Final Report Deliverable

Contract BDV24 562-10

# Phase II - ATTAIN: Improving Vulnerable Road Users' Safety Based on Computer-Vision Technologies and High-fidelity Hybrid Traffic Forecasting Tool for Urban Transportation Networks using Emerging Datasets, Volume 2

Samiul Hasan, Ph.D.

Naveen Eluru, Ph.D.

Tanmoy Bhowmik, Ph.D.

Rezaur Rahman, Ph.D.

Jiechao Zhang

Sudipta Dey Tirtha

Kamol C. Roy, Ph.D.

University of Central Florida

Department of Civil, Environmental & Construction Engineering

Orlando, FL 32816-2450

UNIVERSITY OF CENTRAL FLORIDA

December 2021

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| BDV24 562-10 | | |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| Connecting the East Orlando Communities Project - Phase 2 – Volume II | December 31, 2021 |
| | 6. Performing Organization Code |
| **7. Author(s)**<br>Samiul Hasan; Naveen Eluru; Tanmoy Bhowmik; Rezaur Rahman; Sudipta Dey Tirtha; Jiechao Zhang; Kamol C. Roy | |
| **9. Performing Organization Name and Address**<br>Department of Civil, Environmental & Construction Engineering<br>University of Central Florida<br>12800 Pegasus Drive, Suite 211<br>Orlando, FL 32816-2450 | **10. Work Unit No. (TRAIS)**<br><br>**11. Contract or Grant No.** |
| **12. Sponsoring Agency Name and Address**<br>Florida Department of Transportation | **13. Type of Report and Period Covered**<br>Final Deliverable<br><br>14. Sponsoring Agency Code |
| 15. Supplementary Note | |

16. Abstract

This project has developed a high-fidelity hybrid traffic forecasting tool that takes intersection-level historical traffic movement data and travel demand observations at various spatio-temporal resolutions and employs host of variables as inputs to predict future traffic movements. Traditional transportation planning models forecast traffic from a strategic perspective over a longer-term horizon. These models are also not suited to address high-resolution network structure challenges (e.g., signal times and signal bays) — failing to account for bottlenecks and hidden bottlenecks in the network, often at intersections. To develop the traffic forecasting tool, various datasets have been collected from Seminole County as the study area. These datasets include Automated Traffic Signal Performance Measures (ATSPM) data, Central Florida Regional Planning Model (CFRPM) data to generate the spatio-temporal travel demand, and various socio-demographic, built environment, and land-use characteristics. A graph convolution based deep neural network architecture has been developed to learn traffic movement volume from intersection-level features and trained over one year (2016) of data. The developed model has achieved $R^2$ scores greater than 0.90 for predicting each movement type for the test year (2019) data. To visualize model results, a web-based data visualization platform has been developed. The platform provides intuitive information on the quality of the ATSPM data and validation and prediction results for each intersection included in the model.

| 17. Key Word | 18. Distribution Statement |
|---|---|
| Transportation Network, Machine Learning, Deep Learning, Volume Prediction, Visualization Platform, ATSPM. | |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| | Unclassified | 120 | |

## **EXECUTIVE SUMMARY**: Volume II

The goal of this project is to develop a high-fidelity hybrid traffic forecasting tool that takes intersection-level historical traffic movement data and travel demand observations at various spatio-temporal resolutions and employs host of variables as inputs along to predict future traffic movements. Traffic forecasting for a large-scale network at an intersection level is challenging due to higher computational complexity incurred by the network topology, which requires a robust prediction model, (i) with the ability to capture the spatial correlation of traffic in interconnected roads and (ii) able to predict traffic for a longer term (3-5 years).

To develop the high-fidelity hybrid traffic forecasting model, various of datasets have been collected. *First*, in Seminole County, most of the signalized intersections are equipped with advanced traffic signal controllers on the arterials and each signal provides Automated Traffic Signal Performance Measures (ATSPM). The research team has collected the raw data of the ATSPM signals for years 2015, 2016 and 2019. *Second*, the model also depends on the zonal and sub zonal level features related to traffic demand for high fidelity traffic prediction. Hence, the research team has collected GIS data of Traffic Analysis Zones (TAZs) and compiled Central Florida Regional Planning Model (CFRPM) data to generate the spatio-temporal travel demand. *Third*, a comprehensive set of exogenous variables including socio-demographic, built environment, and land-use characteristics has been considered. Information about these variables is collected from different data sources including American Community Survey (ACS), US Census Bureau, NAVSTREET street data, and Florida Department of Revenue (FDOR) databases.

To develop the traffic forecasting tool, a graph convolution based deep neural network architecture has been developed to learn traffic movement volume from intersection-level features such as volume, zonal or sub zonal level trip attraction and production. The results indicate that, for test year 2019, the $R^2$ score of each movement type is greater than 0.90 indicating that the model can capture the variations of traffic movement volumes well. For each movement types, more than 85% of the predicted volumes have a GEH score less than 5. To better predict the traffic volume for projection year 2023, we have retrained the model with 2019 traffic volume data. We then use the model to predict 2023 traffic volume. The results show that compared to 2019 traffic volume, the 2023 traffic volume will have a 14% - 20% growth in a typical workday and 5% - 9% growth in a typical weekend.

To visualize the results of the model, we have developed a web-based data visualization platform. The visualization platform can be used for providing more intuitive information of the ATSPM dataset to anyone using the data. This can also be used to check the quality of the data and improve the maintenance of the detectors. The platform is divided into three parts: Introduction, Data Quality and Traffic Prediction. In the Introduction part, we briefly introduce the projects problems and the solutions. The Data Quality part is mainly used for providing more intuitive information of the dataset which can be used to check the data quality and improve the maintenance of the detectors. In the Traffic Prediction part, we present the prediction results of the developed traffic forecasting framework.

# CHAPTER 1 INTRODUCTION

Due to the high population and economic growth, Florida's urban regions are facing increasing traffic congestion and other mobility related challenges such as traffic accidents, excessive fuel consumption etc.  To overcome such challenges, we need robust methods for accurate and reliable prediction of traffic to ensure the success of capacity improvement and traffic management projects. However, traditional transportation planning models forecast traffic from a strategic perspective over a longer-term horizon. In addition, traditional forecasting frameworks are not suited to address high-resolution network structure challenges (such as signal times and signal bays) — failing to account for bottlenecks and hidden bottlenecks in the network, often at intersections. In fact, employing traditional methods for identifying capacity needs between intersections could result in recommendations that are unwarranted or potentially less efficient.

Ubiquitous use of sensing technologies such as roadway detectors and GPS devices has created a new opportunity to overcome these challenges and implement deployable modeling approaches to understand network-level traffic dynamics for traffic forecasting. In this context, developing a hybrid framework that improves the traditional forecasting system with a high-fidelity (at the level of an intersection) model of traffic forecasting can better serve the decision-making related to transportation network improvement projects. The developed forecasting framework will need to reflect the influence of land use, demographics, and socio-economic characteristics on travel demand generation while accommodating for high resolution network structure in modeling traffic flows.

The goal of Volume II of the project is to develop a high-fidelity hybrid data-driven traffic forecasting tool that takes intersection-level historical traffic movement data and travel demand observations at various spatio-temporal resolutions and employs host of variables as inputs along to predict future traffic movements. Traffic forecasting for a large-scale network at an intersection level is challenging due to higher computational complexity incurred by the network topology, which requires a robust prediction model, (i) with the ability to capture the spatial correlation of traffic in interconnected roads and (ii) able to predict traffic for a longer term (3-5 years).

This part of the project has the following objectives:

1. Develop a hybrid machine learning framework for high-fidelity traffic forecasting utilizing traffic sensors, road capacity, socio-economic, and land use data;
2. Develop a data pipeline that can feed a range of datasets and deliver forecast outputs to a visualization application; and
3. Create map-based visualization outputs providing traffic flow information (such as volume by link direction) to show future traffic forecasts.

# CHAPTER 2 A REVIEW OF DATA-DRIVEN TRAFFIC PREDICTION METHODS

## 2.1 INTRODUCTION

Optimization of a transportation network largely depends on how accurately we can capture traffic dynamics of a network for the purpose of traffic forecasting. Accurate traffic forecast can ensure proactive decision making and optimal action plans for traffic operations and management. However, network level traffic forecasting is a complex process that involves modeling traffic dynamics to represent the interactions between the demand and supply side models.

Traditionally, network modeling approaches (e.g., static/dynamic traffic assignment) rely on mathematical models to estimate link-level traffic flows. These methods rely on several assumptions of user behavior and knowledge such as: (i) drivers have precise information and knowledge about the underlying network; (ii) drivers make rational choices when choosing a route; and (iii) all drivers are homogeneous. Although some of these assumptions may not hold in a real-world scenario, these methods provide the most reasonable solutions to the traffic assignment problem (the final step of a four-step travel demand model). In addition, these assignment methods are not designed to incorporate large-scale high-resolution data available in transportation networks to capture movements at an intersection.

Emerging sensing technologies such as mobile phones, GPS observations, and automated signal performance measures along with traditional traffic sensors such as loop detectors, magnetic vehicle detection, Bluetooth detector etc. have created a new opportunity for us to deal with the challenges of making high-fidelity predictions for a large-scale network using big data. Recent trends show that researchers are more focused on developing data-driven methods to capture complex non-linearities in traffic dynamics. In addition, data-driven methods are easy to deploy in a real-time context, which will eventually help us make proactive decisions for efficient traffic management.

In past few years, data-driven approaches have gained more attention mainly for three reasons: (i) collecting detailed parameters required for model driven approach is tedious; (ii) the concept of smart city has been reshaping the transportation infrastructure with extensive deployment of traffic sensors and advances in data processing technologies and computational power can deal with big data challenges; and (iii) data-driven approaches do not rely on unrealistic behavioral assumptions, hence, are more applicable for solving real-world practical problems.

This chapter presents a review of the existing data-driven traffic prediction methods and their application areas to solve different transportation problems.

## 2.2 DATA-DRIVEN METHODS IN TRAFFIC PREDICTION

A data-driven approach predicts the future traffic state from the current traffic state and historical traffic patterns, without a detailed representation of network dynamics based on traffic flow mechanisms. Data-Driven traffic prediction methods can be grouped into two categories: parametric and non-parametric approaches (Oh et al., 2017) (see **Figure 2.1**). In the following sections, we describe existing data-driven methods applied for traffic prediction.

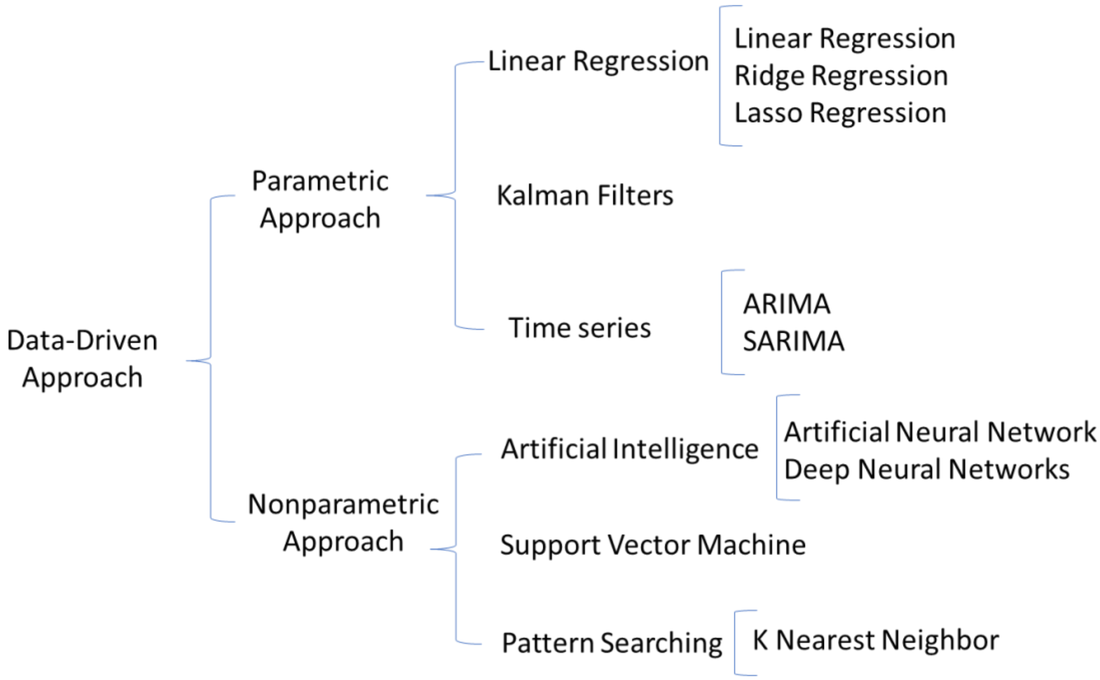



**Figure 2.1: Types of Data-Driven Approaches for Traffic Prediction**

## 2.3 PARAMETRIC METHODS

Parametric models in traffic state prediction aim to find a parametric formulation between known variables such as the previous traffic flow on a segment, and the target variable such as the future flow on the given segment. Parametric methods include Autoregressive Integrated Moving Average method (ARIMA) (Billings and Jiann-Shiou, 2006), Seasonal Autoregressive Integrated Moving Average method (SARIMA) (Williams et al., 2004) and Kalman filter (Chu et al., 2005; Ojeda et al., 2013).

ARIMA model is one of the most popular parametric models due to its simplicity and good performance in forecasting linear and stationary time series, which basically includes a linear function of past observations and random error terms (Billings and Jiann-Shiou, 2006) . SARIMA is an extension of ARIMA model where a seasonality variable is introduced to improve the prediction accuracy for time series prediction. Both ARIMA and SARIMA consider linear relationships between time-lagged variables. However, traffic dynamics is a non-linear phenomenon, hence these models fail to capture the complexity in predicting traffic speed, travel time, and flow variations.

Kalman Filter (KF) based approaches work in a two-step process: first, the prediction step, where it produces estimates of the current state variables and associated uncertainties; second, it tracks the next measurement and the estimates are updated using a weighted average technique, lower weights are given to the estimates with higher uncertainty. The main advantage of a KF is that the recursive framework ensures traffic data is efficiently updated using only data from previous states and not the entire history (Chu et al., 2005). Apart from the popular ARIMA and Kalman Filter based models, other parametric approaches for traffic state prediction include Linear Regression, Ridge Regression and Lasso Regression (Haworth et al., 2014; R. Yu et al., 2017). Due to simplicity, the parametric approaches are highly scalable and transferable, however, they rely on assumption of the data distribution, for example, the ARIMA based model assumes gaussian distribution for the error terms, this assumption may not applicable for nonlinear traffic state variations.

## 2.4 NON-PARAMETRIC METHODS

Non-parametric models have been a popular emerging approach for traffic prediction. Unlike parametric models, non-parametric models do not require any assumptions about the distribution of the data. Nonparametric data-driven methods are divided into two groups: traditional approaches and deep learning approaches. In the following sections, we describe these methods in detail.

### 2.4.1 Traditional Methods

Support Vector Machine (SVM) (Ahn, 2016; C. Wu et al., 2004), K-Nearest Neighbor (KNN) (Cai et al., 2016; Habtemichael and Cetin, 2016; Meng et al., 2015; Myung et al., 2011; Qiao et al., 2013; Yu et al., 2016), and Artificial Neural Network (ANN) (Innamaa, 2005; Lee, 2009; Park et al., 1999; Yu et al., 2008) are some of the most popular non-parametric methods. Although support vector machine was introduced for solving classification problems, it also gained popularity in traffic prediction for its ability to perform well in high dimensional space. SVM is more flexible to define how much error is acceptable in a model. It finds an appropriate line (or hyperplane in higher dimensions) within the error boundary to fit the data. However, one of the major issues with SVM is that it is more sensitive to noise inside the data (Table 2.1). Moreover, it is computationally more expensive compared to other traditional methods.

KNN and ANN are two of the most widely used nonparametric methods for traffic prediction. KNN model relies on matching similar historical patterns with the current one by searching existing data samples and estimate simple average or weighted average to predict the future traffic state. While ANN models recognize and learn the traffic pattern using nonlinear functions, this model is more robust compare KNN in learning traffic patterns, however, due to vanishing gradient problem, the model performance deteriorates in case of long sequence prediction.

### 2.4.2 Deep Learning Methods

Accurate traffic forecasting is one of the major challenges in transportation system management and operation. In general, the complexities in traffic forecasting incurs due to the presence of sharp non-linearities caused by transitions among free flow, breakdown, recovery, and congestion (Polson and Sokolov, 2016). Recently, deep learning has created a new opportunity for the researchers to deal with such nonlinearities in traffic data. Deep learning is a recent non-parametric neural network-based approach for traffic state prediction. The core architecture for modern deep learning methods is based on classical artificial neural networks (ANNs). ANNs are versatile, powerful, and scalable which makes them ideal to tackle large and highly complex machine learning tasks.

Among different architecture for deep learning models, Feed Forward Neural Network (FFNN) has the simplest structure composed of one input layer, two or more hidden layers and one final output layer. However, due to vanishing gradient problem FFNN is weak in processing sequential traffic data. Recurrent Neural Network (RNN) can process sequential data very well (Xu et al., 2017). The basic concept of an RNN is that it stores relevant parts of the input variables and use this information to predict output in the future. RNNs repetitively perform the same computational operation on every element of a sequence and each output is calculated based on the previous computations. Although RNNs can better capture nonlinearity in time series problems, they are weak on learning long-term dependencies due to vanishing of gradient during the backpropagation process (Gers and Cummins, 1999; Hochreiter and Urgen Schmidhuber, 1997) Moreover, traditional RNNs learn a time series sequence based on a predetermined time lag, but it is difficult to find an optimal time window size in an automatic way (Gers and Cummins, 1999; Ma et al., 2015)

To overcome the disadvantages of RNNs, Hochreiter and Schmidhuber proposed the architecture of Long Short-Term Memory Neural Network (LSTM-NN) and an appropriate gradient-based algorithm to solve it (Hochreiter and Urgen Schmidhuber, 1997). The primary objectives of LSTM-NN are to capture long-term dependencies and determine the optimal time lag for time series problems.  LSTM model can capture sharp discontinuities in traffic flows using a combination of non-linear functions (tanh, sigmoid etc.) with gating operation to store valuable information for long data sequences. Hence, applications of LSTM model in transportation have allowed us to deal with more complex problems and big data challenges.

Recently, LSTM and different variants of LSTM models have been used in numerous short-term traffic state prediction problems, such as, traffic speed prediction (Cui and Wang, 2017; Epelbaum et al., 2017; Ma et al., 2015), travel time prediction (Yanjie Duan et al., 2016), traffic flow prediction (Luo et al., 2019; Polson and Sokolov, 2016; Yang et al., 2019), vehicular queue length prediction (Lee et al., 2019; Rahman and Hasan, 2019). In addition to regular traffic condition, studies have also explored the application of LSTM for traffic management during a major event such as hurricane evacuation (Rahman and Hasan, 2018a) and traffic incidents (R. Yu et al., 2017).

LSTM models are designed to learn temporal correlations from sequences of data, and have been widely applied for traffic state prediction for arterials and freeways. However, for large scale network we need to capture both temporal and spatial dependency in traffic state variables. Since, road networks have interconnected road segments, there exists a strong correlation in traffic volumes and speed from nearby road segments. Convolutional Neural Networks can capture this spatial dependency among different traffic state variables; however, in some cases they cannot process the sequential time series data very well, hence researchers have combined convolutional neural network and LSTM model to capture both the spatial and temporal dependency among traffic states (Ma et al., 2017; H. Yu et al., 2017).

Convolutional LSTM neural networks (ConvLSTM) are a promising state-of-the-art approach for network level traffic forecasting. In ConvLSTM model, traffic network is considered as image and network information are extracted using convolution filter, which are then fed into LSTM model to predict future traffic state. While traffic network is converted into an image, a certain amount of noise is included into the spatial relationships of traffic state causing erroneous results in prediction. Moreover, ConvLSTM cannot capture the long-term congestion propagation inside the network for long-term traffic forecasting. To solve these problems, several studies (Cui et al., 2018; Li et al., 2017; Yu et al., 2018) learn the traffic network as a graph and adopt the graph-based convolution operator to extract features from the graph-structured traffic network.

Graph Convolutional Neural Networks (GCNN) generalize traditional neural networks to work on structured graphs (Kipf and Welling, 2017). GCNNs utilize the adjacency matrix of the road network. The adjacency matrix-based graph convolution neural networks incorporate the node adjacency matrix, whose network structures are more flexible. In a traffic prediction problem, traffic networks are considered as directed graph and network features have been processed and learnt by either diffusion graph convolution (Li et al., 2017) or the spectral graph convolutions (Yu et al., 2018) operations, finally the sequential time dependent correlation is captured with LSTM layer. A summary of different data driven methods are shown in **Table 2.1**.

**Table 2.1: A Summary of the Existing Data-Driven Methods**

| Methods | Pros | Cons |
|---|---|---|
| Linear Regressions (Haworth et al., 2014; R. Yu et al., 2017) | 1. Simple and flexible<br>2. Easy to interpret | 1. Cannot capture the spatial features<br>2. Cannot model the nonlinear spatial and temporal relationships |
| ARIMA(Billings and Jiann-Shiou, 2006) SARIMA (Williams et al., 2004) | 1. More interpretable, less flexible<br>2. Works best on long and stable series | 1. Cannot capture the spatial features<br>2. Not appropriate for non-stationary data<br>3. Cannot model the nonlinear spatial and temporal relationships |
| Kalman Filter (34, 35) | 1. Robust to noisy data<br>2. Work for both stationary and nonstationary data<br>3. Computationally efficient, suitable for real-time traffic flow prediction | 1. Cannot capture the spatial features<br>2. Cannot model the nonlinear spatial and temporal relationships |
| KNN (Qiao et al., 2013; Yu et al., 2016) | 1. Computationally efficient, suitable for real-time traffic flow prediction | 1. Not interpretable<br>2. Cannot deal with high dimensional data<br>3. Cannot capture the spatial features<br>4. Cannot model the nonlinear spatial and temporal relationships |
| Support Vector Machine (SVM) (Ahn, 2016; C. Wu et al., 2004) | 1. More powerful<br>2. Works well in high dimensional space | 1. Computationally expensive<br>2. Difficult to choose optimal parameters<br>3. Highly sensitive to noise data |
| ANN (Yasdi, 1999) | 1. More powerful and scalable<br>2. Ideal to tackle large and highly complex machine learning tasks. | 1. Cannot model the nonlinear spatial and temporal relationships<br>2. Due to vanishing gradient problem, the model performance deteriorates in case of long sequence prediction |
| LSTM (Cui and Wang, 2017; Rahman and Hasan, 2018b; Yang et al., 2019) | 1. Accounts for long-term dependencies<br>2. Accounts for the spatial and temporal correlations | 1. Parameters are hard to interpret and visualize<br>2. Cannot capture the unique spatial features |
| CNN (Ma et al., 2017) | 1. Accounts for spatial correlations by using a spatial matrix | 1. Fails to capture temporal correlations<br>2. Cannot capture state transition in time series data |
| Convolutional LSTM (Ai et al., 2019; Cui et al., 2018; H. Yu et al., 2017) | 1. Accounts for the spatial and temporal correlations<br>2. Deals with high dimensional transportation network data | 1. Parameters are hard to interpret and visualize<br>2. Considers transportation network as image, cannot capture state transition in spatial traffic flow<br>3. Comparatively complex and computationally expensive |

| | | |
|---|---|---|
| Graph Convolution (Li et al., 2017; Yu et al., 2018) | 1. Accounts for the spatial and temporal correlations<br>2. Deals with high dimensional transportation network data<br>3. Explicitly captures the stochastic nature of traffic dynamics | 1. More complex and computationally expensive<br>2. Parameters are hard to interpret and visualize |

## 2.5 APPLICATION AREAS FOR DATA-DRIVEN METHODS

In this section, we discuss the applications of data-driven traffic prediction methods. Considering the project scope, the application areas are divided into four categories: (i) Corridor level traffic prediction, (ii) network level traffic prediction, (iii) traffic prediction at an intersection, and (iv) annual average daily traffic (AADT) prediction for different types of areas.

### 2.5.1 Traffic Prediction at A Corridor Level

Several methods have been proposed for traffic forecasting in freeways or interconnected highway segments. These involve forecasting traffic in terms of speed (Cai et al., 2016; Ma et al., 2015; Polson and Sokolov, 2017; Rahman and Hasan, 2018a), volume (Luo et al., 2019; Yasdi, 1999) , and travel time (Billings and Jiann-Shiou, 2006; Chien and Kuchipudi, 2003; Elhenawy et al., 2014; Innamaa, 2005; Myung et al., 2011; C. H. Wu et al., 2004; Zhang and Haghani, 2015). However, in almost all the cases the traffic prediction problem is formulated as a simple time series prediction problem that disregards the spatial dependencies among traffic state variables. Also, they do not consider the variations in travel demand when predicting the future traffic. Hence, these applications are mostly limited to short-term (less than 1 hr) traffic prediction. A summary of the studies predicting traffic at a corridor level is presented in Table 2.2. From the table, we find that corridor-level travel time and traffic speed prediction have gained more attention compared to traffic flow prediction. Moreover, recent trends in solving data-driven traffic prediction problem showed that studies are more focused on application of neural network-based approaches such as LSTM, Feed Forward Neural Network etc. compared to traditional approaches.

**Table 2.2: A Summary of Studies for Traffic Prediction at a Corridor Level**

| Authors | Methods | Predicted (or Target) Variable | Study Scope | Input Data Type | Prediction Horizon |
|---|---|---|---|---|---|
| (Yasdi, 1999) | ANN | Traffic volume | Single highway segment | Loop detector | weekly, daily, and hourly prediction |
| (Chien and Kuchipudi, 2003) | Kalman Filtering | Travel time | Multiple freeway segments | Roadside Terminals: Vehicle Tag Identification | 5-min interval |
| (C. H. Wu et al., 2004) | Support vector regression. | Travel time | Highway segments with variable lengths | Loop detector | 3-min interval for peak hours |
| (Innamaa, 2005) | ANN | Travel time | Inter-urban highway | Video Camera and Loop Detector | 3-5 min interval |
| (Billings and Jiann-Shiou, 2006) | ARIMA | Travel time | Arterial with 10 signalized intersections. | GPS-equipped probe vehicles | 15-20 min interval |
| (Myung et al., 2011) | KNN | Travel time | Freeway segments | VDS and Automatics Toll Collection | 5-min interval |
| (Qiao et al., 2013) | KNN | Travel time | Freeway segments | Bluetooth Detector | 5-min interval |
| (Elhenawy et al., 2014) | Genetic programming and clustering | Travel time | Freeway segments | GPS-equipped probe vehicles | 5-min interval |
| (Zhang and Haghani, 2015) | Gradient Boosted Regression Tree | Travel time | Freeway segments | GPS-equipped probe vehicles | 5-min interval |
| (Ma et al., 2015) | LSTM | Speed | Freeway segment | MVDS | 2-8 min |
| (Cai et al., 2016) | KNN | Speed | Mixed roadway types | Floating Car Speed | 5-60 min |
| (Polson and Sokolov, 2017) | Deep Neural Network | Speed | Freeway segments | Loop detector | 5-40 min |

| | | | | | |
|---|---|---|---|---|---|
| (Rahman and Hasan, 2018a) | LSTM | Speed | Freeway segments | MVDS | 5-15 min |
| (Luo et al., 2019) | KNN and LSTM | Traffic flow | Four expressway segments | Loop detector | 5-min interval |

### 2.5.2 Large Scale Network Level Traffic Prediction

Understanding traffic evolution and congestion propagation for an entire road network rather than on a single road will be more helpful for the traffic manager in proactive decision making (Zhang et al., 2011). However, large-scale network level traffic forecasting is more challenging due to higher computational complexity incurred by the network topology, which requires a robust prediction model, with the ability (i) to capture the spatial correlation of traffic in interconnected roads and map it to a two-dimensional plane, and (ii) to predict traffic for a long-term to reflect congestion propagation. But traditional traffic prediction models treat traffic state variables as sequential data, thus cannot deal with high dimensionality of the data to learn spatial correlation.

Convolutional LSTM methods are the initial attempt to model spatial and temporal correlation among the traffic states for network level traffic prediction. A few studies (Ma et al., 2017) Cui and Wang, 2017) have implemented the convolutional LSTM model for network level traffic speed prediction. Although this model outperforms existing state of art data-driven model, however it fails to capture the stochastic nature of traffic dynamics of a transportaion network.

