UTCM

Improving the Quality of Life by Enhancing Mobility

University Transportation Center for Mobility

DOT Grant No. DTRT06-G-0044

Development of Statistical Models to Forecast Crossing Times of Commercial Vehicles

Final Report

Dong Hun Kang, Cesar Ochoa, Rajat Rajbhandari, and Francisco Garcia

Performing Organization
University Transportation Center for Mobility™
Texas Transportation Institute
The Texas A&M University System
College Station, TX

Sponsoring Agency
Department of Transportation
Research and Innovative Technology Administration
Washington, DC

Texas Transportation Institute

UTCM Project #10-09-60
July 2011

Technical Report Documentation Page

1. Project No.	2. Government Accession No.	3. Recipient's Catalog No.
UTCM 10-09-60		

4. Title and Subtitle	5. Report Date
Development of Statistical Models to Forecast Crossing Times of Commercial Vehicles	July 2011
	6. Performing Organization Code Texas Transportation Institute
7. Author(s) Dong Hun Kang, Cesar Ochoa, Rajat Rajbhandari, and Francisco Garcia	8. Performing Organization Report No. UTCM 10-09-60
9. Performing Organization Name and Address University Transportation Center for Mobility™ Texas Transportation Institute The Texas A&M University System 3135 TAMU College Station, TX 77843-3135	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTRT06-G-0044
12. Sponsoring Agency Name and Address Department of Transportation Research and Innovative Technology Administration 400 7th Street, SW Washington, DC 20590	13. Type of Report and Period Covered Final Report 9/1/10 - 3/31/11 14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program

16. Abstract

Border crossing time measurement systems for commercial vehicles are being implemented throughout the U.S.-Mexico border. These systems are based on radio frequency identification (RFID) technology. With funding from the Federal Highway Administration, the Texas Transportation Institute (TTI)/Battelle team installed an RFID-based system at the Bridge of the Americas (BOTA) in El Paso, Texas, to measure and archive crossing times of commercial vehicles. The RFID system at BOTA is already operational, and current truck crossing time information is relayed and archived in a centralized repository. In addition, with funding from the Texas Department of Transportation, TTI deployed RFID systems on the Pharr-Reynosa Bridge. These systems measure the current crossing time and provide the information to users; however, there is no system in place to predict the crossing times of trucks. In fact, there are no systems in place at the U.S.-Mexico border to predict traffic conditions including crossing times of trucks. In this project, statistical models were developed to predict crossing times of trucks over a short range of time. The statistical prediction models use historic data and take into account empirical relationships between border-crossing-related parameters and truck crossing times.

17. Key Word	18. Distribution Statement
United States-Mexico Border, Truck Crossing Times, Real Time Information, Travel Time, Radio Frequency Identification, Commercial Vehicles, Research Projects	Public distribution

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	86	n/a

Development of Statistical Models to Forecast Crossing Times of Commercial Vehicles

Dong Hun Kang, Ph.D.
Associate Transportation Researcher
System Planning, Policy and Environmental Research Group
Texas Transportation Institute, The Texas A&M University System

Cesar Ochoa
Doctoral Student at the University of Texas at El Paso
Graduate Research Assistant
Center for International Intelligent Transportation Research
Texas Transportation Institute, The Texas A&M University System

Rajat Rajbhandari, Ph.D., P.E.
Associate Research Engineer
Center for International Intelligent Transportation Research
Texas Transportation Institute, The Texas A&M University System

and

Francisco Garcia, Ph.D.
Assistant Professor
University of Madrid, Spain

Final Report
Project 10-09-60

University Transportation Center for Mobility™
Texas Transportation Institute
The Texas A&M University System

July 2011

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Acknowledgment

Support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers Program to the University Transportation Center for Mobility™ (DTRT06-G-0044).

The authors would like to acknowledge Crystal Jones (Federal Highway Administration) for providing valuable comments to the project.

The authors would also like to acknowledge funds provided by the Center for International Intelligent Transportation Research (CIITR) at the Texas Transportation Institute's El Paso office to support several related research projects.

Table of Contents

Page

List of Figures

NOTE: Figures include color which may not reproduce well on black and white copiers. A PDF copy of this report with color figures may be accessed via the UTCM website at http://utcm.tamu.edu or on the Transportation Research Board's TRID database at http://trid.trb.org.

Page

List of Tables

Page

Executive Summary

Commercial vehicle crossings at the U.S.-Mexico and U.S.-Canada borders play an important role in the local and regional economies. Truck trade has increased at an annual average growth rate of 7.5 percent for the last decade, resulting in an increase in the volume of commercial vehicles entering the United States through various border crossings. Freight shippers and carriers use the predicted crossing times as part of their pre-trip information to plan a trip from origin to destination. Measuring and reporting crossing times are of considerable interest to a wide range of stakeholders that interact at the border between Mexico and the United States. This project aims to create a tool to help shippers and freight carriers make smart route-choice decisions by providing predicted crossing times at the U.S.-Mexico border. The prediction model has the potential to be implemented at several border crossings where intelligent transportation system technology is being deployed.

From the given conditions of radio frequency identification (RFID) border crossing data, a structured procedure to predict border crossing times of commercial trucks within a short range of time was developed based on statistical models. Currently, RFID reader systems that measure northbound crossing times of commercial vehicles are being implemented throughout the ports of entry along the U.S.-Mexico border. The RFID tags with time stamps at the entry point on the Mexican side and at the exit point of U.S. federal and state inspection facilities are collected and stored to measure the crossing times between the two reader stations. The current RFID system does not have the capability of identifying detailed information about the trucks, such as Free and Secure Trade (FAST)/non-FAST and loaded/empty, and the different types of trucks should have different characteristics of crossing times due to the physical and operational layout of border ports of entry. Therefore, the Gaussian Mixture Model was used to define the set of parameters of each class. Then the Expectation Maximization Algorithm was applied to iteratively estimate the unknown parameters of the FAST, non-FAST-empty, and non-FAST-loaded truck classes.

After the data classification step, the clustered data were processed by the Weighted Moving Average Window to consider the time dependency of crossing times. As described in Chapter 2, average truck crossing times are varied over time. In order to incorporate the dynamic nature of the crossing time variations by time of day, weights are determined in such a way that recent observations have more weight. The weights are also calibrated by the membership functions of the different truck classes.

As a main step in the prediction procedure, regressive functions are determined by the Gaussian Process (GP), which is a widely used stochastic method for pattern recognition. GP is easy to implement and flexible to change for the given conditions. The results obtained after fitting the regressive model with the data collected between 11/3/2009 and 11/7/2009 are shown in plots for different classes of trucks in Chapter 4 and Appendix A. In the plots, most of the prediction curves show good fit to the actual RFID observations. The accuracy of the model depends on the size of the data and the kernel.

At times, unexpected incidents may occur, such as RFID reader failures that hinder RFID data collection. To deal with the unavailability of current crossing time data to be used in the short-

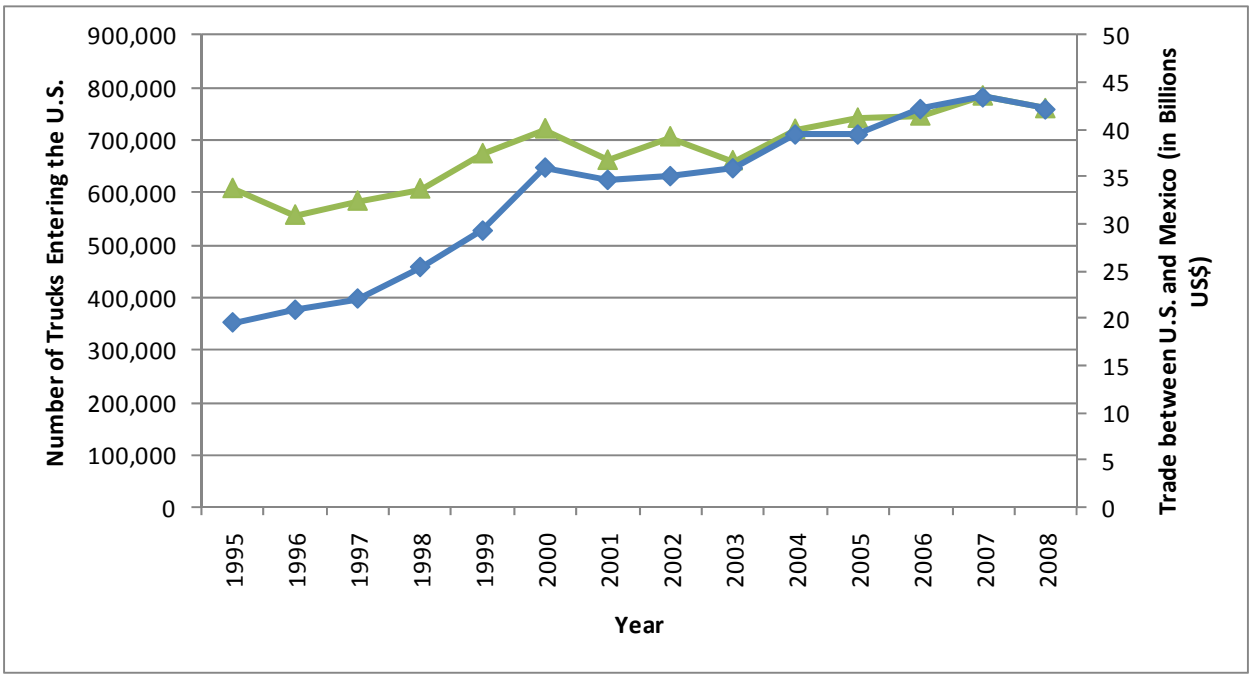
term prediction model, the Bootstrap Aggregating (BAGGING) technique was used to exploit the fitted models from the historical data and improve the predictions by using the combination of the existing fitted models. Actual prediction results with the assumption of no current observation data are presented in Chapter 5. Based on the historical data collected during the first two weeks of November 2009, two different approaches of the BAGGING technique were used. In the first approach, the ensemble of the models from the same day of the week was used. In the second approach, crossing times of Monday, November 16, 2009, were predicted based on the ensemble of models from the whole two weeks of observations. Both of the approaches showed similar results. The approaches both showed errors of about 5 minutes for FAST trucks, 10 minutes for non-FAST-empty trucks, and 15 minutes for non-FAST-loaded trucks.

The Texas Transportation Institute plans to improve the reliability of the developed system. A third sensor will be placed at the Bridge of the Americas, and this sensor will improve the performance of the reading system and the forecasting algorithm. The sensor is planned to be placed between the current sensors, perhaps near the inspection booth. After placement of the third sensor, the current procedure will be readjusted to exploit the additional information for better forecasting.

Chapter 1. Background

1.1. Freight Movement at Land Border Crossings

Land border crossings are important gateways for foreign trade, contributing significantly to the national economy. Commercial vehicle crossings at the U.S.-Mexico-Canada border also play an important role in the local and regional economies. Truck trade has increased at an annual average growth rate of 7.5 percent for the last decade resulting in an increase in the volume of commercial vehicles entering the U.S. through various border crossings. The historic data also confirm that trucks are significantly more important than rail in transporting freight. Hence, performance of freight travel crossing the border has suffered resulting in decreasing reliability. FIGURE 1 shows the increasing trend of commercial vehicles entering the U.S. from Mexico via ports of entry (POEs) in the El Paso, Texas, area and the direct correlation with the trade between the two countries.



Source: U.S. Bureau of Transportation Statistics

FIGURE 1 Annual Total Number of Trucks Entering the U.S. from Mexico between 1995 and 2008

1.2. Performance of Land Border Crossings

Freight travel crossing the border has many of the same performance concerns as passengers traveling between the U.S and Mexico, especially pertaining to reliability. Given the influence of deteriorating performance of commercial vehicles entering the U.S. and its impact on the freight commerce, stakeholders at all levels of government are expressing urgency to monitor the performance of individual land border crossings closely. Given the wide range and diversity of available measures, it is important to have a clear basis for assessing performance of freight

movement crossing the border. However, measures or indices to assess and compare the performance of border-crossing measures should:

- Take into account the local operational characteristics.
- Be used potentially as real-time travel time information for shippers and carriers.
- Be able to evaluate changes in operating practice.
- Be responsive to changes to the infrastructure, e.g., capacity increase.

Clearly, the key performance measure of border crossings that most concerns manufacturers, shippers, and freight carriers is the door-to-door travel time (Texas Transportation Institute 2008). Travel time can then be used to generate several other indices such as delay, reliability etc.

In the context of bi-national freight movement, time needed to cross the U.S.-Mexico or U.S.-Canada border could constitute a significant portion of the door-to-door travel time. The crossing time of a commercial vehicle (or a passenger vehicle) consists of the wait time to reach the primary inspection after joining the queue (wait time) and inspection times at the U.S. and Mexican federal and state facilities. From the perspective of implementing systems at the border, measuring crossing time of individual vehicles is easier compared to measuring the wait time or inspection times. At the same time, crossing time also satisfies the previously mentioned requirements for a performance measure. Unfortunately, there is no reliable system in place to measure and report crossing times of commercial vehicles. However, that is changing slowly because of several systems already in place in Texas and other border crossings along the U.S.-Mexico border.

1.3. Information Needs of Stakeholders

Measuring and reporting crossing times are of considerable interest to a wide range of stakeholders that interact at the border between Mexico and United States (1). Travel time between origin in Mexico and destination on the U.S. side of the border is of high significance to the trade industry. The time it takes to cross the border for trade industry is an important element in making freight logistics related decisions.

For a research project, the Texas Transportation Institute (TTI) met with several stakeholders in the El Paso-Ciudad Juárez region to identify stakeholder requirements regarding a variety of traveler information related to border crossings. The stakeholders identified that information required to make “smart” decisions regarding the best departure time, crossing times at border crossings, and selection of one border crossing over the other, would be highly valuable.

Freight shippers and carriers use the predicted crossing times as part of the pre-trip information to plan a trip from origin to destination. Once the trip starts, travelers could use information to modify predetermined routes to adjust to current and predicted travel conditions and determine an optimal route that would reduce travel time between origin and destination. Local media outlets can relay the predicted crossing times through traditional means of radio and television.

In addition to travelers, public and private agencies operating at the border also use the information to monitor current conditions at and around border crossings for impromptu modification of resources to increase the efficiency of operation.

1.4. Deployment of Intelligent Transportation System (ITS) Technologies

For a study funded by the Federal Highway Administration (FHWA), researchers from TTI and Battelle analyzed six different technologies that could sustain the automatic collection of border crossing times for commercial freight vehicles (2). The full report can be found at http://tti.tamu.edu/documents/TTI-2007-1.pdf. The six technologies examined in the study were:

- Automatic Vehicle Identification (AVI).
- Automatic License Plate Recognition (ALPR).
- Vehicle Matching.
- Automatic Vehicle Location (AVL).
- Mobile Phone Location.
- Inductive Loop Detectors.

Factors such as cost, accuracy of readings, availability, and reliability were analyzed for each of the technologies listed above. From the initial analysis, the study concluded that only three of these six technologies could support a system that would sustain long-term data collection, be easily transferable to other POEs along the southern and northern borders of the United States, and utilize a technology that could allow crossing times for passenger vehicles to be measured. AVI (specifically with radio frequency identification [RFID]), AVL (specifically with global positioning system [GPS]), and ALPR all had the requisite characteristics for consideration for a crossing time measurement system.

After analyzing the advantages and disadvantages of both RFID and GPS technologies, the study team recommended that RFID be selected for the border crossing time measurement system at the Bridge of the Americas (BOTA). In addition, Customs and Border Protection (CBP) is using RFID technology to read freight and driver related information. As part of the Free and Secure Trade program, freight carriers are required to carry transponders provided by the CBP on the windshield. Similarly, the Texas Department of Public Safety (DPS) is using the RFID to monitor truck movement within its safety inspection facility.

As an example of ongoing efforts to measure border crossing time of commercial vehicles, with funding from the Federal Highway Administration, TTI/Battelle team installed RFID-based system at the Bridge of the Americas in El Paso, Texas. The systems at both border crossings are already operational and collect current crossing times of commercial vehicles. In addition, with funding from the Texas Department of Transportation (TxDOT), TTI deployed RFID systems on the Pharr-Reynosa Bridge. TxDOT has also acquired the services of TTI to deploy similar RFID-based systems at World Trade Bridge and Colombia Bridge border crossings in the Laredo area. The Arizona Department of Transportation recently selected a team comprising of TTI and Battelle to install an RFID-based system to measure crossing times of commercial vehicles at the Mariposa port of entry.

1.5. Goals and Objectives of the Project

The project is aimed at creating a tool for shippers and freight carriers to make "smart" route-choice decisions by providing predicted crossing times at the U.S.-Mexico border. The tool will also provide federal inspection agencies to allocate resources for opening and closing lanes based on predicted crossing times. The tool will provide the carriers, shippers, and agencies operating at the border to make efficient travel related decisions ranging from route-choice to allocating resources to operate the border-crossing infrastructure. The prediction model has the potential to be implemented at several border crossings where ITS technology is being deployed. The prediction model will be implemented at the border crossings in El Paso and Pharr and will be demonstrated to TxDOT, other state departments of transportation, FHWA, and other bi-national stakeholders.

Systems to measure crossing time of commercial vehicles are being implemented throughout the U.S.-Mexico border. These systems measure the current crossing time and provide the information to users, however, there is no system in place to predict the crossing times of commercial vehicles. In fact, there are no systems in place at the U.S.-Mexico border to predict any kind of traffic conditions.

In this project, statistical models will be developed to predict crossing times of trucks over a short range of time (for example, an hour forecast). The statistical prediction models will use historic data (specifically the temporal behavior of crossing times) and take into account empirical relationships between other border-crossing-related parameters (such as inspection lanes open, wait times relayed by the Customs and Border Protection, Homeland Security Threat Level, etc.) and the truck crossing times collected from the RFID readers. The prediction model developed in this research will be implemented at the border crossings in El Paso and Pharr where RFID systems are installed. The short-term prediction of truck crossing times will facilitate the carriers and shippers to make efficient travel related decisions.

Chapter 2. Analysis of Truck Crossing Times Data

2.1. Factors Influencing Crossing Times

Crossing times of commercial as well as passenger vehicles are influenced by a wide variety of factors. Some factors are related to operational changes related to federal and state inspection processes while others are external ones such as approaching volume, major incidents around the border crossings, etc. These factors include:

- Time of day and week — Commercial vehicles crossing the border do follow a temporal trend and show a distinct peak and off-peak volume. Intuitively, temporal trend of crossing times should follow similar trend for volumes.
- Shipment type — Depending on the type of shipment, crossing times could vary significantly, especially the ones that are empty and are enrolled in the Free and Secure Trade (FAST) program. If border crossings have separate lanes entering the inspection station for FAST vehicles, the crossing times are reduced further. Unfortunately, the

systems currently deployed in El Paso and Pharr do not have the capability to distinguish the shipment type. However, it is anticipated that the historic data may "reveal" clustered crossing times data that can eventually be attributed to vehicles that went through the primary and secondary inspections.

- Approaching volume of vehicles — The volume of vehicles approaching the federal inspection station has a significant impact on crossing times since there is a limited number of booths to inspect the vehicles. In addition, the number of lanes approaching the inspection facility is constant, which also helps in creating lengthy queues. Approaching volume of vehicles is also a function of time of day and week, special events and holidays, and other factors.
- Special events, incidents, and holidays — Special events (such as concerts, games, etc.) on one side of the border also create heavy passenger vehicle traffic at the border. However, this factor is mostly attributed to creating high crossing times for passenger vehicles, but may influence crossing times of commercial vehicles if the physical layout of the border crossing does not allow significant separation of the two types of vehicles. Volume of commercial vehicles crossing the border does decrease significantly during major holidays, such as Christmas. Local and national threat level advisories exacerbate the condition due to increased inspection time.
- Physical layout of the border crossing — All land border crossings differ from one another in terms of layout of inspection areas, access and egress areas, time of operation, geometric layout, merging and diverging of commercial vehicles with passenger vehicles, etc. It is important to recognize that physical layout does influence crossing times of commercial vehicles. As an example, if the border crossing does not have lanes that separate commercial and passenger vehicles and if there is a significantly higher proportion of passenger vehicles then crossing times of commercial vehicles increase.
- Inspections related factors — The number of inspection lanes open or active during any given time is highly correlated with crossing times of vehicles. All vehicles go through the primary inspection process. Vehicles that go through the secondary inspection have a much higher crossing time than the rest.
- Daily currency exchange rates and fuel prices — Heavy fluctuation and significant change in currency exchange rates have been known to produce heavy passenger vehicle traffic trying to cross the border — mostly for shopping.

Forecasting models have to take into account various factors that influence crossing times of commercial vehicles. However, it is intuitive that accommodating the impact of all these factors in the forecasting model would create an overtly complex model. In addition, questions arise whether all factors have to be considered to obtain a reliable forecasting model.

2.2. Characteristics of the Border Crossings

Commercial vehicles entering the U.S. from Mexico can be divided into multiple categories based on type of shipment — enrolled in the FAST program versus shipments not enrolled in the program. Policies govern that shipments enrolled in the FAST program are cleared quickly and have shorter crossing times than shipments not enrolled in the program. Moreover, most of the larger commercial border crossings have dedicated FAST lanes where crossing times might be

significantly shorter since these vehicles do not have to mix with vehicles not enrolled in the program.

Vehicles entering the U.S. may also be empty (approximately 28 percent at BOTA) and would not require going through the secondary inspection. Data are available in the public domain as to what percentage of total shipments is empty. Obviously, vehicles that have to go through the secondary inspection have much longer crossing times than the rest. However, it is not clear as to what percentage of total vehicles go through the secondary inspection. The Department of Homeland Security has not released such information to the public.

Unfortunately, the systems currently deployed in El Paso and Pharr do not have the capability to distinguish the shipment type. However, it is anticipated that the historic data may "reveal" clustered crossing times data that can eventually be used to obtain probabilistic behavior of vehicles that went through the primary and secondary inspections.

All land border crossings differ from one another in terms of layout of inspection areas, access and egress areas, time of operation, geometric layout, merging and diverging of commercial vehicles with passenger vehicles, etc. Understanding the influence of these factors on crossing times of commercial vehicles is crucial. However, almost all border crossings have common operational characteristics in terms of inspection and processing of shipments — as governed by state and federal policies. Hence, for the purposes of this research, geometric and operational characteristics of the two border crossings are described in detail in subsequent sections.

The Bridge of the Americas land border-crossing facility is located in the El Paso–Ciudad Juárez region. FIGURE 2 shows the map of the El Paso–Ciudad Juárez region and the location of the border crossing, which is used by both commercial and passenger vehicles to cross the U.S.-Mexico border. Commercial vehicles access the facility from Mexico through Cuatro Siglos (a street on the Mexican side of the border), and are directed to specific lanes by road signs in order to separate the two types of vehicular traffic. Once on the physical bridge, commercial and passenger vehicles are separated by a concrete barrier. Commercial vehicle traffic is handled by two dedicated outside lanes on each bridge structure. Commercial vehicles, after clearing the federal and state inspection facilities merge onto the Gateway Boulevard North, which provides access to US54 or IH10.

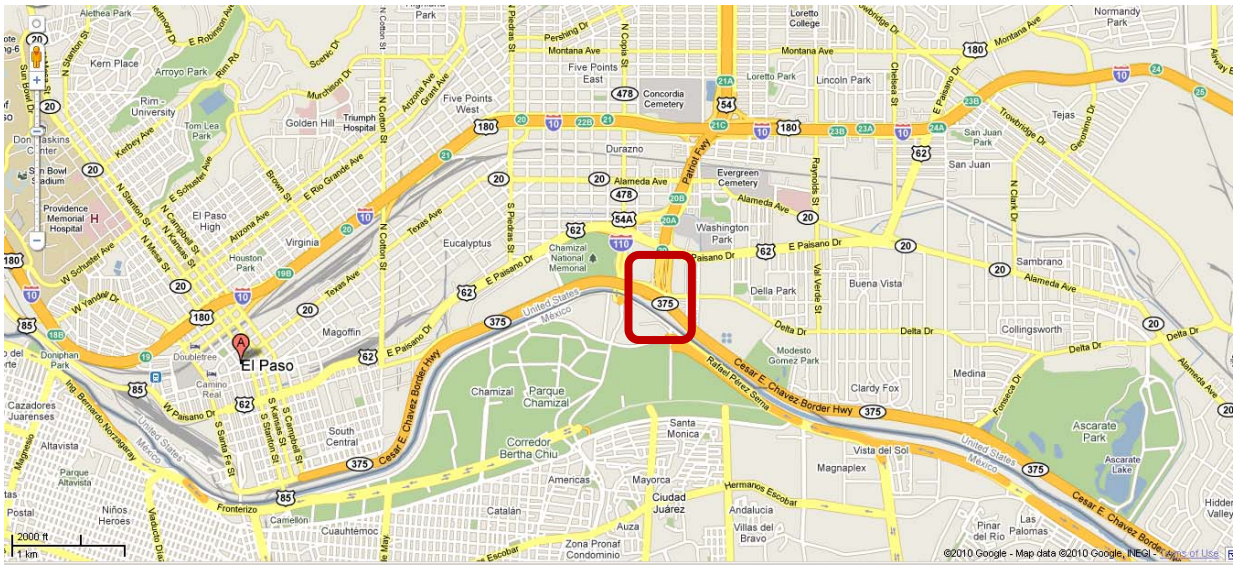


FIGURE 2 Map of Ciudad Juárez–El Paso Region, Including Location of the Bridge of the Americas

There are several key agencies involved in the border crossing process for northbound commercial trucks. These agencies include:

- U.S. Federal Agencies:
  - Customs and Border Protection (CBP).
  - Federal Motor Carrier Safety Administration (FMCSA).
- U.S. State Level Agencies:
  - Texas Department of Public Safety (DPS).
- Mexican Federal Agencies:
  - Aduanas (Mexican Customs).

The northbound commercial freight border crossing process begins at the Aduanas facility on the Mexican side of the border (also referred to as the Mexican Export Lot). After clearing customs on the Mexican side, a truck crosses the physical bridge structure. Immediately upon entering the United States, the truck proceeds to the U.S. Federal Inspection Compound. Entrance to the Federal Inspection Compound is accessed through one of six primary inspection booths. At these primary inspection booths, CBP agents determine whether the truck requires secondary inspection and direct the driver to it, or otherwise instruct the driver to proceed to the exit. Final clearance to exit the federal inspection facility is given at one of two booths at the exit, with the instructions to proceed to the state inspection facility.

The federal and state inspection facilities are connected by a one-lane access road that passes under US54. After exiting the U.S. federal inspection facility, commercial vehicle trucks continue on the one-lane access road to the primary inspection area of the state facility, where they may be required to undergo a secondary inspection.

The Bridge of the Americas facility operates from 6:00 AM to 6:00 PM Monday through Friday and from 6:00 AM to 2:00 PM on Saturdays. Empty truck traffic prefers using this free bridge to

avoid paying the toll at the Ysleta–Zaragoza Bridge. Only empty containers are permitted to cross between the hours of 6:00 AM and 8:00 AM. On October 27, 2003, one of BOTA's two northbound lanes was converted to a designated "FAST" lane. This lane, part of the Free and Secure Trade Program, allows cargo that meets specified security requirements to be expedited through U.S. Primary inspection. The FAST lane occupies the outside lane on the physical bridge structure at BOTA (the farthest right lane for northbound traffic).

Approximately 15 percent of the total northbound truck volume at this crossing is now expedited across the border in this lane. FIGURE 3 shows the location of federal and state inspection facilities at the border crossing. The figure also shows the location of RFID stations relative to the inspection facilities.

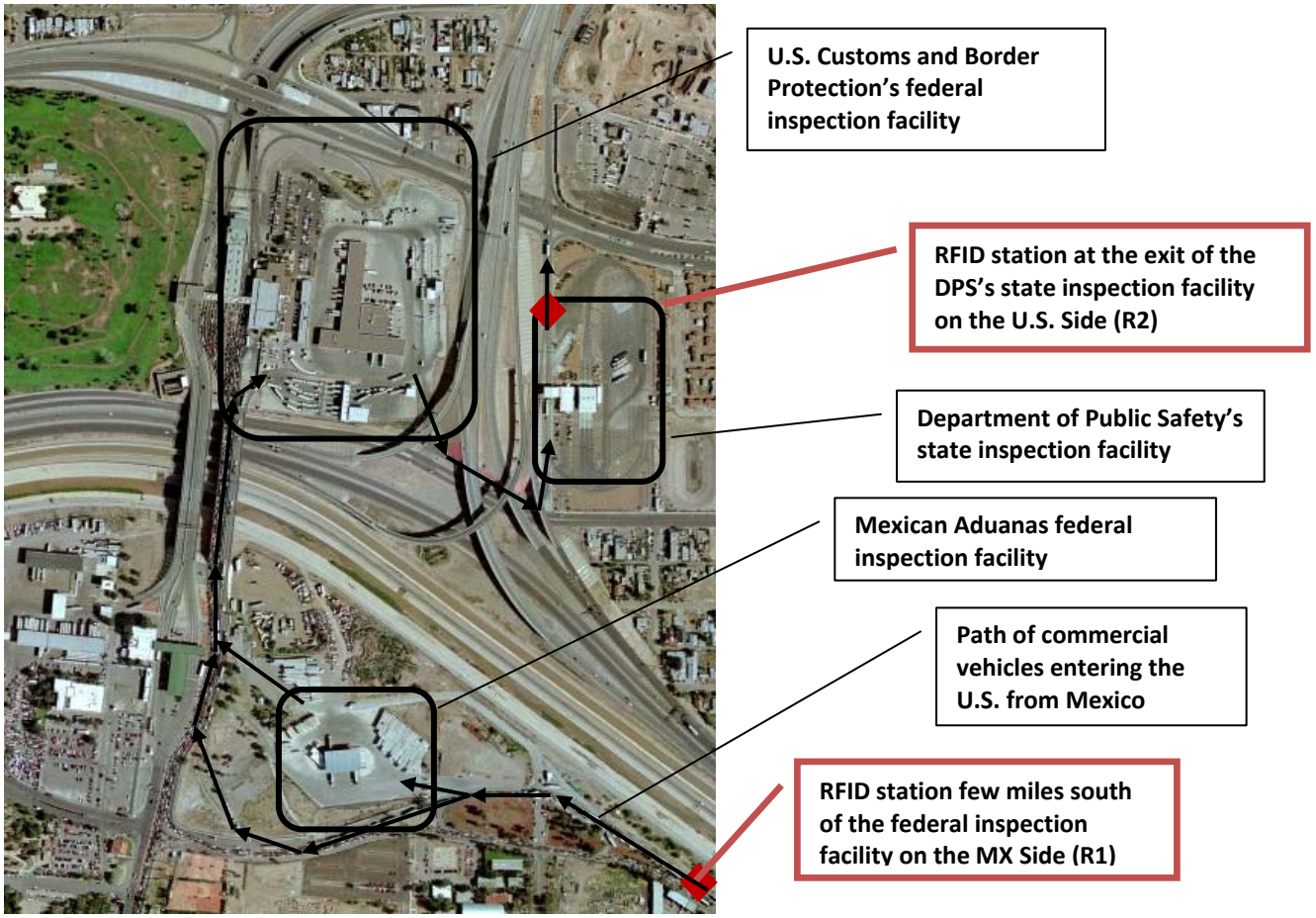


FIGURE 3 State and Federal Inspection Facilities and RFID Stations at the Bridge of the Americas in El Paso, Texas

2.3. Collection and Processing Raw Crossing Times

Northbound commercial vehicles pass RFID reader stations installed at both sides of the border. The readers read the unique identifier associated with each transponder (also called tags)

attached to a vehicle. The reader station applies a time stamp to the tag read and forwards the resulting data record to a central location for further processing via a data communication link.

A central server receives data packets consisting of the unique identifier and the time stamp associated with each identifier from RFID stations. The server stores all inbound raw reader station data and subsequently processes data in an archive for future access and use. The raw data are processed to match tag reads of individual trucks at the entrance point on the Mexican side and the exit point on the U.S. side. The difference in time stamps yields a single vehicle travel time between the RFID stations — referred to hereafter as raw crossing time data. FIGURE 4 illustrates the collection and processing of raw crossing time data.

The central server also includes a process by which average crossing time is calculated every 15 minutes using a 2-hour time window. The average travel times between the readers are determined using the following procedure: The procedure uses 120 minutes as the time window, meaning this value is used as a maximum travel time that could occur at any given segment and total crossing time. For example, to calculate average travel time between R1 and R2 at 9 AM, all the tags that were read between 7 AM and 9 AM are matched and travel times of matched tags are averaged (simple mean).

