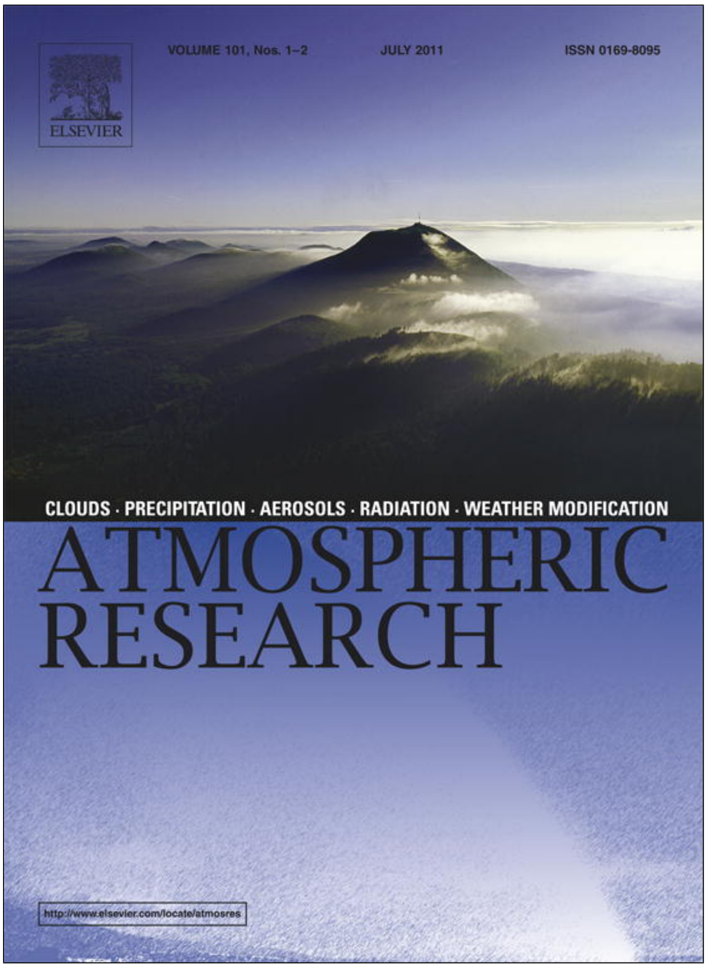
VOLUME 101, Nos. 1–2
JULY 2011
ISSN 0169-8095
ELSEVIER
CLOUDS · PRECIPITATION · AEROSOLS · RADIATION · WEATHER MODIFICATION
ATMOSPHERIC RESEARCH
http://www.elsevier.com/locate/atmosres

Atmospheric Research 101 (2011) 314–326

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmos

Application of cellular neural network (CNN) to the prediction of missing air pollutant data

Ülkü Alver Şahin [a,*], Cuma Bayat [b], Osman N. Uçan [c]

[a] Istanbul University, Engineering Faculty, Environmental Eng. Dept. 34 320, Avcilar, Istanbul, Turkey
[b] Arel University, Letters and Science Faculty, Tepekent-Büyükçekmece, Istanbul,Turkey
[c] Aydin University, Engineering and Architecture Faculty, Electrical-Electronics Eng. Dept. Florya, Istanbul, Turkey

A R T I C L E I N F O

Article history:
Received 29 August 2010
Received in revised form 20 March 2011
Accepted 21 March 2011

Keywords:
Missing data
Air quality
Particulate matter (PM)
Sulfur dioxide (SO_2)
Meteorology
Cellular Neural Network (CNN)

A B S T R A C T

For air-quality assessments in most major urban centers, air pollutants are monitored using continuous samplers. Sometimes data are not collected due to equipment failure or during equipment calibration. In this paper, we predict daily air pollutant concentrations (PM_{10} and SO_2) from the Yenibosna and Umraniye air pollution measurement stations in Istanbul for times at which pollution data was not recorded. We predicted these pollutant concentrations using the CNN model with meteorological parameters, estimating missing daily pollutant concentrations for two data sets from 2002 to 2003. These data sets had 50 and 20% of data missing. The results of the CNN model predictions are compared with the results of a multivariate linear regression (LR). Results show that the correlation between predicted and observed data was higher for all pollutants using the CNN model (0.54–0.87). The CNN model predicted SO_2 concentrations better than PM_{10} concentrations. Another interesting result is that winter concentrations of all pollutants were predicted better than summer concentrations. Experiments showed that accurate predictions of missing air pollutant concentrations are possible using the new approach contained in the CNN model. We therefore proposed a new approach to model air-pollution monitoring problem using CNN.

1. Introduction

The main sources of air pollution in Istanbul are the combustion of poor quality coal, increased traffic load and industrial activities. In the last two decades, many scientists have focused on the air pollution problems of Istanbul-Turkey (Erturk, 1986; Tayanç, 2000; Saral and Ertürk 2003; Sahin, 2005; Im et al., 2008, Hanedar et al., 2011). During the winter, sulfur dioxide (SO_2) and particulate matter (PM) are the major air pollutants affecting regional air quality. Missing data, which may be due to insufficient sampling and errors in measurements or problems with data acquisition, presents a problem that is frequently encountered in environmental research. Regardless of the reasons for missing data, discontinuities in data pose a significant obstacle to time-series prediction schemes, which generally require continuous data as a condition for their implementation.

The substitution of mean values for missing data is commonly suggested, and is still used in many statistical software packages (Junninen et al., 2004). A slightly better approach is to impute the missing elements from an ANOVA model or similar statistical method. Another approach to the problem is to use a simplistic interpolation method, such as assuming the season's average concentration at the time of day for which data are missing, or to linearly interpolate between values of the previous and following to obtain continuous data sets. Neither of these methods is ideal, because the meteorology on the missing day may have been significantly different from the days on which the interpolation is based, leading to unrealistic predictions (Dirks et al., 2002). Clearly, a complementary method is required.

There are many deterministic and stochastic approaches to modeling the concentrations of air pollutants. The well-

* Corresponding author.
E-mail address: ulkualver@istanbul.edu.tr (Ü.A. Şahin).

0169-8095/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.atmosres.2011.03.005

known machine-learning approach is Artificial Neural Networks (ANN). That is concerned with the design and development of algorithms that allow computers to empirically learn the behavior of data sets. Machine learning approaches have been used and applied to the correction of bias for various environmental problems and weather prediction since 1990. Neural networks are suitable for the application of these areas due to their ability to model non-linear mechanism. A recent paper by Manzato (2007) and Fernandez-Ferrero et al. (2009) studied different statistical downscaling methods applied to different numerical weather forecasting. These paper results have shown the ANNs proved to be a powerful statistical method, but special care must be used to prevent over fitting.

In many studies, ANNs are applied to predict SO_2 and PM_{10} concentrations (Boznar et al., 1993; Mok and Tam, 1998; Saral and Ertürk, 2003; Chelani et al., 2002; Onat et al., 2004; Sahin et al., 2005, Yildirim and Bayramoğlu, 2006). Gardner and Dorling (1998) have published a comprehensive review of studies using an ANN approach for environmental air pollution modeling. Kukkonen et al. (2003) have studied five neural network (NN) models, a linear statistical model and a deterministic modeling system for the prediction of urban NO_2 and PM_{10} concentrations. Sahin et al. (2004) used a multi-layer neural network model to predict daily CO concentrations, using meteorological variables, in the European side of Istanbul, Turkey. Kurt et al. (2008) also developed an online air pollution forecasting system in Istanbul using NN. Another NN model developed by Saral and Ertürk (2003) was also used to predict regional SO_2 concentrations. Junninen et al. (2004) applied regression-based imputation, nearest neighbor interpolation, a self organizing map, a multi-layer perceptron model and hybrid methods to simulate missing air quality data. Nagendra and Khare (2006) studied the usefulness of NNs in understanding the relationship between traffic parameters and NO_2 concentrations. Recently, several researchers used NN techniques to predict airborne PM concentrations: e.g. Ordieres et al. (2005) Hooyberghs et al. (2005), Perez and Reyes (2006) and Slini et al. (2006). These days, some scientist use machine learning approaches to modeling the satellite data (Lary et al., 2009; Gupta and Christopher, 2009). All of these studies reported that ANN could be used to develop efficient air-quality analysis and forward-looking prediction models. But in ANNs, the training process becomes increasingly complex and requires longer time durations as the number of weighting coefficients of the ANN rise into the millions due to the complexity of the environmental study.

To reduce the number of weighting coefficients, Chua and Yang (1988) introduced another machine learning approach, Cellular Neural Network (CNN) in 1988. Because each cell of the CNN is represented by a separate analog processor, and because each cell is locally interconnected to its neighbors by matrix A and gets a feedback from them by matrix B, this configuration results in a very high-speed tool for parallel dynamic processing of 2-D structures (Cimagalli, 1993; Guzelis and Karamahmut, 1994; Ucan et al., 2001; Grassi and Grieco, 2002). CNN approaches have been applied to air pollution modeling by a number of researchers, with excellent results (Sahin, 2005; Ozcan et al., 2007; Thai and Cat, 2008).

