Use of Weather Inputs in Traffic Volume Forecasting

Shane Butler*, **John Ringwood*** and **Damien Fay****

*Department of Electronic Engineering**
NUI Maynooth, Maynooth
Co. Kildare, Ireland

shane.butler@eeng.nuim.ie
john.ringwood@eeng.nuim.ie

*Department of Mathematics***
NUI Galway
Co. Galway, Ireland

damien.fay@nuigalway.ie

Abstract — **In this paper, an examination of the effect of including rainfall inputs in the forecasting of daily vehicular traffic volumes is undertaken. A case study is carried out at a busy intersection in Dublin city to examine if any reduction in forecasting error can be obtained by the incorporation of rainfall inputs. This paper also demonstrates the value of incorporating lessons learned from linear time series modelling to the non-linear analysis undertaken.**

Keywords — **Traffic forecasting, time series modelling, neural networks.**

I Introduction

Intelligent Transportation Systems (ITS) have been identified as the key to solving many of the increasing congestion problems present nowadays on most urban roads. Intelligent Transportation Systems have the potential to smooth traffic flow, optimise traffic signals, provide valuable information to travellers, reduce travel time and decrease fuel consumption, ultimately reducing congestion [1].

One shortcoming identified in literature, is the need for weather responsive traffic management [2]. Most traffic management systems are designed assuming clear conditions and suffer from lack of flexility during inclement weather conditions. Issues such as traffic signal timing which may be optimised assuming clear conditions may not be performing optimally during periods of adverse weather.

It has been demonstrated previously that changes in precipitation intensity impacts the speed, headways and capacity of roadways [3]. In this paper, an examination of the effects of precipitation on traffic volume forecasting is undertaken to investigate if an improvement in forecasting accuracy can be achieved by its inclusion. It has been shown previously that precipitation does not impact on mean daily traffic volumes [4], however, it could be reasoned that the reduction of speeds, headway and capacity of roads may manifest itself in the form of fewer vehicles passing the same point on a roadway during inclement rainfall conditions. This would also explain the increased congestion experienced by road users during periods of significant rainfall. By including rainfall in the forecast of traffic volumes it will be investigated if any improvement in forecasting accuracy can be obtained.

This paper examines the application of neural networks to modelling daily traffic volumes. Neural networks have been applied, successfully, to traffic forecasting problems in numerous publications [5, 6, 7]. Self-Organising feature maps for day-type identification will also be employed to identify the number of different day-types present over a standard week.

In addition to the traditional black-box approach to neural network modelling, an effort is also made to incorporate information on an effective input and model structure suggested by linear time series modelling.

II Data Availability

a) *Traffic Data*

To undertake the modelling exercise presented in this paper, loop detector data was gratefully provided by Dublin City Council for a busy junction in Dublin city center. The junction is TCS 172. The approach to the junction which was modelled, consists of two lanes of traffic on the south to north approach to the junction.

Traffic volumes passing through each of the two lanes modelled was available in 15 minute intervals over a period of approximately four years, from 2002 to 2006. However, in order to tabulate the data with available meteorological data, the traffic data had to be summed to hourly intervals. The volumes passing through each of the two lanes on the approach to the junction were also summed. A plot of traffic flow on a random weekday is presented in Figure 1 below, showing the traffic volume per hour against time.

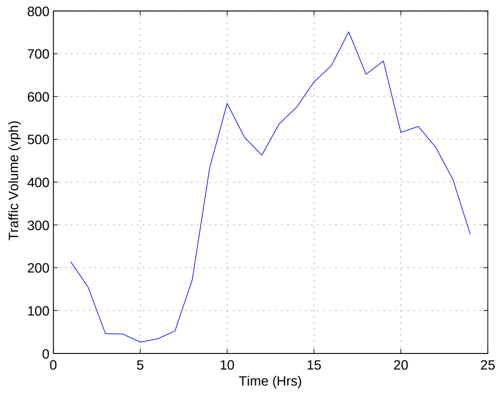

Fig. 1: Traffic Flow for a Typical Weekday at Junction TCS 172

b) Weather Data

Weather data for the period of 2000 to 2006 was made available by Met Eireann. The data consists of hourly observations for temperature, humidity, etc. and hourly means or summations for rainfall, wind, etc. At this stage, only the rainfall data has been used as a causal input in the forecasting of daily traffic volumes.

III Data Preprocessing

Prior to undertaking the modelling exercise it was necessary to identify all outliers and missing data in the traffic data. The weather data required no further preprocessing as no outliers or missing data was present.

