a 50 percent reduction in mean errors than with the method proposed by Vigos et al. (2008).

2.4.2 Off-ramps queues

The negative impacts of queue spillback from off-ramps onto the main freeway were investigated in a number of studies. Newell (1999), as a special case of a general theory of freeways with special lanes proposed by Daganzo (1997), analyzed the consequences of having a queue from an exit ramp back onto a freeway utilizing a graphical solution. He considered through vehicles as type 1 vehicles and those that want to exit the freeway as type 2 vehicles. Type 1 vehicles could travel in all lanes, but type 2 vehicles were supposed to drive in the right lane. In his delay evaluation method, traffic was presumed as a continuous deterministic fluid ignoring the integer nature of vehicle counts and probable variations. In a first come, first served order, the delay of n vehicles caused by the exit queue was determined from the horizontal distance at height n between the arrival (A) and departure (D) curves for both types of vehicles (Figure 2.6). This graphical solution demonstrated the issue of queue spillover and its impact on freeway flow (Newell, 1999).
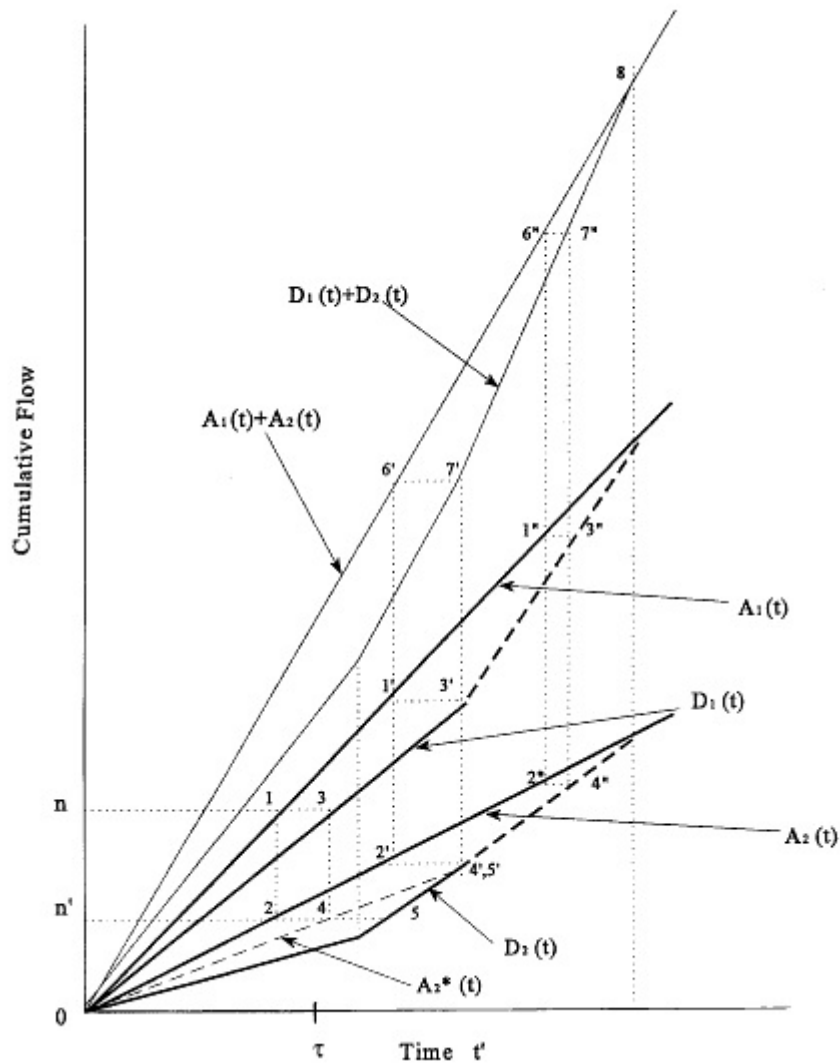
Figure 2.6. Cumulative arrival and departure curves at an upstream off-ramp (source: Newell, 1999)

The comportment of multi-lane freeway traffic on an oversaturated off-ramp was also analyzed by Manuz and Daganzo (2002) in three different situations using kinematic waves in practical examinations. They found that on wide freeways an off-ramp queue can spread across all lanes and involve through vehicles in a first-in-first-out (FIFO) system with similar speeds in all lanes. They illustrated that the flow decrease caused by FIFO off-ramp bottlenecks can be relieved by changes in the origin destination mix. Multi-lane traffic that travels at different

velocities was observed directly upstream and downstream of the FIFO queue. Semi-congested traffic situations, where some lanes are queued and others are not, also formed downstream of the FIFO queue. The outcomes demonstrated that drivers are inclined to take longer routes to their destination to avoid congested areas; smart control strategies were proposed to alleviate the negative effects of FIFO.

An off-ramp with a downstream signalized intersection can experience congestion and create a long queue that can cause a spillback onto the freeway. In such situations, the traffic flow is not homogeneous along the link. Thus, the spatial–temporal relationship between time occupancy, collected by detectors, and space occupancy is not directly useable to calculate the queue length. Qian et.al (2012) developed a method based on queuing features and queue profile on signalized links. This approach assumes an almost stationary time occupancy for the free flow and for the congested traffic situation with a linear increase in between two constants. In this method, time occupancy readings were collected from two detectors installed at an upstream entrance and in the middle of the link. The queue length was estimated based on aggregated time occupancy from the area below the occupancy profile and converting it to the link density. The accuracy of the model was reported to be dependent on the length of the queue and the verity of vehicle lengths (Qian et.al, 2012). Simplifying assumptions in this approach restricts the model's application in congested traffic conditions in which the free flow traffic is not observed for a number of consecutive time intervals and also for detecting long queues.

2.5 Discussion and Motivation

Based on the reviewed literature, the majority of studies focused on queue estimation at signalized intersections and very few dealt with metered on-ramps. Despite the significant

impacts of congested off-ramps on the performance of freeways, there is only one study in the literature focused on off-ramp queue length estimation (Qian et al., 2012). This thesis develops a novel real-time queue length estimation for freeway off-ramps with a signal downstream.

The specific theoretical basis and data requirements for prevalent queue estimation techniques restrict their application for being directly usable in off-ramp queue length estimations. The fundamental input-output models are mainly applicable for under-saturated traffic conditions and require a known arrival flow profile, which make them unsuitable in the case of long, congested off-ramps, especially for real-time queue length estimations. The shockwave-based methods are based on a robust theoretical foundation and produce acceptable estimations of queue lengths at signalized intersections according to the detectability of shockwaves in signal cycles. However, on freeway off-ramps, especially in congested traffic situations, the shockwaves may not occur for a number of consecutive time intervals.

In this thesis, a learning machine algorithm, CBR, is utilized to produce an initial measurement of queue length by comparing real-time detector occupancy percentages with a pre-defined library. A KF estimation process is also applied to correct and update measurements and also to predict the queue length in the next time step. This framework is expected to be able to reproduce reliable queue estimates for both saturated and non-saturated traffic conditions.

CHAPTER THREE: DEVELOPED COMBINED CASE-BASED REASONING AND

KALMAN FILTERING QUEUE ESTIMATION APPROACH

Figure 3.1 provides an overview of the off-ramp queue length estimation/prediction framework that combines the use of CBR and KF, which are described in detail in subsections 3.1 and 3.2. The term 'estimation' in this thesis refers to the measurement of queue length at the end of each time interval based on collected information during that time period. The term 'prediction' is used for a queue length which is anticipated to be formed by the end of the next time interval when a new set of information is not still available.
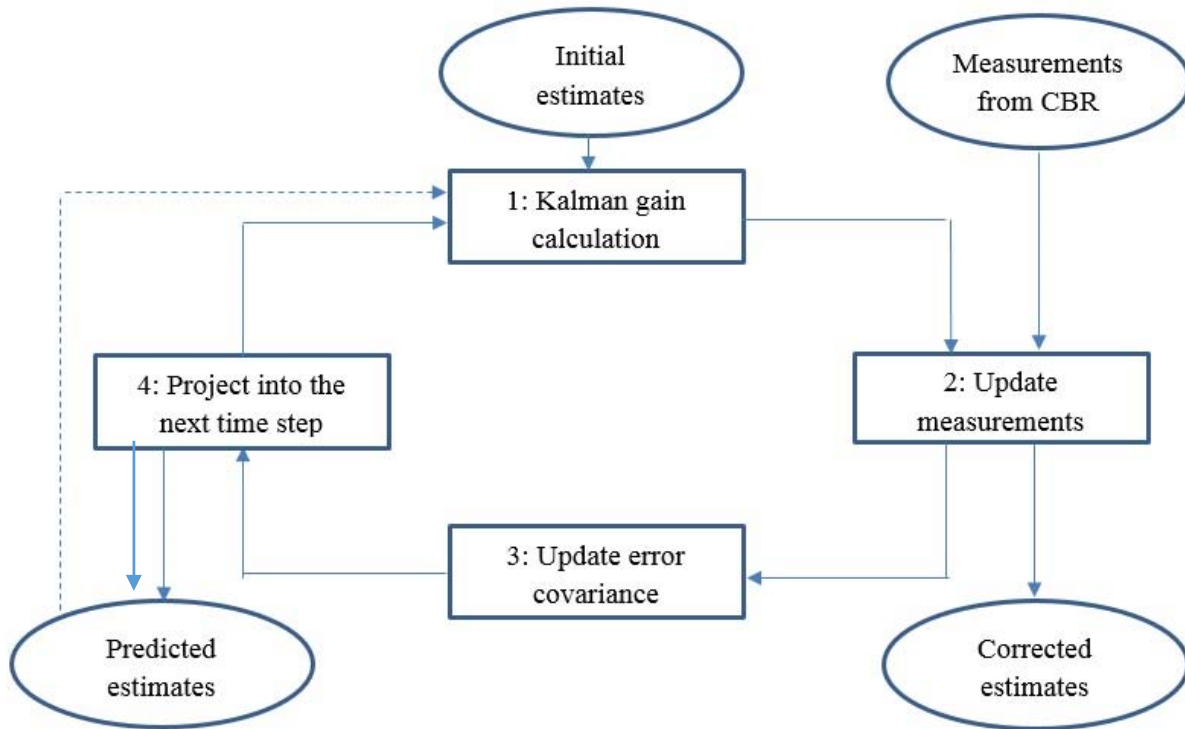
Figure 3.1. CBR-based KF Queue Length Estimation/Prediction Process

The basis of the proposed approach uses real-time occupancy readings, which indicate the percentage of time intervals in which each detector is occupied; three detectors are positioned

upstream, mid-block, and downstream. The first estimation is based on the CBR technique, which provides an indirect measurement of the queue length based on each series of occupancy readings from detectors in real time compared with a historical library of data/events; an estimate is produced according to the most similar set of data. The output of the CBR for each time interval is updated/corrected using a KF. A KF, as discussed briefly in the literature, is an optimal estimator that produces estimations from indirect, inaccurate, and noisy observations. It is a recursive process that allows new measurements to be updated/corrected once they become available.

As Figure 3.1 indicates, the KF process starts with an initial estimate of queue length for the first time interval and incorporates the measured CBR output produced by occupancy observations. The process continues to update the error covariance and to predict the queue length for the next time step. Thus, every cycle in the recursive process generates two outputs: 1) an updated estimate for the current time interval and 2) a predicted queue length for the next time interval.

In addition, the estimation prediction method developed in this thesis is implemented in a rolling horizon fashion, which means that only the estimated queue length corresponding to the early estimations are considered final. Thus, the predicated queue lengths are re-estimated in the succeeding steps in a rolling horizon fashion.

