Proceedings of the Third International Symposium on Air Quality Management at Urban, Regional and Global Scales. pp. 1567-1581, 26-30 September 2005, Istanbul – Turkey

PREDICTION OF MISSING SO_2 AND PM_{10} CONCENTRATIONS USING CELLULAR NEURAL NETWORK (CNN)

Cuma Bayat[1], Osman N. Ucan[2], Ulku Sahin[1*] and Kurtulus Ozcan[1]

[1]Istanbul University, Engineering Faculty, Environmental Eng. Dept. 34 850, Avcilar, Istanbul, Turkey * ulkualver@istanbul.edu.tr
[2]Istanbul University, Engineering Faculty, Electrical-Electronics Eng. Dept. 34 850, Avcilar, Istanbul, Turkey

ABSTRACT

Air pollutant, monitoring using continuous samplers are carried out in most major urban centers in the world and generally forms the basis for air quality assessments. Such assessment less reliable as the proportion of data missing due to equipment failure and periods of calibration increases. Missing data, i.e incomplete data matrices, are a problem that is repeatedly encountered in environmental research. In this paper, we predict missing concentrations of PM_{10} and SO_2 air pollutant from Istanbul Yenibosna and Ümraniye air pollution measurement stations with meteorological parameters: temperature, pressure, sunshine, cloudy, rainfall, wind speed and wind direction relative humidity etc.., from Istanbul Florya and Göztepe meteorological stations using CNN model. We consider one-neighborhood relation and guarantee stability of CNN by choosing symmetric feedback and feed-forward cloning matrices of A, B. Here, the total number of different assigned elements of matrices (A,B) and bias constant (I), are limited to only 11 and our air pollution problem is optimized by altering these 11 elements. Then, we correlate CNN outputs with actual measured values by using three statistical performance criteria. In this paper, we have measurement index of agreement (*d*) between 0.69 and 0.92 for all pollutant. CNN model for PM_{10} are producing considerably better predictions than SO_2. As a result, CNN-based approaches can be considered as a compromising approach in air pollutant prediction.

Key Words: Missing data, Air Quality, Particulate Matter (PM), Sulfur Dioxide (SO_2), Meteorology, Cellular Neural Network (CNN).

1. INTRODUCTION

Istanbul has the risk of air pollution, since it is the most industrialized city of Turkey with lots of factories. The main air pollutants of Istanbul are Particulate Matter (PM), Sulfur Dioxide (SO_2), which threaten human health and corrupt air quality. It is necessary that measured continuously these air pollutants for the regional authorities. As the incomplete data, missing data is occur, air quality assessment is less reliable. Missing data are a problem that is repeatedly encountered in environmental research. The situation may be the result of insufficient sampling, errors in measurements or faults in data acquisition. Whatever the reasons, discontinuities pose a significant obstacle for time-series prediction schemes, which generally require continuous data as a condition for their use.

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 something similar. Another approach to the problem is to use a simplistic interpolation method such as assuming the season's average concentration for the particular time of day that is missing, or linearly interpolating between the previous and following day, in order to obtain continuous data sets. Neither of these methods is ideal since the meteorology on the missing day may have been significantly different from the days on which the interpolation was based, leading to unrealistic predictions (Dirks et al., 2002). Clearly, a complementary method is required.

There are many deterministic and stochastic approaches in modeling of the air pollutants. As a well-known stochastic approach, Artificial Neural Network (ANN) has been applied to various environmental problems since 1990 and some satisfactory results are obtained. In many studies, ANN is applied to predict SO_2 concentration (Boznar et al., 1993; Mok and Tam, 1998; Saral, 1999; Chelani et al., 2002). Perez et al. (2000) have compared the $PM_{2.5}$ predictions produced by three different methods: multi-layer neural networks, linear regression and persistence. Gardner and Dorling (1998) have examined all main studies and summarized the use of ANN in environmental air pollution. Kukkonen et al. (2003) have studied five neural network 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) have applied the multi-layer neural network model to predict daily CO concentrations using meteorological variables as predictors for the European part of Istanbul, Turkey. Junninen et al., (2004) have applied regression based imputation, nearest neighbor interpolation, self organizing map, multi-layer perceptron model and hybrid methods to simulate missing air quality data. In all study, it is reported that ANN could be used to develop efficient air-quality analysis and prediction models in future. But in ANN, the training process becomes more complex and needs long time durations as the number of weight coefficients of ANN rise up to millions due to the complexity of environmental study.

