Task 5 Deliverable – Final Report
Contract BDV24-562-05

Study of Intersection Turning Movement Estimation

Mohamed A. Abdel-Aty, Ph.D., P.E.

Qing Cai, Ph.D.

Nada Mahmoud

Jinghui Yuan, Ph.D.

University of Central Florida

Department of Civil, Environmental & Construction Engineering

Orlando, FL 32816-2450

UNIVERSITY OF
CENTRAL FLORIDA

July 2020

Disclaimer

"The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation."

UNITS CONVERSION

APPROXIMATE CONVERSIONS TO SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
LENGTH				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
AREA				
in^2	squareinches	645.2	square millimeters	mm^2
ft^2	squarefeet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
VOLUME				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: volumes greater than 1000 L shall be shown in m^3				
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
TEMPERATURE (exact degrees)				
ºF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	ºC
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
ILLUMINATION				
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
FORCE and PRESSURE or STRESS				
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
LENGTH				
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
AREA				
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
VOLUME				
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
MASS				
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
TEMPERATURE (exact degrees)				
ºC	Celsius	1.8C+32	Fahrenheit	ºF
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
ILLUMINATION				
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
FORCE and PRESSURE or STRESS				
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

1. Report No. **16207134**	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle **Study of Intersection Turning Movement Estimation**		5. Report Date **July 31, 2020** 6. Performing Organization Code
7. Author(s) **Mohamed A. Abdel-Aty, Ph.D, PE; Qing Cai, Ph.D; Nada Mahmoud; Jinghui Yuan, Ph.D.**		
9. Performing Organization Name and Address **Department of Civil, Environmental & Construction Engineering** **University of Central Florida** **12800 Pegasus Drive, Suite 211** **Orlando, FL 32816-2450**		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address **Florida Department of Transportation**		13. Type of Report and Period Covered **Final Deliverable (Draft)** 14. Sponsoring Agency Code
15. Supplementary Note		

16. Abstract

This project aims at estimating real-time turning movement counts at signalized intersections using traffic data from adjacent intersections (upstream and downstream intersections) to emulate traffic detector systems. In order to achieve the objective, first, an extensive literature review was conducted to summarize different methodological approaches as well as data sources to predict/estimate traffic parameters. The reviewed research also determined the commonly utilized performance measures to evaluate the developed prediction/estimation models. Afterwards, the GRIDSMART system data was utilized to develop different turning movement estimation algorithms. Moreover, sequence of intersections was determined considering several factors: correlation, coefficient of variation, distance between intersections, number of access points/stop-controlled intersection between two consecutive intersections, and data availability. As a result, real time data were collected from 19 intersections along two main corridors in Orange county, Florida. They were divided into four groups on the US 17/92 corridor and seven groups on the US 441 corridor. The collected data was divided into two parts: one is to develop algorithms and the other is to validate the developed algorithms. Each group consists of three consecutive intersections. The data was processed and aggregated at the cycle level and utilized in developing, tuning, and validating six models: negative binomial model, finite mixture model, multivariate adaptive regression spline, neural networks, random forest, and gradient boosting. The models were trained to estimate one traffic movement at a time. Then, an extensive comparison was carried out to determine the best methodological approach to estimate turning movement estimation algorithms for individual groups, each corridor, and for the whole study area. The results showed that Gradient Boosting (XGBoost) was the best model, which could provide the state-of-the-art accuracy to estimate cycle-level through and left-turn movements counts at signalized intersections. Furthermore, the developed generic estimation model was tested for transferability in the abnormal traffic conditions. The model was able to capture the traffic change due to an accident instantaneously. A list of intersections were suggested to apply the developed algorithms. Hence, the proposed method could increase the movement data coverage as well as reduce the sensor cost. Finally, Miovision TrafficLink data were analyzed and validated by comparing it to ATSPM data. The results illustrated that Miovision detected at least 9.06% less through volume than the ATSPM data. However, for left-turn volume, Miovision could help identify more vehicles in most cases.

17. Key Word	18. Distribution Statement		
Intersection movement estimation, Gridsmart, Miovision,			
19. Security Classif. (of this report)	20. Security Classif. (of this page) **Unclassified**	21. No. of Pages 129	22. Price

EXECUTIVE SUMMARY

Accurate turning movement counts at intersections are important for transportation engineers or planners to conduct different types of traffic and transportation studies, including transportation planning, geometric design, level of service analysis, capacity analysis, signal timing calculations, and traffic signal coordination. The traditional method to obtain the vehicle volumes at intersections is based on manual counts by observers, which is useful but requires extensive labor and can involve human errors from the observers. Recently, advanced traffic sensors such as camera detectors are available to continuously provide detailed and accurate vehicle turning movements. However, it is very expensive to fully instrument every intersection with the sensors. According to a previous study (Zhu et al., 2014), the traffic patterns of the adjacent intersections should be highly correlated. Hence, it should be feasible and cost-effective to install traffic sensors at several intersections to detect the turning movements and develop algorithms to estimate the turning movements at the adjacent intersections.

This research aimed to develop algorithms to estimate turning movements at signalized intersections using traffic data from adjacent intersections. Aiming to achieve this objective, previous studies were reviewed to conclude the most efficient methodological approaches to predict/estimate traffic parameters. It also illustrates the utilized data in literature to develop parametric and non-parametric (machine learning models) as well as the performance measures that were commonly used to evaluate the developed models. The traffic movements and signal data from the GRIDSMART system were utilized to develop different turning movement estimation algorithms (generic, corridor, and individual groups models). Moreover, sequence of intersections was determined considering several factors: correlation, coefficient of variation, distance between intersections, number of access points/stop-controlled intersection between two

consecutive intersections, and data availability. As a result, a total of 19 intersections were divided into four groups on US 17/92 and seven groups on US 441. The collected data was divided into two parts: one is to develop algorithms and the other is to validate the developed algorithms. Different approaches that were widely employed in previous short-term traffic forecasting studies (parametric models, machine learning models, and deep learning models) were followed to develop a comprehensive turning movement estimation model. Six models were applied, tuned, and tested: negative binomial model, finite mixture model, multivariate adaptive regression spline, neural networks, random forest, and gradient boosting. The models were developed to estimate corridor-level (North-South direction) through movements. They were trained to estimate one traffic movement at a time. Three performance measures were utilized to evaluate the developed models: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). An extensive comparison was carried out to determine the best methodological approach to estimate turning movement estimation algorithms for individual groups, each corridor, and for the whole study area. The Gradient Boosting Decision Tree (GBDT) model was chosen to estimate turning movements as it outperformed other parametric and nonparametric algorithms. Thus, totally 56 models were trained and tested using the GBDT model. The results show that implementing a specific estimation algorithm for a certain group improves the performance measures. Moreover, for the developed generic models using all day data, MAPE values of AM and PM peak periods were higher than off-peak and night-time periods for both through and left-turn movement models. Hence, specific models were developed using refined subsets of data for peak periods only. The groups average MAPE was 7% for through movement models and 2% for left turn movement models. Besides, the performance measures for all the developed models outperformed the models in the most recent literature.

The study concludes that the developed Turning Movement Estimation (TMC) algorithms could emulate GRIDSMART at most intersections. To achieve the best performance, it was recommended to apply the developed algorithms at six intersections (two at US 17/92 and four at US 441) while the GRIDSMART system could be used to detect movements at the rest intersections and provide the required data for the algorithms. Moreover, it was recommended to use peak period models to estimate through and left-turn movement at the AM and PM peaks and the generic model could be used for other time periods. The right-turn movements were not estimated in this study since it is not controlled by the signal timing. Both models for individual intersection group and all groups were developed. The two types of models could provide the state-of-the-art accuracy compared to the previous studies. While the models for individual intersection group could provide more accurate estimation results, it is less convenient to program models for each intersection. Hence, the adoption of the models should depend on the practitioners' needs.

Afterwards, Miovision TrafficLink data was analyzed and validated by comparing it to the ATSPM data. One-minute aggregated turning movement data was utilized in validating Miovision system using ATSPM data. Hence, ATSPM data at upstream (advanced) and stop-line detectors was downloaded and aggregated in one-minute increments for the same time period. The ATSPM data at the advanced and stop line detectors were compared. It shows that the stop-line detectors could detect over 10% less vehicles for most cases, compared to the advanced detectors. Finally, the results illustrated that Miovision detected at least 9.06% less through volume than the ATSPM data. However, for left-turn volume, Miovision could help identify more vehicles in most cases.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

As the rapid advancement of technology is changing the way we travel, Intelligent Transportation Systems (ITS) technologies are becoming vital in enhancing traffic network management (Chen & Muller, 2001), which could provide various traffic information to users about real-time traffic conditions (e.g., traffic flow and travel time). The information provided to users through ITS technologies should represent future traffic conditions. Moreover, it should be updated in real-time to guarantee the reliability and efficiency of the traffic system. Short-term traffic estimation has been widely employed in ITS applications to feed the system with the required traffic information.

In general, accurate turning movement counts at intersections are important for transportation engineers or planners to conduct different types of traffic and transportation studies, including transportation planning, geometric design, level of service analysis, capacity analysis, signal timing calculations, and traffic signal coordination. Thus, the main objective of this project is to develop an intersection turning movement estimation algorithm(s) to help expand the turning movement data coverage. Traffic detectors could be deployed at some intersections and then several short-term traffic estimation approaches could be employed to calculate the required traffic information of the other intersections without detectors. This will help decrease the expensive cost of collecting traffic information at signalized intersections by using sensors. The basic idea is to estimate turning movements at one intersection using real-time data from adjacent intersections. Hence, the estimated turning movements could emulate real-time counts from traffic movement detectors. This study is the first attempt to estimate cycle-level aggregated through and left turn movements for major roads at signalized intersections.

In this project, two types of movement detection systems were investigated:

GRIDSMART in Orange county and MioVison in Seminole county. The GRIDSMART detectors were installed on two major corridors (US 441 and US 17/92), which provides detailed information of each vehicle at intersections and signal timing. The MioVison detectors were installed several intersections, providing turning vehicle counts by a minute. The Automated Traffic Signal Performance Measures (ATSPM) data were also available at the intersections where the Miovison detectors were installed. Given the available detectors, the main objectives of this project are summarized as follows:

1. Review the traffic flow/movement prediction/estimation methods
2. Develop movement estimation algorithms for GRIDSMART considering the sequence of intersections and estimation algorithms
3. Compare the Miovison data with ATSPM for different movements at intersections

This report is organized as follow: the second chapter illustrates the literature review. It includes different input data sources and time intervals for training and testing different forecasting approaches, the methodological framework for predicting/estimating short-term traffic parameters, and different accuracy evaluation measures. Chapter 3 focuses on the GRIDSMART system, which included GRIDSMART data collection, preparation, and validation. It comprised the methodologies and the developed turning movement estimation models. At the end of the chapter, the results were illustrated and summarized to conclude the best developed models based on different calculated performance measures. It also listed at which intersections the developed algorithm(s) could emulate the detectors systems based on the model performance. Chapter 4 summarized data from the Miovision system. The Miovision data were validated and compared based on the ATSPM data. The performance of movement data from Miovison was investigated extensively. The last chapter concludes summary of data collection, data validation and the estimation models that were developed for intersections with

the GRIDSMART and Miovision systems. Based on the models' performance, the corresponding application suggestions were provided.

CHAPTER 2. LITERATURE REVIEW

For predicting/estimating short-term traffic parameters (e.g., traffic flow, speed, and travel time), several methodologies were addressed in previous research. Research studies were mostly based on two modelling approaches: parametric and nonparametric models. In addition, some simulation models were also utilized to forecast short-term traffic parameters. The selection of the most appropriate methodology in estimating traffic parameters is important. Generally, both parametric and nonparametric approaches have been widely employed in previous short-term traffic forecasting studies. The parametric models could be divided into statistical parametric techniques (such as time-series models) and state space models (Qiao et al., 2013). Further, the nonparametric models comprise regression techniques and neural networks' approaches (Vlahogianni et al., 2004). Moreover, some simulation models have also been used in the short-term traffic estimation. A comprehensive literature review was conducted. According to the tasks in the project scope, the literature review covers the following concepts:

- Data sources for traffic volume estimation.
- Different methodological approaches that been adopted to predict/estimate traffic parameters.
- Different accuracy evaluation measures

2.1 Data Source for Traffic Volume Estimation

To estimate the short-term traffic, various input data are needed. Many input traffic data sources were utilized in previous traffic forecasting research, including loop detectors, Global Positioning System (GPS) trajectory data, etc. Among them, the most commonly used data source is loop detectors. Table 1 shows the summary of data sources utilized in previous research.

Table 1: Summary of data sources

Data Source	Dependent Variables	Independent Variables	Spatial Relationship	Temporal	Study	Facility Type
Loop Detectors	Traffic volume and travel speed	Traffic volume and travel speed	Adjacent roads upstream the target road	Data aggregated by 5-min / forecasting 5 minutes ahead	(Chang et al., 2000)	Urban Network
	Traffic volume, day of week	Traffic volume	No	Data aggregated by 10-min/ forecasting 24 hrs ahead and horizons smaller than 80 minutes	(Thomas et al., 2010)	Urban Network
	Traffic volume	Traffic volume	No	Data aggregated by 5-min / forecasting 5 minutes ahead	(Yoon and Chang, 2014)	Urban Network
	Link volume; time of day; day of week; weekend or not	Traffic volume	No	Data aggregated by 5-min / forecasting 5 minutes ahead	(Mackenzie et al., 2018)	Urban Network
	Traffic flow	Traffic volume	No	Data aggregated by 15-min	(Williams and Hoel, 2003)	Freeways
	Traffic flow	Traffic volume	No	Data aggregated by 15-min	(Ghosh et al., 2007, 2005)	Urban Network
	Traffic volume and occupancy (normal condition and abnormal events)	Traffic volume	No	Data aggregated by 15-min	(Guo et al., 2013)	Urban Network
	Traffic flow and travel time (normal condition and abnormal events)	Traffic volume	No	Data aggregated by 5-min and 15-min	(Guo et al., 2018)	Urban Network
	One-intersection model: (current volume); Three-intersection model: (current, upstream, and downstream volumes)	Traffic volume	Upstream and downstream intersection	Data aggregated by 15-min	(Zhu et al., 2014)	Urban Network

Data Source	Dependent Variables	Independent Variables	Spatial Relationship	Temporal	Study	Facility Type
	Traffic volume	Traffic volume	No	Data aggregated by 15-min	(Zhan et al., 2018)	Suburban Network
	45 minutes (30 steps) volume; normal days or holiday; time of day	Traffic volume	Multivariate (multiple locations)	Data aggregated by 90-sec	(Zheng and Su, 2014)	Urban Network
	Univariate (current volume data); Multivariate (current volume and upstream volume data)	Traffic volume	Current and upstream intersection	Data aggregated by 3-min/ forecasting 15 minutes ahead	(Vlahogianni et al., 2005)	Urban Network
	Multivariate (current volume and upstream volume data)	Traffic volume	Current and upstream loop detectors	Data aggregated by 3-min	(Stathopoulos and Karlaftis, 2003)	Urban Network
	Traffic Volume	Traffic Volume	Upstream detectors (weight matrix for first-order, second-order, and third-order)	Data aggregated by 90-sec	(Kamarianakis and Prastacos, 2005)	Urban Network
Observed Vehicle Counts	Traffic volume	Traffic volume and turning volume	Road network	Not defined	(Chen and Chen, 2011)	Urban Network
	Inbound and outbound traffic volume	Turning movements	No	Data aggregated by one hour/ forecasting one hour ahead	(Ghanim and Shaaban, 2018)	Suburban Network
Taxi Trajectory Data	Time Domain, upstream road segment, downstream road segment	Taxi Volume	Both upstream and downstream segments	Forecasting 5 minutes ahead	(Xia et al., 2016)	Urban Network
GPS Trajectory Data	Vehicle arrival information, signal status	Traffic volume	Coordinated intersections	Not defined	(Zheng and Liu, 2017)	Urban Network
Bluetooth Data	Bluetooth average travel time	Travel time	No	Data aggregated by 5-min	(Qiao et al., 2013)	Freeways

2.1.1 Loop Detectors Data Sources

From the previous literature along the past two decades, traffic data from loop detectors were the most commonly used in estimating traffic parameters. Data were usually aggregated at 5-min or 15-min intervals. For example, Chang et al. (2000) collected 1,339 time-series data points in the 5-minute interval and then fed into combined Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN) models to predict traffic flow and travel speed. Moreover, to estimate speed data up to 5-min in the future, Dia (2001) developed an object-oriented neural network model using data from four inductive loops on a section of the Pacific Highway in Queensland, Australia. In addition, 5-min interval vehicle detection data collected by inductive loop detectors at about 20 signalized intersections were utilized in the study by Thomas et al. (2010). The data measurements were provided in 5-min intervals and then aggregated in 10-min intervals to be used in their model. Also, Yoon & Chang (2014) utilized traffic volume data collected in every 5 minutes from dual loop detectors at a signalized intersection on the south Beltway, one of the main Signalized arterials in Seoul, South Korea. Furthermore, Mackenzie et al. (2018) used lane-specific traffic count data which were collected from dual-loop detectors every 5 minutes. These data were used to evaluate the performance of different short-term traffic flow prediction algorithms.

Also, many studies adopted 15-min intervals in the data collection for short-term traffic forecasting. Williams & Hoel (2003) utilized loop detectors' data aggregated at 15-min discrete time intervals in developing models and forecasting the short-term traffic flow. Ghosh et al. (2005) and Ghosh et al.(2007) obtained 15-min data from loop detectors at 183 junctions in the city center of Dublin to predict traffic flow using thee different time series models. Also, Guo et al. (2013) utilized 15-min traffic data collected by loop detectors from Central London. In their recent study, Guo et al. (2018) collected two kinds of traffic flow data from two major roads in central London

using loop detectors, which were aggregated in 5-min and 15-min intervals, respectively. They also used link-based 5-min interval travel time data to test their models. Zhu et al. (2014) estimated short-term traffic data at intersections in Baotou City of China based on 128 datasets, which were collected at 15-min interval from loop detectors. Moreover, Zhan et al. (2018) utilized 14 loop detectors to collect data from an arterial in Arcadia, Canada in 2015. The raw data were processed into time series flow data that separated by 15-min time interval. Those data were sent to the control center every hour to forecast the flow in the next upcoming hour.

Traffic data from loop detectors were also processed in other time intervals. Zheng & Su (2014) utilized data collected every 1.5 minutes from seven loop detectors on a signalized arterial near downtown Athens, Greece. While for the same signalized arterial, Vlahogianni et al. (2005) developed their models based on 3-minute interval univariate and multivariate data. However, they only focused on estimating traffic using a single optimized model, while ignoring the effect of time interval. Also, Stathopoulos & Karlaftis (2003) developed their model based on 3-min interval detectors data from spatial standpoint at the same location (Alexander Venue).

2.1.2 Traffic Flow Data Sources

Kamarianakis & Prastacos (2005) built their model using flow data from 25 loop detectors located at urban streets in the downtown of Athens, Greece. The data were 7.5-min average traffic flow data. On another hand, Clark (2003) aggregated traffic data in 10-min and used them to develop a nonparametric regression model. Moreover, Jiang et al. (2005) utilized traffic flow data from North Carolina Department of Transportation to validate their model. The utilized time series data were lane-specific average hourly traffic flow on Durham freeway in the state of North Carolina.

Two studies tried to use link volume data to estimate turning movements at intersections. Chen et al (2010) developed estimation models using link volume data collected from observed vehicle counts at an arterial street connected by a series of signalized intersections. In the arterial

network, link flows on all approaches of both major and minor roads are available. The traffic counts on links originating from and terminating to the external station are used as observation constraints for the developed model. Ghanim & Shaaban (2018) developed an ANN model to estimate the hourly turning movement by using link volume which was also collected hourly. The data were collected from 847 signalized intersections (including 691 four-leg intersections and 156 three-leg intersections) in Palm Beach County, Florida. The data collection was conducted during AM and PM periods between the years of 2010-2014. All three-leg intersections were treated as four-leg intersections, where the in-traffic and out-traffic turning movements of the missing fourth leg are set to zero.

2.1.3 Other Data Sources

Moreover, other approaches were adopted to collect data that used in developing different forecasting models. Xia et al. (2016) used Global Positioning System (GPS) data to obtain real time big taxi trajectory data. Further, Zheng & Liu (2017) also used GPS trajectory data from Connected Vehicles (CV) or navigation devices to estimate traffic volumes. Finally, Qiao et al., (2013) used continuous data that were collected by using Bluetooth devices to predict real time travel time for stochastic freeway applications. Those devices provided Bluetooth average travel times every 5 minutes.

The majority of the previous studies focused on predicting traffic flow at intersection approaches, while only few tried to estimate turning movements. To develop estimation models, the turning movements of all surrounding intersections were used as dependent variables and other data such as link volume were employed as independent variables in the previous studies. In this project, the main objective is to develop algorithms to estimate real-time turning movements at intersections where no turning movement data are available. The turning movement data of adjacent intersection(s) will also be used as input, together with other traffic information (e.g.,

traffic volume, signal timing, etc.). Noteworthy, this would be the first attempt to include turning movement data of adjacent intersections to estimate the turning movements.

2.2 Methodology

For short-term traffic parameters prediction/estimation, several methodologies were addressed in previous research. Generally, both parametric and nonparametric approaches have been widely employed in previous short-term traffic volume (5 minutes to 15 minutes) prediction studies (Qiao et al., 2013). The parametric models could be divided into statistical parametric techniques (such as time-series models) and state space models (Xie et al., 2007; Qiao et al., 2013). The nonparametric models comprise regression techniques and machine learning approaches (Vlahogianni et al., 2004). In addition, some simulation models have also been utilized.

2.2.1 Studies predicting/estimating traffic volume based on parametric models

Parametric models are divided into statistical parametric techniques (e.g., historical average model and time series models) and the state space models (e.g., Kalman Filtering (KF)). The time series models include the Autoregressive model (AR), Moving average model (MA), Autoregressive Moving Average model (ARMA), Autoregressive Integrated Moving Average (ARIMA) and its derivatives.

Stathopoulos & Karlaftis (2003) developed a multivariate time-series state space model to forecast the real-time traffic flow at downstream locations from upstream detector data. The model was developed using 3-min traffic count data from detectors located at five locations along a signalized arterial (Alexander Avenue) near downtown Athens, Greece. The data were divided into two subsets: 70% to calibrate the model, and 30% to test it. The multivariate approach was able to model variety of univariate models as well (e.g. ARIMA model). The results showed that using multivariate state space model was recommended in the urban roadway system. It also showed that ARIMA approach is very helpful to predict real-time traffic flow at entry points of

the study area.

Seasonal Autoregressive Integrated Moving Average (SARIMA) model was developed by Williams & Hoel (2003) to predict traffic flow. They estimated short-term traffic parameters at a certain location using previous information at the forecast location in the network. Two datasets from two different types of highways were utilized, one in United States and another in United Kingdom. The SARIMA model achieved a Mean Absolute Percentage Error (MAPE) of 8.9%. However, it was recommended to optimize the model's parameters to improve the MAPE.

Moreover, Guo et al. (2013) introduced Grey System Model (GM) to reduce the dependency on model training. They also used the technique of Singular Spectrum Analysis (SSA) in a novel two-stage prediction structure to enhance forecasting accuracy. In their study, the accuracies of the estimated traffic parameters for both SSA and non-SSA model structures were compared. It was concluded that SSA data smoothing techniques before applying either machine learning, or statistical prediction methods could enhance the accuracy of the traffic parameters. Moreover, GM and SARIMA prediction models were compared in terms of the accuracy of estimated traffic parameters. Traffic flow data from a corridor in Central London were used to train and calibrate those models. Traffic parameters were estimated for 15 minutes ahead using both GM and SARIMA methods, not only under normal traffic conditions but also in case of incidents. The results showed that the accuracy of GM method outperformed SARIMA model under both conditions.

In order to obtain short-term traffic flow, Ghosh et al. (2005) developed and compared three different time-series models: random walk model, Holt-Winters' exponential smoothing technique, and SARIMA model. They used the 15-min interval traffic data obtained from loop detectors at a four-legged junction at the city center of Dublin to train their models. Three models were developed for predicting short-term traffic flow and were evaluated by comparing the Root

Mean Square Error (RMSE) and the MAPE. The estimated traffic flow using Holt-Winters' exponential smoothing technique and SARIMA model matched significantly the observed traffic flow data during peak hours. However, in case of severe changes in the observed traffic data, the models could fail to match significantly the estimated data. Furthermore, using 15-min interval traffic flow data that were collected at the same junction (TCS-183) at the city center of Dublin, Ghosh et al. (2007) proposed a Bayesian SARIMA time-series model. They used the Markov Chain Monte Carlo (MCMC) technique to estimate the distribution of the parameters of the model. The model was developed and trained using the traffic flow data collected from inductive loop-detectors. They concluded that the Bayesian approach in SARIMA model was very beneficial in the case of modeling traffic behavior with extreme peaks and fluctuations.

Another application of time-series models was the Space–Time Autoregressive Integrated Moving Average (STARIMA) model developed by Kamarianakis & Prastacos (2005) to forecast traffic flow. A three-stage iterative space-time model was built based on the traffic flow data from roads at the city center of Athens, Greece, which were measured every 7.5 minutes by using 25 loop-detectors. Also, separated ARIMA models were fitted to the datasets, a comparison between the average Standard Errors (SE) of the ARIMA models and the RMSE of the SARIMA model was conducted. The results showed that the values were quite close.

Min et al. (2009) proposed a hybrid spatio-temporal method, Dynamic Space-Time Autoregressive Integrated Moving Average (Dynamic STARIMA) model, to predict short-term traffic flow and enhance traffic prediction accuracy and efficiency. This model combines both STARIMA model and Dynamic Turn Ratio Prediction (DTRP) model, which was trained using traffic flow data collected from 50 sensors on the 2nd Ring Road of Beijing, China. They proved that the performance of Dynamic STARIMA model was impressive in predicting short-term traffic flow with higher accuracy when compared to competitive methods. Moreover, the model

demonstrated that it could predict traffic flow very fast for the specified locations on the network. However, the proposed Dynamic STARIMA model only showed its efficiency and accuracy on urban road network. Thus, it couldn't be implemented into other transportation systems like irregular suburban road network. In the following year, they introduced a novel approach in the ARIMA family, which was the Generalized Space-Time Autoregressive Integrated Moving Average (GSTARIMA) model (Min et al., 2010). They compared the proposed GSTARIMA model and the traditional STARIMA model using the same traffic flow data that were collected from sensors in the 2nd Ring Road of Beijing, China. Their results showed that GSTARIMA model's accuracy was better than the traditional STARIMA model for this urban expressway. However, the model needs more time for calculations and the parameter estimation is more difficult than that in STARIMA model. Moreover, GSTARIMA model requires more historical data to predict traffic parameters.

A Spatial-Temporal Random Effect (STRE) model was presented by Wu et al. (2016) with less complexity due to mathematical dimension. In total, 105 detectors in the downtown area of the city of Bellevue, Washington were used to collect the data for model development. The data were sent to the traffic management center every minute and the data were aggregated into 5 minutes interval in order to reduce the random noise effect. The STRE model estimated the traffic volumes effectively with the MAPE of 16%. The model comparison results indicated that STRE model outperformed the ARMA model, the spatiotemporal ARMA model, and Artificial Neural Networks (ANN) model.

Further, Zhu at al. (2016) utilized speed information, while considering spatial and temporal dimensions of traffic, to predict short-term traffic counts. Linear Conditional Gaussian Bayesian Network (LCG-BN) model was used for their prediction. The performance of the developed model was tested and compared with different approaches (i.e., continuous Bayesian

Network (BN), ARIMA, and one hidden layer Neural Network (NN)) using a microscopic traffic simulation dataset generated by SUMO1 under different time intervals of 5, 10, and 15 minutes. The traffic network consisted of 8 external links and 11 internal links. It was concluded that the prediction accuracy increased significantly when utilizing speed data along with spatial data.

State-space models have also been utilized in estimating short-term traffic parameters. The most widely used state-space model is the Kalman Filter (KM) model. Chen & Chien (2001) carried out dynamic travel time prediction using the KM model. They used probe vehicles' real-time reports to develop their models. Their study was on a segment of I-80 corridor in New Jersey based on collected travel-time data from probe vehicles. Two prediction approaches were addressed in this study: path-based and link-based travel time prediction methods. The study proved that under normal flow conditions path-based method outperformed the link-based one. Moreover, they also proved that the prediction accuracy could be improved by increasing probe vehicles' penetration rate. On the other hand, Xie et al. (2007) investigated predicting short-term traffic volume using the application of Kalman Filter with discrete wavelet analysis (Wavelet Kalman Filter prediction model). Traffic flow data were collected from detectors deployed on three highways: Interstate 5 (I-5), Interstate 90 (I-90), and Interstate 405 (I-405) in Seattle. The original data were denoised by dividing into several approximate and detailed data using discrete wavelet decomposition analysis. Thus, the prediction accuracy can be improved when using the Kalman Filter model with the prepared denoised data. The results showed that the investigated Wavelet Kalman Filter models outperformed the direct Kalman Filter model in terms of MAPE and RMSE.

Guo et al. (2014) investigated Adaptive Kalman Filter (AKF) model that can update the process variances. They used 15-min aggregated traffic flow data that collected from four highway

[1] SUMO: is an open source microscopic simulator mainly developed by employees of the Institute of Transportation Systems at the German Aerospace Center.

systems: the motorway system in the United Kingdom, and the metropolitan freeway systems in Maryland, Minnesota, and Washington State of the United States. The AKF model was proved to be adaptable in the case of unsteady traffic flow. Moreover, it was concluded that the state-space models developed base on the stochastic structure could affect the performance of AKF model.

Another parametric approach was proposed by Zhan et al. (2018). They developed an ensemble Time Decay Error-Correction (TDEC) model based on three core ideas: learn from base models' mistakes that had taken place in the past; balance the diversity and accuracy of the model; remove the outliers in traffic forecasts. In order to test the developed model, data were collected from arterial sensors in Arcadia, CA in 2015 and processed into 15-mins interval traffic flow time series. The measured traffic flow data were sent to the traffic control center every hour and were used for the following hour. The developed ensemble model achieved an improvement of 16.3% in mean of absolute error and 17% in standard deviation, which outperformed two ensemble prediction schemes based on Ridge Regression and Lasso. Moreover, it produced traffic flow predictions that were more robust and accurate than other ensemble models' predictions.

For studies that attempted to forecast the turning movements at intersections, Chen et al. (2010) developed a model to estimate turning movements counts at intersections by using the nonlinear Path Flow Estimator (PFE). The estimations were based on the entry and exit traffic flow counts from arterial network. The intersections' turning movement estimations were tested through two case studies. The first case study was at a main arterial that contains eight signalized intersections. They were tested to estimate the turning movements from linear traffic network considering all travel delay patterns which were identical for all the intersections. The second case study was at the St. Helena network in Napa Valley in California that contained 113 links. The results showed that the error was acceptable. However, it was concluded that the accuracy of the estimated turning movements could increase when more traffic counts and turning movements'

data are available.

2.2.2 Studies predicting/estimating traffic volume based on non-parametric (machine learning) approaches

Machine learning nonparametric models were widely used in traffic parameters predictions and estimations (e.g., nonparametric regression, Neural Networks (NN) techniques, and the K-Nearest Neighbor (KNN) techniques). The first KNN model for traffic volume prediction was developed by Davis et al. (1991). They developed their model using freeway data and compared it to a simple univariate linear time-series method. The KNN method performed comparably to, but not better than, the linear time-series approach. Moreover, it was concluded that larger databases may improve the accuracy of the KNN method.

Clark (2003) adopted the KNN technique (multivariate extension of nonparametric regression technique) in traffic estimation. The model was trained using 10-min interval traffic data from the London orbital motorway M25. His model didn't rely on either the formulation of the parametric model or assumptions of traffic state, while it only required modest traffic data to implement. Moreover, the accuracy was higher than using a naïve method. It was also mentioned that using 15-min interval traffic data to train the model could result in lower RMSE and MAPE than that using 10-min interval traffic data.

In the case of estimating short-term travel time using Bluetooth traffic data, Qiao et al. (2013) developed four different prediction models (i.e., historical average, ARIMA, Kalman Filter, and KNN methods) . The study was carried out on a 1.18-mile freeway segment located on Virginia Route I-66 Eastbound. The data were collected using Bluetooth sensors that provided avergae travel time of 5-min interval. Morover, they also introduced a modified nonparametric model K-Nearest Neighbor–Trend Adjustment Model (KNN-T) which included the effects of traffic trends in the forecasting model. Their results proved that the KNN and KNN-T models outperformed the

parametric models (i.e., historical average, ARIMA, and Kalman Filter), where the MAPE values were reduced by 10% for the off-peak periods, and by 20% for the peak periods. Focusing on nonparametric techniques only, their results showed that the KNN-T model had 2.5% and 4.8% lower MAPEs than that in the KNN model for off-peak and peak periods, respectively. Furthermore, as the KNN-T model captured the effects of traffic trends (time-varying trend), it could provide more precise forecasted traffic parameters.