3.1 Case-Based Reasoning

Case-based reasoning (CBR) is a machine learning algorithm in AI and was developed based on the notion that similar problems should have similar solutions (Burkhard, 2001). Unlike the majority of machine learning techniques, CBR does not find a generalized relationship

between inputs and outputs of the system. The CBR process requires a library of data of past experiences including several problems (i.e. cases) and their successful solutions. When a new problem arises, the system searches in the library to find the most similar case to the new case and reuses the corresponding solution to solve it. The main advantage of CBR compared to other machine learning algorithms is that the new solved problems can be retained in the library to reduce the calculations in following iterations and to refresh the library with recent data. In other words training can be initially conducted offline using synthesized data (e.g. data generated from microsimulation) and then can resume and fine tuned online using real observed field data. The fundamental tasks of the CBR method are summarized in Figure 3.2 and follow the steps outlined below (Hossein, 2010):

Step 1: CBR retrieves similar past cases to solve the new problem;

Step 2: CBR reuses the retrieved case with minimal re-computation;

Step 3: CBR revises the suggested solution (if necessary) and finally;

Step 4: CBR retains the new experience in the system.
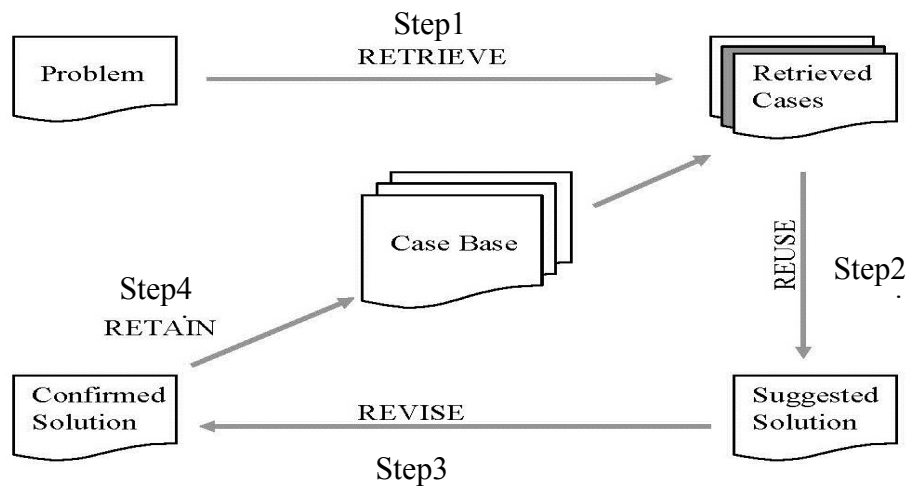
Figure 3.2. Case-Based Reasoning Process (source: Hossein, 2010)

This method was initially introduced by Schank (1983) to describe the concept of "dynamic memory," which recalls earlier patterns in the learning process. Thereafter, CBR was developed and applied in different fields, especially in the field of law. CBR was also applied to solve diverse types of transportation and traffic management problems. Instances include analyzing and evaluating the performance of various traffic control scenarios such as high-occupancy vehicle (HOV) lanes (Khattak and Renski, 1999), ramp metering, dynamic route guidance, and variable speed limits (Schutter et al., 2003). Khattak and Kanafani (1996) also developed a powerful CBR-based framework for ITS planning.

In addition, CBR has various applications in ITS planning and implementation decision-making. Dehlgren et al. (2004) combined the CBR process with an expert system to develop a knowledge-based website. Three ITS technologies were stored in the library of a CBR including a transportation demand management mechanism, automatic vehicle location/computer aided dispatch, and freeway service patrol. The users were able to access information about the locations and advantages of a specific technology and find an optimum solution (Dehlgren et al., 2004, Khattak et al., 2006). A responsive signal control design used the CBR principles to dynamically update the signal timing by recognizing the traffic pattern and assigning the appropriate signal (Hossein and Kattan, 2011).

In this research, a CBR algorithm is utilized to produce queue length measurements using occupancy readings from three loop detectors. For this purpose, a library of occupancy percentages from the same detectors with a corresponding queue length for each set of data is provided. Because the real network data was not available for this study, the library was generated using a micro-simulated network as described in section 4. Occupancy readings and queue length in the library were based on observations and calculations performed in a Paramics

micro-simulated network. Several hours of simulation were performed on the model to produce a wide variety of cases in the library.

In the proposed CBR process, every real-time set of occupancy times from the detectors is treated as a new pattern and is compared to all cases in the library. The degree of similarity for each case is calculated using equation 3.1. The solution to the equation returns a number in the range of [0, 1] where 0 and 1 represent the minimum and maximum degree of similarity, respectively.

$$S_i = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{1 + \left| T_{i_j} - T_j \right|} \qquad \text{for } i=1 \text{ to m} \qquad (3.1)$$

Where,

m is the number of cases in the library,

n is the number of detectors,

S_i is the degree of similarity for case number i,

T_{i_j} is the occupancy time for detector j in case i, and

T_j is the real-time occupancy for detector j.

The most similar case is identified by the maximum degree of similarity and the queue length is determined accordingly. This indirect measurement of queue length is prone to errors arising from the operation of the detectors and the method of calculation. As explained in the next section, a KF is used to update noisy measurements through a recursive computation process for each time interval and to predict the queue length over the next interval.

3.2 Kalman Filtering

A Kalman filter is named after Radulf E. Kalman, who developed a recursive algorithm to solve a discrete-data linear filtering problem in 1960. KF is a type of predictor-corrector estimator that minimizes the estimated error covariance based on certain predetermined conditions. KF is a very robust tool capable of controlling noisy input measurement data to produce less noisy output. As discussed in the literature, a KF is an effective algorithm to enhance the accuracy of queue length estimations.

As indicated in Figure 3.3, the KF process begins by providing an a priori estimate of the state (i.e. predicted queue length) using the inflow and outflow counts and the state equation (3.2), which represents the relationship between states in two consecutive time steps. The initial estimate is updated using the measurement to an a posteriori estimate of the state (i.e. estimated queue length); then, the next time step utilizes the a posteriori estimate to predict the state in the next time interval.

Initialize $\hat{x}_k{}^-$, P_k

Prediction

Project to the next state

State equation:

$$\hat{x}_{k+1}{}^- = \varphi \, \hat{x}_k + \beta u_k + w_k$$

Update covariance:

$$P_{k+1}{}^- = \varphi P_k \varphi^T + Q$$

Correction

Kalman gain calculation:

$$K_k = P_k{}^- H^T (H P_k{}^- H^T + R)^{-1}$$

Update the estimate via measurement:

$$\hat{x}_k = \hat{x}_k{}^- + K_k (Z_k - H \hat{x}_k{}^-)$$

Update error covariance:

$$P_k = (I - K_k H) P_k{}^-$$
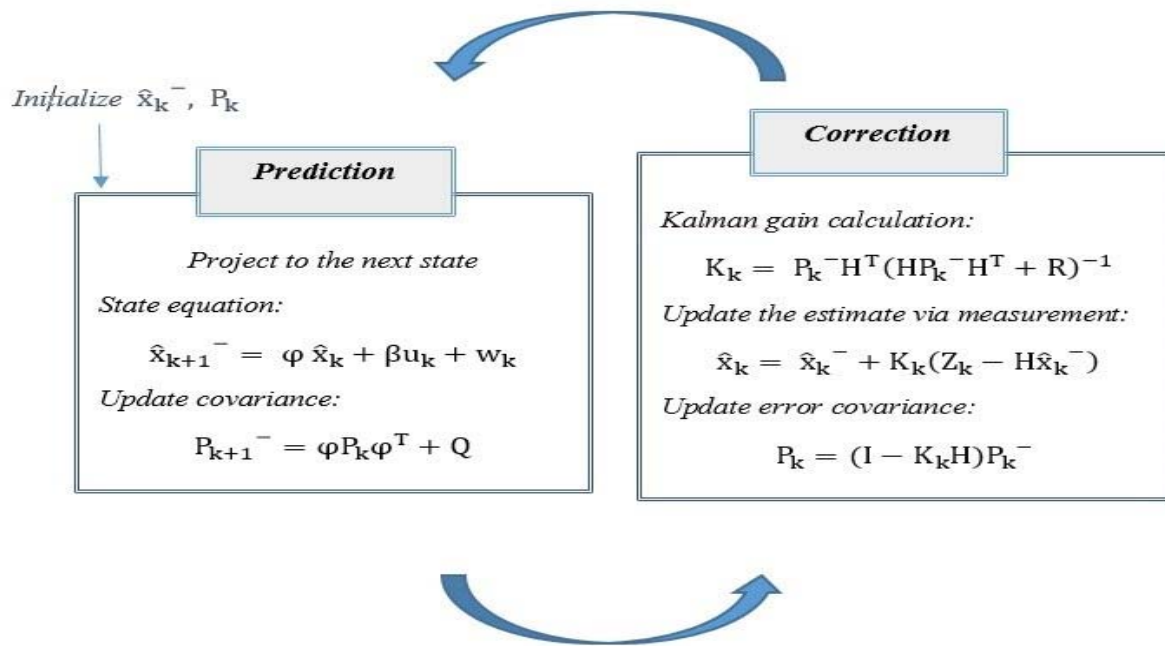
Figure 3.3. Kalman Filter Process

The state equation is written as:

$$x_{k+1} = \varphi x_k + \beta u_k + w_k \qquad (3.2)$$

Where,

x_k is the state vector of the process at time step k,

φ is the state transition matrix of the process from the state at k to the state at $k + 1$, which is assumed stationary over time,

u_k is the optional control input, which can be either constant or vary by time according to the problem,

β is the relation matrix between optional control input at k to the state at $k + 1$, and

w_k is the process white noise with known covariance of Q and probability distribution of $N \sim (0, Q)$.

$$Q = E[w_k w_k^T] \qquad (3.3)$$

Once the next measurement is observed, the KF updates the estimates using the measurement equation.

$$Z_k = H x_k + v_k \qquad (3.4)$$

Where,

Z_k is the actual measurement of x at time k.

H is the noiseless connection between the state vector, which is assumed to be stationary over time.

v_k is the measurement white noise with a known covariance of R and a probability distribution of $N\sim(0,R)$.

$$Q = E[v_k v_k^T] \tag{3.5}$$

The mean squared error (MSE) is equivalent to:

$$P_k = E[e_k e_k^T] = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] \tag{3.6}$$

Where,

P_k is the error covariance matrix at time step k,

$\hat{x}_k$ is the a posteriori estimate of the state,

The a priori estimate of $\hat{x}_k$ is called $\hat{x}_k^-$, and an updated equation is written for the new estimate combining the old estimate with the measurement data:

$$\hat{x}_k = \hat{x}_k^- + K_k(Z_k - H\hat{x}_k^-) \tag{3.7}$$

Where K_k is the Kalman gain calculated by equations (3.8) and (3.9) and (3.10):

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \tag{3.8}$$

$$P_{k+1}^- = \varphi P_k \varphi^T + Q \tag{3.9}$$

Where P_k^- is an a priori estimate of error covariance. The Kalman gain is reused in equation (3.10) to produce an a posteriori estimate of error covariance.

$$P_k = (I - K_k H) P_k^- \tag{3.10}$$

In the queue length estimation problem, the a priori estimate of the state is determined based on the fundamental traffic flow conservation equation:

$$\hat{x}_{k+1}^- = \hat{x}_k + \left(f_{in_k} - f_{out_k} \right) + w_k \tag{3.11}$$

In which $\varphi = 1$ and $\beta = 1$ and $\left(f_{in_k} - f_{out_k} \right)$ represents u_k where f_{in_k} and f_{out_k} are the inflow and outflow count on the ramp from upstream and downstream detectors, respectively.