To reduce weight coefficients, Chua and Yang (1988) have introduced Cellular Neural Network (CNN) in 1988. Since each cell of the CNN is represented by a separate analog processor and since 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; Uçan et al., 2001; Grassi, 2002).

In this study, we have applied CNN to predict the daily mean missing concentrations of PM_{10} and SO_2 pollutants in the Yenibosna and Ümraniye-Istanbul region of Turkey. PM_{10} and SO_2 pollutants and meteorological parameters are measured from Yenibosna and Ümraniye air pollution monitoring stations, and Istanbul-Florya and Göztepe meteorological stations.

2. MATERIALS AND METHODS

2.1. Study Area and Data

The study area is in Istanbul metropolitan city, which is located 41°N and 29°E. The Bosphorus channel separates this city into two parts, the European and the Asian side. The total area of the all parts of city is about 5700 km^2. More than 12 millions people are living and more than 40 % of the heavy industrial activities of Turkey are located in Istanbul. For this reason, air pollution problems are important in Istanbul. For this reason, air pollution problems are important for this city. Istanbul Greater Metropolitan Municipality, Directorate of Environmental Protection (IGMM-DEP) has made air pollution measurements in 10 stations placed in various points of Istanbul considering the topography of the city since 1992. In this study, the daily SO_2 and PM_{10} concentration data was measured by two stations located in Yenibosna and Ümraniye and the daily meteorological data was measured by two stations located in Florya and Göztepe as shown in Figure 1.

Figure 1. Location of the air quality measurement stations in Istanbul and study area (Y-AQMS: Yenibosna Air Quality Measurement Station, Ü-AQMS: Ümraniye Air Quality Measurement Station).

Air pollution measurement station, Yenibosna is in Bahcelievler County of the European side of Istanbul, and Ümraniye is in the Asian side of Istanbul. 5 % and 5.5 % of Istanbul population is lives in Bahçelievler and Ümraniye, respectively. In heating, 60 % natural gases and 40 % fossil fuel are used. Goncaloglu (2000) has investigated the overall industrial factories in Istanbul and emission inventory is listed. Our working area, Bahcelievler and Ümraniye has found to be the polluted county of Istanbul due to the emission ratios of SO_2 and PM_{10} and the usage of fuel oil 4 and coal.

In this study, the daily SO_2 and PM_{10} data were taken by IGMM-DEP and measured using AF 21 M and MP 101 M sensors, respectively, produced by the Environmental Inc. We have evaluated data measured during 2002 and 2003 years. The numbers of total data is 1460 per one air pollutant for Yenibosna and Ümraniye AQMS during 2002-2003. The monitoring data is designed to meet the requirements of training and testing CNN.

The climate of Istanbul is of Mediterranean type. The summers tend to be hot and winters are cold and wet. The Bosphorus, Marmara and Black Sea influence the climate of Istanbul. Here, the General Directorate of the Turkish State Meteorological Services (GDTSMS) in Istanbul provided the daily meteorological data. There are 17 meteorology stations in various points of Istanbul. We have used Florya in European side and Göztepe in Asian side Meteorological Stations data because of its being close to our working stations, Yenibosna and Ümraniye. To predict the missing air pollutant concentration, the meteorological parameters are used and their notations and daily statistical evaluation during 2002-2003 shown in Table 1.

Table 1. The minimum, mean and maximum values of meteorological model parameters during 2002 and 2003 years.

Parameters	Notations	Units	Minimum		Mean		Maximum	
			F-AQMS	G-AQMS	F-AQMS	G-AQMS	F-AQMS	G-AQMS
Temperature	T	^oC	-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	1.5	1.7	31.8	61.9

F-AQMS: Florya Air Quality Measurement Stations, G-AQMS: Göztepe Air Quality Measurement Stations,

2.2. Cellular Neural Network (CNN)

Most neural Networks fall into two main classes: (1) Memoryless Neural Networks and (2) Dynamical Neural Networks. As in Hopfield Networks and Cellular Neural Network (CNN), Dynamical Neural Networks have usually been designed as dynamical systems where the inputs are set to some 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 processes signals in real time (Chua and Yang, 1988). Like cellular automata, the CNN is made of a massive aggregate of regularly spaced identical circuits, called cells, which communicate with each other directly only through their nearest neighbors (Figure 2 and 3).