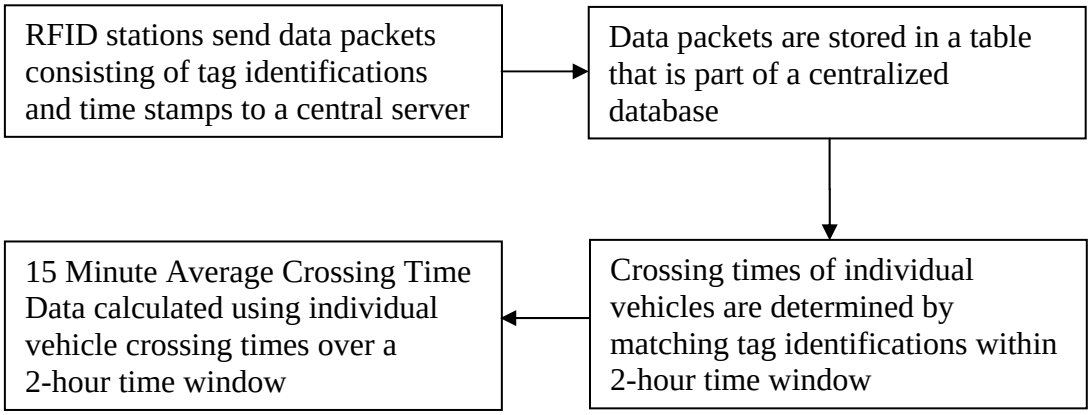


FIGURE 4 Flowchart Showing Determination of Individual and Average Truck Crossing Times

Trucks crossing the U.S.-Mexico border carry many different types of transponders issued by federal, state, and toll collection agencies. Not all trucks are equipped with RFID transponders. Trucks that are enrolled in FAST programs carry transponders provided by the CBP. Some trucks carry transponders used for tolls. The RFID systems currently deployed at both border crossings are not able to distinguish the type/provider/source of transponder since they do not maintain a database whereby a transponder's identification number can be tied with the federal/state/toll concessionaire agency that issued it.

On average, 1500 trucks cross BOTA daily. There is a general consensus that over 50 percent of the trucks have some form of RFID transponder; however, the exact percentage of trucks with "readable" transponders is not known. Analysis of archived RFID data showed that on average the system reads between 600-1000 transponders per day both on the Mexican side and the U.S. side. Many of these transponders are identified several times a day, which is due to the fact that a

large number of trucks cross the border several times a day. The system re-identifies about 150-250 transponders per day for BOTA. Considering the fact that BOTA is open 13 hours a day during weekdays and 8 hours a day during Saturday, the system produces approximately 15-25 matches per hour. FIGURE 5 shows monthly variation of sample size of crossing times.

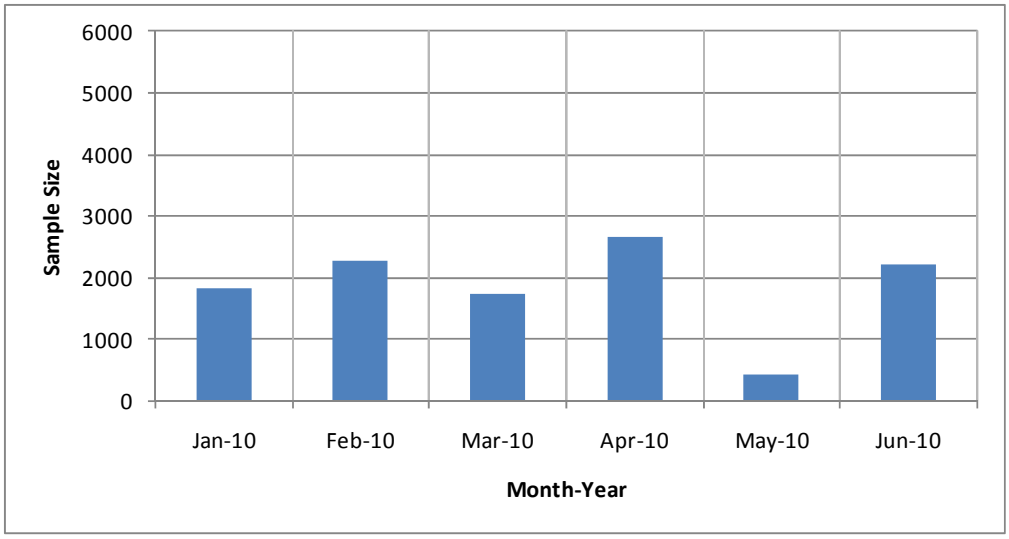


FIGURE 5 Monthly Variations of Sample Size of Crossings Times Obtained from the RFID System

2.4. Variation of Crossing Times of Trucks at International Border Crossings

In an increasingly just-in-time manufacturing economy, unpredictable and highly variable crossing times for trucks at the border act as a barrier to trade that slows and inhibits cross-border economic investment opportunities. Hence, knowing current (and predicted) crossing time accurately and reliably is of considerable interest to a wide range of stakeholders at the border (3).

The issue of crossing time variability is therefore very timely, and it is anticipated that the issue will become more pressing for the industry. A recent study concluded that the variability in border-crossing times reduces productivity of the carriers (4). The study analyzed the variability of crossing times on the U.S.-Canada border using limited sample data collected through field surveys. The study used results of in-person qualitative interviews with carriers that frequently move goods across the border and showed that that carriers use a range of strategies to reduce crossing time variability.

In terms of providing an advanced traveler information system at the border, knowledge of variability is important (5). Since variability directly impacts the accuracy of predicted crossing times, a high degree of variability indicates that the travel time would be unpredictable and the traveler information service would be less reliable (6). From the motorist's perspective, decrease in travel time variability reduces the uncertainty in decision-making about departure time and

route choice as well as the anxiety and stress caused by such uncertainty (7). The reduction in variability is as valuable as the reduction of mean travel time or even more valuable in some situations (8). All these conclusions can be equally applied to border-crossing-related real-time information.

The histograms of individual crossing times of trucks at both border crossings (described in subsequent sections) show a very high variability. As a cautionary note, trucks that have RFID transponders may not necessarily experience shorter (or longer) crossing times. Obviously, trucks carrying shipments pre-cleared by the FAST program require less crossing time than the ones that are not. However, non-FAST trucks may also carry transponders issued by the CBP. Being able to distinguish the truck that went through FAST versus non-FAST lanes by the RFID readers will provide a better way of distinguishing crossing times for different types of trucks.

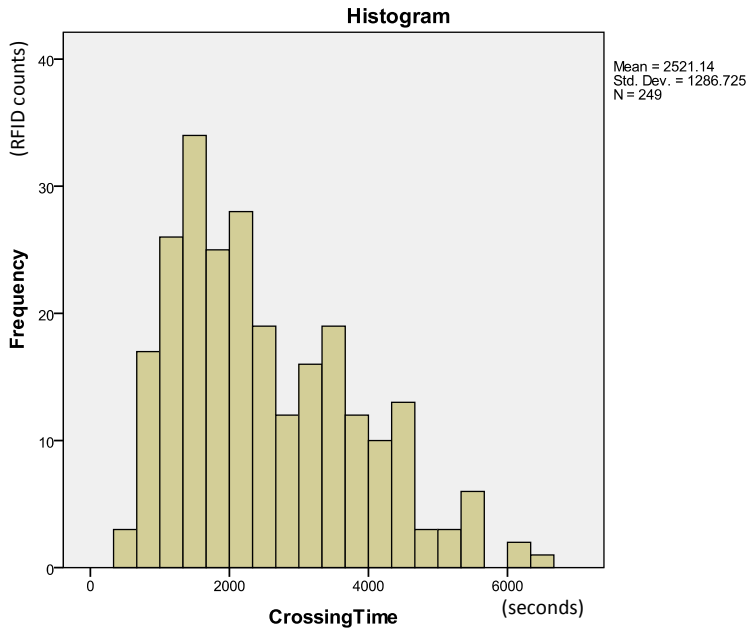
Travel time variability is comprised of three distinct components: variance of travel time from day to day, from time of day, and from vehicle to vehicle, in the case of freeway or arterial segments (9). Day to day and time of day indicate the demand on the segment, while vehicle-to-vehicle travel time difference is introduced by driver behavior, such as aggressiveness and lane choice decisions. In the case of border crossings, vehicle-to-vehicle travel time difference has to be the function of the types of shipment, randomness in inspection of trucks, driver aggressiveness, etc. Shipment type can be easily identified with right deployment, e.g., additional RFID readers over FAST and non-FAST lanes.

Also, there is a huge difference in crossing times of trucks that go through the secondary inspection at CBP and/or DPS from the ones that do not. Trucks carrying empty containers or without any containers also have shorter inspection times than loaded trucks. Ideally, if the system can distinguish the trucks based on the following, then the variability of crossing times can be distinguished and explained in much clearer terms and differentiated by shipment type:

- Trucks that go through FAST lanes.
- Shipments that go through primary inspection only.
- Shipments that go through primary and secondary inspection.

2.5. Temporal Variation of Raw Truck Crossing Times Data

The histogram of crossing time of individual commercial vehicles (travel times between the first and the last RFID stations) shows the highly variable crossing time. Because the field system cannot distinguish between types of shipments carried by the vehicles, large variations in crossing times are obvious. The variation in crossing times is also attributed to the temporal pattern, meaning time-of-day and day-of-week trends. The histogram in FIGURE 6 shows that on a typical weekday at both border crossings, 95 percent of trucks take approximately 6000 seconds (100 minutes) or less to cross the border, and 50 percent of trucks require approximately 3000 seconds (50 minutes) or less to cross the border.



Bridge of the Americas

FIGURE 6 Histogram of Truck Crossing Times on a Typical Weekday

Histograms of truck crossing times categorized by different hours of a typical weekday for both border crossing are shown in FIGURE 7. Distribution of crossing times of commercial vehicles at the Bridge of the Americas during the early part of the morning until noon shows a left shifted bell curve with a high variability of crossing time. However, after 3 PM the distribution seems to follow a Poisson curve.

Distribution of crossing times of commercial vehicles at the Pharr-Reynosa Bridge shows a similar trend of shift in distribution but the Poisson type distribution starts much earlier in the day. This behavior is attributed to the fact that the Pharr-Reynosa Bridge has a much higher number of northbound trucks crossing than BOTA. This indicates that there is a substantial presence of lengthy queue of trucks waiting to cross the border at the Pharr-Reynosa Bridge in the morning period and throughout the day. Both of these observations have to be incorporated into the forecasting models.

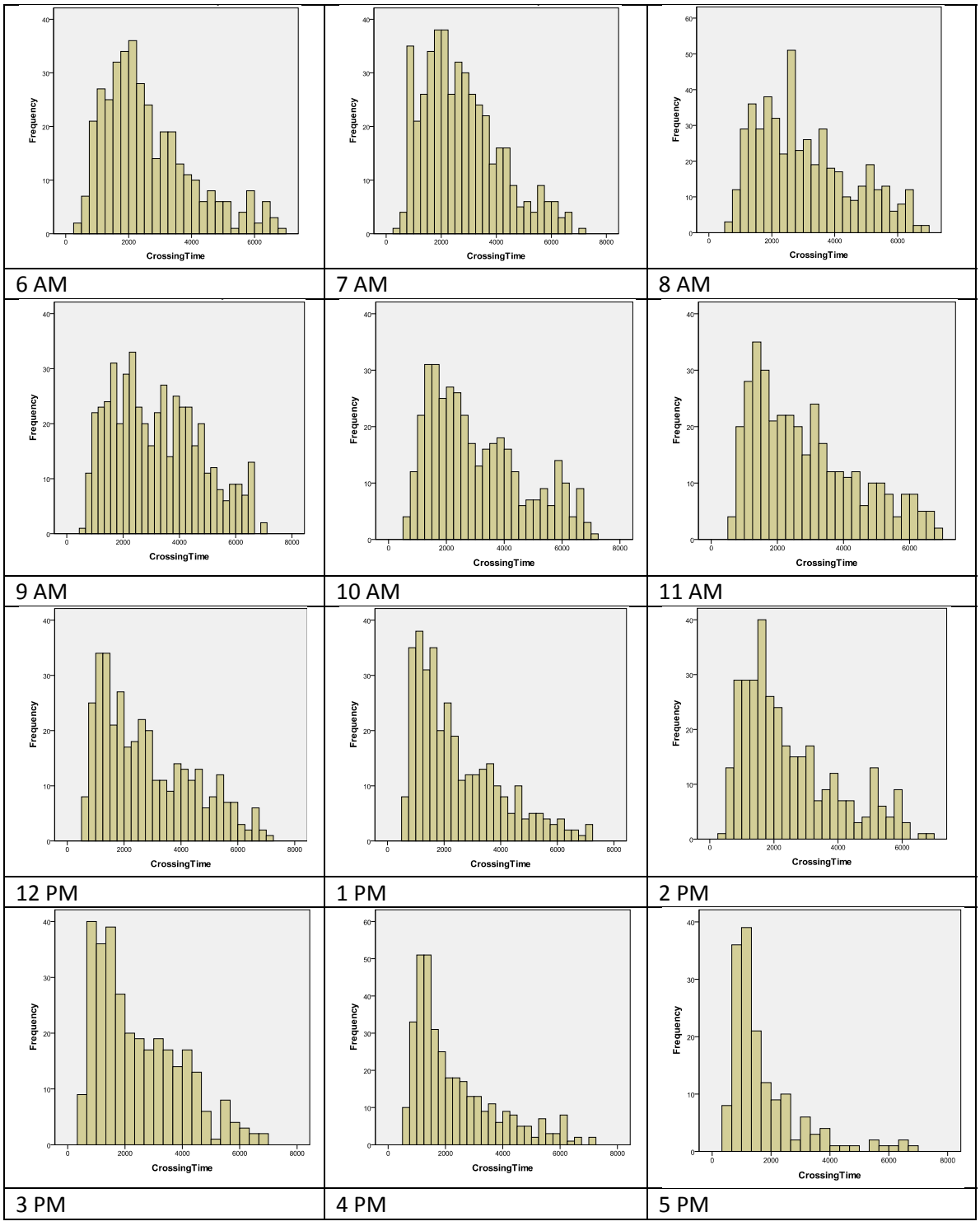


FIGURE 7 Histogram of Truck Crossing Times at the Bridge of the Americas

Box plots of individual crossing times (shown in FIGURE 8) over different time of day (times when the border crossings are open) also show highly variable crossing times at both crossings.

However, hourly average crossing times (shown by a line connecting individual boxes) do follow historically known (anecdotal) peak and off-peak trends.

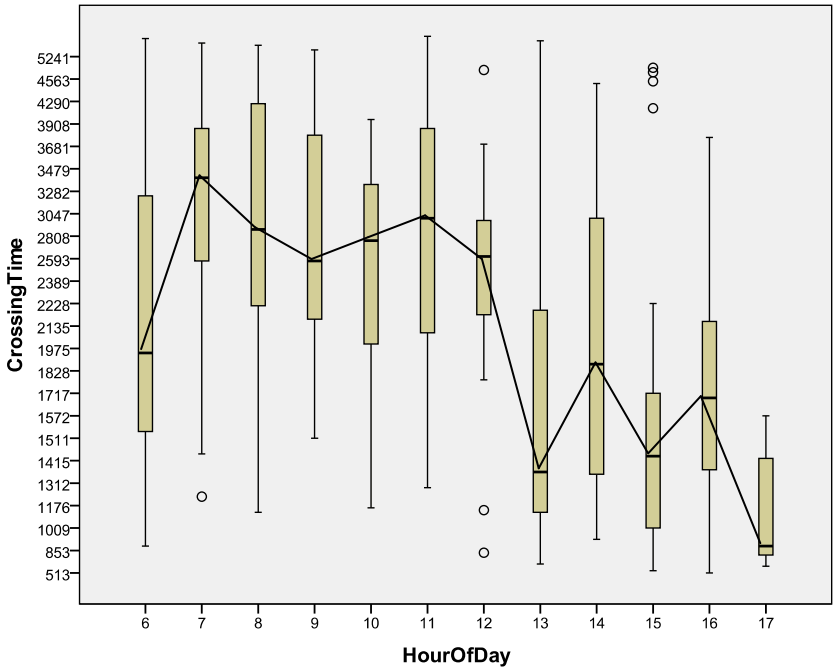


FIGURE 8 Box Plots of Truck Crossing Times at the Bridge of the Americas

2.6. Temporal Variation of 15-Minute Average Crossing Times

FIGURE 9 shows a snapshot of hourly and daily variation of average crossing times of the northbound commercial vehicles at BOTA for Monday through Saturday. The Bridge of the Americas facility operates from 6:00 AM to 6:00 PM Monday through Friday and from 6:00 AM to 2:00 PM on Saturdays. Also, the bridge closes at 3:00 PM on Saturdays, after which there are no data. Empty truck traffic prefers using this free bridge to avoid paying the toll at the Ysleta–Zaragoza Bridge.

Crossing times on Fridays are substantially higher, as Fridays are normally the busiest days at the border crossing. This increase in crossing times on Fridays occurs because shippers rush to close out the week's sales by getting as many orders off their docks as possible, which causes northbound truck traffic at BOTA to increase. Average crossing times are lower on Mondays and Saturdays, which are typically the least busy days of the week at BOTA. Also, BOTA does not seem to have a spike in average crossing times in the morning, the phenomena that is typical at the Pharr-Reynosa Bridge.

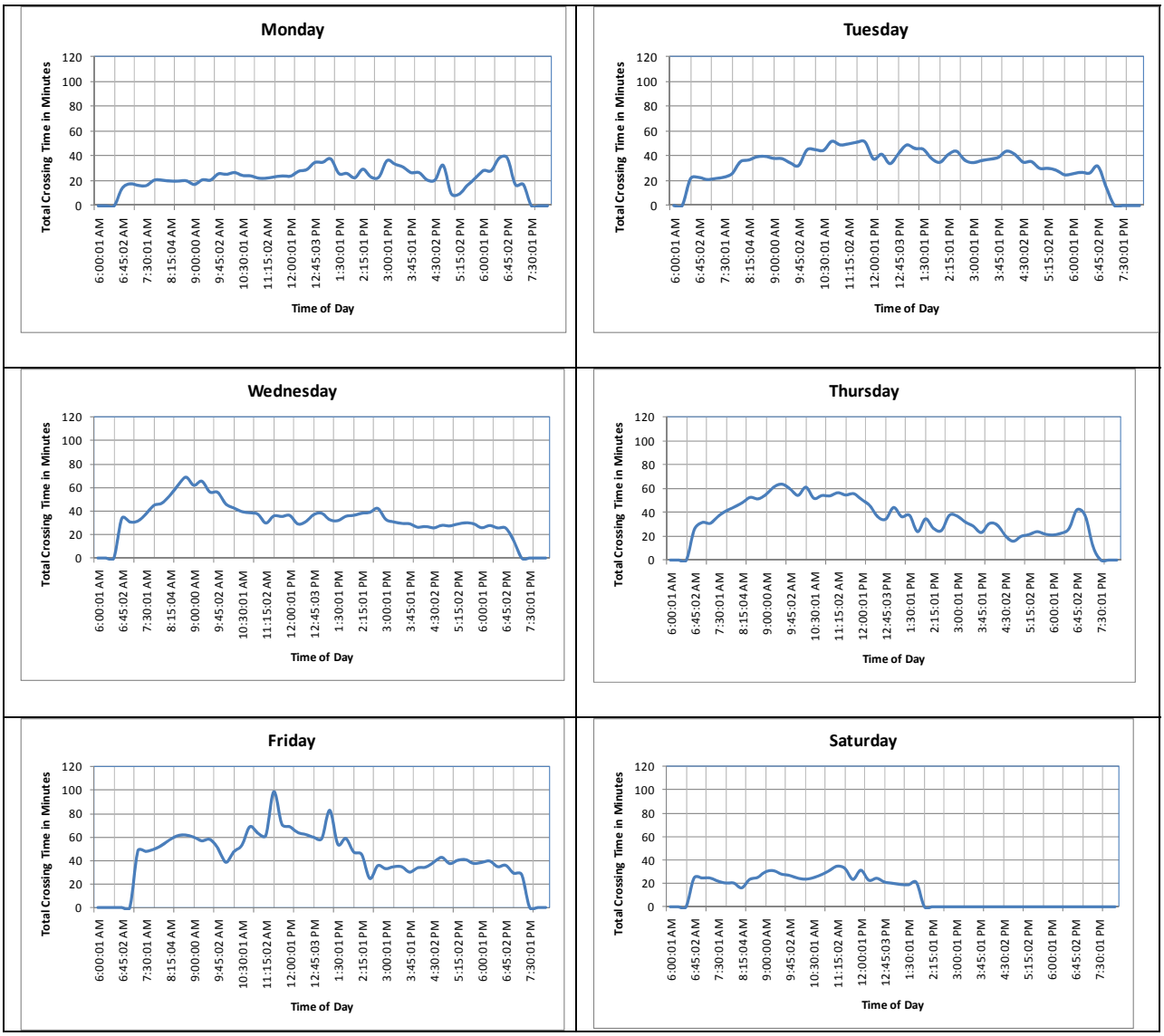


FIGURE 9 Temporal Variation of Average Crossing Times at the Bridge of the Americas on the week of September 07, 2009

FIGURE 10 shows the variation of the average crossing times at time periods when trucks arrive on the Mexican side of the border crossings. The graph is basically a histogram of frequency of crossing times at different times of day, derived from when the transponders appeared on the Mexican side. The graphs add to the fact that trucks take longer to cross Pharr-Reynosa International Bridge during early morning periods than to cross BOTA.

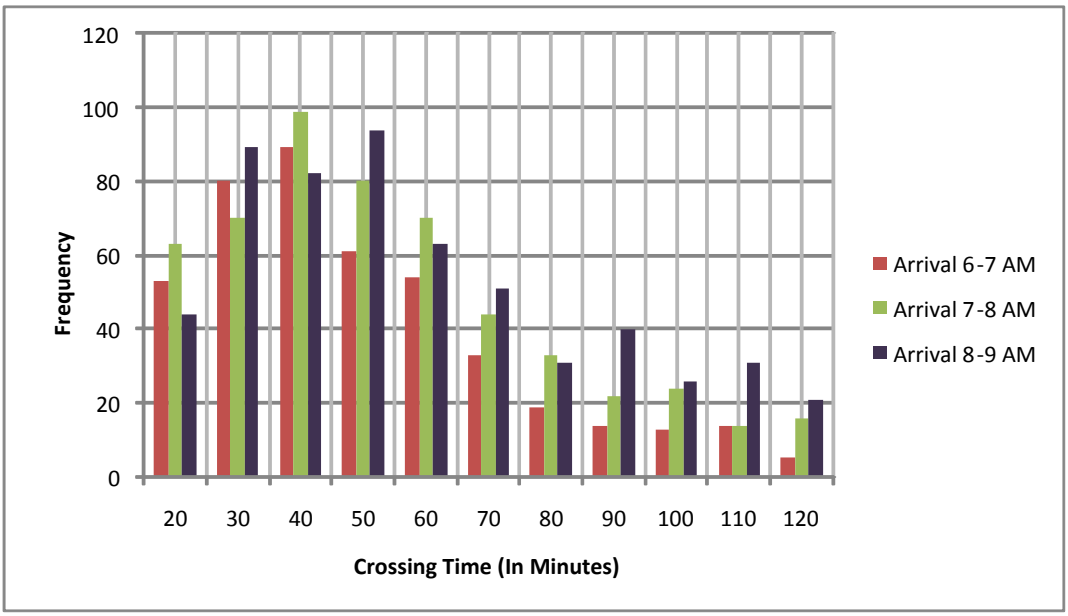


FIGURE 10 Variations of Average Crossing Times at Different Arrival Times

Chapter 3. Review of Forecasting Techniques

3.1. Background

The review focused on three main tasks: 1) gathering information on a wide range of short-term prediction models used to predict freeway travel time using RFID or automated vehicle identification (AVI) systems, 2) gathering information on parameters that affect the crossing times of commercial vehicles at the border crossings, and 3) identifying appropriateness of forecasting techniques to predict crossing times of trucks in the context of this project. Researchers used a wide variety of literature sources including on-line databases, compendium of papers, etc. The goal of the literature review was to identify data requirements, strength, and capabilities of various statistical models.

Researchers also identified appropriate techniques to filter outlier data, which may include trucks that went through secondary inspections, trucks that passed through upstream (northbound) detector but not the downstream (southbound) detectors on the U.S. side. Based on the literature review, various statistical forecasting techniques (artificial neural network, support vector machine, etc.) were developed to predict crossing times of trucks in an on-line environment.

3.2. Forecasting Techniques

Forecasting is the estimation of an expected value of variable of interest at some specified future date or time. Prediction is a similar, but more general term. However, for the purposes of this research, both terms will be used interchangeably.

Quantitative method of forecasting uses formal mathematical modeling using historical observations of a variable(s) to predict its future values. Quantitative method can be categorized into two methods — statistical and non-statistical. Statistical methods involve mathematical formulations and based on statistical properties of the model, forecast of variables can be determined. These forecasting methods are based on the assumptions that it is possible to identify the underlying factors that influence the variable that is being forecast. If the causes are understood, projections of the influencing variables can be made and used in the forecast. Examples of these methods include regression models, autoregressive models and autoregressive integrated moving average models. Non-statistical methods include forecasting models that do not possess formal statistical properties. Examples of these types of models include moving average models and various smoothing techniques.

Forecasting itself is loosely categorized as short-term and long-term. Even though difference between the two is highly subjective, short-term methods are designed to forecast only one period ahead and long-term forecasting includes multiple periods ahead. Researchers also use absolute amount of time to differentiate between the two. However, for the purposes of this research, forecasting crossing times of commercial vehicles will be for a short-term and limited to one hour ahead.

3.3. Time Series Methods

Time series forecasting methods rely on the fact that the data follow a sequence of measurements that follow non-random order. The analysis of time series is based on the assumption that successive values in the data file represent consecutive measurements taken at equally spaced time intervals. There are two main tasks while analyzing time series data — identifying the temporal trend of the phenomenon represented by the time series of observations, and predicting future values of the variable. Both of these goals require that the pattern of observed time series data is identified and more or less formally described.

3.4. Artificial Neural Networks

Artificial Neural Networks (ANNs) are complex mathematical structures that try to imitate a biological brain and its way of thinking. They are able to learn from a series of examples and apply that knowledge to unknown situations. These structures have a series of interconnected elements, known as artificial neurons and those connections are responsible for storing knowledge. One of the many types of ANN most used is the perceptron. It is made up of three layers, known as input layer, hidden layer and output layer. Different layers of the ANN are illustrated in FIGURE 11. The input layer receives the initial values of the variables, the output layer shows the results of the network for the input, and the hidden layer makes all the operations to get the results.

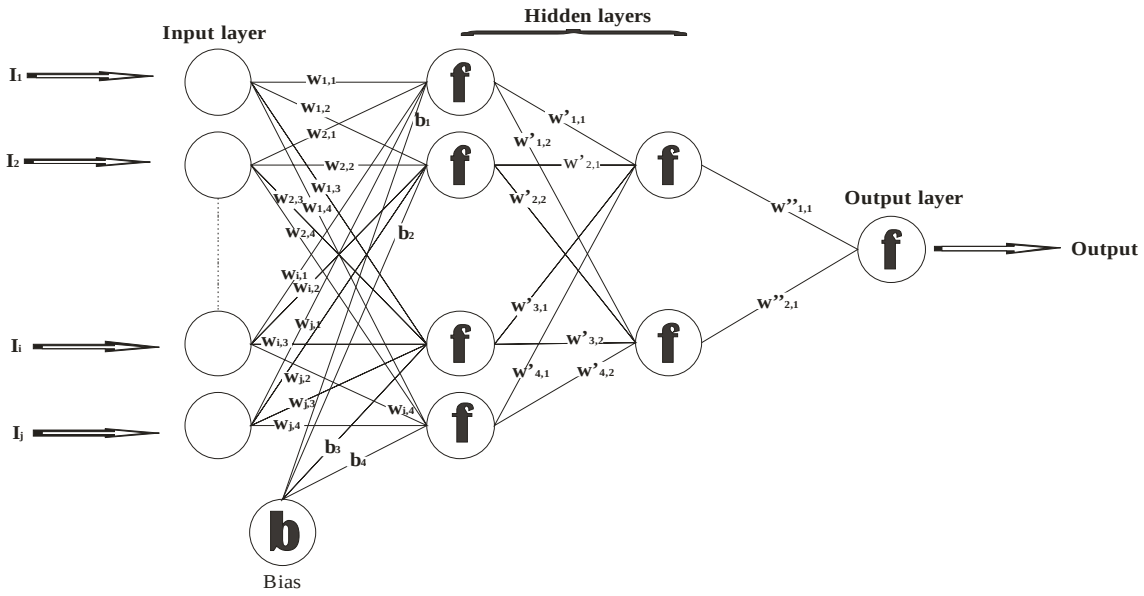


FIGURE 11 General Structure of the Artificial Neural Network

The number of neurons in the input layer is the same as independent variables, and the number of neurons in the output layer is the same as dependent variables. However, there is not a single procedure to define the number of neurons (and layers) the hidden layer ought to have. This means it is very difficult to choose a model, even for an experienced user. In general, the model is obtained by a trial and error process. There are some general recommendations about the final structure as it must be a pyramidal type or about the maximum number of neurons as a function of the number of examples, but they are only recommendations and not rules.

Advantages of ANNs are that they are universal nonlinear functional approximators. Data are not required to be in a normal distribution as in other statistical models. Their principal disadvantages are that they need a large amount of data for the training process, and the final model is a non-interpretable one, where connections between variables are unknown. In addition, the training process can be very laborious depending on the size of the original data, which is often represented as a matrix. Hence, ANNs are one of the best ways to model processes if getting output response is more important than knowing how variables are connected to one another. ANNs have been widely used to predict traffic flow variables (10) (11) (12) (13).

ANNs have been widely used in many engineering fields, especially when the relation between the variables involved in the process is not so important as to find a suitable solution to the problem. From that point of view, ANNs have become a very important tool to model specific industrial processes, but also others kind of complex processes such as environmental changes, climate, tress growing, financial markets, and traffic flow.

The ANN method was tested by including multiple variables that influenced the crossing time. Independent variables included:

- Wait time of trucks.
- Number of lanes open.

- Wait time for FAST trucks.
- Number of FAST lanes open.
- 15 minute total transponders counted by the reader on the Mexican side.
- 15 minute total transponders counted by the reader on the U.S. side.
- 60 minute total transponders counted by the reader on the Mexican side.
- 60 minute total transponders counted by the reader on the U.S. side.

Dependent variable was the average crossing time refreshed every 15 minutes. The number of lanes open was clearly an important variable that influenced the time for crossing. The total number of transponders read during the last 15 and 60 minutes and the average crossing time provides an idea about the traffic flow while crossing the border. Preliminary principal component analysis indicates that all of those variables contributed more than 2 percent of the total variation in the data set.

A multilayer perceptron structure was chosen for the ANN. It is one of the models most used in engineering applications and its nature as universal approximator makes it highly appropriate for modeling very complicated relationships. The hyperbolic tangent sigmoid function (Eq. 1) was used as the transfer function. This is equivalent to the hyperbolic tangent function and also improves network performance by producing an output more quickly. To improve the ANN results all the data were normalized according to Eq. 2. The transfer function produces an output in the interval (-1, +1), which means that data normalization is highly appropriate for improving network performance.

$$f(\theta) = \frac{2}{1+e^{(-2\theta)}} - 1 \qquad [1]$$

f(θ): Output value of the neuron

θ: Input value of the neuron

$$\theta' = \frac{\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}} \qquad [2]$$

θ': Value after normalization of vector X

θmax y θmin: Maximum and minimum values of vector X.

The training method chosen was supervised training. The initial data set was divided into three subsets at random without repetition. The training set had 2911 elements (more or less 67 percent of the total set), the validation set had 755 elements (more or less 17 percent of the total set), and the test set had 700 elements (more or less 16 percent of the total set).

To avoid the problem of overfitting during the training phase, the early stopping technique was used. Overfitting occurs when the error in the validation set starts to increase while decreasing in the training set; it is a clear indication of a generalizing loose capacity. To prevent this situation and design the ANN structure, a specific program was developed using the Neural Network Toolbox® ver. 4.0.2, from the MATLAB® Program Ver. 6.5.0. Release 13 (fig. 1). This program generates different perceptrons with different neurons in their inner layers, compares the training error and validation error every 100 training epochs, and also compares all the preceptrons generated between them.

The end result showed that the model had 70 percent error. The main reason for the poor accuracy is the great variability of the output data. Hence, researchers abandoned the idea of applying the ANN method to predict crossing time due to poor accuracy and extremely lengthy processing time, which were not appropriate for on-line prediction of crossing times.

3.5. Travel Time Prediction and Estimation Models

Research on using short-term forecasting techniques for predicting travel time on roadway segments is widespread. Past studies include application of various forms of forecasting models such as improved adaptive exponential smoothing (IAES), artificial neural network (ANN), non-parameter regression (NPR), and auto regression integrated moving average (ARIMA) models. These studies use various vehicle re-identification techniques to obtain travel time data either by using license plate recognition, transponder identification numbers, or location-based services. One study showed a strong performance of IAES is superior to other models in shorter forecasting horizon and the IAES is capable of dealing with abnormal traffic condition (14).

