Analyzing Vehicular Pollutant Concentration Using Neural Network

by Tao Feng**, Akimasa Fujiwara***, Junyi Zhang****, Zhongzhen Yang*****

1. Introduction

Transportation environment pollution, as one of the negative effects caused by road traffic, has become a major concern of policy makers and researchers because of its more and more serious influences on peoples living condition in a world wide. To control and alleviate the traffic environmental load, the related policies and measures of transportation demand control and environment protection have been put forth in different countries. The level of pollutant concentration is an important indicator to capture the effects of these policies/measures. Therefore, the prediction of the roadside pollutant concentration seems quite important for both environment management and transportation planning.

Traditionally the levels of emitted pollutants are first predicted based on emission models subject to traffic conditions, and then the predicted concentration is put into dispersion model to estimate the pollutant concentration. Since two stages are necessary in this approach, then errors in each stage might inevitably lead to deviation between the estimated and actual values, which may deteriorate the precision of concentration prediction. In fact, it seems not an easy task to precisely predict it from the perspectives of both the determination of the amount of pollutants and the selection of influential factors related to the dispersion. Many affecting factors such as land use patterns, traffic conditions and meteorological conditions will lead to a obvious change of dispersion in some degree. Therefore, in spite of the existence of many concentration models, most of them could not satisfactorily fit the actual situations because of the complex uncertain relationship between pollutant concentration and the influencing factors. The diverse constraints make models be applied properly only under certain conditions.

Recently, artificial neural network approach has been shown in the literature as one of the best alternatives of modeling the pollutant concentration, especially for short-term forecasting, possibly because they can approximate almost any function, regardless of its degree of nonlinearity and without prior knowledge of its functional form. However, among numerous neural network based models, few of them integrate the whole information which affect the pollutant concentration together, and especially, none of them consider the concentration level of CO_2.

Therefore, the current paper attempts to simultaneously incorporate multiple factors affecting pollutant emission and dispersion and to build the new prediction model of pollutant concentration. The model integrates main factors such as meteorological conditions, traffic conditions and land use patterns to simulate the emission and dispersion mechanisms of vehicular pollutant, CO_2. To reflect the complicated nonlinear relationship between the affecting factors and the concentration, an artificial neural network (ANN) based prediction model of pollutant concentration by road traffic is developed, and the model performance is verified based on a case study in which the dataset are detected from the main road segments of Dalian City, China.

Remainder of this paper is arranged as follows: A brief review on existing studies is made in section 2 first, followed by description of the objectives and contributions of this paper. Section 3 introduces the data used in the model and corresponding experiment scheme. Section 4 empirically examines the effectiveness of the neural network prediction model. Finally, the whole study is summarized and some future studies are also mentioned in section 5.

2. Study Purpose

Up to now, many models have been developed and applied for analyzing traffic environmental load, in which the most popular one is Gauss model that takes highway as study object, regarding the initial dispersion parameters of traffic, as well as considering indirectly the onflow effect caused by the traffic (Wang, 2005). Subsequently, a series of line source models based on Gauss model, such as HIWAY, CALINE, GM, CAL3QHC, etc., have been developed to improve the limitation of the Gauss model, but still have many disadvantages in representing the actual influences on pollutant concentration based on traffic flow characteristics. Nagendra (2002) presented a review relating to line source models in which the research progress and history can be referred to in details.

Due to the complex relationship between concentration and the factors, and the difficulty to flexibly reflect such non-linear mechanism, models based on neural network technique attract more and more scholars to further the estimation of pollutant concentration. The ANN-based model, compared with the traditional models, can combine the calculation of the emissions and the dispersion together with great flexibility. The studies on pollutant concentration, such as CO, NO_X, NO_2, could be traced from most of the previous literatures (Drozdowicz et al, 1997; Moseholm et al, 1996; Gardner et al, 1999; Nagendra et al, 2004.).

Comparing to the existing literatures, the main contribution of the present paper is as follows:

- It builds the neural network model for roadside pollutant concentration, CO_2, which were not found to be studied

*Keywords: Artificial Neural Network, Pollutant Concentration, Traffic flow, and Meteorological Conditions.

**Student Member of JSCE, D. Sc., ***Member of JSCE, Dr. Eng., Prof. ****Member of JSCE, Dr. Eng., Associate Prof., Graduate School for International Development and Cooperation, Hiroshima Uni. (1-5-1 Kagamiyama, Hiroshima, Japan. TEL: 082-424-6921. FAX: 082-424-6921)
*****Member of JSCE, Dr. Eng., Prof. College of Transportation and Logistics, Dalian Maritime University, China.

based on the database of internet literatures.