Another example of KNN techniques was adopted by Yoon and Chang (2014). They proposed a K-Nearest Neighbor Non-Parametric Regression (KNN-NPR) algorithm to predict short term-traffic volume at signalized intersections. The traffic volume data were collected every five minutes using dual-loop detectors on a signalized urban arterial located on the South Beltway, one of main arterials in Seoul. The model estimated the direction of the future traffic volume without a time-delayed response. It superiorly outperformed both Kalman Filter model and ARIMA model.

Furthermore, Zheng & Su (2014) proposed a new two-step algorithm based on K-Nearest Neighbor enhanced by Least Square Probabilistic Classification (KNN-LSPC). First, time constraints, and local minimal were utilized in choosing the neighbor; to reduce overlapping between candidates. Second, KNN-LSPC algorithm was developed for estimating traffic volume using data from seven loop detectors on a signalized arterial (Alexander Venue) near downtown Athens, Greece. They used traffic data collected by every 1.5 mins and divided them into two groups (weekends and weekdays). The research proved that parameters of KNN-LSPC can be analytically optimized, which justified the outperformance of the enhanced KNN-model. By comparing KNN-LSPC algorithm with the benchmark models (six KNN algorithms and two Kalman filter algorithms), the KNN-LSPC model outperformed them in most cases but its performance wasn’t surprisingly huge compared with Kalman Filter algorithms. However, it was

indicated the data-driven KNN-LSPC algorithm was free of assumptions on data distribution and it had higher flexibility and extendibility than the Kalman Filtered algorithms.

In order to promote the accuracy and efficiency of short-term traffic flow prediction, Xia et al. (2016) developed a Spatial-Temporal Weighted K-Nearest Neighbor (STW-KNN) model. They implemented it in a MapReduce framework on a Hadoop platform. STW-KNN was used to estimate the traffic volume of the following five minutes. The traffic flow correlation and weight were considered in the study. The model was tested by using real-time big taxi trajectory data (contain recorded GPS trajectories by 12,000 taxis in a period of 15 days in November 2012) in Sanlihe E. Rd. in the city of Beijing. The developed model was compared with Artificial Neural Networks (ANNs), conventional KNN, Naïve Bayes (NB), and Random Forest (RF) models. It was proved to have significantly higher efficiency and scalability than the other models. Moreover, the STW-KNN model outperformed the other models by reducing MAPE value more than 11.59%. Specifically, it improved the accuracy to 89.71% with MAPEs between 3.34% and 6% in space and time domain.

To achieve the aim of improving traffic forecasting accuracy, Guo et al. (2018) developed a fusion-based framework that based on combined multiple standalone predictors. The framework was evaluated using three different strategies: average, weighted, and KNN methods. Those methods were applied to three different machine learning methods: NN, Support Vector Regression (SVR), and Random Forest (RF). Traffic flow data for 5-min interval that collected from inductive loop detectors were used to develop their framework. Data were collected from Cromwell Road in the Royal Borough Kensington and Chelsea, and Marylebone Road in the center of London, United Kingdom. The fusion model's accuracy in the cases of disrupted traffic and incidents were evaluated. The results indicated that the average and weighted fusion methods were more accurate than the individual methods. Moreover, the KNN fusion based method superiorly

outperformed when traffic conditions were disruptred. However, calibrating multiple prediction models in parallel was required in developing the proposed fusion framework.

Another commonly used machine learning model in traffic parameters prediction is Gradient Boosting Decision Trees (GBDT). For instance, a GBDT method that combines the correlation between the search of sliding time windows and feature extension was developed to forecast traffic flow at intersections (Xia and Chen, 2017). Furthermore, the GBDT model was proposed to predict short-term traffic volume on freeways (Yang et al., 2017). They proved that GBDT outperformed the Support Vector Machine (SVM) and Back Propagation Neural Networks (BPNN) models. Different studies have utilized the GBDT model to predict different travel time horizons (5 min ahead, 10 min ahead, and 15 min ahead). The developed models outperformed both SVM and BPNN models. It also improved the prediction accuracy and model interpretability (Zhang and Haghani, 2015; Cheng et al., 2019; Li and Bai, 2017; Zhang et al., 2016).

An alternative approach to estimate short-term traffic is the deep learning nonparametric models (e.g., Long Short-Term Memory network). In recent research, deep learning models were found to be promising in many research fields due to their ability to capture nonlinearity in models. Many other studies adopted neural network models to forecast different traffic parameters. Clark et al., (1993) developed the first neural network model based on a basic three-layered back-propagation Multilayer Perceptron (MLP) model. The model consisted of one input layer, one hidden layer, and one output layer.

Vlahogianni et al., (2005) concentrated on developing a neural network model for forecasting short-term traffic volume at signalized intersections on congested urban arterials. The model was developed based on error back-propagation feed-forward MLP techniques and genetic algorithms by considering both spatial and temporal traffic flow characteristics. In order to test the model, univariate and multivariate data from signalized urban arterials were used to evaluate the

developed neural network. Data from Alexander Avenue (a major signalized arterial in Athens, Greece that connected between two congested arterials) were utilized to train and test the developed algorithm. The available data were collected by a 3-minute interval. However, the time period was ignored while developing the model, and the focus was on estimating traffic using a single optimized model. The results showed that the multivariate data were helpful in the case of multiple-step forecasting. Moreover, the results showed that neural networks model performed better than the univariate ARIMA and multivariate State-Space models in forecasting short-term traffic flow.

In order to enhance the performance of traffic forecasting, Chang et al. (2000) combined an ARIMA model and ANN model. Furthermore, they compared ARIMA and ANN models' performance in forecasting traffic flow. Their results showed that the combined ARIMA and ANN model produced more accurate estimated traffic flows than using a single model (ARIMA or ANN). However, they also concluded that the ANN model could not always outperform the ARIMA model in forecasting traffic flow.

A novel nonparametric model was developed to estimate traffic flow on freeways by Jiang et al. (2005). They presented a dynamic time-delay recurrent Wavelet Neural Network (WNN). Average hourly traffic flow data from Durham freeway in North Calorina were used to train and validate the model. The authors only had an hourly traffic flow data but they recommended to use one-minute data or data aggregated in 2, 5, or 10-min interval in order to carry out an effective short-term traffic flow prediction. The results showed that the accuracy of estimated traffic flow utilizing WNN was within 10% accuracy, which was considered exceptionally excellent as they considered the complication of the forecasting problem.

Zhu et al. (2014) developed a Radial Basis Function Neural Network (RBFNN), a three-layered feed-forward neural network, and compared it with the Box-Jenkins model (ARIMA

model). To develop the model, 128 datasets with 15-min time interval were utilized. The data were split into 125 datasets for developing the model and three datasets for testing its accuracy. The study was conducted in Baotou City of China to predict short-term traffic data at intersections using traffic data collected by loop detectors. The area consisted of three adjacent intersections along Shaoxian Rd. The results showed better accuracy for the RBFNN model. Moreover, they concluded that for adjacent intersections, the parameters of speed, density, and traffic volumes could affect each other.

The Back Propagation (BP) and Radial Basis Function Neural Networks (RBF-NN) were combined linearly by Zheng at al. (2006). They developed a Bayesian Combined Nerual Network (BCNN) model based on 15-min interval traffic flow rates that were collected from Singapore's Ayer Rajah Expressway. Two different datasets were used in the experimental study: one dataset to train the single neural network models and the other dataset to evaluate the performance of the models and to compare between them. The evaluation and comparison were carried out using MAPE, Variance of Absolute Percentage Error (VAPE), and Probabiloty of Percentage Error (PPE). The findings of this research showed that the BCNN model outperformed the single neural network models (BP and RBF) in predicting traffic flow.

Rong et al. (2015) compared three prediction models for predicting traffic parameters: ARIMA model, BPNN model, and nonparametric regression (KNN) model. Their study was executed by collecting traffic volume, speed, and occupancy data by every 5 minutes. These data were collected from two sites: Jianguomen and Jimen Bridges in Beijing, China. Absolute Error (AE), RMSE, error distribution, and model probability were calculated to evaluate the three models. The results showed that in case of gentle time series fluctuation, the three models can give the same accuracy. However, the KNN model seemed to be significantly more accurate when time series fluctuate dramatically.

Dia (2001) aimed to predict speed by five minutes in the future with a high degree of accuracy. Thus, he developed an object-oriented Time Lag Recurrent Network (TLRN) to forecast traffic speed parameter. Speed and traffic flow data collected by inductive loops on a section of the Pacific Highway in Queensland, Australia were utilized to train, validate, and test the developed model. He anticipated the Mean Squared Error (MSE) as measure of effectiveness to test and compare different neural network models: MLP, Hybrid, Recurrent and TLRN. The results showed that the dynamic neural networks outperformed MLP. Also, he concluded that those neural networks can predict speed for the upcoming 5 minutes with the accuracy between 90 and 94%. However, they could also be used to predict travel time up to the following 15 minutes with the accuracy of 93-95%, which could be considered as high degree of accuracy.

Recently, Long Short-Term Memory models were commonly used in predicting traffic parameters. For instance, Zhao et al., (2017) utilized LSTM for traffic volume forecast concluded that LSTM is robust in predicting traffic parameters (e.g., traffic volume, travel time, traffic speed and occupancy). They integrated a designed Origin-Destination Correlation (ODC) matrix with LSTM. Moreover, Liu et al. (2017) developed Singular Point Probability LSTM (SPP-LSTM) model that was integrated with ARIMA model to predict high accuracy and stability long-term traffic flow. Li and Ban (2019) also developed a Convolutional Neural Network LSTM (CNN-LSTM) model to predict real-time traffic volume at signalized intersections. Traffic volumes were considered as an input to the CNN by transferring them into a 2D image, then LSTM was trained using the output of the CNN. On another study, Dia (2001) aimed to predict speed by five minutes in the future with a high degree of accuracy. Thus, he developed an object-oriented Time Lag Recurrent Netwrok (TLRN) to forecast traffic speed parameter. Speed and traffic flow data collected by inductive loops on a section of the Pacific Highway in Queensland, Australia were utilized to train, validate, and test the developed model. He anticipated the MSE as measure of

effictivness to test and compare different neural network models: MLP, Hybrid, Recurrent and TLRN. The results showed that the dynamic neural networks outperformed MLP. Also, he concluded that those neural networks can predict speed for the upcoming 5 minutes with the accuracy between 90 and 94%. However, they could also be used to predict travel time up to the following 15 minutes with the accuracy of 93-95%, which could be considered as high degree of accuracy. Finally, Hierarchical Temporal Memory (HTM) was proposed by Mackenzie et al. (2018) to invistigate and evaluate it over real-world Sydney Coordinated Adaptive traffic system data. They used time series data that contained a timestamp and traffic data counts per lane. The data were collected from two sperated links, i.e., one was along a corridor and the other was at an intersection. The results of HTM were compared to the Long Short-Term Memory (LSTM) method. The results showed that HTM outperformed LTSM in terms of MAPE, RMSE, and Mean GEH2.

2.2.3 Studies predicting/estimating turning movements

In the literature, limited studies attempted to estimate/predict turning movements at signalized intersections, with the availability of total entering and exiting volumes at all approaches. For instance, Chen et al. (2012) developed a model to estimate turning movement counts at intersections by using a nonlinear Path Flow Estimator (PFE). They utilized grouped inbound and outbound peak hour volumes as an input in their model to derive the vehicles' path and calculate the turning movements' volumes. In addition, they assumed a constraint that the estimated volume cannot exceed the capacity of the approach. Their results showed that the error was acceptable for main corridors, but not good for minor ones. They concluded that the accuracy of the estimated turning movements could increase when more traffic counts and turning movements' data are

[2] GEH: a statistic commonly used to evaluate predictive traffic models, named after its creator, Geoffrey E. Havers.

available.

Another study aimed to estimate a reliable hourly turning movement estimation model using machine learning (Ghanim and Shaaban, 2018). An Artificial Neural Network (ANN) model was developed to estimate turning movements at a high level of accuracy using only approach volumes as an input. The model was trained to analyze the relationship between the approach volumes and the corresponding turning movements. Their results showed that the estimated through movements is close in different traditional estimation approaches. However, the right and left turn movements estimations were significantly better when using ANN model. Their developed model was limited to peak hours.

Finally, a latest study aimed to develop short-term prediction algorithm of movements at intersections based on the Partial Least Square model (PLS) (Li et al., 2020). Trajectory data was utilized in developing their prediction model. This data represents traffic conditions by providing different features (number of sampled trajectories, number of stops, and average speed) per 15 minutes. Their performance measures of Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) were 8.34 and 16.68%, respectively. The developed PLS algorithm was compared to different prediction models (i.e., ARIMA, K-NN and Support Vector Regression) and outperformed them. All these studies estimate/predict the movements in 15 minutes or one hour. To our best knowledge, no study attempted to estimate movement counts at intersections at the cycle level.

Past research focused on predicting traffic volume at intersections and freeways. Limited studies have been conducted to estimate/predict turning movement counts at signalized intersections. Table 2 summarizes the different methodologies that were adopted in the literature to estimate/predict traffic parameters, the estimation/prediction duration, and the achieved accuracy. To the best of the authors' knowledge, no one attempt to estimate short-term turning

movement counts at signalized intersections at the cycle level. Hence, this research aims to extend past research and bridge the gap, as previous studies focused on estimating 15 minutes or one-hour traffic movements at signalized intersections with the total entering and exiting movements considered as inputs in the developed models.

Table 2: Summary of the adopted estimation/prediction methodologies in literature

Reference	Objective	Method	Time Interval	Performance Measures
Chen et al. (2012)	Predicting turning movements at urban signalized intersections	Path Flow Estimator	One hour	RMSE = 28.89
Ghanim and Shaaban (2018)	Predicting turning movements at urban signalized intersections	Artificial Neural Network	One hour	RMSPE = 4.7%
Li et al., (2020)	Predicting turning movements at urban signalized intersections	Partial Least Square Model	15-min	RMSE = 8.34 MAPE = 16.68%
Stathopoulos and Karlaftis, (2003)	Predict traffic volume at urban arterials	Multivariate Time-Series State Space Model	15-min	MAPE = 15%
Williams and Hoel, (2003)	Predict short term traffic flow at fixed locations on freeways	Seasonal ARIMA Model	15-min	MAPE = 8.6%
Ghosh et al., (2005)	Predict traffic volume at urban signalized intersections	Seasonal ARIMA Model	15-min	RMSE = 42.89 MAPE = 28.5%
Ghosh et al., (2007)	Predict traffic volume at urban signalized intersections	Seasonal ARIMA With Bayesian Inference Model	15-min	APE = 5.4%
Kamarianakis and Prastacos, (2005)	Predict traffic volume at urban signalized intersections	STARIMA Model	7.5-min	RMSE = 42.11
Min et al., (2009)	Predict traffic volume at urban signalized intersections	Dynamic STARIMA Model	5-min	MRE = 9.49%
Min et al., (2010)	Predict traffic volume at urban signalized intersections	GSTARIMA model	15-min	MRE = 11.49%
Clark, (2003)	Predict traffic volume along a motorway network	K-Nearest Neighbor	10-min	RMSE = 6.6 MAPE = 10.4%
Guo et al., (2018)	Predict traffic volume at fixed locations on urban corridors	KNN Fusion Based Model	5-min 15-min	MAPE = 6.28% (normal conditions 5-min) MAPE = 19.4% (incident period 15-min)
Xia et al., (2016)	Predict traffic volume at urban signalized intersections	Spatial-Temporal Weighted KNN	5-min	MAPE = 28.56%
Yoon and Chang, (2014)	Predict traffic volume at urban signalized intersections	KNN-Nonparametric Regression	5-min	MAPE = 12.84%
Zheng and Su, (2014)	Predict traffic volume at urban signalized intersections	KNN-LSPC	1.5-min	MAPE = 11.31%
Qiao et al., (2013)	Predict travel time on freeways	KNN-Trend	5-min	MAPE = 9.74%
Xia and Chen, (2017)	Predict traffic volume at urban signalized intersections	GBDT	10-min	MAPE = 9.74%
Yang et al., (2017)	Predict traffic volume at fixed locations on freeways	GBDT	5-min	MAPE = 7.4%
Zhang and Haghani, (2015)	Predict traffic volume at fixed locations on freeways	GBDT	5-min	MAPE = 2.9% (non-peak) MAPE = 9.9% (peak)
Cheng et al., (2019)	Predict travel time at fixed locations on freeways	GBDT	5-min 10-min 15-min	MAPE = 2.45% (5-min) MAPE = 3.94% (10-min) MAPE = 4.66% (15-min)
Zhang et al., (2016)	Predict travel time on congested urban road networks	GBDT	30-min	MAPE = 9.6%
Jiang et al., (2005)	Predict traffic volume at fixed locations on freeways	Dynamic Wavelet Neural Network Model	One hour	Absolute error = 9%
Zheng et al., (2006)	Predict traffic volume at fixed locations on freeways	Bayesian Combined Neural Network model	15-min	MAPE = 6.10%
Zhu et al., (2014)	Predict traffic volume at urban signalized intersections	Artificial Neural Network	15-min	MAPE = 13.58%
Zhao et al., (2017)	Predict traffic volume at fixed locations on freeways	LSTM	15-min	MRE = 6.21

2.2.4 Simulation Models

Using simulation models to predict traffic parameters was adopted by some researchers. For example, Head (1995) estimated a model to estimate traffic flows. The estimated model was used in developing proactive real-time traffic-adaptive signal control logic and traffic flow theory. A simulation model was built using the TRAF-NETSIM traffic simulation model to demonstrate the prediction model. The prediction model was implemented to develop real-time traffic-adaptive signal control logic (hierarchical-distributed logic called RHODES) as a part of research. The model was implemented to predict data at one intersection, based on a traffic data given by detectors upstream this intersection. Moreover, the model was tested using Durbin-Watson (D-W) statistic measures. The research concluded that communications between adjacent intersections were required to share information about detectors and signal timings, and a better right and left turn permitted movements should be included in the model. Also, Lin et al. (2008) established a model that can simulate the traffic movement in the urban traffic network, and forecast the traffic flow states. He used microscopic model CORSIM (which was exploited by the Federal Highway Administration) as the practical traffic system. The model can simulate the current traffic movements in an urban network. Moreover, it can predict the future traffic flow states. In their research, they utilized the macroscopic Urban Traffic Network (UTN) model to forecast 5 minutes traffic flow using CORSIM. Their results proved that the model could compute the traffic status of the UTN accurately. However, the suggested method had too many parameters to be fixed and it heavily relied on the precision of detectors. Another example for using macroscopic UTN model and CORSIM simulation to predict short-term traffic flow was carried out by Kong et al. (2013). Their experiment verified the superior performance of the proposed prediction model.

To specify, parametric models are easy to implement and understand. Also, the models could be able to deal with spatial effects with other intersections and multivariate observations.

Non-parametric models are totally data-driven and thus free of assumptions on data distribution. While the models especially deep learning models could provide accurate estimation results, the model structures are complex. The models established based on the simulation could provide reasonable results and could be extended to develop signal control plans. However, the simulation needs to be calibrated for a specific network. Hence, it is required to calibrate the network including the target intersections first. As in this project we plan to estimate the turning movements at intersections, both parametric and non-parametric models will be considered. If possible, a hybrid model will be developed to incorporate the advantages of the two approaches. In addition, intersections are located on several arterials. For some intersections lack of sufficient observed turning movement data, the simulation method will be attempted to validate the developed model if necessary.

shows the summary of traffic volume prediction methodologies in the literature. To specify, parametric models are easy to implement and understand. Also, the models could be able to deal with spatial effects with other intersections and multivariate observations. Non-parametric models are totally data-driven and thus free of assumptions on data distribution. While the models especially deep learning models could provide accurate estimation results, the model structures are complex. The models established based on the simulation could provide reasonable results and could be extended to develop signal control plans. However, the simulation needs to be calibrated for a specific network. Hence, it is required to calibrate the network including the target intersections first. As in this project we plan to estimate the turning movements at intersections, both parametric and non-parametric models will be considered. If possible, a hybrid model will be developed to incorporate the advantages of the two approaches. In addition, intersections are located on several arterials. For some intersections lack of sufficient observed turning movement data, the simulation method will be attempted to validate the developed model if necessary.

2.3 Accuracy evaluation

In various research that aimed to predict different traffic flow parameters, many measures of effectiveness were utilized to evaluate and test different proposed models. The most commonly used accuracy measures are Mean Absolute Error (MAE), RMSE, and MAPE. To compare the prediction models for a single data set, MAE and RMSE are frequently used. However, MAPE is commonly used in case of comparing the performance of prediction models for two different data sets (Hyndman, 2010). Moreover, Hyndman & Koehler (2006) proposed another alternative for scaled percentage errors to evaluate prediction accuracy Nevertheless, many other measures of effectiveness have been used to evaluate and compare different models in previous literature such as:

- Absolute Percentage Error (APE)
- Mean Squared Error (MSE)
- Mean Percentage Error (MPE)
- Mean Absolute Relative Error (MARE)
- Mean Relative Percent Error (MRPE)
- Mean Absolute Deviation (MAD)
- Variance of Absolute Percentage Error (VAPE)
- Probability of Percentage Error (PPE)
- Root Relative-square Error (RRSE)
- Mean Absolute Scaled Error (MASE)

The following Table 3 and 4 summarize the measures of effectiveness that were used to evaluate different parametric and non-parametric models consecutively. It could be noticed that the most common measures are MAE and MAPE. In addition, Head (1995) used MAE while Zhu et al. (2016) used MAPE for simulation models.

Table 3: Summary of the used measure of effectiveness in parametric models

Model	Citation	Developed Model	Measures of Effectiveness
	Chang et al. (2000)	ANN and ARIMA	MSE and MAE

Parametric models	Williams & Hoel (2003)	SARIMA	MAPE
	Ghosh et al. (2005) and (2007)	SARIMA	APE
	Guo et al. (2013)	Singular Spectrum Analysis (SSA)-Grey System Model (GM)	MPE, MAPE, and RMSE
	Zhan et al. (2018)	TDEC Ensemble model	MAE
	Stathopoulos & Karlaftis (2003)	State space and ARIMA	MAPE
	Stathopoulos & Karlaftis (2003)	State space and ARIMA	MAPE
	Kamarianakis & Prastacos (2005)	STARIMA	MSE
	Chen et al (2010)	PFE	RMSE and MSE

Table 4: Summary of the used measure of effectiveness in non-parametric models

Model	Citation	Developed Model	Measures of Effectiveness
Non-parametric models	Chang et al. (2000)	ANN and ARIMA	MSE and MAE
	Yoon & Chang, (2014)	KNN	MAPE and MAE
	Mackenzie et al, (2018)	LSTM and HTM	MAPE and RMSE
	Guo et al. (2018)	Fusion (average, weighted, and KNN) model, (NN, SVR, and RF)	MPE, MAPE and RMSE
	Zhu et al. (2014)	RBF-NN and ARIMA	MAD, MAPE and RMSE
	Zheng & Su (2014)	KNN-LSPC	MAPE
	Vlahogianni et al. (2005)	MLPs and genetic algorithm	MAE, MSE, and MRPE
	Ghanim & Shaaban (2018)	ANN	MSE, SE, RMSPE and R^2
	Xia et al. (2016)	Spatial–temporal-weighted-KNN	MAPE, RMSE, MSE, and MAE
	Qiao et al., (2013)	Historical average, ARIMA, KNN and modified KNN-T	RMSE, RRSE, MAE, MAPE, and MARE

2.4 Summary

From the literature review, it was shown in many studies that the nonparametric models performed better than the parametric models. However, the parametric models also have their characteristics that were proven to have superior performance in many cases. The literature illustrated that the multivariate data were helpful in case of multiple-steps-ahead forecasting, and the ARIMA approach was very helpful in case of entry points of the study area. It was also concluded that the Bayesian SARIMA models can improve the modeling of the traffic behavior in reality with rapid fluctuations and extreme peaks. Finally, for parametric models, the STRE model outperformed the autoregressive moving average (ARMA). Moreover, Random walk model, Holt-Winters' exponential smoothing techniques, and the seasonal ARIMA model failed to match well in cases of some abrupt changes in the observed traffic flow data. Further, previous studies verified the importance of the interactions between adjacent intersections in case of short-term traffic forecasting.

With respect to nonparametric (machine learning) techniques, previous studies adopted

using KNN, NN, GBDT, and LSTM models. Also, many comparisons between parametric and nonparametric models were carried out. For the KNN models, both KNN and KNN-T were found to have better performance than ARIMA and Kalman Filter. Also, studying the trend effects of travel time patterns has the potential to improve the prediction accuracy. Moreover, the KNN fusion based method achieved significantly superior results, especially during disrupted traffic conditions. However, the method requires calibrating multiple prediction models in parallel. As for the KNN-NPR model, it outperformed Kalman Filtering and ARIMA models. Another KNN model is the KNN-LSPC, that outperformed KNN and Kalman Filter algorithms. However, KNN-LSPC is free of assumptions on data distribution and has high flexibility and extendibility more than Kalman Filter algorithms. Finally, STW-KNN outperformed ANNs, KNN, NB, and RF. From the previous results, it could be concluded that various KNN models are beneficial methods in case of short-term traffic flow forecasting. It was indicated that the KNN model outperformed several parametric models.

The other commonly used approach of nonparametric techniques is neural networks models. Neural networks models were proven to be efficient in case of forecasting turning movements at intersections with high level of accuracy. They outperformed the nonlinear PFE method which required more traffic counts and turning movements' data to increase the prediction accuracy. Moreover, for estimating short-term traffic flow, neural networks model performed better than the univariate ARIMA and multivariate State-Space models. Previous literature also recommended using RBFNN model to calculate traffic flows to increase prediction accuracy. While in the case of using LCG-BN, previous studies recommended utilizing speed data instead of spatial data to get a significant higher prediction accuracy. Also, it was indicated that the Bayesian combined neural network (BCNN) model performed better than Back-Propagation (BP) and Radial Basis (RFB) neural network predictors.

Gradient Boosting Decision Tree (GBDT) model was adopted in many studies to predict traffic parameters (traffic volume and travel time) at urban intersections and freeways. In some studies, it was found to have better performance than Support Vector Machine (SVM) and Back Propagation Neural Networks (BPNN) models. It also improved the prediction accuracy and model interpretability. An alternative approach to estimate short-term traffic is the deep learning nonparametric models (e.g., Long Short-Term Memory network). In recent research, deep learning models were found to be promising in many research fields due to their ability to capture nonlinearity in models.

As shown in Table 1, the majority of previous studies concentrated on urban networks. The prediction accuracy of both parametric and nonparametric prediction models was proved to be acceptable on urban networks. In addition to urban networks, there exists a large proportion of suburban networks which are more eager for traffic movements prediction with insufficient traffic monitoring devices. Also, there exist huge differences between suburban and urban networks in terms of road network density, travel patterns, traffic speed, volume, and density.

In general, past research focused on predicting traffic volume at intersections and freeways. Limited studies have been conducted to estimate/predict turning movement counts at signalized intersections. To the best of the authors' knowledge, this study is the first attempt to estimate/predict short-term turning movement counts at signalized intersections at the cycle level. It extends past research and aims to bridge the gap, as previous studies focused on estimating 15 minutes or one-hour traffic movements at signalized intersections with the total entering and exiting movements considered as inputs in the developed models. Moreover, in this study, it was assumed that the target intersection has no traffic volume data. Thus, implementing estimation/prediction algorithms at target intersections to provide traffic volumes will save the extensive cost of data collections; as detector systems cost more than $20,000 per intersection.

CHAPTER 3. ANALYSIS OF THE GRIDSMART DATA

In the previous tasks, an extensive literature review was conducted, concluding different methods, data, and performance measures that were previously utilized to estimate different traffic parameters at signalized intersections. Moreover, data was collected and explored to determine whether it will be utilized in developing the turning movement estimation algorithm(s) or not. As a result, two main detectors systems data in the study area were collected: GRIDSMART and Miovision systems. This chapter discusses the development of turning movements estimation algorithm(s) for the identified intersections using traffic data from adjacent intersections based on the GRIDSMART system. The collected data were split into training and test data to develop the required turning movement estimation algorithms, and to validate the developed algorithms.

3.1 Data Collection and Preparation

Input data are crucial to provide accurate estimation results. In recent years, with the advancement of Big Data, abundant data can be used for the better vehicle movement estimation. Traffic data from different locations could provide data such as volume, travel time, and speed. Among the traffic data, traffic volume data is the basic data source for vehicle counts estimation. Several studies tried to incorporate signal status data with traffic volume data to better in estimating vehicle counts (Kong et al., 2013; Zheng and Liu, 2017). GRIDSMART is one of the detectors systems that could provide detailed volume data considering different turning movements (i.e., though, right-turn, left-turn, and U-turn movements) as well as signal timing data.

GRIDSMART is a single-camera system that gathers traffic data. GRIDSMART empowers traffic engineers to adjust signal timing and traffic flow strategies and enables real-time monitoring and visual assessment. Automated video data analysis is used by GRIDSMART cameras to identify vehicles crossing through user-defined regions at intersections. Processing the

resulting data could generate volumes per approach and turning movement counts. The system contains a fisheye camera mounted high above the respective location to gather data. It could provide a full intersection view, including the center where cars and trucks. The horizon to horizon approach offers highly accurate turn counts, views and functionality from the center of the intersection. This section will define the study area, the different groups of intersections, and data processing and aggregation for algorithms development.

3.1.1 Study area and data collection

The GRIDSMART data were collected along three main corridors on Orange county, Florida: Orange Blossom Trail (US 441), Mills Avenue (US 17/92), and Colonial Dr. (SR 50). **Error! Reference source not found. Error! Reference source not found.** shows an overview to the study area where GRIDSMART data was collected. IP addresses for 32 intersections along the three corridors were provided through the GRIDSMART software. Twelve intersections are along Orange Blossom Trail corridor, one of them is the intersection between the US 441 and SR 50 corridors. There are 17 intersections along US 17/92, one of them intersects with the SR 50 corridor. Finally, five intersections are along SR 50, two of which are the intersections with SR 17/92 and US 441.

Figure 1: Locations of intersections

The data were explored for all months in 2018 and 2019 to check the availability and completeness in the GRIDSMART system. The data of January 2019 was chosen as the month had the most complete data for US 441 (some data were missing in other months, while some cameras were not installed yet, and some intersections were not connected to the GRIDSMART system). For US 17/92, data in May 2019 was selected as it was the most complete and consistent data for this corridor (on other months some data were missing, some cameras were not installed yet, and some intersections weren't yet connected to GRIDSMART). However, there was no data for consecutive intersections on SR 50. Hence, the turning movement algorithms were developed for the first two corridors (US 441 and US 17/92). Through the GRIDSMART system, the turning movements counts data for the selected dates were downloaded by using the IP addresses. Further,

the real-time videos from GRIDSMART cameras were manually observed for each intersection which were used to match the downloaded data with the user-defined lanes. Figure 2 shows GRIDSMART system interface.

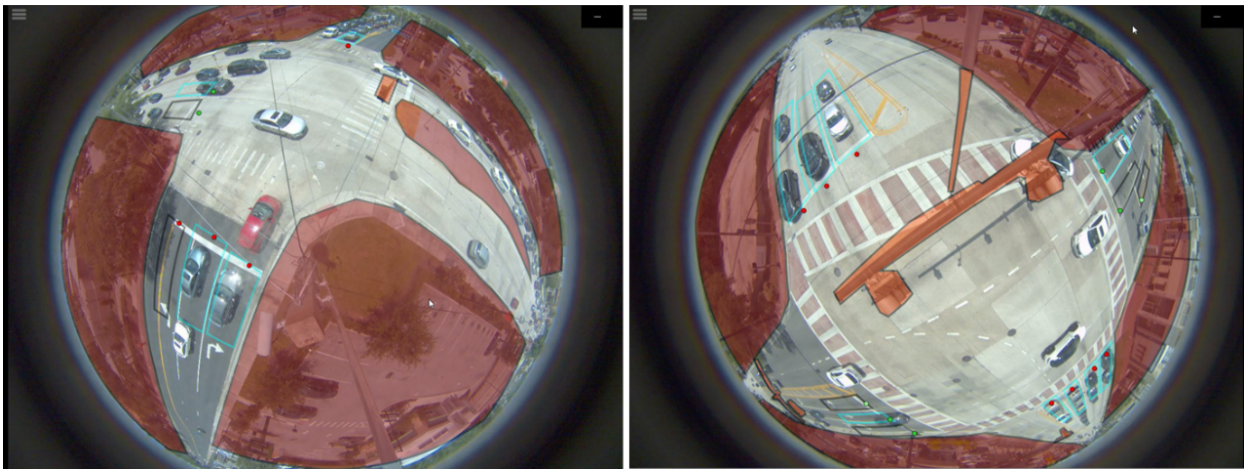

Figure 2: GRIDSMART system interface

After checking all the GRIDSMART data the following problems were identified:

- The intersection between US 441 and South St. can connect to GRIDSMART but we cannot view video on the system or download its real-time data.
- Three intersections cannot connect to GRIDSMART:
    1- The intersection between US 17/92 and SR 50.
    2- The intersection between SR 50 and Magnolia St.
    3- The intersection between US 17/92 and Packwood Avenue.
- For the Intersection of SR 50 & Westmoreland St., since 23rd of January, there is no counts for all turning movements (all counts are zeros). Therefore, it will be excluded from the study.
- For Intersection US 17/92 & Princeton St., the data for the curb Southbound lane is classified as left turn movement (the turns are only left) while it is a right turn lane.
- For Intersection US 441 & Long St., the Northbound data files doesn't accurately match

the video.

