Arab J Sci Eng (2015) 40:303–312
DOI 10.1007/s13369-014-1438-3

RESEARCH ARTICLE - CIVIL ENGINEERING

Impact of Stock Market Indices and Other Regional Exogenous Factors on Predictive Modeling of Border Traffic with Neural Network Models

El-Sayed M. El-Alfy · Nedal T. Ratrout · Uneb Gazder

Received: 2 April 2014 / Accepted: 31 August 2014 / Published online: 28 November 2014

Abstract This paper analyzes the impact of stock market indices, as indicators of political and economic stability, and other regional exogenous factors on the performance of predictive modeling of border traffic using neural network models. To prove the concept, the Saudi–Bahrain corridor through King Fahd causeway is selected as our area of study. These two countries have strong cultural ties and a wide variety of variables affects the incoming and outgoing traffic flows. Various models of artificial neural networks are constructed for different prediction horizons and look-back periods using a dataset prepared for the period from 2003 till 2013. In our study, stock market indices are proposed, for the first time, to be used in border traffic forecasting. These indices are added as a surrogate measure of the political and economic conditions of the countries which are under study. Their effects on models with varying ranges of time-series inputs and different prediction horizons are studied in detail. It is found that including stock market indices and other most relevant local factors has generally improved the prediction performance of the neural network models in all cases. Additional reduction in the prediction error is achieved by the proposed ensemble model trained with different time lags. Yet, the degree of improvement depends on the look-ahead horizon for prediction.

Keywords Intelligent transportation system · Predictive traffic modeling · Time-series forecasting · Neural networks and ensembles · Multivariate border traffic analysis

1 Introduction

Modeling and prediction of travel behavior pattern, traffic flow volume, and speed variability are crucial components for intelligent transportation systems and road network planning. Several approaches have been proposed and investigated in the literature for short-term and long-term scenarios under normal and abnormal conditions [1]. Most of the approaches focus on traffic within a city and/or a single-variable time-series analysis. Statistical regression methods have been one of the early tools that have been applied to this research area. For example, Moorthy and Ratcliffe [2] investigated the application of time-series analysis for short-term traffic count forecasting. Mai et al. [3] used additive seasonal vector auto-regressive moving average to predict short-term traffic flow considering spatial dependency among multiple sites in the city center of Dublin, Ireland. Guo et al. [4] described a two-stage short-term traffic prediction structure and used traffic flow from a corridor in Central London to calibrate and evaluate their proposed method.

Due to the modeling complexity involved in this field, especially when the number of variables increases dramatically, machine-learning paradigms, such as artificial neural networks and support vector machines, attracted the attention of several researchers. They have the ability to approximate linear and nonlinear relationships with a very high accuracy and to work from historical data without prior knowledge

E.-S. M. El-Alfy is on leave from the College of Engineering, Tanta University, Egypt.

E.-S. M. El-Alfy (✉)
College of Computer Sciences and Engineering,
King Fahd University of Petroleum and Minerals,
Dhahran 31261, Saudi Arabia
e-mail: alfy@kfupm.edu.sa

N. T. Ratrout · U. Gazder
Department of Civil Engineering, King Fahd University
of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
e-mail: nratrout@kfupm.edu.sa

U. Gazder
e-mail: unebgazdar@gmail.com

about the form of the input–output relationship. Karlaftis and Vlahogianni [5] presented a review of the differences, similarities and insights of statistical methods as compared to neural networks in transportation research. Combining neural networks with fuzzy logic, Yin et al. [6] described an approach to predict the traffic flows in an urban street network. In [7], an adaptive hybrid fuzzy rule-based system is presented for modeling short-term traffic forecasting in urban arterial networks. It was demonstrated that better results can be achieved than using conventional statistical methods. Zhang and Liu [8] presented a method for time-series traffic forecasting based on a least square support vector machine (LS-SVM) which applies linear least square criteria to the loss function instead of quadratic programming [9].

The assessment of some exogenous factors, such as meteorological measurements, traffic mix, and socioeconomic attributes, has been addressed in the context of travel mode changes. For instance, weather conditions has been considered for active mode trips within the city of Toronto using multinomial logit modeling [10] and for pedestrian counts in the downtown of Montpelier, Vermont [11]. In another study, the travel mode choice probabilities were quantified using cluster analysis of weather variables in Sydney [12]. For traffic speed forecasting, Tsirigotis et al. [13] evaluated the effects of weather and traffic mix over Attica Tollway, in Athens Greece, using different auto-regressive integrated moving average (ARIMA) models. Some researchers have explored similar problems for other types of traffic. For instance, Dougherty and Cobbett [14] used back-propagation neural networks for inter-urban short-term traffic forecasts in the Netherlands. Lingras and Mountford [15] used a genetic algorithm optimized time-delay neural network to estimate the inter-city traffic volume along a strategic highway between Calgary and Edmonton in Alberta.