In this study, we have applied a CNN approach to the problem of predicting the daily mean missing concentrations of PM_{10} and SO_2 pollutants in the Yenibosna and Umraniye-Istanbul regions of Turkey. This paper is organized as follows: In Sections 2.1 and 2.2 the Cellular Neural Network (CNN) and Multiple Linear Regression (LR) modeling techniques are defined. In order to evaluate model prediction, statistical performance indices are explained in Section 2.3. The study area and database are explained in Section 2.4. Model construction is described in Section 2.5. In Section 3.1, PM_{10} and SO_2 pollution in Istanbul is explained and in Section 3.2, the CNN is tested on real data and the results are presented and compared to LR technique. In Section 4, the results of the study are evaluated.

2. Materials and methods

2.1. Architecture of CNN

Most neural networks fall into two main classes: (1) memoryless neural networks and (2) dynamical neural networks. As in Hopfield Networks and CNNs, dynamical neural networks are usually designed as dynamic systems in which the inputs are set to constant values and the path approach to a stable equilibrium point depends upon the initial state. A CNN is composed of large-scale nonlinear analog circuits which process signals in real time (Chua and Yang, 1988). The basic unit of a CNN is called a cell, and these units communicate with each other directly only through their nearest neighbors. Adjacent cells can therefore interact directly with each other. Cells not directly connected together affect each other indirectly because of the propagation effects of the continuous real-time dynamics of the CNN. The structure of a two-dimensional (2-D) 3×3 CNN is shown in Fig. 1.

The Cellular Neural Network used in this study consisted of M rows and N columns ($M\times N$). In this structure, the ith line and jth column are designated cell (i,j) and denoted by $C(i,j)$. A typical example of a cell is shown in Fig. 2. In Fig. 2, u_{ij}, y_{ij} and x_{ij} correspond to the input, the output and the state variable of the cell, respectively. The node voltage v_{xij} of $C(i,j)$ is defined as the state of the cell whose initial condition is assumed to have a magnitude less than of equal to 1. Each cell contains one independent current source, one linear-

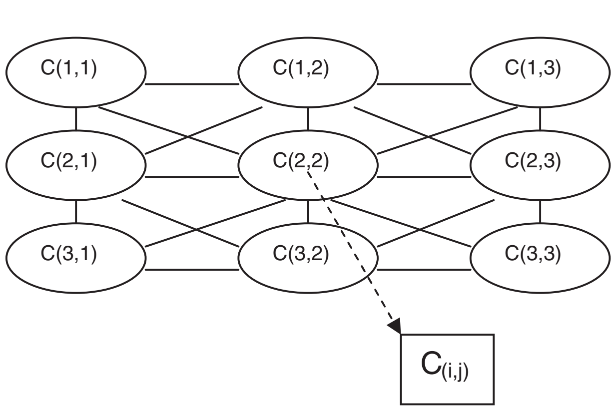

Fig. 1. A 2-D cellular neural network of size 3×3. Links between the cells (ellipse) indicate interactions between the linked cells.

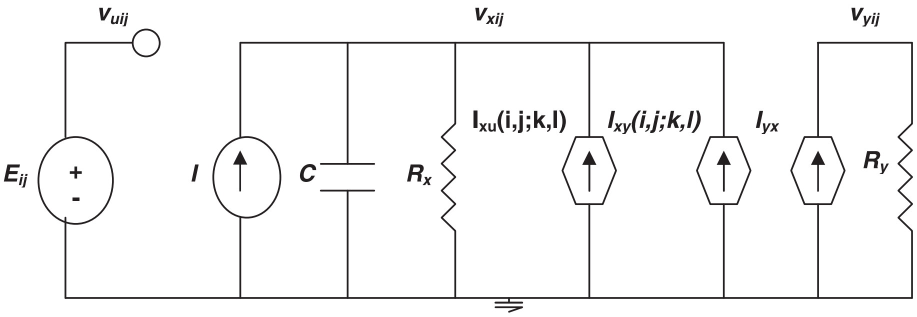

Fig. 2. A classical CNN cell scheme.

capacitor C, two linear resistors Rx and Ry and linear voltage controlled current sources ($I_{xy}(i,j{:}k,l)$), which are coupled to its neighbor cells via the controlling input voltage and the feedback from the output voltage of each neighboring cell $C(k,l)$. The constant coefficients $A(i,j;k,l)$ and $B(i,j;k,l)$ are known as the cloning templates, and these are the parameters linking cell $C(i,j)$ to its neighbor $C(k,l)$. The equivalent block diagram of a CNN cell is shown in Fig. 3. The first-order nonlinear equation defining the dynamic of a CNN can be derived as follows (Arena et al., 1997; Hadad and Piroozmand, 2007; Thai and Cat, 2008):

The r-neighborhood of a cell $C(i,j)$ in a cellular neural network is defined by:

$$N_r(i,j) = \{C(k,l)/max(/k-i/,\, l-j/ \le r, \quad 1 \le i \le M\,;\, 1 \le j \le N\}. \tag{1}$$

A general form of the cell dynamical equations may be written as follows:

$$C\frac{dv_{xij}(t)}{dt} = -\frac{1}{R}v_{xij}(t) + \sum_{C(k,l)\in N_r(i,j)} A(i,j;k,l)v_{ykl}(t) + \sum_{C(k,l)\in N_r(i,j)} B(i,j;k,l)v_{ukl} + I. \tag{2}$$

In the CNN system, (A,B,I) are the local connective weighting values of each cell $C(i,j)$ to its neighbors. Each cell of the CNN is represented by a separate analog processor, and each cell is locally interconnected to its neighbors by matrix A and gets a feedback from them by matrix B. This configuration results in a very high-speed tool for parallel dynamic processing of 2-D structures

$$A = \begin{bmatrix} a_{-1,-1} & a_{-1,0} & a_{-1,1} \\ a_{0,-1} & a_{0,0} & a_{0,1} \\ a_{1,-1} & a_{1,0} & a_{1,1} \end{bmatrix},\ B = \begin{bmatrix} b_{-1,-1} & b_{-1,0} & b_{-1,1} \\ b_{0,-1} & b_{0,0} & b_{0,1} \\ b_{1,-1} & b_{1,0} & b_{1,1} \end{bmatrix}, I. \tag{3}$$

The output is related to the state by the nonlinear equation. Characteristic of the output function $v_{yi,j} = f(v_{xi,j})$ is as follows:

$$v_{yij}(t) = \frac{1}{2}\left(\left|v_{xij}(t)+1\right| - \left|v_{xij}(t)-1\right|\right) \qquad v_{yij} = \begin{cases} -1 & when \quad v_{xij} < -1 \\ v_{xij} & when -1 < v_{xij} < 1 \\ 1 & when \quad v_{xij} > 1 \end{cases}. \tag{4}$$

The network behavior of a CNN depends on the initial state of the cells, namely the bias I, and the weighting values of the A and B matrices, which are associated with the connections inside the well-defined neighborhood of each cell. CNNs are arrays of locally and regularly interconnected neurons or cells whose global functionalities are defined by a small number of parameters (A, B and I) that specify the operation of the component cells as well as the connection weights between them. The CNN can also be considered as a nonlinear convolution with the template. Since their introduction in 1988 by Chua, CNNs have attracted a lot of attention. Not only do these systems have a number of attractive properties from a theoretical point of view, but they also have many well-known applications such as image processing, motion detection, pattern recognition and simulation. Albora et al. (2001) applied this contemporary approach to the separation of regional and residual magnetic anomalies on synthetic and real data. Hadad and Piroozmand (2007) applied the CNN to modeling and solving the nuclear reactor dynamic equations. Here, we have predicted air pollution parameters using a CNN approach. To evaluate the prediction results of the CNN, statistical performance indices have been calculated as described in Section 2.3.

2.2. Multiple linear regression model

Linear regression (LR) models have been used as a reference for comparison with the neural network models in several

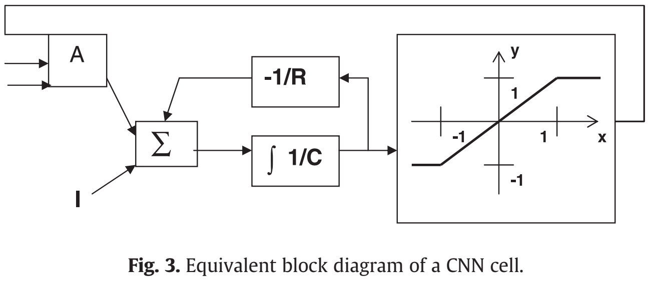

Fig. 3. Equivalent block diagram of a CNN cell.

studies (Nunnari et al., 2004; Grivas and Chaloulakou, 2006, Agirre-Basurko et al., 2006). This model is one of the most cost-effective approaches for time series analysis, and many authors have been inspired to apply this technique, after appropriate modifications, in developing pollutant forecasting models.

