

Appendix A – Literature Summary

- 1 Currency movements and international border crossings (2000)
- 2 Unified Nogales/ Santa Cruz County Transportation 2000 Plan (2000)
- 3 Estimating Texas-Mexico North American Free Trade Agreement Truck Volumes (2001)
- 4 Specification of a borderplex econometric forecasting model (2001)
- 5 Cross Border Cargo Vehicle Flows (2002)
- 6 Assessment of Automated Data Collection Technologies for calculation of commercial motor vehicle border crossing travel time delay (2002)
- 7 El Paso Customs District Cross-Border Trade Flows (2003)
- 8 Borderplex Bridge and Air Econometric Forecast Accuracy (2004)
- 9 Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study: Strategic & Geographic Area Overview (2004)
- 10 Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study: Existing and Future Travel Demand (2004)
- 11 Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study : Travel Demand Analysis Process (2004)
- 12 Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility : Study Partnership of Transportation Problems and Opportunities Report (2004)
- 13 Traffic Forecast Based on Real Data (2004)
- 14 An error correction analysis of U.S.-Mexico trade flows (2005)
- 15 Analyzing highway flow patterns using cluster analysis (2005)
- 16 Tradeoffs between security and Inspection Capacity: Policy Options for Land Border Ports of Entry (2006)
- 17 AZ Multimodal Freight TM1: Analysis of Freight Dependent Industries (2007)
- 18 AZ Multimodal Freight TM2: Assessment of Arizona's Existing Freight Infrastructure (2007)
- 19 AZ Multimodal Freight TM3: Strategic Directions for Freight Planning (2007)
- 20 Use of Box and Jenkins Time Series Technique in Traffic Volume Forecasting (2007)
- 21 Nogales Railroad Small Area Transportation Study (2007)
- 22 Bottleneck Study of Mariposa POE (2008)
- 23 Mariposa/I-19 Connector Route Study (2008)
- 24 Socioeconomic determinants of Mexican Circular and permanent Migration

1. Currency movements and international border crossings (2000)

This report analyzed the responses of international commuter flows to exchange rate valuation shifts. This study was conducted around one of the main border crossing points between the US and Mexico which was the El Paso, Texas and Ciudad Juarez, Mexico border. Rather than focusing on the movements of physical goods, this report focused on the movement of commuters as a result of the US dollar/Mexican Peso exchange rate. This was particularly significant to our study as we found that the exchange rate was one of the external factors that considerably impacted the movement of truck crossings through the Nogales POE.

2. Unified Nogales/ Santa Cruz County Transportation 2000 Plan (2000)

This plan was provided to the city of Nogales by Kimley-Horn and Associates, Inc. Due to the growth of the Nogales/Santa Cruz County region there arose a need for changes within the current transportation infrastructure so that it could efficiently handle the growing amounts of traffic in the area. The primary purpose of this report was to provide a five-year (short term), ten-year (mid-term), and twenty-year (long term) transportation plan that addresses the needs of the area. This report includes a summary of the existing conditions of Nogales/Santa Cruz County as well as its transportation needs.

3. Estimating Texas-Mexico North American Free Trade Agreement Truck Volumes (2001)

This study addressed the issue of accommodating NAFTA truck traffic along international highway trade corridors and border points of entry linking the US and Mexico. There were two data sources available for estimating the number trucks crossing the border, one being the counts of trucks crossing the bridges and the second being the US international trade data. Based on each of these two data sets the study developed two methods of estimation.

The first method which used cross-border volumes was theoretically weakest because it required a number of assumptions to be made. The second method used densities and volumes in explaining why results show that the average value per truckload per port varies significantly. In conclusion, the methods developed can be used successfully to estimate commodity truck volumes; however, a large effort was still required to analyze, match and use current trade statistics and calibrate results. Further analysis of trailer loads in relation to commodity densities and volumes that were crossing border ports would enable more accurate standardized ETT(Equivalent Trade Truck) volumes to be determined. This data was important because it could possibly be incorporated into federal and state planning actions that address maintenance, rehabilitation, and reconstruction of international trade corridors in the United States.

4. Specification of a borderplex econometric forecasting model (2001)

This article used the El Paso-Ciudad Juarez borderplex as the framework for assessing the whether or not the traditional approach used for econometric modeling could also be applicable to border metropolitan areas. This question was addressed by first estimating an econometric model and testing it under various currency conditions as well as comparing the results of the model simulation to extrapolations from a Bayesian vector autoregressive model. The results showed that the traditional approach was in fact applicable for analysis of business trends in international border regions.

Although the traditional method was sufficient to analyze international border regions there were still several areas in which it could be improved. More specifically, inclusion of data regarding maquiladoras and their impact on the El Paso-Ciudad Juarez borderplex would be helpful due to their prominence in the Ciudad Juarez region. This study also included a discussion of a slight alteration in the traditional model to account for the fact that economic conditions in Ciudad Juarez frequently diverge from the conditions in the rest of the Mexican macro economy.

5. Cross Border Cargo Vehicle Flows (2002)

This article was written by Thomas M. Fullerton Jr. and Roberto Tinajero and was funded by the NAFTA Intermodal Institute at the Public Policy Research Center at the University of Texas at El Paso. With the growth of the maquiladora sector in Northern Mexico there had been an increase in merchandise trade and thus the volume of cargo across international borders had also increased. The study conducted by Fullerton and Tinajero attempted to determine if the trends caused by these changes could be modeled through analysis of the short-term time series characteristics of the cargo vehicle flows through El Paso, Texas. In this report the econometric methodologies involved both univariate and transfer function ARIMA analysis.

After analyzing the data the factors which were found to impact the monthly fluctuations in border region cargo vehicle flow traffic were as follows: the maquiladora employment in Ciudad Juarez, the non agricultural employment in El Paso, inflation adjusted exchange rate and both Mexican and U. S. industrial activity.

6. Assessment of Automated Data Collection Technologies for calculation of commercial motor vehicle border crossing travel time delay (2002)

The data in this report was prepared by Batelle for the Office of Freight Management and Operations of the Federal Highway Administration of the U.S. Department of Transportation. The purpose of this study was to test various technologies that could potentially be used as a means to collect truck travel time data instead of the onsite data collectors that were at the border crossings of U.S. to its neighbors. Specifically the report highlighted which characteristics did and did not make them suited to collect this data as well as their maturity for deployment. Furthermore this study also investigated technologies

that were still in their development stages. The results of this analysis, however, did not provide detailed specifications of each of the technologies and were meant to be used simply as guidelines.

The findings of this study did not return any clear winner as the best type of technology to be used due to the fact that there were so many variables that needed to be considered. When choosing which to use the authors of the report suggested taking into account the following: the technology's functionality, cost, purpose of operations, maturity and availability as well several secondary issues such as susceptibility to vandalism.

The study defined what criteria that would make a selected technology a viable candidate to use to collect truck travel times. In the appendix of this study, each of the tested technologies was listed and summarized with its advantages and disadvantages of using each.

7. El Paso Customs District Cross-Border Trade Flows (2003)

With an emphasis on the linkage between El Paso and Ciudad Juarez this paper analyzed the monthly trade flows through the El Paso Customs District. The study differentiated itself from previous studies in that earlier studies focused on US demand for aggregate imports and exports while this study focused on bilateral trade flows with Mexico. It also provided an analysis of the empirical regularities associated with monthly exports and imports of merchandise trade in the El Paso Customs District area. Also within this report a time series methodology was used to potentially quantify cross-border trade flows through El Paso.