For border traffic modeling, the underlying process is more complex and can involve a huge number of variables. The best set of variables to be used in border transport modeling is a burning question for the planners because these variables can give an insight on the travelers' behavior which can in turn change the policy decisions. These decisions may include provision of alternative modes, route and fare changes, and frequency. Political and economic situation of the neighboring countries often affects the trips generated or attracted across their borders. Political turmoil in any country degrades its economy and security conditions which in turn reduce its attraction of recreational as well as commercial trips from other countries. Stock market indices tend to fluctuate with the economic and political conditions of the country and can be used as a useful surrogate indicator of its overall stability. This study is the first to propose using stock market indices in boarder traffic forecasting as an indicator of political stability and economic prosperity of the countries under consideration. We chose the corridor of King Fahd causeway, which provides a land access between Saudi Arabia and the island of Bahrain, as the field of our study. We evaluated the impact of several factors on border traffic prediction, including variability of the stock market indices, weather conditions, and religious-related events and vacations. We built a variety of neural network models for predicting the traffic with different time lags and prediction horizons. We also proposed and evaluated ensembles of neural networks trained with different time lags.

The rest of the paper is organized as follows. Section 2 describes the adopted methodology for collecting and preparing data and for building predictive models with neural networks and ensembles. Section 3 describes the evaluation measures and experimental work and discusses the results. Finally, Sect. 4 concludes the paper and stresses upon the major outcomes of this study and future work.

2 Methodology

2.1 Data Collection and Analysis

2.1.1 Background

The border traffic data have been collected along King Fahd causeway, which is a vital land connection for travelers between Saudi Arabia and Bahrain; see Fig. 1. The total length of the causeway is about 25 km and has two lanes in each direction of a total width of about 25 m. These two countries have strong social, political, and economic ties which encourage their residents to travel more frequently between them. There are two modes of transportation operating over the causeway: private vehicles and public transport. In 2012, the average number of vehicles crossing the causeway was more than 22,400 per day in both directions [16].

This causeway is not only used by Saudi and Bahrain travelers, but it is also the only land road connecting Bahrain to other Gulf countries. Since there is no railway connection between Bahrain and Saudi Arabia, the air and the causeway links are very important for travelers to and from Bahrain. All land travelers from the Gulf countries go to Bahrain through this causeway passing Saudi Arabia. Bahrain airport also acts as a hub for connecting international flights to many countries of the world. So, the people from the eastern region in Saudi Arabia go to Bahrain through this causeway for catching flights to their destinations. This is one of the reasons that buses are being frequently operated on an hourly basis between Bahrain Airport in Manama, the capital of Bahrain, and Dammam and Al-Khobar cities in the eastern region of Saudi Arabia. The Gulf airlines, which is the national airline of Bahrain, also provides its bus service for the airline passengers between the airport and the Saudi eastern region.

Fig. 1 Location of KFC on the map and its view from the east (N. S. Brahmavar (http://creativecommons.org/licenses/by/3.0)

Considering the fact that the study area is across the borders of two countries, economic prosperity and political stability of the countries also become important factors affecting travel demand. We included stock market indices in both countries as related indicators. The Saudi Stock Exchange, also known as Tadawul, is the only stock exchange in Saudi Arabia. It is supervised by a government organization called Capital Market Authority. In 1984, a Ministerial Committee composed of the Ministry of Finance and National Economy, Ministry of Commerce, and Saudi Arabian Monetary Agency (SAMA) was formed to regulate and develop the market. 156 publicly traded companies were listed (as of September 2, 2012). In Bahrain, the Bahrain Bourse (BHB) was established as a shareholding company for the year 2010 to replace Bahrain Stock Exchange (BSE) which was established in 1987. Currently, 50 companies are listed on the exchange. The BSE operates as an autonomous institution supervised by an independent Board of Directors, chaired by the Governor of the Central Bank of Bahrain.

2.1.2 Dataset Description

In this research, a recent dataset is prepared and utilized for the period from 2003 till 21st of October 2013. There are a total of 19 variables. A description of all the variables included in the dataset is given in Table 1, and the statistics for the numerical variables are presented in Table 2.

In addition to the day and name of the month and the name of weekday, the variables include daily traffic counts of incoming (entering Saudi Arabia) and outgoing (exiting Saudi Arabia) vehicles. They also include for the same period weather-related variables (average daily temperature and humidity) and the daily average stock market indices in both countries. The air-travel data give the number of flights and passengers traveling daily between King Fahd International Airport and Bahrain International Airport. This has been included as an alternate mode for travelers passing the causeway which may therefore be interrelated with the causeway travel demand.

Other included variables are related to holidays and vacations in this region. For example, the dummy variables Hajj and Ramadan are included because they represent dates announced as annual holidays in both countries. Similarly in summer vacations, educational institutes especially schools are closed. These variables were included based upon the assumption that travel demand will be affected by weekly or yearly holidays and some unique pattern can be observed in these periods. Moreover, the dummy variable S represents the period in which salaries are distributed to government employees in KSA at the end of each Islamic month which does not always coincide with the end of the Gregorian month. The purpose of including this variable was to test whether or not people would like to travel more when they receive their salaries for activities like shopping and recreation.

