8th WCTR

ARTIFICIAL NEURAL NETWORK AND STATISTICAL MODELLING OF TRAFFIC FLOWS - THE BEST OF BOTH WORLDS

STEPHEN D. CLARK
Institute for Transport Studies
University of Leeds
LEEDS, LS2 9JT
England

HAIBO CHEN
Institute for Transport Studies
University of Leeds
LEEDS, LS2 9JT
England

SUSAN M. GRANT-MULLER
Institute for Transport Studies
University of Leeds
LEEDS, LS2 9JT
England

Abstract

This paper uses motorway traffic flow data to evaluate the performance of a number of forecasting techniques. These techniques are: naïve methods; ARIMA time series methods; Artificial Neural Network methods and a composite SOM/ARIMA method. The performance was evaluated using the root mean square and mean percentage forecast error and additionally a weighted index. The SOM/ARIMA method was found to perform well on all measures.

INTRODUCTION

The UK Highways Agency is responsible for the management and control of the UK trunk road and motorway networks. The Agency has recently embarked on a programme of extensive instrumentation of the London orbital Motorway, the M25 (see figure 1) as part of the MIDAS programme (Nuttall, 1995 and Maxwell and Beck, 1996). Lane specific detector loops are implemented at 500m intervals and information on the traffic state is collected from these loops every minute, 24 hours a day.

This information is used by the Highways Agency to assess the state of traffic and, when and where appropriate, engage a system of Variable Speed Regulation signs to advise drivers of a new speed limit. An enhancement to this approach would be to anticipate the conditions and engage the system so as to forestall the breakdown in traffic conditions.

The Highways Agency commissioned the Institute for Transport Studies at the University of Leeds to develop techniques to produce accurate short term forecasts of the traffic state. The work involves two main strands, firstly the use of solely Artificial Neural Network techniques to produce forecasts of traffic flows and travel times and secondly, a combined Artificial Neural Network and statistical time series methodology to forecast traffic flows alone. It is this latter strand which is considered in this paper.

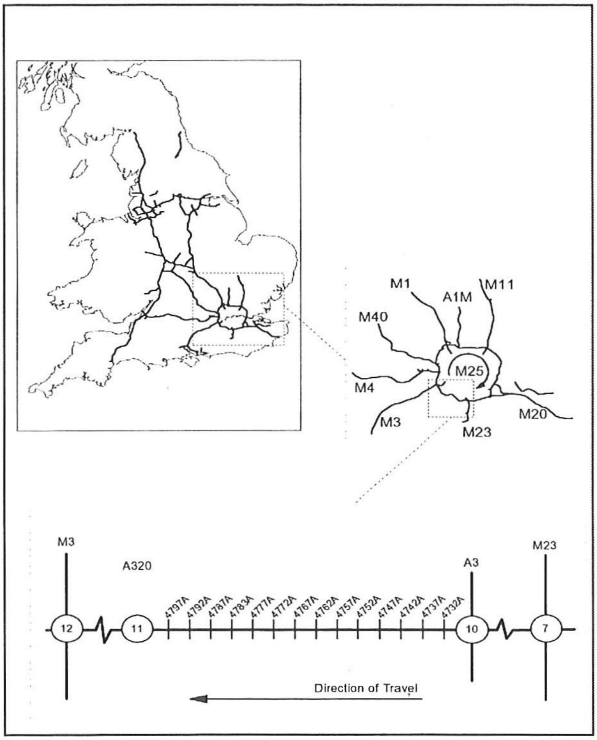

Figure 1 - Location of test site

MODELLING APPROACHES

A considerable body of work exists on the modelling of traffic flows using traditional and innovative modelling techniques (Clark, Dougherty and Kirby, 1993 and Conner, Martin and Atlas, 1994 provide a flavour of the work).

Traditional techniques include time-distance and statistical modelling. Time distance techniques use information on the spatial configuration of a network of roads to project forward the movement in traffic across space and time to arrive at a forecast, displaced in time and space from the reference sight. In the simple case of a junction free stretch of road, a vehicle travelling at 60 kilometres per hour could be expected to be 5 km's further downstream in five minutes. Statistical modelling techniques can include classes of naive techniques, regression, smoothing, decomposition and Box-Jenkins (1976) approaches.

More innovative techniques include the use of artificial neural networks. These two class of techniques have their own advantages and disadvantages. The advantage of the statistical techniques is that their structure is largely transparent and its behaviour may be rationalised. An advantage of artificial neural networks is that they can capture complex, potentially non-linear, relationships in a system.