Another study implemented the prototype travel time prediction system for Taipei urban network by utilizing the taxi dispatching system as their location-based services data source. Real-time, historical and linear combination predictors are evaluated and compared. Location-based services provide appropriate location for the users in different locations through the mobile communication network. In this research, three predictors — current-time predictor, historical traffic pattern predictor, and weighted combination predictor — had been implemented and compared. Researchers found that current-time predictor performed better than historical predictor (15).

In addition to vehicle re-identification as a method to obtain travel time on roadway segments and thereby predicting travel time, there are studies that have used vehicle volume from vehicle detectors for predicting freeway travel times. The model incorporates real-time data provided inductive loops, which is used as raw model input and to update key model parameters in real time. The model was evaluated on a simulation test bed to examine its accuracy and robustness before field validation. The simulation testing was conducted using the Paramics microscopic simulation software. The estimated travel times by the recursive model were compared to the measured travel times of individually simulated vehicles. Field validation of the new algorithm was undertaken using data from two operational freeways in Melbourne, Australia (16). The measured travel times were compared to those predicted by three models: the recursive model, an instantaneous speed model and the Drive Time model. The recursive algorithm tracks the average travel times more closely than the other two algorithms for both test freeway sections. The instantaneous and Drive Time models produce up to 20 percent prediction error in some

cases; the recursive model produces 13 percent error and in a number of cases less than 10 percent. The recursive model is capable of accurately predicting travel times for operational freeways.

3.6. Support Vector Machine

Support Vector Machine (SVM) is a machine learning method used in many regression and classification applications. There are several applications in science and engineering that require methods to extract patterns and SVM is a recent technique to solve this type of problem. There are several different algorithms for pattern recognition but most of them focus on developing techniques to minimize misclassification/smoothing error of given observations. SVM objective is to minimize the error given the training data without overfitting it. The main advantage of SVM, over other algorithms, including ANN, is that the classifier returned by SVM takes into consideration the error minimization of new observations. SVM uses a regularization that allows some flexibility in the final outcome. The regularization terms allow researchers to accept some error deviation in the Regressive model. Usually the curve to fit the observations does not pass over all possible time-observations. However researchers need to decide which observations are important to fit. The regularization is a kind of filtering technique that basically tends to reduce the effect of outliers in the regression model.

One of the principal features of SVM is that the solution hyperplane can be found using Quadratic Programming (QP) optimization methods. Due to the duality theory any SVM problem can be cast into a QP one. This optimization method is simple but its time complexity is extremely high due to matrix operations and gradient computations needed to solve this problem. Furthermore, SVM under training of two sets with slight changes might return extremely different results, and this is one of the main drawbacks of SVM.

Very few past studies have used SVM to predict travel time. One such study using SVM obtained significant performance and high accuracy in predicting travel times (17). However, SVM does not take into account the effect of random or unexpected events. SVM finds the global solution of the problem but it sometimes encounters the problem of overfitting. Overfitting the data means that all data points lie exactly in the regressive function, which in fact is a curve replicating the observations. This is not favorable even though the error is nearly zero. The problem with such a model is that outliers become important.

SVM certainly is a state-of-the-art methods; the problem is that this method is expensive in terms of computational operations during training to find the optimal Regression Model (Kernel Classifier). Another reason the researchers did not chose this method is that the model should be dynamic; therefore it is desirable to keep the computational load as small as possible. SVM needs to compute the kernel and other constants repetitively when SVM is trained. Even though SVM is a very good machine learning tool and it is efficient for large data sets, it is not the best fit to our problem characteristics. Our intention is to split all the data collected during two weeks and build independent models for each day. The objective is to obtain a system that is efficient in terms of prediction but also efficient in serving time.

3.7. Gaussian Process

Gaussian Process (GP) is stochastic methods widely used for pattern recognition. There are various applications of GP in data classification and regression. Since GP is a supervised method it is not an optimal clustering method. However, GP is an extremely flexible tool for regressive models.

GP is Kernel Methods, and the book by Rasmussen and Williams (2006) has a very detailed explanation of kernel theory and GP. The kernel is a function that defines the relation between observations and constrains the solution space. The kernel is parameter dependent and therefore the fit of the solution depends on the selection of these parameters. The optimal value of these parameters is obtained using the Conjugate Gradient (CG) minimization algorithm. The main advantage of CG over QP relies on the computational simplicity of the algorithm and the number of unknowns. GP requires estimation of the maximum covariance and the length-scale parameters. The noise parameter is left constant since the noise parameter in fact reflects the uncertainty in the predicted value.

In FIGURE 12, a graphical representation of GP is shown and the variables are defined as follows: the x values are the influencing features, the y values are the given observations and the f values are the predictions given by the GP. The relationship that exists between variables is such that the presence of similar input features increases the relation between variables increases. Usually the *Radial Basis Function* is used to define the relation between variables and in the current project this function is the desired kernel. This function estimates the relation between events occurring in near real time. For example, assuming at 9:00 AM there is a high demand of incoming trucks; then it is very likely that the observations obtained at the same time will be very influential in the outcome for the next two hours. However, as time passes by the observation at 9:00 AM has little or no significant influence on the outcome.

It was decided that the GP would be appropriate for this project because hidden unobserved events can be implied in the decision function. Besides, the set of parameters for one model are almost the same for different models. The method used to estimate the parameters is not computationally expensive and in this case there is little knowledge of a good initial solution. Choosing a good starting point makes the algorithm converge faster, therefore the serving time is reduced.

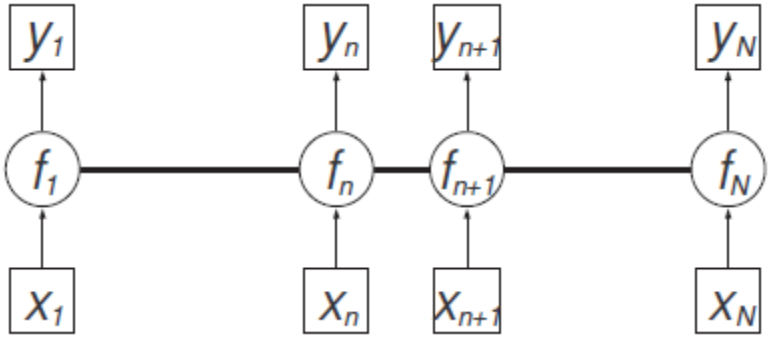


FIGURE 12 Graphical Representation of Gaussian Process

Tsuyoshi et al. proposed a solution using GP and developed a predictive system to forecast the travel time between two highway segments (18). Using a similar approach, the GP method was used to find the regression model in this project for the following reasons: the hyper-parameter estimation is simple, prediction is done by conditioning on the observed variables, and the solution space depends on the kernel covariance function. The theory behind GP is relatively easy to implement and to manipulate. This algorithm can also be implemented in high-level programming languages.

Chapter 4. Prediction Methodology

4.1. Framework of the Prediction Model

In the previous chapter several methods to forecast nonlinear time series were described. In this chapter, detailed methodology to forecast crossing times is described. The methodology primarily focuses on the Gaussian Process. FIGURE 13 illustrates key steps used in the proposed methodology. The flow of data between the steps is also described in subsequent paragraphs.

Most of the earlier studies previously described use forecasting method on data that have a unique class. For this project, three classes of data have been defined to distinguish between different types of trucks crossing the border. However, there is no prior knowledge of which class a given observation of crossing time belongs to. And hence, before application of forecasting technique, it is crucial to "cluster" crossing time observations. The main motivation is to determine these clusters dynamically on real-time data and provide accurate predictions.

Basically GP provides the knowledge hidden by the observed data, which means that GP conforms the *ensemble of experts* and the goal is to determine which of these *experts* have enough knowledge to determine future behavior, average crossing time of trucks at a given border crossing.

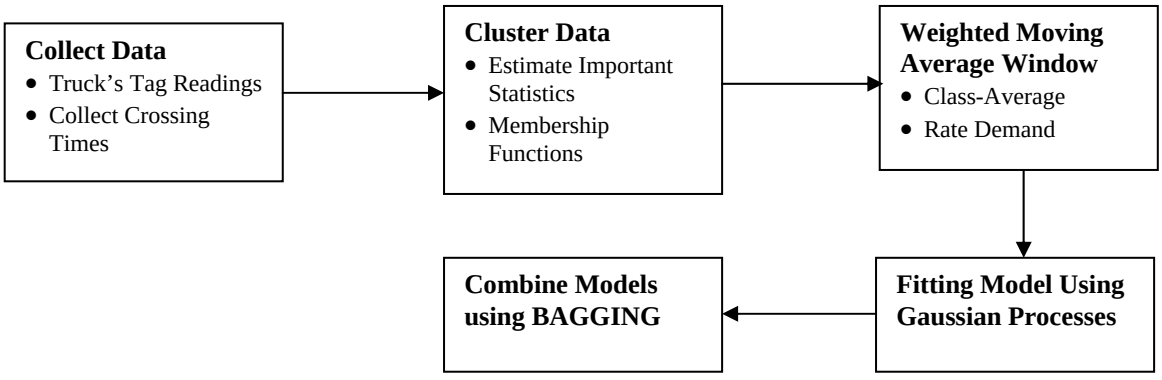


FIGURE 13 Key Steps of the Prediction Model

4.2. Brief Description of Key Steps

The methodology to predict the crossing times of commercial vehicles at the El Paso-Ciudad Juárez Bridge of the Americas is composed of multiple stages. A complete description of each step is described in subsequent sections. Truck crossing times have an extremely unpredictable behavior and coupled to this there is no prior knowledge of which vehicle class (out of three classes mentioned previously) each crossing time observation belongs to. In FIGURE 14, sample truck crossing times data observed at the border crossing are shown. The data were collected on Monday 11/2/2009 between 6:00 AM to 6:00 PM. The crossing times of trucks were obtained from RFID stations deployed on both sides of the border crossing.

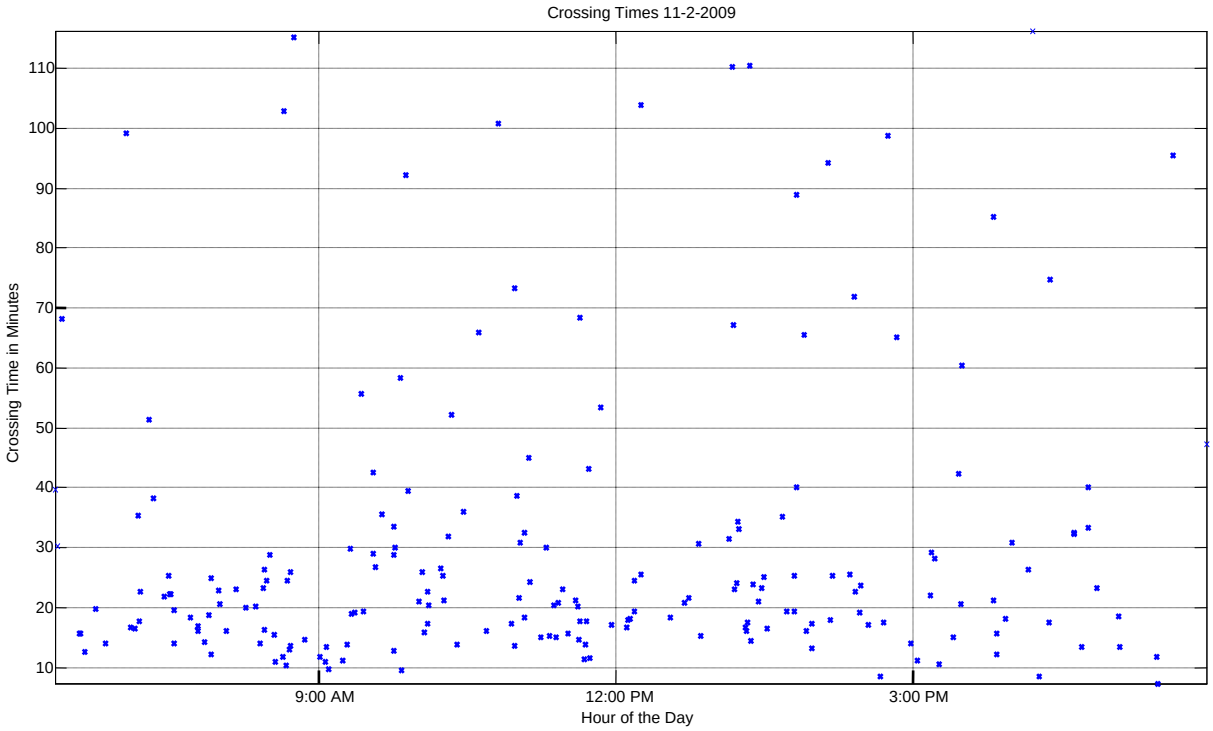


FIGURE 14 Raw Truck Crossing Times Data Collected on 11/02/2009

In FIGURE 14, the range of observed truck crossing times is between 10 and 120 minutes. In this situation, to reduce forecasting uncertainty the observed data were clustered in three different known classes (or types) of trucks. The classification of each observed data was obtained using the unsupervised statistical methods about the crossing times. In order to determine the classifier, data collected during the previous two weeks from the day to be forecasted was used. By clustering the data, it was possible to predict truck crossing times with higher accuracy and efficiency.

In the clustering step, observed data were “separated” according to the assigned class and before applying prediction algorithm, a preprocessing step was applied to the semi-cluster of the observed data. In the preprocessing step, a weighted average of the observed data was computed. The weights were determined using a class membership value obtained in the classifier and a

measure of the relevance of a given observation to the dynamics in the model. The main advantage of this approach was that the most recent data inside the window had much more relevance than data obtained several hours earlier. Using the weighted average approach, sharp transitions could be easily observed between classifications.

In the next step, a smooth fitting function was obtained to determine the characteristics of the crossing times. The Gaussian Process was used to obtain such a fitting function. The function defines the relationships (cross-covariance) between observed values based on some similarity measure of the data. One of the properties of GP is that the space of possible fitting functions is restricted by a Prior knowledge distribution and another property is the kernel function used to determine the shape of the function. The kernel determines the space of possible functions and through this function the smoothness assumption is implied.

Through the aforementioned steps, models fitting the given data for each day were constructed. Once a family of models was available, Bootstrap Aggregating or (BAGGING) method was used to combine all models by a Voting Scheme, which essentially consults from various experts (Fitting Models).

In this methodology, feature vectors of new incoming data were used to determine weights that were based on the relation between previous values and the current observations.

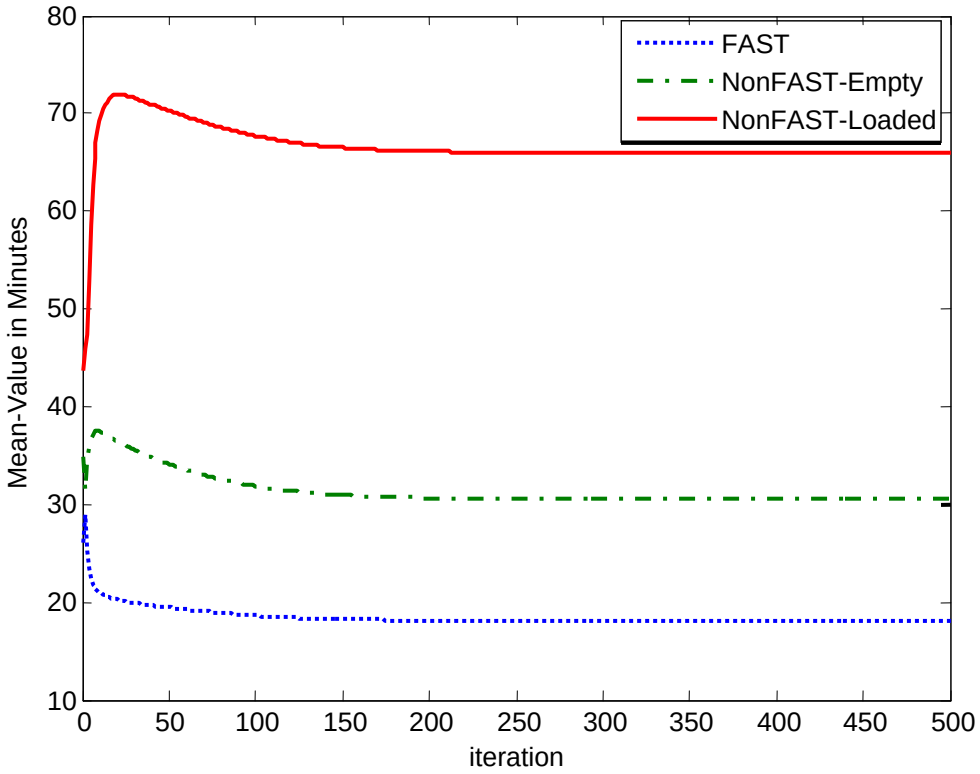
4.3. Data Classification

In general, FAST trucks have the shortest crossing time among all three classes of trucks and non-FAST-loaded trucks have the longest crossing times. Since no information about truck class is known from the RFID data, a machine learning method of Unsupervised Classification was used on two weeks of sample data.

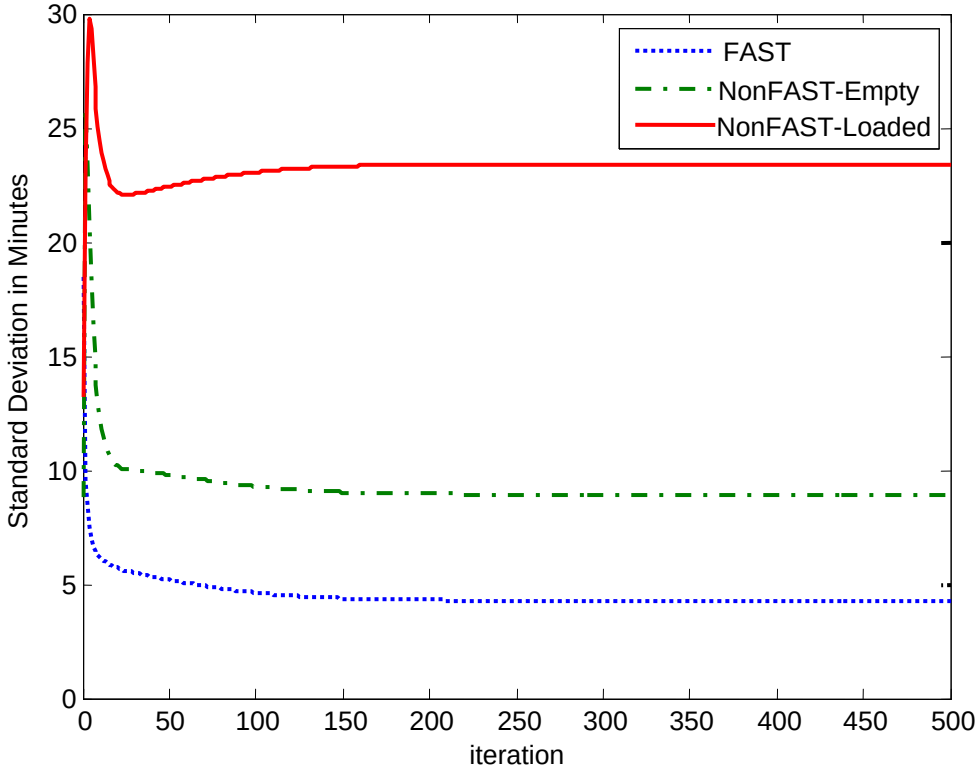
The objective of this step was to develop a “generic” classifier that would classify newly available crossing times data. The classifier was developed by statistical and optimization methods, identifying the probability density that best fits the histogram of the observed data. The Gaussian Mixture Model (GMM) was defined to estimate the set of parameters that maximizes the likelihood of the observed data falling into weighted mixture of three normal distributions. The Expectation Maximization (EM) algorithm was used iteratively to estimate unknown parameters in the process. The EM algorithm maximizes the likelihood of the information by finding the parameters that best describe the observed data.

In FIGURE 15, three different plots of parameters are shown and each plot corresponds to estimated parameters after simulation of the EM algorithm. FIGURE 15(a) shows average crossing time of the trucks for each class according to the EM algorithm. The average crossing times are as follows: trucks enrolled in the FAST program take on average around 20 minutes, empty trucks take around 30 minutes, while loaded trucks take around 65 minutes.
FIGURE 15(b) shows the actual deviation of the given observations with respect to their corresponding average crossing times. The standard deviation for FAST and EMPTY type trucks is below 10 minutes, which indicates that the samples are clustered around the mean value. Meanwhile the deviation between mean crossing times of FAST and LOADED trucks is above

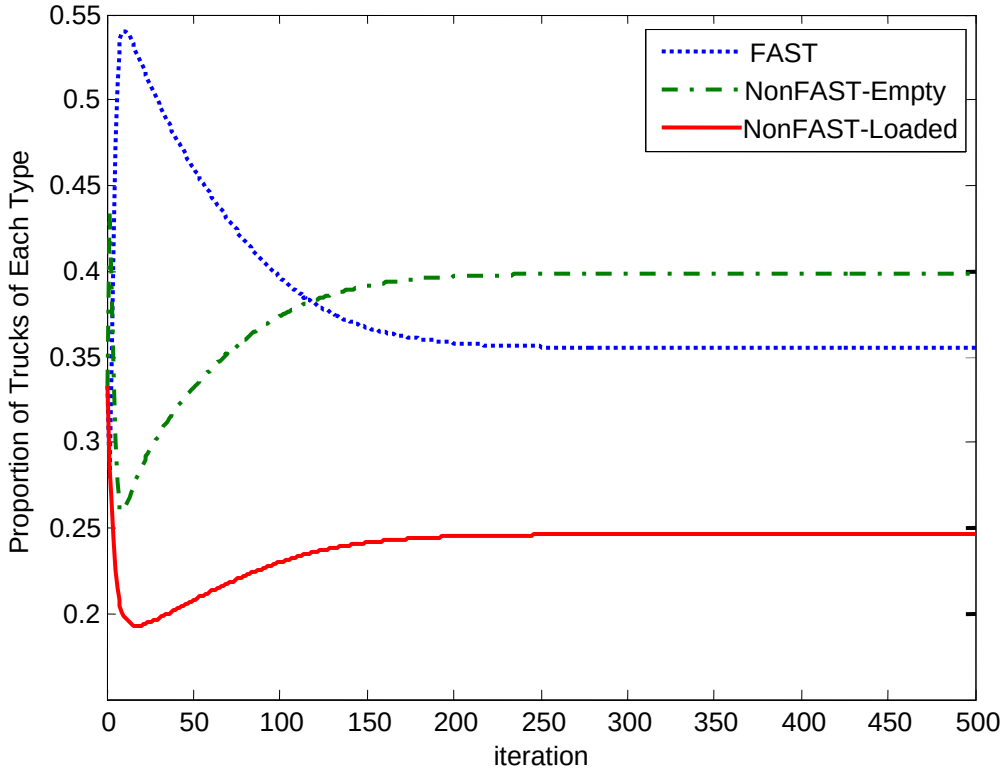
20 minutes. Basically the third Class (LOADED trucks) deviation is high for two reasons: the first reason is that the number of samples of this class is small compared to the other two classes and the second one is that EM approximates the density of a set of observations. In FIGURE 15(c) the results obtained reflect an approximation of the proportion of the number of trucks of each class.



(a) Mean Crossing Time



(b) Standard Deviation



(c) Proportion

FIGURE 15 Plots after Data Classification Step

The classifier obtained using the result from the EM algorithm is shown in FIGURE 16, which shows normalized histogram of all observed crossing times between 11/2/2009 to 11/14/09. The plot also includes approximating density and the weighted generating normal distributions. The number of samples used to estimate the classifier was 912 samples collected over two weeks. The approximating density is the combination of the three normal distributions with mean, standard deviation and mixing proportion obtained using the EM. The approximating density tends to fit the data histogram as much as possible.

The advantage of this method is the generation of three Weighted Generating Normal Distributions. These three distributions define the membership value of each observation, and the larger the value the more likely the observation to belong to that particular class. When the classifier is subjected to new observations it generates three values instead of specifying a single class value and these values are used to obtain the weights to compute the mean crossing times for individual truck classes. The classifier shown in red is the combination of the parameters seen in FIGURE 16. The three green dotted bell curves determine the degree of membership of the observations.

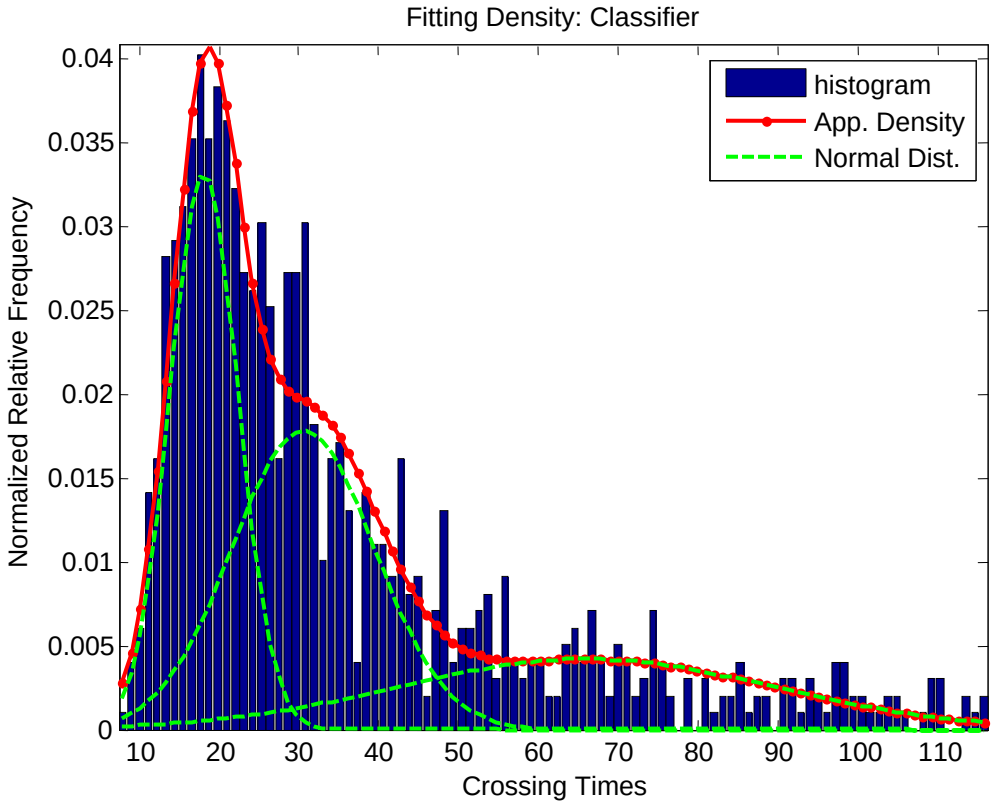


FIGURE 16 Normalized Histogram, Density Approximation, and Normal Distribution

4.4. Data Preprocessing

The proposed method to compute the average crossing time for each truck class is illustrated in FIGURE 17. The vertical blue line represents the classifier axis. The horizontal red line represents the Weighted Moving Average Window (WMAW) and the width is 120 minutes and

it moves in increments of 15 minutes. Using this approach, biased average crossing times are reduced.

Prior to determining the weights, entry time t_{entry} read at Mexico's reader and an exit time t_{exit} read after the truck crosses the border crossing are noted for each observation. The time difference t^* of these two times is the crossing time of trucks entering the queue.

$$t^* = t_{exit} - t_{entry} \quad [3]$$

Two different weights were used — one that depends on t^* and one that depends on t_{exit}. The moving average window has three parameters: starting time t_{init}, ending time t_{final}, and shifting time t_{shift}. The weighting factors are described using the following equations below. Let $w_{classifier}$ be the weight related to the classifier (the blue line) and let w_{window} the weight related to the WMAW.

$$w_{classifier} = Membership(t^*) \quad [4]$$

$$w_{window} = \left(1 - \frac{(t_{final} - t_{exit})}{(t_{final} - t_{init})}\right) \quad [5]$$

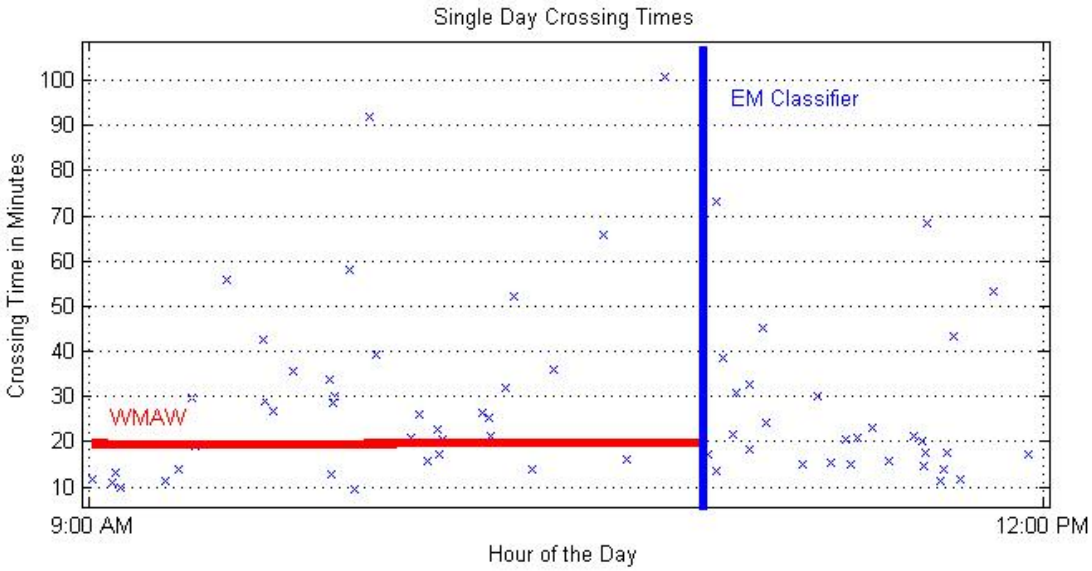
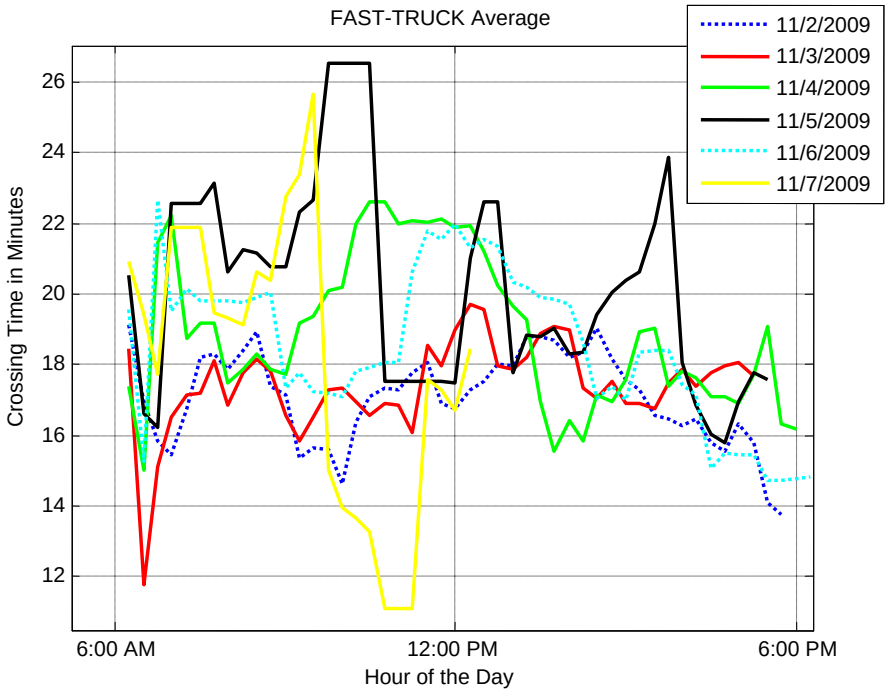


FIGURE 17 Weighted Moving Average Window

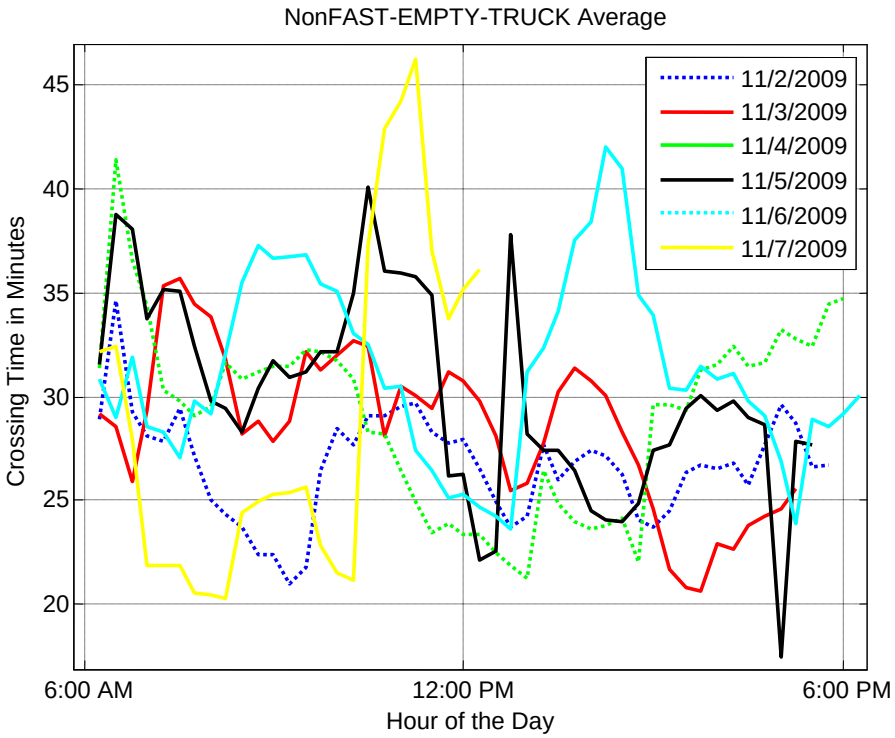
Certainly, the weights of each observation are re-scaled to maintain the laws of probability but this is just a vector normalization operation. FIGURE 18 shows the results obtained after applying the WMAW algorithm to observations obtained during the period 11/2/09 and 11/7/09. The membership functions, which are the three Weighted Generating Normal Distributions (WGND), separate the simplest sample mean into three different average values. The idea is that using the WGND, the weights of the observations depend on the value return by these

distributions. In FIGURE 18(a) most of the curves oscillate around 19 minutes, which is close to the mean value obtained using EM algorithm. Furthermore, the mean values do not deviate too much and this is due to the 5-minute deviation also obtained in the EM. The other two plots have the same effect; hence, it can be concluded that using WMAW accurate average crossing times for each class can be estimated.