Table 1 Summary of the variables in the dataset

Variable	Description
S	A dummy variable indicating the period from 24th to 1st of each Islamic month where salaries are disbursed (0 or 1)
Ramadan	A dummy variable indicating period from 24th of the 9th Islamic month to the 7th of the 10th Islamic month (0 or 1)
Hajj	A dummy variable indicating period from the 1st to the 14th of the 12th Islamic month (0 or 1)
Month	Gregorian month labeled by integers (1–12)
Vacation	A dummy variable indicating months of summer vacations for June and July (0 or 1)
Day	Day of the Gregorian month (1–31)
Weekday	Name of the weekday labeled by integers (1–7)
SA_DIndex	Daily index of the Saudi stock market
SA_Temp	Daily temperature recorded in Dhahran (KSA) in °C
SA_Humidity	Percentage daily humidity recorded in Dhahran (KSA)
BH_Temp	Daily temperature recorded at Bahrain International Airport in °C
BH_Humidity	Percentage daily humidity recorded at Bahrain International Airport
In_Traffic	Total incoming traffic, i.e., the total number of vehicles per day entering KSA through King Fahd causeway
Out_Traffic	Total outgoing traffic, i.e., the total number of vehicles per day exiting KSA through King Fahd causeway
BH_DIndex	Daily index of the stock market in Bahrain
Arr_Flights	Total number of flights per day arriving at King Fahd International Airport from Bahrain International Airport
Arr_Passengers	Total number of passengers per day arriving at King Fahd International Airport from Bahrain International Airport
Dep_Flights	Total number of flights per day departing from King Fahd International Airport to Bahrain International Airport
Dep_Passengers	Total number of passengers per day departing from King Fahd International Airport to Bahrain International Airport

Table 2 Statistics of the numerical variables in the dataset

Variable	Mean	Minimum	Maximum	SD
In_Traffic	8,434	754	17,100	2,403.55
Out_Traffic	10,088	1,024	27,285	3,470.33
BH_Temp	27.4	10.3	39.6	6.82
BH_Humidity	56.1	14	95	12.86
SA_Temp	27.5	7.8	40.7	7.87
SA_Humidity	46.85	10	96	17.53
SA_DIndex	7,656.79	2,500	19,600	3,158.72
BH_DIndex	1,704.38	1,035.3	2,902.68	526.16
Arr_Flights	2	0	14	1.43
Arr_Passengers	190	0	1,402	141.2
Dep_Flights	2	0	16	1.52
Dep_Passengers	181	0	1,968	147.77

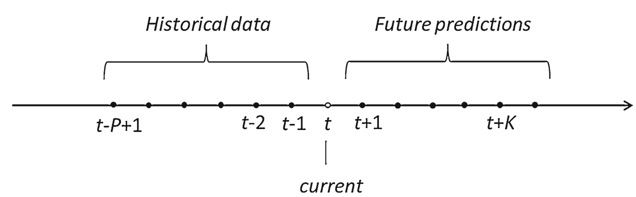

Fig. 2 Using data at current and past instances to predict the future

2.2 Predictive Traffic Modeling

Given historical data for traffic flow and other variables over a period of time, it can be used to construct models to predict the traffic in the future. Mathematically, each variable is denoted as a time series $\{v_1, v_2, \ldots, v_t, \ldots, v_N\}$. At time instant t, it is required to use the measurements for the current measurements as well as the past measurements at $t-1, t-2, \ldots, t-P+1$, where P is called lag period or window size, to predict the traffic at a future instant, $t+K$, where K is the look-ahead period; see Fig. 2.

In our study, we constructed three categories of models: (1) predictive models for the incoming traffic flow by considering only the traffic flow as a time series, (2) predictive models that take into consideration the incoming traffic flow as well as the stock market indices in both countries, and (3) predictive models that uses all the most relevant variables in addition to the incoming traffic flow. To demonstrate the strength of the relationship between various variables and the travel demand on the corridor under study, a correlation analysis between these variables and the vehicle traffic counts is conducted.

In each case, the selected variables were then used to build various artificial neural network models for daily, weekly, monthly, quarterly, and yearly prediction horizons. Moreover, we built more accurate models using ensemble of seven components of the same type of artificial neural networks (ANNs) but with different look-back periods ranging from 1 to 7 days. For training and testing of each case, we used tenfold stratified cross validation.

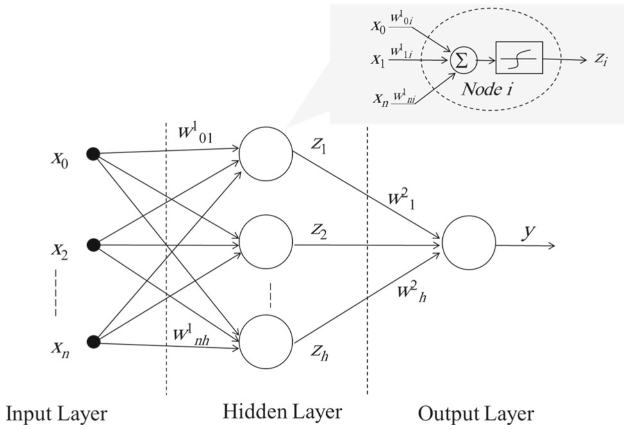

Fig. 3 A typical architecture of MLP neural network model

2.2.1 Neural Network Models

Neural networks are known for their attractive capability of function approximation with high accuracy. A typical example of a neural network structure is shown in Fig. 3, a feed-forward neural network or multilayer perceptron (MLP). It has a set of layers; each consists of a set of processing units (neurons). The output from each layer is fed as input to the subsequent layer. The first layer is called input layer and is composed of a number of neurons equal to the number of input variables. The last layer is called output layer and it generates an approximate value for each dependent variable or variables. Other layers that may exist between the input and the output layers are called hidden layers. The links connecting neurons from one to the next layers have weights and each neuron has an activation function. Using a training dataset and a learning algorithm, the optimal link weights (network parameters) can be determined.