- For Intersection US 441 & Robinson, some data were missing on January 13th.

Thus, five intersections were excluded from data collection (US 441 & South St., US 17/92 & SR 50, SR 50 & Magnolia St., US 17/92 & Packwood Avenue, SR 50 & Westmoreland St., Only the counts data for 27 intersections were obtained from the GRIDSMART system including 15 intersections along US 17/92, 11 intersections along Orange Blossom Trail corridor, and 2 intersections along SR 50 (including the intersection with US 441). The Intersection of SR 50 & Orange Avenue will be excluded because it is far from other intersections. Thus, it will not be beneficial to be used in developing the algorithm. Therefore, to develop the turning movement estimation algorithm(s), 26 intersections' counts data along two main corridors (i.e., US 17/92 and US 441) will be explored to consider which intersections' data will be utilized.

Traffic counts data

The downloaded counts data from GRIDSMART contains 14 variables: Count Version, Site Version, Local Timestamp, UTC offset (minutes), Turn, Vehicle Length (ft), Vehicle Speed (MPH), Light State, Seconds in Zone, Vehicles Remaining in Zone, Seconds of Light State, Seconds since Green, Zone Recent (FFS), and Zone Calibration (FFS). Each file in the downloaded counts folder contains several excel files, each represents a lane at the intersection where a virtual detector was implemented in the system. Each file represents the data per lane per day. The data contains many records, each record represents one vehicle that occupied the intersection at a specific lane and approach. Since files names don't define their corresponding lane, ten minutes videos were recorded for all the intersections in order to match the ground-truth lanes with the implemented virtual detectors to the different counts' files. Hence, the corresponding lane for each ID could be identified. Afterwards, data were combined in order to represent the total turning movements counts for each single approach (e.g., NB-S, NB-R, NB-L). Note that the 'S' in the

files represents the through movements in the GRIDSMART system.

The previously downloaded real-time traffic counts data from GRIDSMART were aggregated into one-hour traffic counts in order to determine the peak and off-peak periods. Thus, the average hourly total traffic volumes along each corridor were calculated with respect to weekdays and weekends. The following Figure 3 and Figure **4** and show the average hourly traffic counts of all intersections along US 17/92 and US 441, respectively. The figures show that the hourly traffic volume trend was the same for both corridors. Thus, for weekdays, four periods were identified. The AM peak period were found to be from 07:00 to 9:00, while the PM peak period is from 16:00 to 18:00. The off-peak periods are from 06:00 to 07:00 and from 09:00 to 15:00. Finally, the Nighttime is from 18:00 to 06:00. Moreover, the weekends graphs show only one peak from 12:00 to 13:00. As weekdays have higher traffic volumes, this study will focus on weekdays.

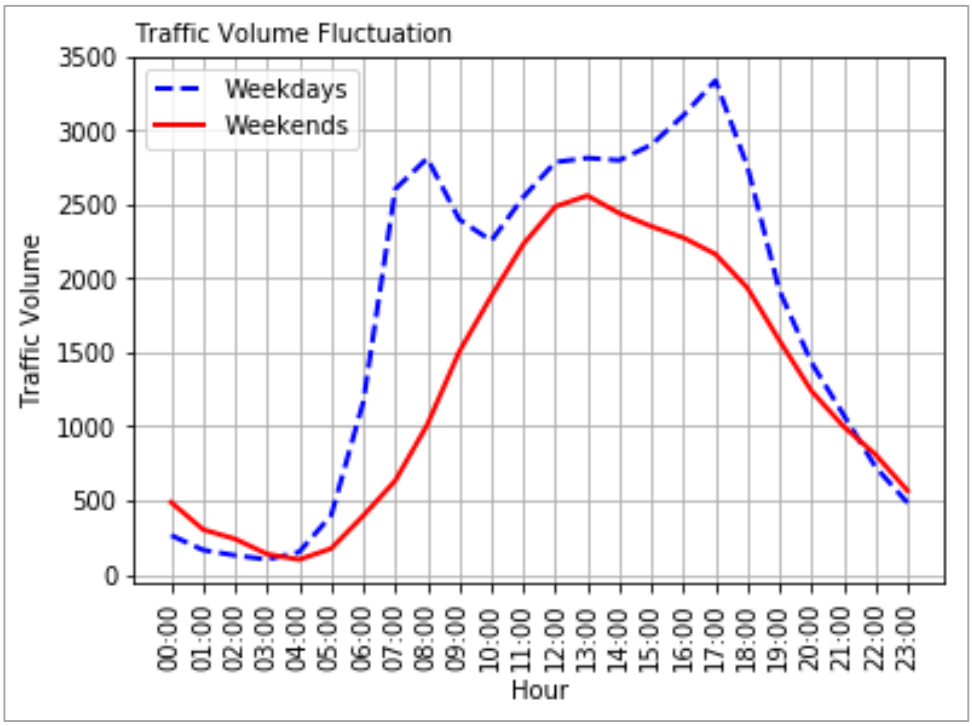

Figure 3: Average hourly traffic volumes along US 17/92

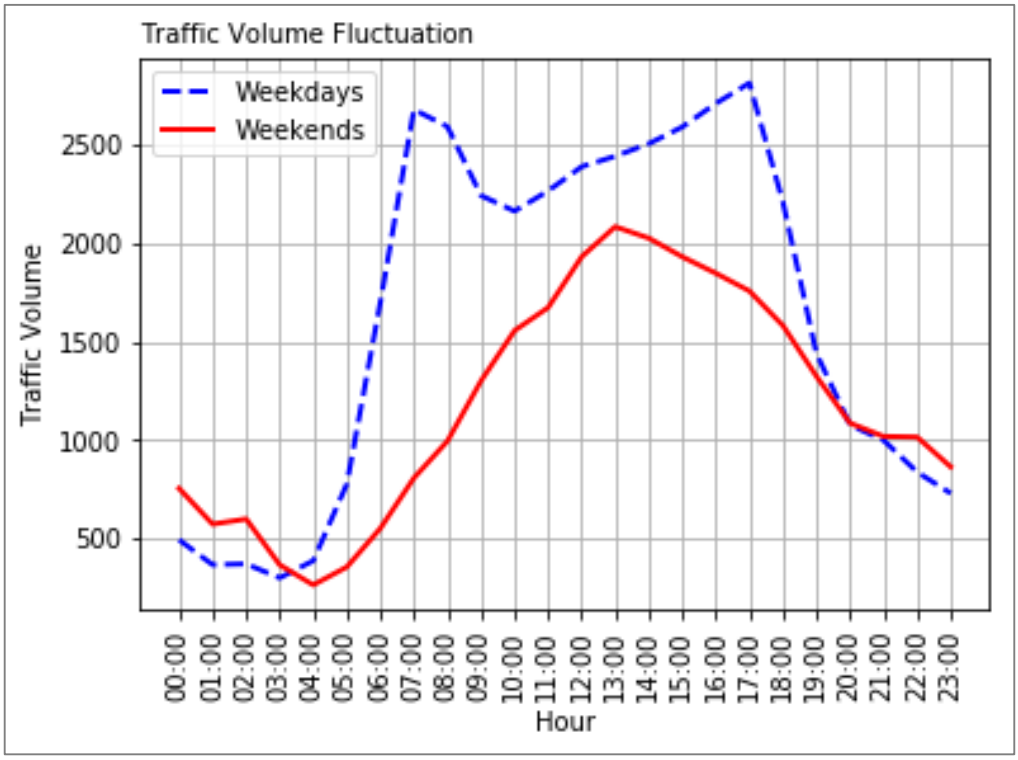

Figure 4: Average hourly traffic volumes along US 441

Moreover, Figure 5 and Figure **6** show the average hourly traffic volumes for both Northbound and Southbound directions for US 17/92 and US 441, respectively. The graphs were plotted for the pre-defined time periods (AM peak, PM peak, off-peak, and Nighttime). The graphs show the same trend for each time period for a certain corridor, with some changes in the Northbound and Southbound volumes according to different time periods.

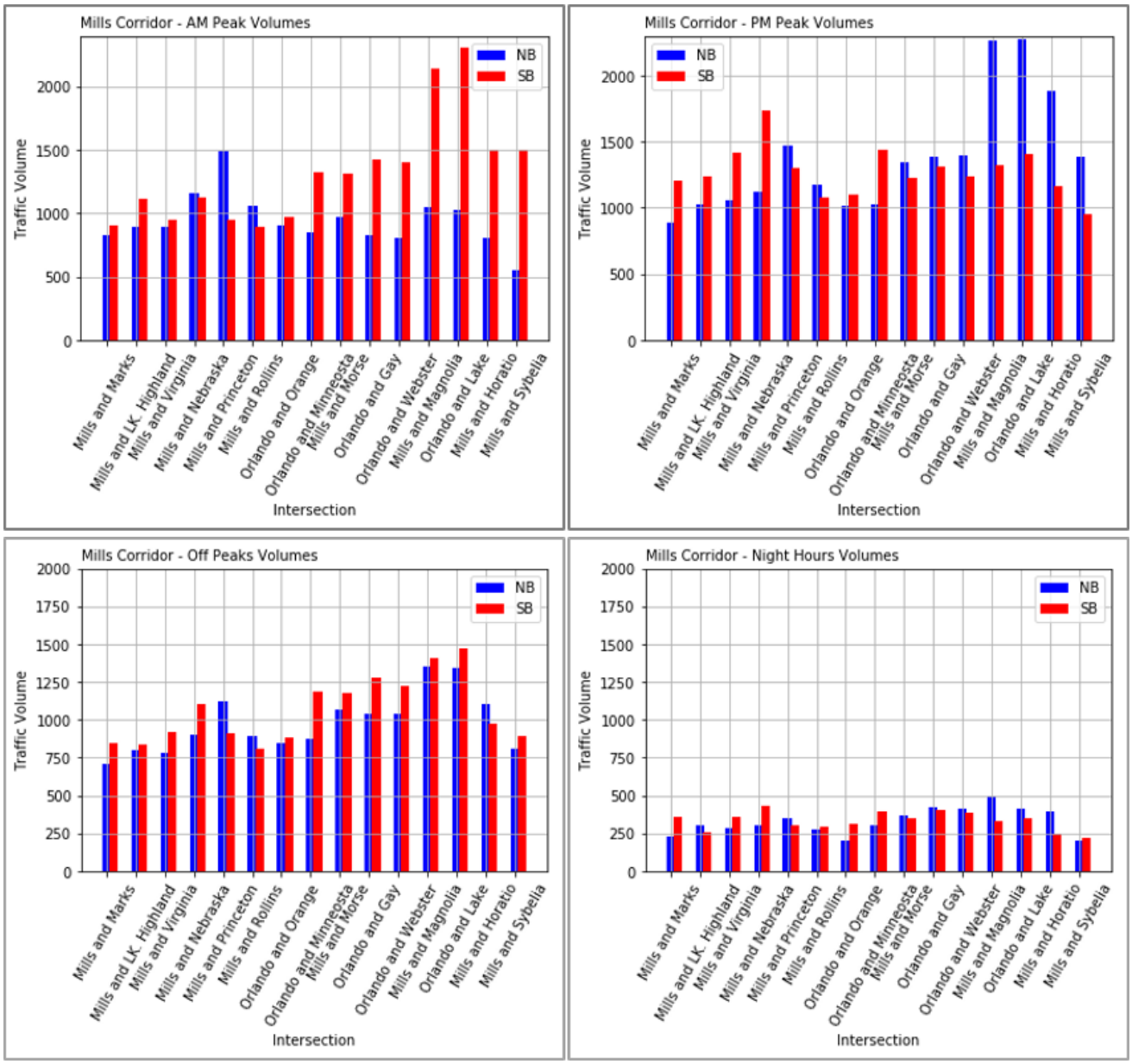

Figure 5: Average hourly traffic counts for Northbound and Southbound directions - US 17/92

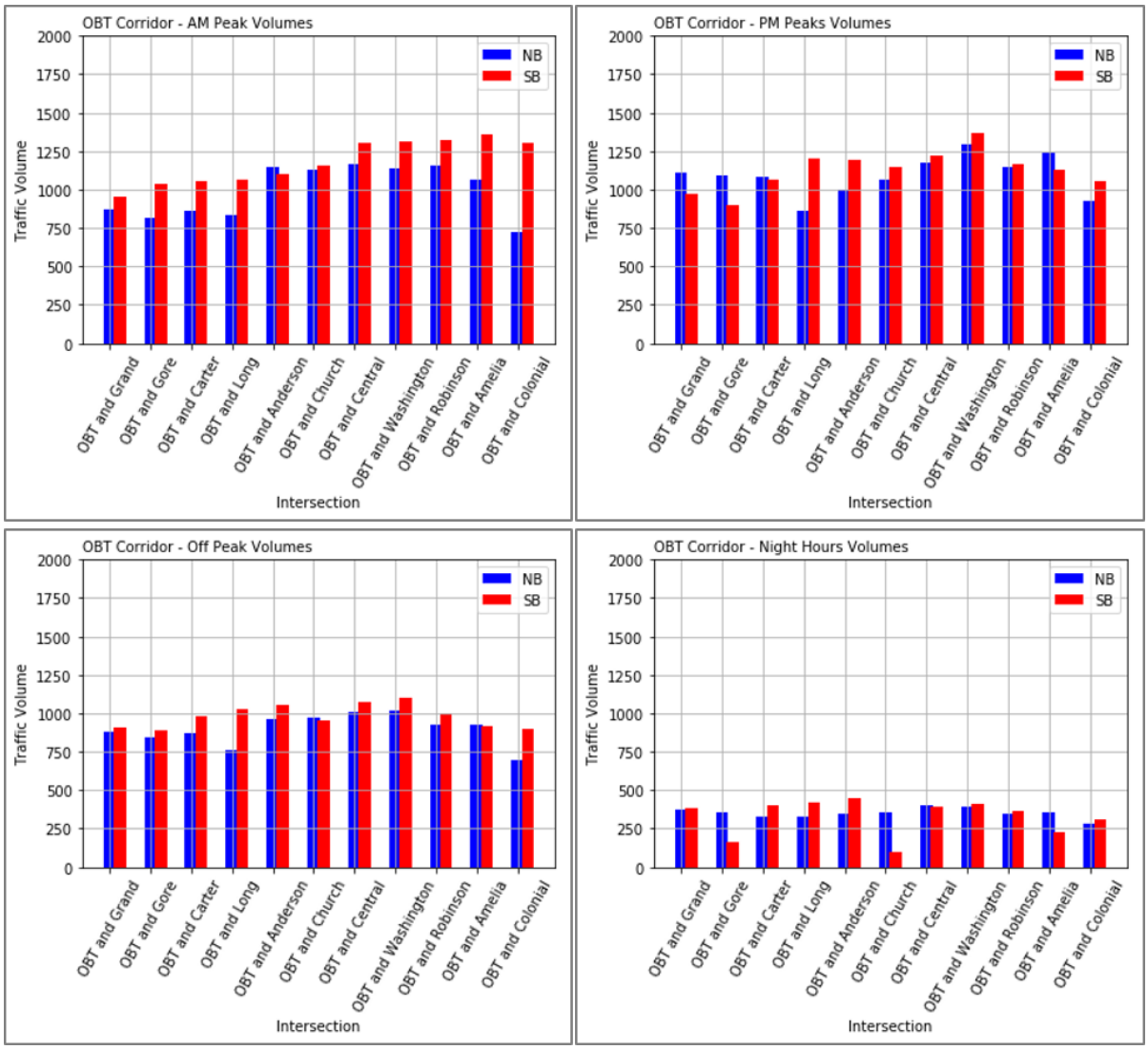

Figure 6: Average hourly traffic counts for Northbound and Southbound directions – US 441

The data counts for the two corridors were aggregated into five minutes to calculate the Coefficient of Variation (CV), CV is the standard deviation divided by the mean volume. The CV could be used to measure the variability of weekdays traffic counts for the four approaches at each intersection in both US 17/92 and US 441.

Figure 7 andFigure **8** show the CVs at the pre-defined time periods for both US 17/92 and US 441, respectively. At US 17/92, the coefficient of variation was relatively small in the AM and PM peaks for Northbound and Southbound directions compared to the Eastbound and Westbound directions. Moreover, for the US 441, there were no big variations in CVs for AM peak, PM peak,

and off-peak periods. The CVs for both US 17/92 and US 441 showed different variations for different time periods. Thus, this leads to the consideration of developing different turning movement estimation algorithm(s) for different time periods in the next task of the project.

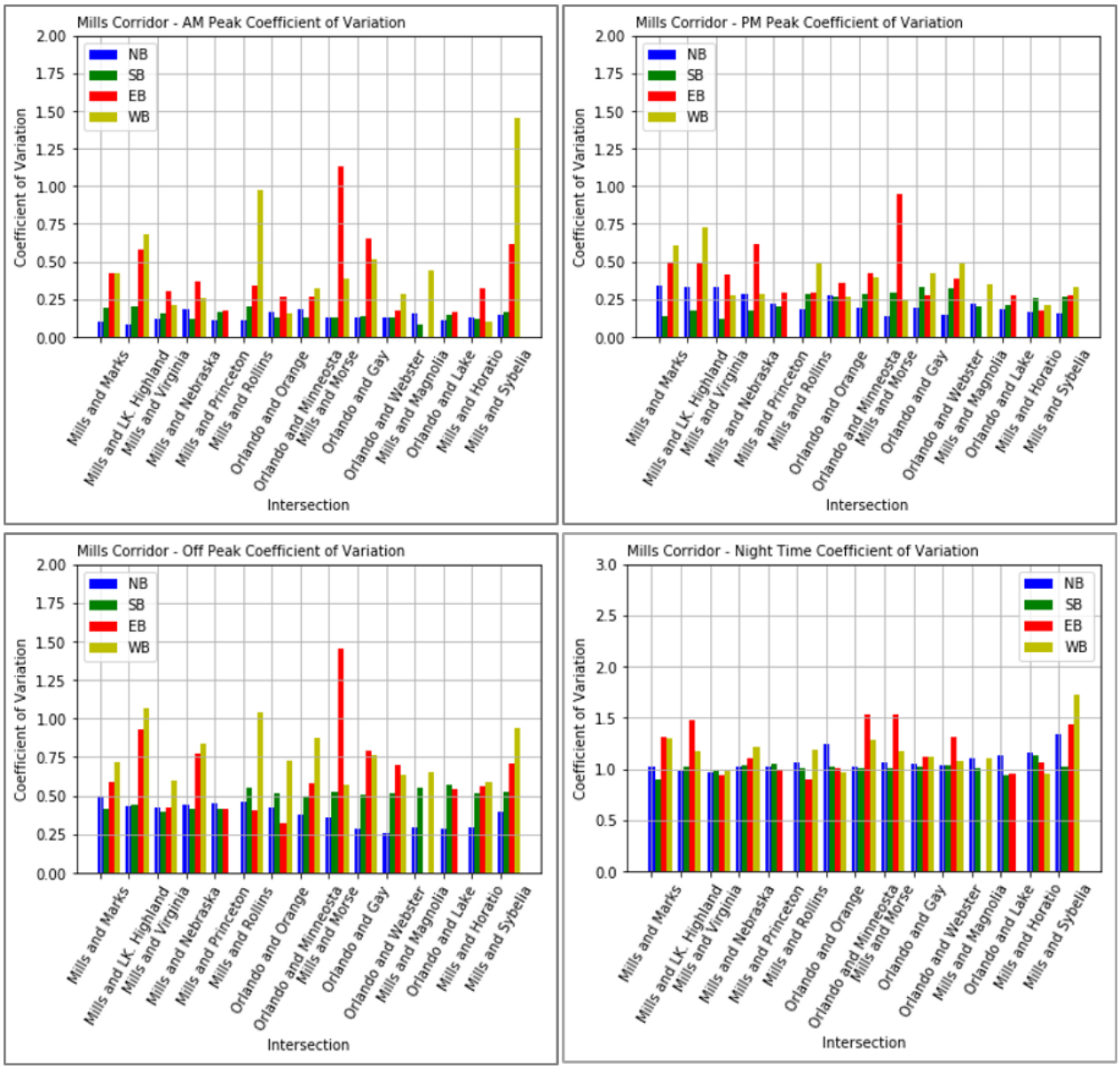

Figure 7: Coefficient of Variation at US 17/92

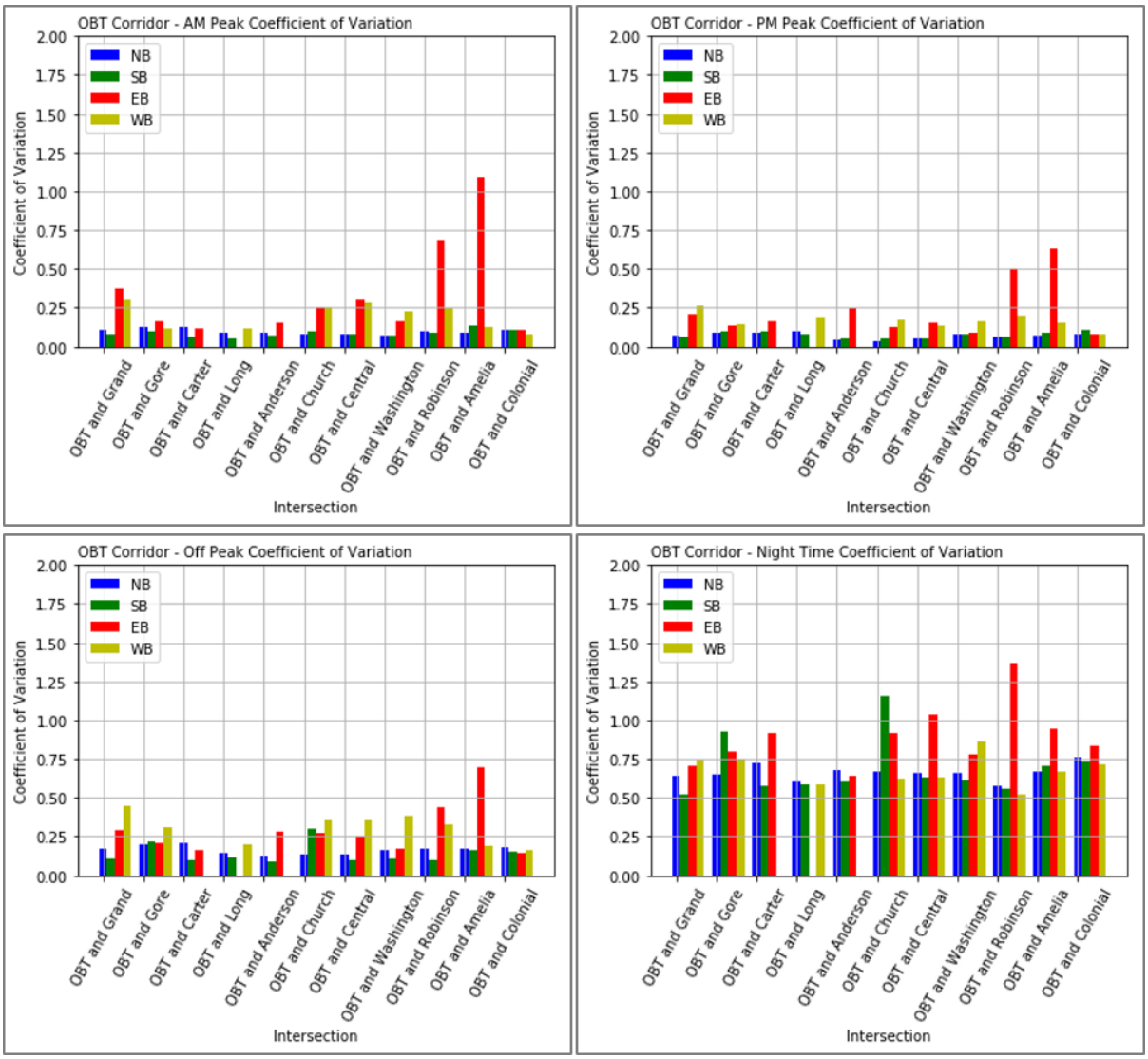

Figure 8: Coefficient of Variation at US 441

The coefficient of variation for both Northbound and Southbound were plotted for different intersections along each corridor in Figure 9 and Figure **10**, respectively. The CV values indicate the stability of the entering vehicles by different movements. The very high CV is associated to unstable entering traffic at certain intersections during certain time periods, which might be difficult to estimate. For US 17/92, the Coefficients of Variation (CV) were relatively low, especially for AM and PM peaks. However, at Nighttime, CVs significantly increased. Two intersections (US 17/92 & Rollins St. and US 17/92 & Orange Avenue) had large CV values during nighttime. Also, the CVs between US 17/92 intersections with Marks St. and Lake Highland St.

was relatively high at the same period (Nighttime period). At the US 441, the CVs gave the same trend as in the Mills corridor. The values for both AM and PM peaks were relatively low. However, the CVs increased in the off-peak period slightly. Then, the CVs significantly increased at Nighttime. Moreover, the CV values of Intersections US 441 Gore and US 441 & Church were high at Nighttime.

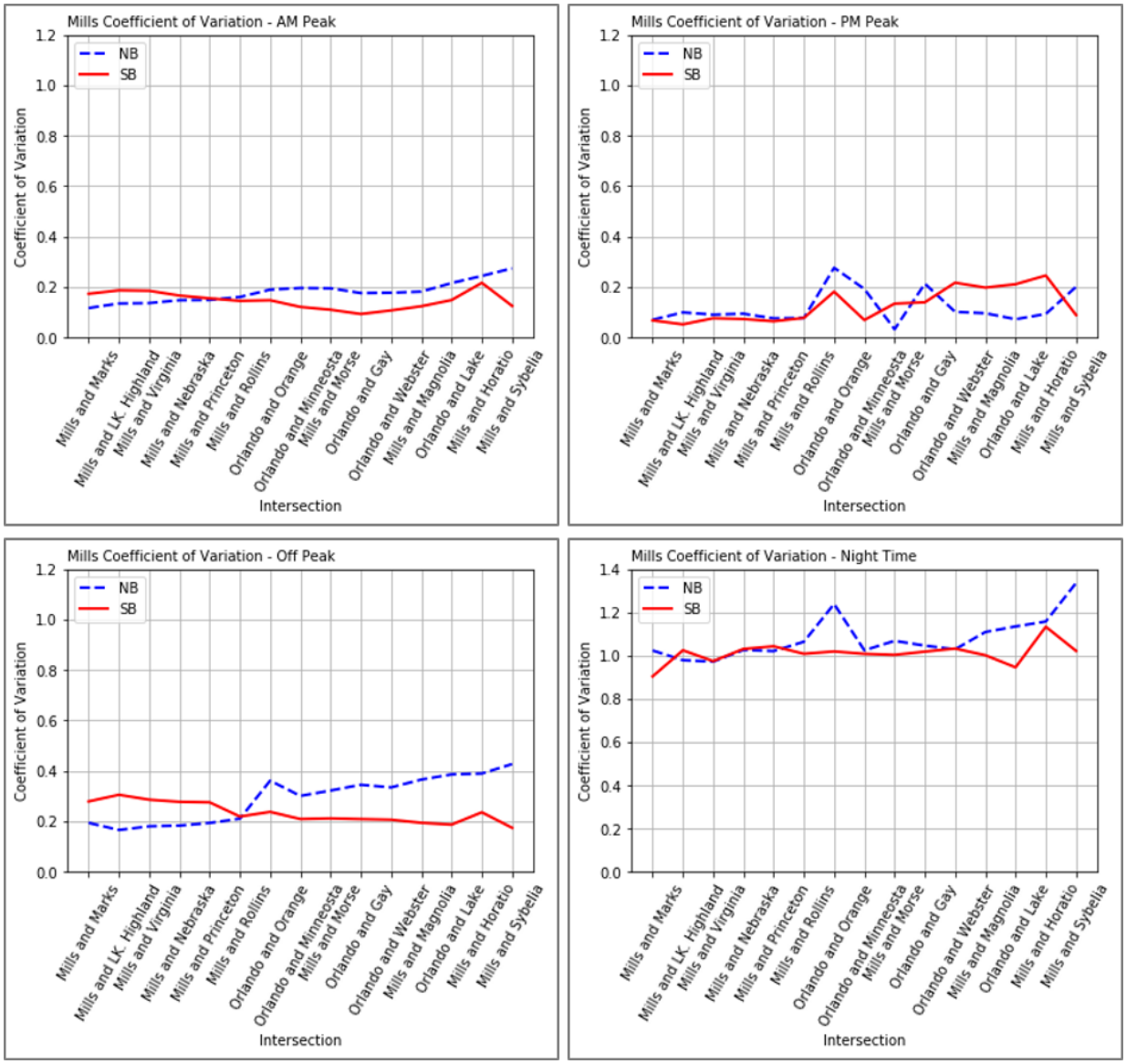

Figure 9: Coefficient of Variation for Northbound and Southbound at US 17/92

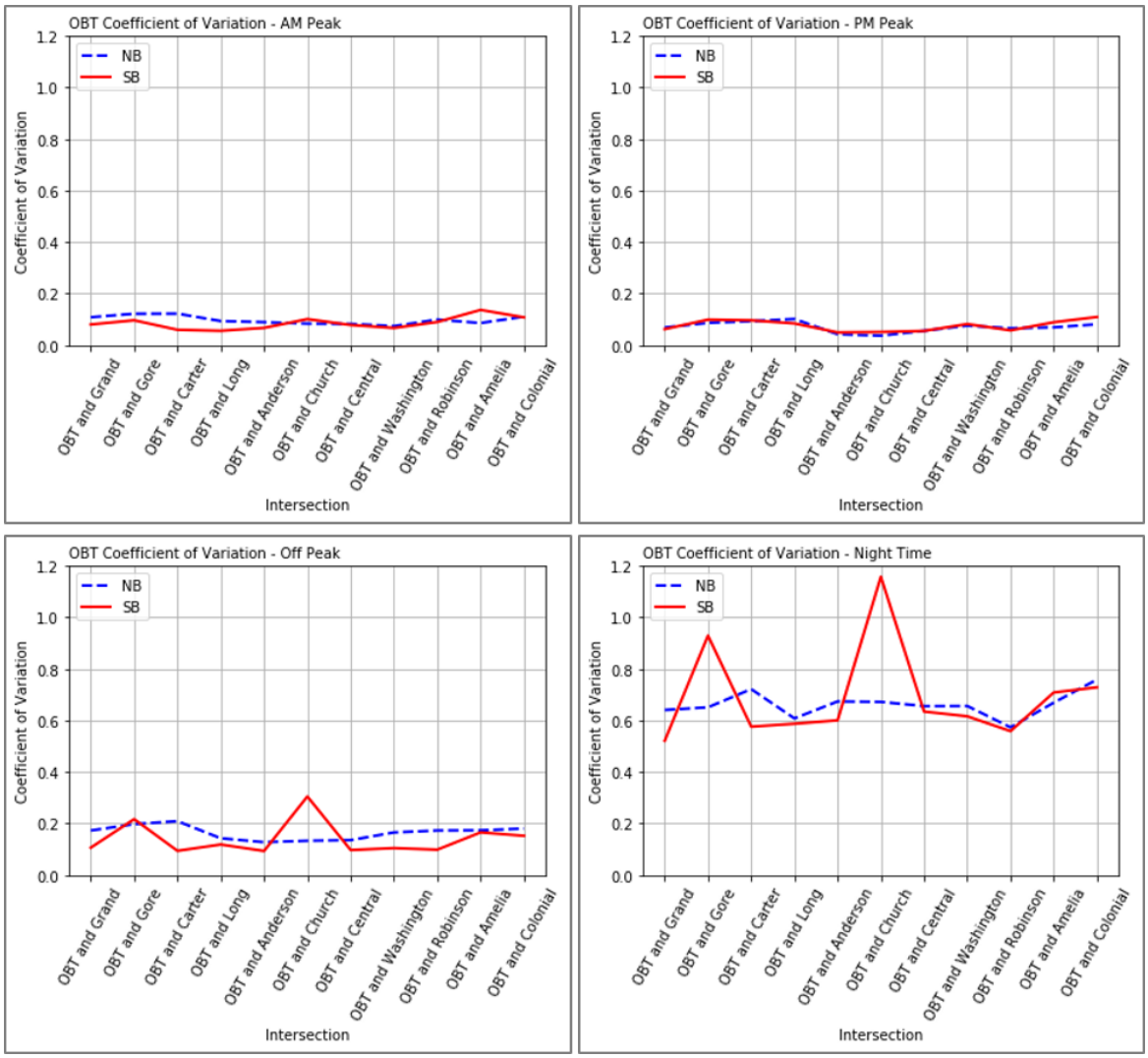

Figure 10: Coefficient of Variation for Northbound and Southbound at US 441

Signal data

The signal data were downloaded from the GRIDSMART system along with the traffic counts data. The events files contain timestamps, codes, and data for certain system events such as reboot, publish, flash, and phase changes. **Error! Reference source not found.** shows the GRIDSMART interface with the signal state of different phases. By exploring the cycle time for each intersection, the cycle times were found to vary for the same intersection. Thus, intersections' traffic signals along the two corridors were found to be fully actuated. However, one problem was detected while investigating downloaded events files from the GRIDSMART system. Two intersections which

are Intersections US 441 & Carter St. and Intersection US 441 & Long St. are controlled from the same cabinet. Thus, the downloaded data files from GRIDSMART are from the same directory. However, only one events file for signal data is downloaded and it's not defined to which intersection it belongs to. Hence, these two intersections could not be used in this report.