The results of the analysis showed that borderplex international trade did respond fairly rapidly to economic activity changes as well as changes in cross-border relative prices. It was concluded that a better understanding of border trade and regional economic performance could be gained through additional empirical analysis.

8. Borderplex Bridge and Air Econometric Forecast Accuracy (2004)

This study focused on the borderplex area surrounding El Paso, Texas. At the time of the study there were two sets of transportation equations in the borderplex model (Fullerton 2001). One equation was used for the northbound traffic at the El Paso/Ciudad Juarez international bridge and the other was used for the passenger, cargo and mail flows at the El Paso international airport. This study was focused on the borderplex transportation variable forecasts that were published during the years 1998 to 2003. Each of these studies was measured based on its predictive accuracy relative to a random walk benchmark. At the time this study was conducted there had not previously been any formal accuracy assessments on the transportation variables included in the border region system of simultaneous equations. Also included in the report was information on regional econometric forecasting research, borderplex model attributes, as well as suggestions for future research topics.

The results showed that air freight, airmail, bridge auto, and bridge pedestrian series forecasts were somewhat more accurate than random walk benchmarks over the course of the sample period in question. For air passenger and bridge cargo random walk benchmarks tended to work better.

9. **Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study: Strategic & Geographic Area Overview (2004)**

This report was prepared by the IBI group for URS Canada and was part of a series of several technical working papers which served to devise a long term strategy to provide safe and efficient travel of people and goods to and from Southeast Michigan and Southwest Ontario. This report was the first in the series seeks to evaluate the current conditions of the transportation routes used to connect Southeast Michigan and Southwest Ontario and thus determine any areas of need. More specifically this is done through a study of the geographic and demographic characteristics of each of these regions, the infrastructure of the border crossings and traffic flows at the border crossings.

10. **Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study: Existing and Future Travel Demand (2004)**

This report was prepared by the IBI group for URS Canada and was part of a series of several technical working papers which served to devise a long term strategy to provide safe and efficient travel of people and goods to and from Southeast Michigan and Southwest Ontario. This particular report described, analyzed and evaluated the current and future rail and road conditions with the ultimate goal of providing both qualitative and quantitative perception of the travel demands in this region. The modes of transportation covered in this report including passenger cars, commercial vehicles, rail and marine. Although all types of transportation were covered there was a particular emphasis on the vehicle traffic on the three bridges and tunnel crossings connecting Southwest Ontario and Southeast Michigan due to the extensive delays that occur in these areas.

11. **Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study : Travel Demand Analysis Process (2004)**

This report was prepared by the IBI group for URS Canada and was part of a series of several technical working papers which served to devise a long term strategy to provide safe and efficient travel of people and goods to and from Southeast Michigan and Southwest Ontario. This report provided an analysis of estimated future demand at both the existing border crossings between these two regions as well as potential new crossings. These potential new crossings were also analyzed based on their impact on the existing transportation systems (i.e. any diversions they might cause).

In particular this paper included:

- A review of available data and modeling and forecasting techniques

- A description of how to develop a travel demand analysis process covering the various means of travel (highways, roads, railroads) as well as a model to forecast regional travel demand
- A description of the Validation of the aforementioned model to conditions in the base year (2000)

12. Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility : Study Partnership of Transportation Problems and Opportunities Report (2004)

This report was prepared by the IBI group for URS Canada and was part of a series of several technical working papers which served to devise a long term strategy to provide safe and efficient travel of people and goods to and from Southeast Michigan and Southwest Ontario. This particular report served to give a general overview of the type of transportation issues that would be addressed in this planning study, the extent of these issues and lastly the characteristics of the area with regards to transportation and socioeconomic conditions.

Upon completion of this analysis three major transportation problems were defined that needed to be addressed in this study. The first problem was lack of alternative routes for border crossings in the case of operational failure, heavy congestion, major incidents or any other type of disruption that could block or slow down one of the crossing routes. The second problem defined was that Windsor-Detroit roadway crossings do not have sufficient roadway capacity to efficiently handle the forecasted future travel demand. The third problem found was also at the Windsor-Detroit border crossing and is caused by lack of border crossing capacity at this crossing to handle both existing and future travel demands.

There were also several opportunities for improvement found at in this region that were addressed in this study, they are listed as follows:

- Development of a multi-modal strategy for a balanced transportation system that provides more transportation choices
- Protection of future required right-of-way
- Optimization of existing infrastructure
- Facility rehabilitation to avoid or delay replacement
- Partnerships with other proponents to co-operatively address common problems and/or shared objectives
- Revenue generation and/or cost reduction
- Support for provincial, state and national economic and planning objectives

13. Traffic Forecast Based on Real Data (2004)

Within this report different methods to forecast traffic were analyzed and discussed. The analysis was based on the traffic at a highway in Duisburg, Germany. Data was collected over a 2 year period and then separated into four basic classes. Two models for short term forecasting were examined, the

constant and the linear model. It was concluded that for short term horizons and constant model provided a good prediction and for more long term forecasts it was better to use a heuristic.

14. An error correction analysis of U.S.-Mexico trade flows (2005)

This report was written by Thomas M. Fullerton Jr. and Richard L. Sprinkle to estimate the bilateral trade elasticities between the U.S. and Mexico. Most previous studies had focused on analyzing aggregate trade flows. One of the key goals of this particular study was to gain an understanding of the response of bilateral trade to changes in relative prices and changes in income. More specifically there was an emphasis on determining if the bilateral trade flows react to domestic prices, foreign prices or the exchange rate between Mexican Pesos and U.S. dollars. Also, an error connection approach was used, taking into account both short-run dynamic adjustments and long-term determinants of trade flows, to estimate import and export demand functions. The findings of this study validated the fact that error correction model could be used to analyze trade flows and that in the case of U.S. and Mexico these flows responded heterogeneously to prices, income and their exchange rate.

15. Analyzing highway flow patterns using cluster analysis (2005)

This paper investigated the determination of historical traffic by patterns by means of Ward's hierarchical clustering pattern. Using this procedure days were classified into working and non-working days and more specifically the working days were classified by the day of the week and whether they were in a vacation period. After analyzing the traffic flow it was concluded that a pre-classification into working days and non-working days significantly improved the clustering result. Four types of working days were classified: Mondays, core week days, Friday's and days within vacation periods. The working day patterns discovered could be used as input for macroscopic traffic models and as a basis for traffic management scenarios.

It is important to note that the cluster analysis was only executed for one location and due to the fact that different highways serve different traffic types precautions should be taken when translating the results of this study to other locations.

16. Tradeoffs between security and Inspection Capacity: Policy Options for Land Border Ports of Entry (2006)

This study was conducted in response to the terrorist attacks that occurred on September 11, 2001. Its purpose was to propose solutions that help increase national security (i.e. more rigorous inspections at international borders) and also avoid causing issues with existing regional traffic networks. This report explored three potential options in order to have more thorough primary inspections at the border crossing between El Paso, Texas and Ciudad Juarez, Mexico. After analyzing each of the three proposed

options it was concluded that none of them were suitable solutions because they would not allow for a more detailed inspection at the border crossing without also causing extra delays in local traffic flows. The study did, however, suggest that these options have the potential to be implemented successfully if they were to be implemented in increments thus alleviating their interference with local traffic.

