PERGAMON

Atmospheric Environment 33 (1999) 709–719

ATMOSPHERIC ENVIRONMENT

Neural network modelling and prediction of hourly NO_x and NO_2 concentrations in urban air in London

M.W. Gardner*, S.R. Dorling

School of Environmental Sciences, University of East Anglia, Norwich, UK

Received 26 March 1998; accepted 1 June 1998

Abstract

Multilayer perceptron (MLP) neural networks were trained to model hourly NO_x and NO_2 pollutant concentrations in Central London from basic hourly meteorological data. Results have shown that the models perform well when compared to previous attempts to model the same pollutants using regression based models. This work also illustrates that MLP neural networks are capable of resolving complex patterns of source emissions without any explicit external guidance. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Air quality modelling; Nitrogen oxides; Primary pollutant; Multilayer perceptron; Artificial neural network

1. Introduction

Nitrogen oxides ($NO_x = NO + NO_2$) are emitted into the urban atmosphere primarily from vehicle exhausts. Primary NO_x emissions are mostly in the form of nitric oxide (NO) which then reacts with ozone (O_3) to form nitrogen dioxide (NO_2). Much work has been carried out to determine the factors which control NO_x and NO_2 concentrations in order to enable the development of tools to aid in the forecasting of pollutant concentrations. One approach to predict future concentrations is to use a detailed atmospheric diffusion model. Such models aim to resolve the underlying physical and chemical equations controlling pollutant concentrations and therefore require detailed emissions data and meteorological fields. Collet and Oduyemi (1997) provide a detailed review of this particular type of model. The second approach is to devise statistical models which attempt to determine the underlying relationship between a set of input data (predictors) and targets (predictand). Regression modelling is an example of such a statistical approach and has been applied to air quality modelling and prediction in a number of studies (Shi and Harrison, 1997; Ziomass et al., 1995) . One of the limitations imposed by linear regression models is that they will underperform when used to model non-linear systems (Gardner and Dorling, 1998). Artificial neural networks can model non-linear systems and have been used with some success to model sulphur dioxide concentrations in Slovenia (Boznar et al., 1993). Comrie (1997) has compared multilayer perceptron models with more traditional regression models for ozone forcasting. In this work, since the relationship between NO_x, NO_2 and meteorology is complex and extremely non-linear, artificial neural networks were used to model and predict hourly NO_x and NO_2 concentrations from readily observable local meteorological data.

2. Artificial neural network models

There exists a plethora of neural network architectures (Jain et al., 1996). In this work multilayer perceptron

* Corresponding author.

PII: S1352-2310(98)00230-1

(MLP) neural networks were used, details of which can be found in Gardner and Dorling (1998). MLP neural networks are capable of modelling highly non-linear relationships and can be trained to accurately generalise when presented with new, unseen data. MLP neural networks learn to model a relationship during a supervised training procedure, when they are repeatedly presented with series of input and associated output data. In the case of modelling pollutant concentrations the input data could consist of measurements of meteorological conditions and the output would be the pollutant concentration. To use the trained MLP neural network for prediction involves presenting the network with a set of forecast meteorological data. For this reason it is important to use meteorological data that are routinely forecasted.

In this study the MLP neural network models were trained on data for a particular site and can therefore only be used with confidence at the site in question. The local meteorological conditions that determine the processes controlling the pollutant behaviour will vary between sites. Knowledge of the spatial and temporal scales of pollutant behaviour is important in determining the extent to which the model predictions can be extrapolated (Rao et al., 1997; Vukovich, 1997). Shi and Harrison (1997) suggest that the results from their linear regression model developed using data from a Central London site were representative over at least the whole Central London area.

The MLP neural networks used in this study were trained using the scaled conjugate gradient algorithm (Moller, 1991), which is implemented in the Stuttgart Neural Network Simulator. This simulator allows neural networks to be easily designed and trained, runs under the UNIX operating system and is freely available via the internet (ftp: //ftp.informatik.uni-stuttgart.de). The scaled conjugate gradient algorithm is efficient and has a high probability of finding a good solution and unlike many neural network training algorithms it is not sensitive to the choice of training parameters. This algorithm has been shown to be considerably faster than the commonly used standard backpropagation algorithm, as used by Strachan and Stedman (1997), and also to be faster than other conjugate gradient methods (Moller, 1993).

When training MLP neural networks it is important to avoid overtraining of the networks. Overtraining occurs when the model learns the noisy details in the training data which results in the model having poor generalisation capabilities when presented with new data. The objective when training the neural network is to extract the underlying patterns in the data. This assumes that the training data is adequately extensive and also representative. Neural networks perform well when used for interpolation, but poorly in cases of extrapolation. The training data must fully represent all cases about which the network is required to generalise. In order to avoid overtraining the networks it is sensible to divide the data into three partitions – a training, validation and test set. The training set forms the bulk of the data. The validation set is used during training in order to check the generalisation performance. Training can be stopped when the performance on the validation data reaches a maximum. Finally the test data are the data upon which the final model is tested. In this work, due to limited availability of meteorological data, the models were trained on data from 1990. Data from 1991 were used as both the validation and test data sets. It is noted that without truely independant test data the model performance statistics will be artificially biased. In order to determine the degree of bias the generalisation performance was assessed by calculating various difference statistics and also bootstrap estimates of standard error. These estimates were calculated by randomly resampling the test data 1000 times, with replacement, following the methodology described in Wilks (1995).

Throughout this work all the MLP neural networks were of the same architecture with two hidden layers each of which contained 20 nodes. These networks were larger than necessary for several reasons. Since theoretically only one hidden layer is required. (Hornik et al., 1989) it is highly probable that smaller networks could be successfully trained. However, networks with more than one hidden layer can be trained faster and are more likely to avoid local (inefficient) solutions. The networks all require one input per input variable and all have a single output representing the pollutant concentration. The transfer function used in the hidden layer nodes was the hyperbolic tangent function, and for the input layer and output layer the function was the unbounded identity function. Using large networks to speed up the training time requires that care is taken to avoid overtraining (Sarle, 1995). Additional issues relating to the practical use of multilayer perceptrons are addressed in Gardner and Dorling (1998).