The first task undertaken was to identify and remove all missing data in the dataset and to remove all bank holidays, holiday periods (Christmas, Easter, etc.). Any other inconsistent days identified, possibly due to roadworks, lane closures, school holidays, etc were also removed.

The next task was to identify any trends in traffic volumes over each of the seasons. Examination of the data clearly showed fluctuations in the daily traffic volumes and profiles over the year. Data for period of February to April over three years was finally selected for modelling, which demonstrated relatively consistent daily profiles and volumes.

IV Day Type Identification

In this section Self-organising feature maps are employed for day type identifications, in order to classify and compare the trends over each of the days of the week for the dataset being modelled.

The objective of this exercise was to classify each of the days of the week and determine the number of separate models that would be required to model each of the day types identified.

Self-organising feature maps or Kohonen maps, have been applied successfully in other fields such as electrical load forecasting [8] to assist in identifying day types within large datasets. In this application, the algorithm used by Hsu and Yang [9] is employed to carry out the day-type identification exercise.

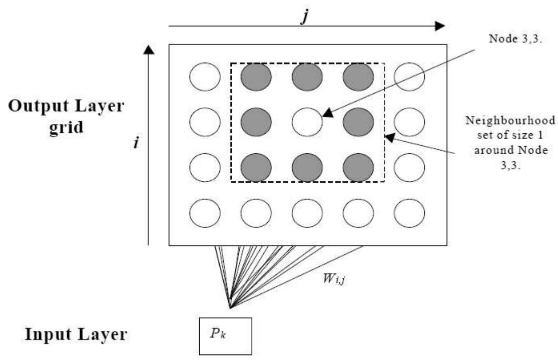

Fig. 2: Kohonen Map Structure [8]

The Kohonen map network consists of a grid of output nodes connected to the inputs via a set of weights as shown in Figure 2. When presented with the k^{th} input vector $\mathbf{P}_k$, ε R^{1xn} the network calculates the activation of each node by $\mathbf{P}_k$ as:

$$a_{i,j,k} = W_{i,j} P_k \tag{1}$$

The inputs are said to be *mapped* onto the network node with the highest activation. After several inputs have been presented, similar inputs are mapped to the same or adjacent nodes of the network, i.e. within a small neighbourhood.

Having carried out the necessary preprocessing and training described in [8], the next step is to assign day types to the triggered nodes. In this case, each of the days of the week, Monday to Sunday are assigned. Figure 3 shows the nodes that have been triggered for Tuesday to Sunday respectively (Monday has not been included due to space constraints). It is clear from examining the results, that each of the weekdays have triggered the same nodes, and that separate nodes

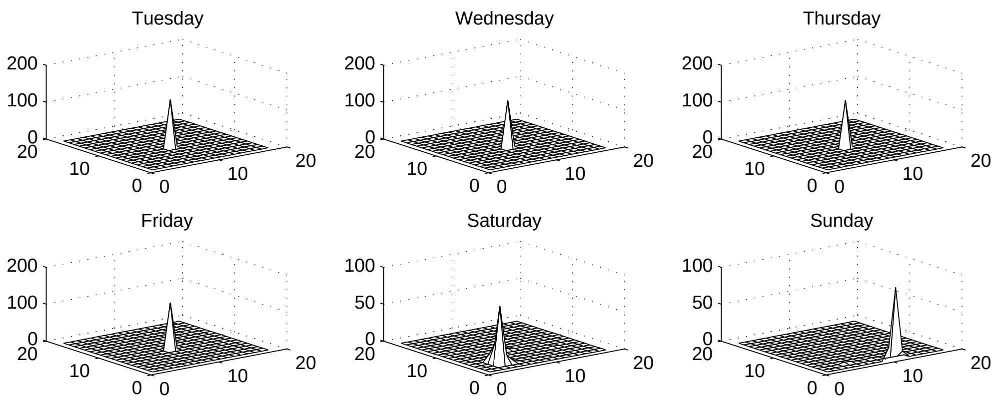

Fig. 3: Nodes Triggered by Respective Daily Traffic Profiles

have been triggered for both Saturday and Sunday (The nodes triggered for Monday were the same as each of the other weekdays). Effectively, each of the weekdays within the dataset examined, represent a single 'Day-Type'. Thus, the dataset needs to be disaggregated into weekdays, Saturdays and Sundays for model building. Only weekdays have been considered for the modelling exercise presented here.

V Neural Network Modelling

a) Introduction

Neural networks provide an unconstrained nonlinear modelling technique where a general nonlinear mapping is formed between some subset of past time series values and a future time series value. Temporal information may be presented to the network using a time lagged vector of time series data at the input, known as a tapped delay line, with the current value of the time series presented at the output.

