ASSESSMENT AND PREDICTION OF BENZENE CONCENTRATIONS IN A STREET CANYON USING A VARIETY OF MODELS, IN THE MEANS OF ENVIRONMENTAL MANAGEMENT PURPOSES.

P.A. KASSOMENOS[1,*]
S.P. KARAKITSIOS[2]
G.A. PILIDIS[2]

[1] *University of Ioannina, Department of Physics*
Laboratory of Meteorology
GR-45110, Ioannina, Greece
[2] *University of Ioannina*
Department of Biological applications and technologies
GR-45110, Ioannina, Greece

*to whom all correspondence should be addressed:
e-mail: pkassom@cc.uoi.gr

ABSTRACT

Urban air quality nowadays is one of the major environmental issues, due to its impact to various health problems, caused by the daily exposure of the population in dangerous air pollutants. The highest levels of air pollutants are usually observed in street canyons, emitted by urban traffic. To understand the way that pollutants dispersed in a street canyon environment, various modeling techniques are used. Both the accuracy of the predictions and the quickness of the calculations, are significant factors to adopt a modeling technique.

In the present study, we used benzene as an indicator of traffic pollution. Two different modeling techniques, an artificial neural network (ANN) and a semi empirical deterministic model (DET) are used, to predict benzene concentrations in a street canyon environment.

The ANN was based on a training procedure using measurements collected in a specific street canyon (benzene concentrations, traffic density, vehicle's type distribution), while the DET model was based on road traffic emission rate, wind speed and direction, and the geometrical characteristics of the road.

After the validation of models, their response to preselected "what if" scenarios was attempted.

Although both models produced very good results, given the limited amount of data available, the ANN succeeded slightly better than DET in predicting benzene concentrations. On the contrary, ANN presents lower response in predicting the effect of significant changes in traffic flow patterns on benzene concentrations.

The results from the simulations indicate that the ANN is a promising technique for benzene modeling in an urban environment, since DET seems to be ideal for environmental management purposes.

KEYWORDS: Benzene, modeling, artificial neural networks, street canyon.

1. INTRODUCTION

Although there is quite a lot of work done towards the direction of studying the procedures that describe the distribution of VOCs in a street canyon (Xie *et al.,* 2003 among others), only recently some modeling studies of benzene in urban street canyons have appeared. Up to now all papers referring to benzene modeling (Carr *et al.*, 2002, Vardoulakis *et al.*, 2003) use traffic flow as the determinative parameter for calculating road emissions.

Artificial neural networks have been used to predict atmospheric pollutants' concentrations during the last decade due to their ability (Nunnari *et al.* 1998) to handle non linear systems. The non reactive pollutants are non linear systems (in a much smaller degree than photoreacting pollutants) since they are highly depend on atmospheric dispersion conditions and vehicles emissions.

Recently, non reactive pollutants forecasting via ANNs (Dorling *et al.* 2003), has gained an increased interest, because ANNs can handle the large amounts of data, routinely acquired through modern monitoring techniques. If these data (e.g. air quality, meteorological and traffic) are used for training an ANN, this can lead to reliable prediction of atmospheric concentrations of emitted pollutants in a way easier and faster than a deterministic model (Nagendra *et al.*, 2004).

In our case DET, as though constructed according to the basic modeling principles, uses traffic density as a basic parameter. This parameter can describe the size of traffic volume in the road in a more realistic way. Compared to traffic flow, density allows the estimation of Road Emission Rate (RER) for several traffic conditions, ranging from zero values of traffic density (free flow), up to a maximum theoretical value (traffic jam conditions). Traffic density also represents a better measure of the emissions, because at each value of traffic flow correspond two values of traffic density, with consequently significantly different road emission rates (Kassomenos *et al.,* 2004). Moreover complicated emission modes, requiring complex calculations like the start-stop conditions, can be adequately approached.