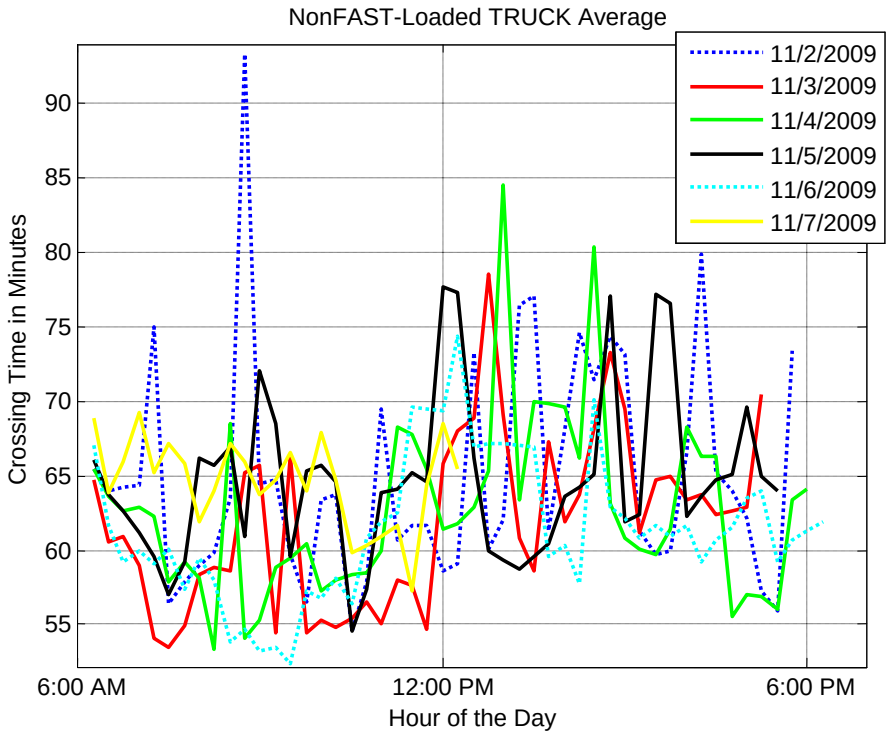
It is important to point out that all three plots reflect the estimated mean and standard deviation values obtained using the EM algorithm. Non-FAST-loaded trucks have large standard deviation but due to the preprocessing step we can see that average crossing times oscillate around 60 minutes and they have an approximate max-to-min range of 20 minutes. This is a good estimate for this type of truck since the sample set is small and its dynamics are extremely unstable.



(a) Crossing Times of FAST Trucks



(b) Crossing Time of EMPTY Trucks



(c) Crossing Time of LOADED Trucks

FIGURE 18 Crossing Times of (a) FAST, (b) EMPTY, and (c) LOADED Truck Classes

In FIGURE 19, the crossing times collected on 11/2/09 and the curves of the average crossing times are plotted. One observation that is not visible in the plot is that the average crossing times are influenced by the most recent observations of truck crossing times. The observation is given more weight if $t_{final} - t_{exit}$ is small; this means that the observation is recent.

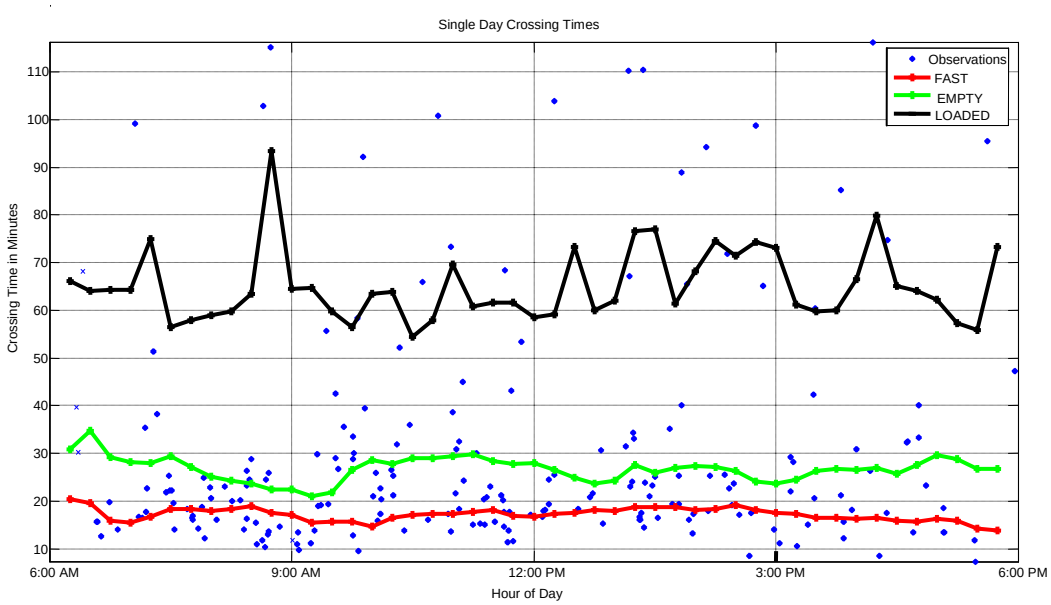


FIGURE 19 Moving Average and Crossing Times 11/2/09

The prediction of unobserved data is highly dependent on the number of trucks entering the queue. Therefore queuing rates were used as one of the key features for the prediction model. The rate of trucks going IN/OUT of the border crossing was computed using the window approach. The count of trucks was done based on the number of transponders read by the RFID stations. All transponders with time t_{entry}/t_{exit} that fell between t_{final} and t_{init} were counted, and the values were saved. A possible future problem that needs to be considered is the number of trucks carrying readable transponders. If the number of trucks without transponders increases significantly over time, then the percentage of trucks with transponders becomes smaller resulting in reduction in performance of the prediction model. In FIGURE 20, on Monday the number of trucks carrying transponders was larger compared to all other days. In order to improve accuracy of the prediction model, it needs at least, in a 2-hour time interval, five trucks with transponders. This statement is not a constraint on the forecasting method but on the prediction accuracy.

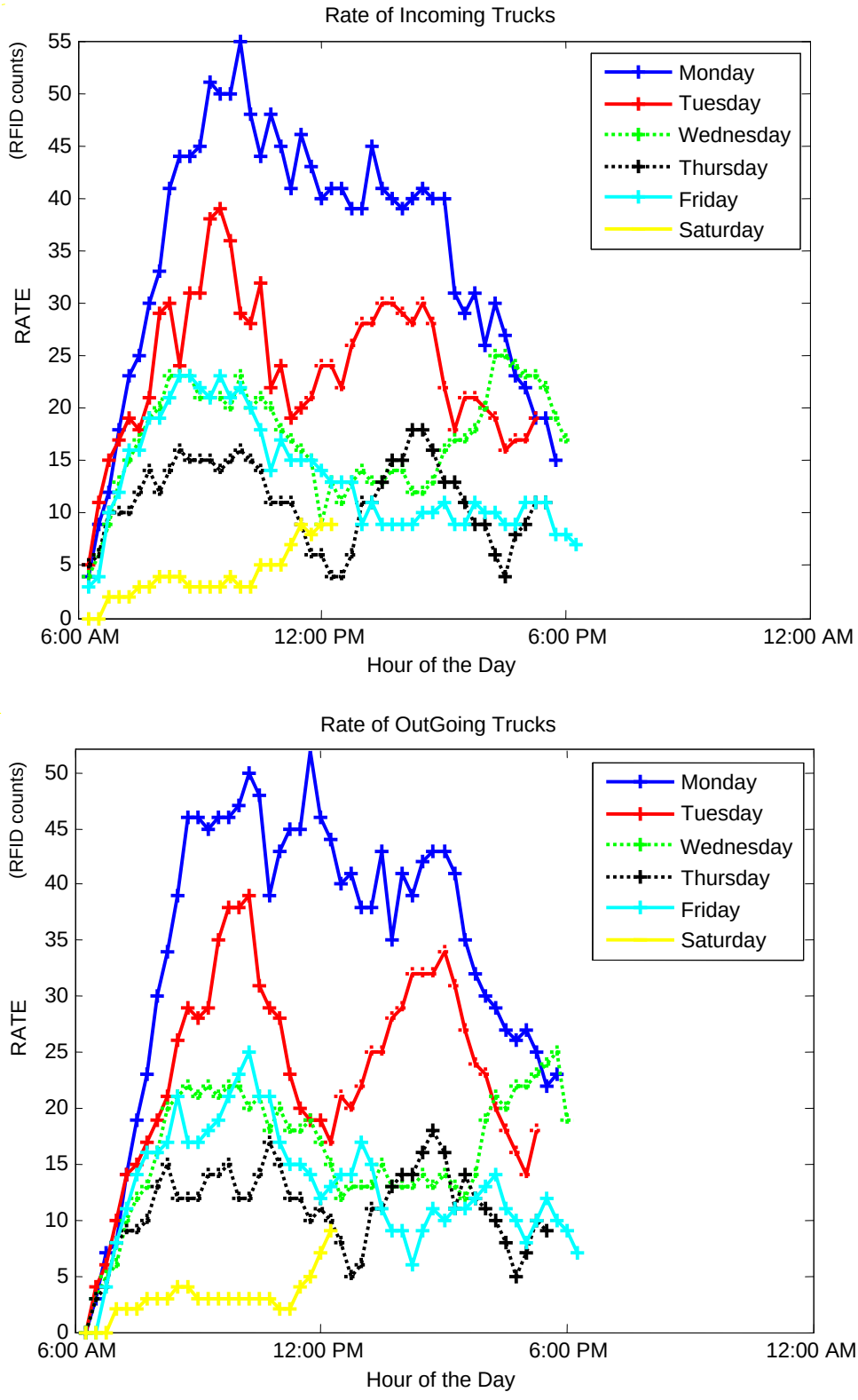


FIGURE 20 Rates of Trucks with Transponders Going In/Out of the Border Crossing

In this section basic filtering process before data fitting was explained. By filtering it is meant that we smooth the variance of all observations due to variability of the data in three different types. The techniques applied in this section basically create weights for all observations in the sliding time window and the weights filter out trucks' crossing times.

4.5. Regressive Fitting Models Using Gaussian Process

The regressive functions are determined using Gaussian Process (GP). One of the properties of regression functions using GP is that data are correlated through the reproducing kernel. The simple statistical model is assumed to be a function corrupted by normally distributed noise. There are several factors that influence the high variability in the crossing times and those are absorbed in the assumption of noise. The noisy model is as follows:

$$y = f(x) + \mathcal{N}(0, \sigma_n^2) \quad [6]$$

GP Regression (GPR) is a supervised method and the data relation is location dependent. GPR restricts the space of possible solutions and the space depends on the kernel chosen. The kernel below depends on the distance between two observations. Observations occurring nearby have higher correlation, while distant observations have negligible relation. The distance relation $(x - x')^2$ and the $x - values$ are the times of the day at which the observations are obtained. The rest of the parameters determine the maximum correlation σ_f^2 between observations, the correlation due to noise σ_n^2 and the lengthscale parameter L^2 that weights the distance between observations.

$$k(x, x') = \sigma_f^2 e^{\left[-\frac{(x-x')^2}{2L^2}\right]} + \sigma_n^2 \delta(x, x') \quad [7]$$

The models are built based on daily observations because a single model for the set of all observations will pass through the average of those. In order to build the regressive models, a set of given observations of the 15-minute average crossing times determined by using the functions described in the previous section was used. A model corrupted by Gaussian Noise is also Gaussian and new observations can be computed using the conditional distribution theorem. The assumption is as follows; suppose we want to predict y^* at time x^* and we are given the times y. The joint Normal distribution is as follows:

$$\begin{bmatrix} \boldsymbol{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right) \quad [8]$$

In order to determine the value of a new observation, the joint distribution was conditioned. The normal distribution of y^* is conditioned on y being given and the best prediction is the mean of the distribution and the variance is the uncertainty.

$$y_* | \boldsymbol{y} \sim \mathcal{N}(K_* K^{-1} y, K_{**} - K_* K^{-1} K_*^T) \quad [9]$$

The mean of the distribution gives the best prediction according to the observed data. The mean is defined as follows:

$$\bar{y}_* = K_* K^{-1} y \quad [10]$$

And the variance of the distribution is defined as follows:

$$var(y_*) = K_{**} - K_* K^{-1} K_*^T \quad [11]$$

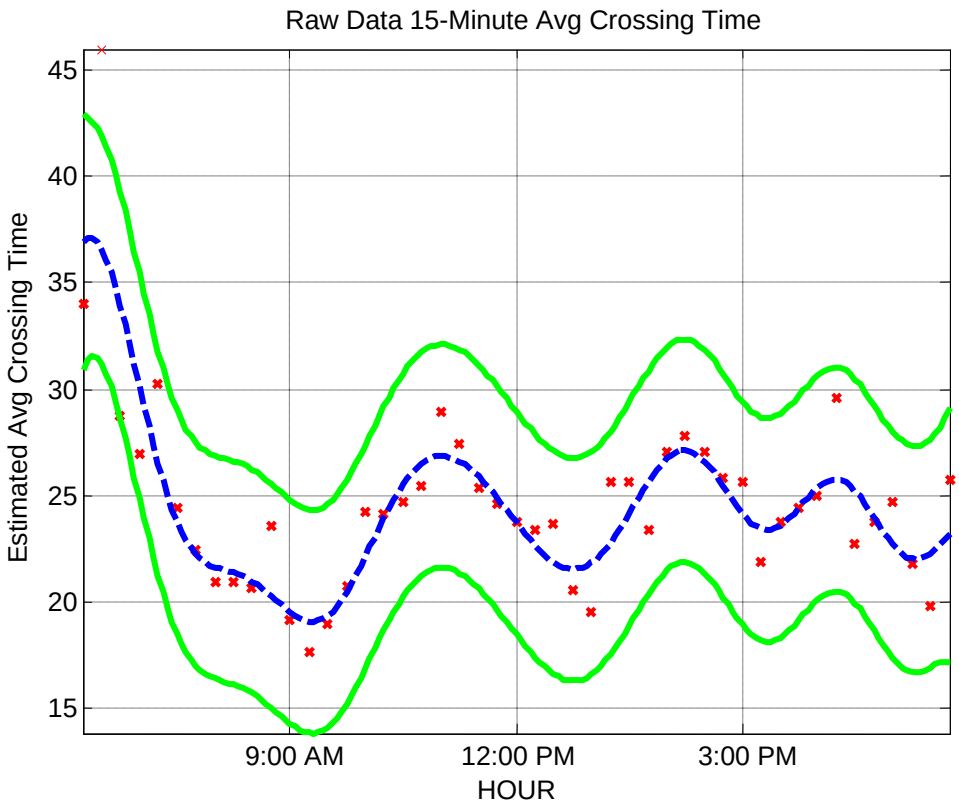
The variance determines the confidence intervals, and in our case we set this interval for 5–12 minutes.

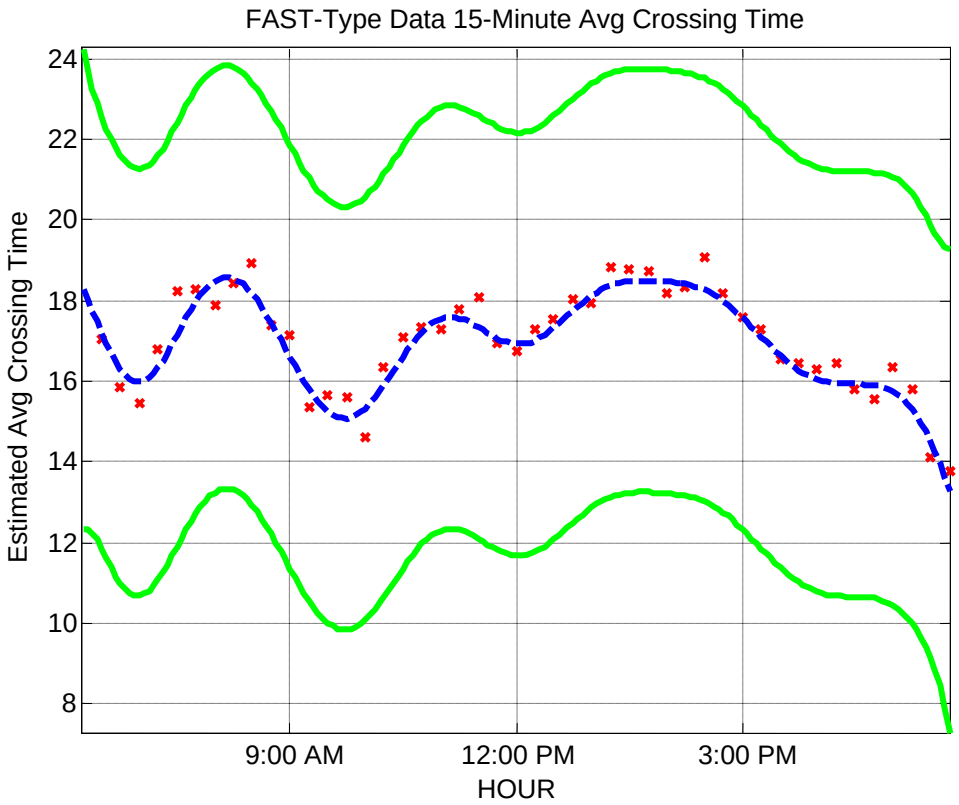
$$var(y_*) = K_{**} - K_* K^{-1} K_*^T \quad [12]$$

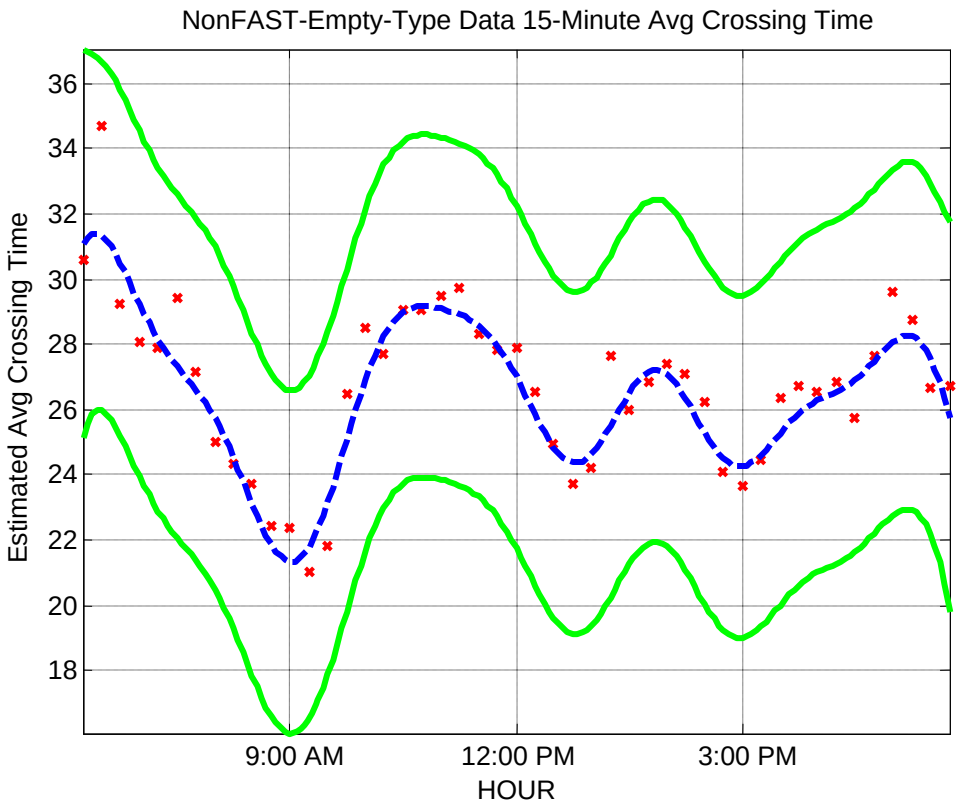
The variance determines the confidence intervals and in this case these intervals were allowed to oscillate around 5-12 minutes from the mean predicted value.

In FIGURE 21, the results from the regressive models are illustrated. The models are for the data collected on 11/2/2009, the first Monday in November. In Appendix A, results from the models for all days between 11/3/2009 and 11/7/2009 are shown. There are four different plots and each plot describes the regression model of the 15-minute average crossing times. The classes are: Unclassified average, FAST trucks, non-FAST-empty trucks, and non-FAST-loaded trucks. It is easily observed that almost all curves fit nicely over the observed values. The curves confidence intervals deviate around 5 minutes for FAST type, 7 minutes for non-FAST-empty trucks, and 10-12 minutes for non-FAST-loaded trucks. In the figure, dotted center blue curves represent prediction curve and the solid line green curves represent upper and lower confidence intervals.

The accuracy of the model depends on the size of the data set and the kernel chosen. This method differs from other methods such as Least Square in that it minimizes the error influenced by outliers. Outliers do not have a firm effect on regressive models.


Raw Data 15-Minute Avg Crossing Time
Estimated Avg Crossing Time
45
40
35
30
25
20
15
9:00 AM
12:00 PM
3:00 PM
HOUR


FAST-Type Data 15-Minute Avg Crossing Time
Estimated Avg Crossing Time
24
22
20
18
16
14
12
10
8
9:00 AM
12:00 PM
3:00 PM
HOUR



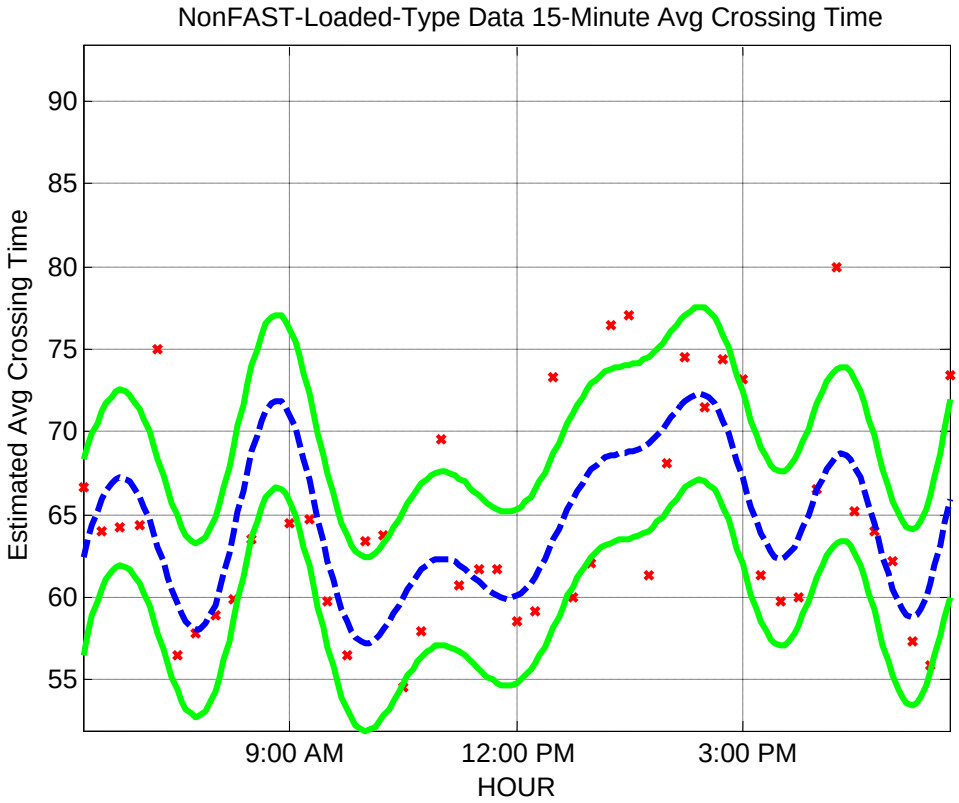


FIGURE 21 Results of Regression Models

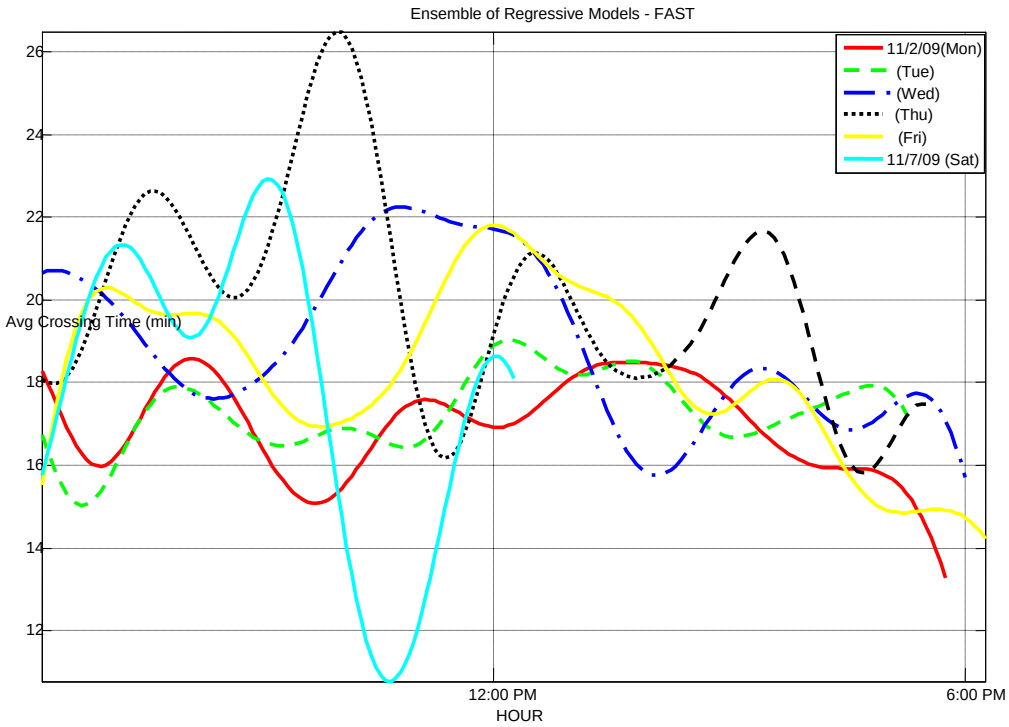
4.6. Application of Ensemble Models: Bootstrap Aggregating (BAGGING)

Usually in machine learning concepts a classifier is a system that predicts future outcomes to new incoming inputs. There are several hidden factors that change truck crossing times at border crossings. These hidden factors include above average inspection time, secondary inspection of trucks, unexpected mechanical failures of trucks, delays in the queue, and many more. The prediction model also depends on the reliability of the RFID readers and the number of trucks carrying transponders. The ideal situation is to have readable transponders on each and every truck, which would certainly improve the estimates of the average crossing times.

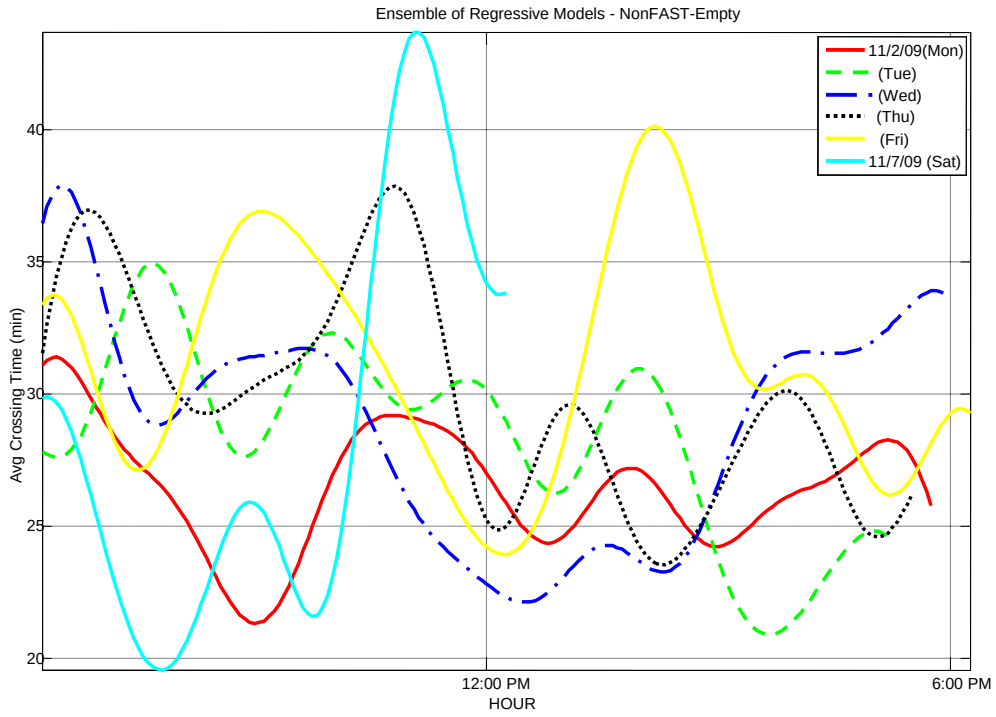
The main objective of the ensemble methods is to overcome those difficulties mentioned in the above. It is "impossible" for a single classifier to capture most of these hidden factors, and second the classifier will fit to the global average of the data collected all day. This occurs in any Statistical Learning method. Therefore, in order to improve performance, multiple classifiers are trained with different data to obtain different learners. The theory behind this idea is that a family of experts is built, and based on their knowledge, input for new observations are predicted. This technique is called Bootstrap Aggregating (BAGGING), which was used in the prediction model.

Sometimes an unexpected failure of RFID readers may occur, resulting in a partial set of data. If these events occur, ensemble models are used to build a prediction model. There are three possible situations. If enough data are collected, then a different prediction model is used than if partial data are collected, in which case a hybrid method is used that includes weighted vote of the classifiers and a feature distance to predict the crossing time. If no data are observed at the prediction instant then a weighted vote is used to predict the crossing time.

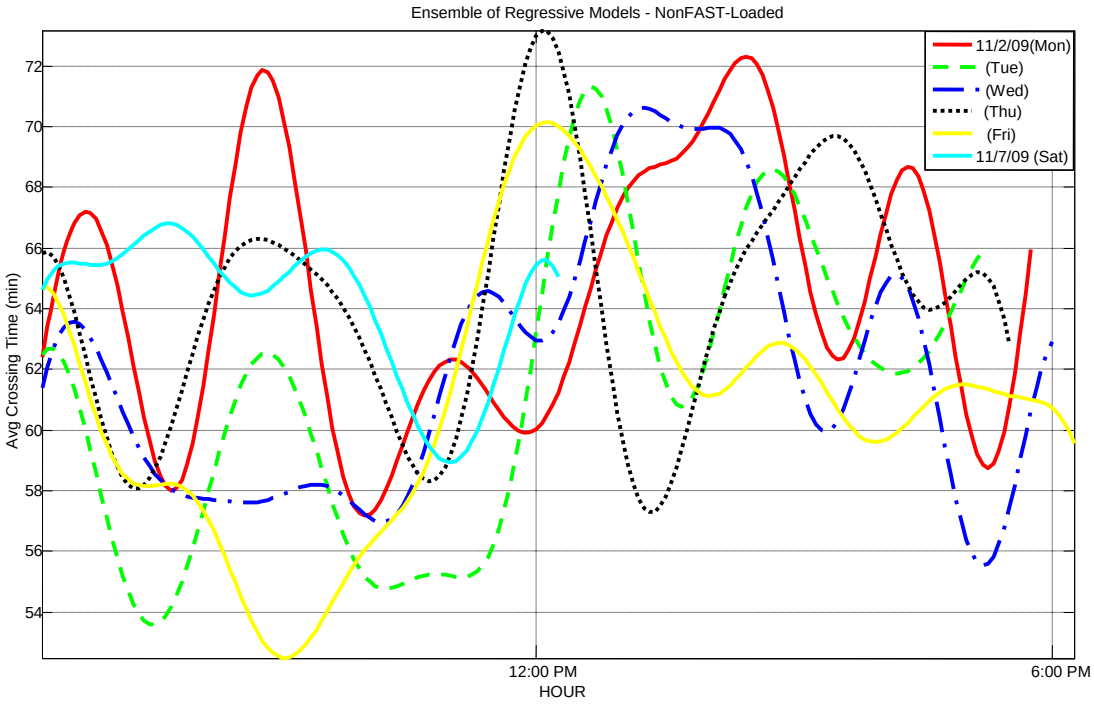
The following illustrations (FIGURES 22 through 24 and Appendix B) show the family of classifiers and fitting models (for each truck class) for the data collected between 11/2/2009-11/7/2009 and 11/9/2009-11/14/2009.