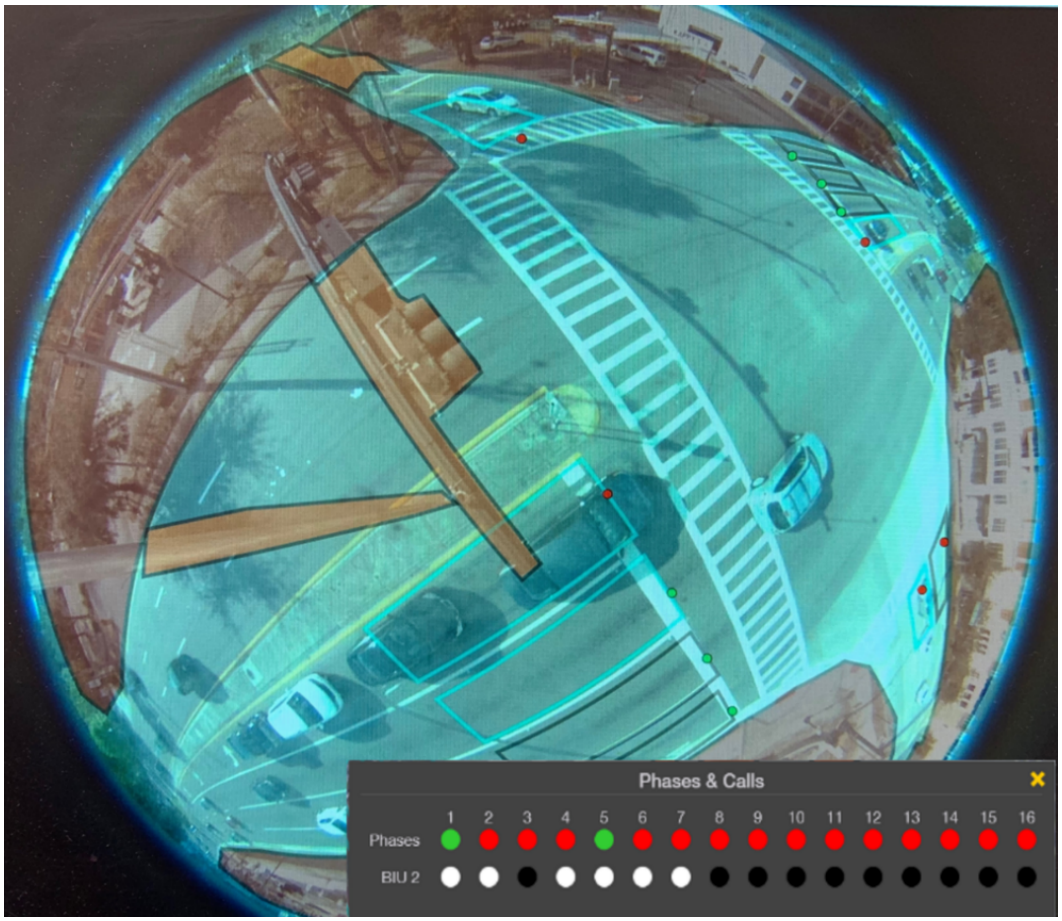

Figure 11: GRIDSMART signal state for different phases

Finally, all intersections with missing or corruption of data, low correlation, high coefficient of variations, large segment's distance between intersections, or large number of access points/minor roads intersections were excluded. Finally, the data for 11 intersections on US 441 corridor and 8 intersections on US 17/92 were downloaded. Figure 12 shows the final intersections to be utilized in developing turning movements estimation algorithms on both corridors. Furthermore, the number of through and left turn lanes for each intersection is shown in

Table 5. This shows the difference in geometric characteristics that might affect the estimation accuracy.

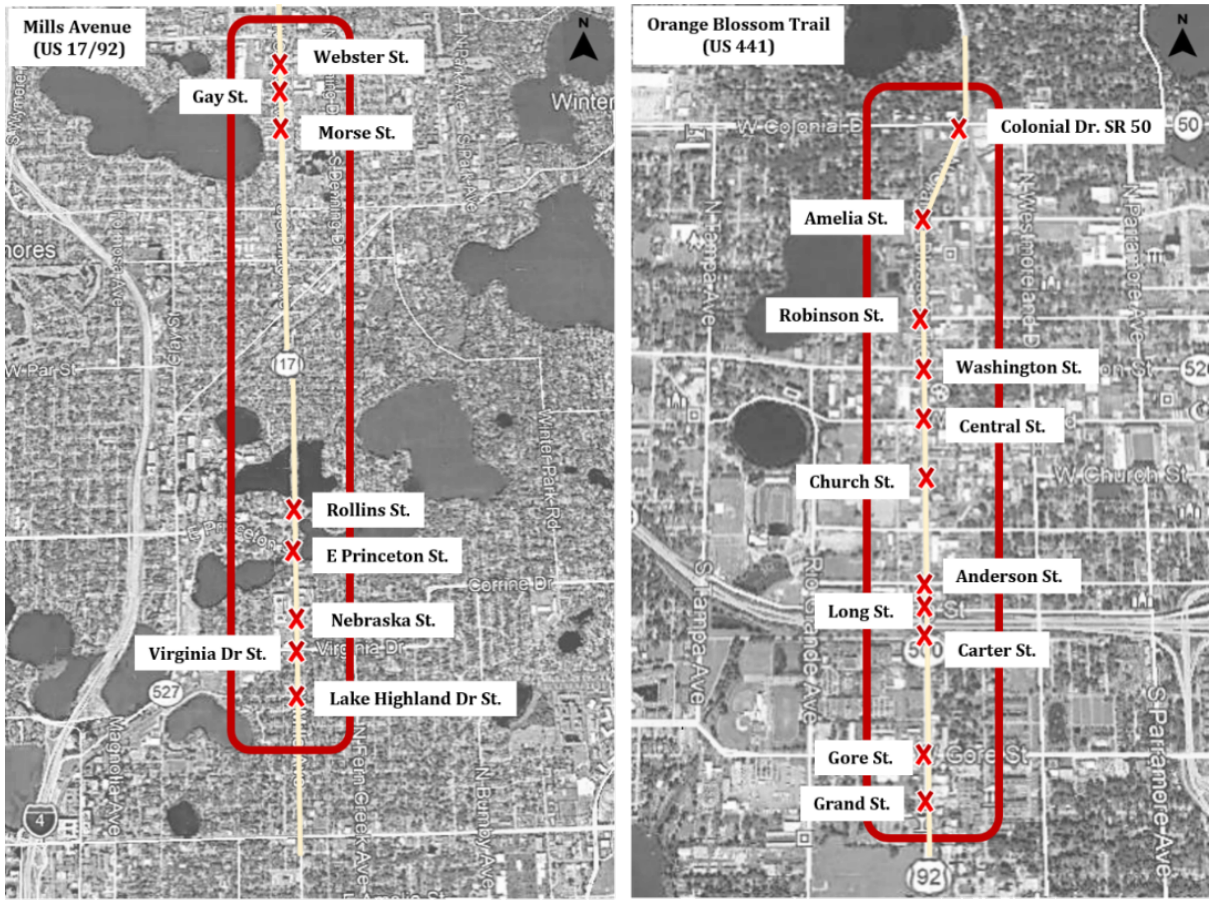

Figure 12: Intersections with consistent and complete data on GRIDSMART

Table 5: Number of through and left turn lanes for each intersection

Mills Avenue US 17/92								
Intersection	Northbound		Southbound		Eastbound		Westbound	
Movement	Through	Left	Through	Left	Through	Left	Through	Left
Rollins	2	1	2	1	1	1	1	1
Princeton	2	2	2	--	--	1	--	--
Nebraska	2	1	2	1	1	1	1	1
Virginia	2	1	2	2	1	1	1	1
Lk. Highland	2	1	2	1	**		**	
Webster	2	1	2	*	1	1	1	1
Gay	2	1	2	1	1	1	1	1
Morse	2	1	2	1	2	2	2	2
OBT Corridor US 441								
Intersection	Northbound		Southbound		Eastbound		Westbound	
Movement	Through	Left	Through	Left	Through	Left	Through	Left
Colonial Dr.	2	1	2	1	2	1	2	1
Amelia	2	1	2	1	**		1	1
Robinson	2	1	2	1	**		1	1
Washington	2	1	2	1	1	1	1	1
Central	2	1	2	1	1	1	1	1
Church	2	1	2	1	1	1	1	1
Anderson	2	*	2	1	**		--	--
Long	2	1	2	--	--	--	2	***
Carter	2	--	2	1	2	1	--	--
Gore	2	1	2	1	1	***	1	***
Grand	2	1	2	1	**		**	

* indicates no left turn is allowed
** indicates one shared lane
*** indicates one lane shared with through

3.1.2 Data grouping

The segments between adjacent intersections could have different access points/minor roads, layouts, geometric characteristics, signal control plans, and data availability. To ensure the accuracy of estimation results, it is necessary to divide the study intersections into several groups considering certain factors. In this section, intersections were grouped depending on the distance between intersections, number of access points/minor roads between two consecutive intersections, and data availability. Table 6 summarizes the distances as well as the number of access points/minor roads between each pair of consecutive intersections for both corridors (US 17/92 and US 441).

Table 6: Distances and number of access points/minor roads between two consecutive intersections

Mills Avenue US 17/92			
Intersection 1	**Intersection 2**	**Distance between intersections (miles)**	**Number of minor roads between intersections**
Rollins	Princeton	0.2	0
Princeton	Nebraska	0.35	1
Nebraska	Virginia	0.1	1
Virginia	Lk. Highland	0.2	2
Webster	Gay	0.15	0
Gay	Morse	0.3	7
OBT Corridor US 441			
Intersection 1	**Intersection 2**	**Distance between intersections (miles)**	**Number of minor roads between intersections**
Colonial Dr.	Amelia	0.3	2
Amelia	Robinson	0.25	1
Robinson	Washington	0.13	2
Washington	Central	0.13	2
Central	Church	0.13	2
Anderson	Long	0.06	0
Long	Carter	0.06	0
Carter	Gore	0.25	3
Gore	Grand	0.3	3

The intersection groups were determined based on the distances between intersections, number of access points, minor roads, and data availability. Each group consists three consecutive intersections which includes an upstream intersection, a downstream intersection, and a middle intersection at which turning movement volumes are estimated. A total of 19 intersections were divided into four groups on US 17/92 and seven groups on US 441. It should be noted that an intersection could be included into multiple groups as different intersection types. Those groups were utilized in developing turning movement estimation algorithms which will be discussed in the upcoming sections. Figure 13 and Figure **14** show different groups of intersections on US 17/92 and OBT Corridor US 441, respectively.

Figure 13: Groups of intersections on US 17/92

Figure 14: Groups of intersections on US 441

3.1.3 Data processing and aggregation

The downloaded GRIDSMART data included two main types of data sources: signal data and traffic volume data. The signal data is represented by 16 different digits that corresponds to different signal phases at an intersection. Further, each record represents a change in an event (change in any of the 16 digits). Those 16 digits show whether the signal phase is red (R), yellow (Y), or green (G). Moreover, the downloaded traffic volume data includes several files, and each file represents a lane at the intersection where virtual detector was implemented in the system. Thus, one file includes the data per lane per day. The traffic raw data contains many records, and each record represents one vehicle that arrives at the intersection from an approach. Figure 15 shows three main data sources that were utilized to achieve the project aim: traffic volume, signal data, and general features.

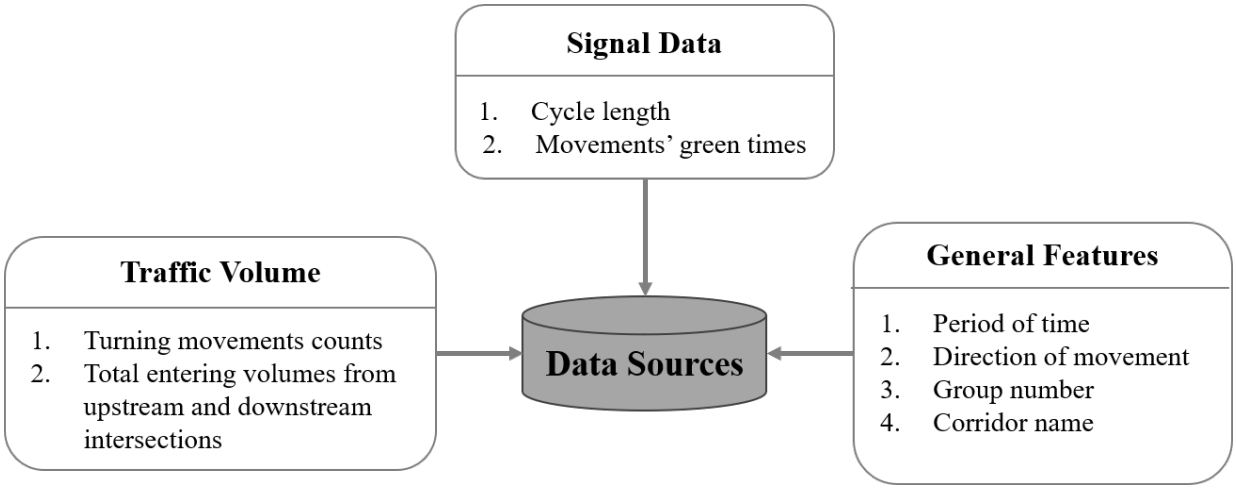

Figure 15: Data sources

Specifically, traffic volume data from different intersections could provide data such as counts of turning movement and total entering volumes calculated from upstream and downstream intersections movement counts. Besides, cycle length and movements green time could be calculated from the signal data sources. Finally, general features such as time periods (AM peak, off peak, PM peak or nighttime), directions of movements, group number, and corridor name were

also considered while developing the estimation algorithms.

The collected data was preprocessed by two steps. First, the travel direction was identified. Second, the clockwise method was utilized to rename the other approaches based on the travel direction in a nomenclature as shown in Figure 16 (Yuan and Abdel-aty, 2018). For example, if the travel direction is towards north, the northbound approach was labeled with number “1” and other approaches were labelled as well in the clockwise direction (eastbound: “2”, southbound: “3”, westbound: “4”). Similarly, when the travel direction is southbound, it was labelled with number “1” and other approaches were labelled clock wisely as well (westbound: “2”, northbound: “3”, eastbound: “4”). This process was to generalize the movement estimation at the corridor level.

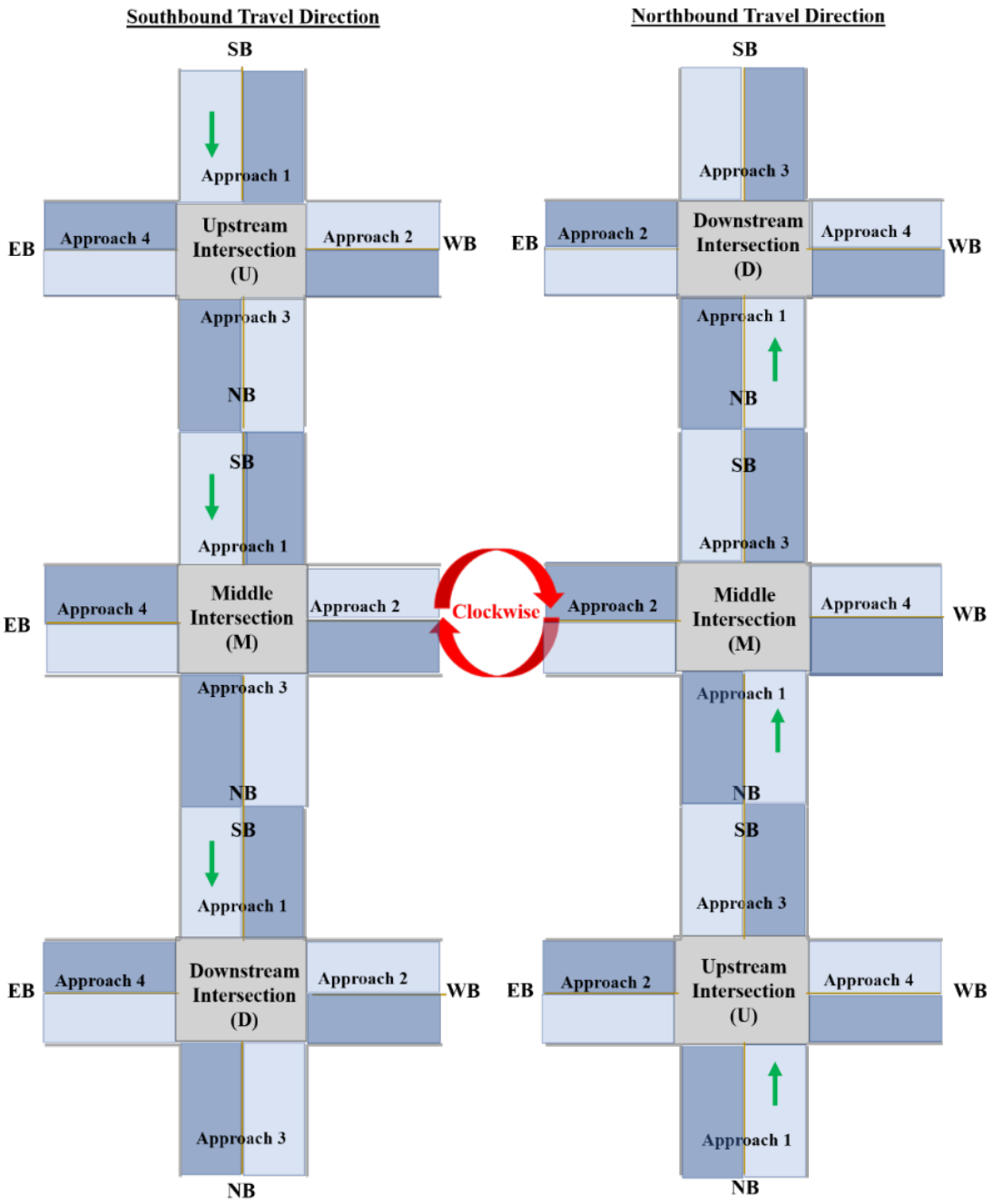

Figure 16: The nomenclature of the four approaches

For developing models, the raw data was processed and aggregated in the cycle level. The data processing and aggregation procedures were shown in Figure 17. First, for the raw signal data the coordination phase at intersection was determined and movements phases were identified. Thus, the start and end of cycle timestamps could be calculated based on the coordination phase of the intersection. Afterwards, the intersection's cycle lengths along the whole month as well as

the green time for each movement per approach per cycle were calculated. Second, the approach direction (Northbound, Southbound, Eastbound, or Westbound) and movements types (through, left, right, U-turn) of the raw traffic counts data were determined for each record. Then, the downloaded data by each day were combined in one file.

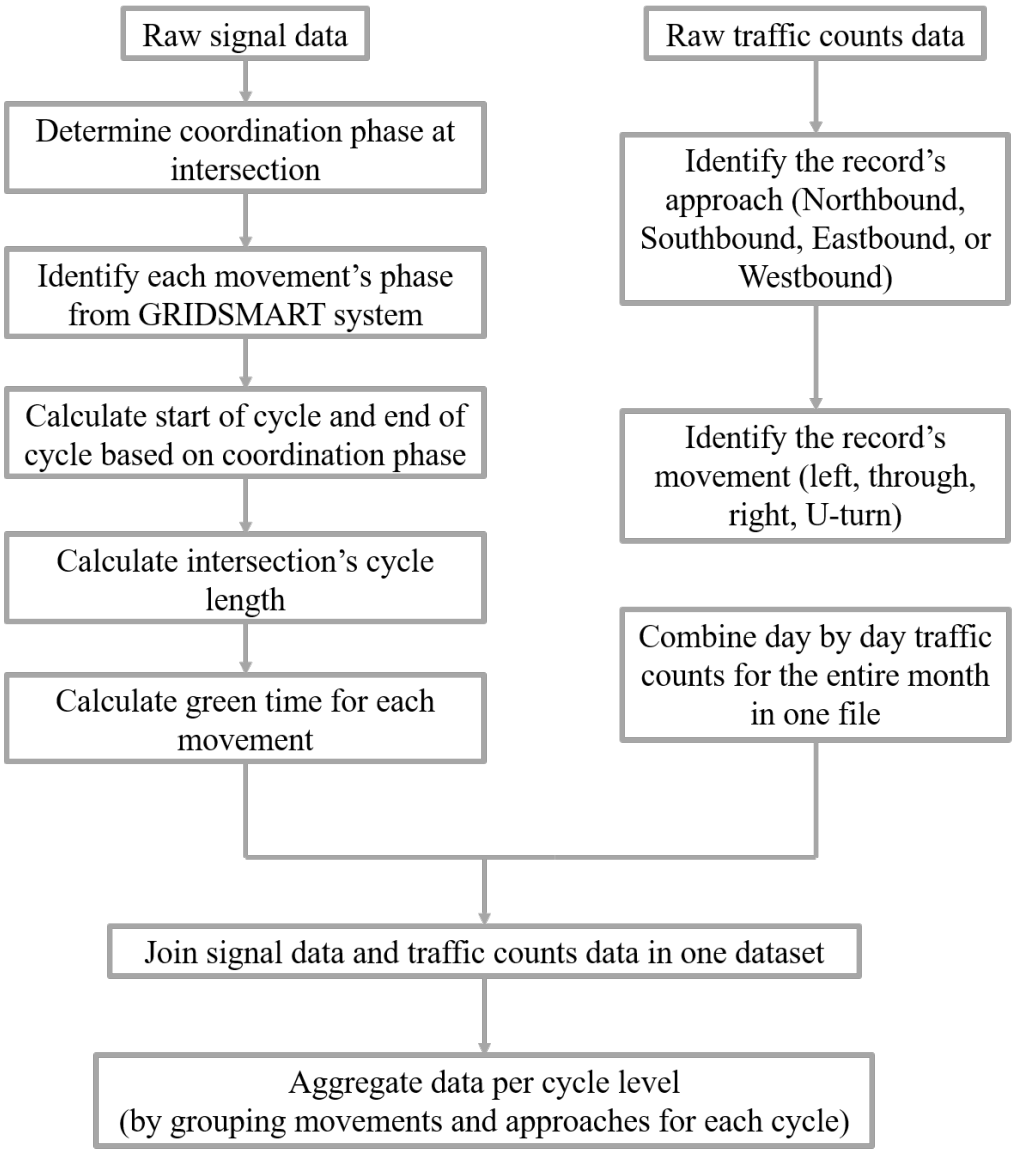

Figure 17: Data processing and aggregation

The two processed datasets for both signal and traffic counts data were merged into one big dataset. This dataset was aggregated by grouping the counts of each movement per cycle. As a result, a final dataset for each intersection including turning movements counts, cycle lengths,

and green times for each movement was prepared. The data of both Northbound and Southbound directions were combined into one dataset.

Individual dataset for each group defined in Figure 13 and Figure 14 was constructed using its intersections data. Each dataset included the signal data and turning movements aggregated counts of upstream and downstream intersections. It also included signal data for the middle intersection (the target intersection at which turning movements were estimated). Further, weekends (Saturdays and Sundays) were removed from the dataset as their traffic fluctuation differs from weekdays. The three intersections were merged together based on the previously defined cycle lengths for the target (middle) intersection. Moreover, two signal cycles were merged together (current cycle and previous cycle) for developing the movement estimation algorithms. Figure 18 shows the procedure of defining cycles in a certain group of intersections.

Finally, general features such as time period (am peak, off peak, pm peak or nighttime), direction of movement, group number, and corridor name were included in the datasets in order to identify the records in the general models. Besides, the total entering volume for Northbound and Southbound approaches were calculated per cycle using the turning movements counts for the upstream and downstream intersections.

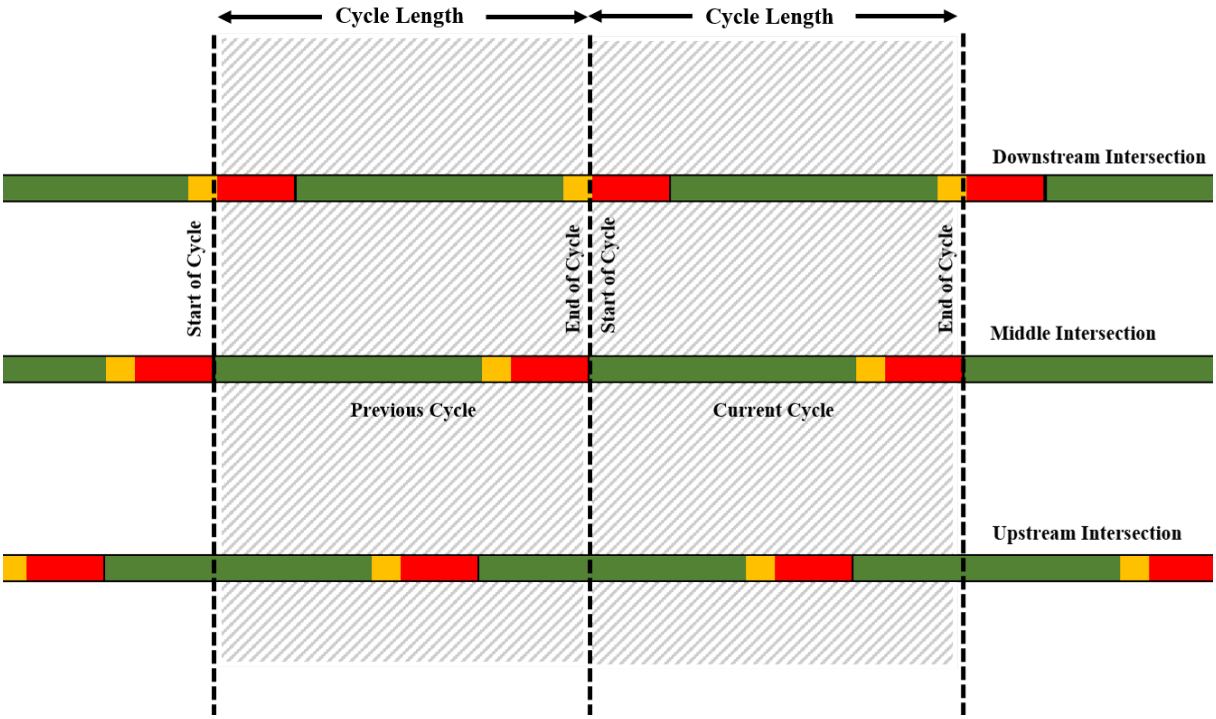

Figure 18: Defining signal cycles in one group

To develop the generic models, the data of groups were combined for the US 17/92 and US 441 corridors, respectively. Besides, all data of the two corridors were further combined to develop generic models based on all data. Figure 19 shows a flowchart that summarizes arrangements of different datasets. Finally, totally 14 datasets were prepared and compiled to develop generic turning movement estimation algorithms as well as individual algorithms for each specific corridor and group. Each dataset includes 110 features (traffic counts, green times, cycle length, and general features), and the description of those features were summarized in Table 7. Further,

Table 8 shows the final datasets and their corresponding number of records.