The general form of a multiple linear regression is:

$$Y_i = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots\dots\dots + \beta_p X_{ip} + \varepsilon_i \quad (5)$$

where, for a set of *i* observations, Y_i is the predictand variable, β_0 is a coefficient, $\beta_1, \beta_2, \dots, \beta_p$ are the coefficients of the independent variables (predictors) $X_{i1}, \dots, X_{ip}$ and ε_i is the residual error.

The hypotheses required to apply multiple linear regressions are: (i) the predictor variables must be independent, and (ii) the residual errors ε_i must be independent and they must be normally distributed, with mean *0* and constant variance σ^2 (Agirre-Basurko et al., 2006).

The observations $\{X_{i1}, X_{i2}, \dots, X_{ip}, Y_i\}_{i=1,2,\dots,n}$ form the calibration set and are helpful in estimating the fitting parameters $\beta_1, \beta_2, \dots, \beta_p$. The least-squares method is the usual technique used to estimate the parameters. Hence the equation for the predicted value is:

$$\hat{Y}_i = b_o + b_1 X_{i1} + b_2 X_{i2} + \dots\dots\dots + b_p X_{ip} \quad (6)$$

where, *bi* is the estimate of each β_i parameters and $\hat{Y}_i$ is the predicted value.

The goal of the regression analysis is to determine the values of the parameters of the regression equation and then to quantify the goodness of the fit with respect to the dependent variable *Y*.

2.3. Statistical performance indices

In this study, in order to objectively evaluate model prediction, five statistical performance indices were computed: the correlation coefficient (r), and the index of agreement (d), the mean bias error (Bias), the mean absolute error (MAE) and the root mean squared error (RMSE). These indices are based on the deviations between predicted and original observation values. RMSE summarizes the difference between the observed and the imputed concentrations and was used to quantify the average error of model. Moreover, the MAE and RMSE were included in the comparison as more sensitive measures of residual error. Bias is the degree of correspondence between the mean prediction and the mean observation. Lower values of Bias are optimal, while bias values <0 indicate under-forecasting. Evaluation can also be undertaken by considering measures of agreement, such as the Pearson product moment correlation coefficient (r). The index of agreement is a bounded, relative measure that is capable of measuring the degree to which predictions are error-free. The denominator accounts for the model's deviation from the mean of the observations as well as the observation deviation from their mean. In a good model d and r should approach to 1 (Nunnari et al., 2004; Kukkonen et al., 2003). All these indices are formulated as follows;

$$r = \sqrt{1 - \frac{\sum_{i=1}^{N} (O_i - T_i)^2}{\sum_{i=1}^{N} \left(O_i - \hat{O}\right)^2}} \quad (7)$$

$$d = 1 - \frac{\sum_{i=1}^{N} (P_i - O_i)^2}{\sum_{i=1}^{N} \left(|P_i - \bar{O}| + |O_i - \bar{O}|\right)^2} \quad (8)$$

$$Bias = \frac{1}{N} \sum_{i=1}^{N} (O_i - P_i) \quad (9)$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |O_i - T_i| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (O_i - T_i)^2} \quad (11)$$

where, O_i and P_i are the observed and predicted pollution values, respectively, in $i = 1., 2., \dots, N$ days, $\hat{O}$ is the mean of the observed times series and *N* is the total number of observations. In addition, the standard deviations (σ) of the predicted time series (*P*) have been calculated.

2.4. The study area and database

The study area is the metropolitan city of Istanbul which is located at 41°N and 29°E. The Bosphorus separates Istanbul into two parts, the European and the Asian sides. The total area of both parts of the city is approximately 5700 km^2. More than 12 million people are living in Istanbul and more than 40% of Turkey's heavy industry is located in the city. For this reason, air pollution problems are of prime importance in Istanbul. The Greater Istanbul Metropolitan Municipality's, Directorate of Environmental Protection (GIMM-DEP) has conducted air pollution measurement at 10 observation stations located at various key topographic points around the city since 1992. In this study, the daily SO_2 and PM_{10} concentration data were measured at two stations located in Yenibosna and Umraniye, and daily meteorological data were measured at two stations located in Florya and Goztepe as shown in Fig. 4. We have categorized the sampling sites using the criteria proposed by the European Environmental (EU) Agency (EU) and shown in Table 1 (Dingenen et al., 2004). Table 1 shows the specific pollution sources near the air quality monitoring stations. Among these criteria are the distance of the stations from large pollution sources such as cities, power plants and major motorways, and the traffic volume.

In this study, daily SO_2 and PM_{10} data were collected by GIMM-DEP and measured using AF 21 M and MP 101 M sensors, respectively (Environmental Inc.) We evaluated data measured during 2002 and 2003. The 1460 data points for each air pollutant for the Yenibosna and Umraniye AQMS during this period. To predict the missing air pollutant concentration data, we used daily meteorological data

Fig. 4. Location of the air quality measurement and meteorology stations in Istanbul.

provided by the General Directorate of the Turkish State Meteorological Services (GDTSMS) in Istanbul. We used the Meteorological Stations located at Florya on the European side and Goztepe on the Asian side. The meteorological parameters used to predict the missing air pollutant concentrations in this study, as well as their notations and daily statistical evaluation during 2002–2003 are shown in Table 2.

2.5. Building the models

We estimated the daily missing concentrations of PM_{10} and SO_2 parameters during 2002–2003 for Yenibosna and Umraniye air pollution monitoring stations. These data sets were organized as follows: one of them was formed assuming a missing data percentage of 50. In this data set, it was assumed that data measurements were not performed every other day. The other, was formed using a missing data percentage of 20. It was assumed that data measurements were not performed for one out of every five days in this data set. The use of these assumptions in training and testing studies is explained in Table 3 in detail.

The most important factor in the establishment of the CNN model is neighboring relations. For this reason, we have calculated correlations between meteorological and pollution parameters using the statistical software package SPSS11.5. To improve prediction performance, the CNN model was set with side by side high correlation coefficients among the data values. In our CNN model, the elements of the input (u) and output (y) matrix structure are shown in Fig. 5. The elements of the input matrix consist of daily SO_2 and PM_{10} concentrations to be predicted for the 20% and 50% missed data. In the CNN training study, the elements of the output matrix consisted of all daily observed concentrations. We have designed a MATLAB 7.0 code on Pentium IV computers for our CNN model.

Table 1
Specific pollution sources and category by EU of the air pollutant sampling sites.

AQ stations	Pollution sources			Categorized by EU	
	Commercial	Industrial	Traffic	Urban background [a]	Kerbside [b]
Yenibosna	x	x	x	x	x
Umraniye	x	x		x	

[a] Urban backgound: <2500 vehicles/day within a radius of 50 m.
[b] Kerbside: within street canyons.

3. Results and discussion

3.1. PM_{10} and SO_2 pollution in Istanbul

Summary statistics of daily PM_{10} and SO_2 data between 1999 and 2003 at the Yenibosna and Umraniye stations are given in Table 4. The daily PM_{10} and SO_2 concentrations for each station are given in Fig. 6. The PM_{10} and SO_2 concentrations recorded at the Yenibosna station were higher than those at the Umraniye station. In Yenibosna, traffic, industry and residential populations are quite dense. The five-year average SO_2 concentration measured at the Yenibosna station was one and a half times higher than the concentration measured at the Umraniye station. As shown in Fig. 6, at both monitoring stations the results recorded in winter were five times higher than those measured in summer. The 24-hour PM_{10} limit of 50 μg/m^3 was exceeded on many days (more than 80%) for all stations. But the 24-hour SO_2 limit of 125 μg/m^3 was exceeded on only a few days (about five days) for all stations. Before 1995, the average SO_2 level was 250 μg/m^3 in Istanbul (Tayanç, 2000). After 1995, the use of natural gas instead of coal became more widespread and SO_2 levels have therefore begun to decrease. After 1999, the average SO_2 concentration was 25 μg/m^3. However, PM_{10} levels have not effectively decreased over this period. There is no significant difference in PM_{10} pollution

Table 2
The minimum, mean and maximum values of meteorological model parameters during 2002 and 2003 years.

Parameters	Notations	Units	Minimum		Mean		Maximum	
			Florya	Goztepe	Florya	Goztepe	Florya	Goztepe
Temperature	T	°C	−2.2	−2.2	14.7	14.7	31.2	32
Wind speed	WS	m/s	0.3	0.2	2.2	2.5	6.2	7.3
Sunshine	S	Hour	0	0	6.7	6.3	13.8	12.9
Rel. humidity	RH	%	43.3	38.7	72.2	74.8	95.7	96
Pressure	P	mbar	990.9	988.8	1012.5	1012.6	1031.4	1032.7
Cloudy	C	m	0	0	4.4	6.3	10	10
Wind direction	WD	North (N), South (S), West (W), East (E)	WSW		–		NNW	
Rainfall	R	mm	0	0	–	–	31.8	61.9

levels between winter and summer. The effect of long distance transport should be considered as well as the anthropogenic pollution sourced from industry, heating and transport (Karaca et al., 2009; Kindap, 2008).