(a) FAST Trucks

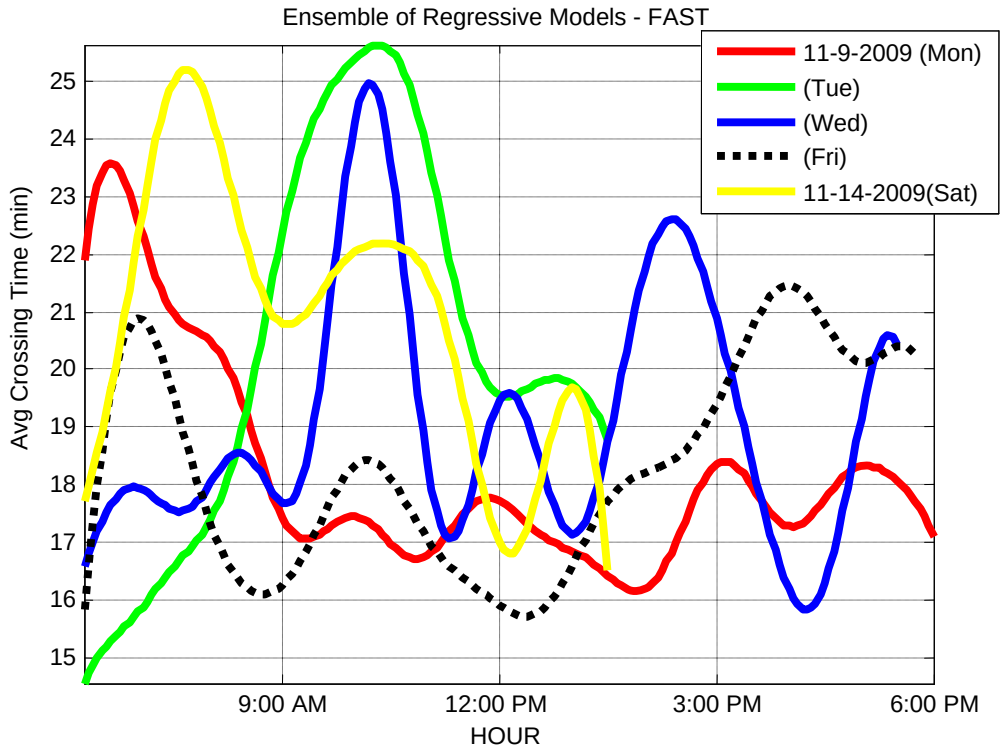


(b) Non-FAST-Empty Trucks

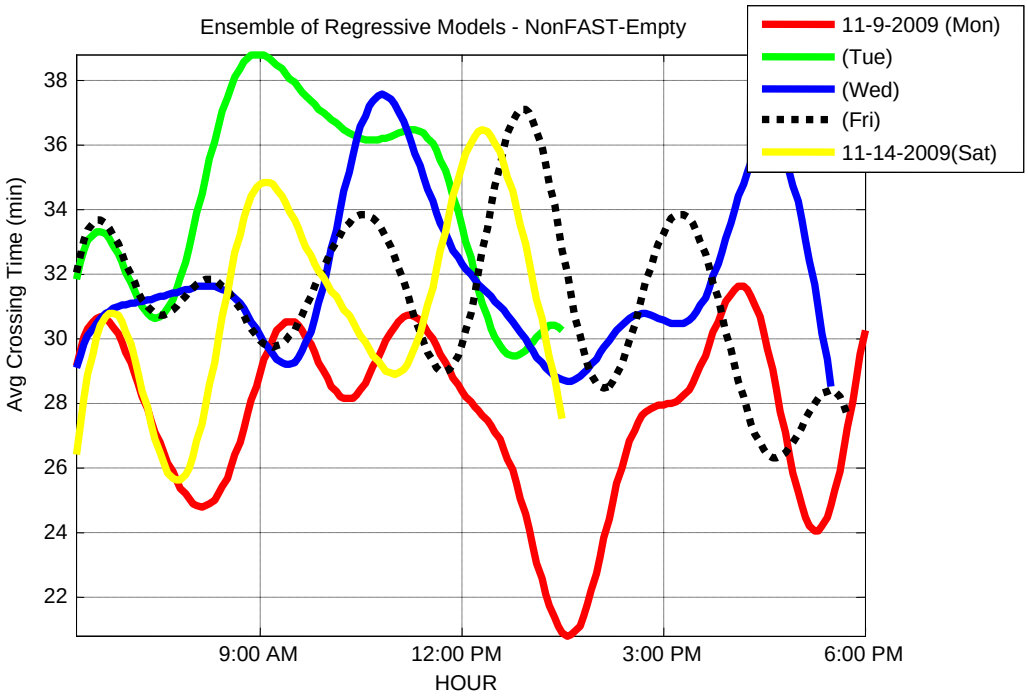


(c) Non-FAST-Loaded Trucks

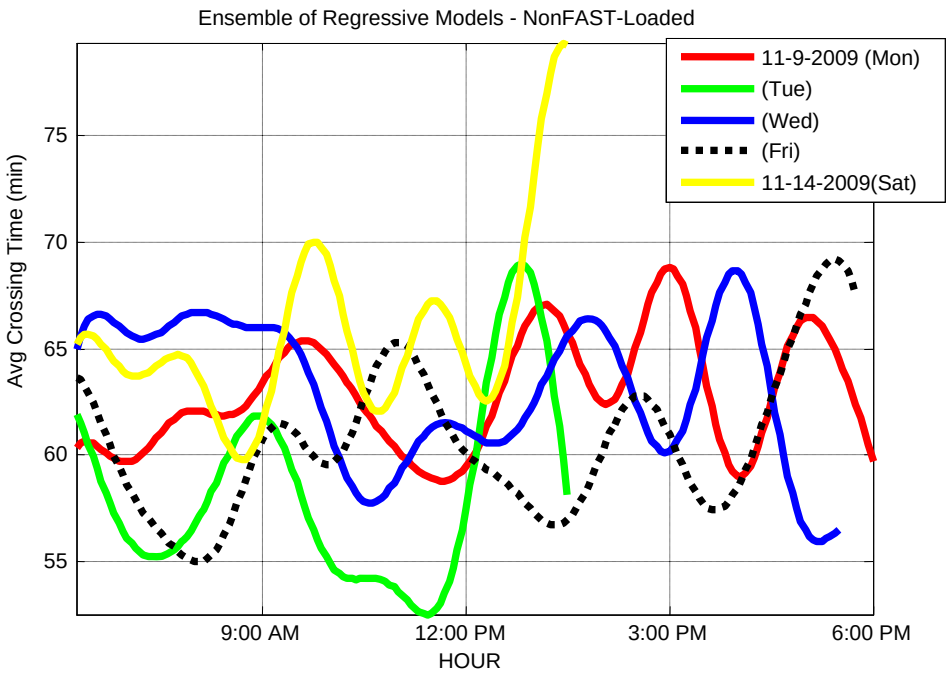
FIGURE 22 Ensemble of Fitting Models for Different Truck Classes for the Dates 11/2/2009-11/7/2009



(a) FAST Trucks

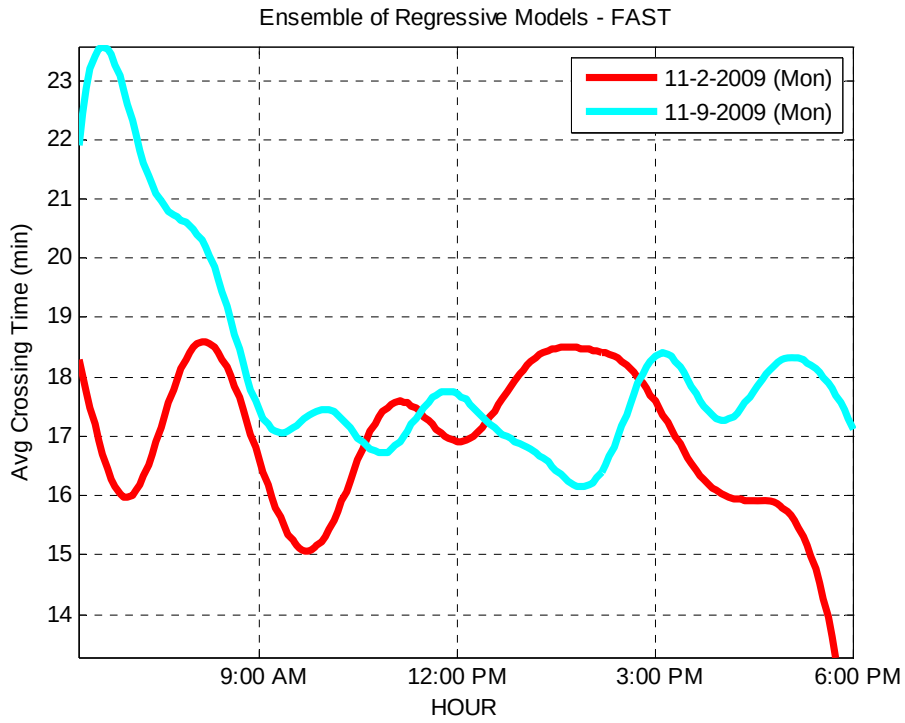


(b) Non-FAST-Empty Trucks

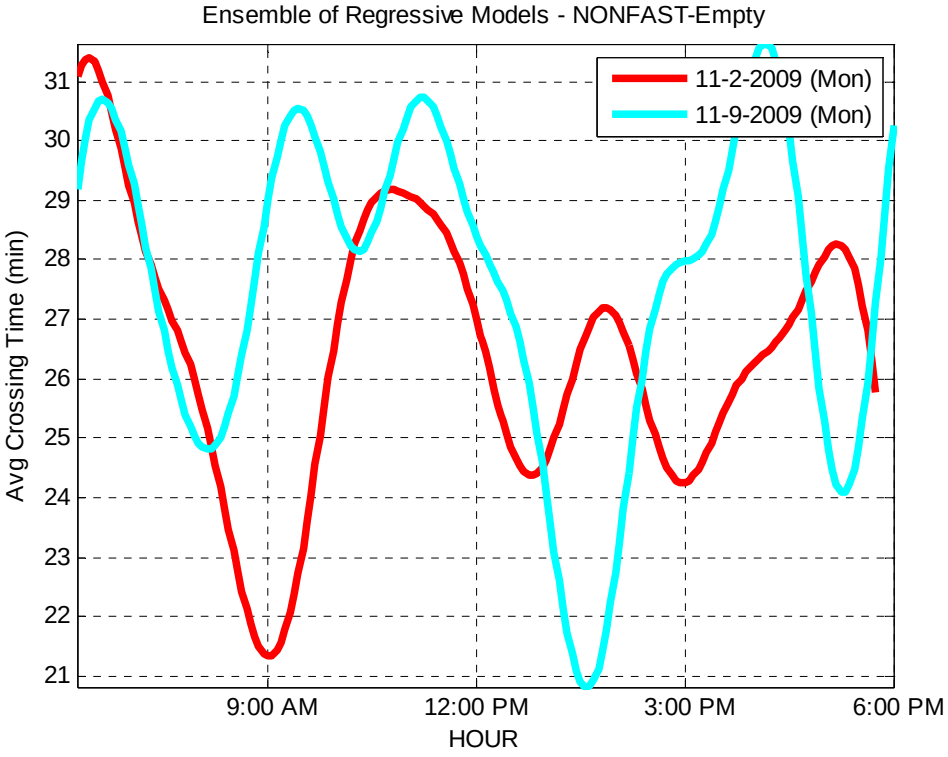


(c) Non-FAST-Loaded Trucks

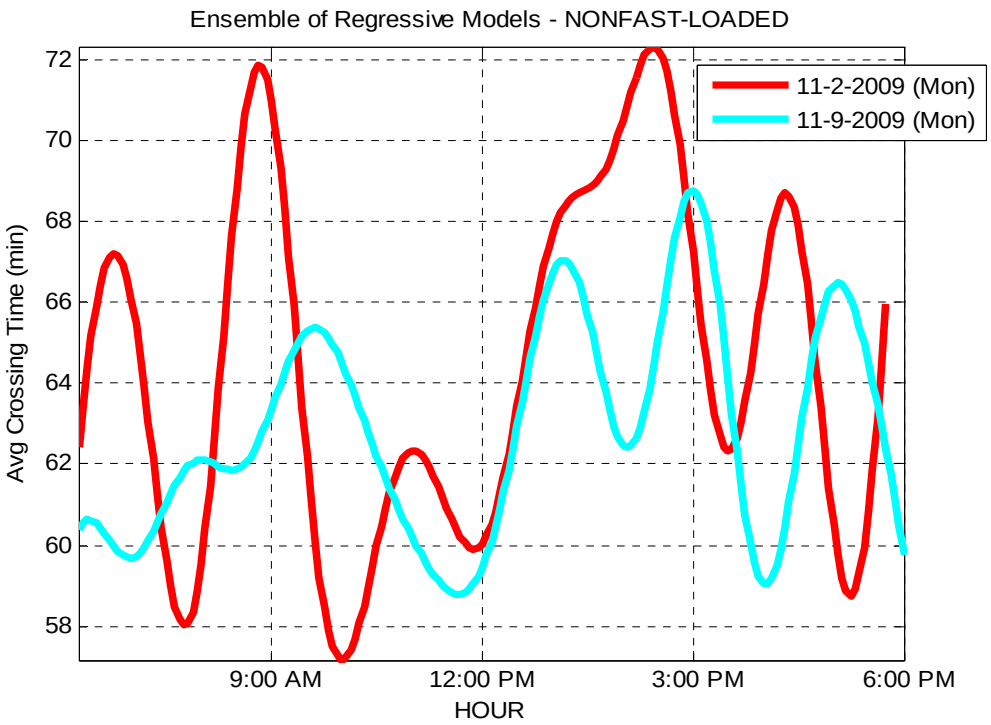
FIGURE 23 Ensemble of Fitting Models for Different Truck Classes for the Dates 11/9/2009-11/14/2009



(a) FAST Trucks



(b) Non-FAST-Empty Trucks



(c) Non-FAST-Loaded Trucks

FIGURE 24 Ensemble of Fitting Models for Different Truck Classes for Two Consecutive Mondays

Chapter 5. Prediction Results

At the end of the last chapter, an ensemble of experts was built, which defined different patterns in the observed data. In this chapter, methods to combine the knowledge of each expert are defined. In Appendix A we show patterns observed during the first two weeks of November 2009 and we plot the data of a specific day of the week in a single plot.

5.1. Prediction Methods under Different Situations

Unexpected events can occur at a border crossing and those events might affect the crossing times of trucks or the readers may fail and crossing time data may be unavailable for a prediction model to process. Hence, prediction models should be sensitive enough for such unexpected events and still be able to provide predicted crossing times with some confidence.

While predicting crossing time during such events, it is still necessary to assume that RFID stations are working properly for the previous two weeks and the day after the event for which the crossing time data would be available. With no input data on the day of the unexpected event given to the prediction model, weighted vote of ensemble models can be used to predict the crossing times.

In the event that crossing times are not available for a short period of time (a few hours) due to temporary hardware failure, then for that time period no crossing time observations are available at all. To predict crossing times during this period, the prediction model will use the same weighted approach done with no observed data and then combine it with the weights based on the historically observed data. Finally, in the event of normal operation the data collected during the day is used to feed the prediction model and forecast the average crossing time for each truck class.

The following sections analyze the performance of the prediction model during the abovementioned events.

5.2. BAGGING Using 2-Experts on the Same Day of Week (No Observed Data)

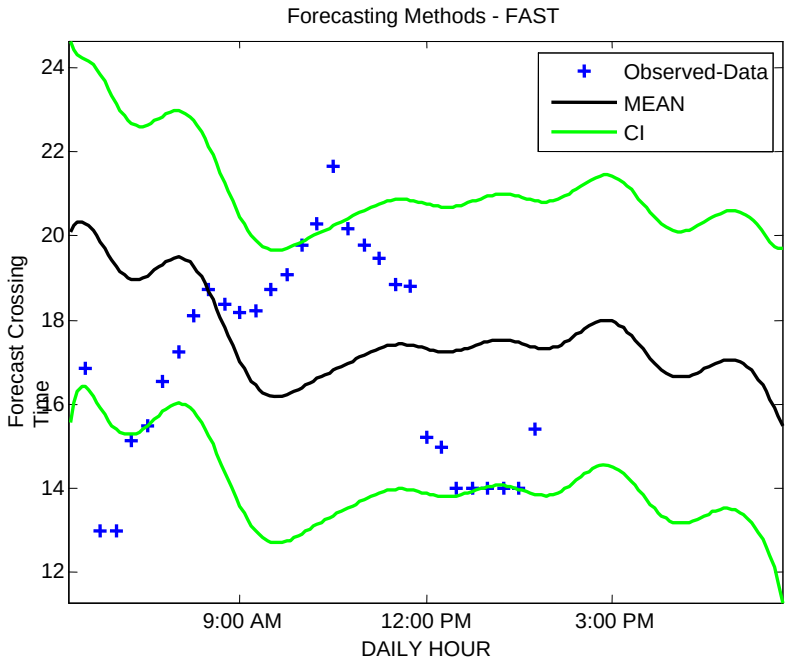
BAGGING, based on two experts includes combining the models for two days. This method is used in a situation when there is a need to predict the crossing time on a particular day of the week but due to an unexpected event sensors are not collecting data. Hence, there is no knowledge of the data on the current day. In this method we predict the crossing time using only two experts, which is shown in Equation 11.

$$E\left(Y_{prediction}^{current}\right) = \frac{1}{Total\ Number\ of\ Days} \sum_{Days=\{WeekDays\}}^{Total\ DAYS} Y_{historic}^{previous} \quad [13]$$

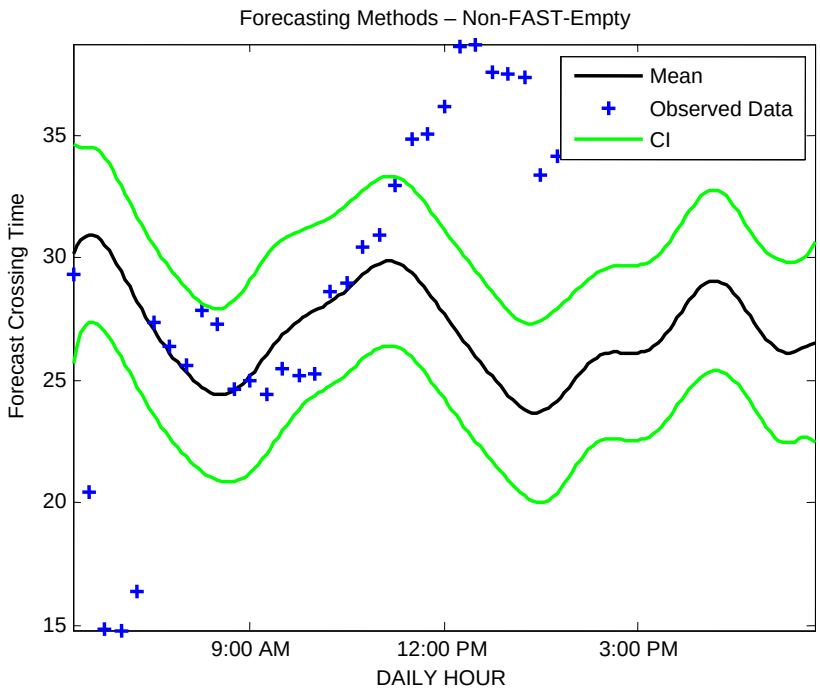
Equation 11 assumes that for a particular day of forecast, e.g., Monday, data from the previous two Mondays are the best indicator and hence other historic data can be disregarded. TABLE 1 shows the results of the BAGGING method applied to the crossing time data observed during 11/16/2009 to 11/21/2009. Performance measures include minimum and maximum absolute error and the mean square error. The most relevant measure is the maximum absolute error since it is the largest error between the prediction model and the actual observed average crossing time. In FIGURE 25 and Table 1, the prediction model, the actual observed data, and upper-lower confidence intervals (CI) are shown. The results show prediction errors of $0-7$ minutes for FAST, $0-12$ minutes for non-FAST-empty, and $0-15$ minutes for non-FAST-loaded trucks. Certainly this is not achieved using the simplest approach, though we did not use any data to forecast the crossing times.

Table 1 Performance Measures from BAGGING (Using 2 Experts) Method

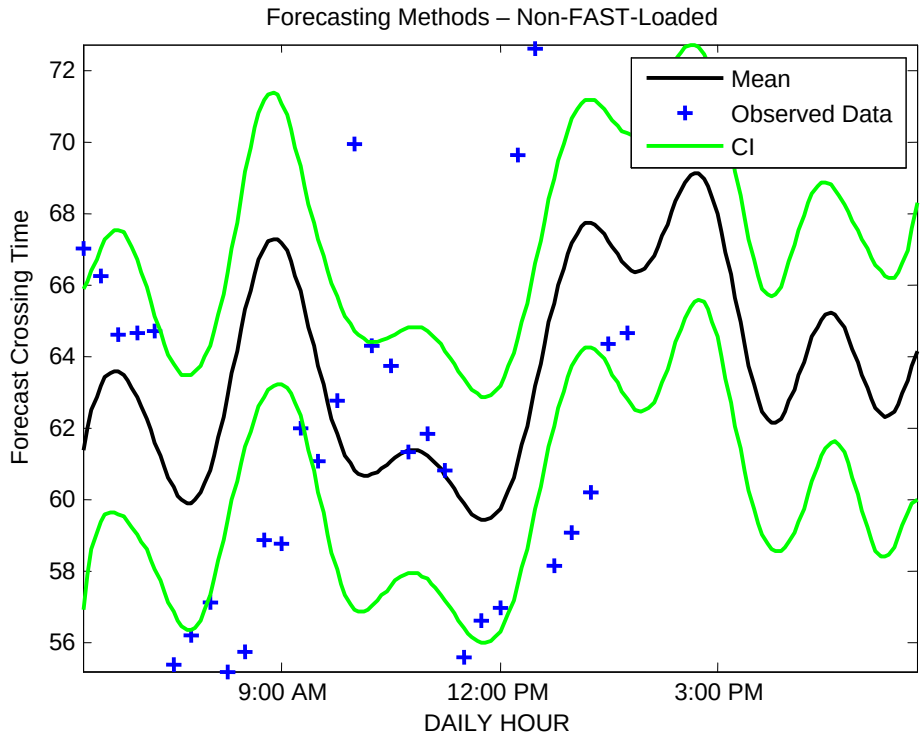
Performance Measure	FAST	Non-FAST-Empty	Non-FAST-Loaded
11/16/2009-Monday			
Minimum Absolute Error	0.04	0.43	0.43
Maximum Absolute Error	7.34	16.04	12.62
Mean Square Error	408.62	2190	702.85
11/17/2009-Tuesday			
Minimum Absolute Error	0.14	0.10	0.09
Maximum Absolute Error	12.19	16.40	16.60
Mean Square Error	1042.7	1742.8	1265
11/18/2009-Wednesday			
Minimum Absolute Error	0.07	0.03	0.36
Maximum Absolute Error	9.26	18.47	23.98
Mean Square Error	503.58	1140.9	1834.9
11/20/2009-Friday			
Minimum Absolute Error	0.16	0.08	0.05
Maximum Absolute Error	7.25	25.87	24.19
Mean Square Error	512.02	5888.8	5363.6
11/21/2009-Saturday			
Minimum Absolute Error	0.77	0.02	0.01
Maximum Absolute Error	7.59	30.61	6.70
Mean Square Error	293.58	3415.6	315.39



(a) FAST Trucks



(b) Non-FAST-Empty Trucks



(c) Non-FAST-Loaded Trucks

FIGURE 25 Prediction Results using BAGGING Method (Using 2 Experts)

5.3. BAGGING Using Two Weeks Ensemble of Experts (No Observed Data)

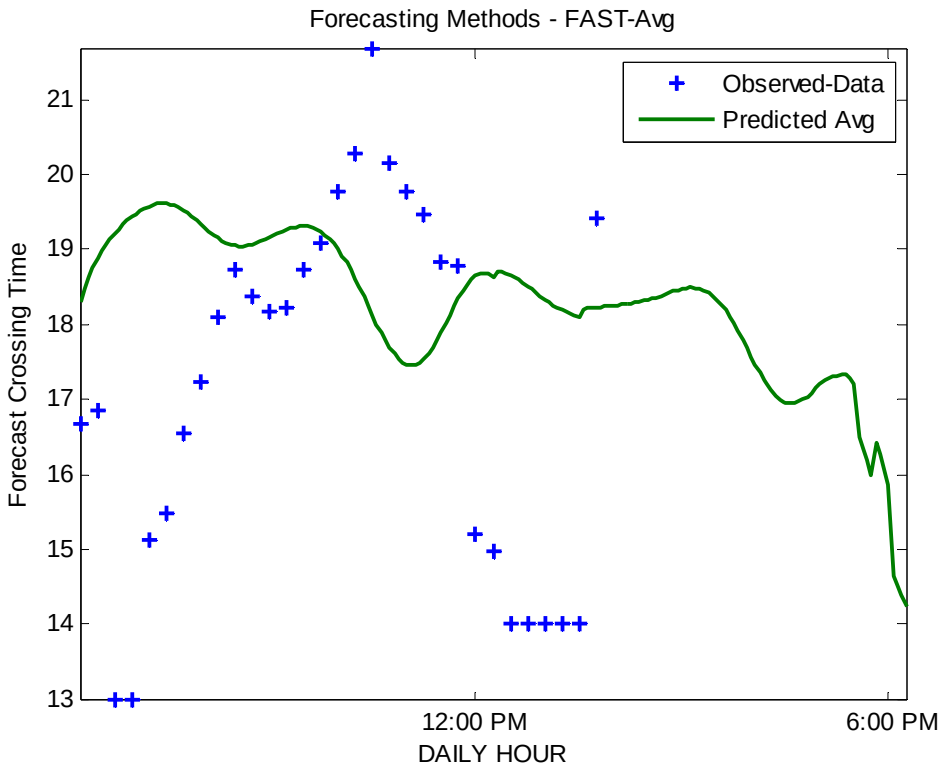
The previous method combines two models for previous two days only for predicting crossing time on a particular day. In this method, all models for the period of two weeks are combined to cast a weighted vote for each model in the ensemble. The weights are determined using a uniform random number generator and by normalizing the values to maintain probability laws. Weights were randomly generated and simulated resulting in slight improvement in performance measures compared to the previous approach. The BAGGING equation used in this method is given below.

$$E\left(Y_{prediction}^{current}\right) = \sum_{Days=\{WeekDays\}}^{Total\ DAYS} w_{day} * Y_{historic}^{previous} \quad [14]$$

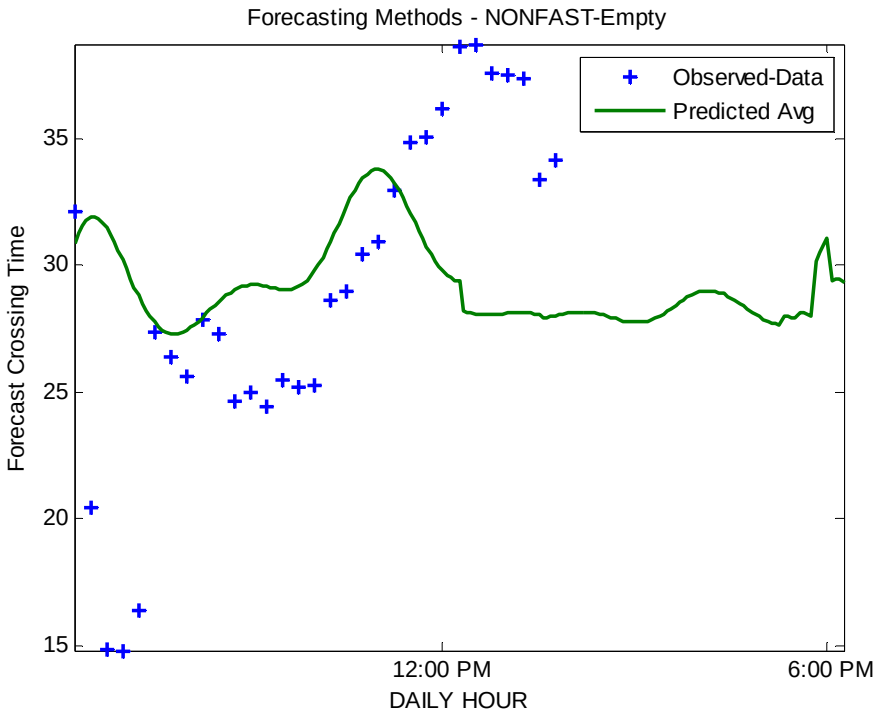
In FIGURE 26 and TABLE 2, performance measures for predicting the crossing time on Monday, 11/16/2009, are shown. Compared to the results in TABLE 1 for the same day, the improvements are minor.

