ELSEVIER
International Journal of Forecasting 13 (1997) 43–50
international journal of forecasting

Should we use neural networks or statistical models for short-term motorway traffic forecasting?

Howard R. Kirby*, Susan M. Watson, Mark S. Dougherty[1]
Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK

Abstract

This article discusses the relative merits of neural networks and time series methods for traffic forecasting and summarises the findings from a comparative study of their performance for motorway traffic in France. Whilst it was possible to get a good performance with both neural networks and traditional Auto-Regressive Integrated Moving Average (ARIMA) models when forecasting up to an hour ahead using data supplied in 30-min intervals, a purpose-built pattern based forecasting model known as ATHENA, developed by INRETS, out-performed both these methods somewhat. The ways in which these models relate to the structure of traffic data are discussed and alternative paradigms are proposed.

Keywords: Traffic forecasting; ARIMA models; Neural networks; ATHENA model; Motorway flows

1. Introduction

In studies at the University of Leeds (for example Dougherty et al., 1994), neural networks have been found to be a convenient tool for developing relationships between streams of input data and streams of output data, not only for the pattern recognition kinds of tasks with which they are usually associated, but also for a range of modelling situations. As the functional relationship within a neural network is non-linear, it is of a more sophisticated form than many traditional linear statistical models. However, the transparency of defining an explicit functional form by the modeller at the outset of the modelling process, as occurs with traditional statistical modelling, can be perceived as an advantage. This raises the question of what the relative advantages and disadvantages of the different approaches are.

An initial comparative study of the different approaches was reported by Clark et al. (1993), for a set of 20 links in an urban road network (the city of Leicester). Inter-urban roads present perhaps a more challenging forecasting situation to address (partly because traffic can travel such great distances in half an hour, so there is greater variety in the extent and the density of information that could contribute to a forecast half an hour ahead). In this article, we report findings from a compatative study for an inter-

* Corresponding author.
[1] Present address: Centre for Research on Transportation and Society, Högskolan Dalarna, S78188 Borlänge, Sweden. Tel: +46 23 77 85 35; fax: +46 23 77 85 01.

PII S0169-2070(96)00699-1

urban motorway, at a site near Beaune, France. This site was chosen for the study because of the work already done by a team from Institut National pour la Recherche sur les Transports et leur Sécurité (INRETS), Paris, in developing a short-term forecasting model (known as ATHENA) for the site; the UK Science and Engineering Research Council (SERC) funded a study with the aim of comparing neural network techniques with those developed by the French team. The study was extended in scope to include the use of standard Auto-Regressive Integrated Moving Average (ARIMA) models.

The data sets and the experiments performed with them are described in Section 2 and Section 3, respectively. Section 4 describes the results obtained using the different modelling techniques. While the relative merits of neural networks versus traditional linear statistical models and clustering based methods such as ATHENA can be addressed just from the point of view of empirical performance with specific data sets, there are also some philosophical (and professional) questions. Following the conclusions in Section 5, the issue of whether these methods can yet handle the mathematical structures appropriate to the traffic situation being studied, is discussed in Section 6.

2. The data sets

The data sets came from four sites near Beaune (Fig. 1).

The data were provided by sensors of the

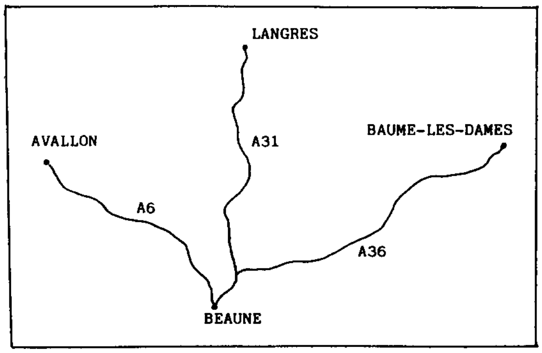

Fig. 1. Location of the French sites.

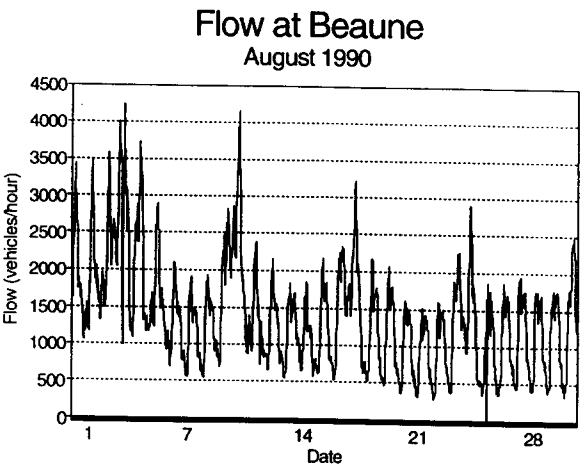

Fig. 2. Traffic flow data for August 1990.