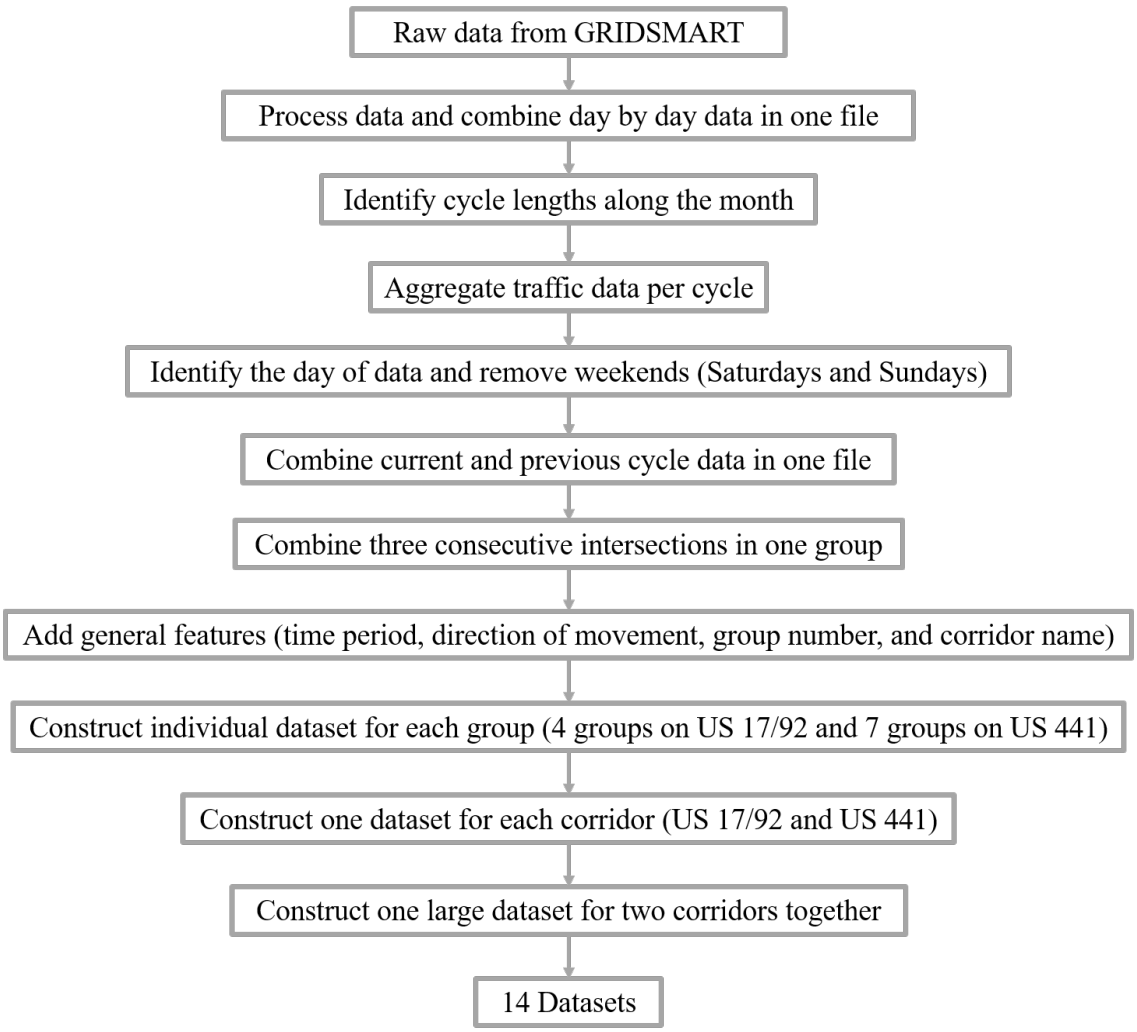

Figure 19: Summary of datasets arrangements

Table 7: Features labelling and description

Variable	Description
CycleLengthM	Cycle length at middle intersection
1LGM	Left turn green time at middle intersection (direction of movement)
1SGM	Through movement green time at middle intersection (direction of movement)
3SGM	Through movement green time at middle intersection (opposed to direction of movement)
2SGM	Through movement green time at middle intersection (East-West direction)
CycleLengthD	Cycle length at downstream intersection
1LGD	Left turn green time at downstream intersection (direction of movement)
1SGD	Through movement green time at downstream intersection (direction of movement)
3LGD	Through movement green time at downstream intersection (opposed to direction of movement)
3SGD	Through movement green time at downstream intersection (East-West direction)
2LGD	Left turn green time at downstream intersection (East-West direction)
2SGD	Through movement green time at downstream intersection (East-West direction)
4LGD	Left turn green time at downstream intersection (West-East direction)
4SGD	Through movement green time at downstream intersection (West-East direction)
1LD	Left turn volume at downstream intersection (direction of movement)
1RD	Right turn volume at downstream intersection (direction of movement)
1SD	Through movement volume at downstream intersection (direction of movement)
3LD	Left turn volume at downstream intersection (opposed to direction of movement)
3RD	Right turn volume at downstream intersection (opposed to direction of movement)
3SD	Through movement volume at downstream intersection (opposed to direction of movement)

Variable	Description
2LD	Left turn volume at downstream intersection (East-West direction)
2RD	Right turn volume at downstream intersection (East-West direction)
2SD	Through movement volume at downstream intersection (East-West direction)
4LD	Left turn volume at downstream intersection (West-East direction)
4RD	Right turn volume at downstream intersection (West-East direction)
4SD	Through movement volume at downstream intersection (West-East direction)
CycleLengthU	Cycle length at Upstream intersection
1LGU	Left turn green time at upstream intersection (direction of movement)
1SGU	Through movement green time at upstream intersection (direction of movement)
3LGU	Through movement green time at upstream intersection (opposed to direction of movement)
3SGU	Through movement green time at upstream intersection (East-West direction)
2LGU	Left turn green time at upstream intersection (East-West direction)
2SGU	Through movement green time at upstream intersection (East-West direction)
4LGU	Left turn green time at upstream intersection (West-East direction)
4SGU	Through movement green time at upstream intersection (West-East direction)
1LU	Left turn volume at upstream intersection (direction of movement)
1RU	Right turn volume at upstream intersection (direction of movement)
1SU	Through movement volume at upstream intersection (direction of movement)
3LU	Left turn volume at upstream intersection (opposed to direction of movement)
3RU	Right turn volume at upstream intersection (opposed to direction of movement)
3SU	Through movement volume at upstream intersection (opposed to direction of movement)
2LU	Left turn volume at upstream intersection (East-West direction)
2RU	Right turn volume at upstream intersection (East-West direction)
2SU	Through movement volume at upstream intersection (East-West direction)
4LU	Left turn volume at upstream intersection (West-East direction)
4RU	Right turn volume at upstream intersection (West-East direction)
4SU	Through movement volume at upstream intersection (West-East direction)
Period	AM peak, PM peak, Off peak, Night-time
1Ent	Total volume entering intersection (direction of movement)
3Ent	Total volume entering intersection (opposed to direction of movement)
1XCycleLengthM	Cycle length at middle intersection for previous cycle for previous cycle
1X1LGM	Left turn green time at middle intersection (direction of movement) for previous cycle
1X1SGM	Through movement green time at middle intersection (direction of movement) for previous cycle
1X3SGM	Through movement green time at middle intersection (opposed to direction of movement) for previous cycle
1X2SGM	Through movement green time at middle intersection (East-West direction) for previous cycle
1XCycleLengthD	Cycle length at downstream intersection for previous cycle
1X1LGD	Left turn green time at downstream intersection (direction of movement) for previous cycle
1X1SGD	Through movement green time at downstream intersection (direction of movement) for previous cycle
1X3LGD	Through movement green time at downstream intersection (opposed to direction of movement) for previous cycle
1X3SGD	Through movement green time at downstream intersection (East-West direction) for previous cycle
1X2LGD	Left turn green time at downstream intersection (East-West direction) for previous cycle
1X2SGD	Through movement green time at downstream intersection (East-West direction) for previous cycle
1X4LGD	Left turn green time at downstream intersection (West-East direction) for previous cycle
1X4SGD	Through movement green time at downstream intersection (West-East direction) for previous cycle
1X1LD	Left turn volume at downstream intersection (direction of movement) for previous cycle
1X1RD	Right turn volume at downstream intersection (direction of movement) for previous cycle
1X1SD	Through movement volume at downstream intersection (direction of movement) for previous cycle
1X3LD	Left turn volume at downstream intersection (opposed to direction of movement) for previous cycle
1X3RD	Right turn volume at downstream intersection (opposed to direction of movement) for previous cycle
1X3SD	Through movement volume at downstream intersection (opposed to direction of movement) for previous cycle
1X2LD	Left turn volume at downstream intersection (East-West direction) for previous cycle
1X2RD	Right turn volume at downstream intersection (East-West direction) for previous cycle
1X2SD	Through movement volume at downstream intersection (East-West direction) for previous cycle
1X4LD	Left turn volume at downstream intersection (West-East direction) for previous cycle
1X4RD	Right turn volume at downstream intersection (West-East direction) for previous cycle
1X4SD	Through movement volume at downstream intersection (West-East direction) for previous cycle
1XCycleLengthU	Cycle length at Upstream intersection for previous cycle
1X1LGU	Left turn green time at upstream intersection (direction of movement) for previous cycle
1X1SGU	Through movement green time at upstream intersection (direction of movement) for previous cycle
1X3LGU	Through movement green time at upstream intersection (opposed to direction of movement) for previous cycle
1X3SGU	Through movement green time at upstream intersection (East-West direction) for previous cycle
1X2LGU	Left turn green time at upstream intersection (East-West direction) for previous cycle
1X2SGU	Through movement green time at upstream intersection (East-West direction) for previous cycle
1X4LGU	Left turn green time at upstream intersection (West-East direction) for previous cycle
1X4SGU	Through movement green time at upstream intersection (West-East direction) for previous cycle
1X1LU	Left turn volume at upstream intersection (direction of movement) for previous cycle

Variable	Description
1X1RU	Right turn volume at upstream intersection (direction of movement) for previous cycle
1X1SU	Through movement volume at upstream intersection (direction of movement) for previous cycle
1X3LU	Left turn volume at upstream intersection (opposed to direction of movement) for previous cycle
1X3RU	Right turn volume at upstream intersection (opposed to direction of movement) for previous cycle
1X3SU	Through movement volume at upstream intersection (opposed to direction of movement) for previous cycle
1X2LU	Left turn volume at upstream intersection (East-West direction) for previous cycle
1X2RU	Right turn volume at upstream intersection (East-West direction) for previous cycle
1X2SU	Through movement volume at upstream intersection (East-West direction) for previous cycle
1X4LU	Left turn volume at upstream intersection (West-East direction) for previous cycle
1X4RU	Right turn volume at upstream intersection (West-East direction) for previous cycle
1X4SU	Through movement volume at upstream intersection (West-East direction) for previous cycle
1X1Ent	Total volume entering intersection (direction of movement) for previous cycle
1X3Ent	Total volume entering intersection (opposed to direction of movement) for previous cycle
ID	Group ID
Direction	Direction of Estimation
Corridor	Corridor Name
1SM	Left turn volume at middle intersection (direction of movement)
3SM	Right turn volume at middle intersection (direction of movement)
2SM	Through movement volume at middle intersection (direction of movement)
4SM	Left turn volume at middle intersection (opposed to direction of movement)
1LM	Right turn volume at middle intersection (opposed to direction of movement)
3LM	Through movement volume at middle intersection (opposed to direction of movement)
2LM	Left turn volume at middle intersection (East-West direction)
4LM	Right turn volume at middle intersection (East-West direction)
1RM	Through movement volume at middle intersection (East-West direction)
3RM	Left turn volume at middle intersection (West-East direction)
2RM	Right turn volume at middle intersection (West-East direction)
4RM	Through movement volume at middle intersection (West-East direction)

Table 8: Final datasets and their corresponding number of records

Dataset	Group	Number of Records
Two Corridors	--	318,899
Mills Corridor	--	116,036
Mills Corridor	Grp1	34,670
	Grp2	25,572
	Grp3	35,612
	Grp4	20,182
OBT Corridor	--	202,863
OBT Corridor	Grp1	20,858
	Grp2	33,206
	Grp3	30,574
	Grp4	22,854
	Grp5	32,872
	Grp6	31,556
	Grp7	30,943

3.2 Data Validation

It was essential to understand GRIDSMART vehicle counts' accuracy before using its data in developing turning movement estimation algorithm(s). Thus, GRIDSMART accuracy was first

validated. Manual turning movements counts per approach per intersection were carried out over a given time period. Afterwards, the manual counts were compared with the corresponding GRIDSMART system data. Consequently, three intersections along US 17/92 (intersections at Nebraska St., Princeton St, and Rollins St.) were chosen to validate GRIDSMART data counts. The intersections were chosen as the distance between them were relatively short and there are only few other access points between intersections. The validation was carried out twice, the first time was on March 8th, 2019 before starting the model development and the second time was on December 23rd, 2019. The two validation efforts were conducted for different traffic conditions: the first one reflected the traffic during a regular weekday while the second one indicated the traffic during holiday conditions.

(1) First validation

To conduct the first validation, manual counts were carried out for 30 minutes during both AM and PM peak periods. Videos were recorded at the same time. Table 9 shows the AM and PM peak times in which the vehicles were manually counted at each intersection. Moreover,

Table 10 and **Error! Reference source not found.** show the average five minutes turning movements for both Manual counts and the GRIDSMART system for the selected intersections.

Table 9: AM and PM counts intervals

Intersection Name	AM Peak Counts Interval	PM Peak Counts Interval
US 17/92 and Nebraska St.	07:45 to 08:15	04:45 to 05:15
US 17/92, and Princeton St.	08:15 to 08:45	05:15 to 05:45
US 17/92 and Rollins St.	08:45 to 09:15	05:45 to 06:15

Table 10: Turning Movements' Counts - AM peak

Movement	Mills and Nebraska		Mills and Princeton		Mills and Rollins	
	Manual	GRIDSMART	Manual	GRIDSMART	Manual	GRIDSMART
NB-Through	104	107	93	91	69	81
NB-Right	1	1	0	0	6	1
NB-Left	2	2	43	44	2	9
NB-Uturn	0	0	0	0	0	0
SB-Through	78	84	76	73	82	66
SB-Right	3	4	7	7	0	6
SB-Left	16	18	0	6	8	2
SB-Uturn	0	0	0	0	0	0
EB-Through	2	2	0	0	3	3
EB-Right	1	1	45	31	2	4
EB-Left	1	1	7	7	1	3
EB-Uturn	0	0	0	0	0	0
WB-Through	2	2	0	0	2	3
WB-Right	28	21	0	0	5	0
WB-Left	1	1	0	0	2	1
WB-Uturn	0	0	0	0	0	0

Table 11: Turning Movements' Counts - PM peak

Movement	Mills and Nebraska		Mills and Princeton		Mills and Rollins	
	Manual	GRIDSMART	Manual	GRIDSMART	Manual	GRIDSMART
NB-Through	77	76	78	79	99	76
NB-Right	2	2	0	0	6	2
NB-Left	5	7	34	33	3	10
NB-Uturn	0	0	0	0	0	0
SB-Through	117	111	95	95	77	97
SB-Right	10	9	4	5	2	6
SB-Left	25	28	0	9	9	3
SB-Uturn	0	0	0	0	0	0
EB-Through	4	4	0	0	2	2
EB-Right	3	3	59	53	1	6
EB-Left	4	4	9	9	1	5
EB-Uturn	0	0	0	0	0	0
WB-Through	3	2	0	0	2	2
WB-Right	28	27	0	0	8	1
WB-Left	3	3	0	0	4	1
WB-Uturn	0	0	0	0	0	0

Manual counts comprise vehicles' turning movements in five minutes intervals. Mean Absolute Error (MAE) between manual counts data and GRIDSMART system data was calculated. The following

Table **12** and Table 13 show the MAE value for all possible turning movements for the three intersections for both AM peak and PM peak respectively.

Table 12: Turning Movements' Mean Absolute Error - AM peak

Turning Movement	Mills and Nebraska	Mills and Princeton	Mills and Rollins
NB-Left	0.67	4.83	7.00
NB-Right	0.00	--	5.17
NB-Through	10.33	4.83	13.00
NB-Uturn	0.00	0.00	0.00
SB-Left	2.50	--	6.50
SB-Right	1.33	0.33	4.83
SB-Through	8.17	2.67	15.83
SB-Uturn	0.00	0.00	0.00
EB-Left	0.33	0.67	2.00
EB-Right	0.00	13.33	2.33
EB-Through	0.50	--	1.50
EB-Uturn	0.00	0.00	0.00
WB-Left	0.00	--	1.33
WB-Right	10.00	--	4.00
WB-Through	1.00	--	1.17
WB-Uturn	0.00	--	0.00

Table 13: Turning Movements' Mean Absolute Error - PM peak

Turning Movement	Mills and Nebraska	Mills and Princeton	Mills and Rollins
NB-Left	2.00	1.33	7.00
NB-Right	0.00	--	3.67
NB-Through	2.17	1.17	23.00
NB-Uturn	0.00	0.00	0.00
SB-Left	4.00	--	6.67
SB-Right	0.83	0.17	3.50
SB-Through	5.83	0.83	21.67
SB-Uturn	0.00	0.00	0.00
EB-Left	0.50	0.50	3.67
EB-Right	0.33	7.00	5.33
EB-Through	0.17	--	0.67
EB-Uturn	0.00	0.00	0.00
WB-Left	0.33	--	2.83
WB-Right	0.83	--	6.67
WB-Through	0.83	--	0.83
WB-Uturn	0.00	--	0.00

The results showed that for Intersection US 17/92 & Nebraska St. are relatively low. However, for the other two intersections (US 17/92 & Princeton St. and US 17/92 & Nebraska St.), some turning movements counts have high MAE (e.g., NB-Left, SB-Through, and EB-Right).

For more interpretations of MAE values, the relation between the average manually counted turning movements in five minutes and the corresponding MAE values were plotted for both AM and PM peaks in Figure 20. The figures show that most of the high MAE values occurred when the volumes were too small. However, MAE values decreased in case of high five minutes volumes. The results suggested that for the movements with high volume, the GRIDSMART system could provide acceptable detection accuracy. However, for some movements with low traffic volume, large errors might be found.

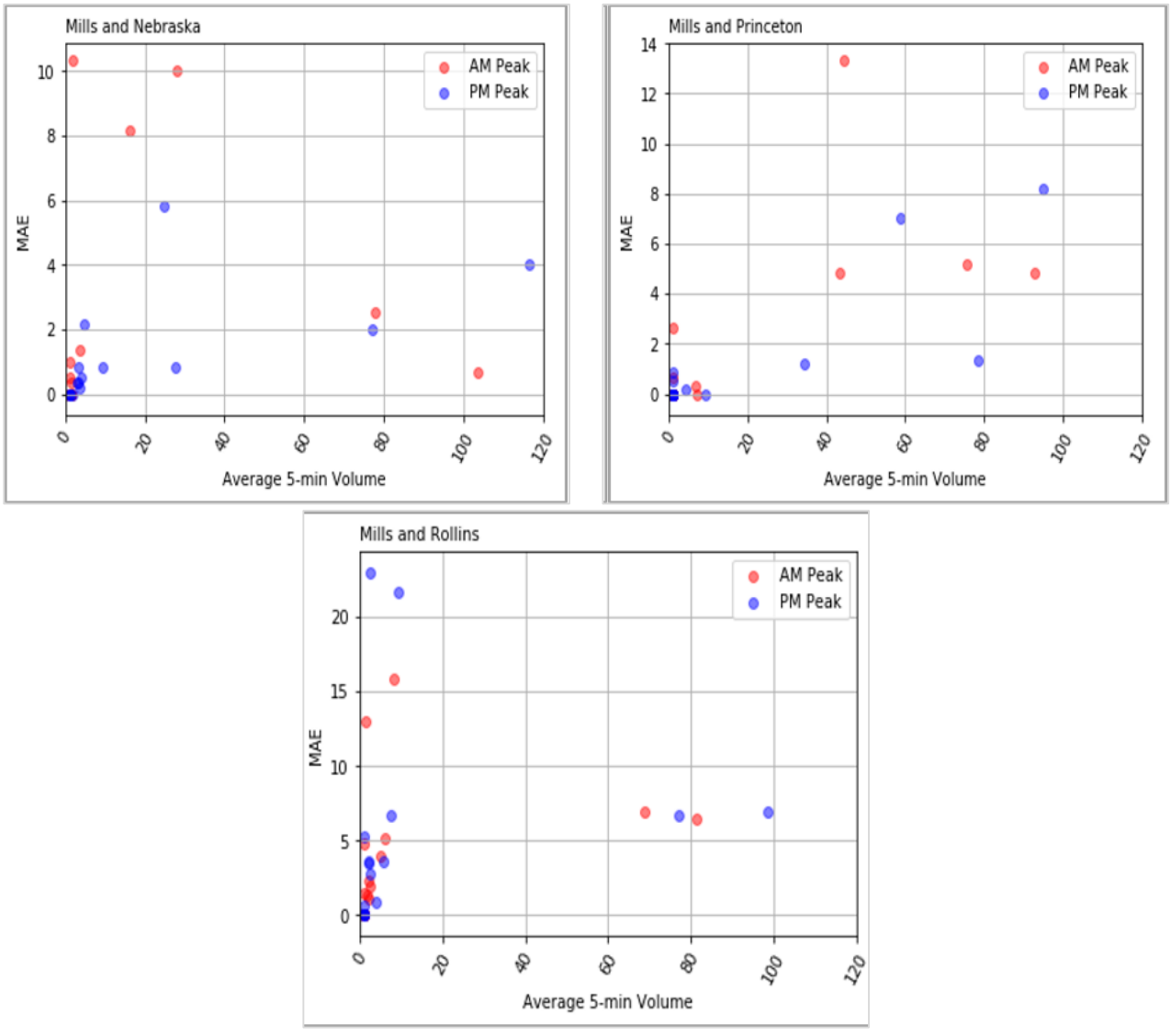

Figure 20: MAE for average five minutes volumes

(2) Second validation

Another validation was carried out by comparing turning movement counts manually per approach per intersection for two hours in the AM peak period as the ground truth data with the corresponding GRIDSMART data. Four intersections along the US 17/92 corridor (intersections at Nebraska St., Princeton St, Rollins St., and Virginia St.) were chosen for this validation. Three intersections were common in the first and the second validation: Nebraska St., Princeton St, and Rollins St. The intersections were chosen as the distance between them were relatively short and there are only few other access points between intersections. The videos were recorded on 23rd of December from 7:00 to 9:00 AM.

GRIDSMART data for the same two-hour period was downloaded, processed, and aggregated at the cycle level. Then, turning movements were counted for each cycle using the same defined start and end of cycles timestamps for the aggregated data. Cycle level turning movement counts for Northbound and Southbound were plotted in

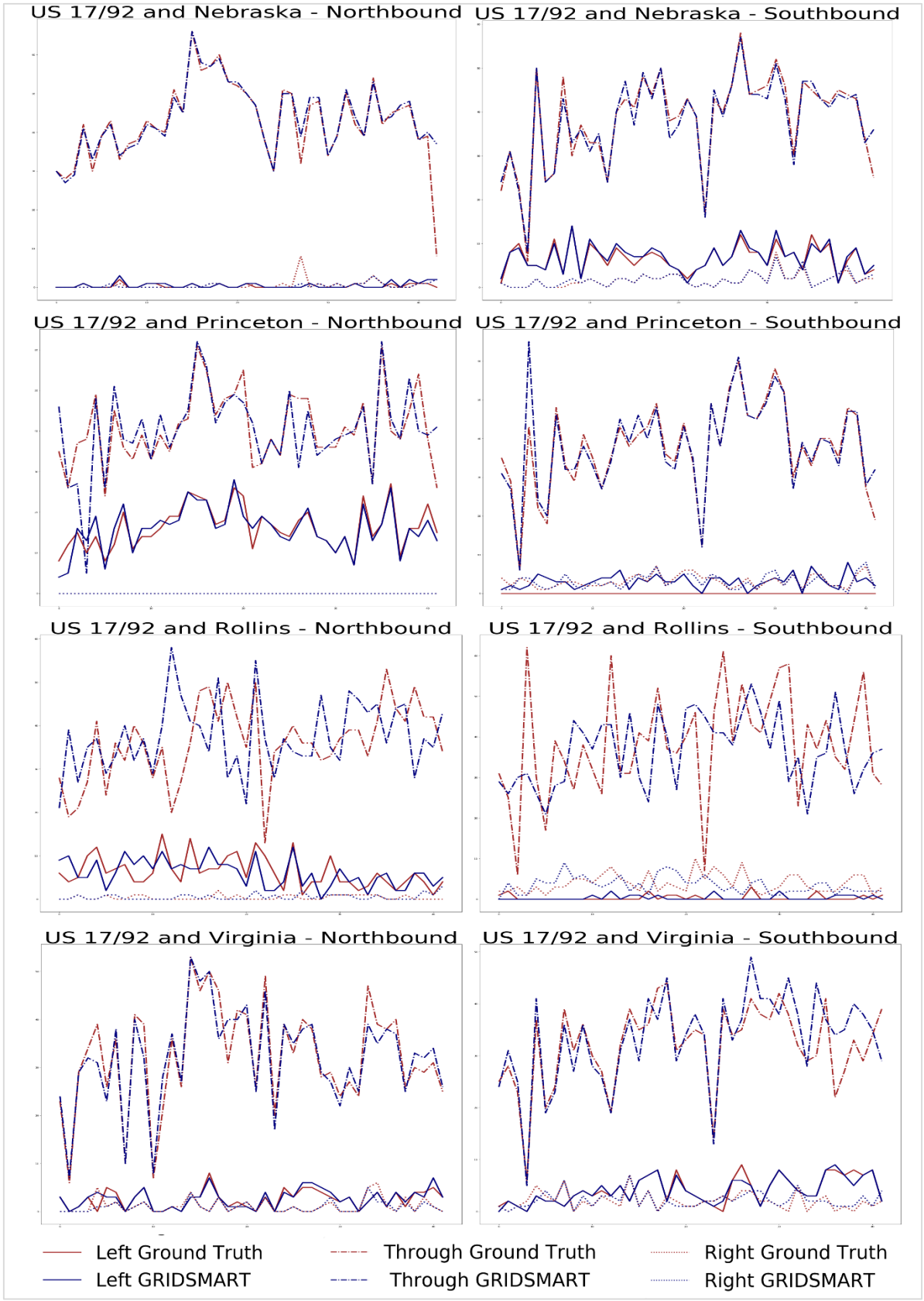
US 17/92 and Nebraska - Northbound
US 17/92 and Nebraska - Southbound
US 17/92 and Princeton - Northbound
US 17/92 and Princeton - Southbound
US 17/92 and Rollins - Northbound
US 17/92 and Rollins - Southbound
US 17/92 and Virginia - Northbound
US 17/92 and Virginia - Southbound
Left Ground Truth
Left GRIDSMART
Through Ground Truth
Through GRIDSMART
Right Ground Truth
Right GRIDSMART

Figure **21** for both the ground truth and GRIDSMART data. It shows that GRIDSMART could provide reasonable counts for most intersections. However, there existed several significant errors for several movements. For example, significant difference could be observed between the ground truth data and the GRIDSMART data for the northbound and southbound movements at the intersection US 17/92 and Rollins street.

Finally, in order to evaluate and validate the GRIDSMART data, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) for each movement were calculated. Table 14 shows the comparison results. The table shows that some intersections have relatively high MAEs for certain movements such as the northbound through movement for the intersection with Rollins St. and southbound through movements for the intersections at Rollins St. and Virginia St. Moreover, the MAPE values were relatively high for right and left turns movements since the volumes are so low (i.e., lower than 10 vehicles per cycle) and MAPE is more sensitive to the error. Finally, the MAPE values for Rollins intersection were found to be relatively high in all movements. MAPEs for other intersections were around 10%.

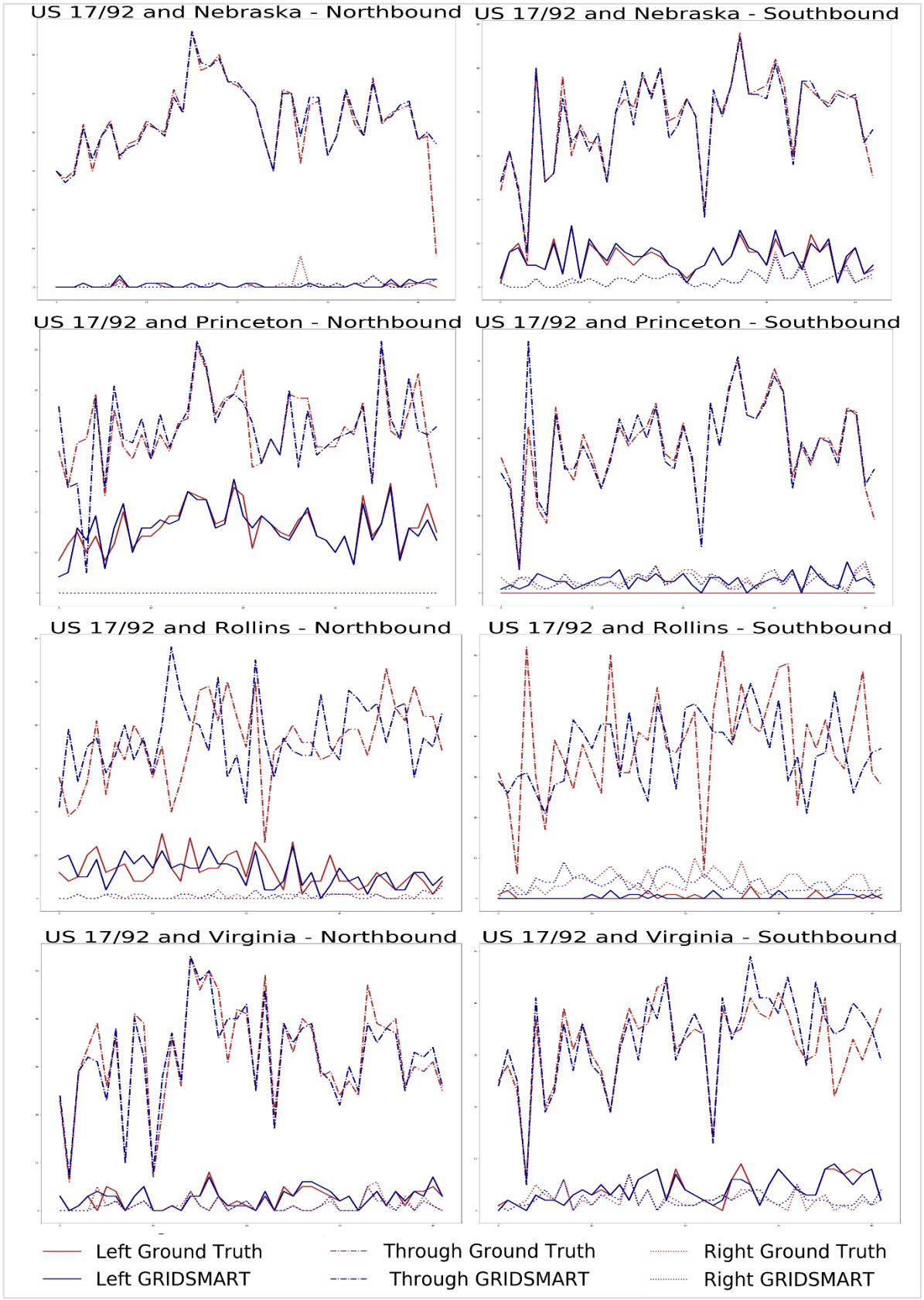

Figure 21: Cycle level turning movements volume fluctuation on US 17/92 corridor

Table 14: Performance measures of GRIDSMART data based on the ground truth data

Mean Absolute Error (MAE)				
Turning Movement	Nebraska	Princeton	Rollins	Virginia
NB - Through	1.56	4.21	9.02	2.55
NB - Left	0.19	1.67	2.76	0.45
NB – Right	0.30	--	0.62	0.31
SB – Through	1.56	2.17	11.05	3.67
SB – Left	0.49	--	0.69	0.29
SB – Right	0.14	0.76	2.57	0.55
EB – Through	0.19	--	0.62	0.64
EB – Left	0.19	0.50	1.71	0.43
EB – Right	0.30	3.19	1.95	0.64
WB – Through	0.12	--	1.12	1.02
WB – Left	0.09	--	0.81	0.62
WB - Right	1.21	--	0.50	1.31
Mean Absolute Percentage Error (MAPE)				
Turning Movement	Nebraska	Princeton	Rollins	Virginia
NB - Through	10.67%	10.89%	30.46%	8.32%
NB - Left	26.92%	11.98%	57.35%	13.46%
NB – Right	20.54%	--	60.00%	21.06%
SB – Through	5.12%	6.61%	46.35%	11.79%
SB – Left	9.27%	--	100.00%	5.52%
SB – Right	6.29%	26.17%	80.23%	21.00%
EB – Through	11.76%	--	72.22%	9.70%
EB – Left	11.67%	17.07%	79.43%	29.29%
EB – Right	23.33%	21.99%	63.72%	20.72%
WB – Through	4.76%	--	66.67%	10.86%
WB – Left	5.00%	--	77.78%	23.80%
WB - Right	14.03%	--	100.00%	14.93%

The MAEs on AM peak periods in the two validation studies are shown in

Figure **22.** The same color indicates the data from the same intersection while the solid and dotted lines were the data in different dates. Generally, the validation results were consistent for both March19th and December 23rd Validations on AM peak period. However, the MAE values for the latter were lower as the data were aggregated in cycle level. Besides, it was concluded that some intersections have relatively high MAEs for certain movements such as the northbound and southbound through movements for the intersection with Rollins St. Further, in case of low right and left turn movements

(i.e., lower than 10 vehicles per cycle), the MAE values were found to be relatively high.

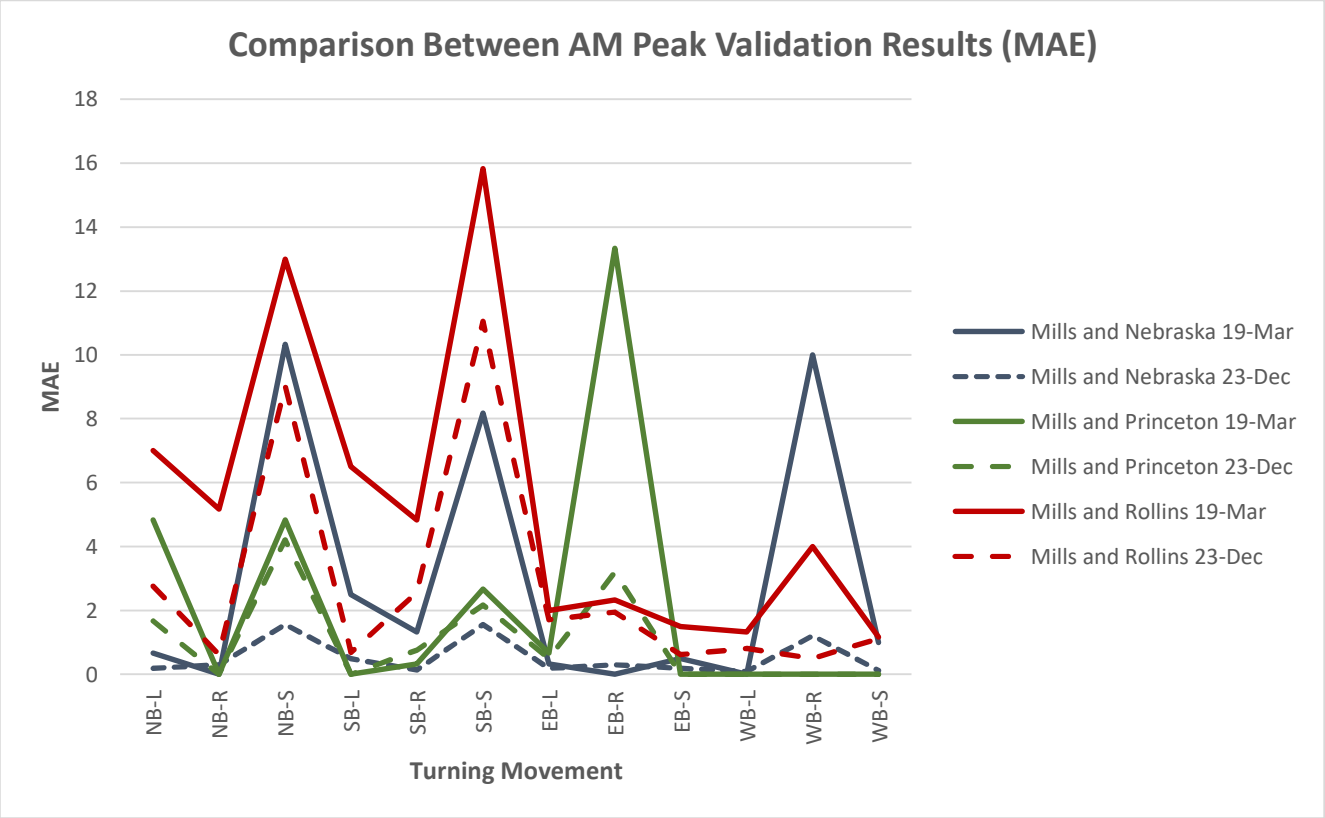

Figure 22: Comparison between AM peak validation results

3.3 Methodology

Various traffic prediction approaches were developed and introduced in the previous literature to predict future traffic parameters (e.g., traffic flow, speed, and travel time). The selection of the most appropriate methodology in estimating traffic parameters is important. Different approaches have been widely employed in previous short-term traffic forecasting studies (parametric models, machine learning models, and deep learning models). Those approaches were followed to develop a comprehensive turning movement estimation model using the processed dataset for the US 17/92 corridor. The models were developed and compared based on the data of AM peak periods (7:00 to 9:00 am). The following models were applied, tuned, and tested:

1. Negative binomial model

2. Finite mixture model
3. Multivariate adaptive regression spline
4. Neural networks
5. Random forest
6. Gradient boosting

The models were developed to estimate corridor-level (North-South direction) through movements. They were trained to estimate one traffic movement at a time. For the East-West direction, it's hard to estimate turning movements as there is no data at the upstream and downstream intersections could be used for the model. Once detectors are installed at the east-west intersections, the estimation models could be extended to the east and west approaches. Furthermore, each dataset was split into the training dataset and test dataset with a ratio of 4:1.