3. Data

In order to enable a direct comparison with the regression models developed by Shi and Harrison (1997) data were selected that were as close as possible to those used in their original study. Hourly NO_x and NO_2 data were obtained from the Department of the Environment, Transport and the Regions (DETR) automatic monitoring network between 1990 and 1991 for two monitoring sites in Central London (http: //www.aeat.co.uk/netcen/aqarchive/auto.html). The data from both sites were combined to produce one series of data. This was necessary due to the fact that the Central London Laboratory monitoring site switched to Bridge Place during July 1990. The sites are very close to each other and data from both sites were treated in this manner by Shi and Harrison (1997).

Hourly meteorological data were obtained for the same period for London Weather Centre from the British Atmospheric Data Centre (http: //www.badc.rl.ac.uk/). It was not possible to exactly match the meteorological variables used by Shi and Harrison (1997). However, in our opinion those used matched closely enough to enable a fair comparison of results and would, in any case, be considered to be a less comprehensive set of predictor variables. The meteorological variables used in this work were as follows:

- Low cloud amount (LOW): oktas.
- Base of lowest cloud (BASE): presented to the network in the United Kingom Meteorological Office (UKMO) synoptic code shown in Table 1.
- Visibility (VIS): presented to the network in the UKMO synoptic code shown in Table 2.
- Dry bulb temperature (DRY): °C.
- Vapour pressure (VP): mbar.
- Wind speed (WS): m s^{-1}.

The meteorological variables used by Shi and Harrison (1997) were wind speed, wind direction, air temperature, relative humidity, total radiation, the Pasquill Stability Category and the boundary layer depth. In this paper the emphasis has been placed upon using a minimal set of meteorological predictors that are readily observed at almost all meteorological stations in the UK. Variables such as boundary layer depth, radiation and stability indices are not readily available even though they have been shown to be important in determining NO_x and NO_2 concentrations (Derwent et al., 1995; Ziomass et al., 1995). It was anticipated that the provision of the base of the lowest cloud, windspeed and visibility would together act as proxies for boundary layer depth. Similarly the provision of cloud amount and time of the day could indicate likely total radiation levels.

Shi and Harrison (1997) went to some effort to approximate the diurnal nature of NO_x emissions which they naturally considered as an important predictor for both the NO_x and NO_2 concentrations. As a comparison exercise neural networks were trained with and without this emissions factor. Instead of the emissions factor a network was given two additional time of day inputs consisting of the sine and cosine of the time of day normalised between 0 and 24 h. The provision of these two components allows the cyclical nature of the variable to be realised by the MLP neural network. Details of the derivation of the emissions factor are given in appendix A.

In order to be used with the MLP neural networks all the data were normalised into the range $-1.0 + 1.0$. This was carried out by determining the maximum and minimum values of each variable over the whole data period and calculating normalised variables using the following formula:

$$x_{\mathrm{norm}} = 2 * \left(\frac{(x - x_{\mathrm{min}})}{(x_{\mathrm{max}} - x_{\mathrm{min}})} \right) - 1.0.$$

The data were later returned to original units using the formula:

$$x = \left(\frac{(x_{\mathrm{norm}} + 1.0) * (x_{\mathrm{max}} - x_{\mathrm{min}})}{2} \right) + x_{\mathrm{min}}.$$

Observations of the height of the low cloud base were not made during times of very poor visibility. In this work missing low cloud base information is substituted by a value of $+1.0$ which is the maximum value the normalised value can attain and represents a cloud base above 2500 m or no cloud. It was assumed that in general there would be little low cloud when visibility was poor in Central London – typically under anticyclonic conditions when radiative cooling and a stable boundary layer would enable fog or mist to form, although very occasionally this might not be true.

Table 1
UKMO synoptic code for reporting height of lowest cloud

Code	Height/(m)
0	0–50
1	50–100
2	100–200
3	200–300
4	300–600
5	600–1000
6	1000–1500
7	1500–2000
8	2000–2500
9	Above 2500 m or no clouds

Table 2
UKMO synoptic code for reporting visibility

Code	Visibility
00	Visibility <0.1 km
01–50	Visibility over the range 0.1–5.0 km, at 0.1 km intervals e.g. 01 = 0.1 km, 02 = 0.2 km and 50 = 5.0 km
60–80	Visibility over the range 10–30 km, at 1 km intervals e.g. 60 = 10 km, 70 = 20 km and 80 = 30 km
81–88	Visibility over the range 35–70 km, at 5 km intervals e.g. 81 = 35 km, 82 = 40 km and 88 = 70 km
89	>70 km

4. MLP neural network models – results and discussion

4.1. No emissions factor

Two MLP neural networks (MLP1 and MLP2) were trained on the data as shown in Table 3. These networks were given no explicit information concerning the likely emissions of NO_x or other pollutants. It was expected that the neural networks would be able to learn the pattern of emissions using the time of day inputs.

Table 4 shows the performance statistics of the trained networks when used to predict pollutant concentrations for 1991. The correlation statistics were similar for both the NO_2 and NO_x models and suggest that 47% of the variability in the NO_2 and 54% of the variability in the NO_x concentrations can be attributed to the variation in the meteorological predictors. Shi and Harrison (1997) do not provide any indication of the magnitude of the predictive errors for their models and it is not clear whether their correlations refer to the logarithmically transformed pollutant concentrations or whether the data have been converted back into concentration units. It should be pointed out that simple correlation statistics are meaningless without any indication of the magnitude of residual errors (Willmott, 1982; Willmott et al., 1985). However, in terms of the reported correlation coefficient, it is clear that models MLP1 and MLP2 were outperforming the ordinary least squares (OLS) models described in their paper. The OLS models also used the emissions factor, something that was not included here, and for NO_2 predictions, O_3 and NO concentrations were used as input predictors to their models. For comparison the OLS models could explain 32% and 16% of the NO_x and NO_2 variability, respectively. It is also interesting to note that the MLP2 model outperformed an autoregressive model developed by Shi and Harrison (1997) which included the previous hour lagged NO_x concentrations as an input variable.