JOINT APPROACH

Other papers (Van der Voort, Dougherty and Watson, 1996) have demonstrated that a combined used of traditional statistical and artificial neural network techniques may enhance the performance of a forecasting system. Following the work of Van der Voort, an artificial neural network based Self Organising Map (SOM) (Kohonen, 1988) was deployed to classify a series of traffic patterns. A reasonable ARIMA model was then associated with each of the resultant clusters and the performance evaluated.

SELF ORGANISING MAP

As mentioned above, the detection infrastructure on the M25 is able to measure traffic flow (vehicles/hour), vehicle speed (km/hour), detector occupancy (percentage of time occupied by a vehicle) and vehicle headways (seconds) at regular 500m intervals in each lane, both clockwise and anti-clockwise. This information is available on a minute by minute basis. In order to reduce the effect of individual missing observations and produce useful time horizon forecasts these one minute lane measurements were combined with a weighted mean to produce a single observation per 15 minute period, per link. The measures used to capture the traffic state were flow, speed and occupancy at a triplet of adjacent detectors. An example of these three measures at a site for a 6:00am to 21:00pm period for five days (excluding Saturdays and Sundays) in March 1997 is shown in figures 2 to 4.

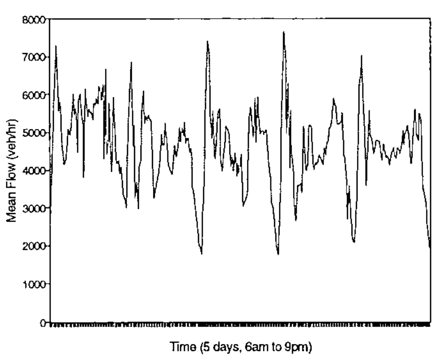

Figure 2 – Flow (vehicles/hour)

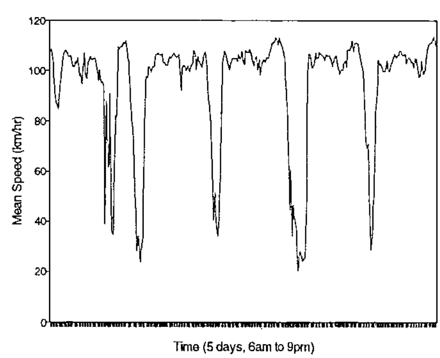

Figure 3 – Speed (km/hour)

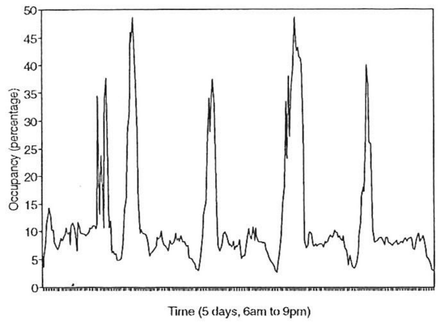

Figure 4 – Occupancy (percent)

A SOM network was trained using three weeks of 15 minute aggregated traffic data from October 1996. The SOM map, shown in figure 5, formed four distinct clusters, one large, one medium and the other two of a smaller size. Bi-variate scatter plots of the three measures where the cluster membership is denoted by a 1, 2, 3 or 4 are given in figures 6 to 8.