3.3.1 Negative binomial model

Negative Binomial (NB) has been widely used to estimate the over-dispersed count data (Daraghmi et al., 2012). Overdispersion occurred when the mean and the variance are not equal. Hence, the mean and the variance were calculated for through and left turn movements. For through movement, the calculated mean and variance were 22 and 466, respectively. Further, for left turn movements, the mean and variance were 2.2 and 15. The generalized negative binomial equation (4) can be written as:

$$\ln(Y) = \alpha + \beta i X i + \in \quad (1)$$

Where Y is the response variable, α is the intercept, X_i is a predictor (i = 1, 2, 3, … ,n), n is the number of predictors, β is the regression coefficient, and ∈ is the regression error vector.

3.3.2 Finite mixture model

The finite mixture model is considered as a highly flexible approach that has been widely used in

considerable applications. Usually, the finite mixture model deals with the stagnant stochastic processes as it considers the mixed components as Gaussian distributions. Nevertheless, it can deal with nonstationary processes by applying the finite mixture model on several time intervals. The generalized equation (5) of the finite mixture model is as follow:

$$f(y_i) = \sum_{k=1}^{K} \pi_k f_k(y_i) \quad (2)$$

Where π_k is the mixing probability of component k, $0 \leq \pi_k \leq 1$, and $\sum_{k=1}^{K} \pi_k = 1$ for k = 1, 2, 3, … K. Moreover, $f_k(y_i)$ could follow any probability distribution (Chen et al., 2014).

3.3.3 Multivariate adaptive regression spline

Multivariate Adaptive Regression Splines (MARS) is a nonparametric regression approach that was widely adopted in various data analyses including traffic flow prediction (Ermagun and Levinson, 2018). It could model non-linearities as well as high-dimensional predictors' and responses interactions (Friedman, 1991). The MARS model aims to build a regression function using a combination of basis functions. The sum of the basis functions represents the regression function (Xu et al., 2013). Generally, the MARS model is shown in equation (6):

$$Y = f(X) + \in = \beta_o + \sum_{j=1}^{r} \beta_m h_m(X) + \in \quad (3)$$

Where Y is the response variable, X is a predictor, β_o is the intercept, each $h_m(X)$ is a basis function or a product of two or more such functions, β_m are the coefficients which usually estimated by minimizing the sum of squares error and $\in$ is the regression error.

3.3.4 Neural networks

Neural Network (NN) is a non-parametric flexible approach that could model non-linearities and

complex relations between predictors and response variables (Ghanim and Shaaban, 2018). A three layered NN that contains two hidden layers and one output layer was utilized to estimate the traffic movement counts at the middle intersection. Input vectors were processed by the input layer, and then weights and biases were applied by the hidden layer to the signals received from the input layer. The NN model adjusts weights and biases in order to improve the performance of the model. Moreover, Rectified Linear Unit (ReLU) was utilized as an activation function. After several iterations, the first hidden layer was set to contain 100 neurons, while the second one contained 40 neurons. Finally, the output layer received the signal from the second hidden layer and transferred it into the outcome of the NN which represented the predicted movement. Figure 23 shows the structure of the utilized NN.

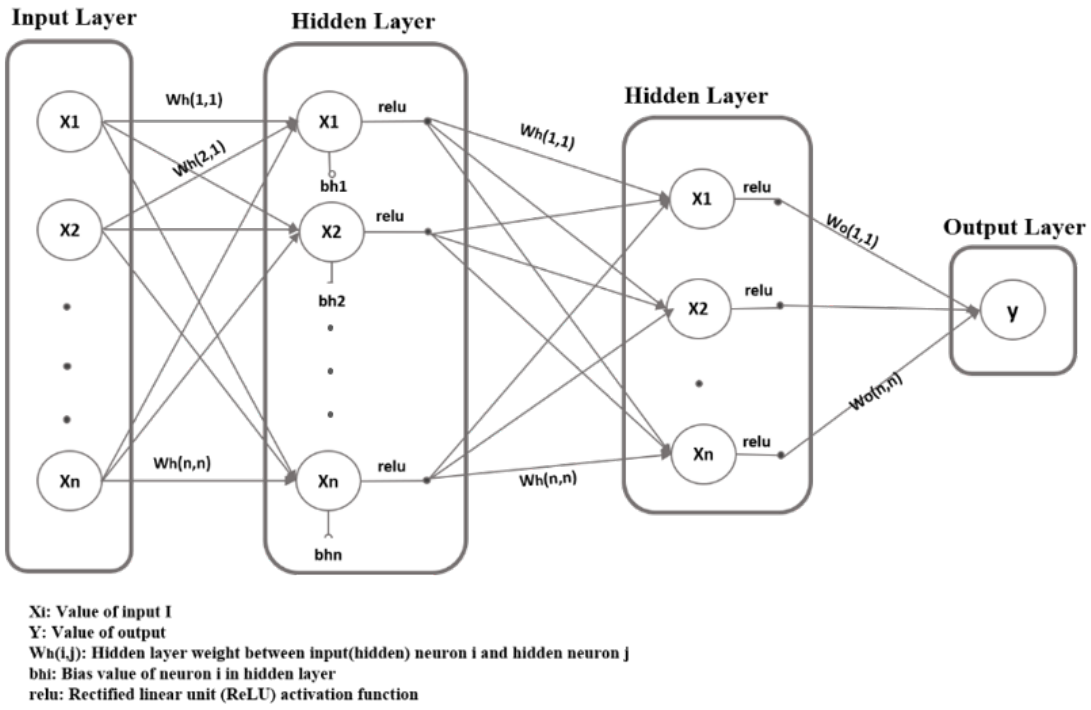

Figure 23: Artificial Neural Network's Layout Structure

3.3.5 Random forest

The Random Forest algorithm is a nonparametric approach based on decision trees that assumes a joint distribution. A n-dimensional random vector X = (X1, X2, … Xn)T represents the predictor

variables and Y represents the response variable (Zhang and Ma, n.d.). Random Forest assumes a joint distribution $P_{XY}(X,Y)$. In order to predict Y, a prediction function is determined by minimizing the expected values of the loss function $f(X)$. The utilized loss function in case of regression is commonly the squared error loss as shown in equation (7).

$$L\big(Y, f(X)\big) = (Y - f(X))^2 \tag{4}$$

The number of trees are tuned based on the performance measures in equations 5, 6, and 7(2)(2)(3)(2). The best measures corresponding to 60 decision trees were utilized in the developed model.

3.3.6 Gradient Boosting Decision Trees

Gradient Boosting is an ensemble machine learning approach based on sequentially trained decision trees. It fits the negative gradients (residual errors) in order to learn the decision trees (Ke et al., 2017). This approach emphasizes the incorrectly estimated training records in the model. Thus, it generates multiple models in sequence. Incorrect estimates in previous base models appear more frequently in the training data than the correctly estimated ones. Consequently, the main aim of a new base model is to correct the errors of the previous models. The main concept is to combine several weak models to create one high accurate model (Zhang and Haghani, 2015a).

When using Gradient Boosting Decision Trees (GBDT), it is considered as an optimization approach that adds a base model in each step to minimize a certain loss function (mean absolute error or the mean squared error loss functions). In boosting, the loss function measures the deviation of the predicted values from the corresponding ground truth. Shortly, the GBDT develops consecutive models. In each model, training data is resampled in order to provide the model with the most useful information. Then, weights are adjusted in each training step based on the produced error from the previous step. Thus, incorrectly estimated records have higher weights

in the consecutive model.

3.3.7 Performance Measures

Three performance measures were utilized to evaluate the developed models: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). Those measures are defined in Equations (5) (6), and (7).

$$MAE = \frac{1}{n}\sum|y_{actual} - y_{pred}| \quad (5)$$

$$RMSE = \sqrt[2]{\frac{1}{n}\sum|y_{actual} - y_{pred}|^2} \quad (6)$$

$$MAPE = \left(\frac{1}{n}\sum\left|\frac{y_{actual} - y_{pred}}{y_{actual}}\right|\right) * 100 \quad (7)$$

Where y_{actual} is the movement data from GRIDSMART, y_{pred} is the estimated movement's volume, and n is the number of observations. The performance measures results for testing data were relaxed by assuming the estimated movement's volumes correct if they are below minimum green time's volume for through and left turn movements for urban signalized intersection (Urbanik et al., 2015).

3.3.8 Model Selection

The models were compared using the previously defined performance measures in equations (5) (6), and (7). The compared performance measure was summarized to conclude the best modelling approach to be utilized in developing turning movement estimation algorithms for individual groups, each corridor, and for the whole study area. Table 15 shows performance measures for the developed models. It indicates that the GBDT model had the lowest values for all the performance measures among all the developed models. The MAE, RMSE, and MAPE for gradient boosting

were 4.4%, 6.0%, and 13.10%, respectively. Hence, the GBDT model was chosen to be applied to the rest of the processed datasets.

Table 15: Performance measures for the developed models

Through Movement			
Model	MAE	RMSE	MAPE
Negative Binomial	9.02	18.60	26.20%
Finite Mixture Model	5.24	6.90	14.60%
Multivariate Adaptive Regression Spline	4.66	6.08	14.21%
Neural Networks	18.27	22.75	34.90%
Random Forest	5.10	7.68	14.20%
Gradient Boosting Decision Trees	4.40	6.00	13.10%

3.4 Developing Turning Movement Estimation Models

In this section, the development of turning movements estimation models were illustrated. The developed models were utilized in estimating cycle-level through and left turn movement counts, which could be utilized as inputs to adjust traffic signal. As a result of the performance measures comparison in Table 15, the GBDT model was applied to the processed datasets. The processed datasets include a large dataset that combines two corridors (US 17/92 and US 441 corridors), individual datasets for each corridor, and individual datasets for each group of intersections. Different models were developed for different movements separately. In total, the GBDT model was applied to 14 datasets that were shown previously in Table 8.

First, a model that comprises all variables from each dataset. Model's hyperparameters (number of trees, interaction depth, and shrinkage) were tuned. Number of trees corresponds to the number of basis functions and iterations in the additive expansion. Interaction depth enumerates the maximum allowed variable interactions of each tree. Shrinkage describes the learning rate or the step-size reduction. Moreover, cross-validation was implemented in the GBDT model using tuned number of folds to avoid model's overfitting (Greenwell et al., 2019).

The GBDT outputs relative influence value for each variable that indicates its importance in the model. This value is computed based on Mean Squared Error (MSE) for each split in each

tree. Then, each variable's improvement is averaged across all the trees. A variable is considered most important when having largest average decrease in MSE. Variables with zero feature importance were excluded from the model. The remaining variables were utilized to train the model. Further, those variables were reduced gradually using the backward method. Variables were excluded from the dataset one by one. In each iteration the performance measures for training and testing datasets were calculated. If they remained constant or improved thus the variable should be removed from the dataset. Moreover, if they got worsened, the variable should be kept in the dataset. Subsequently, the final model was trained. Finally, marginal effects were calculated for all variables as they could capture the changes in the dependent variable corresponding to a unit change in the response variables (Cai et al., 2016).

Traffic volume fluctuations for peak periods (AM and PM peaks) varies from night-time and off-peak period. Thus, individual models were trained based on subsets of datasets that includes only AM and PM peak periods data. AM and PM periods were considered from 7:00 to 10:00 and from 15:00 to 18:00 respectively. The following subsections illustrates the developed generic models, corridor models, and group models for all processed datasets, respectively.

3.4.1 Generic Model

Generic GBDT models were developed using datasets that combines the two corridors (US 17/92 and US 441 corridors). Individual models were developed for through and left-turn movements, respectively as the movements have corresponding signal phases. It is difficult to estimate the right-turn volume because the right-turn vehicles are not restricted by the signal timing. Variables reduction was carried out for each model based on the above-mentioned methodology. The number of utilized variables to train the generic through and left turn movement models were 23 and 53, respectively. Similarly, GBDT models were developed for the peak periods. Furthermore, marginal effects were calculated for each variable in each model in Table 16 Table 17 for through

and left-turn movements, respectively. Finally, the estimated counts were plotted along with their corresponding true volumes from GRIDSMART system. Figure 24 shows plots of estimated and observed through and left turn movements.

Table 16: Through movement models marginal effects

Variable	Marginal Effect	Variable	Marginal Effect
X1Ent	0.2296	**X1X1LD**	0.0902
X1SD	0.2486	**X1LD**	0.3439
X4SD	0.0289	**X3SGM**	0.0046
X3Ent	-0.0036	**X2LGM**	0.0408
CycleLengthM	0.0620	**X1SGM**	0.0133
X1X1Ent	0.0742	**X3SU**	-0.0071
Period	--	**CycleLengthU**	-0.0248
X1X3Ent	0.0287	**CycleLengthD**	-0.0095
X1SU	0.0828	**X3LGD**	-0.0052
X4LD	0.0045	**X3SD**	-0.0080
X1X1SU	0.0196	**X4RU**	0.0212
X4SU	-0.0101		

Table 17: Left turn movement models marginal effects

Variable	Marginal Effect	Variable	Marginal Effect
X1Ent	0.0215	**X3LGM**	-0.0025
X1LGM	0.0613	**X1LGD**	0.0017
X4LGU	0.0014	**X1X3SU**	0.0028
X1X4LGU	0.0003	**X1LU**	-0.0226
X1SD	-0.0001	**X3SGD**	-0.0017
X1X1Ent	0.0072	**X4SD**	-0.0048
X1SGU	-0.0063	**X1X3SGD**	-0.0005
X1X1LGM	-0.0015	**X2SGD**	0.0041
X4SGU	-0.0098	**X3SGM**	-0.0031
X2SGM	0.0044	**X1X2SGD**	0.0052
X1X4RU	0.0279	**X1X4SGM**	0.0026
X3LU	0.0491	**X1SU**	0.0055
X4SGM	0.0090	**X1X4LGD**	-0.0001
X4RU	0.0456	**X2LGM**	-0.0051
X1SGM	0.0037	**X3LD**	0.0048
X1X1SGU	-0.0031	**X1X2RU**	0.0079
X1X3LU	0.0114	**X3SU**	-0.0005
Period	--	**X1X4LD**	0.0143
X3SGU	-0.0005	**X1X4RD**	-0.0147
X1X2SGM	0.0022	**Direction**	--
X2RU	0.0157	**X1X1LGD**	0.0005
X1X1SD	-0.0004	**X1X3SGU**	-0.0003
X4LD	0.0269	**CycleLengthU**	-0.0015
X1RD	0.0172	**X1RU**	0.0157
X4SGD	0.0000	**X3LGU**	-0.0021
CycleLengthM	0.0048	**X1X3LGM**	0.0001

Variable	Marginal Effect	Variable	Marginal Effect
X1LGU	-0.0017		

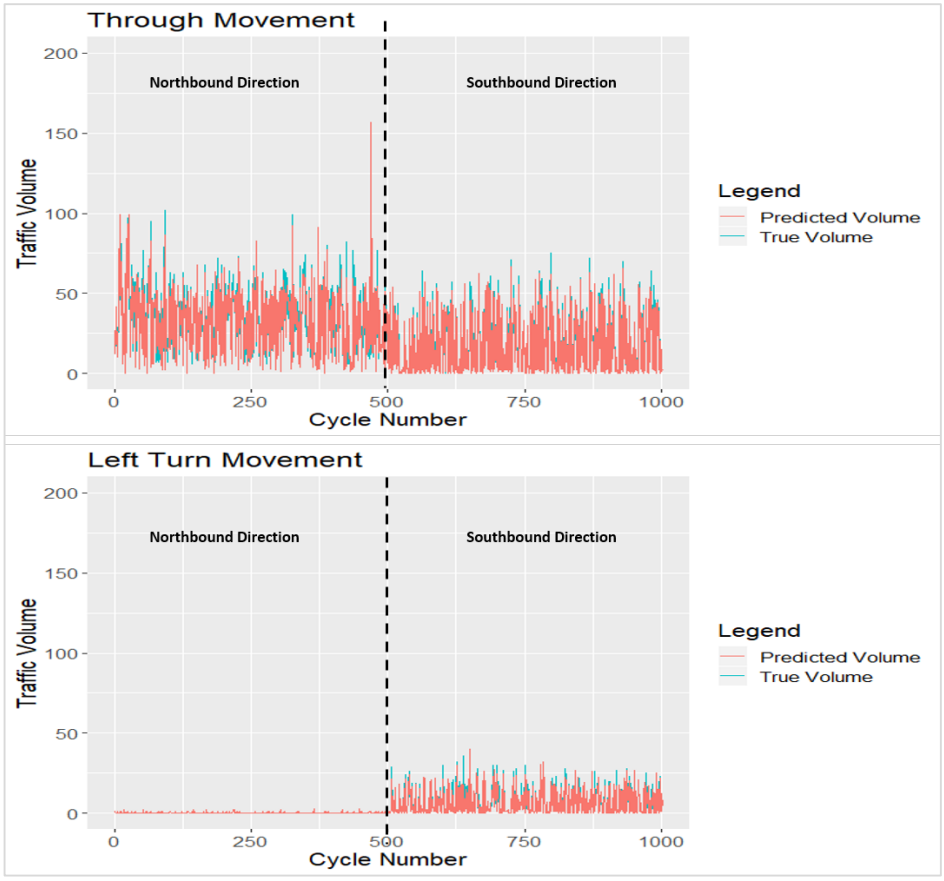

Figure 24: Estimated turning movement counts and their corresponding GRIDSMART counts

3.4.2 Corridors Models

The corridor models consist of eight different models. First, a generic model was developed for each corridor (US 17/92 and US 441) for both through and left-turn movements. Then, turning estimation models were developed for peak periods. The following models were developed for US 17/92 and US 441 corridors:

- Model to estimate through movement using generic dataset.
- Model to estimate left turn movement using generic dataset.
- Model to estimate through movement in peak periods (AM and PM peaks).
- Model to estimate left turn movement in peak periods (AM and PM peaks).

Likewise, for each model, variables were chosen based on relative influence values and the effects of excluding the variable from the model. Twenty-nine variables were utilized in training both the generic models of through movement for US 17/92 and US 441 corridors. Moreover, 40 and 59 variables were utilized to train the generic left-turn movement models for US 17/92 and US 441 corridors, respectively.

The marginal effects were calculated for each variable to capture its effect in the model. Table 18 to Table 21 show the calculated marginal effects for the developed generic through and left turn movement models for the US 17/92 and US 441 corridors. Finally, the estimated turning movement counts were plotted against the GRIDSMART counts for both corridors. Figure 25 and Figure **26** show these temporal plots of generic models for both through and left-turn movements, respectively.

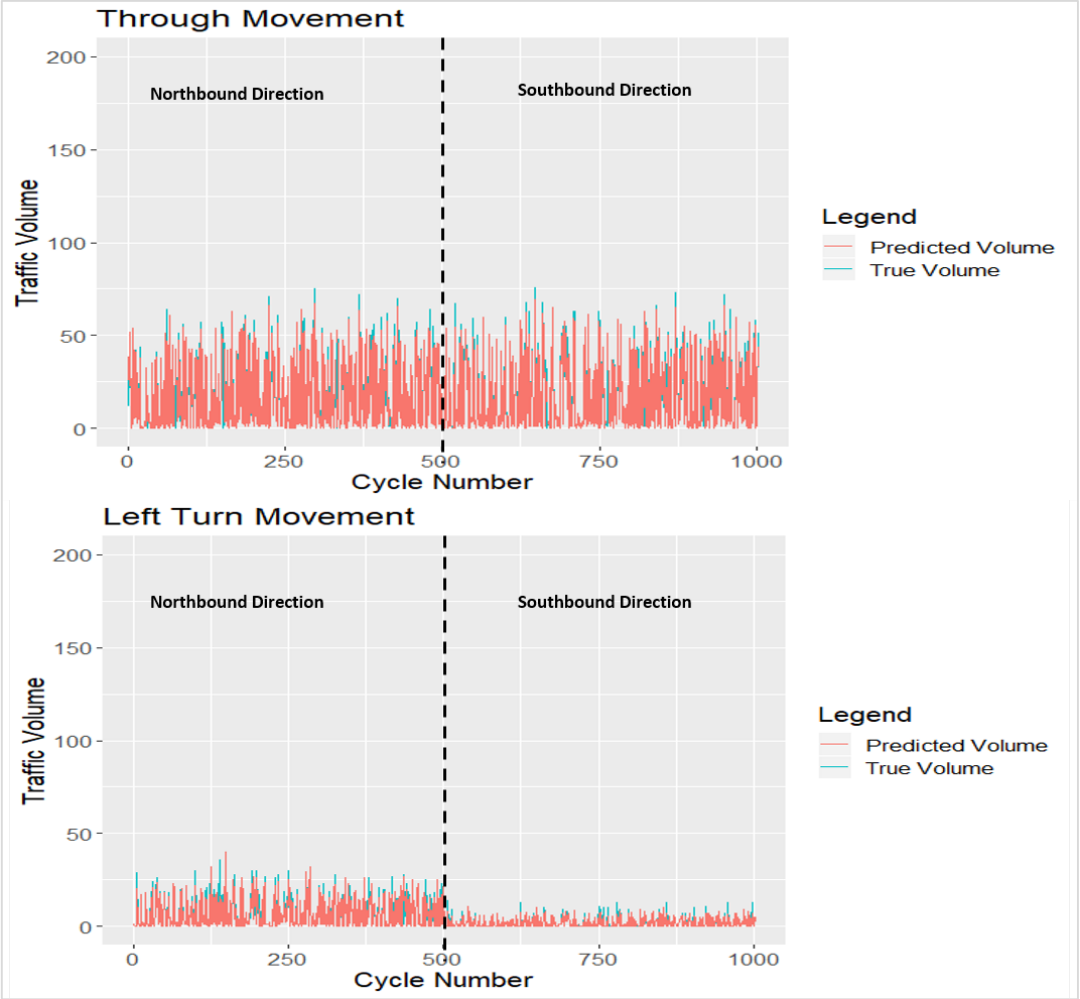

Figure 25: Estimated turning movement counts and their corresponding GRIDSMART counts (US 17/92)

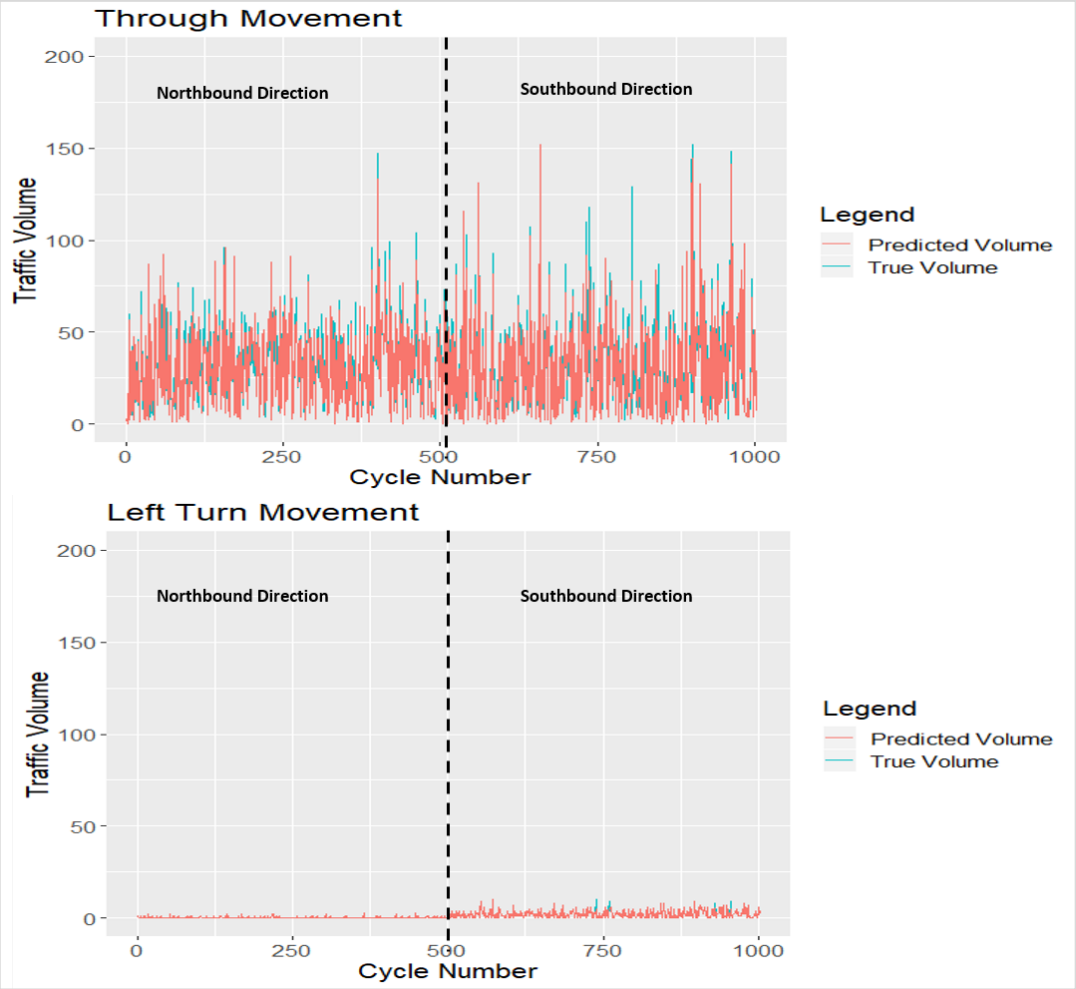

Figure 26: Estimated turning movement counts and their corresponding GRIDSMART counts (US 441)

Table 18: Generic through movement models marginal effects – US 17/92 corridor

Variable	Marginal Effect	Variable	Marginal Effect
X1Ent	0.1967	**X1XCycleLengthM**	-0.0044
X1SD	0.2167	**X4LGM**	0.0075
X1X1Ent	0.0641	**X3SGM**	0.0049
X1X1SD	0.0361	**X1X1SU**	0.0098
CycleLengthM	0.0509	**X1X4LU**	0.0165
X1LD	0.1812	**X1X3SGM**	-0.0032
X1X3SU	-0.0034	**X1LGM**	0.0158
X1X2LGM	0.0168	**X4LU**	0.0740
X1SU	0.0180	**X4SU**	-0.1115
X1SGM	0.0063	**X3LGD**	-0.0054
X2LGM	0.0075	**X1X3LGM**	0.0065
X1X1LD	0.1320	**X4RU**	0.0928
X1X4LGM	0.0208	**CycleLengthD**	-0.0080
X1X3Ent	0.0055	**CycleLengthU**	-0.0159
X1X2RU	0.2612		

Table 19: Generic left turn movement models marginal effects – US 17/92 corridor

Variable	Marginal Effect	Variable	Marginal Effect
X1LGM	0.0745	**X4SD**	0.0200
X1Ent	0.0185	**X2SGM**	0.0024
X1X1Ent	0.0201	**X1X3SGU**	-0.0015
X1X4LGU	0.0035	**X1X4LD**	0.0045
X4LGU	-0.0002	**X3SGM**	-0.0073
X1X1LGM	-0.0063	**X2SGD**	0.0018
Direction	--	**X1X1SGD**	-0.0011
X1SD	-0.0007	**Period**	--
X3SU	0.0035	**X1X2RU**	0.0058
X1SGM	0.0054	**X1X3LGM**	0.0012
X1SU	-0.0036	**X3LGM**	0.0003
X4RU	0.0629	**CycleLengthM**	0.0038
X3LGU	-0.0021	**X3SGU**	-0.0014
X1SGU	-0.0052	**X1X2SGD**	0.0014
X1X4RU	0.0196	**X1X1SGU**	-0.0032
X1X1SU	0.0008	**X1RD**	0.0172
X1X3LGD	0.0013	**X3LU**	0.0134
X4LD	0.0369	**X1X2LGU**	0.0005
X1X3SU	0.0011	**X1X4SD**	0.0126
X1X3LGU	-0.0005	**CycleLengthU**	-0.0024

Table 20: Generic through movement models marginal effects – US 441 corridor

Variable	Marginal Effect	Variable	Marginal Effect
X1Ent	0.1336	**X2SGM**	0.0228
X1SD	0.2656	**X1X4RU**	0.0379
X1X1Ent	0.0433	**X3SU**	-0.0053
X3Ent	-0.0081	**CycleLengthD**	-0.0121
CycleLengthM	0.0489	**X1RD**	0.2245
X1SU	0.1938	**X3LGD**	-0.0079
X4LD	0.0283	**X1X3RD**	0.1007
X4SU	-0.0033	**X1X4LU**	0.0125
X1X1SU	0.0262	**X1X4SD**	-0.0576
X4SD	0.0426	**X3SGD**	-0.0043
Period	0.0000	**X1X4LD**	-0.0074
X1X3Ent	0.0150	**X1X3SU**	-0.0034
X1SGM	0.0304	**X3SD**	-0.0065
X3SGM	0.0152	**CycleLengthU**	-0.0344
X1LD	0.3930		

Table 21: Generic left turn movement models marginal effects – US 441 corridor

Variable	Marginal Effect	Variable	Marginal Effect
X4SGM	0.0170	**X1X4RU**	0.0190
X1LGM	0.0503	**X4LGM**	-0.0051
X1Ent	0.0103	**X1X3LGU**	-0.0003
X1X4SGM	0.0085	**X1X1SD**	0.0002
X4SGU	-0.0118	**X3Ent**	-0.0005
X1SD	0.0039	**X1X3SGD**	0.0006
Period	0.0000	**X4SGD**	-0.0016
X4LD	0.0457	**X1X4SGD**	-0.0010
X1SU	0.0201	**X2SU**	0.0001
X1X2LGU	0.0031	**X2SGD**	-0.0001
X3LGM	0.0023	**X4LGU**	0.0009
X3LGU	-0.0006	**X4SU**	-0.0006
X1LGD	0.0025	**X1X2SGM**	0.0029
Direction	0.0000	**X1LGU**	-0.0022
X1SGM	0.0041	**X2LGD**	0.0002
X3SU	-0.0019	**X1X4SGU**	0.0000
X1RU	0.0416	**X1X2SU**	0.0013
X1X4SU	0.0001	**X1X3SGM**	-0.0013
X4RU	0.0150	**X1SGD**	-0.0020
X3SGU	0.0000	**X1X1LGM**	-0.0030
X1RD	0.0183	**X2LU**	0.0116
X1X2LGM	-0.0070	**X1X3LGM**	0.0001
X2SGM	0.0046	**X1X4LU**	0.0120
X2RU	0.0022	**X3LGD**	0.0001
X2LGM	-0.0066	**CycleLengthD**	0.0000
X3LU	0.0049	**CycleLengthU**	-0.0014
X1X1LGD	0.0009	**X1SGU**	0.0000
X4SD	0.0047	**X1X1RD**	0.0124
CycleLengthM	0.0029	**X1X1RU**	0.0119
X3SGM	-0.0003		

3.4.3 Groups Models

After training several models to estimate the turning movement counts for each group, it was found that the performance measures got improved when the utilized datasets were more refined. Thus, turning movement estimation models were trained separately for each group in each corridor following the same GBDT methodology using peak periods datasets. A total of 11 groups datasets were processed, i.e., four groups at the US 17/92 corridor and seven groups at the US 441 corridor. These datasets were utilized to develop generic models as well as peak periods models for each group. Note that each developed model estimates one turning movement at a time. Hence, a total

of 44 models were trained. Figure 27 shows the developed models for each estimated movement.