Istanbul needs a long-term plan to address its air pollution problems since the city is home to the majority of industries and the population in Turkey. A continuous data record is very important. One problem with this record is the high number of missing data points in the AQM stations' records for the period analyzed, from 1999 to 2003. As shown in Table 4, the missing data rates for SO_2 and PM_{10} are 19.7% and 38.8%, respectively, at the Yenibosna station and 13.7% and 22.1%, respectively, at the Umraniye station. Because of the technical difficulties, the missing data fraction is as high as 50% for some one-year period. That's why the idea of such a study can be completed with the missing data, where it could be developed and implemented. Five years of data were examined and the 2002 to 2003 period was selected, because this period had a missing data fraction less than 8%. With this method (CNN), the pixel belonging to the time zone of missing data can be defined as 2-dimensional; not only the past data but also today and the day after data can be affected to predict missing data, effectively.

3.2. Analysis of CNN model

The data recording was performed as a continuous and periodic determination of air pollutant parameters at measurement stations located in different areas of the city. However, it was not always possible to obtain continuous data due to a malfunctioning of measurement devices, power failures or environmental factors. Data missing due to these factors negatively affect the results of modeling and pollutant tracing studies. In this study, a CNN model structure was tested under scenarios of 20% and 50% missing data fractions for SO_2 and PM_{10} concentrations measured on the Anatolian (Asian) and European sides of Istanbul during 2002 and 2003.

The CNN training process required approximately 4 and 5 min, respectively, to predict daily mean pollutant concentrations based on data from the Yenibosna and Umraniye AQM Stations. The processes were stopped when the error reached a value of 2.10^{-4}. Testing of the CNN approach with the optimized A,B,I templates occurred in real time. In training the CNN model using u and y matrices, we obtained A, B and I templates for each study as follows:

To predict 50% missing PM_{10} concentration data in Yenibosna:

$$A = \begin{bmatrix} 0.0426 & 0.0545 & 0.0521 \\ 0.0644 & 1.2885 & 0.0644 \\ 0.0521 & 0.0545 & 0.0426 \end{bmatrix}$$

$$B = \begin{bmatrix} -0.0615 & -0.0618 & -0.0613 \\ -0.0616 & -0.0614 & -0.0616 \\ -0.0613 & -0.0618 & -0.0615 \end{bmatrix} \qquad I = [-0.0614]. \tag{12}$$

To predict 20% missing PM_{10} concentration data in Yenibosna:

$$A = \begin{bmatrix} -0.0054 & 0.0075 & -0.0280 \\ 0.0048 & 1.005 & 0.0048 \\ -0.0280 & 0.0075 & -0.0054 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0065 & 0.0068 & 0.0069 \\ 0.0068 & 0.0069 & 0.0068 \\ 0.0069 & 0.0068 & 0.0065 \end{bmatrix} \qquad I = [0.0069] \tag{13}$$

To predict 50% missing SO_2 concentration data in Yenibosna:

$$A = \begin{bmatrix} 0.0537 & 0.0732 & 0.0636 \\ 0.0592 & 1.4524 & 0.0592 \\ 0.0636 & 0.0732 & 0.0537 \end{bmatrix} \tag{14}$$

$$B = \begin{bmatrix} -0.0731 & -0.0734 & -0.0733 \\ -0.0734 & -0.0734 & -0.0734 \\ -0.0733 & -0.0734 & -0.0731 \end{bmatrix} \qquad I = [-0.0754]$$

Table 3
The structure of SO_2 and PM_{10} data for the CNN and LR model training and testing.

	Training data sets during 2002–2003.	Testing data sets during 2002–2003.
Missing data percentage of 50	O_t, M_{t+1}, O_{t+2}, M_{t+3}, O_{t+4}, M_{t+5}, O_{t+6}, M_{t+7}, O_{t+8}, M_{t+9}, O_{t+10},…… O_{t+730},	M_t, O_{t+1}, M_{t+2}, O_{t+3}, M_{t+4}, O_{t+5}, M_{t+6}, O_{t+7}, M_{t+8}, O_{t+9}, M_{t+10},…… O_{t+730},
Missing data percentage of 20	O_t, O_{t+1}, O_{t+2}, O_{t+3}, O_{t+4}, M_{t+5}, O_{t+6}, O_{t+7}, O_{t+8}, O_{t+9}, M_{t+10},…… O_{t+730},	O_t, M_{t+1}, O_{t+2}, O_{t+3}, O_{t+4}, O_{t+5}, M_{t+6}, O_{t+7}, O_{t+8}, O_{t+9}, O_{t+10},…… O_{t+730},

O: Observed data; M: Missing data; t: daily.

$$u=\begin{bmatrix} RH_{t1} & RH_{t2} & RH_{t3} & \cdots & RH_{t730} \\ R_{t1} & R_{t2} & R_{t3} & \cdots & R_{t730} \\ C_{t1} & C_{t2} & C_{t3} & \cdots & C_{t730} \\ WS_{t1} & WS_{t2} & WS_{t3} & \cdots & WS_{t730} \\ MAP_{t1} & MAP_{t2} & MAP_{t3} & \cdots & MAP_{t730} \\ T_{t1} & T_{t2} & T_{t3} & \cdots & T_{t730} \\ P_{t1} & P_{t2} & P_{t3} & \cdots & P_{t730} \\ S_{t1} & S_{t2} & S_{t3} & \cdots & S_{t730} \\ WD_{t1} & WD_{t2} & WD_{t3} & \cdots & WD_{t730} \end{bmatrix} \quad y=\begin{bmatrix} RH_{t1} & RH_{t2} & RH_{t3} & \cdots & RH_{t730} \\ R_{t1} & R_{t2} & R_{t3} & \cdots & R_{t730} \\ C_{t1} & C_{t2} & C_{t3} & \cdots & C_{t730} \\ WS_{t1} & WS_{t2} & WS_{t3} & \cdots & WS_{t730} \\ AP_{t1} & AP_{t2} & AP_{t3} & \cdots & AP_{t730} \\ T_{t1} & T_{t2} & T_{t3} & \cdots & T_{t730} \\ P_{t1} & P_{t2} & P_{t3} & \cdots & P_{t730} \\ S_{t1} & S_{t2} & S_{t3} & \cdots & S_{t730} \\ WD_{t1} & WD_{t2} & WD_{t3} & \cdots & WD_{t730} \end{bmatrix}$$

Fig. 5. Input (u) and output (y) matrices of our CNN model. (AP: Air Pollutant, SO_2 and PM_{10} in observed; MAP: air pollutant with missing data).

To predict 20% missing SO_2 concentration data in Yenibosna:

$$A=\begin{bmatrix} 0.0619 & 0.0697 & 0.0674 \\ 0.0570 & 1.4848 & 0.0570 \\ 0.0674 & 0.0697 & 0.0619 \end{bmatrix} \tag{15}$$

$$B=\begin{bmatrix} -0.0752 & -0.0755 & -0.0754 \\ -0.0754 & -0.0754 & -0.0754 \\ -0.0754 & -0.0755 & -0.0752 \end{bmatrix} \quad I=[-0.0754]$$

To predict 50% missing PM_{10} concentration data in Umraniye:

$$A=\begin{bmatrix} -0.0024 & 0.0400 & 0.0308 \\ 0.0541 & 1.0350 & 0.0541 \\ 0.0308 & 0.\ 0400 & -0.0024 \end{bmatrix} \tag{16}$$

$$B=\begin{bmatrix} -0.0238 & -0.0237 & -0.0240 \\ -0.0238 & -0.0239 & -0.0238 \\ -0.0240 & -0.0237 & -0.0238 \end{bmatrix} \quad I=[-0.0239]$$

To predict 20% missing PM_{10} concentration data in Umraniye:

$$A=\begin{bmatrix} -0.0054 & 0.0097 & -0.0260 \\ 0.0038 & 1.005 & 0.0038 \\ -0.0260 & 0.0097 & -0.0054 \end{bmatrix} \tag{17}$$

$$B=\begin{bmatrix} 0.0090 & 0.0089 & 0.0089 \\ 0.0088 & 0.0089 & 0.0088 \\ 0.0089 & 0.0088 & 0.0090 \end{bmatrix} \quad I=[0.0089]$$

To predict 50% missing SO_2 concentration data in Umraniye:

$$A=\begin{bmatrix} 0.0142 & -0.0182 & 0.0277 \\ 0.0157 & 1.1553 & 0.0157 \\ 0.0277 & -0.0182 & 0.0142 \end{bmatrix} \tag{18}$$

$$B=\begin{bmatrix} -0.0271 & -0.0273 & -0.0271 \\ -0.0272 & -0.0273 & -0.0272 \\ -0.0271 & -0.0273 & -0.0271 \end{bmatrix} \quad I=[-0.0273]$$

To predict 20% missing SO_2 concentration data in Umraniye:

$$A=\begin{bmatrix} 0.0402 & 0.0131 & 0.0289 \\ 0.0276 & 1.2505 & 0.0276 \\ 0.0289 & 0.0131 & 0.0402 \end{bmatrix} \tag{19}$$

$$B=\begin{bmatrix} -0.0429 & -0.0433 & -0.0433 \\ -0.0433 & -0.0433 & -0.0433 \\ -0.0433 & -0.0433 & -0.0429 \end{bmatrix} \quad I=[-0.0433]$$

Here, neighborhood (r) is chosen as 1. To guarantee stability of the CNN, the templates are symmetric. We have replaced the template values obtained in Eqs. (12)–(19) with those from Eqs. (2)–(3). In the optimization process, all template coefficients were chosen to four significant figures. We have especially chosen a linear region of the piece-wise linear function as in Fig. 3. Thus, we have obtained multilevel CNN outputs between −1 and +1 values. Furthermore, we have mapped CNN output values to actual measured values over the range of 0–250 μg/m^3 for SO_2 and 0–500 μg/m^3 for PM_{10}. As a result, we have reached precise results that are relatively close to the desired concentrations.