The indirect measurement of queue length for each time interval (Z_k) is available from the CBR output as described in section 3.1, and it is directly related to the state; thus, in equation 3.4, H is equal to 1.

$$Z_k = x_k + v_k \tag{3.12}$$

The process noise w_k arising from the error in detector count measurements is assumed to be randomly distributed around zero with a covariance of Q and be independent of the count value. In this research, Q is assumed to be approximately equal to 1, which is the amount of error of the simulated detector in count measurements (Bahuleyan and Kattan, 2015). However, the impacts of changing the value of Q are assessed later as a part of the sensitivity analysis. Although this simplifying assumption may not reflect the real error, it avoids the state equation being nonlinear, which is required to develop the more complicated extended KF and does not necessarily improve the overall accuracy of the model (Vigos et al., 2006). If the real-time actual queue lengths were available, it would be possible to calculate the process noise covariance Q dynamically. Because such information is not provided, Q is assumed to be constant over time, which stabilizes the Kalman gain K and the error covariance P after a few iterations.

The measurement noise covariance (R) is even more complicated to produce according to different causes of errors. Measurements may include noises caused by modeling errors, accuracy and adequacy of the case library, and detector measurement errors. Hence, R is assumed to be a constant and not dependent on the measured value of queue length, which fixed *K* and *P* after a few iterations. In this research, different R values are also examined to find a value that produces more accurate results.

CHAPTER FOUR: **NETWORK SIMULATION AND RESULTS ANALYSIS FOR CASE STUDY**

In this chapter, the developed framework is examined on a simulated network in Quadstone Paramics Modeller software. The results of the simulations evaluate the accuracy of the model and its application in real-world traffic networks. Paramics Analyser is also utilized to calculate the maximum queue length in each time interval to produce the library of cases. Different values of KF parameters are used to examine their impacts on the estimation results. A sensitivity analysis is performed to optimize the KF parameters and to evaluate the effects of various measurement errors on the accuracy of the results. Moreover, changes in traffic demand and its impact on the accuracy and reliability of the model are examined. The performance of the proposed method is also evaluated for changes in time intervals and number of detectors to determine how the duration of time steps and number of detectors affect the accuracy of predictions and estimations.

4.1 The Study Area

The proposed method is examined through a micro-simulated model in Paramics software using a freeway off-ramp on Deerfoot Trail in Calgary, Alberta, Canada. Deerfoot Trail is the major north-south transportation freeway through the City of Calgary and composes a section of the Queen Elizabeth II Highway, also known as Highway 2. The parameters of the Paramics model (i.e., the mean headway factor and the mean reaction time) were previously calibrated by the transportation group at the University of Calgary (Kattan and Saidi, 2013). A 2006 seed origin-destination matrix provided by the City of Calgary was also calibrated using the Paramics Modeler to obtain the 2012 vehicle count provided on the Alberta Transportation website. An illustration of the study area is presented in Figure 4.1.

Figure 4.1 The Study Area: (southbound Deerfoot Trail to Westbound McKnight Boulevard, Calgary, Alberta, Canada)

The target ramp connects southbound Deerfoot Trail to westbound McKnight Boulevard, and has a length of 650 meters. The signal downstream controls the intersection of McKnight Boulevard and 4 Street NE with a pre-timed five phase signal plan and a cycle length of 106 seconds with a green phase of 75 seconds for the westbound traffic. This intersection is coded in Paramics. For the purpose of the experiment, three detectors are installed in Paramics at the two ends and at the middle of the ramp on the model; the detectors record the occupancy and volume count in predefined time intervals. The downstream detector was installed on the ramp at a distance of 450 meters from the signal followed by the mid-block and upstream detectors at distances of 800 meters and 1050 meters from the signal, respectively. The model was also able to measure the queue length on the ramp to provide an accurate simulated measure referent to be used as a basis of comparison. The simulated network, including the location of detectors and the downstream pre-timed signalized intersection, is represented in Figure 4.2.
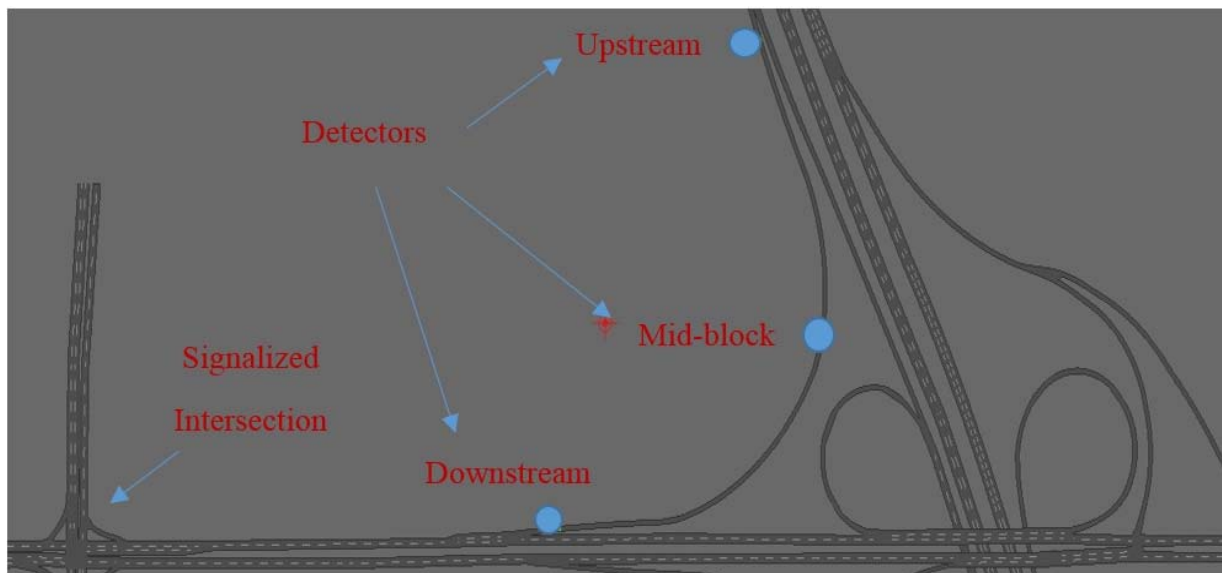
Figure 4.2 Testbed simulation network (southbound Deerfoot Trail to westbound McKnight Boulevard, Calgary, Alberta, Canada)

4.2 Development of the Queue Length Case Library

The CBR algorithm requires a library of cases with corresponding solutions, which, in this study, consist of occupancy readings from detectors and the corresponding queue length. Because the historical data of such measurements was not available, a set of cases was produced and updated utilizing the Paramics model over 20 hours of simulation. Paramics Modeller gathers the occupancy time in seconds, every second, from three detectors and reports the readings in required time intervals. A simple application programming interface (API) is developed to directly produce the occupancy percentages based on the reporting period duration. The maximum queue length on the links that form the ramp are calculated utilizing Paramics Analyser for the same time intervals; the case library is produced by combining the two sets of reports for 10 simulation runs. The main advantage of the CBR library in the proposed model is its simple structure, which allows the library to easily update and reproduce the cases for various traffic conditions or estimation intervals.

Loop detectors must be installed at the desired locations on the ramp to gather occupancy readings to develop the case library for real-world applications. For queue length estimations based on a CBR library, it is not necessary to have real-time estimations. Any off-line queue detection method that produces reliable estimates of queue lengths in a specific time interval would be applicable. Video–based queue detection methods, such as using existing Closed-Circuit TeleVision (CCTV) cameras or portable cameras that are capable of being processed, are an example of such estimation techniques. The developed library can be periodically updated based on traffic pattern variations such as seasonal fluctuations. It is recommended that traffic conditions during special events also be recorded and added to the library, which makes the model applicable for queue length estimation in the case of non-recurrent traffic congestions.

4.3 Experiment Design

The combined CBR and KF process is implemented with a Microsoft Visual C++ 10 platform and integrated with Paramics using an API to read the required data from the simulated network dynamically. The CBR algorithm assigns a queue length from the library to each set of real-time occupancy readings, which are used as the measurement equation in KF. Thereafter, the measurement is corrected through the filter and reported as the real-time queue length estimation. The a posterior estimations that are less than 1 are set equal to zero to avoid unnecessary corrections on zero queue lengths.

The fundamental analysis is conducted to estimate the queue length based on a morning peak hours demand matrix, which was previously calibrated by the transportation group at the University of Calgary. The simulation was run for 2 hours to examine the performance of the developed queue estimation/prediction approach using 1 minute estimation intervals.

As discussed in subsection 3.2, the measurement noise covariance (R) is assumed to be constant over time and not dependent on the measured value of queue length that fixes K and P after a few iterations. In this research, the first set of experiments is designed to find the optimum R value based on the minimum error in the results. For this purpose, various R values from 1 to 5 are examined to fine tune the KF process. The process noise covariance (Q), which mainly arises from the error in detector count measurements, is assumed to be equal to 1, which is the number of errors that occur in the simulated detector in terms of count measurements (Bahuleyan and Kattan, 2015). As the value of Q changes through the range 1 to 5, the effects are examined by conducting a set of experiments. For real-world applications, the value of Q can be assumed to be approximately equal to the error of count measurements based on the specification of detector devices.

A third set of experiments is designed to evaluate the impacts of the level of congestion on the accuracy of estimations. For this purpose, simulation is conducted based on high vehicular demand of non-recurrent congestion in the entire network, which is assigned equal to 120% of the total demand matrix. The same analysis is also performed for a low level of demand by releasing 80% of AM peak demand to the simulated network; the results are compared to the main analysis outputs. The duration of the time interval is also an essential parameter in the queue length estimation process; thus, the performance of the proposed method is evaluated using different time intervals.

The mean absolute error (MAE) and the mean absolute percentage error (MAPE) are utilized to evaluate the accuracy of the method as well as to compare the results from other experiments. The mentioned measurements of performance are calculated based on equations 4.1 and 4.2.

$$MAE = \frac{1}{n} \sum_n |Obsevation - Estimation| \tag{4.1}$$

$$MAPE = \frac{1}{n} \sum_n \left| \frac{Obsevation - Estimation}{Observation} \right| \times 100\% \tag{4.2}$$

4.4 Model Results for AM Peak Demand

The fundamental analysis is conducted for AM peak demand using 60 seconds time intervals based on the corresponding queue length case library. As discussed in subsection 4.3, the initial analysis is designed to examine the effects of the measurement noise covariance (R) and process noise covariance (Q) on the MAE; the optimum values are used in the remaining experiments.

4.4.1 Adjusting Kalman Filter parameters

A KF as an estimator/predictor process requires the accurate value of actual measurements of the estimated parameter to calculate the real noise covariance in consecutive time intervals; the actual measurements are not available in queue estimation problems. Vigos et al. (2006) showed that the Kalman gain in each time interval does not depend on precise values of Q and R, but only depends on the ratio of $\frac{T^2 Q}{R}$ where T is the duration of time interval; thus, they directly examined the various values of K to improve the KF. In this thesis, it is suggested that the optimum value of R be calculated because the value of Q is assumed to be known and is mainly dependent on detector errors, which are identifiable.

First, the value of Q is set equal to 1 based on the detector count measurement error and the value of R is increased from 1 to 5 in every simulation. As illustrated in Figure 4.3, the analysis of the simulation results shows that when R=2, the minimum MAE is obtained for estimated queue lengths and predicted queue lengths and is equal to 3.15 and 4.31 vehicles, respectively. This value of R stabilizes the Kalman gain (K) at 0.5 and the value of mean error covariance (P) at 1. Thus, R=2 is utilized in all experiments and the sensitivity analysis in the rest of this thesis.