For $n-h-1$ MLP, there are n inputs at the input layer, h neurons in the hidden layer, and one neuron in the output layer. Referring to Fig. 3, the output is calculated as a nonlinear function of the inputs as follows:

$$y = \sum_{j=0}^{h} f\left(w_j^{(2)} z_j - b\right) \tag{1}$$

$$z_j = \sum_{i=0}^{n} f\left(w_{ij}^{(1)} x_i - b_j\right) \tag{2}$$

where $x_0 = 1, z_0 = 1, b$ and b_j are biases, and $f(\cdot)$ is an activation function. The network parameters, a.k.a. weights and biases, are determined using a learning algorithm such as back propagation or conjugate gradient learning algorithm.

2.2.2 Ensemble of Neural Networks

A more intelligent approach for using neural networks for traffic prediction is to use ensembles instead of using a single-network model. This approach relieves the problems pertinent to a single network such as falling into local minima and determining the number of neural cells in hidden layers [17,18]. An ensemble model combines the outputs of a group of neural network predictors. The straightforward approach to combine the outputs from various components in the first layer is to compute their average value. The average can be a simple average or a weighted average. The component models of an ensemble can differ in their training algorithms, input parameters, or network architecture.

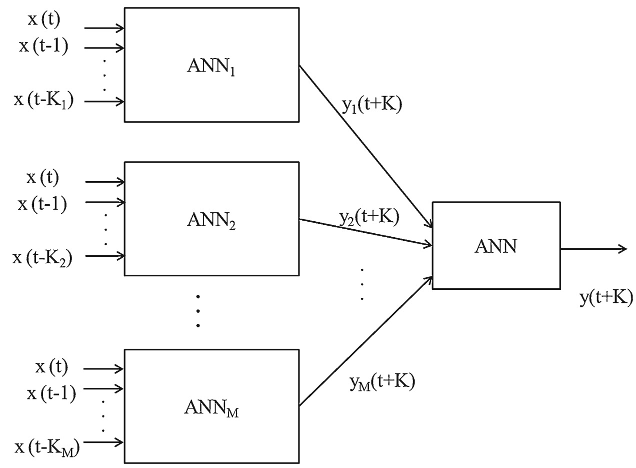

Fig. 4 Ensemble model for traffic prediction combining a number of neural networks trained with different time lags

In our study, we developed seven neural network models in the first layer of the ensemble trained using different time lags, $K_1 = 0, K_2 = 1, \ldots,$ and $K_7 = 6$. Then, we built another neural network, ANN, at the output layer of the ensemble to combine the outputs from the networks in the first layer. The structure of this proposed model is shown in Fig. 4.

3 Experimental Evaluation

3.1 Evaluation Measures

In order to assess and compare various predictive models, two performance measures were used. These measures are the root mean square relative error (RMSE) and the mean absolute percentage error (MAPE) [19,20]. The smaller the values of these measures are, the better the performance is. The definitions of these measures are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum\nolimits_{t=1}^{N} \left(\frac{y_t - \hat{y}_t}{y_t}\right)^2} \tag{3}$$

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \frac{|y_t - \hat{y}_t|}{y_t} \tag{4}$$

Table 3 Pearson's correlation and its significance at 5 % probability

Variable	Correlation	Significance
S	0.05	Yes
Hajj	0.08	Yes
Summer	0.06	Yes
Weekday	−0.33	Yes
SA_Humidity	−0.01	No
BH_Temp	0.02	No
Out_Traffic	0.63	Yes
Arr_Flights	0.33	Yes
Dep_Flights	0.33	Yes
Ramadan	−0.03	No
Month	0.00	No
Day	0.00	No
SA_Temp	0.03	No
SA_DIndex	0.12	Yes
BH_Humidity	−0.12	Yes
BH_DIndex	0.13	Yes
Arr_Passengers	0.32	Yes
Dep_Passengers	0.30	Yes

No, not significant and not selected for modeling; Yes, significant and selected for modeling

where N is the number of time instants, and y_t and $\hat{y}_t$ are the actual and predicted traffic flow values at time instant t, respectively.

3.2 Experiments and Results

3.2.1 Correlation Analysis

We started by a correlation analysis for all the variables mentioned previously with the daily incoming traffic count. The reason for this analysis was to identify the most correlated factors to the travel behavior on the causeway, which can subsequently help construct more effective and simplified models. For this purpose, we used the Pearson's product-moment correlation coefficient (PMCC), which is denoted as r and is calculated as follows [21]:

$$r = \frac{\sum_{t=1}^{N}(x_t - \mu_x)(y_t - \mu_y)}{\sqrt{\sum_{t=1}^{N}(x_t - \mu_x)^2}\sqrt{\sum_{t=1}^{N}\left(y_t - \mu_y\right)^2}} \tag{5}$$

where μ_x and μ_y are the mean values of the variables x and y for which correlation is to be calculated and $r \in (1.0, -1.0)$. When $|r| = 1$, it means perfect correlation (dependence). The closer the value of $|r|$ to 1, the stronger is the relationship between the two variables.