In this work, both an ANN and a DET (Kassomenos *et al.,* 2004) model are developed and evaluated in comparison with measurements which have been recorded in a street canyon. After that, the two models were applied to three preselected "what if" scenarios to examine their efficiency to act as tools for management of atmospheric environment in a city.

2. DATA

Data (traffic flow, type of vehicle, air quality and meteorological) were collected during an experimental campaign. All measurements were made near the side of a busy road located in the city of Ioannina, NW Greece (Figure 1). This road is a typical street canyon having a width of 10m and surrounded by buildings of about 15 m height. Four measuring campaigns, were carried out lasting one week each: Friday 18/4/03 to Thursday 24/4/03 for the spring season, Saturday 26/7/03 to Friday 1/8/03 for summer, Thursday 13/11/03 to Wednesday 19/11/03 for autumn and Thursday 12/2/04 to Wednesday 18/2/04 for winter. The measurements were recorded for 12.5 hours on a daily basis (09:00- 21:30 Local Time).
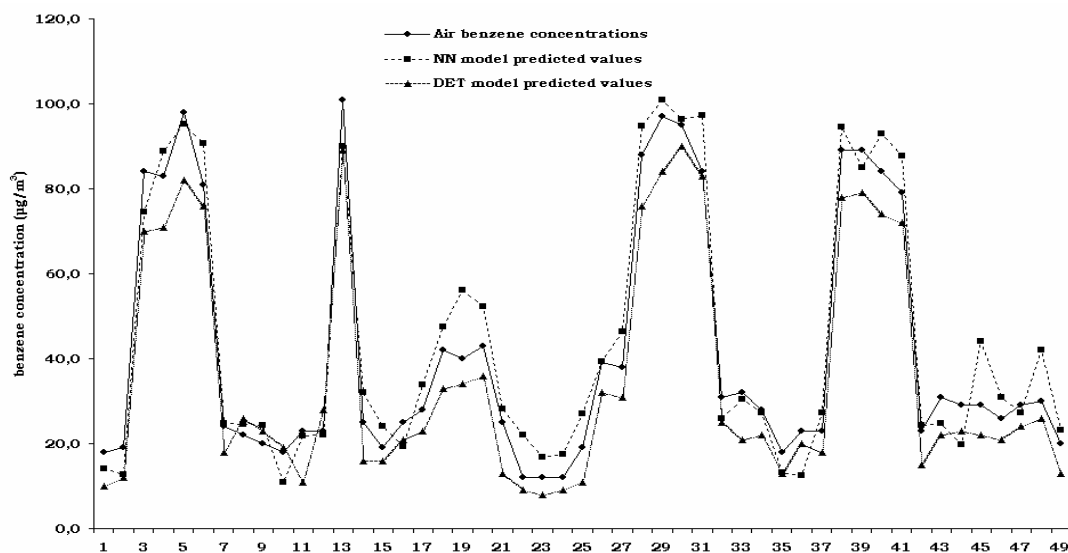

Figure 1. Observed vs ANN and DET models predicted benzene concentrations

The following parameters were measured:
- Traffic flow from two independent observers. Vehicles were registered in seven main categories (catalytic passenger vehicles, non catalytic passenger vehicles, diesel passenger vehicles, light duty vehicles, heavy duty vehicles, buses and motorcycles) every half an hour. It was found that the maximum traffic density value (D_j) in the specific road, is nearly 200 vehicles/km, while the maximum theoretical speed is 50 km h^{-1}, equal to the speed limit for the road. In order to estimate the traffic density the vehicles' speed

was calculated as the ratio of the distance (a part of the street canyon) to the time needed for the vehicle to cross this part of the road.

- Wind speed (m s^{-1}) and direction (degrees) on an hourly basis on the roof of the same building in front of which the rest of the observations were made. The threshold of the anemometer was 0.3 m s^{-1}.

- 2-hourly air benzene concentrations, at a height of 1.8 m and at a horizontal distance of 3 m from the center of the nearest traffic lane. The sampling procedure (active sampling) included adsorption of benzene on TENAX TA (60/80 mesh) tubes using a personal pump (SKC, model 222). The sampling flow rate was set to 100 ml min^{-1} and the sampling time was set to 30 min. Approximately 3 lt of air were pumped through the tube. Totally 196 samples were collected.

