PERGAMON

AE International – Europe

Atmospheric Environment 37 (2003) 4539–4550

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modelling system and measurements in central Helsinki

Jaakko Kukkonen[a,*], Leena Partanen[a], Ari Karppinen[a], Juhani Ruuskanen[b], Heikki Junninen[b], Mikko Kolehmainen[b], Harri Niska[b], Stephen Dorling[c], Tim Chatterton[c], Rob Foxall[d], Gavin Cawley[d]

[a] *Finnish Meteorological Institute, Air Quality Research, Sahaajankatu 20E, Helsinki FIN-00810, Finland*
[b] *Department of Environmental Sciences, University of Kuopio, PL 1627, Kuopio FIN-70211, Finland*
[c] *School of Environmental Sciences, University of East Anglia, Earlham Road, Norwich NR4 7TJ, UK*
[d] *School of Information Systems, University of East Anglia, Earlham Road, Norwich NR4 7TJ, UK*

Received 17 February 2003; received in revised form 1 July 2003; accepted 3 July 2003

Abstract

Five neural network (NN) models, a linear statistical model and a deterministic modelling system (DET) were evaluated for the prediction of urban NO_2 and PM_{10} concentrations. The model evaluation work considered the sequential hourly concentration time series of NO_2 and PM_{10}, which were measured at two stations in central Helsinki, from 1996 to 1999. The models utilised selected traffic flow and pre-processed meteorological variables as input data. An imputed concentration dataset was also created, in which the missing values were replaced, in order to obtain a harmonised database that is well suited for the inter-comparison of models. Three statistical criteria were adopted: the index of agreement (IA), the squared correlation coefficient (R^2) and the fractional bias. The results obtained with various non-linear NN models show a good agreement with the measured concentration data for NO_2; for instance, the annual mean of the IA values and their standard deviations range from 0.86 ± 0.02 to 0.91 ± 0.01. In the case of NO_2, the non-linear NN models produce a range of model performance values that are slightly better than those by the DET. NN models generally perform better than the statistical linear model, for predicting both NO_2 and PM_{10} concentrations. In the case of PM_{10}, the model performance statistics of the NN models were not as good as those for NO_2 over the entire range of models considered. However, the currently available NN models are neither applicable for predicting spatial concentration distributions in urban areas, nor for evaluating air pollution abatement scenarios for future years.

Keywords: Neural network; Urban air; NO_2; PM_{10}; Model evaluation

1. Introduction

During recent years, statistical models including those based on artificial neural networks (NN) have been increasingly applied and evaluated for the regression analysis and forecasting of air quality. In their overview of applications of NN in the atmospheric sciences,

*Corresponding author. Tel.: +358-9-1929-5405; fax: +358-9-1929-5403.
E-mail address: jaakko.kukkonen@fmi.fi (J. Kukkonen).

1352-2310/$ - see front matter
doi:10.1016/S1352-2310(03)00583-1

Gardner and Dorling (1998) concluded that NN generally give as good or better results compared with statistical linear methods, especially where the problem being analysed includes non-linear behaviour. The NN methods can also be used in combination with traditional deterministic modelling techniques.

Gardner and Dorling (1999) tested the benefits of using a multi-layer perceptron (MLP) NN approach to model NO_2 concentrations in London, relative to other statistical modelling approaches. They found out that the temporal variation of emissions could be represented by using the input variables of time of day and day of week. In addition, simple meteorological input variables were used, providing some indication of atmospheric stability, without the need for processing of the measured meteorological data. The MLP models consistently outperformed a linear regression approach.

Kolehmainen et al. (2001) evaluated various computational models using hourly concentration time series of NO_2 and basic meteorological variables collected for the city of Stockholm in 1994–1998. They concluded that the MLP NN yielded more accurate regression analysis and forecasting of air quality, compared with the results obtained using the self-organising map (SOM) or a linear time series method. Gardner and Dorling (1999) obtained a similar result using a different set of data.

The review by Gardner and Dorling (1998) does not include any applications of NN to the modelling of particulate matter. Perez et al. (2000), however, compared the forecasting of air quality for fine particulate matter produced by three different methods: a multilayer NN, linear regression and persistence methods (the latter assigns hourly values for the subsequent day to be equal to the equivalent values for the current day). The three methods were applied to the hourly averaged $PM_{2.5}$ data for the years 1994–1995 measured at one location in the downtown area of Santiago, Chile. The NN gave the best results overall in the forecasting of the hourly concentrations of $PM_{2.5}$.

Gardner (1999) undertook a model inter-comparison using linear regression, MLP and classification and regression tree (CART) approaches for hourly PM_{10} modelling in Christchurch, New Zealand, for the period 1989–1992. The MLP method outperformed CART and linear regression across the range of performance measures employed. The most important predictor variables in the MLP approach were time of day, temperature, vertical temperature gradient and wind speed.

The study presented here is part of an EU-funded research project "Air Pollution Episodes: Modelling Tools for Improved Smog Management-APPETISE" (2000–2002; http://www.uea.ac.uk/env/appetise/), reviewed by Greig et al. (2000). The APPETISE project has represented the first concerted attempt to undertake a model inter-comparison exercise between advanced statistical and present day deterministic air quality modelling approaches. The final aim of the project was to produce recommendations on the suitability of various models, or various classes of models, for specific applications. The APPETISE project concentrated on four pollutants: nitrogen oxides, particulate matter, tropospheric ozone and sulphur dioxide.

Schlink et al. (2003) have performed a model inter-comparison exercise within the APPETISE project for the statistical regression of tropospheric ozone, using 14 different statistical modelling techniques and a deterministic Lagrangian trajectory model including chemistry. Ten measurement sites were selected, located in Germany, Italy, United Kingdom and the Czech Republic. The authors recommended those methods that are able to model static non-linearities; these include NN and generalised additive models. The best predictions were obtained for multi-variate approaches using observed meteorological data.

This paper focuses on the model evaluation and inter-comparison for NO_2 and PM_{10} concentrations in urban areas. We only address the "now-casting" of air quality, i.e., air quality forecasting using numerical weather forecasting models is outside the scope of the study. The main motivation of this study was practical, regarding the comparison of the numerical performance of various statistical methods and the deterministic modelling system (DET). Both kinds of methods are currently used in regulatory air pollution forecasting by the local authorities worldwide. However, there have been no rigorous inter-comparisons in previous scientific literature of the methods based on artificial NN and the deterministic modelling methodologies for this purpose.

The term "prediction" is used in this paper to mean establishing the relationship between observed independent variables (predictors, such as meteorological variables) and an observed dependent variable (predictand; in this case concentration). When the predictors are forecast by some method, we can "forecast" or "predict" the predictand. Computations using statistical methods could be interpreted as fitting the data with predictors.