The results of the correlation analysis are shown in Table 3. For the selection of the most relevant variables, a significance level of 5 % probability was set for the coefficient similar to [22,23]. It can be observed from these results that outgoing traffic count and air-travel passenger and flight information are the most significant variables. Apart from that, the name of weekdays and vacation periods (both religious and summer) have significant effect on travel demand which supports the assumption that travel demand changes on weekly and yearly holidays. Stock market indices also show significant correlation with the traffic volume which enforces the assumption that they can be useful in predictive traffic modeling. On the other hand, the variable S, indicating salary disbursement period, is of less significance. Using the results of the correlation analysis, 12 variables are indicated as correlated to the incoming traffic flow and will be used in predictive modeling.

3.2.2 Modeling of Incoming Traffic Flow

Using the StatSoft Statistica Software, we conducted several modeling experiments. In the first experiment, we used only the incoming traffic flow variable as a time series in modeling. We constructed different multilayer perceptrons (MLPs) and ensembles of MLPs using tenfold stratified cross validation for five prediction horizons and seven time lag steps. We used 3-layer perceptrons consisting of input layer, hidden layer, and output layer. In each case, we tested differ-

Table 4 Best configurations for various P values of the MLPs and ensembles in terms of number of hidden neurons and type of activation functions

P	Traffic		Traffic + stock		Al	
	#Hidden	Activation	#Hidden	Activation	#Hidden	Activation
1	5	Hyperbolic	3	Hyperbolic	7	Hyperbolic
2	4	Logistic	7	Logistic	7	Logistic
3	4	Hyperbolic	9	Hyperbolic	5	Hyperbolic
4	4	Hyperbolic	11	Hyperbolic	5	Hyperbolic
5	6	Logistic	8	Hyperbolic	6	Hyperbolic
6	8	Hyperbolic	9	Hyperbolic	8	Hyperbolic
7	7	Hyperbolic	8	Hyperbolic	16	Logistic
Ensemble	9	Logistic	9	Logistic	9	Logistic

ent number of neurons in the hidden layer and two types of activation functions (logistic and hyperbolic). We also ran experiments using two-hidden layers. Due to the limitations on the space, we only report the best model configuration, in terms of number of hidden neurons and type of activation functions used, for each case in the second and third columns of Table 4. We considered various time window sizes (time lag of 1 day, 2 days, 3 days, …, and 7 days). For each time lag, as indicated in the first column of Table 4 by the variable P, the same configuration is used for predicting the traffic at different time horizons (1-day, 1-week, 1-month, quarter-year, and 1-year). The corresponding perfor-

Table 5 Incoming traffic flow prediction using MLPs and ensembles with best configurations for various P and K values

K	P	Traffic		Traffic + stock		All	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
1	1	0.05	0.134	0.048	0.136	0.018	0.074
	2	0.048	0.133	0.043	0.129	0.017	0.074
	3	0.033	0.111	0.033	0.116	0.017	0.072
	4	0.036	0.109	0.033	0.113	0.017	0.075
	5	0.031	0.102	0.031	0.108	0.017	0.074
	6	0.032	0.1	0.03	0.105	0.017	0.073
	7	0.019	0.076	0.018	0.079	0.017	0.073
	Ensemble	0.0177	0.074	0.019	0.075	0.012	0.061
7	1	0.1	0.123	0.096	0.123	0.095	0.119
	2	0.097	0.12	0.096	0.12	0.09	0.111
	3	0.097	0.119	0.088	0.119	0.085	0.122
	4	0.095	0.118	0.088	0.12	0.082	0.117
	5	0.097	0.122	0.088	0.122	0.08	0.116
	6	0.092	0.121	0.087	0.12	0.075	0.112
	7	0.092	0.118	0.075	0.119	0.075	0.119
	Ensemble	0.09	0.117	0.066	0.114	0.06	0.103
30	1	0.193	0.22	0.189	0.212	0.169	0.16
	2	0.174	0.213	0.168	0.207	0.144	0.152
	3	0.172	0.163	0.167	0.199	0.14	0.153
	4	0.173	0.154	0.162	0.2	0.138	0.158
	5	0.174	0.212	0.161	0.196	0.135	0.153
	6	0.177	0.184	0.16	0.176	0.154	0.16
	7	0.171	0.178	0.151	0.165	0.143	0.146
	Ensemble	0.134	0.171	0.13	0.1511	0.129	0.13
90	1	0.193	0.22	0.189	0.212	0.169	0.16
	2	0.174	0.214	0.174	0.177	0.167	0.151
	3	0.176	0.186	0.17	0.175	0.16	0.161
	4	0.175	0.186	0.166	0.173	0.158	0.159
	5	0.175	0.212	0.163	0.175	0.155	0.155
	6	0.177	0.184	0.16	0.176	0.154	0.16
	7	0.172	0.181	0.157	0.173	0.15	0.16
	Ensemble	0.165	0.18	0.151	0.162	0.15	0.15
365	1	0.209	0.222	0.21	0.2	0.171	0.149
	2	0.208	0.221	0.18	0.185	0.169	0.151
	3	0.204	0.189	0.18	0.184	0.168	0.15
	4	0.204	0.186	0.168	0.176	0.165	0.151
	5	0.199	0.215	0.181	0.177	0.162	0.151
	6	0.19	0.196	0.18	0.169	0.162	0.155
	7	0.189	0.164	0.164	0.159	0.161	0.15
	Ensemble	0.188	0.158	0.148	0.147	0.145	0.132

mance measures are shown in the third and fourth columns of Table 5.

