Transportation Research Part C 63 (2016) 182–194

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Short-term prediction of border crossing time and traffic volume for commercial trucks: A case study for the Ambassador Bridge

Md Moniruzzaman [a,*], Hanna Maoh [a,b,1], William Anderson [a,c,2]

[a] University of Windsor, Cross-Border Institute, 401 Sunset Ave., Windsor, ON N9B 3P4 Canada
[b] University of Windsor, Department of Civil and Environmental Engineering, 401 Sunset Ave., Windsor, ON N9B 3P4 Canada
[c] University of Windsor, Department of Political Science, 401 Sunset Ave., Windsor, ON N9B 3P4 Canada

ARTICLE INFO

Article history:
Received 4 August 2015
Received in revised form 30 November 2015
Accepted 4 December 2015
Available online 5 January 2016

Keywords:
Short-term forecast
Artificial Neural Networks
Cross-border
Traffic flow
Crossing time
Ambassador Bridge

ABSTRACT

Short-term forecasting of traffic characteristics, such as traffic flow, speed, travel time, and queue length, has gained considerable attention from transportation researchers and practitioners over past three decades. While past studies primarily focused on traffic characteristics on freeways or urban arterials this study places particular emphasis on modeling the crossing time over one of the busiest US–Canada bridges, the Ambassador Bridge. Using a month-long volume data from Remote Traffic Microwave Sensors and a yearlong Global Positioning System data for crossing time two sets of ANN models are designed, trained, and validated to perform short-term predictions of (1) the volume of trucks crossing the Ambassador Bridge and (2) the time it takes for the trucks to cross the bridge from one side to the other. The prediction of crossing time is contingent on truck volume on the bridge and therefore separate ANN models were trained to predict the volume. A multilayer feedforward neural network with backpropagation approach was used to train the ANN models. Predicted crossing times from the ANNs have a high correlation with the observed values. Evaluation indicators further confirmed the high forecasting capability of the trained ANN models. The ANN models from this study could be used for short-term forecasting of crossing time that would support operations of ITS technologies.

1. Introduction

Short-term forecasting of traffic characteristics, such as traffic flow, speed, travel time, and queue length, has gained considerable attention from transportation researchers and practitioners over past three decades (Chen and Rakha, 2014; Lin et al., 2014a; Qi and Ishak, 2014; Vlahogianni et al., 2014; Zhang et al., 2014). The primary goal of the forecasting is to feed predicted information into Intelligent Transportation Systems (ITS). ITS, for instance – Advanced Traveller Information Systems (ATMS) and Advanced Traffic Management Systems (ATMS), provide advanced information to travellers, about traffic conditions or scheduling for a trip. These systems are also used to post travel time and/or expected delays via on-route variable message signs (VMS). Quick processing of real-time data that is fed into VMS has been possible with the progression

* Corresponding author. Tel.: +1 519 253 3000x4531; fax: +1 519 971 3686.
E-mail addresses: monirm@uwindsor.ca (M. Moniruzzaman), maohhf@uwindsor.ca (H. Maoh), bander@uwindsor.ca (W. Anderson).
[1] Tel.: +1 519 253 3000x4987.
[2] Tel.: +1 519 253 3000x2366.

http://dx.doi.org/10.1016/j.trc.2015.12.004
0968-090X/

achieved in computational power and the availability of more sophisticated methodologies (Fei et al., 2011; Vlahogianni et al., 2005).

A broad array of methods was used in past research to model real-time traffic data to perform short-term predictions (e.g. 10 min). These methods could be classified as parametric, nonparametric methods or hybrid of both. Parametric methods include autoregressive integrated moving average (ARIMA) based models (Chandra and Al-Deek, 2009; Smith et al., 2002; Vlahogianni et al., 2004; Wang and Papageorgiou, 2007; Williams and Hoel, 2003). On the other hand, nonparametric data driven approaches include a variety of Artificial Neural Network (ANN) models, nearest neighbor nonparametric regression, simulation models, Bayesian models, and Support Vector Regression (SVR) (Fei et al., 2011; Lin et al., 2014a; Smith et al., 2002; Vlahogianni et al., 2005). Hybrid integration models that make use of fuzzy logic and genetic algorithms were also used in past studies (Dimitriou et al., 2008; Ishak and Alecsandru, 2004).

Although there are ample research on short-term forecasting of traffic flow, travel time, and/or speed, most of the existing efforts focused on freeways or urban arterials. In comparison, little has been done to model traffic conditions at border crossings (Huang and Sadek, 2009; Lin et al., 2014a). Typically, cross-border flows are different from those observed on major road arterials. In North America, the increased security and inspection procedures in aftermath of the attacks of 9/11 have emphasized the importance of researching land border facilities and their performance (Anderson, 2012; Anderson et al., 2014). In this paper, we focus on the analysis of border traffic volume and crossing times at the Ambassador Bridge, which connects Windsor, Ontario to Detroit, Michigan. The Ambassador Bridge stands as one of the busiest Canada–US crossings as it handles over 26% of the total Canada–US trade (Anderson, 2012).

ANN models (per traffic direction) were designed, trained, and validated to perform short-term predictions of (1) the volume of trucks crossing the Ambassador Bridge and (2) the time it takes for the trucks to cross the bridge from one side to the other. A multilayer feedforward neural network with backpropagation approach was used to train the ANN models. The training of the first set of models is based on a one month worth of Remote Traffic Microwave Sensors (RTMS) data pertaining to trucks crossing the bridge from Canada to the US and vice versa during the months of May and June in 2015. On the other hand, Global Positioning System (GPS) data from a large sample of Canadian trucks that crossed the bridge during a yearlong period (September 2012–August 2013) were employed to model the crossing times on the bridge. To our knowledge, no previous study utilized the type of detailed data employed in this research to develop short-term predictive models of traffic volumes and crossing times for a land border crossing.

The remainder of this paper is organized as follows: the next section provides an overview of the existing efforts that have been performed to forecast traffic characteristics at border crossings. Section three highlights the importance of the Ambassador Bridge and the size of daily truck traffic it handles. It also describes the datasets used in the performed analysis. Section four presents the ANN approach used in the modeling of the data. It also highlights the evaluation indicators employed to validate the trained networks. Section five reports the results and discusses the performance of the developed ANN models. Finally, the last section provides a conclusion to our study.

2. Forecasting traffic characteristics on border crossings

To our knowledge, there is only one attempt in the past that paid particular attention to forecasting traffic flow on a border crossing. In that study, Lin et al. (2014a) used an innovative approach entitled "Enhanced Spinning Network" to forecast hourly traffic volumes at the Peace Bridge that connects Niagara Falls, Canada and Buffalo, US. The method builds on an original method known as the Spinning Network (SPN) algorithm, which was first proposed by Huang and Sadek (2009). According to the authors, an SPN algorithm attempts to mimic how the human memory works to remember information to come up with an exact answer.

Typically, the process of remembering something specific is based on retrieving fragments of information that would be stored in different layers of memory. These layers could be represented by concentric rings in which rings with larger radii represent distant memory in the mind. At first, an individual spends time spinning through the elements of her fragmented and distant memory to piece out meaningful information by comparing the various fragments. Once meaningful information is pieced together, it is placed in a more recent context (i.e. passed to a smaller ring) and compared to the memory fragments found in that recent layer to piece together further meaningful details. Joined information is then passed to a more recent context and so forth until the individual is able to paint a full picture about what she needs to remember.