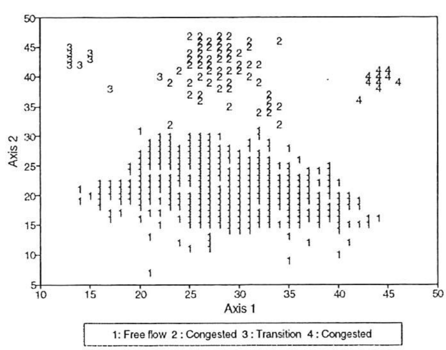

Figure 5 – Formation of SOM Clusters

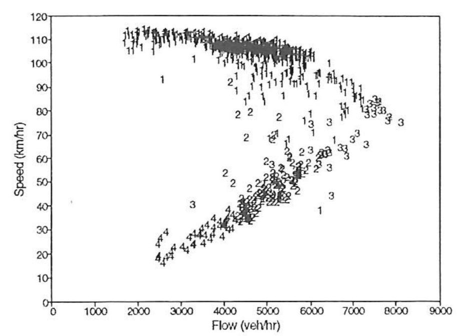

Figure 6 – Speed / Flow relationship

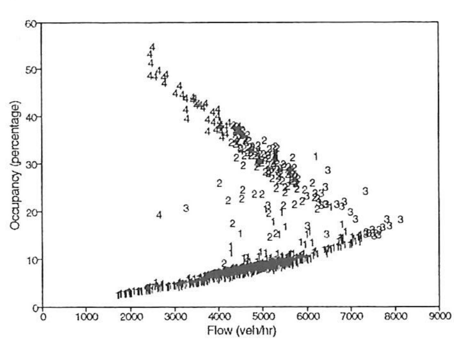

Figure 7 – Occupancy / Flow relationship

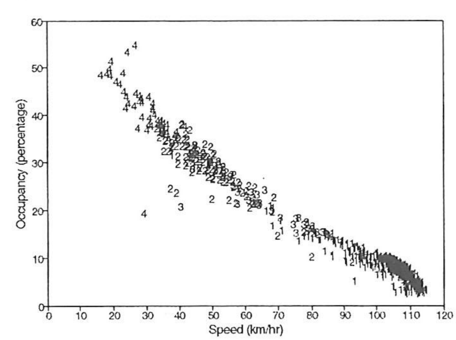

Figure 8 – Occupancy / Speed relationship

Reassuringly these graphs conform to recognised relationships in traffic flow theory. Also the clusters now have an interpretation. Cluster 1 is associated with traffic in free-flowing conditions, cluster 3 is a transition phase between free-flow and congested conditions and also the point of maximum throughput for the network. Cluster 2 is the initial stages of congested conditions and cluster 4 is high congestion (low flow and speeds but high link occupancies). Inspection of the transition from one state to another also shows plausible results ie the system does not move from cluster 1 state to or from cluster 4 state without first passing through states 3 or 2 or both first.

GOODNESS OF FIT

To assess the performance of a forecasting methodology some measure of goodness of fit is required. At the heart of our performance measure is the residual, r_t.

$$r_t = f_t - v_t \tag{1}$$

Where f_t is the forecast flow at time t;
v_t is the observed flow at time t;

A simple and intuitive measure of performance would be to sum the residuals over all time periods. The obvious drawback to this is that the positive and negative residuals will cancel out, giving a distorted, optimistic view of the performance. A better approach would be to sum the absolute value of the residuals. A further distortion still exists in the approach. A given value of a residual will be more significant when the underlying value of the series, v_t, is low. Thus a residual of 10 when the observed flow is 100 will be more serious than when the observed flow is 1,000. To overcome this, the percentage absolute error may be more appropriate. Formally this can be written as eqn (2). Thus the mean error per forecast can be calculated by dividing E by the number of forecasts made, N.

$$E = 100 \sum_{t=1}^{N} \left| \frac{f_t - v_t}{v_t} \right| \tag{2}$$

NAIVE MODELS

Before embarking on a complex modelling exercise consideration should be given as to whether a simpler, almost naive, method may be more effective. For this study five such naive methods have been proposed.

NAIVE - 1

$$v_{t+1} = v_t \tag{3}$$

This simple method assumes that the best information about what is to happen is what is currently happening.

NAIVE - 2

$$v_{t+1} = 2\, v_t - v_{t-1} \tag{4}$$

This method applies the growth seen in the previous two time periods to the current period, in effect continuing a trend.

NAIVE - 3

$$v_{t+1} = 0.5\, v_t + 0.5\, v_{t-1} \tag{5}$$

Similar to naive method one, but the average of the two most recent observations is used, making it less sensitive to sudden changes in flow.

NAIVE - 4

$$v_{t+1} = 0.5\ v_t + 0.25\ v_{t-1} + 0.25\ v_{t-2} \tag{6}$$

A slight variant on method 3, but greatest weight is given to the most recent observation, and a residual weight to the remaining previous two observations.

NAIVE - 5

$$v_{t+1} = 0.25\ v_t + 0.25\ v_{t-1} + 0.25\ v_{t-2} + 0.25\ v_{t-3} \tag{7}$$

This is a naive method will a "long memory".

A danger is that any further variants on the above methods would no longer fulfil the criterion that the method is unsophisticated in structure, simple to apply and easy to comprehend.

STATISTICAL MODELLING

A range of statistical modelling techniques are available. This project selected the Box-Jenkins methodology for a number of reasons. The autoregressive component of the ARIMA model encompasses the technique of regression whilst the moving average component can provide a smoothing approach. Decomposition was not thought to be appropriate since no systematic components were apparent in the data.

