

# Air Pollutants Tracing Model using Perceptron Neural Network and Non-negative Least Square

Suk-Hyun Yu<sup>†</sup>, Hee-Yong Kwon<sup>\*\*</sup>

## ABSTRACT

In this paper, air pollutant tracing models using perceptron neural network(PNN) and non-negative least square(NNLS) are proposed. When the measured values of the air pollution and the contribution concentration of each source by chemical transport modeling are given, they estimate and trace the amount of the air pollutants emission from each source. Two kinds of emissions data are used in the experiments : CH<sub>4</sub> and N<sub>2</sub>O of Geumgo-dong landfill greenhouse gas, and PM<sub>10</sub> of 17 areas in Northeast Asia and eight regions of the Korean Peninsula. Emission values were calculated using pseudo inverse method, PNN and NNLS. Pseudo inverse method could be used for the model, but it may have negative emission values. In order to deal with the problem, we used the PNN and NNLS methods. As a result, the estimation using the NNLS is closer to the measured values than that using PNN. The proposed tracing models have better utilization and generalization than those of conventional pseudo inverse model. It could be used more efficiently for air quality management and air pollution reduction.

**Key words:** Air Pollution Tracing Model, Perceptron, Non-negative Least Square

## 1. INTRODUCTION

After the Industrial Revolution, air pollution has been increasing due to rapid industrialization and urbanization. Air pollutants can cause physical damage to the human body directly and/or indirectly and it can lead to social and economic loss.

Air pollutants can be divided into primary pollutants such as sulfur dioxide(SO<sub>2</sub>), carbon monoxide(CO), and lead(Pb), and secondary pollutants such as nitrogen dioxide(NO<sub>2</sub>), particulate matter(PM<sub>10</sub>), and ozone(O<sub>3</sub>). The contamination of primary pollutants have been reduced by the air

pollution regulatory policy intensively promoted since the 1990s. The contamination of secondary pollutants, however, have been increased due to automobiles, industrial activities, and the construction sites[1]. Among them, greenhouse gases causing a large number of complaints and PM<sub>10</sub> being harmful to the human body should be managed with particular emphasis.

The greenhouse gases could be generated by energy industries, industrial processes, agriculture, waste, land use, and forestry. In order to reduce the gases, IPCC (Intergovernmental Panel on Climate Change) has presented the guideline for greenhouse gas inventories[2,3].

In Korea, in order to survey a landfill, not only FOD (First Order Decay) method presented by IPCC, but also greenhouse gases direct measuring method using flux chamber to derive the value of its reaction constant, k has been used. It, however, has a defect in that measuring time is limited, emissions from landfill surface are not uniform, and the cost is high. Recently, mainly in the developed countries, not only FOD method

---

\* Corresponding Author : Hee-Yong Kwon, Address : 310, Arikwan, Anyang Univ., Anyang 5-dong, Manangu, Anyang-si, Gyeonggi-do, 430-714, Korea, TEL : +82-031-467-0878, FAX : +82-031-463-1249, E-mail : hykwon@anyang.ac.kr

Receipt date : Oct. 24, 2013, Revision date : Nov. 8, 2013  
Approval date : Nov. 15, 2013

<sup>†</sup> Dept. of Information & Communications Engineering, Anyang Univ.

(E-mail: shy@anyang.ac.kr)

<sup>\*\*</sup> Dept. of Computer Science Engineering, Anyang Univ.

but also new estimating techniques using micro-climate data around landfills, methane concentration and the application of modeling techniques have been studied [4].

PM is a complex mixture of solid and liquid particles that vary in size and composition, and remain suspended in the air. PM10 is defined as particulate matter with a diameter less than 10  $\mu$  m. Many health effect studies have shown an association between exposure to PM10 and increase in daily mortality and symptoms of certain illnesses such as asthma, chronic bronchitis, decreased lung function, and premature death[5]. The social damage cost of the metropolitan area due to particulate matter is estimated to be worth 10 trillion won per year [6].

It is important to estimate the exact emissions in order to reduce the damage of all of these pollutants. In Korea, the city of Seoul constructs a national and international emissions database with the research results of the National Institute of Environmental Research[7,8], Moon et al. calculate yellow dust emissions associated with the 5 expression of experience and the ADAMS model equations of National Weather Service respectively, and calculate and check the PM10 concentrations using WRF\_CMAQ.