The enhanced SPN used in Lin et al. (2014a) extends the original SPN by using the Dynamic Time Warping (DTW) algorithm. It is argued that the DTW is more suitable for time series analysis since it provides the optimal pairing of the elements of two time series sequences. The advantage of the DTW is its ability to perform a comprehensive comparison of all the elements of two time series sequences to pair elements that have a more similar pattern. The performance of the enhanced SPN was compared to the outcome obtained from three other methods which included the original SPN, SARIMA (Seasonal-ARIMA) model, and SVR model. Hourly crossing volumes during 2009 and 2010 on the Peace Bridge were used to compare the predicting accuracy of the four approaches. Hourly volumes were predicted for five distinct periods: Mondays through Thursdays, Fridays, weekends, holidays, and game days. The Mean Absolute Percentage Error (MAPE) was used to compare the results. It was found that the DTW-SPN outperformed all other approaches except for game days where SVR performed slightly better.

A handful of studies have been conducted in the past to predict the delays of land border crossings (Khan, 2010; Lin and Lin, 2001; Lin et al., 2014b; Paselk and Mannering, 1994). Paselk and Mannering (1994) used a hazard duration model to study the determinants of border delays at the US–Canada border crossing between Washington State and the province of British Columbia. The model was then applied to predict instantaneous delays of the formed queues at the border. Lin and Lin (2001) used simulated delay data to develop a model that could provide quick and realistic estimates of border delays. The model took into account the time of day, volume/capacity ratio, number of available inspection booths, and vehicle processing capacity of a booth to predict delays at the border.

Khan (2010) used the VISSIM microscopic simulation software to estimate delays at the Ambassador Bridge. The delays were obtained from repeated runs of a stochastic microsimulation traffic model. Generated delays were then used to train an ANN model to forecast delays on the bridge. The results from the trained ANN model were then compared to the delays produced in VISSIM. With an R^2 greater than 0.97, the ANN model was able to produce comparable results to the ones obtained in VISSIM. More recently, Lin et al. (2014b) used transient multi-server queueing model to predict border crossing delay at the Peace Bridge. A two-step forecasting approach was used in the analysis. Firstly, a multi-model combined forecast method, which was first reported in (Lin et al., 2012), was used to predict likely traffic volume on the bridge. Next, the forecasted volume was used in two queueing models, namely – Erlang service times and exponential inter-arrival times ($M/E_k/n$) queueing model, and Batch Markovian Arrival Process (BMAP) and phase type (PH) services ($BMAP/PH/n$) queueing model, to predict delays in the border. Results from the queueing models were compared to estimates obtained via VISSIM. While the queueing models and VISSIM produced comparable results, the former proved more efficient with respect to the computer run-time requirements.

3. Context and data

3.1. Cross-border economy and the Ambassador Bridge

Ontario's economy is highly dependent on the cross-border transportation links it shares with the US. Ontario ranks fourth among the largest trading partners of the US after Canada (as a whole), Mexico, and China. In 2013, 27% of Ontario's total Gross Domestic Products (GDP) were exported to the US while 39% of its products were imported from the US. The bridges (notably – Ambassador, Blue Water, Peace, and Lewiston–Queenston) connecting Ontario to the US are therefore very important and often referred to as "critical infrastructure" (Anderson, 2012; Ontario Chamber of Commerce, 2004). The primary mode of transportation for trade between the US and Canada is truck. Around 55% of the total Canada–US trade in 2013 was moved by trucks and the percentage is even higher for Ontario's trade with Central US (76%) (Transport Canada, 2013).

Delays in border crossing times on the bridges connecting Ontario to the US can have negative impacts on Ontario's economy. According to the Ontario Chamber of Commerce (2004), border delays cost every Ontarian taxpayer a total of $1100 per year. Also, the overall cost due to border delays is estimated at $13.6 billion for the US and Canada, of which Ontario itself bears 38%. Given the importance of the border to the North American economy, special attention is paid to commercial trucks using the Ambassador Bridge. This bridge is chosen because of its economic significance to the overall Canada–US trade relationship. For instance, this bridge accounts for over one third of the US–Canada truck-borne trade (Haughton and Isotupa, 2012). Also, the bridge is vital for the US and Canadian automotive industries. This is the case since the manufacturing of motor vehicle and its parts comprise 30% of Ontario's exports and 23% of its imports from the US (Industry Canada, 2014). The Ontario Chamber of Commerce (2004) forecasted that Canada may lose over 70,000 jobs by 2030 due to potential delays at the Ambassador Bridge. Moreover, a four hour delay at the Ambassador Bridge will translate in 7 million loss in Ontario's production. It is therefore important to have a modeling system which could provide accurate traffic predictions, especially crossing times, to truck drivers using the bridge to avoid unnecessary delays at the border.

3.2. Cross border truck data

One of the primary datasets used in this study is a large GPS database that was lent to us by Transport Canada. The data were collected by Shaw Tracking, a GPS vehicle tracking company which operates in North America. The GPS database contains GPS pings that were generated from 40,650 individual Canada-based trucks that belong to 750 carriers. The records in the database summarize the movement of the trucks for a full year (September 2012 – August 2013). A subset of these records was taken for this study resulting in approximately 11,000 trucks pertaining to 354 carriers that crossed the Ambassador Bridge during the study period. Along with the truck and its carrier identifiers each GPS ping in the database comes with its geographic location (latitude, longitude), and a time stamp. To ensure privacy, the carrier identifiers were kept anonymous to us. Truck's elapsed and dwell time was calculated using the time stamp of the sequential GPS pings.

While the GPS dataset captures the movement of a large number of trucks on the Ambassador Bridge, it only provides a sample of all the trucks that crossed the bridge when the data were collected. Consequently, the GPS records cannot be used to depict the actual border related truck volumes over time. Therefore, we utilized a total truck volume database that was compiled from a network of RTMS that are operated by the Cross Border Institute (CBI) at the University of Windsor on Canadian side of the bridge. The RTMS data were used to model truck volumes on the bridge while the GPS data were used to model the crossing times on the bridge. The time lag between the two datasets should not be limiting given that the recent monthly trends of truck crossings from the US into Canada, and Ontario have been systematic (Statistics Canada, 2015).