SNRD (Système National de Recueil des Données, National Data Collection System), which provides traffic volume data for each half-hour. A total of 48 data points is thus used for each day, comprising the total number of vehicles of all types passing the collection point in a southbound direction. Although the system comprises 81 stations, only data for four sites were used in the original ATHENA model and thus in this study. The first was from a site downstream of a point where the flow along three feeder motorways converges onto a single motorway. Each feeder motorway had a single detector site approximately 30 km upstream of the convergence point. The data are essentially split into two sets, the first being July and August 1984–1989 and the second a 'test' set comprising the same months in 1990. No further pre-processing was deemed necessary for these data. An example of the flow data for August 1990 is given in Fig. 2.

3. The models

Traffic control and information systems for inter-urban highways need methods of forecasting a short period ahead. Such forecasts can be used for differing purposes depending on whether the context is urban, inter-urban or motorway traffic. In a motorway context, this could ensure that, for example, appropriate variable message signs are displayed, or in the case of severe congestion that traffic may be re-routed. The purpose of applying the models here was to make a prediction of flow at the downstream site. It was postulated that the behaviour observed upstream might be of help in making such a prediction, since the predictor would have advance warning of the impending approach of a block of traffic with different characteristics to that currently under observation. A standard time series method was compared to a neural network approach, and with the ATHENA pattern based forecasting method previously developed for the Beaune data set.

3.1. The ATHENA model applied to the Beaune data

Essentially the ATHENA approach uses an initial processing of the data, applying the *k*-means method (a mathematical clustering procedure minimising the Euclidean distance) to separate the data into homogeneous groups according to the shape of their historic hourly profiles before fitting linear regression models. Further details of the method are contained within Danech-Pajouh and Aron (1991). The unit of data used in the procedure is the same as that used for ARIMA and neural network modelling, i.e. the half hourly traffic flow at Beaune. The approach is novel in its prior treatment of data, as the conventional approach in analysing traffic patterns is to assume that homogeneous groups would be formed using 'calendar classification', i.e. that all Mondays would be inherently similar in the pattern of traffic flows, all Tuesdays would also form their own pattern and so on. However, it was found that such 'calendar' clustering was not as efficient as a method for separating the idea into homogeneous groups as a mathematical grouping based on the *k*-means method. It should be noted that the results for the ATHENA method presented in this article were supplied by the authors of the method, whose assistance we gratefully acknowledge, whilst the neural network and ARIMA modelling was carried out at Leeds.

3.2. Neural network modelling with the Beaune data

The neural network that was used for prediction, was a back-propagation technique (usually with one hidden layer), and was implemented on a commercially available simulation package, NeuralWorks Professional II Plus. The input layer was of varying width, depending on the number of input parameters. The output layer had just one processing element, the output of which was an estimate of the quantity being predicted. A suitable number of nodes in the hidden layer was determined by experiment; a typical number was about 30. Sigmoid transfer functions were used, and the delta learning rule.

A simple historical pattern of the previous six values of each data point was used as input to the network (i.e. the six preceding values of both the three upstream and one downstream data, yielding 24 data points). The use of six historical values was found in practice to be optimum as the improvement gained by using a longer historical look-back was negligible (see the article of Dougherty and Cobbett, 1996, in this issue for a description of the approach used to determine the optimum input set). The model was developed to forecast just one data point ahead; in this case this was equivalent to a half-hour look-ahead horizon.

3.3. ARIMA time series modelling with the Beaune data

The ATHENA model is a clustering method, its approach is different to that of classical time-series techniques. It was therefore considered desirable to compare its performance, and that of neural networks, to the performance of more standard time series models. Given that the data consists of half hourly flows for the Beaune downstream point which is supplemented with corresponding flows for three upstream points, the following ways of modelling the data were considered.

The ARIMA models reported here rely in their simplest form on the relationships between traffic flows through time at different lags. Although the model could easily be extended to include seasonal and other effects, in practice this did not have a substantial impact on the results. The model was thus given by a linear function of the previous Beaune flows through time. This could be classed as an univariate procedure. Where the model included a function of historical Beaune flows together with historical flows from other upstream points, that would be classed as a multiple procedure.