2. Materials and methods

2.1. Models

2.1.1. Emission inventory and deterministic models

We have used an emission inventory of NO_x in the Helsinki Metropolitan Area (Karppinen et al., 2000a) that has been updated to include data for the years 1996 and 1997. The inventory includes the emissions from various mobile sources (road traffic, harbours and marine traffic, and aviation) and stationary sources (power plants, other point sources and residential heating). The vehicular emission factors are based on the LIISA modelling system (Mäkelä et al., 1996; Laurikko, 1998). The computations included approximately 5000 line sources, 169 point sources, area sources and the regional background concentrations.

We have utilised a modelling system developed for evaluating the traffic volumes, emissions from stationary and vehicular sources, and atmospheric dispersion of pollution in an urban area. The dispersion modelling is based on a combined application of the road network dispersion model CAR-FMI (e.g., Kukkonen et al., 2001), applied for evaluating the dispersion of pollution originating from vehicular traffic, and the Urban Dispersion Modelling system UDM-FMI (e.g., Karppinen et al., 2000a), for evaluating the dispersion from stationary sources. Both of these models are so-called new-generation multiple source Gaussian urban dispersion models. The relevant meteorological parameters for the models are evaluated using boundary layer scaling variables produced by a meteorological pre-processing model, adapted for the urban environment (Karppinen et al., 2000c).

The performance of this DET has been previously evaluated by comparing the model predictions with results of the air quality-monitoring network in the Helsinki Metropolitan Area in 1993 (Karppinen et al., 2000b), and in 1996 and 1997 (Kousa et al., 2001). The performance of individual models within the system has been evaluated also against the results of several field measurement campaigns (e.g., Kukkonen et al., 2001; Tiitta et al., 2002).

2.1.2. NN and other statistical models

We present here a brief review of the artificial NN models utilised in this study. For a more detailed description of the properties of NN, the reader is referred to textbooks such as Bishop (1995), Haykin (1999) and Hecht-Nielsen (1991).

We have applied five NN models and one linear statistical model. The NN methods are more complex and flexible than linear statistical models, and as a result these are able to model strongly non-linear dependencies characterising the data. A second advantage of NN models is their ability to learn complex and even a priori unknown relationships directly from the training data. On the other hand, statistical linear methods are computationally simpler and easier to implement.

Regarding the NN models, we have selected the feed-forward back-propagation MLP for use in this study; this model category is abbreviated simply as MLP in the following. This choice was based on previous studies, such as Gardner and Dorling (1999) and Kolehmainen et al. (2001). The main criterion for selecting the MLP model for air quality prediction was its accuracy and reliability, compared with other available NN model categories.

An MLP model consists of a network of simple processing elements and connections. The processing elements, called neurons, are arranged in layers; these are the input layer, the so-called "hidden" layer or layers, and the output layer. Each neuron computes a weighted sum of the inputs it receives from neighbouring neurons, processes this using an "activation" function, and distributes the result to the subsequent layer. The results obtained using a MLP model depend on the architecture of the network and model parameters, such as the size of the NN, the training algorithm, and transfer functions used in the hidden and output layers. We have therefore conducted the computations in this study using several types of MLP models.

Before NN models can be utilised for predictions, these models require an initial learning step that optimises the connection weights between neurons so as to minimise a cost function, which measures the importance of the difference between the desired and the actual output of the network. The teaching mechanism used in the MLP model is normally based on some form of gradient descent minimisation of the cost function. The most common cost function, the sum-of-squares error, is based on the assumption that the observed data is the result of the observation of some deterministic system that is corrupted by a Gaussian noise process with constant (homoscedastic) variance. If the variance of the noise process is assumed to vary with time, an additional output can be used to model the local variance of the noise process, as described in Foxall et al. (2002).

If the variance of the noise term varies in time in the training step of an NN, the model is called an NN with heteroscedastic Gaussian noise; the corresponding acronym that is used in the following is NN-HeG. The NN-2HeG and NN-3HeG are corresponding models assuming two- and three-component mixture heteroscedastic Gaussian noise, respectively. In the so-called homoscedastic NN models, the training is based on the assumption of a constant variance of the noise signal; this is called as a homoscedastic NN with Gaussian noise; NN-HoG. In NN-Lapl the noise term is assumed to follow a Laplacian distribution. In the linear model (LIN) the predicted outputs are simply a linear combination of the inputs.

The LIN assumes that the pollutant concentration can be approximated by a weighted sum of the predictor variables. The use of a LIN as the benchmark for model comparison purposes allows us to assess whether the use of non-LINs, such as artificial NN, is justified. A pure persistence model can be expressed as a special case of the LIN, and so generally we should expect the LIN to out-perform the pure persistence model. The LIN therefore provides a more rigorous benchmark than the pure persistence model.

2.2. Experimental data

The location of the air quality monitoring stations in the Helsinki Metropolitan Area and the pollutants monitored in 1999 are presented in Fig. 1. The measurements at these stations were conducted by the Helsinki Metropolitan Area Council. Most of these stations were located with the purpose of monitoring "hot spots" in the vicinity of the busiest traffic environments, or major local energy production sources. The network contains five permanent multi-component stations, located at Luukki, Leppävaara, Töölö, Tikkurila and Vallila. Mobile stations are re-located each calendar year.

The stations represent urban traffic (Töölö, Vallila, Sörnäinen, Herttoniemi) and suburban traffic (Leppävaara, Tikkurila and Myyrmäki) environments, suburban residential environments (Tapiola) and urban background environments (Kallio 2). Regional background concentrations were monitored in a rural environment at Luukki, approximately 20 km to the northwest of Helsinki town centre.

We have selected the stations of Töölö and Vallila to be used in this study. The criteria for selecting these stations were that these are permanently located and represent the most polluted parts of the city, i.e., urban traffic environments in central Helsinki. However, neither of these stations is located in a street canyon. The modelling of air quality at urban traffic stations is more demanding, compared with that for suburban or urban background stations (Kousa et al., 2001).

The station of Töölö is situated in a small square that is situated in the middle of a busy junction, surrounded by several buildings. This station is situated at a distance of less than 10 m from busy streets, with traffic volumes of approximately 50 000 vehicles per day. The station of Vallila is situated in a small park, at the distance of about 14 m from a street with a traffic volume of 13 000 vehicles per day. Both of the traffic volumes mentioned are average values during weekdays in 1999.

The input data includes traffic flow, concentrations of NO_2 and PM_{10}, and meteorological data. The datasets consist of sequential hourly time series of concentrations and meteorological variables. The time period selected covers four years, from 1996 to 1999. All traffic flow, air quality and pre-processed meteorological data are presented as hourly averaged values.