The general form of an ARIMA(p,d,q) model is:

$$v_t - \mu_v = \sum_{i=1}^{p} \phi_i (v_{t-i} - \mu_v) - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{8}$$

where v_t is the flow differenced by d periods at time t;
μ_v is the mean flow;
ϕ_i, θ_i are parameters to be estimated;
ε_t is a $N(0,\sigma^2)$ noise term.

The scheme adopted for this study is to use 24 observations to estimate the parameters in the equation and then use this equation to produce forecasts for the next two observations. Thus observations 1 to 24 are used to forecast 25 and 26. Then observations 2 to 25 are used to forecast 26 and 27. This stepping continues until the end of the series is reached.

Data exploration

A sample of 50, twenty-four observation series were selected at random and a range of models were fitted to the data. It soon became clear that the series were non-stationary in their means and variances. The remedy for this situation was to take the first difference of the data and then apply a logarithmic transformation.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the transformed series were largely inconclusive. Little pattern in significant correlations emerged at low lags (1 or 2). Random significant spikes occasionally happened at large lags but given the short nature of the series, little value was given to these spikes.

Since the ACF and PACF were uninformative, three initial models were hypothesised:

ARIMA(1,1,0)   ARIMA(0,1,1)   ARIMA(1,1,1)

If any of these were good (ie significant parameter estimates) then an over-parameterised model would be evaluated:

ARIMA(2,1,0)   ARIMA(0,1,2)   ARIMA(2,1,1)   ARIMA(1,1,2)

and then judged against the corresponding first model set, using the significance of the estimates and the Akaike's Information criterion (AIC) (Akaike, 1974) and Schwarz's Bayesian criterion (SBC) (Schwarz, 1978) statistics. All analysis was performed using SAS 6.1 for Windows with Unconditional Least Squares estimation of the parameters.

The higher order models were inferior in all cases because:

(a) They were unstable and the estimation procedure failed to converge;
(b) The second order parameter was insignificant;
(c) Both the first and second order parameters became insignificant;
(d) The second order parameter was significant but the first order became insignificant; or
(e) The AIC/SBC values were poor in comparison to the lower order model.

This meant that the lower order models were always preferable.

The ARIMA(1,1,1) suffered from the following features:

(a) They were unstable and the estimation procedure failed;
(b) Estimates were produced at the invertibility region (ϕ_1 » 1.0 or θ_1 » 1.0); or
(c) Large correlations were found (>0.80) between the ϕ_1 and θ_1 parameters.

This meant that it was inferior to the pure (ie un-mixed) models.

Candidate Models

Six candidate models were identified to fit a logarithmic transformation of the series. These model were: ARIMA(0,0,1); ARIMA(1,0,0); ARIMA(1,0,1); ARIMA(0,1,1); ARIMA(1,1,0); ARIMA(1,1,1). Ten attempts can be made to fit some of these models to the available data. The first six are to assume that one model is correct for all observations in the series.

The seventh is to fit the model which gives the smallest sum of squares errors in the estimation procedure. This approach tends to favour the selection of models with the most parameters, which may not necessarily be the best at forecasting. An attempt to overcome this drawback is to derive a measure which penalises those models with a large number of parameters, such a measure is the Bayesian Information Criterion (BIC) (Mills, 1990).

$$BIC = \ln\left(\frac{SSE}{n}\right) + \frac{(p+q)\ln(n)}{n} \tag{9}$$

where p is the number of AR parameters;
q is the number of MA parameters;
n is the number of observations (n=24).

The appropriate choice is then the model of the six candidates which gives smallest BIC. This is criteria eight.

The ninth criteria for model selection is termed "hindsight". Here the model to fit at time t is the model which should have been fitted at time t-1. This means that when a new observation is available, its value is compared with the six forecasts made in the previous step. The model which gives the smallest one step ahead forecast should, with the benefit of hindsight, have been fitted. It is now too late to use this information but it can be assumed that the current step should use the model. In essence this is hypothesising that a model which is best for the period t to t+24 is also best for t+1 to t+25.

The tenth criteria is a scoring method (Beale, 1997a). Once a new observation is made available the model which, in the previous step, forecast closest to the observed value is accumulated a score of 5, the second best forecast accumulates a score of 4 and thus to a zero accumulated score for the worst performance. Then the model which has the highest accumulated score is selected for the current time step. The model with best overall performance will eventually be the most selected model.