In addition, although it was possible to analyse July and August data as one series, it was decided to treat them separately at this stage in order to minimise the seasonal effect within the model. Data were originally recorded at half hourly intervals and this was used to produce initial modelling results. However, as past experience has suggested that hourly data may be a more appropriate unit for analysis, the original data was then subsequently aggregated to form hourly units for analysis.

4. Results of fitting the model

4.1. Diagnostic measures

The diagnostic measures used were of two kinds; those used *within* a given kind of modelling techniques, to aid selection of the most appropriate model of that kind; and those used *between* different techniques, to evaluate performance. Those used within the neural network technique are referred to in the article of Dougherty and Cobbett (1996); and those used with the ATHENA technique are described in Danech-Pajou and Aron (1991).

Diagnostics obtained in back-forecasting tests indicated that some improvements could still be expected on the ARIMA models, as some of the systematic variation remains within each data set. This was not only reflected in the non-zero mean and skewed nature of the residual series, but also in the χ^2 tests for white noise and examination of the correlation structure. In fact a simple time-plot of the residual series showed in each case some remaining cyclical structure and that several outlying (extreme) points may be exerting undue influence on model parameters.

For the ARIMA model fitting and assessment procedure, standard statistical diagnostics were employed. These included the autocorrelation function, inverse autocorrelation function, partial autocorrelation functions, χ^2 tests for white noise, t-statistics, correlation between parameter estimates and Akaike's information criterion (AIC) values. Differencing to accommodate trend and seasonal effects was also attempted, although this only proved effective for the half hourly July data. A similar underlying structure was found to be appropriate for both August and the hourly July data sets (i.e. autoregressive).

To compare the forecasts by different modelling methods, it is easy to suggest a range of possible diagnostic measures, but hard to choose between them. In order to produce a useful basic summary, the absolute percentage errors were computed for the forecasts from each method, where:

$$\text{Abs error} = \text{Absolute}\left[\frac{\text{residual}}{\text{observed}} \times 100\right]$$

Danech-Pajouh and Aron (1991) used this measure as the basis for comparing the ATHENA predictions with observed outcomes, and for that reason it was felt to be useful here. However, the question of what is the most appropriate comparative diagnostic deserves fuller study.

4.2. Results of the comparative evaluation

The results from the ARIMA models are compared with those obtained with neural networks and the ATHENA method, in Table 1.

As can be seen from Table 1, very little improvement was achieved by including the upstream data for the ARIMA fits and this was also found to be the case when seasonal factors were introduced. It is also clear that the ATHENA model fitted to the 60-min data marginally outperforms the ARIMA model, and easily outperforms the neural network model. Indeed, the 30-min forecast calculated using a neural network performs only slightly better than

Table 1
Summary of main performance results

Forecasting technique			% Frequency of absolute percentage error	
			0–20	20+
30 min	neural nets		91.5	8.5
	ARIMA (Jul)	Univariate	97.0	3.0
		Multiple	97.1	2.9
	ARIMA (Aug)	Univariate	95.4	4.6
		Multiple	95.5	4.5
60 min	neural nets		84.8	15.2
	ARIMA (Jul)	Univariate	93.6	6.4
		Multiple	94.8	5.2
	ARIMA (Aug)	Univariate	90.5	9.5
		Multiple	93.5	6.5
	ATHENA		97.5	2.5
120 min	ATHENA		89.1	10.9

the 2-h forecast calculated using the ATHENA method.

5. Conclusion

A comparison has been made here between the application of neural network, ARIMA and pattern-based ATHENA forecasting models to French motorway data. In many respects, the findings obtained are similar to findings from previous studies such as that carried out by Clark et al. (1993). Our experiments on the Beaune data led us to draw the following conclusions:

- Neural networks perform slightly worse than the traditional ARIMA time series when predicting 30 min ahead, and the improvement to ARIMA projections achieved by including upstream data is only slight.
- Neural networks also perform less well than time series models of the ARIMA type for 60-min projections using 30-min time-bands; and noticeably worse than pattern based time series technique of the kind used in the ATHENA model, which continues to do well for forecasts of up to 2 h ahead.
- The training data used here is particularly 'thin' for models of a neural network type and the poor performance by contrast with other techniques is not surprising (indeed, it was reassuring that the technique did as well as it did).
- The ATHENA model is particularly heavily specified, with a separate model being fitted for each hourly group profile; that is, it essentially consisted of 24 models. In contrast, only a single neural network was trained; and for the ARIMA models, the only clustering was into July/August months. In a study undertaken subsequent to that reported here (see Dochy et al., 1995) the ATHENA approach was extended, such that 24 neural networks (rather than regression-type models) were used following the pattern based clustering. The accuracy obtained was better than either the single neural network or the standard ATHENA method (as described here). This raises a more general issue about the degree of specification/parametrisation which is both appropriate and practical for the context in which the models are being applied.
- A further issue remains regarding the degree of expertise and automation in the modelling method. It can be argued that a certain amount of expertise is needed for each of the methods applied here, but that possibly the neural network is the most 'automatic' after the initial training period. This may imply it has certain practical advantages which are not matched by the ATHENA or ARIMA models.

