International Journal of Civil & Structural Engineering – IJCSE
Volume 1 : Issue 1 *[ISSN 2372 – 3971]*
Publication Date : 09 January 2014

Traffic Forecasting for King Fahd Causeway: Comparison of Parametric Technique with Artificial Neural Networks

Ashar Ahmed
PhD Student, School of Civil Engineering
University Sains Malaysia
Pulao Penang, Malaysia
ashar.ue17@gmail.com

Uneb Gazder
PhD Student, Department of Civil and Environmental Engineering
King Fahd University of Petroleum and Minerals
Riyadh, Kingdom of Saudi Arabia
unebgazdar@yahoo.com

***Abstract*—Traffic prediction involves forecasting traffic in terms of Annual Average Daily Traffic (AADT), Design Hour Volumes (DHV) and Directional Design Hour Volumes (DDHV). These forecasts are used for a wide variety of purposes from the planning to the design and operational stages of the highway network. The forecasting needs the historical traffic data as well as the systems characteristics, apart from that choice of an appropriate model or technique is also an important consideration. This paper gives an overview of the traffic forecasting process and the models that are used for this purpose with emphasis on the use of Artificial Neural Networks (ANNs) and other modern techniques. ANNs are being compared with the traditional Parametric techniques used in this regard by applying linear regression analysis and ANNs for daily traffic forecasting on King Fahd causeway. It was observed from the estimated error values of both techniques that ANNs have better accuracy than linear regression technique for predicting daily traffic.**

Keywords— Traffic Forecasting, Artificial Neural Networks, Linear Regression

I. Introduction

Traffic prediction involves forecasting traffic in terms of Annual Average Daily Traffic (AADT), Design Hour Volumes (DHV) and Directional Design Hour Volumes (DDHV). These traffic forecasts are used at the design stage to determine the geometric and pavement design, in the operation stage to investigate the operational efficiency of the roadway in terms of delay or level of service, and for other purposes like impact assessment, incident management, etc. The following factors must be addressed when traffic forecasts are to be made:

- Methodology and level of effort
- Analysis year: Both present and future
- Analysis period: Daily, Hourly, Peak hour
- Forecasting Model

For the purpose of this review, we only focused on the forecasting models which are used in this regard. A typical process of traffic forecasting is depicted in **Error! Reference source not found.**.

The existing data used for traffic forecasts includes; traffic counts, pedestrian and bicycle usage, transit ridership, and traffic related factors such as directional, seasonal and peak hour factors and heavy traffic fractions. Apart from that, system related factors also affect the traffic demand and have to be determined, they include roadway characteristics and level and quality of service, and land-use and demographic patterns. In some cases, it may not be possible to find out the traffic counts for substantially long periods on continuous basis, the following equations are used to convert the available data for use.

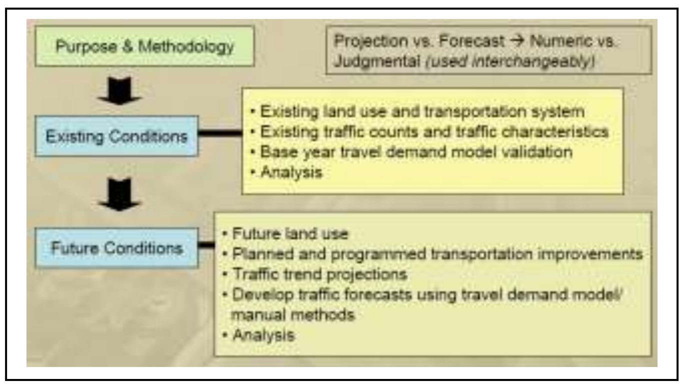


Figure 1. Travel Forecasting Process

$$AADT = ADT \text{ X } SF \text{ X Axle Correction Factor} \quad (1)$$
$$DHV = AADT \text{ X } K_{30} \quad (2)$$
$$DDHV = DHV \text{ X } D_{30} \quad (3)$$

The values of above mentioned factors are prominently dependent upon type of road and the normal ranges of these factors which are adopted by FDOT are given below in **Error! Reference source not found.Error! Reference source not found.**. However, care must be taken to check the reasonableness of these values by comparing with similar conditions or local sample values [1].