A geo-fence over the bridge, as shown in Fig. 1, was utilized to calculate the border crossing time that each truck required every time it crossed the Ambassador Bridge. This geo-fence, which was delineated by Transport Canada, consists of four zones. Areas within zones 1 and 4 captures the US and Canadian inspection plazas, respectively. The total time it takes a

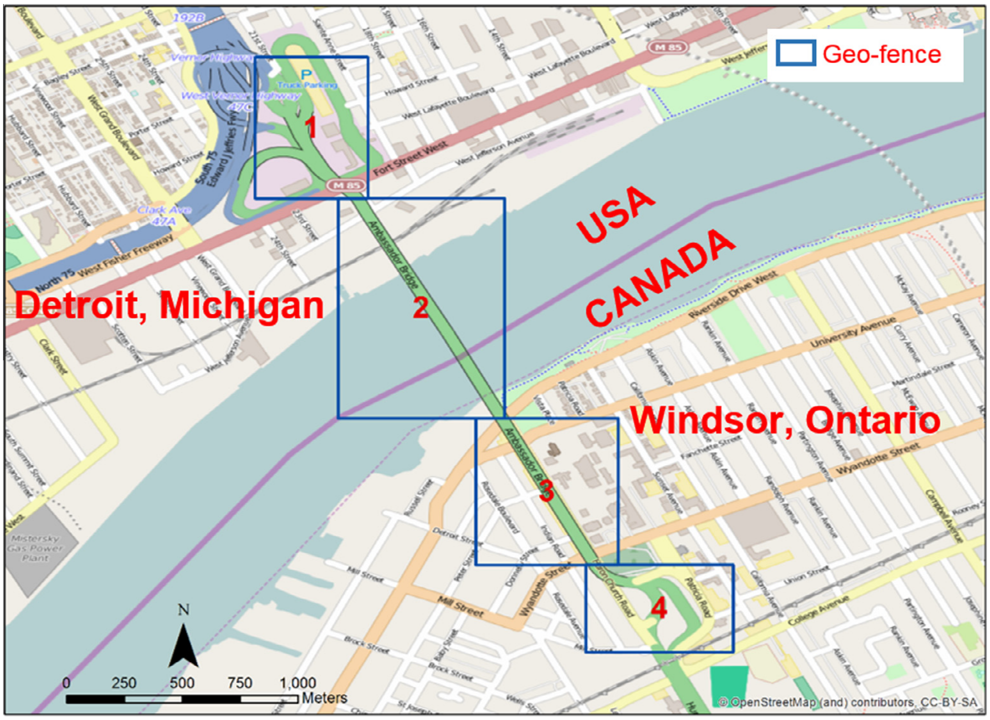

Fig. 1. Geo-fence over the Ambassador Bridge.

truck to cross the geo-fence is composed of three elements: (1) truck's driving time on the bridge, (2) the delay of the truck due to the queue on the bridge, and (3) the inspection time. Given that the secondary inspection facility on the Canadian side is not with the inspection plaza, the inspection time for trucks travelling from the US to Canada includes only the time spends in primary inspection (i.e. in zone 4). By comparison, inspection time for trucks entering the US includes both primary and secondary inspection times.

Exploration of the GPS pings within the geo-fence revealed that trucks returning from the US to Canada frequently stop at the duty free shops located in zone 1 (i.e. trucks will have a dwelling time in that zone). As a result, for Canada-bound trucks, we subtracted the dwelling time observed in zone 1, if any, from the total crossing time to obtain the actual time it took a truck to cross the bridge. Similarly dwelled time in zone 4 was subtracted from the total crossing time for the US bound trucks, albeit the latter did not happen frequently.

4. Methods of analysis

ANN models have a long application history in transportation research and other disciplines. It is one of the most widely used data mining techniques for modeling complex relationships in data. Its advantage over classical statistical models is in its ability to capture the nonlinear relationship between dependent and independent variables without the need for any a *priori* knowledge about the nonlinear relationship. Unlike many data mining techniques, ANN demonstrates superior performance when dealing with both categorical and continuous variables (Francis, 2001). It has been structured in such a fashion that is analogous to the human brain. That is, an ANN processes data within hidden layers in the same way neurons in the human brain do.

A large number of ANN approaches have been reported in the literature including the multilayer perception networks, spectral-based neural networks, radial basis function networks, resource allocating networks, recurrent networks, and wavelet networks. This study utilizes a multilayer feedforward neural network to model crossing times and cross border truck volumes. A multilayer feedforward neural network consists of three layers: input, hidden, and output. Each independent variable in a given layer is represented by a node. Nodes from one layer are connected to all nodes in the following layer and each connection carries some weight associated with it (Fig. 2). The weights reflect the strength of the connections and are updated continuously as described in the following paragraph (Babel and Shinde, 2011; Francis, 2001; Raju et al., 2011).

A backpropagation approach is typically used to estimate the best set of weights for the connections in a neural network. In this approach, predicted values are updated in each iteration, often known as epochs, and errors are computed and compared for successive iterations until no more significant reduction of the errors is possible. This entire process is known as learning or training the network. Interestingly, the training process uses a statistical optimization technique to minimize

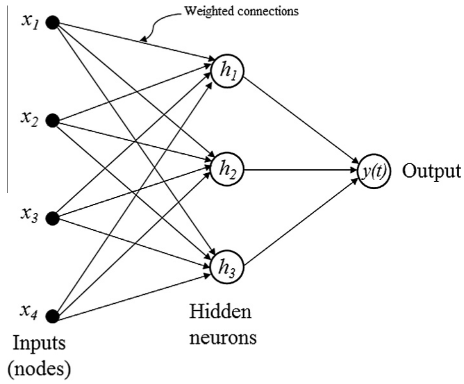

Fig. 2. A multilayer feedforward neural networks with 4 nodes and 3 hidden neurons (Warner and Misra, 1996).

the errors (viz. minimizing the sum of the squared errors). However, because of the nonlinearity in neural networks, there is no closed form solution to the minimization problem (Francis, 2001; Warner and Misra, 1996).

To calculate a weighted sum, a linear equation is formed with the independent variables and their associated weights. The hidden layer receive the weighted sum and, using a transfer function, it processes the sum in the layer. In the hidden layer, different types of transfer function (for instance – log-sigmoid or tan-sigmoid) could be applied to fit data. The output layer then receives the new weighted sum passed through the hidden neurons and, using another transfer function, the output layer recalculates the sum and this process continues.

The Neural Network Toolbox of MATLAB (MATLAB, 2015) was used to train and validate the ANN models in this paper. Furthermore, the Neural Tools of PALISADE (Palisade Corporation, 2015) was used to calculate individual variable impacts used in the training of the network. A tan-sigmoid transfer function with 10 hidden neurons was utilized as this is the recommended function for time series ANN models in MATLAB. However, in the output layer, a purely linear function is used by the toolbox. Different algorithms could be used to solve the minimization problem of the ANN model. MATLAB has three algorithms for handling time series ANN models: Levenberg–Marquardt (LM), Bayesian Regularization, and Scaled Conjugate Gradient. In this study, we used the LM algorithm as it is computationally more efficient and capable of producing comparable results to the other two algorithms used by MATLAB.

Following the approach described above, two ANN models were designed and trained to predict crossing time for traffic crossing the bridge in both directions. Similarly, two more ANN models were designed and trained to predict truck volumes on the bridge per direction. The latter ANN models were needed since the predictions of crossing time is dependent on truck volumes on the bridge. For a given direction, say US bound, the trained models seek to provide a truck heading towards the bridge with a forecast for the time the driver should expect to cross the border when the truck arrives at the bridge. It would be ideal to perform predictions for every minute of a given hour of a given day of the year. However, the minute-to-minute stochastic variability in the crossing time for a yearlong worth of data makes it impractical to train the ANN model at this level of detail, let alone the computational power needed to train the model. Following common practice, every hour of the day was broken into 4 quarters to predict crossing times in a resolution of 15 min. Here, the individual observations (i.e. crossing time) for the whole year were averaged for every quarter by their hours and days. This resulted in a dataset with 672 observations (i.e. 4 quarters $\times$ 24 h $\times$ 7 days). It should be noted that observations were not grouped by months, instead only by hour of day and day of week, due to a small variation in average crossing time across the months for which the data existed. Variability in average border crossing time by hours, days, and months is elaborately discussed in the next section of this paper.