Since the MLP models developed in this work did not use exactly the same input meteorological data as Shi and Harrison (1997) used, multiple linear regression (LR) models were developed using the same data upon which the MLP models were trained – as listed in Table 3. The LR models were designed in the most simple manner and no attempt was made to transform the variables. The performance statistics for the LR models were calculated by testing the models using the independant data from 1991 and are presented in Table 5. These statistics can be directly compared with those presented in Table 4. Surprisingly, the simple LR models developed here appear to outperform those developed by Shi and Harrison (1997). There are several possible explanations for this. Firstly, contrary to our earlier expectations, the variables used in this work could be better predictors of NO_2 and NO_x than the variables used in Shi and Harrison (1997). Secondly the logarithmic data tranformations they adopted could degrade the models rather than helping. Finally, the period upon which the models were tested could

Table 3
Training data for multilayer perceptron (MLP) models and multiple linear regression (LR) models

Model ID	Input data	Output data
MLP1,LR1	sin(h), cos(h), LOW, BASE, VIS, DRY, VP, WS	NO_2
MLP2,LR2	sin(h), cos(h), LOW, BASE, VIS, DRY, VP, WS	NO_x
MLP3,LR3	LOW, BASE, VIS, DRY, VP, WS, Q	NO_2
MLP4,LR4	LOW, BASE, VIS, DRY, VP, WS, Q	NO_x
MLP5,LR5	sin(h), cos(h), LOW, BASE, VIS, DRY, VP, WS, $NO_2(t-1)$	$NO_2(t)$
MLP6,LR6	sin(h), cos(h), LOW, BASE, VIS, DRY, VP, WS, $NO_x(t-1)$	$NO_x(t)$
MLP7,LR7	sin(h), cos(h), LOW, BASE, VIS, DRY, VP, WS, $NO_2(t-24)$	$NO_2(t)$
MLP8,LR8	sin(h), cos(h), LOW, BASE, VIS, DRY, VP, WS, $NO_x(t-24)$	$NO_x(t)$

Table 4
MLP model performance statistics calculated over the whole of 1991 comparing actual and predicted pollutant concentrations. MAE – mean absolute error/ppb, RMSE – root of the mean-squared error/ppb, r^2 – coefficient of determination. Subscript week MLP denotes the statistics calculated for the MLP models over the 7 day period 25th–31st May 1991, whilst subscript week SH denotes the performance of the models of Shi and Harrison (1997) also in that week. Bootstrap estimates of standard error are given in brackets

Model	MLP1	MLP2	MLP3	MLP4
MAE	9.8(0.3)	38.5(1.5)	9.9(0.3)	38.5(1.4)
RMSE	18.2(1.8)	78.6(6.0)	17.9(1.8)	77.5(6.0)
r^2	0.47(0.04)	0.54(0.03)	0.47(0.04)	0.55(0.03)
r^2_{weekSH}	*n/a*	*n/a*	0.16	0.32
$r^2_{weekMLP}$	0.44	0.42	0.47	0.46
$MAE_{weekMLP}$	8.5	27.1	8.5	27.8
$RMSE_{weekMLP}$	14.1	51.1	13.7	49.5
Model	**MLP5**	**MLP6**	**MLP7**	**MLP8**
MAE	4.3(0.1)	18.1(0.6)	9.5(0.3)	39.0(1.5)
RMSE	7.3(0.6)	33.8(2.3)	17.1(1.7)	80.2(5.1)
r^2	0.91(0.02)	0.92(0.02)	0.51(0.04)	0.62(0.04)
r^2_{weekSH}	0.69	0.42	*n/a*	*n/a*
$r^2_{weekMLP}$	0.87	0.88	0.44	0.28
$MAE_{weekMLP}$	4.3	14.6	8.8	29.5
$RMSE_{weekMLP}$	6.7	23.8	13.9	55.6

Table 5
Multiple linear regression model performance statistics calculated over the whole of 1991 comparing actual and predicted pollutant concentrations. MAE – mean absolute error/ppb, RMSE – root of the mean-squared error/ppb, r^2 – coefficient of determination. Subscript week LR denotes the statistics calculated for the LR models over the 7 day period 25th–31st May 1991, whilst subscript week SH denotes the performance of the models of Shi and Harrison (1997) also in that week. Bootstrap estimates of standard error are given in brackets

Model	LR1	LR2	LR3	LR4
MAE	10.9(0.3)	44.9(1.7)	10.3(0.3)	43.5(1.7)
RMSE	19.3(1.8)	97.0(7.7)	18.7(1.8)	95.7(7.8)
r^2	0.38(0.03)	0.34(0.02)	0.41(0.04)	0.35(0.02)
r^2_{weekSH}	*n/a*	*n/a*	0.16	0.32
r^2_{weekLR}	0.29	0.29	0.37	0.26
$\mathrm{MAE}_{\mathrm{weekLR}}$	9.6	31.8	9.0	32.8
$\mathrm{RMSE}_{\mathrm{weekLR}}$	15.5	55.2	14.5	56.2
Model	**LR5**	**LR6**	**LR7**	**LR8**
MAE	4.6(0.1)	18.6(0.6)	10.1(0.3)	42.5(1.6)
RMSE	7.4(0.5)	34.3(2.3)	17.7(1.7)	89.5(6.8)
r^2	0.90(0.01)	0.91(0.01)	0.46(0.04)	0.47(0.02)
r^2_{weekSH}	0.69	0.42	*n/a*	*n/a*
r^2_{weekLR}	0.84	0.89	0.29	0.24
$\mathrm{MAE}_{\mathrm{weekLR}}$	4.6	13.1	9.8	33.3
$\mathrm{RMSE}_{\mathrm{weekLR}}$	7.2	21.5	15.3	57.1