17. AZ Multimodal Freight TM1: Analysis of Freight Dependent Industries (2007)

This report is the first in a series of three reports prepared by Wilbur Smith Associates for the Arizona Department of Transportation that analyzed Arizona's multimodal freight system. This first report summarized trends within the industry as well as Freight flows specific to the Arizona Freight market. It also presented the findings of surveys and interviews conducted among various businesses and transportation providers in Arizona. This study built on the work of previous studies in order to make Arizona's freight system a more integral part of the long range transportation planning process. It was split into three sections which served to identify both key industries that deal with freight systems as well as trends and key issues in those industries.

18. AZ Multimodal Freight TM2: Assessment of Arizona's Existing Freight Infrastructure (2007)

This report was the second in a series of three reports prepared by Wilbur Smith Associates for Arizona Department of Transportation that analyzes Arizona's multimodal freight system. This study analyzed Arizona's freight infrastructure which affects Arizona's freight movement. Key areas focused on in this study including the number of lanes, traffic congestion areas, and locations of steep grades.

19. AZ Multimodal Freight TM3: Strategic Directions for Freight Planning (2007)

This report was the third in a series of three reports prepared by Wilbur Smith Associates for Arizona Department of Transportation that analyzes Arizona's multimodal freight system. One of the main objectives of this three part study was to come up with a way to make freight analysis an integrated part of Arizona's long range planning process. This report suggested 6 different strategic directions for freight planning in Arizona as well as specific tactics for how to implement them and test their performance. The six strategic directions determined by this study are as listed as follows:

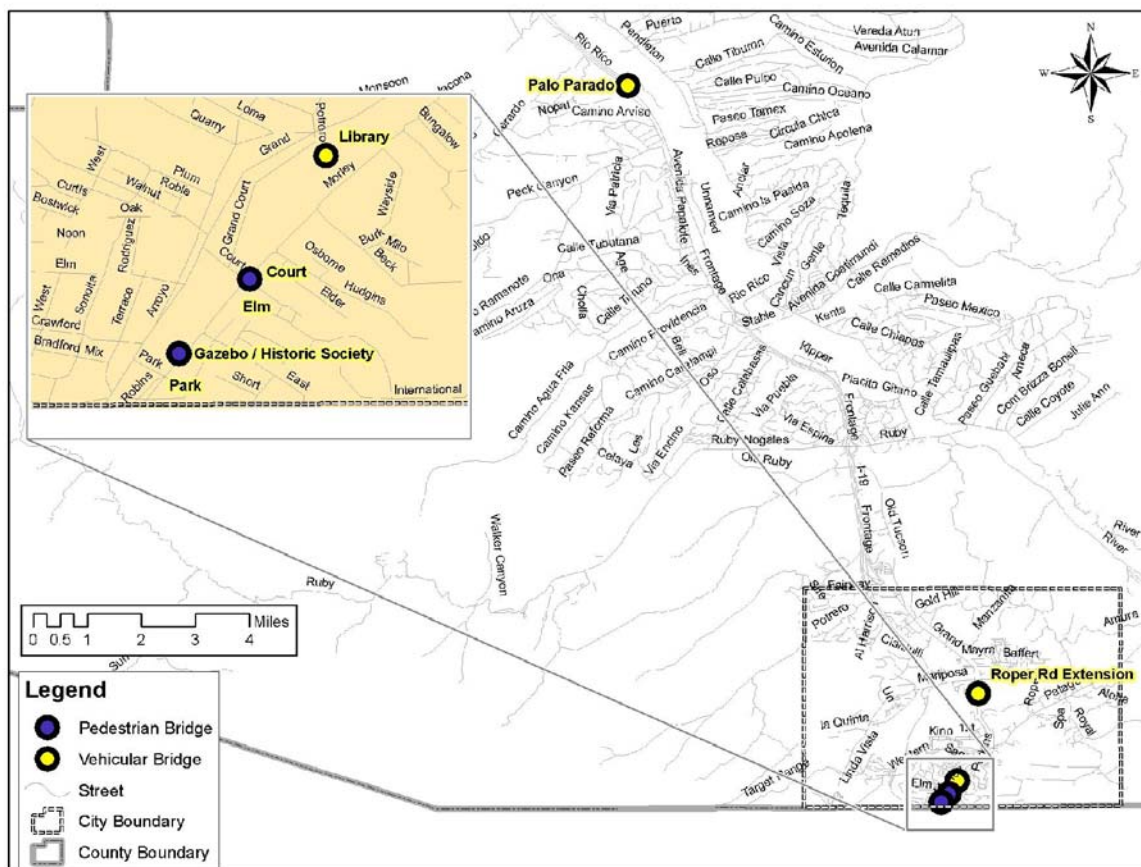
1. Strengthen the planning relationship between freight transportation and economic development
2. Coordinate freight planning with local land use planning
3. Preserve and prioritize key freight infrastructure
4. Seek Opportunities to Improve Freight Operations
5. Enhance Freight System Safety and Security
6. Promote Environmental Preservation and Energy Efficiency in Freight Operations

20. Use of Box and Jenkins Time Series Technique in Traffic Volume Forecasting (2007)

This paper investigates the application of analysis techniques developed by Box and Jenkins to predict a daily traffic volume for a certain Egyptian intercity road. Data was collected for 14 years starting in 1990 and was used to forecast for the year 2003 and this forecast was compared with actual traffic volume in 2003.

21. Nogales Railroad Small Area Transportation Study (2007)

This report was prepared by Kimbley-Horn and Associates for the City of Nogales. The purpose of this report was to analyze the existing situation of the Freight Rail Operations that run through both Nogales, Arizona and Nogales, Sonora. This study provided recommendations as to possible locations for both pedestrian and vehicular bridges over the existing railroads in Nogales. For each suggestion there was also an analysis of its affect on the minority populations surrounding the area and proposed solutions to keep the public informed and thus avoid discriminating against any minority populations. The findings for possible bridge locations can be seen on the map below.



22. Bottleneck Study of Mariposa POE (2008)

This study was conducted at the Mariposa Port of Entry at Nogales in order to determine bottleneck areas which would cause significant delays in the movement of people or goods across the border and to suggest solutions to eliminate these delays. This was done through traffic data collection and analysis to identify the location and nature of bottlenecks both going into and leaving the Mariposa POE. Lastly, the study proposed improvements to alleviate congestion and provides estimates of associated costs.

23. Mariposa/I-19 Connector Route Study (2008)

This study was conducted by Wilbur Smith Associates for the Arizona Department of Transportation Planning Committee. The purpose of this study was to plan for adequate roadway capacity and safe traffic movements to and from the Mariposa Port of Entry to I-19 while keeping in mind the mix of commercial versus privately owned vehicles and the mix of Port versus local traffic.

The traffic moving through the Mariposa Port of Entry primarily used Arizona State Route 189 (SR189), Mariposa Road, to access Interstate Highway 19 (I-19) or Grand Avenue located on the east side of I-19. There was no convenient alternative route to access I-19 and the only possibility would be to use Target Range Road. This route was unsuitable for most travelers due to its circular path as well as the fact that it passed through residential neighborhoods and by the local hospital to access I-19 at the Western Avenue Interchange.

At the time of the study it was believed that SR189 would be unable to adequately handle the anticipated traffic volume in both a safe and effective matter therefore this study explored various options to address this problem. It was determined that if no improvements were made to handle the increasing traffic volumes at the Mariposa POE, extreme congestion could be expected in the area of Frank Reed Road to the Mariposa Road/I-19 TI and beyond to Grand Avenue.