Let $\widehat{V}(t)$ represents the dependent variable $y(t)$ in the ANN model. The training of the network used to predict truck volumes, $\widehat{V}(t)$, relied only on the time lags of the dependent variable (i.e. $V(t-1)$, $V(t-2)$, ..., $V(t-d)$).

$$\widehat{V}(t) = f(V(t-1), V(t-2), \ldots, V(t-d))$$

where d is the number of time lags. In this network, explanatory variables pertaining to hours and days were not used since those did not improve the accuracy of predictions as confirmed by the three evaluation indicators described later in this paper. The number of lags used was decided based on the correlations between dependent variable and its sequential time lags. Since the correlation gradually decreases with no abrupt change, a correlation greater than 0.5 was used as a threshold to limit the number of lags in the model. As such, d was determined to be 6. Nonlinear Autoregressive (NAR) functionality of the Neural Network Toolbox was used to train ANN models for $\widehat{V}(t)$.

Unlike the ANNs for truck volume, along with the lags of crossing time $C(t-d)$, three explanatory variables were used in developing the ANN models for crossing time. The variables (truck volume on the bridge $V(\cdot)$, hours of day $H(\cdot)$, and days of

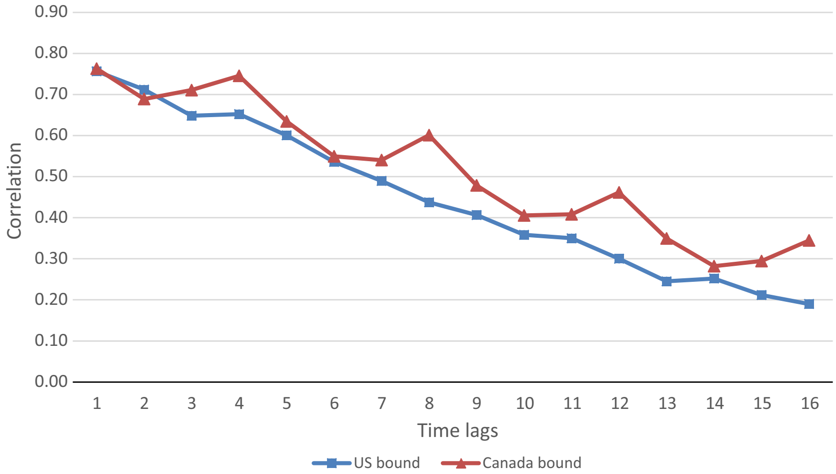

Fig. 3. Correlation between the dependent variable and its time lags.

week $D(\cdot)$) were considered since they significantly improved the accuracy of predictions. The general form of the ANN model for the crossing time $\widehat{C}(t)$ can be expressed as follows:

$$\widehat{C}(t) = f(C(t-1), C(t-2), \ldots, C(t-d), \widehat{V}(t), V(t-1), V(t-2), \ldots, V(t-d), H(t), H(t-1), H(t-2), \ldots,$$
$$H(t-d), D(t), D(t-1), D(t-2), \ldots, D(t-d))$$

The number of time lags d was decided in the same way as it was done in the case of the truck volume models (i.e. based on a correlation test). However, different number of lags was used for Canada and US bound crossing time ANN models. As shown in Fig. 3, the correlation was greater than 0.5 for the first eight time lags in the Canada bound case. In comparison, the 0.5 correlation was maintained for the first six time lags in the US bound case. Therefore, d was set to be 8 for the Canada bound ANN model and 6 for the US.

It is interesting to see the difference in trend between Canada and the US in Fig. 3. While the correlation is gradually decreasing for the US bound trucks the trend oscillates for trucks returning to Canada. Moreover, for Canada bound trucks, there is an interesting trend in the correlations where a peak happens in every fourth time lag and the correlations in between two peaks follow a similar trend. This pattern is indicative of high correlation in hourly truck volume (as four 15 min time lags are equal to 1 h). ANNs for crossing time were also trained in MATLAB. Nonlinear Autoregressive with Exogenous Input (NARX) functionality on MATLAB's Neural Network Toolbox was utilized to train these networks.

The performance of the ANN models was evaluated by means of three indicators: MAPE also known as Mean Relative Percentage Error or Mean Absolute Relative Error, Variance Absolute Percentage Error (VAPE), and Probability of Percentage Error (PPE). These indicators reflect accuracy, stability, and reliability of the obtained predictions, respectively (Zheng et al., 2006). The equations for MAPE, VAPE, and PPE are as follows:

$$MAPE = \frac{1}{N}\sum_{i=0}^{N-1}\left(\frac{\mathrm{abs}(Y-\widehat{Y})}{Y}\right) \times 100$$

$$VAPE = \sqrt{\frac{1}{N(N-1)}\sum_{i=0}^{N-1}\left(\frac{\mathrm{abs}(Y-\widehat{Y})}{Y}\right)^2 - \left[\sum_{i=0}^{N-1}\left(\frac{\mathrm{abs}(Y-\widehat{Y})}{Y}\right)\right]^2} \times 100$$

$$PPE = \Pr((Y-\widehat{Y}) < \pm 10\%) \times 100$$

where N is the number of observations, Y is the observed value and $\widehat{Y}$ is the predicted value.

5. Analysis and results

5.1. Filtering and descriptive analysis of data

Processing of the yearlong GPS pings resulted in 96,536 records on the trucks travelled from Canada to the US and 92,897 on trucks in the opposite direction. The analysis started by examining the trends in crossing time over the 24 h of a day, 7 days of a week, and the 12 months for which the data were provided (Fig. 4). Based on the observed trends, the variability in hourly crossing times is more pronounced when compared to the daily and monthly trends. The trend starts to decrease