- Whilst the absolute percentage errors were used as a comparative measure, it is yet to be determined whether a single measure like this is sufficient for use with transport time series where several indicators may be more appropriate.
- The level of aggregation in the data can affect different methods in different ways. For example, a coarse level of aggregation will remove high frequency variation in the time series, and so make it more likely that a simple statistical model will suffice to describe the variation that remains. However, the coarser the level of aggregation, the fewer the number of data points; and some valuable information about the structure in the data might be lost by the aggregation. Techniques such as neural networks, which require a large training set to 'learn' the relationships, are particularly penalised by coarser levels of aggregation. There were therefore *a priori* grounds for suspecting that, with comparatively little training data, an exceptionally good forecasting performance by neural networks might not be expected using the Beaune data.
- Whilst the numerical accuracy of the forecasts is obviously of great importance in deciding upon a particular forecasting methodology, it is by no means the only factor to be considered. Other issues such as development time, the level of skill required to extract good performance from the technique, and transparency of the model are also significant.

It is not possible to generalise from the results of this study to suggest which approach may be best in circumstances. The arguments are finely balanced, and affected by characteristics of the chosen data set. The Beaune data set is substantially different from the Dutch data sets which had been used to develop the neural network methodology (see the article by Dougherty and Cobbett, 1996); the latter provided data at 1-min intervals, compared with the 30-min aggregates of the French data. Thus, the relatively poor performance of neural networks with the 30-min French data may not have been found had the comparatively study been based on this 1-min Dutch data. So an overall conclusion is that more extensive comparative evaluation studies would be worthwhile.

6. Discussion

One of the issues often discussed in any comparison of neural network methods with conventional statistical models, is the supposed advantage that the latter enjoy, by virtue of having a transparent and explicit structure; whereas neural networks are often maligned as being a 'black box' with their structure hidden from the user. (For a discussion of these and related issues see Ripley, 1992, 1994). In the context of this traffic application, it is worth considering this issue of 'structure' more deeply. Consider a section of road, on which traffic observed at a point i will have passed census points on preceding sections $(i-1, i-2, \ldots, i-H$ and so on), and that one has flow estimates $y_{i,t}$ for all i, and at intervals t, $t-1$ and so on.

The conventional time series approach would propose a relationship of the form

$$y_{i,t} = \sum_{h=0}^{H} (\phi_{i-h,1} y_{i-h,t-1} + \phi_{i-h,2} y_{i-h,t-2} + \cdots + \phi_{i-h,p} y_{i-h,t-p}) + \varepsilon_t$$

However, such an approach takes no account of the structural relationship that exists between a measurement at one point and a measurement at another; namely, that vehicles observed at one site move physically to another, *in a time that varies with traffic conditions*; and moreover that that time can be estimated from the speed estimates obtained at the sensor sites.

In other words, the lags, k, (in the expressions of the kind $\phi_{i,k} y_{i,t-k}$) should *not* be regarded as fixed quantities, but as ones that are related to the speed measurements.

This clearly complicates the set of forecasting equations considerably, providing a model form which existing techniques would not be able to fit. Hence we see that, no other how well time series techniques might appear to fit the traffic data, their functional form is not necessarily one that is consistent with traffic characteristics.

This suggests of course that an alternative approach, perhaps one based on modelling the dispersion of traffic, may provide a better modelling paradigm.

There is one other respect in which the scope of the comparative analyses needs to be extended. Given that, if forecasting, say, 30 min ahead, one very soon has feed-back on the adequacy of the forecast, methodologies have been developed (for both neural networks and statistical models) that allow the parameters of a fitted model to be continuously adapted to reflect changes in the performance of the model. For a review of time series techniques that have been developed for this purpose (for example, using Bayes methods, adaptive forecasting and Kalman filtering) see Watson (1993). To some extent, the comparisons have been a little unfair to the neural networks and Box–Jenkins methodology, because neither of these techniques provided 'adaptive forecasts', whereas the ATHENA model was (in part at least) an adaptive forecasting procedure.