Henze et al. at the University of Columbia evaluated PM2.5 emissions in the United States using the GOES-Chem Adjoint model[9], Jacob, Harvard University's Professor, evaluated the global CO emissions using MOPITT satellite data[10]. Domestic PM10 is largely affected by a long-range trans-boundary movement in the Northeast Asia region. Thus, for the fine metropolitan air quality management, it requires a complex analysis for the long-distance movement and local sources. For this, it is essential to obtain accurate domestic and Northeast Asia emissions data.

In this paper, we have collected the INTEX-B emission inventory data of greenhouse gases in

Geumgo-dong landfill and East Asia region. Then, we propose an air pollutant sources tracing model to estimate greenhouse gases, CH<sub>4</sub>, N<sub>2</sub>O and PM10. The purpose of research is the estimation of emissions of each source with greenhouse gas data in domestic and East Asia regions. When measured emissions and contribution concentrations of each source by a chemical transport modeling are given, it is an optimization problem finding emission values of each source to minimize the error between the measured values and the estimated ones.

In this paper, three methods were studied. First, emissions of each source was estimated using the pseudo inverse method. There was a small error between the measured emissions and the estimated ones, but the negative estimate of the emissions occurred, so it should be improved. Second, we use PNN which shows excellent learning capability, given a target value. PNN learning may have negative emissions, so we improve the network adding the learning conditions to limit the negative weights. Then we resolve the pseudo inverse method's defect of negative emission problem. Third, in order to reduce the error between measured emission values and the ones calculated by the proposed PNN, the NNLS method was applied. The results calculated by the NNLS are closer to the measured emissions than those of the PNN.

In section 2, we describe the proposed models. Experimental results are shown in section 3, and concluded in section 4.

## 2. PROPOSED TRACING MODEL

### 2.1 The definition of air pollutant tracing model

Given measured emissions and contribution concentrations of each source by a chemical transport modeling, the model should estimate emissions of each source. It is an optimization

problem finding emission values,  $Q$  of each source to minimize the error between the measured values and the estimated ones, can be expressed as (1).

$$C = F(x) \times Q \tag{1}$$

( $C$ : measured emissions,  $F(x)$ : contribution concentrations by chemical transport modeling,  $Q$ : emissions of each source)

Also, each term in (1) can be defined as follows.

$$C = [c_1 c_2 \dots c_i \dots c_n]' \tag{2}$$

( $c_i$ :  $i$ -th measured value,  $i$ : measuring site,  $n$ : site number)

$$F(x) = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1m} \\ f_{21} & f_{22} & \dots & f_{2m} \\ \dots & \dots & f_{ij} & \dots \\ f_{n1} & f_{n2} & \dots & f_{nm} \end{bmatrix} \tag{3}$$

( $f_{ij}$ : contribution concentration of  $i$ -th measuring value of  $j$ -th site,  $n$ : number of measurements,  $m$ : number of sources)

$$Q = [q_1 q_2 \dots q_i \dots q_m]' \tag{4}$$

( $q_i$ : emission value of  $i$ -th source,  $m$ : number of sources)

### 2.2 Emission estimation using Pseudo Inverse Method

Emission estimation problem can be regarded as estimating  $Q$  to fit  $F(x)$  with measured values  $C$ , which are described in (1). It can be solved as follows (5), and is called pseudo inverse method[12].

$$C = FQ \tag{5}$$

$$F^T C = F^T F Q$$

$$(F^T F)^{-1} F^T C = (F^T F)^{-1} (F^T F) Q$$

$$Q = (F^T F)^{-1} F^T C$$

However, there exist negative emissions when the pseudo inverse method is applied with the problem. As the negative values are not compat-

ible with the physical world, it should be compensated with non-negative values. Thus we introduce PNN and NNLS to overcome the limitation of pseudo inversion.

### 2.3 Emission estimation using Perceptron Neural Network (PNN)

The human brain performs calculation to react to an external stimulus of environmental factors by the physical interconnection of the cells called neurons. Neural Network model is an intelligent system which mimics the human brain's neurons[11].