Table 2 Performance Measures from BAGGING Method (Using 2 Weeks Ensemble of Experts)

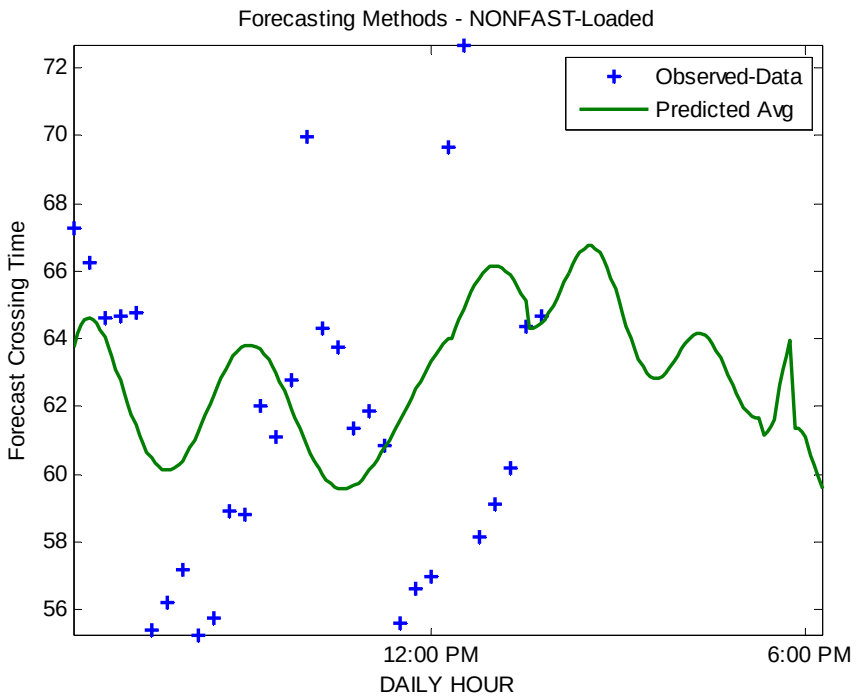
Performance Measures	FAST	NON-FAST EMPTY	NON-FAST LOADED
11/16/2009-Monday			
Minimum Absolute Error	.15	.18	.11
Maximum Absolute Error	5.77	16.96	12.51
Mean Square Error	304.16	2065	817.34



(a) FAST Trucks



(b) Non-FAST-Empty Trucks

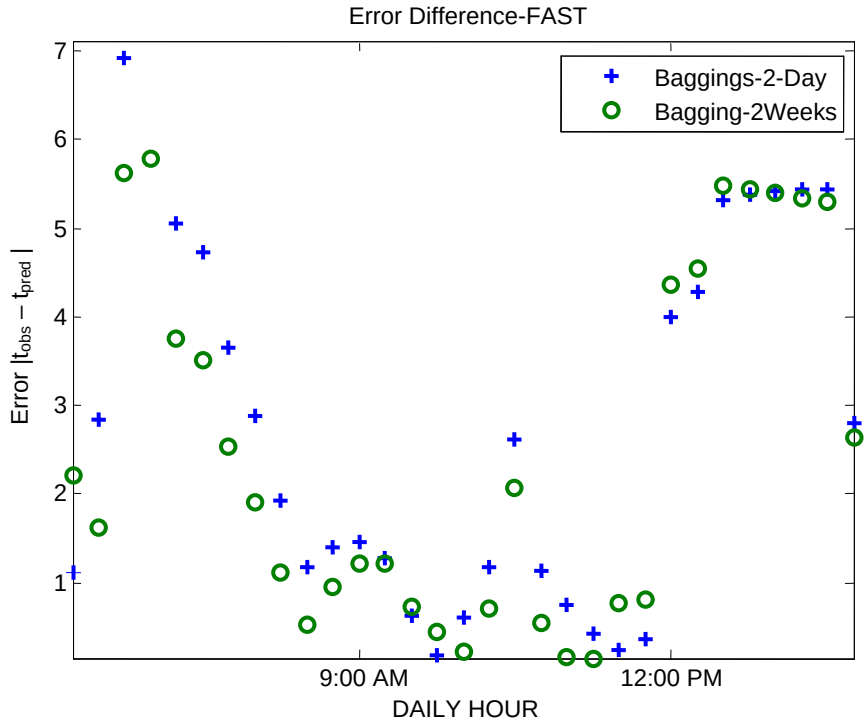


(c) Non-FAST-Loaded Trucks

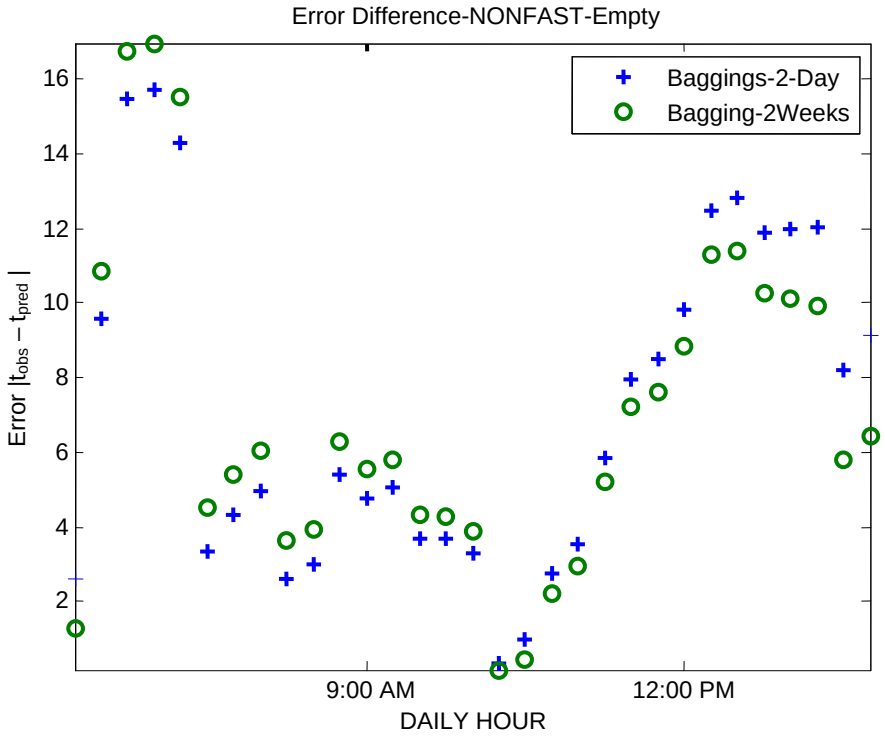
FIGURE 26 Prediction Results using BAGGING Method (Using 2 Weeks Ensemble of Experts)

5.4. Performance of BAGGING Methods (No Observed Data)

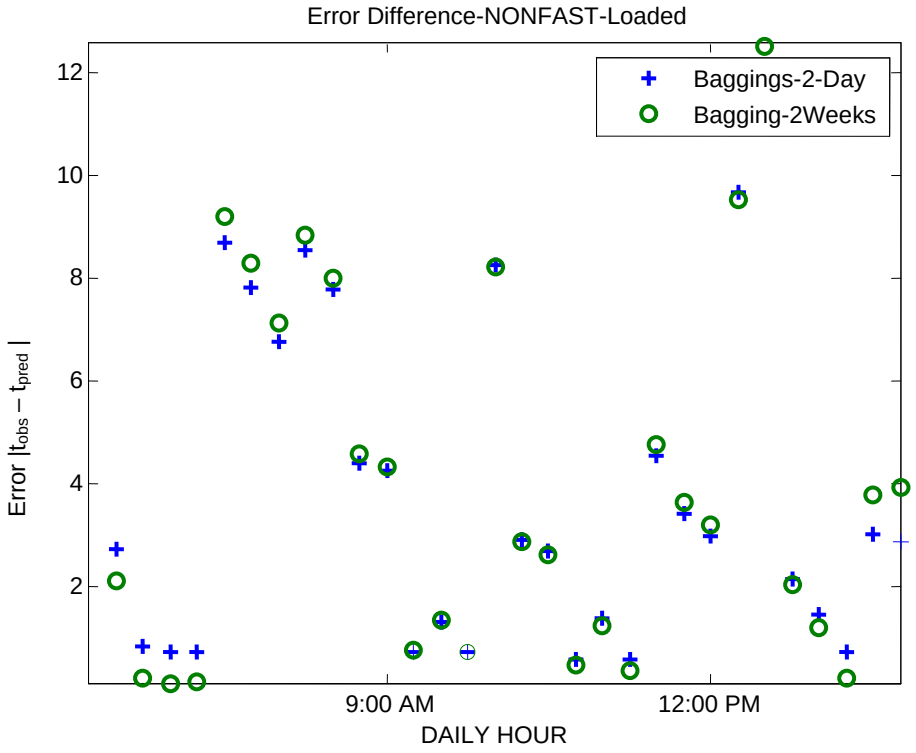
In the previously described BAGGING methods it was assumed that hidden patterns can be obtained using historic data for the same days in the previous two weeks or two previous weeks of full data. FIGURE 27 illustrates prediction error differences between two methods for the same day for truck classes and the errors are around 5 minutes for FAST, 10 minutes for non-FAST-empty trucks, and 15 minutes for non-FAST-loaded trucks.



(a) FAST Trucks



(b) Non-FAST-Empty Trucks



(c) Non-FAST-Loaded Trucks

FIGURE 27 Error Differences between Two Bagging Methods for the Same Day Data

5.5. Observations

In the figure below we can see the relative count of Tag-ID trucks detected during November 2009. The average count is around 80 observations, which is an acceptable subset of the total amount of trucks crossing BOTA. In this plot we can see that on Saturday there were less than 20 observations and under this constraint our prediction is difficult since we will have about 3 or fewer observations per hour in average. In Appendix C, FIGURE C.4, the predicted crossing times do not deviate by much from the actual observations. In all three types of forecast averages we have the following largest errors: 9 minutes for FAST, 16 minutes for non-NFAST-empty and 8 minutes for non-FAST-loaded type trucks. In the mentioned plots, we also show the confidence intervals and those intervals indicate a ± 5 minutes uncertainty. In those plots we can see that most of the actual observed crossing times are nearly forecasted, except for some outliers that minimized the performance of our algorithm.

As we observed in the sample case of 11/21/2009 data and forecasting, even with extremely few observations and the assumption that our RFID tag readers were broken for that specific day, our results show very consistent and reliable forecasts. If we recall the way we compute the average

we can reduce the effect of outliers by changing the parameters used to estimate the average of each truck type.

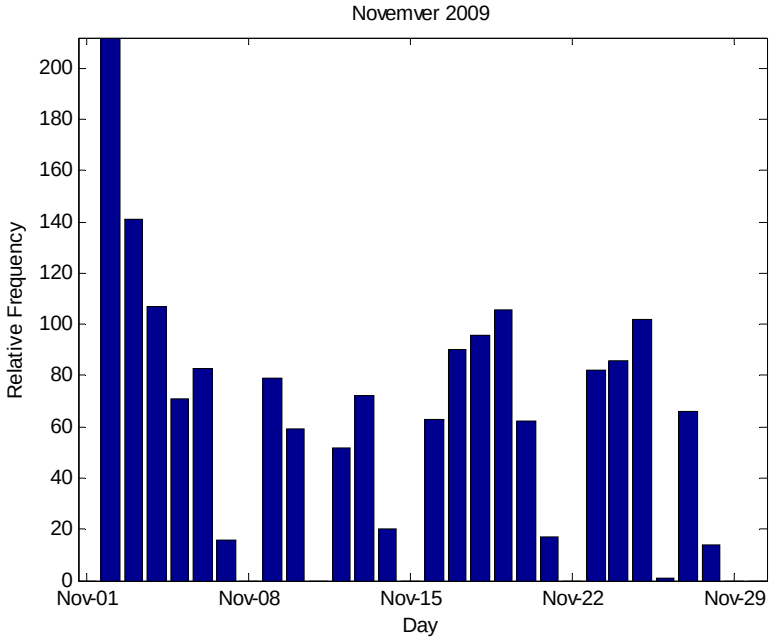


FIGURE 28 Histogram of Tag-ID trucks at BOTA

We already describe that on every Saturday of November we have few observations. In FIGURE 28 we can see how all these observations are distributed and the data are sparse. There are some assumptions we can infer on the observed data; we can tell that very few FAST type trucks used BOTA on Saturdays and most of the samples in this set are likely to be the non-FAST-loaded type, even though the set is small we over-exploit the data to forecast three different types of trucks. Ideally we would like to have a similar amount of observations as those observed on Monday, November 2, 2009, since our forecasting will be reflective near the actual observed behavior of the lines.

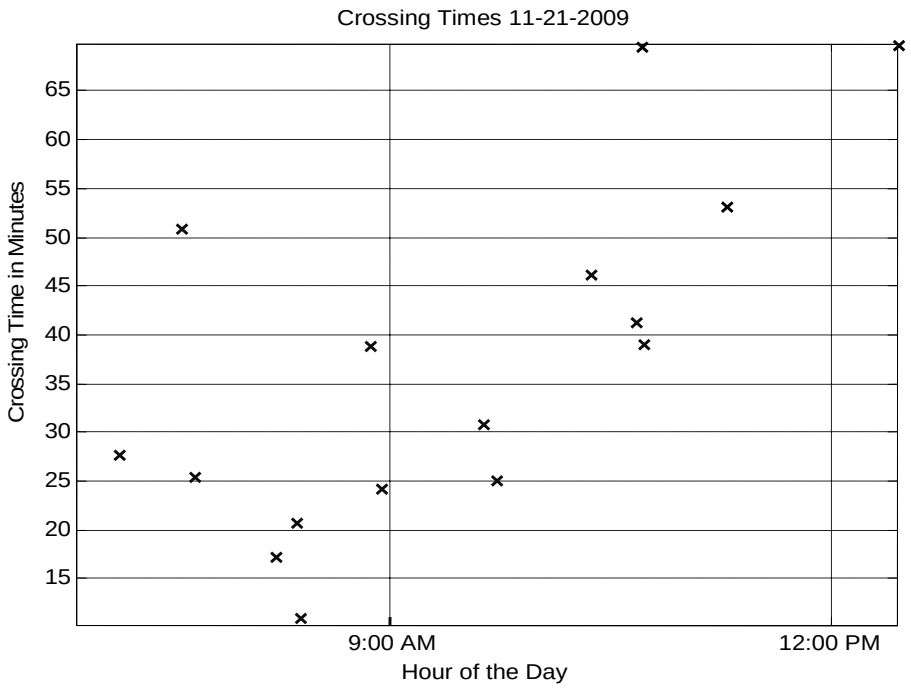


FIGURE 29 Crossing Time Observations on 11/21/2009

5.6. Conclusions

A structured procedure to predict border crossing times of commercial trucks within a short range of time was developed based on statistical models. Currently RFID reader systems that measure northbound crossing times of commercial vehicles are being implemented throughout the POEs along the U.S.-Mexico border. The RFID tags with time stamps at the entry point on the Mexican side and at the exit point of federal and state inspection facilities are collected and stored to measure the crossing times between the two reader stations. Since the current RFID system does not have the capability of identifying detailed information about the trucks such as FAST/non-FAST, and loaded/empty, and the different types of trucks are supposed to have different characteristics of crossing times due to the physical and operational layout of border ports of entry, Gaussian Mixture Model was used to define the set of parameters of each class. Then the Expectation Maximization Algorithm was applied to iteratively estimate the unknown parameters of the FAST, non-FAST-empty, and non-FAST-loaded truck classes.

After the data classification step, the clustered data are processed by the Weighted Moving Average Window to consider the time dependency of crossing times. As described in Chapter 2, average truck crossing times are varied over time. In order to incorporate the dynamic nature of the crossing time variations by time of day, weights are determined in such a way that recent observations have more weights and the weights are also calibrated by the membership functions of the different truck classes.

As a main step of the prediction procedure, regressive functions are determined by Gaussian Process that is a widely used stochastic method for pattern recognition. GP is easy to implement and flexible to change for the given conditions. The results obtained after fitting the regressive

model with the data collected between 11/3/2009 and 11/7/2009 are shown in plots for different classes of trucks in Chapter 4 and Appendix A. In the plots, most of the prediction curves showed very nice fit to the actual RFID observations. The accuracy of the model depended on the size of the data and the kernel.

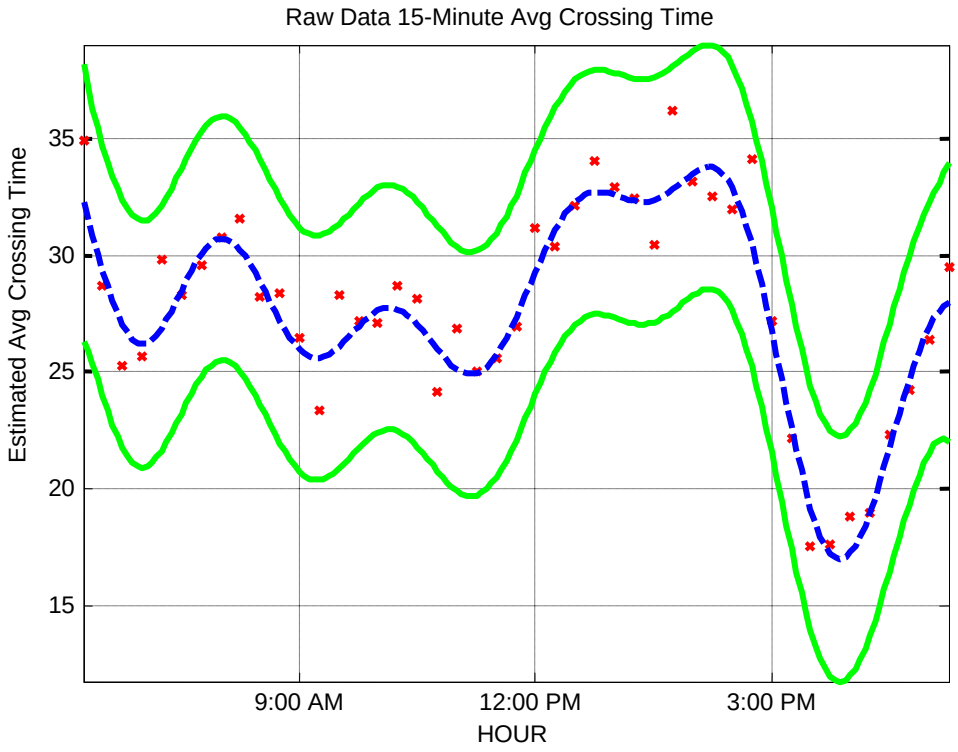
Sometimes unexpected incidents might occur, such as RFID reader failures that hinder the RFID data collection. In order to deal with the unavailability of the current crossing time data to be used in the short-term prediction model, the Bootstrap Aggregating (BAGGING) technique was used to exploit the fitted models from the historical data and improve the predictions by using the combination of the existing fitted models. Actual prediction results with the assumption of no current observation data were presented in the previous sections of this Chapter. Based on the historical data collected during the first two weeks of November 2009, two different approaches of the BAGGING technique were illustrated. First, the ensemble of the models from the same day of the week was used. Table 1 shows the results by several performance measures and FIGURE 25 shows the plots. In the second approach, crossing times of Monday, November 16, 2009, were predicted based on the ensemble of models from the whole two weeks of observations. Both of the approaches showed similar results, summarized in FIGURE 27. Both of the methods showed the errors around 5 minutes for FAST trucks, 10 minutes for non-FAST-empty trucks, and 15 minutes for non-FAST-loaded trucks.

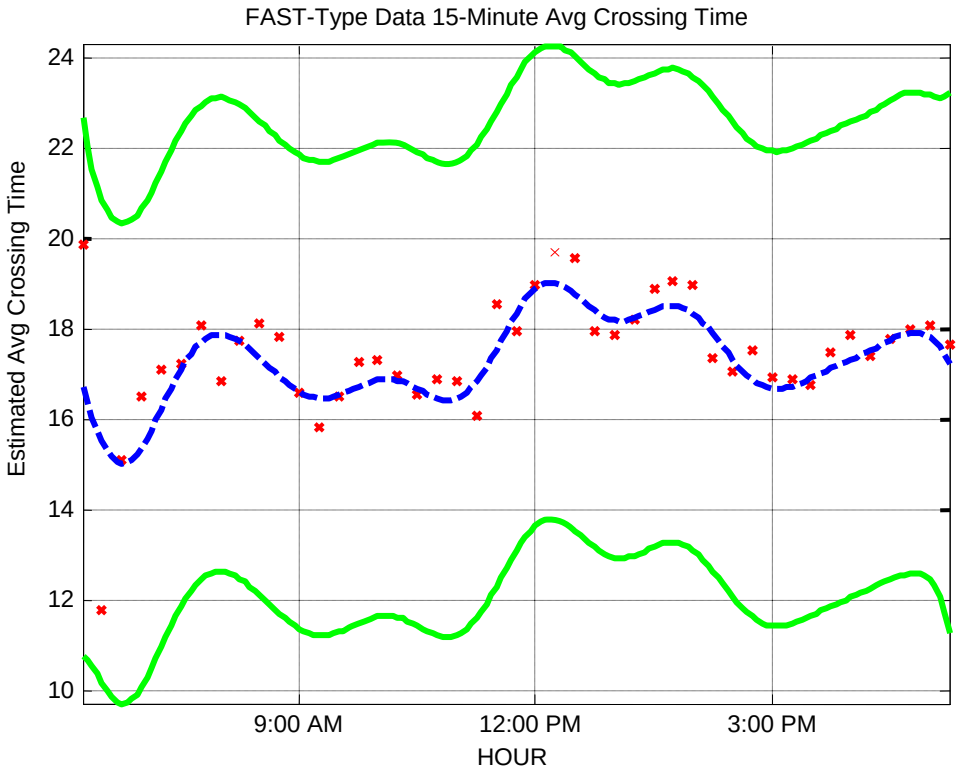
5.7. Future Work

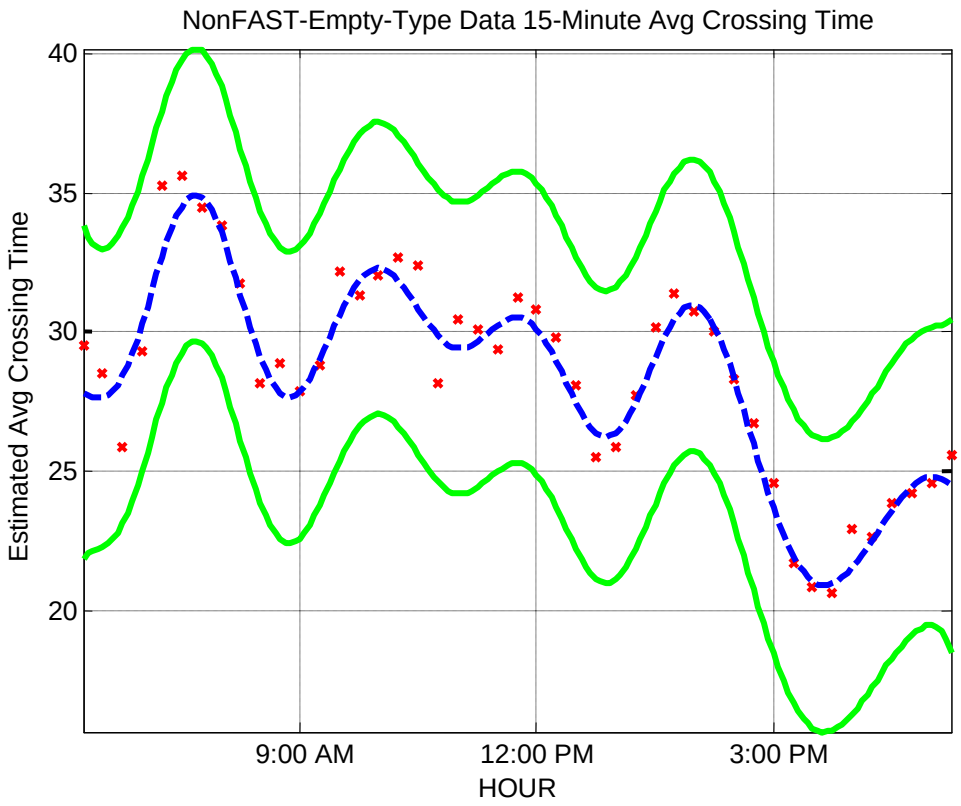
In the future, the main objective of the Texas Transportation Institute is to improve the reliability of our system. A third sensor will be placed at BOTA, and this sensor will improve the performance of the reading system and also the forecasting algorithm. The sensor is planned to be placed between the current sensors, perhaps near the inspection booth. This sensor will have the same operational capabilities of the other sensors. First, the sensor will read a Tag from a truck carrying it and it will collect the time of reading. Second, the sensor will submit the read time to the collection database. Basically the operations are reading, collecting, and transmitting a packet of time stamps with RFID.

The benefits obtained from the placement of the third sensor are various and we want to exploit them as much as possible. Our forecasting algorithm operates under the difference of two observation times, the Entry time and the Exit time. After placement of the third sensor we will have to adapt our algorithm to a new observation. Our algorithm can be easily scaled to this new sensor.

Appendix A: Plots of Regression Models of Gaussian Process







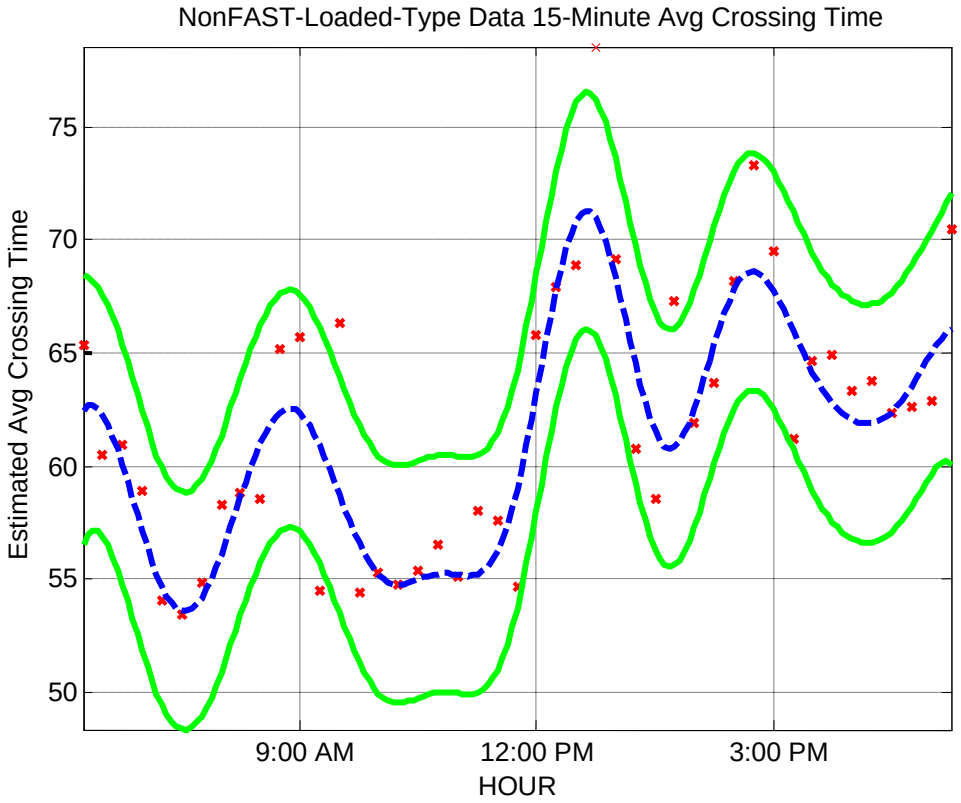
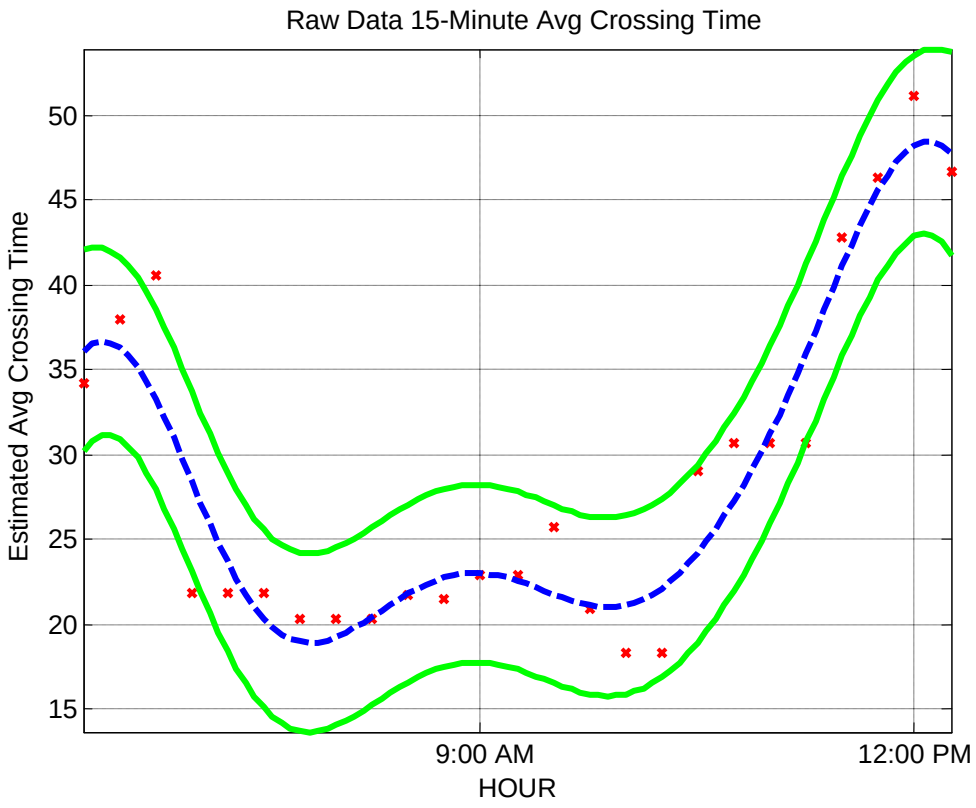
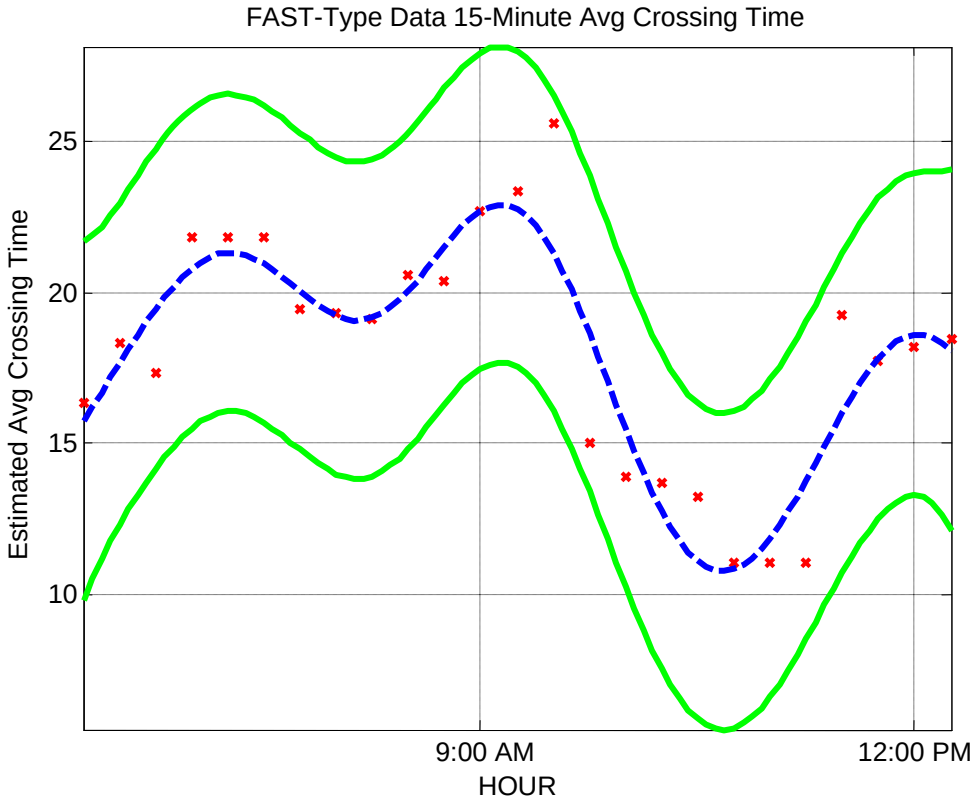
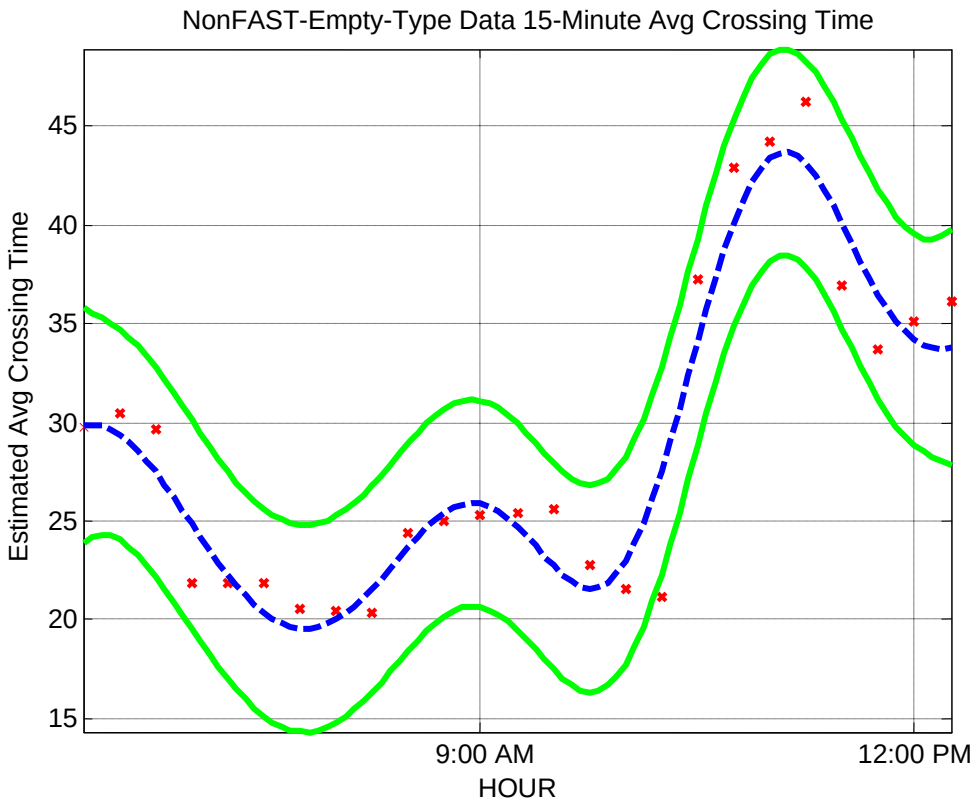


FIGURE A-1 Regression Models of Tuesday, 11-3-2009


Raw Data 15-Minute Avg Crossing Time
Estimated Avg Crossing Time
50
45
40
35
30
25
20
15
9:00 AM
12:00 PM
HOUR


FAST-Type Data 15-Minute Avg Crossing Time
Estimated Avg Crossing Time
25
20
15
10
9:00 AM
12:00 PM
HOUR