Recently, graph convolution neural network has been emerging as a new approach to overcome the challenge of dimensionality when dealing with large scale network level travel demand prediction.  Graph convolutional neural network approaches utilize the concepts of graph theory along with deep neural network architectures to model the stochastic traffic dynamics inside a network. These approaches aim at learning the interactions between roadways in the traffic network to forecast network-wide traffic states. However, the application of such a neural network architecture hardly exists for a large-scale transportation network. As shown in **Table 2.3**, several studies have applied the concept of graph convolution to represent traffic network as a generalized graph for traffic state prediction (Cui et al., 2018; Li et al., 2017). These studies have focused on learning short term correlations among network wide traffic states such as speed to predict traffic states 5 to 20 min ahead of current time.

**Table 2.3: A Summary of Data-driven Studies for Predicting Traffic at a Network Level**

| **Authors** | **Methods** | **Predicted (or Target) Variable** | **Study Scope** | **Input Data Type** | **Prediction Horizon** |
|---|---|---|---|---|---|
| (Ma et al., 2017) | Convolution Neural Network | Speed | Two small scale sub networks | GPS probe vehicle | 10-20 min |
| (Li et al., 2017) | Diffusion Convolutional LSTM | Speed | Small highway network | Loop detector | 15-60 min |
| (Cui and Wang, 2017) | Convolutional LSTM | Speed | Small freeway network | Loop detector data | 5 min |

### 2.5.3 Traffic Prediction at Intersections

Data-driven methods for intersection level traffic prediction mostly involve traffic flow (Alajali et al., 2018), movement volume prediction (Li et al., 2020) and traffic signal queue length prediction (Chang and Su, 1995; Lee et al., 2019; Rahman and Hasan, 2020) (**Table 2.4**). These studies are highly data intensive; hence they either use a simulation-based approach or hybrid approach to develop intersection-level traffic prediction models. For instance, Chang and Su (Chang and Su, 1995) developed a data-driven neural network model for predicting queue length at short time step (3s). They used the data from simulation experiments to train the model for queue prediction. Lee et al. (Lee et al., 2019) developed a deep learning model for queue length estimation. They relied on traffic simulations to generate the training data and used real-world driving data from the Federal Highway Administration's Next Generation Simulation (NGSIM) program to test the approach. However, one limitation is that these approaches do not consider the coordination among multiple intersections; these models are based on isolated intersection.

Recently, (Alajali et al., 2018) applied gradient boosted decision tree based model to predict intersection traffic volume for large scale network which covers intersections at central business district (CBD) area of Melbourne, Australia. This study proposed an online and offline training approach to deal with the limitation in computation power for large scale data. However, the proposed method was only limited to aggregated traffic volume prediction at intersection level rather than traffic movement volume. A recent study by (Li et al., 2020) proposed a deep learning method to predict intersection level traffic movement volume. This study utilizes Convolutional LSTM model to capture spatio-temporal dependency among network-wide traffic states. However, the proposed approach does not consider travel demand variation, hence only limited to short term traffic movement prediction (i.e., 5-15 min ahead of current time).

**Table 2.4** summarizes the existing methods for intersection level traffic prediction. We find that most of studies adopt deep learning models, while some studies consider micro-level directional traffic volume. Moreover, since deep learning models capture the spatio-temporal dependency of interconnected roadway segments at an intersection, hence these models outperform the traditional models for intersection level traffic prediction.

**Table 2.4: A Summary of the Studies Covering Traffic Prediction at Intersection/s**

| Authors | Methods | Predicted (or Target) Variable | Study Scope | Input Data Type | Prediction Horizon |
|---|---|---|---|---|---|
| (Chang and Su, 1995) | ANN | Queue length | Isolated intersection | Micro Simulation | 3s - 9s |
| (Alajali et al., 2018) | Gradient Boosted Decision Tree | Intersection traffic volume | Intersections at CBD area Melbourne Australia | Traffic sensor data, Traffic incident data | 15-min interval |
| (Lee et al., 2019) | ConvLSTM | Lane-based queue length | Isolated intersection | Aimsum Simulation and NGSIM data | 100s cycle length |
| (Li et al., 2020) | ConvLSTM | Traffic movement volume | Three connected signalized intersections | Loop Detector | 15 min interval |
| (Rahman and Hasan, 2020) | LSTM | Lane based queue length | Nine connected intersections at an arterial corridor | Adaptive Traffic Signal control | 120s cycle length |

### 2.5.4 Annual Average Daily Traffic Prediction

Traffic demand generation is a fundamental factor in transportation applications, which can enhance the success of capacity improvement and traffic management plans. Traffic demand volume is estimated as Annual Average Daily Traffic (AADT), as the average daily traffic flow at a specific location over a year (Zhao and Park, 2004). However, due to the high cost of developing and maintaining the traffic monitoring equipment such as Automatic Traffic Counters (ATCs), the coverage of AADT values in the roads is always limited by the counting techniques (Transport, 2013, Transport, 2014). Thus, modeling and predicting AADT has received extensive research interest in the past decade (Zhao and Chung, 2001, Zhao and Park, 2004, Eom et al., 2006, Fu et al., 2017, Sharma et al., 2001). A wide range of algorithms have been applied into the prediction of AADT, which mainly contains three different types – linear regression (Pun et al., 2019, Zhao and Chung, 2001, Doustmohammadi and Anderson, 2016, Xia et al., 1999, Shen et al., 1999), spatial statistics (Eom et al., 2006, Shamo et al., 2015, Wang et al., 2013, Ma et al., 2019, Zhao and Park, 2004) and machine learning (Pun et al., 2019, Castro-Neto et al., 2009, Gastaldi et al., 2013, Sfyridis and Agnolucci, 2020, Wu and Xu, 2019, Sharma et al., 2001, Khan et al., 2019, Khan et al., 2018, Fu et al., 2017, Gastaldi et al., 2014, Duddu and Pulugurtha, 2013).

In terms of linear regression models, Xia et al. (Xia et al., 1999) developed a multiple linear regression model to estimate the AADT for nonstate roads in urbanized areas in Florida using the traffic data extracted from 450 count stations with 12 different predictors. The study indicates that the most important variables in predicting the AADT are road characteristics, such as lane number, function and area type. However, socioeconomic features such as population and automobile ownership show less importance in the model. Since the area type and size have significant influence on the estimation of AADT, Zhao et al. (Shen et al., 1999) applied the linear regression model in four different types of areas – state wide, rural, small-medium urban and large metropolitan area. The results indicate that the linear regression model can work better on the small-medium urban and large metropolitan areas. The poor performance of the model in statewide and rural areas reveal that adequate variables can improve the predictive power of the models. In the study (Doustmohammadi and Anderson, 2016), using more detailed variables combining the roadway characteristics and socioeconomic features within a quarter-mile buffer of a small-medium community, the researchers developed a linear regression model with high accuracy. Besides, Zhao and Chung (Zhao and Chung, 2001) further developed a linear regression model with accessibility measurements based on travel time to the regional center in the entire Broward County in South Florida.

However, linear regression model cannot capture the spatial dependencies when predicting the AADT which limit the predictive power in the network with spatial correlations of different areas (Shamo et al., 2015). To consider the spatial dependency in the model, Zhao and Park (Zhao and Park, 2004) proposed a geographically weighted regression methods (GWR) to predict AADT which can estimate the model parameters locally rather than globally showing accuracy than the linear regression models in Broward County, Florida. Besides, researchers (Eom et al., 2006, Selby and Kockelman, 2011) developed a spatial statistical method called kriging to predict the AADT in nonfreeway and highway network. Shamo et al. (Shamo et al., 2015) further applied three kriging methods – simple kriging, ordinary kriging and universal kriging to determine the most suitable predictive model. There are also limitations with kriging since it uses the correlation function as the sole predictor of spatial characteristics which limits the predictive power of the methods in predicting complex transportation network (Ma et al., 2019). To overcome the limitations of kriging, Ma et al. (Ma et al., 2019) proposed a copula-based model combining the spatial dependency and marginal distribution to predict the AADT in California state highway network which shows higher accuracy than the traditional kriging models.

In recent years, machine learning models such as random forest (Pun et al., 2019, Sfyridis and Agnolucci, 2020) and artificial neural network (Gastaldi et al., 2013) have been widely applied for predicting AADT. Sharma et al. (Sharma et al., 2001) developed a neural network model to predict the AADT of low-volume roads in Alberta, Canada with the predictors of road classification and trip purposes. Castro-Neto et al. (Castro-Neto et al., 2009) used a support vector regression model (SVR) to estimate the AADT for both urban roads and rural roads in Tennessee using nearly 20 years of historical AADT data. To better capture the spatial characteristics, Gastaldi et al. (Gastaldi et al., 2013) applied fuzzy set theory to represent the fuzzy boundaries of road class when predicting the AADT in the province of Venice, Italy with neural networks. Sfyridis and Agnolucci (Sfyridis and Agnolucci, 2020) further developed a support vector regression (SVR) and random forest (RF) to estimate AADT with more comprehensive variables such as socioeconomic and land use characteristics in England and Wales across the full road network. Wu and Xu (Wu and Xu, 2019) compared the performance of multiple regression model, random forest model and neural network with the functional class and the lane number of road network for both minor roads and major roads showing that the neural network outperforms the other two prediction models. Khan (Khan et al., 2018) used SVR and artificial neural network (ANN) to predict AADT for different functional class roadway in South Carolina. It shows that the SVR has higher accuracy than ANN across all types of roadway. Khan et al. (Khan et al., 2019) further applied deep learning models – simple recurrent neural network (RNN), gated recurrent neural network model and the long short-term memory model – to predict the AADT in highway network in South Carolina.

In summary, the previous research modelled the annual average daily traffic in different types of areas, such as arterial, highway and the entire traffic network. It indicates that the socioeconomic, demographic and roadway characteristics have significant influence on the AADT prediction.

**Table 2.5: A Summary of the Studies for AADT Prediction**

| Authors | Methods | Predictors | Area |
|---|---|---|---|
| (Al-Masaeid et al., 1998) | Linear Regression | Socioeconomic and demographic | Full road networks in Mafraq City |
| (Xia et al., 1999) | Linear Regression | Roadway characteristics and socioeconomic variables | Nonstate roads in Broward County, Florida |
| (Shen et al., 1999) | Linear Regression | Roadway characteristics and socioeconomic variables | Statewide area, rural area, small-medium urban area and large metropolitan area |
| (Zhao and Chung, 2001) | Linear Regression | Roadway characteristics, socioeconomic | Interstate highways, expressways, urban |

| | | | |
|---|---|---|---|
| | | characteristics, accessibility to expressways and regional accessibility to employment | roads and rural roads in South Florida |
| (Doustmohammadi and Anderson, 2016) | Linear Regression | Function classification of the road, number of lanes, population, retail employment and non-retail employment | Small or medium sized community |
| (Zhao and Park, 2004) | Geographically weighted regression | Number of lanes, regional accessibility to employment, population and employment in the buffer area and direct access to expressways | Broward County, Florida |
| (Eom et al., 2006) | Kriging | Roadway Characteristics and socioeconomic | Nonfreeway in Wake County, North Carolina |
| (Shamo et al., 2015) | Simple kriging, ordinary kriging and universal kriging | Roadway Characteristics | Highway in Washington state |
| (Ma et al., 2019) | Copula-based model | Roadway Characteristics and adjacent measured AADT data | California state highway network |
| (Castro-Neto et al., 2009) | Support vector regression (SVR) | Roadway Characteristics | Roadway network for both rural area and urban area in Tennessee state |
| (Gastaldi et al., 2013) | Fuzzy set theory and neural network | Volume data | Rural road network in province of Venice, Italy |
| (Sfyridis and Agnolucci, 2020) | Support vector regression (SVR) and random forest (RF) | Road information, socioeconomic characteristics and urban area polygons | Major and minor roads in England and Wales |
| (Wu and Xu, 2019) | Linear regression, random forest and neural network | Functional class and the number of traffic lanes on the major road and minor road | Major roads and minor roads in Washington State |
| (Sharma et al., 2001) | Neural network | Functional class of roadway and land use | Low-volume rural roads in Alberta, Canada |
| (Khan et al., 2019) | Simple recurrent neural network (RNN), gated recurrent model (GRU) and long short-term memory (LSTM) | Functional class of roadway | Interstate, arterial and local roads for both rural and urban area in South Carolina |
| (Khan et al., 2018) | Support vector regression (SVR) and | Functional class of roadway | Interstate and arterial for both |

| | | | |
|---|---|---|---|
| | artificial neural network (ANN) | | rural and urban area in South Carolina |
| (Gastaldi et al., 2014) | Artificial neural network (ANN) | Volume data | Rural road network in province of Venice, Italy |
| (Duddu and Pulugurtha, 2013) | Statistical model and neural network | Demographic characteristics and land use | City of Charlotte, in Mecklenburg County, North Carolina |

## 2.6 DATA-DRIVEN METHODS IN PRACTICE

In practice, most of the transportation projects are focusing on hybrid approaches by combining dynamic traffic assignment (DTA) method with data-driven approaches to forecast traffic state. We reviewed several projects that provide insights on the state-of-the-art methods to develop demand sensitive model for real time traffic forecasting. For instance, (Mahmassani et al., 2014; Serulle et al., 2017) adopted traditional simulation based DTA model to understand the demand variation in transportation network in response to different adverse weather conditions. These studies are highly data intensive and utilize advanced traffic sensors and real time weather information to predict the operating traffic conditions of a network including travel time variations at a higher resolution.

In addition (Bartley et al., 2020) developed a short-term prediction model for a roadway network in metropolitan Perth both with prediction horizon of 15 min and 75 min using the information of traffic conditions, exogenous variables (such as rainfall and holidays) and road characteristics. (Tao et al., 2012) also proposed a prediction model for travel time reliability of highway based on the roadway network and historical traffic information. (Tran and Shahabi, 2020) applied a deep learning model for the traffic flow prediction and bus arrival time estimation in Los Angeles using trajectory data and incidents data (e.g. accidents, traffic hazards, road closures and ramp meters).

Recent trends in traffic operational practices show that transportation agencies are more focused on data-driven approaches to solve different transportation problems, especially real-time traffic forecasting for incident management (**Table 2.6**). This is because data-driven traffic prediction methods can achieve state-of-the-art prediction accuracy with less complexity and computation time, while these methods are more flexible and easier to deploy in real-world applications.

**Table 2.6.: A Summary of Practical Implementations of Traffic Prediction Methods**

| **Authors** | **Sponsors** | **Scope** | **Area** | **Data Type** | **Application** |
|---|---|---|---|---|---|
| (Mahmassani et al., 2014) | US DOT | Weather Responsive Traffic Estimation and Prediction System | - Riverdale road network, UTAH: 80 freeways, 31 highways, 35 ramps and 1012 arterials<br>- Total 510 nodes, 174 zones | - Weather data: Automated Surface Observing, Road Weather Information System<br>- Traffic data: loop detectors, radar detectors, Signal timing plans<br>- Time Dependent OD Matrix: (static or historical OD matrix for the planning time horizon); and time-dependent traffic counts on selected links. | Predict hourly traffic demand at different operating condition |
| (Serulle et al., 2017) | US DOT | Integrated Modeling for Road Condition Prediction | suburban Kansas City study | - Traffic: Kansas City Scout Advanced Traffic Management System detectors<br>- Incident and work zone data: KC Scout's event feed<br>- Hydrology: Storm Watch hydrological stations<br>- Weather data: National Oceanic and Atmospheric Administration (NOAA), National Weather Service's (NWS) National Digital Forecast Database (NDFD) | Predict pavement condition and travel time for different weather condition; updated every 1 min |

| | | | | | |
|---|---|---|---|---|---|
| (Bartley et al., 2020) | PATREC & iMOVE CRC | Short-term Traffic Speed Prediction for Perth Roads | Main roads network in metropolitan Perth | - Traffic conditions: vehicle detection stations data; GPS speed; vehicle counts on freeways; arterial volumes<br>- Exogenous variables: rainfall data; road space bookings; school holidays; public holidays<br>- Road characteristics：IRIS road link definitions | Predict the average speed for road sections with prediction horizon of 15 minutes and 75 minutes |
| (Tao et al., 2012) | Second Strategic Highway Research Program | Forecasting and Delivery of Highway Travel Time Reliability Information | I-66 | - Static data: network; sensor locations<br>- Historical data with real-time updates: sensor data; incident data; weather data; work zone data | Predict travel time for a given route based on both historical patterns and current conditions |
| (Tran and Shahabi, 2020) | USDOT | Deep-Learning Traffic Flow Prediction of Public Transportation Systems | Los Angeles | - Trajectory data<br>- Incidents data: accidents; traffic hazards; road closures; ramp meters | Traffic flow forecasting and bus arrival time estimation in Los Angeles |

### 2.7 SUMMARY

In this chapter, a review is presented on literatures covering existing data-driven traffic prediction methods, their application areas, and practical implementations. The reviews suggest that data-driven modeling for large scale transportation networks are less common. Several studies have proposed spatial and temporal learning-based methods predicting traffic of large-scale networks. However, these methods have not been tested in any realistic networks. Rather they have used several interconnected point-based detectors to create a small network and validated the proposed algorithm on that network. There remain several unanswered questions: Are these methods scalable? How will they process the spatial and temporal dependency of traffic states for a large network? A limitation of the existing data-driven approaches is that they only rely on short-term correlation between past and present traffic states to predict future traffic state. Thus, these approaches will not perform well for solving long-term traffic prediction problems.

Moreover, the reviews also include existing literatures on intersection-level traffic prediction which involves traffic signal queue length prediction, intersection-level aggregated traffic volume prediction, and traffic movement volume prediction. These studies provide critical insights on the challenges to deal with intersection level traffic prediction problem, especially while dealing with a large-scale network. Since deep learning models can capture the spatio-temporal dependency of interconnected roadway segments at an intersection, most of the studies adopt deep learning models while considering micro level directional traffic volume. These approaches outperform the traditional models for intersection level traffic prediction. Additionally, the reviews reveal recent trend in traffic operational practices, which show that transportation agencies are more focused on data-driven traffic forecasting approaches for real-time incident management. From these case studies, it is evident that data-driven approaches are getting more attention mostly because of their flexibility and deployability in real-time applications for traffic prediction.

# CHAPTER 3 A REVIEW OF EXISTING TRAVEL DEMAND MODELS AND THEIR COMPONENTS

## 3.1 INTRODUCTION

Florida's urban regions are facing increasing traffic congestion. The increasing auto travel and its adverse environmental impacts have led in the past decade to the serious consideration and implementation of travel demand management strategies (for example, enhancing existing public transportation services, building new services such as light rail transit, and improving non-motorized infrastructure such as bicycle lanes and sidewalk). The main objective of these demand management strategies is to encourage the efficient use of transportation resources by influencing travel behavior. Travel demand management strategies offer flexible solutions that can be tailored to meet the specific requirements of a particular urban region. Travel demand models offer analytical frameworks to evaluate the effectiveness of the aforementioned management strategies. The broad area of regional travel demand modeling approaches can be classified into two major groups: (1) trip-based modeling approaches and (2) activity-based modeling approaches (also includes tour-based models). Trip-based approaches focus on trips as the basic unit of analysis and activity-based models are focused on individual level activity participation with tours representing the basic element of analysis. While trip-based approaches have been applied in practice for several decades, activity-based models have been extensively applied in practice in the last decade. The two approaches are widely applied for regional travel demand modeling across the US. The inherently higher resolution of the activity-based models (with emphasis on individual level demand prediction), results in a behaviorally rich yet computationally intensive model framework (relative to the trip-based model). While activity-based models are likely to improve sensitivity to policy changes; developing these models are resource intensive. Therefore, to date, trip-based models with various enhancement of the traditional four-step approach remain the most commonly employed framework for transportation planning in urban metropolitan regions.

In this chapter, we review existing literature on different travel demand models (both trip-based and activity-based) and provide a summary of the various model components. Further, to complement these regional travel demand models, we also review alternative less resource intensive methods employed in the literature to generate travel demand outputs. The remaining chapter is organized as follows: the next section describes an overview of trip-based travel demand model with details on the existing travel demand models and its four basic components.

### 3.2 INTRODUCTION TRIP-BASED TRAVEL DEMAND MODELS

The history of demand modeling for planning projects at various levels (such as local, regional, or state level) has been dominated by the traditional travel demand modeling approach referred to as the four-step or the trip-based approach. The traditional trip-based models focus on individual trips and employ a sequential approach of statistical planning with the following four steps - trip generation, trip distribution, mode choice and trip assignment. A schematic diagram providing a brief overview of the trip-based travel demand model steps is provided in figure 3.1.

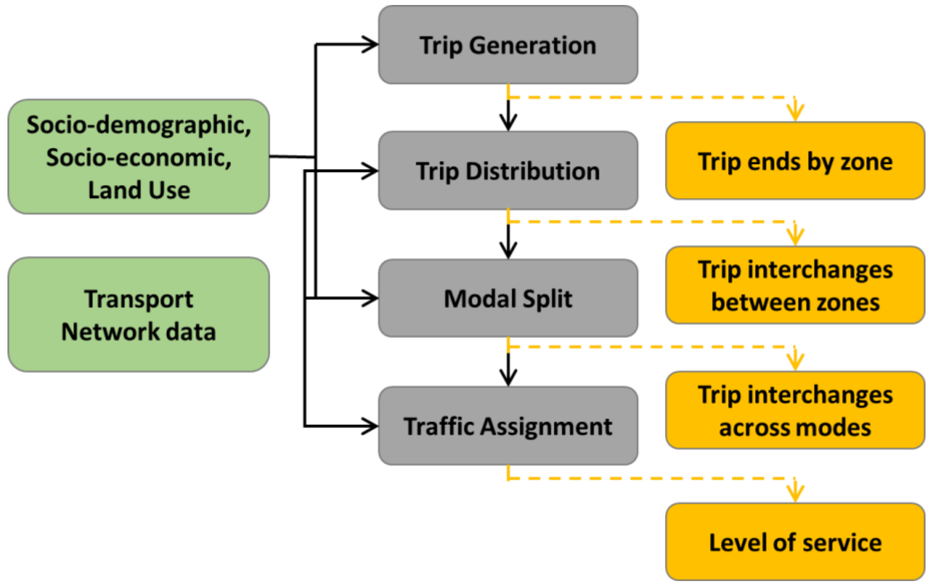



**Figure 3.1: Steps of a Trip Based Travel Demand Model**

The trip generation step involves the estimation of the number of trips produced and attracted by different trip purposes to each spatial unit in the study area. In the trip distribution step, we focus on determining the trip interchanges (such as number of trips from one zone to another, i.e. in each pair of zones). In the mode choice step, we convert the person-trips between each zonal pair as trips by travel mode. Finally, the trip assignment step assigns the vehicle trips to the roadway network to obtain the link level vehicular volumes and travel times. The first three steps of the trip-based method typically constitute the transportation demand side, while trip assignment normally represents the transportation supply side. Thus, the trip-based method accommodates for interaction between transportation demand and supply when executed with full feedback.

### 3.2.1 Existing Literatures

For the sake of brevity and ease of presentation, we have selected eight regional planning models developed for Central Florida, Atlanta, San Diego, Northern New Jersey, North Central Texas, Lincoln, Chicago, and Delaware. To offer an easy to review of earlier research, a concise summary of these regional trip-based travel demand models is presented in tabular format in **Table 3.1**. The table provides information on the study region, dataset used for the analysis, tool used for developing the model, different kind of exogenous variables used for the models, and finally the spatial (at what spatial unit, the travel demand model is developed) and temporal resolution (at what temporal unit, the traffic is predicted) of the studies (see Eluru et al., 2018 for detail).

Several observations can be made from Table 3.1. First, in terms of modeling tool, the common forecasting packages employed across the frameworks is the TransCAD transportation planning software (3 out of 8) while in the Central Florida region, CUBE Voyager is used as the primary transportation modeling tool. Second, four broad categories of independent variables are considered in these regional models including socio-demographic (such as population, and household size), socio-economic (such as employment, and income), land use (such as residential, institutional, and industrial areas) and transportation infrastructure related attributes (such as roadway networks, and transit facilities). Third, the most prevalent spatial unit considered for aggregating the trips is the Traffic Analysis Zones (TAZs) (7 out of 8 regions). Fourth, in terms of the temporal resolution, the traffic was being forecasted by different time periods (morning peak, evening peak, off peak hours) while the models for Atlanta region predicted traffic flow on a daily basis.

### 3.2.2 Component of Trip-Based Travel Demand Models

With respect to modeling components, the traditional four-step modeling process is enhanced with different supplemental model steps for different regions. Therefore, to keep the discussion relevant and brief, we limit ourselves to discussing the basic four-steps including: trip generation, trip distribution, mode choice and trip assignment of trip-based demand models for different regions in the subsequent sections.

#### 3.2.2.1 Trip Generation

A summary of trip generation component of the selected travel demand models is presented in Table 3.2. The information provided in Table 3.2 include flow unit of daily trip, explanatory variables employed, and trip purpose considered in the models. From Table 3.2, we can see that in terms of flow unit, the estimated number of trips is considered by mode (both motorized and non-motorized mode) in all models. Further, the model for San Diego also considered transit trips separately from motorized trips. In terms of modeling framework, the cross-classification approach is commonly used in majority of the regions (except for Atlanta and Chicago). With respect to trip purpose, it is evident from Table 3.2 that CFRPM considers the fewest number of trip purposes compared to other demand models presented. Moreover, from Table 3.2, we can also see that CFRPM employs only the zonal characteristics in trip generation. Unlike other models, household attributes, roadway characteristics or accessibility measures were not explored.

**Table 3.1: Overview of Existing Trip-based Travel Demand Models for Different Regions**

| Study Region | Dataset | Modeling Tool | Input Features | Spatial Resolution | Temporal Resolution |
|---|---|---|---|---|---|
| Central Florida (CFRPM v 5.0; Flemming, 2010) | National Household Travel Survey (NHTS) data (2008) | CUBE Voyager | • Socio-demographic<br>• Socio-economic | TAZs (5,350 TAZs) | AM Peak, Midday, PM Peak, and Evening off-peak |
| Atlanta (ARC, 2011) | Household travel survey data compiled by ARC | ALOGIT | • Socio-demographic<br>• Socio-economic | TAZs | Daily level |
| San Diego (SANDAG, 2011) | SANDAG Household Travel Surveys (2006) | TransCAD, ArcInfo | • Socio-demographic<br>• Socio-economic<br>• Land use<br>• Regional control variables | TAZs | Peak and off-peak hour |
| Northern New Jersey (NJRTME, 2018) | Regional Household Travel Survey (RHTS) (2010-2011) | CUBE Voyager | • Socio-demographic<br>• Socio-economic<br>• Transportation infrastructure | TAZs (2900 TAZs) | Peak and off-peak hour |
| North Central Texas (NCTCOG, 2009) | Household Survey data (1996) | TransCAD (version 4.8) | • Socio-demographic<br>• Socio-economic<br>• Transportation infrastructure | Travel survey zones (4,874 TSZs) | AM peak, PM peak and off-peak period |
| Lincoln (LMPO, 2011) | Household travel survey by Colorado Front Range | TransCAD, GISDK | • Land use<br>• Roadway network data<br>• Transportation network | TAZs | AM peak, PM peak and off-peak period |
| Chicago (CMAP, 2014) | 2011 "Travel Tracker" survey | Emme, PopSyn | • Socio-demographic<br>• Socio-economic<br>• Highway network<br>• Rail network | TAZs (1,944 TAZs) | AM peak, PM peak and off-peak period |
| Delaware (DVRPC, 2011) | DVRPC 2000 household travel survey | VISUM, Python | • Socio-demographic<br>• Socio-economic<br>• Roadway network data | TAZs | AM peak, PM peak and off-peak period |

### Table 3.2: Trip Generation Overview for Different Regions

| Area | Flow unit of daily trip | Method used | Explanatory variables | Trip purpose |
|---|---|---|---|---|
| Central Florida | All Trips | Cross classification | Population classified by single family and multi-family, dwelling units classified by single family and multi-family, number of apartments, mobile homes, recreational vehicle spaces, hotel, number of employees by type, and school location, external trips, spatial trip generator | Home-based work<br>Home-based shopping<br>Home-based social recreational<br>Home-based other<br>Non-home-based |
| Atlanta | All trips including walking and bicycling trips | Multinomial logit model | Household (HH) size, HH income group, number of workers, number of children, number of autos, highway accessibility measure, transit accessibility measure, density of household, distance access measure, transit accessibility, total employment, employment class | Home based work<br>Home based shopping<br>Home based grade school<br>Home based university<br>Home based other<br>Non-home based |
| San Diego | Motorized trips, transit trips | Cross classification | Dwelling unit by structure type, population by age category, land use acres by land use type, employment by land use type, unique generator trips, external trips, trip rates, regional control variables | Home based work<br>Home based college<br>Home based shopping<br>Home based other<br>Work based other<br>Other-other<br>Serve passenger<br>Regional airport |
| Northern New Jersey | All Trips | Cross classification | Life Cycle, income, household size, number of workers, area type, population/employment density, intersection/network density, pedestrian restrictive network, availability of autos to the traveler, street network connectivity | Home-based Work Direct<br>Home-based Work Strategic<br>Home-based Shop<br>Home-based Other<br>Home-based University<br>Work-based Other<br>Non-Home Non-Work<br>Airport, Truck trips |
| North Central Texas | All Trips | Cross classification | Income quartiles, Household size, area type, employment type and number of households | Home based work<br>Home based non-work<br>Non-home based<br>Other |
| Lincoln | All Trips | Cross classification | Income, household size, employment, land use | Home-Based Work<br>Home-Based Shop<br>Home-Based Recreational<br>Home-Based Other<br>Home-Based University<br>Work-Based-Other<br>Other-Based-Other |

| | | | | |
|---|---|---|---|---|
| Chicago | All Trips | Matched survey household data | Number of Adults, Workers and Children fifteen years and younger, Income quartiles, Age of householder, vehicle ownership | Trip ends by purpose for working adults, non-working adults and children |
| Delaware | All Trips | Cross classification | Employment, employment type, vehicle ownership, area type, trip purpose | Home-based work<br>Homebased non-work<br>Non-home-based<br>External transit trips |

### 3.2.2.2 Trip Distribution

A summary of trip distribution step for the selected demand models are presented from the perspective of methodology, impedance variables, input and output of trip distribution step in Table 3.3. In terms of methodology, from Table 3.3 we can see that Gravity model is employed for trip distribution step in all of our selected travel demand models. Due to the simple formulation of Gravity model, it is easy to estimate and calibrate. The term impedance is a measure of the cost of travel between two zones. With respect to impedance variables, it is evident from Table 3.3 that the models for Central Florida have considered level-of-service variables (time, distance and cost) related to auto mode only to calculate impedance measures for the input in the trip distribution step. Composite impedance measures (combination of impedance measures for both motorized and non-motorized modes) are not explored in the models for these areas. From the input column of Table 3.3, we can see that, in general, the zone input for trip distribution step includes trip production and attraction by zones, impedances and friction factors. In order to adjust trip distribution patterns of zones, some of the models (San Diego, and Northern New Jersey) have also used k-factor (also known as balancing factor). However, k-factor might interfere with future travel prediction ability of demand models, and hence are recommended to be employed cautiously. In general, the output of trip distribution step is production-attraction zonal trips by purpose.