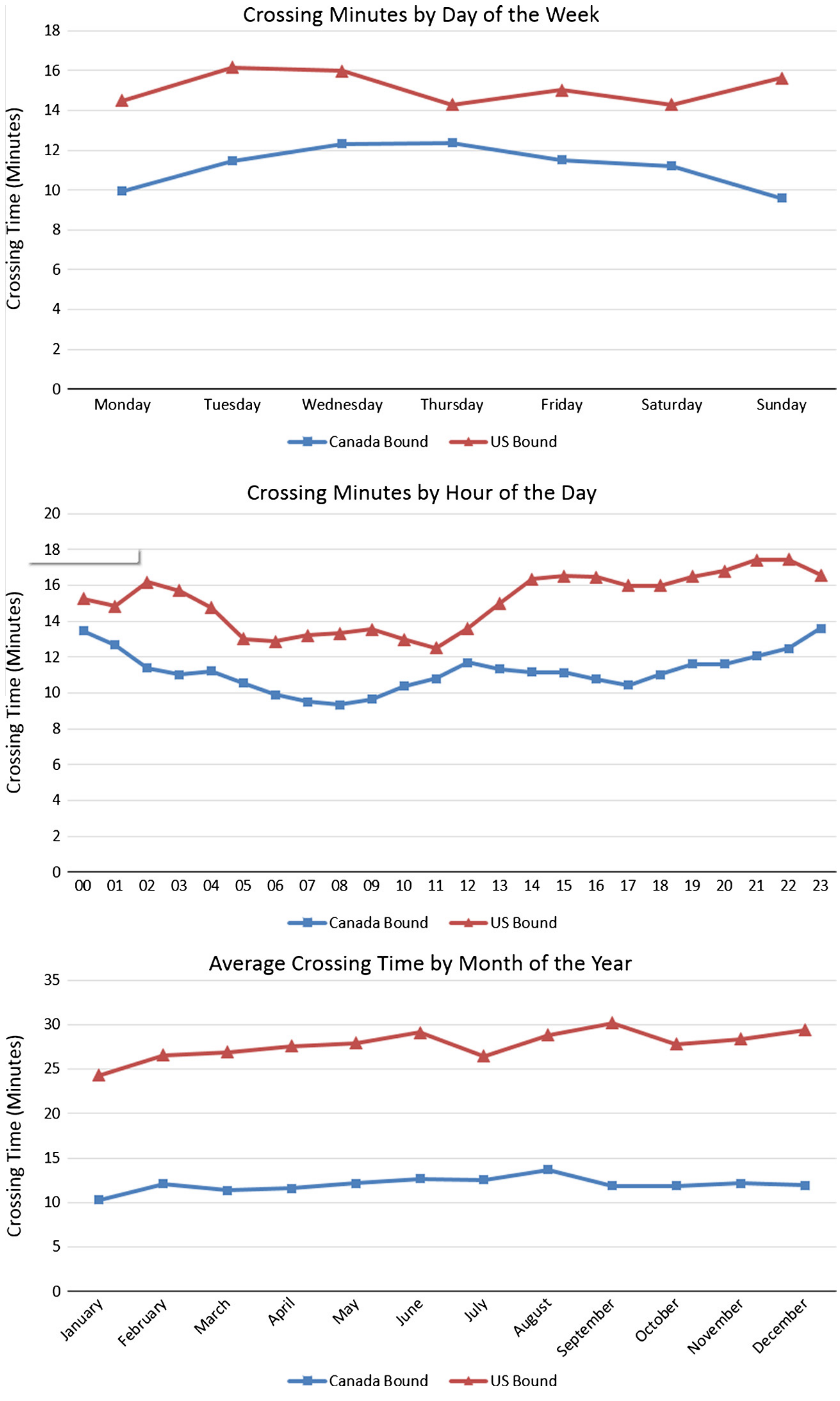

Fig. 4. Crossing time of the trucks by hours, days, and months.

from mid night until 8 am and then gradually increases throughout the day. It is interesting to see that crossing time is greatest, for both Canada and the US bound trucks for the period 10 pm–12 am. This could be due to the fact that there are fewer number of inspection booths that are open during these hours resulting in longer wait time. Crossing time for the US bound trucks is longer on Tuesday and Wednesday and is relatively the same from Thursday to Monday except a peak on Sunday. The trend is however little different for Canada bound trucks. Crossing time starts increasing after Monday and continues to increase until Thursday and it is lowest on Sunday. Nonetheless, the lowest variability in crossing time was observed by months. The crossing time on January and July is lower whereas it higher in September and tends to be the same for the remaining months.

Although there are similarities and dissimilarities in the crossing time trends between the US and Canada bound trucks, the US bound trucks tend to always have higher crossing time than Canada bound trucks. This is likely to be the case due to the fact that, unlike Canada, the US bound crossing time includes time spent by some trucks in secondary inspection. More on that, the average crossing time for all individual records of Canada bound crossings (i.e. 96,536) is 13.5 min whereas, for the US (i.e. 92,897 records), it is 20.1 min.

High standard deviations in crossing time, in clusters of every quarter hour, among the individual crossing records required further investigation to identify the cause. We found that in some cases two or more trucks arriving at the bridge within a same time window, e.g. within 15 min, have significantly different crossing times. This could be due to a number of reasons including the variability in truck's inspection time. It was therefore necessary to clean the initial observations before grouping them in their quarters for training purposes. For cleaning the data, we assumed that multiple trucks arriving in the same 15 min time window should have comparatively same crossing time. Consequently, we calculated the average crossing time of trucks in every 15 min time window and any truck that took longer than 125% of the average of that time window was dropped from the dataset. In case there is only one truck in the time window, the crossing time was compared to the overall average crossing time. While the initial mean and standard deviation among the 672 observations before cleaning the records were 12.73 and 2.06 min for Canada bound trucks and 19.17 and 3.25 min for the trucks returning to the US, respectively, those statistics dropped in light of the data treatment. As such, the mean crossing time of the 672 observations that were generated from the cleaned data dropped to 11.2 min and the standard deviation to 1.97 min for Canada bound trucks, and to 15.11 and 2.40 min, respectively, for the trucks returning to the US.

It should be noted that we tried other data cleaning approaches as well. For instance, crossing time by Industrial Classifications, length of trips, frequency of crossing for same truck over the year were used to treat the raw data. Nonetheless, the correlations between crossing time and industry type or trip distance or frequency were close to zero and the cleaning did not produce any smoothed observations.

5.2. Forecast of crossing time on the Ambassador Bridge

To fit ANN models, MATLAB splits the observations into three groups: training (80%), validation (10%), and testing (10%). The first group of observations is used for training the ANN whereas observations under validation are used to measure

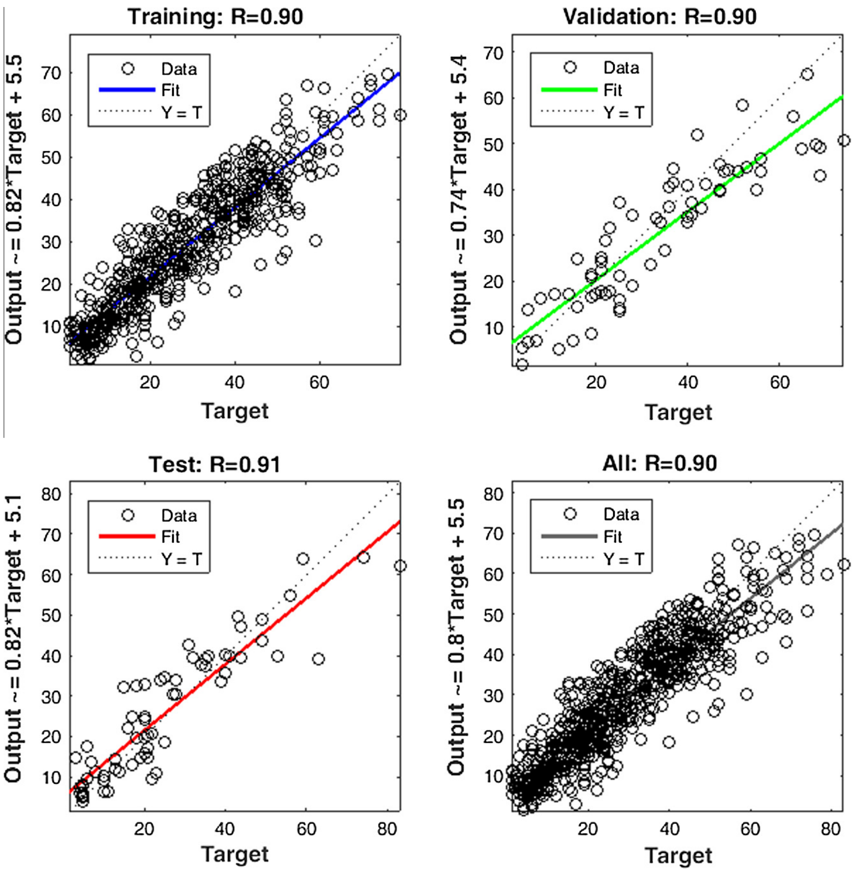

Fig. 5. Linear model fit between observed (target) and predicted (output) volumes for the US bound trucks.