Table 4
Summary statistics of daily PM_{10} and SO_2 concentrations (μg/m^3) of each station between 1999 and 2003 years.

Stations/pollution	Valid data number	Missing data number	Mean	Std. deviation	Min	Max	Median
YENİBOSNAPM$_{10}$	1118	708	64.2	33.8	10	272	55.5
ÜMRANİYE PM$_{10}$	1422	403	53.8	30.6	5	287	46.4
YENİBOSNASO$_2$	1467	359	30.2	28.6	0	205	23.0
ÜMRANİYE SO$_2$	1575	250	20.2	20.4	0	166	13.6

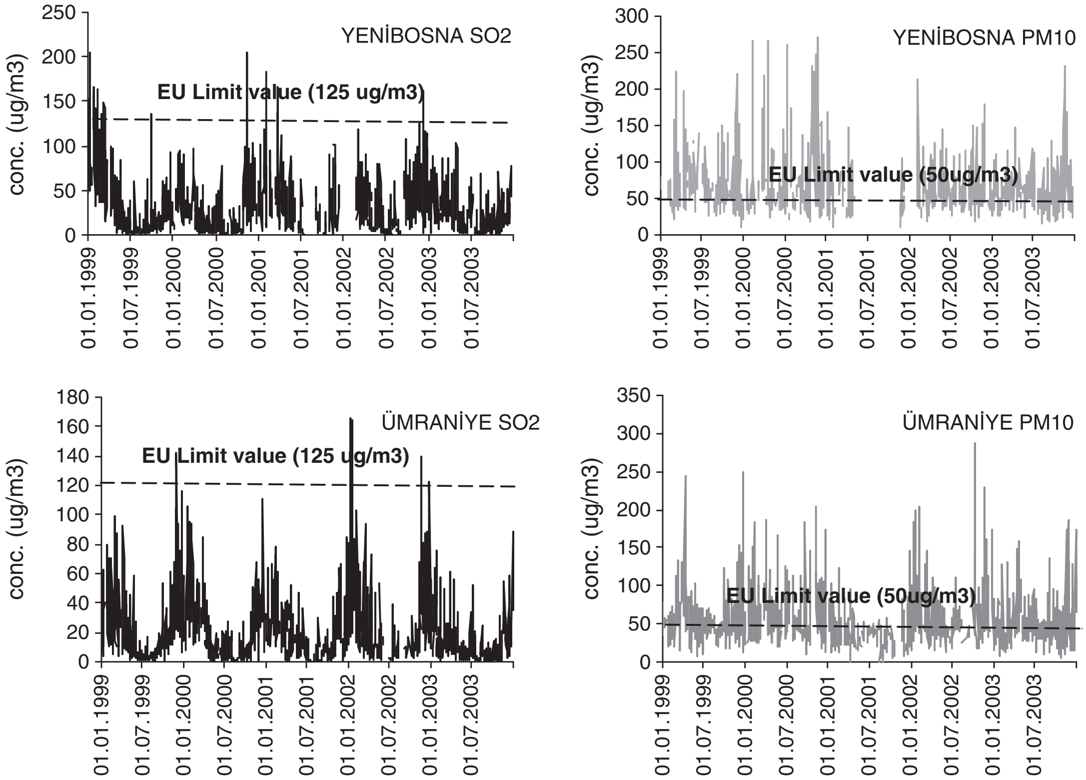

Fig. 6. Variations of daily average PM_{10} and SO_2 concentrations of the Yenibosna and Ümraniye air quality stations in Istanbul.

In this study, the Multiple Linear Regression (LR) model was developed as a comparison to the performance of the CNN-based approach. Linear Regression Equations were derived by using the training data sets defined in Table 3. Pollutant concentrations were calculated by using these derived equations. The equations were given below;

To predict 50% missing PM_{10} concentration in Yenibosna:

$$PM_{10t+1} = 0.164\,PM_{102t} - 0.065\,T_t + 0.259\,R_t - 5.420\,WS_t + 0.000\,WD_t - 0.039\,RH_t + 1.105\,P_t - 1.465\,C_t - 0.577S_t - 1045.152. \quad (20)$$

To predict 20% missing PM_{10} concentration in Yenibosna:

$$PM_{10t+1} = 0.259\,PM_{102t} - 0.341\,T_t + 0.083\,R_t - 5.187\,WS_t + 0.006\,WD_t - 0.044\,RH_t + 0.703\,P_t - 0.954\,C_t - 0.034\,S_t - 642.105. \quad (21)$$

To predict 50% missing SO_2 concentration in Yenibosna:

$$SO_{2t+1} = 0.381\,SO_{2t} - 0.863\,T_t + 0.097\,R_t - 3.463\,WS_t + 0.008\,WD_t - 0.197\,RH_t + 0.210\,P_t - 1.739\,C_t - 1.003\,S_t - 147.250. \quad (22)$$

To predict 20% missing SO_2 concentration in Yenibosna:

$$SO_{2t+1} = 0.416\,SO_{2t} - 0.990\,T_t + 0.19\,R_t - 4.212\,WS_t + 0.004\,WD_t - 0.285\,RH_t + 0.254\,P_t - 0.358\,C_t - 0.068\,S_t - 195.177. \quad (23)$$

To predict 50% missing PM_{10} concentration in Ümraniye:

$$PM_{10t+1} = 0.316\,PM_{102t} - 0.162\,T_t + 0.364\,R_t - 4.372\,WS_t + 0.028\,WD_t - 0.052\,RH_t + 0.708\,P_t - 1.008\,C_t - 0.841\,S_t - 656.624. \quad (24)$$

To predict 20% missing PM_{10} concentration in Ümraniye:

$$PM_{10t+1} = 0.280\,PM_{102t} - 0.653\,T_t + 0.120\,R_t - 5.765\,WS_t + 0.011\,WD_t - 0.053\,RH_t + 0.276\,P_t - 1.555\,C_t - 0.588\,S_t - 203.686. \quad (25)$$

To predict 50% missing SO_2 concentration in Ümraniye:

$$SO_{2t+1} = 0.444\,SO_{2t} - 0.377\,T_t - 0.166\,R_t - 2.602\,WS_t + 0.001\,WD_t - 0.154\,RH_t + 0.549\,P_t - 0.480\,C_t - 0.762\,S_t - 515.177. \quad (26)$$

To predict 20% missing SO_2 concentration in Ümraniye:

$$SO_{2t+1} = 0.377\,SO_{2t} - 0.677\,T_t + 0.046\,R_t - 3.259\,WS_t - 0.001\,WD_t - 0.207\,RH_t + 0.327\,P_t - 0.283\,C_t - 0.409\,S_t - 282.587. \quad (27)$$

The data sets including all available measurements were used to train the CNN model, then the coefficients obtained in the training process were used to test, the model under the scenarios of 20 percent and 50 percent missing data. The correlation coefficients obtained after training the CNN and LR models are given in Table 5. CNN training results had much higher correlation coefficients than the LR training results in both the 20% and 50% data deficiency cases. Additionally, the

Table 5
Training and validation results for CNN and LR model.

Site	Missing data (%)	Pollution	Model	Correlation (r), training set	Correlation (r), validation set
Yenibosna	50	PM_{10}	CNN	0.66	0.62
			LR	0.48	–
		SO_2	CNN	0.72	0.70
			LR	0.69	–
	20	PM_{10}	CNN	0.89	0.85
			LR	0.51	–
		SO_2	CNN	0.88	0.80
			LR	0.68	–
Umraniye	50	PM_{10}	CNN	0.73	0.64
			LR	0.51	–
		SO_2	CNN	0.75	0.69
			LR	0.71	–
	20	PM_{10}	CNN	0.87	0.84
			LR	0.52	–
		SO_2	CNN	0.90	0.73
			LR	0.70	–

highest SO_2 prediction value was calculated as 0.90 in Umraniye and the lowest value was calculated as 0.72 in the Yenibosna stations. The highest PM_{10} prediction value was calculated as 0.89 and the lowest value was calculated as 0.72 at the Yenibosna stations.