## Table 3.3: Trip Distribution Overview for Different Regions

| Area | Methodology | Impedance variables | Input | Output |
|---|---|---|---|---|
| Central Florida | Gravity model | In-vehicle travel time, prohibited movements, penalized movements, toll cost, toll service time | Trip productions and attractions by TAZ, travel impedance is travel time, terminal time and toll cost are also considered as additional travel impedance, trip length frequency (represented by friction factors) | Production-attraction zonal trip tables by purpose |
| Atlanta | Gravity model | Transit travel time, auto travel time | Composite impedance (auto and transit), trip production and attraction by zone and trip purpose, friction factor | Production-attraction zonal trip tables by purpose and by market groups (combination of car ownership and income) |
| San Diego | Gravity model | Auto impedances (travel time, travel distance, toll cost), transit impedances (number of transfers, cash fare, first wait time, transfer wait time, transfer walk time, in-vehicle travel time, main mode indicator), non-motorized mode impedance (travel distance, elevation, walkability factor) | Trip production and attraction by zone and trip purpose, composite impedances (auto, transit and non-motorized) for peak and off-peak conditions, friction factors, Balancing factor for zones | Production-attraction zonal trip tables by purpose |
| Northern New Jersey | Gravity model | Auto travel time, toll cost, in-vehicle travel time (in-vehicle and drive access) for transit, out-of-vehicle time (walk and wait) for transit, transit fare, park and ride cost | Composite impedance (auto and transit), zonal trip ends (productions and attractions), friction factors, and the specific zone-to-zone adjustment factor (k-factor) | Production-attraction zonal trip tables by purpose and income group combination |
| North Central Texas | Gravity model | Travel time in minutes, income, trip purpose | The gravity model has three types of inputs: a friction factor table, trip production and attraction totals, and an impedance matrix. | Production-attraction zonal trip tables by purpose |
| Chicago | Doubly Constrained Intervening opportunities Model | Travel time, travel cost | Trip production and attraction totals, balancing co-efficient and an impedance matrix | Production-attraction zonal trip tables by purpose |

| | | | | |
|---|---|---|---|---|
| Lincoln | Gravity model | Travel time, spatial separation | Friction factor table, trip production and attraction totals, and factor adjustment. | Production-attraction zonal trip tables by purpose |
| Delaware | Gravity model | In-vehicle time, out-of-vehicle time, auto toll, auto distance | Trip production and attraction totals, and an impedance matrix. | Production-attraction zonal trip tables by purpose |

#### 3.2.2.3 Mode Choice

A summary of mode choice step for the selected travel demand models is presented in Table 3.4. The information provided in Table 3.4 includes modes considered, modeling framework, explanatory variables, and output of mode choice step. With respect to mode considered, it is evident from Table 3.4 that non-motorized modes (pedestrian and bicycle) are considered in the mode choice models for San Diego, North New Jersey and Lincoln only. However, auto (reflecting occupancy levels) and transit (reflecting transit access mode and transit ride mode) modes are considered in all models. In terms of modeling framework, mode choice models are estimated by using a nested logit model in all regional models considered in our literature review. From Table 3.4, it is evident that level-of-service variables (travel time, travel distance, and travel cost for available modes) are the most commonly used explanatory variables in developing mode choice models. Trip maker attributes, household characteristics, land-use variables, roadway network characteristics, and infrastructure for transit/non-motorist are rarely considered. For example, household and transit infrastructure characteristics are considered in the model for Atlanta, while land-use variables are considered in the model for San Diego. In general, the output of the mode choice step of a travel demand model is a set of person-trip tables by available mode and trip purposes explored. However, mode choice trip tables are also quantified by different time periods (peak and off-peak periods) in the models for different areas (Central Florida and San Diego). Quantifying mode choice separately by different time periods reflects the quality of service provided by different available modes across different time periods.

#### 3.2.2.4 Trip Assignment

A summary of trip assignment across the models is presented in Table 3.5. The information provided in Table 3.5 include step components, model structure, input and output of trip assignment. From Table 3.5, we can see that for all regions (other than Atlanta, Lincoln and Chicago), this step includes two components: highway and transit assignment components. User equilibrium assignment process is mostly used (other than San Diego and Chicago) for trip assignment. The output from the assignment process is the daily traffic volume associated with highway assignment process and daily ridership of transit from transit assignment process.

## Table 3.4: Mode Choice Overview for Different Regions

| Area | Modes considered | Modeling framework | Explanatory variables | Output |
|---|---|---|---|---|
| Central Florida | Drive alone, person shared ride (2,3+), Walk to transit, Park and ride, Kiss and ride, Local bus, Express bus, Urban rail, Commuter rail | Nested logit model | Transit walk time, highway terminal time, transit auto access time, transit run time, highway run time, transit wait time, transit transfer time, transit number of transfers, transit fare, highway auto operating costs, highway parking costs, HOV time difference | Mode choice for two time periods (peak and off-peak hours) and trip purposes |
| Atlanta | Drive alone, 2 person shared ride, 3 person shared ride, 4+ person shared ride, Walk to transit, Drive to transit, Non-premium trips, Premium trips, Park and ride, Kiss and ride | Nested logit model | Auto ownership, number of workers in household, initial wait time, walk time, drive time, in-vehicle time, transfer time, block per square mile | Mode choice by trip purpose and market groups (combination of car ownership and income) |
| San Diego | Drive alone non-toll trip, Drive alone toll trip, shared-ride non-toll non-high occupancy vehicle (HOV) trip (2,3+), Two person shared-ride non-toll HOV trip (2,3+),, Bicycle trip, Pedestrian trip, Walk to transit, Drive to transit, Drop off at transit station, Local bus, Express bus, BRT/rapid bus, Light rail/street car, Commuter rail | Nested logit model | Trip distance, auto travel time, transit travel time, highway tolls, non-motorized distance, transit fares, auto operating cost, parking cost, terminal time, transit transfer time, wait time, land use density, employment density, intersection density | Mode choice for two time periods (peak and off-peak hours), three income levels (low income, middle income and high income) and six trip purposes |
| Northern New Jersey | Drive alone, person shared ride (2,3,4+), Walk to transit, Drive to transit, Rail, PATH system, Bus, Ferry, LRT, LDF | Nested logit model | In-vehicle travel time for auto and transit, walk time, bike time, drive access time, drive cost, transit fare, transit distance, use of subway, distance between different transit mode | Mode choice by trip purpose |
| North Central Texas | Drive alone, Share ride (2,3+ occupants), Transit with walking access, Transit with driving access | Multinomial logit and nested logit model | Level-of-service variables, zonal land use variables and household demographic variables | Mode choice by trip purpose |

| | | | | |
|---|---|---|---|---|
| Lincoln | Automobile, transit, non-motorized modes | Does not include a mode choice step | Based on trip distance and proximity to transit service | Mode choice by trip purpose |
| Chicago | Drive alone, Two-person auto, Three or more-person auto, Transit | Multinomial logit model | Transit in-vehicle travel times, Transit out-of-vehicle travel times, Transit fare, Highway travel times, Highway travel distances | Mode choice by trip purpose |
| Delaware | Highway, transit-walk, transit-auto | Nested logit model | In-vehicle time, out-of-vehicle time, auto toll, auto distance, parking cost, terminal Time, heavy rail portion, airline distance | Mode choice by trip purpose and car ownership |

**Table 3.5: Trip Assignment Overview for Different Regions**

| Area | Components | Model structure | Input | Output |
|---|---|---|---|---|
| Central Florida | Highway assignment, Transit assignment | User equilibrium assignment process | Vehicle trips, volume delay function, highway network | Daily traffic volume Daily boardings of transit |
| Atlanta | Highway assignment | User equilibrium technique | Vehicle trips by mode, volume delay function, value of time factors, highway network | Daily traffic volume by facility types |
| San Diego | Highway assignment, Transit assignment | Multi-modal multi-class assignment | Vehicle trips between zones, highway network, volume-delay function, traffic counts | Daily and peak hour traffic volume Daily and peak hour boardings of transit |
| Northern New Jersey | Highway assignment, Transit assignment | User equilibrium assignment process | Vehicle trips between zones, highway network, volume-delay function, traffic counts | Daily traffic volume Weekday boardings of transit |
| North Central Texas | Highway assignment, Transit assignment | User equilibrium generalized cost method | Vehicle trips between zones, highway network, volume-delay function, roadway link operating and toll cost, generalized Cost and value-of-time | A.M. peak period, P.M. peak period, and off-peak period traffic and the transit assignment |
| Lincoln | Highway assignment | Equilibrium and stochastic equilibrium assignment methods | Vehicle trips between zones, highway network, volume-delay function, cost function | A.M. peak period, P.M. peak period, and off-peak period traffic assignment |
| Chicago | Highway assignment | Path-based algorithm | Volume-delay function, time of the day, link speed | Eight periods of the day including A.M and P.M peak. |
| Delaware | Highway assignment, Transit assignment | Equilibrium assignment method | In-vehicle time, out-of-vehicle time, auto toll, auto distance, cost access/egress time, walk time at transfer, wait time | Peak, midday and evening period traffic and the transit assignment |

## 3.3 ACTIVITY-BASED TRAVEL DEMAND MODELS

The traditional trip-based approach to transportation modeling has dominated the planning process for nearly three decades. However, there are certain number of limitations associated with the trip-based approach. For instance, in traditional trip-based approaches, trips are the primary unit of analysis and separate individual models are developed for different trip purpose while ignoring the dependences across different trips. From behavioral perspective, it is hard to justify as the number of home-based and non-home-based trips are not independent of each other (for example, see (Axhausen and Gärling, 1992; Bhat et al., 2004; Bhat and Koppelman, 2006; Vovsha and Bradley, 2006)). However, this interrelationship is not considered in any of the steps in the trip-based model (trip generation, distribution, mode choice and traffic assignment). This has led to an increasing shift towards an active stream of research that examines alternative paradigms for predicting travel demand and supply by incorporating more behaviorally realistic methodologies. Specifically, researchers have attempted to overcome the conceptual and behavioral inadequacy of the trip-based approach through the use of an activity-based modeling (ABM) paradigm.

The objective of an activity-based travel modeling framework is to micro-simulate individual activity and travel participation over the course of certain time interval, which usually is a typical weekday. The framework considers travel as a means to pursue activities distributed in time and space. The activity-based modeling framework essentially attempts to replace the statistical travel prediction focus of the traditional trip-based model framework with a more behavioral approach that explicitly recognizes the fundamental role of activity participation as a precursor to travel. In doing so, the activity-based framework offers several advantages from a travel modeling perspective. First, it ensures that there is consistency across the different choices (for example, mode choices, time-of-day choices, and destination choices) assigned to successive trips of an individual. Second, the modeling of daily patterns allows a more accurate analysis of the changes to travel patterns in response to travel demand management strategies and policy initiatives, because of the explicit recognition of the interplay among the several activity participation-related choices of an individual. Third, activity-based models allow the incorporation of the influence of inter-household interactions on travel. Finally, the activity-based framework generally considers a finer resolution of time and space (compared to the traditional trip-based framework), and thus is more suited to providing travel inputs needed for vehicle emissions and air quality modeling.

### 3.3.1 Existing Literatures

Federal transportation agencies and metropolitan organizations (MPOs) in the United States and Europe have started to recognize the advantages of the activity-based framework and are investing substantially in the development and deployment of activity-based models. Examples of urban regions in the United States that have an operational or research based activity model include: North Florida (NERPM-AB v.1.0, 2015), San Francisco (SFCTA, 2002), Mid-Ohio (MORPC, 2005), New York (NYMTC, 2010), Sacramento (SACSIM), Dallas-Fort-Worth (CEMDAP) and Southeast Florida (FAMOS). A brief review of these models is provided in Table 3.6. Specifically, the table presents details of the study region, dataset used for the analysis, conceptual methods adopted in the study along with the tool used for developing the model, different kind of exogenous variables used for the models and finally the spatial (at what spatial unit, the travel demand model is developed) and temporal resolution (at what temporal unit, the traffic is predicted) of the studies.

Several observations can be made from the Table 3.6. First, in terms of modeling tool, the common forecasting package employed across the frameworks is the CUBE transportation planning software (4 out of 7). Second, a host of independent variables are considered in these regional models including socio-demographic (such as population and household size), socio-economic (such as employment and income), land use, activity-travel infrastructure, and physical environment related attributes. Third, similar to the trip-based approach, Traffic Analysis Zones (TAZs) is the most prevalent spatial unit in activity-based approaches too (6 out of 7 regions). Fourth, in terms of the temporal resolution, the traffic was being forecasted by different time period (morning peak, evening peak, off peak hours) in majority of the models presented in Table 3.6.

### 3.3.2 Alternative Approaches

The proposed research module is motivated by the need to develop practical approaches that can be employed to reflect regional changes in modeling traffic volumes at high fidelity. While travel demand modeling tools described in section 3.2 and 3.3 offer a mechanism to conduct an evaluation of regional impacts on travel demand, they are resource intensive. Hence, to explore for alternative and less resource intensive research methods, we reviewed the literature for alternative methods to generate travel demand outputs.

A summary of the review is presented in Table 3.7. From the table we can make several observations. First, alternative approaches that do not require running the full travel demand model have been employed in literature. These approaches include post processing, direct modeling, sketch planning and linear interpolation methods that develop travel demand outputs using simpler regression approaches. Second, the application of linear interpolation is more common in roadway infrastructure project (such as bridges, tunnels, and roadways) while sketch planning tool is more commonly used for transit planning. For instance, the study by Lane et al., 2006 incorporated the reverse commute trips and relevant transportation system variables to forecast the light and commuter rail ridership in 17 US regions using regression based multivariate sketch planning models. Finally, a number of independent variables are considered in these models including socio-demographic and socio-economic characteristics (such as population composition, employment status, and migration), economic and policy conditions (such as employment, GDP, and housing conditions) and transportation infrastructure attributes (such as transportation facilities, traffic signal plans, vehicle fleet composition, mode preferences, and travel characteristics). Please see the recent report for more detailed description of using alternative methods in project planning forecasts (Eluru et al., 2020).

**Table 3.6: Overview of Existing Activity-based Travel Demand Models for Different Regions**

| Study Region | Dataset | Modeling Tool | Input Features | Spatial Resolution | Temporal Resolution |
|---|---|---|---|---|---|
| North Florida (NERPM-AB v.1.0, 2015) | NHTS data | DaySim, Cube Voyager | • Socio-demographic<br>• Socio-economic<br>• Land use<br>• Transportation infrastructure | TAZs (2,522 TAZs) | AM Peak, Midday, PM Peak, and Evening off-peak |
| San Francisco (SFCTA, 2002) | Household survey data in San Francisco (1999) | Cube, C++, Java | • Socio-economic<br>• Pedestrian environment factors<br>• Parking data<br>• Built environment | TAZ (1,739 TAZs) | Daily and time of the day (early, AM peak, midday, PM peak, late) |
| Mid-Ohio (MORPC, 2005) | Household Interview Survey (HIS) (1999) | Cube, Java | • Socio-demographic<br>• Socio-economic<br>• Land use | TAZ (1,800 TAZs) | Every 1-hour period |
| New York (NYMTC, 2010) | Regional Travel – Household Interview Survey (RT-HIS) | TransCAD, CUBE Voyager | • Socio-demographic<br>• Socio-economic | TAZ (4,631 TAZs) | AM peak, midday, PM peak, late |
| Sacramento (SACG, 2015) | SACOG 2000 Household Travel Survey | DaySim | • Land use<br>• Socio-demographic<br>• Socio-economic<br>• Transportation infrastructure | TAZ | AM peak, midday, PM peak, evening |
| Dallas–Fort Worth (Pinjari et al., 2006) | 1996 DFW household activity survey | CEMDAP | • Socio-demographics<br>• Activity-travel environment characteristics<br>• Socioeconomics, Land-Use<br>• Transportation System | TAZ (4,874 TAZs) Persons (6,166 individuals) | AM Peak, PM Peak and off-peak |
| Southeast Florida (Pendyala et al., 2005; FAMOS) | Household activity and travel data collected in the Southeast Florida region of Florida in 1999 | Florida Activity Mobility Simulator (FAMOS) | • Socio-demographic<br>• Socio-economic<br>• Zonal network level of service | Person | Peak and off-peak period |

**Table 3.7: Literature for Alternative Methods to Generate Travel Demand Outputs**

| Study | Study Region | Infrastructure Analyzed | Dimensions Analyzed | Methodology | Factors Considered | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | *Socio Demographic* | *Economic and Policy* | *Transportation Infrastructure* | *Emerging Factors* |
| Cervero, 2006 | Greater Charlotte, the San Francisco Bay Area, and south St. Louis County | Rail Transit Projects | Transit Forecasts | Post processing and Direct modeling | √ | -- | √ | -- |
| Lane et al., 2006 | 17 US region: Chicago, New Jersey etc. (Commuter and light rail) | Rail Transit Projects | Ridership Forecast | Sketch Planning | √ | √ | √ | -- |
| FDOT, 2015 | Us 301 (Gall Boulevard) from SR 56; to SR 39 (Pasco county, Florida) | Roadway Infrastructure | Traffic Forecasts | Linear Interpolation | -- | -- | √ | -- |
| Conway et al., 2017 | Randstadt, Netherlands (Transit and land-use accessibility) | Roadway Infrastructure | Transit Forecasts | Sketch Planning | -- | √ | √ | -- |
| Lasley et al., 2017 | Texas (highway) | Roadway Infrastructure | Congestion benefits | Sketch Planning | -- | √ | √ | -- |
| FDOT, 2017 | US 98/John Singletary Bridge | Roadway Infrastructure | Traffic Forecasts | Linear Interpolation | -- | -- | √ | -- |
| VDOT, 2017 | I-95 Express Lane Extension | Roadway Infrastructure | Traffic Forecasts | Linear Interpolation | -- | -- | √ | -- |
| Arndt et al., 2009 | Texas (Light rail) | Rail Transit Projects | Ridership Forecast | Sketch Planning | -- | √ | √ | -- |
| ST, 2010 | Seattle, WA (Light rail) | Rail Transit Projects | Ridership Forecast | Sketch Planning | √ | -- | √ | -- |
| Upchurch and Kuby, 2014 | Phoenix (Light rail) | Rail Transit Projects | Ridership Forecast | Sketch Planning | √ | √ | √ | -- |
| Duggal, 2016 | Edmonton city, Canada (Light rail) | Rail Transit Projects | Ridership Forecast | Sketch Planning | √ | -- | √ | -- |

### 3.4 SUMMARY

In this chapter, we present an overview of research for generating travel demand modeling outputs. The three approaches include: (a) trip-based models, (b) activity-based models and (c) alternative research methods.

The first two methods, while resource intensive, are suited for all time scales. On the other hand, the alternative approaches such as sketch planning are appropriate in the short term. To cover these two timescales, we intend to test two approaches: (1) the trip-based model for Central Florida - Central Florida Regional Planning Model (CFRPM) and (2) a linear regression-based approach to generate travel demand outputs for the short term. The travel demand outputs generated from these approaches will serve as inputs to the hybrid traffic prediction framework.

# CHAPTER 4 STUDY AREA AND DATA COLLECTION

## 4.1 INTRODUCTION

The proposed hybrid model for high fidelity traffic prediction depends on the roadway network characteristics to learn how the flow will propagate at different links. In this study, the network characteristics are defined in terms of the distribution of the nodes (or intersections) over the network and connectivity among adjacent nodes. The underlying road network will be represented using features such as roadway capacity, speed limits, AADT, number of lanes etc. This chapter includes a detail description of the study area, spatial distribution of the ATSPM signals in the study area, and characteristics of the roadway network to be considered in the model.

## 4.2 STUDY AREA

The research team, in consultation with the project oversight committee, selected Seminole County as the study area to test the proposed high-fidelity model **(Figure 4.1)**. As shown in Figure 4.1, the research team collected the information on all the local streets, major arterials and highways of the study area. In Seminole County, most of the signalized intersections are equipped with advanced traffic signal controllers on the arterials and each signal provides Automated Traffic Signal Performance Measures (ATSPM). The research team has collected the raw data of the ATSPM signals for years 2015, 2016 and 2019. **Table 4.1** presents the number of signals for which data are available in each of these years. **Figure 4.1** shows the spatial distribution of ATSPM signals for year 2015. The proposed high-fidelity model will take network information as input to learn how traffic flows propagate along different roadway links. Hence, the research team will create a network graph considering the locations of different intersections **(Figure 4.1)** and the connectivity among these intersections. The connectivity is defined as the roadway links among different nodes.

**Table 4.1: Number of ATSPM signals for each year**

| Year | Number of ATSPM Signals |
|---|---|
| 2015 | 253 |
| 2016 | 318 |
| 2019 | 325 |

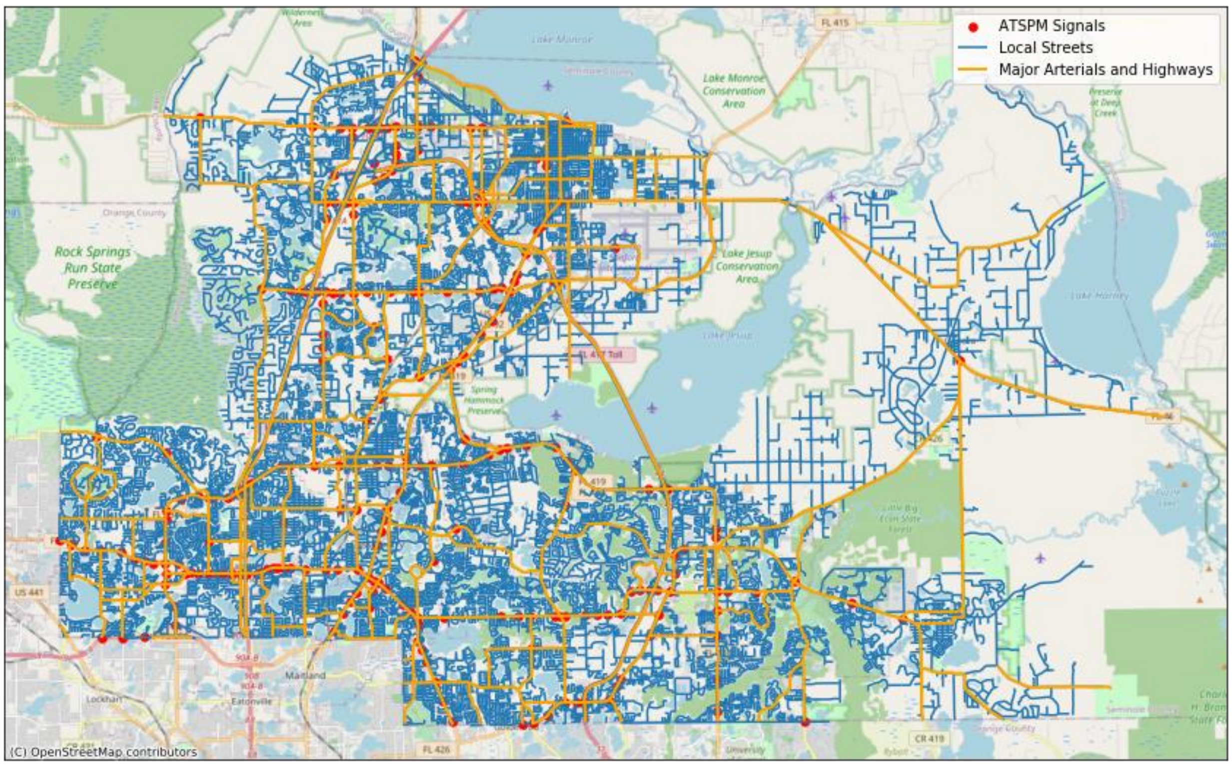



**Figure 4.1: The study area (Open Street Map)**

## 4.3 TRAFFIC ANALYSIS ZONES

The proposed hybrid modeling approach also depends on the zonal and sub zonal level features related to traffic demand for high fidelity traffic prediction. Hence, the research team has collected GIS data of Traffic Analysis Zones (TAZs) for the study area. The data is retrieved from the

Central Florida Regional Planning Model (CFRPM) version 6.0 for base year 2010 **(see Figure 4.2)**. Seminole County includes 230 TAZs. In **Figure 4.2,** the dark lines show the boundaries of different TAZs. Based on the boundaries of the TAZs the research team will collect different features related to zonal travel demand such as land use characteristics, population etc. The research team also explored the distribution of ATSPM traffic signals within each TAZ. This will help the research team to develop subzones insides each TAZ to extract features related to traffic demand features at a micro level and assign the related information to different intersections equipped with ATSPM signals. In the model implementation phase, we will employ the updated TAZ boundaries for this project.