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-time dynamics of CNN. An example of a two-dimensional (2-D) CNN is shown in Fig. 2. Now let us define the neighborhood of *C(i,j)*.

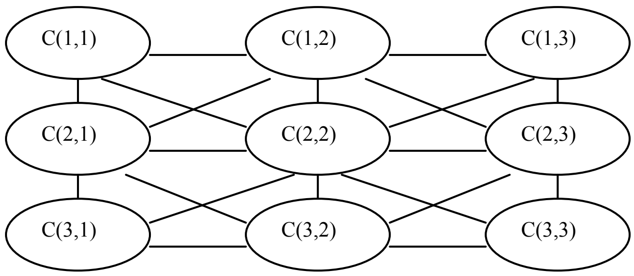

Figure 2. A 2-D cellular neural network. The circuits size 3x3. The link between cells (ellipse) indicates interactions between the linked cells.

We consider a cellular neural network, which consist of *M* line and *N* column (*MxN*). In this structure i^{th} line and j^{th} column are named (*i,j*) cell and explained as C*(i,j)*. Figure 3 shows the neighborhoods of the *C(i,j)* cell (located at the center and shaded) for neighborhood of first second and third (r=1,2,3). In addition, the neighborhood has the property of symmetry (if C*(k,l)* $\in N_r(i,j,$ C*(i,j)* $\in N_r(k,l)$). The r-neighborhood of a cell *C(i,j)* in a cellular neural network is defined by:

$$N_r(i,j) = \{C(k,l) / max (/i-k /, /j-l / \leq r, \qquad 1 \leq i \leq M ; 1 \leq j \leq N\} \qquad (1)$$

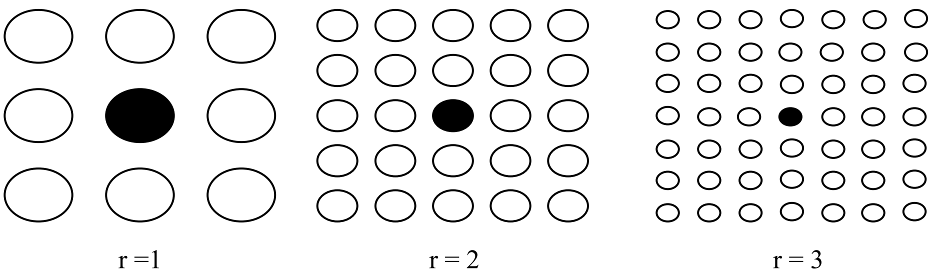

Figure 3. The neighborhood of cell *C(i,j)* for r = 1, r = 2 and r = 3, respectively.

Cells are multiple-input single-output nonlinear processors described by one, or one among several different, parametric functional. A cell is characterized by a state variable, which is generally not observable as such outside the cell itself. It contains linear and nonlinear circuit elements, such as linear resistors, capacitors and nonlinear controlled sources (Fig. 4). Every cell is connected to other cells within its neighborhood. In this scheme, information is only exchanged between neighboring neurons and this local information characteristic does not prevent the capability of obtaining global processing. The CNN is a dynamical system operating in continuous or discrete time.

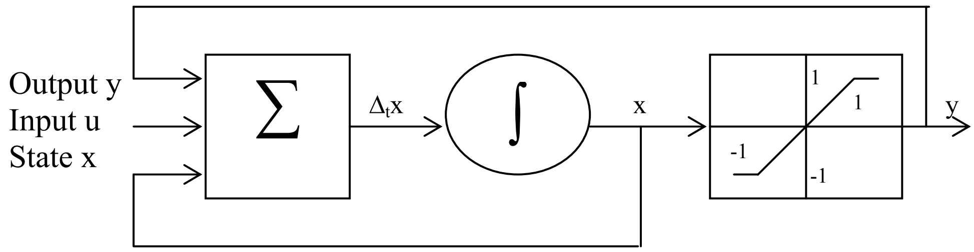

Figure 4. Functional block diagram of a CNN cell.

Cells can be characterized by a functional block diagram that is typical of neural network theory: Fig. 4 depicts a two-stage functional block diagram of a cell, composed of a generalized weighted sum (in general nonlinear with memory) integration, output nonlinear function / functional (Cimagalli, 1993; Albora et al., 2001). Data can be fed to the CNN through two different ports: initial conditions of the state and input *u*. Bias value *I* may be used as a third port.