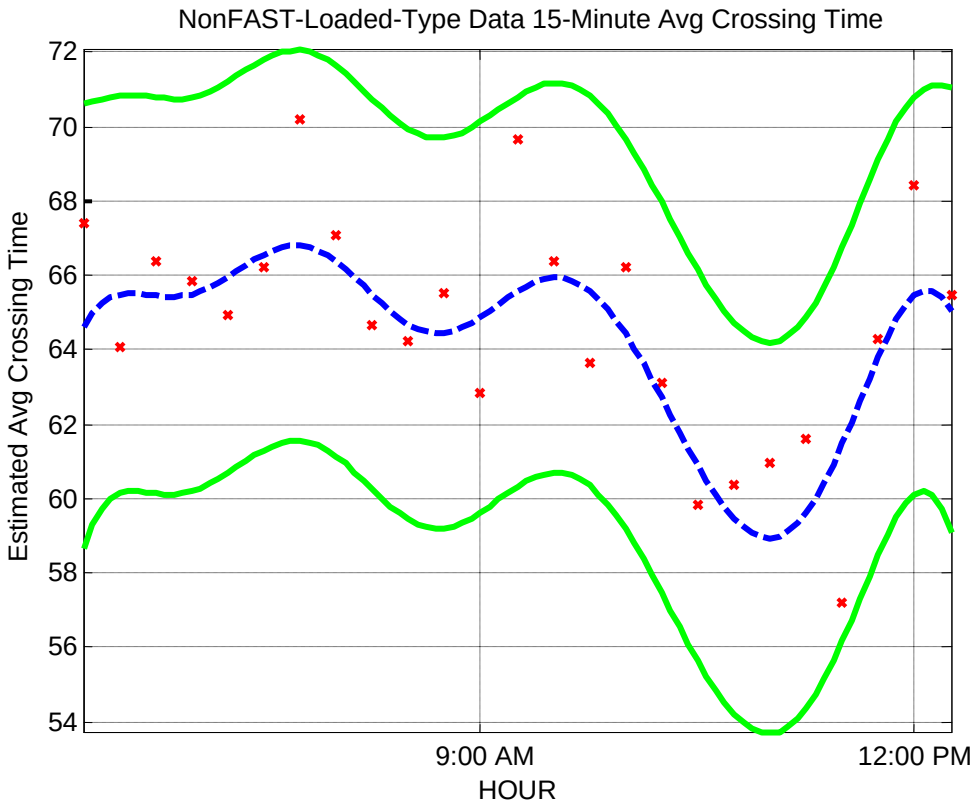
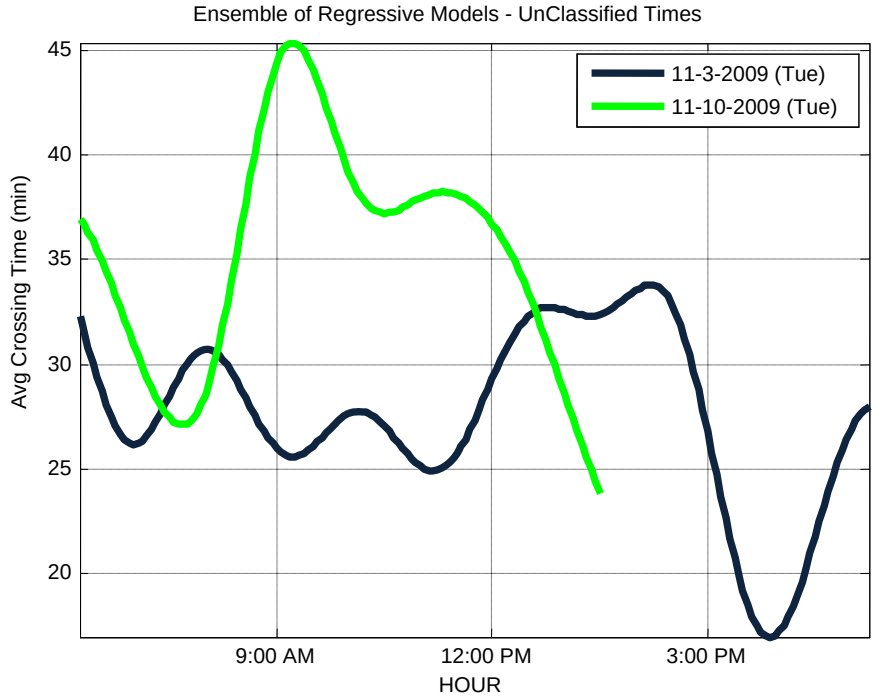
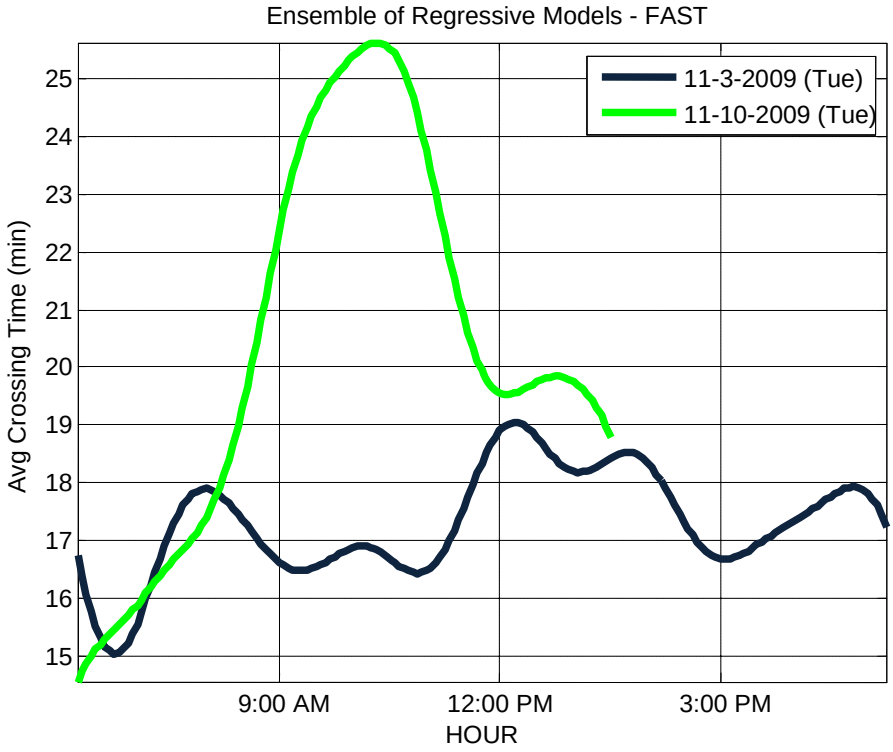


FIGURE A-2 Regression Models of Saturday, 11-7-2009

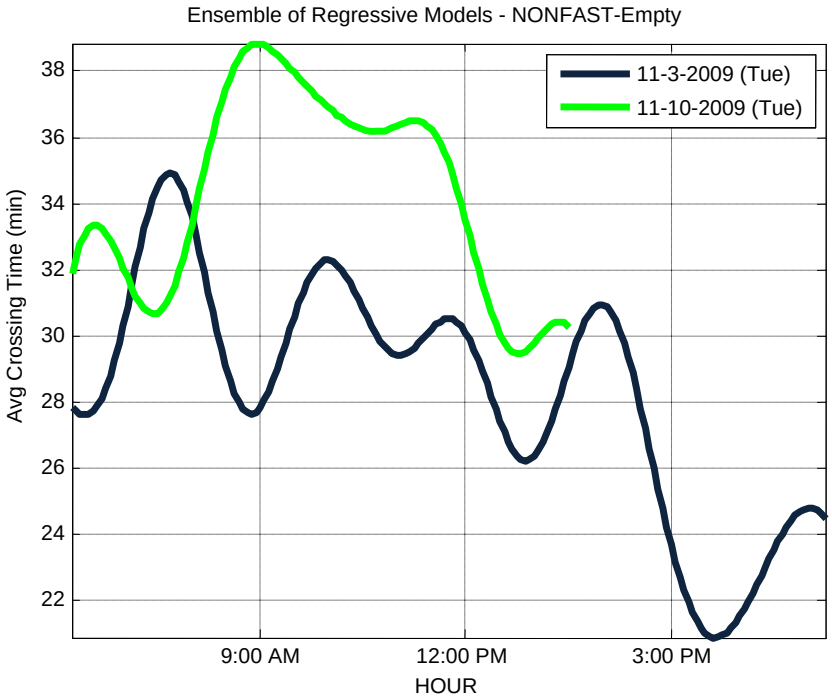
Appendix B: Plots of Ensemble of Regression Models



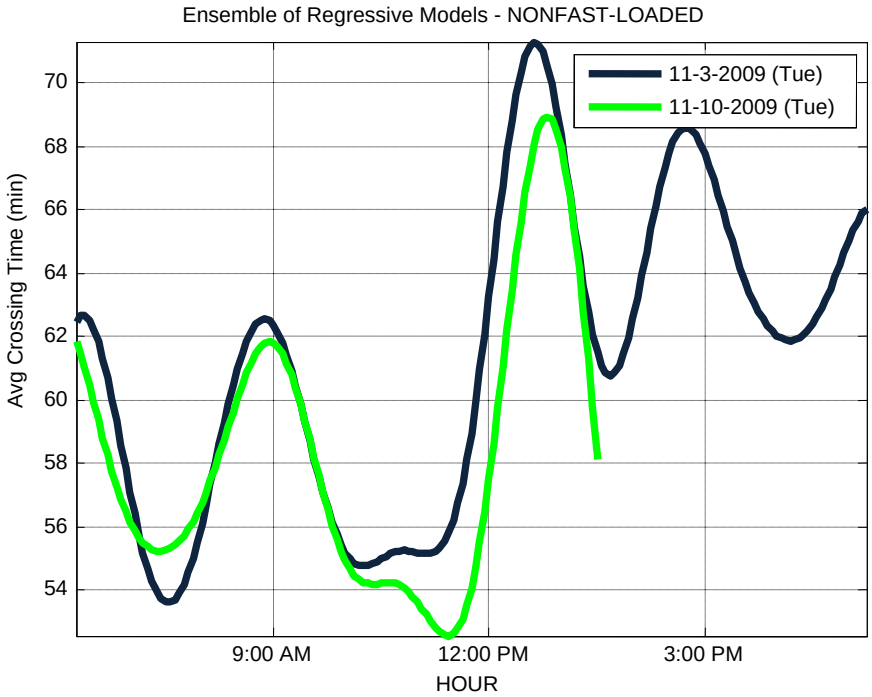
(a) Unclassified



(b) FAST

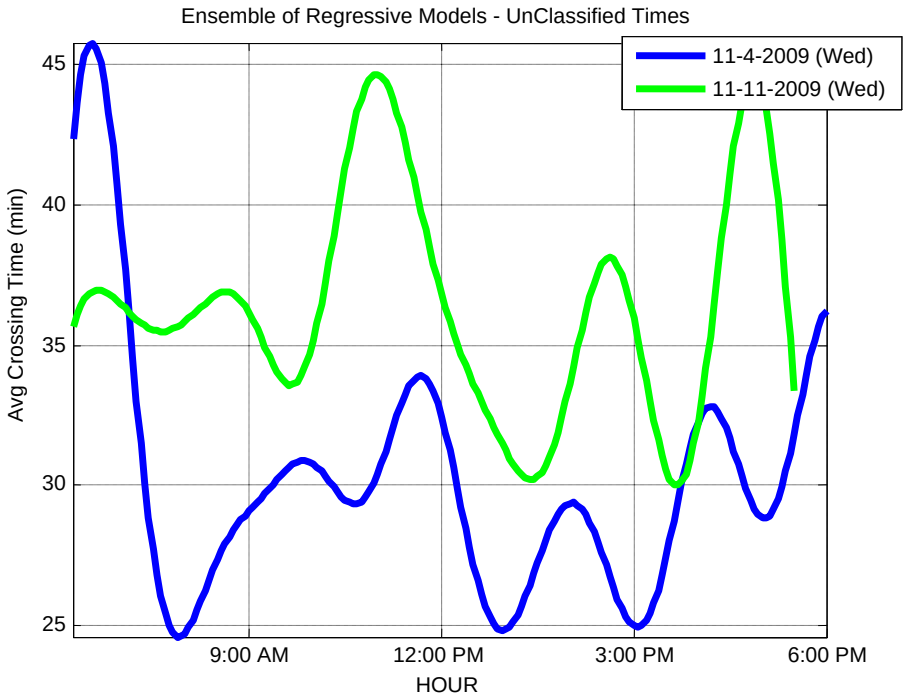


(c) Non-FAST-Empty

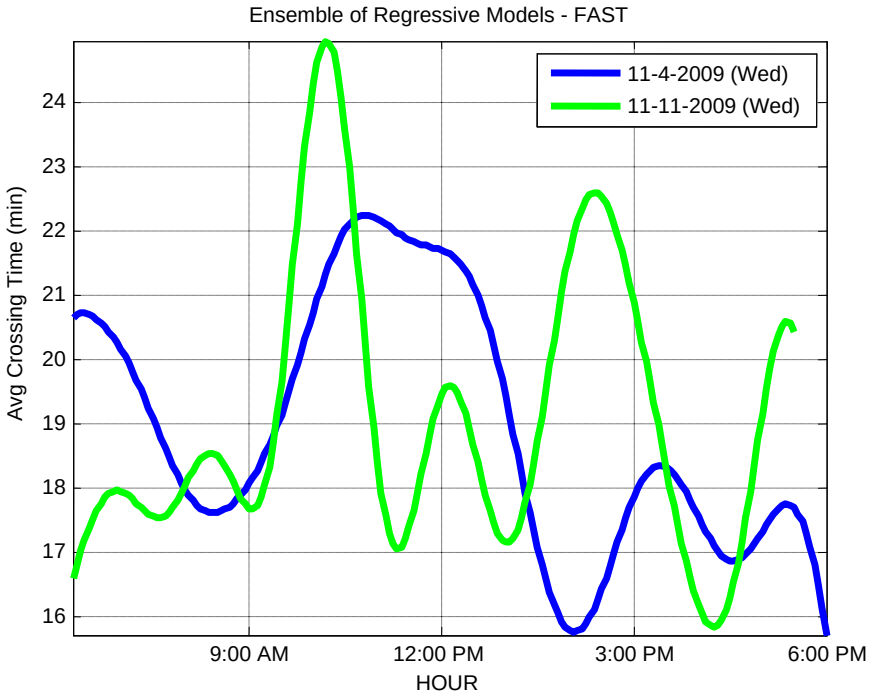


(d) Non-FAST-Loaded

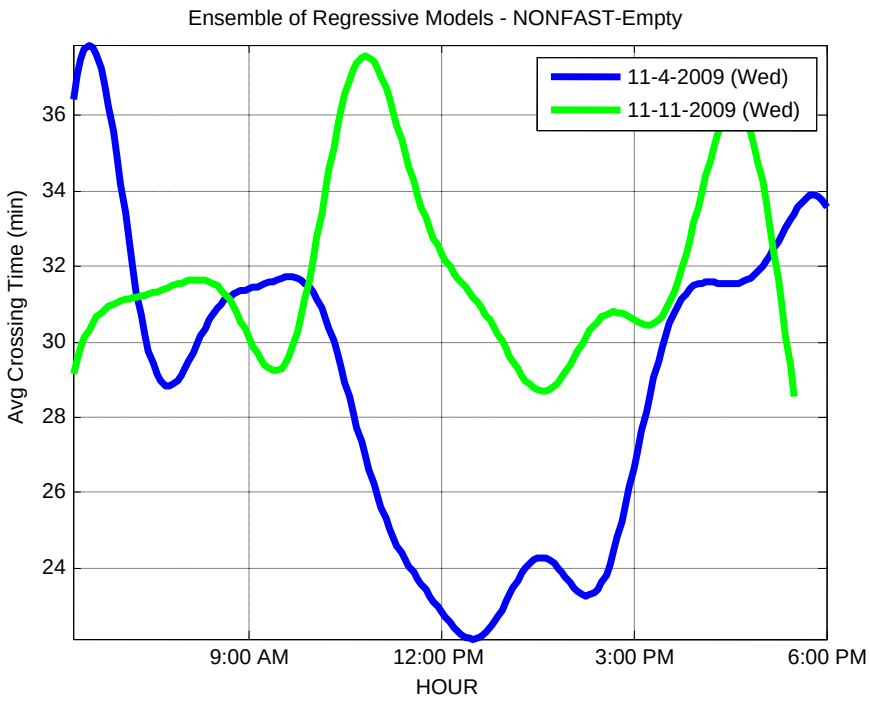
FIGURE B-1 Ensemble Models (Tuesday) 11/3/2009 and 11/10/2009 (a) Unclassified, (b) FAST, (c) Non-FAST-Empty, and (d) Non-FAST-Loaded



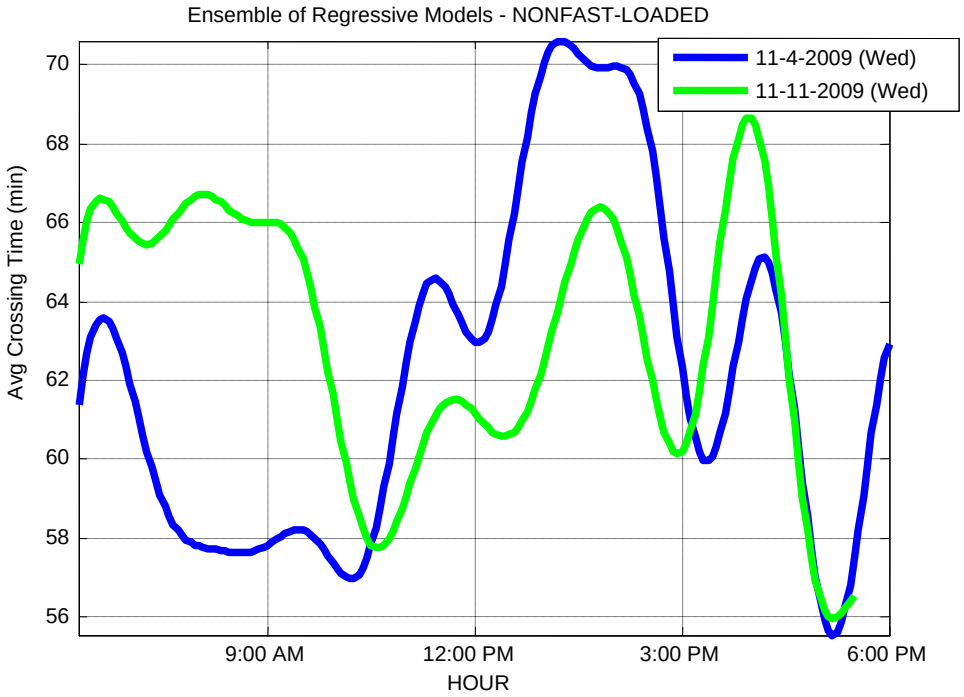
(a) Unclassified



(b) FAST

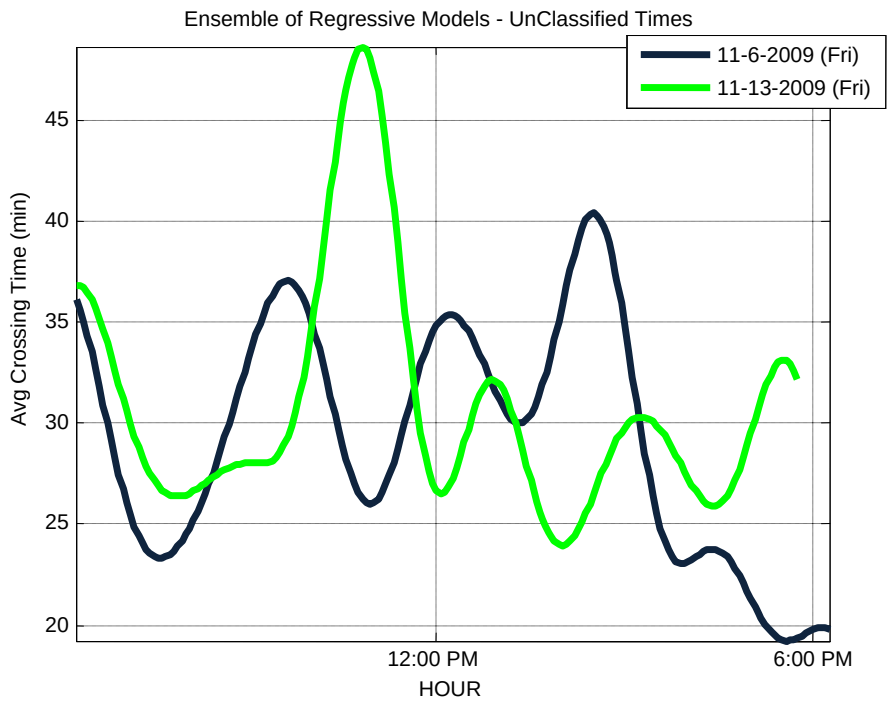


(c) Non-FAST-Empty

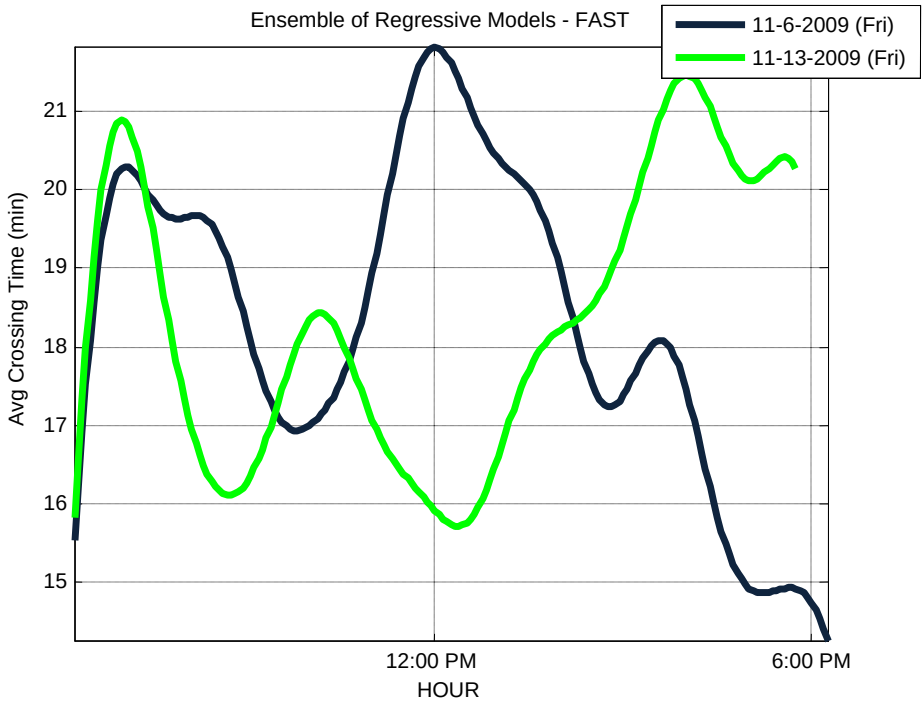


(d) Non-FAST-Loaded

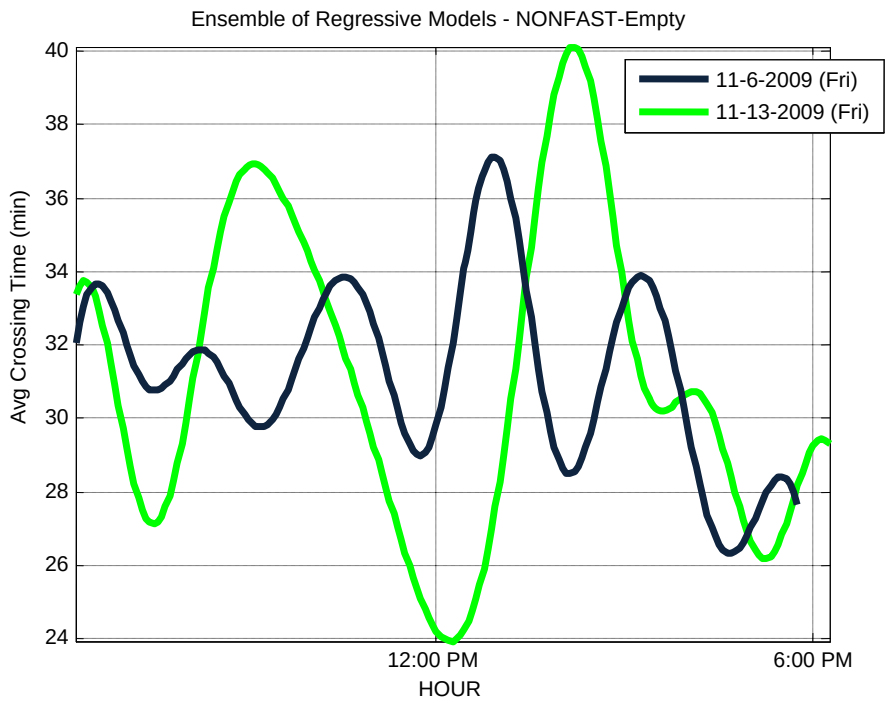
FIGURE B-2 Ensemble Models (Wednesday) 11/4/2009 and 11/11/2009 (a) Unclassified, (b) FAST, (c) Non-FAST-Empty, and (d) Non-FAST-Loaded



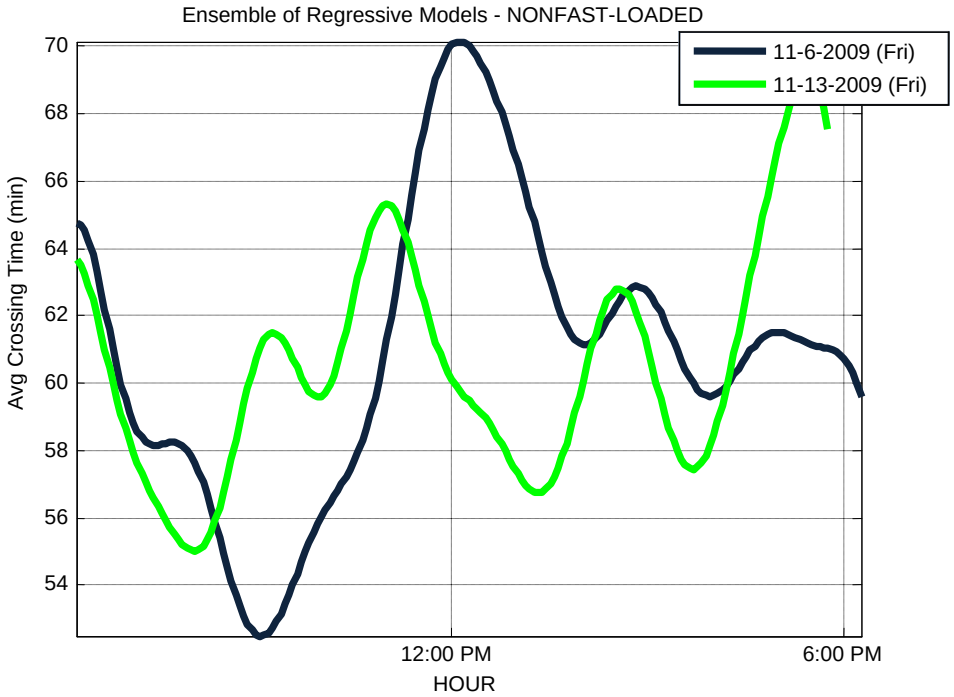
(a) Unclassified



(b) FAST

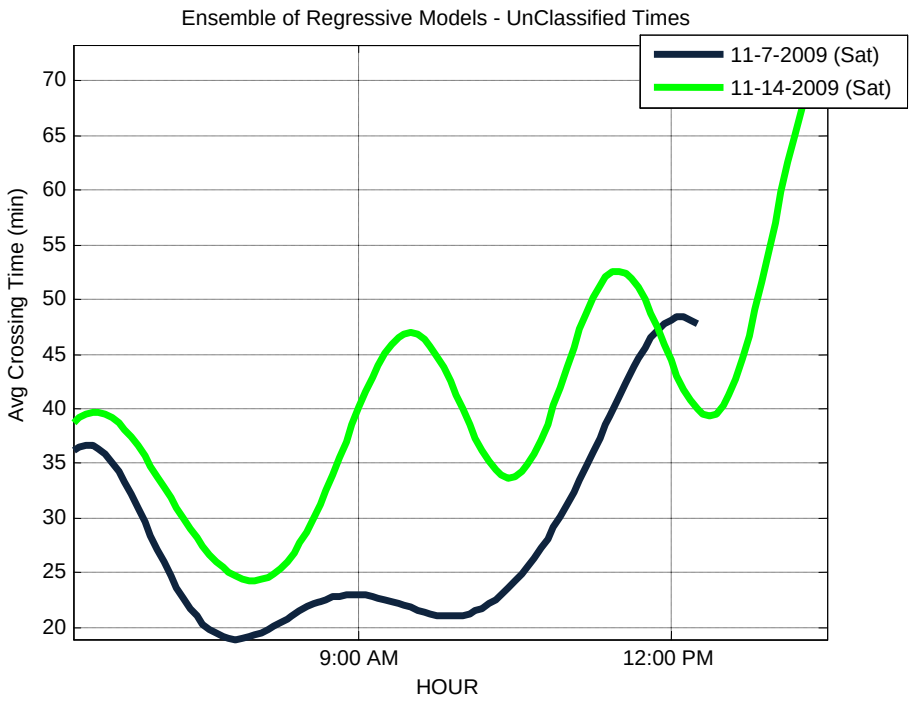


(c) Non-FAST-Empty

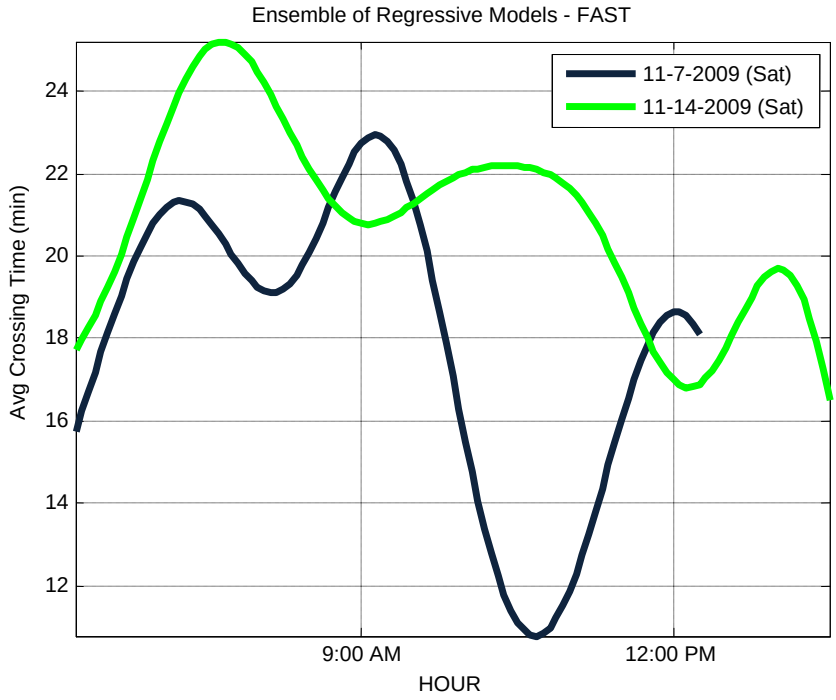


(d) Non-FAST-Loaded

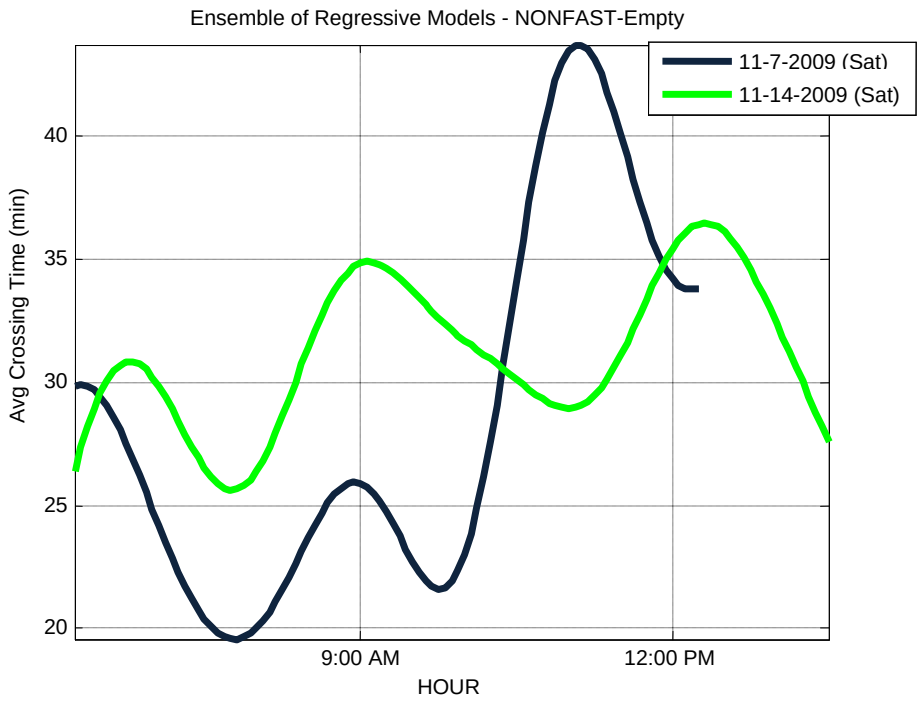
FIGURE B-3 Ensemble Models (Friday) 11/6/2009 and 11/13/2009 (a) Unclassified, (b) FAST, (c) Non-FAST-Empty, and (d) Non-FAST-Loaded