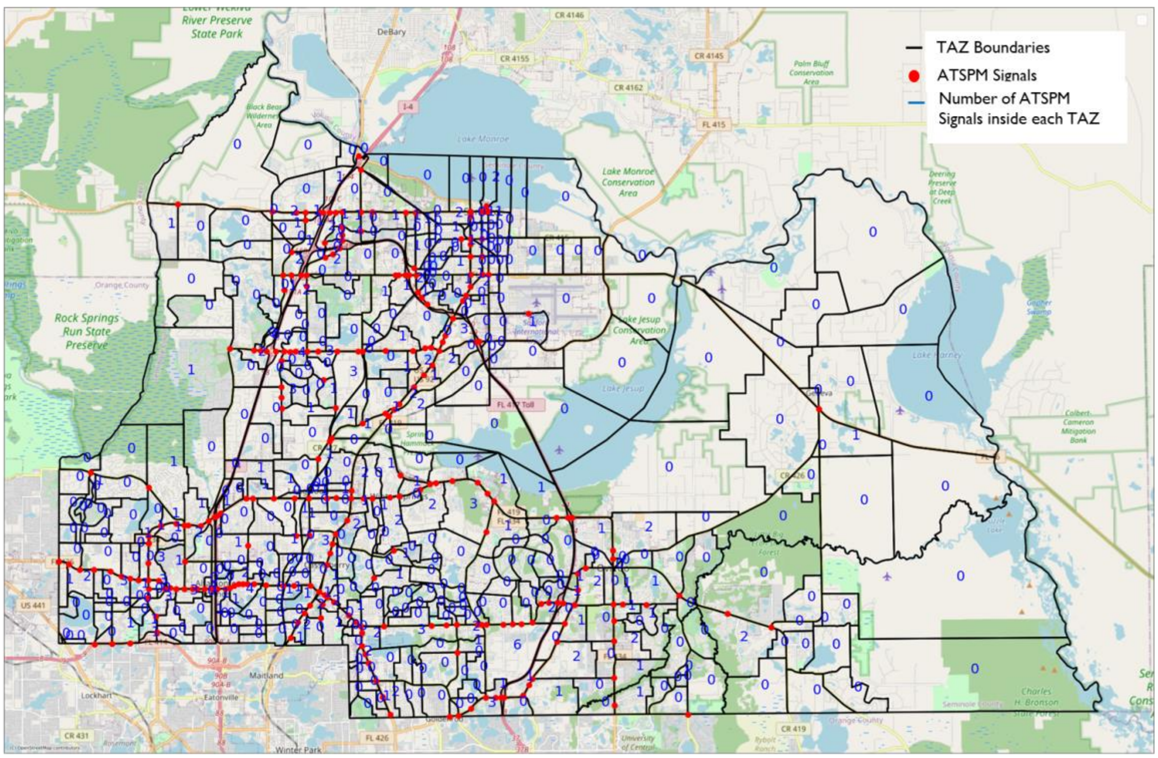



**Figure 4.2 : Distribution of ATSPM signals inside different Traffic Analysis Zones (TAZs)**

## 4.4 DEMAND CHARACTERISTICS

In the Central Florida region, CFRPM is employed as the trip-based model system. Hence, we have compiled CFRPM data to generate the spatio-temporal travel demand. Further, for the current project, we consider a comprehensive set of exogenous variables including socio-demographic, built environment and land-use characteristics. Information about these variables is collected from different data sources including American Community Survey (ACS), US Census Bureau, NAVSTREET street data and Florida Department of Revenue (FDOR) databases. A summary of the variables with its' source, the timeframe up to which we collected the data, and the access websites are provided in **Table 4.2**.

For socio-demographic attributes, we collect information on total population, number of people by gender, age, race, income, proportion of means of transportation used by commuter for their work trips (car, transit, bike and walk), household density and number of households by different vehicle ownership (0, 1, 2, 3 and 4 or more). We gather the information about the demographic attributes from the ACS data. The variables that are compiled within built environment category includes number of law enforcement offices, business center, commercial center, school, hospital, recreational center, restaurant and shopping center. These variables are gathered in Geographical Information System (GIS) shape file format. Information about the number of law enforcement offices within built environment category are downloaded from Florida Geographic Data Library (FGDL) database and the rest are gathered from NAVSTREET street data. To capture the land use characteristics, parcel data (for 2016-2019) obtained from Florida Department of Revenue (FDOR) were utilized. Each parcel is assigned a unique ID (Parcel ID) linking it with equivalent parcel level attribute information such as property/feature value, land value, land area in square feet, land use codes (DOR-UC), owner name, owner address, physical address, physical zip code, building details and so on contained in the Name-Address-Legal (NAL) file. For our analysis purpose, we consolidated the land use categories reported by DOR into 12 land use categories. These are Single Family Residential, Multi-Family Residential, Retail/Office, Industrial/Manufacturing, Agriculture, Institutional/Infrastructure, Public, Recreational, Water, Vacant, and Others.

**Table 4.2: List of demand variables and their Sources**

| Variable/ Timeframe | Source/Link |
|---|---|
| ***Dependent Variable*** | |
| CFRPM Data (2010, 2015) | Florida Standard Urban Transportation Modeling Structure (FSUTMS) https://www.fsutmsonline.net/index.php?/model_pages/comments/cfrpm_version_61 |
| ***Independent Variables*** | |
| Socio-Demographics (2014-2019) | American Community Survey (ACS Census data) https://www.census.gov/data.html |
| Built Environments (2014-2018) | NAVSTREETS Street Data; Florida Geographic Data Library (FGDL) https://ocean.floridamarine.org/acp/stpacp/Geodata/Metadata/NAVTEQ_Roads_FL.html https://www.fgdl.org/download/ |
| Land-Use Variables (2014-2019) | Florida Department of Revenue (FDOR) https://floridarevenue.com/property/Pages/DataPortal_RequestAssessmentRollGISData.aspx |

## 4.5 SUMMARY OF DATASETS

The research team has gathered all the data related to network characteristics, travel demand, and intersection level traffic movement volume from multiple sources to create a comprehensive database**. Table 4.3** presents a list of all of the datasets gathered so far under this project. The team has also written required Python scripts to process this data, visualize them, and generate appropriate features to be used as inputs to the proposed framework. The databases and Python scripts developed under this task have been extensively used throughout the project. Moreover, while completing the subsequent tasks of the project the team has further refined these Python scripts when needed.

**Table 4.3: List of datasets gathered and their sources**

| Data Description | Sources | Year |
|---|---|---|
| Transportation network of Seminole County | FDOT:https://gis-fdot.opendata.arcgis.com/search?collection=Dataset | 2020 |
| | Seminole County: http://www.seminolecountyfl.gov/departments-services/information-services/gis-geographic-information-systems/gis-data.stml | 2020 |
| Seminole County Traffic Analysis Zone | FDOT District 5: Central Florida Regional Planning Model Version (CFRPM) 6.0 | 2010 |
| | FDOT District 5: Central Florida Regional Planning Model Version (CFRPM) 7.0 | 2015 |
| Traffic network characteristics (speed limits and the number of lanes) | FDOT:https://gis-fdot.opendata.arcgis.com/search?collection=Dataset | 2020 |
| ATSPM event log data | FDOT District 5 | 2015, 2016, & 2019 |
| CFRPM Data | Florida Standard Urban Transportation Modeling Structure (FSUTMS) https://www.fsutmsonline.net/index.php?/model_pages/comments/cfrpm_version_61 | 2010 & 2015 |
| Socio-Demographics | American Community Survey (ACS Census data) https://www.census.gov/data.html | 2014-2019 |
| Built Environments | NAVSTREETS Street Data; Florida Geographic Data Library (FGDL) https://ocean.floridamarine.org/acp/stpacp/Geodata/Metadata/NAVTEQ_Roads_FL.html https://www.fgdl.org/download/ | 2014-2018 |
| Land-Use Variables | Florida Department of Revenue (FDOR) https://floridarevenue.com/property/Pages/DataPortal_RequestAssessmentRollGISData.aspx | 2014-2019 |

## 4.6 SUMMARY

This chapter reports the selection of the study area, road network characteristics, and the distribution of the traffic signals within the road network. It also includes information on TAZs and distribution of ATSPM equipped traffic signals for each TAZ. A brief description is also provided on all the potential demand variables required for running the model. A list of datasets and their sources has been provided.

# CHAPTER 5 EXTRACTING TRAFFIC VOLUME INFORMATION FROM ATSPM DATA

## 5.1 INTRODUCTION

In this chapter, we describe the process of calculating traffic volume for each intersection from Automated Traffic Signal Performance Measures (ATSPM) data. The ATSPM data contain real-time and historical performance at signalized intersections. We also report the results of preliminary data exploration of ATSPM data. The research team has collected the raw ATSPM data from FDOT District 5 office.

## 5.2 ATSPM EVENTS

Traffic signal controller manufacturers (e.g., Econolite, Intelight, Siemens, McCain, TrafficWare and some others) developed a “data-logger” program that runs in the background of the traffic signal controller firmware. *The Indiana Traffic Signal Hi Resolution Data Logger Enumerations* f encode events to a resolution to the nearest 100 milliseconds. For each signal, all events generated by a signal controller, or a loop detector are recorded in four bytes: two bytes for the timestamp identifying when the event occurred, one byte for event code providing the information of event type, and one byte for event parameters, signifying detector ID and signal phases. The event code is important for determining the type of reported activity, which includes, for example, phase initiation or termination, and detector on/off. **Table 5.1** shows sample data collected at the intersection of SR 434 and US 17-92 and the description according to the corresponding event codes and event parameters.

**Table 5.1: Sample data collected at the intersection SR 434 and US 17-92**

| Sample Raw Data | | | Description |
|---|---|---|---|
| **Timestamp** | **Event Code** | **Event Parameter** | |
| 7/25/2019 11:59:33 | 81 | 20 | Detector 20 Off |
| 7/25/2019 11:59:36 | 81 | 2 | Detector 2 Off |
| 7/25/2019 11:59:38 | 82 | 2 | Detector 2 On |
| 7/25/2019 11:59:39 | 81 | 2 | Detector 2 Off |
| 7/25/2019 11:59:41 | 82 | 21 | Detector 21 On |
| 7/25/2019 11:59:41 | 81 | 21 | Detector 21 Off |
| 7/25/2019 11:59:42 | 8 | 1 | Phase 1 Begin Yellow |
| 7/25/2019 11:59:42 | 7 | 1 | Phase 1 Green Termination |
| 7/25/2019 11:59:43 | 82 | 18 | Detector 18 On |
| 7/25/2019 11:59:43 | 81 | 18 | Detector 18 Off |
| 7/25/2019 11:59:47 | 82 | 22 | Detector 22 On |
| 7/25/2019 11:59:47 | 9 | 1 | Phase 1 End Yellow |

## 5.3 EXTRACTING TRAFFIC VOLUME FROM ATSPM DATA

Each signal uses 5 types of potential detectors: setback detector 1, setback detector 2, stop bar detector, right turn detector, and exit lane detector (see **Figure 5.1**). For the 325 signals in Seminole County, we have found 1700 detectors recorded in the ATSPM data. Based on the distance from stop bar detector, we have found that all of these detectors are either setback detector 1 (330 ft upstream of the stop bar detector) or setback detector 2 (170 ft upstream of the stop bar detector). We have also checked that, in the given dataset, each lane of the signal has only one detector.

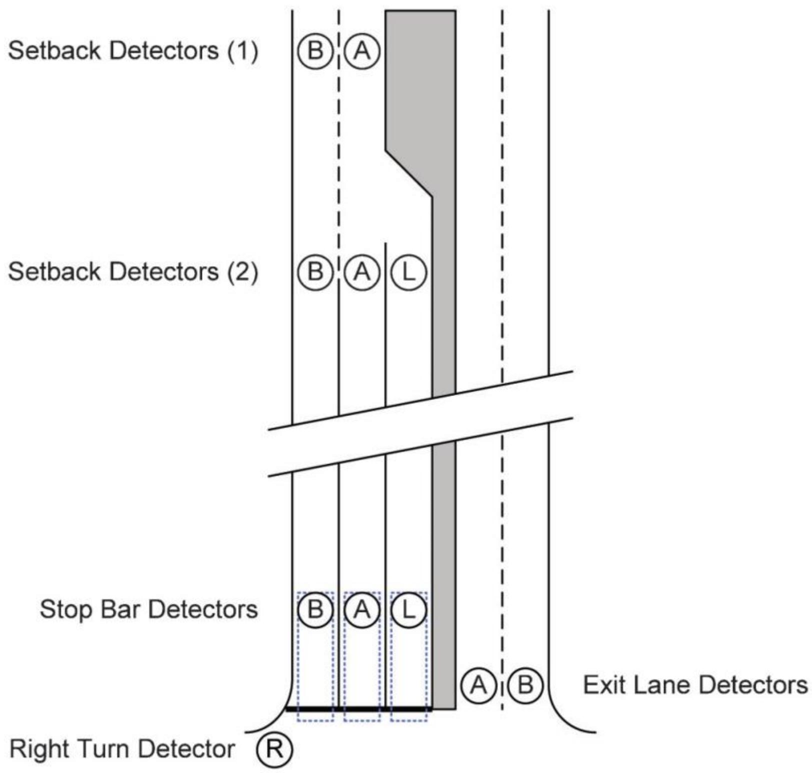
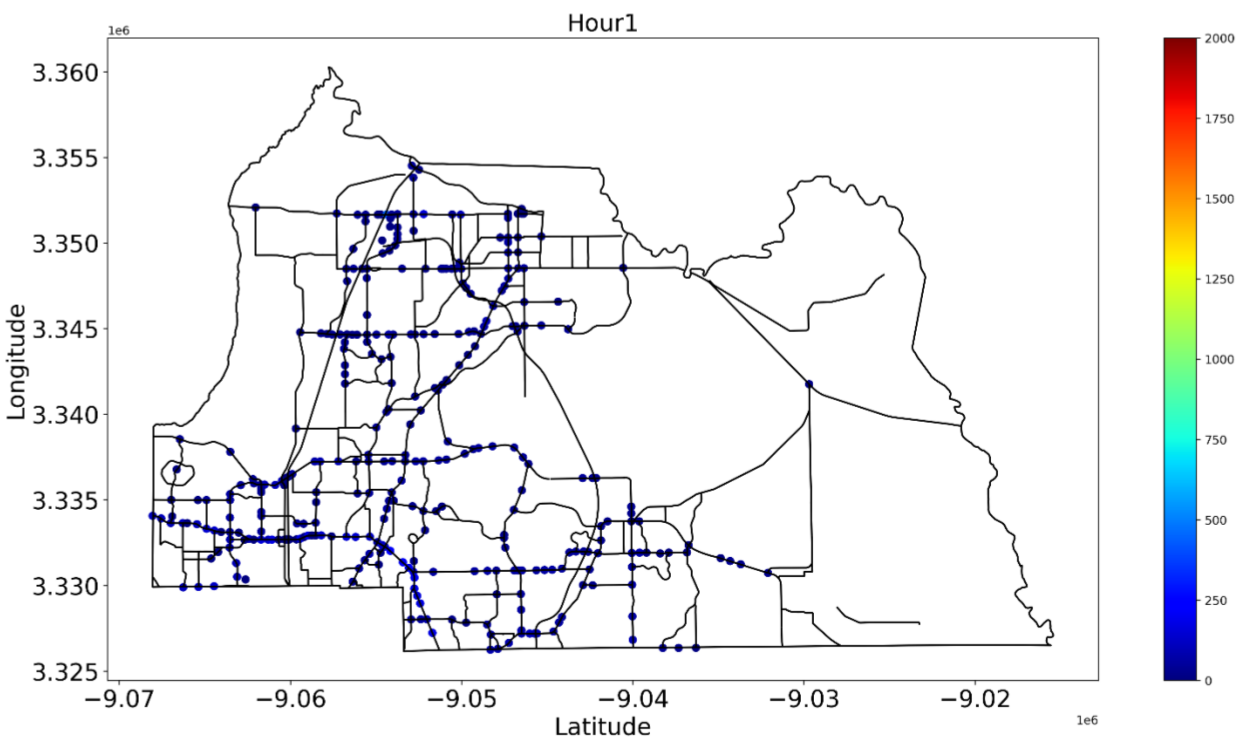


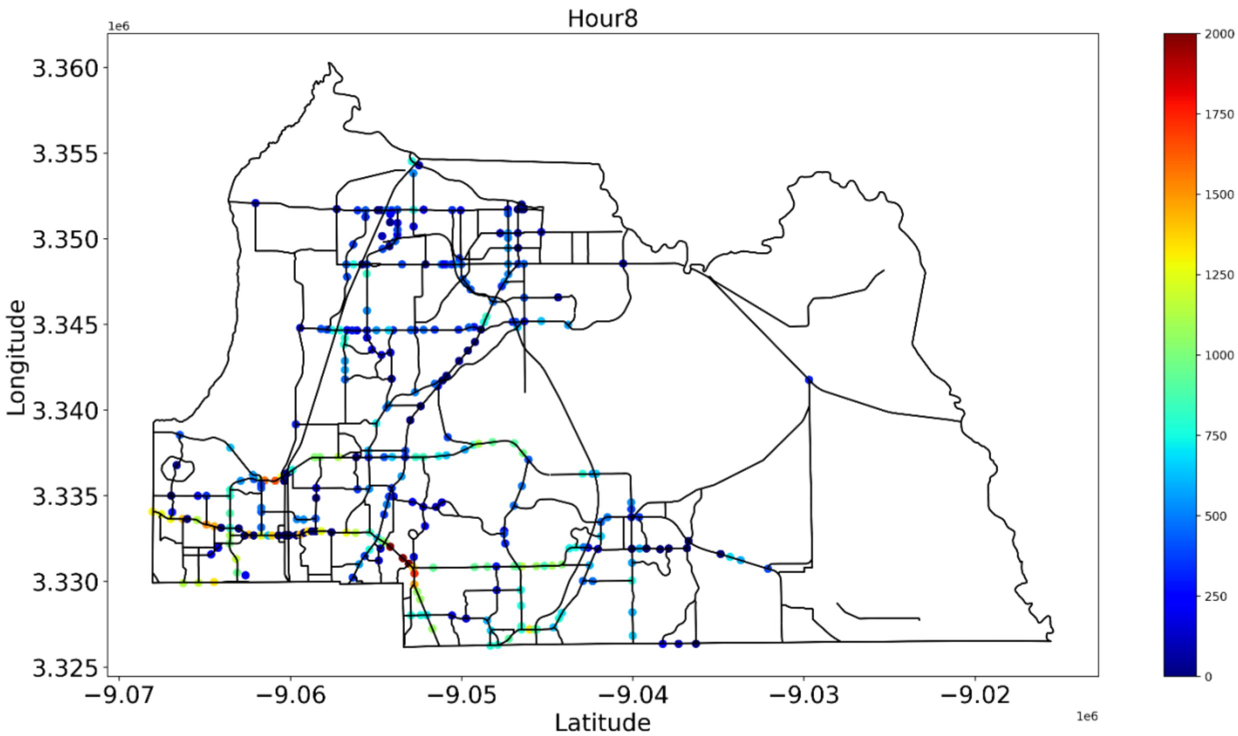

**Figure 5.1: Location of the potential detectors (Day et al., 2014)**

We have calculated the traffic volume of every movement type for each signal from ATSPM database. In the database, the event code ‘82’ represents the event ‘detector on’ which means that a vehicle is going through the intersection; this information is used to count the traffic volume. According to the configuration of each detector, we can identify the direction type and movement type of each detected vehicle from each signal using the event parameter. When the event code is ‘82’, the corresponding event parameter represents a detector channel. This event parameter (or detector channel) provides information of both the movement type and direction type for a given signal. For example, for the signal US-17/92 and SR 434: the event parameter 18 represents the movement type ‘go through’ and direction ‘northbound’ and the event parameter 8 represents the movement type ‘go through’ and direction ‘southbound’.

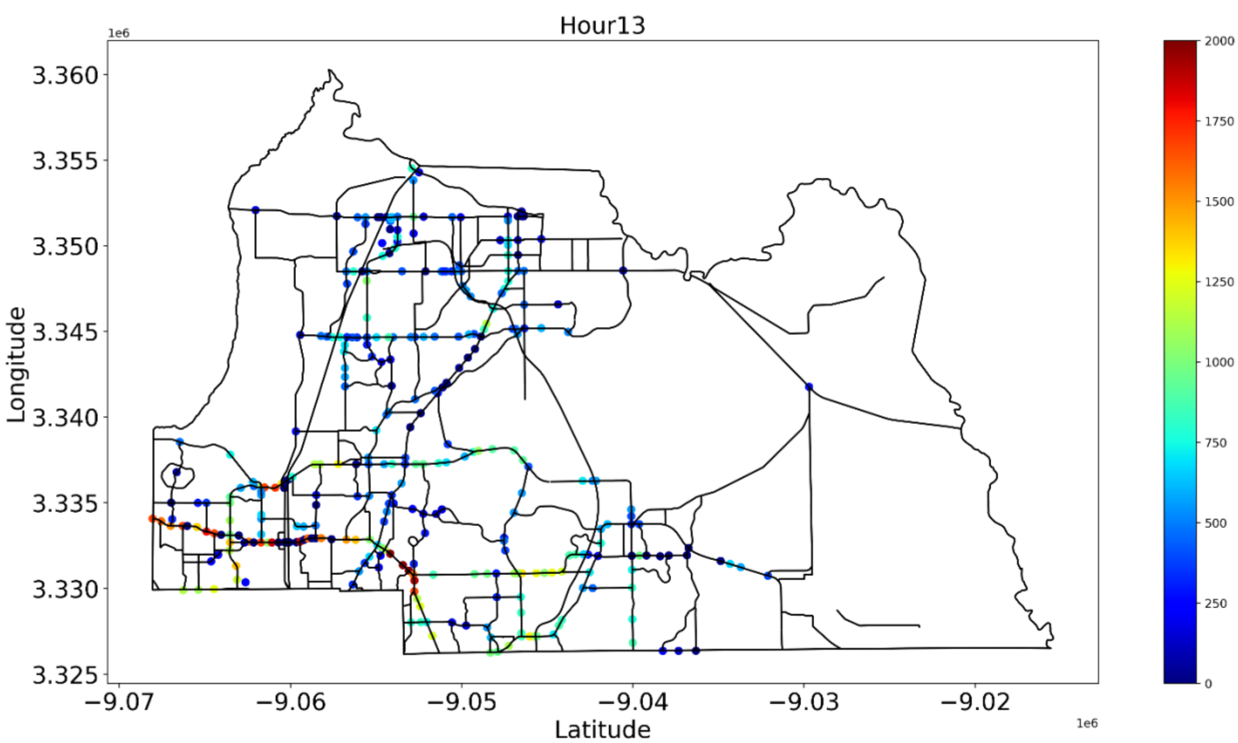
To visualize the calculated traffic volume, we aggregate the average hourly traffic volume of each signal. **Figure 5.2** shows the spatial distribution of average hourly traffic volume for all the signals in 4 time periods – midnight (1 am), morning peak hour (8 am), midday peak hour (13 pm), and afternoon peak hour (18 pm). From the figure we can clearly distinguish the average hourly traffic volume during peak and off periods. As shown in **Figure 5.2 (a),** for off peak period (in our case at midnight), the hourly traffic volume (less than 500 vehicles per hour) is less than peak period hourly traffic volumes. During morning peak hours, the traffic volume starts to increase, and many signals have traffic volume higher than 1000 vehicles per hour. However, for given day traffic volumes are highest at midday and afternoon peak hours. Moreover, we find that most of the signals can accommodate more than 1500 vehicles per hour. Overall, our analysis reveals the spatiotemporal behavior of intersection level hourly traffic volume from ATSPM data.



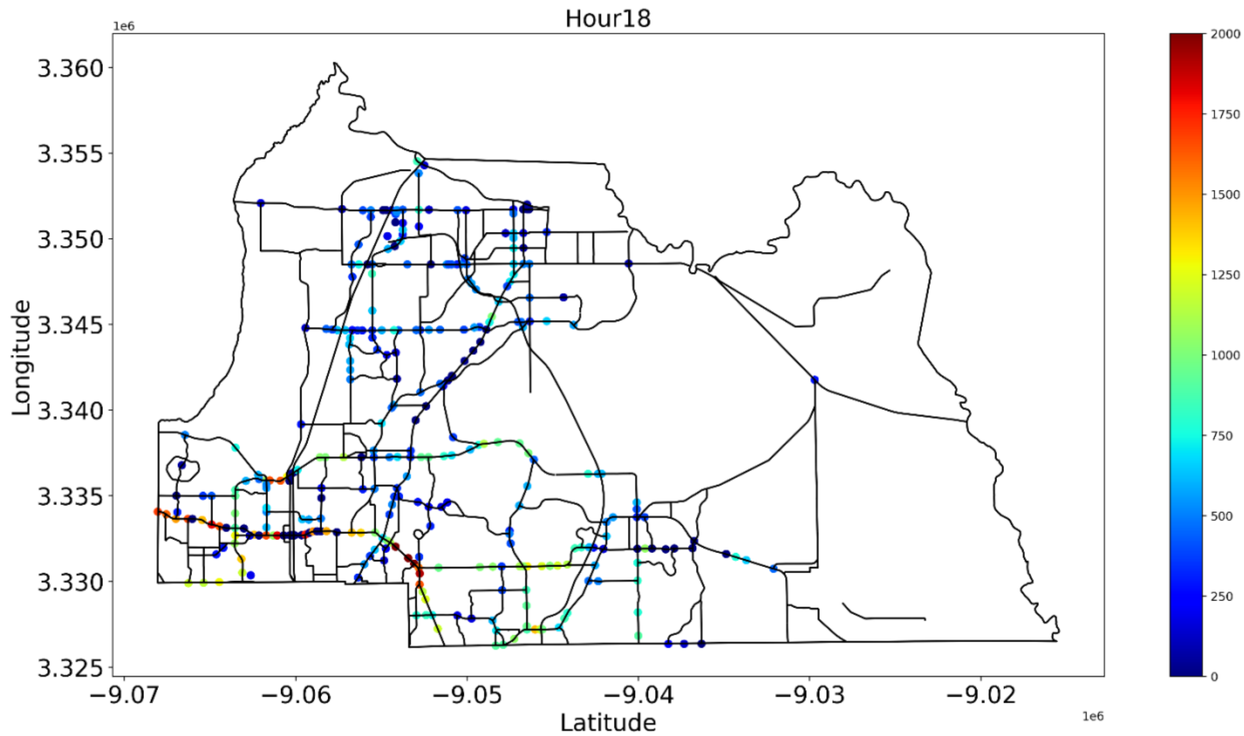

(a) Hourly traffic volume at midnight (1 am)




(b) Hourly traffic volume at morning peak hour (8 am)



(c) Hourly traffic volume at midday peak hour (13 pm),




(d) Hourly traffic volume at evening peak hour (13 pm),

**Figure 5.2: The spatial distribution of average hourly traffic volume across all signals**

In addition, we select one specific signal (SR 434 & US 17-92) to demonstrate temporal patterns (using time series plot) of traffic volume for each movement types. As shown in figure 5.3, the signal is located at SR 434 and US 17-92 intersection, connecting two major arterials in Seminole county. We exacted data for all the 4 movement directions from ATSPM database.

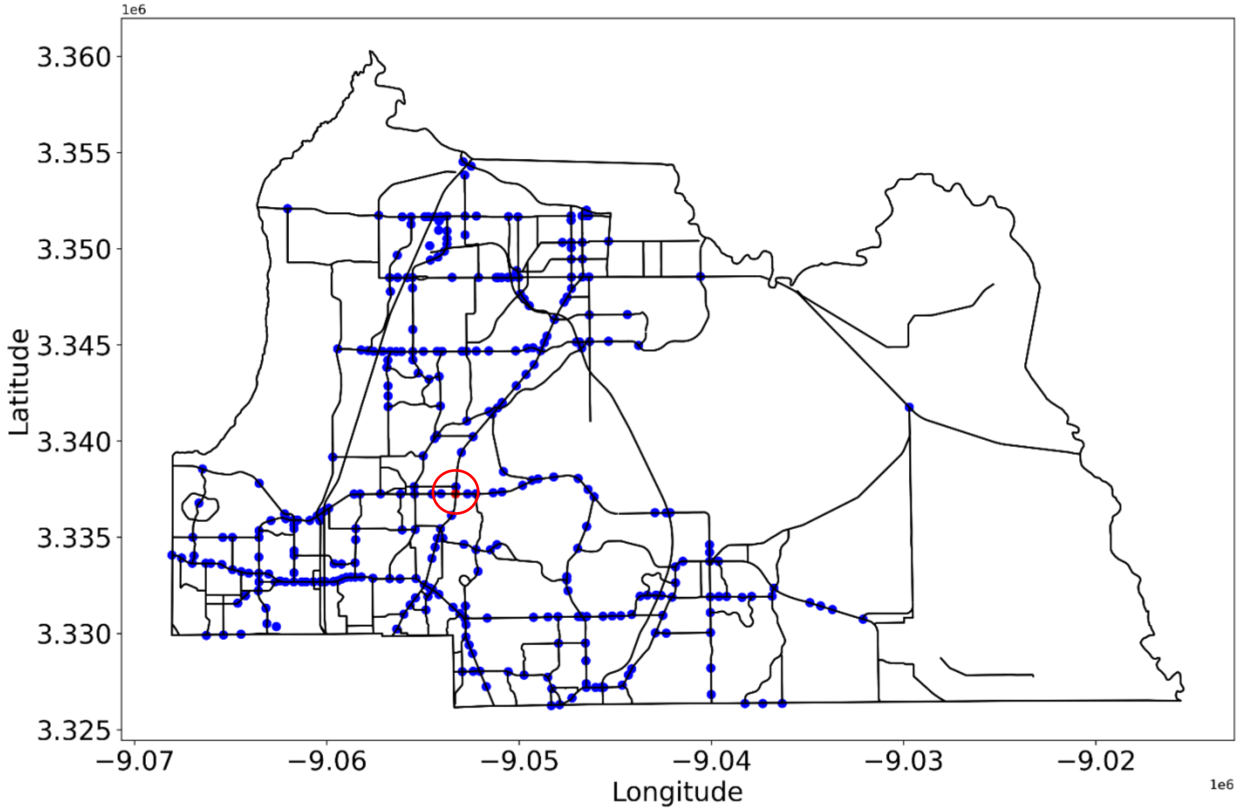



**Figure 5.3: The position of selected signal (SR 434 and US 17-92)**

**Figure 5.4** shows the distribution of daily traffic volume over one-year period for the selected signal. From the figure we observe that, daily volume has a typical weekly periodicity which means weekdays have more traffic compared to the weekend days. We also observe a few irregularities in traffic pattern correspond to some typical holiday events such as Thanksgiving Day and Christmas Day. Furthermore, our analysis reveals travelers' response to COVID-19 restrictions. We find that after March 2020, traffic volume significantly decreases due to decrease in travel demand caused by COVID-19 restrictions and traffic tend to increase after April 2020.