represent a "difficult" week and may not be representative of the more long term model behaviour as calculated in this work. To assess these effects the model performance statistics were calculated using the same one-week test period adopted by Shi and Harrison (25th–31st May 1991). These statistics are presented in the lower portion of the results Table 5. Clearly, this particular period was one in which the models experienced some difficulty, as indicated by the relatively poor performance compared with the annual statistics. However the important point to note is that the regression models developed here performed in all but one instance better than Shi and Harrisons' models.

To ensure that the MLP model performance statistics are comparable with the linear regression models of Shi and Harrison the performance statistics for the MLP models were also calculated for the 7 day period. These statistics are presented at the bottom of Table 4, and illustrate that the MLP models perform more consistently throughout the test year than the LR models.

This exercise has demonstrated that neural network models are producing considerably better predictions than conventional linear regression models using both similar and identical input meteorological data. It appears that the lack of any information concerning diurnal NO_x source strengths does not seriously hinder the MLP neural network model. It is likely that the MLP neural network is picking up the diurnal variation in source strength by associating time of day with emissions. This would be unlikely to occur in other traditional models where such a complex interaction between predictor variables is not permitted. As a comparison MLP models were developed in the next section that use an emissions factor rather than the time of day inputs.

4.2. With emissions factor

Two MLP models (MLP3 and MLP4) were trained without time of day inputs and using a NO_x emissions factor as an additional input. The details of these models are shown in Table 3. Multiple linear regression models were also developed (LR3 and LR4). The emission factor derivation is outlined in appendix A. For each time of day the estimated emissions factor was inserted into the training data. Weekday and weekend differences were also accounted for. Shi and Harrison (1997) argue that the variation in the source strength must be included in this type of model.

The trained MLP models were tested on the data for 1991 and the performance statistics are presented in Table 4. The performance of both models (MLP3 and MLP4) were extremely similar to the performance of the models with no emissions factor (MLP1 and MLP2). This leads to the conclusion that the emissions factor could be replaced by time of day inputs when using MLP neural networks without any detrimental effects – in this work it would appear marginally beneficial. A major advantage of not using an emissions factor but allowing the model to learn the nature of emissions is the ease with which the model can be constructed. There is no requirement to attempt to quantify the nature of emissions from statistical averages. If the nature of emissions change over time then the emissions factors and regression models will require recalculating. Also there may be more subtle meteorological influences determining the nature of emissions, for example more people may drive to work when it is cold and wet. It is possible that this type of interaction between the input variables is one reason why the neural network approach is outperforming linear regression.

4.3. General discussion

Fig. 1 shows the performance of the MLP1 and MLP2 models during December 1991, a month including a severe smog episode in London (Bower et al., 1994; Derwent et al., 1995) making it an interesting time period over which to look at model performance. It is apparent from Fig. 1 that both the models were performing well during non-episode conditions. However during the episode between the 11–17 December the NO_2 model and to a lesser extent the NO_x model were

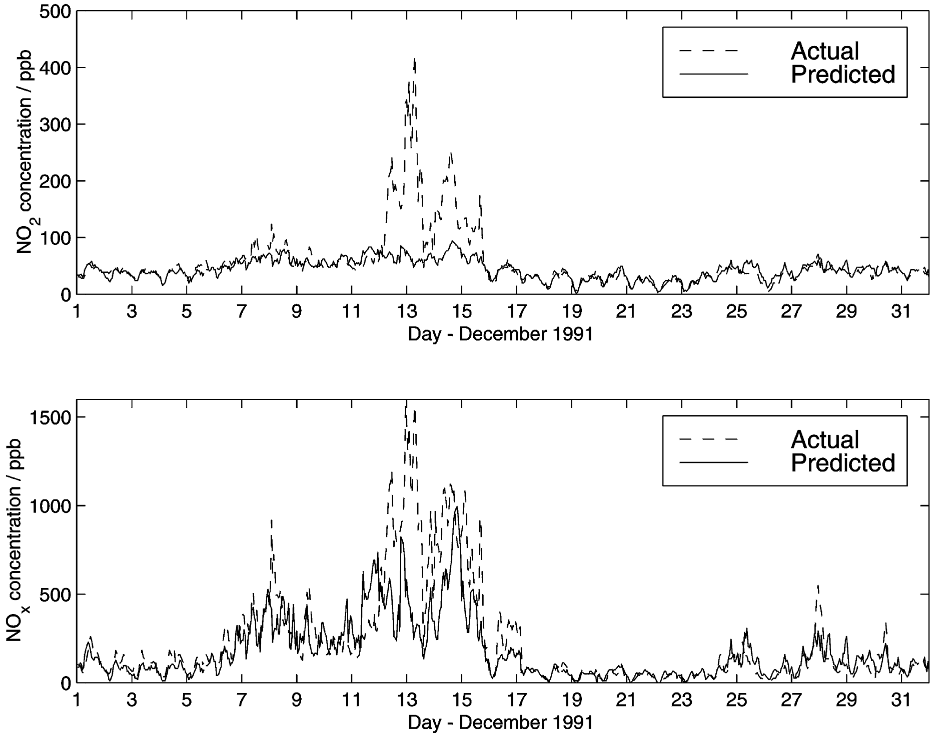

Fig. 1. Comparison of MLP1 and MLP2 models measurements. For NO_2 the correlation between actual and predicted concentrations is 0.62 and for NO_x the correlation is 0.77.

underpredicting the severity of the episode. Modelled concentrations of NO_2 did not exceed 100 ppb and NO_x did not exceed 1100 ppb, compared with actual concentrations of the order 420 ppb and 1570 ppb, respectively. This illustrates one of the drawbacks of statistical modelling techniques – the models will only perform well when used to predict within the bounds of the training data. The exceptional conditions that resulted in this pollution episode were uncommon; similar conditions did not occur during 1990 when the model was trained. It is now known that unusual NO_x chemistry was taking place due to the extreme cold and high NO concentrations (Bower et al., 1994). It is important therefore to ensure that the training data is adequately representative and extensive so that such events can be resolved. Plots displaying the models performance throughout the whole of 1991 are available via http: //www.uea.ac.uk/env/pubs/mwgø1/.