Neural network for the emission estimation problem can be designed with single-layer perceptron network model, which consists of input layer and output layer of neurons. Contribution concentrations,  $F(x)$  are the input for the input layer, measured values  $C$  are presented as the output, target for the output layer. With these neural networks, the weights connecting input layer with output are learned, which ultimately is the emissions of each source. Figure 1 shows the proposed neural network model. In the figure, an output pattern is calculated by summation of the results of multiplication weights( $w$ ) with contribution concentration input( $F(x)$ ). The difference between the output and the measured values( $C$ ) is the error for the pattern. PNN regulates

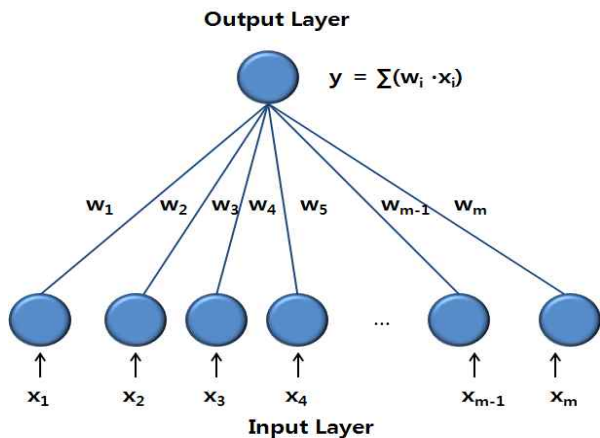


Fig. 1. Single Layer Perceptron.

the weights ( $w$ ) with the error and learns the emission values of each source. It is important to include a condition that when a weight has negative value, it should be limited with non-negative value. Because the weights are emission values and there is no negative emission in the real world, it can not be negative. Neural network learning with such condition can improve the negative emission calculated by pseudo inverse method. Table 1 shows the parameters of the proposed PNN.

Table 1. Neural Network Parameter

Data	Number of Layer	Learning Rate	Loop
CH <sub>4</sub> , N <sub>2</sub> O	1	0.01	25,000,000
PM <sub>10</sub>	1	0.1	50,000,000

#### 2.4 Emission estimation using Non-negative Least Square (NNLS)

Using PNN, the negative emission of pseudo inverse method could be resolved. However, NNLS with constraints of (6) is more efficient than PNN for the problem. In (6),  $f$  denotes measured emissions,  $E$  denotes contribution concentrations obtained by the chemical transport modeling, and  $x$  estimated emissions which should be calculated. Namely, NNLS estimates  $x$  that approximates an  $E$  to  $f$ , so it is suitable for the problem. Also, it implements the rule that negative weights should not be allowed like PNN [13-21].

$$\text{Minimize } \|Ex - f\| \quad \text{subject to } x \geq 0 \quad (6)$$

### 3. EXPERIMENTS AND RESULTS

#### 3.1 Experimental environment

There are two experimental data : one is domestic data for Geumgo-dong landfill greenhouse gas, and the other is global data for PM10 of 17 areas in Northeast Asia and 8 regions of the Korean Peninsula. Greenhouse gas measurements

were performed on the bank which separates the landfill into two areas, active and inactive, at intervals of 10m. 5 data sets are collected. The emission gases to be estimated are CH<sub>4</sub> and N<sub>2</sub>O.

Second experiments are performed for the global PM10 which is affected by long-distance travel over the Northeast Asia territories. For more accurate calculations, we use INTEX-B emission inventory. INTEX-B is an emissions data to the region since 2001, reflecting the rapid growth in Asia to complement the TRACE-P emissions in 2000. Target year is 2006, the  $0.5^\circ \times 0.5^\circ$  spatial resolution of emissions data of which target is anthropogenic sources, the target materials are SO<sub>2</sub>, NO<sub>x</sub>, CO, NMVOC, OC, BC, PM10, and PM2.5. Source categories are Power, Industry, Residential, and Transportation. There are included PM10 and PM2.5, but excluded NH<sub>3</sub> and biological combustion emissions from the calculation compared with the TRACE-P emissions.

In this paper, we use the PM10 emissions of 17 regions and countries of the Northeast Asia and 8 domestic areas in INTEX-B[22].

#### 3.2 Experimental results

Three experiments are performed : pseudo inverse method, perceptron neural network(PNN), non-negative least square(NNLS).

As a result, pseudo inverse method has the smallest error between measured emissions and the model's output, but it has a negative emission problem. To deal with the problem, we use PNN and NNLS. Eventually, NNLS has smaller error than PNN, and is the most appropriate method among the three.