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

$$\frac{dx_{ij}(t)}{dt} = -x_{ij}(t) + \sum_{C(k,l)\in N_r(\mathrm{i,j})} A(i,\ j;k,l) y_{\mathrm{kl}}(t) + \sum_{C(k,l)\in N_r(\mathrm{i,j})} B(i,\ j;k,l)\ u_{\mathrm{kl}}(t) + I \tag{2}$$

$$y_{ij}(t) = f\left[x_{ij}(t)\right] = \frac{1}{2}\left(\left|x_{ij}(t)+1\right| - \left|x_{ij}(t)-1\right|\right) \tag{3}$$

$$A = \begin{bmatrix} a_{-1,-1} & a_{-1,0} & a_{-1,1} \\ a_{0,-1} & a_{0,0} & a_{0,1} \\ a_{1,-1} & a_{1,0} & a_{1,1} \end{bmatrix},\ B = \begin{bmatrix} b_{-1,-1} & b_{-1,0} & b_{-1,1} \\ b_{0,-1} & b_{0,0} & b_{0,1} \\ b_{1,-1} & b_{1,0} & b_{1,1} \end{bmatrix}, I \tag{4}$$

where; *x*, *y*, *u*, *I* denote respectively cell state, output, input, bias and *j* and *k* are cell indices. CNN parameter values are assumed to be space-invariant and nonlinear function is chosen as piece-wise linear (Fig. 4). *A*, *B* and *I*, the cloning matrices, are identically repeated in the neighborhood of every neuron.

The network behavior of a CNN depends on the initial state of the cells, namely the bias *I*, and the weights values of *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 functionality are defined by a small number of parameters (*A, B, I*) that specify the operation of the component cells as well as the connection weights between them. CNN can also be considered as a nonlinear convolution with the template. Since their introduction in 1988 by Chua, the CNN has attracted a lot of attention. Not only from a theoretical point of view these systems have a number of attractive properties, but also furthermore, there are many well-known applications like image processing, motion detection, pattern recognition, simulation. Albora et al., 2001 applied this contemporary approach for the separation of regional and residual magnetic anomalies, on synthetic and real data. Here, we have predicted air pollution parameters using CNN approach. To evaluate the prediction results of CNN, statistical performance indices have calculated and shown in Section 3.

2.3. Statistical Performance Indices

In this study, in order to objectively evaluate model prediction, three statistical performance indices are computed: the mean bias error (*Bias*), correlation coefficient (*R*), and the index of agreement (*d*). These are based on the deviations between predicted and original observation values. *Bias* is the degree of correspondence between the mean prediction and the mean observation. Lower numbers of *Bias* are the best and values of b*ias* <0 indicate under-forecasting. Evaluation can also be undertaken by considering measures of agreement, such as the Pearson product moment correlation coefficient (*R*) values. The index of agreement, bounded, relative measure that is capable to measure the degree of which predictions are error-free. The denominator accounts for the model's deviation from the mean of the observations as well as to the observations deviation from their mean. In a good model *d* and *R* should approach to one (Kukkonen et al., 2003). All these indices are formulated as follows;

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(O_i - P_i) \qquad (5)$$

$$r = \sqrt{1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}} \qquad (6)$$

$$d = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(\left|P_i - \overline{O}\right| + \left|O_i - \overline{O}\right|)^2} \qquad (7)$$

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

3. RESULTS AND DISCUSSION

In this paper, missing PM_{10} and SO_2 concentration values are predicted using Cellular Neural Networks in Istanbul–Yenibosna and Istanbul-Ümraniye air pollution station. We have estimated two data sets for SO_2 and PM_{10}. One of them, data set of PM_{10} and SO_2 is formed missing data percentage of 50, another them; it is formed missing data percentage of 20. %50 and %20 of daily mean observed SO_2 and PM_{10} concentrations are changed average concentration value of all SO_2 and PM_{10} data, respectively. These missing data is predicted using CNN model.

We have calculated correlations between meteorological and pollution parameters by statistical package program SPSS11.5. In our CNN model, the elements of input (*u*) and target (*T*) matrices are shown in Fig.5. The elements of input matrix consist of 20% and 50% missed data of SO_2 and PM_{10} to be predicted. We arrange *u* and *T* matrices' elements regarding to correlation coefficient information to improve prediction performance. After training CNN using *u* and *T* matrices, we have obtained A, B and I templates for each study as shown in equation 8-15.