The data set was tested using the A, B, I (Eqs. (12)–(19)) in terms of the CNN model obtained after training. We estimated the daily mean missing PM_{10} and SO_2 concentrations during 2002 and 2003, and compared these estimates to observed data (Figs. 7 and 8) from the Yenibosna and Umraniye stations, respectively. The CNN and LR model results were also checked by calculating five different statistical indices, given in Eqs. (7)–(1), which are based on the deviations between predicted values and original observations. The final results of statistical model evaluation for the daily mean missing PM_{10} and SO_2 concentrations during 2002 and 2003 have been presented in Table 6. For both pollutants and both missing data assumptions, the results have been

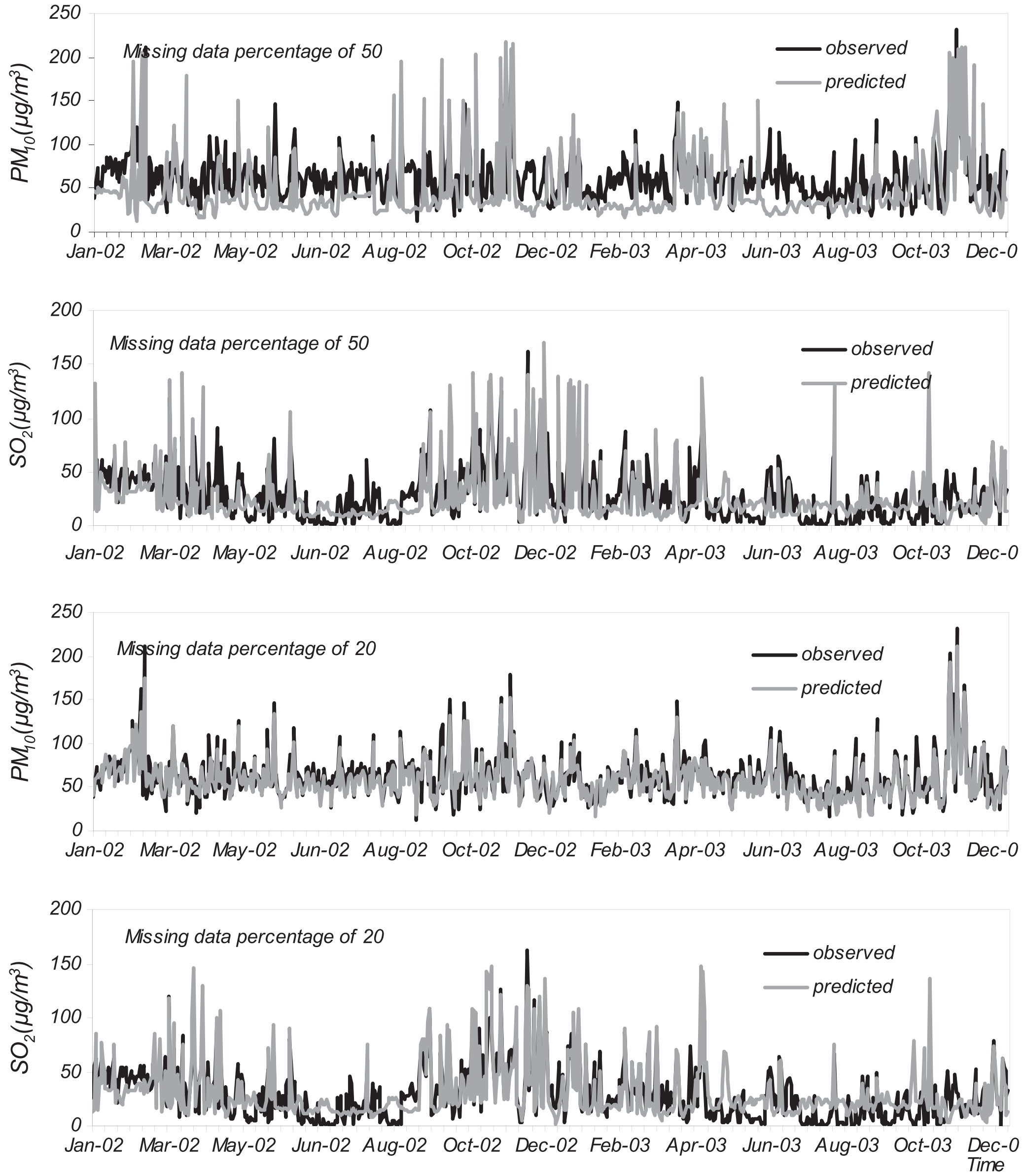

Fig. 7. Two years of observed and CNN model predicted daily mean PM_{10} and SO_2 concentrations at the Yenibosna Station.

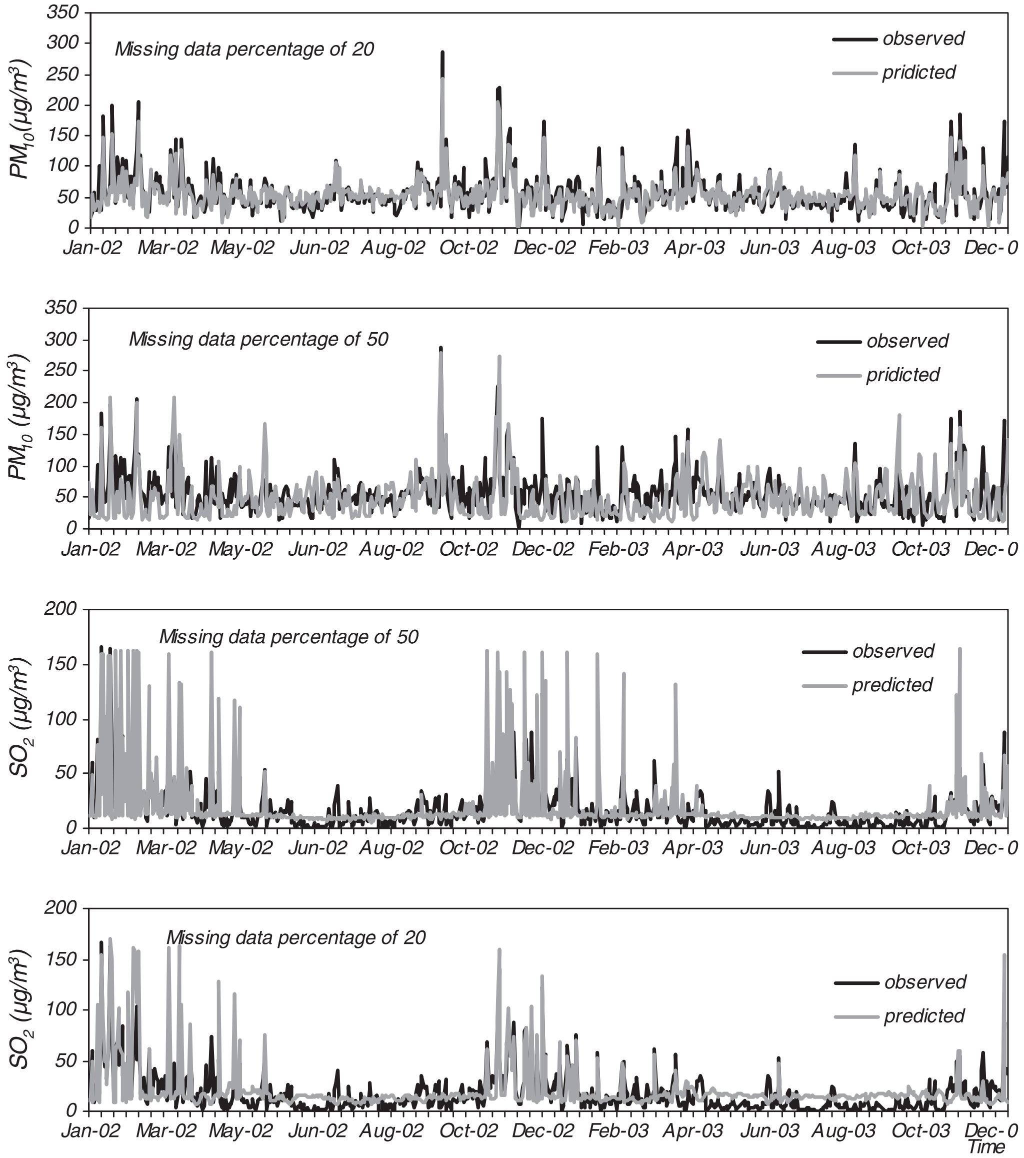

Fig. 8. Two years of observed and CNN model predicted daily mean PM_{10} and SO_2 concentrations at the Umraniye Station.