We also calculate the hourly traffic flow volume over one-week period, shown in **Figure 5.5**. To better understand the temporal patterns of hourly traffic flow volume, we select data from two weeks – one week before COVID-19 (7/22/2019-7/28/2019) and another week after COVID-19

(4/20/2020-4/26/2020). In general, the weekly distribution also reveals that the traffic volume is higher on weekdays than weekends. We also find that on weekdays the southbound volumes dominate during the a.m. peak hour and the northbound volumes dominate during the p.m. peak hour for both the periods before and after the COVID-19. In the next task, we will validate the automated traffic count from ATSPM data by comparing with manual counts.

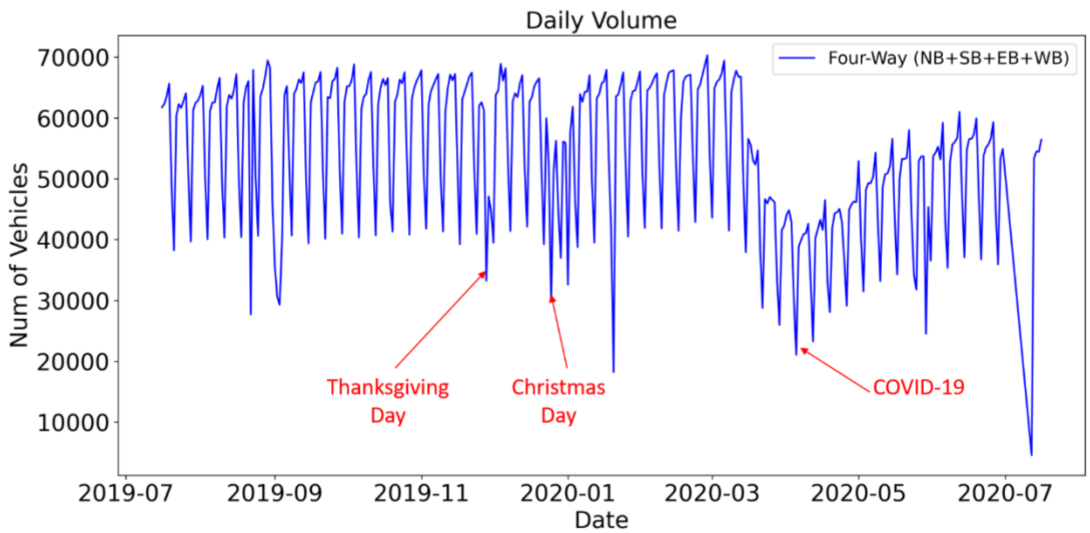



**Figure 5.4: Daily volume distribution of a selected signal (SR 434/US 17-92)**

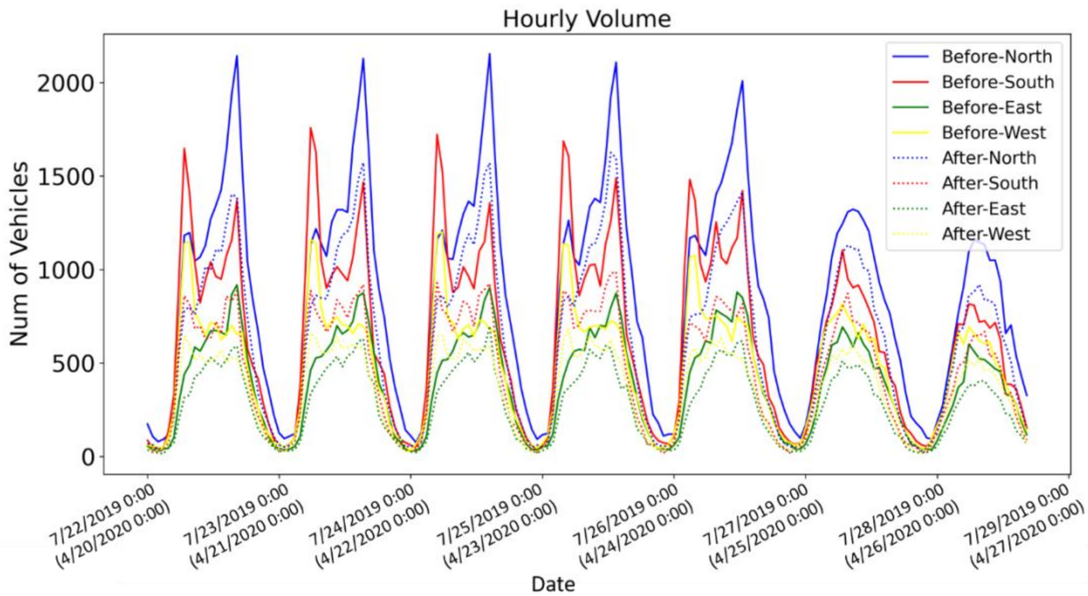



**Figure 5.5: Hourly volume distribution of a selected signal (SR 434/US 17-92)**

For each direction, traffic volume can be divided into different movement types, such as through, left-turn, and right-turn. It should be noted that not all signals have data available for all movement types and directions. For instance, if a signal has no exclusive right-turn lane, right-turn volume will not be available. **Figure 5.6** shows the average hourly traffic volume for each movement type of each direction for the selected signal (SR 434/US 17-92) over the time periods. The distribution indicates that 'through' movement in northbound direction has dominant volumes followed by the 'through' movement in southbound direction. The distribution shows commuting characteristics in the northbound and southbound directions; in the morning peak hour, the southbound direction has more volume and in the evening peak hour, the northbound direction has more volume. Traffic volume in the westbound and eastbound directions show similar trends.

Both results of the daily volume and the hourly volume of a particular signal indicate that ATSPM data can adequately capture the temporal patterns of the traffic demand and flows, which will be suitable for developing a high-resolution traffic prediction model. However, for some signals, detector data are not available for all directions and all movement types. Appropriate assumptions need to be made to account for such missing data.

## 5.4 DATABASE AND CODE DEVELOPMENT

In this project, we report the results based on the ATSPM data over three-year periods (2015, 2016, 2019) collected in the Seminole County. We also have data from January to July of 2020. However, due to significant disruptions in traffic patterns due to COVID-19, the team has decided to exclude the data from 2020 from the analysis.

To process the data, Python scripts have been written. Since the size of the raw dataset is too large to be directly processed, we first write codes to divide the raw dataset into multiple smaller pieces. Then, to make it easier to analyze the patterns of each signal, we convert the raw data pieces into signal-based files (i.e., one data file for one signal). Finally, we count the traffic volume of each signal based on the event code and event parameter available in the data. The database and codes developed under this task are summarized in **Table 3.2**.

**Table 5.2: List of database and codes developed**

| **Database** | |
|---|---|
| **Name** | **Description** |
| ATSPM Data | Three years (2/21/2015 – 12/31/2016, 2019/1/1 - 1/16/2020) of ATSPM data in Seminole County. The raw data size is 2,334,948.38 Mb with 59,584,620,138 events for 325 signals. |
| **Code** | |
| **Name** | **Description** |
| ATSPM_Data_Piece.py | Divide the raw dataset into multiple pieces. |
| ATSPM_Filter_Signal.py | Filter all the data outside of Seminole County. |
| ATSPM_Signal.py | Convert the raw data pieces to signal based files. |

| | |
|---|---|
| ATSPM_Hourly.py | Calculate hourly traffic volume for each signal. |
| ATSPM_Hourly_Information.py | Exploratory analysis and visualization of hourly traffic volume for each signal. |

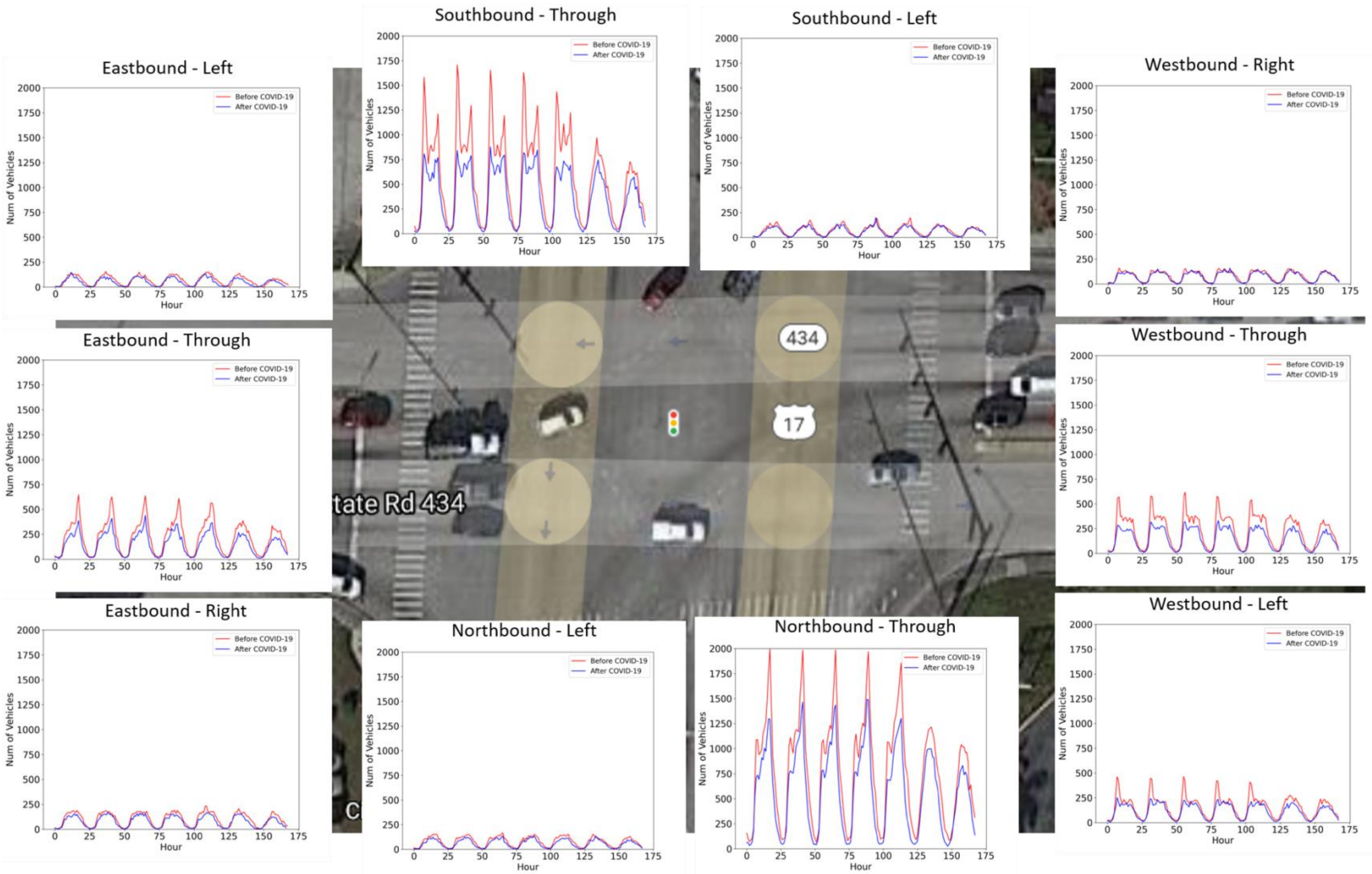



**Figure 5.6: Distribution of hourly traffic volume of an intersection (SR 434/US 17-92)**

## 5.5 DATA VALIDATION

To validate the quality of ATSPM data, we have compared the ATSPM data with manually counted data from FDOT project. We have selected seven signals – **US 17-92 & SR 434**, **SR 434 at Chapman Rd**, **SR 434 at Carrigan Ave**, **SR-426 at Oviedo Mall Blvd**, **SR-426 at Winter Springs Blvd**, **SR 419-434 at Tuskawilla Rd**, **SR 426 at Red Bug Lake Rd** from two time periods (2016, 2019) to make the comparison. The information of detector number for each signal is included in **Table 5.3**.

**Table 5.3: Number of detectors available in ATSPM data in each signal, cross-checked with TMC data**

| Signal ID | Signal Name | Date | Direction | | | |
|---|---|---|---|---|---|---|
| | | | Northbound | Southbound | Eastbound | Westbound |
| 1130 | US 17-92 & SR 434 | 01/16/2016 | 4 | 4 | 2 | 2 |
| 1405 | SR 434 at Chapman Rd | 09/19/2019 | 3 | 3 | 0 | 0 |
| 1410 | SR 434 at Carrigan Ave (2016) | 10/20/2016 | 3 | 3 | 0 | 0 |
| 1410 | SR 434 at Carrigan Ave (2019) | 09/10/2019 | 3 | 3 | 0 | 0 |
| 1830 | SR-426 at Oviedo Mall | 10/29/2016 | 2 | 2 | 0 | 0 |
| 1835 | SR-426 at Winter Springs | 10/29/2016 | 2 | 2 | 0 | 0 |
| 1365 | SR 419-434 at Tuskawilla Rd | 10/26/2016 | 0 | 0 | 1 | 1 |
| 1825 | SR 426 at Red Bug Lake Rd | 09/19/2019 | 2 | 2 | 2 | 2 |

We apply the GEH statistic as evaluation metric to compare the two sets of traffic volume. The detail of GEH statistic can be seen in (TfL, 2010). For traffic modelling work in the "baseline" scenario, a GEH of less than 5.0 is considered a good match between the modelled and observed hourly volume. The formula of GEH score can be seen as equation (5.1).

$$GEH = \sqrt{\frac{2(M-C)^2}{M+C}} \tag{5.1}$$

Where M is hourly traffic volume counted by ATSPM database and C is manually count data from FDOT. In the comparison, the total volume represents the movement type ‘through’ in ATSPM data and ‘through & right’ in FDOT count data. When the ATSPM data can identify the movement type ‘left’ and ‘right’, we will compare all the movement types, respectively.

**Table 5.4** demonstrates the comparison results (in terms of GEH score) for ATSPM data against FDOT manual count data for seven intersections. Among them, three of the intersections: ***SR 434 & US 17-92, SR-426 & Oviedo Mall (2016), SR-426 & Winter Springs (2016)*** have GEH scores less than 10 across all the data samples (traffic movement count) collected for different movement types and directions. That means for all these intersections traffic volume counted by ATSPM detectors match well with manually counted traffic volume, in other words all the ATSPM detectors are properly functioning for these intersections.

In case of ***SR 434 & Chapman Road*** intersection GEH scores are less than 10 for all the ATSPM data samples collected from northbound detectors, for southbound detectors a few data samples (11% of the total samples) have GEH score greater than 10. Overall, for ***SR 434 & Chapman Road*** intersection the ATSPM detector-based count data comply with the manual count **(Table 5.4)**. For ***SR 434 & Carrigan***, we compare the ATSPM data samples with FDOT count data collected on 10/20/2016. The results show that all the ATSPM data samples are well matched with FDOT count data (GEH score is less than 10) for both northbound and southbound detection. However, for the same intersection when we compare the ATSPM data with FDOT count data collected on 9/10/2019, only northbound ATSPM detectors’ data samples comply (GEH scores less than 10) with manual counts. However, for the southbound ATSPM detectors, the GEH score is higher than 10 for all the data samples. We also check the traffic volume counted by each detector in southbound of SR ***434 & Carrigan Avenue***, and we found one of three detectors work well but the other two detectors cannot detect the traffic volume which indicates the two detectors may have problems.

For ***SR 419-434 & Tuskawilla Road*** intersection we compare the ATSPM data samples with FDOT manual count data collected on 10/26/2016. The comparison results indicate that the GEH scores for all the data samples are higher than 10, such anomalous results indicate the chances of detector malfunction. Thus, we check the detector number of this signal. We found both of eastbound and westbound on ***SR 419-434&Tuskawilla Road*** have two lanes. However, in ATSPM database, only 1 detector/direction are used, which means that the traffic volume counted by ATSPM data are nearly half of the real-world traffic volume. So, when we adjust the traffic volume by multiplying the ATSPM count data samples by a factor of 2, the GEH scores become less than 5 for all the samples **(Tables 2.4)**. Similarly**,** we find the comparison results of ***SR 426 & Red Bug Lake Road*** are bad since all the GEH score are higher than 10, and we find all the detectors of ***SR 426 & Red Bug Lake Road*** can identify the traffic volume indicating the detectors may have problems.

**Table 5.4 Summary of the comparison results**

| Intersection Location | Date | Calculation References | Direction | % of Samples with GEH < 5 | % of Samples with GEH < 10 | Remarks |
|---|---|---|---|---|---|---|
| *SR 434 & US 17-92 (2016)* | 1/16/2016 | Appendix A Table 3.5 | Northbound | 83.3 | 100 | All the detectors were properly functioning |
| | | | Southbound | 50 | 100 | |
| | | | Eastbound | 50 | 100 | |
| | | | Westbound | 44.4 | 100 | |
| *SR 434 & Chapman Road (2019)* | 9/19/2019 | Appendix A Table 3.6 | Northbound | 100 | 100 | All the detectors were properly functioning |
| | | | Southbound | 44.4 | 88.9 | |
| *SR 434 & Carrigan Avenue (2016)* | 10/20/2016 | Appendix A Table 3.7 | Northbound | 100 | 100 | All the detectors were properly functioning |
| | | | Southbound | 100 | 100 | |
| *SR 434 & Carrigan Avenue (2019)* | 9/10/2019 | Appendix A Table 3.8 | Northbound | 100 | 100 | Two of the detectors in southbound direction were malfunctioning |
| | | | Southbound | 0 | 0 | |
| SR-426 & Oviedo Mall (2016) | 10/29/2016 | Appendix A Table 3.9 | Northbound | 100 | 100 | All the detectors were properly functioning |
| | | | Southbound | 100 | 100 | |
| *SR-426 & Winter Springs (2016)* | 10/29/2016 | Appendix A Table 3.10 | Northbound | 100 | 100 | All the detectors were properly functioning |
| | | | Southbound | 100 | 100 | |
| | 10/26/2019 | Appendix A Table 3.11 | Eastbound | 0 | 0 | |
| | | | Westbound | 0 | 0 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| *419-434 & Tuskawilla Road (2016)* | | | Eastbound* | 100 | 100 | One of the detectors was malfunctioning in each direction, missing about half of the traffic. |
| | | | Westbound* | 100 | 100 | |
| *SR 426 & Red Bug Lake Road (2019)* | 9/19/2019 | Appendix A Table 3.12 | Northbound | 0 | 12.5 | All the detectors were malfunctioning |
| | | | Southbound | 0 | 0 | |
| | | | Eastbound | 0 | 0 | |
| | | | Westbound | 0 | 0 | |

* multiplied the ATSPM count by a factor of 2 to adjust the missing traffic in each direction

### 5.6 ATSPM Movement Volume Data Preprocessing

The raw data collected from traffic detectors are subjected to errors. Several factors such as detector's malfunctioning, false encoding during storing the data into the server, overlapping of multiple entries, duplicate entries, bad weather conditions etc. can cause errors. Therefore, before proceeding to any data analysis, we need an extensive data cleaning and quality checking**. Figure 5.7** shows the framework for the data processing steps.

To check the quality of the data, we followed several steps starting with checking the percentage of missing data for different detectors. We consider the detectors having a higher percentage of missing values as unreliable ones. Moreover, data imputation is not feasible for the detectors with too many missing values as it will produce unrealistic data distribution. Considering this issue, we retain the detectors having missing values less than 20% of total data samples. We apply three techniques to check the quality of the data and detecting outliers.

First, we compare vehicle per hour for each movement types (i.e., Left, Through, Right) with Federal Highway Administration (FHWA) guidelines on maximum capacities at signalized intersection. According to the FHWA guidelines for left and right turn movement the capacities vary from 150-350 veh/hr, while for through movement the capacity varies from 1600-2100 veh/hr depending on number of available lanes (FHWA, 2004). We observe the distribution of the hourly turning movement volume for different movement types. Almost all the data samples, except a few (less than 0.1%), have hourly volume less than capacity. We consider the samples with values greater than capacity as outliers.

Second, we apply a machine learning algorithm to detect the outlier from temporal pattern of the traffic movement volume. The algorithm learns traffic pattern for different hours of the day and days of the week and isolates the outliers which show unusual pattern.

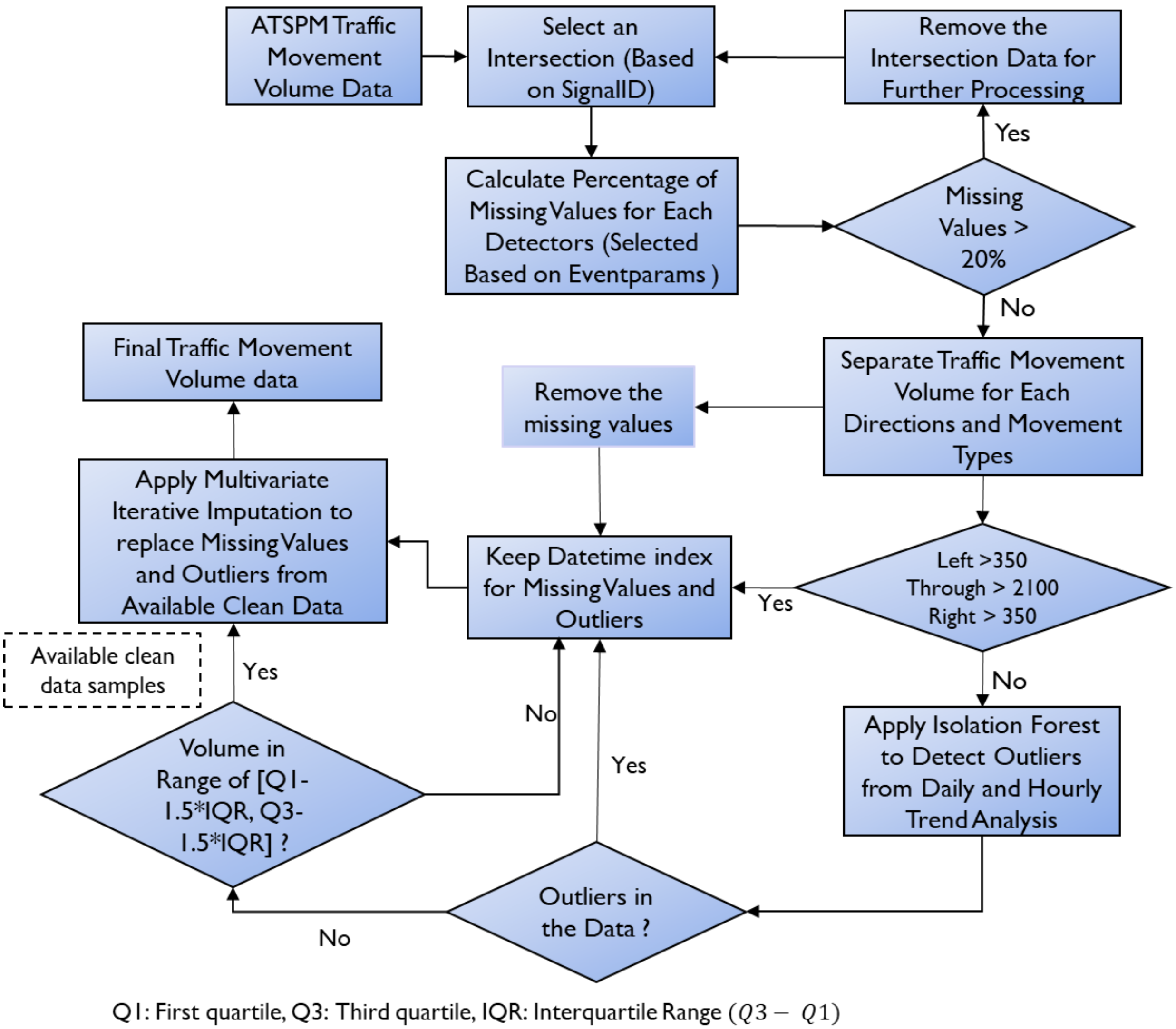



**Figure 5.7**. Framework for data processing

Third, we check if the turning movement volume remains within the range between $(Q1 - 1.5 * IQR)$ and $(Q3 - 1.5 * IQR)$, where $Q1$ and $Q3$ indicates the first and third quartile and IQR indicates the interquartile range $(Q3 - Q1)$.

Finally, we use a technique known as multivariate iterative imputation adapting Bayesian ridge regression as estimator to impute the missing values and outliers. To fit the estimator, we use time of the day (hour), day of the week (day), and volume with missing values as inputs. For each imputation the algorithm takes a sample from gaussian posterior of the fitted estimator. We use Python scikit learn (Pedregosa et al., 2011) library to implement the algorithm. The details about the data imputation algorithm are provided in reference (van Buuren and Groothuis-Oudshoorn, 2011). **Figure 5.8** shows the distribution of through movement volumes before and after data processing.

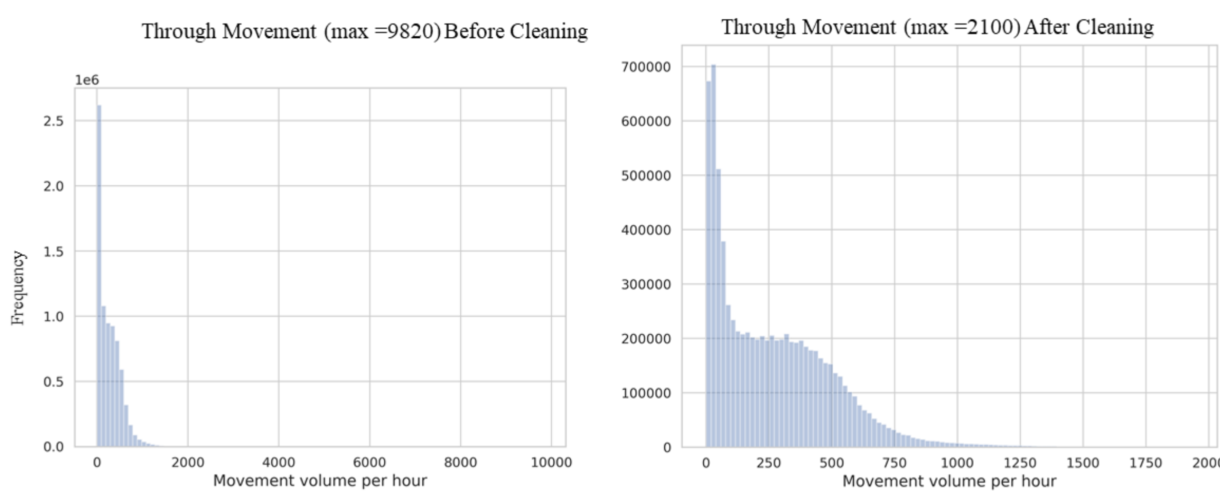



**Figure 5.8** Distribution of data samples based on vehicles per hour per lane

From the distribution, we find that some approaches may have more detectors compared to the number of available lanes due to the presence of multiple detectors in a lane. Since we are determining the values of the target variables (volume of different movement types) by individual detectors, data from multiple detectors in a given lane are likely to overestimate the values of target variables for that lane. To remove the additional detector in a lane, we first identified the approaches likely to have multiple detectors and then compared distances from stop bar to remove the additional detectors. All the removals have been manually checked to ensure that we keep only one detector per lane for each approach. **Table 5.4** gives a sample of the signals with multiple detectors.