- It simultaneously combines variety of factors such as meteorological condition, traffic condition and road space condition into together to reflect the interactions actually.
- It uses the dataset with 30 seconds' interval which make it much suitable for shot-term or real-time forecasts in estimating the pollutant concentration.

Moreover, from the methodological view of artificial neural network, the ANN based model avoids the potential errors happened in the two-step forecast process, and makes it much easier to use in practical applications.

3. Research Framework

As the strong capability in approximating nonlinear relationship, BP (Back-Propagation) algorithm which has been applied widely is adopted in this paper. The single output of the model is the concentration level of CO_2. The inputs composed of three categories: meteorological condition, traffic condition and road spatial characteristics which can be concretely expressed as temperature, humidity, wind speed, wind direction, traffic volume, travel speed, flow composition, source strength and height level.

To get sufficient data required in network training, a real experiment designed for obtaining data is carried out. The relevant data were collected from five main road segments selected in a case study area of Dalian, China. Characteristics of those links are shown in Table 1.

Table 1 Characteristics of surveyed road segments

Site Names	HBR	HHR	JFR	ZNR	ZSR
Lane numbers	6	6	6	6	6
Road width (m)	21	21	18	21	21
Buildings on roadside (Average Height, m)	18	24	24	10	12
Distance from buildings to roadside (m)	7	6	11	>30	16
Distance from observation station to roadside (m)	0.5	0.5	0.5	0.5	0.5

Here, HBR (HuaBei Road), HHR (HuangHe Road), JFR (JieFang Road), ZNR (ZhongNan Road) and ZSR (ZhongShan Road) are abbreviations of five different sites respectively. All sites selected here are links with versatile characteristics. To obtain the needed data simultaneously, various monitors, except the video camera which located on top of the buildings, are set on certain level (0.8m or 1.8m) within 0.5m distance to roadside. The experiment could be depicted in details as Fig.1.

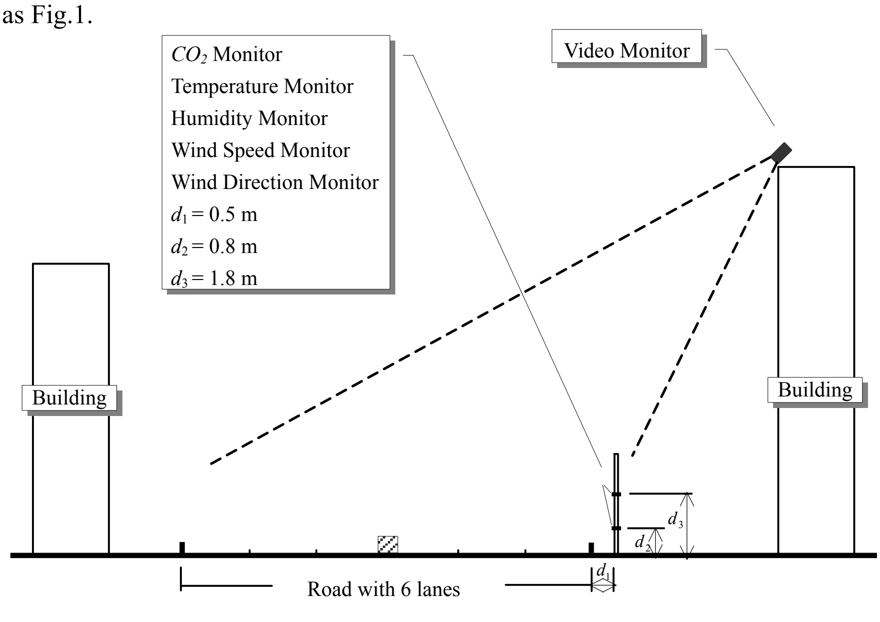

Fig.1 Experimental scheme design for data observation (cutaway view)

3.1 Data Analysis

Totally 1,200 sets of data are obtained from five survey stations, of which 1,000 sets are valid. Data are collected with an interval of thirty seconds. Table 2 shows some result of the elementary data analysis.

It is shown that CO_2 concentration changes obviously with the height level. The average CO_2 concentration values at 0.8m in all stations are higher than the values at 1.8m. The same phenomenon is seen in terms of the temperature, although the temperature changes a little. It is notable that although the data reflect some rules, such relationship is not satisfied enough as expected. One of the possible reasons maybe the high nonlinearity existed between various factors affecting the pollutant concentration.