TABLE I. AVERAGE VALUES OF K30 ADOPTED BY FDOT

Road Type	Low	K_{30} Average	High	Standard Deviation
Rural Freeway	9.60	11.8	14.6	1.43
Rural Arterial	9.40	11.0	15.6	1.42
Urban Freeway	9.40	9.7	10.0	0.28
Urban Arterial	9.20	10.2	11.5	0.92

TABLE II. AVERAGE VALUES OF K30 ADOPTED BY FDOT

Road Type	Low	D_{30} Average	High	Standard Deviation
Rural Freeway	52.3	54.8	57.3	1.73
Rural Arterial	51.1	58.1	79.6	6.29
Urban Freeway	50.4	55.8	61.2	4.11
Urban Arterial	50.8	57.9	67.1	4.60

II. Selection of Model

Application of a particular model in real time projects is sometimes dependent upon the adoptability of that model by the local organizations. In most of the cases, locally recognized models are preferred. In addition to that, following issues should also be checked prior to use of a particular model:

- Use the latest version of a model
- Review the accuracy of the model for the base year in term of Root Mean Square Error (RMSE)
- Review the previous models applied to the same region/corridor
- Calibrate the model for parallel facilities, competing facilities, transit services, network revisions, disaggregation of zones, and socioeconomic data

FDOT has established the accuracy levels for selection of model based upon the type of facility and volume of traffic which are given in Table III.

TABLE III. ACCURACY LEVELS FOR MODEL SELECTION

Statistic	Standards: Acceptable	Standards: Preferable
RMSE: LT 5,000 VPD	100%	45%
RMSE: 5,000-9,999 VPD	45%	35%
RMSE: 10,000-14,999 VPD	35%	27%
RMSE: 15,000-19,999 VPD	30%	25%
RMSE: 20,000-29,999 VPD	27%	15%
RMSE: 30,000-49,999 VPD	25%	15%
RMSE: 50,000-59,999 VPD	20%	10%
RMSE: 60,000+ VPD	19%	10%
RMSE Areawide	45%	35%

Volumes from the selected model for the future years should be reviewed for logical traffic growth rates. Modelled traffic volumes should be compared with the prevalent general growth trends in the area. If any ill-logical growth pattern exists then the model specification, network coding and traffic assignment has to be revised.

III. Models for Traffic Forecasting

In this section we would describe some traditional methods of traffic forecasting. These models are mainly developed using the historical growth trends and regression analysis.

A. *Traffic Forecasting Using Growth Trends*

If any traffic model is not available or the application does not require the use of complex models, traffic forecasts can be made using traffic growth trends. The historical trends of traffic demand can be established from one of the following resources:

- Historical Data
- Land-Use Management Maps (LUMS)
- Gas Sales Record

The data is plotted on a graph with AADT on y-axis and year of count on the x-axis. A least square regression analysis is done to establish the growth trends. For traffic forecasts; either the empirically developed regression equations are used or the graph can be extended linearly to the future year as shown in Figure 2[2].

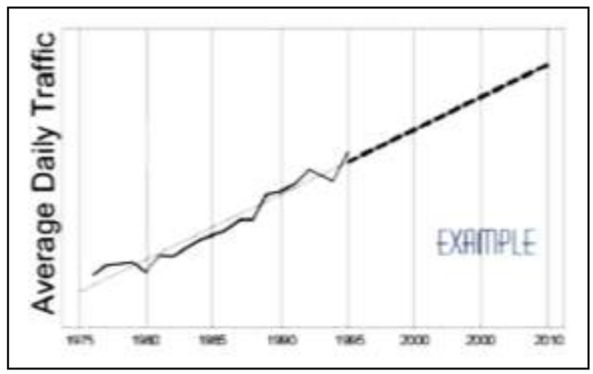


Figure 2. Graphical Method of Traffic Forecasting

B. *Site Specific Models*

Site-specific models are developed based upon the characteristics of the traffic generators and attractors without taking into consideration the state of the road network. NCHRP developed regression models for trip generation and attraction as a function of land-use character, intensity and location. The factors included in these models are:

- Automobile ownership
- Income
- Household size
- Density
- Type of development

Institute of transport engineers developed their own rates for trip generation based upon the land-use per unit of factor related to the type of land-use.