Process noise covariance is also examined through several simulation experiments to evaluate the effects of the changes in Q on the MAE. As expected, the MAE increases with an increase in the process error covariance. For Q=1, which means that the process noise is normally distributed with a mean value of 0 and a covariance of 1, the estimation accuracy is ± 3.15 vehicles. This value increases to ± 4.29 for Q=5 as illustrated in Figure 4.4
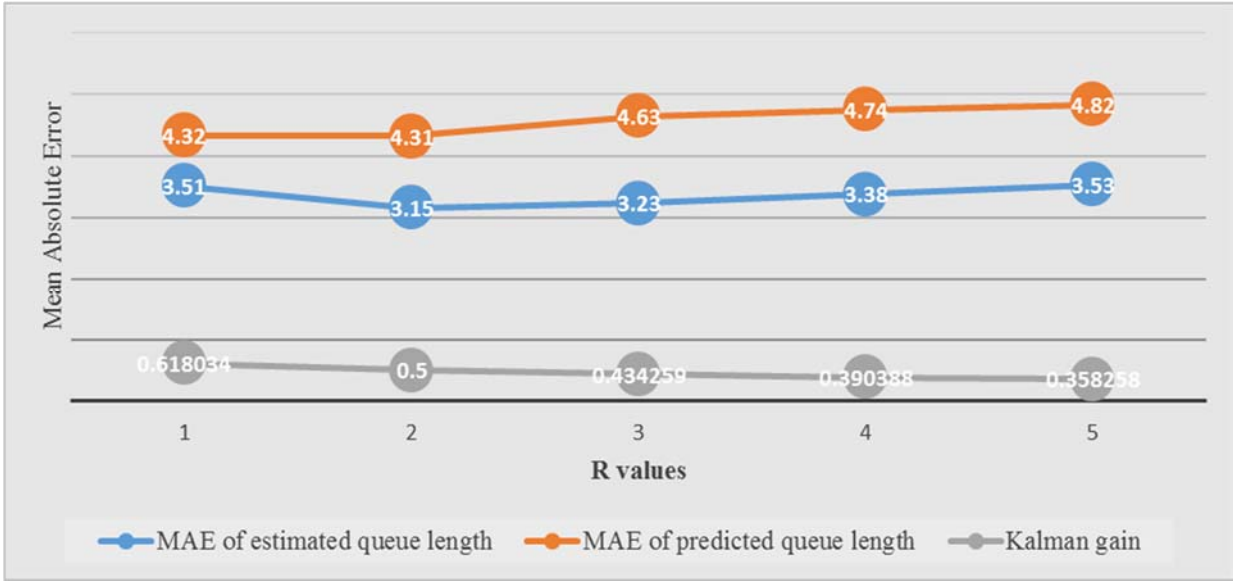
Figure 4.3 Mean absolute errors and Kalman gains for various values of measurement noise covariance (R)

.
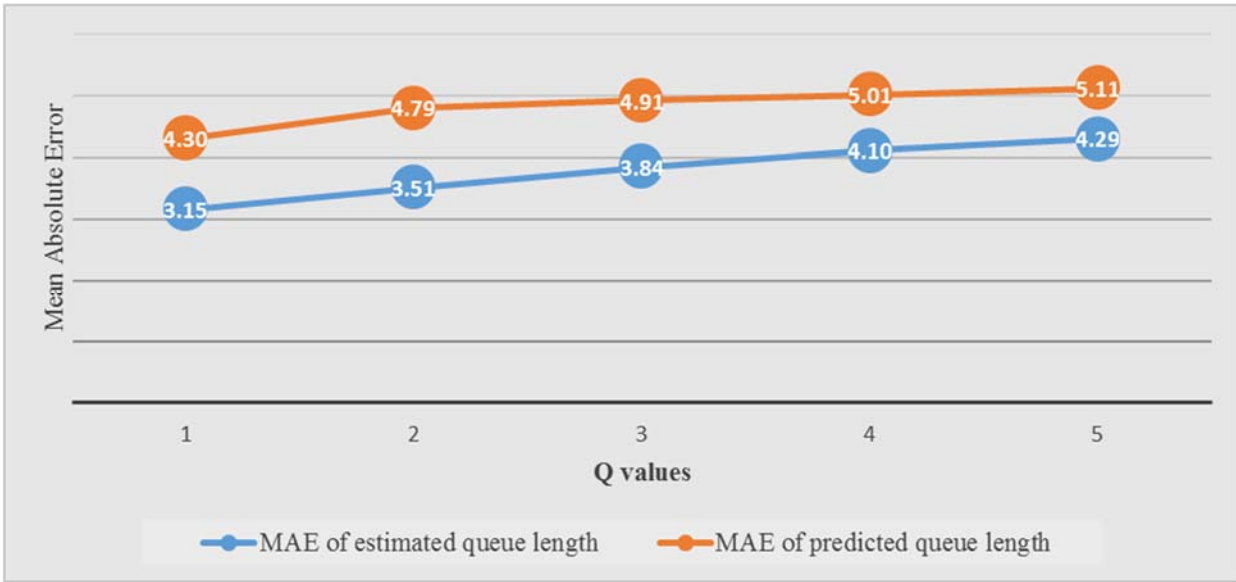
Figure 4.4 Mean absolute errors for various values of process noise covariance (Q)

4.4.2 Model Results For AM Peak Demand

The simulation is conducted from 6:30 to 8:30 AM and for the AM peak demand in 1 minute time intervals. An analysis of the results shows some fluctuations in the queue length during the first hour of simulation with a maximum queue length of 20 vehicles on the ramp. During the second hour, traffic is congested with a queue length varying between 60 to 95 vehicles on the ramp.

The queue lengths produced by the CBR algorithm show the MAE is equal to 5.63, which means that the number of vehicles in the queue is measured with an accuracy of ±5.63 vehicles on the entire ramp. The KF process significantly improves the estimation by reducing the MAE to 3.15 vehicles in the two hour simulation. Standard deviation of errors for estimated queue length is equal to 4.3 vehicles. This level of accuracy is satisfactory for a ramp with a length of 650 meters with a maximum queue of about 98 vehicles.

The simulation results, including the CBR estimations, KF improvements, and predicted and actual queue lengths, are illustrated in Figure 4.5. As shown in the figure, the model is capable of reliably estimating the queue length in the time period after 7:20 AM, which represents a congested traffic condition with a particularly long queue. The developed model is also capable of predicting the queue length for next time interval through a KF process. The MAE, comparing actual queue length and predicted queue length, is equal to 4.30 vehicles, which is acceptable to be used in proactive control strategies. Standard deviation of errors for predicted queue length is equal to 5.2 vehicles.
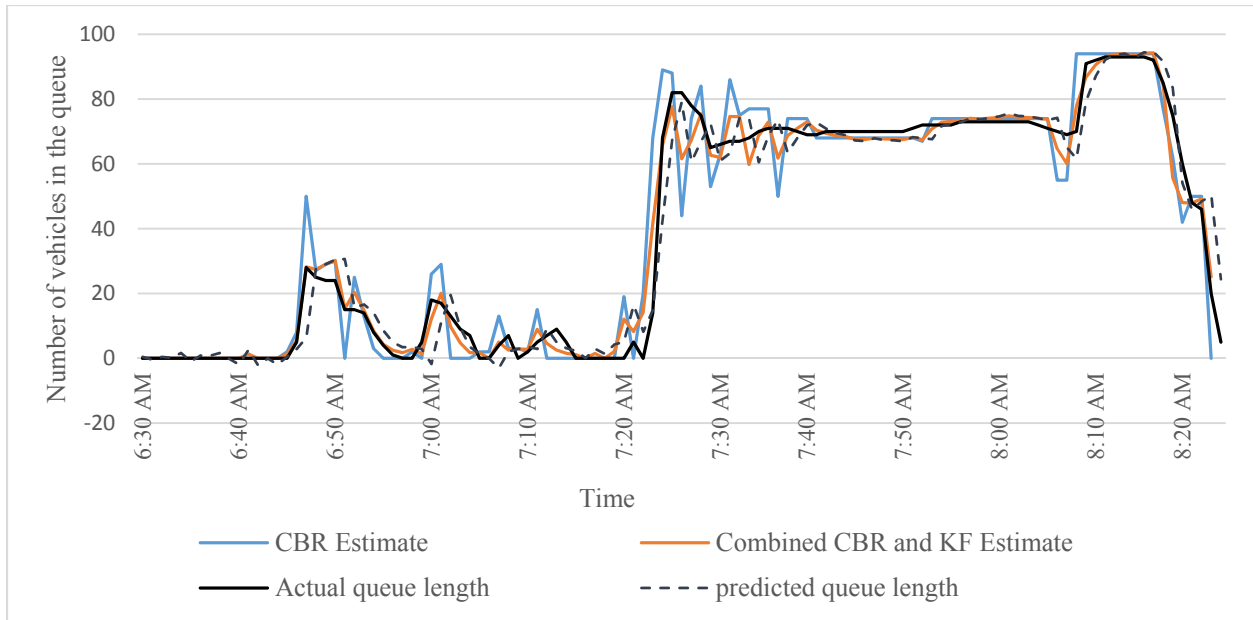
Figure 4.5 Estimated and predicted queue length using CBR and combined CBR and KF vs Actual queue length

There is only one analogous study in the literature, conducted by Qian et al (2012), on queue length calculations for off-ramps. In this queue-profile based method, Qian reported a MAPE of 24% on a ramp with a length of 100 meters in 60 seconds time intervals. For comparison purposes, the MAPE, calculated for the proposed method using equation 4.2, is 14.44% for an off-ramp with a length of 650 meters, which is more than six times longer than in the analogous study; the same time intervals were used.

CBR algorithm is able to update the case library after every successive estimation by adding the new set of data and corresponding solution (i.e. queue length). In this case study the real-time occupancy readings from three detectors and the queue length estimated by combined CBR and KF method are added to the library in every time step as a new case. Dynamically updating the library improved the accuracy of CBR measurements and KF estimates about 5% to a MAE of 5.34 and 3.0 vehicles respectively.

However, in real world applications, various issues may affect the applicability of dynamic updates in the library. For instance, growing the library increases the computation and estimation time and it requires to be limited by a specific threshold. On the other hand, one may control the size of the library by replacing the preliminary recorded cases with the new produced estimations, which may result in inserting the estimation errors into the case library. Thus, it is recommended to update the case library utilizing accurate measurements of queue length in required time periods rather than dynamically updating based on estimations.

4.5 Sensitivity Analysis for Changes in the Level of Demand

Simulation results are analyzed based on 120% and 80% of demand to evaluate the effect of the level of demand on the accuracy of the method. The results of the analysis are compared to the main analysis, which is based on 100% of AM peak demand.

4.5.1 Model Results For Non-recurrent traffic Congestion

In the case of high volume traffic, the simulation results show the formation and dissipation of queues with a maximum length of 40 vehicles until 8:00 AM. After that time the queue becomes more congested and rapidly grows to a length of 98 vehicles on the ramp; these high numbers block the ramp and continue to the end of the simulation period.

Figure 4.6 illustrates the estimated and predicted queue length compared with the actual queue length. The results shows that MAE = 5.62 for the CBR queue measurements and MAE = 4.10 for the combined CBR and KF estimations. The CBR based queue measurements have the same accuracy for both 100% and 120% because the pattern matching basis of CBR is based on comparing occupancy percentages and does not depend on various traffic characteristics or demand levels. Thus, the algorithm produces the same results for the same set of occupancies

from either level of traffic volumes.