24. Socioeconomic determinants of Mexican Circular and permanent Migration

This study determined which factors had impacted the duration of the Mexican temporary and permanent migrants to the United States. Through surveys the following determinants were found to have a significant impact on Mexican circular and permanent migration: socioeconomic, human capital, migration experience, social capital and labor. We reviewed this study in particular because we had originally hypothesized that the factors that impact the origin and destination of migration across the U.S./Mexico border would be significant in forecasting the amount of POV and Pedestrian crossings.

Appendix: Statistical details

Time Series

The set of data points observed at successive times we call time series data. Let Y be the time series data, then $y_t, t = 1, 2, \dots$ represent the data points of the time series Y .

If the time series contains no trend, it can be represented as:

$$y_t = \beta_0 + \dot{\alpha}_t \quad (1)$$

where β_0 is the unchanging average level of the series, $\dot{\alpha}_t$ is a random variable representing irregular fluctuations around the average level at time t .

If the time series contains a trend, it can be written as:

$$y_t = T_t + \dot{\alpha}_t \quad (2)$$

where $T_t = f(\beta_0, \beta_1, \dots; t), t = 1, 2, \dots$, is the expected value (trend) of the series at time t , $f(\beta_0, \beta_1, \dots; t)$ is an increasing or decreasing function describing the trend pattern. $\dot{\alpha}_t$ is a random variable representing irregular fluctuations around the trend at time t .

Seasonal Series

Time series observed at shorter than yearly intervals often display a regular pattern of fluctuations that repeat from year to year. The commercial truck data in our study is a good example of this kind of pattern. This periodic pattern is usually described as **seasonal movement**, **seasonality** or **seasonal**. In this section, we will briefly introduce what seasonality is and how it is represented. There are two types of seasonality, *Additive* and *Multiplicative*. In our example we use “year” to describe the interval of seasonality, however, the interval of seasonality is not limited to being described in years.

A time series Y , observed L times per year at time $t = 1, 2, \dots, L$ is said to have constant seasonality if the average value of Y changes over time such that

$$\text{Additive:} \quad E(y_t) = f(\beta_0, \beta_1, \dots; t) + S_t \quad (3)$$

$$\text{Multiplicative:} \quad E(y_t) = f(\beta_0, \beta_1, \dots; t) \square S_t \quad (4)$$

where

$$S_t = S(t + L) = S_{t+2L} = \dots$$

and

$$\text{Additive:} \quad \sum_{t=1}^L S_t = 0$$

Multiplicative

$$\sum_{i=1}^L S_i = L$$

and $f(\beta_0, \beta_1, \dots; t)$ is a function describing the trend. Each observation period is called a season, and L , the length of the seasonality, is the number of seasons in a year. The S_i values are called seasonal indexes.

Let \hat{y}_t be the irregular fluctuation away from the trend and seasonal effects at time t , and $T_t = f(\beta_0, \beta_1, \dots; t)$. Then the general seasonal models can be written as

Additive

$$y_t = T_t + S_t + \hat{y}_t$$

Multiplicative

$$y_t = T_t \cdot S_t \cdot \hat{y}_t$$

Additive Holt-Winter's Model

Holt-Winter's model is one type of exponential smoothing model, which is capable of handling the trends and seasonality of our data. There are two kinds of Holt-Winter's models, one is additive and the other is multiplicative. We only consider the additive Holt-Winter's model here.

In additive Holt-Winter's model, we decompose y_t in the following way:

$$y_t = a_t + b_t t + s_t + \hat{y}_t \quad (5)$$

where a_t is the unseasoned level of time series at time t , b_t is the slope of the trend at time t , s_t is index of season $i, i = 1, 2, \dots, L$. i is corresponding to the season of current t .

The additive Holt-Winters prediction function (for time series with period length L) is

$$\hat{y}_{t+h} = a_t + b_t h + s_{t+1+(h-1) \bmod L}$$

where a_t , b_t and S_t are given by

$$\text{Level:} \quad a_t = \alpha(y_t - s_{t-L}) + (1 - \alpha)(a_{t-1} + b_{t-1}) \quad (6)$$

$$\text{Trend:} \quad b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \quad (7)$$

$$\text{Seasonality} \quad s_t = \gamma(y_t - a_t) + (1 - \gamma)s_{t-L} \quad (8)$$

All the three equations are updated dynamically. The equations are intended to give more weight to recent observations and less weight to observations occurring further in the past.

In the result of Holt-Winter's model, we will report the value of α , β and γ used in the model as well as the estimated values for the level, trend and seasonal components, which are shown a , b and

$s_i, i = 1, \dots, L$. Due to the fact that the level and trend are constantly changing throughout the model building process, time is represented using the subscript t . However, the level and trend in the result is the estimated overall level and trend.

ARIMA models

ARIMA stands for autoregressive-integrated-moving average model. ARIMA models are one of the most commonly used time series model. Different ARIMA models are identified by the number of autoregressive parameters (p), the degree of differencing (d), and the number of moving average parameters (q). For short, we write it as $ARIMA(p, d, q)$. If there is not differencing term, i.e. $d=0$, we call it $ARMA(p, q)$. We define the ARIMA model as below:

y_t	=	The actual value of the series at time t
y_{t-i}	=	The value of the series at time $t-i$
$\hat{\epsilon}_t$	=	The error term at time t
$\hat{\epsilon}_{t-i}$	=	The error term at time $t-i$
ϕ_i	=	The autoregressive parameter for y_{t-i}
θ_i	=	The moving average parameter for $\hat{\epsilon}_{t-i}$
w_t	=	$\Delta^d y_t$, the y_t series differenced d times

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \hat{\epsilon}_t + \theta_0 - \theta_1 \hat{\epsilon}_{t-1} - \theta_2 \hat{\epsilon}_{t-2} - \dots - \theta_q \hat{\epsilon}_{t-p} \quad (9)$$

Note that $\hat{\epsilon}_t, \hat{\epsilon}_{t-1}, \hat{\epsilon}_{t-2}, \dots$ are uncorrelated to each other, also the meaning of $\Delta^d y_t$ is shown as below:

$$\Delta^d y_t = y_t - y_{t-d}$$

To build an ARIMA model, we need to choose the value of (p, d, q) first, which determine the structure of the model. After that, we need to estimate all the parameters ϕ_i and θ_i . In the result of ARIMA model, we will report all the ϕ_i 's and θ_i 's. Sometimes, we may use ARi to represent ϕ_i and MAi to represent θ_i . The (p, d, q) values are specified by the model builder, and the computer program takes them as input.

For those who have more interest on this topic, we recommend the building ARIMA model section in Shumway and Stoffer's book.

Forecasting of the ARIMA model

When we have identified the model, we need to use the model to forecast the future values. One can refer to the Updating Equations for ARIMA Forecasting section in Farnum and Stanton's book for mathematical details for parameter estimating and updating.

We give the general procedure for forecasting the ARIMA model here and take the *ARIMA*(1,1,1) model as an example to illustrate it. According to equation (9), *ARIMA*(1,1,1) can be written as:

$$w_t = \phi_1 w_{t-1} + \theta_0 + \dot{\theta}_t - \theta_1 \dot{\theta}_{t-1} \quad (10)$$

where $w_t = \Delta y_t = y_t - y_{t-1}$.

step 1: Rewrite the estimated model and solve for y_t as a function of all other terms in the model.

Equation (10) can be rewritten as

$$y_t = \theta_0 + (1 - \phi_1) y_{t-1} - \phi_1 y_{t-2} + \dot{\theta}_t - \theta_1 \dot{\theta}_{t-1} \quad (11)$$

step 2: Replace t with $t + j$ in the equation resulting from step 1.