It was interesting to note that the addition of an hourly sunshine input to the MLP models did not significantly improve the models performance. This indicates that our choice of proxies for radiation levels, boundary layer stability and depth, as described earlier, are working as intended. This is an important result since it demonstrates that the MLP models can, to a certain extent, make up for deficiencies in the input data. The lack of intuitively important variables does not necessarily restrict the likelihood of developing a reasonable model given the availability of a number of suitable proxies. This is less likely to be the case when developing regression models where interactions between input variables are not permitted.

4.4. *Auto-regressive models*

Plots of the autocorrelation function (Fig. 2) of the MLP1 and MLP2 model residuals show considerable autocorrelation of the residuals. This suggests that a poor (or good) model prediction will tend to be followed by another. One way to correct for this behaviour is to supply the model with information concerning previous pollutant concentrations. This way the model will be constrained by the previous pollutant concentrations and series of poor predictions should be avoided.

Shi and Harrison (1997) use an autoregressive (AR) model with an extra 1 h lagged pollutant predictor variable. As a comparison to these AR models two neural network models (MLP5 and MLP6) were trained with the same data as models MLP1 and MLP2 but with an additional 1 h lagged concentration input. Simple linear regression models (LR5 and LR6) were developed using the same data. The performance statistics of the trained network are presented in Table 4. Fig. 3 shows the autocorrelation function for the MLP5 and MLP6 model residuals from which it can be seen that there is now very little autocorrelation of the residuals. The temporal information provided by including previous concentration data to the model has corrected the tendency of the model to make a run of poor (or good) predictions.

Both MLP5 and MLP6 models performed extremely well when used for prediction, explaining 91% and 92% of the variability in the hourly NO_2 and NO_x concentrations, respectively. The performance during December 1991 is shown in Fig. 4. Unfortunately 1 h ahead forecasts are of little practical use (except perhaps in nowcasting) and bearing in mind the high lag-1 h autocorrelation apparent in the NO_x and NO_2 time series these results are not surprising. Two more neural network models (MLP7 and MLP8) were developed to assess the performance of forecasts that could be made using 24 h lagged pollutant concentrations. Such models should give a more reasonable indication of how these MLP neural network models would perform in a practical forecasting role. Obviously there will be some further degradation of performance due to the errors that will be inherent in the forecast meteorological data required by the models to predict pollutant concentrations.

The performance of models MLP7 and MLP8 are listed in Table 4. It is evident that these models performed better than the models with no previous concentration data but they do not perform to the levels of MLP5 and MLP6 which had the 1 h lagged pollutant data as an additional input. Fig. 5 shows the model performance during December 1991. This figure shows the slight improvement made in the predictions during the pollution episode compared to MLP1 and MLP2 shown in Fig. 1.

Fig. 6 shows the autocorrelation function of the residuals for the models MLP7 and MLP8. It is apparent from this plot that there is still some degree of autocorrelation of the residuals however this is to a lesser extent than with models MLP1 and MLP2. The main difference is the removal of the 24 h lag correlation peak and the more rapid decrease in the autocorrelation function.

5. Conclusions

This work has shown that MLP neural networks can accurately model the relationship between local meteorological data and NO_2/NO_x concentrations in

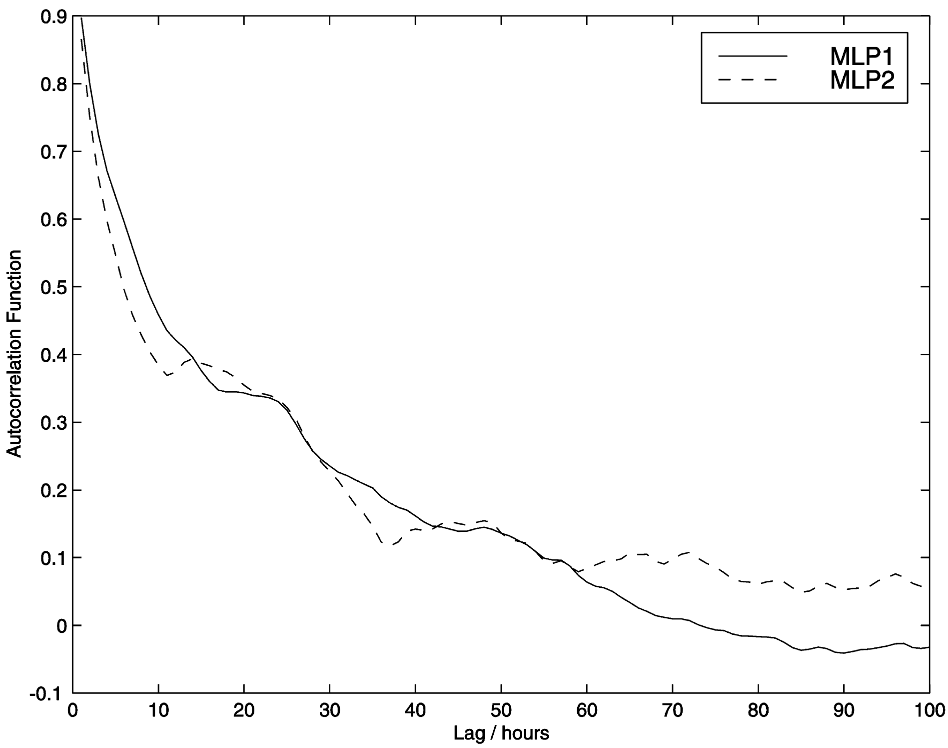

Fig. 2. Residual autocorrelation for MLP1 and MLP2 models.