2.2.1. Replacement of missing concentration values

Clearly, the conclusions based on comparisons of model predictions against measured data are dependent on the quality of the data. The fraction of missing concentration data ranged 1–5% for all sets of data measured during any specific year and station. Although

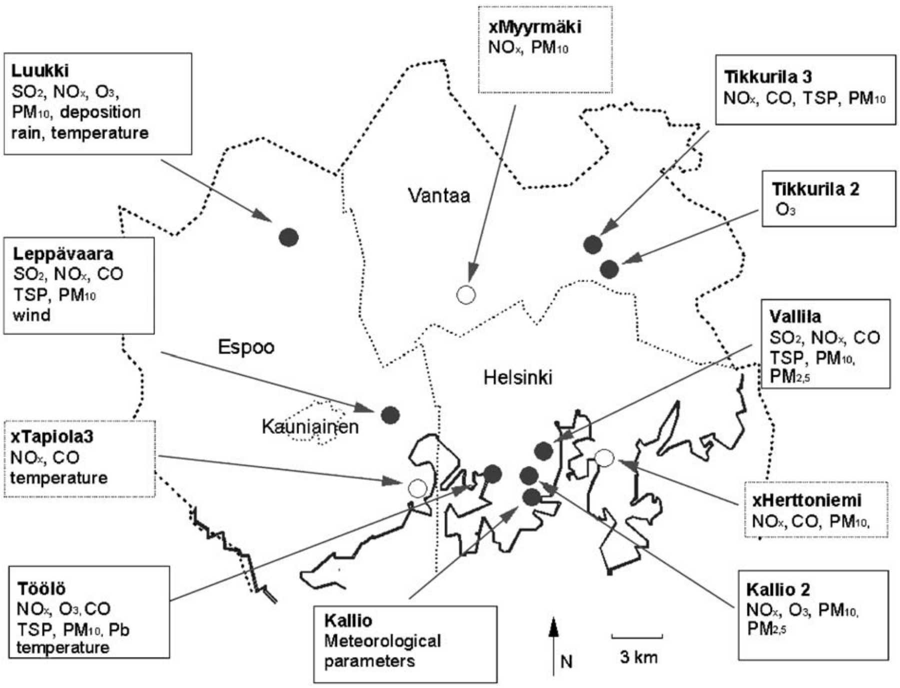

Fig. 1. Location of the air quality monitoring stations in the Helsinki Metropolitan Area in 1999. The legends show the name of the station and the pollutants that are measured continuously. The figure also shows the location of an urban meteorological station (Kallio). The open circles correspond to mobile stations; these are re-located each calender year. The other stations are permanently located. Reference: Helsinki Metropolitan Area Council.

the fraction of missing data is fairly small, using the original, so-called raw data, could result in the various NN models neglecting those segments of the data. We have therefore replaced the missing values in the datasets, using imputed values, in order to obtain a uniform dataset that is the same for all models to be evaluated. The purpose of using imputed data is to allow a consistent and fair model comparison exercise rather than to improve the absolute performance of the models.

The missing concentration data have been replaced using the hybrid method, i.e., a combination of linear interpolation and SOM; for a detailed description of this method, the reader is referred to Kolehmainen et al. (2000). Temporally short and longer gaps of missing data were interpolated by linear interpolation and SOM, respectively.

The computations with deterministic models have utilised only the raw concentration data, and those using NN models have utilised both the raw and imputed data.

2.2.2. Traffic flow data

The relevant traffic flow data was obtained using the EMME-2 traffic planning model (INRO, 1994), in the vicinity of the stations of Vallila and Töölö. This data contains hourly traffic volumes and average driving speed, separately for various vehicle categories (light-duty traffic, busses and other heavy-duty traffic). There have been no substantial changes in the traffic flows near the considered stations during the period 1996–1999. We have therefore applied only one hourly time series of traffic flow data for all the years considered.

2.2.3. Concentration data

The hourly concentrations of NO_2 and PM_{10} were available for the stations of Vallila and Töölö during the selected years. The concentrations of nitrogen oxides were measured with chemiluminescence monitors (Horiba APNA 360, with an accuracy of 1% and a precision of 1% of the full scale). The correct operation of the analysers was verified daily. Manual calibration was performed once a month using a NO cylinder (with an accuracy of 3%); the span drift of NO_x analysers between calibrations was lower than 3%.

At the station of Töölö, the concentration of PM_{10} was measured with a TEOM (Tapered Element Oscillating Microbalance). At the station of Vallila, the concentration of PM_{10} was measured with an Eberline FH 62 I-R which is based on the β-attenuation method. The flow rate of the PM analysers was calibrated twice a year and the mass measurement once a year. The daily average concentrations of PM_{10} and $PM_{2.5}$, determined by Eberline FH 62 I-R analysers, were compared with the corresponding results obtained by virtual impactors; the agreement was good (Viidanoja et al., 2002). Corresponding field inter-comparisons have also been conducted for TEOM monitors; these have also indicated a good agreement of measured PM_{10} values obtained with these methods (Viidanoja et al., 2002).

2.2.4. Meteorological data

We have selected the pre-processed meteorological data computed for the location of central Helsinki to be used in this study, as it is best representative for the whole of the urban area (compared with, e.g., utilising solely data measured at the airport), and contains also relevant derived meteorological parameters, such as, e.g., the Monin–Obukhov length and the mixing height. The pre-processed meteorological data is based on a combination of the data from the synoptic stations at Helsinki-Vantaa (about 15 km north of central Helsinki) and Helsinki-Isosaari (an island about 20 km south of central Helsinki). The mixing height of the atmospheric boundary layer was evaluated using the meteorological pre-processor, based on the sounding observations at Jokioinen (90 km northwest) and the routine meteorological observations. The pre-processed meteorological data is based on three-hourly synoptic measurements; it has been interpolated into an hourly time series. A single time series of meteorological data was assumed to be spatially representative for the whole of the study area.

3. Results

3.1. Compliance of measured concentrations with air quality guidelines and limit values

The national guidelines are defined on a monthly basis, as the 99th percentile of the hourly values for NO_2 and the second highest daily mean value, both for NO_2 and PM_{10}. The national guideline value for both pollutants is $70\,\mu g/m^3$. During the period 1996–2001 at the stations of Töölö and Vallila, there have been nine exceedences of the guideline threshold levels both for NO_2 and PM_{10} (during 72 months). For NO_2, most of these have occurred in winter, early spring or late summer, and for PM_{10}, in winter or spring. The measured concentrations of NO_2 and PM_{10} have not exceeded the limit values set by the EU during the period considered.

3.2. Details of model computations and model evaluation procedures

The NN model computations utilised both the raw and imputed data. The NN model computations were performed with either the pre-processed meteorological data during the 24 h period to be forecasted, or no meteorological data at all during this period.