By doing this in equation (11), we get

$$y_{t+j} = \theta_0 + (1 - \phi_1) y_{t+j-1} - \phi_1 y_{t+j-2} + \dot{\theta}_{t+j} - \theta_1 \dot{\theta}_{t+j-1} \quad (12)$$

step 3: Use actual values of the series when possible; otherwise use forecasts. Use estimates (residuals) for past error terms, and 0 (their expected values) for future error terms. Suppose we are at time t , and going to forecast for $j = 1, 2, 3$ periods ahead of the *ARIMA*(1,1,1) model.

a. $t+1$: $\hat{y}_{t+1}(t) = \hat{\theta}_0 + (1 + \hat{\phi}_1) y_t - \hat{\phi}_1 y_{t-1} - \hat{\theta}_1 e_t,$

$\dot{\theta}_{t+1}$ is replaced by 0, $\dot{\theta}_t$ is replaced by the estimate, e_t .

b. $t+2$: $\hat{y}_{t+2}(t) = \hat{\theta}_0 + (1 + \hat{\phi}_1) \hat{y}_{t+1} - \hat{\phi}_1 y_t$

both the error terms are in the future and are replaced by 0. y_{t+1} is also a future value, thus replaced by \hat{y}_{t+1}

c. $t+3$: $\hat{y}_{t+3}(t) = \hat{\theta}_0 + (1 + \hat{\phi}_1) \hat{y}_{t+2} - \hat{\phi}_1 \hat{y}_{t+1}$

Every forecast from this point on depends on only the previous two forecasts.

All the above mentioned procedures are usually done by a computer package. In our study, we use the R system (R Development Core Team 2009) computer package to do the parameter estimation in model building. Shumway and Stonffer point out some issues of using R for time series models, which are very useful when using R to build ARIMA models.

Seasonal ARIMA mode

The ARIMA model can be easily extended to handle seasonality. The same methods as the in the nonseasonal ARIMA models are used. However, we have two sets of parameters, one set to deal with the season-to-season movement and the other to deal with the movement within the seasons. We use (P, D, Q) corresponding to (p, d, q) in the nonseasonal ARIMA model. P is number of seasonal

autoregressive parameters, D is the degree of seasonal differencing, Q is the number of moving average parameters. All the parameters for the seasonal part are changed to upper case. For example, the seasonal autoregressive parameter is Φ_i and the seasonal moving average parameter is Θ_i . We use $ARIMA(p, d, q)(P, D, Q)_L$ to represent a multiplicative seasonal ARIMA model with nonseasonal orders (p, d, q) , seasonal orders (P, D, Q) and length of seasonal L . In the nonseasonal ARIMA model, the one order differencing is performed as $\Delta y_t = y_t - y_{t-1}$, while the one order seasonal differencing should take over a span of L periods rather than one period. Thus $\Delta_L y_t = y_t - y_{t-L}$. The following paragraphs will give some examples of the equations of seasonal ARIMA model.

There are two methods to combine the seasonal part and nonseasonal part together, one is Additive, and the other is Multiplicative. In the additive method, we just add all the seasonal terms and nonseasonal terms together, while in the multiplicative method, the combined model is formed by applying the nonseasonal model to the terms we get from a purely seasonal model.

We use the ARIMA model with one seasonal MA term ($P=1$) and one nonseasonal MA term ($p=1$) as an example. A purely seasonal model (without nonseasonal terms) first-order moving average model would be written:

$$y_t = \Theta_0 + \dot{\Theta}_t - \Theta_1 \dot{\Theta}_{t-L} \quad (13)$$

A purely nonseasonal first-order moving average model would be written:

$$y_t = \theta_0 + \dot{\theta}_t - \theta_1 \dot{\theta}_{t-1} \quad (14)$$

By combining additively, we write

$$y_t = \Theta_0 + \dot{\Theta}_t - \theta_1 \dot{\Theta}_{t-1} - \Theta_1 \dot{\Theta}_{t-L} \quad (15)$$

Note that the Θ_0 in equation (15) is different from that in equation (13).

Using the multiplicative way, we write the seasonal MA part as:

$$y_t = \Theta_0 + u_t - \Theta_1 u_{t-L} \quad (16)$$

then applying the nonseasonal MA term to the errors u , we have

$$u_t = \theta_0 + \dot{\theta}_t - \theta_1 \dot{\theta}_{t-1} \quad (17)$$

Substituting equation (17) back into (16) gives

$$\begin{aligned} y_t &= \Theta_0 + (\theta_0 + \dot{\theta}_t - \theta_1 \dot{\theta}_{t-1}) - \Theta_1 (\theta_0 + \dot{\theta}_t - \theta_1 \dot{\theta}_{t-1}) \\ &= \Theta_0 + \dot{\theta}_t - \theta_1 \dot{\theta}_{t-1} - \Theta_1 \dot{\theta}_{t-L} + \theta_1 \Theta_1 \dot{\theta}_{t-L-1} \end{aligned} \quad (18)$$

Here we have the equation for y_t with both seasonal and nonseasonal parameters incorporated.

The parameter identification for the seasonal ARIMA model is really a two step process, one to determine the nonseasonal parameters, and the other to determine the seasonal parameters. In the result of seasonal ARIMA model, in addition to ϕ_i 's and θ_i 's, we also report Φ_i 's and Θ_i 's. In some cases, we may report Φ_i as SARi and Θ_i as SMAi.

ARIMA Model with exogenous variables

When taking exogenous variables into consideration. For example, if we take another series X in to consideration. We build the ARIMA model on the data set $Y - X$. Take the $ARIMA(1,1,0)$ model for example, the original model is

$$\nabla y_t = \phi \nabla y_{t-1} + \hat{\epsilon}_t$$

After incorporating the X series, the new model becomes

$$\nabla(y_t - \beta x_t) = \phi \nabla(y_{t-1} - \beta(x_{t-1})) + \hat{\epsilon}_t$$

Where β is another parameter we need to estimate during the model building stage, which will be reported in the result if we use any exogenous variable.

Explain of some statistics

In this part, we will explain some statistics we used in our report. We will cover the ACF (sample autocorrelation function), PACF (sample partial autocorrelation function), Theil's U statistic, R^2 (coefficient of determination), and VIF (Variance Inflation Factor).

ACF and PACF

When evaluating the autocorrelation coefficients evaluated at lag 1, lag 2, lag3, ...lag k ,... and graphed versus k , we get the sample autocorrelation function (ACF). Suppose we have the data series $Y = \{y_1, y_2, \dots, y_i, \dots, y_n\}$, the sample autocorrelation coefficient of lag k , r_k can be calculated by the following way. Take $(y_1, y_{k+1}), (y_2, y_{k+2}), \dots, (y_{n-k}, y_n)$ pairs of the data, and denote the subset $\{y_1, y_2, \dots, y_{n-k}\}$ as Y_1 , the subset $\{y_{k+1}, y_{k+2}, \dots, y_n\}$ as Y_2 . The correlation between Y_1 and Y_2 is r_k .

$$r_k = \frac{\sum_{i=k+1}^n (y_{i-k} - \bar{y}_1)(y_i - \bar{y}_2)}{\sigma_1 \sigma_2}$$

Where \bar{y}_1 and \bar{y}_2 are the means of Y_1 and Y_2 respectively, σ_1^2 and σ_2^2 are the variances of Y_1 and Y_2 respectively.