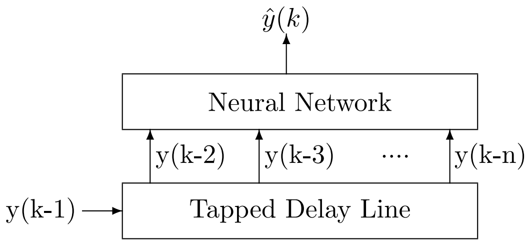

Fig. 4: Time Series Neural Network

Figure 4 above describes the univariate case. To extend this to the multivariate case is relatively straight forward task, whereby the input is now made up of past values of the time series for which the forecast is required, but also on present and past values of other exogeneous variable time series, in this case, rainfall. Such a network is suitable for performing single step predictions, however, to perform multi-step predictions, it is necessary to introduce feedback into the network, whereby another tapped delay line is introduced through which the output of the network is fed iteratively back to the input.

b) Box-Jenkins Input Structure

The application of a total black-box approach to neural network modelling of dynamic systems, would generally utilise a model of the form shown in Figure 5 below, with tapped delay lines for input and output variables forming the input to the network. Using a 'standard' autoregressive (AR) model of form shown in Figure 5, it would be usual to choose inputs which span a single season of the dataset, in this case 24 inputs for the previous 24 hours of the day.

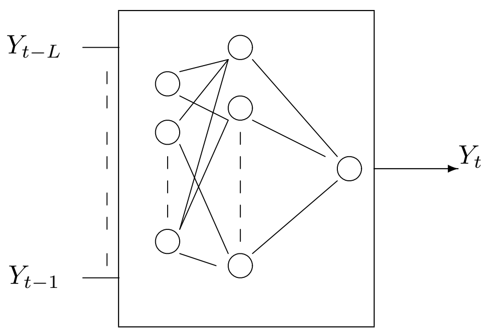

Fig. 5: Network with 'Standard' Autoregressive Input Structure

The use of such a total black-box approach may however, disregard structural information available from linear time series analysis. In this study, an effort is made to incorporate information on an effective input and model structure suggested by linear time series modelling. The approach undertaken here involves the use of the Box-Jenkins [10] methodology. A similar approach was previously carried out in [11], for electrical load forecasting, which resulted in a significant improvement in forecasting accuracy in comparison to using the 'standard' autoregressive input structure.

The general procedure for this linear modelling approach is as follows:

1. Determination of seasonality of time series and application of seasonal differencing
2. Application of further differencing transformations to make the time series stationary
3. Investigation of significant inputs to use a causal variables in model
4. Determination of orders of season and non-seasonal regressors
5. Identification of model parameters

The univariate Box-Jenkins model is derived from the general SARI (seasonal autoregressive integrated) model of the form:

$$\Phi_p B \Phi_P B^L \nabla_L^D \nabla^d Y_t = a_t \tag{2}$$

where:

Y_t is the time series

$\nabla_L^D \nabla^d = (1 - B^L)^D (1 - B)^d$ is a differencing transformation required if the data is nonstationary, d is the degree of non seasonal differencing, D is the degree of non seasonal differencing, and L is the season length (in this case its the number of hours in a day),

B is the backward difference operator (Time Domain),

$\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p)$ is the non seasonal autoregressive operator of order p,

$\Phi_P(B^L) = (1 - \phi_{1.L} B^L - \phi_2 B^{2.L} - .... - \phi_{p.L} B^{PL})$ is the seasonal autoregressive operator of order P.

a_t is the forecast error,

The lags p and P are generally determined using correlation analysis, however in this case the values for p and P were determined based upon the multi-step forecasting performance obtained in testing. The seasonality of the data, L is 24, i.e. one day.

Expansion of the univariate Box-Jenkins model described by equation 2 gives:

$$\begin{aligned} 1 - \phi_1 B - ... - \phi_p B^p - \phi_{1,L} B^L + \phi_1 \phi_{1,L} B^{L+1} \\ ... + \phi_p \phi_{1,L} B^{L+p} - ... - \phi_{P,L} B^{PL} + \phi_1 \phi_{P,L} B^{PL+1} \\ ... + \phi_p \phi_{P,L} B^{PL+p})(1 - B^L)^D (1 - B)^d Y_t = a_t \end{aligned} \tag{3}$$

Defining Z_t as:

$$Z_t = (1 - B^L)^D (1 - B)^d Y_t \tag{4}$$

a Box-Jenkins neural network model of the form:

$$g(Z_y, ..Z_{t-p}, Z_{t-L}..Z_{t-L-p}, .., Z_{t-PL}..Z_{t-PL-p}) = a_t \tag{5}$$

is produced, where the inputs to the neural network have already been subject to seasonal and one-step differencing. To determine the final forecast value from the neural network it is necessary to appropriately integrate the output, using seasonal and one step integration. Figure 6 below shows a neural network with an input structure of the form of equation 5.