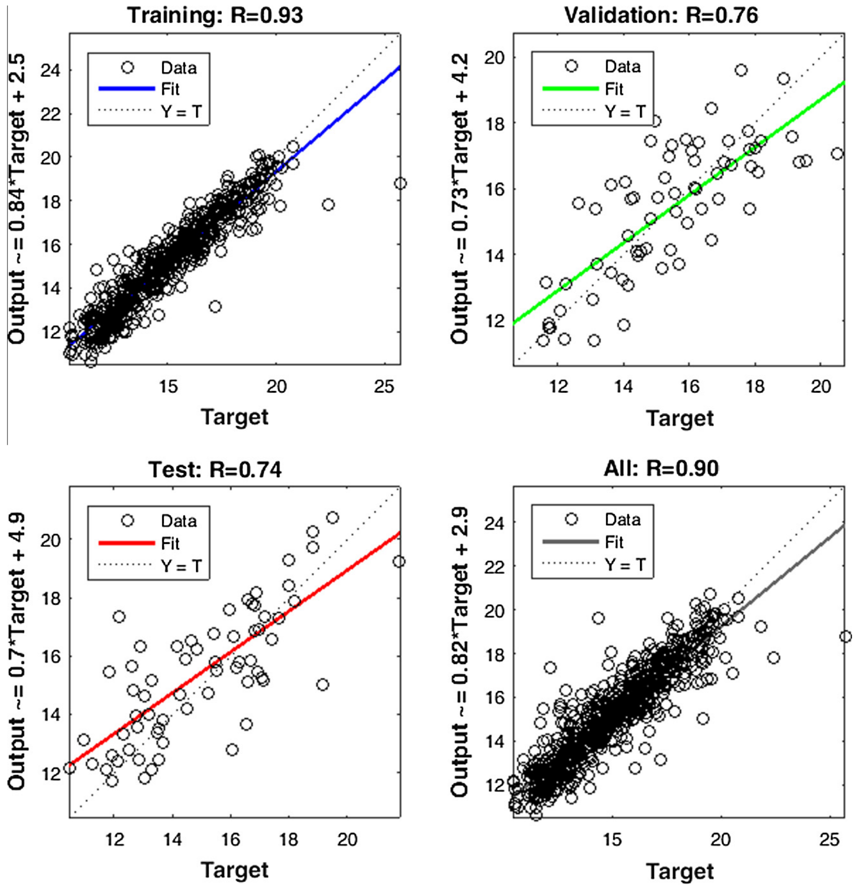

Fig. 6. Linear model fit between observed (target) and predicted (output) crossing time for the US bound trucks.

network generalization. The training of ANN is stopped when no further generalization is possible. Finally the observations under testing are used to provide independent measures of performance for the trained network and have no impact on training the ANN (MATLAB, 2015). Fig. 5 shows the linear fit between actual and predicted border truck volumes for each of the three groups as well as for all observations. High correlations (>0.90) between actual and predicted volumes in all three groups indicate superior predictive capability of the US bound ANN model. Better results were achieved for the Canada bound case with correlations greater than 0.95 for all three data splits (i.e. training, validation, and testing). The diagrams of linear model fits are only reported for the US case for brevity.

Similar to the ANN models for truck volumes, crossing time data were split into three groups: training (80%), validation (10%), and testing (10%). Figs. 6 and 7 show linear model fit between the observed and predicted crossing time for the US and Canada bound trucks, respectively. As can be seen, the correlation between observed and predicted crossing time is very high in the network's training stage (>0.93). The fit of the models is very good as well, with small constants (2.5 and 1.4) and slopes close to 1 (0.84 and 0.87), respectively for the US and Canada bound trucks. However, the correlation and model fit are little smaller for the validation and testing stages of the ANNs. The readers should be aware that retraining the ANNs would change the correlation and fit as the splitting of the data elements into the three groups is done at random. We ran the training of the ANNs multiple times (20 runs) to achieve higher correlations for the training stage. This was done since better trained networks are likely to perform better predictions. The results reported in Figs. 6 and 7 present the results for the best correlation case.

Fig. 8 shows overlay of observed and predicted crossing time for both the US (top) and Canada (bottom) bound trucks. It is evident from the figure that the ANN models captured the trend in the data well and have been able to predict crossing time.

To examine the extent of the errors associated with the predictions, the three evaluation indicators MAPE, VAPE and PPE were used and the results are reported in Table 1. The MAPE represents the percentage deviation of error relative to actual or observed values and hence a lower value is expected for this indicators. The MAPE for the US and Canada bound crossing time in this study is 3.93% and 4.52%, respectively, which are lower compared to values reported in past studies (Li and Rose, 2011; Smith et al., 2002; Vlahogianni et al., 2005; Zheng et al., 2006).

The VAPE characterizes the stability of the predictions and a lower value is expected as well (i.e. the lower value is, the higher stability of the prediction). Unlike MAPE and VAPE, PPE is expected to be higher for better prediction accuracy. The