Table 2 Result of Data Analysis

Site Names	Units	HBR	HHR	JFR	ZNR	ZSR
Traffic Speed (Average)	km/h	28	14	25	26	27
Traffic Volume (Average)	pcu	22	28	23	29	32
Average Temperature (d_3)	centigrade	23.5	21.4	16.4	16.8	19.5
Average Temperature (d_2)	centigrade	23.6	21.5	16.5	17.2	21.1
Average Flow Composition	—	0.23	0.13	0.07	0.46	0.14
CO_2 Average Concentration (d_3)	ppm	395	392	348	358	367
CO_2 Maximum Concentration (d_3)	ppm	418	430	430	380	397
CO_2 Minimum Concentration (d_3)	ppm	375	367	327	326	343
CO_2 Average Concentration (d_2)	ppm	501	496	478	488	479
CO_2 Maximum Concentration (d_2)	ppm	785	574	536	539	522
CO_2 Minimum Concentration (d_2)	ppm	442	457	441	466	440

3.2 Model Specification

The data are divided into three parts, in which 600 sets are used for training, 100 sets for cross-validation and 300 sets for testing. Due to the big differences in scale among the data, all data is re-scaled and randomized within a standard range [0,1] before training. Here, the data of wind direction means the angle between real wind direction and vertical direction to lanes, recorded as α, and $\sin(\alpha)$ is used for training and simulation. Flow composition is denoted by the ratio of big vehicles to small vehicles. Big vehicles include truck and big/medium class cars, which can accommodate more than 12 persons; other types belong to small vehicles.

After various scenario trainings, it is found that the 3 layers neural network model with 3 neurons in a single hidden layer could yield the best forecast result and the quickest convergence. Here, the activation function in the hidden layer adopts Sigmoid function, and linear function in output layer. Fig.2 depicts the structure of network model.

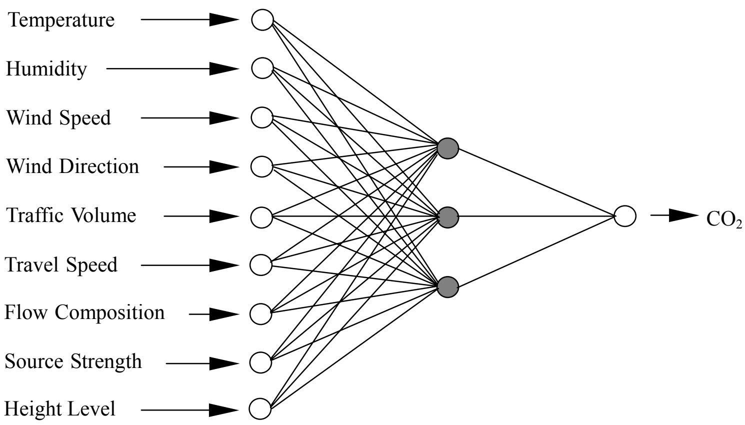

Fig.2 Model Structure of Concentration Prediction Model

4. Result

After iterated training process, the connector weights of the neural network model have been confirmed and are used to examine the general capability of the developed model. The main indicators used to evaluating model performance are average output $\overline{P}$ (actual value), average output $\overline{T}$ (estimated value), mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and good-of-fitness (r^2). The performance results are shown in Table 3.

Table 3 Performance result of the model

Model	$\overline{P}$	$\overline{T}$	Minimum	Maximum	ME (ppm)	MAE (ppm)	RMSE (ppm)	r^2
CO_2	432	431	355	501	138	14.8	19.34	0.90

From the main indicator results revealed in Table 3, it can be known that the ANN-based model could forecast the pollutant concentration with high goodness-of-fit. Here, the ME values of CO_2 is larger than 0, which means that the forecasted results are higher than the observed ones. In addition, P (432) and T (431) of CO_2 are very close, which means that the forecasted accuracy of CO_2 model is much high. At this moment, r^2 (0.90) means that the fit of the model to the data is reasonably good. The comparisons between the observed values and estimated concentration of CO_2 are shown in Fig.3. Based on the comparison results of CO_2 and the performance evaluation, the estimated result by this model is almost identical to the real values, which satisfies the practical prediction requirement.