However, the results of the combined CBR and KF method are more accurate than the CBR method alone; the accuracy of the CBR and KF method is ±4.10 vehicles for 120% of demand compared to the MAE of 3.15 for the case of 100% of demand. The predicted queue length also has an accuracy of ±5.17 vehicles, which is not as accurate as the reported prediction accuracy for the main analysis case.
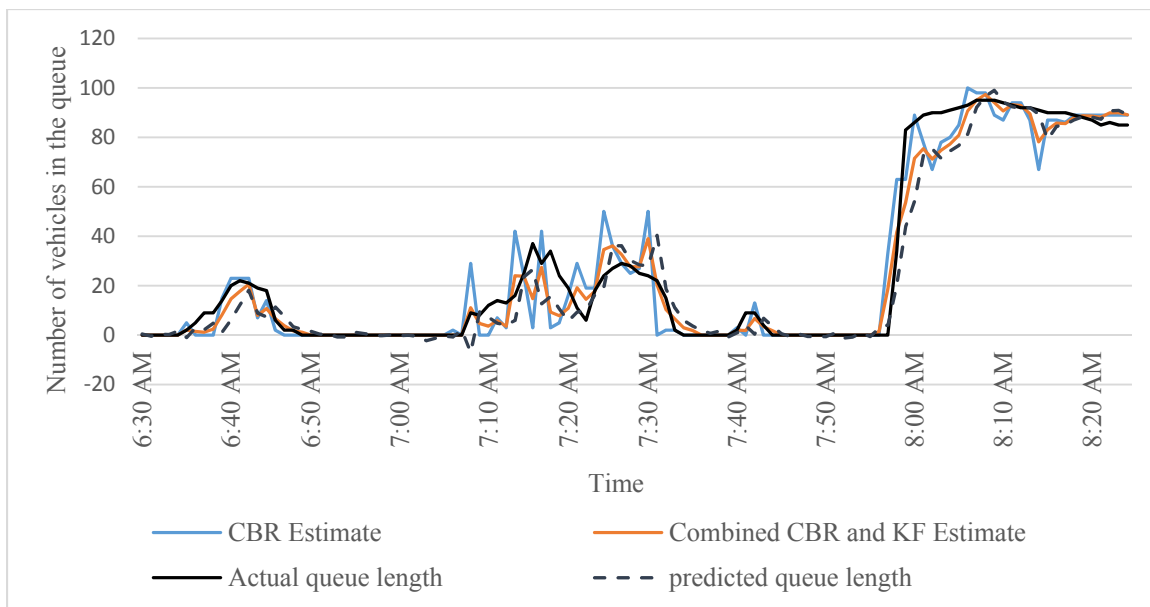
Figure 4.6 Estimated and predicted queue length for non-recurrent congestion

The developed model underestimates the queue length for high congestion levels as illustrated in Figure 4.6 because the case library is mainly developed based on the AM peak demand in this research; thus, it may not include a sufficient range of congested cases to match the real-time occupancy readings in highly congested traffic situations. The library needs to be updated by adding more cases of queue length for various levels of traffic volume to obtain better results.

4.5.2 Model Results For 80% of AM Peak Demand

Figure 4.7 illustrates the queue length estimations in the case of 80% of AM peak demand. In this experiment, the queue starts to form at 7:18 AM and grows to a maximum of 67 vehicles for a short period of five minutes and then completely dissipates at approximately 7:40 AM. The proposed method estimates the queue length in this time period with an accuracy of ± 2.88 vehicles using the CBR algorithm, and a significantly improved MAE of 1.55 is calculated for the combined CBR and KF estimation. The predictions of queue length also show an acceptable accuracy of ± 2.39. As previously mentioned, the library in this research includes more cases of regular AM peak demand than under load and overload situations. Therefore, the model overestimates the queue length in low traffic demand, but has an acceptable level of accuracy. A comparison of estimation errors for the discussed traffic situations are represented in Table 4.1.

Table 4.1 Mean absolute error of estimations for various levels of demand

Level of demand	CBR estimation	Combined CBR and KF estimation	KF prediction
80%	2.80	1.55	2.39
100%	5.63	3.15	4.30
120%	5.62	4.10	5.17
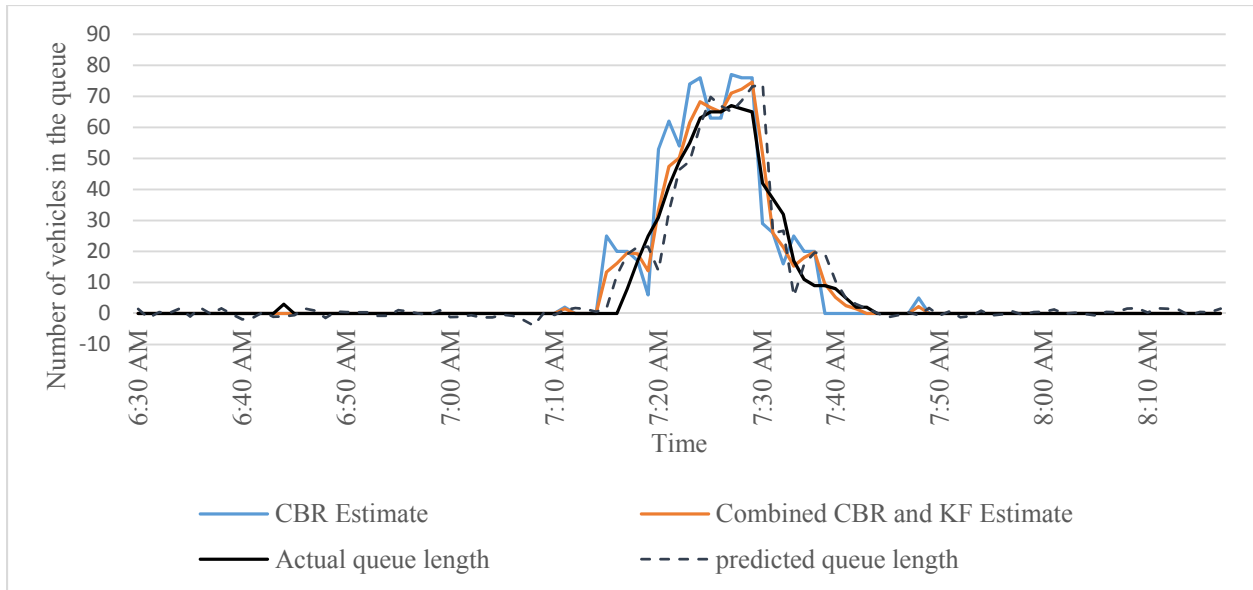
Figure 4.7 Estimated and predicted queue length for 80% of demand

4.6 Sensitivity Analysis for Changes in Estimation Time Intervals

The developed queue estimation method is examined for 106 seconds time intervals which is the cycle length of the downstream signal and also for two minutes time intervals to evaluate the performance of the model in longer time step estimations. As discussed in subsection 4.4.1, the Kalman gain is directly related to the squared time interval; thus, changing the time step may affect the optimum value of R. First, the KF is adjusted and improved using the same approach that was utilized for one minute time intervals and the optimum R value is calculated to be equal to 1 for both experiments.

4.6.1 Model Results 106 Seconds Time Intervals

Figure 4.8 illustrates the analysis results for queue estimations in the time steps with the duration of downstream signal cycle which is 106 seconds with 75 seconds of green time for westbound traffic including freeway exiting vehicles. The analysis of errors shows the MAE = 7.27 for CBR estimations and 5.07 and 7.51 for the combined method estimations and

predictions, respectively. The accuracy of ± 5 vehicles is acceptable for estimating a long queue with more than 90 vehicles. However, increasing the length of time intervals reduces the accuracy for predicted queue length which may not be acceptable for precise queue prediction purposes.
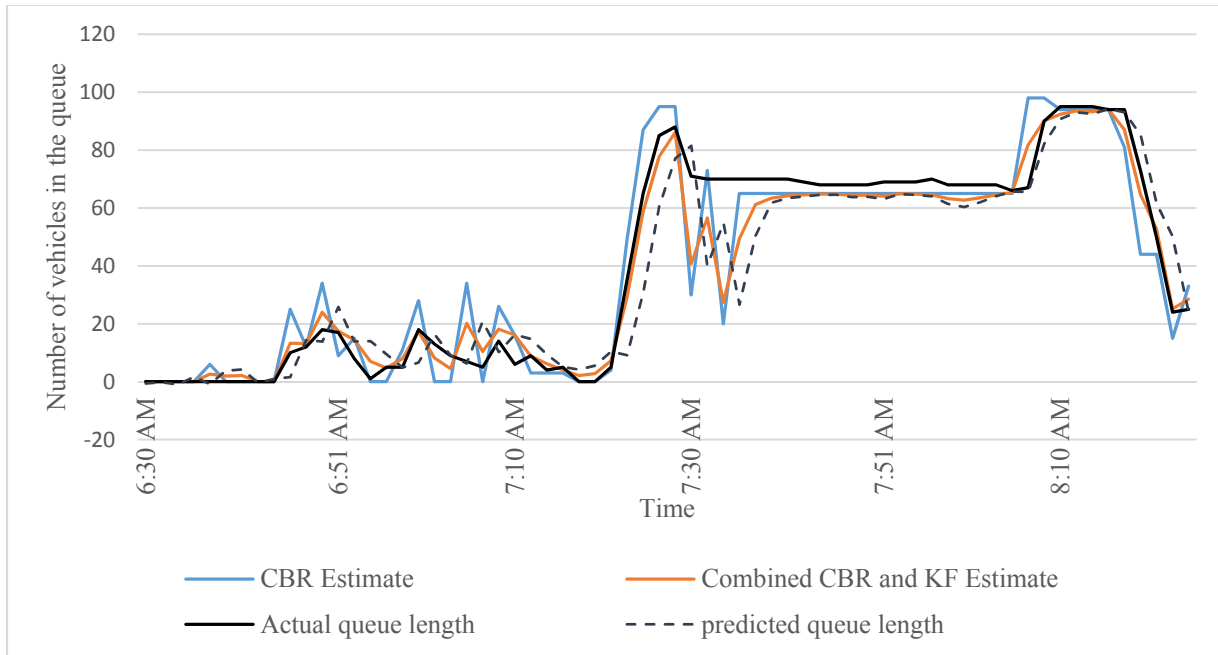
Figure 4.8 Estimated and predicted queue length for 106 seconds time intervals

4.6.2 Model Results 120 Seconds Time Intervals

The analysis of errors shows the MAE = 8.74 for CBR estimations and 7.51 and 8.93 for the combined method estimations and predictions, respectively. In the majority of prior experiments, the accuracy of predictions was reported to be higher than the accuracy of CBR based measurements. However, in the case of two minutes time intervals, the accuracy of predictions is slightly lower than the accuracy of the CBR measurements, which shows that the KF performs better for more frequent estimations. The estimated and predicted queue lengths for

2 minute time intervals are illustrated in Figure 4.9.