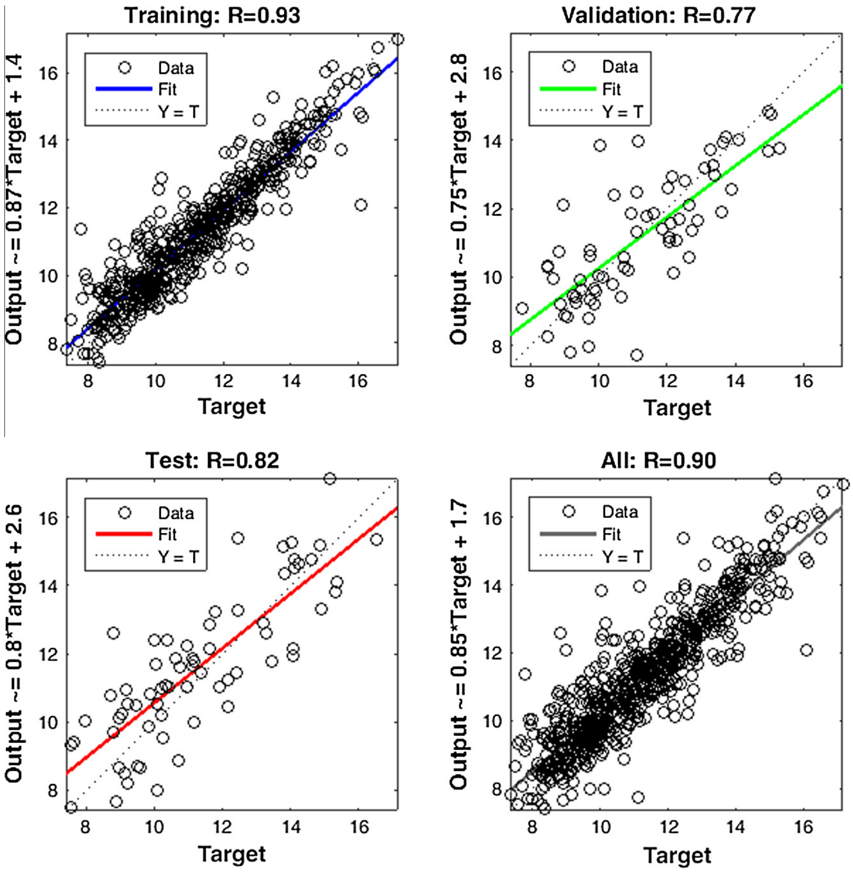

Fig. 7. Linear model fit between observed (target) and predicted (output) crossing time for Canada bound trucks.

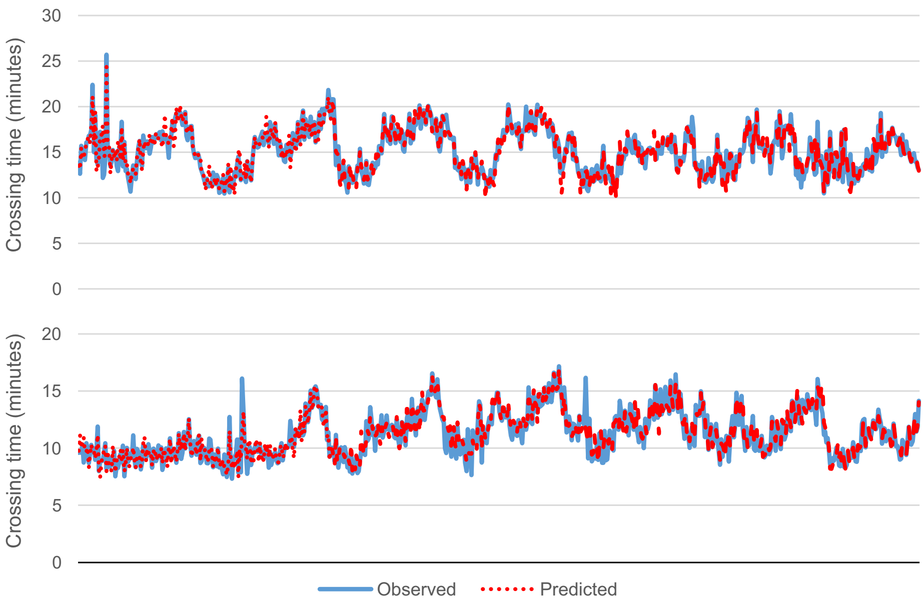

Fig. 8. Observed and predicted crossing time for the US (top) and Canada (bottom) bound crossing time.

Table 1
Evaluation indicators for the US and Canada bound crossing time prediction.

	US bound (%)	Canada bound (%)
MAPE	3.93	4.52
VAPE	6.29	6.63
PPE	90.81	88.55

Table 2
Comparison of mean absolute percentage error across four modeling approaches.

Modeling approach	MAPE (%)	
	US bound	Canada bound
Multilayer feedforward network	3.93	4.52
ARIMAX	7.20	7.18
Multilayer perceptron network	6.66	6.97
Radial basis function network	10.55	10.39

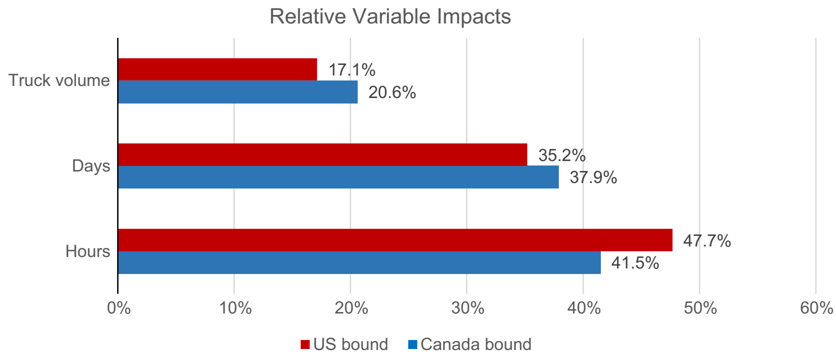

Fig. 9. Relative impacts of independent variables on the US and Canada bound crossing time.

PPE of 90.81% and 88.55%, respectively for the US and Canada bounds, prove that our ANN models are capable of forecasting with high level of accuracy. More specifically, approximately 90% of the predicted crossing times are within ±10% of the actual crossing time.

To compare the prediction accuracy of our crossing time ANN models, we also estimated ARIMAX (ARIMA with Exogenous Inputs) and two other structures of ANN models, namely – multilayer perceptron network and radial basis function network. As can be seen in Table 2, the prediction from the multilayer feedforward ANN model has the lowest MAPE among the modeling approaches utilized in this paper.

Besides the validation of the results, we also estimated the relative impact of the independent variables in predicting the crossing time over the Ambassador Bridge. This was done using the Neural Tools of the PALISADE software since the Neural Network Toolbox of MATLAB did not have this functionality. It should be noted, however, that the two software packages produce comparable results for the trained network. The relative importance of the independent variables is useful as it provides a better understanding of the role that each independent variable play when it comes to training the ANN models. As can be seen in Fig. 9, the hours of day is the most important determinant in predicting the crossing time for both the US and Canada bound trucks. This is followed by the days of week and then the volume of trucks on the bridge. These results are interesting as they suggest that the temporal variation (hours and days) play better role when attempting to predict the crossing time. However, truck volumes remain an important factor that should not be overlooked when performing short term predictions.

6. Conclusions and future studies

Demand for ITS services is rapidly increasing and therefore short-term prediction of traffic characteristics, such as traffic flow, speed, and travel time, are becoming very important for efficient traffic management and operation. While past studies primarily focused on traffic characteristics on freeways or urban arterials this study places particular emphasis on modeling the crossing time over one of the busiest US–Canada bridges, the Ambassador Bridge.

A yearlong GPS database from about 11,000 Canada-based trucks formed the basis for this analysis. Crossing time for trucks that travelled through the bridge was calculated from time stamps of the GPS data. To obtain smoothed crossing times and overcome the minute-to-minute noise in data, the initial observations were averaged by quarters of hour, hours of the day, and days of the week. Multilayer feedforward ANN with backpropagation approach was utilized for the analysis. Predicted crossing times from the ANNs have a high correlation with the observed values. Evaluation indicators further confirmed the high forecasting capability of the trained ANN models. The ANN models from this study could be used for short-term forecasting of crossing time that would support operations of ITS technologies. Since traffic volume proved to be an important factor when training the ANN models for crossing time, another set of ANN models was trained to predict the truck traffic volumes at the bridge. The predicted volumes enable the crossing time ANN models to predict better. The training of the truck volume ANN models was based on a month-long truck volume dataset that was collected via a network of RTMS devices. The predictive ability of the truck volume models is superior to the past studies.