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

Figure 5. Input (*u*) and target (*T*) matrices of our CNN model.(AP: Air Pollutant, SO_2 and PM_{10}; MAP:Air Polutant with missing data).

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

$$A=\begin{bmatrix} 0.0426 & 0.0545 & 0.0521 \\ 0.0644 & 1.2885 & 0.0644 \\ 0.0521 & 0.0545 & 0.0426 \end{bmatrix} \quad B=\begin{bmatrix} -0.0615 & -0.0618 & -0.0613 \\ -0.0616 & -0.0614 & -0.0616 \\ -0.0613 & -0.0618 & -0.0615 \end{bmatrix} \quad I=\left[-0.0614\right] \tag{8}$$

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

$$A=\begin{bmatrix}-0.0054 & 0.0075 & -0.0280\\ 0.0048 & 1.005 & 0.0048\\ -0.0280 & 0.0075 & -0.0054\end{bmatrix}\quad B=\begin{bmatrix}0.0065 & 0.0068 & 0.0069\\ 0.0068 & 0.0069 & 0.0068\\ 0.0069 & 0.0068 & 0.0065\end{bmatrix}\quad I=[0.0069]$$
(9)

To predict 50 % missing SO_2 concentration in Yenibosna:

$$A=\begin{bmatrix}0.0537 & 0.0732 & 0.0636\\ 0.0592 & 1.4524 & 0.0592\\ 0.0636 & 0.0732 & 0.0537\end{bmatrix}\quad B=\begin{bmatrix}-0.0731 & -0.0734 & -0.0733\\ -0.0734 & -0.0734 & -0.0734\\ -0.0733 & -0.0734 & -0.0731\end{bmatrix}\quad I=[-0.0734]$$
(10)

To predict 20 % missing SO_2 concentration in Yenibosna:

$$A=\begin{bmatrix}0.0619 & 0.0697 & 0.0674\\ 0.0570 & 1.4848 & 0.0570\\ 0.0674 & 0.0697 & 0.0619\end{bmatrix}\quad B=\begin{bmatrix}-0.0752 & -0.0755 & -0.0754\\ -0.0754 & -0.0754 & -0.0754\\ -0.0754 & -0.0755 & -0.0752\end{bmatrix}\quad I=[-0.0754]$$
(11)

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

$$A=\begin{bmatrix}-0.0024 & 0.0400 & 0.0308\\ 0.0541 & 1.0350 & 0.0541\\ 0.0308 & 0.0400 & -0.0024\end{bmatrix}\quad B=\begin{bmatrix}-0.0238 & -0.0237 & -0.0240\\ -0.0238 & -0.0239 & -0.0238\\ -0.0240 & -0.0237 & -0.0238\end{bmatrix}\quad I=[-0.0239]$$
(12)

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

$$A=\begin{bmatrix}-0.0054 & 0.0097 & -0.0260\\ 0.0038 & 1.005 & 0.0038\\ -0.0260 & 0.0097 & -0.0054\end{bmatrix}\quad B=\begin{bmatrix}0.0090 & 0.0089 & 0.0089\\ 0.0088 & 0.0089 & 0.0088\\ 0.0089 & 0.0088 & 0.0090\end{bmatrix}\quad I=[0.0089]$$
(13)

To predict 50 % missing SO_2 concentration in Ümraniye:

$$A=\begin{bmatrix}0.0142 & -0.0182 & 0.0277\\ 0.0157 & 1.1553 & 0.0157\\ 0.0277 & -0.0182 & 0.0142\end{bmatrix}\quad B=\begin{bmatrix}-0.0271 & -0.0273 & -0.0271\\ -0.0272 & -0.0273 & -0.0272\\ -0.0271 & -0.0273 & -0.0271\end{bmatrix}\quad I=[-0.0273]$$
(14)

To predict 20 % missing SO_2 concentration in Ümraniye:

$$A=\begin{bmatrix} 0.0402 & 0.0131 & 0.0289 \\ 0.0276 & 1.2505 & 0.0276 \\ 0.0289 & 0.0131 & 0.0402 \end{bmatrix} \quad B=\begin{bmatrix} -0.0429 & -0.0433 & -0.0433 \\ -0.0433 & -0.0433 & -0.0433 \\ -0.0433 & -0.0433 & -0.0429 \end{bmatrix} \quad I=\left[-0.0433\right]$$
(15)

Here, neighborhood (r) is chosen as 1. To guarantee stability of CNN, templates are symmetric. We have replaced the template values obtained in Equation (8-15) to Equations (2-3). Then we have compared CNN predicted and actual {PM_{10} and SO_2} concentrations as in Figure 6 and 7 during 2002 and 2003.