To explore the impact of stock market indices on the prediction of incoming traffic flow, we repeated the experiment using multilayer perceptrons (MLPs) and ensembles of MLPs. Again, different models have been constructed using tenfold cross validation for various forward and backward steps. The best configurations are shown in the fourth and fifth columns of Table 4, and the corresponding performance measures are shown in the fifth and sixth columns of Table 5.

In a third experiment, we explored the impact of stock market indices on the prediction of incoming traffic flow. We used all the most relevant variables, as indicated by the correlation analysis in Table 3, and repeated the experiment using multilayer perceptrons (MLPs) and ensembles of MLPs. Again, different models have been constructed using tenfold cross validation for various forward and backward steps. For this final case, the best configurations are shown in the sixth and seventh columns of Table 4, and the corresponding performance measures are shown in the seventh and eighth columns of Table 5.

3.2.3 Discussion of the Results

As a preliminary motivation for including stock market indices in border traffic prediction, we observed a growth trend in traffic and the Bahrain stock market index over the same time period. A typical chart is shown in Fig. 5 for the historical data during the month of January 2003. This observation provides some justification for our conjecture that a relationship is found between the variations of stock market and the border traffic volume. This finding is also confirmed by two other results. The first result is the statistically significant correlation between the daily stock market indices and the outgoing traffic volume as shown in Table 3. The second result is the improvement in the prediction accuracy due to the inclusion of stock market indices as show in Table 5. The relationship between stock market indices on one hand and political stability and economic prosperity on the other hand is intuitive. Economic prosperity is an indirect measure of the infrastructure integrity and tourism services of the attraction countries. It is also an indirect measure of the travelers' expenditure power on trips. Similarly, political instability, especially when it is life-threatening, is expected to affect trip generation. Therefore, it can be claimed that political stability and economic prosperity affect both border traffic generation and stock markets. Since stock market indices are readily available, they can be used as surrogate measures of political stability and economic prosperity which are difficult to quantify and measure in this area of study.

Two more findings are observed as follows. When all most relevant variables, as indicated from the correlation analysis in Sect. 3, are taken into consideration, superior performance is achieved as demonstrated in the last two columns in Table 5. Moreover, the ensemble model resulted in more reduction in the error for each case. To illustrate the percentage improvement, we considered the results for the case with only the traffic flow as a base case. Mathematically, this improvement is given by,

$$\text{Improvement} = \frac{\text{RMSE}_{\text{base}} - \text{RMSE}_x}{\text{RMSE}_{\text{base}}} \times 100\,\% \tag{6}$$

where $\text{RMSE}_{\text{base}}$ refers to the base case when only the traffic flow is used in the modeling, and RMSE_x refers to one of the other cases, i.e., 'traffic + stock' or 'All'. This improvement

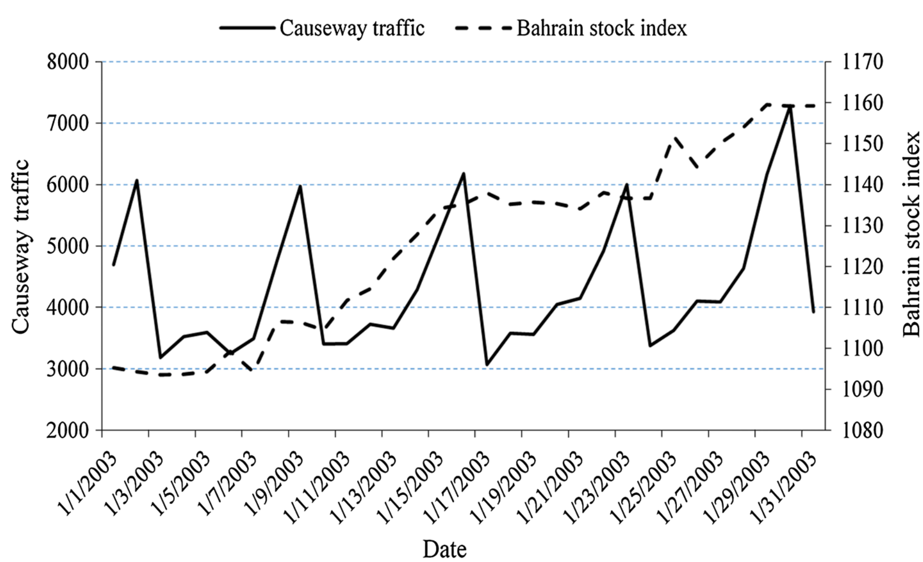

Fig. 5 Variations of the daily Bahrain stock market index and the causeway daily traffic for the month of January 2003