3. METHODOLOGY
3.1 ANN model
In this work the data available (196 records) were randomly divided into two groups. The first group (75% of data) was used to train the model. The model developed by the first procedure was used then to predict the benzene concentrations for the second group (25% of the data). The input parameters were: flow of each vehicle category, vehicles' speed and wind speed. The output parameter was benzene concentration.

ANN architecture consisted of three layers. The first layer consists from seven neurons, one for each input, the second layer consists from eight neurons (experimentally found to yield best results) and the third layer includes only one neuron, as much as the output parameters (benzene concentration). To reduce the number of input neurons for improving the modeling results, diesel passenger vehicles and light duty vehicles were integrated in to the same category, named light diesel vehicles, because of their almost similar benzene emission factors. For the same reasons, heavy duty vehicles and buses were integrated into one category, named heavy diesel vehicles.

Prior to conducting the network training operation for ANN model using the backpropagation paradigm (Rumelhart and McClelland, 1986), a training set consisting of 147 cases had to be obtained from the environmental data. They were then divided into two distinct sets, training set, and a testing set, using a random process. The training set consisted of 90% of the data used to train the model or 134 cases. The remaining 10%, or 15 cases, were excluded as a test set.

The training process of this ANN model was performed using the NeuroShell simulator developed by Ward Systems Group. The simulator iteratively adjusts the weights until the error between the output data and the observed data is minimised. Several network factors such as momentum, learning rate, number of hidden nodes, and thresholds were tested during the training process to improve network generalisation and prediction accuracy. The best results (i.e. the error is minimised) were obtained when the network completed 44,281 epochs in 0:11:29 hours of continuous training.

3.2 DET model
The DET model is extensively described in Kassomenos *et al.* (2004). It consists of three modules: the traffic, the emissions and the dispersion module. The main objective of this arrangement (i.e. the three traffic modules) was the development of a model in which benzene RER depends on traffic density. Traffic density depends on traffic flow, vehicles speed and the theoretical maximum density of the road travelled. To calculate RER, traffic flow was divided into the seven main aforementioned vehicle categories based on their type and fuel used. In the last stage of the methodology, the STREET dispersion model was used to predict benzene concentrations (Johnson *et al.* 1973). In this model, pollutant's concentration is a function of RER, urban background concentration, wind speed and direction and street's canyon height and width.

4. RESULTS AND DISCUSSION
The validation of each season's predictions by the DET model are presented in Kassomenos *et al.* (2005). In this study a comparison between the measurements of benzene and the

predictions between the two models (ANN and DET) was done. The comparison refers to the aforementioned 49 randomly selected values which were chosen to validate the ANN model results.

The benzene values predicted by DET model were very close to the measured ones (Figure 1).

It is obvious from the analysis, presented in Figure 1, that DET model slightly underestimates the measurements. This is possibly due to the underestimation of the emission factors compared to the actual ones. There are many reasons that the emission factors are higher than the theoretical factors: degradation of the catalyst, insufficient maintenance of the engine, bad fuel quality, use of unleaded petrol in motorcycles etc. DET model was able to follow the variations of the measured benzene concentrations when fed by real traffic data. Such variations could be caused by changes in the composition of the fleet even, when only a very small change in traffic density was observed. This is the reason of the high R^2 values computed (Table 1).

In addition, DET model can describe conditions in which traffic flow and fleet composition seems to have insignificant variation but they lead to a completely different RER, due to the fact that for the same traffic flow, two completely different density conditions exist

Table 1. DET vs ANN evaluation parameters

	ANN model	DET model
Mean Absolute Error	6	7
Minimum Absolute Error	0	1
Maximum Absolute Error	16	18
RMSE	7. 084	7.976
RRMSE	1.2715	1.2414
R^2	0.9497	0.9792

On the other hand, benzene values predicted by the ANN model are very close to the measured ones (Figure 1) and its evaluation parameters are similar to those of DET model (Table 1). Unlike to DET model, ANN model does not present any specific trend and does not overestimate or underestimate the observed values.