Model results have also been checked by calculation five different statistical indices, given in Equation 5-7, which are based on the deviations between predicted values and original observations. The final results of statistical model evaluation for SO_2 and PM_{10} in Yenibosna and Ümraniye during 2002 and 2003 years have been presented in Table 3. For both pollutants, the results have been presented separately for each Air Quality Measurement Stations and each percentage of missing data. For PM_{10} in Yenibosna, index of agreement (d) of CNN is 0.69 and 0.92 for missing data percentage of 50 and 20, respectively. We have measured similar results for PM_{10} in Ümraniye, d is 0.74 and 0.91 for missing data percentage of 50 and 20, respectively. For SO_2 in Yenibosna and Ümraniye, d and r are the same value, 0.85 and 0.73, respectively, for missing data percentage of 20. For PM_{10} in Yenibosna and Ümraniye, *Bias* value is positive and for SO_2, it is mostly negative. This result has demonstrated that the prediction concentration of PM_{10} is less than observed PM_{10} concentration. However, the prediction concentration of SO_2 is high than observed SO_2 concentration. Table 2 and Figure 6,7 has demonstrated that CNN model for PM_{10} are producing considerably better predictions than SO_2. Furthermore, CNN model for missing data percentage of 20 are producing better predictions than missing data percentage of 50.

Table 2: Model performance indices for the CNN model. The results differ by the missing data percentage, Yenibosna and Ümraniye air quality stations and PM_{10} and SO_2 pollutions.

AQMS	AP	MDP (%)	Statistical performance indices						
			Max.	Min.	Avrg.	σ	r	d	Bias
Yenibos na	P M$_1$	50	218	13	47	36,1	0,57	0,70	16,3
		20	211	15	59	23	0,87	0,92	4,2
	S O$_2$	50	170	0,3	26	26,1	0,67	0,81	2,76
		20	148	2,2	30	25,5	0,73	0,85	-0,91
Ümraniye	PM_{10}	50	279	10	49	34,3	0,54	0,74	6,87
		20	244	2,5	54	24,3	0,86	0,91	1,95
	SO_2	50	164	5,5	21	30,3	0,67	0,73	-2,43
		20	170	6	21	24,3	0,73	0,85	-2,54

AQMS: Air Quality Measurement Stations, AP: Air Pollutans, MDP: Missing data percentage

4. CONCLUSION

In this paper, main air pollutants of Istanbul, Particulate Matter (PM), Sulfur Dioxide (SO_2) are estimated using CNN approach. There are many studies for air pollutant modeling. One of the frequently used methods is Artificial Neural Network (ANN). In ANN, the training process time increases as the problem becomes complex. In 1988, Chua and Yang have introduced Cellular Neural Network (CNN) as a new non-linear, dynamic neural network structure. In CNN, the correlations between neighbor pixels are modeled by cloning templates with limited number of elements in solving complex problems.

In previous similar study, index of agreement value has changed between 0.20 and 0.80 for other model techniques, for example Multilayer Neural Network, Regression Analysis etc. (Junninen et al., 2004). In this paper, we have measurement *d* between 0.69 and 0.92 for all study and for all pollutant. The elements of climatic system are commonly nonlinear, irregular and highly complex. As a result, CNN-based approaches can be considered as a compromising approach in air pollutant prediction.