NN models can be used for the forecasting of air quality in time, based on past concentration and other data, without utilising the predictions of numerical weather forecasting models. We have selected a forecasting period of 24 h for practical regulatory reasons; shorter time forecasts are of minimal value for air quality management purposes. This means that it was not allowed to use the hourly concentration data that has been measured at the forecast time or those during the past 23 h as input for NN models.

The variables used as inputs of the NN models of the University of East Anglia (UEA) for NO_2 are listed in Table 1. These contain temporal, concentration and pre-processed meteorological variables. The input data for PM_{10} are the same, except that the input concentration variable is PM_{10} (instead of NO_x, NO_2 and O_3). The inputs of the NN model of the University of Kuopio are the same, except for some minor differences; these models also utilised the CO concentrations (measured at both the stations of Töölö and Vallila) and those of SO_2 (measured only at the station of Vallila) for the prediction of NO_2.

We have especially tried to avoid over-fitting the training data. For the models developed and trained at University of East Anglia, we have used the Bayesian regularisation scheme, using a Laplace prior, propounded by Williams (1995) and discussed in Foxall et al. (2002).

For one of the MLP models, so-called boosting was applied. Boosting refers to increasing the frequency of the episodes in the training data in order to improve the ability of the NN to identify their characteristics. The fraction of the hourly concentration values, which were equal or higher than the 99th hourly percentile, was increased to 5% of the data. For this specific model, the network weights were initialised using SOMs. The acronym of MLP-UKU has been used for this model; UKU refers to the University of Kuopio, where this model was developed and applied. The other MLP models were developed and executed at the University of East Anglia.

We have cross-validated (Stone, 1974) the NN models sequentially on an annual basis. This means that the data corresponding to each of the above-mentioned years is used in turn for the testing of models, and the data from the other 3 years is used for the training of models. This procedure has the advantage of yielding four separate sets of statistical model performance parameters. Comparing the results obtained during several years provides confidence that the conclusions will not be dependent on the specific annual meteorological conditions.

The output of all models consists of hourly concentration time series of NO_2 and in part of the cases of PM_{10}, computed at the stations of Vallila and Töölö, from 1996 to 1999. In the case of the NN models, these include four sequential testing periods with a duration of 1 year.

Three statistical parameters were computed: the index of agreement (IA), the squared correlation coefficient

Table 1
The list of variables used as inputs of the neural network models of the University of East Anglia for NO_2

Input variable	Unit	Input time
Temporal variables		
Weekday	—	$T+24$
Sine of year day	—	$T+24$
Cosine of year day	—	$T+24$
Hour	—	$T+24$
Concentration variables		
NO_x	$\mu g/m^3$	T
NO_2	$\mu g/m^3$	T
O_3	$\mu g/m^3$	T
Meteorological variables		
Pressure	kPa	$T+24$
Temperature	K	$T+24$
Humidity	%	$T+24$
State of ground	—	$T+24$
Cloudiness	0–8	$T+24$
Dewpoint	K	$T+24$
Wetbulb	K	$T+24$
Rain	mm	$T+24$
Visibility	—	$T+24$
Weather	—	$T+24$
Weather of previous hour	—	$T+24$
Weather of previous 3 h	—	$T+24$
Amount of clouds	—	$T+24$
Type of clouds	—	$T+24$
Height of low clouds	m	$T+24$
Type of middle clouds	—	$T+24$
Type of high clouds	—	$T+24$
Sine of direction of flow	—	$T+24$
Cosine of direction of flow	—	$T+24$
Wind speed	m/s	$T+24$
Sunshine	h	$T+24$
Albedo	—	$T+24$
Solar elevation	rad	$T+24$
Solar radiation	W/m^2	$T+24$
Moisture parameters	—	$T+24$
Monin–Obukhov length	m	$T+24$
Temperature scale	K	$T+24$
Friction velocity	m/s	$T+24$
Turb. heat flux	W	$T+24$
Net radiation	W/m^2	$T+24$
Latent heat flux	W	$T+24$
Mixing height	m	$T+24$
Convective velocity scale	m/s	$T+24$
Gradient of the potential temperature	K/m	$T+24$

The input time $T+24$ is the time, for which the forecast applies. The input data regarding the concentrations corresponds to the time T.

(R^2) and the fractional bias (FB); these have been discussed, e.g., by Willmott (1981) and Petersen (1997), see also Karppinen et al. (2000b). The parameters IA and R^2 are measures of the correlation of two time series of values, while FB is a measure of the agreement of the mean values. The lower IA and R^2 are, the weaker is the agreement between two time series of values. FB is defined as the difference between the means of two series divided by the average of the means of the two series, i.e.,

$$\mathrm{FB} = \frac{\overline{C_\mathrm{P}} - \overline{C_\mathrm{O}}}{0.5(\overline{C_\mathrm{P}} + \overline{C_\mathrm{O}})}, \tag{1}$$

where C_P and C_O are the predicted and observed concentrations, respectively. The overbar refers to the average over all hourly values.

3.3. Results of the statistical evaluation of model performance

The final results of the statistical model evaluation have been presented in Tables 2–5. The results for NO_2 are presented in Tables 2 and 3, and those for PM_{10} in Tables 4 and 5. For both pollutants, the results have been presented first separately for each year; these can be used for evaluating the year-to-year variation of the results (Tables 2 and 4). For readability, the results have also been shown as average values over the period considered, and the standard deviations of the yearly averaged values (Tables 3 and 5).

The models could be implemented using the original concentration data series or the imputed data series, in which the missing values have been replaced; this choice has therefore been indicated in the tables for each model computation. As expected, examination of the tables demonstrates that using the raw and imputed data yields approximately the same model performance parameters, irrespective of pollutant and model.

All the models have been executed using pre-processed meteorological input data, except for in one specific case, using no meteorological data at all. The use of pre-processed meteorological data during the forecasting period improved substantially the performance of the NN models, compared with the predictions obtained using no meteorological data. Clearly, this result was physically expected. In the following, we consider only the results obtained using pre-processed meteorological data.