**Table 5.5** List of selected intersections with multiple detectors per lane

| Signal ID | Primary Name | Movement Types | Number of Lanes | Number of Detectors | Remarks |
|---|---|---|---|---|---|
| 1155 | SR-436 & US 17-92 (ADT) | ET | 3 | 6 | Traffic Volume higher than capacity |
| | | WT | 3 | 6 | |
| 1165 | US-17/92 & Prairie Lk (ADT) | ST | 3 | 6 | Traffic Volume higher than capacity |
| 1585 | SR-436 & Oxford | ET | 3 | 7 | Traffic Volume higher than capacity |
| | | WT | 3 | 7 | |
| 2080 | Lake Mary Blvd & Sun Dr (ADT) | ET | 3 | 6 | Traffic Volume higher than capacity |
| | | WT | 3 | 6 | |
| 2095 | Lake Mary Blvd & I-4 East (ADT) | ET | 3 | 6 | Traffic Volume higher than capacity |

### 5.6 SUMMARY

In this chapter, we describe the processing of ATSPM dataset which has been used to count the traffic volume of each signal. Then, we calculate the average hourly traffic volume for each signal over the one-year period (2019.7 – 2020.7) and plot the distribution of average hourly traffic volume of each signal for different periods. Moreover, we also investigate the temporal patterns of traffic volume for a given signal. The results of our exploratory analysis indicate that the ATSPM data are adequate to reveal spatio-temporal patterns of traffic volume for each signal.

For some signals, traffic data are not available for all directions (e.g., northbound, southbound, westbound, and eastbound) and/or all movement types (e.g., go-through, right, left). When implementing the proposed modeling framework, appropriate assumptions need to be made to account for the missing information.

To evaluate the quality of ATSPM dataset, we compare the ATSPM-based traffic volume with FDOT count data. The results show that some of detectors of signals can work well and some of the detectors may have problem when counting the traffic volume. With all the traffic volume data, we also described the framework of the potential demand prediction model.

# CHAPTER 6 FEATURE ENGINEERING

## 6.1 Introduction

Prior to implementing a model, it is important to find the appropriate combination of input features that produces best results. In this study, we have considered different spatiotemporal input features to represent the travel demand, demographic characteristics and built environment characteristics of the study area. We have considered these inputs at different spatiotemporal resolution. So, it is critical for us to establish a feature selection process to find the best model for long term traffic volume prediction. This chapter describes different input features we used for training the model. It also includes a feature selection process to finalize the best model.

## 6.2 Demand Related Features

To generate the travel demand for the study area, we employ Central Florida Regional Planning Model (CFRPM 6.1) daily model which outputs average weekday trip production and trip attraction at the Traffic Analysis Zone (TAZ) level by trip purpose and special generators. Our study area (Seminole County, Florida) includes 230 TAZs, so, we can aggregate the weekday trips to estimate total production and total attraction for each of these 230 zones. **Figure 6.1** demonstrates the spatial distribution of trip attraction and production for the year of 2015. However, each TAZ has multiple intersections within its boundary. Hence, to link this demand information with operational characteristics of an intersection, we need to partition these TAZs into a finer spatial resolution.

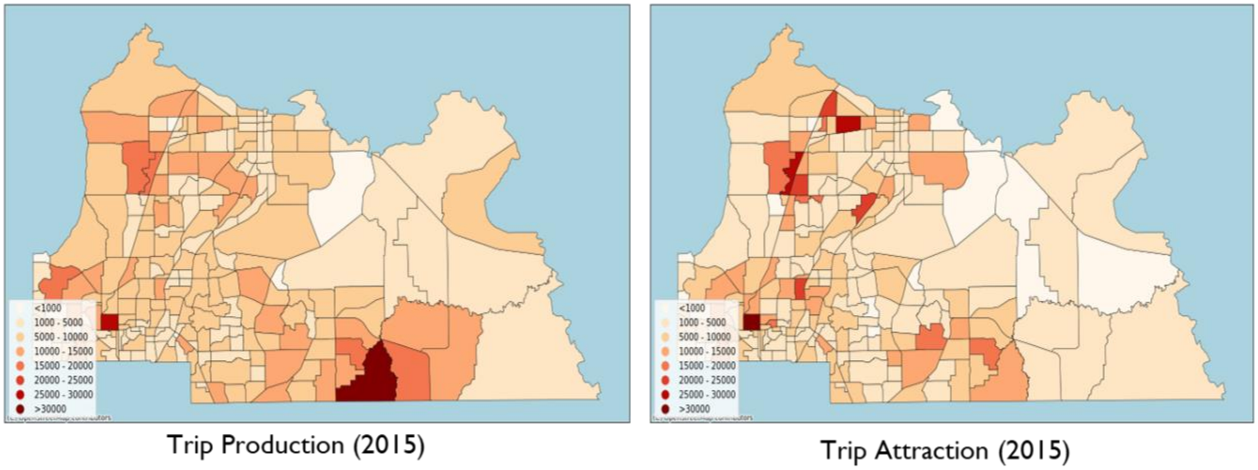



**Figure 6.1** TAZ level trip attraction and trip production

Considering this issue, we develop a technique to partition the TAZs into multiple subzones to generate demand at a finer resolution. Based on the frequency of the intersections within each TAZ, we divide them into 5 subzones. To create the subzones, we follow a three steps process: Firstly, we identify the centroid of each zone and the points in the perimeter that is used to create the polygon, i.e., TAZ boundary. Secondly, depending on the number of subzones (i.e. 5) for each TAZ, we identify the group of points that needed to be connected with the centroid. Finally, we continue this process until reaching to the last point. In total we create 1150 subzones from 230 TAZs. **Figure 6.2** show the TAZs and generated Subzones for the study area. From the CFRPM model we estimate average weekday trip attraction and production at a subzone. Afterwards, we expand the daily demand into hourly demand using hourly distribution factors provided by CFRPM (see (30) for more details).

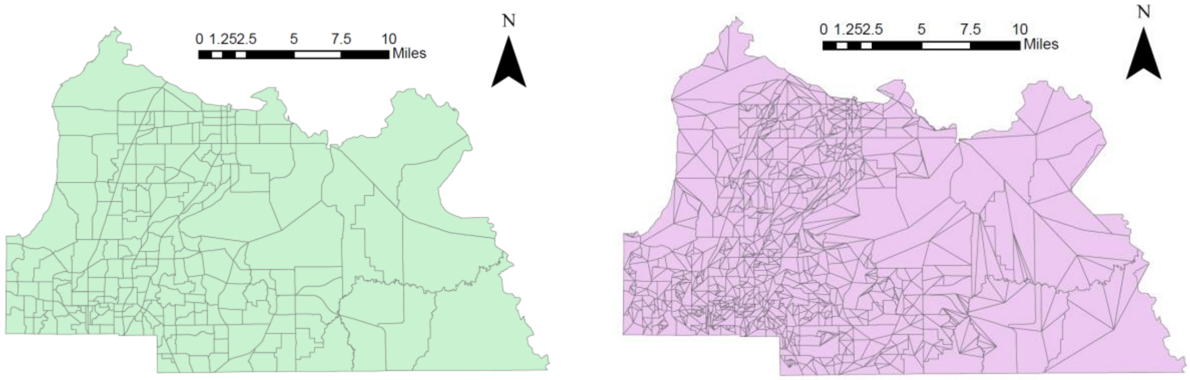



**Figure 6.2** TAZs and Sub-TAZs for the Analysis

We have also extracted exogenous variables including demographic, built environment and land-use characteristics for each subzone. The sociodemographic variables are sourced from American Community Survey (ACS) and include population, median income, and employment. Built environment characteristics variables are processed from NAVSTREET data and include number of restaurants, shopping centers, business centers, entertainment establishments and educational institutions. Land-use characteristics are processed using high resolution parcel level land-use data sourced from Florida Department of Revenue. Each parcel is assigned a unique ID (Parcel ID) linking it with equivalent parcel level attribute information such as property/feature value, land value, land area in square feet, land-use codes (DOR-UC), owner name, owner address, physical address, physical zip code, building details and so on contained in the Name-Address-Legal (NAL) file. Further, we also prepare "Land use mix" variable which is compute as $\frac{-\Sigma_k(p_k(\ln p_k))}{\ln N}$, where $k$ is the category of land-use, $p$ is the proportion of the developed land area, $N$ is the number of land-use categories within a zone/subzone. The sociodemographic, land use and built environment data has been generated at the zone, sub-zone and intersection level (0.5-mile buffer) resolution to evaluate the most appropriate features for model development.

### 6.3 Feature Selection

To prepare the final dataset for running the model we merge the instance of travel demand features with traffic movement volume data. While merging the TAZ or Subzone level data, we search the nearest TAZ/subzone from a given intersection based on Euclidian distance. Finally, we include total trip attraction, production, sociodemographic and built environment related features with the intersection volume. We also extract the seasonality related features such as time of the day, day of the week, workdays, and time period day (early morning, morning, peak hour etc. Thus, the input dataset has different features (independent variables) at different spatial resolution (TAZ, subzone and intersection level), Trip attraction and trip production, sociodemographic characteristics, built environment characteristics and features to capture seasonal/temporal trends (time of day, day of week, etc.).

**Table 6.1 List of spatiotemporal input features**

| Feature Types | Features Description | Resolution |
|---|---|---|
| CFRPM | Trip attraction and production | TAZ and Subzone level for average weekday and hourly |
| Socio Demographic Characteristics | Population, Employment | TAZ and Subzone level |
| Built environment characteristics | Industrial, Institutional, Recreational, Retail/Office, Others, Land mix, Urban Highway, Rural highway, Other road types. | TAZ, Subzone and Intersection level (0.5-mile buffer) |
| Temporal Features | Time of the day, day of the week, time periods, workday | N/A |

Now to finalize the input features we check the importance of all the available features by running an advanced Tree based model know as Xtreme Gradient Boosting (XGBoost). The model takes all the spatial and temporal features at different spatiotemporal resolution as input and predicts aggregated traffic volume for a given intersection. We train and test the model with input data from 2016. To find the best model, we use root mean square error (RMSE), mean absolute error (MAE) and $R^2$ score. **Figure 6.3** shows the feature importance for the best model. The RMSE, MAE and $R^2$ score for the model are 207.354, 124.875, 0.961 respectively. From this analysis, we find that sub zonal level features are more relevant compared to TAZ, and intersection level aggregation of the spatial demand related features. Moreover, built environment characteristics are highly important to predict intersection level traffic movement volume.

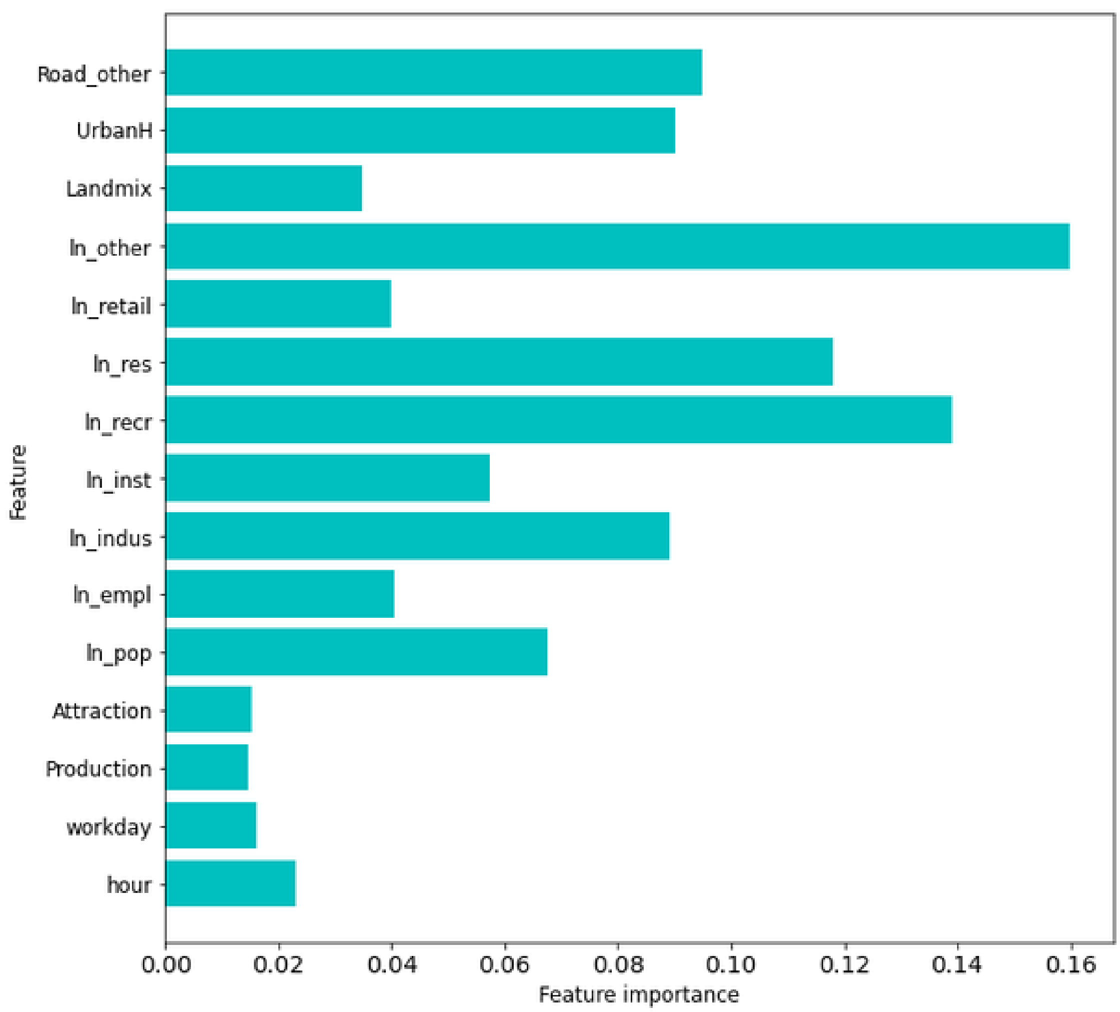



**Figure 6.3** Feature Importance from Extreme Gradient Boosting Model

## 6.4 Summary

This chapter provides details on data preprocessing techniques for model development, which includes description on multiple outlier detection techniques such as comparison with capacity, isolation forest to isolate the anomaly and interquartile range to check the data distribution. Finally, it describes the data fusion and feature extraction techniques to prepare the final dataset for model implementation

# CHAPTER 7 MODEL DEVELOPMENT AND VALIDATION

## 7.1 INTRODUCTION

In this chapter, we describe the model validation procedure for the test year (2019) data. We rigorously test the model based on the proposed performance metrics: RMSE, MAE, and $R^2$ values. We also measure the GEH score distribution to understand the model sensitivity over the variations of predicted values.

## 7.2 MODELING FRAMEWORK

In this component we develop a graph convolution based deep neural network architecture to learn traffic movement volume from traffic intersection level features such as volume, zonal or sub zonal level trip attraction and production. In this architecture we model the flow diffusion process considering drivers' route choice behavior. In the convolutional operation the traffic flow from the origin intersection to destination intersection is modeled using a graph convolution filter. We define convolutional filter based on weighted adjacency matrix, where we assign the weights based on travel distance between two nodes. Finally, we map this diffusion process using a long short-term memory recurrent neural network to predict traffic movement volume for a given time horizon. **Figure 7.1** shows the develop deep learning framework for predicting traffic movement volumes for the entire network.

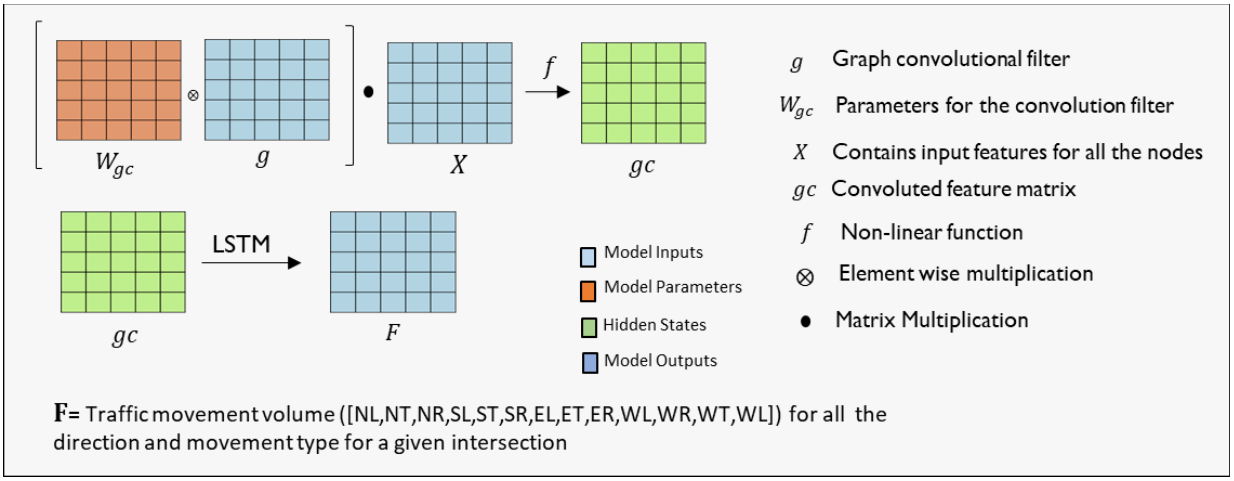



**Figure 7.1** Deep Learning Modeling Framework

The proposed high-fidelity model two layers, in the first layer, we define a graph convolutional operation based on the structure of the network and travel distance between two different nodes. So, in the first layer the model learns the network features and demand distribution for different nodes. We perform a convolution operation between the input features and graph convolution filter ($g$). While training the model, we will estimate the graph convolutional parameters ($W_{gc}$) for this convolutional filter.

The convolutional filter ($g$) represents the diffusion process of the traffic flows from the origin node towards destination nodes. We define convolutional filter based on weighted adjacency matrix, where we assign the weights based on links' travel distance. Output ($gc$) from the first layer is fed into a LSTM model to predict traffic movement volume ($F$) for all the possible directions (i.e., left, though, and right) for each approach (e.g. four approaches for a four-legged intersection) of an intersection. **Figure 7.2** demonstrates the input features and target features of the implemented model for long-term traffic prediction.

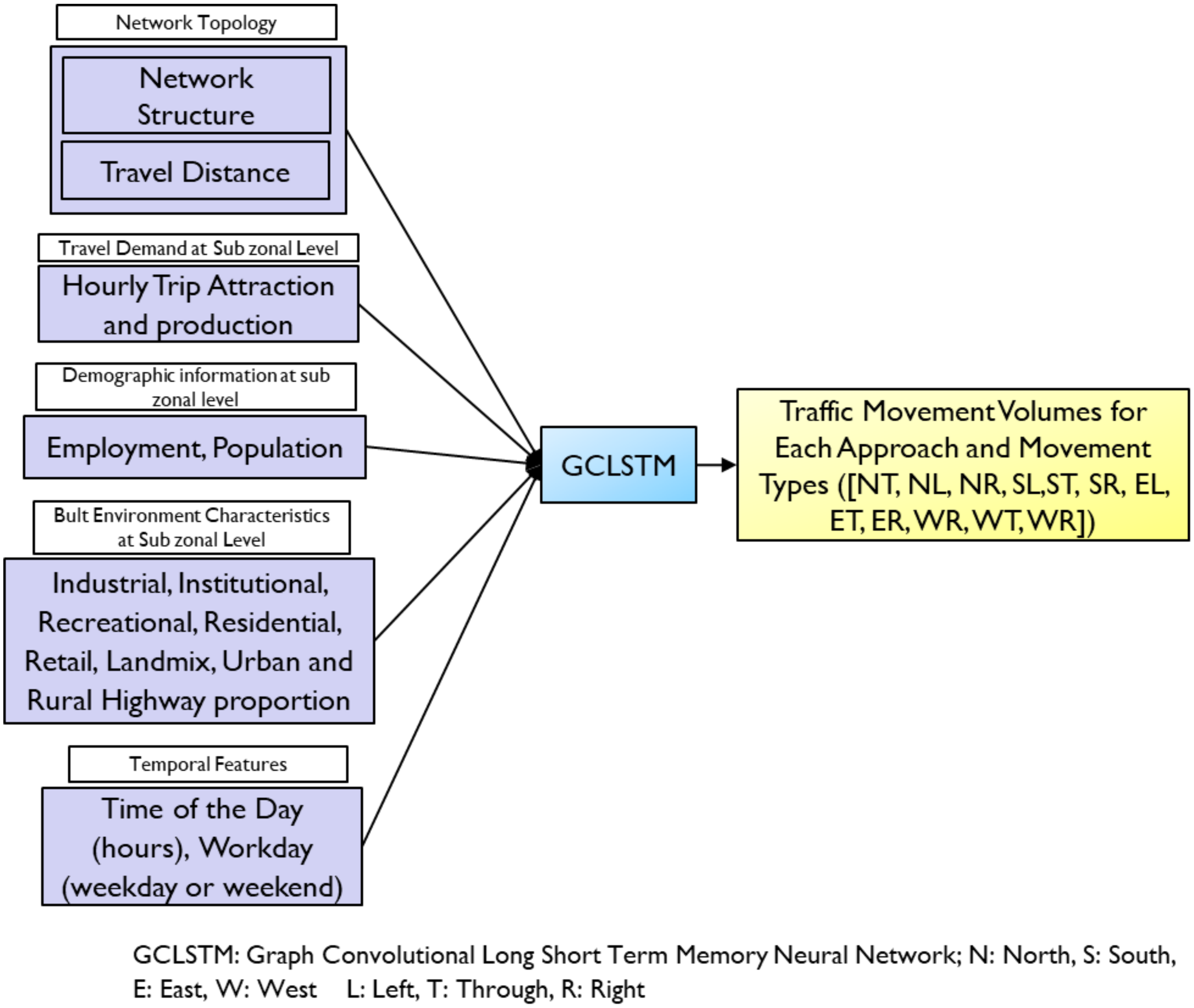



**Figure 7.2** Implemented prototype model framework for traffic movement volume prediction

## 7.3 MODEL TRAINING

We have retrained the final model using 2016 traffic movement volume data. We use 80% of the data for training (learning the parameters), 20% for validation (tuning hyper-parameters). Based on the validation accuracy we tune the hyper-parameters such as learning rate, types of activation functions (i.e., tanh, sigmoid etc.), maximum number of iterations. Once the final model parameters are fixed, we test it on 2019 data set to understand the model performance for long term prediction year data.

We train the model using mean squared error (MSE) as the loss function. At each iteration, the model estimates the MSE for the estimated traffic movement volume ($\hat{F}_m^{ij}$) and the actual movement ($F_m^{ij}$) volume for a given intersection ($m$), direction ($i$), and movement type ($j$). Afterward, the gradient of the loss function is backpropagated to adjust the weights to reduce loss function value. The loss function can be defined as,

$$L_m = Loss\left(F_m^{ij}, \hat{F}_m^{ij}\right) \tag{1}$$

$$MSE = \frac{1}{m*i*j} \sum_{m=1}^{m} \sum_{i=1}^{i} \sum_{j=1}^{j} (F_m^{ij} - \hat{F}_m^{ij})^2 \tag{2}$$

where, $Loss(.)$ is the function to estimate the error between the actual ($F_m^{ij}$) and estimated values ($\hat{F}_m^{ij}$) and $i$ denotes the set of links for the network. In this study, we estimate mean squared error (MSE) as a loss function. It takes about 100 iterations for the model to converge to a stable solution **(Figure 7.3)**, after that there is merely any variation in loss values.

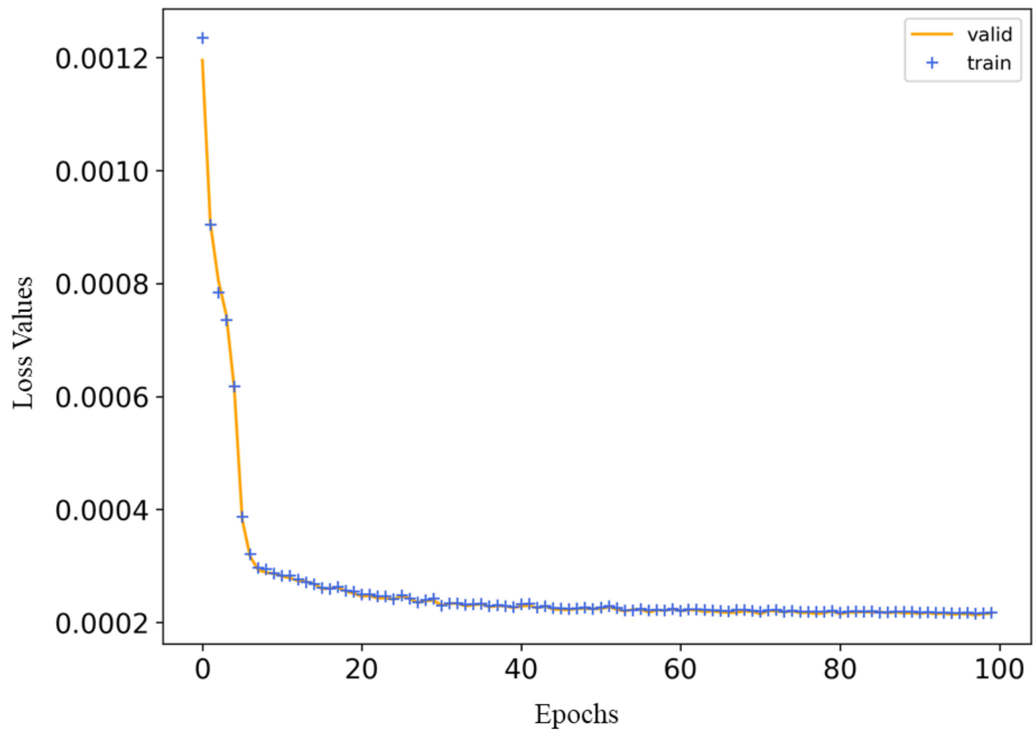
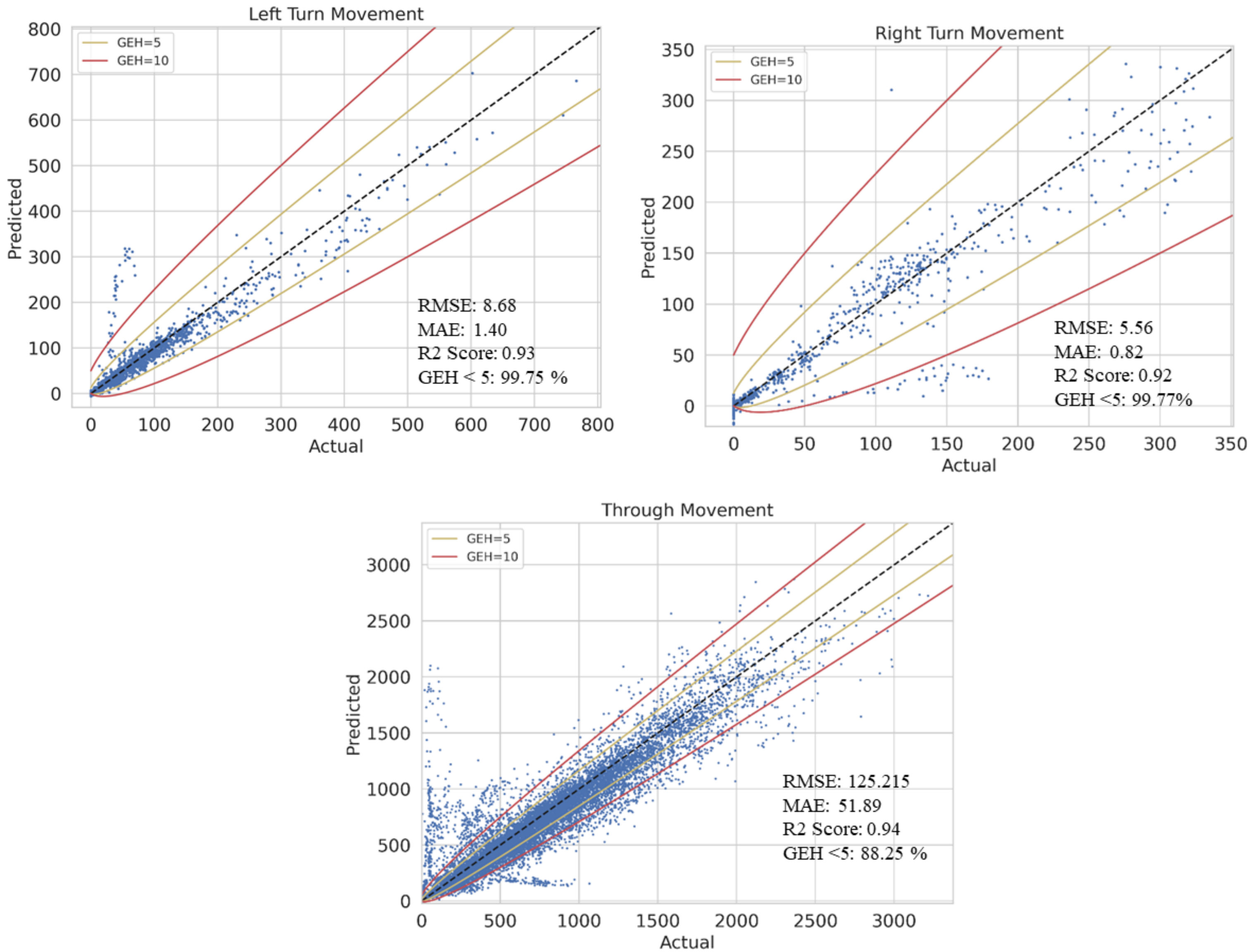