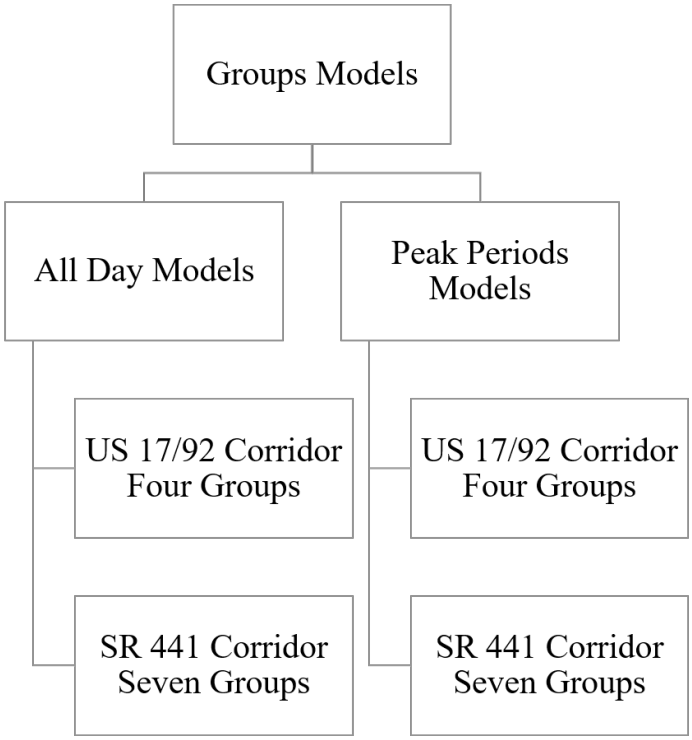

Figure 27: Developed groups models for each estimated movement

3.5 Results

3.5.1 Summary of Performance for All Conditions

After training 56 models to estimate short-term through and left-turn movements at signalized intersections, the estimated counts were compared to movement counts from the GRIDSMART system. Performance measures (MAE, RMSE, and MAPE) were calculated for testing datasets for each developed model for different time periods: AM peak, Off peak, PM peak, and Night-time. Table 22

Table 23 shows the calculated performance measures for the through movement models. The model performance measures for the individual groups were averaged as well. The results show that the performance measures of the groups models are better than the generic models (i.e., two corridors model, US 17/92 corridor model, and US 441 corridor model). Thus, implementing a specific estimation algorithm for a certain group could improve the estimated through volumes. Similarly, Table 23 shows the model performance for left-turn movements. For left-turn movements estimation, the generic models measures were found to have around 10% percentage errors and even close to the group average values along different time periods.

Table 22: Performance measures for through movement models

Test Dataset		AM Peak			Off Peak			PM Peak			Night-Time		
Corridor	**Group**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
Two Corridors	**--**	5.5	7.4	15.4%	4.6	6.3	15.0%	5.9	8	17.4%	1.8	4.5	11.2%
Mills Corridor	**--**	5.5	7.3	15.6%	5.1	6.7	13.4%	6.1	7.9	13.1%	1.4	3.7	8.8%
Mills Corridor US 17/92	**Grp1**	3.7	4.8	9.3%	3.3	4.3	8.7%	5.7	7.8	12.6%	0.7	2.1	3.9%
	Grp2	5.5	7.1	13.2%	5	6.5	12.3%	5	6.5	12.4%	6.5	8.4	11.2%
	Grp3	5.5	7.1	19.0%	5.4	7	16.0%	6.2	7.9	16.0%	1	3.1	4.6%
	Grp4	5.8	7.8	10.3%	4.7	6.2	8.4%	5.8	7.2	9.4%	2.3	4.9	13.3%
OBT Corridor	**--**	5.1	7	14.2%	4.1	5.6	14.4%	5.4	7.3	17.0%	1.7	4	10.3%
OBT Corridor US 441	**Grp1**	5.7	8.1	12.0%	3.7	5.1	9.9%	5.6	7.3	10.5%	2.5	4.7	13.0%
	Grp2	4.6	6.2	10.8%	3.6	5.1	10.3%	5.4	8.4	11.1%	1.4	3.6	10.0%
	Grp3	4.9	6.4	12.8%	4	5.6	12.2%	5.7	7.4	12.0%	1.6	3.4	11.0%
	Grp4	5.3	7.2	12.5%	4.1	5.5	12.9%	5.1	6.9	12.4%	3	5	13.6%
	Grp5	4.5	6.2	12.1%	3.6	5.1	12.5%	5	9.6	10.5%	1.4	3.2	8.6%
	Grp6	4.5	6.1	11.8%	3.6	5.1	12.5%	5	6.8	14.1%	1.4	3.1	8.8%
	Grp7	5	6.7	16.7%	3.8	5.1	11.1%	4.5	6	12.1%	1.4	4	9.3%
Groups Average		5.00	6.70	12.8%	4.07	5.51	11.5%	5.36	7.44	12.1%	2.11	4.14	9.8%

Table 23: Performance measures for left turn movement models

Test Dataset		AM Peak			Off Peak			PM Peak			Night-Time		
Corridor	**Group**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
Two Corridors	--	0.6	1.6	11.4%	0.5	1.5	10.0%	0.9	2.1	11.3%	0.1	0.6	1.5%
Mills Corridor	--	0.9	1.8	11.2%	1.1	2.1	14.5%	1.6	2.9	19.0%	0.2	0.8	3.0%
Mills Corridor US 17/92	**Grp1**	1.4	2.7	7.3%	1.2	2.3	9.7%	1.9	3.6	15.5%	0.1	0.8	1.3%
	Grp2	0.9	1.9	15.0%	1.1	2.1	14.5%	1.2	2.3	9.1%	0.2	1	3.7%
	Grp3	0.7	1.4	15.6%	0.7	1.5	8.5%	1	2.1	9.1%	0.1	0.4	1.1%
	Grp4	0.15	0.7	6.0%	0.6	1.4	11.9%	0.7	1.5	17.0%	0.12	0.7	2.3%
OBT Corridor	--	0.5	1.4	11.2%	0.3	1.1	5.8%	0.5	1.4	6.0%	0.1	0.5	1.0%
OBT Corridor US 441	**Grp1**	0.1	0.1	0.4%	0	0.1	0.1%	0.2	0.8	8.0%	0	0.1	0.1%
	Grp2	0.5	1.6	9.7%	0.01	0.1	0.5%	0.1	1	4.6%	0	0.1	0.1%
	Grp3	0.7	1.5	14.2%	0.9	1.9	14.5%	1.1	2.3	13.3%	0.1	0.5	2.0%
	Grp4	0.4	1.2	9.8%	0.1	0.2	0.4%	0.1	0.2	1.0%	0	0	0.0%
	Grp5	0.2	0.7	0.5%	0.2	0.6	0.0%	0.2	0.6	0.1%	0	0.1	0.1%
	Grp6	1.5	2.3	--	1.5	2.4	--	2.2	3	--	0.2	0.8	--
	Grp7	0.7	1.6	16.0%	0.2	0.7	4.2%	0.1	0.3	0.8%	0.1	0.4	1.0%
Groups Average		5.00	6.70	12.8%	4.07	5.51	11.5%	5.36	7.44	12.1%	2.11	4.14	9.8%

The MAPE for AM and PM peak periods were higher than off-peak and night-time periods for both through and left turn movement models. Hence, specific models were developed using refined subsets of data for peak periods only. Table 24

Table 25 show the calculated performance measures for through and left-turn movement models for peak periods, respectively. The measures show better performance measures for generic models as well as groups models. The groups average MAPE was 7% for through movement models and 2% for left turn movement models. Moreover, the average MAE was 4.8 for through models and 0.8 for left turn movement models. Finally, the average RMSE values were 7.08 and 1.58 for through and left turn movement models, respectively.

The results show that Group 3 at US 17/92 corridor consistently had bad performance during peak and off-peak periods. Thus, the developed TMC algorithm is not recommended to emulate the GRIDSMART system at this intersection. Further, the GRIDSMART system should be kept at some intersections to provide the algorithm with the required data from upstream and downstream intersections. Hence, the study concludes that the developed TMC algorithm could emulate GRIDSMART at some intersections based on the performance measures in Tables 13 to 16. Figure 28 shows the suggested locations to use the developed TMC algorithm. It shows that the developed TMC algorithm could be implemented at six intersections (two at US 17/92 and four at US 441). Moreover, it is recommended to implement peak period models for through and left-turn movements at the AM and PM peaks and use the generic models for other time periods.

The performance measures were compared to a latest study that aimed to develop short-term prediction algorithm of movements at intersections based on the Partial Least Square model (PLS). The developed PLS algorithm was compared to different prediction models (ARIMA, k-NN and SVR). The RMSE and MAPE in their study were 8.34% and 16.68% respectively (Li et al., 2020). The developed turning movement estimation models in this study have better performance. Moreover, our study is the first attempt to estimate cycle-level aggregated through and left-turn movements at signalized intersections.

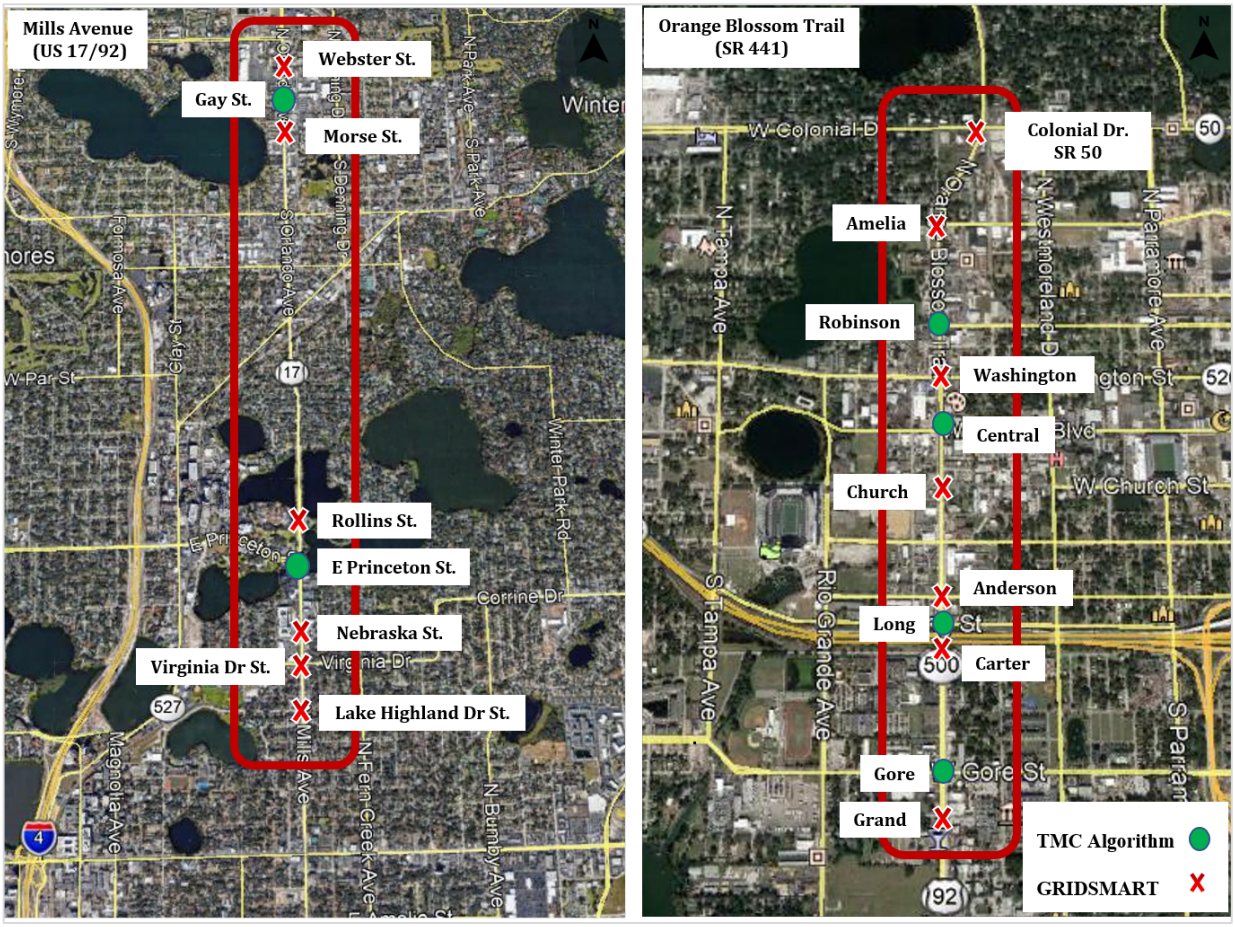

Figure 28: Suggested locations to utilize developed TMC algorithm

Table 24: Performance measures for through movement models (peak periods)

Test Dataset		Average			AM Peak			PM Peak		
Corridor	**Group**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
Two Corridors	--	5.3	7.1	14.0%	5.1	7	13.7%	5.4	7.2	14.6%
Mills Corridor	--	5.7	7.5	14.4%	5.4	7.1	15.5%	6	7.9	13.1%
Mills Corridor US 17/92	**Grp1**	4.5	6.4	10.6%	3.7	5	9.1%	5.5	7.7	12.4%
	Grp2	5.6	7.3	11.6%	5.1	6.8	12.8%	5.9	7.7	10.0%
	Grp3	5.8	7.4	17.0%	5.3	6.9	18.0%	6.3	7.9	16.0%
	Grp4	5.9	8.1	12.0%	5.9	8.3	13.5%	5.9	7.7	9.6%
OBT Corridor	--	4.8	6.6	13.0%	4.7	6.5	12.1%	5	6.7	13.9%
OBT Corridor US 441	**Grp1**	5.7	7.8	11.5%	5.8	8.3	12.0%	5.5	7.4	10.5%
	Grp2	2.3	4.7	9.2%	4.6	6.2	10.7%	5.3	8	10.9%
	Grp3	5.1	6.9	12.3%	4.7	6.3	13.1%	5.5	7.3	11.4%
	Grp4	4.8	6.7	10.5%	4.9	6.9	11.3%	4.8	6.5	9.7%
	Grp5	4.4	10.7	10.0%	4	5.4	10.5%	4.7	14	9.4%
	Grp6	4.4	5.9	11.9%	4.2	5.8	11.5%	4.5	6.1	12.1%
	Grp7	4.4	6	13.2%	4.6	6.3	15.0%	4.1	5.7	11.1%
Groups Average		4.81	7.08	11.8%	4.80	6.56	12.5%	5.27	7.82	11.2%

Table 25: Performance measures for left turn movement models (peak periods)

Test Dataset		Average			AM Peak			PM Peak		
Corridor	**Group**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
Two Corridors	--	0.7	1.9	10.2%	0.5	1.5	9.6%	0.9	2.2	10.9%
Mills Corridor	--	1.2	2.4	14.5%	0.8	1.8	10.8%	1.6	2.9	18.0%
Mills Corridor US 17/92	**Grp1**	1.7	3.3	9.6%	1.5	2.9	7.9%	1.9	3.7	11.5%
	Grp2	1.2	2.3	12.3%	1	2	13.2%	1.4	2.7	11.3%
	Grp3	0.9	1.8	13.2%	0.7	1.4	13.0%	1.2	2.2	13.5%
	Grp4	0.4	1.2	8.0%	0.2	0.7	3.4%	0.8	1.6	13.4%
OBT Corridor	--	0.5	1.4	9.2%	0.5	1.4	10.9%	0.5	1.4	6.9%
OBT Corridor US 441	**Grp1**	0.2	0.7	4.2%	0.1	0.2	0.4%	0.3	0.9	7.7%
	Grp2	0.4	1.4	6.8%	0.5	1.5	9.3%	0.2	1.2	4.4%
	Grp3	0.8	1.8	11.4%	0.7	1.5	11.0%	1.1	2	11.6%
	Grp4	0.2	0.9	4.9%	0.4	1.2	8.2%	0.1	0.3	1.6%
	Grp5	0.1	0.3	0.3%	0.1	0.2	0.5%	0.1	0.3	0.1%
	Grp6	1.7	2.6	26.0%	1.5	2.3	30.0%	2.2	3	18.0%
	Grp7	0.4	1.1	7.0%	0.7	1.5	13.2%	0.1	0.2	0.2%
Groups Average		0.73	1.58	9.4%	0.67	1.40	10.0%	0.85	1.65	8.5%

3.5.2 Validation of Model Performance under Abnormal Conditions

The transferability of the developed models was further validated for abnormal situations. Thus, accident data were collected during the same time period of the GRIDSMART data. Two accidents were selected for analysis, one at each corridor. The data for an hour that includes time of the accidents was utilized to test the developed model. Figure 29 illustrates the estimated turning movements when an accident occurred at the US 17/92 corridor and Princeton street intersection. It shows that the estimated model instantaneously captured the accident. Moreover,

Table **26** shows the corresponding performance measures for the test hour. It shows both through and left turn movements models has good performance with MAPE of around 10%.

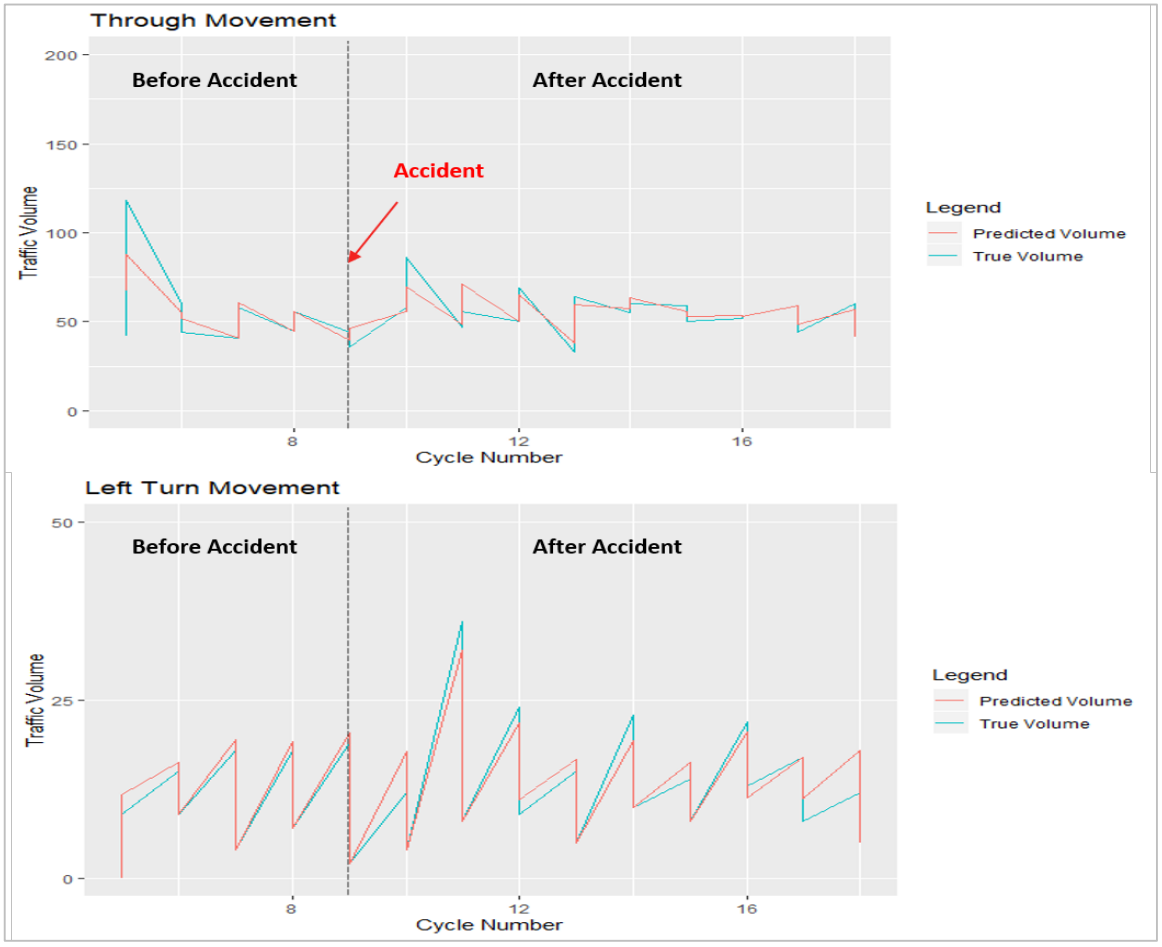

Figure 29: Estimated turning movements when an accident occurred at US 17/92 and Princeton street

Table 26: Performance measures for the model's response to accident at US 17/92 and Princeton street

Performance Measure	Through Movement	Left Turn Movement
MAE	5.6	1.5
RMSE	9.3	2.2
MAPE	10%	10.50%

Similarly, Figure 30 illustrates the estimated turning movements when an accident occurred at US 441 corridor and Robinson street. It also shows that the estimated model instantaneously captured the accident. Moreover, Table 27 shows the corresponding performance measures for the tested hour. The MAPE values were 8.7% and 6.7% for through and left turn movements models respectively.

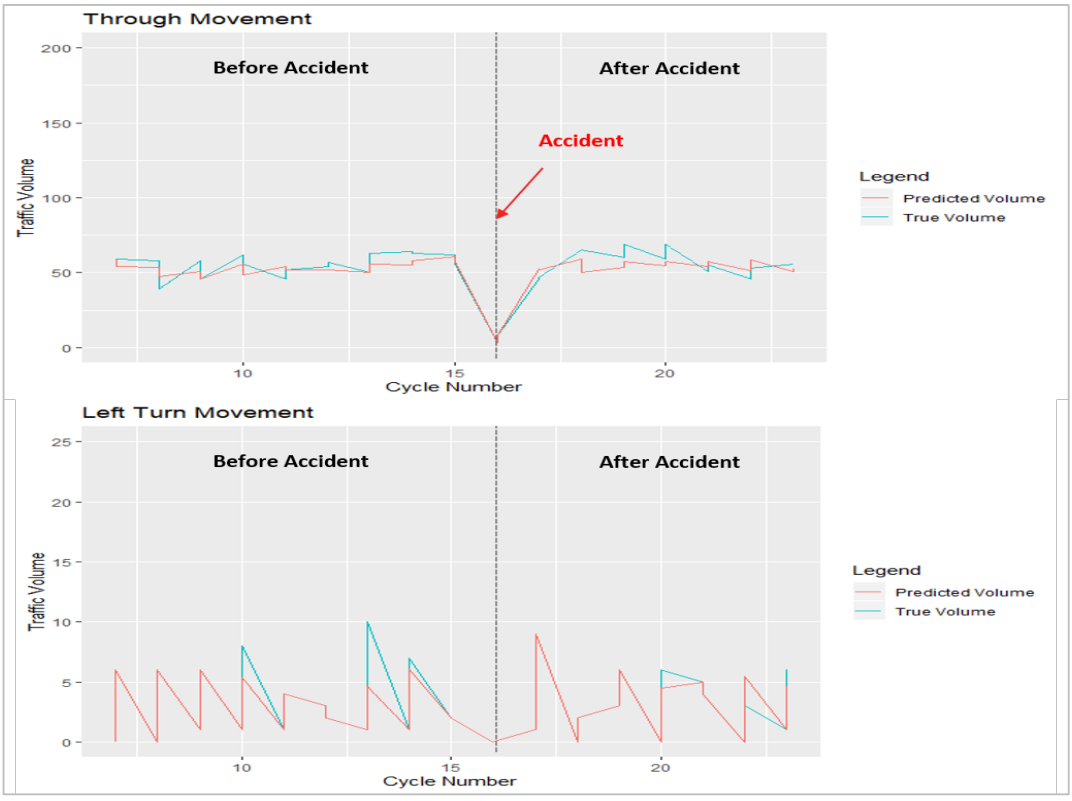

Figure 30: Estimated turning movements when an accident occurred at US 441 and Robinson street

Table 27: Performance measures for the model's response to accident at US 441 and Robinson street

Performance Measure	Through Movement	Left Turn Movement
MAE	4.9	0.4
RMSE	6.1	1.1
MAPE	8.7%	6.70%

3.6 Summary

Data for 11 intersections on US 441 corridor and 8 intersections on US 17/92 were downloaded from the GRIDSMART system. The intersections were divided into eleven groups considering distance between intersections, number of access points, minor roads, and data availability. Each group consists of three consecutive intersections. Traffic volume and signal data from the GRIDSMART system was processed and aggregated at the cycle level for weekdays. Afterwards, general features such as time periods, directions of movements, group number and corridor name were added to the processed dataset. Finally, fourteen datasets were prepared and complied, each dataset includes 110 features. The datasets were utilized in developing general turning movement estimation algorithms as well as individual algorithms for each specific corridor and group.

The GRIDSMART data was validated by comparing it with ground truth data that were collected from four intersections along the US 17/92 corridor (intersections at Nebraska St., Princeton St., Rollins St., and Virginia St). Data were processed and compared at the cycle level using MAE and MAPE as performance measures. The validation results show some misclassification in left-turn movements at Princeton St. Moreover. Meanwhile, for the intersection at Rollins St., there were big variation between ground truth data and GRIDSMART data. Finally, the average MAPE for other intersections was around 10%.

Based on previous short-term traffic forecasting studies, parametric and non-parametric models were applied, tuned, and tested (i.e., negative binomial model, finite mixture model, multivariate adaptive regression spline, neural networks, random forest, and gradient boosting). The models were compared and the GBDT model with the best performance was chosen to be utilized in estimating short-term turning movements at signalized intersections. The processed datasets include a large dataset that combines two corridors (US 17/92 and US 441), individual

datasets for each corridor and individual datasets for each group of intersections. GBDT models hyperparameters were tuned and cross-validation was implemented to avoid overfitting. Moreover, variables selection was carried out based on relative influence values and marginal effects and SE were calculated to quantify the effects of variables. Finally, a total of 56 models were trained and compared to estimate turning movement counts at signalized intersections (28 models for all time periods datasets and 28 models for AM and PM peak periods datasets). The groups models' performance measures were better than the generic models for through movement prediction. However, for left turn movements, the generic models could have around 10% percentage errors and close to the group average values along different time periods.

The transferability of developed models for the abnormal traffic was validated. Two accidents were selected for analysis, one at each corridor. The results show that the estimated model could instantaneously capture the accident. Moreover, the MAPEs for both through and left turn movements models were 10% at US 17/92. For the US 441 corridor, MAPEs were 8.7% and 6.9% for through and left turn movements models respectively. The results concluded that the developed TMC algorithm could emulate GRIDSMART at several intersections based on the performance measures shown in Tables 13 to 16. Six intersections (two at US 17/92 and four at US 441) were recommended to use the developed algorithms. Meanwhile, the GRIDSMART system could be installed at other intersections to detect and provide required data.

CHAPTER 4. ANALYSIS OF THE MIOVISION DATA

4.1 Data Collection and Preparation

Miovision is another camera-based detection system, which could detail movement counts at intersections in Seminole County. One of Miovision company's products is TrafficLink, which is a platform that helps traffic engineers create more responsive and efficient traffic networks. This platform was utilized to access traffic data for signalized intersections.

4.1.1 Study Area and Miovision Data Collection

Miovision TrafficLink was utilized to download traffic counts data at signalized intersections. On the TrafficLink platform, only four intersections along the SR 434 corridor have available data. Figure 31 shows an overview where Miovision data were collected. The system provides 5-minute aggregated turning movements and the data of only four intersections at the SR 434 corridor (Lake Brantley Dr., Wekiva Springs Road., Springs Blvd., and Sanlando Springs) were available. Table 28 summarizes the availability for approach volume data from September to December of 2019. The table shows that the data availability was not stable. September's data was downloaded and explored for two intersections: SR 434 at Lake Brantley Dr. (SEM-1230) and SR 434 at Wekiva Springs Rd. (SEM-1240). Two main issues were observed in the downloaded data. First, the timestamp variable in the dataset refers to the date the data was downloaded, not the corresponding date of the data. Second, the timestamp of the downloaded data changes based on the time zone, which requires an adjustment. Preliminary analysis was carried out using the downloaded data from the TrafficLink platform.

Figure 31: Miovision Study Area

Table 28: Availability of Miovision data

Intersection	September	October	November	December
SR 434 and E Lake Brantley Dr. (SEM - 1230)	From 1st to 12th	All volumes are zeros	From 1st to 7th	All volumes are zeros
SR 434 and Wekiva Springs Rd. (SEM - 1240)	Whole month	Whole month	From 8th to 30th	From 1st to 16th and from 20 to 24th
SR 434 and Springs Blvd,/Gum St. (SEM - 1245)	Whole month	From 1st to 15th	--	--
SR 434 and Sanlando Springs (SEM - 1250)	From 1st to 12th	All volumes are zeros	From 1st to 7th	All volumes are zeros

Due to missing data from the initial download, the UCF team and FDOT coordinated with Miovision to obtain one-minute aggregated data for September 2019 at the same intersections. The new information was provided directly by Miovison. The data was provided on March 18th, 2020. The new Miovision data contains Coordinated Universal Time (UTC), Local Time Zone (LTZ), and turning movements counts per one minute. This data was utilized to validate Miovision system by comparing it with Automated Traffic Signal Performance Measures (ATSPMs) data. Local Time Zone (LTZ) was used to indicate the Miovision time.

4.1.2 Automated Traffic Signal Performance Measures (ATSPMs) Data

The ATSPM data for SEM-1230 and SEM-1240 for the same time period were downloaded from FDOT District 5 backup data. Originally, ATSPM provides events data, it contains four

variables: Signal ID, Timestamp, Event Code, and Event Parameter. The Event Code indicates the signal phase while the Event Parameter indicates the detector. The Event Parameter provides both upstream (advanced) detector and stop-line detector data for through movement. Hence, ATSPM data at the upstream and stop-line detectors were aggregated into one minute. The aggregated data at the main approaches (i.e., eastbound and westbound) of the two intersections during the PM peak periods were plotted in Figure 32. It shows that the advanced detector volume and stop-line detector volume are consistent.

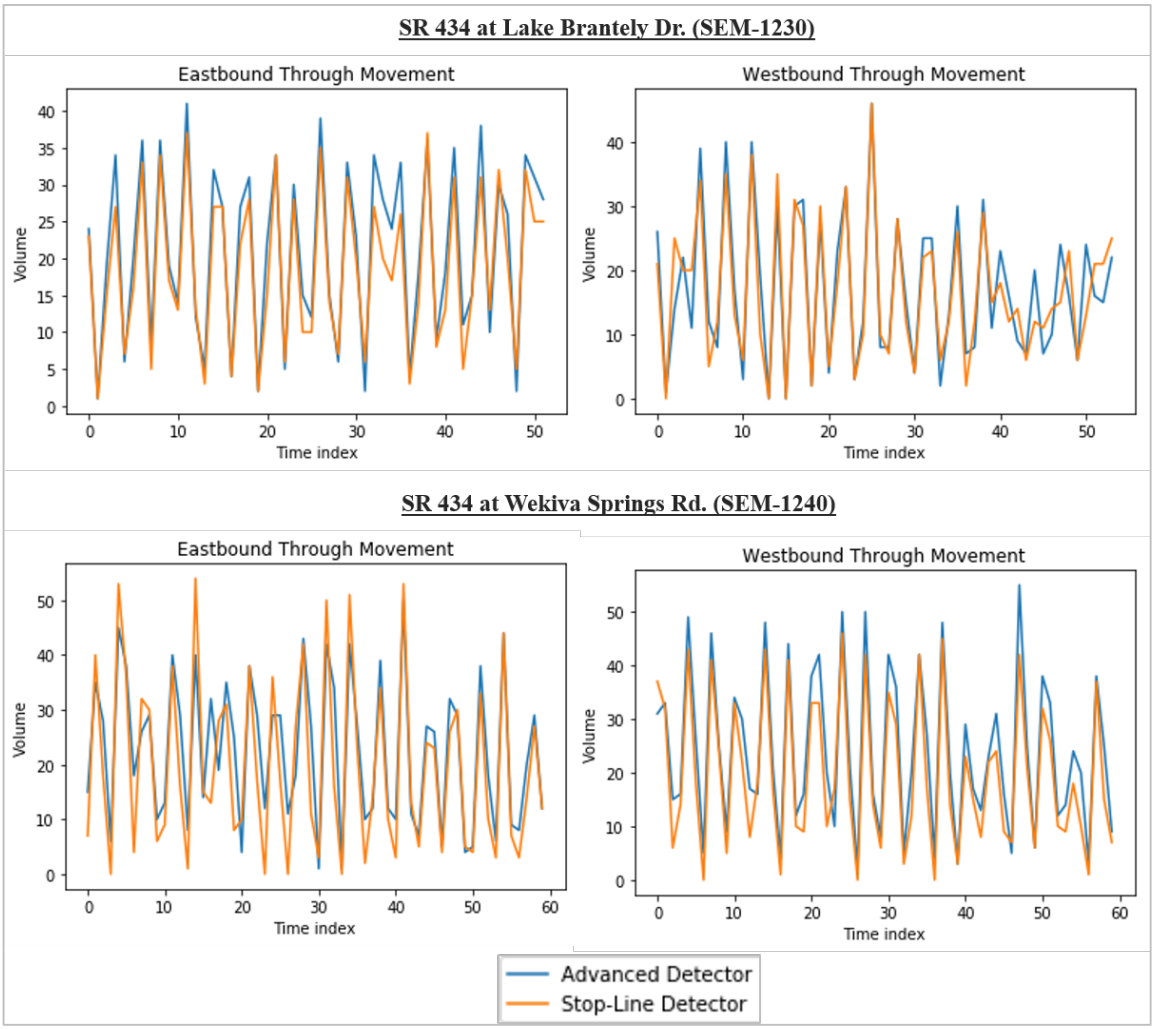

Figure 32: One-minute aggregated data for upstream (advanced) and stop-line detectors

Further, mean absolute difference and mean percentage difference were calculated in order to compare the upstream (advanced) detector and stop-line detector data. Equations (8) and (9) define these comparison measures, respectively.

$$Mean\ Absolute\ Difference = \frac{1}{n}\sum|y_1 - y_2| \quad (8)$$

$$Mean\ Percentage\ Difference = \left(\frac{1}{n}\sum\frac{y_1 - y_2}{y_2}\right) * 100 \quad (9)$$

where y_1 indicates the stop-line detector data and y2 indicates the advanced detector data. Table 29 shows the calculated comparison measures. It shows that the stop-line detectors could detect over 10% less vehicles for most cases, compared to the advanced detectors. It is expected because the stop-line detector could miss some vehicles due to its size and the advanced detector could also catch the left-turn vehicles. Both of them were compared to the Miovision data.