An eleventh criteria is a refinement to the scoring method of Beale, where the best method is allocated a score of 10, second best a score of 6, third best 4, fourth best 3, fifth best 2 and worst 1. This scoring mechanism mimics that used in grand-prix racing.

Of course the best which can be achieved is when the model which gives the smallest one step ahead error is always chosen. In practice this method is not available since when the models are fitted to data from time t to t+24 (to forecast t+25), the observation at t+25 is not available. Using the privileged knowledge of t+25 enables a lower bound for E to be established and no better than this can be achieved, given the candidate models and data. In practice this privileged knowledge will not be available.

Association between model selections

For each of a representative sample of 1062 traffic situations, the candidate model which gave the least one-step ahead absolute percentage error in forecast was identified. This selection was then cross classified with the SOM cluster identified for the same observation. The cross classification is given in table 1.

Table 1 - Cross classification of best ARIMA model and SOM cluster

	(0,0,1)	(1,0,0)	(1,0,1)	(0,1,1)	(1,1,0)	(1,1,1)	ALL
Free flow	127	130	104	129	**196**	177	863
Congestion	23	**38**	14	9	16	19	119
Transition	6	6	2	**12**	4	3	33
Congestion	6	**15**	9	8	6	3	47
ALL	162	189	129	158	222	202	1062

In free flowing conditions, no one model clearly identifies itself as a suitable choice. The choice most often appropriate is an ARIMA(1,1,0). In the transition phase from free-flowing to congested conditions an ARIMA(0,1,1) is most appropriate. For both congested conditions an ARIMA(1,0,0) is appropriate. This table has therefore established a correspondence between which cluster an observation in placed in by the SOM and the type of ARIMA model to apply (given in **bold**). In testing, it is clearly inappropriate to apply this empirical relationship on the same data set on which it was established.

TESTING THE METHODOLOGY

A set of five day's data from three sites in late October 1996 was chosen for testing. Site 1 is the same as that used in establishing the relationship (but the *training* data was not used in this exercise). Site 2 is downstream of site 1 and site 3 is upstream of site 1. A fourth site is located within a junction, some distance from the above three and is from March 1997.

For each site a set of eighteen results are presented. The first five are the naive methods. The next six are the uniform, one model fits all, methods. The minimum SSE is the twelfth and the thirteenth is the minimum BIC. The fourteenth is the hindsight method and fifth and sixteenth are the scoring and grand-prix methods. Seventeen is the best possible result using privileged information. The last result is the SOM selected cluster. In table 2, the figure in the first column for each site is the total absolute percentage error over the whole series whilst the second column is the mean percentage error per observation. Within each site, the best performance statistic is highlighted in **bold**.

For the three within junction sites the SOM method has marginally produced the best performance, whilst it is only the fourth best for the within junction site. This may suggest that the correspondence between a SOM cluster and ARIMA model established using data from a between junction site may not be appropriate for a within junction site.

Table 2 - Performance of all methods using total absolute percentage error

	Site 1		**Site 2**		**Site 3**		**Site 4**	
Naive 1	3468	11.8	3622	12.3	3348	11.4	3143	**10.7**
Naive 2	4669	15.9	5193	17.7	4380	14.9	4678	15.9
Naive 3	3980	13.5	4075	13.9	3871	13.2	3412	11.6
Naive 4	4195	14.3	4300	14.6	4085	13.9	3673	12.5
Naive 5	4888	16.6	4991	17.0	4813	16.4	4439	15.1
ARIMA (0,0,1)	7567	25.7	7716	26.2	7425	25.3	7168	24.4
ARIMA (1,0,0)	3543	12.1	3731	12.7	3424	11.6	3233	11.0
ARIMA (1,0,1)	3685	12.5	3967	13.5	3552	12.1	3416	11.6
ARIMA (0,1,1)	3703	12.6	3972	13.5	3468	11.8	3739	12.7
ARIMA (1,1,0)	3469	11.8	3659	12.4	3345	11.4	3267	11.1
ARIMA (1,1,1)	3939	13.4	4471	15.2	3792	12.9	3617	12.3
Min SSE	3978	13.5	4481	15.2	3824	13.0	3788	12.9
Min BIC	3890	13.2	4388	14.9	3761	12.8	3916	13.3
Hindsight	3882	13.2	4254	14.5	3917	13.3	3611	12.3
Scoring	3565	12.1	3689	12.5	3433	11.7	3452	11.7
Grand-prix	3504	11.9	3699	12.6	3381	11.5	3546	12.1
Best	2151	7.3	2266	7.7	2072	7.0	1968	6.7
SOM	3342	**11.4**	3482	**11.8**	3254	**11.1**	3287	11.2