The overall performance of the model for 2 minutes time intervals may not be acceptable for precise queue estimation purposes. However, the model is applicable for queue management strategies, which do not require the precise queue length.
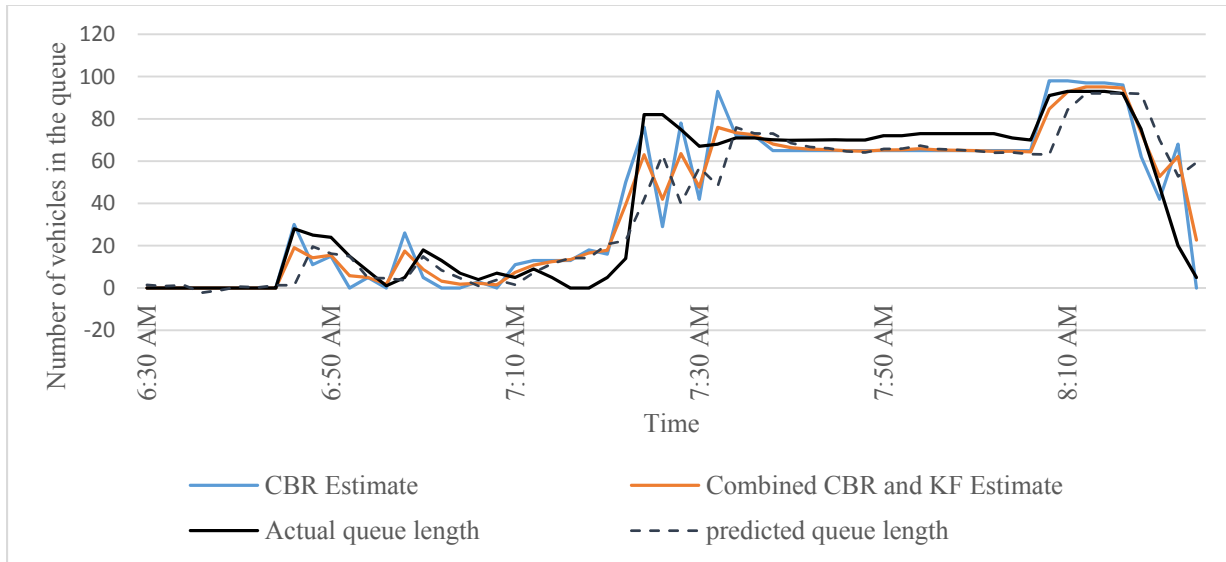
Figure 4.9 Estimated and predicted queue length for 120 seconds time intervals

4.7 Sensitivity Analysis for Changes in the Number of detectors

The number of detectors is increased to 4 detectors on the ramp to examine its impact on the accuracy of queue estimations and predictions. Analysis of the results shows a significant decrease in the accuracy of CBR measurements to ±20 vehicles which is not acceptable. Reduction in CBR measurement accuracy results in low level of performance for the combined CBR and KF utilizing 4 detectors with a MAE= 18.5. These findings might seem counterintuitive as increase observability is expected to results in increase accuracy. This reduced accuracy may be explained by the pattern matching process of the CBR which is more

complicated when more parameters are utilized in a set of data to be fitted to the real-time cases. In addition, the CBR algorithm, used in this thesis, calculates the degree of similarity by treating all occupancies in a similar way. However, adding more information such as priority weights and/or more parameters such as vehicle length speed or information from other sources, for training the CBR may improve the performance of the algorithm. It is to be noted that similar results were reported by Bahuleyan and Kattan (2015) who used a hybrid ANN and KNN based method. Vigos et. al., (2008) examined three to nine detectors to estimate the queue length and reported fluctuation in the amount of error by increasing the number of detectors with an increase in the error with six detectors. These findings show the limitation of machine learning algorithms and their focus on merely pattern matching rather than establishing causality relation. . Thus, the optimal number of detectors for using the proposed model is three detectors in off-ramp queue length estimation.

4.8 Summary of the Results

Chapter 4 examined the proposed CBR and KF queue estimation method on a freeway off-ramp in Calgary, Alberta, Canada, which was simulated in Quadstone Paramics microsimulation software. The estimated queue in this research is the maximum queue length on the ramp for a specific time interval. The CBR case library was developed utilizing data from 20 hours of simulation.

The main experiment was conducted for AM peak traffic demand and based on 1 minute time intervals. The Kalman filter was first fine-tuned and the estimation results were then compared to the actual queue length as calculated in the simulation. It was found that the model is capable of estimating and predicting the queue length with an accuracy of 3.15 and 4.3

vehicles, respectively.

The results were compared to an off-ramp queue length estimation study, the only study found in the literature, and it showed a significantly higher performance for the developed method in this research. Further, a longer ramp with a more congested traffic condition was examined in the proposed method.

Several sensitivity analyses were performed to evaluate the model performance at various traffic congestion levels and for different time intervals. The results showed that the accuracy of the estimations is acceptable for either a low or high level of congestion. However, longer time intervals are not recommended for precise queue length estimation purposes using the proposed method. Increase in the number of detectors was also examined and the results showed a low level of accuracy when using 4 detectors.

In summary, the above experiments demonstrate that the combined CBR and KF queue estimation model produces robust estimations and predictions of long queues on freeway off-ramps. In addition to its superior performance compared with the one other method, results from the developed model do not depend on the traffic characteristics and the cause of congestion. Lastly, the accuracy of the results changes for a high level of traffic congestion.

CHAPTER FIVE: **CONCLUSIONS, CONTRIBUTIONS, AND FUTURE WORKS**

This chapter focuses on the outcomes, concluding comments, and contributions of this research and suggests directions for future research and further developments. Section 5.1 presents a brief description of the research. Overall conclusions and a summary of the results are discussed in this section. Section 5.2 explains the implications and novel contributions of the research. Finally, in section 5.3, the author's perspective on development directions and future research are presented.

5.1 Research Summary and Conclusions

This study developed a framework that integrated case-based reasoning (CBR) as a machine learning algorithm and a Kalman filter (KF) estimator to provide real-time queue length estimations and predictions for freeway off-ramps. The proposed model uses occupancy percentages from three loop detectors to provide real-time queue length estimations. An initial estimation was provided through CBR using a library of occupancy and queue length data. A KF improved the CBR estimations by integrating the flow reservation equation and the state equation into the calculation and also by predicting the queue length for the next time interval.

The proposed framework was examined on a simulated network in Quadstone Paramics using the exit ramp from southbound Deerfoot Trail to westbound McKnight Boulevard, which has a length of 650 meters. A case library of occupancy percentages and queue lengths was developed and updated from a simulated network; this library was used in the CBR process. Several experiments were conducted to improve the Kalman filter parameters and also to

evaluate the performance of the model in various traffic conditions and for different time intervals and change in number of detectors.

The analysis of the results from the case study represented an accuracy of ±3.15 vehicles for the estimated queue lengths using the combined CBR and KF method for a ramp with a length of 650 meters. The KF algorithm also predicted the queue length for the next one minute interval with the mean absolute error (MAE) of 4.3, which is acceptable to be used in proactive signal design for the signal downstream and other dynamic queue management strategies. The results were compared to another off-ramp queue estimation approach, and a mean absolute percentage error (MAPE) of 14.44 % was determined for the proposed method versus 24 % for the other approach.

In the case of non-recurrent congestion, loading the network with 120% of AM peak demand blocked the off-ramp; however, the results showed the MAE to be equal to 4.10 for the target ramp, which is satisfactory for the long test-bed ramp in congested traffic conditions. Decreasing the demand to 80% of the regular demand significantly reduced the MAE to 1.55 for 60 seconds time intervals. The performance of the model was also examined for longer time intervals of 106 seconds which is the signal cycle length and also for 120 seconds, which reduced the accuracy of the estimations and predictions but still applicable for queue management strategies.

The results showed the model performed well when estimating long queues on freeway off-ramps at various congestion levels. In addition to the reliability of the model, the proposed model had the advantage of avoiding assumptions regarding vehicle length or type. It also did

not require the application of a specific trend or equation in the model, which allowed it to be applied in different traffic situations and combinations.

The main limitation of proposed framework is the considerable data requirement to produce and update the CBR case library. However, this data is important to improve the accuracy of the estimation. Another limitation of the study is the reliance on simulated data as the basis of the analysis.

5.2 Research Contributions

Previous queue length estimation techniques mainly focused on signalized intersections that were dependent on signal timing with a recursive process of evolution and dissipation. Few studies in the literature discuss on-ramps and only one study proposes a method for off-ramp queue length estimation. However, in the off-ramp study, only short ramps are examined, and they are found to be very sensitive to the level of congestion and traffic behavior. In addition, the approaches developed for signalized intersections are not directly applicable to long off-ramp queues that extend beyond the upstream detector.

The proposed model in this thesis is a novel real-time queue length estimation and prediction method targeting long freeway off-ramps. The integrated model produces reliable estimations of the length of queues in congested and uncongested traffic conditions. Thus, this thesis makes the following contributions to the literature in the field of transportation engineering:

i. *Estimating and predicting the queue length on long freeway off-ramps in the case of recurrent congestions and non-recurrent events corresponding to extreme fluctuations in traffic demand*

The proposed framework relies on occupancy and queue length information in the case library and it does not require the development of a generalized equation that may not be applicable in various traffic conditions. If the corresponding information of samples of recurrent or non-recurrent events is recorded in the case library, the model is able to estimate and predict the queue length for the long target ramp.

ii. *Utilizing the CBR machine learning algorithm to produce indirect measurements of queue length based on a library of cases*

In the literature, CBR algorithms are utilized to solve transportation engineering problems and ITS technology; however, their application in queue length estimation is examined for the first time in this thesis. Machine learning algorithms such as ANN and KNN are previously used in the literature for queue estimation purposes, but such techniques are more complicated in the training, implementation, and updating process compared to CBR.

iii. *Updating and correcting CBR measurements dynamically by integrating a KF process to improve the accuracy of the model*

The CBR algorithm estimates the queue length by finding a similar case without the need to incorporate the traffic flow of previous or current time intervals. Integrating a KF process augments the fundamental flow conservation equation in the model to control and update the CBR measurements at every time step.

iv. *Predicting the queue length over a short period of time*

The KF state equation predicts the queue length in the next time interval based on the queue length in the current interval and inflow and outflow traffic counts; thus, the prediction time horizon is as short as the duration of the time interval.

v. *Conducting sensitivity analysis to examine the performance of the developed approach for various levels of measurement errors, demand fluctuations, estimation intervals, and number of detectors*

Several experiments are designed and implemented to improve and evaluate the performance of the proposed queue estimation method. The KF is adjusted through five simulation runs for five different values of measurement noise covariance and, then, the impact of an increase in process error covariance is examined. The performance of the model is also evaluated at various levels of demand, changing estimation intervals and increasing the number of detectors.

vi. *Possibility of dynamically updating the library of events to expand the application of the model as well as to reduce the computational iterations and improve the performance of the model.*

An experiment was conducted to improve the accuracy of CBR measurements by updating the case library in every time step based on real-time occupancies and estimated queue length. The limitations of dynamic updates in the real world applications were also discussed.

5.3 Future Research Scope and Recommendations

Future research should consider testing and validating the algorithm in the field using real-world data. The queue library in real-world applications may be generated using any off-line, but precise, queue estimation techniques combined with real loop detectors installed on the ramp. The time duration for gathering the cases and also the frequency of updating the library

depend on the weekly and annual traffic fluctuations. Performance of the CBR algorithm can be improved by adding other parameters to the pattern including vehicle length, speed, collected by other devices as well as information on weather and road surface conditions.

In addition, the operation of the downstream traffic signal is clearly reflected in the behavior of the queue; thus, the timing plan is the main factor affecting the size of the generated queue. Therefore, the integration of the developed queue length prediction algorithm with a responsive or adaptive downstream signal control to manage the queue will be examined in future research.

Furthermore, the queue length estimation techniques developed in the literature for metered and unmetered on-ramps were shown to be very sensitive to the length of the ramp and the traffic volume and behavior. Because the proposed method in this thesis follows a simple and independent process, it is suggested that long on-ramps be examined to produce real-time estimations of long queues.

REFERENCES

Albiol, A., & Mossi, J. M., 2011, "Video-based traffic queue length estimation". In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference, 1928-1932.

Anusha, S., Vanajakshi, L. D., & Sharma, A., 2013, "A Simple Method for Estimation of Queue Length". International Journal of Innovation Research in Science, Engineering and Technology, 2 (1), 1-5.

Badillo, B. E., Rakha, H., Rioux, T. W., & Abrams, M., 2012, "Queue length estimation using conventional vehicle detector and probe vehicle data". Paper presented at the Intelligent Transportation Systems (ITSC), 2012 International IEEE Conference, 1674-1681.

Bahuleyan, H., & Kattan, L, 2015, "Real-time Estimation of Ramp Queue Length Based on a Hybrid Kalman Filter and Machine Learning Approach". Paper presented at the Transportation Research Board 94th Annual Meeting.

Ban, X., Hao, P., & Sun, Z., 2011, "Real time queue length estimation for signalized intersections using travel times from mobile sensors". Transportation Research Part C: Emerging Technologies, 19(6), 1133-1156.