Table 29: Comparison measures for advanced detector and stop-line detector through movement data

Intersection ID	Approach\Movement	Mean Absolute Difference	Mean Percentage Difference
SEM-1240	Eastbound	3.23	-11.31%
	Westbound	3.28	-20.00%
SEM-1230	Eastbound	2.36	-16.73%
	Westbound	1.56	-0.31%

4.2 Data Validation

The main objective of this task is to validate the Miovision data by comparing it to the ATSPM data at the same signalized intersections. Thus, aggregated one-minute data for both ATSPM and Miovision (new one-minute aggregated data that was recently provided) were plotted for the main corridor SR 434 (East-West direction) at intersections Lake Brantley Dr. (SEM-1230) and at Wekiva Springs Rd. (SEM-1240) in Figure 33 and Figure **34**, respectively. Note that "ATSPM_adv" refers to the ATSPM data at the advanced detector and "ATSPM_stp" refers to ATSPM data at the stop line. The plots show data for the PM peak hour (17:00-18:00) and one off-peak hour (13:00-14:00). Moreover, mean absolute difference and mean percentage difference as defined in equations (8) and (9) were calculated. Note that y1 indicates Miovision data and y2 indicates ATSPM data in the two equations. Table 30 summarizes the results.

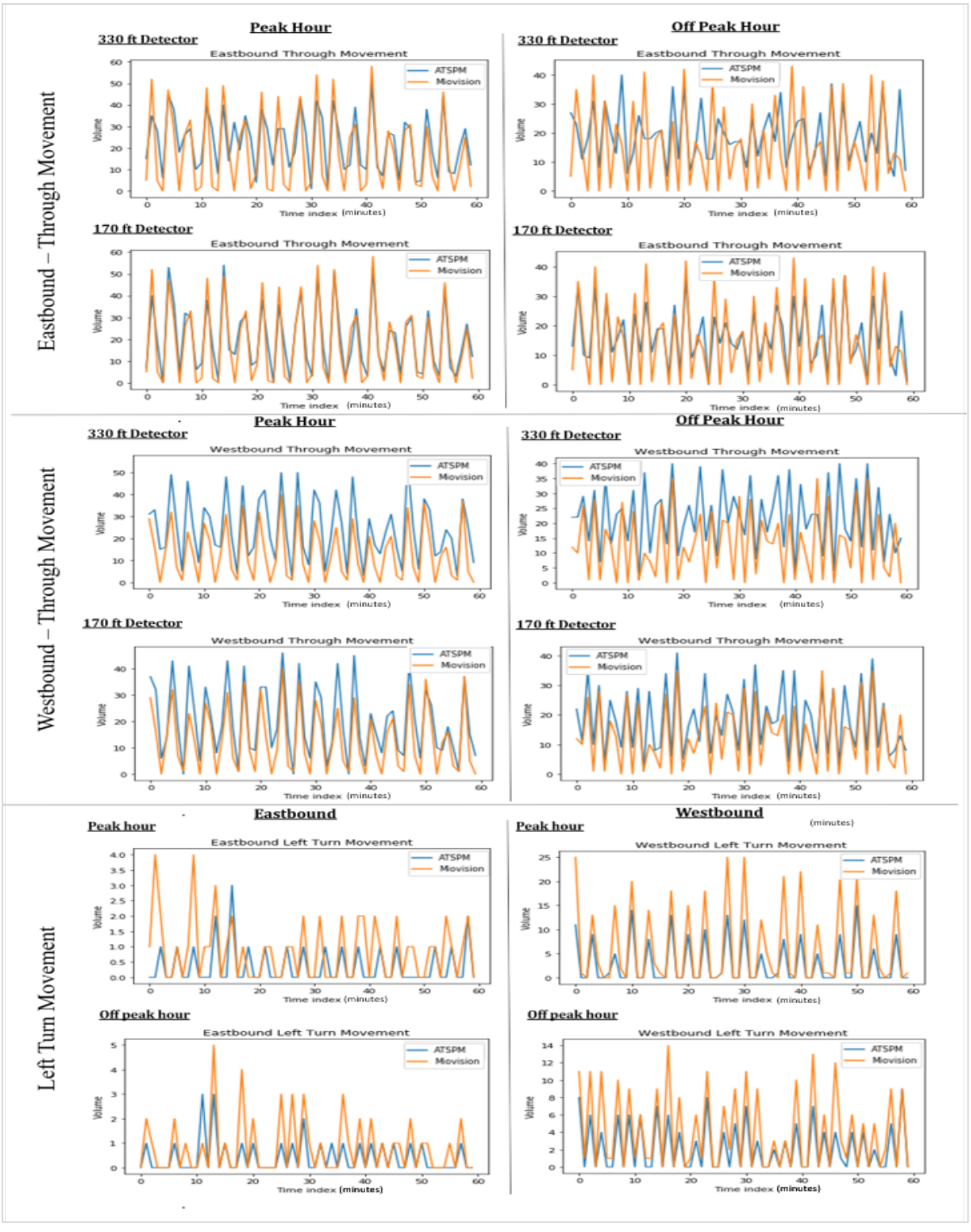

Figure 33: Miovision and ATSPM data – SR 434 and Lake Brantley Dr. Intersection (SEM-1230)

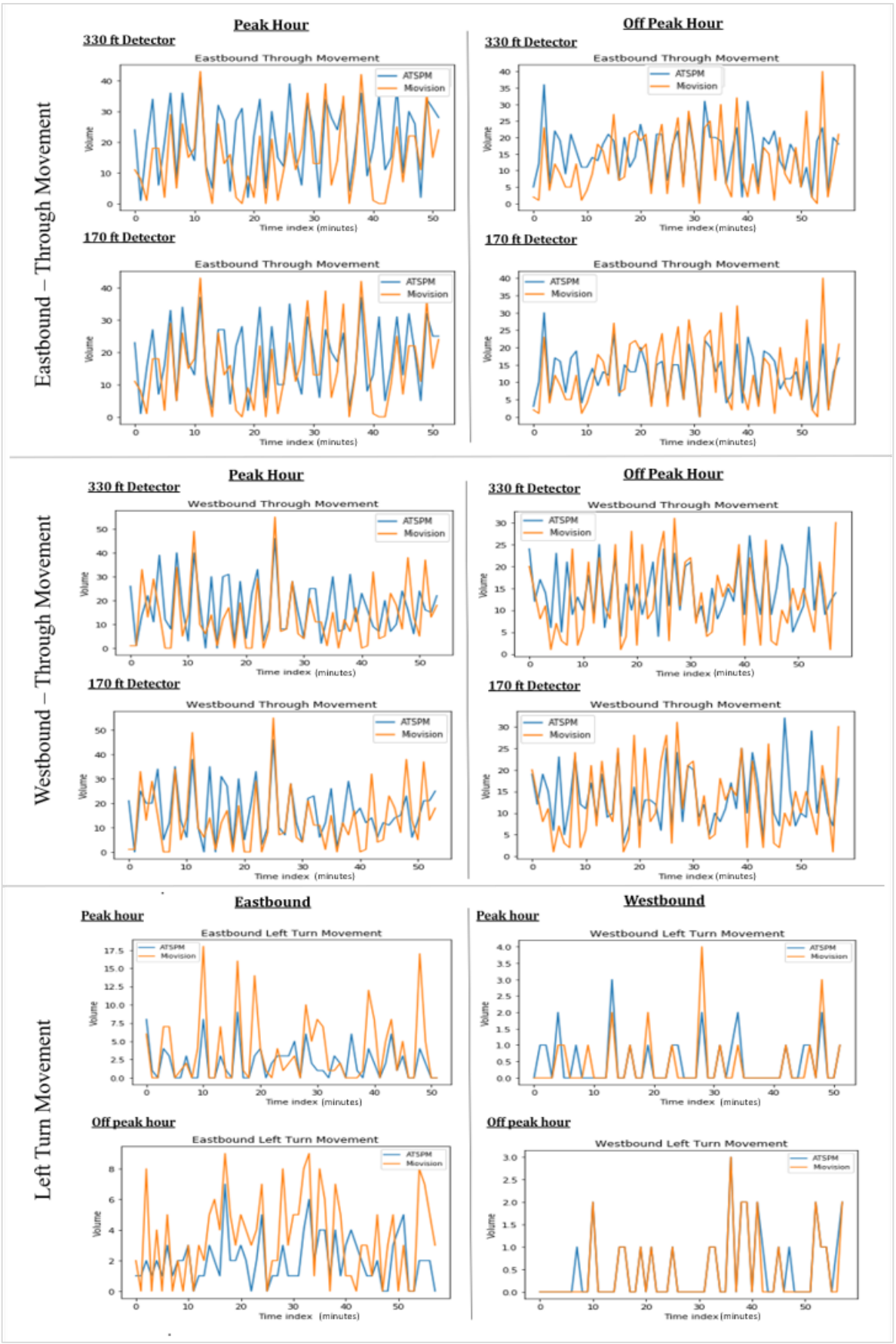

Figure 34: Miovision and ATSPM data – SR 434 at Wekiva Springs Rd. Intersection (SEM-1240)

The results show that the through movement data from Miovision is closer to the data of the ATSPM stop-line detector. Thus, it was found that Miovision could detect at least 9.06% less through volume than the ATSPM data. Moreover, it could identify more vehicles for left-turn volume for most cases. The result is expected since the stop-line detector could miss the left-turn vehicle due to the large loop.

Table 30: Comparison measures for advanced detector and stop-line detector through movement

Intersection ID	Approach\ Movement	Mean Absolute Difference			Mean Percentage Difference		
		Through Stop line	Through Advanced	Left Turn	Through Stop line	Through Advanced	Left Turn
SEM-1240	Eastbound	4.87	7.09	0.85	23.24%	24.35%	-5.32%
	Westbound	5.73	8.14	2.58	56.86%	63.33%	-47.11%
SEM-1230	Eastbound	3.67	4.9	1.4	9.06%	21.39%	-27.52%
	Westbound	3.86	4.66	0.41	14.28%	11.33%	19.64%

4.3 Summary

Miovision five minutes aggregated data was downloaded from TrafficLink platform at only four signalized intersections in Seminole County. Preliminary validation was carried out using this data. For the results please see Appendix D.

Afterwards, new one-minute aggregated turning movement data for September 2019 was provided at the same locations on March 18th, 2020. This data was utilized in validating Miovision system using ATSPM data. Hence, ATSPM data at upstream (advanced) and stop-line detectors was downloaded and aggregated in one-minute increments for the same time period for two intersections with corridor SR 434, at Lake Brantley Dr. (SEM-1230) and at Wekiva Springs Rd. (SEM-1240).

First, the ATSPM data at the advanced and stop line detectors were compared. It shows that the stop-line detectors could detect over 10% less vehicles for most cases, compared to the advanced detectors. Afterwards, both datasets were compared with the aggregated one-minute Miovision data in order to validate it. The validation indicated that the Miovision detect at least 9.06% less through vehicles compared to ATSPM. The results also showed that Miovision could detect similar or less through volume than the ATSPM data with average mean absolute

difference of 4.5 vehicles per minute. Moreover, it identifies more vehicles for left-turn volumes for most cases with average mean absolute difference of 1.37 vehicles per minute.

CHAPTER 5. SUMMARY AND CONCLUSIONS

5.1 Summary

In this research project, the main objective was to develop algorithms to estimate turning movements at signalized intersections using traffic data from adjacent intersections. Aiming to achieve this objective, three main tasks were accomplished:

1. **Comprehensive literature review to illustrate the studies that aimed to predict/estimate different traffic parameters**

Previous studies were reviewed to conclude the most efficient methodological approaches to estimate traffic parameters. It also summarized the utilized data in literature to develop parametric and non-parametric (machine learning models) as well as the performance measures that were commonly used to evaluate the developed models.

2. **Data validation of the GRIDSMART system**

Turning movements data from the GRIDSMART system was validated by comparing its data with the corresponding data counted manually. The results suggested that the detection accuracy for movements with high volumes are acceptable. As a result, the GRIDSMART data was utilized to develop different turning movement estimation algorithms. Moreover, sequence of intersections was determined considering several factors: correlation, coefficient of variation, distance between intersections, number of access points/stop-controlled intersection between two consecutive intersections, and data availability. As a result, a total of 19 intersections were divided into four groups on US 17/92 and seven groups on US 441. The collected data was divided into two parts: one is to develop algorithms and the other is to validate the developed algorithms. The accuracy of sequence-level and generic algorithms was validated and compared for each sequence of intersections.

3. Turning movement estimation algorithms development using the GRIDSMART data

Different approaches that were widely employed in previous short-term traffic forecasting studies (parametric models, machine learning models, and deep learning models) were followed to develop a comprehensive turning movement estimation model. Six models were applied, tuned, and tested: negative binomial model, finite mixture model, multivariate adaptive regression spline, neural networks, random forest, and gradient boosting. The models were developed to estimate corridor-level (North-South direction) through and left-turn movements. They were trained to estimate one traffic movement at a time. Three performance measures were utilized to evaluate the developed models: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). An extensive comparison was carried out to determine the best methodological approach to estimate turning movement estimation algorithms for individual groups, each corridor, and for the whole study area. Based on the model performance, a list of intersections where the developed algorithms could be applied was provided.

4. Analysis of the Miovision data

The Miovision five-minute aggregated data was downloaded from TrafficLink platform at only four signalized intersections in Seminole County. Preliminary validation was carried out using this data. Afterwards, new one-minute aggregated turning movement data was utilized in validating Miovision system using ATSPM data. Hence, ATSPM data at upstream (advanced) and stop-line detectors was downloaded and aggregated in one-minute increments for the same time period. The ATSPM data at the advanced and stop line detectors were compared. It shows that the stop-line detectors could detect over 10% less vehicles for most cases, compared to the advanced detectors. Afterwards, both datasets were compared with the aggregated one-minute Miovision data in order to validate it. The results also showed that Miovision could detect similar or less through volume than the ATSPM data with average mean

absolute difference of 4.5 vehicles per minute. Moreover, it identifies more vehicles for left-turn volumes for most cases with average mean absolute difference of 1.37 vehicles per minute.

5.2 Conclusions

In the last decade, research focused on predicting traffic volume at intersections and freeways. Limited studies have been conducted to estimate/predict turning movement counts at signalized intersections. To the best of the authors' knowledge, this study is the first attempt to estimate short-term turning movement counts at signalized intersections at the cycle level. It extends past research and aims to bridge the gap, as previous studies focused on estimating 15 minutes or one-hour traffic movements at signalized intersections with the total entering and exiting movements considered as inputs in the developed models. Moreover, in this study, it was assumed that the target intersection has no traffic volume data. Thus, implementing estimation/prediction algorithms at target intersections to provide traffic volumes will save the extensive cost of data collections, since detector systems cost more than $20,000 per intersection.

Two main data sources were explored: GRIDSMART and Miovision TrafficLink systems. First, GRIDSMART data was downloaded, processed, and aggregated at the cycle level. Then, intersections were combined in 11 different groups, and each group consists of three consecutive intersections. The Gradient Boosting Decision Tree (GBDT) model was chosen to estimate turning movements as it outperformed other parametric and nonparametric algorithms. Thus, totally 56 models were trained and tested using the GBDT model. The results show that implementing a specific estimation algorithm for a certain group improves the performance measures. Moreover, for the developed generic models using all day data, MAPE values of AM and PM peak periods were higher than off-peak and night-time periods for both through and left-turn movement models. Hence, specific models were developed using refined

subsets of data for peak periods only. The groups average MAPE was 7% for through movement models and 2% for left turn movement models. Besides, the performance measures for all the developed models outperformed the models in the most recent literature.

The study concludes that the developed Turning Movement Estimation (TMC) algorithms could emulate GRIDSMART at most intersections. To achieve the best performance, it was recommended to apply the developed algorithms at six intersections (two at US 17/92 and four at US 441) while the GRIDSMART system could be used to detect movements at the rest intersections and provide the required data for the algorithms. Moreover, it was recommended to use peak period models to estimate through and left-turn movement at the AM and PM peaks and the generic model could be used for other time periods. The right-turn movements were not estimated in this study since it is not controlled by the signal timing. Both models for individual intersection group and all groups were developed. The two types of models could provide the state-of-the-art accuracy compared to the previous studies. While the models for individual intersection group could provide more accurate estimation results, it is less convenient to program models for each intersection. Hence, the adoption of the models should depend on the practitioners' needs.

The Miovision system provides 5-minutes aggregated turning movements data. The data was downloaded at the four locations in Seminole county considering the data availability. However, some difficulties were faced while analyzing the data. First, the timestamp variable in the dataset refers to the date when the data was downloaded. Also, the timestamp changes based on the time zone. Afterwards, one-minute aggregated data was provided for the same locations. The data was compared to one-minute aggregated ATSPM data. For through movements, the Miovision data was compared with aggregated data at both upstream (advanced) detector and at stop-line validation. The results illustrated that Miovision detected at least 9.06% less through volume than the ATSPM data. However, for left-turn volume,

Miovision could help identify more vehicles in most cases.

The developed algorithms could be used estimate the through and left-turn movements on the major approach as the detectors are installed along corridors. It is recommended to aggregate turning movement counts at the cycle level instead of 5 or 10 minutes, to obtain a better estimation performance. Besides, it is suggested to install detectors at roads intersecting with the corridors to detect vehicles at the network level. It could help extend the algorithms to the minor approach and improve the accuracy of algorithms.

REFERENCES

Cai, Q., Lee, J., Eluru, N., Abdel-aty, M., 2016. Macro-level pedestrian and bicycle crash analysis : Incorporating spatial spillover effects in dual state count models. Accid. Anal. Prev. 14–22. https://doi.org/10.1016/j.aap.2016.04.018

Chang, S.C., Kim, R.S., Kim, S.J., Ahn, M.H., 2000. Traffic-flow forecasting using a 3-stage model. Proc. IEEE Intell. Veh. Symp. 2000 451–456. https://doi.org/10.1109/IVS.2000.898384

Chen, A., Chootinan, P., Ryu, S., Lee, M., Recker, W., 2012. An intersection turning movement estimation procedure based on path flow estimator. J. Adv. Transp. 46, 161–176. https://doi.org/10.1002/atr

Chen, C.F., Chen, C.W., 2011. Speeding for Fun? Exploring the Speeding Behavior of Riders of Heavy Motorcycles Using the Theory of Planned Behavior and Psychological Flow Theory. Accid. Anal. Prev. 43, 983–990. https://doi.org/10.1016/j.aap.2010.11.025

Chen, H., Grant-Muller, S., 2001. Use of sequential learning for short-term traffic flow forecasting. Transp. Res. Part C Emerg. Technol. 9, 319–336. https://doi.org/10.1016/S0968-090X(00)00039-5

Chen, M., Chien, S., 2001. Dynamic Freeway Travel-Time Prediction with Probe Vehicle Data: Link Based Versus Path Based. Transp. Res. Rec. J. Transp. Res. Board 1768, 157–161. https://doi.org/10.3141/1768-19

Chen, P., Yin, K., Sun, J., 2014. Application of Finite Mixture of Regression Model with Varying Mixing Probabilities to Estimation of Urban Arterial Travel Times. Transp. Res. Rec. J. Transp. Res. Board 96–105. https://doi.org/10.3141/2442-11

Cheng, J., Li, G., Chen, X., 2019. Research on travel time prediction model of freeway based

on gradient boosting decision tree. IEEE Access 7, 7466–7480. https://doi.org/10.1109/ACCESS.2018.2886549

Clark, S., 2003. Traffic Prediction Using Multivariate Nonparametric Regression. J. Transp. Eng. 129, 161–168. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:2(161)

Clark, S.D., Dougherty, M.S., Kirby, H.R., 1993. The use of neural networks and time series models for short term traffic forecasting: a comparative study. PTRC 21st Summer Annu. Meet. 151–162. https://doi.org/10.1007/978-1-4612-0533-3_7

Daraghmi, Y., Yi, C., Chiang, T., 2012. Space-time Multivariate Negative Binomial Regression for Urban Short-Term Traffic Volume Prediction. 2012 12th Int. Conf. ITS Telecommun. 35–40. https://doi.org/10.1109/ITST.2012.6425198

Davis, B.G.A., Nihan, N.L., 1991. Nonparametric regression and short-term freeway traffic forecasting. J. Transp. Eng. 117, 178–188.

Dia, H., 2001. An object-oriented neural network approach to short-term tra c forecasting. Eur. J. Oper. Res. 131, 253–261.

Ermagun, A., Levinson, D., 2018. Spatiotemporal traffic forecasting : review and proposed directions. , Transp. Rev. 786–814. https://doi.org/10.1080/01441647.2018.1442887

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. Ann. Stat. 19, 1–67.

Ghanim, M.S., Shaaban, K., 2018. Estimating Turning Movements at Signalized Intersections Using Artificial Neural Networks. IEEE Trans. Intell. Transp. Syst. 1–9. https://doi.org/10.1109/TITS.2018.2842147

Ghosh, B., Basu, B., O'Mahony, M., 2007. Bayesian time-series model for short-term traffic flow forecasting. J. Transp. Eng. 133, 180–189. https://doi.org/10.1038/010398d0

Ghosh, B., Basu, B., O'Mahony, M., 2005. Time-series modelling for forecasting vehicular traffic flow in Dublin. 84th Transp. Res. Board Conf. Natl. Acad. Washingt. D.C. https://doi.org/10.1109/JPHOTOV.2015.2434597

Greenwell, B., Boehmke, B., Cunningham, J., 2019. Generalized Boosted Regression Models.

Guo, F., Krishnan, R., Polak, J., 2013. A computationally efficient two-stage method for short-term traffic prediction on urban roads. Transp. Plan. Technol. 36, 62–75. https://doi.org/10.1057/s41309-017-0025-x

Guo, F., Polak, J.W., Krishnan, R., 2018. Predictor fusion for short-term traffic forecasting. Transp. Res. Part C Emerg. Technol. 92, 90–100. https://doi.org/10.1016/j.trc.2018.04.025

Guo, J., Huang, W., Williams, B.M., 2014. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. Transp. Res. Part C Emerg. Technol. 43, 50–64. https://doi.org/10.1016/j.trc.2014.02.006

Head, K.L., 1995. Event-Based Short-Term Traffic Flow Prediction Model. Transp. Res. Rec. J. Transp. Res. Board 45–52.

Hyndman, R.J., 2010. Measuring forecast accuracy. J. Bus. Forecast. Methods Syst. 14. https://doi.org/10.1007/s11336-007-9039-7

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Jiang, X., Adeli, H., Asce, H.M., 2005. Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting. J. Transp. Eng. 131, 771–779.

Kamarianakis, Y., Prastacos, P., 2005. Space-time modeling of traffic flow. Comput. Geosci. 31, 119–133. https://doi.org/10.1016/j.cageo.2004.05.012

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T., 2017. LightGBM : A Highly Efficient Gradient Boosting Decision Tree. 31st Conf. Neural Inf. Process. Syst. 1–9.

Kong, Q.J., Xu, Y., Lin, S., Wen, D., Zhu, F., Liu, Y., 2013. UTN-model-based traffic flow prediction for parallel-transportation management systems. IEEE Trans. Intell. Transp. Syst. 1541–1547. https://doi.org/10.1109/TITS.2013.2252463

Li, W., Ban, X., 2019. A Deep Learning Approach for Lane-Based Short-Term Traffic Volume Prediction at Signalized Intersections, in: Transportation Research Board 98th Annual Meeting.

Li, W., Wang, J., Fan, R., Zhang, Y., Guo, Q., Siddique, C., Jeff, X., 2020. Short-term traffic state prediction from latent structures : Accuracy. Transp. Res. Part C 72–90. https://doi.org/10.1016/j.trc.2019.12.007

Li, X., Bai, R., 2017. Freight Vehicle Travel Time Prediction Using Gradient Boosting Regression Tree. 2016 15th IEEE Int. Conf. Mach. Learn. Appl. 1010–1015. https://doi.org/10.1109/icmla.2016.0182

Lin, S., Xi, Y., Yang, Y., 2008. Short-term traffic flow forecasting using macroscopic urban traffic network model. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 134–138. https://doi.org/10.1109/ITSC.2008.4732567

Liu, B., Cheng, J., Cai, K., Shi, P., Tang, X., 2017. Singular point probability improve LSTM network performance for long-term traffic flow prediction, in: National Conference of Theoretical Computer Science. pp. 328–340.

Mackenzie, J., Roddick, J.F., Zito, R., 2018. An Evaluation of HTM and LSTM for Short-Term Arterial Traffic Flow Prediction. IEEE Trans. Intell. Transp. Syst. 1–11.

https://doi.org/10.1109/TITS.2018.2843349

Min, X., Hu, J., Chen, Q., Zhang, T., Zhang, Y., 2009. Short-term traffic flow forecasting of urban network based on dynamic STARIMA model. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 461–466. https://doi.org/10.1109/ITSC.2009.5309741

Min, X., Hu, J., Zhang, Z., 2010. Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 1535–1540. https://doi.org/10.1109/ITSC.2010.5625123

Qiao, W., Haghani, A., Hamedi, M., 2013. A nonparametric model for short-term travel time prediction using bluetooth data. J. Intell. Transp. Syst. Technol. Planning, Oper. 17, 165–175. https://doi.org/10.1080/15472450.2012.748555

Rong, Y., Zhang, X., Feng, X., Ho, T.K., Wei, W., Xu, D., 2015. Comparative analysis for traffic flow forecasting models with real-life data in Beijing. Adv. Mech. Eng. 7, 1–9. https://doi.org/10.1177/1687814015620324

Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. Transp. Res. Part C Emerg. Technol. 11, 121–135. https://doi.org/10.1016/S0968-090X(03)00004-4

Thomas, T., Weijermars, W., Berkum, E. Van, 2010. Predictions of Urban volumes in single time series. IEEE Trans. Intell. Transp. Syst. 11, 71–80. https://doi.org/10.1109/TITS.2009.2028149

Urbanik, T., Tanaka, A., Lozner, B., Lindstrom, E., Lee, K., Quayle, S., Beaird, S., Tsoi, S., Ryus, P., Gettman, D., Sunkari, S., Balke, K., Bullock, D., 2015. National Cooperative Highway Research Program Report 812: Signal Timing Manual.

Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting:

Overview of objectives and methods. Transp. Rev. 24, 533–557. https://doi.org/10.1080/0144164042000195072

Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. Transp. Res. Part C Emerg. Technol. 13, 211–234. https://doi.org/10.1016/j.trc.2005.04.007

Williams, B.M., Hoel, L.A., 2003. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. J. Transp. Eng. 129, 664–672. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)

Wu, Y.J., Chen, F., Lu, C.T., Yang, S., 2016. Urban Traffic Flow Prediction Using a Spatio-Temporal Random Effects Model. J. Intell. Transp. Syst. Technol. Planning, Oper. 20, 282–293. https://doi.org/10.1080/15472450.2015.1072050

Xia, D., Wang, B., Li, H., Li, Y., Zhang, Z., 2016. A distributed spatial-temporal weighted model on MapReduce for short-term traffic flow forecasting. Neurocomputing 179, 246–263. https://doi.org/10.1016/j.neucom.2015.12.013

Xia, Y., Chen, J., 2017. Traffic Flow Forecasting Method based on Gradient Boosting Decision Tree. Adv. Eng. Res. 130, 413–416. https://doi.org/10.2991/fmsmt-17.2017.87

Xie, Y., Zhang, Y., Ye, Z., 2007. Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition. Comput. Civ. Infrastruct. Eng. 22, 326–334. https://doi.org/10.1111/j.1467-8667.2007.00489.x

Xu, Y., Kong, Q., Liu, Y., 2013. A Spatio-Temporal Multivariate Adaptive Regression Splines Approach for Short-Term Freeway Traffic Volume Prediction. 16th Int. IEEE Conf. Intell. Transp. Syst. 217–222. https://doi.org/10.1109/ITSC.2013.6728236

Yang, S., Wu, J., Du, Y., He, Y., Chen, X., 2017. Ensemble Learning for Short-Term Traffic

Prediction Based on Gradient Boosting Machine. J. Sensors 2017, 1–15. https://doi.org/10.1155/2017/7074143

Yoon, B., Chang, H., 2014. Potentialities of Data-Driven Nonparametric Regression in Urban Signalized Traffic Flow Forecasting. J. Transp. Eng. 140, 548–556. https://doi.org/10.1061/(ASCE)TE.1943-5436

Yuan, J., Abdel-aty, M., 2018. Approach-level real-time crash risk analysis for signalized intersections. Accid. Anal. Prev. 274–289. https://doi.org/10.1016/j.aap.2018.07.031

Zhan, H., Gomes, G., Li, X.S., Madduri, K., Sim, A., Wu, K., 2018. Consensus Ensemble System for Traffic Flow Prediction. IEEE Trans. Intell. Transp. Syst. 1–12.

Zhang, C., Ma, Y., n.d. Ensemble Machine Learning Methods and Applications. Springer New York Dordrecht Heidelberg London.

Zhang, F., Zhu, X., Hu, T., Guo, W., Chen, C., Liu, L., 2016. Urban link travel time prediction based on a gradient boosting method considering spatiotemporal correlations. ISPRS Int. J. Geo-Information 5. https://doi.org/10.3390/ijgi5110201

Zhang, Y., Haghani, A., 2015a. A gradient boosting method to improve travel time prediction. Transp. Res. Part C 58, 308–324. https://doi.org/10.1016/j.trc.2015.02.019

Zhang, Y., Haghani, A., 2015b. A gradient boosting method to improve travel time prediction. Transp. Res. Part C Emerg. Technol. 58, 308–324. https://doi.org/10.1016/j.trc.2015.02.019

Zhao, Z., Chen, W., Wu, X., Chen, P.C.V., Liu, J., 2017. LSTM network: A deep learning approach for short-term traffic forecast. IET Image Process. 11, 68–75. https://doi.org/10.1049/iet-its.2016.0208

Zheng, J., Liu, H.X., 2017. Estimating traffic volumes for signalized intersections using

connected vehicle data. Transp. Res. Part C 347–362.

Zheng, W., Lee, D.-H., Shi, Q., 2006. Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. J. Transp. Eng. 132, 114–121. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:2(114)

Zheng, Z., Su, D., 2014. Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. Transp. Res. Part C Emerg. Technol. 43, 143–157. https://doi.org/10.1016/j.trc.2014.02.009

Zhu, J.Z., Cao, J.X., Zhu, Y., 2014. Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections. Transp. Res. Part C Emerg. Technol. 47, 139–154. https://doi.org/10.1016/j.trc.2014.06.011

Zhu, Z., Peng, B., Xiong, C., Zhang, L., 2016. Short-term traffic flow prediction with linear conditional Gaussian Bayesian network. J. Adv. Transp. 50, 1111–1123. https://doi.org/10.1002/atr