Neural network modelling can also assess the importance of each of the input variables by using the network weights. The method for partitioning the connection weights proposed by Garson (1991) was used. The technique involves partitioning the hidden– output connection weights of each neuron into components associated with each input neuron (Goh, 1995). The results of the calculations are shown in Figure 2, displayed as columns representing the relative importance of the various input variables.

From this figure, it is evident that wind speed is the most significant factor affecting air quality in an area with a relative weight of almost 21%, followed by the speed of the vehicles (~18%) and the not catalytic cars (~17%). Since, air quality models usually consist of an emission and a meteorological part, the inputs were grouped in two main categories: meteorological-input parameters (represented in our case by wind speed) and emission-input parameters (e.g. vehicle speed, vehicle type etc.)

Among the emission parameters vehicle speed, plays the most significant role. As it was shown above vehicle speed is the key factor in the calculation of traffic density and consequently RER.

The remaining emission parameters refer to the total traffic flow. The relative contribution of each is a combination of the emission factor and the percentage of the total traffic flow, a specific vehicle category represents.

Non catalytic passenger cars have the highest relative importance among the traffic flow components. They account for almost 25% of the traffic flow contribution while represent only 13% of the total traffic flow. Catalytic passenger cars, on the other hand account for a significant lower part of the traffic related contribution (9%), although they represent almost 60% of the total traffic flow.
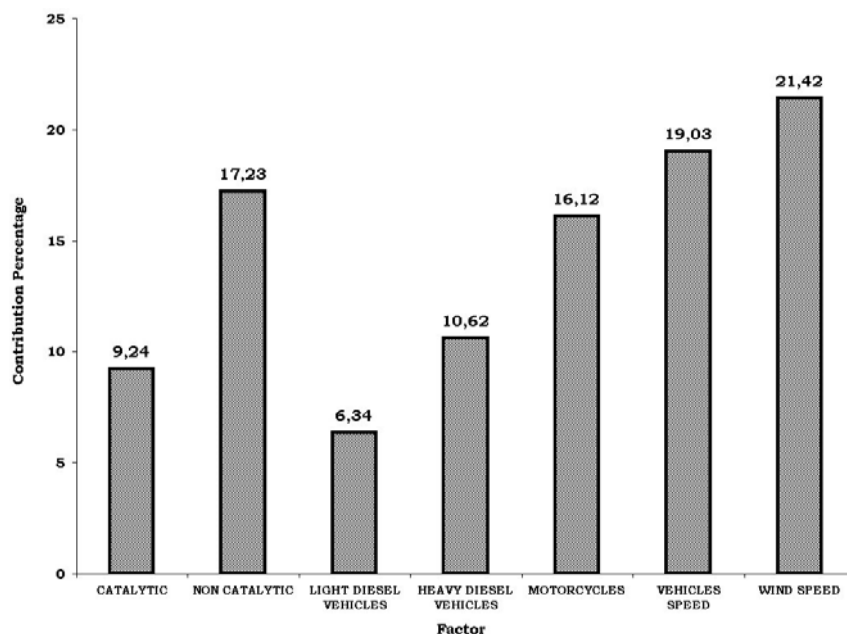
Figure 2. Relative importance of the various input variables in the overall prediction of benzene concentration by the ANN model

Heavy diesel vehicles despite their small number (2% of the total fleet) have a contribution factor slightly higher than catalytic passenger vehicles. This is due to the high emission factors, lower speed and their size.

Motorcycles, present almost the same contribution with non catalytic passenger cars (16%). This is due to two reasons: they have higher benzene emission factor than other vehicle categories and their fraction (13%) of total traffic composition. The contribution of the motorcycles to benzene concentrations was also regarded by the seasonality observed. Finally, light duty vehicles have the lowest contribution factor (6%), a fact that was expected taking in mind the percentage of traffic flow they represent (14%) and the lowest benzene emission factor among all vehicle categories they have.