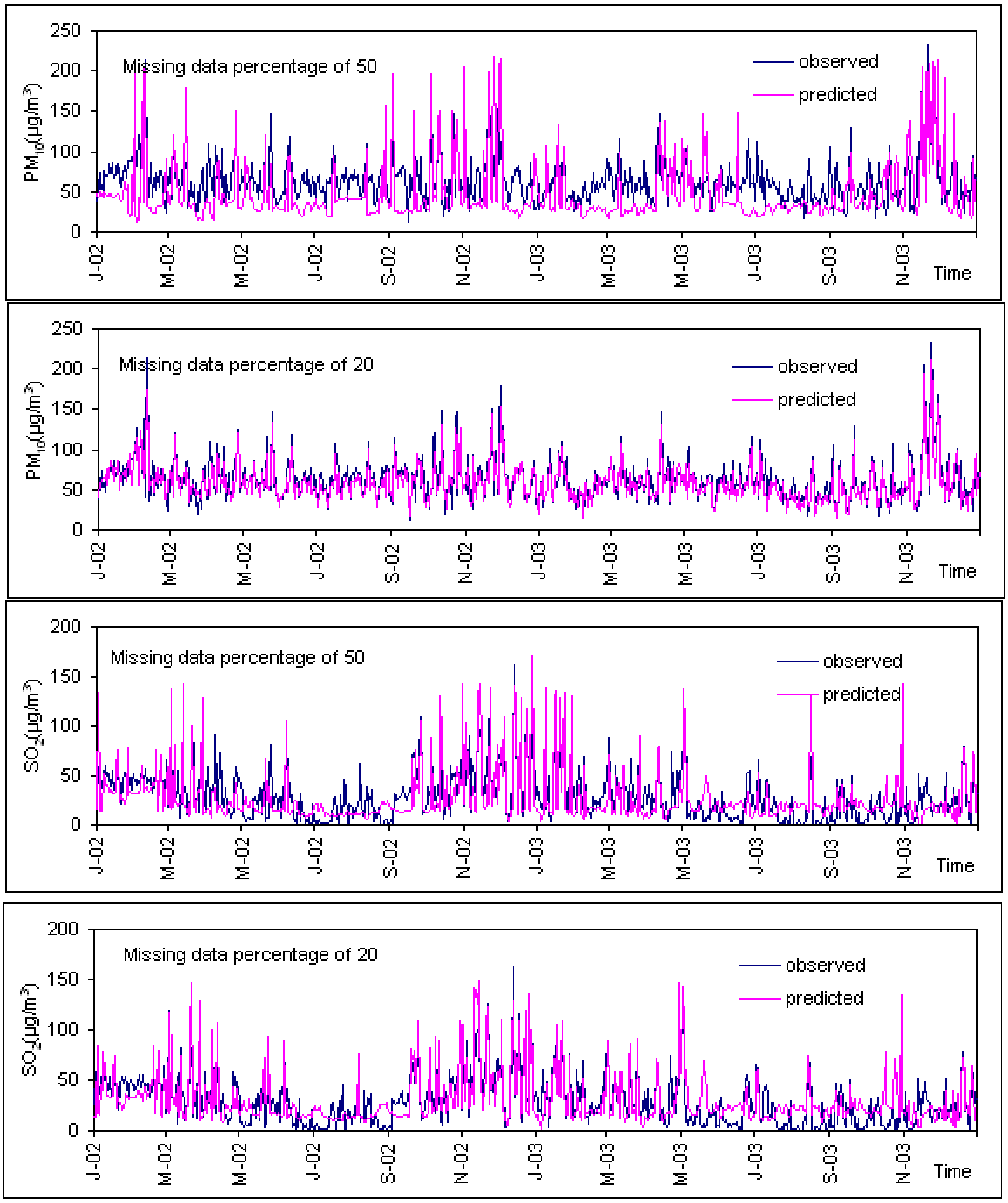

Figure 6: Two years of observed and CNN model predicted daily mean PM_{10} and SO_2 concentrations at the Yenibosna AQM Stations.