Table 6
Model performance indices for the CNN model. The results differ by the missing data percentage, Yenibosna and Umraniye air quality stations and PM_{10} and SO_2 pollution.

Stations	Air pollutant	MDP (%)	Model	Statistical performance indices								
				Max.	Min.	Avrg.	σ	r	d	Bias	MAE	RMSE
Yenibosna	PM_{10}	50	CNN	218	13	47	36.1	0.57	0.70	16.3	27.5	34.4
			LR	112	21.5	63	12.9	0.50	0.61	0.03	16.0	22.6
		20	CNN	211	15	59	23	0.87	0.92	4.2	9.9	13.6
			LR	121	26	63	13.1	0.51	0.62	0.31	15.7	22.5
	SO_2	50	CNN	170	0.3	26	26.1	0.67	0.81	2.76	15.5	20.5
			LR	96	−2.2	29	15.6	0.66	0.77	0.01	13.1	17.7
		20	CNN	148	2.2	30	25.5	0.73	0.85	−0.91	13.6	18.0
			LR	102	−2.9	28	16.1	0.67	0.78	0.52	12.8	17.6
Ümraniye	PM_{10}	50	CNN	279	10	49	34.3	0.54	0.74	6.87	24.4	32.3
			LR	126	8.9	56	16.8	0.53	0.64	−0.35	18.4	26.7
		20	CNN	244	2.5	54	24.3	0.86	0.91	1.95	11.1	16.6
			LR	122	10.9	55	16.5	0.54	0.65	0.11	18.1	26.6
	SO_2	50	CNN	164	5.5	21	30.3	0.60	0.73	−2.43	13.3	24.3
			LR	103	−5.1	18	15	0.70	0.80	−0.12	9.5	15.0
		20	CNN	170	6	21	24.3	0.75	0.85	−2.54	11.5	16.5
			LR	92	−5.8	18	14	0.71	0.80	0.30	9.3	15.0

MDP: Missing data percentage.

presented separately for each Station. It can be inferred from the bias values in Table 6 that all of the CNN model predictions are less than the measured concentration values except for the data set with 20% deficiency for SO_2 at the Umraniye station. When the RMSE values are controlled, RMSE values are mostly equal or close to LR model values in the case of 20% data deficiency. Index of agreement values reached a maximum of 0.81 in the case of 20% data deficiency and reached a maximum of 0.92 in the case of 50% data deficiency. Reducing the data deficiency in the data set boosts the performance of the CNN model. When all of the model results are evaluated in general, the indexes of agreement and correlation in the LR predictions are found to be less than those of the CNN predictions.

We also evaluated the effects of seasonal changes on CNN predictions. The observed versus CNN-predicted daily mean PM_{10} and SO_2 values for the 2002–2003 winter and summer periods are plotted in Fig. 9 and the correlation coefficients are presented in Fig. 9. It is evident that the CNN provides more reliable predictions of the daily mean PM_{10} and SO_2

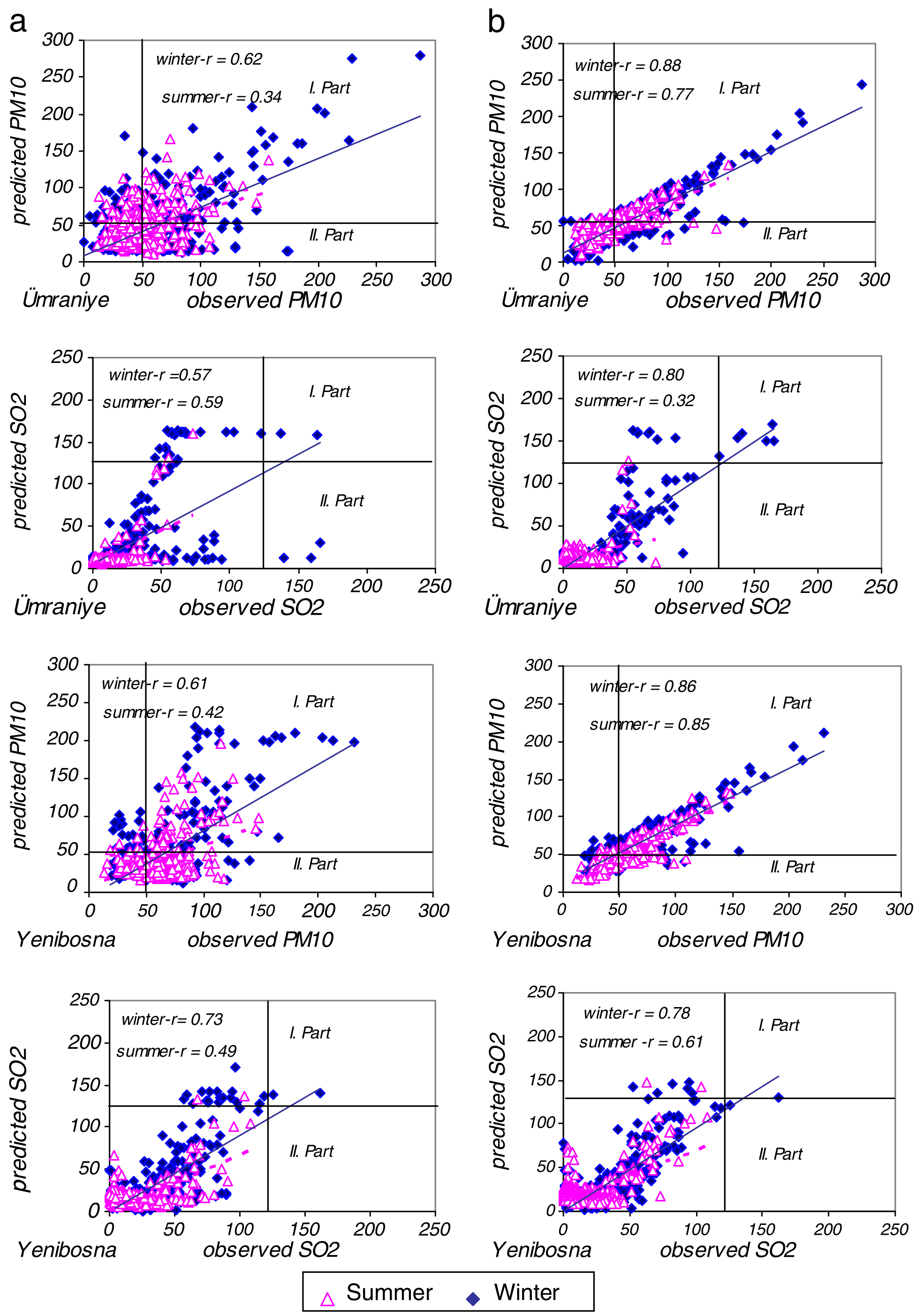

Fig. 9. Scatter plots of predicted versus observed concentrations of SO_2 and PM_{10} at Yenibosna and Umraniye on the CNN test data. a): Missing data percentage of 50 and b): Missing data percentage of 20. I. Part: events exceeding the attention level correctly predicted; II. Part: events exceeding the attention level not correctly predicted.

concentrations at all stations during the winter as compared to the summer, as shown by the greater agreement of observed and predicted values for the winter seasons.

The relevant levels of daily mean SO_2 and PM_{10} concentrations, according to EU legislation (see the EC Normative-Council directive 1999/30/EC of 22 April 1999 relating to limit concentration limits for sulfur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air) are 125 μg/m^3 and 50 μg/m^3, respectively, and are not to be exceeded more than 3 and 35 times a year, respectively. The environmental laws in Turkey are being revised according to guidelines of the European Union. When Draft Air Pollution Control Laws are considered, it will be necessary to assert the EU limit values. There are two parts presented in Fig. 9: Part I demonstrates the correctly predicted events exceeding the attention level, and Part II demonstrates the events exceeding the attention level that were not correctly predicted. Three values in Part II were determined in the prediction of SO_2 for the Umraniye Station data set with 20% data deficiency; the majority of the SO_2 concentrations were lower than the EU-mandated level. Approximately 50% of the measured concentration values are higher than the limit values. This situation was observed in the CNN model prediction, and the studies with 20% data deficiency yielded predictions with 82% success (predicted data in Part I).

4. Conclusion

In this study, the major air pollutants of concern for the city of Istanbul, particulate matter (PM) and sulfur dioxide (SO_2), were estimated using a CNN approach. There are many computational methods available for air pollutant modeling. One of the frequently used methods is the use of an Artificial Neural Network (ANN). In ANN modeling, the training process time increases as the problem becomes increasingly complex. To reduce the complexity of the calculations used by the ANN, Chua and Yang introduced the Cellular Neural Network (CNN) in 1988 as a new non-linear, dynamic neural network structure. In a CNN, the correlations between neighboring pixels are modeled by cloning templates with a limited number of elements and using these pixels for solving complex problems.