As long as the number of data points in the data series is large compared to k , substitution of the total means (\bar{y}), and sums of square ($\sum (y_i - \bar{y})^2$) for partial ones does not greatly change the value of r . Thus, the r_k can be approximately calculated as :

$$r_k = \frac{\sum_{i=k+1}^n (y_{i-k} - \bar{y})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

However, one needs to use the partial ones when then number of data points in the data series is not large enough.

PACF is another important tool for ARIMA model building. The partial autocorrelation coefficient of lag k , denoted as ϕ_{kk} , is a measure of the correlation between y_t and y_{t-k} after adjusting for the presence of all the y_t 's of shorter lag, i.e. $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$. The purpose for this adjustment is to see if there is any additional correlation between y_t and y_{t-k} above and beyond that introduced by the correlation y_t has with $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$. Graphing the ϕ_{kk} with $k = 1, 2, \dots$, we get the sample partial autocorrelation function (PACF). A common way for most computer packages to compute the partial autocorrelation function is to recursively use the knowledge of the autocorrelation coefficients r_k . For more details on the procedure, one can refer to the PACF section in Farnum and Stanton's book.

We use ACF as one of the main tools to build the ARIMA model. A slow decay in ACF is an indication that differencing may be needed. If ACF Cuts off after lag q , we may put a $MA(q)$ term in the ARIMA model. If the PACF cuts off after lag p , probably we need to put a $AR(p)$ term in the ARIMA model.

Table 1 summarizes the behavior of ACF and PACF for the ARMA/ARIMA models. Additionally, we use the ACF to decide whether to use differencing or not in ARIMA model.

Table 1 Behavior of the ACF and PACF for ARMA models

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ACF	Tail off	Cuts off after lag q	Tail off
PACF	Cuts off after lag p	Tails off	Tail off

Theil's U statistic

Theil's U statistic compares the forecast of a model to the "no-change" model. The "no-change" model refers to the model that assumes that the values are relatively stable, $\hat{y}_{t+1} = y_t$. Mathematically, the Theil's U statistic can be written as:

$$U = \frac{\sqrt{\sum e_t^2}}{\sqrt{\sum (y_t - y_{t-1})^2}} \quad (19)$$

$$= \frac{\sqrt{MSE(model)}}{\sqrt{MSE("no-change" model)}} \sqrt{\frac{n}{n-1}} \quad (20)$$

$$= \frac{\sqrt{\sum (A_t - P_t)^2}}{\sqrt{\sum A_t^2}} \quad (21)$$

where

- \hat{y}_t = Estimated value using the model being compared
- e_t = $y_t - \hat{y}_t$ = the forecast error of the model being compared
- P_t = $\hat{y}_t - y_{t-1}$ = the predicted change in period t
- A_t = $y_t - y_{t-1}$ = the actual change in period t

A zero valued U statistic means the model being compared forecasts exactly to the real value, while a one valued U statistic means the model being compared is as good as the "no-change" model. A U statistic greater than one means the model being compared is worse than the "no-change" model, while a U statistic less than one means the model being compared is better than the "no-change" model. Note that the value of the U statistic is always greater than zero; therefore, we prefer the models that can render U statistics as small as possible.

Coefficient of Determination (R square)

The quantity

$$R^2 = \frac{SS_{model}}{SS_{total}} \quad (22)$$

$$= 1 - \frac{SS_{err}}{SS_{total}} \quad (23)$$

is called the coefficient of determination, and is usually called R square for short. The quantities involved in the above equations are defined as below:

$$SS_{total} = \sum_i (y_i - \bar{y})^2$$

$$SS_{model} = \sum_i (\hat{y}_i - \bar{y})^2$$

$$SS_{err} = \sum_i (y_i - \hat{y})^2$$

The R^2 value can be interpreted as the proportion of variation explained by the model. But the R^2 value should be used with caution, since it is always possible to make the R^2 value of a regression model as large as possible. For the regression model, the R^2 value is between 0 and 1, however, for other models, this cannot be guaranteed.

The two definition equations for the R^2 are not always equal. When the model used is a linear regression model, we have

$$SS_{total} = SS_{model} + SS_{err} \quad (24)$$

thus the two definition equations are equal to each other. However, when using some other models, where equation (24) does not hold, the values defined by the two definition equations are different. Depending on which definition is used, the R^2 can be below 0, if we use equation (23), or exceeding 1, if we use equation (22).

For the data with strong nonlinear relationship exists, the R^2 may be not a good measure of how well the model fit the data. When this happens, we would like to use the correlation between the predicted value and the real value to evaluate how well the model fitted value is.

VIF (Variance Inflation Factors)

The Variance Inflation Factors are used for a multicollinearity diagnostic. Multicollinearity among the regression variables implies near-linear dependence among the regressors (regression variables). The presence of near-linear dependencies can dramatically impact the ability to estimate the regression coefficients. Suppose now we have p regressors, then the VIF for the j^{th} regression coefficients can be written as

$$VIF_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the coefficient of multiple determination obtained from regressing x_j on the other regressors. Clearly, if x_j is nearly linearly dependent on some of the other regressors, then R_j^2 will be near unity and VIF_j will be large. VIFs larger than 10 imply serious problems with multicollinearity. Note the VIF values are greater than 1 and less than infinity.

Apparently, if there is only one regressor, we do not have the problem of multicollinearity. If we have two regressors x_1 and x_2 , VIF_1 will be equal to VIF_2 . For more than two regressors, according to the calculation method of VIF_j , their VIF s are usually not equal to each other.

This section is only a brief introduction to multicollinearity and VIF values. For more details about this topic, we recommend the reader refer to the textbook by Montgomery, Peck, and Vining.

Statistical Tests

Ljung-Box test(Farnum and Stanton)

The Ljung-Box test is used to test autocorrelations for several different lags of a time series data. The Hypothesis of the test is as below:

$$H_0: \rho_k = 0 \text{ for all } k \leq m$$

$$H_a: \rho_k \neq 0 \text{ for some value of } k \leq m$$

The test statistic:

$$Q_m = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k},$$

where m is the number of coefficients being tested and n is the number of observations in the series.

The decision rule:

Reject H_0 if $Q_m > \chi^2_{\alpha}(m)$, otherwise, do not reject H_0 . $\chi^2_{\alpha}(m)$ is the upper $\alpha \times 100\%$ points of the chi-square distribution with m degree of freedom.

Conclusion:

If H_0 is rejected, we conclude that with approximately $(1-\alpha) \times 100\%$ confidence that the series is nonrandom.

If H_0 is not rejected, we have some support for a random no-trend model.

In the process of building an ARIMA model, we usually apply the Ljung-Box test on the residual to check if the residual is random and contains no trend. For an ARIMA $(p,d,q)(P,D,Q)_L$ model, the n in the Q_m statistic is the number of terms in the series **after** differencing and m is the number of autocorrelation coefficients being tested. The number of degrees of freedom used in the chi-square test is: $m - (p+1) - (P+Q)$

References

- Farnum, Nicholas R., and LaVerne W. Stanton. Updating Equations for ARIMA Forecasting. In *Quantitative Forecasting Methods*, 487-488.
- . Partial Autocorrelation Function(pacf). In *Quantitative Forecasting Methods*, 290-294.
- . Ljung-Box test. In *Quantitative Forecasting Methods*, 84.

- Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. 2006. Variance Inflation Factors. In *Introduction to Linear Regression Analysis*, 111,335-340. 4th ed. WILEY, July.
- R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. <http://www.R-project.org>.
- Shumway, Robert H., and David S. Stoffer. 2006. Building ARIMA Models. In *Time Series Analysis and Its Applications: With R Examples*, 143-164. 2nd ed. Springer, May 25.
- Shumway, Robert H., and David S. Stonffer. R Time Series Issues. Book website. *Time Series Analysis and Its Applications: With R Examples*. <http://www.stat.pitt.edu/stoffer/tsa2/Rissues.htm>.