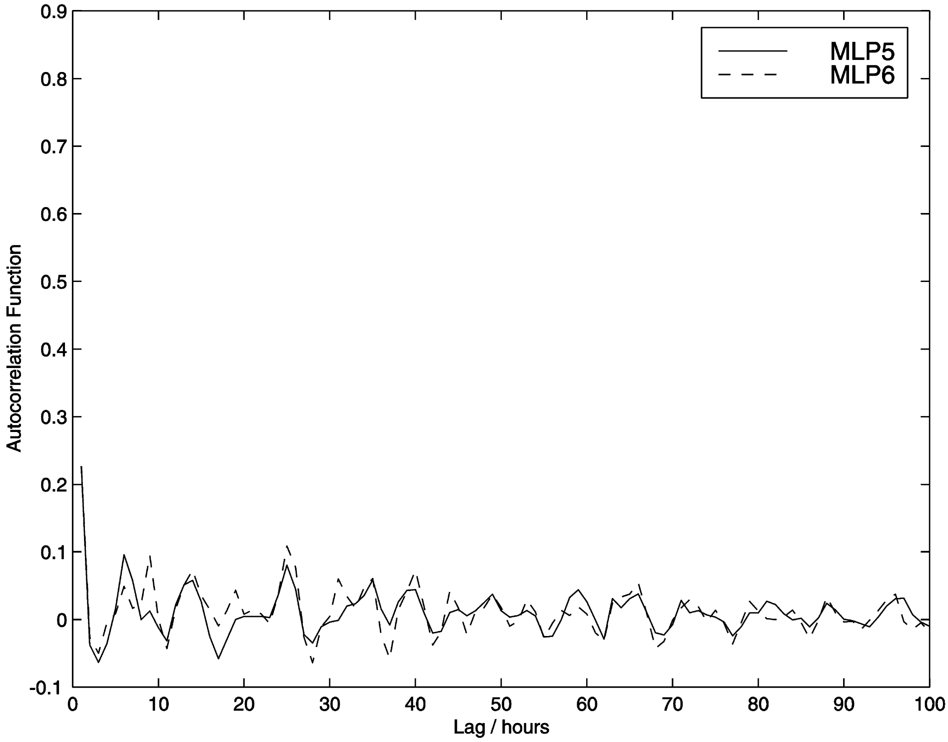

Fig. 3. Residual autocorrelation for MLP5 and MLP6 models.

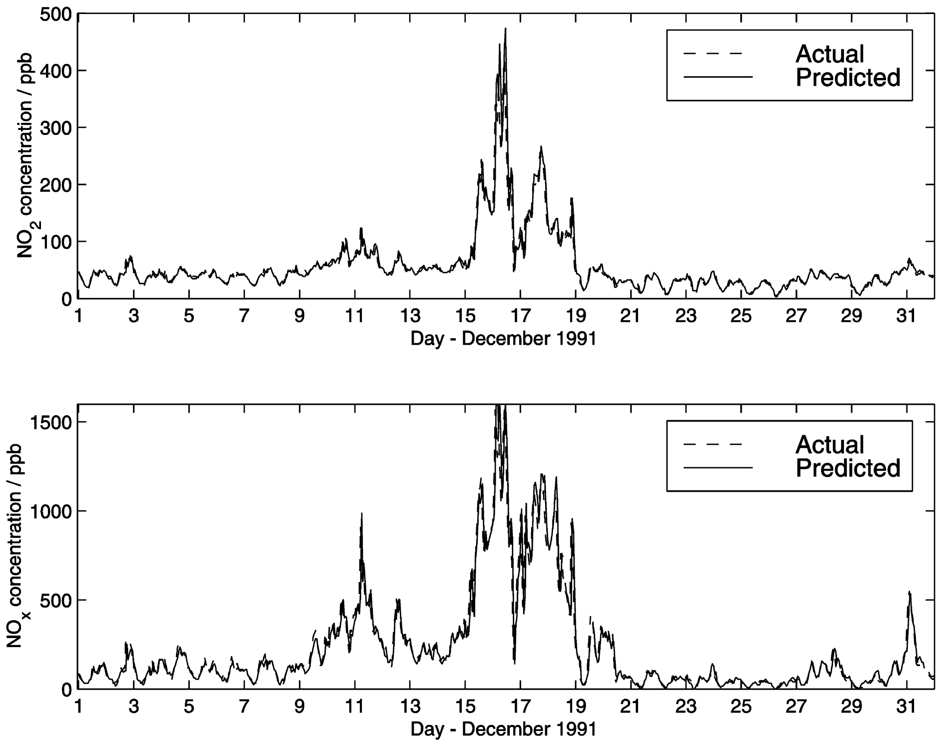

Fig. 4. Comparison of MLP5 and MLP6 models with measurements. For NO_2 the correlation between actual and predicted concentrations is 0.96 and for NO_x the correlation is 0.97.