Burkhard, H.-D., 2001, "Similarity and distance in case based reasoning". Fundamenta Informaticae, 47(3-4), 201-215.

Cao, J., Hadiuzzaman, M., Qiu, T. Z., & Hu, D., 2015, "Real-time queue estimation model development for uninterrupted freeway flow based on shockwave analysis". Canadian Journal of Civil Engineering, 42(3), 153-163.

Cassidy, M. J., 2001, "Recent findings on simple attributes of freeway queue formation and propagation". In Intelligent Transportation Systems, Proceedings 2001 IEEE, 531-535.

Cassidy, M. J., Anani, S. B., & Haigwood, J. M., 2002, "Study of freeway traffic near an off-

ramp". Transportation Research Part A: Policy and Practice, 36(6), 563-572.

Cetin, M., & Comert, G., 2007, "Estimating Queues at Signalized Intersections: Value of Location and Time Data from Instrumented Vehicles". Paper presented at the Intelligent Vehicles Symposium, 2007 IEEE.

Chang, G.-L., & Su, C, 1995, "Predicting intersection queue with neural network models". Transportation Research Part C: Emerging Technologies, 3(3), 175-191.

Cheek, M. T., 2007, "Improvements to a queue and delay estimation algorithm utilized in video imaging vehicle detection systems". Doctoral dissertation, Texas A&M University.

Cheng, Y., Qin, X., Jin, J., Ran, B., & Anderson, J., 2011, "Cycle-by-cycle queue length estimation for signalized intersections using sampled trajectory data". Transportation Research Record: Journal of the Transportation Research Board, (2257), 87-94.

Comert, G., 2013, "Effect of stop line detection in queue length estimation at traffic signals from probe vehicles data". European Journal of Operational Research, 226(1), 67-76.

Comert, G., 2013, "Simple analytical models for estimating the queue lengths from probe vehicles at traffic signals". Transportation Research Part B: Methodological, 55, 59-74.

Comert, G., & Cetin, M., 2009, "Queue length estimation from probe vehicle location and the impacts of sample size". European Journal of Operational Research, 197(1), 196-202.

Comert, G., & Cetin, M., 2011, "Analytical Evaluation of the Error in Queue Length Estimation at Traffic Signals from Probe Vehicle Data". Intelligent Transportation Systems, IEEE Transactions on, 12(2), 563-573.

Daganzo, C. F., 1997, "A continuum theory of traffic dynamics for freeways with special lanes". Transportation Research Part B: Methodological, 31(2), 83-102.

Dahlgren, J., Khattak, A., McDonough, P., Banerjee, I., Orrick, P., & Sharafsaleh, A., 2004,

"ITS Decision Enhancements: Developing Case-Based Reasoning and Expert Systems and Incorporating New Material". California Partners for Advanced Transit and Highways (PATH).

Dougherty, M. S., Kirby, H. R., & Boyle, R. D., 1993, "The use of neural networks to recognize and predict traffic congestion". Traffic engineering & control, 34(6).

Erera, A., Lawson, T., & Daganzo, C., 1998, "Simple, generalized method for analysis of traffic queue upstream of a bottleneck". Transportation Research Record: Journal of the Transportation Research Board (1646), 132-140.

Erera, A., Lawson, T., & Daganzo, C., 1998, "Simple, Generalized Method for Analysis of Traffic Queue Upstream of a Bottleneck". Transportation Research Record: Journal of the Transportation Research Board, 1646, 132-140.

Friedrich, B., Matschke, I., Almasri, E., & Mück, J., 2003, "Data fusion techniques for adaptive traffic signal control". Paper presented at the Proceedings of the 10th IFAC Symposium on Control in Transportation Systems.

Geroliminis, N., & Skabardonis, A., 2005, "Prediction of arrival profiles and queue lengths along signalized arterials by using a Markov decision process". Transportation Research Record: Journal of the Transportation Research Board, (1934), 116-124.

Geroliminis, N., & Skabardonis, A., 2005, "Prediction of arrival profiles and queue lengths along signalized arterials by using a Markov decision process". Transportation Research Record: Journal of the Transportation Research Board, (1934), 116-124.

Geroliminis, N., & Skabardonis, A., 2005, "Prediction of arrival profiles and queue lengths along signalized arterials by using a Markov decision process". Transportation Research Record: Journal of the Transportation Research Board, (1934), 116-124.

Geroliminis, N., & Skabardonis, A., 2005, "Prediction of arrival profiles and queue lengths along signalized arterials by using a Markov decision process". Transportation Research Record: Journal of the Transportation Research Board, (1934), 116-124.

Geroliminis, N., & Skabardonis, A., 2011, "Identification and Analysis of Queue Spillovers in City Street Networks". Intelligent Transportation Systems, IEEE Transactions on, 12(4), 1107-1115.

Gilmore, J. F., & Abe, N., 1995, "Neural Network Models For Traffic Control and Congestion Prediction". Journal of Intelligent Transportation Systems, 2(3), 231-252.

Herrera, J. C., Work, D.B., Ryan H., Ban, X., Jacobson, Q., & Bayen, A.M., 2010, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment", Journal of Transportation Research Part C: Emerging Technologies, 18(4), 568-583.

Hossain, S., Kattan, L., & Radmanesh, A., 2011, "Responsive Signal Control for Non-recurrent Traffic Congestion on an Arterials". Paper presented at the Transportation Research Board 90th Annual Meeting.

Hossain, S., 2010, "Development of Responsive Signal Control for Recurrent and Non-recurrent Traffic Congestion on an Arterials". MSc. Thesis, University of Calgary.

Kalman, R. E., 1960, "Contributions to the theory of optimal control". Bol. Soc. Mat. Mexicana, 5(2), 102-119.

Kattan, L., & Saidi, S., 2013, "Comparative analysis of probe-based ramp metering with detector-based and pre-timed ramp metering". Journal of Advanced Transportation, 47(1), 61-78.

Kaysi, I, 1993, "Integrated approach to vehicle routing and congestion prediction for real-time

Geroliminis, N., & Skabardonis, A., 2005, "Prediction of arrival profiles and queue lengths along signalized arterials by using a Markov decision process". Transportation Research Record: Journal of the Transportation Research Board, (1934), 116-124.

Geroliminis, N., & Skabardonis, A., 2011, "Identification and Analysis of Queue Spillovers in City Street Networks". Intelligent Transportation Systems, IEEE Transactions on, 12(4), 1107-1115.

Gilmore, J. F., & Abe, N., 1995, "Neural Network Models For Traffic Control and Congestion Prediction". Journal of Intelligent Transportation Systems, 2(3), 231-252.

Herrera, J. C., Work, D.B., Ryan H., Ban, X., Jacobson, Q., & Bayen, A.M., 2010, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment", Journal of Transportation Research Part C: Emerging Technologies, 18(4), 568-583.

Hossain, S., Kattan, L., & Radmanesh, A., 2011, "Responsive Signal Control for Non-recurrent Traffic Congestion on an Arterials". Paper presented at the Transportation Research Board 90th Annual Meeting.

Hossain, S., 2010, "Development of Responsive Signal Control for Recurrent and Non-recurrent Traffic Congestion on an Arterials". MSc. Thesis, University of Calgary.

Kalman, R. E., 1960, "Contributions to the theory of optimal control". Bol. Soc. Mat. Mexicana, 5(2), 102-119.

Kattan, L., & Saidi, S., 2013, "Comparative analysis of probe-based ramp metering with detector-based and pre-timed ramp metering". Journal of Advanced Transportation, 47(1), 61-78.

Kaysi, I, 1993, "Integrated approach to vehicle routing and congestion prediction for real-time

driver guidance", Transportation Research Record: Journal of the Transportation Research Board, (1408), 66-74.

Khattak, A., Dahlgren, J., & McDonough, P., 2006, "Tools for Supporting Implementation Decisions of Intelligent Transportation System". Transportation Research Record: Journal of the Transportation Research Board, 1944, 41-50.

Khattak, A., & Kanafani, A., 1996, "Case-based reasoning: a planning tool for intelligent transportation systems". Transportation Research Part C: Emerging Technologies, 4(5), 267-288.

Khattak, A., & Renski, H., 1999, "PLAN-HOV: Case-Based Reasoning Planning Tool for High-Occupancy Vehicle Lane Analysis in a Geographic Information System Environment". Transportation Research Record: Journal of the Transportation Research Board, (1682), 18-27.

Laird, J., Geers, D. G., Yang, W., & Chun Tung, C., 2013, "Integrating video cameras for ALINEA on-ramp queue length estimation". Paper presented at the Intelligent Transportation Systems (ITSC), 2013 International IEEE Conference, 1571-1578.

Lawson, T., Lovell, D., & Daganzo, C., 1997, "Using input-output diagram to determine spatial and temporal extents of a queue upstream of a bottleneck". Transportation Research Record: Journal of the Transportation Research Board, (1572), 140-147.

Ledoux, C., 1997, "An urban traffic flow model integrating neural networks". Transportation Research Part C: Emerging Technologies, 5(5), 287-300.

Liu, H. X., Wu, X., Ma, W., & Hu, H., 2009, "Real-time queue length estimation for congested signalized intersections". Transportation Research Part C: Emerging Technologies, 17(4), 412-427.

Ma, D., Wang, D., Bie, Y., Sun, F., & Jin, S., 2012, "A method for queue length estimation in an urban street network based on roll time occupancy data". Mathematical Problems in Engineering.

Markov, A., 1971, "Extension of the limit theorems of probability theory to a sum of variables connected in a chain". Dynamic Probabilistic Systems, 1: Markov Models, 552-577.

Meyn, S. P., & Tweedie, R. L., 2012, "Markov chains and stochastic stability: Springer Science & Business Media.

Michalopoulos, P. G., & Stephanopoulos, G., 1977, "Oversaturated signal systems with queue length constraints I". Transportation Research, 11(6), 413-421.

Michalopoulos, P. G., & Stephanopoulos, G., 1977, "Oversaturated signal systems with queue length constraints II". Transportation Research, 11(6), 423-428.

MÜCK, J., 2002, "Using detectors near the stop-line to estimate traffic flows". Traffic engineering & control, 43(11), 429-434.

Mucsi, K., Khan, A. M., & Ahmadi, M., 2011, "An Adaptive Neuro-Fuzzy Inference System for estimating the number of vehicles for queue management at signalized intersections". Transportation Research Part C: Emerging Technologies, 19(6), 1033-1047.

Munoz, J. C., & Daganzo, C. F., 2002, "The bottleneck mechanism of a freeway diverge". Transportation Research Part A: Policy and Practice, 36(6), 483-505.

Mystkowski, C., & Khan, S., 1999, "Estimating Queue Lengths by Using SIGNAL94, SYNCHRO3, TRANSYT-7F, PASSER II-90, and CORSIM". Transportation Research Record: Journal of the Transportation Research Board, (1683), 110-117.

Newell, G. F., 1965, "Approximation Methods for Queues with Application to the Fixed-Cycle traffic Light". SIAM Review, 7(2), 223-240.

Newell, G. F., 1999, "Delays caused by a queue at a freeway exit ramp". Transportation Research Part B: Methodological, 33(5), 337-350.

Papageorgiou, M., & Vigos, G., 2008, "Relating time-occupancy measurements to space-occupancy and link vehicle-count". Transportation Research Part C: Emerging Technologies, 16(1), 1-17.

Pecherkova, P., Fli, x, dr, M., & Dunik, J., 2008, "Application of estimation techniques on queue lengths estimation in traffic network". Paper presented at the Cybernetic Intelligent Systems, 2008 IEEE International, 1-6.

Qi, H., Wang, D., Chen, P., & Bie, Y., 2013, "A link traffic model by incorporating both merits of vertical and Horizontal Queue Concept". KSCE Journal of Civil Engineering, 17(5), 1117-1129.