Appendix: Simulation Model Detail

Introduction

The simulation created for this model was based on a model created for a previous study also conducted by Arizona State University for ADOT entitled *Logistics Study of the Guaymas-Tucson Corridor* (Villalobos et al.). For specific details regarding the baseline structure of the simulation we will refer back to this study. The simulation for this study simulation was a follow up step for the Mariposa POE visit conducted in May of 2009. The data gathered during this visit was used to modify the previously created simulation model to make it an accurate representation of the current state of the Mariposa POE. Once the simulation was validated the forecasted values for commercial vehicles were used as inputs to the system to test the current capacity of the POE against predicted future traffic demands.

As previously mentioned, the underlying purpose of the simulation analysis of the Mariposa Port of Entry is to test the ability of the current setup of this POE to handle the increased traffic demands predicted for trucks in five, ten and fifteen year time periods into the future. Some of the measures of the ability of the POE to handle the varying traffic inputs (measured per commercial vehicle) are the following:

- Average time a truck spends in the system
- Time required to process all trucks and how many of those hours exceed the typical 11 hours workday
- Maximum number of trucks in queue
- Bottleneck locations
- Approximate Utilization of bottleneck locations

The ability of the current layout of the Mariposa POE in allowing commercial vehicles to cross the border in a timely matter is determined through the present simulation. The model is based on the ProModel® V6.0 Montecarlo simulation package and it is aimed to obtain the most accurate representation of the current systems flexibility in coping with forecasted future demands for commercial vehicles. By using a simulation model we are able to generate and test several scenarios of our forecasted daily demands of commercial vehicle traffic, in order to analyze and compare all potential outcomes.

For details about the structural elements already built into the simulation model please refer to Guaymas Study (Villalobos et al.).

Process Flow

In our simulation model each commercial vehicle is represented as an entity that enters the system according to a predetermined arrival time, which varies for each scenario. Based on historical data of the commercial vehicles each entity is assigned certain attributes which are used to determine the route the vehicle will take as it goes through the system. To begin, trucks are separated into two categories, those going through the fast lane or normal lanes (3 lane choices for normal). For those vehicles which are routed through the fast lane, they proceed directly from the primary inspections to the highway exit. If they are not routed directly to the fast lane it means they must pass through one or any combination of CBP inspections (including Document check, enforce, full inspection and X-ray) or ADOT inspections. Once all the necessary inspections are completed for each truck it is routed to the highway exit and leaves the system. Note before each truck visibly enters the system on the highway in the upper left corner (starting point of the POE) it is assigned specific characteristics to determine which inspections it will require. The percentages of trucks requiring each type of inspection are based on historical data of the Mariposa POE.

The logical flow of entities in the simulation is explained in further detail in the diagram Figure 1 in Appendix H of the Guaymas Study (Villalobos et al.). In summation, the logical process flow is as follows: when a truck enters the system it must pass through all primary inspections then depending on what attributes it has been assigned it will either be routed straight to the highway exit or it will go through additional inspections and then be routed to the highway and exit the system.

The whole system can be divided into four different sections:

1. Pre-Screening and Primary Inspections: These are the first steps in the process and all trucks are required to go through them.
2. Secondary Inspection: Different tasks can be done in this section: normal secondary inspection, Full (100%) inspection, weapons and enforce inspection and others.
3. X-ray: three stations for x-ray inspection.
4. ADOT compound: ADOT's Motor Vehicle Division safety inspection and other Federal inspections are conducted here.

While the trucks move through all the different individually required steps of the inspection process, several institutions work together. A partial list includes:

- Customs and Border Protection (CBP)
- United States Department of Agriculture (USDA)
- Food and Drug Administration (FDA)
- Arizona Department of Transportation (ADOT)
- Federal Motor Carrier and Safety Administration (FMSCA)

The physical movement of the trucks can be observed in the animation of the simulation displayed in Figure 0-1 Graphical interface of the simulation program. Currently the trucks cross the border in four

lanes, one of them being a fast lane and the other three being regular lanes. All trucks will then enter a pre-screening station, follow to one of the four primary inspection super-booths, and then proceed to either Nogales, Arizona (if they were in the fast lane) or else go on for further inspection in a counter clockwise (CCW) motion around the compound.



Figure 0-1 Graphical interface of the simulation program

Assumptions of the System

Given the security requirements at the site of the POE, we were not able to observe and record all the activities that we would have liked. We were unable to obtain clearance from CBP to enter into their inspection area and measure inspection times thus these inspection times were based on those used in simulation model of the previous study of the Guaymas-Tucson Corridor (Villalobos et al.).

Infrastructure

The physical infrastructure of the POE is presented in Figure 0-1 Graphical interface of the simulation program(in previous section). Each part of the facility is described next, these areas are displayed in Figure 0-1:

1. There are four pre-screening lanes and inspection stations (topmost lane is designated as fast and the other three are normal lanes)

2. There are four primary inspection stations (super-booths).
3. There are different quantities of docks in each side of the main compound, where the grading and detailed inspection takes place.
4. There are three X-ray stations.
5. There are two inspection lanes in the ADOT yard.
6. There are 25 parking spaces in the ADOT yard.

The detailed description of the infrastructure in the main compound is presented in Table 0-1. The information shown in this table was last updated in Dec 2005 as was used in a previous study, *Logistics Capacity Study of the Guaymas-Tucson Corridor* (Villalobos et al.). Changes may have occurred since that time but due to security issues this is the information we have to the best of our knowledge:

Table 0-1 Equipment Available at the main compound

<u>Side</u>	<u>Docks Available</u>
North	18
South	20
West	15
East	21

Movement Logic

Trucks move inside the compound in a counter clock wise (CCW) motion and follow the next priorities when several inspections are required:

1. X-ray
2. Enforcement (full inspection, hazardous and weapons inspection, etc.)
3. Document Review
4. ADOT (all inspections done at ADOT yard)

Due to physical limitations the trucks that are released in the fourth (last to the east) super-booth must go into the compound and drive around the compound again before they are able reach the highway exit. Trucks that require an X-ray inspection will go into the ADOT yard and form a line from there towards the X-ray booths.

Processing Detail

The daily volume of traffic for trucks was determined based on the results of our forecasts. The exact number of trucks for each scenario can be seen in

Table 0-4. The percentage of trucks that are assigned to go through each type of inspection can be seen in Table 0-2.

Table 0-2 Distributions of Inspection Procedures

Percentage	Description
100 %	Pre-Screening
100 %	Primary Inspection
30.74 %	Released to enter the US from Primary inspection (fast lane)
69.26 %	Required further inspections and enter the compound (normal lanes)
*Out of the 69.26% that requires more inspection:	
33 %	Required X-Ray
17 %	Required Full Inspection or Hazardous and Weapons Inspection
83 %	Required Documentation Review
20 %	Required to enter the ADOT yard for Inspection

The inspection times for each station were based on those previously determined in the Guaymas-Tucson Corridor Study (Villalobos et al.). The summary of these inspection times is shown in Table 0-3 below. For further details refer to processing details in appendix H of Guaymas Study (Villalobos et al.).