Figure 2 and Figure 3 are the estimating results of the proposed three methods for the greenhouse gases, CH<sub>4</sub> and N<sub>2</sub>O. Local case 1 ~ 5 are the 5 measurement sets in September 2010 of Daejeon Geumgo-dong landfill. Figure 4 is the estimating results of the proposed three methods for the PM10 emissions of 17 regions and coun-

tries of the Northeast Asia and 8 domestic areas. Global case 1 ~ 4 are the 4 PM10 emission data in January 2008, May, August, and October, one for each season. In the graph, x-axis denotes measurement site, y-axis denotes emission. Legends are measured values, pseudo inverse method, NNLS, and PNN respectively.

In addition, Table 2 and Table 3 shows the RMSE of the proposed three methods' estimation.

#### 4. CONCLUSION

In this paper, we propose air pollutant tracing models by minimizing the error between the measured values and the estimated ones, when measured emissions and contribution concentrations of each source by a chemical transport modeling are given. The study is conducted with pseudo inverse method, PNN, and NNLS to estimate the emissions. Pseudo inverse method estimates emissions closest to the measured ones, but there exist negative emissions. We introduce PNN and NNLS to overcome the problem. As a result, both two method resolve the negative emission problem but NNLS'error is smaller than PNN. So NNLS is the best model for emission tracing.

The proposed air pollutant tracing model does not need expensive equipment or personnel commitments for modeling, so it is efficient in terms of cost and time. However, it has a defect that the minimum or maximum value of the measured emissions may fail to be found. Thus, in order to be an effective model, not only reducing the error between measured emissions and estimated ones but also finding the trends in the emissions is eventually required.

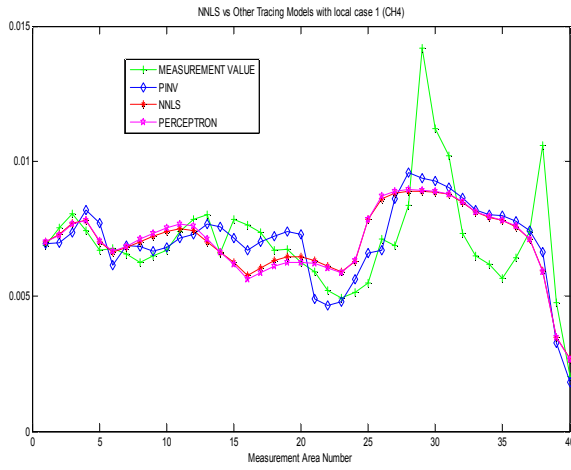
Table 2. RMSE of Three Methods (CH<sub>4</sub>, N<sub>2</sub>O)

$\times 10^{-3}$

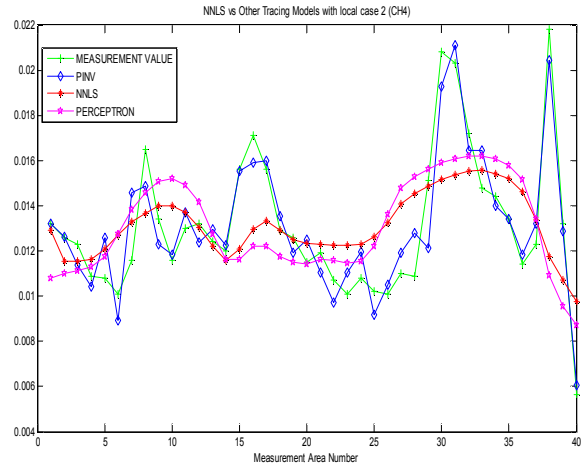
Case	Pollutant	Pseudo Inverse	Perceptron	NNLS
case1	CH <sub>4</sub>	0.219	0.251	0.248
	N <sub>2</sub> O	0.006	0.016	0.012
case2	CH <sub>4</sub>	0.179	0.477	0.441
	N <sub>2</sub> O	0.002	0.038	0.003
case3	CH <sub>4</sub>	0.115	0.269	0.256
	N <sub>2</sub> O	0.006	0.021	0.016
case 4	CH <sub>4</sub>	0.087	0.129	0.125
	N <sub>2</sub> O	0.007	0.016	0.012
case 5	CH <sub>4</sub>	0.125	0.170	0.169
	N <sub>2</sub> O	0.072	0.097	0.088

Table 3. RMSE of Three Methods (PM10)