Acknowledgements

We are grateful to the staff in INRETS, Paris, in particular Thierry Dochy and Mehdi Danech-Pajouh, for provision of the Beaune data and assistance with the ATHENA model.

The findings presented here were previously presented in draft form at the US Transportation Research Board meeting in Washington, January 1994. The authors would like to thank the UK Engineering and Physical Sciences Research Council (EPSRC), whose predecessor body, the Science and Engineering Research Council (SERC) funded the work on which this article is based (grant refs. G/G/27270 and G/H/82549); and partners in the European Commission's General European Road Data Information Exchange Network (GERDIEN) project (DRIVE-II project V2044) for allowing us to continue working with the Utrecht data set on the SERC project.

References

Clark, S.D., M.S. Dougherty and H.R. Kirby, 1993, The use of neural networks and time series models for short term traffic forecasting: a comparative study, in: *Transportation Planning Methods, Proceedings, PTRC 21st Summer Annual Meeting*, September 1993, Manchester (PTRC Education and Research Services Ltd., London).

Danech-Pajouh, M. and M. Aron, 1991, ATHENA: a method for short-term inter-urban motorway traffic forecasting, *Recherche Transports Sécurité*, English Issue 6, 11–16.

Dochy, T., M. Danech-Pajouh and Y. Lechevallier, 1995, Short term road traffic forecasting using neural networks, *Recherche Transports Sécurité*, English Issue 11, 73–82.

Dougherty, M.S. and M.R. Cobbet, 1996, Short-term inter-urban traffic forecasts using neural networks, *International Journal of Forecasting*, 13, 21–31.

Dougherty, M.S., H.R. Kirby and R.D. Boyle, 1994, Using neural networks to recognise, predict and model traffic, in: M. Bielli, G. Ambrosino and M. Boero, eds., *Artificial Intelligence Applications to Traffic Engineering* (VSP, Utrecht, The Netherlands), 233–250.

Ripley, B.D., 1992, Statistical aspects of neural networks, in: *Proceedings of Semstat*, Sandbjerg, Denmark, April 1992, Denmark (Chapman and Hall, London).

Ripley, B.D., 1994, Neural networks and related methods for classification (with discussion), *Journal of the Royal Statistical Society* B56(3), 409–456.

Watson, S.M., 1993, An introduction to adaptive forecasting techniques, *ITS Technical Note 233* (Institute for Transport Studies, University of Leeds, Leeds).

Biographies: Howard R. KIRBY is Deputy Director (Research) at the Institute for Transport Studies, University of Leeds, where he is responsible for the planning, development and management of a wide spectrum of research initiatives. A scientist by training, he has conducted and managed an extensive range of research projects, in three main areas: transport modelling, applications of advanced information technology and safety and environmental issues. In the transport modelling area, principal research interests centre on the statistical and mathematical aspects of modelling aspects of modelling and forecasting travel demand, the use of simplified area-wide models, validation techniques and empirical investigations of factors affecting peoples' travel choices. In the advanced information technology area, principal research interests have been the development of novel methods for capturing data and monitoring traffic and the development and evaluation of artificial intelligence based techniques, such as neural networks.

Susan M. WATSON is a statistician (Ph.D., B.Sc. (Hons), C. Stat.) by training and is currently a lecturer based in the

Institute for Transport Studies, University of Leeds. She has been involved in numerous research projects including the application of time series models within traffic control systems and other transport data, data collection and evaluation methodologies for transport schemes, economic appraisal of new road construction and statistical theory underlying stated preference techniques. Her main interest is the detection and treatment of outlying and missing observations in time series transport data. A chartered statistician, she is also leader of the statistics subject group within the Institute.

Mark S. DOUGHERTY graduated in Computer Science from Cambridge University. Since his appointment as research fellow at the Institute for Transport Studies, University of Leeds, he has worked on a variety of projects applying artificial intelligence to the transport field. He has applied logic programming to constraint satisfaction problems such as traffic signal design. He was involved in the development of the knowledge-based intelligent traffic supervisor CLAIRE. Through several projects he has helped pioneer the use of neural networks in the transport community, with emphasis on the problems of short-term forecasting, parameter estimation and behavioural modelling. Recently he took up an appointment as a senior research fellow in the Centre for Research on Transportation and Society at Borlänge, Sweden. He continues an association with ITS through an appointment as an honorary research fellow.