Considering the average results for the whole period (Tables 2 and 4), the results show improved performance for the NN models, compared with the LIN, irrespective of pollutant. This result is also to be expected, as the

Table 2
The statistical model evaluation parameters of the predicted and measured hourly time series of NO_2 concentrations at the stations of Töölö and Vallila, presented separately for the years from 1996 to 1999

Model and choice of input data	Year 1996			Year 1997			Year 1998			Year 1999		
	FB (%)	IA	R^2	FB (%)	IA	R^2	FB (%)	IA	R^2	FB (%)	IA	R^2
(a) *Töölö* NO_2												
NN-HeG raw	−5.6	0.89	0.66	3.0	0.90	0.67	−3.7	0.91	0.69	−5.8	0.91	0.72
NN-HeG imputed	−5.8	0.89	0.66	4.3	0.90	0.67	−3.4	0.90	0.66	−6.8	0.91	0.71
NN-2HeG raw	−6.6	0.89	0.66	1.7	0.91	0.69	−3.9	0.91	0.71	−6.6	0.91	0.71
NN-3HeG raw	−5.9	0.89	0.66	2.2	0.91	0.68	−3.2	0.92	0.72	−7.2	0.92	0.73
NN-HoG raw	−2.4	0.85	0.54	3.5	0.86	0.56	−3.3	0.87	0.59	−5.1	0.89	0.63
NN-HoG imputed	−5.4	0.87	0.59	4.6	0.85	0.52	−0.4	0.86	0.57	−2.7	0.87	0.59
NN-Lapl raw	−3.9	0.88	0.63	3.6	0.89	0.64	−1.8	0.90	0.67	−4.5	0.90	0.68
NN-Lapl imputed	−6.6	0.87	0.61	3.8	0.90	0.66	−1.4	0.90	0.66	−3.9	0.91	0.70
MLP-UKU imputed, no met	−6.4	0.75	0.43	5.3	0.75	0.36	−0.7	0.78	0.43	1.1	0.79	0.44
MLP-UKU imputed, pre-pr. met	−4.2	0.88	0.64	7.9	0.89	0.66	−1.8	0.91	0.70	−1.3	0.90	0.67
LIN raw	−4.4	0.78	0.41	−3.2	0.84	0.50	−9.1	0.84	0.53	−11	0.81	0.47
LIN imputed	−5.2	0.78	0.41	−3.1	0.83	0.47	−8.2	0.84	0.52	−10	0.82	0.48
DET raw	—	—	—	—	—	—	0.6	0.77	0.36	−4.6	0.75	0.32
(b) *Vallila* NO_2												
NN-HeG imputed	−9.1	0.87	0.61	2.8	0.86	0.56	−1.2	0.88	0.61	−8.3	0.87	0.59
MLP-UKU imputed, no met	−8.6	0.62	0.29	6.6	0.66	0.25	1.2	0.68	0.28	−0.1	0.68	0.28
MLP-UKU imputed, pre-pr. met	−5.2	0.85	0.57	8.6	0.85	0.56	−0.6	0.88	0.63	−0.9	0.86	0.57
LIN imputed	−10	0.83	0.50	−1.7	0.80	0.43	−6.1	0.81	0.45	2.4	0.04	0.01
DET raw	—	—	—	—	—	—	21	0.70	0.27	19	0.68	0.24

The pre-processed meteorological data was applied for all models, except for in one case, "MLP-UKU No met". The terms "raw" and "imputed" refer to the original measured data, and the data, in which the missing values have been replaced computationally, respectively. FB—fractional bias, IA—index of agreement, R^2—correlation coefficient squared.

Table 3
The statistical model evaluation parameters of the predicted and measured hourly time series of NO_2 concentrations at the stations of Töölö and Vallila, presented as average values and their standard deviations for the period 1996–1999

Model and choice of input data	Töölö NO_2						Vallila NO_2					
	Averages, 1996–1999			Standard deviation			Averages, 1996–1999			Standard deviation		
	FB (%)	IA	R^2	FB	IA	R^2	FB (%)	IA	R^2	FB (%)	IA	R^2
NN-HeG raw	−3.0	0.90	0.68	4.1	0.01	0.03	—	—	—	—	—	—
NN-HeG imputed	−2.8	0.90	0.68	4.9	0.01	0.03	−4.0	0.87	0.59	5.7	0.01	0.02
NN-2HeG raw	−3.9	0.91	0.69	3.9	0.01	0.02	—	—	—	—	—	—
NN-3HeG raw	−3.5	0.91	0.70	4.2	0.01	0.03	—	—	—	—	—	—
NN-HoG raw	−1.8	0.87	0.58	3.7	0.02	0.04	—	—	—	—	—	—
NN-HoG imputed	−1.0	0.86	0.57	4.2	0.01	0.03	—		—	—	—	—
NN-Lapl raw	−1.7	0.89	0.65	3.7	0.01	0.02	—	—	—	—	—	—
NN-Lapl imputed	−2.0	0.90	0.66	4.4	0.02	0.04	—	—	—	—	—	—
MLP-UKU imputed, no met.	−0.2	0.77	0.41	4.9	0.02	0.04	−0.2	0.66	0.27	6.3	0.03	0.02
MLP-UKU imputed, pre-pr. met.	0.1	0.89	0.68	5.3	0.02	0.03	0.5	0.86	0.58	5.8	0.02	0.03
LIN raw	−6.9	0.82	0.48	3.7	0.03	0.05	—	—	—	—	—	—
LIN imputed	−6.6	0.82	0.47	3.1	0.03	0.05	−1.0	0.80	0.43	8.4	0.05	0.09
DET raw	−2.0	0.76	0.34	3.7	0.01	0.03	20.0	0.69	0.25	1.4	0.01	0.02

The notation is the same as that in Table 2.

Table 4
The statistical model evaluation parameters of the predicted and measured hourly time series of PM_{10} concentrations at the stations of Töölö and Vallila, presented separately for the years 1996–1999

Model and choice of input data	Year 1996			Year 1997			Year 1998			Year 1999		
	FB (%)	IA	R^2	FB (%)	IA	R^2	FB (%)	IA	R^2	FB (%)	IA	R^2
(a) *Töölö* PM_{10}												
NN-HeG raw	−16	0.71	0.36	−1.5	0.50	0.14	−11	0.79	0.45	−2.2	0.80	0.43
NN-HeG imputed	−16	0.73	0.39	−2.6	0.80	0.44	−12	0.79	0.46	−2.2	0.76	0.38
NN-HoG imputed	−18	0.64	0.26	−1.7	0.76	0.36	−8.7	0.77	0.39	−1.0	0.73	0.31
MLP-UKU imputed, no met.	−7.9	0.64	0.27	2.0	0.66	0.26	−5.8	0.64	0.28	11	0.72	0.30
MLP-UKU imputed, pre-pr.met.	−6.5	0.76	0.40	7.1	0.80	0.45	−6.6	0.75	0.43	5.7	0.76	0.34
LIN raw	−14	0.60	0.22	−14	0.71	0.40	−13	0.37	0.06	−11	0.75	0.38
LIN imputed	−15	0.65	0.25	−14	0.72	0.42	−15	0.71	0.35	−9.3	0.26	0.03
(b) *Vallila* PM_{10}												
NN-HeG imputed	−18	0.70	0.34	−11	0.71	0.32	6.5	0.75	0.33	−2.8	0.77	0.26
LIN imputed	−24	0.64	0.28	−13	0.66	0.27	−0.4	0.47	0.09	−19	0.75	0.20

The notation is the same as that in Table 2.

dependencies of urban NO_2 and PM_{10} concentrations on the corresponding vehicular emissions and relevant meteorological parameters are strongly non-linear. This result is in agreement with previous model inter-comparisons by Gardner and Dorling (1998). The heteroscedastic NN models perform better than both those with constant variance.