C. *Parametric and Non-Parametric Techniques*

For accurate traffic forecasting parametric techniques such as historical average algorithms, smoothing techniques, linear and nonlinear regression, filtering techniques, to autoregressive linear processes including autoregressive moving average (ARMA) have been employed since 1980s.

Some of these techniques, like Holt-Winters and ARIMA, mainly present the future conditions as the repetition of present trends. Some modifications have been made to these techniques to broaden their range by applying "Moving Average" approach. However, they are still based upon a linear combination of past values [3 and 4].

With the growth in development of computation intelligence techniques such as fuzzy logics, machine learning and neural networks; non-parametric techniques are also being employed in this area by many researchers.

Non-parametric techniques mainly focus on identifying the past conditions which are similar to the conditions at the prediction time. They work on the principle of dynamic clustering. The application of these intelligent techniques has shown satisfactory results in traffic forecasting. The trade-off is made between the simplicity of traditional parametric techniques and the effectiveness of intelligent techniques [5].

IV. Application of Neural Networks for forecasting

The process underlying inter-urban traffic flow is difficult to be captured by a single linear statistical algorithm. Artificial Neural Networks have received much more attention in the area of traffic forecasting because of the ability to approximate any degree of complexity and work without prior knowledge of problem solving. ANN is used to determine related number of vehicle and temporal characteristics using the historical flow patterns and it is also suitable for non-linear and dynamic evolutions. Few attempts have been made by researchers in order to enhance the accuracy of ANNs for forecasting which are described below.

Researchers [6] developed a fuzzy-neural model (FNM), integrating fuzzy logics with neural networks for predicting urban traffic flow. Their model contained two modules, gate network (GN) and expert network (EN). GN module applied fuzzy logics to the input data for its classification, and the EN applied neural network for identifying the input–output relationship. The results from their study showed the FNM model provided more accurate forecasts than the normal back propagation algorithms of neural networks [6].

A study [7] employed a genetic algorithm based optimization strategy to determine the appropriate neural network structure. Their study showed that combining the capabilities of a simple static neural network, with genetically optimized step size, momentum and number of hidden units, produces satisfactory results for modeling both univariate and multivariate traffic data [7].

In a study [8] used the simulated annealing algorithm with back propagation NN. This algorithm uses the genetic algorithm to elect, cross and mutate the synaptic weights and thresholds of ANNs. SA algorithm also decides the structure of ANN by identifying the number of input, output and hidden neurons. The results from the above mentioned study show that this integration gives lesser Mean Square Error than the normal Back propagation NN [8].

Another group of researchers [9] integrated the ANN as well as ARIMA models with self-organizing maps (SOMs), which were used to classify the initial flow data in to different states. As a conclusion they found out that hybrid ANN approach gives more satisfactory results than the other. This was in spite of the fact that using the hybrid ARIMA approach increased its accuracy as compared to all individual ARIMA models. They also checked the accuracy of Radial Based NN with BPNN and found out that the former approach gives slightly better results for short term traffic forecasting [9].

As an effort to reduce the iterations required by a normal back propagation NN algorithm, [10] used the Extended Kalman Filter (EKF) algorithm in combination with ANN. This algorithm gauges the weights as per the principle of minimum root mean squared covariance, which reduces the number of iterations required and makes it suitable for rapid dynamic calculations [10].

A traffic dispersion phenomenon is very important for traffic forecasting, especially for signalized networks. The probabilistic approach to simulate dispersion can be very useful for traffic flows under ideal conditions. This approach may not work well in complex cases because it is based upon strict mathematical rules. [11] developed the Neural networks to simulate traffic flows in recent years and their application includes queue prediction and dispersion, classification of highway traffic states and ramp metering. ANNs have the ability to emulate all kinds of traffic violations provided that an appropriate structure is chosen [11].