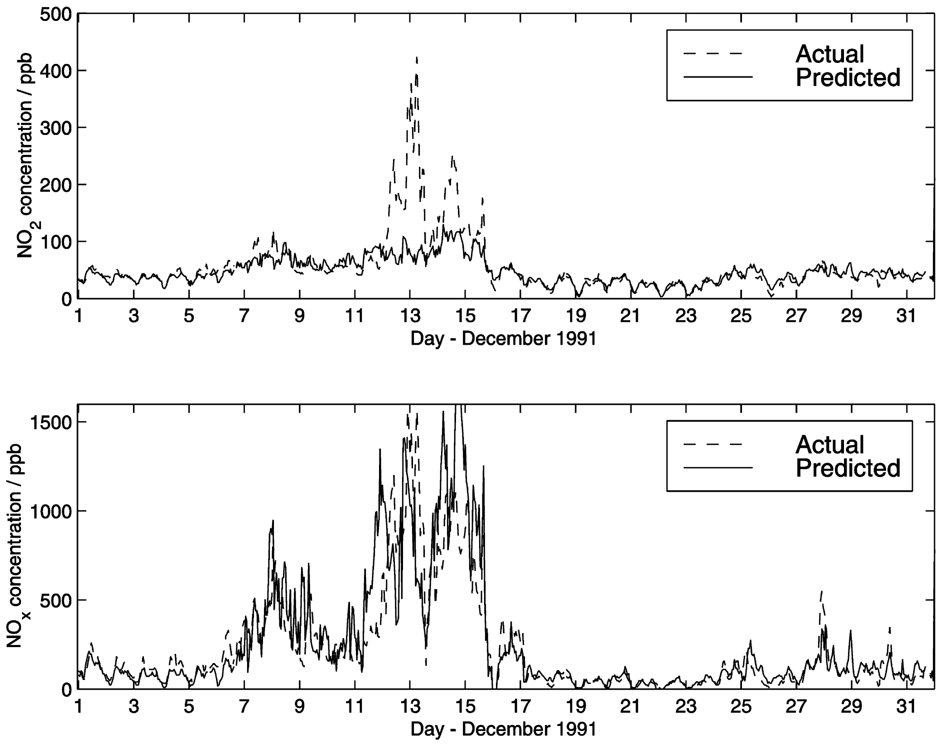

Fig. 5. Comparison of MLP7 and MLP8 models with measurements. For NO_2 the correlation between actual and predicted concentrations is 0.70 and for NO_x the correlation is 0.84.

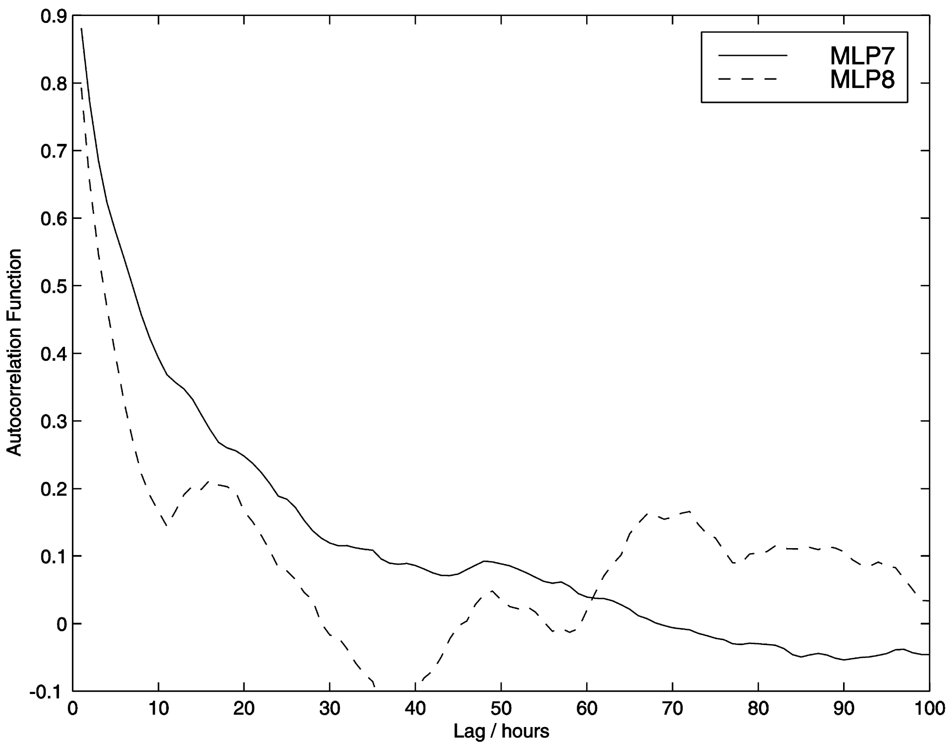

Fig. 6. Residual autocorrelation for MLP7 and MLP8 models.