3.3.1. Model evaluation for NO_2

The standard deviations of the annually averaged model performance parameters can be considered to be moderate; these are usually smaller than the relative differences between various models (Table 3). The performance of the models is therefore not especially sensitive to the particular year which is selected for the model evaluation.

The results obtained with various non-linear NN models show a remarkably good agreement with the measured concentration data for NO_2 at the two stations considered. For instance, the corresponding annually averaged IA values and their standard deviations over the period 1996–1999 range from 0.86 ± 0.02 to 0.91 ± 0.01 for the NN computations. The statistical model performance parameters of NO_2 for the non-linear NN models are slightly better at the stations of Töölö and Vallila, respectively, compared with the corresponding values obtained for the DET.

3.3.2. Model evaluation for PM_{10}

The model evaluation regarding PM_{10} was less extensive than that for NO_2. At the station of Töölö, the model evaluation included three NN models and one linear statistical model. At the station of Vallila, only one NN model was tested. In addition, a linear statistical model was used at both stations; however, no deterministic models were available for predicting reliably the hourly time series of the PM_{10} concentrations.

Table 5
The statistical model evaluation parameters of the predicted and measured hourly time series of PM_{10} concentrations at the stations of Töölö and Vallila, presented as average values and their standard deviations for the period 1996–1999

Model	Töölö PM_{10}						Vallila PM_{10}					
	Averages, 1996–1999			Standard deviation			Averages, 1996–1999			Standard deviation		
	FB (%)	IA	R^2	FB (%)	IA	R^2	FB (%)	IA	R^2	FB (%)	IA	R^2
NN-HeG raw	−7.7	0.70	0.33	7.0	0.14	0.14	—	—	—	—	—	—
NN-HeG imputed	−8.2	0.77	0.42	6.9	0.03	0.04	−6.3	0.73	0.31	9.6	0.03	0.04
NN-HoG imputed	−7.4	0.73	0.33	7.9	0.06	0.06	—	—	—	—	—	—
MLP-UKU no met.	−0.1	0.67	0.28	8.7	0.04	0.02	—	—	—	—	—	—
MLP-UKU pre-pr. met.	−0.1	0.77	0.40	7.5	0.02	0.05	—	—	—	—	—	—
LIN raw	−13.0	0.61	0.24	1.4	0.17	0.16	—	—	—	—	—	—
LIN imputed	−13.3	0.59	0.23	2.7	0.22	0.17	−14.1	0.63	0.20	10.8	0.12	0.09

The notation is the same as that in Table 2.

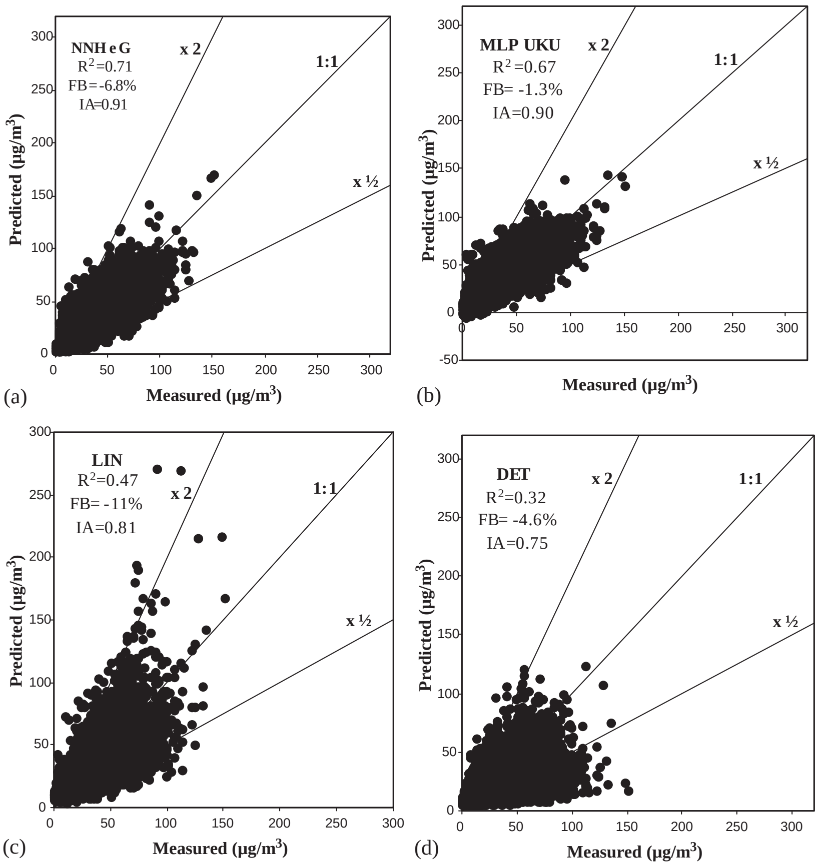

Fig. 2. (a–d) Scatter plots of the measured and predicted concentrations of NO_2 at the station of Töölö in 1999 for two NN models (NN-HeG and MLP-UKU), the LIN and the DET.

In the case of PM_{10}, both the corresponding IA and R^2 values are clearly lower for the whole range of models considered, compared with the corresponding values for NO_2. This is caused by the fact that it is more difficult to predict the particulate matter concentrations in urban areas, compared with those of NO_2. Several source categories influence the PM concentrations, such as combustion and non-combustion traffic sources, suspended particulate matter, urban background concentrations and the regionally and long-range transported contribution.

3.4. Scatter plots and time series of measured and predicted concentrations

The scatter plots of the measured and predicted concentrations for four selected models are presented in Figs. 2a–d. We have presented the results for one of the best models developed by the University of East Anglia (NN-HeG), the model developed by the University of Kuopio (MLP-UKU), the LIN and the DET. The NN and LINs utilise imputed concentration and pre-processed meteorological data. The values of the squared correlation coefficient (R^2), FB and IA are also shown in the figures.

The scatter plots are, as expected, fairly symmetrical for the two NN models presented; however, there is some underprediction for the LIN and DET models (the average bias of predicted and measured values is shown by the values of FB). The scatter plots also give some insight on the model performance in case of the highest concentration values. Part of the highest measured values are substantially overpredicted by the LIN model, and underpredicted by the DET modelling system.