5. MODELS APPLICATION TO ''WHAT IF'' SCENARIOS

After that, three "what if" scenarios were tried. For all the scenarios the same measured data of a typical day of measurements campaign (Saturday 12/2/2004, 11:00-11:30) were taken into consideration for comparison reasons. This data were chosen because of calm conditions (wind speed lower than 0.3 m s^{-1}), so that the effect of traffic parameters should be maximized. In all scenarios, traffic density was kept constant, and when a vehicles category was excluded from the pattern or sustained any change, the appropriate equalizations where made, by increasing or decreasing the number of the other vehicles categories (Table 2).

Table 2. Observed and "what if" scenarios traffic data.

	Observed data	A' scenario	B' scenario	C' scenario
CATALYTIC (veh/30')	322	406	0	161
NON CATALYTIC (veh/30')	84	0	406	42
DIESEL PAS. VEHICLES (veh/30')	39	39	39	242
LIGHT DUTY VEHICLES (veh/30')	30	30	30	33
HEAVY DUTY VEHICLES (veh/30')	2	2	2	2
BUSES (veh/30')	4	4	4	4
MOTORCYCLES (veh/30')	64	64	64	64

In the first scenario, a suggestion is made in which all non catalytic passenger vehicles are replaced by catalytic. This scenario has great interest because gives a conception on future benzene concentration if we accept that traffic volume do not retain significant change.

The second scenario, although does not have an applicative interest, gives the prevailing magnitude of benzene atmospheric pollution if vehicles fleet had the nowadays volume, but without the existence of catalytic technology.

In the third scenario, a more complicated suggestion is made, in which half of the number of each passenger vehicles category (catalytic and non catalytic) is replaced by diesel passenger vehicles. This scenario is meaningful nowadays in Greece, because till now, diesel passenger vehicles were allowed only to taxi drivers but not to civilians and this restriction probably will not exist by the coming of year 2006.

The results of both models predictions are shown in Figure 3. It is obvious that DET model appears to be more flexible in rendering far-out scenarios, in which the number of a vehicle category is set equal to zero or contemporaneously two inputs are out of training limits.
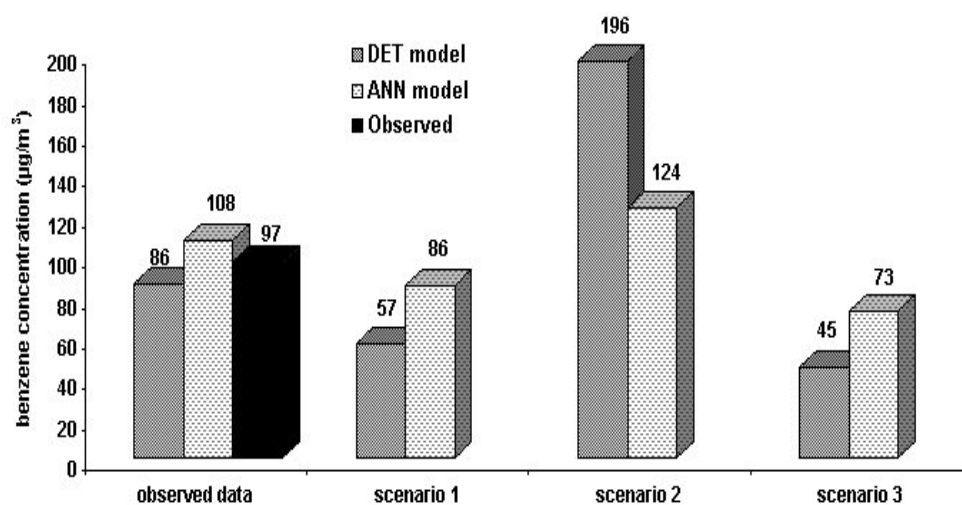

Figure 3. DET and ANN predictions in three different "what if" scenarios

This is due to the inborn debility of ANN models to successfully resolve problems in which, input parameters are not in confines of training set values (Flood and Kartam, 1994).