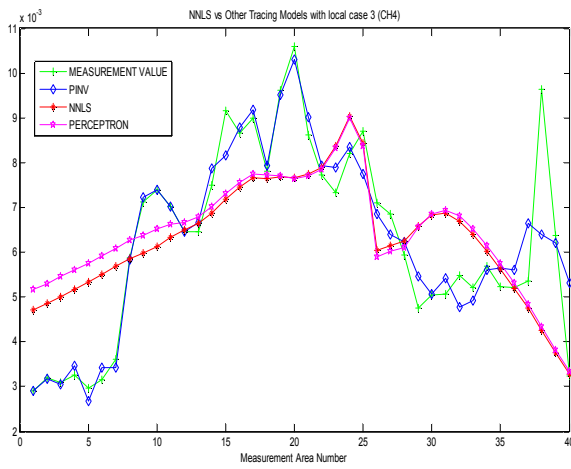
Case	Pseudo Inverse	Perceptron	NNLS
case1	4.15	5.28	4.17
case2	2.65	3.79	2.84
case3	4.77	8.05	5.72
case4	4.49	6.18	6.12



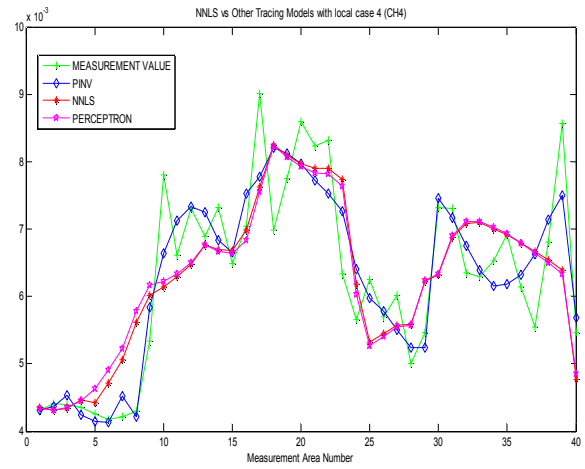
(a) local case 1



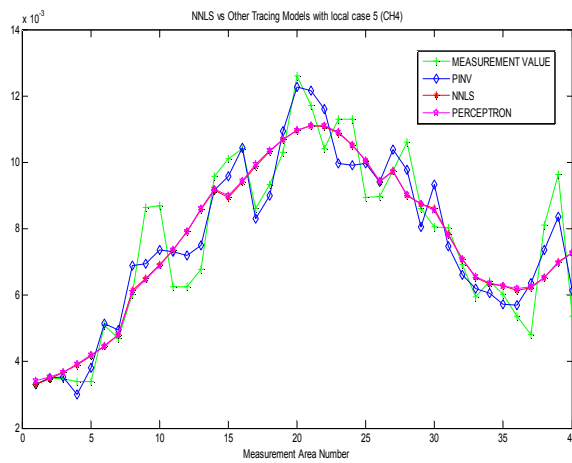
(b) local case 2



(c) local case 3

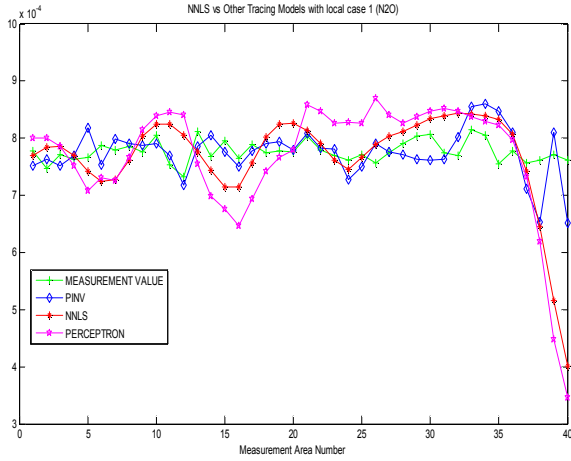


(d) local case 4

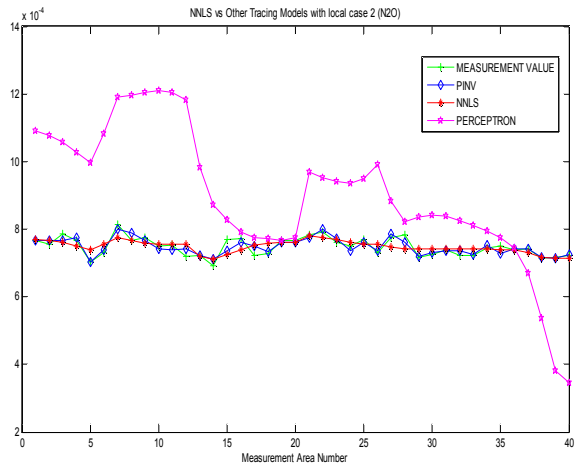