In order to illustrate model performance in case of the highest concentrations, we have selected one air quality episode in March 1999. During this month, the NO_2 concentrations exceeded the national air quality guideline values at the station of Töölö. The predicted concentrations of NO_2 are presented in Figs. 3a–b for the above-mentioned four models, together with the measured data. For this specific episode, both NN models predict the highest concentration values fairly well. The LIN also predicts the highest concentrations

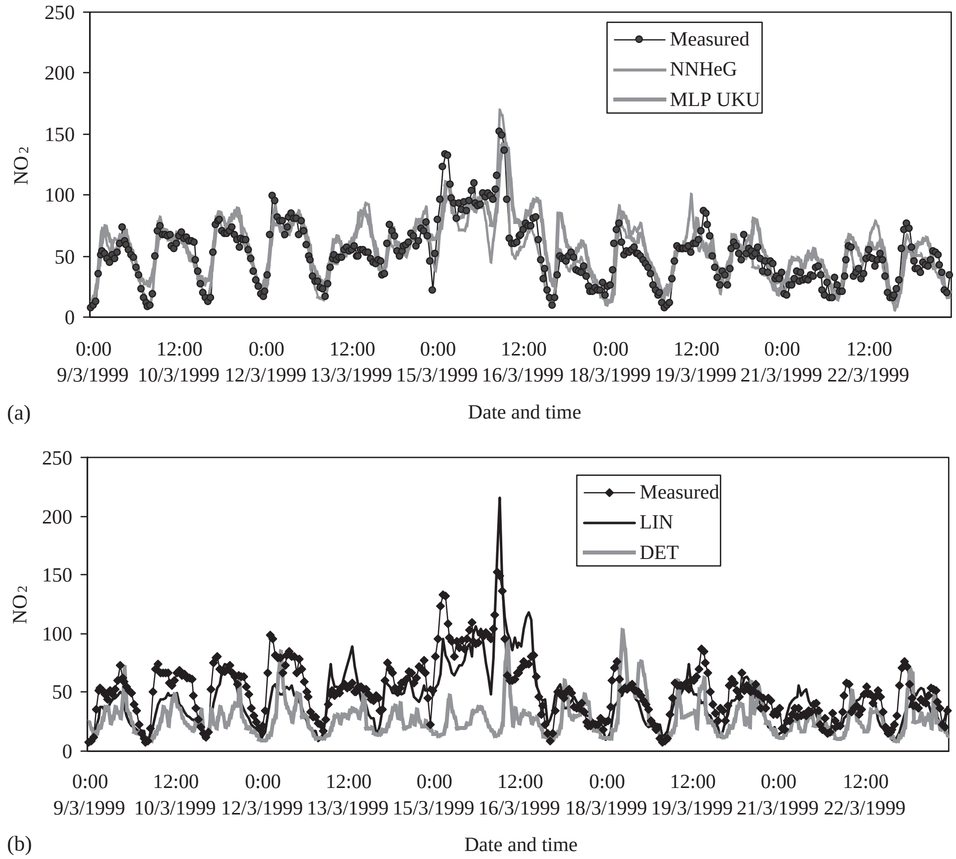

Fig. 3. (a–b) hourly time series of the measured and predicted concentrations of NO_2 at the station of Töölö from 9 to 23 March 1999 for two NN models (NN-HeG and MLP-UKU), the LIN and the DET.

reasonably well, but the DET modelling system tends to underpredict the measured data, especially for the period of the highest measured concentrations from 15 to 17 March 1999.

4. Conclusions

This investigation presents the most extensive evaluation of NN models currently available for the prediction of urban NO_2 and PM_{10} concentrations, both regarding the variety of models to be evaluated, and the amount of experimental data included. Besides non-linear NN models, the evaluation also included a statistical LIN and a DET. Previous studies have not analysed rigorously the relative performance of NN models with regard to DETs. The aim was to produce information on the suitability of various models, or various classes of models, for specific applications.

The deterministic modelling also includes the modelling of traffic flows and emissions within the whole urban area. The comparison of statistical and deterministic models (in the case of NO_2) therefore does not solely concern atmospheric dispersion and meteorological modelling. In particular, major uncertainties in the modelling of traffic-originated emissions (or those in the regional and urban background ozone concentrations) could cause a bias in comparing the results for NO_2 obtained using these two model categories. However, according to the results obtained in previous field measurement campaigns involving NO_x, NO_2 and O_3 (e.g., Kukkonen et al., 2001), the applied emission modelling does not substantially over or underestimate the traffic-originated NO_x emissions.

The study involves the following inherent limitations. We focus on modelling concentrations at two specific spatial locations in central Helsinki; besides, only one DET was used. Helsinki is a Northern European city situated in a fairly flat terrain; clearly, the modelling problem is more complex in cities located in hilly or mountainous terrain. Regarding NN models, the input variables and possibly even the architecture of the models required may be slightly different in various European geographic regions.

Climatic factors may also affect substantially model performance, e.g., the influence of photochemical reactions and meso-scale circulation phenomena is commonly more pronounced in Southern European cities (e.g., Piringer and Kukkonen, 2002). On the other hand, in Northern Europe the frequency of stable atmospheric conditions with low wind speeds tends to be higher than that in lower latitudes; these are commonly the most difficult conditions to model reliably (e.g., Karppinen et al., 2001; Kukkonen, 2001).

All the models to be evaluated had the same available input data. The results obtained with various non-linear NN models show a good agreement with the measured concentration data for NO_2. The non-linear NN models performed slightly better in terms of the model performance values than the DET. The results also show improved performance for most of the NN models, compared with the linear statistical model, both for predicting NO_2 and PM_{10} concentrations. This result is to be expected, as the dependencies of urban NO_2 and PM_{10} concentrations on the corresponding vehicular emissions and relevant meteorological parameters are strongly non-linear. Apparently for the same reason, the heteroscedastic NN models perform better than both those with constant variance cost functions.

There was no DET available that could reliably predict the hourly time series of the PM_{10} concentrations. In the case of PM_{10}, both the corresponding IA and R^2 values were clearly lower compared with those for NO_2, for the whole range of models considered. This was expected, as it is substantially more difficult also for deterministic models to predict the particulate matter concentrations in urban areas, compared with those of NO_2.

A general conclusion from this study is that NN models can be useful and fairly accurate tools of assessment in predicting NO_2 and PM_{10} concentrations in urban areas. After such a model has been trained using appropriate site- and time-specific data, its utilisation requires less effort than performing deterministic model computations. However, NN models also have inherent limitations. The main limitation is the extension of models in terms of time period and location; this always requires training with locally measured data. The NN models cannot therefore be recommended for analysing various air pollution abatement scenarios for future years. The currently available NN models are also not applicable for predicting spatial concentration distributions in urban areas.