V. Traffic Forecasting

King Fahd Causeway provides a link between travelers from Al-Khobar (KSA) and Bahrain by road. There are two modes operating on the causeway; private vehicles and public transport. Since there is no railway connection between Bahrain and KSA so the road and air transport become very important for travelers to and from Bahrain. This causeway also provides an important connection for other parts of the gulf, because Bahrain airport is used as a hub for the connecting flights to Bahrain to all countries. The car and bus mode on the causeway becomes very competitive for the travelers who travel between KSA and Bahrain for using the Bahrain airport for their travel.

In this research, traffic data for the years 1999 – 2001 is used to develop models for predicting daily traffic on King Fahd Causeway by using the input of previous day's traffic.

International Journal of Civil & Structural Engineering – IJCSE
Volume 1 : Issue 1 [ISSN 2372 – 3971]
Publication Date : 09 January 2014

A. *Preprocessing*

Two data sets are used; one set contained total daily traffic including vehicles from different countries and buses. Other set is classified on the basis of cars from each gulf country. Separate models were developed for each data set using regression analysis as well as ANN.

For pre-processing, data was normalized assuming a Gaussian distribution. The data sets were divided in to two parts; 70% was kept for training and the rest was allocated for testing of the networks. During the initial runs of the models it was observed that the model error value was too high for predicting total number of vehicles. So the output traffic volumes were classified in to 3 levels of traffic flow, the extreme levels were considered to be beyond 1.5 times standard deviation from the mean on both sides. This was done according to the assumption regarding Gaussian distribution where majority of the data exists within the above limits.

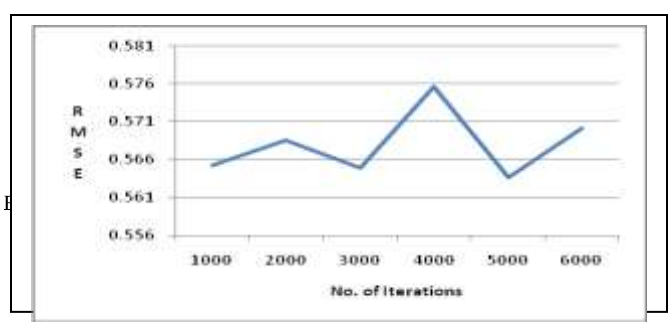
B. *Artificial Neural Networks*

ANNs are applied with different architectures which refers to the different arrangement and interaction between layers of neurons [12]. The simplest and most widely used type of architecture is the multi-layer perceptron (MLP). It consists of an input layer, output layer and one or more hidden layers. Signal flow in MLP is only in the forward direction and each neuron of any layer is connected to all neurons in the preceding layer. MLP has been used for traffic forecasting and has shown better performance as compared to other methods [3 and 4].

The MLPs were trained for both data sets and their performance was compared for different number of hidden neurons, learning rates, activation functions and number of iterations. For all networks, the conjugate gradient descent algorithm was used for optimization. The network parameters for both sets of data (volume-classified and country classified) were iterated and the results can be seen in figures 3 and 4. The networks with minimum acquired root mean square error (RMSE) values for testing data are given in table IV.

TABLE IV. ARTIFICIAL NEURAL NETWORKS

ANN Parameters	Total Volume Data Set	Country-Wise Classified Data Set
Number of Hidden Neurons	10	10
Learning Rate	0.005	0.005
Activation Function	Linear	Linear
Number of Iterations	4000	5000
$RMSE_{Test}$	0.56	0.30