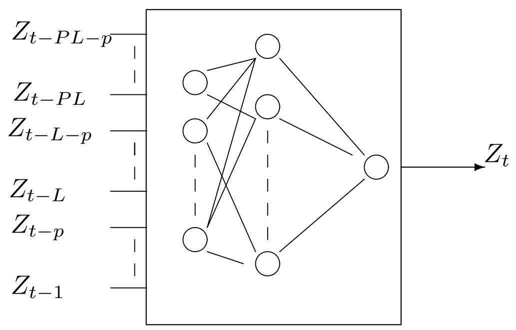

Fig. 6: Network with Appropriately Differenced 'Box-Jenkins' Input Structure

The objective of developing such an input structure given by equation 5 is to focus the network on the most appropriate inputs for forecasting the next output. In general, this also has the effect of reducing the number of inputs required, and hence reduces the required training time. For example, using the 'standard' autoregressive input structure, the previous 24 inputs corresponding to one season length L were employed. Using the Box-Jenkins input structure, with the non-seasonal autoregressive operator p, determined as 9 from the multi-step performance and the seasonal autoregressive operator P determined as 1, results in a total of of 19 inputs to the network.

The integration of rainfall as an input to the neural network involved a doubling of the number of inputs to the network, whereby the total hourly rainfall for each corresponding traffic volume input is presented to the network.

c) Network Training

In time series modelling applications, a network structure that can operate recurrently and produce a continuous output is required, due to the autoregressive nature of time series. In this application, recurrent Multi Layer Perceptrons (MLP's) have been employed. Such a model structures have previously applied successfully to electrical load forecasting problems [11].

A three layer structure was employed with a linear output neuron to remove any restriction on the output range. The number of neurons within each layer was determined from test runs. A standard backpropogation algorithm was employed to train the MLP network.

VI Results

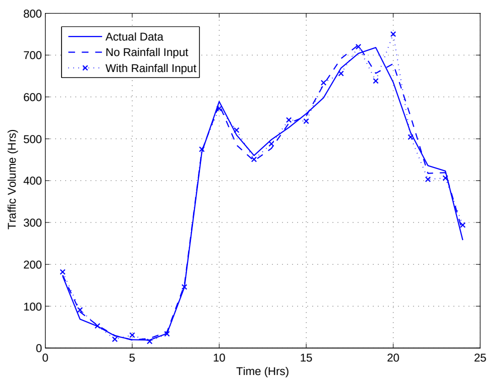

Fig. 7: Single-Step Forecasting Results using Box-Jenkins Input Structure

Model Type	RMSE
Box-Jenkins (No Rainfall Input)	22.07
Box-Jenkins (With Rainfall Input)	28.04

Table 1: Single-Step Forecast Results

Table 1 above compares the results of single-step forecasting of one days traffic volumes using the Box-Jenkins input structure with and without the inclusion of rainfall as a causal input. Interestingly, the inclusion of rainfall in the forecast increases the forecasting error, suggesting that the rainfall input does not influence the volume of traffic recorded over each hour.

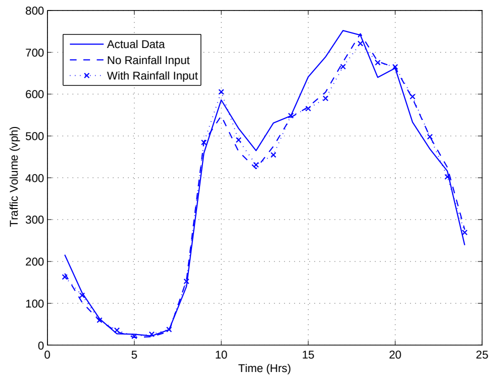

Fig. 8: Multi-Step Forecasting Results using 'Standard' Autoregressive Input Structure

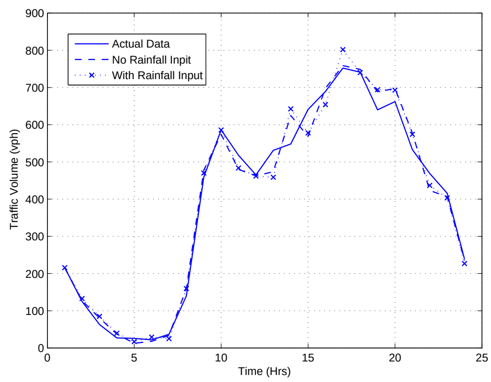

Fig. 9: Multi-Step Forecasting Results using 'Box-Jenkins' Input Structure

Model Type	RMSE
Autoregressive (No Rainfall Input)	39.72
Autoregressive (With Rainfall Input)	41.96
Box-Jenkins (No Rainfall Input)	32.52
Box-Jenkins (With Rainfall Input)	36.18