(e) local case 5

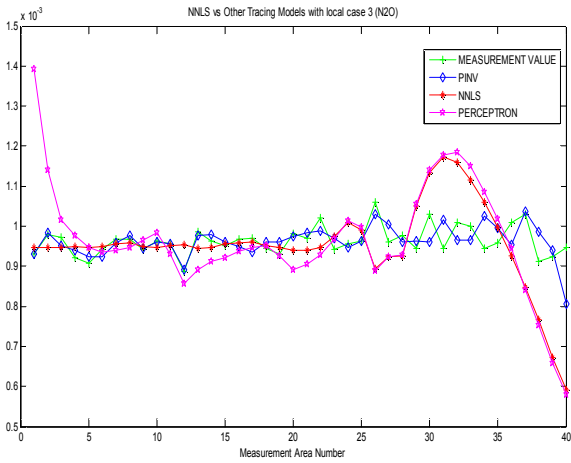
Fig. 2. NNLS vs. Other Methods with local data(CH<sub>4</sub>).



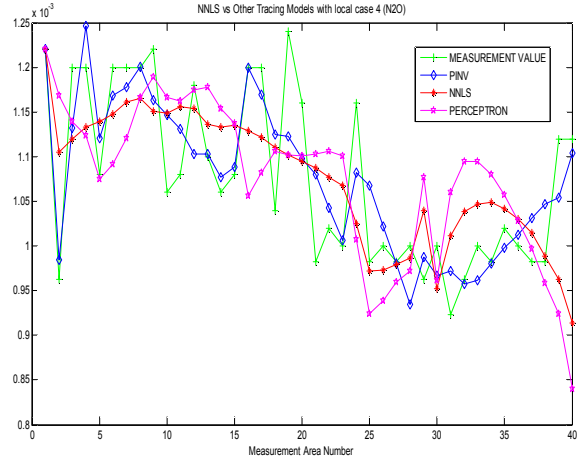
(a) local case 1



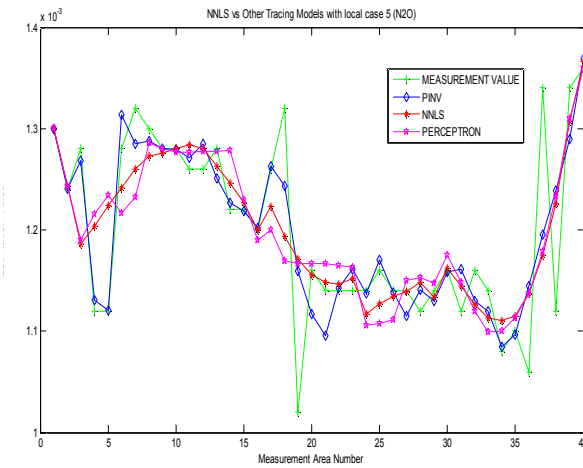
(b) local case 2



(c) local case 3



(d) local case 4



(e) local case 5

Fig. 3. NNLS vs. Other Methods with global data(N<sub>2</sub>O)