A PERFORMANCE INDEX

Of primary importance to the Highways Agency is that the forecasting method should be most accurate during the transition stage between free-flow and congested conditions (typically when the flow is between 4,000 and 6,000 vehicles per link per hour). Also accuracy is not required when the flows are stable or very erratic (ie changes in flows outside the range 10 to 20%). Thus the performance of a forecasting method is based on two components - its score and its weight (Beale, 1997b). The score captures the accuracy of the forecast whilst the weight captures the importance of the forecast. Each of these two components has an element due to the level of flow and also the change and direction of change in flow.

Each time point will yield a score between 0 and 1 on the accuracy of its prediction of the level of flow (S_f) and its prediction of in change of flow (S_d). Corresponding weights between 0 and 1 are also derived (W_f and W_d). The contribution to the performance indicator for this point is given by:

$$p_t = (S_f + S_d) * W_f * W_d \tag{10}$$

and the total value of the indicator, P_o, by:

$$P_o = \sum_{t=1}^{N} p_t \tag{11}$$

If the accuracy of the predictions was 100% then both S_f and S_d will be 1, so a maximum possible value on the performance indicator for a single point will be 2 * W_f * W_d. A corresponding maximum possible indicator for an entire series can then be calculated as P_{max}. The ratio of P_o to P_{max} measures the performance of the forecasting method. Table 3 shows the results of the naive method 5 (chosen by the Highways Agency) compared with the SOM method. Within the table, the percentage gain of the SOM method over the naive method is shown in the final column (% +SOM).

Table 3 - Performance index results

Link	Method	P_o	P_{max}	Index	% +SOM
Site 1	Naive 5	64.22	222.3	0.289	
	SOM	70.81	222.3	0.319	10.3
Site 2	Naive 5	61.93	209.3	0.296	
	SOM	66.26	209.3	0.317	7.0
Site 3	Naive 5	65.70	218.9	0.300	
	SOM	68.89	218.9	0.315	4.9
Site 4	Naive 5	43.02	152.1	0.283	
	SOM	46.15	152.1	0.303	7.3

In all cases the SOM method has produced a higher index value. A more detailed indication is given in figures 9 and 10. The solid line represents the flow whilst the dotted line is the forecast. The bars at the foot of the graph show the individual performance indicator (p_t) for the time point.