Qian, G., Lee, J., & Chung, E., 2012, "Algorithm for queue estimation with loop detector of time occupancy in off-ramps on signalized motorways". Transportation Research Record: Journal of the Transportation Research Board, (2278), 50-56.

Qiu, Z., & Cheng, P., 2007, "State of the art and practice: cellular probe technology applied in advanced traveler information system". Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Ramezani, M., & Geroliminis, N., 2015, "Queue Profile Estimation in Congested Urban Networks with Probe Data". Computer-Aided Civil and Infrastructure Engineering, 30(6), 414-432.

Ross, J. V., Taimre, T., & Pollett, P. K., 2007, "Estimation for queues from queue length data". Queueing Systems, 55(2), 131-138.

Saraf, R. K., 1994, "Adaptive traffic control using neural networks", Transportation Research

Part A: Policy and Practice, 30(1), 57-58.

Schank, R. C., 1983, "Dynamic memory: A theory of reminding and learning in computers and people". Cambridge University Press.

Schutter, B. D., Hoogendoorn, S. P., Schuurman, H., & Stramigioli, S., 2003, "A multi-agent case-based traffic control scenario evaluation system". Paper presented at the Intelligent Transportation Systems, Proceedings. 2003 IEEE.

Sharma, A., Bullock, D., & Bonneson, J., 2007, "Input-output and hybrid techniques for real-time prediction of delay and maximum queue length at signalized intersections". Transportation Research Record: Journal of the Transportation Research Board, (2035), 69-80.

Skalak, D. B., & Rissland, E. L., 1992, "Arguments and cases: An inevitable intertwining". Artificial intelligence and Law, 1(1), 3-44.

Stephanopoulos, G., & Michalopoulos, P. G., 1978, "Real time control of critical intersections based on queue dynamics". Paper presented at the Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference, 1018-1024.

Venkatanarayana, R., Park, H., Smith, B. L., Skerrit Jr, C., & Ruhter, N. W., 2011, "Application of IntelliDrive[SM] to Address Oversaturated Conditions on Arterials". Paper presented at the Transportation Research Board 90th Annual Meeting.

Vigos, G., & Papageorgiou, M., 2010, "A Simplified Estimation Scheme for the Number of Vehicles in Signalized Links". Intelligent Transportation Systems, IEEE Transactions on, 11(2), 312-321.

Vigos, G., Papageorgiou, M., & Wang, Y., 2008, "Real-time estimation of vehicle-count within signalized links". Transportation Research Part C: Emerging Technologies, 16(1), 18-35.

Vigos, G., Papageorgiou, M., & Yibing, W., 2006, "A Ramp Queue Length Estimation Algorithm". In Intelligent Transportation Systems Conference, ITSC06, 418-425.

Viloria, F., Courage, K., & Avery, D., 2000, "Comparison of queue-length models at signalized intersections". Transportation Research Record: Journal of the Transportation Research Board, (1710), 222-230.

Viti, F., & van Zuylen, H., 2004, "Modeling queues at signalized intersections". Transportation Research Record: Journal of the Transportation Research Board, (1883), 68-77.

Viti, F., & Van Zuylen, H. J., 2009, "The Dynamics and the Uncertainty of Queues at Fixed and Actuated Controls: A Probabilistic Approach". Journal of Intelligent Transportation Systems, 13(1), 39-51.

Vlahogianni, E. I., Webber Jr, C. L., Geroliminis, N., & Skabardonis, A., 2007, "Statistical characteristics of transitional queue conditions in signalized arterials". Transportation Research Part C: Emerging Technologies, 15(6), 392-403.

Wirasinghe, S., 1978, "Determination of Traffic Delays from Shock-wave Analysis" . Transportation Research, 12, 343-348.

Worthington, D., 2009, "Reflections on Queue Modelling from the Last 50 Years". The Journal of the Operational Research Society, 60, 83-92.

Wu, J., Jin, X., & Horowitz, A., 2008, "Methodologies for estimating vehicle queue length at metered on-ramps". Transportation Research Record: Journal of the Transportation Research Board, (2047), 75-82.

Yang, L., & Xianfeng, Y., 2014, "Estimating Dynamic Queue Distribution in a Signalized Network Through a Probability Generating Model". Intelligent Transportation Systems, IEEE Transactions on, 15(1), 334-344.

Yi, P., Xin, C., & Zhao, Q., 2001, "Implementation and Field Testing of Characteristics-Based Intersection Queue Estimation Model". Networks and Spatial Economics, 1(1-2), 205-222.

Zhang, H., Ritchie, S., & Lo, Z.-P., 1997, "Macroscopic modeling of freeway traffic using an artificial neural network". Transportation Research Record: Journal of the Transportation Research Board, (1588), 110-119.

APPENDIX

Paramics API code For Combined CBR and KF Queue Length Estimation Model

```cpp
#include <iostream>

#include <stdlib.h>

#include <stdio.h>

#include <string.h>

#include <math.h>

#include <map>

#include <random>

#include <fstream>

#include <cstdlib>

#include <vector>

#include <sstream>

#include <vector>

#include "programmer.h"

using namespace std;
// Variables
double xhat = 1;

double p = 1;

double mean = 0;

double sigma = 1;

default_random_engine rand_gen;

normal_distribution<double> distribution(mean, sigma);
```

```c
int prv_aggr_count1 = 0;

int prv_aggr_count2 = 0;

int prv_aggr_count3 = 0;

int    Time=0;

void  qpx_NET_close(void){

        fclose(flog);

        qps_GUI_printf("END!!!!!!!!!!!\n");

}

void qpx_NET_postOpen(void)

{

        Time = 0;

}

// This is the time during which the detector data is collected. (in Seconds).

        float detector_data_collection_time = 60.0;

        int detector_data_interval = 1; // The time interval to collect the detector data (in

        Seconds)

        void qpx_NET_second(void)

{

// Declaring Variables for Detectors

        DETECTOR* detector1;

        DETECTOR* detector2;

        DETECTOR* detector3;

        LOOP* loop1_lane1;
```

```
        LOOP* loop2_lane1;

        LOOP* loop3_lane1;

//  Finding and assigning the Detectors

        detector1 = qpg_NET_detector("Detector UP");

        loop1_lane1= qpg_DTC_singleLoop(detector1);

        detector2 = qpg_NET_detector("Detector MID");

        loop2_lane1= qpg_DTC_singleLoop(detector2);

        detector3 = qpg_NET_detector("Detector DOWN");

        loop3_lane1= qpg_DTC_singleLoop(detector3);


//  Detector_data_collection_time Loop


        if(Time <= detector_data_collection_time)

        {

                Time++;

        }

        else

        {

                float occ1 = 0;

                float occ2 = 0;

                float occ3 = 0;

                float flow1 = 0;

                float flow2 = 0;
```

```
float flow3 = 0;

int count1 = 0;

int count2 = 0;

int count3 = 0;

float occ11 = 0;

float occ22 = 0;

float occ33 = 0;


//        Finding the occupancy time

        occ1 = qpg_DTL_occupancy(loop1_lane1, APILOOP_COMPLETE);

        occ2 = qpg_DTL_occupancy(loop2_lane1, APILOOP_COMPLETE);

        occ3 = qpg_DTL_occupancy(loop3_lane1, APILOOP_COMPLETE);

//        Calculating occupancy percentage

        occ11 = occ1/detector_data_collection_time;

        occ22 = occ2/detector_data_collection_time;

        occ33 = occ3/detector_data_collection_time;


//      Finding the count data for detectors

        int cur_aggr_count1 = qpg_DTL_count(loop1_lane1,

APILOOP_COMPLETE);

        count1 = cur_aggr_count1 - prv_aggr_count1;

        prv_aggr_count1 = cur_aggr_count1;
```

```cpp
                int cur_aggr_count3 = qpg_DTL_count(loop3_lane1,

APILOOP_COMPLETE);

                count3 = cur_aggr_count3 - prv_aggr_count3;

                prv_aggr_count3 = cur_aggr_count3;


//CBR for queue measurement(Z)

        void reading_case();  //reading case file

        ifstream file;

        file.open("C:\\Users\\...\\ case-library.csv");

// reading the file as string

        vector<int> case_no;

// Same goes for the following variables

        vector<float> occ1_lcase;            //occupancy time in library case for loop detector1

        vector<float> occ2_lcase;            //occupancy time in library case for loop detector2

        vector<float> occ3_lcase;            //occupancy time in library case for loop detector2

        vector<float> Q_lcase;

        string line;

        string tmp;

        int number_of_lines = 0;

        int count = 0;

        while(getline(file, line))//(!file.eof())

        {
```

```cpp
        stringstream linestream(line);

        getline(linestream, tmp, ',');
//      qps_GUI_printf("count: %d", stoi(tmp));
        case_no.push_back(stoi(tmp));

        getline(linestream, tmp, ',');
//      qps_GUI_printf("occ1: %f", atof(tmp.c_str()));
        occ1_lcase.push_back(atof(tmp.c_str()));

        getline(linestream, tmp, ',');
//      qps_GUI_printf("occ2: %f", atof(tmp.c_str()));
        occ2_lcase.push_back(atof(tmp.c_str()));

        getline(linestream, tmp, ',');
//      qps_GUI_printf("occ3:  %f", atof(tmp.c_str()));
        occ3_lcase.push_back(atof(tmp.c_str()));

        getline(linestream, tmp, ',');
//      qps_GUI_printf("Q: %d", stoi(tmp));
        Q_lcase.push_back(atof(tmp.c_str()));

        count++;
```

```cpp
            }

        number_of_lines = count;

        file.close();


// calculating max similarity score

        count = 0;

        int maxCount = 0;

        float * s_score = new float [number_of_lines+1];

        while(count < number_of_lines)//(!file.eof())

        {

                //printf("occ2_lcase:%f", occ2_lcase[count]);

                s_score[count]=(1.0/3)*((1.0/(1+ abs (occ11-occ1_lcase[count])))+(1.0/(1+ abs

(occ22-occ2_lcase[count])))+(1.0/(1+ abs (occ33-occ3_lcase[count]))));

                if(s_score[count] >= s_score[maxCount])

                {

                        maxCount = count;

                }

                //printf("Z:%d", Z);

                count++;

        }

        float Maxs_score = s_score[maxCount];

        int mostsimilarpattern = case_no[maxCount];

        float Z = Q_lcase[maxCount];
```

```cpp
        float occ11_max = occ1_lcase[maxCount];

        float occ22_max = occ2_lcase[maxCount];

        float occ33_max = occ3_lcase[maxCount];

        delete  [] s_score;


//Kalman filter for updating the queue length

        double Q=1;

        double R=2;

        double xhat_minus; //a priori estimate of queue length

        //double xhat; //a posteriori estimate of queue length

        double p_minus; //a priori estimate of error covariance

        //double p; //a posteriori estimate of error covariane

        double k; // Kalman gain

        double w=1;

        //initial guess

        w = distribution(rand_gen);
    //qps_GUI_printf("w: %f", w);


        xhat_minus = xhat + (count1-count3) + w;

        p_minus = p + Q;

        fprintf(flog, " %f,", xhat_minus);
```

```
        //measurement update or correction

        k = p_minus/(p_minus+R);

        xhat = xhat_minus+k*(Z-xhat_minus);

        if (xhat<1){

                xhat = 0;

        }

        p = (1-k)*p_minus;


// update the case library with the new found data

        //ofstream ofile;

        //ofile.open("C:\\Users\\...\\case-library.csv", ios::app);

        //ofile << (number_of_lines+1) << ",";

        //ofile << occ11_max << ",";

        //ofile << occ22_max << ",";

        //ofile << occ33_max << ",";

        //ofile << xhat << "\n";

        //ofile.close();


// Initializing the Time for next step

    Time = 0;

  }

}
```