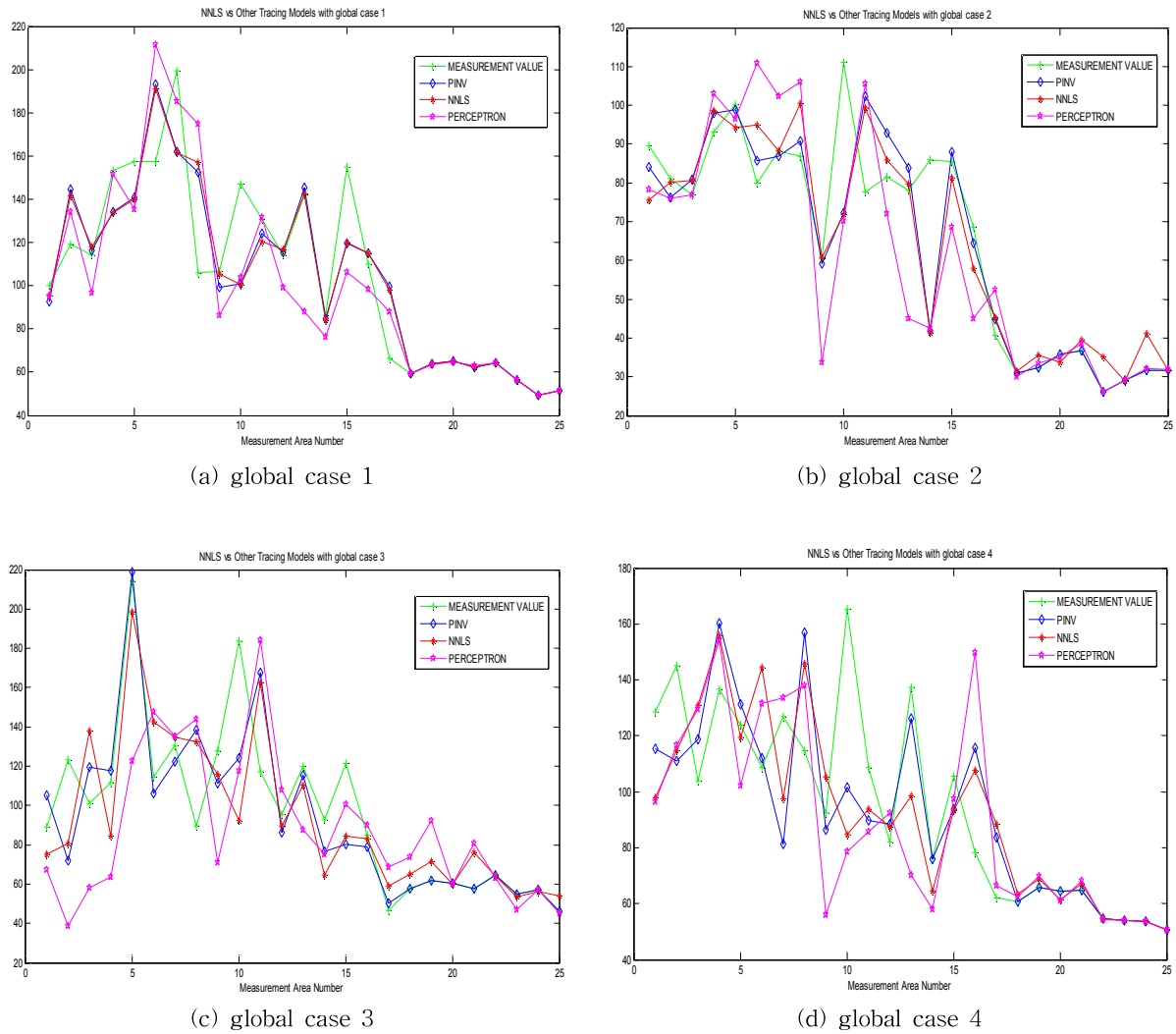


Fig. 4. NNLS vs. Other Methods with global data(PM10).

REFERENCE

[ 1 ] Yong-Pyo Kim, "Air Pollution in Seoul Caused by Aerosols," *Journal of Korean Society for Atmospheric Environment*, Vol. 22, No. 5, pp. 535-553, 2006.

[ 2 ] IPCC, *1996 IPCC Guidelines for National Greenhouse Gas Inventories*, 1996.

[ 3 ] IPCC, *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, 2006.

[ 4 ] Suk-Hyun Yu, *Air Quality Forecasting and Odor Source Tracking Method Using Neural Network and Nonnegative Least Square*, A Doctoral Dissertation of Anyang University, 2011.

[ 5 ] Youn-Seo Koo, Hui-Young Yun, Hee-Yong Kwon, and Suk-Hyun Yu, "A Development of PM10 Forecasting System," *Journal of Korean Society for Atmospheric Environment*, Vol. 26, No. 6, pp. 666-682, 2010.

[ 6 ] Gyeonggi Institute of Health Environment, *Evaluation Report of Air Pollution in Gyeonggi-do*, 2008.

[ 7 ] The Metropolis of Seoul, *Characterization Research of PM10 using Batch Monitoring with Airborne PM10 in Seoul*, 2010.