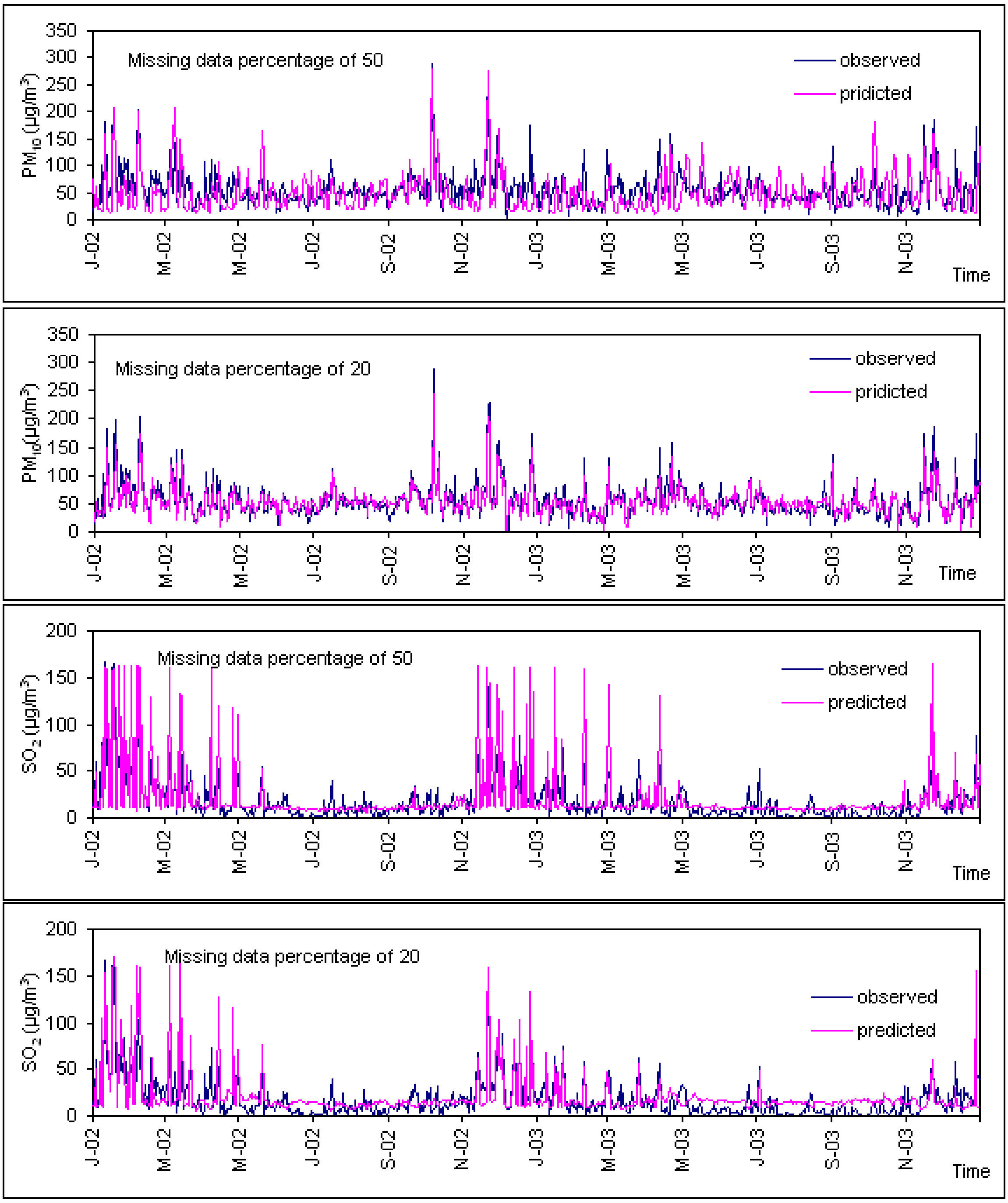

Figure 7: Two years of observed and CNN model predicted daily mean PM_{10} and SO_2 concentrations at the Ümraniye AQM Stations.

5. ACKNOWLEDGEMENTS

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

REFERENCES

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.
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.
Dirks, K.N., Johns, M.D., Hay, J.E., Sturman, A.P., (2002) A simple sem-empirical model for predicting missing carbon monoxide concentrations, Atmospheric Environment 36, 5953-5959.
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.
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.
Grassi, G., Grieco L.A., (2002), Object-Oriented Image Analysis via Analogic CNN Algorithms-Part I: Motion Estimation, 7^{th} IEE International Workshop on Cellular Neural Networks and Their Applications, 172-180.
Goncaloglu, B.I., (2000), Emission inventory of air pollution in Istanbul, Ph.D. Thesis, Yıldız Technical University, Istanbul, Turkey.
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.
Kukkonen, J., Partanen, L., Karppinen, A., et al., (2003), Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modeling system and measurements in central Helsinki, Atmospheric Environment 37, 4539-4550.

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.
Perez P., Trier A., Reyes J., (2000), Prediction of PM2.5 concentration several hours in advance using neural networks in Santiago, Chile, Atmospheric Environmental 34, 1189-1196.
Saral A., (1999), Air Pollution Modeling and Prediction with Artificial Neural Network, Ph.D. Thesis, Yıldız Technical University, Istanbul, Turkey.

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.

Uçan, 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, 223-231, France.