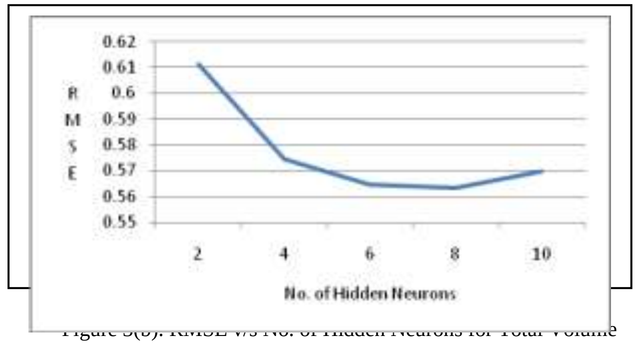
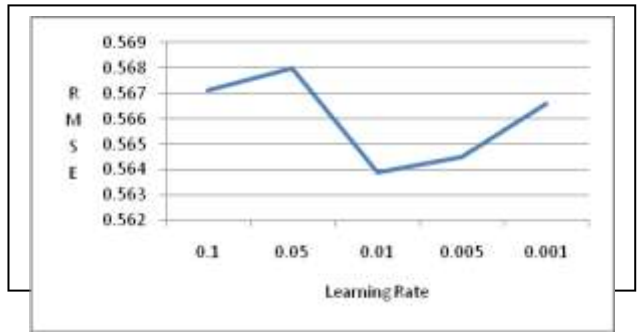


Figure 3(b). RMSE v/s No. of Hidden Neurons for Total Volume



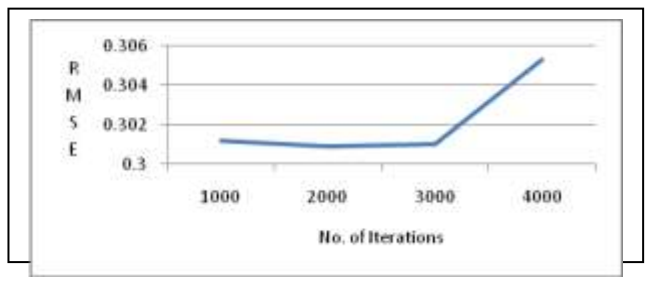


Figure 4(a). RMSE v/s No. of Iterations for Country-wise Classified Data

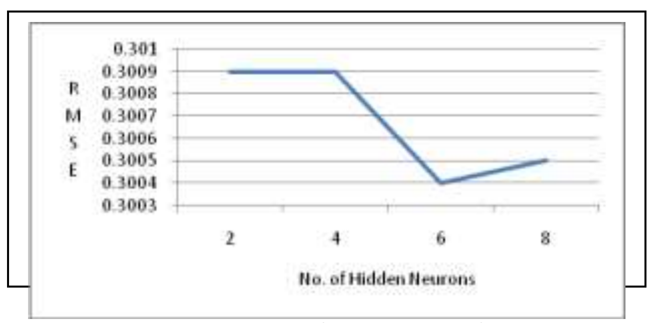
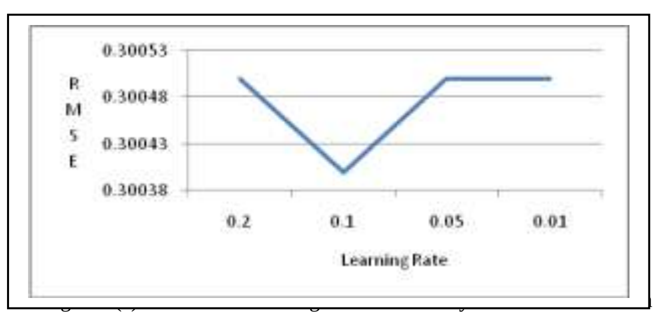


Figure 4(b). RMSE v/s No. of Hidden Neorons for Country-wise Classified Data



C. Linear Regression Model

In the next step a regression model was developed using the same pre-processed data. Least sum of square method was used for developing the model. The results from the regression model are given in table V. It can be seen that class of traffic has a large effect on total volume forecasts while for the country-wise data, the country data seems to have a very small effect on the forecasts. The training and testing set for the regression model was the same as for ANN.

TABLE V. REGRESSION MODEL RESULTS

Model Parameters	Total Volume Data Set	Country-Wise Classified Data Set
Intercept	-0.948955	0.113948
Class/Country	0.865200	-0.031957
Input Traffic Volume	0.479015	0.899961
$RMSE_{Test}$	0.8807531	0.354387

D. Discussion

RMSE values for the regression model are less than the values acquired by ANN, especially for the total volume data set. This may be contributed to the increased randomness of the traffic data on daily basis which may not be captured by a linear regression model. Regression models are developed using a single iteration of the model. Whereas in case of ANNs; different parameters were iterated for improving the accuracy of the model at a small scale. Regression models give an insight to the relationship between the independent variables and forecasted values through the value of coefficients.