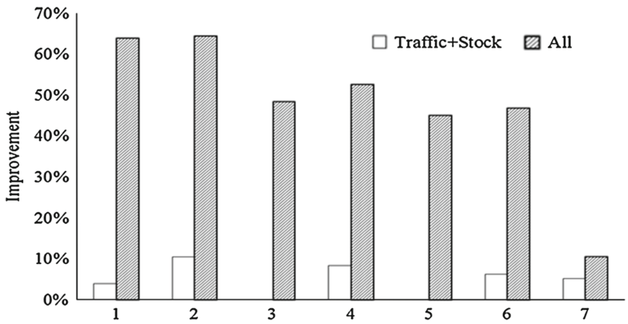

Fig. 6 Percentage improvement for $K = 1$ due to stock market indices and due to using all most relevant variables for different time lag values $P = 1$ to 7

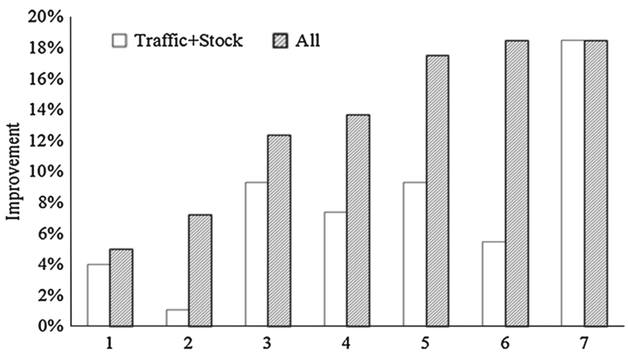

Fig. 7 Percentage improvement for $K = 7$ due to stock market indices and due to using all most relevant variables for different time lag values $P = 1$ to 7

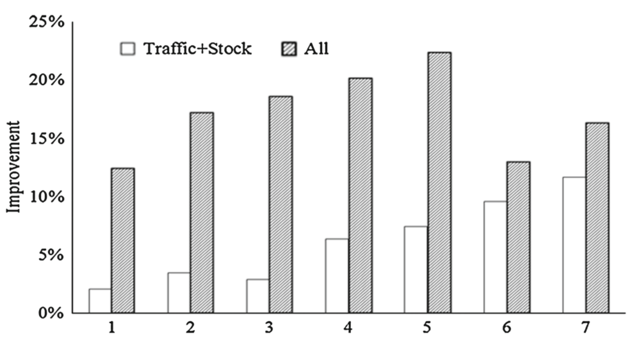

Fig. 8 Percentage improvement for $K = 30$ due to stock market indices and due to using all most relevant variables for different time lag values $P = 1$ to 7

was calculated for each time lag, prediction horizon, and model. The results were drawn as shown in Figs. 6, 7, 8, 9 and 10 for $K = 1, 7, 30, 90$ and 365, respectively. These figures illustrate that the inclusion of stock market indices has generally led to improvement over the case when only the traffic was used, which can reach to more than 18 % in some cases. The degree of improvement depends on the time lag and prediction horizon. Adding other relevant exogenous variables can further improve the accuracy and can reach to more than 60 % in some cases. Yet, the level of improvement

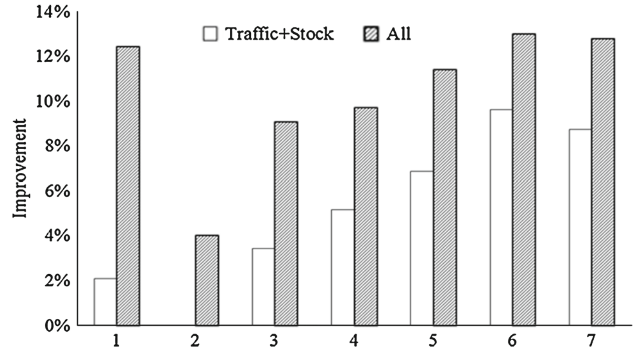

Fig. 9 Percentage improvement for $K = 90$ due to stock market indices and due to using all most relevant variables for different time lag values $P = 1$ to 7

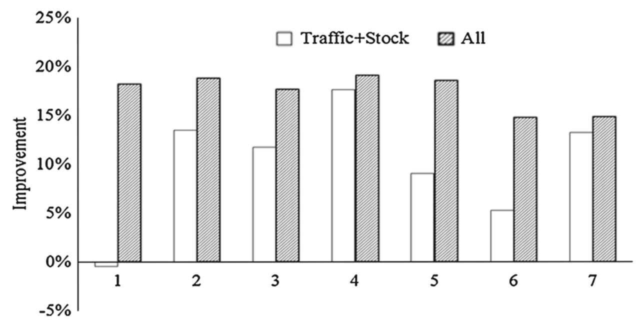

Fig. 10 Percentage improvement for $K = 365$ due to stock market indices and due to using all most relevant variables for different time lag values $P = 1$ to 7

over the case when only stock market indices in addition to traffic are used decays as the time lag increases and as the prediction horizon increases.

4 Conclusions

This study explored the impact of stock market indices and other regional exogenous factors on border traffic prediction. Various models were developed using different look-back periods, P, and for different prediction horizons, K. We also used two types of machine-learning models: neural networks and ensembles. Using a dataset that covers the period from 2003 to 2013, a correlation analysis was conducted to identify the most relevant variables. Then, various prediction models were built and evaluated. The study proposes using stock market indices in border traffic modeling as a surrogate measure of political and economic conditions of the region, these conditions affect the travel demand across borders. This hypothesis was first tested in the correlation analysis which proved that stock market indices have a statistically significant correlation with traffic flow. Moreover, these indices were used as inputs to the traffic forecasting models in addition to time-series traffic data for different prediction horizons. It was observed that the model accuracies tend to improve for all prediction horizons with the inclusion of stock market indices. The results also showed that the proposed ensemble models can have better accuracy than the single multilayer perceptrons. Another observation was that using other exogenous variables (in addition to stock market indices and time-series traffic data) do not show significant improvement in the accuracy for longer than 1-day prediction horizons.