In future work, the NN and deterministic models will be evaluated using forecasted meteorological data, i.e., for the purpose of air quality forecasting. In that case, the relevant archived forecasts of numerical weather prediction models will be used as input values to the models.

Acknowledgements

We wish to acknowledge the financial support of the European Commission for the APPETISE and FUMAPEX projects, and that of the Academy of Finland for the FORECAST project. We also wish to thank the Helsinki Metropolitan Area Council (YTV) for help in the utilisation of the air quality monitoring data, and

W. Burrows, U. Schlink and G. Nunnari for their valuable comments.

References

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

Foxall, R.J., Cawley, G.C., Dorling, S.R., Mandic, D.P., 2002. Error functions for prediction of episodes of poor air quality. Proceedings of the International Conference on Artificial Neural Networks (ICANN-2002), Springer Lecture Notes on Computer Science, Vol. 2415, Madrid, Spain, August 2002, pp. 1031–1036.

Gardner, M.W., 1999. The advantages of artificial neural network and regression tree based air quality models. Ph.D. Thesis, University of East Anglia, UK.

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.

Gardner, M.W., Dorling, S.R., 1999. Neural network modelling and prediction of hourly NO_x and NO_2 concentrations in urban air in London. Atmospheric Environment 33 (5), 709–719.

Greig, A.J., Cawley, G., Dorling, S., Eben, K., Fiala, A.J., Karppinen, A., Keder, J., Kolehmainen, M., Kukkonen, J., Libero, B., Macoun, J., Nironjan, M., Nucifora, A., Nunnari, A., Palus, M., Pelikan, E., Ruuskanen, J., Schlink, U., 2000. Air pollution episodes: modelling tools for improved smog management (APPETISE). In: Longhurst, J.W.S., Brebbia, C.A., Power, H. (Eds.), Air Pollution VIII. Wessex Institute of Technology Press, Bath, UK, pp. 89–98.

Haykin, S., 1999. Neural Networks, A Comprehensive Foundation, 2nd Edition. Prentice-Hall, Englewood Cliffs, NJ, p. 842.

Hecht-Nielsen, R., 1991. Neurocomputing. Addison-Wesley, Reading, MA, USA.

INRO, 1994. EMME/2 User's Manual. INRO Consultants Inc., Montréal, Canada.

Karppinen, A., Kukkonen, J., Elolähde, T., Konttinen, M., Koskentalo, T., Rantakrans, E., 2000a. A modelling system for predicting urban air pollution, Model description and applications in the Helsinki metropolitan area. Atmospheric Environment 34, 3723–3733.

Karppinen, A., Kukkonen, J., Elolähde, T., Konttinen, M., Koskentalo, T., 2000b. A modelling system for predicting urban air pollution, comparison of model predictions with the data of an urban measurement network. Atmospheric Environment 34, 3735–3743.

Karppinen, A., Joffre, S.M., Kukkonen, J., 2000c. The refinement of a meteorological preprocessor for the urban environment. International Journal of Environment and Pollution 14 (1–6), 565–572.

Karppinen, A., Joffre, S.M., Kukkonen, J., Bremer, P., 2001. Evaluation of inversion strengths and mixing heights during extremely stable atmospheric stratification. International Journal of Environment and Pollution 16, 1–6.

Kolehmainen, M., Martikainen, H., Hiltunen, T., Ruuskanen, J., 2000. Forecasting air quality using hybrid neural network modelling. Environmental Monitoring and Assessment 65, 277–286.

Kolehmainen, M., Martikainen, H., Ruuskanen, J., 2001. Neural networks and periodic components used in air quality forecasting. Atmospheric Environment 35, 815–825.

Kousa, A., Kukkonen, J., Karppinen, A., Aarnio, P., Koskentalo, T., 2001. Statistical and diagnostic evaluation of a new-generation urban dispersion modelling system against an extensive dataset in the Helsinki area. Atmospheric Environment 35, 4617–4628.

Kukkonen, J. (Ed.), 2001. COST action 715, Meteorology applied to urban air pollution problems. Working Group 3, Status Report, Directorate-General for Research, Information and Communication Unit, European Commission. Brussels, 73pp.

Kukkonen, J., Härkönen, J., Walden, J., Karppinen, A., Lusa, K., 2001. Evaluation of the dispersion model CAR-FMI against data from a measurement campaign near a major road. Atmospheric Environment 35, 949–960.

Laurikko, J., 1998. On exhaust emissions from petrol-fuelled passenger cars at low ambient temperatures. VTT publications 348, Ph.D. Thesis, Technical Research Centre of Finland, Espoo, Finland.

Mäkelä, K., Kanner, H., Laurikko, J., 1996. Road traffic exhaust gas emissions in Finland—LIISA 95 calculation software (in Finnish). VTT Communities and Infrastructure, Transport Research, Research Notes 1772, Technical Research Center of Finland, Espoo, 45pp. + App, 51pp.

Pérez, P., Trier, A., Reyes, J., 2000. Prediction of $PM_{2.5}$ concentrations several hours in advance using neural networks in Santiago, Chile. Atmospheric Environment 34, 1189–1196.

Petersen, R.L., 1997. A wind tunnel evaluation of methods for estimating surface rougness length at industrial facilities. Atmospheric Environment 31, 45–57.

Piringer, M., Kukkonen, J. (Eds.), 2002. Mixing height and inversions in urban areas. Proceedings of the Workshop, 3 and 4 October 2001, Toulouse, France, COST Action 715, EUR 20451, European Commission, Brussels, 113pp.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M., Doyle, M., 2003. A rigorous intercomparison of ground-level ozone predictions. Atmospheric Environment 37, 3237–3253.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society Series B 36, 111–147.

Tiitta, P., Raunemaa, T., Tissari, J., Yli-Tuomi, T., Leskinen, A., Kukkonen, J., Härkönen, J., Karppinen, A., 2002. Measurements and modelling of $PM_{2.5}$ concentrations near a major road in Kuopio, Finland. Atmospheric Environment 36, 4057–4068.

Viidanoja, J., Sillanpää, M., Laakia, J., Kerminen, V.-M., Hillamo, R., Aarnio, P., Koskentalo, T., 2002. Organic and black carbon in $PM_{2.5}$ and PM_{10}: 1 year data from an urban site in Helsinki, Finland. Atmospheric Environment 36, 3183–3193.

Williams, P.M., 1995. Bayesian regularisation and pruning using a Laplace prior. Neural Computation 7 (1), 117–142.

Willmott, C.J., 1981. On the validation of models. Physical Geography 2, 184–194.