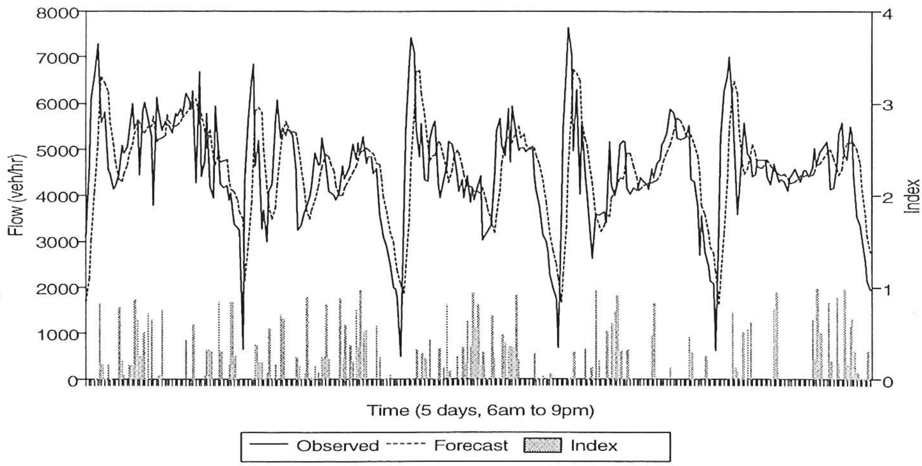

Figure 9 – Performance of naïve method 5

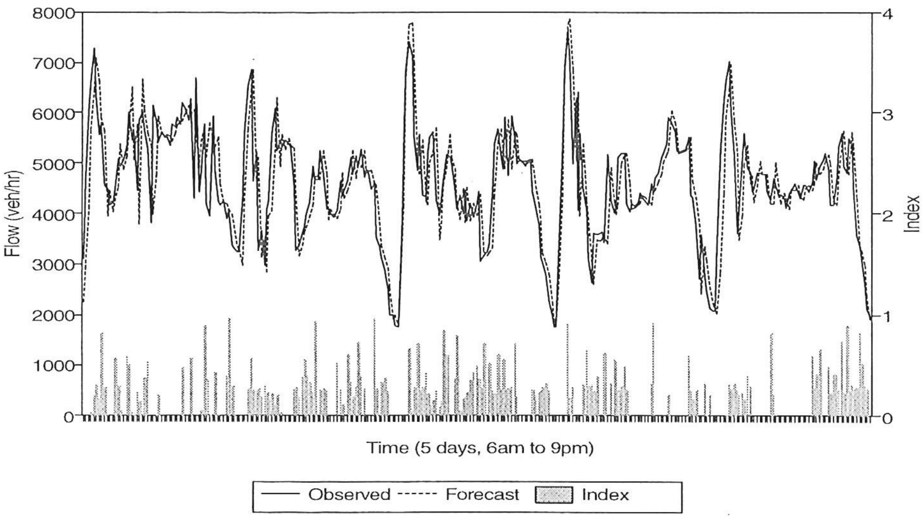

Figure 10 – Performance of SOM method

CONCLUSIONS

A comparison has been formed between the forecasting performances of a range of methods on the M25 traffic flow data, with the following conclusions:

- Using the self-organising map to "pre-cluster" the data prior to fitting an ARIMA model is a technique which has been found to work successfully with the data;

- A comparison between the result for the SOM method and those for a range of other methods indicated that the SOM method has performed best in three of the four sites considered (based on the total absolute percentage error as a measure of accuracy);
- As part of this research it has been possible to derive a new "performance index" which is more appropriate for traffic management needs than the statistical measure of the total absolute percentage error;
- For all four sites, it was found that the SOM method out performs the chosen naive method, based on the performance index;

A number of interesting issues have been raised within the course of the work which are deserving of further research. These include the precise definition of the performance index and the potential of "dynamic" neural networks for data of this type. It is hoped that results from future research into these issues will be reported at a later stage.

ACKNOWLEDGEMENTS

The authors acknowledge the assistance of the United Kingdoms' Highways Agency for the funding of this study and the source for figure 1. They would like to particular mention the help and guidance provided by Mr Stuart Beale of the Telematics division of the Highways Agency. The views expressed in this paper belong solely to the authors and do not necessarily represent those of the Highways Agency.

REFERENCES

Akaike, H (1974). A New Look at the Statistical Model Identification. **IEEE Transactions on Automatic Control**, AC-19, pp716-723.

Beale, S (1997a). Private communication at a Project Meeting.

Beale, S (1997b). Private communication at a Project Meeting.

Clark, SD, Dougherty, MS and Kirby, HR (1993). Use of Neural Networks and Time Series Models for Short Term Forecasting : A Comparative Study. **PTRC Annual Summer Meeting**, Vol p363, pp151-162.

Conner, JT, Martin, RD and Atlas, LE (1994). Recurrent Neural Networks and Robust Time Series Prediction. **IEEE Transactions on Neural Networks**, Vol 5, No 2, pp240-253.

Kohonen, T. (1988). Self-organised formation of topological correct feature maps. **Biological Cybernetics 43**, pp59-69. Reprinted in Anderson, JA and Rosenfeld, E (Eds) (1988). **Neurocomputing: Foundations of Research**. Cambridge, MA. MIT Press.

Maxwell, HA and Beck, I (1996). Traffic Control on the English Motorway Network. **Proceedings of the Eighth International Conference on Road Traffic Monitoring and Control**, 23-25 April 1996. Conference Publication No. 422. 1996. pp136-44.

Mills, TC (1990). **Time Series techniques for economists**. Cambridge University Press, pp138.

Nuttall, I (1995). Slow, Slow, Quick, Quick, Slow: Taking the 'Stop-start' Out of the London Orbital. **Traffic Technology International**. 1995/Winter. pp46-50.

Schwarz, G (1978). Estimating the Dimension of a Model. **Annals of Statistics**, Vol 6, pp461-464.

Van der Voort, M, Dougherty, M and Watson, S (1996). Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow. **Transportation Research C**, Vol 4. No 5 pp307-318.