Table 0-3 - Distribution of processing time of each activity at the POE

Inspection	Time Distribution (min)
Pre-Screening	ERLANG (0.72, 3)
Primary Inspection	ERLANG (1.33, 3)
* 20% including ADOT in Super-Booth	ERLANG (2, 3)
Document Revision	ERLANG (30.745, 3)
Full Inspection	ERLANG (82.2, 3)
Hazardous and weapons Enforcement	ERLANG (82.2, 3)
X-Ray	ERLANG (8.27, 3)
ADOT	TRIANGULAR (25, 30, 35)

The inspection times within the CBP area of the simulation (Document Revision, X-ray, Full, Hazardous and Weapons Enforce) were updated from those used in the Guaymas study (Villalobos et al.) in order to match the overall CBP inspection times collected on our visit to the Mariposa POE. Again due to clearance issues we were not given permission to enter and record data times in this area so their values were changed according to weighted percentages which will be explained more in detail in the validation section of the simulation appendix.

Simulation Scenarios

Ultimately, the goal of creating our simulation model was to test the capacity of the current setup of the Mariposa POE by changing the number of vehicles that enter the system. To do this we first analyzed the

daily traffic data for the year 2008 and found that the percentage of trucks crossing each weekday (Monday-Friday) was a uniform 20% for each day. Then we took our monthly forecasted data and divided it by the average number of workdays in the month of May (20 days) when our visit was conducted to get a daily traffic demand.

To test the various forecasted traffic demands we used the last year in each scenario as our input data for the simulation since this was the largest demand and would greatest test the capacity of our system. All scenarios beginning with the number 1 refer to five year forecasts, those starting with 2 are ten year forecasts and those with 3 are the fifteen year forecasts. The second number after the hyphen refers to which scenario of the five, ten or fifteen year forecast is being used; refer to the section “forecast of commercial vehicles” for further details of what each scenario represents.

Table 0-4 Daily Demand for each Scenario

Scenario	Daily Traffic Demand
1-1	1928
1-2	1800
1-3	1759
1-4	1945
1-5	1816
1-6	1775
1-7	1976
1-8	1847
1-9	1806
2-1	2131
2-2	1969
2-3	1909
2-4	2161
2-5	2000
2-6	1939
3-1	2302
3-2	2139
3-3	2042
3-4	2325
3-5	2159
3-6	2062

Results and Statistical Evaluation

The results of running our simulation under the previously described scenarios are displayed in Table 0-1. In this table, the first two columns show the scenario number and the number of trucks used as a daily demand input for each scenario. The third and fourth columns represent the total number of hours required to process all trucks and how many of those are additional hours over the current 11 hour workday that the port is open. The fifth column shows the average amount of time (in minutes) that a truck will spend in the system. The sixth and seventh columns show the 95% low and high confidence intervals for the maximum number of trucks that will wait in queue on the highway. The last two columns on the right show the bottleneck locations and their approximate utilizations for each scenario.

Table 0-1 Results of running the simulation

Scenario	# Trucks	Required Process time	Extra hours required	Avg. time in system (min)	Max in Queue (low 95%)	Max in Queue (high 95%)	Bottleneck	Approx. Utilization
1-1	1928	15.50	4.50	389.710	1888.26	1893.74	Super-booths	87.70%
1-2	1800	14.66	3.66	368.655	1759.58	1767.42	Super-booths	80.75%
1-3	1759	14.66	3.66	361.327	1719.37	1727.03	X-ray	81.50%
1-4	1945	16.22	5.22	395.704	1904.04	1908.96	Super-booths	81.30%
1-5	1816	15.14	4.14	367.047	1773.80	1781.40	Super-booths	78.84%
1-6	1775	14.64	3.64	362.902	1735.65	1740.15	Super-booths	79.42%
1-7	1976	17.04	6.04	401.391	1934.06	1940.34	Super-booths	84.87%
1-8	1847	15.62	4.62	370.460	1807.50	1813.50	Super-booths	82.69%
1-9	1806	15.15	4.15	363.639	1764.67	1772.33	Super-booths	81.27%
2-1	2131	17.51	6.51	424.579	2091.24	2096.76	Super-booths	84.12%
2-2	1969	16.92	5.92	399.541	1928.55	1936.45	Super-booths	78.84%
2-3	1909	15.60	4.60	387.990	1868.91	1875.69	Super-booths	89.04%
2-4	2161	17.91	6.91	432.178	2121.45	2128.35	Super-booths	84.56%
2-5	2000	16.51	5.51	407.981	1960.05	1964.35	Super-booths	81.25%
2-6	1939	15.89	4.89	388.628	1896.77	1904.83	Super-booths	86.60%
3-1	2302	18.39	7.39	458.475	2262.94	2270.06	Super-booths	87.69%
3-2	2139	17.21	6.21	426.991	2098.73	2107.67	Super-booths	81.43%
3-3	2042	16.65	5.65	412.149	2000.89	2008.31	Super-booths	81.44%
3-4	2325	18.82	7.82	471.270	2285.52	2291.08	Super-booths	87.71%
3-5	2159	17.28	6.28	433.375	2119.19	2127.61	Super-booths	83.49%
3-6	2062	16.70	5.70	416.790	2020.59	2030.21	Super-booths	87.21%

Some of the results that can be obtained from the simulation include the following:

- The maximum number of trucks that will wait in a queue on the highway according to our 95% confidence intervals is within the range of 2119 and 2127 trucks.
- For almost all scenarios the bottleneck location is the super-booths (Insp_PrePri_Norm1, Insp_PrePri_Norm2, Insp_PrePri_Norm3), with the exception of Scenario 1-3 where the bottleneck location is X-ray inspection.
- Based on our forecasts for daily truck traffic we can see that the current system is already at capacity due to the fact that in every scenario additional hours over the typical 11 hour workday are required for all trucks to be processed.

Validation

In order to validate our simulation we used the data times collected on our Mariposa POE visit and updated the simulation created in the Guaymas study in order for its output times to match those we observed. We found that the most accurate manner in which to compare the times recorded in our visit with those in the simulation was by matching the times a truck spent in the CBP area.

Initially we found that time spent in CBP that we measured on our visit of 27.117 minutes was greater than the CBP time of 20.23 minutes in the original version of the simulation. To make up for the 6.887 minute difference we multiplied the inspection times of each area in CBP by a ratio as calculated in Table 0-1 below:

Table 0-1 Changing CBP inspection times

Inspection	% of trucks that receive each inspection	calculation	ratios
DOC	83	$(83/133) \times 6.887$	= 4.298
XRAY	33	$(33/133) \times 6.887$	= 1.708
ENFORCE/FULL	17	$(17/133) \times 6.887$	=0.880
TOTAL	133		6.886

The final inspection times determined for Document, X-ray, Full, and Weapons Enforce Inspections can be found in Table 0-3.

Aside from comparing CBP times we also used some subjective measurements to validate our simulation model. From our visit to the Mariposa POE we determined that if a truck was to enter the system and only pass through primary inspections then exit the system directly afterwards, it would have a total time spent in the system of 4 min 45 seconds or less. Therefore we used this information to change the percentage of trucks in the simulation that are routed through the fast lane and directly leave the system after primary inspections to 30.74% in accordance with our observations (see Table 0-2).

Conclusion

The results of running the simulation model compared the actual inspection times measured in our visit to the Mariposa POE give us confidence in the validity of the results produced by our simulation model.

From our results we can see the given the forecasted future demands of traffic the system is already at capacity and would be unable to handle these traffic demands given the current infrastructure and length of workday (11 hours). We also found that for all but one of our forecasted scenarios the bottleneck of the system occurred at the same location, which we found to be the primary inspection or super-booths.