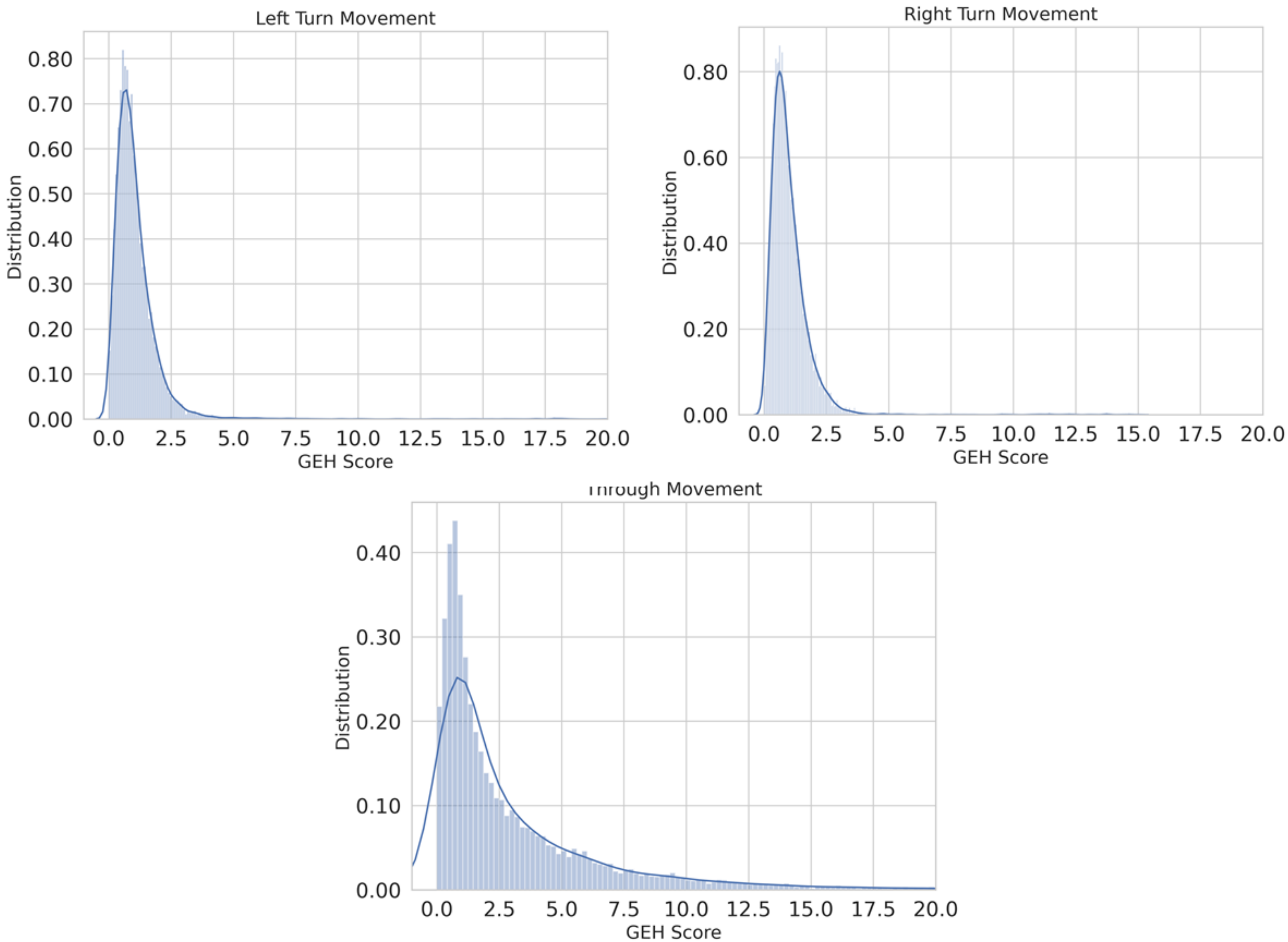

**Figure 7.3** Variation of loss (mean squared error) for different epochs (or iteration)

### 7.4 MODEL PERFORMANCE ON TEST DATA SET (2019)

We apply the final model to predict hourly average traffic movement volume for weekday/weekend. **Figure 7.4 (a)** shows the relation between actual and predicted volumes along with the estimated performance metrics. For each movement type the R2 score is greater than 0.90 which indicates that the model can capture the variations of traffic movement volumes well. Moreover, for through movement, the RMSE and MAE values are higher compared to left and right movement. The possible reason for these results is that overall traffic volume for through movement is very high compared to other two movement types. We also estimate the GEH score to compare the actual traffic with predicted volume, **Figure 7.4 (b)** shows the distribution of GEH score for all the movement types. From the figure we find that, for each movement types more than 85% of the predicted volumes have a GEH score less than 5.




(a) Comparison between actual and predicted traffic volume (training data from 2016, test data from 2019)




(b) Distribution of GEH scores for different movement types

**Figure 7.4** Performance of the model in turning and through movement prediction hourly average for weekday/weekend

## 7.5 TRAFFIC MOVEMENT VOLUME PREDICTION FOR THE FUTURE YEAR (2023)

To better predict the traffic volume in 2023, we have retrained the final model using traffic movement volume data from the most recent year 2019. In the data processing, we remove all the holidays and Hurricane Dorian data (8/24/2019 – 9/10/2019), which are considered as outliers. We use 80% of the data for training (learning the parameters), 10% for validation (tuning hyper-parameters). Based on the validation accuracy we tune the hyper-parameters such as learning rate, types of activation functions (i.e., tanh, sigmoid etc.), maximum number of iterations. Once the final model parameters are fixed, we test it on 2019 data set to evaluate model performance.

We apply the final model to predict hourly average traffic movement volume for weekday/weekend. **Figure 7.5** shows the relation between actual and predicted volumes along with the estimated performance metrics. For each movement type the $R^2$ score is greater than 0.98 which indicates that the model can capture the variations of traffic movement volumes well. In addition, we calculate the distribution of GEH score for each movement. It shows that for through movement type, more than 91% of the predicted volumes have a GEH score less than 5; and for left and right movement types, more than 99% of the predicted volumes have a GEH score less than 5. The results show a higher accuracy compared to the model trained over 2016 data.

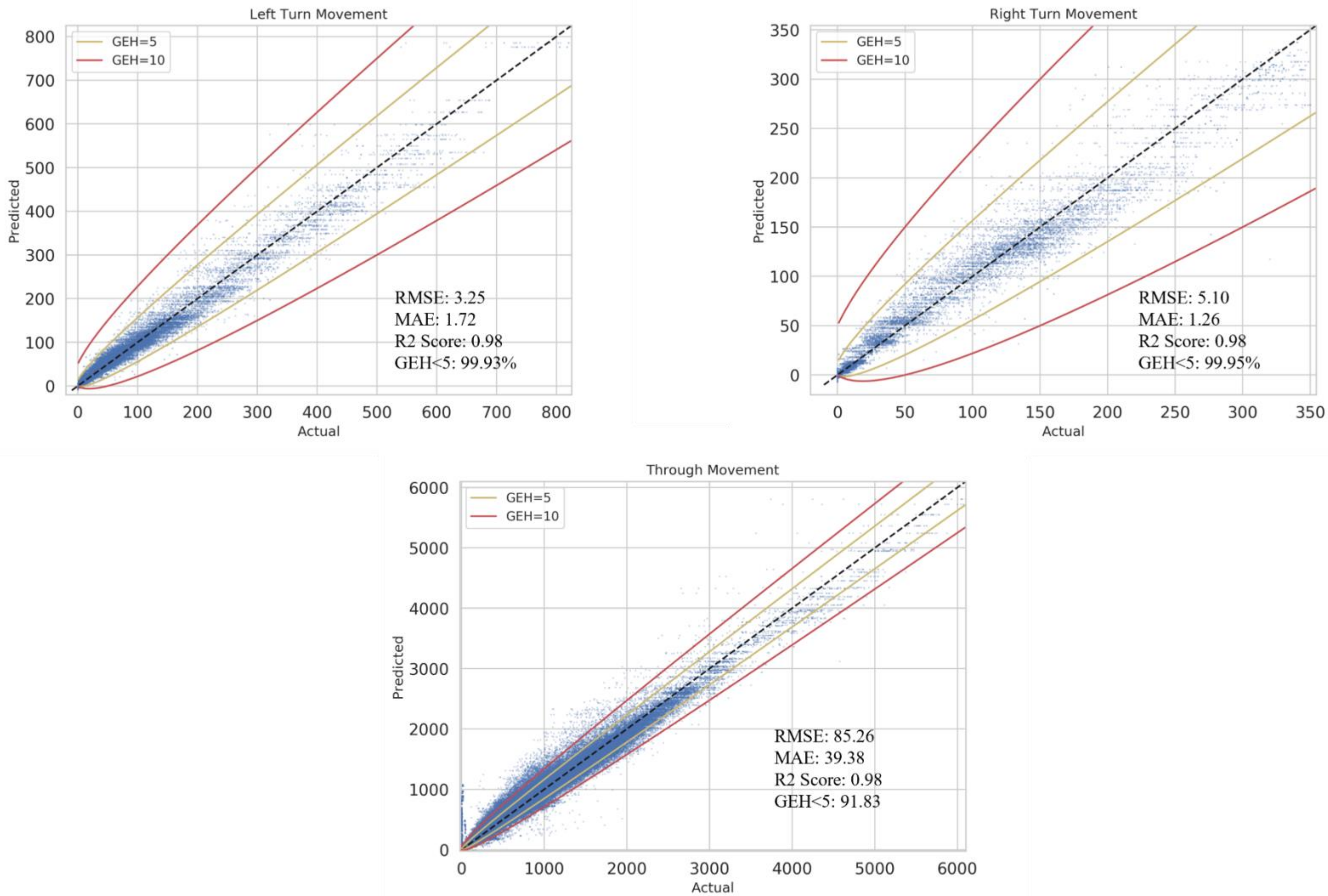



**Figure 7.5** Comparison between actual and predicted traffic volume (training data from 2019, test data from 2019)

We apply the final model to predict hourly average traffic movement volume for 2023. Since the ground truth of traffic volume for 2023 is not available, we evaluate the performance of the model results with the actual value of 2019. **Table 7.1** shows the average traffic volume of each movement type and day type for years 2016, 2019 and 2023 (projection year). The traffic volume of 2016 and 2019 indicate actual values and the traffic volumes of projection year 2023 indicate predicted values. Our projection results indicate that both in weekday and weekend, the traffic volume for each movement will increase in 2023. The increase rates in a typical weekday are from 14%- 20% and in a typical weekend are from 5% - 9%.

**Table 7.1** Summary of actual traffic volume for 2016, 2019, and predicted traffic volume for 2023

| | Movement Type | 2016 - actual value | 2019- actual value | 2016 - 2019 growth rate | 2023 – predicted value | 2019 - 2023 growth rate |
|---|---|---|---|---|---|---|
| Weekday | Left | 212.6 | 267.2 | +25.68% | 320.7 | +20.01% |
| | Through | 1247.6 | 1605.5 | +28.68% | 1833.5 | +14.20% |
| | Right | 104.1 | 137.0 | +34.60% | 163.8 | +19.57% |
| Weekend | Left | 185.4 | 244.6 | +31.93% | 267.5 | +9.32% |
| | Through | 1078.8 | 1480.2 | +37.25% | 1591.2 | +7.50% |
| | Right | 91.9 | 129.3 | +40.79% | 136.9 | +5.89% |

### 7.6 SUMMARY

This chapter provides details on modeling framework, and the performance of model training and testing. We have trained the proposed model with two years traffic volume data – 2016 and 2019 separately. The model trained with 2016 traffic volume data is tested by 2019 traffic volume data which shows the potential of the model for future year prediction. Finally, we trained the model with 2019 traffic volume data and predict 2023 traffic volume data. The model results show the developed model can work well on the future traffic volume prediction.

# CHAPTER 8 DATA VISUALIZATION PLATFORM

## 8.1 INTRODUCTION

To visualize the results of the project, we have developed a data visualization platform. This chapter provides a detailed description of different components of the developed data visualization platform. The visualization platform can be used for providing more intuitive information of the ATSPM dataset to anyone using the data. This can also be used to check the quality of the data and improve the maintenance of the detectors. The platform is divided into three parts: *Introduction*, *Data Quality* and *Traffic Prediction*. In the *Introduction* part, we briefly introduce the projects problems and the solutions. The *Data Quality* part is mainly used for providing more intuitive information of the dataset which can be used to check the data quality and improve the maintenance of the detectors. In the *Traffic Prediction* part, we present the prediction results of the developed traffic forecasting framework.

## 8.2 PLATFORM COMPONENTS

### 8.2.1 PLATFORM INTRODUCTION

At the landing page of the visualization platform, known as the *Introduction part* (shown in **Figure 8.1**), we mention the problems and issues of the project as follows:

- Due to the high population and economic growth, urban regions of Florida are facing increasing traffic congestion and other mobility related challenges such as traffic accidents, excessive fuel consumption etc.
- To overcome such challenges, we need robust methods for accurate and reliable prediction of traffic to ensure the success of capacity improvement and traffic management projects.

- However, traditional transportation planning models forecast traffic from a strategic perspective over a longer-term horizon.
- In addition, these frameworks are not suited to address high-resolution network structure challenges (such as signal times and signal bays) - failing to account for bottlenecks and hidden bottlenecks in the network, often at intersections.
- Employing traditional methods for identifying capacity needs between intersections could result in recommendations that are unwarranted or potentially less efficient.

Besides, we provide our solutions as follows:

We have developed a hybrid travel demand forecasting tool that takes intersection-level historical traffic movement data, OD data from planning models, and a host of land use variables as inputs to predict projected intersection-level hourly traffic movements after 3 years.

In this part, we also present the study area (Seminole county) and all the 166 selected signalized intersections.

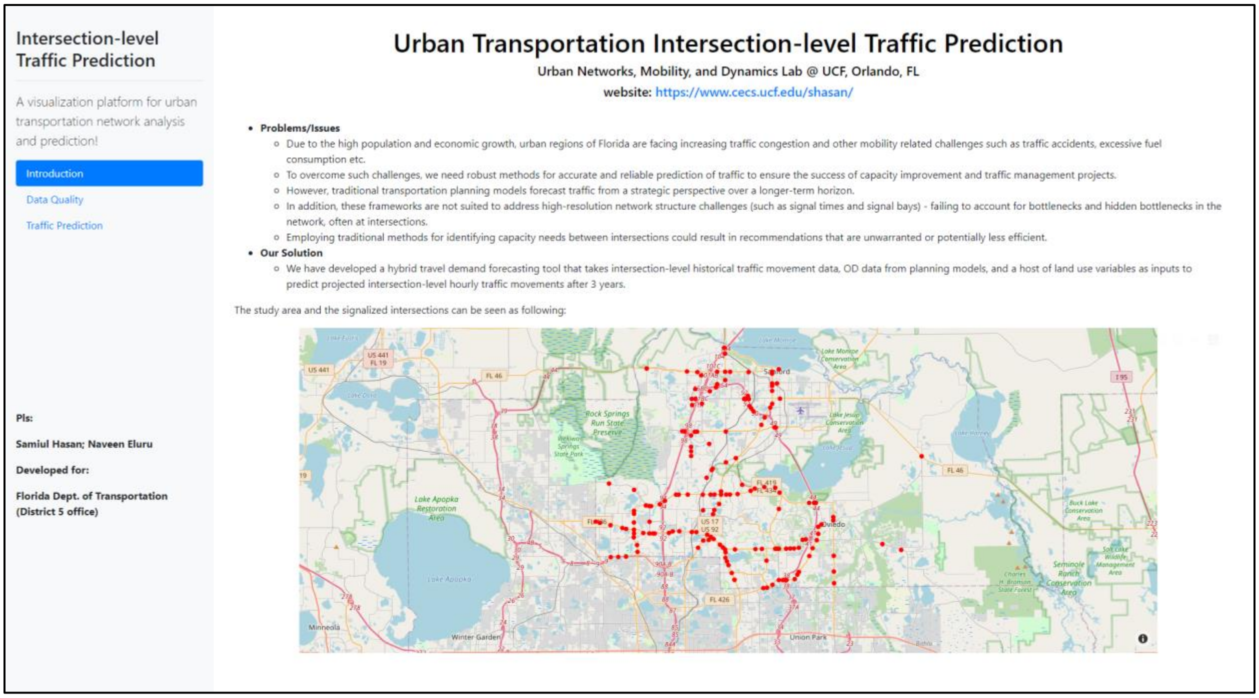



**Figure 8.1** Introduction part of the visualization platform

### 8.2.2 DATA QUALITY

In the *Data Quality* part, we use the ATSPM data in 2016 and 2019. Since the detectors of signal may have problems in detecting the vehicles, we need to evaluate the quality of the data collected by the detectors. We mainly consider two factors identifying whether the detectors work well: hourly volume and GEH score between the detectors with the same direction and movement type. For hourly traffic volume of the signals, we can find whether the signal has outliers in terms of the traffic volume. For example, if the hourly traffic volume of a regular signal is too low (less than 50), we can think that the detectors in the signal cannot detect the accurate traffic volume. Besides, in a signalized intersection, one direction will have different lanes and each lane will have a detector. Thus, to identify whether all the detectors work well in a signalized intersection, we calculate the GEH score between detectors with the same direction and movement type.

The *Data Quality* part mainly contains three parts – input, map and results, shown as **Figure 8.2**. In the input part, we can select the signal ID (with specific location name), the year of the data and the visualization type and corresponding detectors or detector pair. The visualization type can be divided into volume part and GEH score part. If we choose volume part, the 'Detectors/Detector_Pair' will show all the detectors of this signal ID, and if we choose GEH score, 'Detectors/Detector_Pair' will show the detector pairs (two detectors with the same direction type and movement type of the signal). In the map part, it will show the specific location of the signal selected by signal ID in a map.

In the results part, it will plot the corresponding figures based on the input. If we choose 'Volume' in the 'Visualization Type' and specific detector ID in 'Detectors/Detector_Pair', the output part will show the hourly traffic volume heatmap across the selected year. The distribution of the hourly traffic volume can help identify whether the traffic volume of a signal contain any outliers. Besides, if we choose 'GEH Score' in the 'Visualization Type' and specific detector pair in 'Detectors/Detector_Pair', the output part will show the GEH score distribution between the detectors in the selected detector pair over the entire selected year.

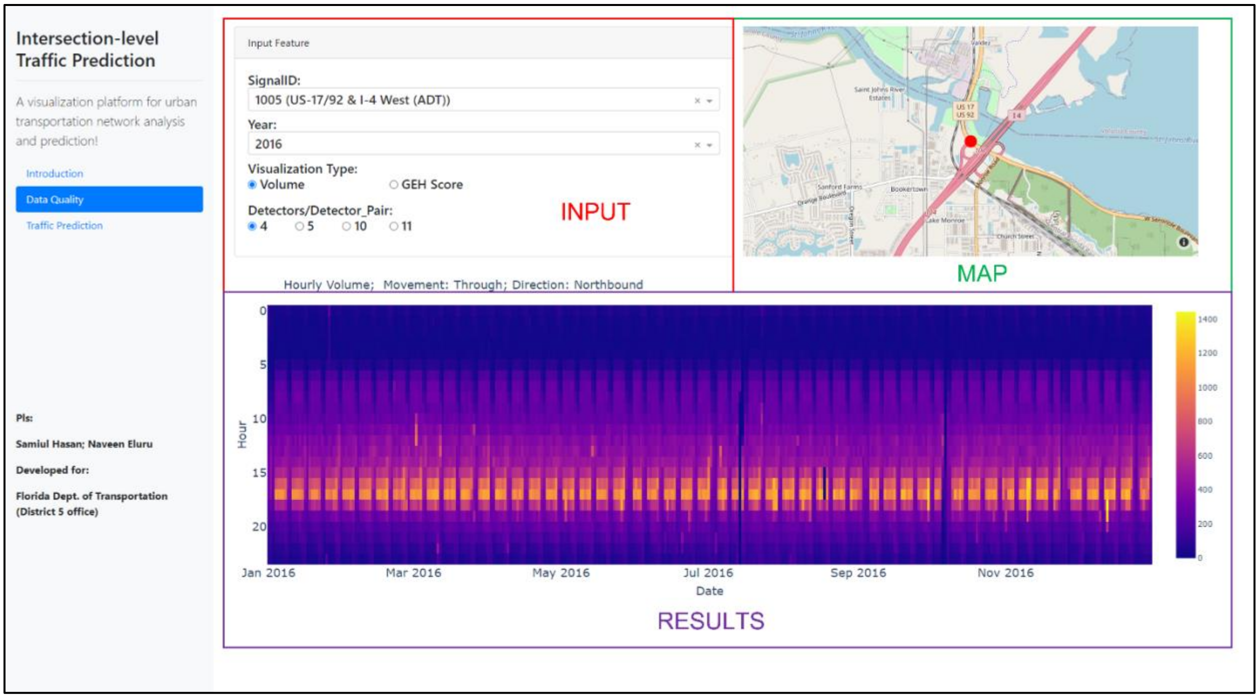



**Figure 8.2** Data quality part of visualization platform

### 8.2.3 TRAFFIC PREDICTION

In the ***Traffic Prediction*** page, we mainly have four parts – input, map, results with standard deviation, results with GEH score boundary, as shown in **Figure 8.3-8.4**. In the input part, we can select the specific intersection which is also displayed in the map part. Based on the modeling framework, the results can be presented in either workday or weekday. The develop model is trained with the 2016 data. From the year options, we can select 2019 which is the test year and the projection year 2023. For each intersection, we should select the specific movement type for showing the results.

In the results part, for the test year (2019), we show the actual value in two ways – actual value with one standard deviation and actual value with GEH score 5 boundaries. For all the study hours, the results show that almost all the predicted values are within one standard deviation of the actual value. And it also shows that the most of the GEH scores of predicted value and actual value are less than 5 which means the developed deep learning model performs well in predicting the next hour traffic volume. For the projection year (2023), we compare the results with the actual value of year 2019. And we also show the actual value of year 2019 with one standard deviation. Besides, we also present the GEH score 5 boundaries for the predicted value of year 2013.

From the results, the developed model can work well on most of the signalized intersections, which shows a big potential for future intersection planning and management.

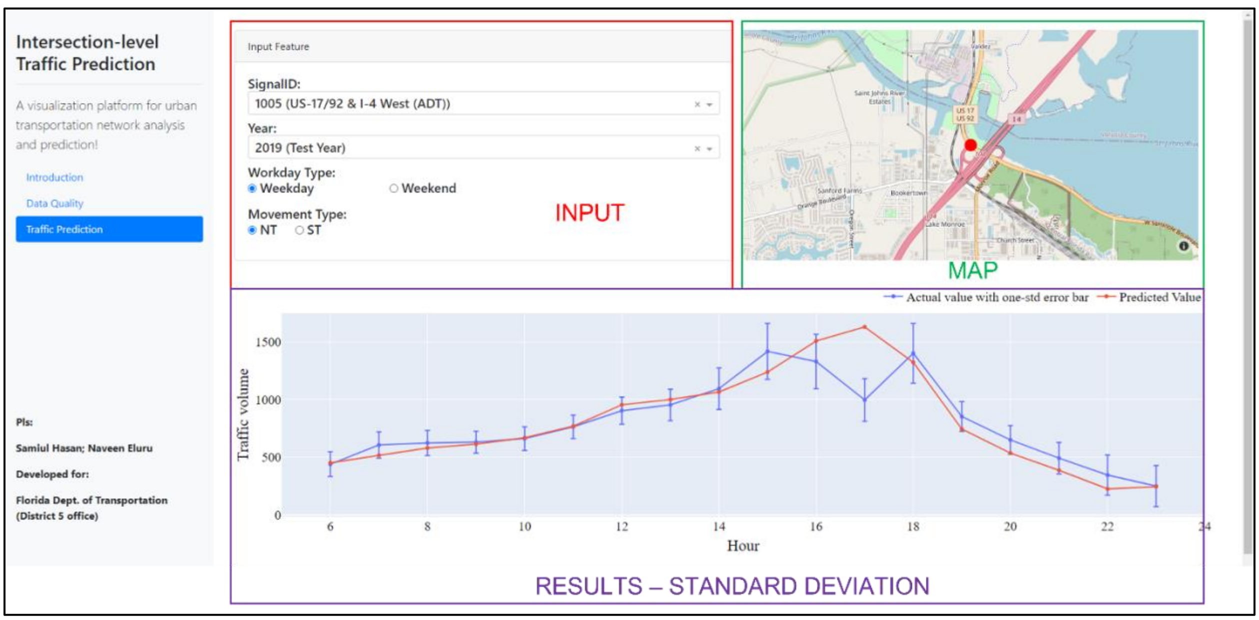



**Figure 8.3** Traffic prediction with one standard deviation of visualization platform

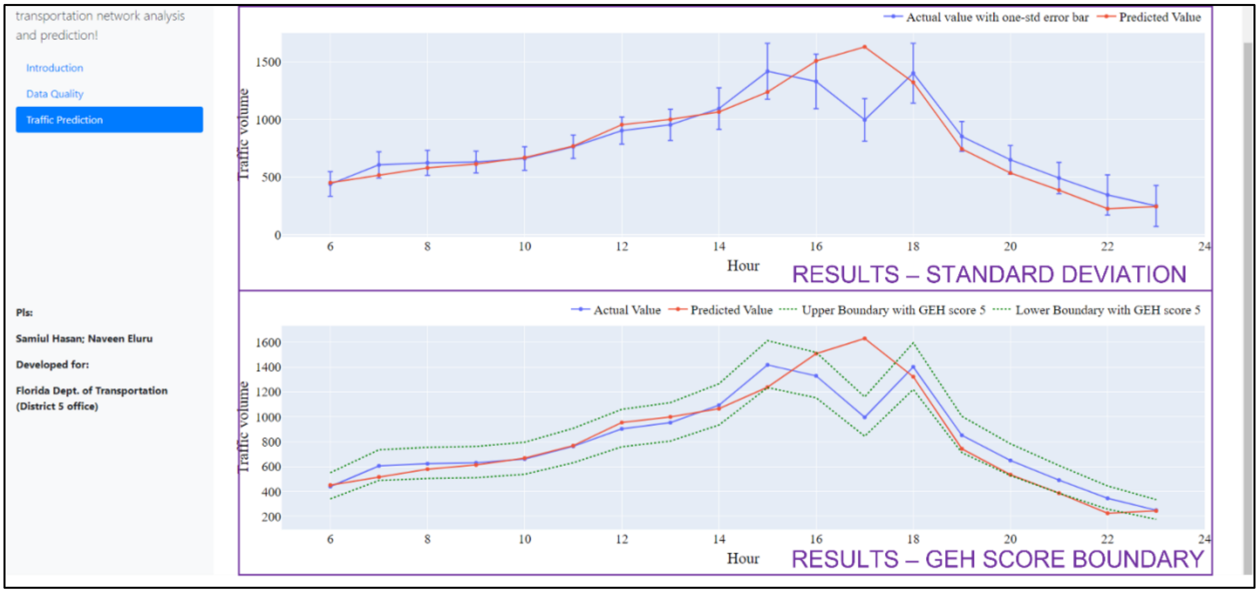



**Figure 8.4** Traffic prediction with GEH score boundary of visualization platform

## 8.3 SUMMARY

In this chapter, we present an overview of the visualization platform. In the platform, we have given a brief introduction about the problems and issues in this project and provide corresponding solutions. We use the hourly traffic flow and GEH score to show the data quality for each intersection. We train the proposed model with 2016 traffic demand data, test the model using 2019 data and predict the 2023 traffic volume using the developed model. In the traffic prediction page, we show the results for test year (2019) indicating how well the model work well for each intersection and for the projection year (2023) indicating the expected traffic volume for the year 2023.