From all the above it is obvious that DET model is a valuable tool for to the assessment of "what if" scenarios concerning possible modifications as much in the fleet composition, as in the traffic characteristics of each road, which in turn are subject to the traffic characteristics of the surrounding roads connected to it. Consequently, the forecast of the changes in the benzene atmospheric concentration measured on a road, which will be the result of changes in the traffic parameters in the area, is possible. Combined with a more extensive traffic study, this model is able to give directions for the improvement of the air quality and constitute a major decision support tool of a master plan concerning a wider urban area.

6. CONCLUSIONS

In the present study two different modeling approaches are used to predict benzene concentrations in an urban street canyon. Both models used as basic input parameters the traffic characteristics of a road (fleet composition, traffic volume, traveling speed) as well as, the wind speed and direction.

The models used, a DET and an ANN, performed in adequate manner. Specifically DET model predicted the measured benzene concentrations for all the seasons of the year with very good results (R^2 about 0.99) but it always underestimates the measurements. ANN succeeded a slowly lower R^2 (about 0.97) but did not present systematic underestimation of the measured values. ANN model provides an easy and reliable methodology to predict benzene concentrations in a street canyon, while DET model is a remarkable tool for environmental management purposes and could be a useful tool for the authorities.

REFERENCES

Carr D., Ehrenstein O., Weiland S., Wagner C., Wellie O., Thomas A., and Mutius E. (2002) Modelling Annual Benzene, Toluene, NO_2, and Soot Concentrations on the Basis of Road Traffic Characteristics, *Environmental Research, Section A*, **90**, 111-118.

Dorling S.R., Foxall R.J., Mandi D.P., Cawley G.C. (2003) Maximum likelihood cost functions for neural network models of air quality data, *Atmospheric Environment*, **37**, 3435–3443.

Flood I., Kartam N. (1994) Neural networks in civil engineering. I:Principles and understanding, *Journal of Computing in Civil Engineering*, **8**, 131–148.

Johnson W.B., Ludwig F.L., Dabberdt W.F., Allen R.J. (1973) An urban diffusion simulation model for carbon monoxide, *Journal of the Air Pollution Control Association,* **23**, 490–498.

Garson G.D. (1991) Interpreting neural-network connection weights, *AI Expert,* **6**, 47–51.

Goh A.T.C. (1995) Back-propagation neural networks for modelling complex systems. *Artificial Intelligence in Engineering,* **9**, 143–151.

Kassomenos P.A., Karakitsios S.P., Pilidis G.A. (2004) A simple semi-empirical approach to modeling benzene concentrations in a street canyon, *Atmospheric Environment,* **38**, 6073-6078.

Kassomenos P.A., Karakitsios S.P., Pilidis G.A. (2005) A methodology to estimate benzene concentrations in a town through a traffic model, *Science of the Total Environment,* **347**(1-3), 272-281.

Nagendra S.M.S., Mukesh Khare M. (2004) Artificial neural network based line source models for vehicular exhaust emission predictions of an urban roadway, *Transportation Research, Part D,* **9**, 199–208.

Nunnari, G., Nucifora, A.F.M., Randieri, C. (1998) The application of neural techniques to the modelling of time-series of atmospheric pollution data, *Ecological Modelling,* **111**, 187–205.

Rumelhart D.E, McClelland J.L. (1986) Parallel Distributed Processing, *Explorations in the Microstructure of Cognition 1-2, MIT Press, Cambridge, MA.*

Xie S., Zhang Y., Qi L., Tang X. (2003) Spatial distribution of traffic-related pollutant concentrations in street canyons, *Atmospheric Environment,* **37**, 3213–3224.

Vardoulakis S., Fisher E.A.B., Pericleous K., Flesca G.N. (2003) Modeling air quality in street canyons: a review, *Atmospheric Environment,* **37**, 155–182.