The Ambassador Bridge was used as a proof of concept in this study. However, the approach used is general and should be applicable to any cross-border bridge. Nonetheless, for real-time prediction and/or implementation purposes, actual crossing time data should be continuously collected and fed into the ANN models. Use of advanced technologies, such as automatic license plate readers, on the bridge would provide real-time crossing time. First reader scans a truck's license plate and records the time as soon as the truck arrives at the bridge. A second reader, located at the end of an inspection plaza, records the time when the truck leaves the plaza. Time stamps from the readers could then be used to calculate the time required by the truck to cross the border. The crossing time is automatically sent to a designated system. The system uses the real-time crossing time and retrains the ANN models. The retrained ANN models then provide short-term forecast of crossing time, as we have shown in this prototype study, and send the information to DMS or to an online page that can be read by drivers from anywhere.

Although this study used the VAPE indicator to calculate the variance in absolute errors (i.e. stability of prediction), a separate study is recommended to deal with the variance in the errors and/or, more specifically, truck-to-truck variability in crossing time. A direction for future study would therefore be towards incorporating the variability in prediction of border crossing time.

Acknowledgements

This research is enabled through a FedDev Ontario grant and a Canada Foundation of Innovation Infrastructure Grant. The authors wish to thank Louis-Paul Tardif and Andrew Carter from Transport Canada for lending the GPS data used in this work. We are also thankful to Kevin Gingerich of the Cross-Border Institute for calculating the truck crossing times from the raw GPS data.

References

Anderson, B., 2012. The Border and the Ontario Economy. Cross-Border Transportation Centre, University of Windsor, Windsor, Ontario.
Anderson, W., Maoh, H., Burke, C., 2014. Passenger car flows across the Canada–US border: the effect of 9/11. Transp. Policy 35, 50–56.
Babel, M., Shinde, V., 2011. Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok. Water Resour. Manage. 25 (6), 1653–1676.
Chandra, S., Al-Deek, H., 2009. Prediction of free traffic speeds and volumes using vector autoregressive models. J. Intell. Transp. Syst.: Technol. Plann. Oper. 13 (2), 53–72.
Chen, H., Rakha, H.A., 2014. Real-time travel time prediction using particle filtering with a non-explicit state-transition model. Transp. Res. Part C: Emerg. Technol. 43 (Part 1), 112–126.
Dimitriou, L., Tsekeris, T., Stathopoulos, A., 2008. Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. Transp. Res. Part C: Emerg. Technol. 16 (5), 554–573.
Fei, X., Lu, C., Liu, K., 2011. A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. Transp. Res. Part C: Emerg. Technol. 19 (6), 1306–1318.
Francis, L., 2001. Neural Networks Demystified. Francis Analytics and Actuarial Data Mining Inc., pp. 253–319.
Haughton, M., Isotupa, K.P.S., 2012. Scheduling commercial vehicle queues at a Canada–US border crossing. Transp. Res. Part E: Logist. Transp. Rev. 48 (1), 190–201.
Huang, S., Sadek, A., 2009. A novel forecasting approach inspired by human memory: the example of short-term traffic volume forecasting. Transp. Res. Part C: Emerg. Technol. 17 (5), 510–525.
Industry Canada, 2014. Trade Data Online (TDO).
Ishak, S., Alecsandru, C., 2004. Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. J. Transp. Eng. 130 (4), 452–465.
Khan, A., 2010. Prediction and display of delay at road border crossings. Open Transp. J. 4, 9–22.
Li, R., Rose, G., 2011. Incorporating uncertainty into short-term travel time predictions. Transp. Res. Part C: Emerg. Technol. 19 (6), 1006–1018.
Lin, F.B., Lin, M.W., 2001. Modeling traffic delays at northern New York border crossings. J. Transp. Eng. 127 (6), 540–545.
Lin, L., Wang, Q., Huang, S., Sadek, A.W., 2014a. On-line prediction of border crossing traffic using an enhanced Spinning Network method. Transp. Res. Part C: Emerg. Technol. 43 (Part 1), 158–173.
Lin, L., Wang, Q., Sadek, A., 2012. Evaluating short-term traffic volume forecasting models based on multiple datasets and data diagnosis measures. In: 92nd Annual Meeting of Transportation Research Board, Washington, DC.
Lin, L., Wang, Q., Sadek, A., 2014b. Border crossing delay prediction using transient multi-server queueing models. Transp. Res. Part A: Policy Pract. 64, 65–91.
MATLAB, 2015. 8.5.0.197613 (R2015a) ed. The MathWorks Inc, Natick, Massachusetts, United States.
Ontario Chamber of Commerce, 2004. Cost of Border Delays to Ontario. Ontario Chamber of Commerce.
Palisade Corporation, 2015. NeuralTools. Ithaca, NY.
Paselk, T., Mannering, F., 1994. Use of duration models for predicting vehicular delay at a US/Canadian border crossing. Transportation 21 (3), 249–270.
Qi, Y., Ishak, S., 2014. A Hidden Markov model for short term prediction of traffic conditions on freeways. Transp. Res. Part C: Emerg. Technol. 43 (Part 1), 95–111.
Raju, M., Srivastava, R.K., Bisht, D., Sharma, H.C., Kumar, A., 2011. Development of artificial neural-network-based models for the simulation of spring discharge. Adv. Artif. Intell. 2011, 11.

Smith, B.L., Williams, B.M., Oswald, R.K., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. Transp. Res. Part C: Emerg. Technol. 10, 303–321.
Statistics Canada, 2015. Number of vehicles travelling between Canada and the United States, monthly (number), CANSIM. Statistics Canada.
Transport Canada, 2013. Transport Canada Statistical Addendum. Transport Canada.
Vlahogianni, E., Karlaftis, M., Golias, J., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transp. Res. Part C: Emerg. Technol. 13 (3), 211–234.
Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting: overview of objectives and methods. Transp. Rev.: Trans. Transdiscipl. J. 24 (5), 533–557.
Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: where we are and where we're going. Transp. Res. Part C: Emerg. Technol. 43 (Part 1), 3–19.
Wang, Y., Papageorgiou, M., 2007. Real-time freeway traffic state estimation based on extended Kalman filter: a case study. Transp. Sci. 41 (2), 167–181.
Warner, B., Misra, M., 1996. Understanding neural networks as statistical tools. Am. Stat. 50 (4), 284–293.
Williams, B.M., Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. J. Transp. Eng. 129 (6), 664–672.
Zhang, Y., Zhang, Y., Haghani, A., 2014. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. Transp. Res. Part C: Emerg. Technol. 43 (Part 1), 65–78.
Zheng, W., Der-Horng, L., Shi, Q., 2006. Short-term freeway traffic flow prediction: Bayesian combined neural network approach. J. Transp. Eng. 132 (2), 114–121.