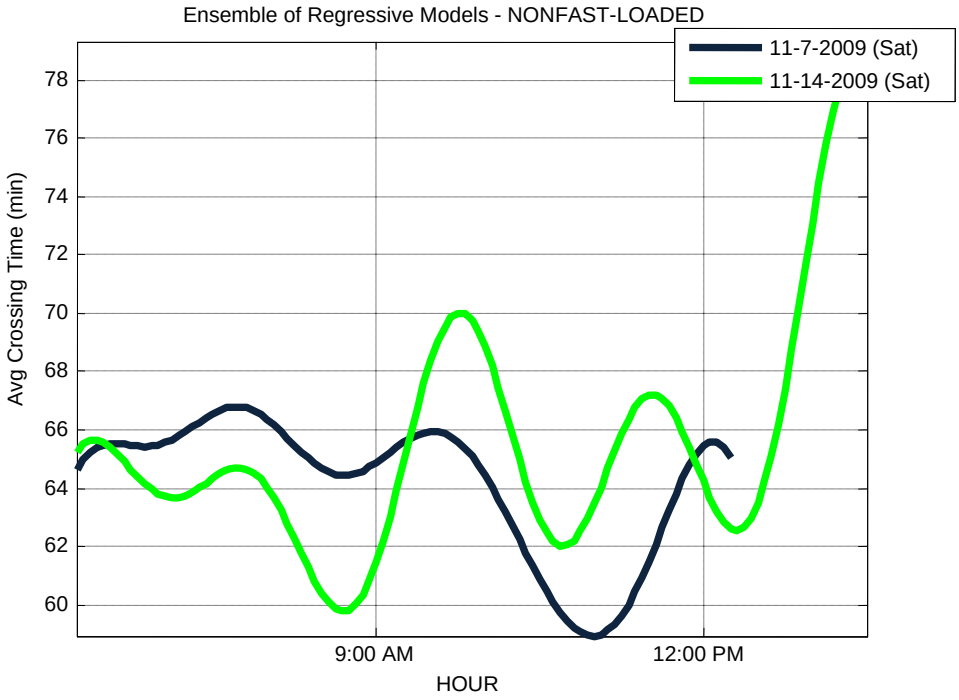
(a) Unclassified



(b) FAST



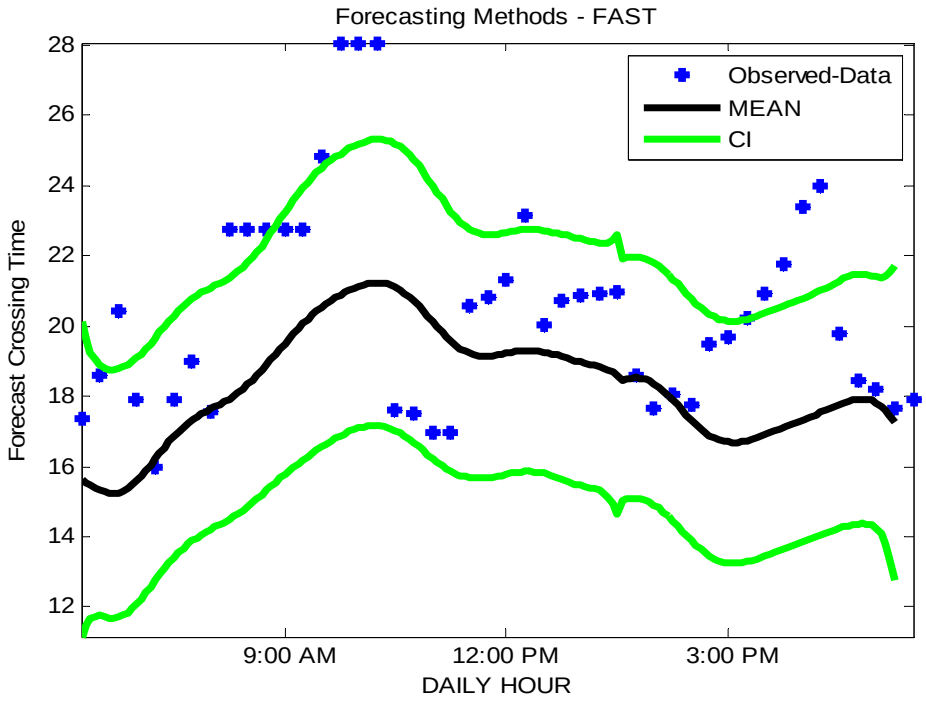
(c) Non-FAST-Empty



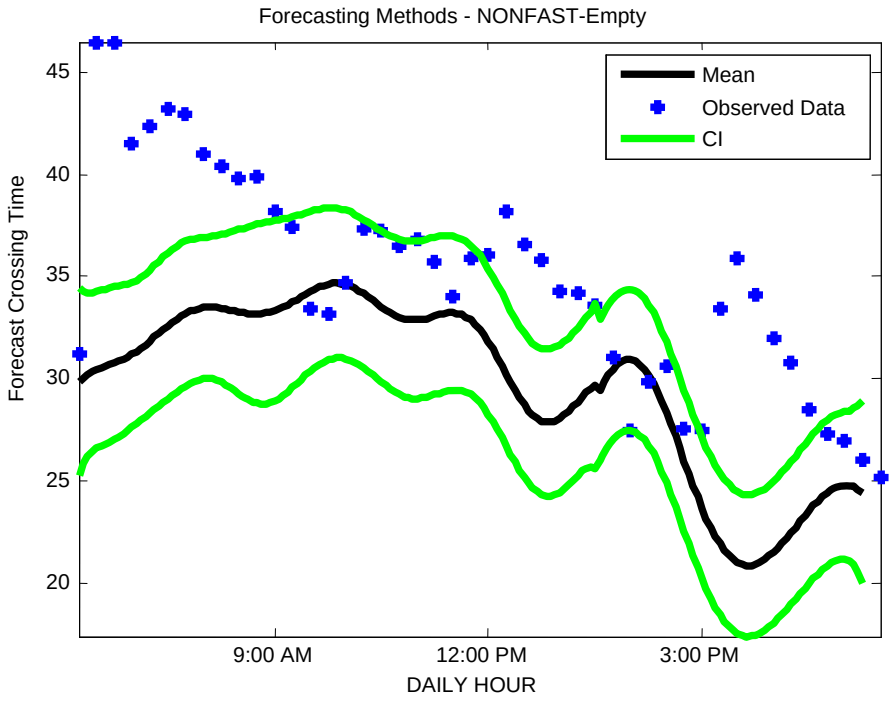
(d) Non-FAST-Loaded

FIGURE B-4 Ensemble Models (Saturday) 11/7/2009 and 11/14/2009 (a) Unclassified, (b) FAST, (c) Non-FAST-Empty, and (d) Non-FAST-Loaded

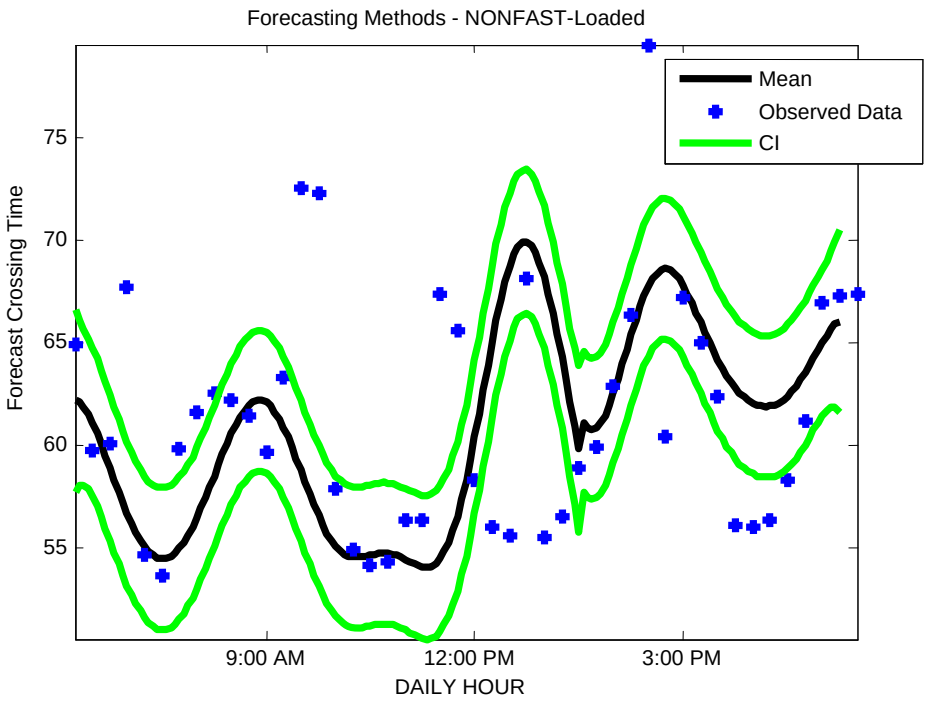
Appendix C: Prediction Results



(a) FAST

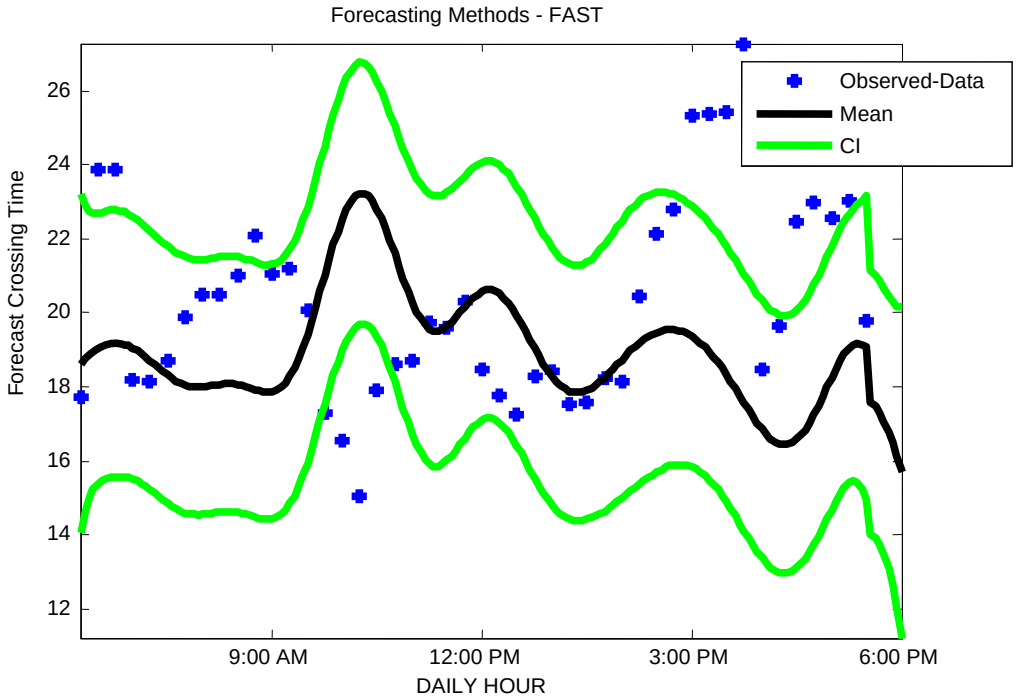


(b) Non-FAST-Empty

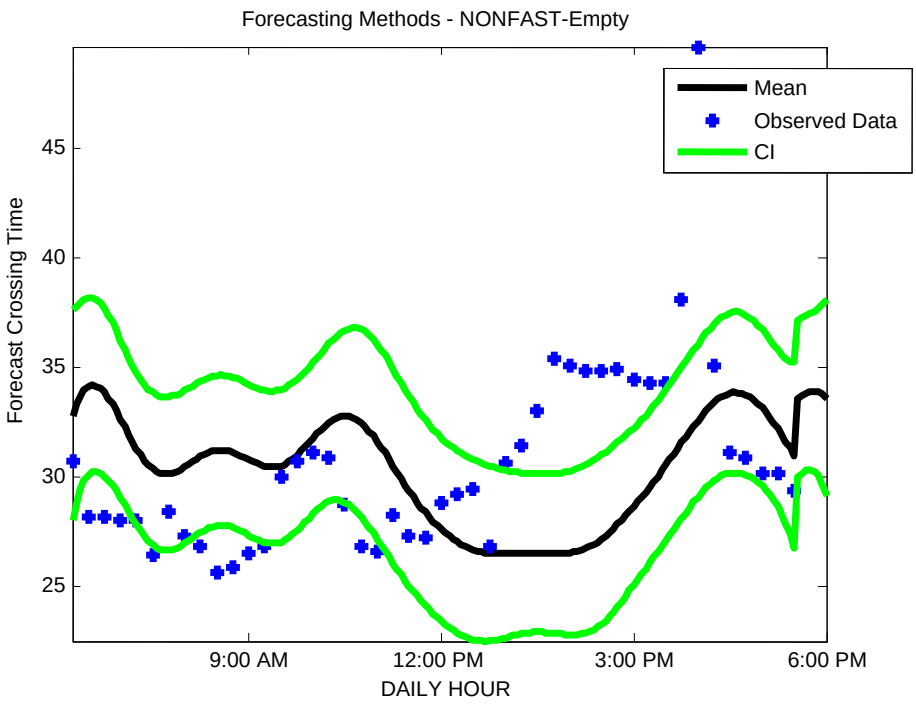


(c) Non-FAST-Loaded

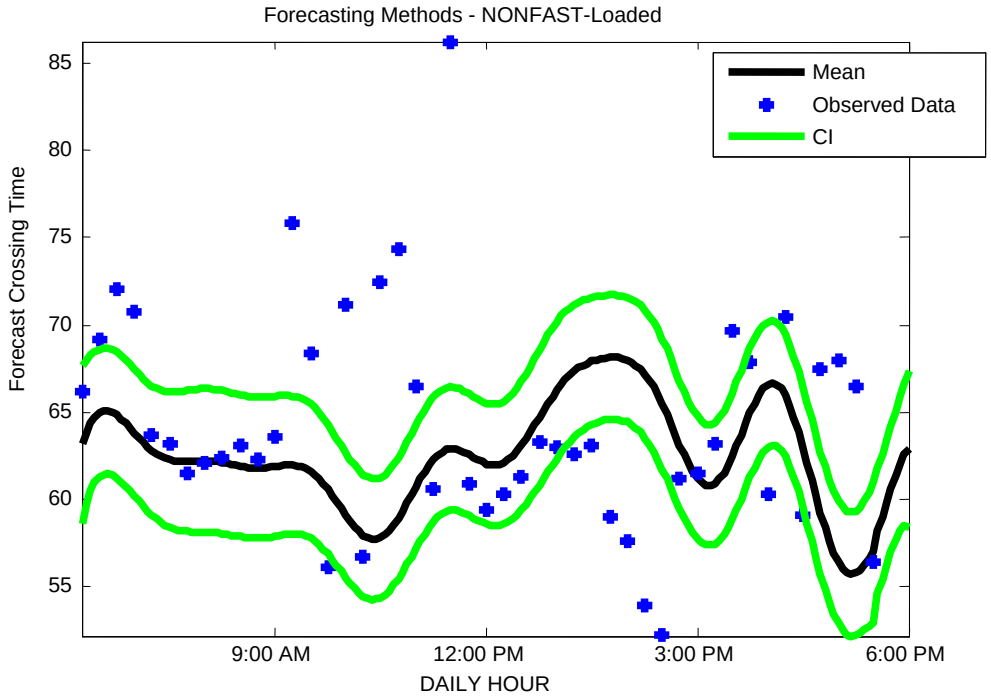
FIGURE C-1 Prediction Results (Tuesday, 11/17/2009) (a) FAST, (b) Non-FAST-Empty, and (c) Non-FAST-Loaded



(a) FAST

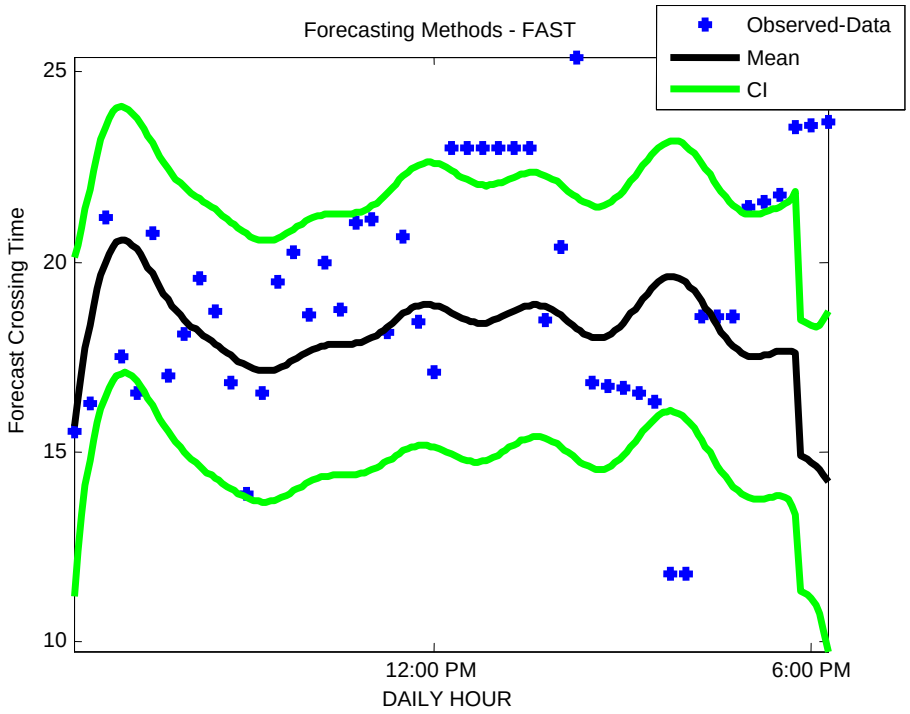


(b) Non-FAST-Empty

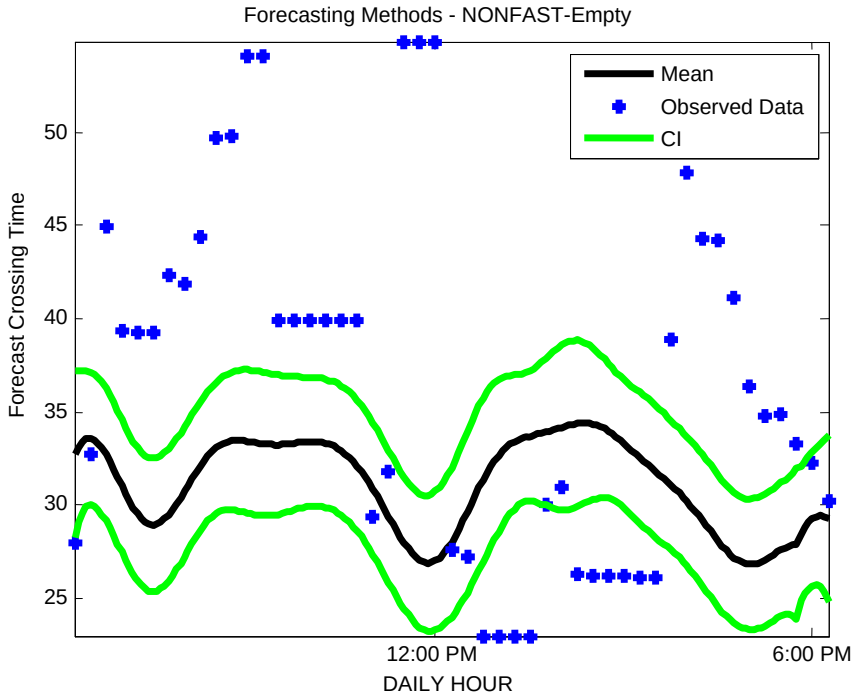


(c) Non-FAST-Loaded

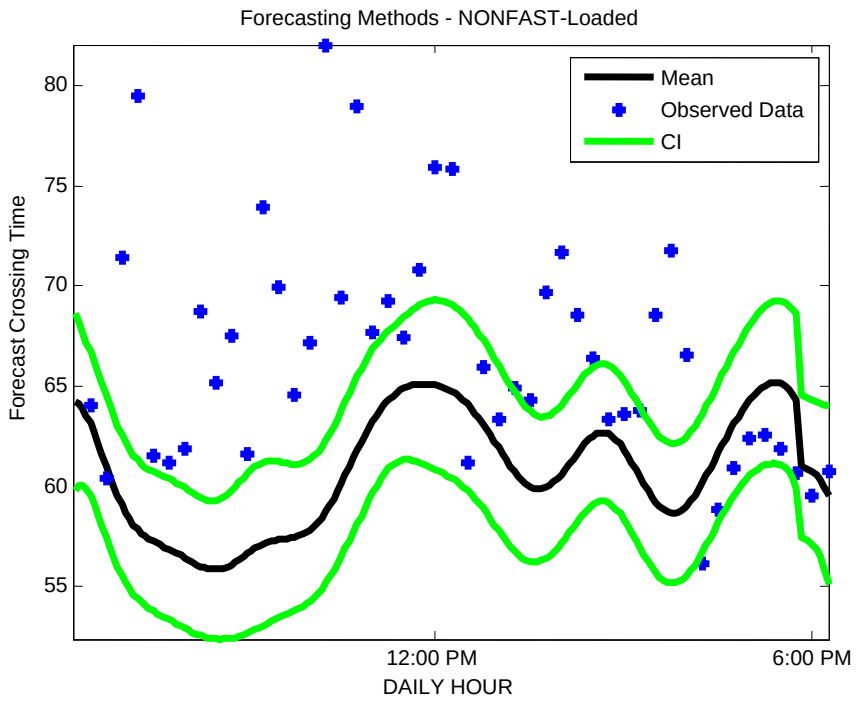
FIGURE C-2 Prediction Results (Wednesday, 11/18/2009) (a) FAST, (b) Non-FAST-Empty, and (c) Non-FAST-Loaded



(a) FAST

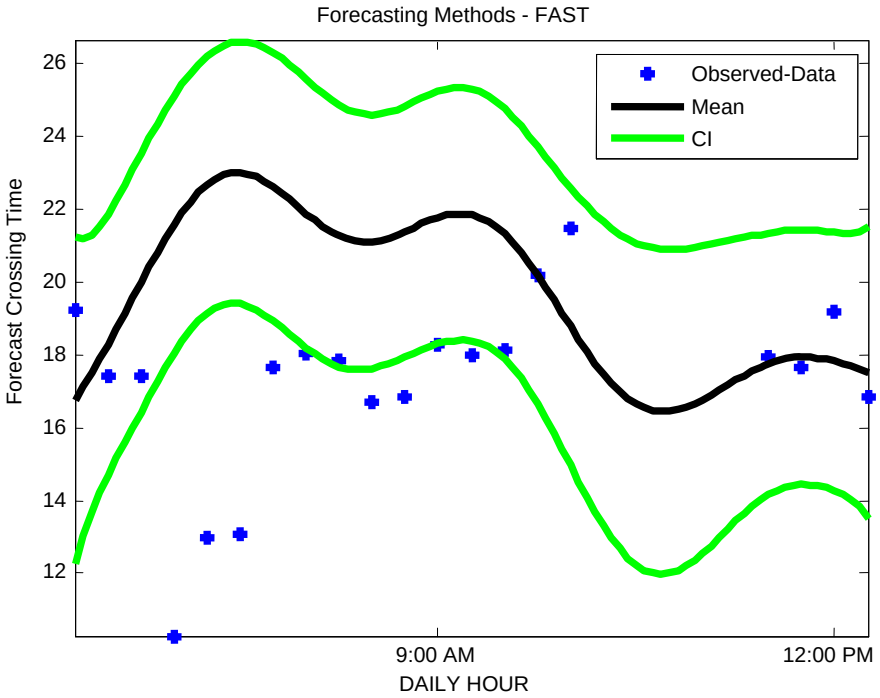


(b) Non-FAST-Empty

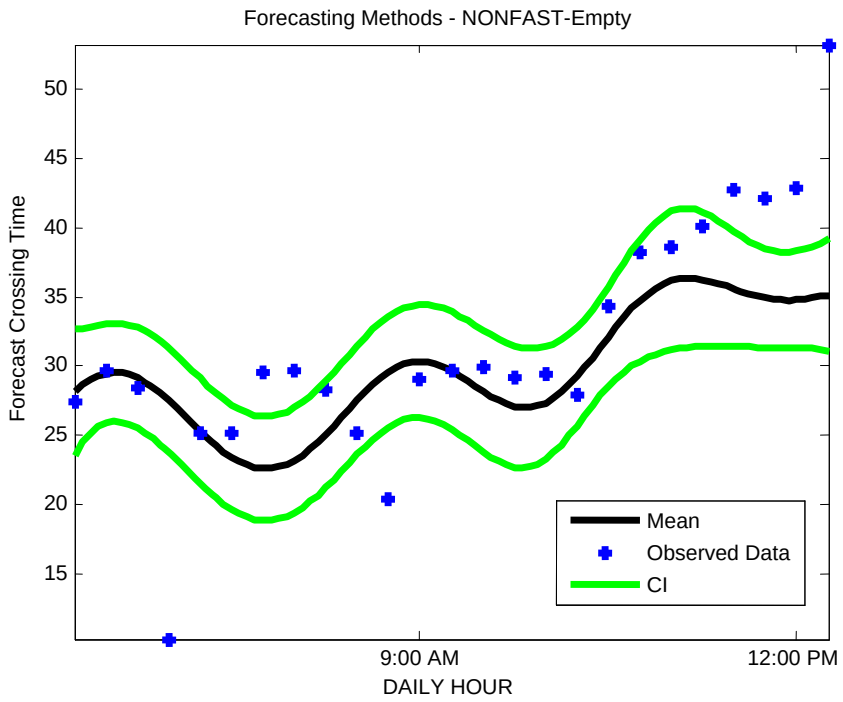


(c) Non-FAST-Loaded

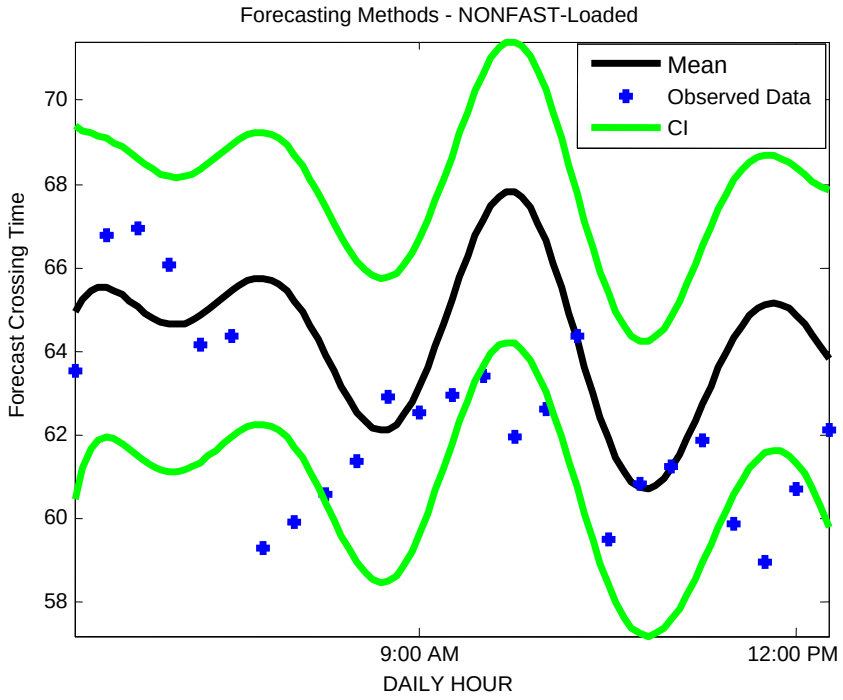
FIGURE C-3 Prediction Results (Friday, 11/20/2009) (a) FAST, (b) Non-FAST-Empty, and (c) Non-FAST-Loaded



(a) FAST



(b) Non-FAST-Empty



(c) Non-FAST-Loaded

FIGURE C-4 Prediction Results (Saturday, 11/21/2009) (a) FAST, (b) Non-FAST-Empty, and (c) Non-FAST-Loaded

References

1. **Rajbhandari, Rajat, Villa, Juan Carlos and Aldrete, Rafael.** *Expansion of the Border Crossing Information System.* s.l. : University Transportation Center for Mobility, Texas A&M University, 2009.
2. **Texas Transportation Institute.** *Measuring Border Delay and Crossing Times at the U.S.-Mexico Border Tasks 1 and 2 Report.* Washington D.C. : Federal Highway Administration, 2007.
3. **Sabean, J., Jones, C.** *Inventory of Current Programs for Measuring Wait Times at Land Border Crossings: Report for Customs and Border Protection, Department of Homeland Security and Canada Border Services Agency.* 2008.
4. *Service Time Variability at Blaine, Washington, Border Crossing and Impact on Regional Supply Chains.* **Goodchild, A., Globerman, S. and Albrecht, S.** Washington D.C. : Transportation Research Record: Journal of the Transportation Research Board, 2008, Vols. Number 2066, pp 71-78.
5. **Li, R.** *Examining Travel Time Variability Using AVI Data.* s.l. : Institute of Transport Studies, December, 2004.
6. *Measuring Variability in Traffic Conditions by Using Archived Traffic Data.* **Turochy, R. E. and Smith, B. L.** Washington D.C. : Transportation Research Record: Journal of the Transportation Research Board, 2002, Vols. Number 1804, pp 168-172.
7. *Vehicle Re-Identification as Method for Deriving Travel Time and Travel Time Distribution.* **Sun, C., Arr, G. and Ramachandran, R. P.** Washington D.C. : Transportation Research Record: Journal of the Transportation Research Board, 2003, Vols. Number 1826, pp 25-31.
8. *The Valuation of Reliability of Personal Travel.* **Bates, J., et al.** Washington D.C. : Transportation Research Part E, 2001, Vols. Volume 37, pp 191-229.
9. *Travel Time Variability: A Review of Theoretical and Empirical Issues.* **Noland, R. B. and Polak, J. W.** s.l. : Transportation Reviews, 2002, Vols. Volume 122, pp 39-54.
10. *Statistical and Neural Classifiers to Detect Traffic Operational Problems on Urban Arterials.* **Khan, S I and Ritchie, S G.** 1998, Transportation Research - Part C, pp. pp. 291-314.
11. *An Urban Traffic Flow Model Integrating Neural Networks.* **Ledoux, C.** 1997, Transportation Research Part C, pp. pp. 287-300.
12. *Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data.* **Dia, H and Rose, G.** 1997, Transportation Research - Part C, pp. pp. 313-331.
13. *Sequential Forecast of Incident Duration Using Artificial Neural Network Models.* **Wei, C H and Lee, Y.** 2007, Accident Analysis and Prevention, pp. pp. 944-954.
14. *An Improved Adaptive Exponential Smoothing Model for Short-Term Travel Time Forecasting of Urban Arterial Street.* **Zhi-Peng, L, Hong, Y and Fu-Qiang, L.** 2008, Acta Automatica Sinica, pp. Vol 34 (11), pp. 1404-1409.
15. *A Knowledge Based Real-Time Travel Time Prediction System for Urban Network.* **Lee, W, Tseng, S and Tsai, S.** 2009, Expert Systems with Applications, pp. pp. 4239-4247.
16. *Arterial Travel Time Estimation Based on Vehicle Re-Identification Using Wireless Magnetic Sensors.* **Kwong, K, et al.** 2009, Transportation Research Part C, pp. pp. 586-606.
17. *Travel Time Prediction with Support Vector Regression.* **Wu, Chun-Sin, Ho, Jan Ming and Lee, D.T.** 4, s.l. : IEEE Transactions on Intelligent Transportation Systems, 2004, Vol. 5.
18. *Travel Time Prediction using Gaussian Process Regression: A Trajectory-Based Approach.* **Tsuyoshi, I. and Kato, S.** s.l. : Proceedings of the Ninth SIAM International Congerence on Data Mining, 2009.

19. **Texas Transportation Institute.** *Implementing a Roadway Freight Performance Measure at the Border.* 2008.
20. *A Recursive Cell Processing Model for Predicting Freeway Travel Times.* **Paterson, D and Rose, G.** 2008, Transportation Research Part C, pp. pp. 432-453.
21. Border Crossing Data. *Bureau of Transportation Statistics.* [Online] U.S. Department of Transportation. http://www.bts.gov/.
22. *Travel Time Estimation Based on Piecewise Truncated Quadratic Speed Trajectory.* **Sun, L, Yang, J and Mahmassani, H.** 2008, Transportation Research Part A, pp. pp. 173-186.
23. *Neural Network Models for Classification and Forecasting of Freeway Traffic Flow Stability.* **Florio, L and Mussone, L.** 1996, Control Engineering Practice, pp. pp. 153-164.

University Transportation Center for Mobility™

Texas Transportation Institute

The Texas A&M University System

College Station, TX 77843-3135

Tel: 979.845.2538 Fax: 979.845.9761

utcm.tamu.edu