VI. Conclusion

In this research, two models were developed to predict daily traffic on King Fahd Causeway on the basis of preceding day's traffic. There were two data sets, one for total daily traffic and another one for daily traffic coming from each country. The models were developed using Artificial Neural Networks (ANNs) and regression model for both data sets. The performances for ANNs were maximized by trying different activation functions, number of iterations, number of hidden neurons and learning rates. The data for total daily traffic was needed to be further classified for improving accuracy of the network. The best performing networks were obtained with a linear activation function; the number of iterations was 5000 and 2000 for total daily volume and country-wise classified volumes respectively. The regression model were developed using the least sum of square method. The Root Mean Square Error (RMSE) values for regression models were higher for both data sets in general. RMSE value for total volume data set for significantly higher for regression model as compared to ANN, which can be attributed to the randomness of the daily traffic.

It was observed that the Root Mean Square Error (RMSE) value for total volume ANN was very high as compared to the ANN for country-wise daily traffic. This value can be improved if more training data can be supplied to the network. Type of network is another issue that can influence the forecasting accuracy. Investigating the significance of data samples and type of network is a possible area for future research in this field.

The overall results of the forecasting analysis show that ANN perform better than linear regression models with any configuration of parameters and the simplest architecture like multi-layer perceptrons. However, ANNs are comparatively less explanatory in describing the significance of the variables on the forecasts.

Acknowledgment

The authors acknowledge the support of Prof. Hashim Al-Madani (University of Bahrain) in conducting this study.

References

[1] FDOT. Resource Guidebook. *FDOT District V, Intermodal Development Unit*. Florida.2006.

[2] Reel, R., Friedman, S., Mysore, V., Tabatabaee, F., & Majano, V. *Project Traffic Forecasting*. Florida Department of Transportation. Retrieved from http://www.dot.state.fl.us/planning/statistics/trafficdata/ptf.pdf 2012.

[3] Cortez, P., Rio, M., Rocha, M., & Sousa, P. Internet Traffic Forecasting using Neural Networks. *2006 International Joint Conference on Neural Networks* (pp. 2635-2642). 2006.

[4] Cortez, P., Rio, M., Rocha, M., & Sousa, P. Multi-scale Internet traffic forecasting using neural networks and time series methods. *Expert Systems*, *29*(2), 143-155. doi:10.1111/j.1468-0394.2010.00568.x 2012.

[5] Zhang, Y., & Liu, Y. Comparison of Parametric and Nonparametric Techniques for Non-peak Traffic Forecasting. *World Academy of Science, Engineering & Technology*, 8-14 2009.

[6] Yin, H., Wong, S. C., Xu, J., & Wong, C. K. Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C*, *10*, 85-98. 2002.

[7] Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies*, *13*(3), 211-234. doi:10.1016/j.trc.2005.04.007 2005.

[8] Nagare, A., & Bhatia, S. Traffic Flow Control using Neural Network. *International Journal of Applied Information Systems (IJAIS)*, *1*(2), 50-52. 2012.

[9] Chen, H., Grant-muller, S., Mussone, L., & Montgomery, F. A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting, 277-286. 2001.

[10] Tang, Y.-H., & Xi, B. Dynamic forecasting of traffic volume based on quantificational dynamics : A nearness perspective, *5*(4), 389-394. 2010.

[11] Qiao, F., Yang, H., & Lam, W. H. K. Intelligent simulation and prediction of traffic flow dispersion. *Transportation Research Part B*, *35*, 843-863. 2001.

[12] Hensher, D. A., & Ton, T. T. A Comparison of the Predictive Potential of Artificial Neural Networks and Nested Logit Models for Commuter Mode Choice. *Transportation Research Part E*, *36*(2000), 155-172. 2000.