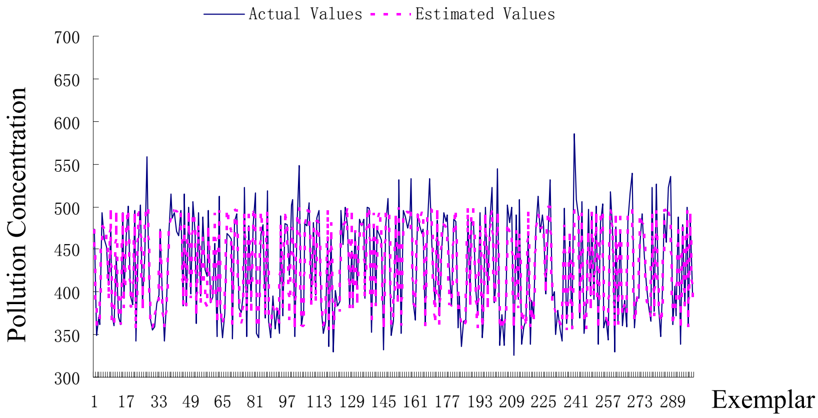

Fig.3 Result comparisons based on validation dataset

5. Summary

Pollutant concentration in transportation is still a complicated problem which is difficult to formulate mathematically. Traditional two-stage method could not ensure the forecasting accuracy and overcome the difficulties in obtaining multiple data required in the process of model validations. Neural network approach has already been verified to be efficient in simulating high non-linear problems.

The neural network approach developed in this paper is a one-step model of pollutant concentration forecasting which avoid some inherent disadvantages existed in traditional models. Moreover, it simultaneously incorporates the meteorological conditions, traffic condition and road spatial characteristics into one model which could be expectedly used comprehensively. The model's performance have been evaluated based on a set of actual observation data collected at five major road segments in the city of Dalian, China. The results revealed that the optimized neural network model can be used to simulate such complex non-linear relationship between the pollutant concentration and the corresponding affecting factors with high goodness-of-fit. Thus, it is suggested to be applied in the actual estimations of pollutant concentration and the policy evaluation relating to transportation environment.

References

1) Dirks, K.N., Murray, D.J., Hay, J.E., Sturman, A.P., 2003. A semi-empirical model for predicting the effect of changes in traffic flow patterns on carbon monoxide concentrations. *Atmosphere Environment* 37, 2719-2724.
2) Drozdowicz, B., Benz, S.J., Santa Cruz, A.S.M., Scenna, N.J., 1997. A neural network based model for the analysis of carbon monoxide contamination in the urban area of Rosario. *Fifth International Conference on Air Pollution Modelling, Monitoring and Management, Air Pollution V. Computational Mechanics Publications*, Southampton, Boston, pp. 677-684.
3) D. Stead., 1999. Relationships between transport emissions and travel patterns in Britain. *Transport Policy* 6, pp247-248.
4) Gardner, M.W., Dorling, S.R., 1999. Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London. *Atmosphere Environment* 33, pp709-719.
5) Moseholm L., Silva J., Larson T., 1996. Forecasting carbon monoxide concentration near a sheltered intersection using video traffic surveillance and neural network. *Transportation Research D*, Vol.1, No.1, pp15-28.
6) Nagendra S.M.S., Khare M., 2002. Line source emission modeling, *Atmospheric Environment* 36, pp2083-2098.
7) Sharad Gokhale, Mukesh Khare, 2004. A review of deterministic, stochastic and hybrid vehicular exhaust emission models, *International Journal of Transport Management* 2, pp59-74.
8) Shiva Nagendra, S.M., Mukesh Khare, 2004. Artificial neural network based line source models for vehicular exhaust emission predictions of an urban roadway. *Transportation Research Part D* 9, pp199-208.
9) Sudarmanto Budi Nugroho, Akimasa Fujiwara and Junyi Zhang, 2006. A comparative analysis of surface ozone patterns at different land use types in Jakarta city based on a structural equation model with latent variables, *the 85th Annual Meeting of the Transportation Research Board*, January 22-26, Washington, D.C.
10) WANG W., CHEN X.W., LU J., 2003. Study on sustainable development theory of urban transportation system. Science Press, Beijing.
11) WANG W., YU L., FEI W.W., YANG F., 2005. Review of Emission Dispersion Models Based on Gaussian Line-Source Model. Environmental Protection in Transportation, Vol.25, No.5, pp7-10.
12) YANG Z.Z., FENG T., 2004. A Neural Network based Car Ownership Model. Proceedings of the 8th International Conference of AATT, ASCE, pp631-635.