[ 8 ] Yun-Seob Moon and Seong-Hwan Lee, "Estimation of Hourly Emission Flux of Asian Bust Using Empirical Formulas in the Source Area," *Journal of Korean Society for*



- Atmospheric Environment*, Vol. 25, No. 6, pp. 539–549, 2009.
- [9] D.K. Haneze, J.H. Seinfeld, and D.T. Shindell, “Inverse Modeling and Mapping us Air Quality Influence of Inorganic PM<sub>2.5</sub> Precursor Emissions using the Adjoint of CEOS-Chem,” *Atmos. Chem. Phys.*, Vol. 9, No. 16, pp. 5877–5903, 2009.
- [10] M. Kopacz, D.J. Jacob, D.K. Hanze, and C.L. Heald, “Comparison of Adjoint and Analytical Bayesian Inversion Methods for Constraining Asian Sources of Carbon Monoxide Using Satellite(MOPITT) Measurements of CO Columns,” *Journal of Geophysical Research Atmospheres*, Vol. 114, Issue D4, pp. 1–10, 2007.
- [11] David E. Rumelhart, James L. McClelland, and the PDP Research Group, *Parallel Distributed Processing*, MIT Press Publishers, Cambridge, Massachusetts., 1987.
- [12] Suk-Hyun Yu, Sang-Jin Park, Youn-Seo Koo, and Hee-Yong Kwon, “Odor Classification and Source Analysis using Pseudo Inverse,” *Journal of Korea Multimedia Society*, Vol. 13, No. 8, pp. 1171–1182, 2010.
- [13] C.L Lawson and R.J Hanson, *Solving Least Squares Problems*, Prentice-Hall Publishers, Englewood Cliffs, New Jersey, 1974.
- [14] D. Lee and H. Seung, “Learning the Parts of Objects by Non-negative Matrix Factorization,” *Nature*, Vol. 401, No. 6755, pp. 788–791, 1999.
- [15] R.B. Bapat and T.E.S. Raghavan, *Nonnegative Matrices and Applications*, Cambridge University Press Publishers, Cambridge, UK., 1997.
- [16] A. Berman and R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press Publishers, New York., 1994.
- [17] N. Vo, B. Moran, and S. Challa, “Nonnegative Least Square Classifier for Face Recognition,” *Lecture Notes in Computer Science*, Vol. 5553, pp. 449–456, 2009.
- [18] G. Bombara, M. Cococcioni, B. Lazzarini, and F. Marcelloni, “S-NNLS: An Efficient Non-negative Least Squares Algorithm for Sequential Data,” *International Journal for Numerical Methods in Biomedical Engineering*, Vol. 27, No. 5, pp. 770–773, 2011.
- [19] Wu Jiaying, J.A. van Aardt, and G.P. Asner, “A Comparison of Signal Deconvolution Algorithms Based on Small-Footprint LiDAR Waveform Simulation,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49, No. 6, pp. 2402–2414, 2011.
- [20] Yong-Deok Kim and Seung-Jin Choi, “Non-negative Tucker Decomposition,” *Journal of KIISE: Computing Practices and Letters*, Vol. 14, No. 3, pp. 296–300, 2008.
- [21] Youn-Seo Koo, Hee-Yong Kwon, Eun-Sung Son, Hyun-Jin Jin, Byoung-Won Jung, and Gwi-Wuk Heo, “Review of the Estimation Methodology of Methane Emission in a Landfill using Inverse Modeling Technique,” *Journal of Korean Society of Odor Research and Engineering*, Vol. 12, No. 3, pp. 111–123, 2013.
- [22] Korea Environmental Industry&Technology Institute, *Development of Technology for Improvement of Air Pollutant Amount Inventory and Reliability*, 2012.



Suk-Hyun Yu

1999 B.S. in Computer Science  
Engineering, Anyang  
University

2002 M.S. in Computer Science  
Engineering, Anyang  
University

2011 Ph.D. in Computer Science Engineering, Anyang  
University

2012- Assistant Professor, Department of Information  
& Communications Engineering, Anyang  
University

Interest: Pattern Recognition, Neural Network, Parallel  
Computer Vision, Processing Application



Hee-Yong Kwon

1983 B.S. in Computer  
Engineering, Seoul  
National University

1985 M.S. in Computer  
Engineering, Seoul  
National University

1993 Ph.D. in Computer Engineering, Seoul National  
University

1986-1995 Korea Telecom Research Center Senior  
Researcher

1995- Professor, Department of Computer Science  
Engineering, Anyang University

Interest: Pattern Recognition, Neural Network, Parallel  
Processing Application