Acknowledgments The authors would like to acknowledge the support provided by the Deanship of Scientific Research (DSR) at King Fahd University of Petroleum and Minerals (KFUPM) for funding this work under Grant No. IN121065.

References

1. Khattak, A.J.; De Palma, A.: The impact of adverse weather conditions on the propensity to change travel decisions: a survey of brussels commuters. Transp. Res. Part A Policy Pract. **31**(3), 181–203 (1997)
2. Moorthy, C.K.; Ratcliffe, B.G.: Short term traffic forecasting using time series methods. Transp. Plan. Technol. **12**(1), 45–56 (1988)
3. Mai, T.; Ghosh, B.; Wilson, S.: Short-term traffic-flow forecasting with auto-regressive moving average models. In: Proceedings of the ICE—Transport (2013)
4. Guo, F.; Krishnan, R.; Polak, J.: A computationally efficient two-stage method for short-term traffic prediction on urban roads. Transp. Plan. Technol. **36**(1), 62–75 (2013)
5. Karlaftis, M.; Vlahogianni, E.: Statistical methods versus neural networks in transportation research: differences, similarities and some insights. Transp. Res. Part C Emerg. Technol. **19**(3), 387–399 (2011)
6. Yin, H.; Wong, S.; Xu, J.; Wong, C.: Urban traffic flow prediction using a fuzzy-neural approach. Transp. Res. Part C Emerg. Technol. **10**(2), 85–98 (2002)
7. Dimitriou, L.; Tsekeris, T.; Stathopoulos, A.: Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. Transp. Res. Part C Emerg. Technol. **16**(5), 554–573 (2008)
8. Zhang, Y.; Liu, Y.: Traffic forecasting using least squares support vector machines. Transportmetrica **5**(3), 193–213 (2009)
9. Suykens, J.; Vandewalle, J.; De Moor, B.: Optimal control by least squares support vector machines. Neural Netw. **14**(1), 23–35 (2001)
10. Saneinejad, S.; Roorda, M.J.; Kennedy, C.: Modelling the impact of weather conditions on active transportation travel behaviour. Transp. Res Part D Transp. Environ. **17**(2), 129–137 (2012)
11. Aultman-Hall, L.; Lane, D.; Lambert, R.R.: Assessing impact of weather and season on pedestrian traffic volumes. Transp. Res. Rec. J. Transp. Res. Board **2140**(1), 35–43 (2009)
12. Clifton, K.J.; Chen, R.B.; Cutter, A.: Representing weather in travel behaviour models: a case study from Sydney, Australia. In: Australasian Transport Research Forum 2011 Proceedings, pp. 28–30 (2011)
13. Tsirigotis, L.; Vlahogianni, E.I.; Karlaftis, M.G.: Does information on weather affect the performance of short-term traffic forecasting models?. Int. J. Intell. Transp. Syst. Res. **10**(1), 1–10 (2012)
14. Dougherty, M.S.; Cobbett, M.R.: Short-term inter-urban traffic forecasts using neural networks. Int. J. Forecast. **13**(1), 21–31 (1997)
15. Lingras, P.; Mountford, P.: Time delay neural networks designed using genetic algorithms for short term inter-city traffic forecasting. In: Engineering of Intelligent Systems, pp. 290–299. Springer, Berlin (2001)
16. KFCA Authority: King Fahd causeway authority passenger statistics (2012). http://www.kfca.com.sa
17. Adhikari, R.; Agrawal, R.K.: A homogeneous ensemble of artificial neural networks for time series forecasting. Int. J. Comput. Appl. **32**(7), 1–8 (2011)
18. Lin, J.; Zhu, B.: A novel neural network ensemble system for economic forecasting. In: Huang, D.S.; Heutte, L.; Loog, M. Advanced Intelligent Computing Theories and Applications. With Aspects of Contemporary Intelligent Computing Techniques, vol. 2, pp. 1227–1233. Springer, Berlin (2007)
19. Hyndman, R.J.; Koehler, A.B.: Another look at measures of forecast accuracy. Int. J. Forecast. **22**(4), 679–688 (2006)
20. Armstrong, J.; Collopy, F.: Error measures for generalizing about forecasting methods: empirical comparisons. Int. J. Forecast. **8**(1), 69–80 (1992)
21. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I.: Pearson correlation coefficient. In: Noise Reduction in Speech Processing, Springer Topics in Signal Processing, vol. 2, pp. 1–4. Springer, Berlin (2009)
22. Al-Ahmadi, H.: Development of intercity mode choice models for Saudi Arabia. J. King Abdulaziz Univ. Eng. Sci. **17**(1), 3–20 (2006)
23. Bullock, D.; Catarella, A.: A real-time simulation environment for evaluating traffic signal systems. Transp. Res. Rec. J. Transp. Res. Board **1634**, 130–135 (1998)