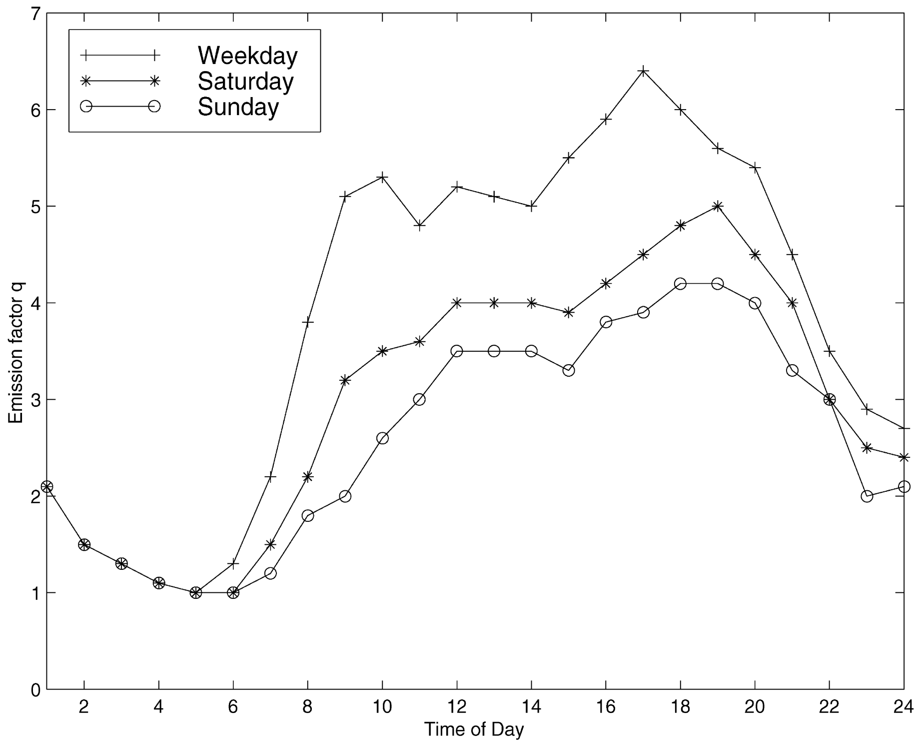

Fig. 7. Derived NO_x emissions factor.

an urban environment. The models are seen to learn the underlying pattern of emissions without any external guidance. This enables the models to be easily constructed.

It has also been demonstrated that MLP neural networks offer several advantages over traditional multiple linear regression models. These include the ability of MLP models to make efficient use of proxy data when the optimum predictor variables are unavailable.

Tests using previous lagged concentrations illustrate that whilst the best predictions are made by models with additional 1 h lagged pollutant concentration inputs, the models with 24 h lagged pollutant inputs manage to predict pollution episodes reasonably well. Such models could be implemented as air quality forecast models to provide 24 h forecasts of pollutant levels. Our additional successful experience in modelling secondary as well as primary pollutants in both urban and rural settings is demonstrating the versatility of the MLP neural network approach. More work is ongoing to assess the degradation of the air quality predictions due to the inherent errors in forecast meteorological data (which would be required in an operational mode) as well as the development of models trained on gridded meteorological model output.

Acknowledgements

We are grateful to BADC and DETR/NETCEN for making suitable data freely available and also the School of Environmental Sciences, University of East Anglia, for supporting this work.

Appendix A

Shi and Harrison (1997) derived an emissions factor which was used as an input to linear regression models developed to predict hourly NO_x and NO_2 concentrations. They argued that it was important to be able to provide the models with some indicator of source strength. By averaging the NO_x concentrations for Central London over a whole year clear diurnal cycles and weekday/weekend differences were observed. These differences can be related to both meteorology and source strength variability. Shi and Harrison (1997) use a simple box model in an inverse manner (Derwent et al., 1995) to calculate approximate emission rates for each time of the day and for each day of the week. This emissions factor was then averaged for weekdays, all of which exhibited similar behaviour and finally the emissions factors were arbitrarily scaled.

The emissions factors calculated by Shi and Harrison (1997) were used in this study since their calculation requires knowledge of the boundary layer depth which was not available in this work and this allowed good comparison of techniques. The emissions factors calculated by Shi and Harrison (1997) from data between June 1989 and May 1990 are shown in Fig. 7 for weekdays, Saturdays and Sundays.

Reference

Bower, J.S., Broughton, G.F.J., Stedman, J.R., Williams, M.L., 1994. A winter NO_2 smog episode in the U.K. Atmospheric Environment 28 (3), 461–475.

Boznar, M., Lesjak, M., Mlakar, P., 1993. A neural network–based method for the short–term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment 27B (2), 221–230.

Collet, R.S., Oduyemi, K., 1997. Air quality modelling: a technical review of mathematical approaches. Meteorological Applications 4, 235–246.

Comrie, A.C., 1997. Comparing neural network and regression models for ozone forecasting. Journal of the Air and Waste Management Association 47, 653–663.

Derwent, R.G., Middleton, D.R., Field, R.A., Goldstone, M.E., Lester, J.N., Perry, R., 1995. Analysis and interpretation of air quality data from an urban roadside location in central London over the period from July 1991 to July 1992. Atmospheric Environment 29 (8), 923–946.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences, Atmospheric Environment 32 (14/15), 2627–2636.

Hornik, K., Stinchcoombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Jain, A.K., Mao, J., Mohiuddin, K.M., 1996. Artificial neural networks – a tutorial. Computer, 31–44.

Moller, M.F., 1991. A scaled conjugate gradient algorithm for fast supervised learning. IEEE Transactions on Systems. Man and Cybernetics 21, 272–280.

Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6, 525–533.

Rao, S.T., Zurbenko, I.G., Neagu, R., Porter, P.S., Ku, J.Y., Henry, R.F., 1997. Space and time scales in ambient ozone data. Bulletin of the American Meteorological Society, 78 (10), 2153–2166.

Sarle, W.S., 1995. Stopped training and other remedies for overfitting. Proceedings of the 27th Symposium on the Interface. Available via ftp: //ftp.sas.com/pub/neural/inter95.ps.Z.

Shi, J.P., Harrison, R.M., 1997. Regression modelling of hourly NO_x and NO_2 concentrations in urban air in London. Atmospheric Environment 31 (24), 4081–4094.

Strachan, I., Stedman, J.R., 1997. The development of a neural network model for forecasting UK air quality, Technical Report, AEAT/2346/20008001/005, Culham: AEA Technology, National Environmental Technology Centre.

Vukovich, F.M., 1997. Time scales of surface ozone variations in the regional, non-urban environment. Atmospheric Environment 31 (10), 1513–1530.

Wilks, D.S., 1995. Statistical methods in the atmospheric sciences: an introduction, Academic Press, International Geophysics Series, San Diego, CA.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society 63 (11), 1309–1313.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. Journal of Geophysical Research 90 (C5), 8995–9005.

Ziomass, I.C., Melas, D., Zerefos, C.S., Bais, A.F., 1995. Forecasting peak pollutant levels from meteorological variables, Atmospheric Environment 29 (24), 3703–3711.