Here, we model missing daily mean PM_{10} and SO_2 air pollutant concentration data in Istanbul. Comparing the results obtained using the CNN model with those obtained using a LR model, we observed that the CNN model provides more reliable predictions. In previous similar ANN modeling studies the correlation coefficient values ranged between 0.50 and 0.80 (Mok and Tam, 1998; Chelani et al., 2002; Sahin et al., 2005; Hooyberghs et al., 2005; Slini et al., 2006). In this paper, the measured r values for the CNN model were found to be between 0.54 and 0.87 for daily mean PM_{10} concentrations and 0.60 and 0.75 for daily mean SO_2 concentrations.

These result shows that the CNN modeling technique can be considered a promising approach for air pollutant prediction. We have proposed a new method for modeling the air-pollution problem using a CNN. In addition, we propose to test the ability of CNN models to model other environmental pollution problems. We specifically propose to apply CNN methods to three-dimensional air pollution modeling problems in the future.

Acknowledgments

We are grateful to the Istanbul Municipality, Environmental Protection Directorate and the Department of Meteorology in Istanbul for their help in obtaining actual data. This work was supported by the Research Fund of the University of Istanbul. Project Number: T-486/25062004.

References

Agirre-Basurko, E., Ibarra-Berastegi, G., Madariaga, I., 2006. Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Balbao area. Environmental Modeling & Software 21, 430–446.

Albora, A.M., Ucan, O.N., Ozmen, A., Ozkan, T., 2001. Separation of Bouguer anomaly map using cellular neural network. Journal of Applied Geophysics 46, 129–142.

Arena, P., Caponetto, R., Fortuna, L., Manganaro, G., 1997. Cellular neural network to explore complexity. Soft Computing 01, 120–136.

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

Cimagalli, V., 1993. Cellular neural networks a review. Proceedings of sixth Italian workshop on parallel architectures and Neural Networks, Vietri Sul Mare, Italy, May 12–14.

Chelani, A.B., Chalapati Rao, C.V., Phadke, K.M., Hasan, M.Z., 2002. Prediction of sulfur dioxide concentration using artificial neural networks. Environmental Modeling & Software 17, 161–168.

Chua, L.O., Yang, L., 1988. Cellular neural networks: application. IEEE Transactions on Circuits and Systems 35 (10), 1273–1290.

Dingenen, R.V., Raes, F., Putaud, J.P., et al., 2004. A European aerosol phenomenology-1; physical characteristics of particulate matter at kerbside, urban, rural and background sites in Europe. Atmospheric Environment 38, 2561–2577.

Dirks, K.N., Johns, M.D., Hay, J.E., Sturman, A.P., 2002. A simple semi-empirical model for predicting missing carbon monoxide concentrations. Atmospheric Environment 36, 5953–5959.

Erturk, F., 1986. Investigation of strategies for the control of air pollution in the Golden Horn Region, Istanbul, using a simple dispersion model. Environmental Pollution B 11, 161–168.

Fernandez-Ferrero, A., Saenz, J., Ibarra-Berastegi, G., Fernadez, J., 2009. Evaluation of statistical downscaling in short range precipitation forecasting. Atmospheric Research 94 (3), 448–461.

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

Grassi, G., Grieco, L.A., 2002. Object-oriented image analysis via analogic CNN algorithms—part I: motion estimation. 7th IEEE International Workshop on Cellular Neural Networks and Their Applications, pp. 172–180.

Grivas, G., Chaloulakou, A., 2006. Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. Atmospheric Environment 40, 1216–1229.

Gupta, P., Christopher, S.A., 2009. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. Journal of Geophysical research, Atmospheres 114 (D20205), 1029–1040.

Guzelis, C., Karamahmut, S., 1994. Recurrent perceptron learning algorithm for completely stable cellular neural networks. Proc. Third IEEE Int. Workshop on Cellular Neural Network and Applications, Rome, Italy. 177–182.

Hadad, K., Piroozmand, A., 2007. Application of cellular neural network (CNN) method to the nuclear reactor dynamics equations. Annals of Nuclear Energy 34, 406–416.

Hanedar, A., Alp, K., Kaynak, B., Baek, J., Avşar, E., Odman, M.T., 2011. Concentrations and sources of PAHs at three stations in İstanbul, Turkey. Atmospheric Research 99, 391–399.

Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., Brasseur, O., 2005. A neural network forecast for daily average PM10 concentrations in Belgium. Atmospheric Environment 39, 3279–3289.

Im, U., Tayançi, M., Yenigün, O., 2008. Interaction patterns of major photochemical pollutants in Istanbul, Turkey. Atmospheric Research 89, 382–390.

Junninen, H., Niska, H., Tuppurainen, K., et al., 2004. Methods for imputation of missing values in air quality data sets. Atmospheric Environment 38, 2895–2907.

Karaca, F., Anil, I., Alagha, O., 2009. Long-range potential source contributions of episodic aerosol events to PM10 profile of a megacity. Atmospheric Environment 43, 5713–5722.

Kindap, T., 2008. Identifying the trans-boundary transport of air pollutants to the city of Istanbul under specific weather conditions. Water, Air, and Soil Pollution 189, 279–289.

Kukkonen, J., Partanen, L., Karppinen, A., et al., 2003. Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modeling system and measurements in central Helsinki. Atmospheric Environment 37, 4539–4550.

Kurt, A., Gülbağcı, B., Karaca, F., Alagha, O., 2008. An online air pollution forecasting system using neural networks. Environmental International 34, 592–598.

Lary, D.J., Remer, L.A., Macneill, D., Roscoe, B., Paradise, S., 2009. Machine learning and bias correction of MODIS aerosol optical depth. IEEE Geoscience and Remote Sensing Letters 6 (4), 694–698.

Manzato, A., 2007. Sounding-derived indices for neural network based short-term thunderstorm and rainfall forecasts. Atmospheric Research 83 (2–4), 349–363.

Mok, K.M., Tam, S.C., 1998. Short-term prediction of SO_2 concentration in Macau with artificial neural networks. Energy and Buildings 28, 279–286.

Nagendra, S.M.S., Khare, M., 2006. Artificial neural network approaches for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. Ecological Modeling 190, 99–115.

Nunnari, G., Dorling, S., Schlink, U., Cawley, G., Foxal, R., Chatterton, T., 2004. Modelling SO_2 concentration at a point with statistical approaches. Environmental Modeling & Software 19, 887–905.

Onat, B., Bayat, C., Sahin, U., 2004. PM_{10} dispersion modelling: urban case study from Turkey. Fresenius Environmental Bulletin 13 (9), 889–894.

Ordieres, J.B., Vergara, E.P., Capuz, R.S., Salazar, R.E., 2005. Neural network prediction model for fine particulate matter ($PM_{2.5}$) on the US–Mexico border in El Paso (Texas) and Ciudad Juarez (Chihuahua). Environmental Modeling & Software 20, 547–559.

Ozcan, H., Bilgili, E., Sahin, U., Ucan, O.N., Bayat, C., 2007. Modelling of tropospheric ozone concentrations using genetically trained multi-level cellular neural networks. Advances in Atmospheric Science 24 (5), 907–914.

Perez, P., Reyes, J., 2006. An integrated neural network model for PM10 forecasting. Atmospheric Environment 40, 2845–2851.

Sahin, U., 2005. Modelling PM and THC Distribution in Istanbul with Artificial Neural Network Techniques. Istanbul University, Science Technology Institute, PhD Thesis.

Sahin, U., Ucan, O.N., Bayat, C., 2005. Modelling of SO_2 distribution in Istanbul using artificial neural networks. Environmental Modeling and Assessment 10, 135–142.

Sahin, U., Ucan, O.N., Soyhan, B., Bayat, C., 2004. Modelling of CO distribution in Istanbul using artificial neural networks. Fresenius Environmental Bulletin 13 (9), 839–845.

Saral, A., Ertürk, F., 2003. Prediction of ground level SO_2 concentrations using artificial neural network. Water, Air, and Soil Pollution 3, 297–306.

Slini, T., Kaprara, A., Karatzas, K., Moussiopoulos, N., 2006. PM_{10} forecasting for Thessaloniki, Greece. Environmental Modelling & Software 21, 559–565.

Ucan, O.N., Bilgili, E., Albora, A.M., 2001. Detection of buried objects on archeological areas using genetic cellular neural network. European Geophysical Society XXVI General Assembly, 3, pp. 223–231. France.

Tayanç, M., 2000. An assessment of spatial and temporal variation of sulphur dioxide levels over Istanbul, Turkey. Environmental Pollution 107, 61–69.

Thai, V.D., Cat, P.T., 2008. Modelling air-pollution problem by cellular neural network. 10th Intl. Conf. on Control, Automation, Robotics and Vision Hanoi, Vietnam, 17–20 December 2008, pp. 1115–1118.

Yildirim, Y., Bayramoğlu, M., 2006. Adaptive neuro-fuzzy based modeling for prediction of air pollution daily levels in city of Zonguldak. Chemosphere 63, 1575–1582.