Table 2: Multi-Step Forecast Results

Table 2 above compares the multi-step forecasting results achieved using both the 'standard' autoregressive input structure and the Box-Jenkins input structure. As was the case in the single-step forecasting, the inclusion of the rainfall inputs again increased the forecasting error. Figures 8 and 9 demonstrate the results of the multi-step (24 step ahead) forecasting.

It is noted, that the use of the Box-Jenkins input structure leads to a significant improvement in forecasting accuracy in comparison to that achieved using just the 'standard' autoregressive input structure.

VII Conclusions

This paper has examined the effect of including rainfall inputs in the forecasting of daily vehicular traffic volumes. Prior to the modelling exercise, self-organising feature maps were employed to carry out day-type identification of the dataset available. It was shown that all of the weekdays map to the same activation nodes of the feature map, which implies that each standard weekday represents a single day-type.

Neural networks were employed to carry out the modelling exercise, which have been shown previously to be suited to the forecasting of traffic volumes. The inclusion of rainfall as a causal input to the model, has been shown not to improve the forecasting error achieved. This suggests that at the junction examined, rainfall does not influence the volume of traffic passing through over an hourly period. It may be possible that during intense rainfall traffic volumes passing through the junction are reduced, however the effect may be averaged out over a full hourly cycle. A suggested next step, is to obtain rainfall data in 15 minute intervals to tabulate with the available traffic data at a smaller sampling interval and investigate its influence on forecasting accuracy. It is also likely that the time of day at which rain falls will have an influence over its effect on traffic volumes and may require further investigation.

It is not immediately apparent the reason why the inclusion of rainfall inputs to the network should result in an increase in RMSE for the results. It is noted however, that the data set used for modelling was not very 'rain rich' and there may have been an insufficient number of rainy days for the training of the network. A future approach may be to obtain a more rain rich data set to examine any effect this may have and also to examine if there is any threshold effect whereby only rain above a certain intensity influences traffic volumes.

It has also been demonstrated in this paper, that by incorporating ideas from linear time series analysis in the application of a non-linear modelling tool, a significant improvement in forecasting accuracy can be achieved.

References

[1] B. L. Smith and R. K. Oswald, "Effects of parameter selection on forecast accurracy and execution time in nonparametric regression," in *IEEE Intelligent Transporation Systems*, vol. 6, Dearborn (MI), USA, Oct. 2000.

[2] P. A. Pisano and L. C. Goodwin, "Research needs for weather responsive traffic management," Transport Research Board (TRB) Annual Report, Tech. Rep., 2004.

[3] M. Agarwal, T. H. Maze, and R. Souleyrette, "Impacts of weather on urban freeway traffic flow characteristics and facility capacity," in *Mid-Continent Transportation Research Symposium*, Ames, Iowa, Aug. 2005.

[4] J. H. Hogema, "Effects of rain on daily traffic volume and on driving behaviour," TNO Human Factors Research Institute, Tech. Rep., 1996.

[5] S. Y. Yun, S. Namkoong, J. Rho, S. Shin, and J. Choi, "A performance evaluation of neural network models in traffic volume forecasting," *Mathematical and Computer Modelling*, vol. 27, 1998.

[6] B. V. Arem, H. R. Kirby, M. J. V. D. Vilst, and J. C. Whittaker, "Recent advances and applications in the field of short-term traffic forecasting," *International Journal of Forecasting*, vol. 13, 1997.

[7] M. S. Dougherty and M. R. Cobbett, "Short-term inter-urban traffic forecasts using neural networks," *International Journal of Forecasting*, vol. 13, 1997.

[8] D. Fay, J. V. Ringwood, M. Condon, and M. Kelly, "Establishing a solution strategy for electrical demand forecasting in Ireland," in *22nd International Symposium on Forecasting*, Dublin, Jun. 2002.

[9] Y. Y. Hsu and C. C. Yang, "Design of artificial neural networks for short-term load forecasting. Part 1: Self organising feature maps for day type identification," in *IEE Proceedings-C, Generation, Transmission and Distribution*, Sep. 1991.

[10] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control.* Prentice Hall, 1994.

[11] J. V. Ringwood and F. Murray, "Forecasting of weekly electricity consumption using neural networks," in *Irish DSP and Control Colloquium*, Dublin, Jun. 1996.