# CHAPTER 9 CONCLUSION

In this project, we developed a high-fidelity hybrid traffic forecasting model for the accurate and reliable prediction of traffic ensure the success of capacity improvement and traffic management projects. By reviewing previous literatures, we found that traditional transportation planning models forecast traffic from a strategic perspective over a longer-term horizon. In addition, traditional forecasting frameworks are not suited to address high-resolution network structure challenges (such as signal times and signal bays) — failing to account for bottlenecks and hidden bottlenecks in the network, often at intersections. Besides, data-driven modeling for large scale transportation networks are less common. Few studies have proposed spatial and temporal learning-based methods predicting traffic of large-scale networks. However, these methods have not been tested in any realistic networks.

To develop the high-fidelity hybrid traffic forecasting model, various of datasets have been collected, such as Automated Traffic Signal Performance Measures (ATSPM), GIS data of Traffic Analysis Zones (TAZs) and compiled Central Florida Regional Planning Model (CFRPM) data, socio-demographic, built environment, and land-use characteristics. With the datasets, a graph convolution based deep neural network architecture has been developed to learn traffic movement volume from intersection level features such as volume, zonal or sub zonal level trip attraction and production. The results indicate that, for test year 2019, the $R^2$ score of each movement type is greater than 0.90 indicating that the model can capture the variations of traffic movement volumes well. For each movement types, more than 85% of the predicted volumes have a GEH score less than 5. To better predict the traffic volume for projection year 2023, we have retrained the model with 2019 traffic volume data. We then use the model to predict 2023 traffic volume. The results show that compared to 2019 traffic volume, the 2023 traffic volume will have a 14% - 20% growth in a typical weekday and 5% - 9% growth in a typical weekend.

To visualize the results of the project, we have developed a visualization platform. The visualization platform can be used for providing more intuitive information of the ATSPM dataset to anyone using the data and to check the quality of the data and improve the maintenance of the detectors.

## APPENDIX A. COMPARISION BETWEEN ATSPM DATA AND FDOT COUNT

### Table A.1: Comparison results of SR 434 & US 17-92 (2016)

| 01/16/2016 | NORTHBOUND | | | | | | SOUTHBOUND | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Time | FDOT COUNT - TOTAL | | ATSPM - TOTAL | | GEH | | FDOT COUNT - TOTAL | | ATSPM - TOTAL | | GEH | |
| | Left | Through | Left | Through | Left | Through | Left | Through | Left | Through | Left | Through |
| 9:00 - 10:00 | 160 | 875 | 105 | 898 | 4.78 | 0.77 | 172 | 781 | 88 | 661 | 7.37 | 4.47 |
| 12:00 - 13:00 | 206 | 1337 | 139 | 1338 | 5.10 | 0.03 | 227 | 1124 | 128 | 963 | 7.43 | 4.98 |
| 13:00 - 14:00 | 251 | 1337 | 148 | 1319 | 7.29 | 0.49 | 238 | 1062 | 132 | 888 | 7.79 | 5.57 |
| 16:00 - 17:00 | 174 | 1259 | 125 | 1275 | 4.01 | 0.45 | 218 | 924 | 115 | 807 | 7.98 | 3.98 |
| 17:00 - 18:00 | 170 | 1142 | 112 | 1147 | 4.88 | 0.15 | 216 | 879 | 119 | 744 | 7.49 | 4.74 |
| 19:00 - 20:00 | 119 | 890 | 76 | 877 | 4.35 | 0.44 | 124 | 585 | 77 | 524 | 4.69 | 2.59 |

| 01/16/2016 | EASTBOUND | | | | | | | | | WESTBOUND | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Time | FDOT COUNT - TOTAL | | | ATSPM - TOTAL | | | GEH | | | FDOT COUNT - TOTAL | | | ATSPM - TOTAL | | | GEH | | |
| | Left | Through | Right | Left | Through | Right | Left | Through | Right | Left | Through | Right | Left | Through | Right | Left | Through | Right |
| 9:00 - 10:00 | 142 | 493 | 186 | 95 | 359 | 153 | 4.32 | 6.49 | 2.53 | 307 | 384 | 136 | 189 | 292 | 123 | 7.49 | 5.00 | 1.14 |
| 12:00 - 13:00 | 226 | 624 | 255 | 123 | 453 | 191 | 7.80 | 7.37 | 4.29 | 387 | 487 | 196 | 237 | 379 | 119 | 8.49 | 5.19 | 6.14 |
| 13:00 - 14:00 | 192 | 569 | 228 | 108 | 445 | 170 | 6.86 | 5.51 | 4.11 | 422 | 516 | 223 | 252 | 441 | 142 | 9.26 | 3.43 | 6.00 |
| 16:00 - 17:00 | 165 | 437 | 201 | 103 | 317 | 157 | 5.36 | 6.18 | 3.29 | 430 | 582 | 168 | 271 | 484 | 136 | 8.49 | 4.24 | 2.60 |
| 17:00 - 18:00 | 133 | 394 | 178 | 82 | 283 | 152 | 4.92 | 6.03 | 2.02 | 461 | 641 | 184 | 279 | 552 | 143 | 9.46 | 3.64 | 3.21 |
| 19:00 - 20:00 | 96 | 313 | 119 | 58 | 230 | 103 | 4.33 | 5.04 | 1.52 | 272 | 374 | 122 | 161 | 296 | 95 | 7.54 | 4.26 | 2.59 |

**Table A.2: Comparison results of SR 434 & Chapman Road (2019)**

| **09/19/2019** | **NORTHBOUND** | | | **SOUTHBOUND** | | |
|---|---|---|---|---|---|---|
| **Time** | **FDOT COUNT - TOTAL** | **ATSPM - TOTAL** | **GEH** | **FDOT COUNT - TOTAL** | **ATSPM - TOTAL** | **GEH** |
| 7:00 - 8:00 | 1111 | 709 | 0.76 | 1929 | 1423 | 10.56 |
| 8:00 - 9:00 | 1181 | 857 | 2.51 | 2197 | 1744 | 6.98 |
| 12:00 -13:00 | 1356 | 1109 | 4.12 | 1205 | 918 | 6.42 |
| 13:00 - 14:00 | 1228 | 1006 | 3.10 | 1113 | 915 | 3.91 |
| 16:00 - 17:00 | 2165 | 1566 | 0.79 | 1150 | 877 | 6.28 |
| 17:00 -18:00 | 2623 | 1775 | 0.50 | 1325 | 927 | 9.59 |
| 19:00 - 20:00 | 1267 | 947 | 0.13 | 1009 | 805 | 4.54 |

**Table A.3: Comparison results of SR 434 & Carrigan Avenue (2016)**

| 10/20/2016 | NORTHBOUND | | | SOUTHBOUND | | |
|---|---|---|---|---|---|---|
| Time | FDOT COUNT - TOTAL | ATSPM - TOTAL | GEH | FDOT COUNT - TOTAL | ATSPM - TOTAL | GEH |
| 8:00 - 9:00 | 1000 | 908 | 2.98 | 2356 | 2486 | 2.64 |
| 10:00 - 11:00 | 960 | 896 | 2.10 | 1076 | 1133 | 1.71 |
| 12:00 - 13:00 | 1317 | 1218 | 2.78 | 1088 | 1150 | 1.85 |
| 14:00 - 15:00 | 1352 | 1279 | 2.01 | 1137 | 1166 | 0.85 |
| 17:00 - 18:00 | 2603 | 2460 | 2.84 | 1410 | 1428 | 0.48 |
| 19:00 - 20:00 | 1281 | 1187 | 2.68 | 1007 | 991 | 0.51 |

**Table A.4: Comparison results of SR 434 & Carrigan Avenue (2019)**

| **09/10/2019** | **NORTHBOUND** | | | **SOUTHBOUND** | | |
|---|---|---|---|---|---|---|
| **Time** | **FDOT COUNT - TOTAL** | **ATSPM - TOTAL** | **GEH** | **FDOT COUNT - TOTAL** | **ATSPM - TOTAL** | **GEH** |
| 7:00 - 8:00 | 980 | 907 | 2.38 | 2105 | 802 | 34.18 |
| 8:00 - 9:00 | 1017 | 914 | 3.31 | 2575 | 993 | 37.45 |
| 12:00 - 13:00 | 1408 | 1291 | 3.18 | 1240 | 506 | 24.84 |
| 13:00 - 14:00 | 1338 | 1234 | 2.90 | 1162 | 491 | 23.34 |
| 16:00 - 17:00 | 2238 | 2147 | 1.94 | 1248 | 534 | 23.92 |
| 17:00 - 18:00 | 2580 | 2489 | 1.81 | 1516 | 634 | 26.90 |
| 19:00 - 20:00 | 1393 | 1318 | 2.04 | 976 | 402 | 21.87 |

**Table A.5: Comparison results of SR-426 & Oviedo Mall (2016)**

| 10/29/2016 | NORTHBOUND | | | SOUTHBOUND | | |
|---|---|---|---|---|---|---|
| Time | FDOT COUNT - TOTAL | ATSPM - TOTAL | GEH | FDOT COUNT - TOTAL | ATSPM - TOTAL | GEH |
| 7:00 - 8:00 | 602 | 597 | 0.20 | 1176 | 1174 | 0.06 |
| 8:00 - 9:00 | 737 | 732 | 0.18 | 1229 | 1223 | 0.17 |
| 12:00 - 13:00 | 749 | 745 | 0.15 | 954 | 956 | 0.06 |
| 13:00 - 14:00 | 818 | 813 | 0.18 | 903 | 898 | 0.17 |
| 14:00 - 15:00 | 854 | 898 | 1.49 | 1007 | 1006 | 0.03 |
| 16:00 - 17:00 | 944 | 934 | 0.33 | 1012 | 1007 | 0.16 |
| 17:00 - 18:00 | 1077 | 1074 | 0.09 | 1068 | 1057 | 0.34 |
| 19:00 - 20:00 | 643 | 646 | 0.12 | 717 | 719 | 0.07 |

**Table A.6: Comparison results of SR-426 & Winter Springs (2016)**

| 10/29/2016 | NORTHBOUND | | | SOUTHBOUND | | |
|---|---|---|---|---|---|---|
| Time | FDOT COUNT - TOTAL | ATSPM - TOTAL | GEH | FDOT COUNT - TOTAL | ATSPM - TOTAL | GEH |
| 7:00 - 8:00 | 414 | 411 | 0.15 | 940 | 938 | 0.07 |
| 8:00 - 9:00 | 456 | 455 | 0.05 | 852 | 848 | 0.14 |
| 12:00 - 13:00 | 553 | 546 | 0.30 | 692 | 686 | 0.23 |
| 13:00 - 14:00 | 565 | 569 | 0.17 | 632 | 642 | 0.40 |
| 14:00 - 15:00 | 600 | 596 | 0.16 | 793 | 790 | 0.11 |
| 16:00 - 17:00 | 720 | 720 | 0.00 | 770 | 765 | 0.18 |
| 17:00 - 18:00 | 879 | 878 | 0.03 | 865 | 863 | 0.07 |
| 19:00 - 20:00 | 537 | 532 | 0.22 | 686 | 679 | 0.27 |

## Table A.7: Comparison results of SR 419-434&Tuskawilla Road (2019)

| 10/26/2016 | EASTBOUND | | | | | WESTBOUND | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Time | FDOT COUNT - TOTAL | ATSPM - TOTAL | ATSPM – TOTAL (double) | GEH | GEH (double) | FDOT COUNT - TOTAL | ATSPM - TOTAL | ATSPM – TOTAL (double) | GEH | GEH (double) |
| 7:30 - 8:30 | 1336 | 714 | 1428 | 19.43 | 2.47 | 1121 | 532 | 1064 | 20.49 | 1.72 |
| 8:30 - 9:30 | 1101 | 620 | 1240 | 16.40 | 4.06 | 916 | 432 | 864 | 18.64 | 1.74 |
| 11:30 - 12:30 | 837 | 457 | 914 | 14.94 | 2.60 | 591 | 291 | 582 | 14.29 | 0.37 |
| 12:30 - 13:30 | 970 | 521 | 1042 | 16.44 | 2.27 | 644 | 332 | 664 | 14.12 | 0.78 |
| 15:00 - 16:00 | 1201 | 647 | 1294 | 18.23 | 2.63 | 829 | 430 | 860 | 15.90 | 1.07 |
| 16:30 - 17:30 | 1591 | 862 | 1724 | 20.82 | 3.27 | 1120 | 555 | 1110 | 19.52 | 0.30 |
| 17:30 - 18:30 | 1496 | 844 | 1688 | 19.06 | 4.81 | 1090 | 548 | 1096 | 18.94 | 0.18 |
| 19:00 - 20:00 | 796 | 455 | 910 | 13.63 | 3.90 | 568 | 284 | 568 | 13.76 | 0.00 |

**Table A.8: Comparison results of SR 426 & Red Bug Lake Road (2019)**

| 09/19/2019 | NORTHBOUND | | | SOUTHBOUND | | |
|---|---|---|---|---|---|---|
| **Time** | **FDOT COUNT - TOTAL** | **ATSPM - TOTAL** | **GEH** | **FDOT COUNT - TOTAL** | **ATSPM - TOTAL** | **GEH** |
| 7:00 - 8:00 | 683 | 535 | 6.00 | 1028 | 50 | 42.13 |
| 8:00 - 9:00 | 892 | 562 | 12.24 | 1059 | 48 | 42.97 |
| 12:00 -13:00 | 883 | 440 | 17.22 | 914 | 52 | 39.22 |
| 13:00 - 14:00 | 866 | 395 | 18.76 | 857 | 44 | 38.30 |
| 14:00 - 15:00 | 804 | 480 | 12.79 | 979 | 44 | 41.34 |
| 16:00 - 17:00 | 1021 | 459 | 20.66 | 931 | 38 | 40.57 |
| 17:00 -18:00 | 1205 | 440 | 26.67 | 996 | 40 | 42.00 |
| 19:00 - 20:00 | 658 | 316 | 15.50 | 700 | 66 | 32.40 |
| **09/19/2019** | **EASTBOUND** | | | **WESTBOUND** | | |
| **Time** | **FDOT COUNT - TOTAL** | **ATSPM - TOTAL** | **GEH** | **FDOT COUNT - TOTAL** | **ATSPM - TOTAL** | **GEH** |
| 7:00 - 8:00 | 1508 | 281 | 41.03 | 1784 | 44 | 57.55 |
| 8:00 - 9:00 | 1911 | 355 | 46.23 | 1742 | 46 | 57.55 |
| 12:00 -13:00 | 1507 | 335 | 38.62 | 1315 | 48 | 56.72 |
| 13:00 - 14:00 | 1570 | 375 | 38.32 | 1325 | 44 | 48.53 |
| 14:00 - 15:00 | 1624 | 332 | 41.31 | 1327 | 46 | 48.96 |
| 16:00 - 17:00 | 1869 | 480 | 40.53 | 1389 | 40 | 48.89 |
| 17:00 -18:00 | 1919 | 526 | 39.84 | 1588 | 40 | 50.47 |
| 19:00 - 20:00 | 1436 | 250 | 40.85 | 1028 | 52 | 54.26 |

## REFERENCES


Ahn, J., 2016. Highway traffic flow prediction using support vector regression and Bayesian classifier. 2016 Int. Conf. Big Data Smart Comput. 239–244. https://doi.org/10.1109/BIGCOMP.2016.7425919

A. Pinjari, N. Eluru, R. Copperman, I.N. Sener, J.Y. Guo, S. Srinivasan, C.R. Bhat, (2006), Activity-Based Travel-Demand Analysis for Metropolitan Areas in Texas: CEMDAP Models, Framework, Software Architecture, and Application Results. URL https://ctr.utexas.edu/wp-content/uploads/pubs/4080_8.pdf

ARC, (2011), The Travel Forecasting Model Set for the Atlanta Region 2010 Documentation, ARC Model Documentation, Atlanta Regional Commission (ARC). http://documents.atlantaregional.com/plan2040/docs/tp_PLAN2040RTP_appendE_0702711.pdf

Arndt, J. C., Morgan, C. A., Overman, J. H., Clower, T. L., Weinstein, B. L., & Seman, M. (2009). Transportation, Social and Economic Impacts of Light and Commuter Rail (No. FHWA/TX-10/0-5652-1).

Axhausen, K.W., Gärling, T., 1992. Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems: Foreign summaries. Transp. Rev. 12 4 , 323–341. doi:10.1080/01441649208716826

Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., Ju, Y., 2019. A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. Neural Comput. Appl. 31, 1665–1677. https://doi.org/10.1007/s00521-018-3470-9

Alajali, W., Zhou, W., Wen, S., Wang, Y., 2018. Intersection traffic prediction using decision tree models. Symmetry (Basel). 10, 1–16. https://doi.org/10.3390/sym10090386

Atwood, J., Towsley, D., 2016. Diffusion-convolutional neural networks. Adv. Neural Inf. Process. Syst. 2001–2009.

Bhat, C.R., Guo, J.Y., Srinivasan, S., Sivakumar, A., 2004. Comprehensive econometric microsimulator for daily activity-travel patterns, in: Transportation Research Record. pp. 57–66. doi:10.3141/1894-07

Bhat, C.R., Koppelman, F.S., 2006. Activity-Based Modeling of Travel Demand, in: Handbook of Transportation Science. pp. 39–65. doi:10.1007/0-306-48058-1_3

Billings, D., Jiann-Shiou, Y., 2006. Application of the ARIMA Models to Urban Roadway Travel Time Prediction-A Case Study. Systems, Man and Cybernetics, 2006. SMC'06. IEEE Int. Conf. 2529–2534.

Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., Sun, J., 2016. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transp. Res. Part C Emerg. Technol. 62, 21–34. https://doi.org/10.1016/j.trc.2015.11.002

Chang, G.L., Su, C.C., 1995. Predicting intersection queue with neural network models. Transp. Res. Part C 3, 175–191. https://doi.org/10.1016/0968-090X(95)00005-4

Chien, S.I.-J., Kuchipudi, C.M., 2003. Dynamic Travel Time Prediction with Real-Time and Historic Data. J. Transp. Eng. 129, 608–616. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(608)

Chu, L., Oh, J.-S., Recker, W., 2005. Adaptive Kalman Filter Based Freeway Travel time Estimation. Transp. Res. Board 2005 Annu. Meet. 1–21.

Cui, Z., Henrickson, K., Ke, R., Wang, Y., 2018. Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting.

Cui, Z., Wang, Y., 2017. Deep Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction 22–25.

Duggal, M., Radakovic, N., Bhowmick, A., & Datla, S. (2016). Sketch Model for Estimating Station Level Ridership for LRT. In TAC 2016: Efficient Transportation-Managing the Demand-2016 Conference and Exhibition of the Transportation Association of Canada.

DVRPC, (2011), Travel Demand Model Upgrade -Travel Improvement Model (TIM) 1.0. URL https://www.dvrpc.org/reports/TR10006.pdf?fbclid=IwAR25t9f726qyoPPqgxBGPLmAPriccyYRFz20vOP7FobDfMDpLNmiITtsac

Elhenawy, M., Chen, H., Rakha, H.A., 2014. Dynamic travel time prediction using data clustering and genetic programming. Transp. Res. Part C Emerg. Technol. 42, 82–98. https://doi.org/10.1016/j.trc.2014.02.016

Epelbaum, T., Gamboa, F., Loubes, J.-M., Martin, J., 2017. Deep Learning applied to Road Traffic Speed forecasting.

Gers, F.A., Cummins, F., 1999. 1 Introduction 2 Standard LSTM 1–19.

Habtemichael, F.G., Cetin, M., 2016. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. Transp. Res. Part C Emerg. Technol. 66, 61–78.

https://doi.org/10.1016/j.trc.2015.08.017

Haworth, J., Shawe-taylor, J., Cheng, T., Wang, J., 2014. Local online kernel ridge regression for forecasting of urban travel times. Transp. Res. Part C 46, 151–178. https://doi.org/10.1016/j.trc.2014.05.015

Hochreiter, S., Urgen Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Innamaa, S., 2005. Short-term prediction of travel time using neural networks on an interurban highway. Transportation (Amst). 32, 649–669. https://doi.org/10.1007/s11116-005-0219-y

Kim, T.S., Lee, W.K., Sohn, S.Y., 2019. Graph convolutional network approach applied to predict hourly bike-sharing demands considering spatial, temporal, and global effects. PLoS One 14, 1–16. https://doi.org/10.1371/journal.pone.0220782

Kipf, T.N., Welling, M., 2017. SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS 1–14.

Lee, S., Xie, K., Ngoduy, D., Keyvan-Ekbatani, M., 2019. An advanced deep learning approach to real-time estimation of lane-based queue lengths at a signalized junction. Transp. Res. Part C Emerg. Technol. 109, 117–136. https://doi.org/10.1016/j.trc.2019.10.011

Lee, Y.L.Y., 2009. Freeway travel time forecast using artifical neural networks with cluster method. 2009 12th Int. Conf. Inf. Fusion 1331–1338.

Li, W., Ban, X.J., Zheng, J., Liu, H.X., Gong, C., Li, Y., 2020. Real-Time Movement-Based Traffic Volume Prediction at Signalized Intersections. J. Transp. Eng. Part A Syst. 146, 1–

15. https://doi.org/10.1061/JTEPBS.0000384

Li, Y., Yu, R., Shahabi, C., Liu, Y., 2017. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting 1–16.

Lin, L., He, Z., Peeta, S., 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach This study proposes a novel Graph Convolutional Neural Network with Data-driven Graph Filter (GCNN-DDGF) model 97, 258–276. https://doi.org/10.1016/j.trc.2018.10.011

Luo, X., Li, D., Yang, Y., Zhang, S., 2019. Spatiotemporal Traffic Flow Prediction with KNN and LSTM 2019.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Yong, Wang, Yunpeng, 2017. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. Sensors (Switzerland) 17, 1–16. https://doi.org/10.3390/s17040818

Ma, X., Tao, Z., Wang, Yinhai, Yu, H., Wang, Yunpeng, 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp. Res. Part C Emerg. Technol. 54, 187–197. https://doi.org/10.1016/j.trc.2015.03.014

Mahmassani, H.S., Hou, T., Kim, J., Chen, Y., Hong, Z., Halat, H., Haas, R., 2014. Implementation of a Weather Responsive Traffic Estimation and Prediction System ( TrEPS ) for Signal Timing at Utah DOT. https://doi.org/FHWA-JPO-14-140

Meng, M., Shao, C., Wong, Y., Wang, B., Li, H., 2015. A two-stage short-term traffic flow prediction method based on AVL and AKNN techniques. J. Cent. South Univ. 22, 779–786.

https://doi.org/10.1007/s11771-015-2582-y

Myung, J., Kim, D.-K., Kho, S.-Y., Park, C.-H., 2011. Travel Time Prediction Using k Nearest Neighbor Method with Combined Data from Vehicle Detector System and Automatic Toll Collection System. Transp. Res. Rec. J. Transp. Res. Board 2256, 51–59. https://doi.org/10.3141/2256-07

Oh, S., Byon, Y.J., Jang, K., Yeo, H., 2017. Short-term travel-time prediction on highway: A review on model-based approach. KSCE J. Civ. Eng. 1–13. https://doi.org/10.1007/s12205-017-0535-8

Ojeda, L.L., Kibangou, A.Y., Wit, C.C. De, 2013. Adaptive Kalman Filtering for Multi-Step ahead Traffic Flow Prediction. 2013 Am. Control Conf. 4724–4729. https://doi.org/10.1109/ACC.2013.6580568

Park, D., Rilett, L.R., Han, G., 1999. Spectral Basis Neural Networks for Real-Time Travel Time Forecasting. J. Transp. Eng. 125, 515–523. https://doi.org/10.1061/(ASCE)0733-947X(1999)125:6(515)

Polson, N., Sokolov, V., 2016. Deep Learning for Short-Term Traffic Flow Prediction 1–29. https://doi.org/10.1016/j.trc.2017.02.024

Polson, N.G., Sokolov, V.O., 2017. Deep learning for short-term traffic flow prediction. Transp. Res. Part C Emerg. Technol. 79, 1–17. https://doi.org/10.1016/j.trc.2017.02.024

Qiao, W., Haghani, A., Hamedi, M., 2013. A Nonparametric Model for Short-Term Travel Time Prediction Using Bluetooth Data. J. Intell. Transp. Syst. 17, 165–175.

https://doi.org/10.1080/15472450.2012.748555

Rahman, R., Hasan, S., 2020. Real-time signal queue length prediction using long short-term memory neural network. Neural Comput. Appl. 9. https://doi.org/10.1007/s00521-020-05196-9

Rahman, R., Hasan, S., 2019. Real-time Signal Queue Length Prediction Using Long Short-Term Memory Neural Network, in: Transportation Research Board 98th Annual MeetingTransportation Research Board.

Rahman, R., Hasan, S., 2018a. Short-Term Traffic Speed Prediction for Freeways During Hurricane Evacuation : A Deep Learning Approach 1291–1296. https://doi.org/10.1109/ITSC.2018.8569443

Rahman, R., Hasan, S., 2018b. Short-Term Traffic Speed Prediction for Freeways during Hurricane Evacuation: A Deep Learning Approach. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 2018-Novem, 1291–1296. https://doi.org/10.1109/ITSC.2018.8569443

Serulle, J.K.G., Mahmassani, H., Gopalakrishna, D., Krueger, B., Jiaqi, M., Fang, Z., Zihan, H., Ostojic, M., Urena, N., 2017. Integrated Modeling for Road Condition Prediction Final Report.

Svozil, D., Kvasnicka, V., Pospichal, J., 1997. Introduction to multi-layer feed-forward neural networks. Chemom. Intell. Lab. Syst. 39, 43–62. https://doi.org/10.1016/S0169-7439(97)00061-0

Williams, B.M., Asce, M., Hoel, L.A., Asce, F., 2004. Modeling and Forecasting Vehicular Traffic

Flow as a Seasonal ARIMA Process : Theoretical Basis and Empirical Results 664–672.

Wu, C., Wei, C., Su, D., Chang, M., Ho, J., 2004. Travel time prediction with support vector regression. Proc. 2003 IEEE Int. Conf. Intell. Transp. Syst. 2, 1438–1442. https://doi.org/10.1109/ITSC.2003.1252721

Wu, C.H., Ho, J.M., Lee, D.T., 2004. Travel-time prediction with support vector regression. IEEE Trans. Intell. Transp. Syst. 5, 276–281. https://doi.org/10.1109/TITS.2004.837813

Xu, J., Rahmatizadeh, R., Turgut, D., 2017. Real-Time Prediction of Taxi Demand Using Recurrent Neural Networks 1–10.

Yang, B., Sun, S., Li, J., Lin, X., Tian, Y., 2019. Traffic flow prediction using LSTM with feature enhancement. Neurocomputing 332, 320–327. https://doi.org/10.1016/j.neucom.2018.12.016

Yanjie Duan, Yisheng Lv, Fei-Yue Wang, 2016. Travel time prediction with LSTM neural network. 2016 IEEE 19th Int. Conf. Intell. Transp. Syst. 1053–1058. https://doi.org/10.1109/ITSC.2016.7795686

Yasdi, R., 1999. Prediction of Road Traffic using a Neural Network Approach. Neural Comput. Appl. 8, 135–142. https://doi.org/10.1007/s005210050015

Yu, B., Song, X., Guan, F., Yang, Z., Yao, B., 2016. k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition. J. Transp. Eng. 142, 04016018. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000816

Yu, B., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. IJCAI Int. Jt. Conf. Artif. Intell. 2018-July, 3634–3640.

Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. Sensors (Switzerland) 17, 1–16. https://doi.org/10.3390/s17071501

Yu, J., Chang, G.-L., Ho, H.W., Liu, Y., 2008. Variation Based Online Travel Time Prediction Using Clustered Neural Networks. 2008 11th Int. IEEE Conf. Intell. Transp. Syst. 85–90. https://doi.org/10.1109/ITSC.2008.4732594

Yu, R., Li, Y., Shahabi, C., Demiryurek, U., Liu, Y., 2017. Deep learning: A generic approach for extreme condition traffic forecasting. Proc. 17th SIAM Int. Conf. Data Mining, SDM 2017 777–785. https://doi.org/10.1137/1.9781611974973.87

Zhang, J., Wang, F.Y., Wang, K., Lin, W.H., Xu, X., Chen, C., 2011. Data-driven intelligent transportation systems: A survey. IEEE Trans. Intell. Transp. Syst. 12, 1624–1639. https://doi.org/10.1109/TITS.2011.2158001

Zhang, Y., Haghani, A., 2015. 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

Zhou, J., Cui, G., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2018. Graph Neural Networks: A Review of Methods and Applications 1–20.